



Research Article

Sparse Gene Coexpression Network Analysis Reveals EIF3J-AS1 as a Prognostic Marker for Breast Cancer

Xin Chen , Zuyuan Yang, Chao Yang, Kan Xie, Weijun Sun, and Shengli Xie

Guangdong Key Laboratory of IoT Information Technology, School of Automation, Guangdong University of Technology, Guangzhou, Guangdong, China

Correspondence should be addressed to Shengli Xie; shlxie@gdut.edu.cn

Received 5 October 2017; Revised 13 May 2018; Accepted 23 May 2018; Published 12 June 2018

Academic Editor: Vittorio Loreto

Copyright © 2018 Xin Chen 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.

Predictive and prognostic biomarkers facilitate the selection of treatment strategies that can improve the survival of patients. Accumulating evidence indicates that long noncoding RNAs (lncRNAs) play important roles in cancer progression, with diagnostic and prognostic potential. However, few prognostic lncRNAs are reported for breast cancer, and little is known about their functions that contribute to cancer pathogenesis. In this paper, we used weighted correlation network analysis (WGCNA) to construct networks containing noncoding and protein-coding genes based on their expression in 1097 breast cancer patients. The differentially expressed genes were significantly overlapped with gene modules regulating cell cycle and cell adhesion. The cell cycle-related lncRNAs were consistently downregulated in breast cancer. One lncRNA, EIF3J-AS1, is significantly associated with clinicopathological characteristics, including tumor size, lymph node metastasis, estrogen receptor (ER), and progesterone receptor (PR) status. Kaplan-Meier survival analysis revealed that EIF3J-AS1, a downregulated lncRNA in breast tumor, is a potential prognostic marker for breast cancer. EIF3J-AS1 may function in an estrogen-independent manner and could be inhibited by the compound FDI-6. Therefore, integrating sparse gene coexpression network and clinicopathological features can accelerate identification and functional characterization of novel prognostic lncRNAs in breast cancer.

1. Introduction

Breast cancer is a highly heterogeneous disease, which is commonly divided into five subtypes, basal-like, HER2, luminal A, luminal B, and normal-like, using histopathological status of either estrogen receptor (ER), human epidermal growth factor receptor (HER2), or a gene expression-based classifier (PAM50) [1]. The use of the mRNA-based prognostic marker, comprised of multiple differentially expressed genes, has been supported by clinical guidelines, which assists the clinical treatment of breast cancer by integrating clinicopathological factors [2, 3].

Gene coexpression networks (GCNs) have been widely used in the studies of cancer for the identification of prognostic signature [4]. GCN from transcriptomic profiles facilitates elucidating gene interactions and exploring regulatory mechanisms [5]. For each gene expression profile, it contains expressions of tens of thousands of genes in detected samples. The coexpression network is constructed based on the

pairwise gene correlation matrix. In the network, each node represents one gene, while each edge represents a pair of genes with highly correlated expression pattern. The large coexpression network is not easy to interpret because of its high dimensionality. Besides, there are few master regulatory genes which basically control the state of the network [6]. It is promising to decompose the sparse network into smaller components [7, 8], which are also referred to as gene modules. GCN is quite sparse with only a few “hub” genes densely connected to each other. For years, the scale-free network model has been supported for biological networks [9]. For example, sparse signal transduction networks follow the scale-free properties. In *E. coli* and *S. cerevisiae*, the degree distribution is $P(k) = k^{-\gamma}$, $r \approx 2$ [10], which implies that majority of the molecules are involved in few interactions and minority of them have many interactions [9, 11].

Long noncoding RNA (lncRNA), with length longer than 200 nt, has been regarded as the dark matter of the genome for decades. However, with the development and application

of next-generation sequencing (NGS), lncRNAs have been found to have a myriad of molecular functions in diseases including cancers [12]. LncRNAs such as HOTAIR and MALAT1 had been reported as a prognostic marker for breast cancer [13, 14]. Differential analysis and coexpression network has been successfully applied to identify prognostic lncRNAs in breast cancer [15]. Therefore, in this study, weighted correlation network analysis (WGCNA) was used to identify modules of highly correlated genes. Then, we focus on those modules significantly enriched by differentially expressed genes, which play important roles in breast cancer. The deregulated lncRNAs were identified by integrating clinicopathological characteristics and further investigated to determine their prognostic potential in biologically meaningful modules.

2. Materials and Methods

2.1. Data Preprocessing. We downloaded the transcriptomic expression profiles of TCGA breast cancer from TCGA (<https://tcga-data.nci.nih.gov/>). The dataset contains primary tumor and adjacent normal samples from primary and metastatic samples. Before constructing the gene coexpression network (GCN), we first filtered out genes with FPKM < 1 in more than half of all the samples. The expression profiling contains the expression value of 12,488 genes in 1202 samples from 1097 patients. In the TCGA BRCA cohort, the number of normal samples is ~10% of the tumor samples.

2.2. Weighted Correlation Network Analysis (WGCNA). WGCNA uses adjacency to measure the similarity between two genes in the network, which is calculated based on the correlation coefficient. In the network, similarity s_{ij} is defined as the absolute correlation coefficient between the profiles of genes i and j : $s_{ij} = |\text{cor}(x_i, x_j)|$. For a traditional unweighted network, adjacency a_{ij} is defined as follows:

$$a_{ij} = \begin{cases} 1 & \text{if } s_{ij} \geq \tau, \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where τ is the hard (fixed) threshold parameter to weigh the edges. However, the unweighted networks do not accord with the continuous characteristics of the coexpression information, which will lead to loss of information. Therefore, weighted networks fit the nature of the continuous coexpression. The corresponding adjacency can be defined as $a_{ij} = s_{ij}^\beta$, where $\beta \geq 1$. The threshold $\beta = 10$ is chosen based on the approximate scale-free topology criterion of the coexpression network. The adjacency a_{ij} is further transformed to a topological overlap matrix (TOM), which is a measure to evaluate how strongly two genes are correlated to the same set of neighboring genes. Then, 1-TOM is used as a dissimilarity measure for hierarchical clustering. In the clustering dendrogram, each branch represents one module, which compromises of genes with highly similar expression pattern. In this way, modules can be defined based on different branch-cutting methods,

for example, the dynamic tree cut methods [16]. For more details, please refer to [17].

2.3. Identifying Differentially Expressed Genes. To identify the differential genes, we only chose the matched tumor and normal samples from 112 patients, in order to avoid the bias caused by unbalanced sample size in the TCGA BRCA cohort. The raw read counts were downloaded from TCGA (<https://tcga-data.nci.nih.gov/>). R package DESeq [18] was used to identify differentially expressed genes (DEGs) in breast cancer. The significance P value was adjusted by Benjamini–Hochberg FDR. The cutoff of significant P value was 0.05.

2.4. Gene Ontology Annotation and Enrichment Analysis. We used the online web tool DAVID [19] v6.8 for functional enrichment analysis. Gene Ontology (GO) defines concepts used to describe gene functions along three aspects: biological process (BP), molecular function (MF), and cellular component (CC). When performing functional enrichment analysis on the genes in each module, we considered GO terms of BP branch. EASE score is a modified Fisher's exact P value to evaluate whether the interested genes are significantly enriched in a specific gene function, which contains a lot of genes to achieve this function. The smaller the P value is, the more enriched the interested genes are. The Benjamini–Hochberg false discovery rate (BH-FDR) was used for correcting multiple comparisons. The enrichment threshold of P value was set to 0.01.

2.5. Associating lncRNAs with Clinicopathological Characteristics. The patients were divided into high and low groups, according to the median expression level of candidate lncRNAs. The patients with the lncRNA expression level larger than its median expression value were assigned into the high group and vice versa. Chi-squared test was used to associate gene expression with clinicopathological features.

