Identification of Long Noncoding RNAs as Predictors of Survival in Triple-Negative Breast Cancer Based on Network Analysis

Xiao-Xiao Li,1,2 Li-Juan Wang,1,2 Jie Hou,1 Hong-Yang Liu,1 Rui Wang,1 Chao Wang,1 and Wen-Hai Xie1,2

1School of Life Sciences, Shandong University of Technology, Zibo, China
2Institute of Biomedical Research, Shandong University of Technology, Shandong Provincial Research Center for Bioinformatics Engineering and Technique, Zibo Key Laboratory of New Drug Development of Neurodegenerative Diseases, Zibo, China

Correspondence should be addressed to Wen-Hai Xie; xiewenhai@sdut.edu.cn

Received 8 October 2019; Revised 31 December 2019; Accepted 21 January 2020; Published 4 March 2020

Copyright © 2020 Xiao-Xiao Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Breast cancer is the most common cancer observed in adult females, worldwide. Due to the heterogeneity and varied molecular subtypes of breast cancer, the molecular mechanisms underlying carcinogenesis in different subtypes of breast cancer are distinct. Recently, long noncoding RNAs (lncRNAs) have been shown to be oncogenic or play important roles in cancer suppression and are used as biomarkers for diagnosis and therapy. In this study, we identified 134 lncRNAs and 6,414 coding genes were differentially expressed in triple-negative (TN), human epidermal growth factor receptor 2- (HER2-) positive, luminal A-positive, and luminal B-positive breast cancer. Of these, 37 lncRNAs were found to be dysregulated in all four subtypes of breast cancers. Subtypes of breast cancer special modules and lncRNA-mRNA interaction networks were constructed through weighted gene coexpression network analysis (WGCNA). Survival analysis of another public datasets was used to verify the identified lncRNAs exhibiting potential indicative roles in TN prognosis. Results from heat map analysis of the identified lncRNAs revealed that five blocks were significantly displayed. High expressions of lncRNAs, including LINC00911, CSMD2-AS1, LINC01192, SNHG19, DSCAM-AS1, PCAT4, ACVR28-AS1, and CNTFR-AS1, and low expressions of THAP9-AS1, MALAT1, TUG1, CAHM, FAM2011, NNT-AS1, COX10-AS1, and RPARD-AS1 were associated with low survival possibility in TN breast cancers. This study provides novel lncRNAs as potential biomarkers for the therapeutic and prognostic classification of different breast cancer subtypes.

1. Introduction

Breast cancer is the most frequently occurring cancer and the second cause of female cancer mortality worldwide. According to a report from the International Agency for Research on Cancer, breast cancer accounts for approximately 11.6% of cancer incidences [1]. Although deaths caused by breast cancer have declined over time due to the development of new diagnostic and therapeutic strategies, the molecular mechanisms underlying this disease remain to be revealed. As breast cancers are notoriously heterogeneous, clinical and morphologic signatures are diversified. Consequently, the exact classification of breast cancer is a challenge for clinicians and scientist. Variations in the expressions of the estrogen receptor (ER), human epidermal growth factor receptor 2 (HER2), and progesterone receptor (PR) can be used to classify breast cancers into four major intrinsic molecular subtypes: triple-negative (TN), luminal A, luminal B, and HER2-positive [2]. TN breast cancers are well-known as lacking ER and PR as well as HER2 [3]. Luminal A breast cancers are defined as ER-positive, HER2-negative, and “low” recurrence risk, partially expressing PR [4]. Luminal B breast cancers are characterized by HER2-positive, ER and PR-negative, and a high histologic grade [5]. HER2-positive breast cancers are defined as HER2-positive and lacking of ER and PR. Along with ER, PR, and HER2, many
other molecular genes also have been used to subtype breast cancers. High expression of Ki67, known as a proliferation marker, was usually been used to define breast cancer subtypes [6]. Molecular classifications are using multigene assays to classify breast cancers into low- and high-risk groups for personalized therapy [7, 8]. Many genes have been found as signatures in the diagnosis and prognosis of breast cancer. Decreased levels of uPA and PAI-1 are associated with lower risk of cancer recurrence, and overexpression of Ki67, cyclin D, cyclin E, p27, and p21 indicates uncontrolled tumor cell proliferation [9]. The molecular interaction regulation mechanisms mediating heterogeneity and different molecular subtypes of breast cancer need to be further probed.

