Journal of Biosciences and Medicines, 2026, 14(11), 531-554
https://www.scirp.org/journal/jbm
ISSN Online: 2327-509X
ISSN Print: 2327-5081

A Four-Long Non-Coding RNA Signature as a Potential Prognostic Biomarker for Breast Cancer

Ying Rao[1*], Linli Zhang[2*], Mingtao Luo[1], Xuanhua Chen[1], Yongle Li[1], Yixin Liu[1], Lihe Jiang[1,3,4#]

[1]School of Basic Medical Sciences, Youjiang Medical University for Nationalities, Baise, China
[2]Department of Cardiology, Liuzhou People's Hospital, Affiliated of Guangxi Medical University, Liuzhou, China
[3]Key Laboratory of Basic Research on Regional Diseases (Guangxi Medical University), Education Department of Guangxi Zhuang Autonomous Region, Nanning, China
[4]Fujian Provincial Key Laboratory of Innovative Drug Target Research, Xiamen University, Xiamen, China
Email: #jianglihe@ymun.edu.cn

How to cite this paper: Rao, Y., Zhang, L.L., Luo, M.T., Chen, X.H., Li, Y.L., Liu, Y.X. and Jiang, L.H. (2026) A Four-Long Non-Coding RNA Signature as a Potential Prognostic Biomarker for Breast Cancer. *Journal of Biosciences and Medicines*, **14**, 531-554.
https://doi.org/10.4236/jbm.2026.1411038

Received: October 20, 2026
Accepted: November 23, 2026
Published: November 26, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

Objective: To provide a basis for patient prognosis prediction by constructing a competing endogenous RNA (ceRNA) regulatory network of breast cancer (BRCA). **Methods:** Differentially expressed (DE) lncRNAs (DElncRNAs), mRNAs (DEmRNAs), and miRNAs (DEmiRNAs) between BRCA and normal breast tissue samples were identified based on TCGA datasets with clinical data. lncRNA-miRNA-mRNA ceRNA network was established. The transcription factor regulatory network and its protein interaction network were mapped to the target genes in the ceRNA network, then gene oncology enrichment analysis of lncRNA interacting DEmRNAs. Univariate Lasso regression and multivariate Cox regression analyses were used and analyzed DElncRNAs, followed by a risk model to stratify patients into high- and low-risk groups, and to validate model validity and survival (OS), and to predict potential binding protein sites of lncRNAs associated with prognostic models of subcellular localization. **Results:** There were 1094 DElncRNAs, 86 DEmiRNAs, and 2130 DEmRNAs in BRCA. The lncRNA-miRNA-mRNA ceRNA network of BRCA was composed of 81 DElncRNAs, 18 DEmiRNAs, and 27 DEmRNAs. DElncRNAs through univariate Lasso regression as well as multivariate regression analysis revealed that four lncRNAs (AC007731.1, ADAMTS9-AS1, C10orf126 and SRGAP3-AS2) constituted a prognostic model associated with survival of BRCA patients. **Conclusions:** A prognostic risk model based on four lncRNAs (AC007731.1, ADAMTS9-AS1, C10orf126, and SRGAP3-AS2)

*Ying Rao and Lin-Li Zhang have been contributed equally to this work.
#Corresponding author.

was constructed and showed potential utility for predicting overall survival in patients with breast cancer. Further validation is required before clinical application.

Keywords

Breast Cancer, Long Noncoding RNA, Prognosis Bioinformatics, CeRNA Networks, Transcription Factor Regulatory Network

1. Introduction

Breast cancer is one of the most common cancers and the leading cause of death due to cancer in women [1]. Despite advances in treatment, including endocrine therapy, CDK4/6 inhibitors, and HER2-targeted therapies, clinical outcomes remain heterogeneous across breast cancer subtypes [2]-[4].

Long noncoding RNAs (lncRNAs) are a class of noncoding RNAs that are more than 200 nucleotides long, which have been reported to be involved in many cellular processes, such as transcriptional and post-transcriptional regulation [5]-[7]. Several lncRNAs have important roles in cancer development, some of which have also been identified in BRCA [8]-[10]. Recent studies have also examined subtype-specific lncRNA patterns and prognostic signatures in breast cancer [11] [12]. An umbrella review identified potential non-coding RNA biomarkers while emphasizing heterogeneity across studies [13]. Moreover, accumulating evidence highlights the role of lncRNAs as competing endogenous RNAs (ceRNAs) in tumorigenesis and progression [14] [15].

LncRNA-miRNA-mRNA ceRNA networks have been investigated in different biological settings, including stem cells and breast cancer [16] [17]. Many lncRNAs have been proposed as ceRNAs for miRNAs. In breast cancer, DDX11-AS1 has been reported to promote progression through the miR-30c-5p/MTDH regulatory axis [18]. The H19/let-7/Lin28 ceRNA network has been reported to promote epithelial-mesenchymal transition in breast cancer cells, accompanied by downregulation of autophagy [19]. Furthermore, ADAMTS9-AS1 has been reported to restrain aggressive breast cancer cell traits by sponging miR-513a-5p [20]. Previous studies have identified lncRNA-based prognostic signatures in breast cancer. Notably, the eight-lncRNA signature reported in reference [10] included AC007731.1, ADAMTS9-AS1, C10orf126, and SRGAP3-AS2, all of which were retained in the present study. Building on previous research, the present study constructed a more concise four-lncRNA prognostic model based on a predicted lncRNA-miRNA-mRNA ceRNA network and further investigated their potential RNA-binding proteins and subcellular localization, providing additional computational insights into their potential biological characteristics.

lncRNAs have been investigated as potential diagnostic and prognostic biomarkers in breast cancer [21]. Several studies have shown lncRNAs are consid-

ered as potential biomarkers in BRCA. For example, a TMPO-AS1-associated ceRNA network has been linked to poor survival in basal-like breast cancer [22].

Therefore, the present study aimed to construct a four-lncRNA prognostic risk model based on a predicted lncRNA-miRNA-mRNA ceRNA network using TCGA breast cancer data. Prognostic lncRNAs were identified using univariate Cox regression, LASSO regression, and multivariate Cox regression analyses, and the resulting risk model was evaluated for its prognostic performance. Furthermore, the potential RNA-binding proteins and subcellular localization of the four selected lncRNAs were predicted using bioinformatics tools.

2. Materials and Methods

2.1. Data Sources

From TCGA database (https://cancergenome.nih.gov/), 1222 breast cancer sample RNA-seq, 1207 breast cancer miRNA-seq data as well as breast cancer-related clinical information were downloaded. Among them, clinical information contains basic information of patients, survival status and prognosis information, etc. Breast cancer samples with mRNA, miRNA, lncRNA sequence data were generated by the sequencing platform named Illumina-HiSeq RNAseq. Gene expression matrix files were generated from downloaded data in TCGA database, and breast cancer sample RNA-seq data were divided into breast cancer group (1109 cases) and paracancerous group (113 cases), and breast cancer miRNA-seq data were divided into breast cancer group (1103 cases) and paracancerous group (104 cases). Human genomic lncRNA, miRNA, and mRNA annotation information was downloaded from the GENCODE database (https://www.gencodegenes.org/). For survival analysis, duplicate sample records were removed during data preprocessing, and primary tumor samples were retained. TCGA sample identifiers were converted into patient-level identifiers, which were used to match gene expression data with the corresponding clinical records. Patients without matched expression data or the required clinical information were excluded. Following data processing and matching, a total of 1057 patients were included in the final survival analysis. Overall survival (OS) was assessed using the clinical diagnosis and follow-up records obtained from the Genomic Data Commons (GDC). For deceased patients, survival time was determined using the recorded days to death, whereas patients who remained alive were right-censored at their last available follow-up, with survival time determined using the days to last follow-up.

2.2. Screening of Differentially Expressed lncRNAs, mRNAs, and miRNAs

The "edgeR" package was employed in R software for gene differential expression analysis (identification criteria were $p < 0.05$ after correction. Differential expression fold changes absolute greater than 2). Volcano plots were drawn using the R software "ggplot2" package and heatmaps were drawn using the "pheatmap" package.

2.3. Construction of ceRNA Network and Functional Enrichment Analysis of DElncRNA Target Genes

Putative lncRNA-miRNA interactions were identified using the miRcode database (http://www.mircode.org/) (https://doi.org/10.1093/bioinformatics/bts344). Candidate miRNA-mRNA interactions were retrieved from the miRDB database (https://mirdb.org/), TargetScan database (https://www.targetscan.org/vert_72/), and miRTarBase (https://mirtarbase.cuhk.edu.cn/) (https://doi.org/10.1093/nar/gkae1072). Only interactions supported by all three databases were retained. The predicted target mRNAs were then intersected with the differentially expressed mRNAs, and the retained interactions were integrated to construct a putative lncRNA-miRNA-mRNA ceRNA network, which was visualized using Cytoscape software. Matched-sample expression correlations were not used as criteria for constructing or filtering the ceRNA network. The retained interactions were based on database predictions and were therefore considered putative regulatory relationships rather than experimentally validated interactions. Gene Ontology (GO) functional enrichment analysis of the target genes of DElncRNAs was performed using the R software "Clusterprofiler" package to explore the potential functions of the target genes.

2.4. DElncRNA Target Gene Transcription Factor Network and Protein Interaction Network

From DAVIDE (https://davidbioinformatics.nih.gov/) database selected "UCSC_TFBS" to extract clinically relevant transcription factors (TFs), and the regulatory network of target genes and TFs was constructed by Cytoscape software to explore the regulatory link between TFs and target genes. DElncRNA target genes were subjected to protein interaction analysis by String (https://www.string-db.org/).

2.5. Univariate Cox Regression Analysis and Lasso Regression

Univariate Cox analysis was performed to evaluate the association between DElncRNA expression levels and the overall survival (OS) of patients, and DElncRNAs with a p-value < 0.05 were selected for further analysis. For prognostic analysis, gene expression levels were normalized using the trimmed mean of M-values (TMM) method and log2-transformed as log2 (CPM + 1). The Lasso model was used to analyze them to further find the significant lncRNAs from prognostic DElncRNAs. Lasso analysis was implemented by the "glmnet" package in R software. Ten-fold cross-validation was performed using partial likelihood deviance as the evaluation criterion, and the optimal penalty parameter (λ) was selected according to the minimum cross-validation error (lambda.min). LncRNAs with nonzero regression coefficients were retained for subsequent analysis.