2.6. Statistic Method for Cross-Dataset Validation and Survival Analysis. The expression difference of candidate lncRNAs is compared by Mann–Whitney U test in cancer versus normal samples. Kaplan–Meier and Cox regression analyses were utilized to assess the prognostic significance of lncRNAs. The statistical analysis was performed using R.

3. Results

3.1. Gene Modules Identified Using WGCNA. WGCNA was used to construct the gene coexpression network (GCN) based on the TCGA BRCA dataset. Only genes with appreciable expression levels (FPKM > 1) in more than half of the samples were considered for further analysis. The gene expression profiles, comprising of 12,488 genes in 1202 samples, were \log_2 transformed and subjected to WGCNA. As shown in Figure 1(a), power 10 was chosen as the soft threshold to identify coexpressed gene modules (for details, see Section 2.2). 16 gene modules were identified, and module names were color-coded including blue, brown, green, grey, red, turquoise, yellow, and black. As the “grey” module was reserved for unassigned genes, we further

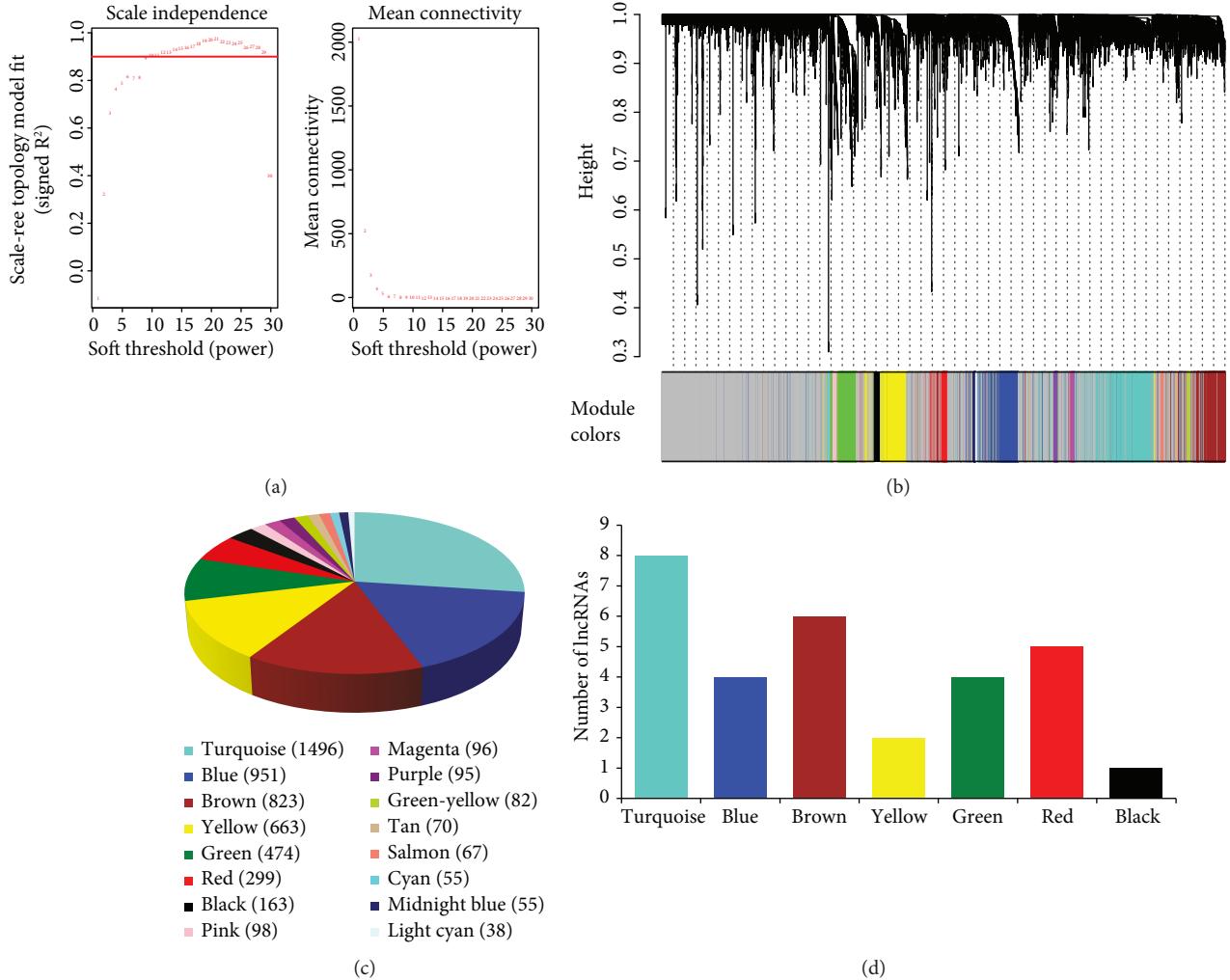


FIGURE 1: Gene modules detected using the weighted correlation network analysis (WGCNA). (a) Scale-free topology index and mean connectivity were used to determine the soft threshold. (b) Clustering dendrogram of genes. The dissimilarity of genes is based on the topological overlap. The genes are assigned to different modules and modules are named using different colors. (c) Number of genes in each module identified from WGCNA. The numbers in the parentheses represent the number of genes in each module. (d) Number of lncRNAs in the modules containing more than 100 genes.

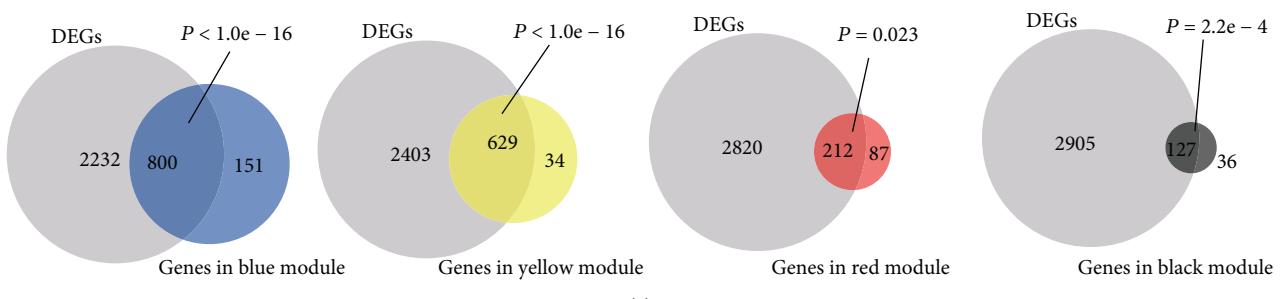
focused on the other modules except the grey module. The clustering dendrogram of genes is shown in Figure 1(b).

As shown in Figure 1(c), the turquoise, blue, and brown modules were the top 3 modules containing the highest number of genes. We further checked the number of lncRNAs in each module. There are 8 lncRNAs in the turquoise module. And in the brown module, there are 6 lncRNAs, while 5 in the red module (Figure 1(d)).

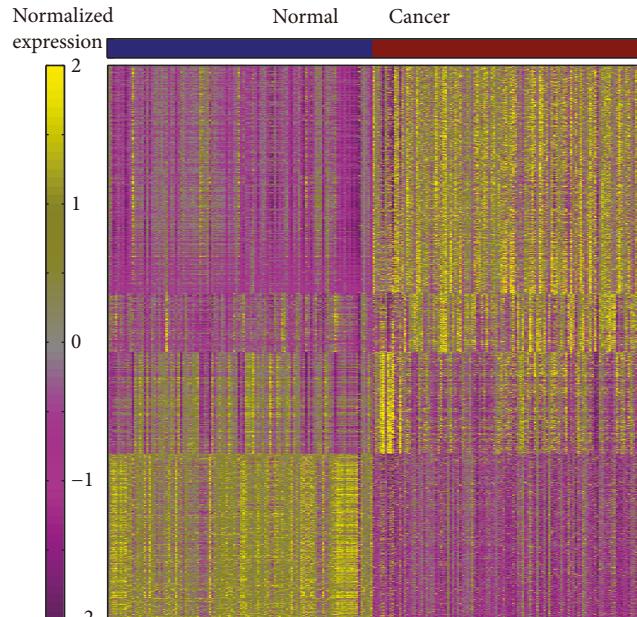
3.2. Modules Significantly Enriched in DEGs of Breast Cancer. The dysregulation of important genes (protein-coding and noncoding) plays important roles in tumorigenesis. We would like to know how many genes are differentially expressed in breast cancer. DESeq [18] was used to determine the DEGs between cancer and normal breast tissue, from the TCGA BRCA dataset. In total, 3032 DEGs were identified. Among these DEGs, an lncRNA Xist has experimentally supported associations with human breast cancer [20, 21].

