Abstract. Growing evidence indicates that microRNAs (miRNAs) play critical roles in the initiation and progression of breast carcinoma (BC) and are promising diagnostic biomarkers. In the present study, we aimed to identify a multi-marker miRNA pool with high diagnostic performance for BC. We collected miRNA expression profiles of BC samples and normal breast tissues from The Cancer Genome Atlas (TCGA) and screened differentially expressed miRNAs by conducting two-sample t-tests and by calculating log2 fold-change (log2FC) ratios. Statistical significance was established at p<0.001 and |log2FC| >1. Then, we generated receiver operating characteristic (ROC) curves, calculated the area under the curve (AUC) with a 95% confidence interval (95% CI), and calculated the diagnostic sensitivity and specificity using MedCalc software. Additionally, we predicted the targets of candidate miRNAs using 10 online databases: TarBase, miRTarBase, TargetScan, TargetMiner, microRNA.org, RNA22, PicTar-vert, miRDB, PITA and PolymiRTS. Target genes that were predicted by at least four algorithms were chosen, and cooperative targets of multiple miRNAs were further selected for GO and KEGG pathway analyses through the DAVID online tool. Eventually, a total of 66 differentially expressed miRNAs were identified after miRNA expression profiles were analyzed in BC and normal breast samples. Of these, we selected nine dysregulated miRNAs as candidate diagnostic markers: seven upregulated miRNAs (hsa-miR-21, hsa-miR-96, hsa-miR-183, hsa-miR-182, hsa-miR-141, hsa-miR-200a and hsa-miR-429) and two downregulated miRNAs (hsa-miR-139 and hsa-miR-145). The ROC curve for the combination of these nine differently expressed miRNAs showed extremely high diagnostic accuracy, with an AUC of 0.995 (95% CI, 0.988-0.999) and diagnostic sensitivity and specificity of 98.7 and 98.9%, respectively. In conclusion, the combination of these nine miRNAs significantly improved the accuracy of breast cancer diagnosis.

Introduction

Breast cancer (BC) has become a modern global epidemic among women as its morbidity has markedly increased over the past few decades (1). According to the Cancer Statistics 2016, BC is the most frequently diagnosed malignancy and the second leading cause of tumor-related death in females in the US, with 246,660 new diagnoses and 40,450 deaths projected to have occurred in 2016 (2). Recently, accumulating evidence indicates that early diagnosis and treatment result in favorable prognoses; however, current clinical tumor biomarkers lack sensitivity and specificity for early detection of BC (3). Consequently, identification of novel diagnostic biomarkers with high sensitivity and specificity for differentiating breast cancer from benign breast disease and normal samples is urgently needed. Previous studies have reported that microRNAs (miRNAs) play vital roles in the initiation and progression of BC and that certain miRNAs can serve as potential diagnostic biomarkers (4,5).

miRNAs are short non-coding RNA molecules of 18-25 nucleotides that play essential roles in regulating gene expression by binding to target mRNAs and directing RNA-induced silencing complexes (RISCs) to degrade target mRNAs or repress mRNA translation (6,7). Notably, miRNAs are reportedly linked to the occurrence and development of a variety of malignant tumors. Numerous studies have indicated that miRNAs could serve as oncogenes or tumor suppressors by regulating cellular biological processes, such as cell differentiation, proliferation and apoptosis (4,8,9). Moreover, miRNAs are also involved in multiple cancer-related cell signaling pathways in various malignancies, including breast cancer (10,11). Previous studies have reported that many miRNAs are aberrantly expressed in BC and thus potentially act as biomarkers for cancer diagnosis. Fu et al (12) showed

Correspondence to: Professor Gang Chen or Professor Dian-Zhong Luo, Department of Pathology, First Affiliated Hospital of Guangxi Medical University, 6 Shuangyong Road, Nanning, Guangxi Zhuang Autonomous Region 530021, P.R. China
E-mail: chen_gang_triones@163.com
E-mail: 13878802796@163.com
*Contributed equally

Key words: breast carcinoma, miRNAs, TCGA, diagnostic marker, ROC curves
that miR-141, miR-145, miR-183, miR-21 and miR-638, which are differentially expressed in BC and normal tissues, function as potential diagnostic indicators to distinguish tumor cases from normal cases. However, most of the previous studies focused on the diagnostic value of a single miRNA. Additionally, the small sample size of current studies and the lack of sensitivity and specificity of candidate miRNAs limit the clinical application of these biomarkers. In recent years, the development of high-throughput technologies and bioinformatics has promoted the development of such diagnostic markers.