2.6. Multivariate Cox Risk Assessment Analysis and Visual Graphics

Multivariate Cox risk assessment was employed to find lncRNA models that could

be used as predictors of patient prognosis. The 12 candidate lncRNAs identified by LASSO regression were included in multivariate Cox regression analysis to further identify prognostic lncRNAs. Four lncRNAs (AC007731.1, ADAMTS9-AS1, C10orf126, and SRGAP3-AS2) were ultimately retained to construct the prognostic risk model. A risk assessment formula was constructed based on the expression levels and regression coefficients of lncRNAs. The eventual risk score was calculated based on the coefficients of every lncRNA as below:

$$\text{risk score} = \sum_{i=1}^{n} \text{coefi} \times \text{id}$$

In this formula, coefi was defined as the coefficient of lncRNAs correlated with survival and it was defined as the expression of lncRNAs. Random forest plots were drawn using the R software "survminer" package. The risk score visual graph was again plotted to draw a risk heat map according to the key lncRNAs expression profiles of high-risk group and low-risk group.

2.7. Evaluation of a Prognostic Risk Score Model

Based on the sensitivity and specificity of survival prediction, the receiver operating characteristic curve (ROC) was drawn by using the R software "timeROC" package, and the time-dependent receiver operating characteristic curve (ROC) within 3 and 5 years was drawn to evaluate its prognostic performance, and a p-value < 0.05 was considered statistically significant. Concordance index calculation (C-index) was performed using the R software "survcomp" package to estimate the accuracy of the prognostic risk model.

2.8. Kaplan-Meier (KM) Survival Prognostic Analysis

Kaplan-Meier survival curves were used to assess prognosis between low-risk and high-risk groups. And the survival curve graph of DElncRNAs in the prognostic model was plotted by "survival" of R software.

2.9. Prediction of lncRNA Potential Binding Protein Sites

Using the NCBI database (https://www.ncbi.nlm.nih.gov/gene) to query the sequence of the lncRNA before using the catRAPID database (https://tools.tartaglialab.com/page/catrapid_omics2_group) to predict the potential binding protein sites of lncRNAs, Cytoscape software was used to construct the protein binding site enrichment network of lncRNAs. Subsequently, the RPISeq database (http://pridb.gdcb.iastate.edu/RPISeq/batch-rna.html) (https://doi.org/10.1186/1471-2105-12-489) was combined, and the scores of the predicted potential binding protein sites of lncRNAs were comprehensively evaluated. Finally, the lncLocator database (http://www.csbio.sjtu.edu.cn/bioinf/lncLocator/) was used to identify the subcellular localization landscape of key lncRNAs.

2.10. Statistical Analysis

Mining of the filtered lncRNA, miRNA, mRNA, and clinical information data

with differential expression was performed using R (version 4.0.4) and Perl (version 5.28.1) software. Measurement data was expressed as mean ± standard deviation (SX), comparison of data between groups was made by t-test or analysis of variance depending on their nature, and comparisons of differences in survival time between groups were made by Cox proportional hazard model, and statistical significance was considered when $p < 0.05$.

3. Result

3.1. DEmRNAs, DElncRNAs, and DEmiRNAs in Breast Cancer

Compared with para carcinoma tissues, 1094 differentially expressed lncRNAs (843 upregulated, 251 downregulated), 86 differentially expressed miRNAs (67 upregulated, 19 downregulated) and 2130 differentially expressed mRNAs (1345 upregulated, 785 downregulated) were found. **Figure 1** shows the Heatmap and volcano plot of DElncRNAs, DEmiRNAs, and DEmRNAs, respectively.

3.2. Characterization of the ceRNA Network and GO Enrichment

The interactions between lncRNAs and miRNAs were predicted using the miRcode database, resulting in 81 DElncRNAs associated with 18 DEmiRNAs. Candidate miRNA-mRNA interactions were identified based on the intersection of predictions from the miRDB, TargetScan, and miRTarBase databases. Subsequently, 27 DEmRNAs were retained to construct a putative lncRNA-miRNA-mRNA ceRNA network, which was visualized using Cytoscape software (**Figure 2**). Ultimately, the ceRNA network included 81 DElncRNAs (61 upregulated and 20 downregulated), 18 DEmiRNAs (14 upregulated and 4 downregulated) and 27 DEmRNAs (9 upregulated and 18 downregulated). The target genes of DElncRNAs were further subjected to GO functional enrichment analysis to explore their potential functions (**Figure 3**). The results indicated that these 27 target genes were associated with the positive regulation of the ERK1 and ERK2 cascades, cellular responses to growth factor stimulation, and stem cell population maintenance.

3.3. Transcription Factor Enrichment and Protein Interaction Networks

To explore the potential molecular mechanisms of the target genes, transcription factors (TFs) were predicted for the 27 target genes (9 up-regulated and 18 down-regulated) (**Table 1**), and a further 5 TFs for the 27 target genes (**Table 1**). Transcription factor network diagram was constructed (**Figure 4(A)**). We next explored the protein interaction relationship between DElncRNAs target genes and their transcription factors (**Figure 4(B)**).

3.4. Initial Identification of Optimal Prognostic Biomarkers

To determine the effect of individual lncRNAs in the ceRNA network on patient prognosis, 81 DElncRNAs in the ceRNA network were subjected to univariate Cox regression analysis. The lasso algorithm was used for feature selection and

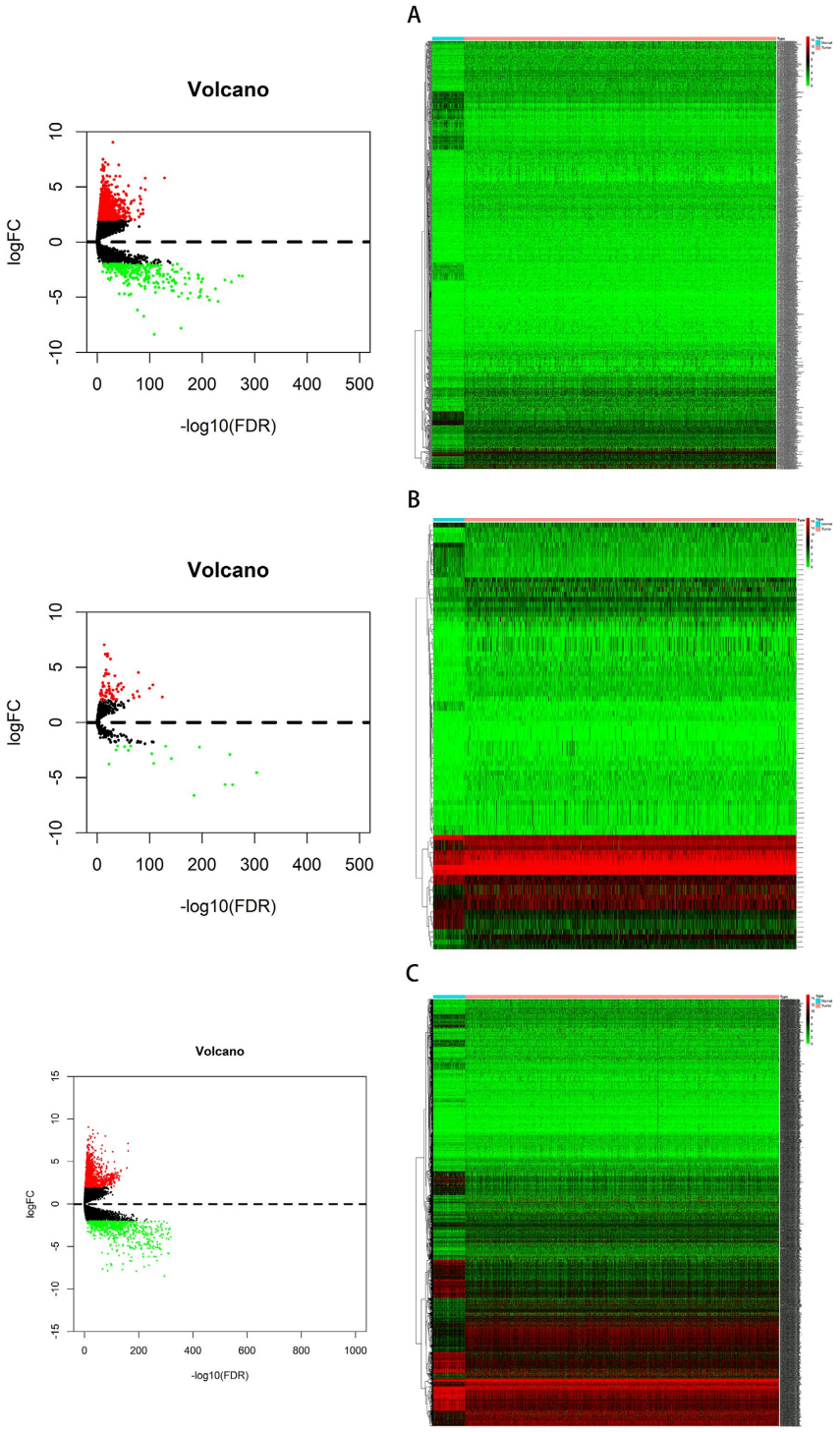


Figure 1. Volcano plots and heatmaps of DElncRNAs, DEmiRNAs, and DEmRNAs. Panels (A)-(C) represent lncRNAs, miRNAs, and mRNAs, respectively. The volcano plots (left panels) show differential RNA expression between breast cancer (BRCA) tissues and adjacent normal tissues. The x-axis represents –log10(FDR), and the y-axis represents logFC. The heatmaps (right panels) show the expression profiles of differentially expressed RNAs, with rows representing RNAs and columns representing samples. Color intensity indicates relative RNA expression levels. Blue and red sample annotations represent normal and BRCA samples, respectively.