We further checked the dysregulated lncRNAs in the coexpressed gene modules. For each module, we calculated the number of DEGs. The hypergeometric test was used to test if the DEGs are significantly enriched in the module. From the P values, we found that the blue, yellow, red, and black modules are significantly overlapped with DEGs ($P < 0.05$, Figure 2(a)). For genes in these four modules, the expression heat map is shown in Figure 2(b). There are more upregulated DEGs in cancer tissues than those in downregulated ones. In our analysis, only 30 lncRNAs were included, excluding those lncRNAs with low expression levels [22]. Among these lncRNAs in the four modules, 5 lncRNAs are upregulated in cancer relative to normal tissue and 9 lncRNAs are downregulated (Figure 2(c)).

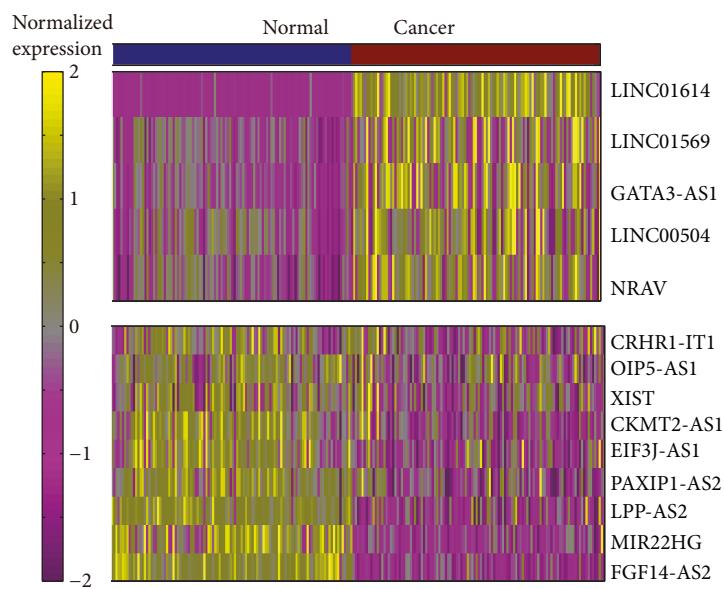
3.3. Genes in Modules Mainly Involved in Cell Cycle and Cell Adhesion. The gene modules in the network are often enriched with specific functions [23], which enable its



(a)



(b)



(c)

FIGURE 2: Differential genes of breast cancer in the modules. (a) Overlapped genes of each module and DEGs. The genes in the blue, yellow, red, and black modules are significantly overlapped with DEGs. (b) The expression heat map of DEGs in the blue, yellow, red, and black modules. (c) Expression levels of 5 upregulated lncRNAs in cancer relative to normal tissue (up) and 9 downregulated lncRNAs (down).

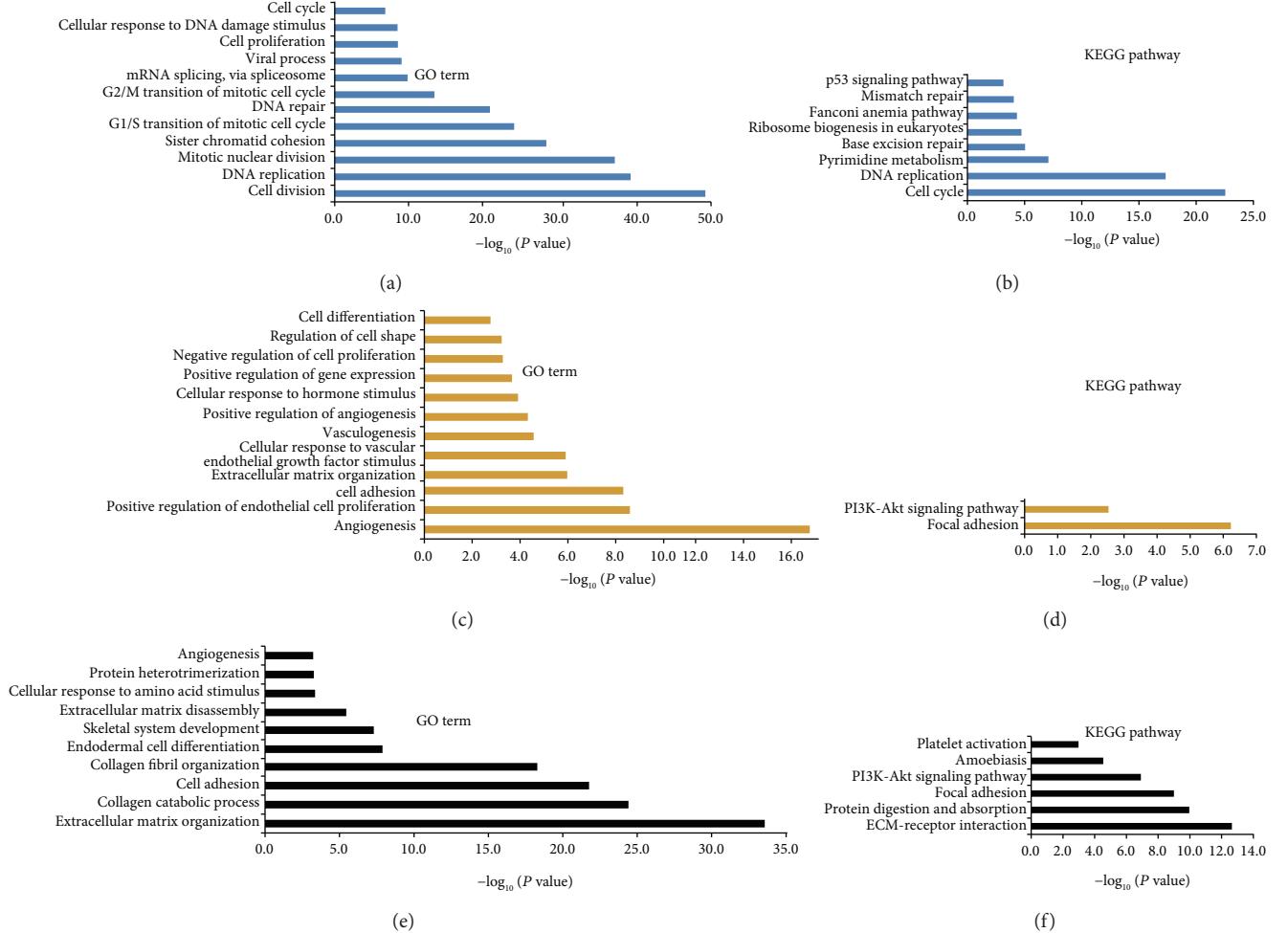


FIGURE 3: Biological function of DEGs in each module. (a) Gene Ontology and (b) KEGG pathway enrichment analysis for the blue module. (c) Gene Ontology and (d) KEGG pathway enrichment analysis for the yellow module. (e) Gene Ontology and (f) KEGG pathway enrichment analysis for the black module.

application in generating the hypothesis of biological significance. Besides, the “coexpression” approach has been used to understand the lncRNA function [24]. To test whether the identified modules are biologically meaningful, information from Gene Ontology (GO) and KEGG pathway was used for function enrichment analysis. GO terms with BH adjusted P value of <0.01 are regarded as significant.

