



Construction of ceRNA Networks Associated With CD8 T Cells in Breast Cancer

Zhilin Chen^{1,2†}, Ruifa Feng^{3†}, Ulf Dietrich Kahlert^{4†}, Zhitong Chen², Luz Angela Torres-dela Roche², Amr Soliman², Chen Miao^{5*†}, Rudy Leon De Wilde^{2*†} and Wenjie Shi^{2,4*†}

OPEN ACCESS

Edited by:

Zhenjian Zhuo,
Guangzhou Medical University, China

Reviewed by:

Kejun Wang,
Henan Agricultural University, China
Yipeng Xu,
Zhejiang Cancer Hospital, China

*Correspondence:

Chen Miao
cm_njmu@163.com
Rudy Leon De Wilde
rudy-leon.dewilde@pius-hospital.de
Wenjie Shi
wenjie.shi@uni-oldenburg.de

[†]These authors have contributed equally to this work and share first authorship

[‡]These authors have contributed equally to this work and share last authorship

Specialty section:

This article was submitted to
Cancer Genetics,
a section of the journal
Frontiers in Oncology

Received: 24 February 2022

Accepted: 12 April 2022

Published: 09 June 2022

Citation:

Chen Z, Feng R, Kahlert UD,
Chen Z, Torres-dela Roche LA,
Soliman A, Miao C, De Wilde RL
and Shi W (2022) Construction of
ceRNA Networks Associated With
CD8 T Cells in Breast Cancer.
Front. Oncol. 12:883197.
doi: 10.3389/fonc.2022.883197

¹ Department of Breast and Thoracic Oncological Surgery, The First Affiliated Hospital of Hainan Medical University, Haikou, China, ² University Hospital for Gynecology, Pius-Hospital, University Medicine Oldenburg, Oldenburg, Germany, ³ Breast Center of The Second Affiliated Hospital of Guilin Medical University, Guilin, China, ⁴ Molecular and Experimental Surgery, University Clinic for General-, Visceral- and Vascular Surgery, University Medicine Magdeburg and Otto-von Guericke University, Magdeburg, Germany, ⁵ Department of Pathology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

Background: The infiltration of CD8 T cells is usually linked to a favorable prognosis and may predict the therapeutic response of breast cancer patients to immunotherapy. The purpose of this research is to investigate the competing endogenous RNA (ceRNA) network correlated with the infiltration of CD8 T cells.

Methods: Based on expression profiles, CD8 T cell abundances for each breast cancer (BC) patient were inferred using the bioinformatic method by immune markers and expression profiles. We were able to extract the differentially expressed RNAs (DEmRNAs, DEmiRNAs, and DElncRNAs) between low and high CD8 T-cell samples. The ceRNA network was constructed using Cytoscape. Machine learning models were built by lncRNAs to predict CD8 T-cell abundances. The lncRNAs were used to develop a prognostic model that could predict the survival rates of BC patients. The expression of selected lncRNA (XIST) was validated by quantitative real-time PCR (qRT-PCR).

Results: A total of 1,599 DElncRNAs, 89 DEmiRNAs, and 1,794 DEmRNAs between high and low CD8 T-cell groups were obtained. Two ceRNA networks that have positive or negative correlations with CD8 T cells were built. Among the two ceRNA networks, nine lncRNAs (MIR29B2CHG, NEAT1, MALAT1, LINC00943, LINC01146, AC092718.4, AC005332.4, NORAD, and XIST) were selected for model construction. Among six prevalent machine learning models, artificial neural networks performed best, with an area under the curve (AUC) of 0.855. Patients from the high-risk category with BC had a lower survival rate compared to those from the low-risk group. The qRT-PCR results revealed significantly reduced XIST expression in normal breast samples, which was consistent with our integrated analysis.

Conclusion: These results potentially provide insights into the ceRNA networks linked with T-cell infiltration and provide accurate models for T-cell prediction.

Keywords: ceRNA, T cell, breast cancer, lncRNA, target

INTRODUCTION

Breast cancer (BC) is by far the most frequently diagnosed kind of cancer and the fifth greatest cause of cancer-related deaths globally (1). About 2.3 million BC patients and 0.68 million deaths were reported in 2020 (1). BC is a very heterogeneous illness, consisting of several biological subgroups with distinct genetic features and therapeutic implications. In 2000, BC was first separated into the luminal, basal-like, HER2+, and normal breast-like subtypes (2). Each subtype has its own set of molecular markers, outcomes, clinical characteristics, and therapy responses. For example, the prognosis of patients with the luminal subtype is relatively good (3), and the primary treatments are mainly based on endocrine therapy (4). However, the prognosis of BC samples in the basal-like subtype is very poor (5), and there is a dearth of effective therapies available (6). Despite advancements in BC treatment, advanced or metastatic BC continues to have a dismal survival rate of 25% (7). Thus, novel therapeutic agents are needed.

Immunotherapy has been utilized to treat advanced breast cancer with metastases in the past few years. Immune-checkpoint inhibitors aim to block suppressive immune receptors and activate dysfunctional T cells, including CD8+ T cells. In a randomized, double-blind, and placebo-controlled clinical trial that contained 1,174 TNBC patients, the complete response rate was 63% (95% CI, 59.5%–66.4%) for patients in the group of pembrolizumab/chemotherapy compared with 56% (95% CI: 50.6%–60.6%) for the group of chemotherapy alone (8). The Food and Drug Administration authorized the use of pembrolizumab with chemotherapy for TNBC patients on July 26, 2021, based on these clinical study findings. This approval, however, brought up a new issue: how to distinguish between cancer patients who are immunotherapy sensitive and those who are immunotherapy insensitive. CD8 T cells are the most potent effectors in the anticancer immune response, and they are the foundation of cancer immunotherapy (9). Several indicators, especially CD8 T lymphocytes (TILs), have been identified to predict immunotherapy response (10). Thus, the prediction of CD8 T cells will contribute to the screening of immunotherapy-responsive BC patients.