Long noncoding RNAs (lncRNAs), which are 200 nucleotides in length and play important roles in various biological pathways, have been shown to be oncogenic or play important roles in cancer suppression and are used as biomarkers for diagnosis and therapy [10, 11]. Aberrant expressions of lncRNAs, such as MALAT1 [12], DANCR [13], PDCD4-AS1 and [14], MIR100HG [15], are closely associated with the occurrence and progression of breast cancer. Overexpression of HOTAIR stimulates the invasion and metastasis of breast cancer cells [16]. PIWI-interacting noncoding RNA especially expressed in breast cancer is a response to estrogen regulation in breast cancer cells [17]. lncRNAs can also mediate the progression of breast cancer by interacting with microRNAs. For example, MEG3 acts as a suppressor to inhibit cell epithelial-mesenchymal transition by sponging miR-421 and targeting E-cadherin in breast cancer [18]. Highly expressed H19/low miR-675 and low NEAT1/high miR-204 could differentiate breast cancer subtypes and could be considered as diagnostic and therapeutic biomarkers [19]. lncRNAs can also act as epigenetic regulators affecting histone modifications and stabilize signal complexes and nuclear structures by recruiting chromatin modification factors [20]. Although a growing volume of lncRNAs has been found playing varied roles in the pathogenesis of breast cancer and drug resistance, their biological participation mechanisms are not yet fully understood.

TN breast cancer is a heterogeneous subtype of breast cancers with poor prognosis and is beginning to be refined by molecular characteristics. Based on gene expression profiles, six TN breast cancer subtypes were classified [21]. As the high heterogeneous of TN breast cancer, it is necessary to disclose its molecular characteristics and identify effective clinical strategy to improve the therapeutic approach. A series of lncRNAs has been found dysregulated in TN breast cancer [22]. lncRNAs are involved in the regulation of most mammalian protein coding genes, which are responsible for various cellular processes such as cell differentiation, development, proliferation, and apoptosis. Exploitation of the function of dysregulated lncRNAs will provide potential clinical applications for TN’s diagnosis and treatment. WGCNA can be used to identify key lncRNAs associated with multiple cancer pathogenesis and progression [23–26]. Survival analysis verified the identified lncRNAs possessing potential indicative roles in TN prognosis. These survival expectation indictors associated with coexpressing coding genes formed coexpressed axes, which could act as attractive therapeutic targets for the treatment of TN breast cancer. The workflow of this study is summarized in Figure 1.

2. Materials and Methods

2.1. Breast Cancer Expression Profiles Downloaded from Gene Expression Omnibus. The public datasets GSE45827 and GSE58812 were downloaded from the NCBI Gene Expression Omnibus (GEO). GSE45827 includes 41 TN, 30 HER2, 29 luminal A, and 30 luminal B breast cancer patients and 11 normal tissue samples [27]. GSE58812 includes 107 triple-negative breast cancer patients [28]. Normal tissue samples were used as a control to identify differentially
expressed coding genes (DEGs) and differentially expressed lncRNAs (DElncRNAs). In order to reduce the bias of different data resources, datasets adopted in this study were both based on the same GPL570 platform. The type of microarray was Affymetrix Human Genome U133 plus 2.0 Array. All raw data preprocessing was conducted in the R software environment. The Affy package was adopted to process the initial datasets [29]. A robust multiarray average function was used to detect and normalize the expression of probes. The Limma package was used for differential expression analysis [30]. The Benjamini and Hochberg method was used to control the false discovery rate (FDR). The adjusted p values were used to control the FDR. The FDR ratio was set as 5%. Adjusted p values < 0.05 and log (fold change) > 1 were selected as the thresholds for DEGs and DElncRNAs.

2.2. IncRNA Classification Pipeline. The pipeline of IncRNA classification was utilized to identify IncRNA expression as previously described [31], with minor modifications. The Affymetrix HG-U133 plus 2.0 probe set ID was mapped to the latest NetAffx annotation files. First, we extracted the probe sets with RefSeq IDs labelled as "NR_". For the selected probe sets annotated with Ensembl gene IDs, we annotated with IncRNA, processed the transcript noncoding and antisense according to Homo_sapiens.GRCH38.96.gtf. The last probe sets without Ensembl gene IDs were filtered using pseudogenes, rRNAs, microRNAs, snoRNA, and tRNAs according to NetAffx annotation files. Finally, we obtained 2,553 Affymetrix probe IDs, which were annotated as IncRNA transcripts (1,870 annotated lncRNAs).

2.3. Construction of a Weighted Gene Coexpression Network and Identification of Modules Associated with Different Subtypes of Breast Cancer. WGCNA was adopted to identify coexpression modules [32]. The signed weighted correlation network was constructed by the first creating a matrix of pairwise correlations between all pairs of genes [33]. The power of seven was interpreted as a soft threshold of the correlation matrix. Depending on the resulting adjacency matrix, we calculated the topological overlap matrix, which measures the interconnectedness of the coexpression network. Then, genes with highly similar coexpression relationships were grouped together by hierarchical clustering based on the topological overlap matrix. The Dynamic Hybrid Tree Cut algorithm was adopted to define modules, meaning the coexpression genes. Module eigengenes were defined as the first principal component to represent each module. Finally, modules were merged based on correlation above 0.85. After identifying the coexpression modules, we associated modules with each sample type. Modules with correlation > 0.5 (p < 0.01) were picked up as respective subtypes of breast cancer.

