

WITHDRAWN: Identification of a novel lncRNA-miRNA-mRNA competing endogenous RNA network associated with prognosis of breast cancer

Xiaohui Wan

The Second Hospital of Jilin University

Shuhong Hao

The Second Hospital of Jilin University

Chunmei Hu

The Second Hospital of Jilin University

Rongfeng Qu

qurongfeng_qr-f@yeah.net

The Second Hospital of Jilin University <https://orcid.org/0000-0002-1224-7452>

Primary research

Keywords: Breast cancer, Competing endogenous RNA, Long noncoding RNA, microRNA, Bioinformatic analysis

Posted Date: February 28th, 2020

DOI: <https://doi.org/10.21203/rs.2.24821/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

EDITORIAL NOTE:

The full text of this preprint has been withdrawn by the authors while they make corrections to the work. Therefore, the authors do not wish this work to be cited as a reference. Questions should be directed to the corresponding author.

Abstract

Background: Breast cancer is one of the most common malignant tumors. Recently, the effects of competing endogenous RNA (ceRNA) on molecular biological mechanism of cancer has aroused great interest. However, research on the pathogenesis and biomarkers of breast cancer is still limited. Thus, this study is aimed to identify the competing endogenous RNA (ceRNA) network related to prognosis of patients with breast cancer.

Methods: The RNA SEQ data and corresponding clinical information were downloaded from the Cancer Genome Atlas (TCGA) database, and the difference analysis was performed after data quality control. The similarity between two groups of genes with traits in the network was analyzed by weighted correlation network analysis (WGCNA). Next, the interaction among lncRNA, miRNA, and mRNA was predicted using miRcode, TargetScan, miRDB, and miRTarBase database, and the lncRNA-miRNA-mRNA ceRNA network was constructed. Finally, the survival model of target mRNA was established by Cox regression analysis.

Results: In total 5056 differentially expressed lncRNAs, 712 differentially expressed miRNAs, and 9878 differentially expressed mRNAs were identified. WGCNA predicted that 823 lncRNAs and 1813 mRNAs were closely related to the breast cancer. The lncRNA-miRNA-mRNA ceRNA network involved in breast cancer was constructed with 27 lncRNA, 14 miRNAs and 4 mRNAs. The AUC of four survival models of target mRNA (ZC3H12B + HRH1 + TMEM132C + PAG1) was 0.609, which suggested the sensitivity and specificity of prognosis prediction of breast cancer.

Conclusion: This study provides insight into the ceRNA network involved in breast cancer biology, which significantly associated with gene regulation and prognosis of breast cancer.

Background

Breast cancer, the most frequently diagnosed malignancy in women, remains the second reason for cancer-related in female around the world [1]. It is reported that there are three major therapies for breast cancer, including chemotherapy, endocrine therapy, and targeted therapy [2]. Breast cancer, a heterogeneous disease, varies greatly among different patients and even within each individual patient [3]. Although the survival rate has gradually increased due to the early detection and advances of therapeutic strategy in recent years, there are still many obstacles in the treatment of breast cancer [4]. Since we still do not fully understand how breast cancer occurs and progresses, it is imperative and meaningful to explore the underlying molecular mechanisms and identify novel prognostic biomarkers and potential therapeutic targets for breast cancer.

It is well known that competing endogenous RNAs (ceRNAs) could indirectly regulate each other by reducing the amount of microRNAs (miRNAs) available to target messenger RNAs (mRNAs), thereby affecting the progression of various disease conditions, including cancers [5]. A recent study has demonstrated that lncRNAs acting as miRNA decoys modulating the derepression of miRNA targets has

been found in multiple human cancers [6]. It has been proved and widely acknowledged that noncoding RNA including miRNA and long noncoding RNA (lncRNA) could regulate gene expression at transcriptional and post-transcriptional level [7]. The aberrant expression of miRNA and lncRNA has been widely reported to be implicated in tumorigenesis and development of various cancers, including breast cancer [8, 9]. A prior study has proposed a ceRNA hypothesis that mRNA, miRNA and lncRNA could achieve cross-talk between each other by forming a regulatory network [10]. Accumulating evidence has elucidated that lncRNAs could serve as sponges for miRNAs through miRNA response elements, resulting in alterations in the miRNAs-regulated mRNA levels, and the lncRNA-miRNA-mRNA ceRNA network is reported to be involved in the development and progression of some tumors, including breast cancer [11]. Although ceRNAs have been applied for biological function annotations, few of them have investigated the ceRNA network on specific disease systematically in breast cancer [12]. Nevertheless, understanding of the pivotal lncRNA-miRNA-mRNA ceRNA networks associated significantly with prognosis of breast cancer remain very limited. Therefore, the present study is aimed to explore the role of the discovered lncRNA-miRNA-mRNA ceRNA network in the pathogenesis of breast cancer and its possible mechanisms is explored, which may help to provide a novel direction for treating breast cancer.

Methods

Datasets selection

The transcriptome and clinical data of breast cancer were retrieved from TCGA database. RNA SEQ, miRNA SEQ and clinical data were downloaded using data transmission tools (provided by GDC apps) (<https://tcga-data.nci.nih.gov/>). The RNA sequencing data consisted of 1109 breast cancer tissues and 113 normal tissues, which are combined into a matrix file by Perl script (<http://www.perl.org/>), and then the ensemble ID of the sample was transformed by the annotation file of GENCODE Gene Set-09.2019 version. The lncRNA and mRNA ensemble ID not included in the GENCODE database are excluded. Finally, the expression matrix of lncRNA and mRNA was obtained. The miRNA SEQ data included 1103 breast cancer tissues and 104 normal tissues, which were combined into a matrix file using Perl language (<http://www.perl.org/>). The downloaded clinical data contained 1097 cases to obtain useful information. Due to the availability of public data in TCGA database, this study does not need moral approval or informed consent.