On the other hand, lncRNAs are a novel class of ncRNAs that have more than 200 nucleotides in length. Dysregulation of lncRNAs has been implicated in the development of BC, involving cell growth, apoptosis, migration, and therapy resistance regulation (11). Salmena et al. proposed the competitive endogenous RNAs (ceRNAs) hypothesis (12) whereby RNA transcripts can communicate with each other *via* miRNA response elements (MREs). By competitively interacting with miRNAs, the lncRNAs can act as ceRNAs and positively regulate mRNAs. These ncRNAs act as ceRNAs to modulate mRNA expression and regulate protein levels, which contributes to the occurrence and development of tumors (13). Studies have shown that each miRNA can control the transcriptional expression levels of hundreds of proteins. Each mRNA contains different MREs, and so it may be targeted by multiple miRNAs (14). Numerous investigations have shown

that ceRNAs could play critical roles in the genesis, progression, and outcome of BC (15). However, the ceRNA network that could influence the CD8 T cells has not been extensively studied so far.

Using bioinformatic analysis, our team discovered numerous lncRNA, miRNA, and mRNA, and constructed ceRNA networks associated with T cells. Machine learning models were constructed to predict the T cells by hub lncRNAs from the ceRNA networks. A risk score model with good prognostic predictive accuracy was constructed by hub lncRNAs from the ceRNA networks. These models may hold considerable promise for personalized therapy and prognosis prediction in BC patients.

MATERIALS AND METHODS

Breast Cancer Dataset

The level 3 mRNA expression profiles of BC were obtained from The Cancer Genome Atlas (TCGA) by the “TCGAbiolinks” R package (16). A total of 1,072 BC samples were found with available lncRNA and mRNA expression profiles, and 1,066 BC samples were found with available miRNA expression profiles. The lncRNA and mRNA expression data were converted from “fragments per kilobase million” into “transcripts per kilobase million”. We gathered clinical data on 1,087 BC patients, including their overall survival and progression-free survival data.

Estimation of Immune Cell Types

The single-sample Gene Set Enrichment Analysis (ssGSEA) model was used to calculate the amounts of immune cells and to convert the genomic data into the values of 28 different categories of human immune cells (17), which included the cells that were thought to be anti-tumor such as “activated CD4 T cell” and “activated CD8 T cell”. Several pro-tumor immunity cells, including “regulatory T cell” and “type-2 T-helper cell”, were also obtained. Two groups of BC patients were established according to the median value of “activated CD8 T cell”. The microenvironment cell populations counter (MCP-counter) method (18), which allows the accurate quantification of immune and stromal cells by genomic data, was used in this study. The ESTIMATE, a tool for predicting tumor purity and the presence of infiltrating stromal/immune cells in tumor tissues using gene expression data, was used (19).

Screening of DElncRNAs, DEmiRNAs, and DEMRNAs Between Low and High CD8 T Cell Groups

The expression profiles of 8,633 lncRNAs were obtained using the annotation profile from the GENCODE program. After excluding 25% of lncRNAs having the lowest mean expression value, 6,475 lncRNAs were retained for further study. The expression profiles of 2,236 miRNAs were obtained by the miRNA annotation from the miRBaseVersions.db R package. After removing 25% of lncRNAs with the lowest mean

expression value, 1,677 miRNAs were retained for further investigation. The expression profiles of 19,618 mRNAs were obtained by the mRNA annotation of the GRCh38 project. After excluding 25% of mRNAs with the lowest mean expression value, 14,714 mRNAs were retained for further study.

The differentially expressed RNAs (DElncRNAs, DEmiRNAs, and DEMRNAs) between two T-cell groups were calculated by the R package edgeR (20). DElncRNAs and DEmiRNAs were filtered using p-value of 0.05 and $|\log_2\text{fold change}| > 0.3$ as criteria. DEMRNAs were filtered using p-value of 0.05 and $|\log_2\text{fold change}| > 0.5$ as cutoffs. Upregulation mRNAs (UmRNAs), down-regulation mRNAs (DmRNAs), upregulation miRNAs (UmiRNAs), downregulation miRNAs (DmiRNAs), upregulation lncRNAs (UlnRNAs), and downregulation lncRNAs (DlnRNAs) were defined by their expression values in the high CD8 T-cell group.

Constructing lncRNA–miRNA–mRNA ceRNA Networks and Enrichment Analysis

Based on ceRNA theory, we constructed lncRNA–miRNA–mRNA ceRNA networks by lncRNA–miRNA and miRNA–mRNA correlations. First, we downloaded all the lncRNA–miRNA and miRNA–mRNA interactions that were downloaded from StarBase (<http://starbase.sysu.edu.cn/>) (21). The first ceRNA network that has a positive correlation value with CD8 T cells was constructed by selecting ncRNA–miRNA and miRNA–mRNA interactions that contained UmRNAs, DmiRNAs, and UlnRNAs. The second ceRNA network that has a negative correlation value with CD8 T cells was constructed by selecting ncRNA–miRNA and miRNA–mRNA interactions that contained DmRNAs, UmiRNAs, and DlnRNAs. In this study, Cytoscape software was utilized to visualize the network.

Since functional enrichment analysis is crucial for interpreting high-throughput omics data in life science, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed in R using the function of clusterProfiler (p-value < 0.05) (22). The lncRNAs were ranked by the number of ncRNA–miRNA interactions. The top 4 lncRNAs were selected from the first ceRNA network, and the top 5 lncRNAs were selected from the second ceRNA network. These nine lncRNAs were defined as the hub lncRNAs.

Machine Learning Models for Prediction of CD8 T Cells

Using the expression profiles of nine hub lncRNAs, machine learning models were built for the prediction of CD8 T cells. BC patients were equally and randomly divided into training and testing samples. Different machine learning models, including decision tree (DT), gradient boosting machines (GBM), generalized linear models (GLM), artificial neural networks (ANN), random forest (RF), and support vector machine (SVM), were chosen to build models to predict the CD8 T-cell group by the R “Caret” package (23). The parameters for each machine learning model were determined by the fivefold cross-validation. In order to

evaluate the robustness and prediction accuracy of constructed models in BC, the “ROCR” package was used to calculate the value of the area under the ROC curve (AUC).

Cox Regression Model for Prediction of Prognosis

These nine hub lncRNAs were used to construct the Cox model. The coefficients for nine lncRNAs were obtained. The risk score of the samples was calculated by the coefficients and lncRNA expression profiles. Based on the median value, BC individuals were categorized into high- and low-risk classes. The correlation analysis of clinical information includes overall survival, age, gender, stage, and T, N, and M classifications in the TNM system. The log-rank test was used for the Kaplan–Meier curves of patient survival analyses.