Among those four modules, the red module has no significantly enriched GO function and KEGG pathway. The genes in the blue module are enriched in the GO terms “cell division,” “DNA replication,” and “cell cycle” (Figure 3(a)). As for the KEGG pathway, “cell cycle” and “DNA replication” were the top enriched pathways (Figure 3(b)). For genes in the yellow module, they are enriched in the GO terms “signal transduction,” “cell adhesion,” and “angiogenesis” (Figure 3(c)). As for the KEGG pathway, “focal adhesion” and “signalling pathway” were highly enriched (Figure 3(d)). Similar to the yellow module, genes in the black module are enriched in the GO terms “cell adhesion” and “collagen catabolic process” (Figure 3(e)). The KEGG pathways “focal adhesion” and

“PI3K-Akt signalling pathway” were highly enriched by genes in the black module (Figure 3(f)). The function similarity of the yellow and black modules can be known from the gene dendrogram from Figure 1(b). We also performed the similarity comparisons of all the 16 modules (Figure S1). From the clustering tree, the black module and yellow module are in the same branch, which further supports their similarity in gene functions.

3.4. Clinical Significance of Deregulated LncRNAs in Breast Cancer. As the results showed (Figure 3), the genes in the blue module are cell cycle-related while genes in the black and yellow modules are related with cell adhesion and signal transduction. We further explored the clinical significance of the deregulated lncRNAs in these modules, according to their relationship with clinicopathological characteristics.

We divided the patients into high and low groups, according to the median expression level of candidate lncRNAs. As shown in Table 1, higher EIF3J-AS1 expression group has more older patients ($P = 0.016$) and more lymph node metastasis ($P = 0.038$). Low EIF3J-AS1 expression

TABLE 1: The relationship between EIF3J-AS1 and clinicopathological features in the TCGA cohort.

Clinicopathological characteristics	Low expression	High expression	χ^2	P value
Age				
≤ 60	320	280		
> 60	224	265	5.8	0.016
Tumor size				
$< 2 \text{ cm}$	119	157		
$2\text{--}5 \text{ cm}$	344	293	10.29	0.006
$> 5 \text{ cm}$	80	93		
Lymph node metastasis				
N0	274	239		
N1	161	203		
N2	65	54	8.41	0.038
N3	39	35		
Distant metastasis				
No	469	435	3.68e	
Yes	11	11	-29	1
Clinical stage				
I-II	406	398		
III-IV	130	134	0.08	0.78
Estrogen receptor				
Negative	198	42		
Positive	327	472	125.9	<2.2e
Progesterone receptor				
Negative	245	99		
Positive	278	414	87.4	<2.2e
			-16	

group has larger tumor size ($P = 0.006$) and more advanced clinical stage and PR-negative patients ($P = 0.002$).

For LPP-AS2 (Table 2), low LPP-AS2 expression group has larger tumor size ($P = 0.008$) and more lymph node metastasis ($P = 6.20e - 07$), advanced clinical stage ($P = 7.86e - 06$), and PR-negative patients ($P = 0.002$).

Another lncRNA in the blue module, CKMT2-AS1, has higher expression in older patients ($P = 0.016$). Lower CKMT2-AS1 expression group has larger tumor size ($P = 0.001$) and more ER-negative ($P < 2.20e - 16$) and PR-negative ($P = 1.00e - 15$) patients (Table S1). For lncRNAs in the yellow and black modules, the low MIR22HG expression group has larger tumor size ($P = 8.40e - 05$) and more ER-negative patients ($P = 0.036$) than the high MIR22HG expression group (Table S2). Similarly, the low FGF14-AS2 group has more lymph node metastasis ($P = 0.0006$) and more advanced clinical stage ($P = 0.029$) and PR-negative patients ($P = 0.029$) than the high FGF14-AS2 expression group (Table S3), which is consistent with the recent study [25]. However, the low LINC01614 expression group has more ER-negative ($P = 1.90e - 07$) and PR-negative patients ($P = 2.30e - 6$) than its high expression group (Table S4). For other clinicopathological characteristics such as tumor size, lymph node metastasis, and clinical

TABLE 2: The relationship between LPP-AS2 and clinicopathological features in the TCGA cohort.

Clinicopathological characteristics	Low expression	High expression	χ^2	P value
Age				
≤ 60	301	299		
> 60	243	266	0.56	0.46
Tumor size				
$< 2 \text{ cm}$	117	159		
$2\text{--}5 \text{ cm}$	341	296	9.62	0.008
$> 5 \text{ cm}$	85	88		
Lymph node metastasis				
N0	230	283		
N1	177	187		
N2	87	32	31.65	6.2e - 07
N3	40	34		
Distant metastasis				
No	466	438		
Yes	16	6	3.06	0.08
Clinical stage				
I-II	370	434		
III-IV	164	100	19.97	7.9e - 06
Estrogen receptor				
Negative	124	116		
Positive	387	412	0.65	0.42
Progesterone receptor				
Negative	193	151		
Positive	318	374	9.07	0.002

stage, there is no significant difference between the low and high LINC01614 expression groups.

The blue module is enriched in cell cycle, which plays important roles in cancer. Moreover, considering EIF3J-AS1 and LPP-AS2 are significantly correlated with most of the clinicopathological characteristics, we further validate their expression and prognostic potential in breast cancer.

3.5. EIF3J-AS1 and LPP-AS2 Are Candidate Biomarkers in Breast Cancer. As shown in Figure 2(c), EIF3J-AS1, LPP-AS2, and CKMT2-AS1 are lowly expressed in tumor samples, compared with their corresponding matched normal breast tissues from the TCGA cohort. It suggests that these three lncRNAs are candidate biomarkers in breast cancer. In Figure 4, it demonstrated that the expressions of EIF3J-AS1 (Figure 4(a)), LPP-AS2 (Figure 4(b)), and CKMT2-AS1 (Figure 4(c)) are indeed significantly lower in tumor.

The dataset GSE31448 contains cancer and normal mammary samples from 353 patients. Another dataset GSE58135 also contains expression profiles of 140 normal and tumor breast tissues. From Figures 4(d) and 4(e) and Figure S2A-B, EIF3J-AS1 is lowly expressed in breast tumor. LPP-AS2 is also lowly expressed in tumor tissues, compared to normal breast tissues (Figures 4(f) and 4(g)).