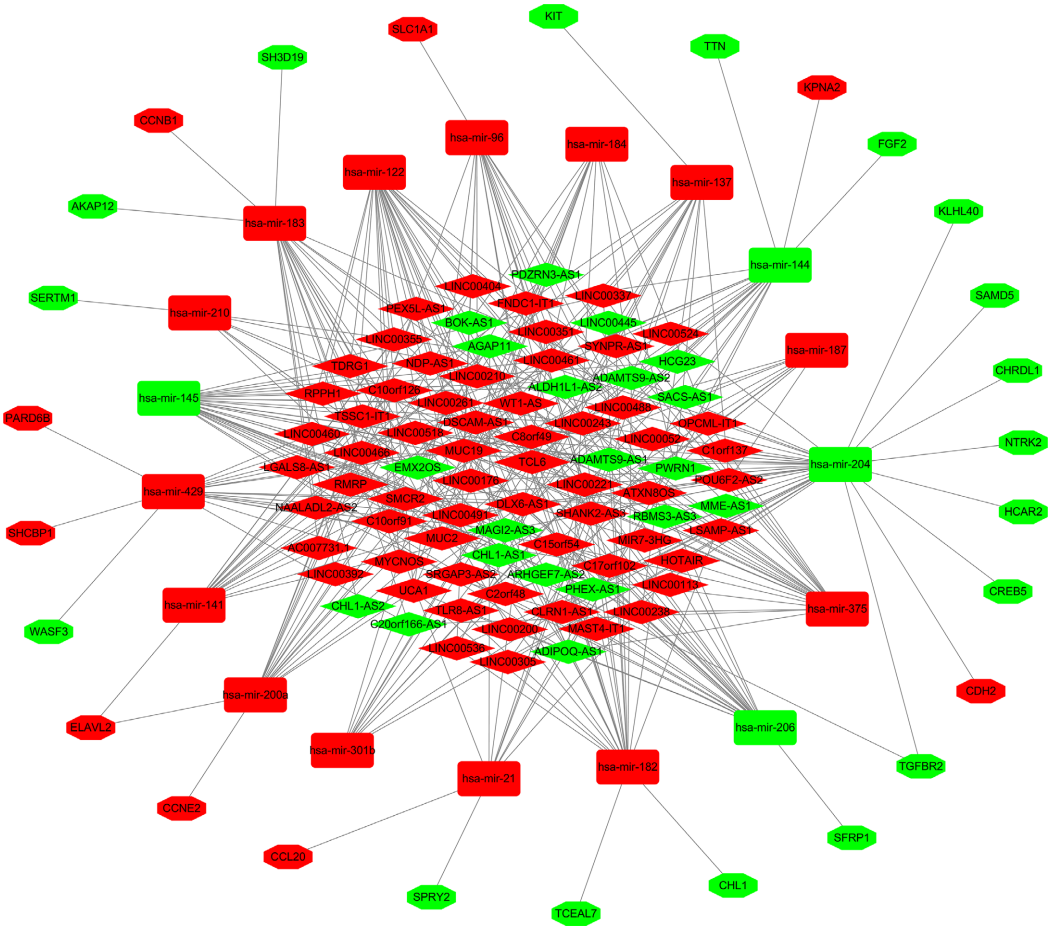


Figure 2. LncRNA-miRNA-mRNA ceRNA network in breast cancer. Diamonds represent lncRNA, round rectangles represent miRNAs, and octagons represent mRNAs. Red and green nodes represent up and down regulation respectively. The whole ceRNA network contained 81 DElncRNAs (61 upregulated and 20 downregulated), 18 DEmiRNAs (14 upregulated and 4 downregulated) and 27 DEmRNAs (9 upregulated and 18 downregulated).

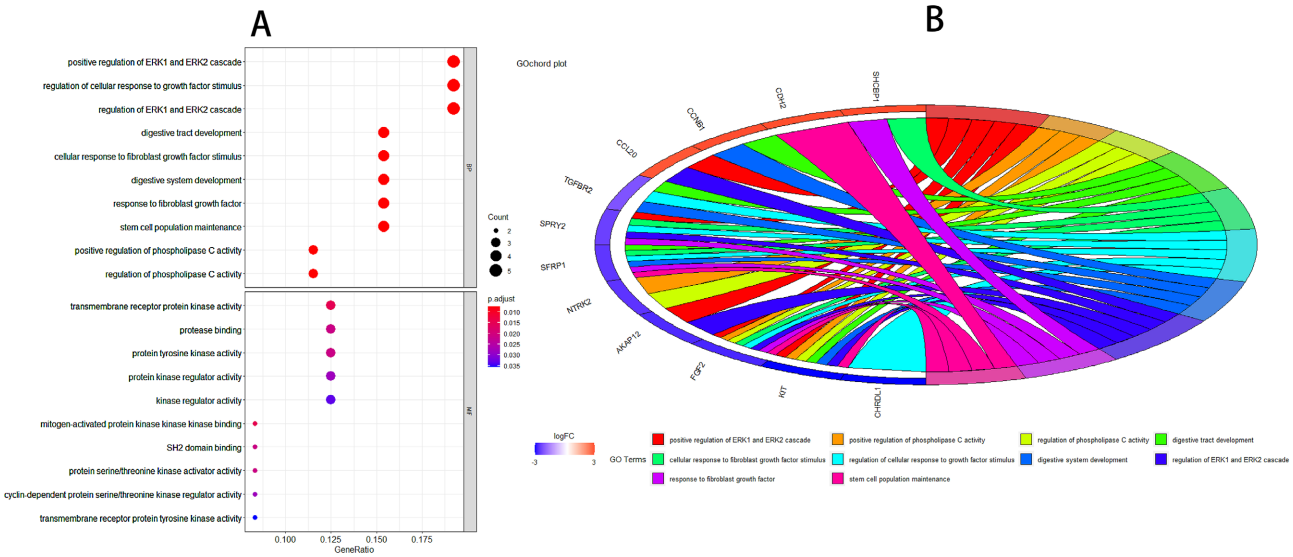


Figure 3. Target gene GO enrichment analysis. GO enrichments included the following. BP: biological process, CC: cellular component and MF: molecular function.

Table 1. Transcription factors enriched in target genes in ceRNAs.

Term	Count	PValue	FDR
FOXO1	20	5.34E−05	0.008063555
FREAC7	21	1.64E−04	0.012367817
NFY	20	7.88E−04	0.026382813
HFH1	19	8.03E−04	0.026382813
LHX3	17	8.74E−04	0.026382813

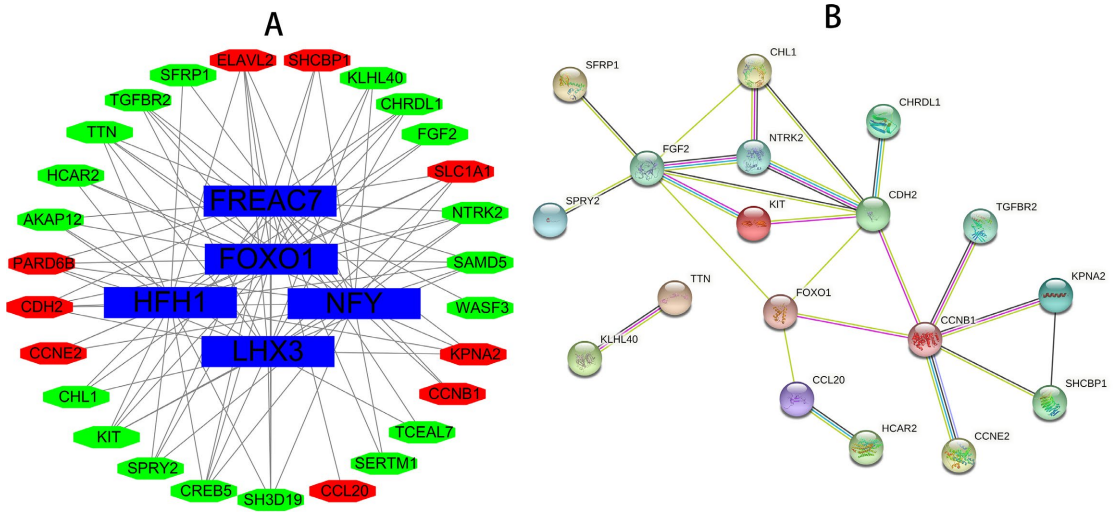


Figure 4. Target genes with transcription factors. (A) DElncRNA target gene transcription factor network. Blue rectangles represent target gene enriched transcription factors (TFs), and octagons represent target genes of DElncRNAs. Red indicates upregulated target genes (9) and green indicates downregulated target genes (18). (B) Protein interaction network of target genes with transcription factors.

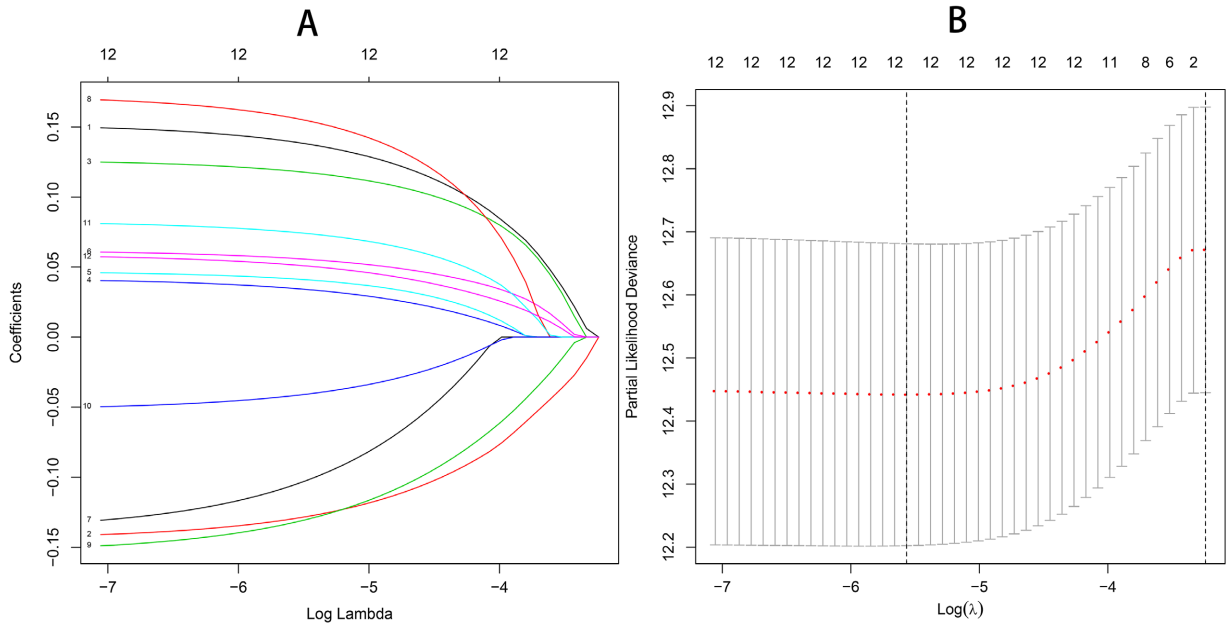


Figure 5. Least absolute shrinkage and selection operator (LASSO) regression. (A) regression coefficients vs λ Schematic of the value relationship. (B) λ Values versus their corresponding lncRNA numbers are schematically represented.

Table 2. The 12 lncRNAs screened by lasso regression analysis.