Real-Time Quantitative PCR

Total RNA was extracted from patients’ breast cancer at the First Affiliated Hospital of Nanjing Medical University using TRIzol reagent (Invitrogen, CA, USA). Isolated RNA was reverse transcribed into cDNA using HiScript II Q RT SuperMix for qPCR (Vazyme Biotech Co., Ltd. Nanjing, China) following standard protocols. Real-time quantitative PCR (qPCR) was performed with synthetic primers and ChamQ SYBR qPCR Master Mix (Vazyme Biotech Co., Ltd. Nanjing, China) with a Quant Studio 5 Real-Time PCR Detection System (Thermo Fisher Scientific, MA, USA). The relative expression levels of XIST were calculated and quantified with the $2^{-\Delta\Delta Ct}$ method after normalization with the reference β -actin expression. All primers used are listed in **Table 1**.

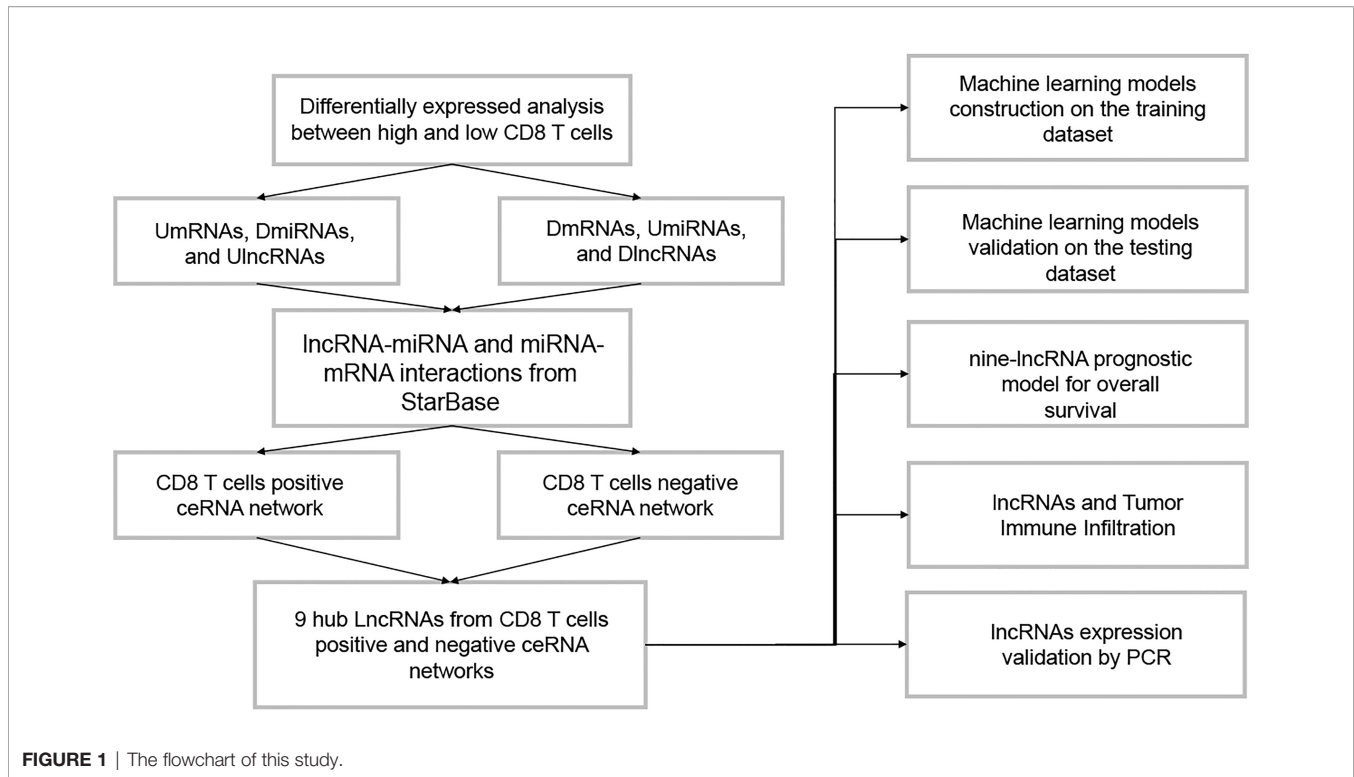
RESULTS

DElncRNAs, DEmiRNAs, and DEMRNA in Low and High CD8 T Cell Groups

The flowchart of this study is presented in **Figure 1**. By the ssGSEA and MCP-counter methods, values of CD8 T cells were calculated. High values of CD8 T cells were correlated with better BC prognosis (**Supplementary Figure 1**). Then, the BC samples were divided into two groups (low and high) by the median value of CD8 T cells from the ssGSEA algorithm. Based on the criteria of p-value < 0.05 and $|\log_2FC| > 0.3$, 1,599 DElncRNAs (811 downregulation and 788 upregulation, **Figure 2A**) and 89 DEmiRNAs (11 downregulation and 78 upregulation, **Figure 2B**) between low and high CD8 T-cell groups were obtained. Based on the criteria of p-value < 0.05 and $|\log_2FC| >$