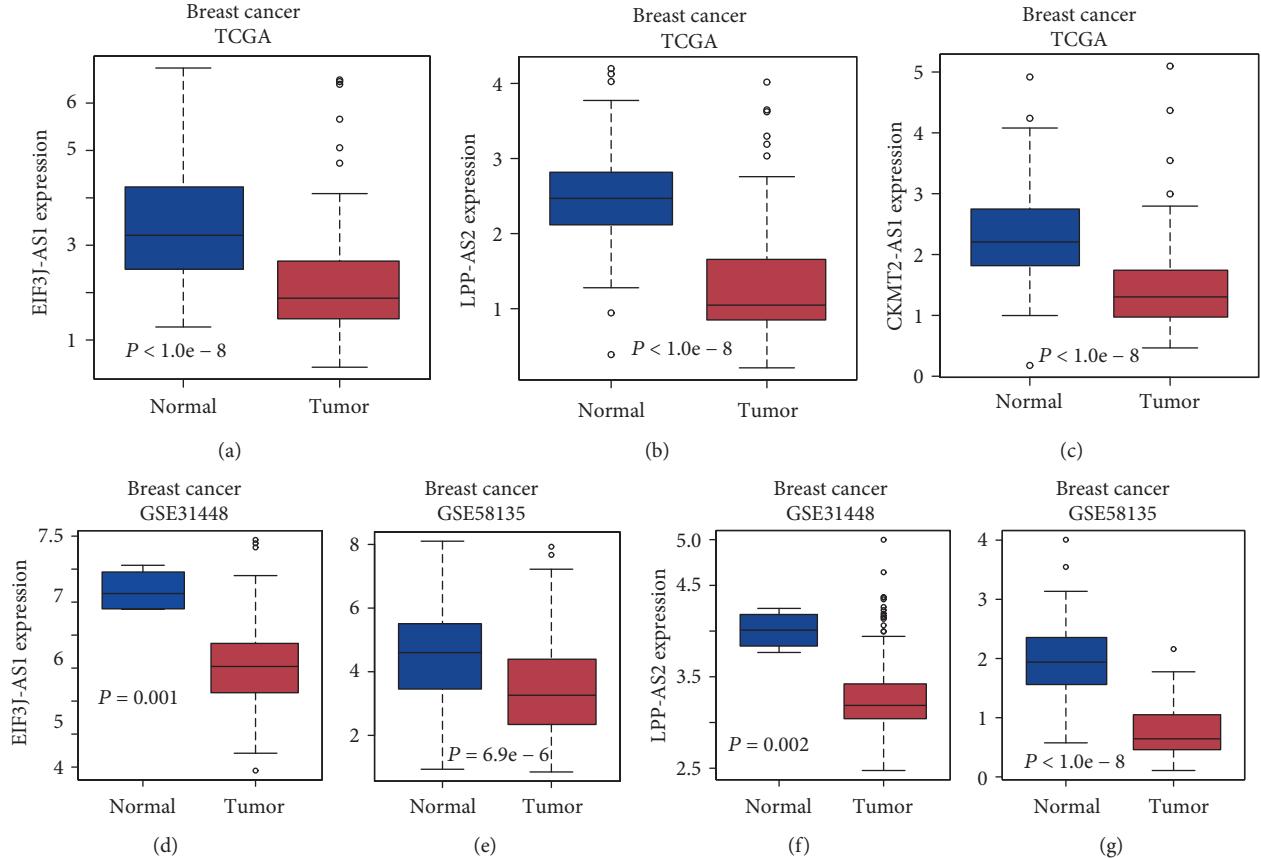


FIGURE 4: Cross-dataset expression of cell cycle-related lncRNAs. (a) EIF3J-AS1, (b) LPP-AS2, and (c) CKMT2-AS1 are downregulated in tumors versus normal tissues in the TCGA breast cancer dataset. (d) EIF3J-AS1 is downregulated in tumor versus normal tissues in the GSE31448 and (e) GSE58135 datasets. (f) LPP-AS2 is downregulated in tumor versus normal tissues in the GSE31448 and (g) GSE58135 datasets.

3.6. The LncRNA EIF3J-AS1 Is a Potential Prognostic Marker in Breast Cancer. As validated in other datasets other than TCGA, EIF3J-AS1 and LPP-AS2, lncRNAs in the blue module, are candidate biomarkers in breast cancer. We further determine their prognostic potency. Using an online survival analysis tool for breast cancer [26], we performed overall survival (OS) and relapse-free survival (RFS) analysis for both EIF3J-AS1 and LPP-AS2. Patients were divided into high- and low-expressed groups, using its median expression value as the cutoff. As shown in Figure 5(a), patients with high expression of EIF3J-AS1 have better OS ($P = 0.0029$). However, the low and high LPP-AS2 expression groups do not exhibit significant difference in OS (Figure 5(b)). Further, we explored the RFS for both lncRNAs. High expression of EIF3J-AS1 suggests better RFS (Figure 5(c), $P < 1.0e - 16$). In contrast to OS, there is significant RFS difference between high and low LPP-AS2 expression groups (Figure 5(d), $P = 0.003$).

3.7. Coexpressed Genes of EIF3J-AS1 Participate in G2/M Phase of Cell Cycle. To dissect the possible mechanism of lncRNAs in breast cancer, we further constructed the subnetwork of the three lncRNAs and their coexpressed genes in the blue module. As shown in the network (Figure 6(a)), EIF3J-AS1 and LPP-AS2 shared coexpressed genes. Using TANRIC

[27], we have identified 812 of these genes showing strong correlations with lncRNA EIF3J-AS1 across the TCGA breast cancer dataset. 54 genes of them have also interacted with EIF3J-AS1 in the GCN of the blue module. As our results showed (Figure 3(a)), the genes in the blue module participate in cell cycle. Therefore, we further found that seven genes (PTTG1, CDC20, BUB1, TTK, CDC45, PLK1, and CCNE1) are known cell cycle genes (NanoString Technologies) and are DEGs in the TCGA cohort. The expression of these seven cell cycle genes is highly correlated with that of EIF3J-AS1.

To elaborate the role of EIF3J-AS1 in cell cycle, we mapped its coexpressed genes to the KEGG pathway “cell cycle.” CCNE1 and CDC45 participate in the G1 and S phases of cell cycle, while the other five genes are involved in the G2/M phase (Figure 6(b)). We speculate that EIF3J-AS1 regulates the later phase of cell cycle, based on the location of its coexpressed genes in the pathway “cell cycle.” Cell cycle assays revealed significantly higher proportions of cells in the G2/M phase, suggesting a cell cycle arrest at the G2/M phase by CKI in MCF7 cells [28]. From Figure S3, we found that downregulated genes after 5FU (Figure S3A) or CKI treatment (Figure S3B-C) were significantly overlapped with DEGs in the blue module. This also supports our conclusion that coexpressed genes of EIF3J-AS1 participate in the G2/M phase of cell cycle.

