Identification of key microRNAs involved in tumorigenesis and prognostic microRNAs in breast cancer

Dongchen Lu†, Wei Han† and Kai Lu*

Department of Thyroid and Breast Surgery, Nanjing Hospital of Chinese Medicine affiliated to Nanjing University of Chinese Medicine, Nanjing 210000, China

† Dongchen Lu and Wei Han contributed equally to this work.

* Correspondence: E-mail: yanhe920363002@163.com; Tel: +8602586369051.

Abstract: Breast cancer is a commonly diagnosed cancer in women, and one of the leading causes of cancer-related death among female patients. However, the key microRNAs involved in its tumorigenesis and microRNAs of prognostic values have not been fully understood. In the present study, we aimed to perform a systematic analysis of microRNA expression profiles to identify some key microRNAs associated with tumor initiation and prognosis. Using TCGA breast cancer datasets, we identified 110 differentially expressed microRNAs. The functional enrichment analysis of the upregulated microRNAs revealed signaling transduction pathways, such as Notch and Wnt signaling pathway, and metabolism-related pathways such as sugar and nucleotide sugar metabolism, and oxidative stress response. Moreover, multivariable Cox model based on three variables of hsa-mir-130a, hsa-mir-3677, and hsa-mir-1247 stratified patients into high-risk and low-risk groups, which showed significant prognostic difference. In addition, we also tested the performance of this model in patient cohorts of any specific breast cancer subtypes or different TNM stages. The high performance in risk prediction was also observed in all of breast cancer subtypes and TNM stages. We also observed that there were highly possible interactions between hsa-mir-130a and seven target genes. Among these target genes, VAV3 and ESR1 were predicted as the target genes of hsa-mir-130a, suggesting that hsa-mir-130a may function by regulating the expression of VAV3 and ESR1 in breast cancer. In conclusion, the stratification based on the multivariable Cox model showed high performance in risk prediction. The dysregulated microRNAs and prognostic microRNAs greatly improved our understanding of the microRNA-related molecular mechanism underlying breast cancer.
Keywords: microRNAs; tumorigenesis; prognosis; breast cancer; hsa-mir-130a; VAV3; ESR1

1. Introduction

Breast cancer has become a major concern worldwide, as it is one of the most commonly diagnosed and leading causes of death among female patients with cancer [1]. The risk factors of breast cancer are of a great range, including pregnancy- and hormone-related factors, alcohol consumption and fat intakes, and environmental exposures [2]. Interestingly, though family history is often considered to introduce increased risks, only one out of nine breast cancer patients has an affected mother, sister, or daughter according to a previous study [3]. Notably, breast cancer is a highly heterogeneous disease. Immunopathologically, breast cancer can be categorized according to the receptor profile as indolent ER+ (ER+, HER2-), HER2+ (ER-, HER2+), triple positive (ER+, PR+, HER2+), and triple negative (TN) (ER-, PR-, HER2-) [4]. Meanwhile, 5 genetically distinct subtypes are reported so far based on mRNA expression profiling [5], including luminal A/B subtype, normal-like subtype, HER2-enriched subtype, and basal subtype [6], and prognostic outcomes exhibit a varying landscape among different subtypes. The luminal-A subtype is the most common and patients with this subtype have the best prognoses, while the prognoses of patients with luminal-B or normal-like subtype are slightly worse when compared with those with luminal-A tumors but still good; those with HER2-enriched subtype have poor prognoses as the tumors are highly proliferative, and the patients with basal-like subtype have the worst prognoses among all subtypes, as they are facing a higher risk of metastasis and early relapse [7]. Particularly, patients with breast cancer died mostly from chemo-resistance and metastasis [8], however, the molecular mechanisms underlying the progression of breast cancer are not fully understood, which is essential for its detection and management.