Differential expression analysis

The "edge" R software package was employed to identify differentially expressed lncRNA, mRNA, and miRNA in breast cancer and normal samples. All q values were corrected for statistical significance of multiple tests using false discovery rate (FDR) with $|\log_2(\text{fold change [FC]})| > 0.5$ and $\text{FDR} < 0.05$. Volcano maps of differentially expressed lncRNA, miRNA, and mRNA were drawn with ggplot2 of R package.

Functional enrichment analysis

The gene ontology (GO) functional enrichment analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Enrichment Analysis (GSEA) pathway enrichment analysis were conducted using the "ClusterProfiler" R software. Go function enrichment analysis was carried out from three aspects of biological process (BP), cell composition (CC) and molecular function (MF). KEGG and GSEA pathway enrichment analysis were carried out at the significant level of $p < 0.05$

Weighted Correlation Network Analysis (WGCNA)

WGCNA is an algorithm for gene co-expression network recognition through mRNA or lncRNA with different characters in high-throughput expression spectrum, which can distinguish genes into multiple clusters. The pairwise Pearson correlation analysis was utilized to evaluate the weighted co-expression relationship among all datasets in the adjacency matrix. In this study, WGCNA was adopted to analyze mRNA and lncRNA to obtain the mRNA or lncRNA most relevant to breast cancer.

Construction of ceRNA network

According to the relationship among lncRNA, miRNA, and mRNA, the ceRNA network was constructed. First of all, based on the miRcode (<http://www.mircode.org/>) database, it was predicted that the significantly upregulated lncRNA in breast cancer has a targeted relationship with the significantly downregulated miRNA in breast cancer and the significantly downregulated lncRNA in breast cancer has a targeted relationship with the significantly upregulated miRNA. The miRNA with correct trend and targeted relationship was selected as the target miRNA. Next, based on the TargetScan (<http://www.targetscan.org/>), miRDB (<http://www.mirdb.org/>), and miRTarBase (<http://mirtarbase.mbc.nctu.edu.tw/php/index.php>), we predicted that the target miRNA with poor expression in breast cancer has a targeted relationship with upregulated mRNA, and the target miRNA with high expression in breast cancer has a targeted relationship with downregulated mRNA. The mRNA with correct trend and retrieved from three databases was selected as candidate mRNA. The combination of predicted lncRNA-miRNA pair and miRNA-mRNA pair was used to construct the lncRNA-miRNA-mRNA regulatory network. Finally, the network was visualized and mapped using the Cytoscape v3.6.1 software.

Cox regression analysis

Single variable Cox proportional risk regression analysis was used to explore the relationship between the expression level of mRNAs and overall survival (OS). Then, multivariate Cox analysis was used to evaluate the contribution of the selected genes. Survival analysis was conducted with R package "survival"

Results

Identification of significant DE-lncRNAs, DE-miRNAs, and DE-mRNAs in breast cancer

To explore the expression levels of DE-lncRNAs, DE-miRNAs, and DE-mRNAs in breast cancer, transcriptome data of breast cancer and adjacent normal tissues were obtained from TCGA database. Table 1 displayed the clinicopathological characteristics of patients with breast cancer. After data quality control, 5056 DE-lncRNAs (3559 upregulated lncRNAs and 1497 downregulated lncRNAs), 712 DE-miRNAs (503 upregulated miRNAs and 209 downregulated miRNAs) and 9878 DE-mRNAs (6906 upregulated mRNAs and 2972 downregulated mRNAs) were obtained through differential analysis of edgeR package. The distribution of lncRNA, miRNA and mRNA in the two dimensions of $-\log_{10}(\text{FDR})$ and $\log_2(\text{FC})$ was shown by volcano map (Fig. 1A-C).

Table 1
The clinicopathological characteristics of patients with breast cancer

	Alive (n = 953)	Dead (n = 144)	Total (n = 1097)
Gender			
FEMALE	942(98.8%)	143(99.3%)	1085(98.9%)
MALE	11(1.2%)	1(0.7%)	12(1.1%)
Age			
Mean(SD)	58.1(12.9)	61(14.8)	58.5(13.2)
Median[MIN, MAX]	58[26, 90]	62[26, 90]	58[26, 90]
FAB classification			
M0	791(83.0%)	115(79.9%)	906(82.6%)
M1	5(0.5%)	17(11.8%)	22(2.0%)
MX	151(15.9%)	12(8.3%)	163(14.9%)
cM0 (i+)	6(0.6%)	0(0%)	6(0.5%)

Functional analysis for the DE-mRNAs

To explore the biological functions of the identified DE-mRNAs, we performed GO term enrichment analysis through “ClusterProfiler” R software. In the BP, upregulated mRNA was significantly enriched in protein activation cascade, complement activation, immunoglobulin mediated immune response, complement activation, classical pathway, and circulating immunoglobulin-mediated humoral immune response (Fig. 2A). The downregulated mRNA was significantly enriched in the process of blood circulation, circulatory system, positive regulation of cytoplasmic calcium concentration, cell adhesion through plasma membrane adhesion molecules, and penile erection (Fig. 2B). It was found the interaction between the enriched pathway of upregulated or downregulated genes and its corresponding mRNAs in the BP (Fig. 2C-D). In addition, through KEGG-GSEA analysis, it was found that the upregulated mRNA was significantly enriched in transcription disorders, systemic lupus erythematosus, and

alcoholism in human cancers (Fig. 2E). These data implied that these significantly different genes may play an important role in the regulation of breast cancer through these enriched pathways.