Sequence	Gene names	Ensemble name	lncRNA types
1	AC007731.1	ENSG00000188280	antisense
2	C10orf126	ENSG00000204365	lincRNA
3	ADAMTS9-AS1	ENSG00000241158	antisense
4	WT1-AS	ENSG00000183242	antisense
5	LINC00536	ENSG00000249917	lincRNA
6	SRGAP3-AS2	ENSG00000228723	antisense
7	TLR8-AS1	ENSG00000233338	antisense
8	PWRN1	ENSG00000259905	lincRNA
9	LINC00200	ENSG00000229205	lincRNA
10	HOTAIR	ENSG00000228630	antisense
11	MME-AS1	ENSG00000240666	antisense
12	SYNPR-AS1	ENSG00000241359	antisense

the lasso regression coefficients are shown in **Figure 5(A)**. A total of 12 lncRNAs were identified by this method (**Figure 5(B)**). To further screen the variables to reduce the number of differential genes, preventing outlying, 12 lncRNAs among them (**Table 2**) were identified to further serve as prognostic markers in breast cancer.

3.5. Establishment and Visualization of a Prognostic Risk Model for Breast Cancer lncRNAs

Using multivariate Cox regression analysis, four lncRNAs (AC007731.1, ADAMTS9-AS1, C10orf126, and SRGAP3-AS2) were identified as comprehensive prognostic biomarkers for breast cancer patients. The risk scores were calculated for all patients, and the expression levels and regression coefficients of these 4 lncRNAs were calculated to establish a prognostic risk score formula with the median risk score as the cutoff. The formula of prognostic risk score was as follows. Risk score = (0.14347) × AC007731.1 + (−0.17224) × ADAMTS9-AS1 + (0.15819) × C10orf126 + (−0.14734) × SRGAP3-AS2, and a risk score was calculated for each patient according to the formula described above. The random forest plot is shown in **Figure 6**. The 1057 patients were divided into high-risk (n= 528) and low-risk (n = 529) groups (**Figure 7(A)**), with high expression of risk genes (AC007731.1 and C10orf126) in the high-risk group and high expression of protective genes (ADAMTS9-AS1 and SRGAP3-AS2) in the low-risk group. And scatter plot (**Figure 7(B)**), finally, the expression profile heat map of each key lncRNA in high-risk group and low-risk group (**Figure 7(C)**) as well as high and low-risk group survival curve (**Figure 7(D)**) were plotted. The results indicated that patients with a high risk score had a worse prognosis than those with a low risk score.

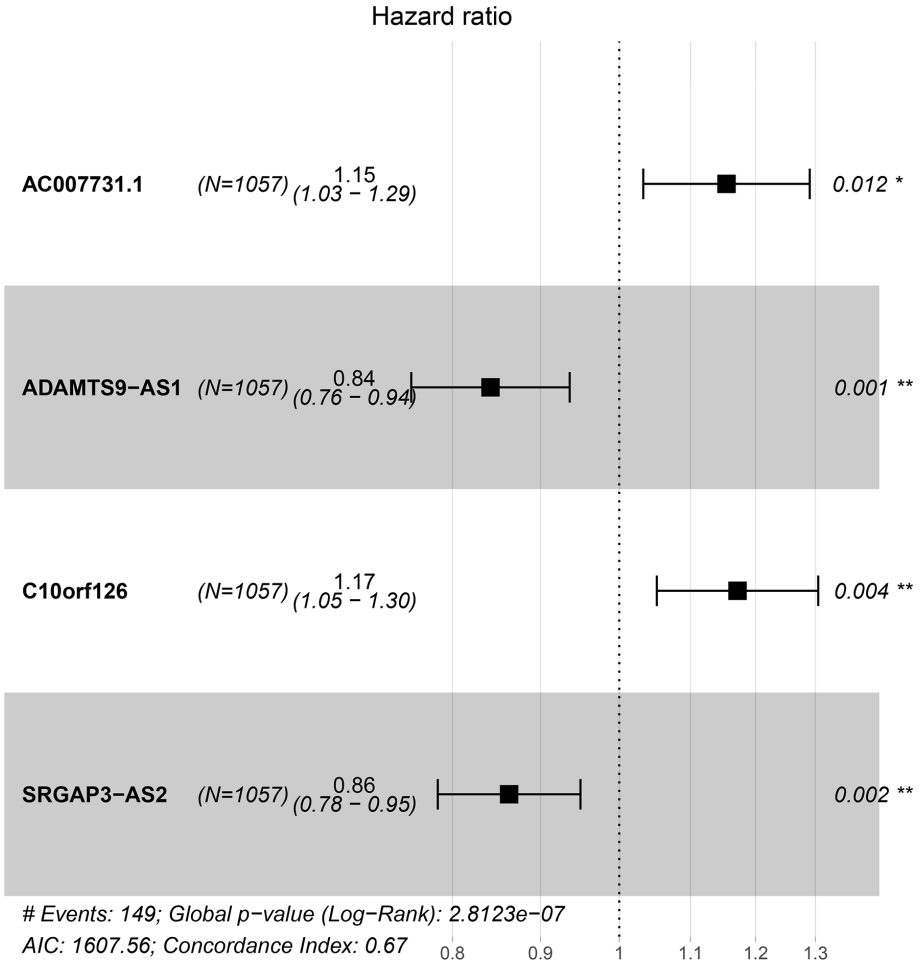
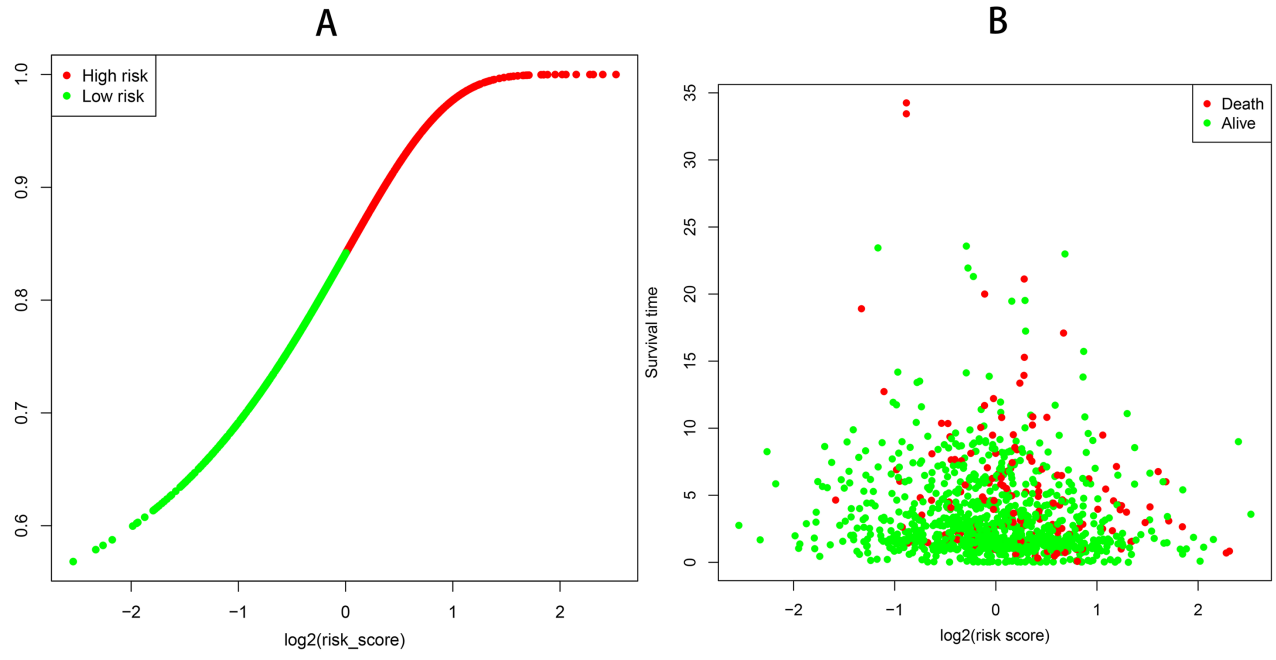


Figure 6. 4 key lncRNA risk assessment rate forest plot. Black squares represent hrs and short horizontal lines represent 95% CIs.



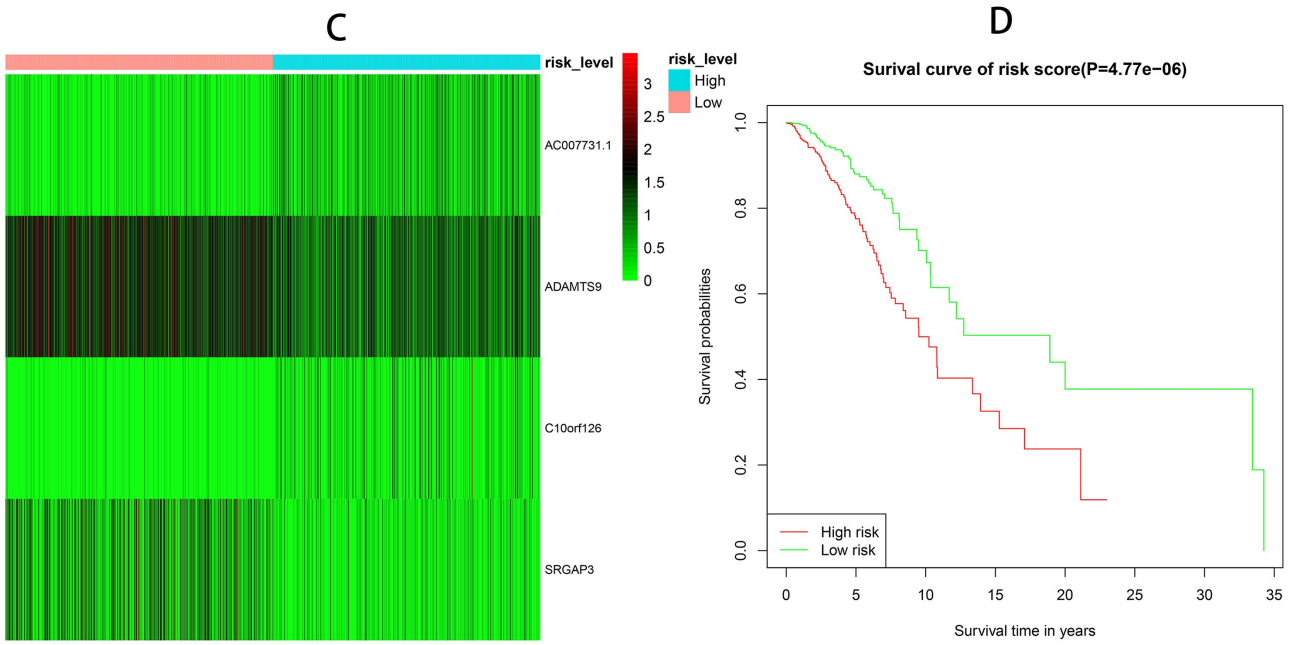


Figure 7. Risk model for breast cancer lncRNAs. (A) cumulative function distribution plot of risk score expression values (upper left). (B) survival status and duration in different groups of patients (top right). (C) incorporated gene expression profile Heatmap (bottom left). (D) survival curves for low - and high-risk groups (bottom right).