However, high-throughput technologies have enabled researchers to examine breast cancer at a closer level, and microRNA (miRNA), as a popular biomarker detected in many diseases, has attracted much attention. MicroRNA is a class of small non-coding RNAs, functioning as an important regulatory molecule in the post-transcriptional regulation. MicroRNAs can control the expression level of their target genes through binding with certain mRNA sequences, which leads to mRNA cleavage or destabilization [9,10]. MiRNAs often play diverse roles in tumor manifestations, as several miRNAs have been identified as tumor suppressors, such as miR-15a and miR-16-1, while some, including miR-21 and miR-155, function as oncogenes [11]. In breast cancer, over 50 miRNAs have been recognized as dysregulated that are associated with different hallmarks of cancers so far [8]. However, the key microRNAs involved in tumorigenesis and prognostic microRNAs have not been fully understood. In the present study, we aimed to perform a systematic analysis of microRNA expression profiles to identify some key microRNAs associated with tumor initiation and prognosis.

2. Materials and methods

2.1. MicroRNA and gene expression data collection

We collected microRNA and gene expression data of TCGA breast cancer (BRCA) tissues and normal tissues from UCSC Xena database [12] (https://xenabrowser.net/datapages/). The
gene and miRNA expression data was normalized and transformed by \( \log_2 (\text{FPKM} + 1) \) and \( \log_2 (\text{RPM} + 1) \), respectively.

2.2. **Differential expression analysis of microRNAs between breast cancer and normal tissues**

The microRNA expression values were first transformed to \( \log_2 (\text{RPM} + 1) \). To avoid the interference by the outliers, we used the Wilcoxon rank-sum test, which is a nonparametric method, and fold change to determine whether a given microRNA was differentially expressed. Bonferroni adjusted \( P \)-value < 0.05 for Wilcoxon rank-sum test and fold change >2 or <1/2 were chosen as the thresholds for differential expression.

2.3. **Functional enrichment analysis**

The upregulated and downregulated microRNAs were used to conduct the microRNA enrichment analysis, respectively. The analysis was implemented in a webserver, MiEEA, miRNA enrichment analysis and annotation (https://ccb-compute2.cs.uni-saarland.de/mieaa_tool/) [13]. The threshold for these gene sets was \( P \)-value < 0.05.

2.4. **MiRNA target prediction in breast cancer**

The mRNA-miRNA interaction network was downloaded from miRTarBase database [14]. The interactions for Homo sapiens were extracted. The expression of microRNAs and their target genes across the tumor tissues must have reverse expression patterns with Pearson correlation coefficient < −0.3 and \( P < 0.0001 \).

2.5. **Cox-regression based survival analysis**

For those dysregulated microRNAs in breast cancer, their expression values were then marked as high or low if the expression values were higher or lower than their corresponding median expression, respectively. The subset of microRNAs to be used in the multivariable Cox model were selected by Maximum Minimum Parents and Children (MMPC) algorithm, which had a higher statistical significance in coefficient estimation and better performance in model fitting than any other combinations [15]. As with the dichotomization in microRNA expression, the samples were stratified based on the median of risk scores estimated by the Cox model. Survival analysis was implemented in R programming software with `coxph` function in `survival` package. The Kaplan-Meier curves were used to visualize the overall survival for each group.

3. **Results**

3.1. **Dysregulated microRNAs in breast cancer**

We collected microRNA expression profiles of 1195 samples, with 1091 tumor tissues and 104 normal tissues, from TCGA breast cancer (BRCA) database. From their microRNA expression profiles, we quantified a total of 1881 microRNAs. To identify key microRNAs involved in
tumorigenesis of breast cancer, we performed differential expression analysis of identified microRNAs. Totally, we identified 110 differentially expressed microRNAs (Figure 1A, Bonferroni adjusted P-value < 0.05 and fold change > 2 or < 1/2), including 46 up-regulated and 64 down-regulated microRNAs. The dysregulated microRNAs exhibited significantly different expression patterns between the tumor and normal tissues (Figure 1B).

Among the dysregulated microRNAs, hsa-mir-21, hsa-mir-96, hsa-mir-183, hsa-mir-592, and hsa-mir-182 were identified as the top-five upregulated microRNAs. Moreover, the top-five downregulated microRNAs consisted of hsa-mir-139, hsa-mir-10b, hsa-mir-99a, hsa-mir-145, and hsa-let-7c (Table 1). Notably, the top-five upregulated microRNAs have been reported to play an oncogenic role in cancers [16–20]. In contrast, the top-five downregulated microRNAs have been reported to act as tumor suppressors [21–25]. These results indicated that the comparison between tumor and normal tissues could reveal some key microRNAs involved in tumorigenesis.