The Cancer Genome Atlas (TCGA) provides open access to many comprehensive miRNA-sequencing datasets. In the present study, we analyzed miRNA expression data published by TCGA to identify miRNAs differentially expressed between breast tumor and normal tissues. Ultimately, we determined a nine-miRNA signature for diagnosing breast cancer. Furthermore, we predicted the targets of these miRNAs through in silico algorithms and conducted gene ontology annotation and pathway enrichment analyses to determine the potential biological functions of the nine miRNAs in the signature.

Materials and methods

Data collection from TCGA. The level three miRNASeq datasets of 1,110 BC and 104 normal samples were collected from the TCGA (http://cancergenome.nih.gov/). The miRNA expression profiles for both breast tumor and normal tissues were analyzed using Illumina HiSeq systems (781 tumor specimens and 87 normal specimens) or Illumina Genome Analyzer systems (324 tumor samples and 17 normal samples). There were no ethical issues since the data were retrieved from TCGA. The present study, was conducted in accordance with the publication guidelines proposed by TCGA (http://cancergenome.nih.gov/publications/publicationguidelines).

Processing of miRNA expression data. After downloading miRNASeq datasets, we first excluded miRNAs with no expression. Then, the expression data of the remaining miRNAs were log2-transformed for downstream analyses. To screen differentially expressed miRNAs, we conducted two-sample t-tests to compare miRNA expression levels in BC and normal breast tissues. In addition, we calculated log2 fold-change (log2FC) ratios. To reduce the false-positive rates (FPR) of our diagnostic tests, we set p<0.001 and llog2FC| >1 (13) as standards for identifying aberrantly expressed miRNAs. Statistical analyses were conducted through SPSS 20 (IBM, Armonk, NY, USA), and graphs were produced by GraphPad Prism 5 (GraphPad Software, Inc., La Jolla, CA, USA).

Selection of candidate diagnostic biomarkers. We selected differentially expressed miRNAs that were analyzed using both Illumina HiSeq and Illumina Genome Analyzer systems. Next, we generated receiver operating characteristic (ROC) curves, calculated the area under the curve (AUC) with a 95% confidence interval (95% CI) and calculated the diagnostic sensitivity and specificity using MedCalc software (14). miRNAs with AUC >0.9 were identified as potential diagnostic biomarkers.

Target prediction and enrichment analysis. The potential targets of candidate miRNAs were predicted by 10 online databases: TarBase, miRTarBase, TargetScan, TargetMiner, microRNA.org, RNA22, PicTar-vert, miRDB, PITA and PolymiRTS. Only target genes predicted by at least four algorithms were chosen, and genes targeted by more than one miRNA were selected for further analysis. Gene Ontology (GO) functional annotation and KEGG pathway analyses were conducted through the DAVID online tool (https://david.ncifcrf.gov/) and visualized by Cytoscape 3.3.0 software and the R package ‘ggplot2’ (http://www.inside-r.org/packages/cran/ggplot2).

Results

Selection of differentially expressed miRNAs and candidate diagnostic biomarkers. A total of 66 differentially expressed miRNAs were identified (p<0.001, |log2FC| >1) after miRNA expression profiles were analyzed in BC and normal breast samples of these, 13 miRNAs were downregulated and 53 miRNAs were upregulated in BC specimens (data not shown). We performed ROC analysis of the 66 miRNAs and selected nine dysregulated miRNAs as candidate diagnostic markers since they showed high diagnostic accuracy (AUC >0.9). As shown in Table I, these nine dysregulated miRNAs included seven upregulated miRNAs (hsa-miR-21, hsa-miR-96, hsa-miR-183, hsa-miR-182, hsa-miR-141, hsa-miR-200a and hsa-miR-429) and two downregulated miRNAs (hsa-miR-139 and hsa-miR-145). Differential analysis of the expression data of the selected nine miRNAs indicated that these miRNAs exhibited significant statistical differences (p<0.0001, llog2FC| >1) (Table I, Fig. 1). Among these nine miRNAs, both hsa-miR-429 and hsa-miR-200a are located on chromosome 1p36.33, while hsa-miR-183, hsa-miR-96 and hsa-miR-182 are located on chromosome 7q32.2.