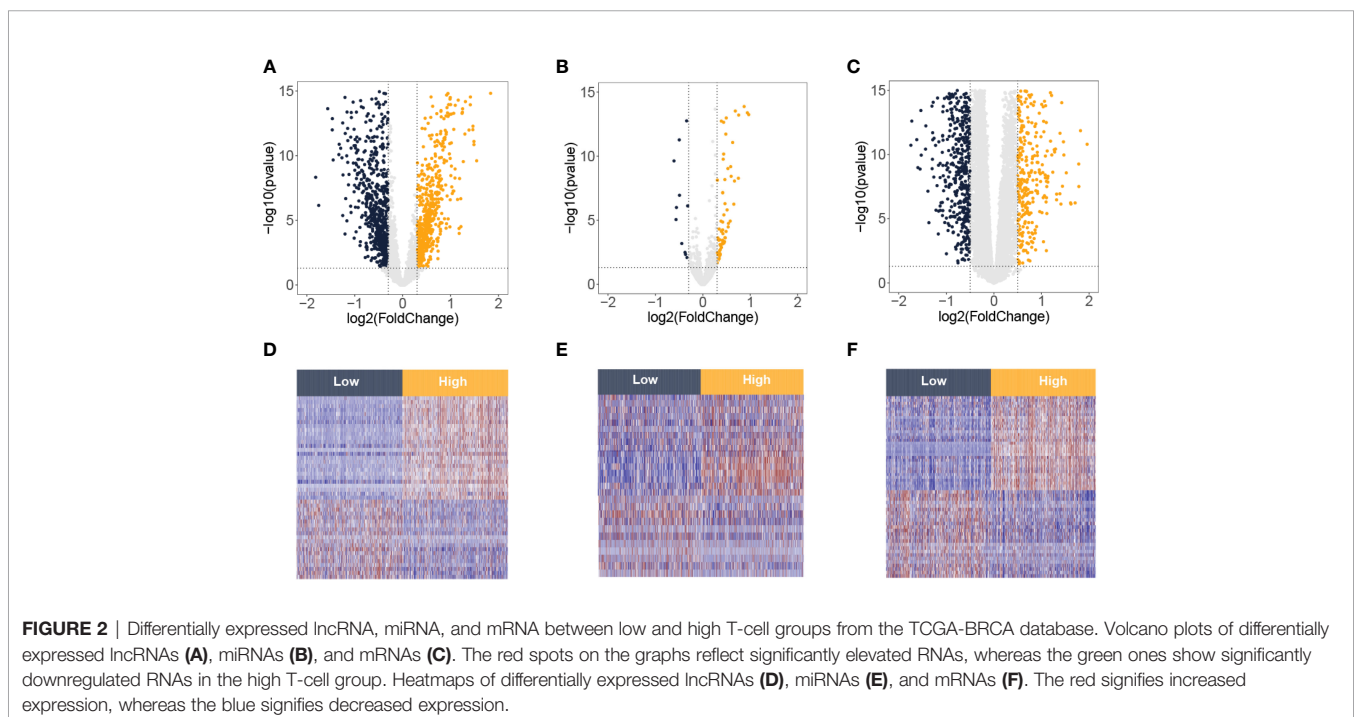
TABLE 1 | Sequences of primers for real-time quantitative polymerase chain reaction.

Gene	Sequence (5' → 3')
Xist	
Forward	TGGATAGAGGACCCAAGCGA
Reverse	CAAGACTGGCCAGGCATAA
β -actin	
Forward	AGGATTCTATGTGGCGAC
Reverse	ATAGCACAGCCTGGATAGCAA



0.5, 1,794 DEmRNAs (659 downregulation and 1,135 upregulation, **Figure 2C**). Heatmaps of the top DElncRNAs, DEmiRNAs, and DEmRNAs are presented in **Figures 2D–F**. Upregulation mRNAs (UmRNAs), downregulation

mRNAs (DmRNAs), upregulation miRNAs (UmiRNAs), downregulation miRNAs (DmiRNAs), upregulation lncRNAs (UlncRNAs), and downregulation lncRNAs (DlncRNAs) were defined by their expression values in the high CD8 T-cell group.



Construction of the ceRNA Network

We downloaded the lncRNAs–miRNAs and miRNAs–mRNAs interactions from the StarBase database. Due to the ceRNA theory, lncRNAs and mRNAs have a positive regulatory interaction, while miRNAs and mRNAs have a negative regulatory interaction. Then, lncRNA–miRNA and miRNA–mRNA interactions that contain UmRNAs, DmiRNAs, and UlncRNAs were used to construct a CD8 T-cell positive ceRNA network (Figure 3A). The lncRNA–miRNA and miRNA–mRNA interactions that contain DmRNAs, UmiRNAs, and DlncRNAs were used to construct a CD8 T-cell negative ceRNA network (Figure 3B).

Gene Set Enrichment Analysis

To outline the potential function of the genes in the ceRNA networks, the pathways from the CD8 T-cell positive ceRNA network and the CD8 T-cell negative ceRNA network were obtained. Top GO terms and KEGG pathways in the CD8 T-cell positive ceRNA network included many immune pathways such as “antigen processing and presentation” (Supplementary Table 1). Top GO terms and KEGG pathways in the CD8 T-cell negative ceRNA network contained many cancer pathways such as “proteoglycans in cancer” (Supplementary Table 2).

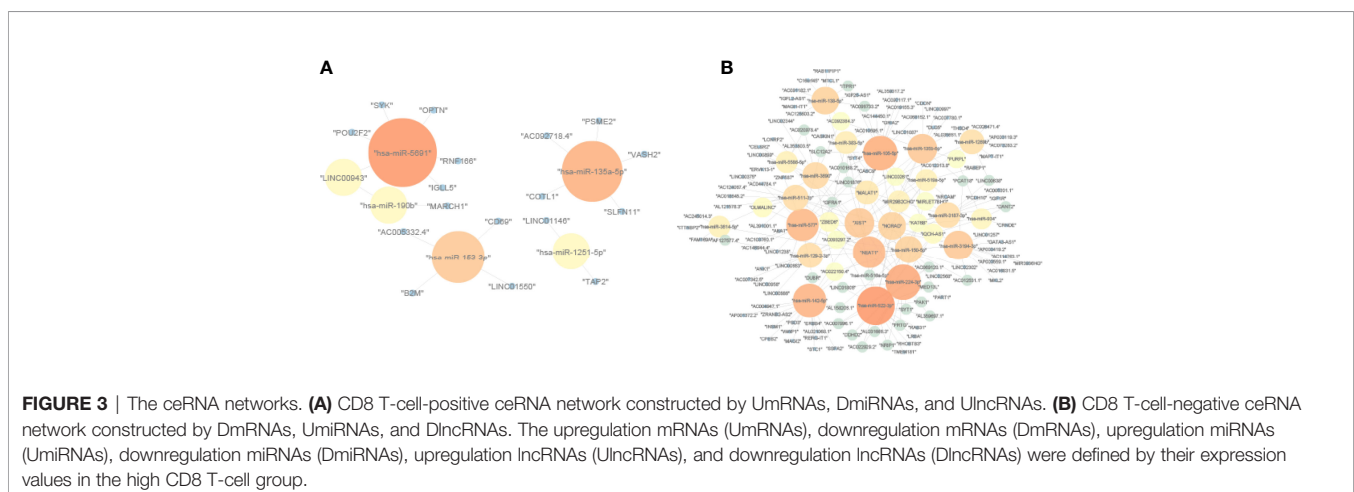
Machine Learning Models for the Prediction of CD8 T Cell Groups