2.4. Construction and Analysis of IncRNA-mRNA Coexpression Network. Based on the Pearson correlation coefficient of lncRNA-mRNA, we constructed breast cancer-related lncRNA-mRNA coexpression networks. The threshold of selection for IncRNA-mRNA pairs was set at an absolute value of the Pearson correlation coefficient above 0.6. Topological analysis of the coexpression network was performed using the Cytoscape (version 3.7.1) software [34]. Hub genes were determined using high connectivity of genes by summing the connection strengths with other module genes.

2.5. Gene Ontology Analysis. Gene ontology analysis of identified modules and core genes was performed using DAVID 6.8 [35]. The corrected FDR of a p value < 0.05 was selected as the threshold for enrichment of the GO_ FAT terms.

2.6. Validation of Identified IncRNA Roles in TN Cancer Classification. The dataset GSE58812 was used to verify the important roles of identified DElncRNAs. Preprocessing of GSE58812 was conducted in the R software as described above. After obtaining the expression matrix of the datasets, these TN cancer samples were classified into groups based on principal component analysis. The clustering was performed by package ggfortify of R [36]. As the raw reference of GSE58812 classified the 107 TN breast cancer samples into three subtypes, we set the cluster number as 3. According to the clinical data of these TN cancer samples, survival analysis was performed using survminer package [37]. The median of DElncRNA expressions among different TN cancer sample groups was adopted to display the panel of predicting survival expectation.

3. Results

3.1. DEGs and DElncRNAs in Different Subtypes of Breast Cancer Patients. After applying IncRNA classification, we annotated the microarray with 17,242 coding genes and 1,870 lncRNAs. We identified 3,514 coding genes and 96 lncRNAs, 3,181 coding genes and 87 lncRNAs, 2,198 coding genes and 60 lncRNAs, and 2,841 coding genes and 70 lncRNAs differentially expressed in TN, HER2, luminal A, and luminal B breast cancers compared to normal tissues (Figure 2). The four different subtypes of breast cancer shared 37 common differentially expressed lncRNAs (Figure 3(a)). Most of the 37 common differentially expressed noncoding RNAs were downregulated in the four different subtypes of breast cancers compared to normal tissues (Table 1). This means that the downregulated lncRNAs may be the tumor suppressor genes in the normal tissues. Only five lncRNAs (HOTAIR, DLEU2, THAP9-AS1, POLR2J4, and LINCO1614) were upregulated (Figures 3(e)–3(f)). LncRNA HOTAIR has been proven to play a significant role in promoting invasion of breast cancer cells [38]. DLEU2 was reported to play a role in response to estrogen regulation in human breast cancer cell line [39]. POLR2J4 has also been used to predict recurrence-free survival in hepatocellular carcinoma [40]. Knockdown of LINCO1614 inhibits lung adenocarcinoma cell progression [41]. THAP9-AS1 was the first time observed to be associated with breast cancer.

3.2. WGCNA Analysis of the Coding Genes and IncRNAs. The combination of these differentially expressed coding genes and IncRNAs formed a matrix dataset including 6,414 genes
and 134 lncRNAs among 141 samples. The network was constructed with the soft threshold power of 7 (Figure 4(a)). We identified 15 gene modules based on a minimum module size of 30 merged modules with eigengene correlation above 0.85. Modules related to different breast cancer subtypes are shown in Figure 4(b). A gene dendrogram was obtained using average linkage hierarchical clustering. The color row was painted according to the expression values of genes in the dendrogram. The turquoise module was highly associated with TN, and midnight blue was related to HER2 and luminal B breast cancers. Luminal B breast cancer was highly associated with the grey modules compared to other subtypes of breast cancers. Luminal A has three relatively high modules including black, green, and magenta. Luminal A and luminal B have the same relationship with the green modules (Figure 4(c)). These related modules will provide more information for subtyping of different breast cancers. In addition, the brown and blue modules were, respectively, positively and negatively correlated with normal tissue samples. The 15 gene modules eigengenes were classified into three groups using hierarchical clustering (Figure 4(d)).

