Guo *et al. Cancer Cell International* (2025) 25:233
https://doi.org/10.1186/s12935-025-03832-9

Cancer Cell International

RESEARCH Open Access

Prognostic value of a lncRNA signature in early-stage invasive breast cancer patients

Rong Guo[1,2†], Xiting Wang[1,2†], Yinju Yang[1,2†], Jiaying Zou[1,2†], Ming Li[1,2], Zeying Li[1,2], Yuan Yan[1,2], Nan Lan[1,2], JianYun Nie[1,2*], Yiyin Tang[1,2*] and Guojun Zhang[1,2*]

Abstract

Background Existing staging approaches fall short in precisely forecasting the likelihood of recurrence and survival outcomes among patients undergoing surgery for early-stage breast cancer (EBC). Our study hypothesized that multivariate long non-coding RNA (lncRNA) expression profiles, when systematically integrated into a composite model, may synergistically refine postoperative risk categorization and enhance prognostic forecasting precision in this patient cohort.

Methods For the discovery set, lncRNA expression profiling associated with breast cancer progression was discovered by analyzing the differential expression profiles in three paired primary breast cancer tumor tissues and liver metastases. We found 12 distinctially expressed lncRNAs. A total of 400 patients were consecutively recruited and randomized to either training group or validation group. We first confirmed the expression of these lncRNAs using qRT-PCR. Subsequently, employing the LASSO Cox regression model with five lncRNA features as covariates, we constructed a five-lncRNA signature. We then validated this signature in an independent cohort to assess its prognostic and predictive capabilities in disease-free survival (DFS) duration.

Results We constructed a classifier using the LASSO model, incorporating five specific lncRNAs: CBR3-AS1, HNF4A-AS1, LINC00622, LINC00993 and LINC00342. Utilizing this tool, we successfully stratified patients into two distinct categories: high- and low-risk groups. Significant differences were observed in both DFS and overall survival (OS) between the two groups. Within the initial patient cohort, significant differences of 5-year DFS was observed across high- and low-risk group (61.1% vs. 92.2%, HR 6.3, 95% CI 3.5–11.6; $P < 0.001$). The 5-year DFS rate was 72.9% and 85.4% for high- and low-risk group respectively in validation cohort (HR 2.6, 95% CI: 1.5–4.5; $P = 0.001$). The 5-lncRNA signature emerged as an independent prognostic indicator, demonstrating superior prognostic value compared to conventional clinicopathological risk factors.

†Rong Guo, Xiting Wang, Yinju Yang, and Jiaying Zou have contributed equally to this work.

*Correspondence:
JianYun Nie
njyvip@sina.com
Yiyin Tang
tyy1485@163.com
Guojun Zhang
zhangguojun@kmmu.edu.cn

Full list of author information is available at the end of the article

Conclusions The integrated model combining 5-lncRNA molecular signature with clinical parameters demonstrates significant prognostic stratification capacity and therapeutic decision-making value in EBC management. It may help patients consult and personalize disease management.

Keywords Breast cancer, LncRNAs signature, Prognosis

Introduction

Most of breast cancer patients are early diagnosed. However, recurrence and metastasis occur after surgery in about 20–30% of these EBC patients [1]. Conventional recurrence risk stratification in EBC primarily relies on clinicopathological parameters including advanced tumor stage, extensive nodal involvement, high histological grade, along with biomarkers such as hormone receptor negativity/HER2 positivity, and younger patient age [2–4]. Nevertheless, current risk stratification systems exhibit limited discriminatory capacity in precisely delineating recurrence risk tiers, while failing to identify subsets requiring therapeutic intensification due to inherent biological heterogeneity in EBC progression. Therefore, incorporating prognostic molecular biomarkers into the existing staging system would enhance the precision of risk assessment and support more tailored therapeutic strategies.

lncRNAs are essential for cancer progression, including growth, metastasis, and drug sensitivity [5, 6]. Integrating clinicopathological factors into a nomogram represents a highly promising strategy, with the potential to significantly influence clinical decision-making [7, 8]. Systematic characterization of lncRNA regulatory networks has been documented in breast oncology research, and examined their potential clinical relevance [9, 10]. The concept of lncRNA signatures has garnered growing interest, as they may serve as potential predictors and prognostic biomarkers [11, 12]. Thus, integration of clinicopathological factors and lncRNA signature into a predicting model would enhance the accuracy and specificity of the model. However, validation in clinical settings is still needed for these biomarkers and have not yet been used in routine clinical practice.