Table 1. The top-five significantly up-regulated and down-regulated miRNAs in breast cancer.

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Primary Tumor (mean)</th>
<th>Primary Tumor (sd.)</th>
<th>Normal (mean)</th>
<th>Normal (sd.)</th>
<th>P-value</th>
<th>Log2 Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-mir-21</td>
<td>17.83</td>
<td>0.62</td>
<td>15.59</td>
<td>0.86</td>
<td>$3.69 \times 10^{-59}$</td>
<td>2.23</td>
</tr>
<tr>
<td>hsa-mir-96</td>
<td>5.05</td>
<td>1.14</td>
<td>2.20</td>
<td>0.87</td>
<td>$1.65 \times 10^{-56}$</td>
<td>2.85</td>
</tr>
<tr>
<td>hsa-mir-183</td>
<td>13.81</td>
<td>1.12</td>
<td>10.89</td>
<td>1.00</td>
<td>$1.82 \times 10^{-56}$</td>
<td>2.92</td>
</tr>
<tr>
<td>hsa-mir-592</td>
<td>2.10</td>
<td>1.15</td>
<td>0.20</td>
<td>0.39</td>
<td>$1.81 \times 10^{-51}$</td>
<td>1.90</td>
</tr>
<tr>
<td>hsa-mir-182</td>
<td>15.23</td>
<td>1.04</td>
<td>12.97</td>
<td>0.96</td>
<td>$1.75 \times 10^{-50}$</td>
<td>2.26</td>
</tr>
<tr>
<td>hsa-mir-139</td>
<td>5.29</td>
<td>1.04</td>
<td>8.50</td>
<td>0.83</td>
<td>$4.79 \times 10^{-61}$</td>
<td>−3.22</td>
</tr>
<tr>
<td>hsa-mir-10b</td>
<td>15.78</td>
<td>0.96</td>
<td>17.95</td>
<td>0.46</td>
<td>$9.93 \times 10^{-59}$</td>
<td>−2.17</td>
</tr>
<tr>
<td>hsa-mir-99a</td>
<td>9.05</td>
<td>1.22</td>
<td>11.44</td>
<td>0.65</td>
<td>$7.37 \times 10^{-56}$</td>
<td>−2.39</td>
</tr>
<tr>
<td>hsa-mir-145</td>
<td>10.14</td>
<td>1.03</td>
<td>12.57</td>
<td>0.87</td>
<td>$2.31 \times 10^{-54}$</td>
<td>−2.43</td>
</tr>
<tr>
<td>hsa-let-7c</td>
<td>10.87</td>
<td>1.12</td>
<td>12.90</td>
<td>0.57</td>
<td>$2.76 \times 10^{-33}$</td>
<td>−2.03</td>
</tr>
</tbody>
</table>

3.2. Functional enrichment analysis of the dysregulated microRNAs

Similar to gene set enrichment analysis (GSEA), to reveal the biological functions of these dysregulated microRNAs, we performed microRNA enrichment analysis (MEA) using miRNA enrichment analysis and annotation [13] (MiEAA, https://ccb-compute2.cs.uni-saarland.de/mieaa_tool/). The upregulated microRNAs were mostly enriched in cancer related signaling transduction pathways (P-value < 0.05, Figure 2A), such as Notch and Wnt signaling pathway, which exhibited high concordance with the gene set enrichment analysis of the dysregulated genes [26,27]. Moreover, metabolism-related pathways such as sugar and nucleotide sugar metabolism, and oxidative stress response were also enriched by these upregulated microRNAs (P-value < 0.05, Figure 2A). In addition to these pathways, the target genes of upregulated microRNAs, such as TUBA1C, SYMPK, RPS2, HSP90AA1, EIF4EBP2, and BCL2, were identified by the MEA (Figure 2B). Particularly, BCL2, a favorable prognostic biomarker in breast cancer [28], was also downregulated in breast cancer samples compared with the normal tissues (P < 0.0001, Figure 2C, Supplementary Table S1), with an opposite expression pattern of those upregulated microRNAs. In contrast, although the downregulated microRNAs were not
significantly enriched in any pathways and no target genes were found, they were significantly enriched in some Gene Ontology (GO) terms, such as COPII vesicle coat, mesenchymal cell development, protein phosphatase type 2a regulator activity, and phosphatidylinositol 3 4 5 trisphosphate binding (Figure 2D, Supplementary Table S2). The mesenchymal cell development indicated that the downregulated microRNAs may participate in the epithelial-mesenchymal transition (EMT). The functional enrichment analysis revealed some breast cancer-related pathways, target genes, and GO terms that the dysregulated microRNAs may be involved in.

