

ORIGINAL ARTICLE

Integrated analysis of co-expression and ceRNA network identifies five lncRNAs as prognostic markers for breast cancer

Yan Yao¹ | Tingting Zhang² | Lingyu Qi² | Chao Zhou³ | Junyu Wei³ | Fubin Feng³ | Ruijuan Liu³ | Changgang Sun^{4,5} 

¹Clinical Medical Colleges, Weifang Medical University, Weifang, China

²College of First Clinical Medicine, Shandong University of Traditional Chinese Medicine, Jinan, China

³Department of Oncology, Weifang Traditional Chinese Hospital, Weifang, China

⁴Department of Oncology, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, China

⁵Department of Oncology, Affiliated Hospital of Weifang Medical University, Weifang, China

Correspondence

Changgang Sun, Department of Oncology, Affiliated Hospital of Weifang Medical University, No. 2428, Yuhe Road, Kuiwen, Weifang, Shandong, China.
Email: zhongliuyike@163.com

Funding information

National Natural Science Foundation of China, Grant/Award Number: 81473513 and 81673799

Abstract

Long non-coding RNAs (lncRNAs), which competitively bind miRNAs to regulate target mRNA expression in the competing endogenous RNAs (ceRNAs) network, have attracted increasing attention in breast cancer research. We aim to find more effective therapeutic targets and prognostic markers for breast cancer. lncRNA, mRNA and miRNA expression profiles of breast cancer were downloaded from TCGA database. We screened the top 5000 lncRNAs, top 5000 mRNAs and all miRNAs to perform weighted gene co-expression network analysis. The correlation between modules and clinical information of breast cancer was identified by Pearson's correlation coefficient. Based on the most relevant modules, we constructed a ceRNA network of breast cancer. Additionally, the standard Kaplan-Meier univariate curve analysis was adopted to identify the prognosis of lncRNAs. Ultimately, a total of 23 and 5 modules were generated in the lncRNAs/mRNAs and miRNAs co-expression network, respectively. According to the Green module of lncRNAs/mRNAs and Blue module of miRNAs, our constructed ceRNA network consisted of 52 lncRNAs, 17 miRNAs and 79 mRNAs. Through survival analysis, 5 lncRNAs (AL117190.1, COL4A2-AS1, LINC00184, MEG3 and MIR22HG) were identified as crucial prognostic factors for patients with breast cancer. Taken together, we have identified five novel lncRNAs related to prognosis of breast cancer. Our study has contributed to the deeper understanding of the molecular mechanism of breast cancer and provided novel insights into the use of breast cancer drugs and prognosis.

KEYWORDS

breast cancer, ceRNA network, overall survival, The Cancer Genomes Atlas, weighted gene co-expression network analysis

Yao and Zhang contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Journal of Cellular and Molecular Medicine* published by Foundation for Cellular and Molecular Medicine and John Wiley & Sons Ltd.

1 | INTRODUCTION

Breast cancer, the most frequently diagnosed carcinoma, is the deadliest form of cancer affecting women worldwide.¹ According to the statistics of the American Cancer Society, more than 268 600 new cases of invasive breast cancer will be diagnosed in 2019, and approximately 41 760 death cases are expected.² Currently, although multiple therapeutic measures are employed, such as chemotherapy, surgery, radiotherapy, endocrine therapy and targeted therapy,³ the majority of breast cancers remain incurable. This therefore highlights the urgent need for identifying the molecules that have an inhibitory role or facilitating role in tumours, determining novel relevant treatment targets and further improving the treatment and prognosis of patients with breast cancer. Long non-coding RNAs (lncRNAs) are a type of RNA that are over 200 nucleotides in length.⁴ Although they do not participate in protein coding, increasing evidence has demonstrated that lncRNAs are involved in the biological processes of cellular proliferation, apoptosis and differentiation.^{5,6} Furthermore, sufficient information has demonstrated that lncRNAs take part in the regulation of tumour progression as well as tumour biological behaviours through interactions with microRNAs (miRNAs) or messenger RNAs (mRNAs).⁷⁻¹⁰ The competing endogenous RNA (ceRNA) hypothesis proposed by Salmena et al have attracted increasing attention.¹¹ This hypothesis has been supported by a large number of experiments. For example, a previous study found that lncRNA-related ceRNAs play a key role in biological processes of glioblastomas.¹² Additionally, Ning et al found that through competing with miR-331-3p, the lncRNA HOTAIR, as a ceRNA, could induce and activate human epithelial growth factor receptor 2 (HER2) cell signalling networks.¹³ Nevertheless, little is known about the role of ceRNA in breast cancer.

Weighted gene co-expression network analysis (WGCNA) is a system biology algorithm commonly used to explore the correlation between gene sets and clinical features via constructing free-scale gene co-expression networks.¹⁴⁻¹⁶ The advantage of WGCNA is that it can identify and cluster highly correlated genes into the same module. Furthermore, these modules also provide external clinical traits with related modules. However, the majority of studies focused on differentially expressed genes and ignored the high relationship between genes. At present, WGCNA plays a significant role in multiple fields, such as cancer, nervous system and genetic data analysis, which is extremely useful for identifying potential candidate biomarkers or novel treatment targets.¹⁷⁻²⁰

Our study utilized the data related to breast cancer from The Cancer Genomes Atlas (TCGA) database to identify novel lncRNAs. Here, we extracted breast cancer expression profiles of lncRNAs, miRNAs and mRNAs from TCGA database and their corresponding clinical data. Subsequently, we selected lncRNAs, mRNAs and miRNAs whose expression levels were in the top 5000 among the 112 matched specimen pairs. Afterwards, these data were used to construct a co-expression network to determine the module related to the clinical trait. Based on the data of the module, ceRNA network analysis was performed using the following databases: miRTarBase, miRcode and TargetScan. Additionally, we also performed survival

analysis to screen lncRNAs most relevant to prognosis. Findings from our study contribute to further understanding of the molecular mechanisms, biological processes and treatment targets of lncRNAs in the field of breast cancer research.

2 | MATERIALS AND METHODS

2.1 | Data acquisition and pre-processing

As the largest cancer gene information database at present, TCGA database provides not only a variety of cancer types but also multi-omics data, including gene transcript, miRNA expression data and DNA methylation. Additionally, it also has the advantage of containing abundant and standardized clinical data, as well as large sample sizes for each cancer type. All data sets were downloaded from TCGA (<https://portal.gdc.cancer.gov/>, accessed in 18 December, 2018) database through the package TCGAbiolinks in R software (version 3.5.1), including lncRNA, mRNA and miRNA expression profiles of breast cancer specimens and the corresponding clinical follow-up data. Notably, we performed data analysis on the basis of 'Level-3' read counts. The data used in this study met the following criteria: (a) there were pairs of cancerous and the corresponding normal tissue in the datasets; and (b) they had specific follow-up times. Based on the TCGA ethics committee, we obtained these relevant data that are open access and public; thus, ulterior approval was not required.