When screening a wide range of biomarkers simultaneously using high-throughput RNA sequencing, the number of covariates often outnumbers the observations. Traditional multivariable analysis methods, such as Cox-based analysis, have been widely used for building covariate models in survival data. However, these traditional multi-variable analysis methods are not suitable for high-dimensional datasets and particularly fail to handle scenarios with the low ratio of sample size to variables [13, 14]. Alternatively, the selection operator method LASSO and least absolute shrinkage method was applied to overcome this limitation [15, 16].

In this research endeavor, we constructed a multi-lncRNA signature using the LASSO Cox regression approach for forecasting DFS and OS for EBC patients who had already received surgery. Patients assigned a low-risk score exhibit better DFS and OS outcomes relative to their high-risk counterparts. Based on the signature, we developed a nomogram integrating a multi-lncRNA profile with clinicopathological variables to predict patient outcomes. Prognostic utility of this combined lncRNA-clinical model underwent systematic appraisal in training and validation cohorts. We further evaluated the prognostic performance of the classifier relative to single lncRNAs and traditional clinicopathological factors.

Methods

Patient enrollment

Consecutive patients who were diagnosed with EBC and underwent breast surgery between January 1, 2019, and December 31, 2019, were recruited for the current cohort study. We collected the primary tissues of those patients prospectively.

Primary samples of breast cancer patients who met all the criteria below: (1) Female patients with a diagnosis of unilateral breast disease. (2) Histopathological verification of invasive carcinoma. (3) Initial pathological diagnosed with ductal invasive breast cancer. (4) Patients without any evidence of distant metastasis at diagnosis. (5) Sufficient tissues conserved in RNA later and available for further research. The elimination criteria include: bilateral breast cancer, breast carcinoma in situ, and T4 stage.

Discovery cohort

Using the RNA-seq technique, we revealed a highly diverse lncRNA expression profile in paired breast cancer tumor and liver metastasis tissues. We performed a significance analysis of RNA-seq data to identify differentially expressed lncRNAs between primary cancer tissues and their paired liver metastases. When the expression level of LncRNAs changes by at least fivefold, we classify them as differentially expressed, with significance determined by a P-value less than 0.05. Then use average linkage method and non central Pearson correlation coefficient for hierarchical clustering.

Model development and validation cohorts

Building upon RNA-seq data, we further scrutinized breast cancer associated lncRNA expression levels via qRT-PCR to analyze in both training and validation

Table 1 Clinical characteristics of patients in the cohorts from breast cancer patients

	Training cohort (*n* = 200)							Internal validation cohort (*n* = 200)							*P* value (Training vs. Validation)
	all (*n* = 200)		high score (*n* = 72)		low score (*n* = 128)		*P* value (high vs. low)	all (*n* = 200)		high score (*n* = 97)		low score (*n* = 103)		*P* value (high vs. low)	
		%		%		%			%		%		%		
Median age, years	54		54		54			54		54		53			
Range	23-87		30-87		23-85			25-88		25-88		28-81			
20-35	8	4.00	1	1.39	7	5.47	0.332*	11	5.50	4	4.12	7	6.80	0.653	0.247
36-45	41	20.50	18	25.00	23	17.97		40	20.00	20	20.62	20	19.42		
46-55	61	30.50	20	27.78	41	32.03		70	35.00	33	34.02	37	35.92		
56-65	69	34.50	23	31.94	46	35.94		50	25.00	28	28.87	22	21.36		
66-	21	10.50	10	13.89	11	8.59		29	14.50	12	12.37	17	16.50		
pT stage															
pT1	62	31.00	21	29.17	41	32.03	0.137	82	41.00	27	27.84	55	53.40	0.001	0.063
pT2	131	65.50	46	63.89	85	66.41		108	54.00	64	65.98	44	42.72		
pT3	7	3.50	5	6.94	2	1.56		10	5.00	6	6.19	4	3.88		
pN stage															
pN0	88	44.00	26	36.11	62	48.44	0.166	99	49.50	42	43.30	57	55.34	0.116	0.672
pN1	60	30.00	23	31.94	37	28.91		53	26.50	33	34.02	20	19.42		
pN2	21	10.50	7	9.72	14	10.94		22	11.00	11	11.34	11	10.68		
pN3	31	15.50	16	22.22	15	11.72		26	13.00	11	11.34	15	14.56		
Stage															
I	36	18.00	9	12.50	27	21.09	0.069	49	24.50	15	15.46	34	33.01	0.002	0.278
II	107	53.50	36	50.00	71	55.47		100	50.00	60	61.86	40	38.83		
III	57	28.50	27	37.50	30	23.44		51	25.50	22	22.68	29	28.16		
ER status															
Positive	139	69.50	47	65.28	92	71.88	0.331	157	78.50	76	78.35	81	78.64	0.960	0.040
Negtive	61	30.50	25	34.72	36	28.13		43	21.50	21	21.65	22	21.36		
PR status															
Positive	158	79.00	56	77.78	102	79.69	0.750	156	78.00	72	74.23	84	81.55	0.211	0.808
Negtive	42	21.00	16	22.22	26	20.31		44	22.00	25	25.77	19	18.45		
HER2 status															
Positive	53	26.50	23	31.94	30	23.44	0.072	60	30.00	28	28.87	32	31.07	0.838	0.191
Negtive	135	67.50	42	58.33	93	72.66		135	67.50	66	68.04	69	66.99		
Unknow	12	6.00	7	9.72	5	3.91		5	2.50	3	3.09	2	1.94		
Grade															
1	4	2.00	0	0.00	4	3.13	0.106*	0	0.00	0	0.00	0	0.00	0.191	0.119*
2	118	59.00	37	51.39	81	63.28		108	54.00	46	47.42	62	60.19		
3	64	32.00	28	38.89	36	28.13		72	36.00	40	41.24	32	31.07		
Unknow	14	7.00	7	9.72	7	5.47		20	10.00	11	11.34	9	8.74		
Surgery for breast															