3.6. ROC Curves and the C-Index of the Survival Model

The sensitivity and specificity of the survival prediction based on the risk score, using the risk score for 3- and 5-year OS, time-dependent receiver operating characteristic curves (ROC) were plotted to evaluate the accuracy of this prognostic model (Figure 8). Among them, the area under the ROC curve was 0.663

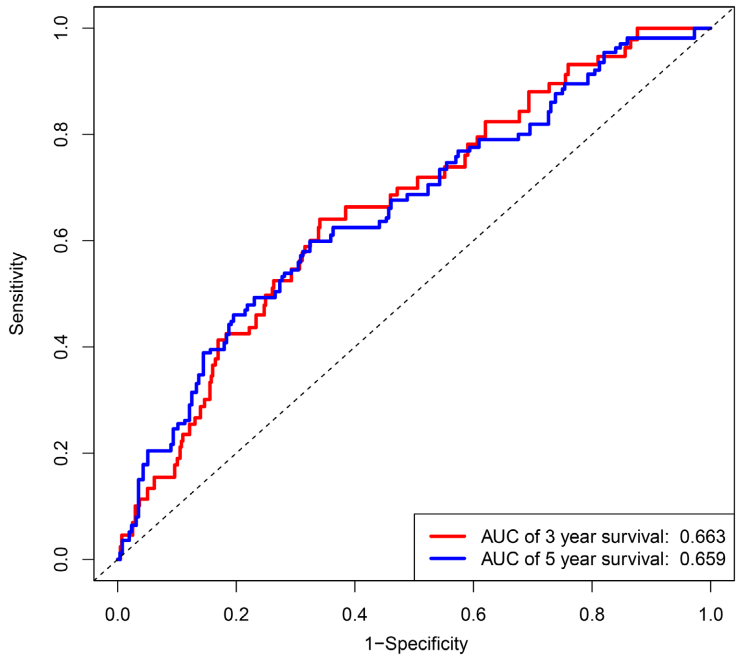


Figure 8. Survival dependent ROC curve for prognostic value.

at 3 years and 0.659 at 5 years, indicating that the risk model had a good predictive value. According to C (C-index) index, again testing its accuracy (Table 3), concordance = 0.667 (SE = 0.025), 95% CI was (0.618 ~ 0.716). The description of C-index also showed that the model was accurate in predicting the prognosis of BRCA in terms of OS, indicating that BRCA patients in the high-risk and low-risk groups were significantly different in prognostic situations (p-value < 0.05), indicating the possibility of the model as a prognostic marker.

Table 3. Concordance index and overall statistical tests for the Cox regression model.

Concordance = 0.667 (se = 0.025)
Likelihood ratio test = 36.06 on 4 df, p = 3e−07
Wald test= 37.51 on 4 df, p = 1e−07
Score (log rank) test =38.43 on 4 df, p = 9e−08

Note: C-index can evaluate the accuracy of various models in predicting outcomes. Among them. C-index. 0.5, without any predictive power, C-index. 0.5 - 0.7, low accuracy, C-index. 0.7 - 0.9, moderate accuracy, and C-index.0.9, high accuracy.95% CI calculated as. C-index −1.96 * number of standards ~ C-index + 1.96 * number of standards.

3.7. Prognostic Analysis of Key lncRNAs

The above 4 key lncRNAs (AC007731.1, ADAMTS9-AS1, C10orf126, and SRGAP3-AS2) were subjected to survival analysis (Figure 9). These 2 lncRNAs, lncRNA AC007731.1, ADAMTS9-AS1, were identified to be associated with the prognosis of breast cancer patients (p value less than 0.05). Among them, low expression of AC007731.1 was significantly associated with good prognosis in breast cancer patients.

3.8. Prediction of Potential Binding Protein Sites

Four lncRNAs, AC007731.1, ADAMTS9-AS1, C10orf126, and SRGAP3-AS2 from the prognostic model had 120 potential binding protein sites predicted by the catRAPID database. Among them, high, medium, and low predictivity profiles are shown in Figure 10. The potential binding protein enrichment profiles of the 4 key lncRNAs are shown in Figure 11. The binding protein sites with predicted interaction strength above 50% and discriminatory power in the top 10, respectively, are shown in Table 4. And the discriminatory power was higher than 75% to examine the accuracy of the prediction results, the scores of the predicted lncRNA potential binding protein sites were separately evaluated using the RPISeq (Table 5) database. The prediction results from combining the two databases indicated that QKI and DPOLA might be the potential binding proteins of lncRNA AC007731.1. HNRDL_HUMAN_151 - 218 may be a potential binding protein site common to lncRNA ADAMTS9-AS1 and SRGAP3-AS2. UT14C_HUMAN_18-736 may be a potential binding protein site for lncRNA C10orf126. Moreover, considering the

cellular localization of lncRNAs determines the potential mechanisms, we analyzed the subcellular localization of 4 key lncRNAs by lncLocator. As shown in Figure 12, AC007731.1 was mainly located in the cytoplasm, ADAMTS9-AS1 was mainly located in the cytoplasm, but the other two lncRNAs (C10orf126 and SRGAP3-AS2) were mainly distributed in the nucleus.

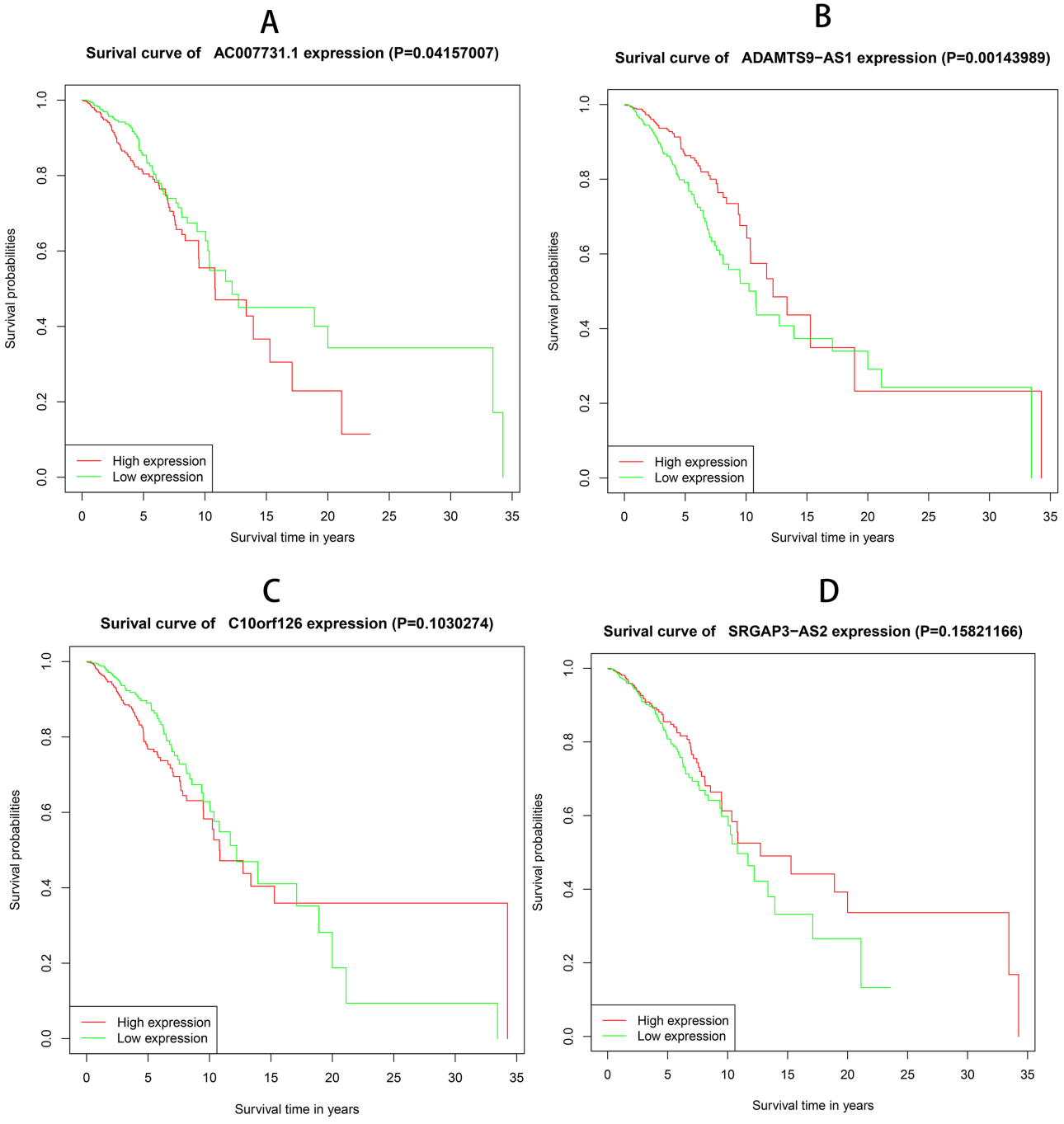


Figure 9. Kaplan-Meier survival curves of the four key lncRNAs in the analyzed breast cancer cohort. (A) AC007731.1. (B) ADAMTS9-AS1. (C) C10orf126. (D) SRGAP3-AS2.

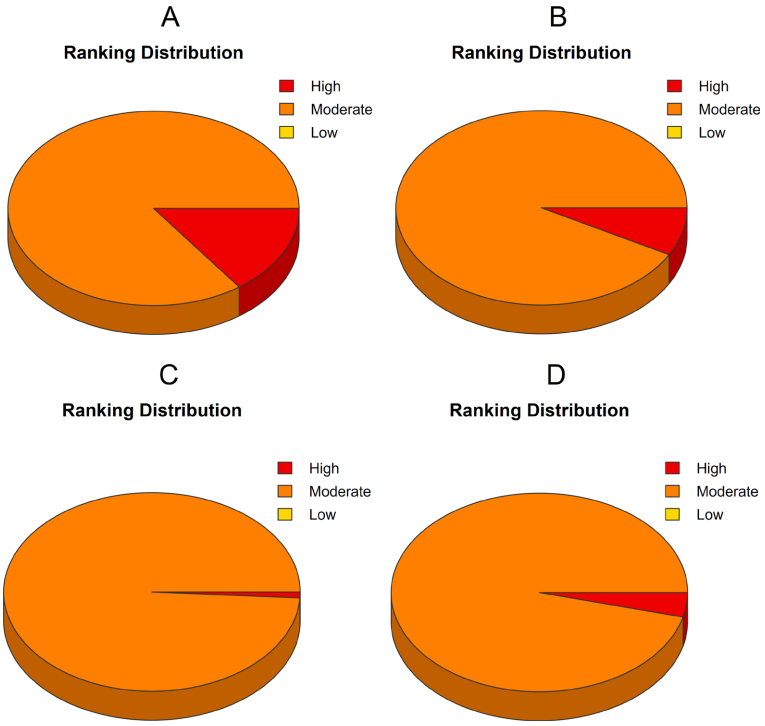


Figure 10. High, medium and low predictivity profiles. Red indicates high predictivity, orange indicates moderate predictivity, and yellow indicates low predictivity. (A) AC007731.1. (B) ADAMTS9-AS1. (C) C10orf126. (D) SRGAP3-AS2.