2.2 | Construction of weighted gene co-expression network

The expression profile of these genes, lncRNAs and miRNAs were applied to construct a gene co-expression network by using the package WGCNA implemented in R software.¹⁴ We performed the same analysis as described previously.¹⁴ The construction process of genes and lncRNAs co-expression network is similar to that of miRNAs co-expression network with the exception of some parameters. For example, we constructed a genes and lncRNAs co-expression network.

This procedure included the following key steps: firstly, in order to eliminate the interference caused by the length of gene and the depth of sequencing, the method of Fragments Per Kilobase Million (FPKM) was applied to standardize the data of 'Level-3' read counts. The formula of FPKM is as follows:

$$\text{FPKM} = \frac{\text{total gene Fragements}}{\text{mapped Reads(Millions)} \times \text{gene Length(KB)}}$$

In this study, FPKM values were >1 in more than 50% of specimen pairs. Then, through drawing cluster trees, outlier samples were removed in order to make the results more credible. Secondly, we calculated Pearson's correlation coefficient (PCC) $\text{cor}(i,j)$ for each pair of mRNAs and lncRNAs. The construction of similar expression matrix is as follows:

$$a_{ij} = (0.5 \times (1 + \text{cor}(i, j)))^\beta$$

In which, a_{ij} refers to the contributions between genes i and j . Subsequently, one needs to introduce a power of β value so that it could transform the similarity matrix into an adjacency matrix. For mRNAs and lncRNAs, the power of β value is 10; however, for miRNAs, the power of β value is 6. On this basis, we constructed a scale-free network and topological overlap matrix (TOM). After that, we also carried out the corresponding dissimilarity of TOM (disTOM), from which hierarchical clustering tree of genes (dendrogram) by function hclust was produced by hierarchical clustering for module detection. The Dynamic Tree Cut method was applied to generate modules with the following major parameters to avoid generation of too many modules: deepSplit of 2 and minModuleSize of 40 (For miRNAs, the minModuleSize was set as 20). The height cut-off was set as 0.25, modules were merged together if their similarity was >0.75. The cut-off was set as 0.20 for miRNAs. Ultimately, these mRNAs and lncRNAs containing co-expression modules are considered to be highly interconnected.

2.3 | Relationship between clinical information and modules

The correlation between modules and clinical information (ie normal or tumour) of breast cancer was identified by PCC. Above all, module eigengenes (MEs) referred to the first principal component of all gene expression levels in the module, and therefore, it was reasonable to consider that MEs represented all genes within a specific module. According to Pearson's correlation tests, we further identified the association between MEs and external clinical information including sample status. If P -value was < .05, it was considered to be a significant correlation.

2.4 | CeRNA network construction and analysis

According to the results of WGCNA, we selected all mRNAs, lncRNAs and miRNAs in the most relevant module to construct a ceRNA network. Briefly, the associated ceRNA network in breast cancer was constructed following three stages. (a) Prediction of lncRNA-miRNA: in order to make lncRNAs and miRNAs map into the interactions successfully, we used the online miRcode (<http://www.mircode.org/>) database.²¹ MiRcode contains 'whole transcriptome' human miRNA target predictions on the basis of the comprehensive GENCODE gene annotation, consisting of more than 10 000 lncRNAs. (b) Prediction of miRNA-mRNA: firstly, the names of DE miRNAs were transformed into human mature miRNA names using starBase v2.0 (<http://starbase.sysu.edu.cn>) online tool. Secondly, three online databases, TargetScan (<http://www.targetscan.org/>),²² miRDB (<http://www.mirdb.org/miRDB/>)²³ and miRTarBase (<http://mirtarbase.mbc.nctu.edu.tw>),²⁴ were used simultaneously for target mRNA prediction. Additionally, we also used the package Venny in Rstudio software to obtain the overlaps for ensuring more credible results. (c) Construction of

lncRNA-miRNA-mRNA ceRNA network: Cytoscape 3.7.0 software was used to construct and visualize the ceRNA network based upon lncRNA-miRNA and miRNA-mRNA pairs.²⁵ Furthermore, for all lncRNAs in the ceRNA network, we used the edgeR package in R software for differential expression. The cut-off threshold was $|\log_2 \text{FC} > 1|$ and adjusted P -value < .05.

2.5 | Gene ontology and pathway enrichment analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID [version 6.8]; <https://david.ncifcr.gov/>)²⁶ has a comprehensive set of functional annotation tools for investigators to more deeply understand the biological meaning behind a large list of genes. DAVID online tool was used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses for overlapping targeted genes. GO consists of the following three parts: biological process (BP), molecular function (MF) and cellular component (CC). Additionally, all significant GO or KEGG terms or genes, with significance of $P < .05$, should be comprised of at least two mRNAs. The results of enrichment analysis were visualized using the package GOplot in R software.

2.6 | Survival analysis

The standard Kaplan-Meier univariate curve analysis was adopted to identify the prognosis of lncRNAs. In addition, we combined analysis with the clinical information of those samples in TCGA database and drew survival curves of all lncRNAs in the ceRNA network. Meanwhile, for overall survival, the high-expression and low-expression cohorts were split for the log-rank test through the package survival in R software. When the P -value of lncRNAs was <.05, it was considered to be statistically significant, which meant that the lncRNA had potential prognosis.

3 | RESULTS

3.1 | Pre-processing of the data sets

Expression data of lncRNAs and mRNAs were collected from 1222 specimens, which were composed of 1109 tumorous and 113 normal samples, and included 112 pairs of matched tumorous and normal samples. Eventually, we obtained a total of 19 983 mRNAs and 14 384 lncRNAs. After obtaining the expression data, we standardized these data sets by using the method of quantile normalization. According to the sum of expression quantity of each gene or lncRNA, we ranked them from largest to smallest and only selected the top 5000 genes and lncRNAs for further analyses. For miRNAs, if their expression levels were higher than 1 in more than 50% of the 102 matched pairs, then they were selected for WGCNA. Next, 665 miRNAs in total were chosen to perform for subsequent analyses.