3.3. Construction of Breast Cancer-Related lncRNA–mRNA Coexpression Networks. Based on the Pearson correlation coefficients of lncRNA–mRNA, we constructed breast cancer-related lncRNA–mRNA coexpression networks using WGCNA. The threshold for the selection of lncRNA–mRNA pairs was set at an absolute value of the Pearson correlation coefficient above 0.6. The TN-related lncRNA–mRNA coexpression network derived from the turquoise module consists of 19 lncRNAs and 218 DEGs (Figure 5(a)). Maternally expressed gene 3 (MEG3) was the hub of lncRNA in the constructed network with 219 interaction links. Previous studies showed that MEG3 was downregulated in various types of cancers including breast cancer. Meta-analysis demonstrated that low expression of MEG3 is associated with poor survival.
in cancer patients [42]. GO analysis of MEG3 linked 219 genes, which enriched cell adhesion, cell proliferation, immune system processes, and other processes. The network could provide useful information for the mechanism of MEG3. CEP55, MELK, and KIF11 were the top 3 degree DEGs in the network. CEP55 is a determinant of cell fate during perturbed mitosis in breast cancer [43].

Luminal A- and luminal B-related IncRNA-mRNA coexpression network derived from the green module consists of 9 IncRNAs and 144 DEGs. LINC01116, LINC01087, AC110619, FOXP4-AS1, and TP53TG1 were the top nodes in this network, while MLPH, AGR2, SMIM4, FOXA1, INPP4B, MAGED, P4HTM, and AGR3 mediated the connection of the whole network.

Normal tissue-related IncRNA-mRNA coexpression network derived from the brown module consists of 23 IncRNAs and 405 DEGs. The top nodes in this network were ADMTS9-AS2, HSD11B1-AS1, LINC01697, MIR497HG, LINC01140, and CARMN. BTNL9 and TNS1 were the top degree DEGs in the constructed network. TNS1 modulated the activation of Cdc42 to regulate cell invasion in breast cancer [44]. The nonnormal-related IncRNA-mRNA

Figure 3: Venn diagram of DElncRNAs among the four molecular subtypes of breast cancer. Venn diagram shows the interactions of DElncRNAs among the four different molecular subtypes of breast cancer (a). The expression of five upregulated DElncRNAs in the four different subtypes of breast cancer and normal tissues was plotted by ggstatsplot (b–f). The expression values of upregulated DElncRNAs were shown as HOTAIR (b), DLEU2 (c), THAP9-AS1 (d), POLR2J4 (e), and LINC01614 (f).
Table 1: The fold changes of core 37 DElncRNAs in breast compared to normal tissue.