Validation of genemodules using WGCNA

WGCNA was performed to analyze the co-expression of the first 80% genes (at least MAD > 0.01) with median absolute deviation. In this study, $\beta = 2$ (scale free R² = 0.91) was selected as the soft threshold parameter to ensure scale-free network (Fig. 3A). According to $\beta = 2$, the proximity matrix was obtained to further verify that the selected memory network approximated scale free (Fig. 3B). There were 11 gene modules by WGCNA. The characteristic genes of all modules were calculated and clustered according to their correlation. The dendrogram of module characteristic gene showed that 10 modules are mainly divided into two clusters, and the results were similar to the heat map of characteristic gene network (Fig. 3C). Due to the large difference coefficient between the modules, there was no need to merge the modules, and finally 11 modules were determined (Fig. 3D). As shown in Fig. 3E, the thermal map of the topological overlap matrix (TOM) presented that there was a high degree of independence between these modules. Subsequently, we analyzed the similarity and adjacency of the co-expression of characteristics and traits (breast cancer tissues and normal tissues) of gene modules of 11 colors. It was found that yellow, brown, and black module genes were positively correlated with breast cancer, while purple, blue and red module genes were negatively correlated with breast cancer, including 1813 mRNA (Fig. 3F). These mRNA were further used in GO-GSEA to represent gene enrichment, gene name and interaction in BP (Fig. 3G-H), which exhibited that these genes were significantly related to the biological processes of anatomical structure development, cell macromolecule metabolism process, cell metabolism process, development process, and multicellular tissue process. In addition, KEGG analysis showed that the genes were significantly enriched in the neuroactive ligand receptor interaction, cytokine cytokine receptor interaction, renin secretion, the interaction between viral protein and cytokine and cytokine receptor, legionellosis, chemokine signaling pathway, and calcium signaling pathway (Fig. 3I).

Screening & validation of lncRNAmolecules using WGCNA

WGCNA was performed to analyze the co-expression of the first 80% lncRNAs (at least MAD > 0.01) with median absolute deviation. In this study, $\beta = 2$ (scale free R² = 0.92) was selected as the soft threshold parameter to ensure scale-free network (Fig. 4A). According to $\beta = 2$, the proximity matrix was obtained to further verify that the selected memory network approximated scale free (Fig. 4B). Eight lncRNA modules were identified by WGCNA. The characteristic genes of all modules were calculated and clustered according to their correlation (Fig. 4C). It was found that the coefficient of dissimilarity among some modules was very small. MEDissThres was set as 0.25 to merge similar modules, and finally 7 modules were determined (Fig. 4D). As shown in Fig. 4, the thermal map of the TOM presented that there was a high degree of independence between these modules. Then, we analyzed the similarity and adjacency of the characteristics and the traits (breast cancer tissues and normal tissues) of lncRNA modules of 7 colors, which revealed that yellow modules were positively correlated with breast cancer, while brown, red, and black modules were negatively correlated with breast cancer (Fig. 4F). The number of lncRNAs in

each module is shown in Fig. 4G. There are 823 lncRNAs in yellow, brown, red and black modules. The target miRNAs of 823 lncRNAs were predicted by miRcode database to obtain the lncRNAs-miRcode-miRNAs relationship. Then, TCGA miRNA SEQ was utilized to analyze the significant difference between miRNA and miRNA in lncRNAs-miRcode-miRNAs axis to screen lncRNAs-miRNAs. Finally, we obtained 355 lncRNA-miRNAs, including 56 lncRNAs and 47 miRNAs. Among these 47 miRNAs, 31 miRNAs were significantly upregulated and 16 were significantly downregulated. The target genes downstream of significantly upregulated and downregulated miRNAs were predicted by TargetScan, miRDB, and miRTarBase, respectively (Fig. 4H). According to the trend, 31 significantly upregulated miRNAs predicted by the three databases were intersected with mRNA (1813) in purple, blue, red, black, brown and yellow modules predicted using WGCNA and significantly downregulated mRNA (2972) predicted using EdgeR, which obtained 16 common genes (Fig. 4I). Subsequently, 16 significantly downregulated miRNAs predicted by the three databases were intersected with mRNA (1813) in purple, blue, red, black, brown and yellow modules predicted using WGCNA and significantly upregulated mRNA (2972) predicted using EdgeR, which obtained 5 common genes (Fig. 4J). Finally, 68 miRNA-mRNA pairs were obtained, including 31 miRNAs and 21 mRNAs. The expression level of these 21 genes in 1222 samples is shown in Fig. 4K.