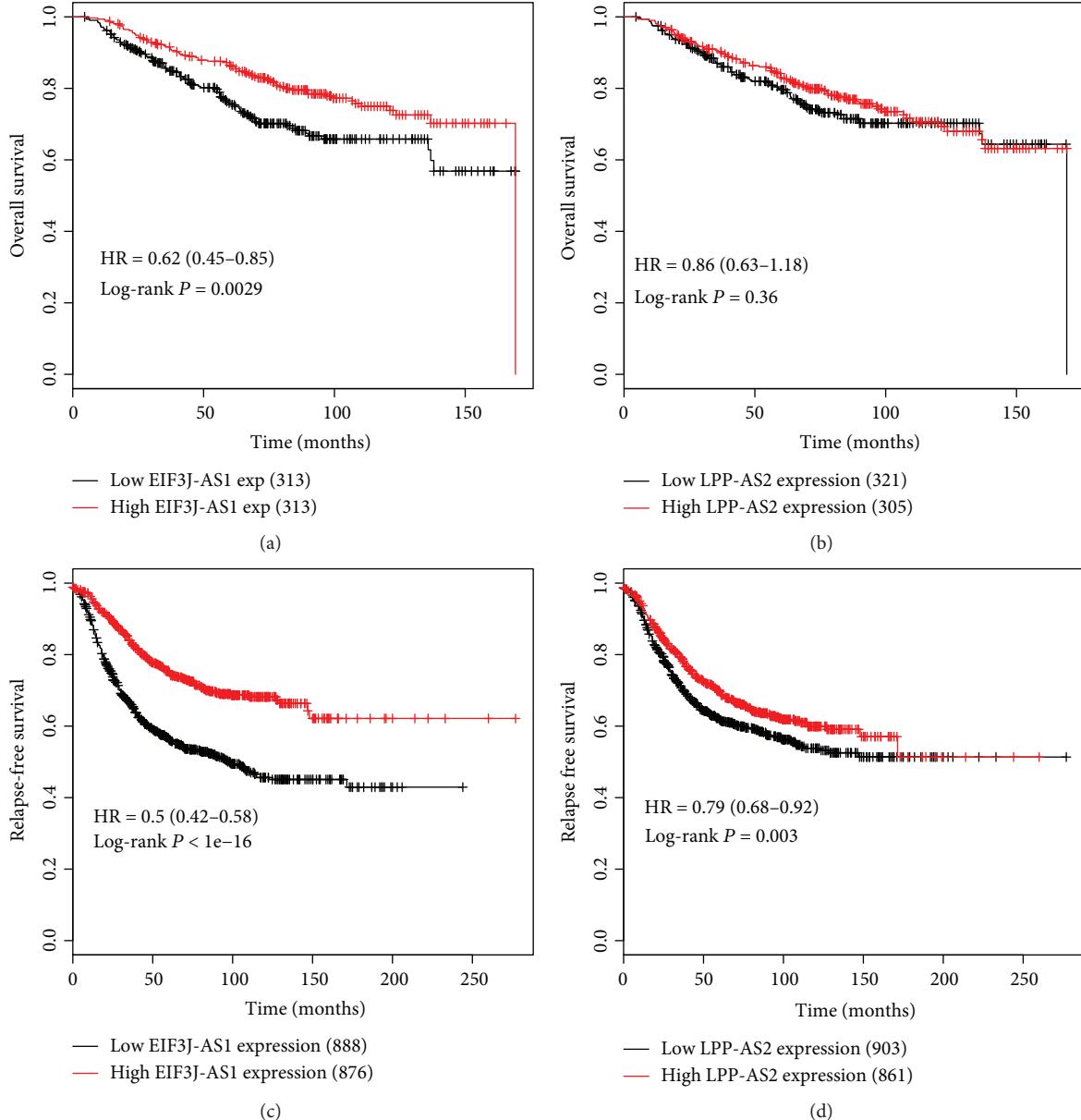
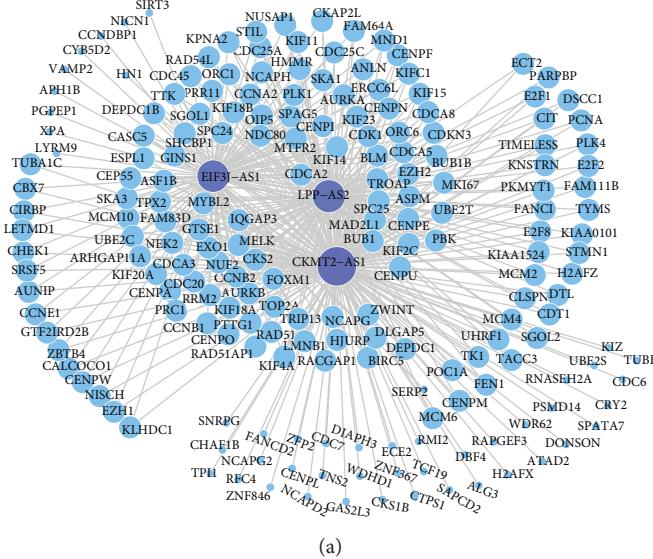


FIGURE 5: Prognostic potential of EIF3J-AS1 and LPP-AS2. (a) According to the Kaplan–Meier plot, patients with high EIF3J-AS1 expression have better overall survival. (b) Patients with high LPP-AS2 expression do not have different overall survival, compared to patients with low LPP-AS2 expression. (c) Patients with high EIF3J-AS1 expression have better relapse-free survival. (d) Patients with high LPP-AS2 expression have better relapse-free survival.

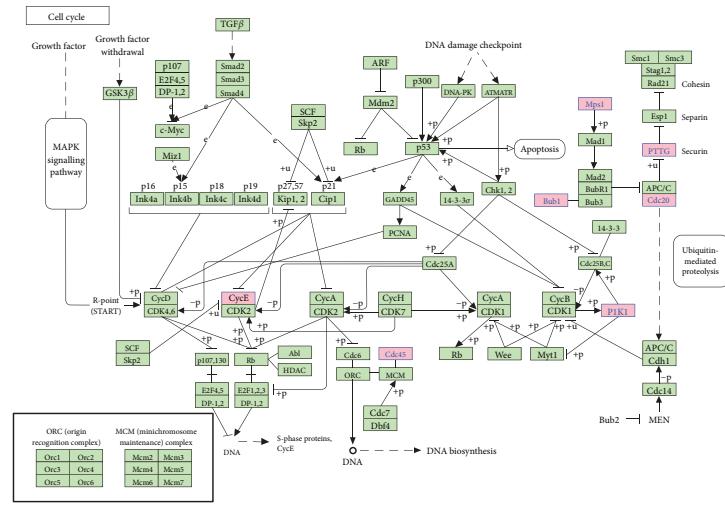
3.8. EIF3J-AS1 May Function in an Estrogen-Independent Manner and Could Be Inhibited by FDI-6. As shown in Figure S2E, EIF3J-AS1 is significantly differentially expressed between ER+ and TNBC patients. Estrogen is the most important regulator of breast cancer. We next check EIF3J-AS1 expression after E2 treatment based on a public dataset (accession ID: GSE62789). Figure 7(a) shows that the expression of EIF3J-AS1 decreased after E2 treatment at early time points. At later time points, its expression level increased gradually. siRNA experiments (Figure 7(b)) of ERα (accession ID: GSE53532) demonstrated that EIF3J-AS1 expression increased after siERα. From E2 treatment

and siERα experiments, EIF3J-AS1 may function in an estrogen-independent manner.

FOXM1 and CCNB1 are coexpressed with EIF3J-AS1 and are included in the blue module. FDI-6 was used as an inhibitor of FOXM1, according to a public dataset (GSE58626). EIF3J-AS1 immediately decreased with FDI-6 treatment: the fold change is around 1.5 (Figure 7(c)). The expression reduction is also expected for its coexpressed genes CCNB1 (Figure 7(d)). Therefore, we speculate that FDI-6 may be a candidate compound that can inhibit EIF3J-AS1 expression, which provides clues for further functional assay on EIF3J-AS1.



(a)



(b)

FIGURE 6: Coexpression network of EIF3J-AS1, LPP-AS2, and CKMT2-AS1. (a) The coexpression network was constructed for differentially expressed lncRNAs and their coexpressed genes in the blue module. Node size is proportional to the number of coexpressed genes. (b) The pink background genes are the seven coexpressed genes of EIF3J-AS1 in the KEGG pathway “cell cycle.”

4. Discussion

LncRNAs have been reported as key players of many important signalling pathways in cancer, including p53 pathway, hypoxia signalling and epithelial-mesenchymal transition (EMT), telomere maintenance, and hormone receptor signalling [12]. The expression of lncRNAs is reported to be specific to tissue and cancer types, which enables lncRNAs as favorable candidate biomarkers for cancer. For example, lncRNA SChLAP1 is cancer- and prostate-specific expressed and is a candidate prognostic marker [29]. Coexpression of lncRNA and protein-coding genes has been utilized to predict the function of uncharacterized lncRNAs [30]. Due to the inherent sparsity of the gene coexpression network, WGCNA was applied to identify the highly connected components (gene modules) from the network. Therefore, the

functions of lncRNAs can be predicted based on their coexpressed protein-coding genes in the module.

Among the deregulated lncRNAs, antisense lncRNAs are a class of long noncoding transcripts from the antisense strand of protein-coding genes. They can function as positive or negative regulators of its paired genes [31]. In the TCGA breast cancer cohort, the expression level of LPP-AS2 is positively correlated with its protein-coding gene LPP ($r = 0.52$). In colorectal cancer, LPP-AS2 has been reported to be repressed by MYC, which is a proto-oncogene-regulating cell proliferation through cell cycle [32]. LPP-AS2 and EIF3J-AS1 are significantly associated with clinicopathological characteristics, such as tumor size, lymph node metastasis, and PR status, which highlight their potency in clinical application.