<table>
<thead>
<tr>
<th>IncRNA symbol</th>
<th>IncRNA Ensembl ID</th>
<th>TN Fold change</th>
<th>Her2 positive Fold change</th>
<th>Luminal A positive Fold change</th>
<th>Luminal B positive Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLEU2*</td>
<td>ENSG00000231607</td>
<td>2.17</td>
<td>2.04</td>
<td>2.21</td>
<td>2.10</td>
</tr>
<tr>
<td>HOTAIR*</td>
<td>ENSG00000228630</td>
<td>3.14</td>
<td>7.90</td>
<td>2.51</td>
<td>5.13</td>
</tr>
<tr>
<td>LINC01614*</td>
<td>ENSG00000230838</td>
<td>5.10</td>
<td>11.35</td>
<td>11.18</td>
<td>16.97</td>
</tr>
<tr>
<td>POLR2J4*</td>
<td>ENSG00000214783</td>
<td>3.44</td>
<td>3.17</td>
<td>3.13</td>
<td>2.87</td>
</tr>
<tr>
<td>THAP9-AS1*</td>
<td>ENSG00000251022</td>
<td>2.68</td>
<td>2.44</td>
<td>2.81</td>
<td>2.43</td>
</tr>
<tr>
<td>AC005064.1</td>
<td>ENSG00000234715</td>
<td>-3.03</td>
<td>-3.08</td>
<td>-2.05</td>
<td>-2.90</td>
</tr>
<tr>
<td>AC099342.1</td>
<td>ENSG00000226097</td>
<td>-4.04</td>
<td>-4.00</td>
<td>-3.79</td>
<td>-3.74</td>
</tr>
<tr>
<td>AC245123.1</td>
<td>ENSG00000277526</td>
<td>-2.21</td>
<td>-2.18</td>
<td>-2.28</td>
<td>-2.23</td>
</tr>
<tr>
<td>ACVR2B-AS1</td>
<td>ENSG00000229589</td>
<td>-3.05</td>
<td>-3.22</td>
<td>-2.88</td>
<td>-3.02</td>
</tr>
<tr>
<td>ADAMTS9-AS2</td>
<td>ENSG00000241684</td>
<td>-5.70</td>
<td>-5.03</td>
<td>-3.06</td>
<td>-5.24</td>
</tr>
<tr>
<td>AP001816.1</td>
<td>ENSG00000254351</td>
<td>-2.42</td>
<td>-2.10</td>
<td>-2.09</td>
<td>-2.05</td>
</tr>
<tr>
<td>AP002518.1</td>
<td>ENSG00000256195</td>
<td>-2.10</td>
<td>-2.20</td>
<td>-2.16</td>
<td>-2.20</td>
</tr>
<tr>
<td>AP003110.1</td>
<td>ENSG00000255084</td>
<td>-2.03</td>
<td>-2.02</td>
<td>-2.10</td>
<td>-2.04</td>
</tr>
<tr>
<td>CARMN</td>
<td>ENSG00000249669</td>
<td>-6.18</td>
<td>-4.67</td>
<td>-3.02</td>
<td>-4.10</td>
</tr>
<tr>
<td>FGF14-AS2</td>
<td>ENSG00000272143</td>
<td>-2.23</td>
<td>-2.90</td>
<td>-2.41</td>
<td>-2.57</td>
</tr>
<tr>
<td>FZD10-AS1</td>
<td>ENSG00000250208</td>
<td>-2.62</td>
<td>-2.66</td>
<td>-2.08</td>
<td>-2.57</td>
</tr>
<tr>
<td>GAS1R</td>
<td>ENSG00000262377</td>
<td>-2.38</td>
<td>-2.68</td>
<td>-2.18</td>
<td>-2.43</td>
</tr>
<tr>
<td>GLIDR</td>
<td>ENSG00000278175</td>
<td>-2.89</td>
<td>-3.65</td>
<td>-2.02</td>
<td>-2.99</td>
</tr>
<tr>
<td>HOTAIRM1</td>
<td>ENSG00000233434</td>
<td>-3.59</td>
<td>-3.80</td>
<td>-3.39</td>
<td>-3.89</td>
</tr>
<tr>
<td>HSD11B1-AS1</td>
<td>ENSG00000227591</td>
<td>-5.50</td>
<td>-5.87</td>
<td>-5.65</td>
<td>-5.94</td>
</tr>
<tr>
<td>LINC00358</td>
<td>ENSG00000229578</td>
<td>-2.25</td>
<td>-2.27</td>
<td>-2.19</td>
<td>-2.21</td>
</tr>
<tr>
<td>LINC00408</td>
<td>ENSG00000262520</td>
<td>-6.38</td>
<td>-6.15</td>
<td>-4.85</td>
<td>-5.02</td>
</tr>
<tr>
<td>LINC00968</td>
<td>ENSG00000246430</td>
<td>-2.17</td>
<td>-2.20</td>
<td>-2.18</td>
<td>-2.21</td>
</tr>
<tr>
<td>LINC01140</td>
<td>ENSG00000267272</td>
<td>-3.15</td>
<td>-3.20</td>
<td>-2.15</td>
<td>-2.65</td>
</tr>
<tr>
<td>LINC01192</td>
<td>ENSG00000241369</td>
<td>-2.38</td>
<td>-2.42</td>
<td>-2.51</td>
<td>-2.57</td>
</tr>
<tr>
<td>LINC01697</td>
<td>ENSG00000232079</td>
<td>-5.65</td>
<td>-5.65</td>
<td>-4.53</td>
<td>-5.73</td>
</tr>
<tr>
<td>LINC01720</td>
<td>ENSG00000231175</td>
<td>-2.05</td>
<td>-2.06</td>
<td>-2.02</td>
<td>-2.03</td>
</tr>
<tr>
<td>LINC02053</td>
<td>ENSG00000241696</td>
<td>-5.70</td>
<td>-5.53</td>
<td>-5.59</td>
<td>-5.07</td>
</tr>
<tr>
<td>LINC02535</td>
<td>ENSG00000234155</td>
<td>-2.09</td>
<td>-2.13</td>
<td>-2.15</td>
<td>-2.12</td>
</tr>
<tr>
<td>MEG3</td>
<td>ENSG00000214548</td>
<td>-6.39</td>
<td>-3.67</td>
<td>-2.37</td>
<td>-3.64</td>
</tr>
<tr>
<td>MIR497HG</td>
<td>ENSG00000267532</td>
<td>-2.69</td>
<td>-2.82</td>
<td>-2.34</td>
<td>-2.68</td>
</tr>
<tr>
<td>PGM5-AS1</td>
<td>ENSG00000224958</td>
<td>-2.35</td>
<td>-3.22</td>
<td>-2.18</td>
<td>-2.35</td>
</tr>
<tr>
<td>PRICKLE2-AS3</td>
<td>ENSG00000226017</td>
<td>-2.26</td>
<td>-2.13</td>
<td>-2.15</td>
<td>-2.31</td>
</tr>
<tr>
<td>SNIHG9</td>
<td>ENSG00000175061</td>
<td>-3.47</td>
<td>-4.40</td>
<td>-2.37</td>
<td>-3.57</td>
</tr>
<tr>
<td>TRHDE-AS1</td>
<td>ENSG00000236333</td>
<td>-2.68</td>
<td>-2.87</td>
<td>-2.63</td>
<td>-2.68</td>
</tr>
<tr>
<td>WARS2-AS1</td>
<td>ENSG00000231365</td>
<td>-2.21</td>
<td>-2.49</td>
<td>-2.18</td>
<td>-2.31</td>
</tr>
<tr>
<td>ZNF436-AS1</td>
<td>ENSG00000249087</td>
<td>-2.88</td>
<td>-3.10</td>
<td>-2.57</td>
<td>-2.75</td>
</tr>
</tbody>
</table>