FIGURE 1 A, Determination of soft-thresholding power in the lncRNAs/ mRNAs WGCNA. B, Determination of soft-thresholding power in the miRNAs WGCNA. Left: Analysis of the scale-free fit index for various soft-thresholding powers (β). Right: Analysis of the mean connectivity for various soft-thresholding powers

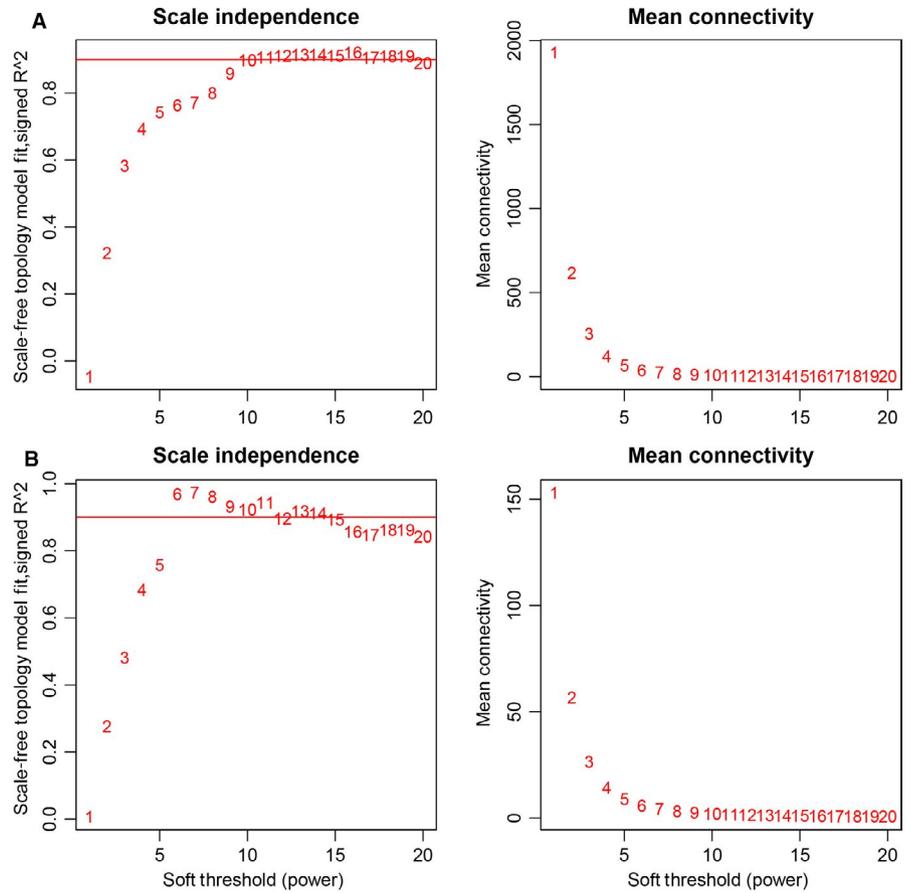
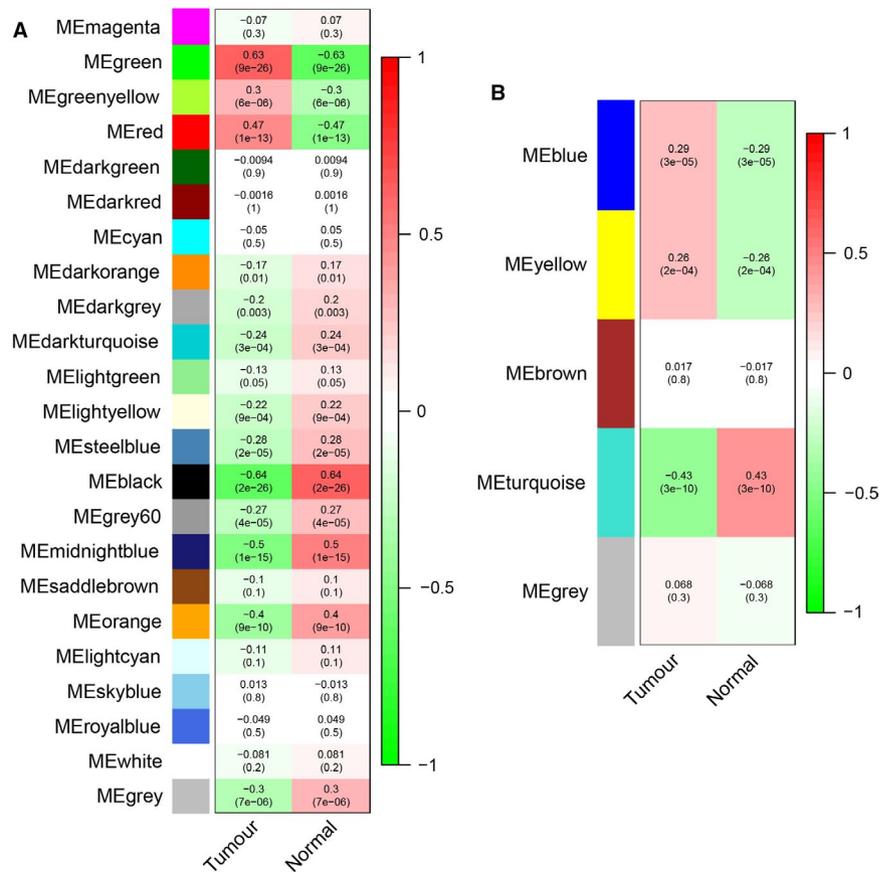


FIGURE 2 A, Clustering dendrogram of lncRNAs and mRNAs. B, Clustering dendrogram of miRNAs. Note: The hierarchical clustering tree was produced by hierarchical clustering based on dissTOM of genes. In the coloured rows below the dendrogram, the two coloured rows represent the original modules and merged modules