![Figure 1](image1.png)

**Figure 1.** Overview of the dysregulated microRNAs in breast cancer. A. The points with red and blue colors represent the upregulated and downregulated microRNAs, respectively. B. Each row of the heatmap represents one of dysregulated microRNAs. The color band on the top represent the original tissues.

3.3. Prognostic performance of the dysregulated microRNAs in breast cancer

Among the dysregulated microRNAs, some may be associated with the patients’ prognosis. Based on univariate Cox regression model, we selected 9 microRNAs significantly associated with the overall survival of breast cancer patients (Figure 3A, Log-rank test, P-value < 0.05). The four upregulated microRNAs were associated with poor overall survival, while the five downregulated microRNAs had a favorable effect on the breast cancer patients (Figure 3A). To build a multivariable Cox regression model, we used MMPC algorithm [15] to select features. Finally, we selected hsa-mir-130a, hsa-mir-3677, and hsa-mir-1247 as the variables of the multivariable Cox model (Figure 3B). The samples were then stratified into high-risk and low-risk groups based on their risk scores estimated by the Cox model. The high-risk group showed significantly worse overall survival
than the low-risk group (Figure 3C, log-rank test, P-value = 0.00017). These results suggested that a combination of the dysregulated microRNAs had a high performance in the prediction of breast cancer prognosis.

Figure 2. Pathway, microRNA target and Gene Ontology enrichment analysis. A. The signaling pathways enriched by the upregulated microRNAs. B. The microRNA targets enriched by the upregulated microRNAs. C. The expression patterns of the BCL2 in TCGA breast cancer and normal tissues. D. GO terms enriched by the downregulated microRNAs.
3.4. The microRNA-based stratification is an independent prognostic factor in breast cancer

To further investigate the performance of the microRNA-based stratification, we also tested their independence from both breast cancer subtypes and TNM stages. Specifically, we observed that the high-risk group showed worse prognosis than the low-risk group in luminal-A (n = 384), luminal-B (n = 110), HER2-positive (n = 33), and triple-negative (n = 99) breast cancers at the significance level of 0.1 (Log-rank test, Figure 4A–D). Notably, despite of minimal sample size, the HER2-positive subtype still had worse prognosis in high-risk group than in the low-risk group at a relatively higher significance level (Figure 4C).
Moreover, we also tested the difference in the overall survival between the high-risk and low-risk groups for a specific TNM stage. Except for samples of the stage IV, which had a smaller sample size (n < 30) and were excluded in this analysis, samples of the other three stages (the sample sizes for I, II, and III: 183, 609, and 242) were used to compare the overall survival of high-risk group with that of low-risk group. Consistently, the high-risk group still exhibited poorer overall survival than low-risk group in all the three TNM stages (Figure 5, P-value < 0.05). These results indicated that the microRNA-based stratification is a prognostic factor independent of subtypes and TNM stages in breast cancer.

Figure 4. The Kaplan–Meier curves for the breast cancer subtypes The Kaplan-Meier curves for luminal-A, luminal-B, HER2-positive, and triple negative subtypes are illustrated in A, B, C and D, respectively.