To predict the CD8 T-cell group of BC samples, we constructed a model based on the hub lncRNAs from the CD8 T-cell-positive/negative ceRNA networks. The top 4 lncRNAs from the CD8 T-cell-positive ceRNA network, and the top 5 lncRNAs from the CD8 T-cell-negative ceRNA network were selected. These nine lncRNAs (NEAT1, XIST, NORAD, MALAT1, MIR29B2CHG, LINC00943, AC005332.4, AC092718.4, and LINC01146) were defined as the hub lncRNAs, and the expression profiles of these hub lncRNAs were kept for further analysis. The expression profiles were evenly and randomly split into a training (50%) and a testing set (50%). Fivefold cross-validation was conducted on the training set to select the best parameters for DT, GBM, GLM, ANN, RF, and SVM. The models were trained on the training set with the selected best parameters. The log2CP value from DT was selected as “−9.65”,

since it achieved the highest AUC value of 0.794 in the fivefold cross-validation (Figure 4A). The interaction depth value from GBM was selected as “7”, since it achieved the highest AUC value of 0.857 in the fivefold cross-validation (Figure 4B). The alpha value from GLM was selected as “0.7”, since it achieved the highest AUC value of 0.838 in the fivefold cross-validation (Figure 4C). The size value from ANN was selected as “11”, since it achieved the highest AUC value of 0.837 in the fivefold cross-validation (Figure 4D). The mtry value from RF was selected as “3”, since it achieved the highest AUC value of 0.842 in the fivefold cross-validation (Figure 4E). The log2C value from SVM was selected as “−1”, since it achieved the highest AUC value of 0.843 in the fivefold cross-validation (Figure 4F). Prediction accuracy was then measured by the AUC value calculated on the test set. On the testing set, the AUC values for DT, GBM, GLM, ANN, RF, and SVM were 0.742, 0.854, 0.848, 0.855, 0.852, and 0.843 (Figures 5A–F). The calculation process of the final model for DT is presented in Supplementary Figure 2.

Establishment of the lncRNA Prognostic Model

Thereafter, we developed a nine-lncRNA prognostic model for overall survival prediction of BC patients. The risk score = $(0.61 * \text{MIR29B2CHG}) + (0.027 * \text{NEAT1}) + (0.006 * \text{MALAT1}) + (-1.21 * \text{LINC00943}) + (0.24 * \text{LINC01146}) + (2.1 * \text{AC092718.4}) + (-1.02 * \text{AC005332.4}) + (0.99 * \text{NORAD}) + (0.193 * \text{XIST})$. The distribution of prognosis, including survival time and status in risk score groups, is presented (Figure 6A). Additionally, the increased expression of AC005332.4 and LINC00943 was observed in the low-risk BC samples. On the other hand, MIR29B2CHG, MALAT1, XIST, NORAD, and AC092718.4 expression levels were greater in the high-risk BC samples (Figure 6B). Low-risk individuals in the BC cohort survived longer than high-risk ones when it came to OS (p -value = 0.001, Figure 6C). MIR29B2CHG, NEAT1, MALAT1, AC005332.4, NORAD, and XIST were elevated in normal breast samples than in breast cancer samples (Figure 6D). LINC01146 and AC092718.4 were more elevated in breast cancer samples than in normal breast samples (Figure 6D).



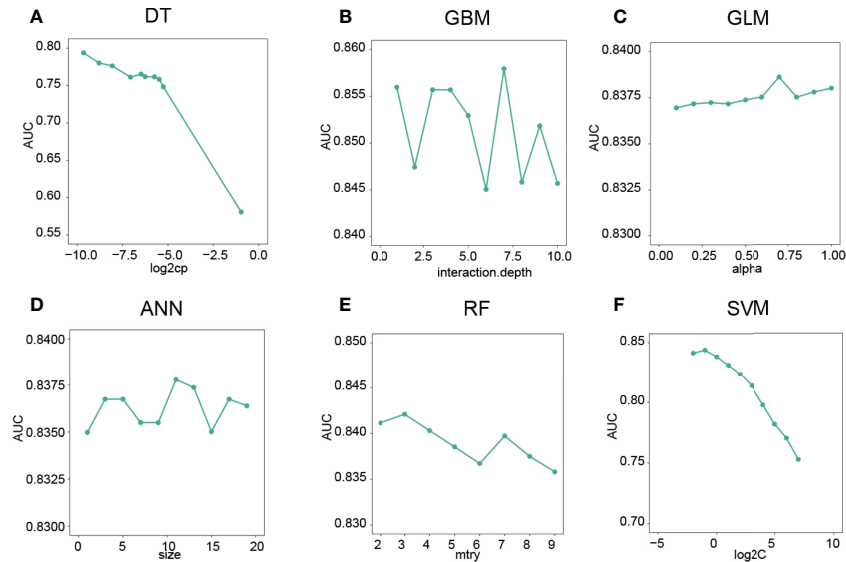


FIGURE 4 | Optimal parameter selection for machine learning models. **(A)** The “log2CP” and corresponding AUC values of the DT algorithm. **(B)** The “interaction depth value” and corresponding AUC values of the GBM algorithm. **(C)** The “alpha value” and corresponding AUC values of the GLM algorithm. **(D)** The “Size” and corresponding AUC values of the ANN algorithm. **(E)** The mtry’ and corresponding AUC values of the RF algorithm. **(F)** The “log2C” value and corresponding AUC values of the SVM algorithm.