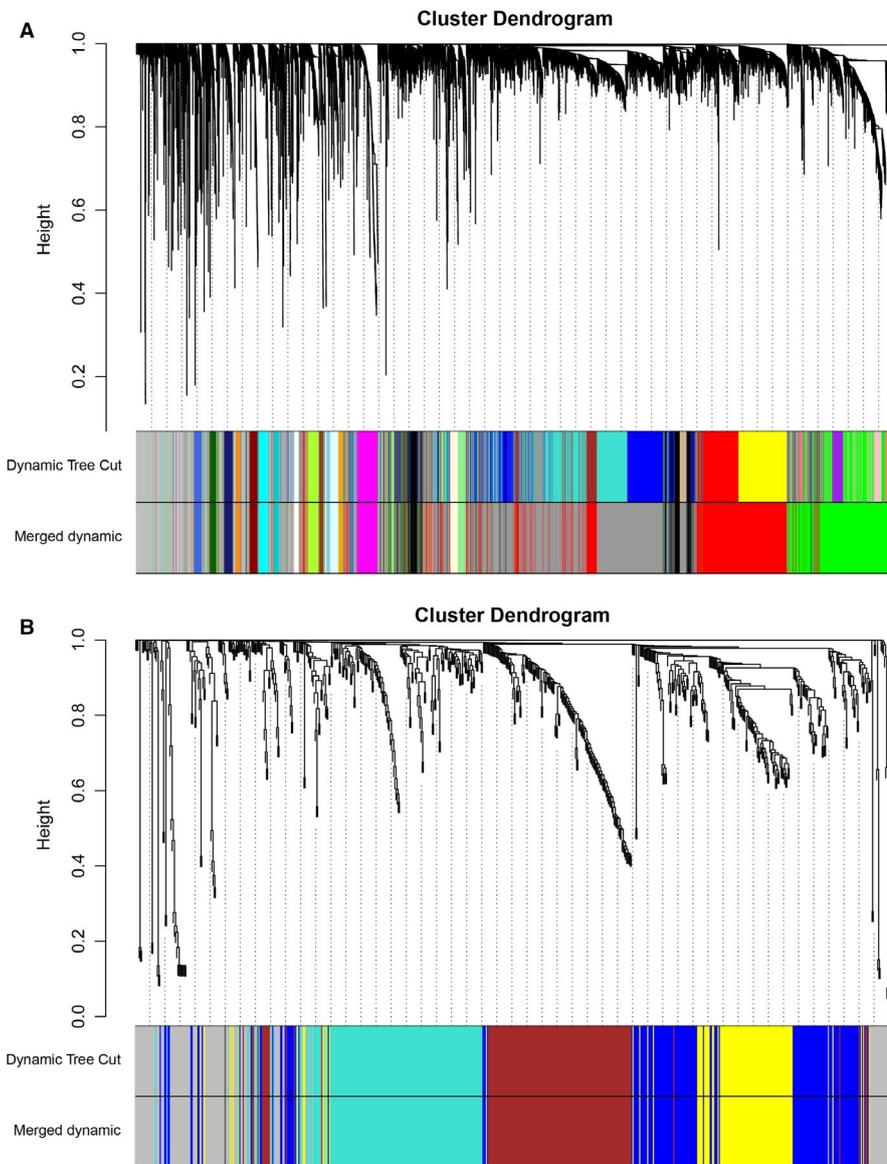


FIGURE 3 A, Module-trait associations of lncRNAs and mRNAs were evaluated by correlations between MEs and clinical traits. B, Module-trait associations of miRNAs were evaluated by correlations between MEs and clinical traits. Note: Each row corresponds to a module eigengene, and each column corresponds to a trait. Each cell contains the corresponding correlation (first line) and *P*-value (second line). The table is colour-coded by correlation according to the colour legend. *P*-values < .05 represent statistical significance

3.2 | Construction of weighted co-expression network and identification of key modules

The expression profiles of 5000 lncRNAs, 5000 mRNAs and 665 miRNAs were obtained for constructing the co-expression network via the package WGCNA in R software. In our study, the most important step is how to select the soft-threshold power. To determine the relative equilibrium between scale independence and mean connectivity, we analysed the network topology with soft-threshold power from 1 to 20. To this end, we eventually confirmed β values of 10 and 6 in lncRNAs/mRNAs and miRNAs co-expression network analysis, respectively (Figure 1A,B). Next, the method of dynamic tree cutting was employed to produce co-expression modules. Afterwards, the minimum number of mRNAs and lncRNAs in each module was set as 30. Additionally, the parameter of MEDissThres was set as 0.25 to merge closely associated modules into larger ones. In the miRNA co-expression network, the minimum number

of each module was set as 20 and the parameter of MEDissThres was set as 0.20. Ultimately, a total of 23 and 5 modules were generated in the lncRNAs/mRNAs and miRNAs co-expression network, respectively (Figure 2A,B). It is worth noting that we calculated and plotted the relation of each module with their corresponding clinical traits. From Figure 3A, we could conclude that the Green module revealed the strongest positive correlation (module-trait weighted correlation = .63) with tumour samples related to the lncRNAs and mRNAs co-expression network. Meanwhile, as shown in Figure 3B, the Blue module has positively correlated (module-trait weighted correlation = .29) with tumour samples related to the miRNAs co-expression network.

3.3 | ceRNA network in breast cancer

Firstly, based on the miRcode online database that matches potential miRNAs with lncRNAs, a total of 257 lncRNA-miRNA pairs