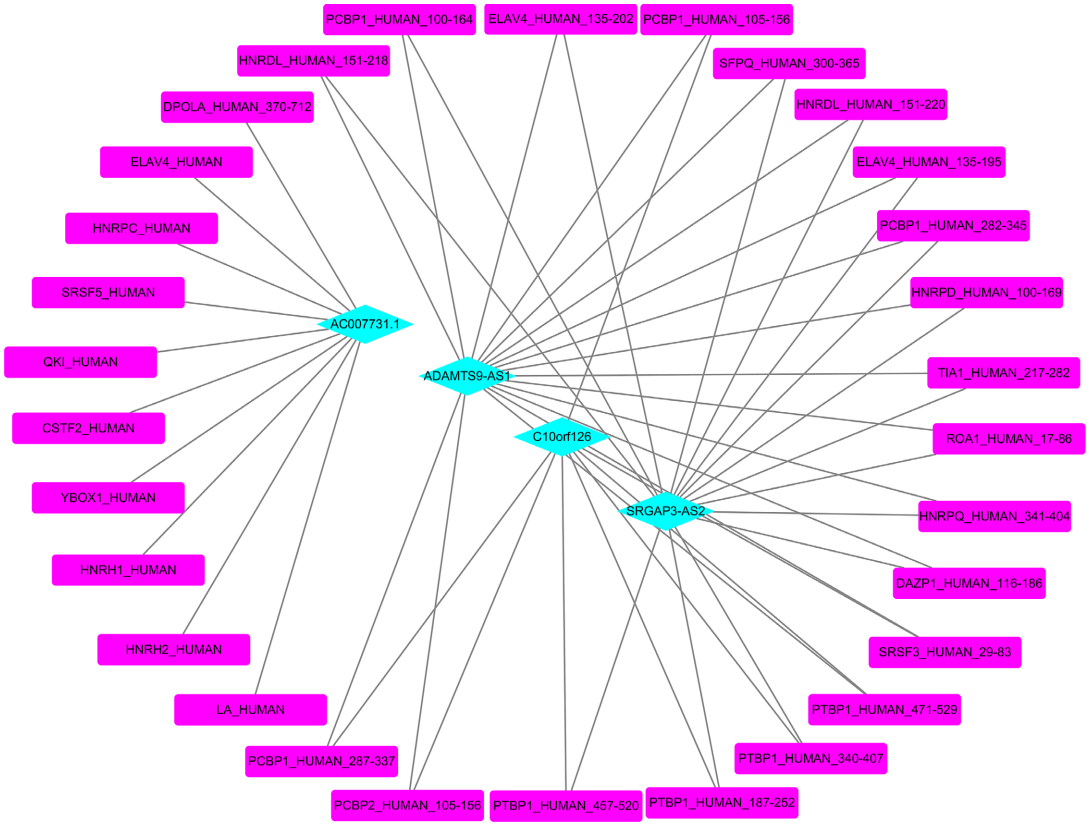


Figure 11. Enrichment of potential binding proteins of 4 key lncRNAs. The blue diamond represents lncRNAs, and the magenta round rectangles represent predicted protein binding sites.

Table 4. The predicted protein binding sites of 4 key lncRNAs in catRAPID database.

Protein binding site of AC007731.1 (AC007731.16)					
Sequence	RNA binding Protein sites	RNA ID	Z-score	Discriminative Power (%)	Interaction Strength (%)
1	DPOLA_HUMAN_370-712	AC007731.16. Hom_1_94744-94926	1.52	99	100
2	QKI_HUMAN	AC007731.16. Hom_1_97078-97267	0.66	94	99
3	YBOX1_HUMAN	AC007731.16. Hom_1_79278-79380	0.72	95	100
4	YBOX1_HUMAN	AC007731.16. Hom_1_153328-153431	0.65	93	100
5	LA_HUMAN	AC007731.16. Hom_1_48677-48868	0.62	92	100
6	LA_HUMAN	AC007731.16. Hom_1_48689-48871	0.42	87	99
7	YBOX1_HUMAN	AC007731.16. Hom_1_62331-62460	0.36	84	100
8	QKI_HUMAN	AC007731.16. Hom_1_48689-48871	0.35	83	97
9	LA_HUMAN	AC007731.16. Hom_1_160229-160420	0.33	81	100
10	YBOX1_HUMAN	AC007731.16. Hom_1_62335-62456	0.33	81	99
Protein binding site of ADAMTS9-AS1					
Sequence	RNA binding Protein sites	RNA ID	Z-score	Discriminative Power (%)	Interaction Strength (%)
1	HNRDL_HUMAN_151-218	NR_110151.1. Hom_1_299-465	−0.15	47	97
2	HNRDL_HUMAN_151-218	NR_110151.1. Hom_1_296-468	−0.17	45	95
5	SFPQ_HUMAN_300-365	NR_110151.1. Hom_1_296-468	−0.18	45	94
6	ELAV3_HUMAN_128-192	NR_110151.1. Hom_1_296-468	−0.18	45	94
7	ELAV3_HUMAN_128-189	NR_110151.1. Hom_1_296-468	−0.19	45	94
3	HNRDL_HUMAN_151-218	NR_110151.1. Hom_1_315-430	−0.26	37	94
4	HNRDL_HUMAN_151-220	NR_110151.1. Hom_1_315-430	−0.28	35	94
8	ELAV3_HUMAN_128-192	NR_110151.1. Hom_1_315-430	−0.27	35	93
9	ELAV4_HUMAN_135-202	NR_110151.1. Hom_1_315-430	−0.27	35	93
10	SFPQ_HUMAN_300-365	NR_110151.1. Hom_1_315-430	−0.27	35	93
Protein binding site of C10orf126					
Sequence	RNA binding Protein sites	RNA ID	Z-score	Discriminative Power (%)	Interaction Strength (%)
1	UT14C_HUMAN_18-736	C10orf126	−0.14	50	56
2	MPP10_HUMAN_18-659	C10orf126	−0.07	54	82
3	RRP44_HUMAN_468-794	C10orf126	−0.24	40	59
4	DNS2A_HUMAN_18-349	C10orf126	−0.27	35	54
5	UT14A_HUMAN_17-737	C10orf126	−0.31	33	44
6	DNS2B_HUMAN_31-362	C10orf126	−0.41	24	38
7	PUR9_HUMAN_136-464	C10orf126	−0.43	24	36

Continued

8	SNR27_HUMAN_2-156	C10orf126	−0.52	20	34
9	LEO1_HUMAN_326-624	C10orf126	−0.58	17	24
10	ZBT44_HUMAN_385-435	C10orf126	−0.57	17	22
Protein binding site of SRGAP3-AS2					
Sequence	RNA binding Protein sites	RNA ID	Z-score	Discriminative Power (%)	Interaction Strength (%)
1	HNRDL_HUMAN_151-218	NR_121663.1. Hom_1_1082-1273	−0.2	42	92
2	HNRDL_HUMAN_151-220	NR_121663.1. Hom_1_1082-1273	−0.23	40	90
3	ROA1_HUMAN_17-86	NR_121663.1. Hom_1_1082-1273	−0.26	37	86
4	ELAV4_HUMAN_135-202	NR_121663.1. Hom_1_1082-1273	−0.27	37	83
5	HNRDL_HUMAN_151-218	NR_121663.1. Hom_1_1089-1271	−0.27	35	84
6	HNRPD_HUMAN_100-169	NR_121663.1. Hom_1_1082-1273	−0.28	35	83
7	ELAV2_HUMAN_128-195	NR_121663.1. Hom_1_1082-1273	−0.27	35	82
8	TIA1_HUMAN_10-79	NR_121663.1. Hom_1_1082-1273	−0.29	35	81
9	SFPQ_HUMAN_300-365	NR_121663.1. Hom_1_1082-1273	−0.29	35	77
10	HNRDL_HUMAN_151-220	NR_121663.1. Hom_1_1089-1271	−0.3	33	81

Note: Only binding protein sites with predicted interaction strength above 50% and discriminatory power in the top 10 are included in the table. Discriminatory power. used to assess the fast interaction propensity, it represents the confidence in the prediction and ranges from 0% (unpredictability) to 100% (predictability). DP values above 50% indicate possible interactions, while above 75% indicate high confidence predictions. Interaction strength was calculated using a reference set consisting of 100 random proteins and 100 random RNA-sequences of the same length as the molecule under study. The intensity ranged from 0% (non-specific) to 100% (specific). Intensity values above 50% indicate high specificity of interaction.

Table 5. The predicted protein binding sites of 4 key lncRNAs in RPISeq database.