Survival analysis

The correlation between the expression level of 21 genes and the OS was explained by single variable Cox proportional risk regression analysis. In total 11 genes were obtained with $p < 0.05$ as the threshold. Further multivariate Cox proportional risk regression analysis was carried out for the above 11 genes (Table 2). Then, an OS survival model with four genes was established: ZC3H12B + HRH1 + TMEM132C + PAG1. According to GEPIA database, ZC3H12B, HRH1, and TMEM132C were poorly expressed, and PAG1 was highly expressed in breast cancer samples (Fig. 5A-D), which were consistent the above findings. The correlation between 4 genes in the gene model is shown in Fig. 5E, F. According to multivariate Cox scores, patients from TCGA were divided into low-risk and high-risk groups (Fig. 5G). In addition, the expression heat map of four genes in the high-risk and low-risk groups is shown in Fig. 5H. Then, the accuracy of 4 genes in predicting survival was estimated. Kaplan Meier survival curve showed that OS of high-risk patients ($n = 507$) was significantly shorter than that of low-risk patients ($n = 507$, $P = 0.012$). The sensitivity and specificity of our model survival prediction was compared using ROC analysis. TCGA dataset showed that the area under the signal characteristic curve (AUC) of the four genes was 0.609, obviously greater than 0.6, indicating that our survival curve was very reliable and the model accuracy was good (Fig. 5I-J).

Table 2
Multivariate Cox proportional risk regression analysis for 11 genes

Gene	Univariate	P	Multivariate	
	HR (95%CI)		HR	P
ZC3H12B	1.00325(1.00151-1.00500)	0.0003	1.00283	0.0021**
JPH2	1.00009(1.00003-1.00016)	0.0066		
OMD	1.00014(1.00003 - 1.00025)	0.0152		
HRH1	1.00041(1.00014 - 1.00068)	0.0033	1.00032	0.0420*
KLF2	1.00005(1.00001 - 1.00010)	0.0158		
AVPR1A	1.00022(1.00009 - 1.00035)	0.0012		
SOCS3	1.00001(1.00000 - 1.00002)	0.0069		
TMEM132C	1.00012(1.00006 - 1.00018)	0.0001	1.00007	0.0663
PAG1	0.99966(0.99935 - 0.99996)	0.0264	0.99958	0.0129*
CCR5	0.99962(0.99932 - 0.99992)	0.0119		
CXCR4	0.99994(0.99989 - 0.99999)	0.022		

Construction of the lncRNA-miRNA-mRNA regulatory network in breast cancer

To predict the relationship between four target genes and their corresponding miRNAs, 19 miRNA-mRNA pairs were obtained, including 13 upregulated miRNAs and 1 downregulated miRNA. In order to obtain the lncRNAs regulating these 14 miRNAs, according to the obtained 355 lncRNA-miRNA, the corresponding lncRNAs of 13 upregulated and 1 downregulated miRNA were extracted, respectively. According to the ceRNA trend, the 13 upregulated miRNA corresponding lncRNA was compared with the 1497 obviously downregulated lncRNA obtained from difference analysis of the breast cancer data in the TCGA database (Fig. 6A). It was found that 21 DE-lncRNAs were obtained. The lncRNA corresponding to 1 downregulated miRNA was compared with 3559 significantly upregulated lncRNAs obtained from the difference analysis of breast cancer data in TCGA database (Fig. 6B), which showed 6 common DE-lncRNAs. Ultimately, 64 lncRNA-miRNA pairs were obtained, including 27 lncRNAs and 14 miRNAs. The lncRNA-miRNA-mRNA ceRNA network was constructed with 27 lncRNAs, 14 miRNAs, and 4 mRNAs (Fig. 6C). Moreover, these miRNAs (miR-429, miR-93, miR-17, miR-106a, miR-338, miR-21, miR-155, miR-98, miR-503, miR-508, miR-506, miR-187, miR-425, and miR-205) were predicted to target 4 mRNAs, respectively. For example, miR-93 could target ZC3H12B and HRH1, and miR-429, miR-93, miR-17, miR-106a, and miR-338 could target ZC3H12B.

Discussion

Breast cancer is a global health threat with its increasing incidence in female population [13]. In spite of recent advances in early detection and cancer therapeutics of breast cancer, contributing to decrease in breast cancer mortality rates, it still remains the main cause for public health concern [14]. Thus, it is of great importance to unveil the molecular switch controlling malignant transformation of breast cancer, which may also contribute to excavate novel prognostic indicators. Taken together, the current study reveals the lncRNA-miRNA-mRNA ceRNA network related to the pathogenesis of breast cancer.

The data in the current study predicted that there are 5056 lncRNAs and 712 miRNAs dysregulated in breast cancer. Similarly, a great deal of research has shown that miRNAs and lncRNAs, plays a vital role in the pathogenesis and progression of cancers, including breast cancer [8, 15]. Nowadays, a growing body of evidence has also implicated that there is a complex and close relationship between miRNAs and lncRNAs [10]. Moreover, emerging evidence has also indicated that ceRNA network could play crucial role in various pathological processes of tumorigenesis, including breast cancer [16]. For instance, it has been reported that lncRNA CCAT1 exerted as a ceRNA and negatively regulated miR-148b expression, thereby reducing cancer cell colony formation and promoting apoptosis to regulate the progression of breast cancer [17]. Lu et al. have exhibited that LINC00511 function as sponge for miR-185-3p to positively recover E2F1/Nanog axis, hence contributing to breast cancer tumorigenesis and stemness [18]. lncRNA ARNILA is proved to promote the invasion and metastasis of triple-negative breast cancer by sponging miR-204 to facilitate expression of its target gene Sox4 [19]. On the contrary, various studies also identified some ceRNA networks which have a suppressive role in breast cancer [20]. Nevertheless, there is still a paucity of studies on comprehensive analysis of ceRNA network in breast cancer, and integrated research is needed to be done.