3.5. The target genes of the microRNAs used in the Cox model

As the microRNAs function by binding to the 3’ untranslated regions (3’ UTR) of their target genes, we then investigated the target genes of the three microRNAs used in the Cox model. We
first collected 322,160 experimentally validated microRNA-mRNA pairs from the miRTarBase database [14]. Given a stringent threshold of P-value < 0.0001 (Correlation test), we only identified 7 pairs of microRNAs and target genes, which consisted of one microRNA (hsa-mir-130a) and seven target genes. The samples were then classified into two groups (high or low expression of hsa-mir-130a), and higher expressions of the seven target genes were observed in the samples with low expression of hsa-mir-130a (Figure 6, Wilcoxon rank-sum test, P-value < 0.0001). Among these target genes, VAV3 and ESR1 have been widely reported to be implicated in breast cancer [29,30], suggesting that hsa-mir-130a may function by regulating the expression of VAV3 and ESR1 in breast cancer.

![Kaplan-Meier curves for breast cancer patients within one of TNM stages](image)

**Figure 5.** The Kaplan-Meier curves for the breast cancer patients within one of TNM stages. The Kaplan-Meier curves for stages I, II, and III are illustrated in A, B, and C, respectively.

4. **Discussion**

Breast cancer is one of the most diagnosed cancers in women. It contributes greatly to cancer-related death among female patients with cancer [1], and raises concern all over the world. However, the key tumorigenic and prognostic microRNAs have not been fully uncovered.
In the present study, we aimed to perform a systematic analysis of microRNA expression profiles to identify some key microRNAs associated with tumor initiation and prognosis. We collected microRNA expression profiles of 1195 samples, with 1091 tumor tissues and 104 normal tissues, from TCGA breast cancer (BRCA) database. Totally, we identified 110 differentially expressed microRNAs (Figure 1A, Bonferroni adjusted P-value < 0.05 and fold change >2 or <1/2), including 46 up-regulated and 64 down-regulated microRNAs. The top-five upregulated microRNAs have been reported to play an oncogenic role in cancers [16–20]. Among the top-five upregulated microRNAs, hsa-mir-21, overexpression in human breast cancer, is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis [31]. Notably, the miR-183/-96/-182 cluster is up-regulated in most breast cancers and increased cell proliferation and migration are observed [32]. The top-five down-regulated microRNAs, hsa-mir-139, hsa-mir-10b, hsa-mir-99a, hsa-mir-145, and hsa-let-7c, were famous tumor suppressors [21–25]. The functional enrichment analysis of the upregulated microRNAs revealed signaling transduction pathways, such as Notch and Wnt signaling pathway, and metabolism related pathways such as sugar and nucleotide sugar metabolism, and oxidative stress response (P-value < 0.05, Figure 2A). Moreover, the downregulated microRNAs were significantly enriched in GO terms, such as COPII vesicle coat, mesenchymal cell development, protein phosphatase type 2a regulator activity, and phosphatidylinositol 3 4 5 trisphosphate binding (Figure 2D). The mesenchymal cell development indicated that the downregulated microRNAs may participate in the epithelial-mesenchymal transition (EMT).
In addition, we also examined the prognostic values of the dysregulated microRNAs. Multivariable Cox model based on three variables of hsa-mir-130a, hsa-mir-3677, and hsa-mir-1247 stratified samples into high-risk and low-risk groups, which exhibited significant prognostic difference. Moreover, we also tested the performance of this stratification method in any specific breast cancer subtype or TNM stage. The high performance in risk prediction was also observed in all of breast cancer subtypes and TNM stages. These results not only indicated that the stratification based on the multivariable Cox model had high performance in risk prediction, but also demonstrated the high efficiency of the algorithm for the feature selection. Among the target genes of hsa-mir-130a, VAV3 and ESR1 have been widely reported to be implicated in breast cancer [29,30], suggesting that hsa-mir-130a may function by regulating the expression of VAV3 and ESR1 in breast cancer.

However, some limitations still existed in the present study. First, these dysregulated microRNAs should be further investigated by experimental validation. Second, more samples are needed to further validate the predictive performance of three microRNAs in the multivariable Cox model. We hope to conduct further research with functional experiments and more samples in the near future. In conclusion, we have identified key microRNAs implicated in breast cancer, which improved our understanding of the microRNA-related molecular mechanism underlying breast cancer.

Conflict of Interest

The authors declare that they have no conflict of interest.

References