Protein binding site of AC007731.1(AC007731.16)				
Sequence	RNA ID	RNA binding protein sites	Prediction using SVM classifier	Prediction using RF classifier
1	AC007731.16	QKI_HUMAN	0.95	0.80
2	AC007731.16	DPOLA_HUMAN	0.92	0.80
3	AC007731.16	HNRH2_HUMAN	0.90	0.95
4	AC007731.16	YBOX1_HUMAN	0.76	1.00
5	AC007731.16	LA_HUMAN	0.70	0.80
Protein binding site of ADAMTS9-AS1				
Sequence	RNA ID	RNA binding protein sites	Prediction using SVM classifier	Prediction using RF classifier
1	ADAMTS9-AS1	HNRDL_HUMAN_151-218	0.99	0.90
2	ADAMTS9-AS1	HNRDL_HUMAN_151-220	0.99	0.90
3	ADAMTS9-AS1	SFPQ_HUMAN_300-365	0.99	0.85

Continued

4	ADAMTS9-AS1	ELAV3_HUMAN_128-192	0.98	0.85
5	ADAMTS9-AS1	ELAV3_HUMAN_128-189	0.98	0.85
Protein binding site of C10orf126				
Sequence	RNA ID	RNA binding protein sites	Prediction using SVM classifier	Prediction using RF classifier
1	C10orf126	UT14C_HUMAN_18-736	0.90	0.60
2	C10orf126	RRP44_HUMAN_468-794	0.97	0.55
3	C10orf126	UT14A_HUMAN_17-737	0.89	0.65
4	C10orf126	DNS2A_HUMAN_18-349	0.85	0.55
5	C10orf126	MPP10_HUMAN_18-659	0.79	0.60
Protein binding site of SRGAP3-AS2				
Sequence	RNA ID	RNA binding protein sites	Prediction using SVM classifier	Prediction using RF classifier
1	SRGAP3-AS2	HNRDL_HUMAN_151-218	0.95	0.80
2	SRGAP3-AS2	HNRDL_HUMAN_151-220	0.95	0.80
3	SRGAP3-AS2	ROA1_HUMAN_17-86	0.90	0.80
4	SRGAP3-AS2	HNRPD_HUMAN_100-169	0.90	0.70
5	SRGAP3-AS2	ELAV4_HUMAN_135-202	0.85	0.75

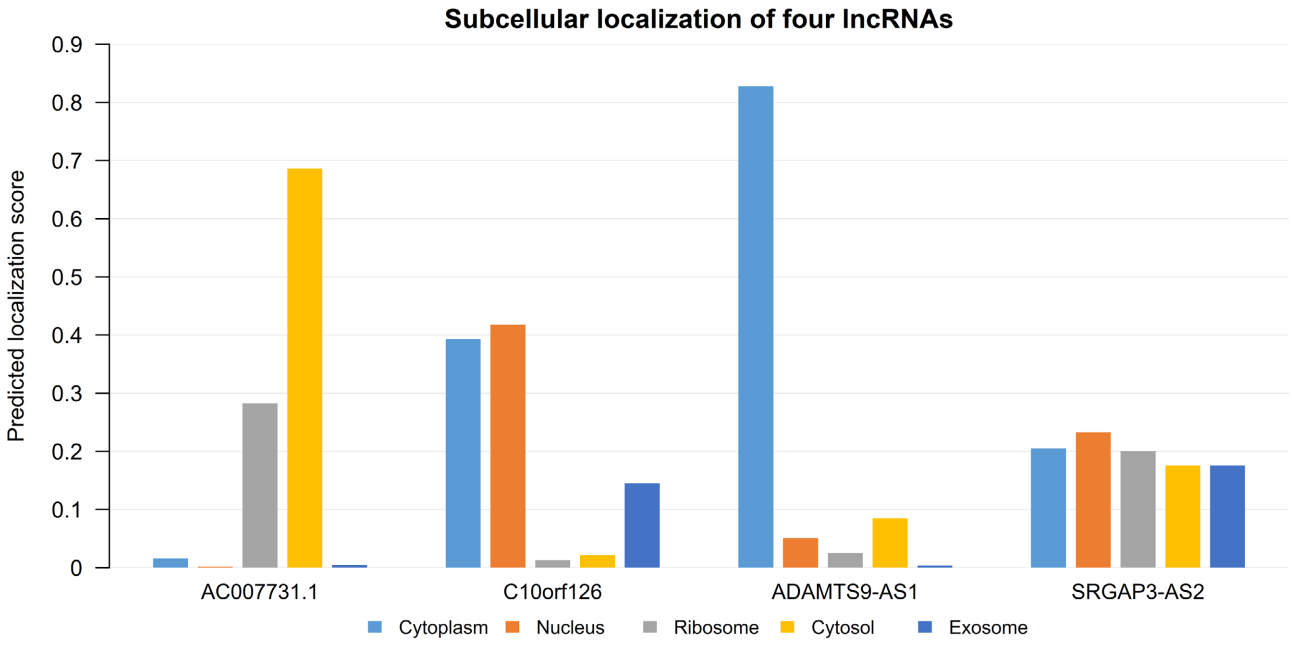


Figure 12. Subcellular localization of 4 key lncRNAs.

4. Discussion

Breast cancer is one of the most commonly diagnosed cancers and a leading cause

of cancer-related mortality among women worldwide.Current treatments for breast cancer include surgery, radiotherapy, chemotherapy, hormonal therapy, and immunotherapy, while CDK4/6 inhibitors and HER2-directed therapies have expanded options for selected breast cancer subtypes [2]-[4]. Recently, a novel mechanistic function of lncRNAs, known as ceRNAs, has been reported. lncRNAs sponge miRNAs to regulate their expression, thereby contributing to various pathological processes, including the development of cancer. lncRNA is also related to the treatment of breast cancer: it has been reported that lncRNA, ARA and ATB are associated with resistance to the breast cancer treatment of doxorubicin and trastuzumab, respectively [23] [24].

In this study, based on TCGA dataset with clinical data information, 1094 DElncRNAs, 86 DEmiRNAs, and 2130 DEmRNAs were identified between BRCA and normal breast tissue samples. To explore the potential relationship between the three, we constructed a unique lncRNA-miRNA-mRNA network based on the ceRNA hypothesis, which includes 81 DElncRNA, 18 DEmiRNA, and 27 DEmRNAs. To investigate the target genes in the ceRNA network, we enriched their transcription factors and the results indicated that the transcription factors (TFs) of DEmRNAs were mainly enriched on 5 genes (FOXO1, FREAC7, NFY, HFH1 and LHX3). The protein interaction relationships between target genes and transcription factors were further explored.

CeRNA-mediated regulation has been proposed and investigated as a potential mechanism in cancer, providing a direction for further research [14] [15] [17]. After GO enrichment analysis using 27 differential mRNAs in our constructed ceRNA regulatory network, we found that 27 mRNAs were closely related to the positive regulation of the ERK1 and ERK2 cascades, regulating cellular response to growth factor stimulation and stem cell population maintenance. These findings suggest that the predicted lncRNA-miRNA-mRNA interactions may be associated with biological processes relevant to breast cancer progression and prognosis. However, the underlying regulatory mechanisms remain to be experimentally validated.

Multiple systems integrating biomarkers may improve predictive accuracy compared to a single biomarker. We used univariate Lasso regression and multivariate Cox regression analyses for DElncRNAs involved in the ceRNA regulatory network of breast cancer. Ultimately, a novel lncRNA was constructed with 4 (AC007731.1, ADAMTS9-AS1, C10orf126, and SRGAP3-AS2) in a prognostic model to predict the OS of BRCA patients. In the TCGA training set, patients were divided into low-and high-risk groups according to the prognostic score calculated from the lncRNA signature. Survival curves indicated a significant difference in OS between the two groups of patients. The four-gene risk score showed modest apparent discrimination for overall survival in the development cohort, with time-dependent AUCs of 0.663 at 3 years and 0.659 at 5 years and a C-index of 0.667. These estimates may be optimistic because performance was assessed in the same cohort used for feature selection and model fitting. Internal validation that

repeats both steps is needed to quantify this optimism. These results suggest that the four lncRNAs of AC007731.1, ADAMTS9-AS1, C10orf126, and SRGAP3-AS2 might serve as prognostic models in BRCA. In the individual survival analyses, overall survival differed significantly between the high- and low-expression groups for AC007731.1 and ADAMTS9-AS1. The differences for C10orf126 and SRGAP3-AS2 were not statistically significant. These analyses compared survival between expression groups for each lncRNA separately. The multivariable Cox model estimated the association between each lncRNA's continuous expression level and overall survival after accounting for the other lncRNAs in the model. Because the two approaches address different questions, statistical significance can differ between them. The four lncRNAs jointly define the prognostic risk score, and whether they cooperate biologically remains unresolved. The predicted shared binding site in HNRDL for ADAMTS9-AS1 and SRGAP3-AS2 provides a testable hypothesis about protein binding. We performed cellular localization analysis of these four lncRNAs, which showed that AC007731.1 was mainly located in the cytoplasm, ADAMTS9-AS1 was mainly located in cytoplasm, and two other lncRNAs (C10orf126 and SRGAP3-AS2) were mainly distributed in the nucleus. The results suggest that these four genes may have an impact on the development of breast cancer at different cell sites.

However, we retrieved and found that little is known about their role in tumorigenesis and progression. Currently, in the study of these 4 lncRNAs, ADAMTS9-AS1 was demonstrated to play an important role in the progression and prognosis of cancer. ADAMTS9-AS1 is an antisense lncRNA, and studies of breast cancer have linked diverse lncRNAs to tumor-associated phenotypes [11] [18]. Other studies have reported associations between lncRNA signatures and breast cancer prognosis [9] [12]. Moreover, lncRNA ADAMTS9-AS1 suppresses invasive features of breast cancer cells by sponging mir-513a-5p [20]. Long non-coding RNA ADAMTS9-AS1 exacerbates cell proliferation, migration, and invasion by triggering the PI3K/Akt/mTOR pathway in hepatocellular carcinoma cells [25]. Long noncoding RNA ADAMTS9-AS1inhibits prostate cancer progression by regulating the mir-142-5p/CCND1 axis [26]. AC007731.1, C10orf126, and SRGAP3-AS2 have also been included in a previously published eight-lncRNA breast cancer prognostic model [10].

In summary, we constructed a four lncRNA significantly associated with OS of BRCA patients (AC007731.1, ADAMTS9-AS1, C10orf126, and SRGAP3-AS2) in a prognostic model. Low prognostic risk patients can be accurately identified from high prognostic risk patients. However, whether these 4 lncRNAs can be novel prognostic markers for breast cancer or novel targets for the disease still needs further validation in clinical samples. Transparent reporting of prediction model development and evaluation is recommended by the TRIPOD+AI statement [27].

5. Conclusion

A prognostic model based on four lncRNAs (AC007731.1, ADAMTS9-AS1, C10orf126, and SRGAP3-AS2) was associated with overall survival in patients

with breast cancer. Significant survival differences were observed between the high-risk and low-risk groups. This four-lncRNA signature may serve as a potential prognostic biomarker, although further validation is needed.

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Availability of Data and Materials

The datasets supporting the conclusions of this article are available in the following public domain resources: TCGA database (https://portal.gdc.cancer.gov/); GENCODE database (https://www.gencodegenes.org/); miRcode database (http://www.mircode.org/) (https://doi.org/10.1093/bioinformatics/bts344); miRDB database (https://mirdb.org/); TargetScan database (https://www.targetscan.org/vert_72/); miRTarBase database (https://mirtarbase.cuhk.edu.cn/)(https://doi.org/10.1093/nar/gkae1072); DAVID database (https://davidbioinformatics.nih.gov/); String database (https://www.string-db.org/); NCBI database (https://www.ncbi.nlm.nih.gov/gene); CatRAPID database (https://tools.tartaglialab.com/page/catrapid_omics2_group); RPISeq database (http://pridb.gdcb.iastate.edu/RPISeq/batch-rna.html) (https://doi.org/10.1186/1471-2105-12-489); LncLocator database (http://www.csbio.sjtu.edu.cn/bioinf/lncLocator/).