In order to systemically explore potential ceRNA regulating these hub genes mentioned above, the upstream miRNAs were predicted first based on a experimentally validated miRNA-target interactions database, miRTarBase. After combining the expression and survival analysis according to the ceRNA hypothesis, 14 candidate miRNAs were identified as key miRNAs (miR-429, miR-93, miR-17, miR-106a, miR-338, miR-21, miR-155, miR-98, miR-503, miR-508, miR-506, miR-187, miR-425, and miR-205) are associated with the prognosis in breast cancer. For example, Ye et al. demonstrated that miR-429 inhibits breast cancer cell progression by targeting ZEB1 and CRKL [21]. A recent study has revealed that miR-93 contributes to inducing EMT and drug resistance of breast cancer cells by targeting PTEN [22]. miR-338p is also found to be downregulated in breast cancer and act as tumor inhibitor in breast cancer by targeting SOX4 [23]. These reports partially increase the credibility of our research. Moreover, these miRNAs are predicted to target 4 mRNAs. For example, miR-93 could target ZC3H12B and HRH1, and miR-429, miR-93, miR-17, miR-106a, and miR-338 could target ZC3H12B. The survival model with four target mRNAs (ZC3H12B + HRH1 + TMEM132C + PAG1) was established by single variable and multivariate Cox proportional risk regression analysis. ZC3H12B, the most enigmatic member of the ZC3H12 protein family, could arrest cells in G2 cell cycle to inhibit proliferation [24], while its role in breast cancer has not been reported. Existing literature has indicated that HRH1 is correlated with the better

prognosis in patients with breast cancer [25]. TMEM132C is reported to be act as the promising diagnostic and prognostic markers in breast cancer [26]. PAG1 is also demonstrated to stimulate tumor progression and chemoresistance of breast cancer [27].

In this present study, we successfully constructed a novel lncRNA-miRNA-mRNA ceRNA network involved in breast cancer with 27 lncRNA, 14 miRNAs and 4 mRNAs. The network of lncRNA-miRNA-mRNA interactions have been investigated to serve as the refine biomarker predictions for developing novel therapeutic approaches in breast cancer [16]. The breast-cancer specific ceRNA network based on the mRNA and miRNA expression data has also been demonstrated in the breast cancer [12]. Moreover, bioinformatics analysis of ceRNA network associated with breast cancer has greatly emerged, a meaningful lncRNA-miRNA-mRNA regulatory axis, namely, LINC00466-hsa-mir-204-NTRK2, plays a vital roles in the prognosis of breast cancer [28], which further hinted that our results is reliable.

Conclusion

In conclusion, we successfully constructed a potential lncRNA-miRNA-mRNA regulatory network in breast cancer. Each component in the ceRNA network was remarkably related to the prognosis of patients with breast cancer. The ceRNA regulatory network is becoming a research hotspot in the field of noncoding RNA. Since our results may provide novel insights into the mechanism underlying pathogenesis of breast cancer, further experimental validation should be done for these findings in the future. In addition, we considered that identifying ceRNA networks associated with metastasis or staging of breast cancer make a lot of sense in clinic and is worthy of further exploration and research.

Abbreviations

ceRNA, competing endogenous RNA; TCGA, Cancer Genome Atlas; WGCNA, weighted correlation network analysis; FDR, false discovery rate; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; CC, cell composition; BP, biological process; OS, overall survival; TOM, topological overlap matrix;

Declarations

Authors' contributions

XW, SH and CH wrote the main manuscript text, RQ collected the data, and XW prepared all figures. All authors reviewed the manuscript.

Author details

Department of Hematology and Oncology, The Second Hospital of Jilin University, Changchun 130022, P. R. China

Acknowledgements

The authors would like to acknowledge the helpful comments on this paper received from the reviewers.

Competing interests

The authors declare that they have no competing interests

Availability of data and materials

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable

Ethics approval and consent to participate

Due to the availability of public data in TCGA database, this study does not need moral approval or informed consent.

Funding

None

References

1. Jiang R, Zhao C, Gao B, Xu J, Song W, Shi P. Mixomics analysis of breast cancer: Long non-coding RNA linc01561 acts as ceRNA involved in the progression of breast cancer. *Int J Biochem Cell Biol* 2018;102:1-9.