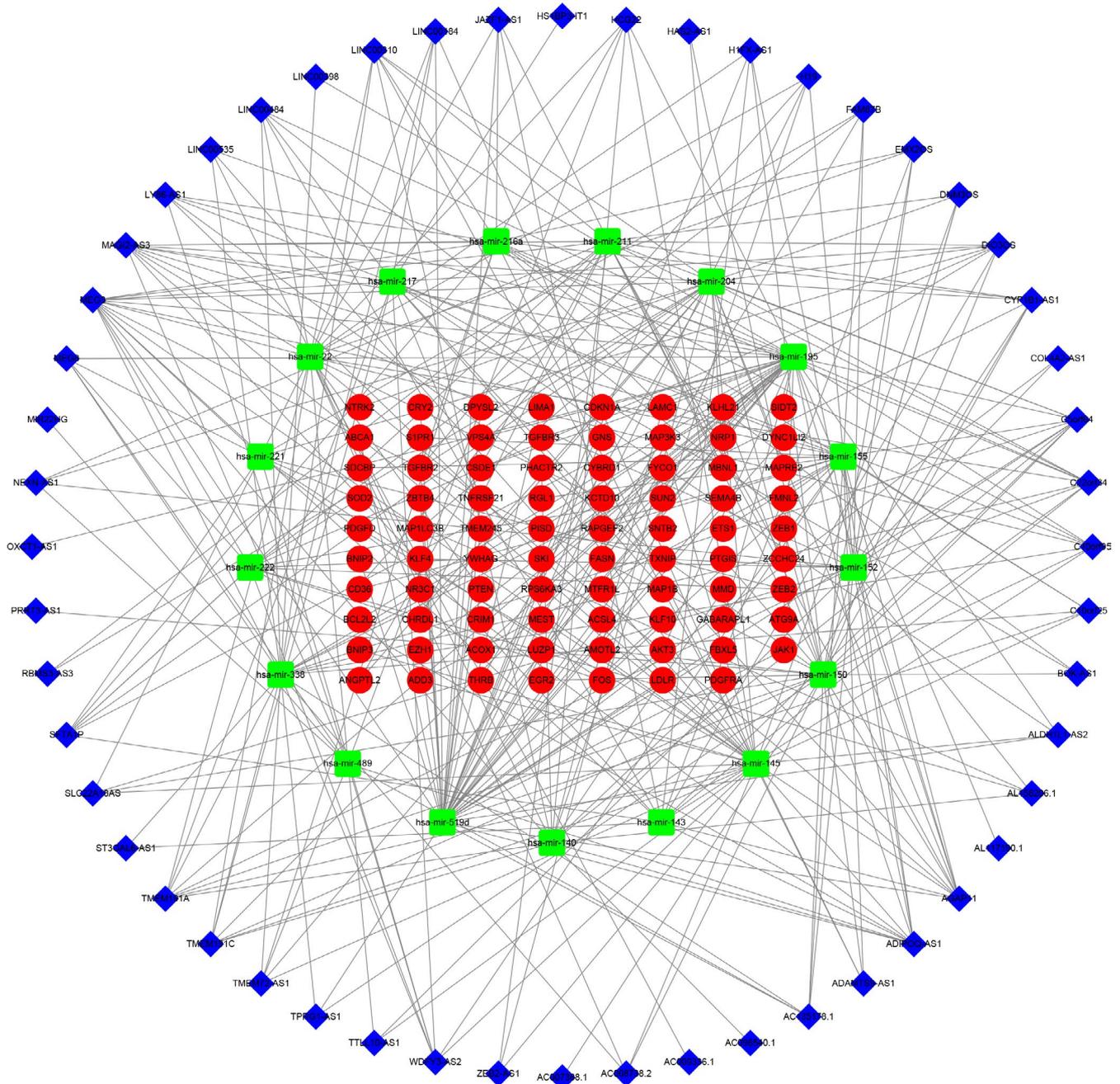


FIGURE 4 Green module of lncRNA and mRNA and Blue module of miRNA ceRNA network. Notes: Blue diamond denotes lncRNA, green square represents miRNA, and red round rectangle represents mRNA

contained 52 lncRNAs (4 up-regulated and 44 down-regulated) and 17 miRNAs. Next, with regard to the target gene predictions of the 17 miRNAs, we used the following online tools: TargetScan, miRDB and miRTarBase, which consisted of 1162 miRNA-mRNA pairs, including 966 target genes. Subsequently, we matched the predicted target gene with the mRNAs in the Green module. Finally, we obtained 79 target mRNAs overall. On the basis of the results above and with the inclusion of 52 lncRNAs, 17 miRNAs and 79 mRNAs, we performed and visualized the lncRNA-miRNA-mRNA network using Cytoscape version 3.7.0 software (Figure 4).

3.4 | Enrichment analysis of Gene Ontology and KEGG pathways

To further clarify the potential biological functions of mRNAs in breast cancer, DAVID online tool was used to perform functional enrichment analysis. From the 79 target mRNAs, a total of 30 GO terms were enriched. In particular, these mRNAs are primarily associated with intracellular signal transduction, positive regulation of programmed cell death, transforming growth factor beta receptor signalling pathway and autophagosome. In addition, from the

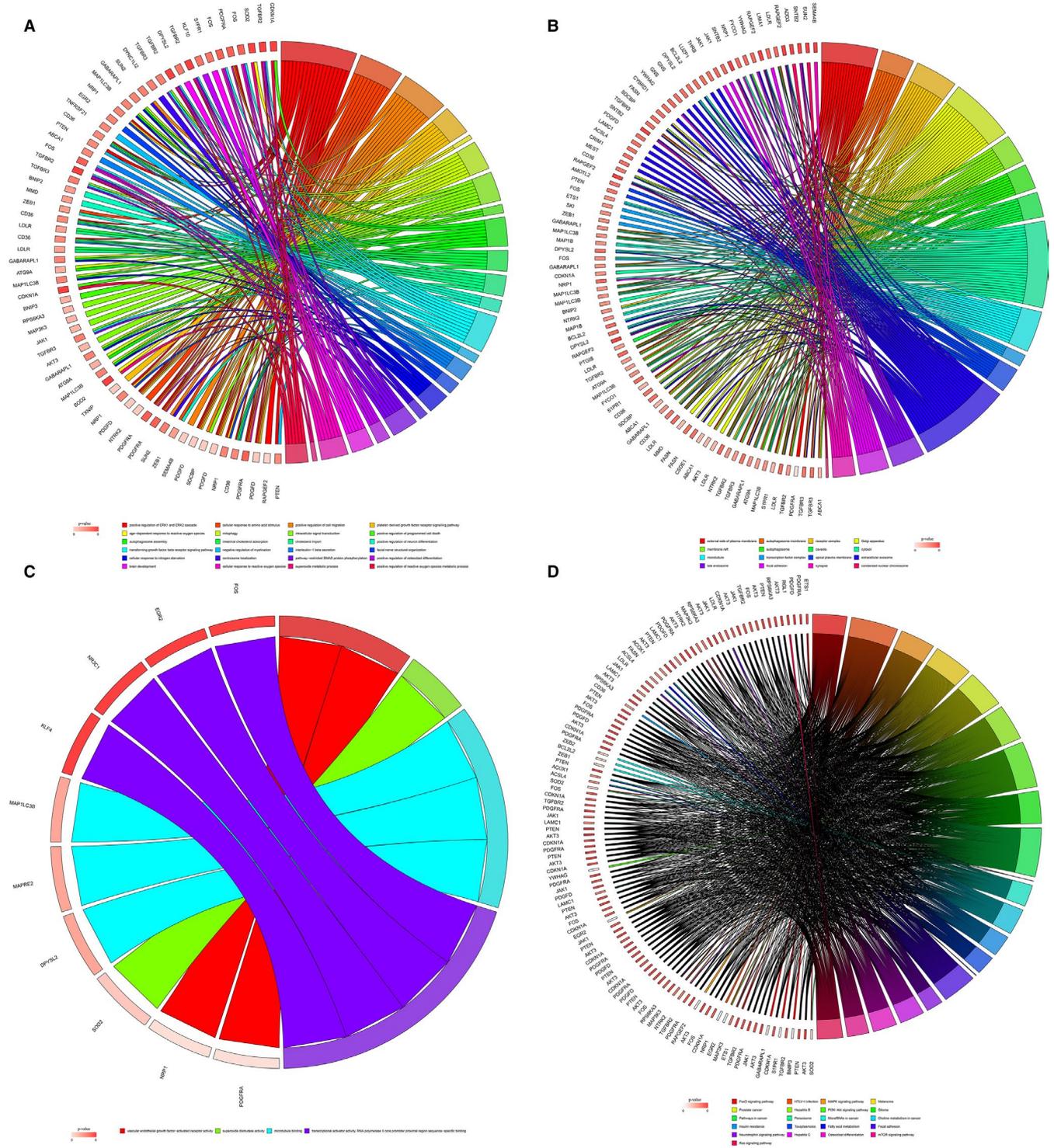


FIGURE 5 Enrichment analyses of overlapped genes. A, Biological process; B, cellular component; C, molecular function; D, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Note: Chord plot displays the relationship between genes and terms

79 target mRNAs, a total of 21 KEGG pathways were enriched. Among these enriched pathways, many are tumour-related, such as MAPK signalling pathway, PI3K-Akt signalling pathway, pathways in cancer, microRNAs in cancer, mTOR signalling pathway and Ras signalling pathway. The visualization results are shown in Figure 5.