The seven genes highly correlated with EIF3J-AS1 are deregulated in cancer, and some are reported as prognostic

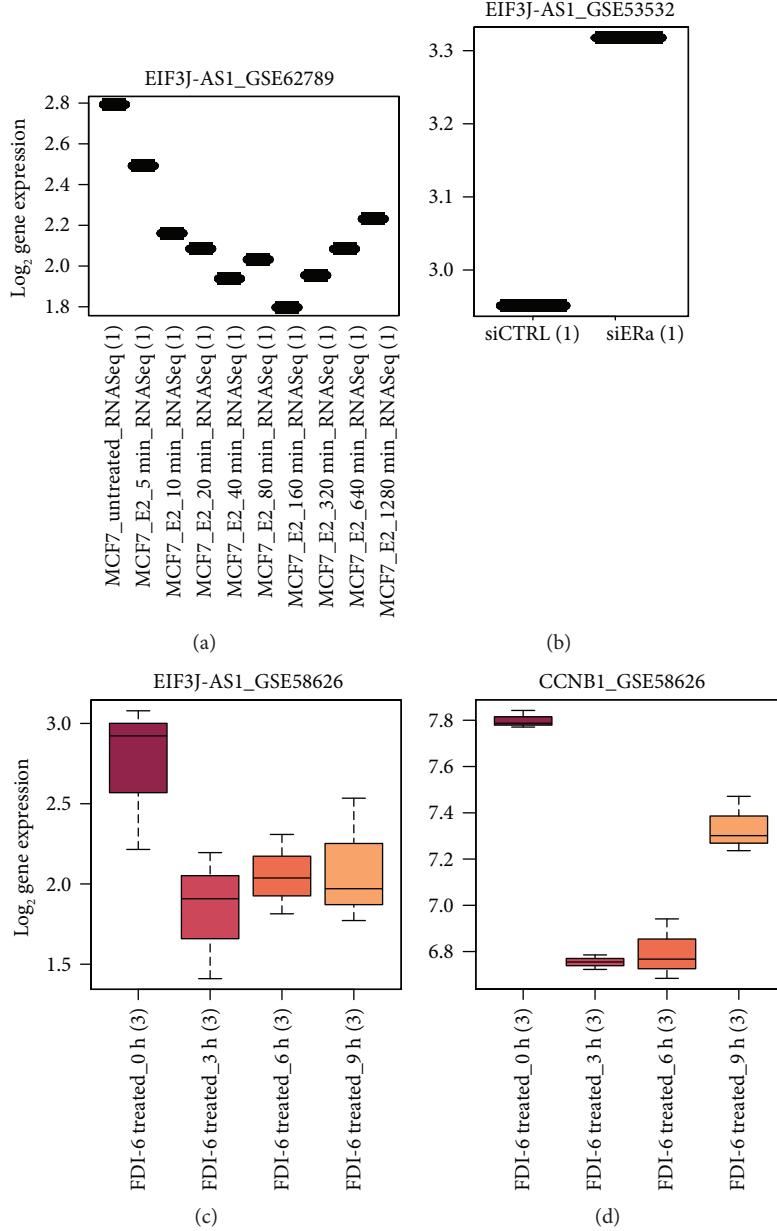


FIGURE 7: Functional study of EIF3J-AS1 via *in vitro* experiments from public datasets. (a) EIF3J-AS1 was downregulated after E2 treatment in breast cancer cell line MCF7. (b) EIF3J-AS1 was upregulated after siERa experiments in MCF7. (c) The expression of EIF3J-AS1 decreased after FDI-6 treatment. (d) CCNB1 is coexpressed with EIF3J-AS1. The expression of CCNB1 decreased after FDI-6 treatment.

markers of breast cancer. According to the results from survival analysis, high CDC20 expression and high PTTG1 indicated aggressive pathological course of breast cancer, particularly TNBC [33]. BUB1 expression is correlated with a poor clinical prognosis in patients with breast cancer [34]. High expression of BUB1 was associated with better OS of low-grade breast cancers [35]. TTK is upregulated in several cancers, especially in TNBC. TTK was also associated with aggressive subgroups, has poor survival, and is a therapeutic target [36]. Moreover, inhibiting TTK has been proposed as a novel strategy for cancer treatment, including TNBC [37]. PLK1 regulates the phosphorylation of RAD51, which promotes the genome stability in breast cancer [38]. PLK1 is

hopefully a target gene in ER-positive breast cancer patients that have acquired resistance to estrogen deprivation therapy [39]. Inhibition of PLK1 is a promising therapeutic approach for patients suffering triple-negative breast cancer (TNBC) [40]. The expression of CCNE1, one cell growth-related gene, has been reduced by the lncRNA LINC00152 via knockdown experiments [41]. Based on the *in vitro* experiments from public datasets, we speculate that EIF3J-AS1 may function in an estrogen-independent manner. EIF3J-AS1 could be inhibited by the compound FDI-6. Therefore, experiments like overexpression and RNA interference (RNAi) of EIF3J-AS1 are needed to further elaborate its regulation of cell cycle via its target genes.

This study constructed and analyzed sparse gene coexpression network based on transcriptomic profiles of the TCGA breast cancer cohort. It identified a prognostic lncRNA that participates in cell cycle process via its coexpressed genes.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was financially supported in part by grants from the National Natural Science Foundation of China (Grant nos. 61603099, 61722304, and 61773127) and the Science and Technology Plan Project of Guangdong (Grant nos. 2014B090907010, 2015B010131014, and 2017B010125002).

Supplementary Materials

Supplementary 1. Figure S1: The similarity of all the 16 gene modules identified by WGCNA.

Supplementary 2. Figure S2: EIF3J-AS1 expression in public datasets. EIF3J-AS1 is downregulated in tumor in (A) GSE52194 and (B) GSE71651. EIF3J-AS1 is also downregulated in tumor versus other cancer types in (C) GSE52194 and (D) GSE71651. (E) EIF3J-AS1 is differentially expressed in ER+ and TNBC patients according to one public dataset in ArrayExpress (E-MTAB-4993).

Supplementary 3. Figure S3: Overlap of 5FU/CKI downregulated genes and DEGs in the blue module. (A) Overlap of 5FU downregulated genes and DEGs in the blue module. (B) Overlap of 2 mg CKI downregulated genes and DEGs in the blue module. (C) Overlap of 1 mg CKI downregulated genes and DEGs in the blue module.

Supplementary 4. Table S1: The relationship between CKMT2-AS1 and clinicopathological characteristics in the TCGA cohort.

Supplementary 5. Table S2: The relationship between MIR22HG and clinicopathological characteristics in the TCGA cohort.

Supplementary 6. Table S3: The relationship between FGF14-AS2 and clinicopathological characteristics in the TCGA cohort.

Supplementary 7. Table S4: The relationship between LINC01614 and clinicopathological characteristics in the TCGA cohort.