2. Ponde NF, Zardavas D, Piccart M. Progress in adjuvant systemic therapy for breast cancer. *Nat Rev Clin Oncol* 2019;16:27-44.
3. Turashvili G, Brogi E. Tumor Heterogeneity in Breast Cancer. *Front Med (Lausanne)* 2017;4:227.
4. Tracy KM, Tye CE, Page NA, Fritz AJ, Stein JL, Lian JB, et al. Selective expression of long non-coding RNAs in a breast cancer cell progression model. *J Cell Physiol* 2018;233:1291-9.
5. Conte F, Fiscon G, Chiara M, Colombo T, Farina L, Paci P. Role of the long non-coding RNA PVT1 in the dysregulation of the ceRNA-ceRNA network in human breast cancer. *PLoS One* 2017;12:e0171661.
6. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 2010;465:1033-8.
7. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet* 2011;12:861-74.
8. Zare M, Bastami M, Solali S, Alivand MR. Aberrant miRNA promoter methylation and EMT-involving miRNAs in breast cancer metastasis: Diagnosis and therapeutic implications. *J Cell Physiol* 2018;233:3729-44.
9. Huarte M. The emerging role of lncRNAs in cancer. *Nat Med* 2015;21:1253-61.
10. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 2011;146:353-8.
11. Huang X, Xie X, Liu P, Yang L, Chen B, Song C, et al. Adam12 and lnc015192 act as ceRNAs in breast cancer by regulating miR-34a. *Oncogene* 2018;37:6316-26.
12. Zhou X, Liu J, Wang W. Construction and investigation of breast-cancer-specific ceRNA network based on the mRNA and miRNA expression data. *IET Syst Biol* 2014;8:96-103.
13. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin* 2018;68:7-30.
14. Lo PK, Wolfson B, Zhou X, Duru N, Gernapudi R, Zhou Q. Noncoding RNAs in breast cancer. *Brief Funct Genomics* 2016;15:200-21.
15. Schmitt AM, Chang HY. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* 2016;29:452-63.
16. Zhang Y, Li Y, Wang Q, Zhang X, Wang D, Tang HC, et al. Identification of an lncRNAmiRNAmRNA interaction mechanism in breast cancer based on bioinformatic analysis. *Mol Med Rep* 2017;16:5113-20.
17. Lai Y, Chen Y, Lin Y, Ye L. Down-regulation of lncRNA CCAT1 enhances radiosensitivity via regulating miR-148b in breast cancer. *Cell Biol Int* 2018;42:227-36.
18. Lu G, Li Y, Ma Y, Lu J, Chen Y, Jiang Q, et al. Long noncoding RNA LINC00511 contributes to breast cancer tumorigenesis and stemness by inducing the miR-185-3p/E2F1/Nanog axis. *J Exp Clin Cancer Res* 2018;37:289.
19. Yang F, Shen Y, Zhang W, Jin J, Huang D, Fang H, et al. An androgen receptor negatively induced long non-coding RNA ARNILA binding to miR-204 promotes the invasion and metastasis of triple-negative breast cancer. *Cell Death Differ* 2018;25:2209-20.

20. Li S, Zhou J, Wang Z, Wang P, Gao X, Wang Y. Long noncoding RNA GAS5 suppresses triple negative breast cancer progression through inhibition of proliferation and invasion by competitively binding miR-196a-5p. *Biomed Pharmacother* 2018;104:451-7.
21. Ye ZB, Ma G, Zhao YH, Xiao Y, Zhan Y, Jing C, et al. miR-429 inhibits migration and invasion of breast cancer cells in vitro. *Int J Oncol* 2015;46:531-8.
22. Chu S, Liu G, Xia P, Chen G, Shi F, Yi T, et al. miR-93 and PTEN: Key regulators of doxorubicin-resistance and EMT in breast cancer. *Oncol Rep* 2017;38:2401-7.
23. Jin Y, Zhao M, Xie Q, Zhang H, Wang Q, Ma Q. MicroRNA-338-3p functions as tumor suppressor in breast cancer by targeting SOX4. *Int J Oncol* 2015;47:1594-602.
24. Wawro M, Wawro K, Kochan J, Solecka A, Sowinska W, Lichawska-Cieslar A, et al. ZC3H12B/MCPIP2, a new active member of the ZC3H12 family. *RNA* 2019;25:840-56.
25. Fernandez-Nogueira P, Noguera-Castells A, Fuster G, Recalde-Percaz L, Moragas N, Lopez-Plana A, et al. Histamine receptor 1 inhibition enhances antitumor therapeutic responses through extracellular signal-regulated kinase (ERK) activation in breast cancer. *Cancer Lett* 2018;424:70-83.
26. de Almeida BP, Apolonio JD, Binnie A, Castelo-Branco P. Roadmap of DNA methylation in breast cancer identifies novel prognostic biomarkers. *BMC Cancer* 2019;19:219.
27. Lu Y, Yang Y, Liu Y, Hao Y, Zhang Y, Hu Y, et al. Upregulation of PAG1/Cbp contributes to adipose-derived mesenchymal stem cells promoted tumor progression and chemoresistance in breast cancer. *Biochem Biophys Res Commun* 2017;494:719-27.
28. Gao C, Li H, Zhuang J, Zhang H, Wang K, Yang J, et al. The construction and analysis of ceRNA networks in invasive breast cancer: a study based on The Cancer Genome Atlas. *Cancer Manag Res* 2019;11:1-11.