Relationship Between Hub lncRNAs and Tumor Immune Infiltration

We analyzed the associations among nine lncRNAs (NEAT1, XIST, NORAD, MALAT1, MIR29B2CHG, LINC00943, AC005332.4, AC092718.4, and LINC01146), risk score, tumor purity, and immune cell infiltration (**Figure 7A**). NEAT1, XIST,

NORAD, MALAT1, and MIR29B2CHG were significantly negatively related to CD8 T cells, cytotoxic lymphocytes, and T cells. These five lncRNAs were significantly positively associated with fibroblasts, endothelial cells, and neutrophils. On the other hand, LINC00943, AC005332.4, AC092718.4, and LINC01146 were significantly positively associated with most types of

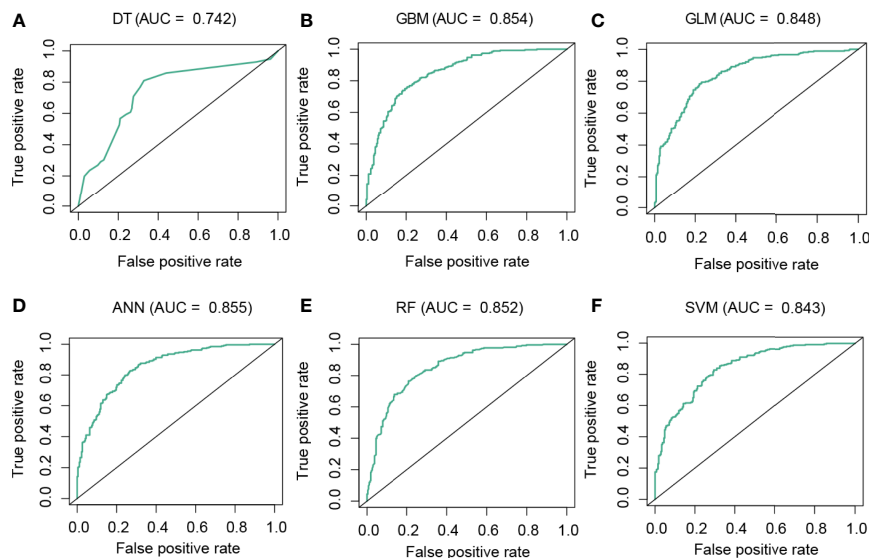


FIGURE 5 | Performance analysis of the models in the test dataset. ROC curve analysis of the performance of DT **(A)**, GBM **(B)**, GLM **(C)**, ANN **(D)**, RF **(E)**, and SVM **(F)**.

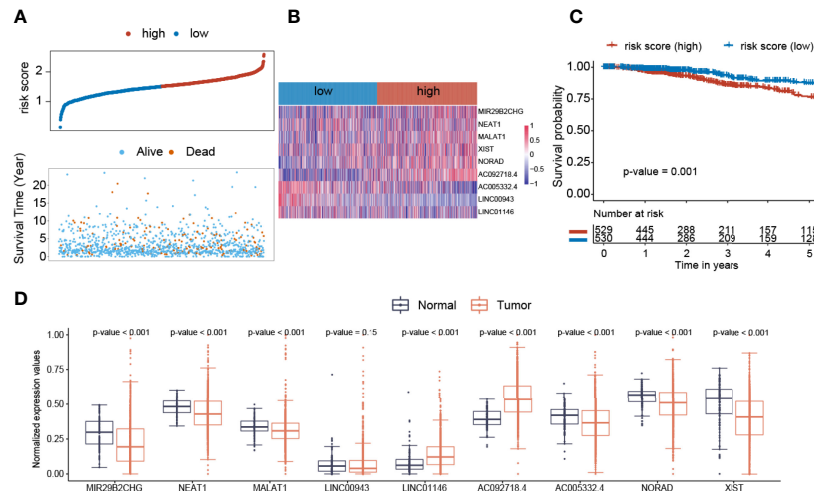


FIGURE 6 | Construction of the Cox regression model. **(A)** The distribution of prognosis including survival status and survival time between two risk score groups. **(B)** The heatmap of nine lncRNAs expression profiles. **(C)** The overall survival analysis of risk score groups. **(D)** The expression distribution values of nine lncRNAs between normal and breast cancer samples.

immune cells, including T and B cells. Risk score value was found to be negatively related to immune cells but positively associated with tumor purity. NEAT1, XIST, NORAD, MALAT1, MIR29B2CHG, and risk value were significantly negatively related to immune checkpoint genes including LAG3, TIGIT, CTLA4, and PDCD1 (Figure 7B). LINC00943, AC005332.4, AC092718.4, and LINC01146 were significantly positively associated with LAG3, TIGIT, CTLA4, and PDCD1.

Expression Patterns of Nine lncRNAs

Among nine lncRNAs, MIR29B2CHG, NEAT1, MALAT1, NORAD, and XIST were elevated in the low CD8 T-cell group than in the high CD8 T-cell group (Supplementary Figure 3). LINC00943, LINC01146, AC005332.4, and AC092718.4 were elevated in high CD8 T cells group than in low CD8 T cells group (Supplementary Figure 3). The associations between the risk model and clinical features were explored. Clinical

characteristics, such as age, stage, T stage, N stage, and M stage, had no effect on the risk score (Supplementary Figures 4A–E). We found that a higher risk score was correlated with a negative prognosis on progression-free survival (Supplementary Figure 4F). Based on the best cutoff values of these nine lncRNAs, the samples were divided into high and low groups. The survival results showed that AC005332.4, AC092718.4, NORAD, and LINC00943 have significant correlations with prognosis (Supplementary Figure 5).

Validation of XIST Expression in Breast Cancer Tissues by RT-qPCR

To validate XIST expression in BC clinical samples, 10 paired BC and normal breast tissues were collected and compared using real-time quantitative PCR (RT-qPCR). As shown in Figure 7C, expression of XIST is significantly lower in BC samples compared to the adjacent normal breast tissues (p-value < 0.05).

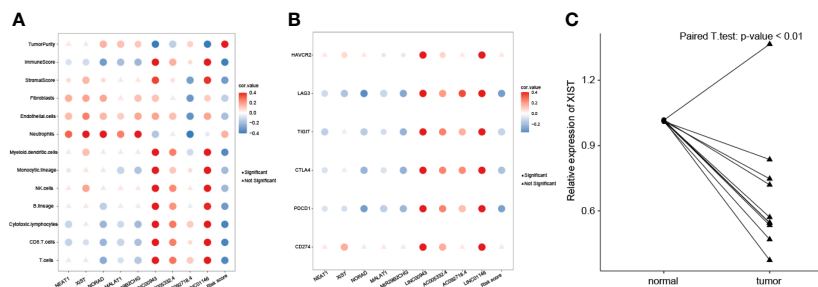


FIGURE 7 | Correlations of lncRNA expression values and immune biomarkers. **(A)** Significant correlations between proportions of immune cells with nine lncRNAs expression profiles. **(B)** Significant correlations between levels of immune checkpoint genes with nine lncRNAs expression profiles. **(C)** The XIST relative expression levels of normal and breast cancer samples were compared.