3.5 | Survival analysis

Kaplan-Meier and log-rank test based on the package survival in R software were used to identify the correlation between all lncRNAs in the ceRNA network and overall survival using $P < .05$ as a cut-off threshold. Of the 52 lncRNAs, Kaplan-Meier revealed that

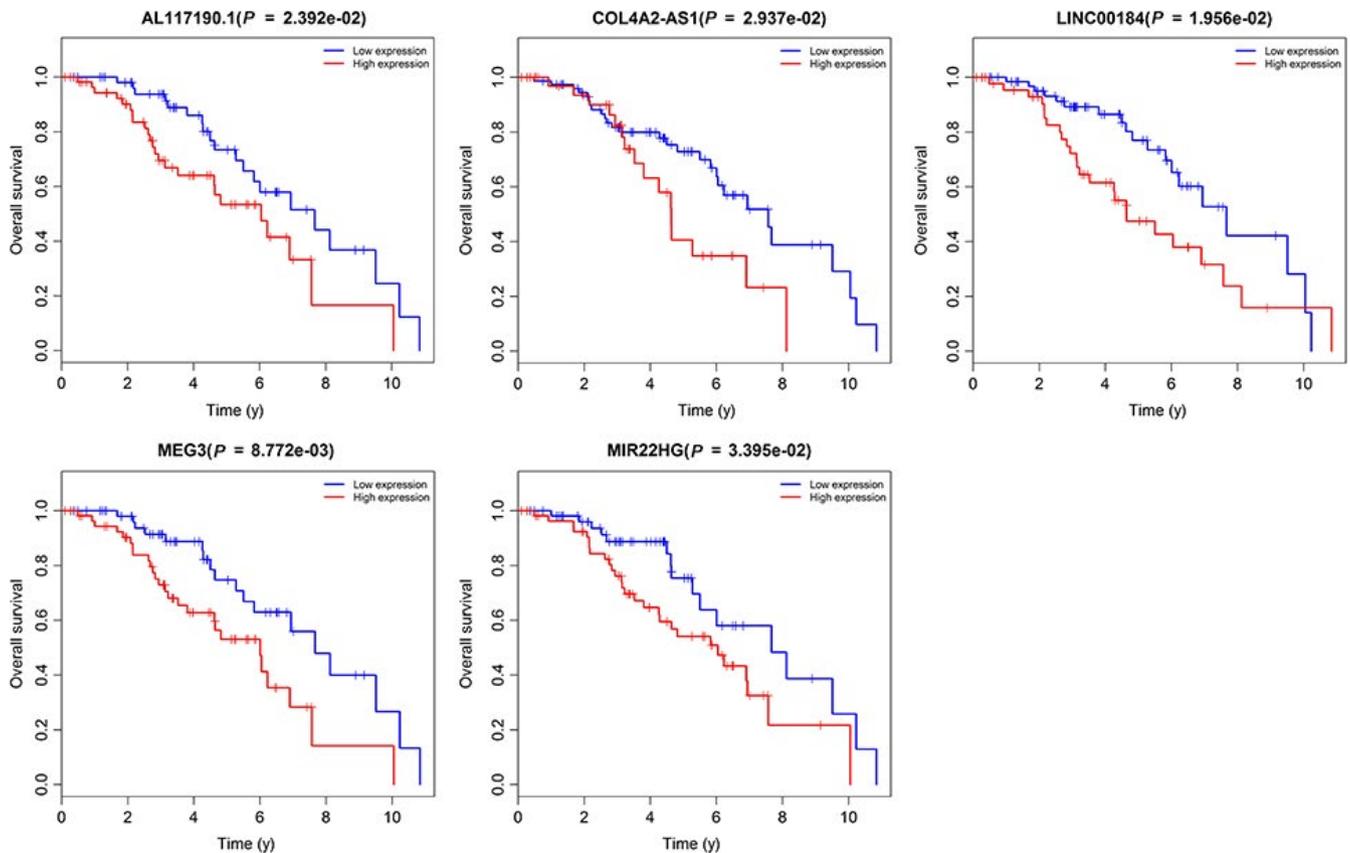


FIGURE 6 Kaplan-Meier survival curves of five lncRNAs in the ceRNA network

5 lncRNAs (AL117190.1, COL4A2-AS1, LINC00184, MEG3 and MIR22HG) were identified as crucial prognostic factors. Notably, low expression of 5 lncRNAs has a better prognosis in patients with breast cancer (Figure 6). When the expression of 5 lncRNAs is high, the survival time is relatively short in patients with breast cancer; thus, these lncRNAs could be related to poor prognosis.

4 | DISCUSSION

Women with breast cancer have a higher rate of distant metastasis and a poorer prognosis than those with other malignant tumours. The current poor overall survival among patients with breast cancer could be due to the lack of more efficient therapeutic targets and prognostic biomarkers. In the present study, we adopted WGCNA to identify modules, including one lncRNAs/mRNAs module (Green) and one miRNA module (Blue) that were most significantly associated with breast cancer tumour status. Then, we constructed a ceRNA network of patients with breast cancer based on lncRNA-miRNA-mRNA interactions to identify potential prognostic lncRNA biomarkers and understand the regulatory mechanisms at a deep level. From this analysis, we conclude that AL117190.1, COL4A2-AS1, LINC00184, MEG3 and MIR22HG act as prognostic biomarkers, whose low expression revealed that patients with breast cancer have better overall survival.