Funding

The project was supported by Key Laboratory of Basic Research on Regional Diseases (Guangxi Medical University), Education Department of Guangxi Zhuang Autonomous Region (Grant No. GXOYJB2025001); Fujian Provincial Key Laboratory of Innovative Drug Target Research (FJ-YW-2025KF06).

Authors' Contributions

Ying Rao: Data curation, Formal analysis, Visualization, Writing—original draft. Lin-Li Zhang: Conceptualization, Methodology, Formal analysis, Writing—original draft. Ming-Tao Luo: Investigation, Data curation. Xuan-Hua Chen: Formal analysis, Visualization. Yong-Le Li: Investigation, Validation. Yi-Xin Liu: Validation, Writing—review & editing. Li-He Jiang: Conceptualization, Supervision, Funding acquisition, Writing—review & editing.

Acknowledgements

We acknowledge TCGA database for providing their platforms and contributors

for uploading their meaningful datasets.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., *et al.* (2024) Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA*: *A Cancer Journal for Clinicians*, **74**, 229-263. https://doi.org/10.3322/caac.21834

[2] Morrison, L., Loibl, S. and Turner, N.C. (2023) The CDK4/6 Inhibitor Revolution—A Game-Changing Era for Breast Cancer Treatment. *Nature Reviews Clinical Oncology*, **21**, 89-105. https://doi.org/10.1038/s41571-023-00840-4

[3] Boscolo Bielo, L., Trapani, D., Nicolò, E., Valenza, C., Guidi, L., Belli, C., *et al.* (2024) The Evolving Landscape of Metastatic HER2-Positive, Hormone Receptor-Positive Breast Cancer. *Cancer Treatment Reviews*, **128**, Article 102761. https://doi.org/10.1016/j.ctrv.2024.102761

[4] Lloyd, M.R., Jhaveri, K., Kalinsky, K., Bardia, A. and Wander, S.A. (2024) Precision Therapeutics and Emerging Strategies for HR-Positive Metastatic Breast Cancer. *Nature Reviews Clinical Oncology*, **21**, 743-761. https://doi.org/10.1038/s41571-024-00935-6

[5] Fatica, A. and Bozzoni, I. (2013) Long Non-Coding RNAs: New Players in Cell Differentiation and Development. *Nature Reviews Genetics*, **15**, 7-21. https://doi.org/10.1038/nrg3606

[6] Mercer, T.R., Dinger, M.E. and Mattick, J.S. (2009) Long Non-Coding RNAs: Insights into Functions. *Nature Reviews Genetics*, **10**, 155-159. https://doi.org/10.1038/nrg2521

[7] Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., *et al.* (2015) The Landscape of Long Noncoding RNAs in the Human Transcriptome. *Nature Genetics*, **47**, 199-208. https://doi.org/10.1038/ng.3192

[8] Qian, D., Zheng, Q., Wu, D., Ye, B., Qian, Y., Zhou, T., *et al.* (2021) Integrated Analysis of Cerna Network Reveals Prognostic and Metastasis Associated Biomarkers in Breast Cancer. *Frontiers in Oncology*, **11**, Article ID: 670138. https://doi.org/10.3389/fonc.2021.670138

[9] Meng, J., Li, P., Zhang, Q., Yang, Z. and Fu, S. (2014) A Four-Long Non-Coding RNA Signature in Predicting Breast Cancer Survival. *Journal of Experimental & Clinical Cancer Research*, **33**, Article No. 84. https://doi.org/10.1186/s13046-014-0084-7

[10] Sun, M., Wu, D., Zhou, K., Li, H., Gong, X., Wei, Q., *et al.* (2019) An Eight-LncRNA Signature Predicts Survival of Breast Cancer Patients: A Comprehensive Study Based on Weighted Gene Co-Expression Network Analysis and Competing Endogenous RNA Network. *Breast Cancer Research and Treatment*, **175**, 59-75. https://doi.org/10.1007/s10549-019-05147-6

[11] Wasson, M.D., Venkatesh, J., Cahill, H.F., McLean, M.E., Dean, C.A. and Marcato, P. (2024) LncRNAs Exhibit Subtype-Specific Expression, Survival Associations, and Cancer-Promoting Effects in Breast Cancer. *Gene*, **901**, Article 148165. https://doi.org/10.1016/j.gene.2024.148165

[12] Yin, F., Zhao, W., Ding, C., Hou, C., Wang, S., Sun, C., *et al.* (2024) A Novel Cellular

Senescence-Related LncRNA Signature for Predicting the Prognosis of Breast Cancer Patients. *Journal of Cancer*, **15**, 4700-4716. https://doi.org/10.7150/jca.96107

[13] Bahramy, A., Zafari, N., Rajabi, F., Aghakhani, A., Jayedi, A., Khaboushan, A.S., *et al.* (2023) Prognostic and Diagnostic Values of Non-Coding RNAs as Biomarkers for Breast Cancer: An Umbrella Review and Pan-Cancer Analysis. *Frontiers in Molecular Biosciences*, **10**, Article ID: 1096524. https://doi.org/10.3389/fmolb.2023.1096524

[14] Salmena, L., Poliseno, L., Tay, Y., Kats, L. and Pandolfi, P. (2011) A Cerna Hypothesis: The Rosetta Stone of a Hidden RNA Language? *Cell*, **146**, 353-358. https://doi.org/10.1016/j.cell.2011.07.014

[15] Karreth, F.A. and Pandolfi, P.P. (2013) Cerna Cross-Talk in Cancer: When Ce-Bling Rivalries Go Awry. *Cancer Discovery*, **3**, 1113-1121. https://doi.org/10.1158/2159-8290.cd-13-0202

[16] Tan, J.Y., Sirey, T., Honti, F., Graham, B., Piovesan, A., Merkenschlager, M., *et al.* (2015) Extensive MicroRNA-Mediated Crosstalk between LncRNAs and MRNAs in Mouse Embryonic Stem Cells. *Genome Research*, **25**, 655-666. https://doi.org/10.1101/gr.181974.114

[17] Shariatmadar Taleghani, A., Zohrab Beigi, Y., Zare-Mirakabad, F. and Masoudi-Nejad, A. (2023) Exploring Cerna Networks for Key Biomarkers in Breast Cancer Subtypes and Immune Regulation. *Scientific Reports*, **13**, Article No. 20795. https://doi.org/10.1038/s41598-023-47816-z

[18] Li, Y., Zhou, M., Yang, L., Liu, S., Yang, L., Xu, B., *et al.* (2024) LncRNA DDX11-AS1 Promotes Breast Cancer Progression by Targeting the miR-30c-5p/MTDH Axis. *Scientific Reports*, **14**, Article No. 26745. https://doi.org/10.1038/s41598-024-78413-3

[19] Xiong, H., Shen, J., Chen, Z., Yang, J., Xie, B., Jia, Y., *et al.* (2020) H19/Let-7/Lin28 Cerna Network Mediates Autophagy Inhibiting Epithelial-Mesenchymal Transition in Breast Cancer. *International Journal of Oncology*, **56**, 794-806. https://doi.org/10.3892/ijo.2020.4967

[20] Fang, S., Zhao, Y. and Hu, X. (2020) lncRNA ADAMTS9-AS1 Restrains the Aggressive Traits of Breast Carcinoma Cells via Sponging miR-513a-5p. *Cancer Management and Research*, **12**, 10693-10703. https://doi.org/10.2147/cmar.s266575

[21] Fan, C., Ma, L. and Liu, N. (2018) Comprehensive Analysis of Novel Three-long Noncoding RNA Signatures as a Diagnostic and Prognostic Biomarkers of Human Triple-Negative Breast Cancer. *Journal of Cellular Biochemistry*, **120**, 3185-3196. https://doi.org/10.1002/jcb.27584

[22] Vats, P., Baweja, B., Nirmal, S., Singh, A. and Nema, R. (2025) TMPO-As1-hsa-let-7b-5p-EZH2-RNA Network Predicts Poor Survival in Basal-Like Breast Cancer Patients. *Reports of Practical Oncology and Radiotherapy*, **30**, 185-201. https://doi.org/10.5603/rpor.105252

[23] Shi, S., Wang, L., Yu, B., Li, Y., Jin, Y. and Bai, X. (2015) LncRNA-ATB Promotes Trastuzumab Resistance and Invasion-Metastasis Cascade in Breast Cancer. *Oncotarget*, **6**, 11652-11663. https://doi.org/10.18632/oncotarget.3457

[24] Jiang, M., Huang, O., Xie, Z., Wu, S., Zhang, X., Shen, A., *et al.* (2014) A Novel Long Non-Coding RNA-ARA: Adriamycin Resistance Associated. *Biochemical Pharmacology*, **87**, 254-283. https://doi.org/10.1016/j.bcp.2013.10.020

[25] Zhang, Z., Li, H., Hu, Y. and Wang, F. (2020) Long Non-Coding RNA Adamts9-as1 Exacerbates Cell Proliferation, Migration, and Invasion via Triggering of the PI3K/Akt/mTOR Pathway in Hepatocellular Carcinoma Cells. *American Journal of Translational Research*, **12**, 5696-5707.

[26] Zhou, Z., Wu, X., Zhou, Y. and Yan, W. (2021) Long Non-Coding RNA ADAMTS9-AS1 Inhibits the Progression of Prostate Cancer by Modulating the miR-142-5p/CCND1 Axis. *The Journal of Gene Medicine*, **23**, e3331. https://doi.org/10.1002/jgm.3331

[27] Collins, G.S., Moons, K.G.M., Dhiman, P., Riley, R.D., Beam, A.L., Van Calster, B., *et al.* (2024) TRIPOD + AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods. *BMJ*, **385**, e078378. https://doi.org/10.1136/bmj-2023-078378

List of Abbreviations

CeRNA	Competing endogenous RNA
DE	Differentially expressed
LncRNA	Long noncoding RNA
DElncRNAs	Differentially expressed lncRNAs
DEmiRNAs	Differentially expressed miRNAs
DEmRNAs	Differentially expressed mRNAs
BRCA	Breast cancer
TCGA	The Cancer Genome Atlas
OS	Overall survival
TFs	Transcription factors
ROC	Receiver operating characteristic curve
C-index	Concordance index calculation
KM	Kaplan-Meier
GO	Gene Ontology
AUC	The area under the ROC curve
FDR	False discovery rate
BP	Biological process
CC	Cellular component
MF	Molecular function