Diagnostic performance of selected miRNAs. Among the seven upregulated miRNAs, the AUC values with 95% CI of hsa-miR-21, hsa-miR-183, hsa-miR-96, hsa-miR-182, hsa-miR-141, hsa-miR-200a and hsa-miR-429 were 0.946 (0.926-0.960), 0.965 (0.951-0.976), 0.960 (0.944-0.972), 0.934 (0.916-0.950), 0.924 (0.905-0.941), 0.921 (0.902-0.939) and 0.920 (0.900-0.937), respectively. For the two downregulated miRNAs, the AUC values with 95% CI of hsa-miR-139 and hsa-miR-145 were 0.941 (0.923-0.956) and 0.910 (0.889-0.928), respectively (Table II, Fig. 2). Furthermore, we constructed a binary logistic regression model to evaluate the combination of these nine miRNAs. The diagnostic performance of the combination was improved significantly. The results showed that the AUC reached 0.995 (95% CI=0.988-0.999), and the diagnostic sensitivity and specificity reached 98.7 and 98.9%, respectively (Table II, Fig. 2).

Target prediction and enrichment analysis. We gathered target genes predicted by at least four algorithms, and then chose those that were targets of multiple miRNAs for GO and KEGG analyses. The results of target predictions for the nine miRNAs are presented as supporting information (data not shown). Eventually, we selected 1,339 target genes for GO annotation and KEGG pathway analyses. The most strongly
enriched GO processes and KEGG pathways are listed in Fig. 3. The top 10 GO functional annotations and KEGG pathways are shown in Fig. 4. GO and KEGG enrichment analyses revealed that the targets of candidate miRNAs were...
Table II. Diagnostic performance of the nine selected miRNAs.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>AUC</th>
<th>95% CI</th>
<th>P-value</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-21</td>
<td>0.946</td>
<td>0.926-0.960</td>
<td>0.0001</td>
<td>90.0</td>
<td>88.5</td>
</tr>
<tr>
<td>hsa-miR-96</td>
<td>0.960</td>
<td>0.944-0.972</td>
<td>0.0001</td>
<td>91.4</td>
<td>93.0</td>
</tr>
<tr>
<td>hsa-miR-183</td>
<td>0.965</td>
<td>0.951-0.976</td>
<td>0.0001</td>
<td>90.0</td>
<td>94.3</td>
</tr>
<tr>
<td>hsa-miR-182</td>
<td>0.934</td>
<td>0.916-0.950</td>
<td>0.0001</td>
<td>87.2</td>
<td>89.7</td>
</tr>
<tr>
<td>hsa-miR-141</td>
<td>0.924</td>
<td>0.905-0.941</td>
<td>0.0001</td>
<td>82.8</td>
<td>89.7</td>
</tr>
<tr>
<td>hsa-miR-200a</td>
<td>0.921</td>
<td>0.902-0.939</td>
<td>0.0001</td>
<td>83.9</td>
<td>88.5</td>
</tr>
<tr>
<td>hsa-miR-429</td>
<td>0.920</td>
<td>0.900-0.937</td>
<td>0.0001</td>
<td>84.5</td>
<td>86.0</td>
</tr>
<tr>
<td>hsa-miR-139</td>
<td>0.941</td>
<td>0.923-0.956</td>
<td>0.0001</td>
<td>81.2</td>
<td>95.4</td>
</tr>
<tr>
<td>hsa-miR-145</td>
<td>0.910</td>
<td>0.889-0.928</td>
<td>0.0001</td>
<td>84.1</td>
<td>86.2</td>
</tr>
<tr>
<td>Combined&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.995</td>
<td>0.988-0.999</td>
<td>0.0001</td>
<td>98.7</td>
<td>98.9</td>
</tr>
</tbody>
</table>

<sup>a</sup>The combination of the nine miRNAs. ROC, receiver operating characteristic; AUC, area under the curve; CI, confidence interval.