In cancers, lncRNAs appear as a prominent layer of previously unrecognized transcriptional regulation, acting simultaneously as an oncogene and tumour suppressor.²⁷ The increasing experimental evidence supports that lncRNAs have crucial roles in numerous biological processes, such as epigenetic regulation, DNA damage, cell cycle regulation and participation in signal transduction.²⁸ For example, overexpressed lncRNA HOTAIR was found to participate in the progression of malignant breast cancer,²⁹ colon cancer,³⁰ liver cancer³¹ and gastrointestinal stromal tumour.³² In addition, dozens of lncRNAs show potential as prognostic markers and therapeutic targets across multiple classes of tumours.³³ Also, the activity of ceRNA is closely associated with the development of cancers,^{34,35} and in the ceRNA network, lncRNAs could competitively bind miRNAs to regulate target mRNA expression.³⁶

Maternally expressed gene 3 (MEG3), also known as GTL2, FP504 or LINC00023, is an imprinted gene with maternal expression, located in human chromosome 14q32.^{37,38} It was reported that MEG3 plays a key role in regulating many functions in cell growth and development via different mechanisms.³⁹ Increasing evidence showed that, in addition to inhibition of cancer cell growth, MEG3 could also play a vital role in stimulation of p53-mediated transcriptional activation and selective activation of p53 target genes.^{40,41} In relevant research of gastric cancer, Sun et al drew an important conclusion that the low-expression level of MEG3 was more common in tumour tissues rather than in adjacent healthy tissues; also,

the reduction of MEG3 expression level in gastric cancer is related with not only poor survival but also promotion of cell proliferation.⁴² Additionally, low MEG3 expression is not only found in gastric cancers, but other cancers as well, because MEG3 acts as a tumour suppressor gene. In breast cancer, reduced expression of MEG3 was correlated with disease-specific survival of breast cancer.⁴³ Our results indicated that expression of MEG3 was low in breast cancer when compared with normal tissues, and low MEG3 expression represented a better overall survival.

Human miR-22 host gene (MIR22HG) was identified and annotated by The Encyclopedia of DNA Elements (ENCODE) project. MIR22HG was first reported by Torimura et al⁴⁴ and is commonly down-regulated in tumour tissues and participates in the inhibition of cell proliferation. Abnormal expression of MIR22HG is associated with a variety of tumours. In patients with lung cancer, MIR22HG functions as a cancer suppressor gene and was associated with poor patient survival.⁴⁵ In addition, low expression of MIR22HG was correlated with hepatocellular carcinoma progression and showed potential as a novel prognostic biomarker and treatment target.⁴⁶ A relevant study has shown that as a novel biomarker of thyroid cancer, the expression level of MIR22HG is associated with overall survival and prognosis. From these previous studies, we can consider that MIR22HG is closely related to the prognosis of the tumour. Drawn from our study, we found that low-expression level of MIR22HR was related to better overall survival of patients with breast cancer. Thus, MIR22HR could be identified as a prognostic independent predictor. However, in terms of its expression and specific function in breast cancer, little is known at present.

According to the consequences of KEGG enrichment pathways, we could reveal the activation role of lncRNAs from the ceRNA network on key pathways that regulate breast cancer development and progression. They were enriched in biological processes closely correlated with breast cancer, such as MAPK signalling pathway, PI3K-AKT signalling pathway and mTOR signalling pathway. Previous studies have demonstrated that mutations in PI3K-AKT and mTOR pathway are related with cell transformation, tumour occurrence and progression.⁴⁷⁻⁴⁹ Based on KEGG pathways, we could dive into more details of these lncRNAs from the ceRNA network in the future.

In our work, we applied an integrated bioinformatics approach, WGCNA, to establish the lncRNA-miRNA-mRNA ceRNA network for breast cancer. Notably, beginning with TCGA RNA transcript profiles collected from breast cancer specimens, via WGCNA and ceRNA bioinformatics analysis, we identified six lncRNAs, of which three (AL117190.1, COL4A2-AS1 and LINC00184) are unknown. Most importantly, low expression of AL117190.1, COL4A2-AS1 and LINC00184 were correlated with better overall survival and the diagnostic and prognostic values need to be further clarified by independent validation. Nevertheless, the verification steps of the experiments use to generate the data utilized in this study lacked the involvement of these lncRNAs, which was a limitation in our study.

In summary, we have identified five novel lncRNAs (AL117190.1, COL4A2-AS1, LINC00184, MEG3 and MIR22HG) related to prognosis of breast cancer, which could act as underlying prognosis

biomarkers for breast cancer. Our study has contributed to further understanding the molecular mechanism of breast cancer and provided novel insights into the use of breast cancer drugs and prognosis.

ACKNOWLEDGEMENT

This study was funded by the National Natural Science Foundation of China (grant numbers 81473513, 81673799).

CONFLICT OF INTEREST

The authors confirm that there are no conflicts of interest.

AUTHORS CONTRIBUTION

Changgang Sun, Tingting Zhang and Yan Yao were involved in the concept and design of the study. Tingting Zhang and Yan Yao drafted the manuscript. All authors participated in acquisition, analysis and interpretation of the data; revised the manuscript; and read and approved the final version.

DATA AVAILABILITY STATEMENT

Data comes from TCGA database, which is a public open platform.

ORCID

Changgang Sun  <https://orcid.org/0000-0002-6648-3602>

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015;65(2):87-108.
2. (n.d.) ACS. How common is breast cancer? American Cancer Society. 2019.
3. Shi H, Zhang L, Qu Y, Hou L, Wang L, Zheng M. Prognostic genes of breast cancer revealed by gene co-expression network analysis. *Oncol Lett.* 2017;14:4535-4542.
4. Li X, Wu Z, Fu X, Han W. Long noncoding RNAs: insights from biological features and functions to diseases. *Med Res Rev.* 2013;33:517-553.
5. Shi X, Sun M, Liu H, Yao Y, Song Y. Long non-coding RNAs: a new frontier in the study of human diseases. *Cancer Lett.* 2013;339:159-166.
6. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet.* 2009;10:155-159.
7. Fachel AA, Tahira AC, Vilella-Arias SA, et al. Expression analysis and in silico characterization of intronic long noncoding RNAs in renal cell carcinoma: emerging functional associations. *Mol Cancer.* 2013;12:140.
8. Hu Y, Chen HY, Yu CY, et al. A long non-coding RNA signature to improve prognosis prediction of colorectal cancer. *Oncotarget.* 2014;5:2230-2242.
9. Li J, Chen Z, Tian L, et al. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut.* 2014;63:1700-1710.
10. Wang F, Yuan JH, Wang SB, et al. Oncofetal long noncoding RNA PVT1 promotes proliferation and stem cell-like property of