DISCUSSION

Breast cancer is a major cause of mortality in the current world population (24). CD8 T cells have been considered to play a central role in immunotherapy (25). Breast cancer immunotherapy is gaining traction as a treatment option, especially for patients with metastatic disease (26). Thus, the prediction of CD8 T cells in the tumor microenvironment (TME) has broad significance for patients and clinicians. Apart from that, the lncRNA-based risk model is gaining popularity as a consequence of its greater predictive potential than other risk models. However, a detailed analysis of a predictive model based on lncRNAs that are associated with T cells has not yet been completed.

It is important to discriminate between immunotherapy-sensitive and immunotherapy-insensitive cancer patients. Several biomarkers for immunotherapy have been identified (10). For example, programmed death-ligand 1 (PD-L1)/programmed cell death-1 (PD-1) levels, tumor mutational burden (TMB), and CD8 T cells can be used to assess the efficacy of immunotherapy. However, since PD-L1 is a changing indicator, there is no established technique for its measurement (27). TMB detection is difficult, costly, and time consuming (28). The deficiency of T cell in TME was the primary factor contributing to immunotherapy tolerance (29). Thus, the prediction or detection of T-cell infiltration within tumors is a promising method for selecting immunotherapy patients.

We began our study by collecting differentially expressed lncRNAs, miRNAs, and mRNAs using differential analysis. Next, we built two ceRNA networks that had positive and negative correlations with CD8 T cells. Among the lncRNAs from two networks, nine lncRNAs were screened to construct a random forest model to predict the CD8 T cells, since these lncRNAs have more interactions with other miRNAs. The random forest model demonstrated excellent prediction ability in discriminating between high and low CD8 T cells, which might serve as a useful reference for personalized immunotherapy.

COX regression was used to construct the prognostic risk model. The value of MIR29B2CHG (C1orf132) was significantly attenuated in BC (30). NEAT1 expression was elevated in BC samples, and the increased expression of NEAT1 was associated with a poor overall survival in BC patients (31). MALAT1 levels were shown to be significantly related to breast cancer development and invasive capacity, indicating that MALAT1 acts as a metastasis suppressor (32). LINC00944 expression has a strong relationship with immune signaling pathways (33). There is not much research describing the roles of LINC00943, LINC01146, AC092718.4, and AC005332.4 in breast cancer. NORAD expression was considerably decreased in BC samples, and its deficiency was linked with poor tumor tissue development (34). XIST expression is significantly reduced in BC samples (35).

This study has some limitations. First, the CD8 T-cell values were predicted by the expression profiles and ssGSEA method. The CD8 T-cell values detected by experiments will increase the credibility of our study. Second, the prognostic values of these lncRNAs should be validated by an independent dataset. Third,

the constructed model for immunotherapy response rate prediction should be tested by an independent dataset that contains the BC patients treated by immunotherapy.

CONCLUSION

In summary, using miRNA/lncRNA/mRNA expression profiles in BC, we constructed two ceRNA networks that are correlated with CD8 T cells. Additionally, we found a combination of nine lncRNAs that may constitute models for CD8 T cells and prognosis prediction. This will serve as a foundation for future research on immunotherapy selection in BC patients.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Hainan Medical University (HYLL-2021-265). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ZlC, UK, and RF: conceptualization, data curation, formal analysis, roles/writing—original draft, and writing—review and editing. ZtC, LR, and AS: roles/writing—original draft. CM, RW, and WS: funding acquisition, methodology, project administration, resources, and supervision.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.883197/full#supplementary-material>

Supplementary Figure 1 | Overall survival analysis of activated CD8 T cells calculated by ssGSEA (A) and MCP-counter (B) in BRCA patients.

Supplementary Figure 2 | The calculation process of the decision tree model.

Supplementary Figure 3 | The expression distribution values of 9 lncRNAs between high and low CD8 T cell groups.

Supplementary Figure 4 | The correlation of risk score with clinical parameters including age (A), stage (B), T (C), N (D), M (E), and (F) progression-free survival.