The five DElncRNAs upregulated in all the four breast cancer subtypes were marked with ***. Coexpression network derived from the blue module was negatively related with normal samples. This cancer-related lncRNA-mRNA coexpression network consists of 38 lncRNAs and 496 DEGs. LINC00408, AC099342.1, LINC02053, LINC01720, and AC245123.1 were the top nodes in this network, while DEFB106A, LECT2, TRIM49, BOLL, OPHN1, ZNF835, and EPPIN were the top degree DEG nodes in the network.

3.4. Function Annotation of Breast Cancer-Related lncRNAs

We performed function annotation of these constructed lncRNA-mRNA coexpression networks using the DAVID
blood vessel development and morphogenesis and the angio-
related lncRNA-mRNA coexpression network enriched
protein transport terms such as Golgi vesicle transport and vesicle organization (Figure 6(d)). That means the brown module is essential for normal tissue maintenance. GO analysis of the normal negative-related lncRNA-mRNA coexpression networks enriched protein transport terms such as Golgi vesicle transport and vesicle organization (Figure 6(d)). These lncRNAs interact with coding genes and mediate cancer cell protein transport and cellular morphology during tumorigenesis.

3.5. DElncRNAs Act as Survival Expectation Indicators Verified Using Public Datasets. The TN breast cancer samples of dataset GSE58812 were classified into three clusters based on the expression of these DElncRNAs (Figure 7(a)). According to
the clinical data of these TN cancer samples, survival analysis of the three clusters displayed different survival probabilities. The KM curves with a $p$ value of pairwise comparisons for the three clusters are shown in Figure 7(b). The survival probability of cluster 3 was 50% lower than those of cluster 1 or 2 (75%). Five blocks that acted as signature predictors of survival in TN breast cancer patients were significantly displayed using heat map analysis (Figure 7(c)). This means that different DElncRNAs play different roles in TN breast cancer progression by promoting or inhibiting cancer progression. The DElncRNAs THAP9-AS1, MALAT1, TUG1, CAHM, FAM2011, NNT-AS1, COX10-AS1, and RPAPP-AS1 were lowly expressed in cluster 3 compared with those of cluster 1 and 2 samples. The analysis of the TCGA database showed that higher expression of TUG1 was associated with better prognosis in breast cancer patients [46]. This is consistent with our bioinformatics prediction. LINC00911, CSM2D2-AS1, LINC01192, SNHG19, DSCAM-AS1, PCAT4, ACVR28-AS1, and CNTFR-AS1 were highly expressed in cluster 3. The relative expression values of DElncRNAs can be used as a predictor for survival expectation.

In order to detect the prediction roles of identified 97 lncRNAs individually, we performed survival analysis of single gene expression and TN breast cancer prognosis. The prediction with significant $p$ values between high and low expression values is displayed in Figure 8. High expression of LINC00911 (Figure 8(a)), H19 (Figure 8(b)), and MIRLRT7BH (Figure 8(c)), respectively, predicted poor survival in TN breast cancer. High expression of MIR155HG (Figure 8(d)), CHAM (Figure 8(e)), and FAM13A-AS1 (Figure 8(f)), respectively, were associated with higher survival probability. This was consistent with the above panel prediction. The individual expression of LINC00911, H19, and MIRLRT7BH can be used as a predictor for poor prognosis. MIR155HG, CHAM, and FAM13A-AS1 can be used as a positive predictor for survival probability.