- hepatocellular carcinoma cells by stabilizing NOP2. *Hepatology*. 2014;60:1278-1290.
11. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA Hypothesis: the rosetta stone of a hidden rna language? *Cell*. 2011;146(3):353-358.
 12. Zhang K, Li Q, Kang X, Wang Y, Wang S. Identification and functional characterization of lncRNAs acting as ceRNA involved in the malignant progression of glioblastoma multiforme. *Oncol Rep*. 2016;36:2911-2925.
 13. Ning S, Zhang J, Wang P, et al. Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. *Nucleic Acids Res*. 2016;44:D980-D985.
 14. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
 15. Wang F, Chang Y, Li J, et al. Strong correlation between ASPM gene expression and HCV cirrhosis progression identified by co-expression analysis. *Dig Liver Dis*. 2017;49:70-76.
 16. Zhou Z, Liu S, Zhang M, et al. Overexpression of topoisomerase 2-alpha confers a poor prognosis in pancreatic adenocarcinoma identified by co-expression analysis. *Dig Dis Sci*. 2017;62:2790-2800.
 17. Chen L, Yuan L, Wang Y, et al. Co-expression network analysis identified FCER1G in association with progression and prognosis in human clear cell renal cell carcinoma. *Int J Biol Sci*. 2017;13:1361-1372.
 18. Yuan L, Zeng G, Chen L, et al. Identification of key genes and pathways in human clear cell renal cell carcinoma (ccRCC) by co-expression analysis. *Int J Biol Sci*. 2018;14:266-279.
 19. Spiers H, Hannon E, Schalkwyk LC, et al. Methyloomic trajectories across human fetal brain development. *Genome Res*. 2015;25:338-352.
 20. Liu Q, Jiang C, Xu J, et al. Genome-wide temporal profiling of transcriptome and open chromatin of early cardiomyocyte differentiation derived from hiPSCs and hESCs. *Circ Res*. 2017;121:376-391.
 21. Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics*. 2012;28:2062-2063.
 22. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005;120:15-20.
 23. Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res*. 2015;43:D146-D152.
 24. Hsu SD, Tseng YT, Shrestha S, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res*. 2014;42:D78-85.
 25. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498-2504.
 26. da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44-57.
 27. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discov*. 2011;1:391-407.
 28. Sahu A, Singhal U, Chinnaiyan AM. Long noncoding RNAs in cancer: from function to translation. *Trends Cancer*. 2015;1:93-109.
 29. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010;464:1071-1076.
 30. Kogo R, Shimamura T, Mimori K, et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Can Res*. 2011;71:6320-6326.
 31. Yang Z, Zhou L, Wu LM, et al. Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. *Ann Surg Oncol*. 2011;18:1243-1250.
 32. Niinuma T, Suzuki H, Nojima M, et al. Upregulation of miR-196a and HOTAIR drive malignant character in gastrointestinal stromal tumors. *Can Res*. 2012;72:1126-1136.
 33. Qi P, Du X. The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. *Mod Pathol*. 2013;26(2):155-165.
 34. Tang J, Zhuo H, Zhang X, et al. A novel biomarker linc00974 interacting with KRT19 promotes proliferation and metastasis in hepatocellular carcinoma. *Cell Death Dis*. 2014;5:e1549.
 35. Rutnam ZJ, Du WW, Yang W, Yang X, Yang BB. The pseudogene TUSC2P promotes TUSC2 function by binding multiple microRNAs. *Nat Commun*. 2014;5:2914.
 36. Wang P, Chen D, Ma H, Li Y. LncRNA MEG3 enhances cisplatin sensitivity in non-small cell lung cancer by regulating miR-21-5p/SOX7 axis. *Onco Targets Ther*. 2017;10:5137-5149.
 37. Benetatos L, Vartholomatos G, Hatzimichael E. MEG3 imprinted gene contribution in tumorigenesis. *Int J Cancer*. 2011;129:773-779.
 38. Zhou Y, Zhang X, Klubanski A. MEG3 noncoding RNA: a tumor suppressor. *J Mol Endocrinol*. 2012;48:R45-53.
 39. da Rocha ST, Edwards CA, Ito M, Ogata T, Ferguson-Smith AC. Genomic imprinting at the mammalian Dlk1-Dio3 domain. *Trends Genet*. 2008;24:306-316.
 40. Zhang X, Zhou Y, Mehta KR, et al. A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells. *J Clin Endocrinol Metab*. 2003;88:5119-5126.
 41. Zhou Y, Zhong Y, Wang Y, et al. Activation of p53 by MEG3 non-coding RNA. *J Biol Chem*. 2007;282:24731-24742.
 42. Sun M, Xia R, Jin F, et al. Downregulated long noncoding RNA MEG3 is associated with poor prognosis and promotes cell proliferation in gastric cancer. *Tumour Biol*. 2014;35:1065-1073.
 43. Zhang Y, Wu J, Jing H, Huang G, Sun Z, Xu S. Long noncoding RNA MEG3 inhibits breast cancer growth via upregulating endoplasmic reticulum stress and activating NF-kB and p53. *J Cell Biochem*. 2019;120(4):6789-6797.
 44. Tani H, Onuma Y, Ito Y, Torimura M. Long non-coding RNAs as surrogate indicators for chemical stress responses in human-induced pluripotent stem cells. *PLoS ONE*. 2014;9:e106282.
 45. Su W, Feng S, Chen X, et al. Silencing of long noncoding RNA MIR22HG triggers cell survival/death signaling via oncogenes YBX1, MET, and p21 in lung cancer. *Can Res*. 2018;78:3207-3219.
 46. Li L, Zhang D, Wu D. Long non-coding RNA miR22HG represses hepatocellular carcinoma cell invasion by deriving miR-22 and targeting HMGB1. *Gastroenterology*. 2016;150:S1045-S1046.
 47. Zhao JJ, Liu Z, Wang L, Shin E, Loda MF, Roberts TM. The oncogenic properties of mutant p110alpha and p110beta phosphatidylinositol 3-kinases in human mammary epithelial cells. *Proc Natl Acad Sci USA*. 2005;102:18443-18448.
 48. Isakoff SJ, Engelman JA, Irie HY, et al. Breast cancer-associated PIK3CA mutations are oncogenic in mammary epithelial cells. *Can Res*. 2005;65:10992-11000.
 49. Kang S, Bader AG, Vogt PK. Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic. *Proc Natl Acad Sci USA*. 2005;102:802-807.

How to cite this article: Yao Y, Zhang T, Qi L, et al. Integrated analysis of co-expression and ceRNA network identifies five lncRNAs as prognostic markers for breast cancer. *J Cell Mol Med*. 2019;23:8410-8419. <https://doi.org/10.1111/jcmm.14721>