Figures

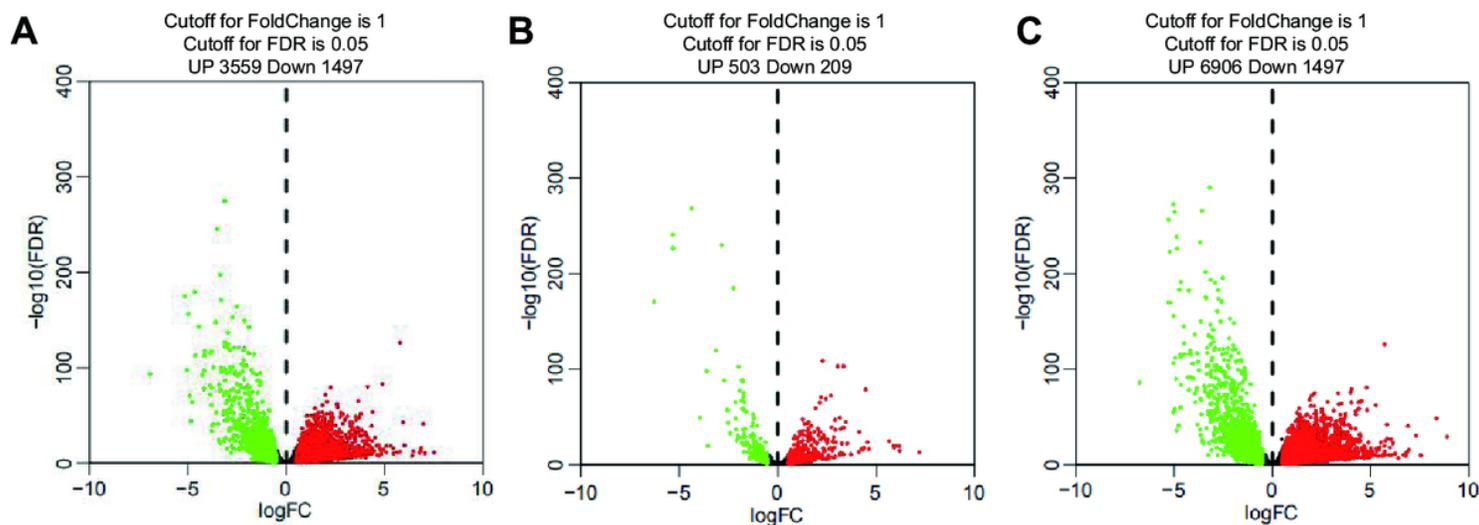


Figure 1

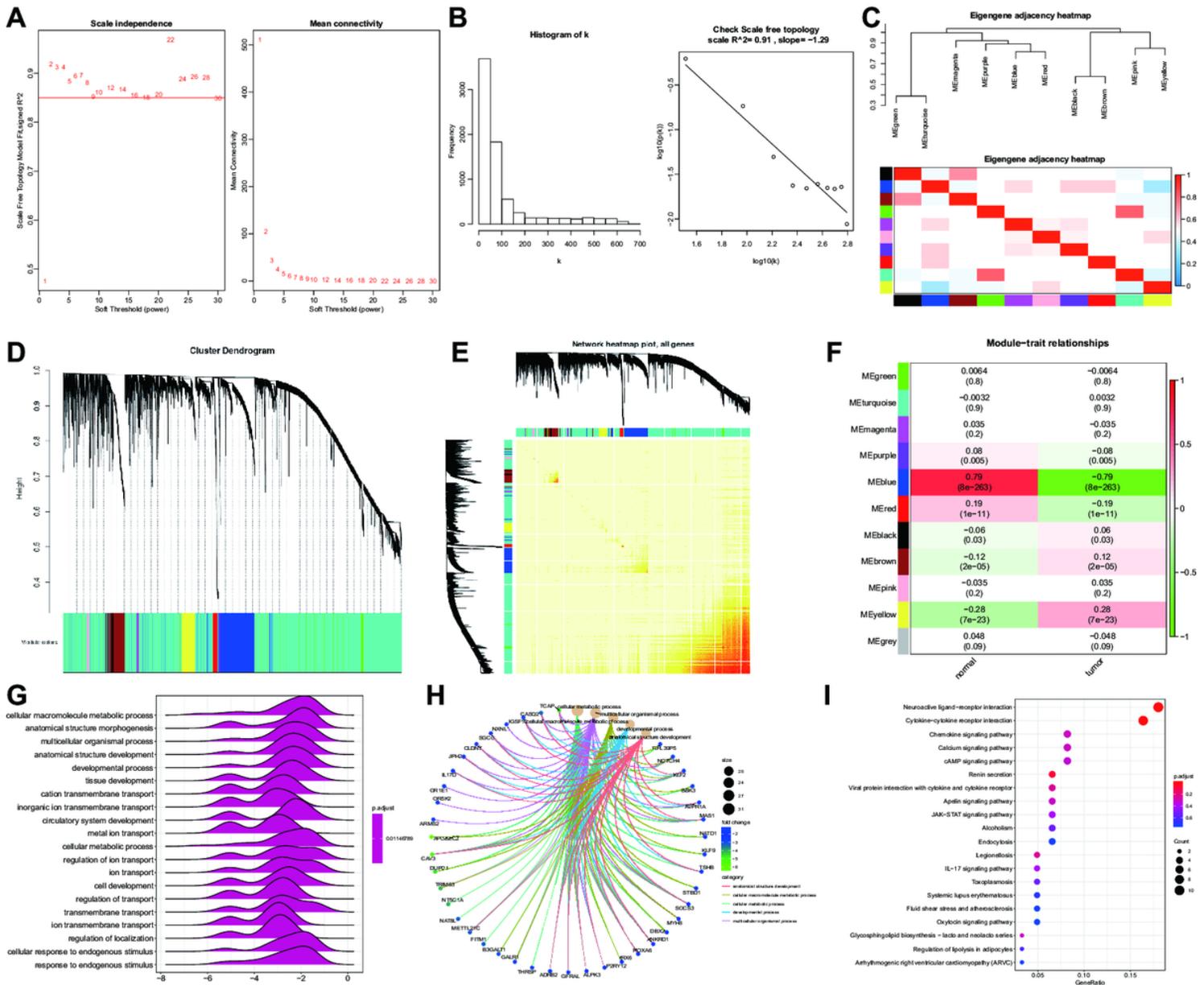


Figure 3

Screening the key genes in breast cancer using WGCNA. (A) Scale free fit index and average connectivity of various soft thresholds (β). (B) Histogram of connectivity distribution and checking scale-free topology when $\beta = 2$. (C) Related heat map of adjacent modules in WGCNA. The rectangles of each row and column represent the characteristic genes of a module. In the related heat map, light blue indicates low adjacency, while red indicates high adjacency. (D) Based on the topological overlap and dissimilarity, a cluster tree was constructed, which showed 11 gene co expression modules in breast cancer. Grey module indicates that there is no co expression between genes. (E) The heat map of TOM in gene network. Different colors of the horizontal and vertical axes represent different modules. The yellow in the middle indicates the connection degree of different modules. There is no significant difference in the interaction between different modules, indicating that there is a high degree of independence between these modules. (F) The heat map of correlation between genes with module characteristics and traits