3.6. The Coexpressing Coding Genes of Survival Expectation Indicators Extracted from the Above lncRNA-mRNA Coexpression Networks. From the above identified TN, luminal A, luminal B, and normal tissue-associated lncRNA-mRNA coexpression networks, useful interaction
information of lncRNAs, and coding genes could be found. LncRNAs LINC00911, LINC01192, PCAT4, and CNTFR-AS1 were listed in the normal samples that were negatively related to the blue module. Extraction of the lncRNA-mRNA interaction network is displayed in Figure 9(a). The four highly expressed lncRNAs associated with 31 coding genes formed a coexpression network. The top degree coding genes were painted with red colors. Several coding genes have been found playing important roles in recurrence, poor diagnosis, and low survival probabilities. Multivariable analyses revealed that LECT2 as one of the predictors for both breast carcinoma recurrence and mortality among smokers [47]. Diagnostically, ZNF835 was associated with ethnicity in colorectal cancer patients [48]. The positive of oligophrenin-1 was significantly correlated with a high Gleason score in prostate cancer [49]. Methyl-DNA binding domain capture technique identified Kcnv1 as a diagnostic marker for early noninvasive detection and subsequent breast cancer surveillance [50]. Brain-derived neurotrophic factor (BDNF) is a potent neurotrophic factor that has been shown to stimulate breast cancer cell growth and metastasis via tyrosine kinase receptors. The methylation of BDNF gene may be a biomarker for suicidal thoughts in patients with breast cancer [51]. BDNF/TrkB pathways activated microRNAs to act as prognostic and predictive biomarkers for detecting patients at a high risk of developing breast cancer [52]. This suggested that lncRNA interacted with Kcnv1 modification methylation levels to mediate survival in TN breast cancer. Increasing expression of SCD5 promoted tumor cell survival in breast cancer [53]. Phosphorylation of hnRNPK S379 participates in the regulation of the migration of triple-negative MDA-MB-231 cells via the EMT signaling pathway [54]. Although there is no information about the function of YWHA, but upregulation of one member of YWHAZ (also known as 14-3-3ζ) had been found associated with poor clinical outcome in breast cancer [55]. Cluster 3 lowly expressed lncRNAs, THAP9-AS1 and TUG1, were also listed in the normal samples negatively related to the blue module. Extraction of the lncRNA-mRNA interaction network displayed that the splicing factor SRSF10 bridged those two lowly expressed lncRNAs to
Figure 7: Verification of the inductive roles of identified DElncRNAs based on public datasets. The TN cancer samples were classified into three clusters (a). Survival analysis of the three clusters showed different survival probability (b). The panels of identified 97 lncRNAs act as signature predictors of survival of TN breast cancer patients (c).
mediate the coexpression network (Figure 9(b)). SRSF10, as a splicing factor, has been found playing important roles in colon and cervical cancer oncogenesis by mediating alternative splicing [56, 57]. Whether SRSF10 plays a similar role in breast cancer or not requires more experimental confirmation. The survival expectation indicators associated with coexpressing coding genes formed coexpressed axes, which could act as attractive therapeutic targets for the treatment of TN breast cancer.

4. Discussion

Although deaths from breast cancer have declined over time, its molecular mechanism remains to be revealed. Owing to
the notorious heterogeneity of breast cancers, it is meaningful to not only classify them based on clinical and morphologic characteristics but also consider intrinsic molecular signatures. Compared with other breast cancers, TN breast cancer was characterized as young onset, high malignancy, easy recurrence, and low survival rates [58]. In addition, TN breast cancer lacks ER and PR as well as HER2 receptor. There are no effective targets of endocrine therapy and targeted therapy for TN breast cancer. It is necessary to identify new molecular targets for TN breast cancer therapy. In our work, we aimed to identify key lncRNAs associated with TN, HER2, luminal A, and luminal B breast cancers. In total, 6,414 coding genes and 134 lncRNAs were found to be dysregulated in different subtypes of breast cancers. A core of 37 lncRNAs was found to be dysregulated in all the four subtypes of breast cancers. Subtypes of breast cancer special modules and lncRNA-mRNA interaction networks were obtained using WGCNA. Moreover, survival analysis of public datasets (GSE58812) verified the identified lncRNAs, which could act as indicators for triple-negative breast cancer prognosis.

Studies on lncRNAs in breast cancer are still in the early stages, and the roles of lncRNAs in breast cancer remain to be elucidated. In the notorious heterogeneity of breast cancers, there is still a core of 37 lncRNAs which dysregulated in all the four major subtypes of breast cancers. Most of the core 37 lncRNAs were downregulated compared to the normal tissues. Only five lncRNAs were upregulated in breast cancer. This indicates that the downregulated lncRNAs may be the tumor suppressor genes in normal tissues. For example, MEG3 is listed in the 32 downregulated core lncRNAs as being downregulated in various types of cancers including breast cancer. Meta-analysis demonstrated that low expression of MEG3 is associated with poor survival in cancer patients [42]. The TN-related lncRNA-mRNA coexpression network derived from the turquoise module displayed MEG3 coexpression pairs. CEP55, MELK, and KIF11 were the top three high degree DEGs in the network. CEP55 is a determinant of cell fate during perturbed mitosis in breast cancer [43]. MELK acts as a potential therapeutic target for TN breast cancer and other aggressive malignancies [59]. A recent study demonstrated that MELK can be upregulated by LINC02418 in colorectal cancer [60]. Recently, a few lncRNAs were reported to function as scaffold molecules recruiting chromatin-modifying complexes to regulate gene expression [61]. lncRNAs can also act as miRNA sponge to reverse miRNA suppression of its target genes [62]. Constructions of breast cancer-related lncRNA-mRNA coexpression networks could provide useful information for lncRNA-mRNA interaction mechanism in breast cancer progression. In this study, we found SRSF10 bridged THAP9-AS1 and TUG1 to mediate the coexpression network. As the splicing factor function of SRSF10, the low expression of THAP9-AS1 and TUG1 in cluster 3 may be related to alternative splicing.

lncRNAs may serve as biomarkers for diagnostic and prognostic purposes and also as potential therapeutic targets in cancer. Several studies have provided insights into the potential clinical implications of lncRNAs in breast cancer. For example, DSCAM-AS1 mediates tumor progression and tamoxifen resistance [63] and HOTAIR reprograms the polycomb repressive complex binding pattern in breast cancer.