Figure 2. ROC curves for the nine-miRNA signature differentiating BC samples from normal breast tissues. The area under the curve (AUC) represents diagnostic performance. ROC curves for (A) hsa-miR-21, (B) hsa-miR-96, (C) hsa-miR-183, (D) hsa-miR-182, (E) hsa-miR-141, (F) hsa-miR-200a, (G) hsa-miR-429, (H) hsa-miR-139 and (I) hsa-miR-145. (J) ROC curves for the combination of these nine miRNAs.
involved in several critical cancer-related pathways, including pathways in cancer, colorectal cancer, the TGF-β, the MAPK and the Wnt signaling pathways.

**Discussion**

Increasing evidence indicates that dysregulation of miRNAs may be closely related to many malignances, including breast carcinoma (BC) (15-17). In the present study, we comprehensively analyzed high-throughput miRNA expression profiles of breast cancer specimens from TCGA and selected nine miRNAs with high diagnostic accuracy, including seven upregulated miRNAs (hsa-miR-21, hsa-miR-96, hsa-miR-183, hsa-miR-182, hsa-miR-200a and hsa-miR-429) and two downregulated miRNAs (hsa-miR-139 and hsa-miR-145).

Among the seven upregulated miRNAs, miR-21 showed the most obvious overexpression in BC specimens. As one of the first mammalian onco-miRNAs identified, miR-21 has been experimentally confirmed to participate in the initiation and progression of various carcinomas. In BC, miR-21 also plays a vital role in carcinogenesis. miR-21 is located on chromosome 17q, a region containing several oncogenes, including HER2, TOP2A and PPM1D (18-20). miR-21 has been reported to function as an onco-miRNA by directly regulating several tumor suppressors, such as tropomyosin 1 (TPM1), programmed cell death 4 (PDCD4), maspin and phosphatase and tensin homolog (PTEN) (21,22). Additionally, it has been reported that miR-21 is involved in certain cancer-related signaling pathways, such as the PTEN/PI3K/AKT, the ERK1/2/MAPK and the miR-21/smad7/ERK signaling pathways, thus, promoting tumor growth and metastasis in vivo and in vitro (21,23-26).

Previous studies have shown that the miR-183/miR-96/miR-182 cluster is overexpressed in various malignant tumors, including BC; thus, this cluster may be an onco-miRNA cluster and could potentially act as a multi-marker for diagnosing malignancies (27). The upregulated miR-183/miR-96/miR-182 cluster inhibits transcriptional factors, such as FOXO1 and FOXO3, and thus, promotes BC cell proliferation, invasion, and migration and reduces cell apoptosis in vivo and in vitro (28,29).
As members of the miR-200 family, miR-141, miR-200a and miR-429 exert multiple complex effects on tumor occurrence and development in different carcinomas. Among the three miRNAs, miR-141 and miR-200a are located on chromosome 1p36.33, while miR-141 is located on chromosome 12p13.31. Several studies suggest that miR-200a, miR-429 and miR-141 may be tumor suppressors since they are downregulated in cancer samples (30-32). However, other studies showed that these three miRNAs may function as oncogenes since they are upregulated in tumor tissues (33-36). In the present study, the expression levels of the three miRNAs were significantly elevated in BC specimens compared with normal breast tissues. These discrepancies may be partially explained by the various sources of the cases studied, different detection techniques and different platforms used. It is reported that members of the miR-200 family can reverse epithelial-mesenchymal transition (EMT) by increasing E-cadherin expression and decreasing ZEB1, ZEB2 and β-catenin expression, further inhibiting cancer cell invasion and migration (31,37-40). Notably, EMT mechanisms are distinct in various malignancies, and the expression of EMT markers (E-cadherin and ZEB) is inconsistent in diverse types of cancers and different histological types of the same cancer (41). This could partially explain the opposing actions of miR-200a, miR-429 and miR-141 on tumorigenesis.