Supplementary Figure 5 | The survival analysis of 9 lncRNAs.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular Portraits of Human Breast Tumours. *Nature* (2000) 406(6797):747–52. doi: 10.1038/35021093
3. Kulkarni S, Augoff K, Rivera L, McCue B, Khoury T, Groman A, et al. Increased Expression Levels of WAVE3 Are Associated With the Progression and Metastasis of Triple Negative Breast Cancer. *PLoS One* (2012) 7(8):e42895. doi: 10.1371/journal.pone.0042895
4. Lee N, Park MJ, Song W, Jeon K, Jeong S. Currently Applied Molecular Assays for Identifying ESR1 Mutations in Patients With Advanced Breast Cancer. *Int J Mol Sci* (2020) 21(22):8807. doi: 10.3390/ijms21228807
5. Gray M, Meehan J, Martinez-Perez C, Kay C, Turnbull AK, Morrison LR, et al. Naturally-Occurring Canine Mammary Tumors as a Translational Model for Human Breast Cancer. *Front Oncol* (2020) 10:617. doi: 10.3389/fonc.2020.00617
6. Shah D, Osipo C. Cancer Stem Cells and HER2 Positive Breast Cancer: The Story So Far. *Genes Dis* (2016) 3(2):114–23. doi: 10.1016/j.gendis.2016.02.002
7. Macia F, Porta M, Murta-Nascimento C, Servitja S, Guxens M, Buron A, et al. Factors Affecting 5- and 10-Year Survival of Women With Breast Cancer: An Analysis Based on a Public General Hospital in Barcelona. *Cancer Epidemiol* (2012) 36(6):554–9. doi: 10.1016/j.canep.2012.07.003
8. Schmid P, Cortes J, Pusztai L, McArthur H, Kummel S, Bergh J, et al. Pembrolizumab for Early Triple-Negative Breast Cancer. *N Engl J Med* (2020) 382(9):810–21. doi: 10.1056/NEJMoa1910549
9. Raskov H, Orhan A, Christensen JP, Gogenur I. Cytotoxic CD8(+) T Cells in Cancer and Cancer Immunotherapy. *Br J Cancer* (2021) 124(2):359–67. doi: 10.1038/s41416-020-01048-4
10. Chen Z, Wang M, De Wilde RL, Feng R, Su M, Torres-de LRL, et al. A Machine Learning Model to Predict the Triple Negative Breast Cancer Immune Subtype. *Front Immunol* (2021) 12:749459. doi: 10.3389/fimmu.2021.749459
11. Zheng S, Li M, Miao K, Xu H. SNHG1 Contributes to Proliferation and Invasion by Regulating miR-382 in Breast Cancer. *Cancer Manag Res* (2019) 11:5589–98. doi: 10.2147/CMAR.S198624
12. 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):353–8. doi: 10.1016/j.cell.2011.07.014
13. Tan JY, Sirey T, Honti F, Graham B, Piovesan A, Merckenschlager M, et al. Extensive microRNA-Mediated Crosstalk Between lncRNAs and mRNAs in Mouse Embryonic Stem Cells. *Genome Res* (2015) 25(5):655–66. doi: 10.1101/gr.181974.114
14. Karreth FA, Pandolfi PP. CeRNA Cross-Talk in Cancer: When Ce-Bling Rivalries Go Awry. *Cancer Discov* (2013) 3(10):1113–21. doi: 10.1158/2159-8290.CD-13-0202
15. Zu F, Liu P, Wang H, Zhu T, Sun J, Sheng W, et al. Integrated Analysis Identifies a Pathway-Related Competing Endogenous RNA Network in the Progression of Pancreatic Cancer. *BMC Cancer* (2020) 20(1):958. doi: 10.1186/s12885-020-07470-4
16. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: An R/Bioconductor Package for Integrative Analysis of TCGA Data. *Nucleic Acids Res* (2016) 44(8):e71. doi: 10.1093/nar/gkv1507
17. Charoentong P, Finotello F, Angelova M, Mayer C, Efreanova M, Rieder D, et al. Pan-Cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep* (2017) 18(1):248–62. doi: 10.1016/j.celrep.2016.12.019
18. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the Population Abundance of Tissue-Infiltrating Immune and Stromal Cell Populations Using Gene Expression. *Genome Biol* (2016) 17(1):218. doi: 10.1186/s13059-016-1070-5
19. Yoshihara K, Shahmoradgol M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring Tumour Purity and Stromal and Immune Cell Admixture From Expression Data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612
20. Robinson MD, McCarthy DJ, Smyth GK. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* (2010) 26(1):139–40. doi: 10.1093/bioinformatics/btp616
21. Li JH, Liu S, Zhou H, Qu LH, Yang JH. StarBase V2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and Protein-RNA Interaction Networks From Large-Scale CLIP-Seq Data. *Nucleic Acids Res* (2014) 42(Database issue):D92–7. doi: 10.1093/nar/gkt1248
22. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118
23. Kuhn M. Building Predictive Models in R Using the Caret Package. *J Stat Software* (2008) 28(5):1–26. doi: 10.18637/jss.v028.i05
24. Swaminathan V, Spiliopoulos MK, Audisio RA. Choices in Surgery for Older Women With Breast Cancer. *Breast Care (Basel)* (2012) 7(6):445–51. doi: 10.1159/000345402
25. Lin J, Yu M, Xu X, Wang Y, Xing H, An J, et al. Identification of Biomarkers Related to CD8(+) T Cell Infiltration With Gene Co-Expression Network in Clear Cell Renal Cell Carcinoma. *Aging (Albany NY)* (2020) 12(4):3694–712. doi: 10.18632/aging.102841
26. Adams S, Gatti-Mays ME, Kalinsky K, Korde LA, Sharon E, Amiri-Kordestani L, et al. Current Landscape of Immunotherapy in Breast Cancer: A Review. *JAMA Oncol* (2019) 5(8):1205–14. doi: 10.1001/jamaoncol.2018.7147
27. Lopez-Beltran A, Henriques V, Cimadamore A, Santoni M, Cheng L, Gevaert T, et al. The Identification of Immunological Biomarkers in Kidney Cancers. *Front Oncol* (2018) 8:456. doi: 10.3389/fonc.2018.00456
28. Fancello L, Gandini S, Pelicci PG, Mazzarella L. Tumor Mutational Burden Quantification From Targeted Gene Panels: Major Advancements and Challenges. *J Immunother Cancer* (2019) 7(1):183. doi: 10.1186/s40425-019-0647-4
29. Lin Z, Huang L, Li SL, Gu J, Cui X, Zhou Y. PTEN Loss Correlates With T Cell Exclusion Across Human Cancers. *BMC Cancer* (2021) 21(1):429. doi: 10.1186/s12885-021-08114-x
30. Shafaroudi AM, Sharifi-Zarchi A, Rahmani S, Nafissi N, Mowla SJ, Lauria A, et al. Expression and Function of *C1orf132* Long-Noncoding RNA in Breast Cancer Cell Lines and Tissues. *Int J Mol Sci* (2021) 22(13):6768. doi: 10.3390/ijms22136768
31. Li S, Hao J, Hong Y, Mai J, Huang W. Long Non-Coding RNA NEAT1 Promotes the Proliferation, Migration, and Metastasis of Human Breast-Cancer Cells by Inhibiting miR-146b-5p Expression. *Cancer Manag Res* (2020) 12:6091–101. doi: 10.2147/CMAR.S252295
32. Kim J, Piao HL, Kim BJ, Yao F, Han Z, Wang Y, et al. Long Noncoding RNA MALAT1 Suppresses Breast Cancer Metastasis. *Nat Genet* (2018) 50(12):1705–15. doi: 10.1038/s41588-018-0252-3
33. de Santiago PR, Blanco A, Morales F, Marcelain K, Harismendy O, Sjoberg HM, et al. Immune-Related lncRNA LINC00944 Responds to Variations in ADAR1 Levels and It Is Associated With Breast Cancer Prognosis. *Life Sci* (2021) 268:118956. doi: 10.1016/j.lfs.2020.118956
34. Liu W, Zhou X, Li Y, Jiang H, Chen A. Long Non-Coding RNA NORAD Inhibits Breast Cancer Cell Proliferation and Metastasis by Regulating miR-155-5p/SOCS1 Axis. *J Breast Cancer* (2021) 24(3):330–43. doi: 10.4048/jbc.2021.24.e32
35. Huang YS, Chang CC, Lee SS, Jou YS, Shih HM. Xist Reduction in Breast Cancer Upregulates AKT Phosphorylation via HDAC3-Mediated Repression of PHLPP1 Expression. *Oncotarget* (2016) 7(28):43256–66. doi: 10.18632/oncotarget.9673

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chen, Feng, Kahlert, Chen, Torres-de-la Roche, Soliman, Miao, De Wilde and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.