Figure 9: The coexpressing coding genes of survival expectation indicators extracted from the identified lncRNA-mRNA coexpression networks. The four highly expressed lncRNAs associated with 31 coding genes to form a coexpression network (a). Top degree coding genes were painted with red color. The splicing factor SRSF10 bridged those two lowly expressed lncRNAs THAP9-AS1 and TUG1 mediated coexpression network (b). SRSF10 was painted with yellow color.
cancer [64]. The relative expression values of DElncRNAs can be used as predictors for survival expectation in TN breast cancer. Interestingly, the expressions of THAP9-AS1 and DLEU2 were relatively lower in cluster 3 and were the only five upregulated lncRNAs in breast cancer samples. There is a paradox between panel lncRNAs and individual lncRNAs prediction of survival probability. H19 and MIRLR77BHG lined in the last block which means high expression predict high survival probability. High expression of H19 and MIRLR77BHG was significant related with poor prognosis, since H19 is an estrogen-inducible gene and plays a key role in cell survival and especially in estrogen-induced cell proliferation such as MCF-7 cells [65]. The expression pattern of H19 in TN breast cancers may be different from ER-positive breast cancer. Disclosing the expression pattern of H19 will provide clues for TN breast cancer diagnosis and progression.

Considering the notorious heterogeneity of breast cancers, exact classification needs more molecular markers. In addition, more efforts are needed to detect the relationship between the expressions of these lncRNAs and breast cancer progression. Jiang and colleagues developed a reliable tool to predict tumor recurrence and the benefit of taxane chemotherapy based on an integrated mRNA-lncRNA signature in TN breast cancers [66]. Moreover, further validation in prospective clinical trials of those identified lncRNAs could provide new therapy strategy for the treatment of TN breast cancers. Based on the data from different subtypes, we could conclude that the expression of lncRNAs may help to divide different breast cancer subtypes, and provide novel knowledge about cancer diagnosis and treatment.

We should also point out that there were some limitations in this study. First, the datasets used to identify the differentially expressed DEGs and DElncRNAs were downloaded from a public database and were not validated by quantitation experiments. Second, due to the limit of the sample size, the KM curves with a p value of pairwise comparisons for the three TN breast cancer clusters were not statistically different enough. Therefore, further validation of the roles of these lncRNAs would be required in future experiments.

5. Conclusions

In conclusion, we identified 134 lncRNAs to be closely related to different subtypes of breast cancers. Based on WGCNA analysis, we constructed breast cancer-related lncRNA-mRNA coexpression networks. The constructed lncRNA-mRNA coexpression networks provided useful information for uncovering the mechanism of lncRNAs in breast cancer genesis. Bioinformatics analysis revealed that lncRNA-mRNA coexpression networks were involved in cell cycle checkpoint, carboxylic acid and organic acid transport, metabolic processes, response to hormone stimulus, Golgi vesicle transport, and vesicle organization. The survival analysis of public datasets verified the identified lncRNAs, which acted as indicators for TN breast cancer prognosis. These survival expectation indicators are associated with coexpressing coding genes to form coexpressed axes, which could act as attractive therapeutic targets for the treatment of TN breast cancer.

Data Availability

The public datasets GSE45827 and GSE58812 were downloaded from the NCBI Gene Expression Omnibus (GEO). The data used to support the finding of this study are available from corresponding author upon request.

Disclosure

X-X Li and L-J Wang performed this study as co-first authors.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors’ Contributions

Xiao-Xiao Li, Li-Juan Wang and Wen-Hai Xie analysed and interpreted the data. Jie Hou, Hong-Yang Liu, Rui Wang, and Chao Wang participated in this study. Li-Juan Wang and Wen-Hai Xie wrote this manuscript. Wen-Hai Xie conceived and designed this study. Wilfing Hai Xie supervised this study and provided financial support. All authors read and approved the final manuscript. Xiao-Xiao Li and Li-Juan Wang contributed equally to this study.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (81501316 and 31500086). We would like to thank Fatima Mecha Grigoriou et al. and Mario Campane et al. who released and shared their datasets on the GEO database.

References


[38] L. C. Liu, Y. L. Wang, P. L. Lin et al., "Long noncoding RNA HOTAIR promotes invasion of breast cancer cells through...