Real-time PCR and miRNA microarray confirm that miR-139 is downregulated in breast cancer (42-44), which is consistent with our results. In vitro, forced expression of miR-139 in the BC cell line MDA-MD-231 was found to suppress cell migration and invasion by targeting genes involved in metastasis-related signaling pathways (42). Functional experiments conducted by Hua et al (45) revealed that increased expression of miR-139 reduced cell proliferation in luminal type BC cells by directly targeting topoisomerase IIα (TOP2α), an oncogene that can promote cell proliferation in vitro.

Both previous studies, and our present study, revealed that miR-145 is downregulated in BC samples (46-50). Acting as an oncogene, miR-145 promoted cell apoptosis (46), and suppressed cell growth, proliferation, migration and invasion (47,51) in vitro. More importantly, the decreased expression of miR-145 was found to be negatively correlated with several clinicopathological parameters, such as ER and HER2 status, tumor size and lymph node metastasis (52), which indicated its potential diagnostic or prognostic value in the clinic.

Current studies have focused on the diagnostic value of dysregulated miRNAs in BC (12,53). However, this is the first identification of a nine-miRNA signature as a potential multi-marker for diagnosing BC found by analyzing high-throughput miRNA expression profiles from TCGA data. In the present study, the ROC curve of the combined nine differentially expressed miRNAs showed an extremely high diagnostic accuracy with an AUC of 0.995 (95% CI=0.988-0.999), and diagnostic sensitivity and specificity of 98.7 and 98.9%, respectively. The diagnostic value of these nine miRNAs combined was higher than that of the individual miRNAs, indicating that the nine-miRNA signature could serve as a potential multi-marker for diagnosing BC.

Previous studies have demonstrated that miRNAs play vital roles in carcinogenesis through participation in cancer-related pathways. Therefore, we performed target prediction of the nine selected miRNAs and subsequently carried out GO and KEGG pathway enrichment analyses to explore the potential
effects of these nine miRNAs on carcinogenesis. GO and KEGG pathways revealed that the targets of these dysregulated miRNAs were involved in many critical cancer-related biological processes and pathways, such as positive regulation of macromolecule biosynthetic process, regulation of transcription, regulation of gene expression, the enzyme-linked receptor protein signaling pathway, pathways in cancer, the TGF-β signaling pathway, and the MAPK signaling pathway. The results of enrichment analyses suggested that these nine miRNAs could regulate important pathways correlated with breast cancer by targeting oncogenes or tumor-suppressor genes.

To the best of our knowledge, this is the first study identifying a nine-miRNA signature as a multi-biomarker for diagnosing BC based on TCGA datasets. Nevertheless, the present study has several limitations. Firstly, our data were retrieved from TCGA and had not been verified in experiments. Therefore, further multicenter clinical trials are needed in the future to confirm our findings. Secondly, the expression profiling of these miRNAs was only based on tissue. It may be of more clinical value to investigate noninvasive diagnostic markers.

In conclusion, in the present study, we identified a tumor-specific miRNA signature in BC containing seven upregulated miRNAs (hsa-miR-183, hsa-miR-96, hsa-miR-21, hsa-miR-182, hsa-miR-141, hsa-miR-200a and hsa-miR-429) and two downregulated miRNAs (hsa-miR-139 and hsa-miR-145) by analyzing high-throughput data from a TCGA dataset. We discovered that these nine aberrantly expressed miRNAs may play important roles in the initiation and progression of BC. More importantly, we assessed the combination of these nine miRNAs, and found that an integrated analysis of these nine miRNAs improved the accuracy of breast cancer diagnosis. However, multicenter studies with large sample sizes are needed to validate our findings before the clinical application of this nine-miRNA signature.

Acknowledgements

The present study was supported by grants from the Natural Science Foundation of Guangxi, China (no. 2015GXNSFAA139187) and the Guangxi Provincial Health Bureau Scientific Research Project (no. Z2014057).

References