the significantly downregulated genes in TCGA database. (J) According to the trend of ceRNA, the intersection between the target mRNA predicted by significantly downregulated miRNA and breast cancer-related mRNA using WGCNA and the significantly upregulated genes in TCGA database. (K) The heat map of differentially expressed 21 target genes in breast cancer data of TCGA database.

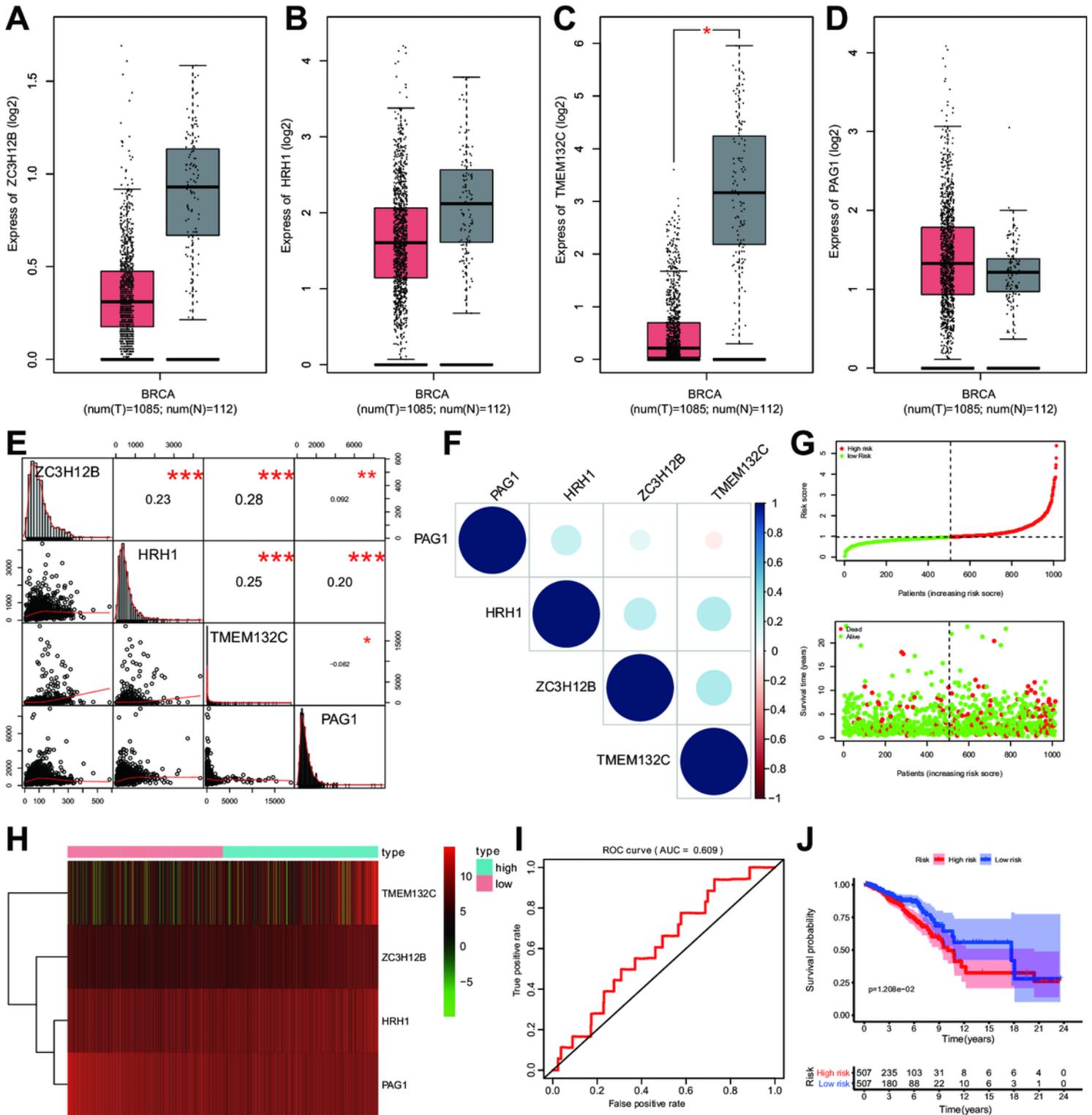


Figure 5

Construction of the lncRNA-miRNA-mRNA ceRNA network in breast cancer. (A) Prediction of target lncRNAs of 13 upregulated miRNA. Blue circle represents the prediction result of miRcode database, and red circle represents the 1497 significantly downregulated lncRNAs from the difference analysis of breast cancer data in TCGA database. (B) Prediction of target lncRNAs of 1 downregulated miRNA. Blue circle represents the prediction result of miRcode database, and red circle represents the 3559 significantly upregulated lncRNAs from the difference analysis of breast cancer data in TCGA database. (C) Construction of the lncRNA-miRNA-mRNA ceRNA network consisting of 27 lncRNAs, 14 miRNAs, and 4 mRNAs.