References

- [1] C. M. Perou, T. Sørlie, M. B. Eisen et al., “Molecular portraits of human breast tumours,” *Nature*, vol. 406, no. 6797, pp. 747–752, 2000.
- [2] E. Senkus, S. Kyriakides, S. Ohno et al., “Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up,” *Annals of Oncology*, vol. 26, Supplement 5, pp. v8–v30, 2015.
- [3] A. S. Coates, E. P. Winer, A. Goldhirsch et al., “Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the primary therapy of early breast cancer 2015,” *Annals of Oncology*, vol. 26, no. 8, pp. 1533–1546, 2015.
- [4] C. Clarke, S. F. Madden, P. Doolan et al., “Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis,” *Carcinogenesis*, vol. 34, no. 10, pp. 2300–2308, 2013.
- [5] L. Zhu, Y. Ding, C. Y. Chen et al., “MetaDCN: meta-analysis framework for differential co-expression network detection with an application in breast cancer,” *Bioinformatics*, vol. 33, no. 8, pp. 1121–1129, 2017.
- [6] M. Nazarieh, A. Wiese, T. Will, M. Hamed, and V. Helms, “Identification of key player genes in gene regulatory networks,” *BMC Systems Biology*, vol. 10, no. 1, p. 88, 2016.
- [7] Z. Yang, Y. Xiang, K. Xie, and Y. Lai, “Adaptive method for nonsmooth nonnegative matrix factorization,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 4, pp. 948–960, 2017.
- [8] Z. Yang, Y. Zhang, Y. Xiang, W. Yan, and S. Xie, “Non-negative matrix factorization with dual constraints for image clustering,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–10, 2018.
- [9] A. H. Tong, G. Lesage, G. D. Bader et al., “Global mapping of the yeast genetic interaction network,” *Science*, vol. 303, no. 5659, pp. 808–813, 2004.
- [10] G. A. Pavlopoulos, M. Secrier, C. N. Moschopoulos et al., “Using graph theory to analyze biological networks,” *BioData Mining*, vol. 4, no. 1, 2011.
- [11] N. Guelzim, S. Bottani, P. Bourgine, and F. Kepes, “Topological and causal structure of the yeast transcriptional regulatory network,” *Nature Genetics*, vol. 31, no. 1, pp. 60–63, 2002.
- [12] J. R. Evans, F. Y. Feng, and A. M. Chinnaiyan, “The bright side of dark matter: lncRNAs in cancer,” *The Journal of Clinical Investigation*, vol. 126, no. 8, pp. 2775–2782, 2016.
- [13] L. Lu, G. Zhu, C. Zhang et al., “Association of large non-coding RNA HOTAIR expression and its downstream intergenic CpG island methylation with survival in breast cancer,” *Breast Cancer Research and Treatment*, vol. 136, no. 3, pp. 875–883, 2012.
- [14] D. Meseure, S. Vacher, F. Lallemand et al., “Prognostic value of a newly identified MALAT1 alternatively spliced transcript in breast cancer,” *British Journal of Cancer*, vol. 114, no. 12, pp. 1395–1404, 2016.
- [15] L. Wu, Q. Xu, H. Zhang et al., “A new avenue for obtaining insight into the functional characteristics of long noncoding RNAs associated with estrogen receptor signaling,” *Scientific Reports*, vol. 6, no. 1, 2016.
- [16] P. Langfelder, B. Zhang, and S. Horvath, “Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R,” *Bioinformatics*, vol. 24, no. 5, pp. 719–720, 2008.
- [17] B. Zhang and S. Horvath, “A general framework for weighted gene co-expression network analysis,” *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, 2005.
- [18] S. Anders and W. Huber, “Differential expression analysis for sequence count data,” *Genome Biology*, vol. 11, no. 10, p. R106, 2010.

- [19] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2009.
- [20] Y. S. Huang, C. C. Chang, S. S. Lee, Y. S. Jou, and H. M. Shih, "Xist reduction in breast cancer upregulates AKT phosphorylation via HDAC3-mediated repression of PHLPP1 expression," *Oncotarget*, vol. 7, no. 28, pp. 43256–43266, 2016.
- [21] A. Vincent-Salomon, C. Ganem-Elbaz, E. Manie et al., "X inactive-specific transcript RNA coating and genetic instability of the X chromosome in BRCA1 breast tumors," *Cancer Research*, vol. 67, no. 11, pp. 5134–5140, 2007.
- [22] L. Pang, J. Hu, G. Zhang et al., "Dysregulated long intergenic non-coding RNA modules contribute to heart failure," *Oncotarget*, vol. 7, no. 37, pp. 59676–59690, 2016.
- [23] R. Sharan, I. Ulitsky, and R. Shamir, "Network-based prediction of protein function," *Molecular Systems Biology*, vol. 3, 2007.
- [24] M. Huarte, M. Guttman, D. Feldser et al., "A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response," *Cell*, vol. 142, no. 3, pp. 409–419, 2010.
- [25] F. Yang, Y. H. Liu, S. Y. Dong et al., "A novel long non-coding RNA FGF14-AS2 is correlated with progression and prognosis in breast cancer," *Biochemical and Biophysical Research Communications*, vol. 470, no. 3, pp. 479–483, 2016.
- [26] B. Gyorffy, A. Lanczky, A. C. Eklund et al., "An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients," *Breast Cancer Research and Treatment*, vol. 123, no. 3, pp. 725–731, 2010.
- [27] J. Li, L. Han, P. Roebuck et al., "TANRIC: an interactive open platform to explore the function of lncRNAs in Cancer," *Cancer Research*, vol. 75, no. 18, pp. 3728–3737, 2015.
- [28] Z. Qu, J. Cui, Y. Harata-Lee et al., "Identification of candidate anti-cancer molecular mechanisms of compound Kushen injection using functional genomics," *Oncotarget*, vol. 7, no. 40, pp. 66003–66019, 2016.
- [29] J. R. Prensner, M. K. Iyer, A. Sahu et al., "The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex," *Nature Genetics*, vol. 45, no. 11, pp. 1392–1398, 2013.
- [30] D. He, J. Wang, Y. Lu et al., "LncRNA functional networks in oligodendrocytes reveal stage-specific myelination control by an IncOL1/Suz12 complex in the CNS," *Neuron*, vol. 93, no. 2, pp. 362–378, 2017.
- [31] T. Li, J. Xie, C. Shen et al., "Upregulation of long noncoding RNA ZEB1-AS1 promotes tumor metastasis and predicts poor prognosis in hepatocellular carcinoma," *Oncogene*, vol. 35, no. 12, pp. 1575–1584, 2016.
- [32] T. Kim, R. Cui, Y. J. Jeon, P. Fadda, H. Alder, and C. M. Croce, "MYC-repressed long noncoding RNAs antagonize MYC-induced cell proliferation and cell cycle progression," *Oncotarget*, vol. 6, no. 22, pp. 18780–18789, 2015.
- [33] H. Karra, H. Repo, I. Ahonen et al., "Cdc20 and securin over-expression predict short-term breast cancer survival," *British Journal of Cancer*, vol. 110, no. 12, pp. 2905–2913, 2014.
- [34] J. Y. Han, Y. K. Han, G.-Y. Park, S. D. Kim, and C. Geun Lee, "Bub1 is required for maintaining cancer stem cells in breast cancer cell lines," *Scientific Reports*, vol. 5, no. 1, 2015.
- [35] A. Mukherjee, C. Joseph, M. Craze, E. Chrysanthou, and I. O. Ellis, "The role of BUB and CDC proteins in low-grade breast cancers," *The Lancet*, vol. 385, p. S72, 2015.
- [36] F. Al-Ejeh, P. T. Simpson, J. M. Sanus et al., "Meta-analysis of the global gene expression profile of triple-negative breast cancer identifies genes for the prognostication and treatment of aggressive breast cancer," *Oncogenesis*, vol. 3, no. 4, p. e100, 2014.
- [37] J. M. Mason, X. Wei, G. C. Fletcher et al., "Functional characterization of CFI-402257, a potent and selective Mps1/TTK kinase inhibitor, for the treatment of cancer," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 12, pp. 3127–3132, 2017.
- [38] K. Yata, J. Y. Bleuyard, R. Nakato et al., "BRCA2 coordinates the activities of cell-cycle kinases to promote genome stability," *Cell Reports*, vol. 7, no. 5, pp. 1547–1559, 2014.
- [39] N. E. Bhola, V. M. Jansen, S. Bafna et al., "Kinome-wide functional screen identifies role of PLK1 in hormone-independent, ER-positive breast cancer," *Cancer Research*, vol. 75, no. 2, pp. 405–414, 2015.
- [40] V. Maire, F. Nemati, M. Richardson et al., "Polo-like kinase 1: a potential therapeutic option in combination with conventional chemotherapy for the management of patients with triple-negative breast cancer," *Cancer Research*, vol. 73, no. 2, pp. 813–823, 2013.
- [41] S. Feng, J. Zhang, W. Su et al., "Overexpression of LINC00152 correlates with poor patient survival and knockdown impairs cell proliferation in lung cancer," *Scientific Reports*, vol. 7, no. 1, p. 2982, 2017.