Table 1 (continued)

	Training cohort (*n* = 200)							Internal validation cohort (*n* = 200)							*P* value (Training vs. Validation)
	all (*n* = 200)		high score (*n* = 72)		low score (*n* = 128)		*P* value (high vs. low)	all (*n* = 200)		high score (*n* = 97)		low score (*n* = 103)		*P* value (high vs. low)	
BCS	18	9.00	7	9.72	11	8.59	0.789	19	9.50	6	6.19	13	12.62	0.121	0.863
Mastectomy	182	91.00	65	90.28	117	91.41		181	90.50	91	93.81	90	87.38		
LVI															
Negtive	102	51.00	36	50.00	66	51.56	0.966	116	58.00	53	54.64	63	61.17	0.633	0.137
Positive	93	46.50	34	47.22	59	46.09		75	37.50	39	40.21	36	34.95		
Unknow	5	2.50	2	2.78	3	2.34		9	4.50	5	5.15	4	3.88		

ER, estrogen recepter; PR, progestrone recepter; Her2, epidermal growth factor receptor-2; BCS, breast conserving surgery; LVI, lymphovascular invasion

P value from Chi-squared test or Fisher's exact test for nominal categories

*Fisher's exact test

cohort. As a result, a cohort of 400 consecutive breast cancer specimens were included. The Cohort was divided into two equal groups: 200 participants assigned to the training cohort, and another 200 participants assigned to the validation cohort. DFS was defined as the durition from the time of surgery to either the date of tumor recurrence or metastasis. For patients who did not experience any relapse during follow-up period, DFS was censored at the last follow-up time or time of death from other causes, respectively. We defined OS as the durition from the time of surgery until the time of death.

Fresh tissue specimens from patients were obtained during surgical procedures and preserved in RNAlater at -20℃. Extracting RNA using TRIzol® reagent. Vazyme Reverse Transcriptase enzyme was used to synthesize cDNA. Vazyme SYBR qPCR Master Mix was used to quantitate relative RNA levels. The qRT-PCR assay for each lncRNA was performed in triplicate for each patient. The sequences of specific primers for the lncRNA used can be found in Supplementary Table 1.

Statistical analysis

SPSS (v26; IBM Corp., USA) and R software (v3.5.2; R Foundation) were used for data analyses. Categorical variables statistics using χ^2 and Fisher's exact tests. Survival outcomes (DFS/OS) correlations were analyzed via Kaplan-Meier survival curves with log-rank comparisons. Multivariable analysis utilized Cox proportional hazards models. Optimal lncRNA expression cutoffs were determined through X-tile plots by maximizing DFS association. X-tile v3.6.1 (Yale School of Medicine) was used to generate X-tile plots.

LASSO Cox regression provides a regularization approach in high-dimensional data settings [15]. To select the top list relevant prognostic markers from the progression-associated lncRNAs identified in the discovery cohort, LASSO Cox regression was applied. A multi-lncRNA classifier was subsequently developed in the training set to predict DFS in patients included. The "glmnet" package in R software was used for LASSO analysis.

A nomogram was constructed with lncRNA signature and clinicopathological factors to predict prognosis of EBC patients. Calibration curves were constructed to evaluate concordance between nomogram-predicted probabilities and observed clinical outcomes. The 45° reference line indicates perfect model calibration (predicted-actual outcome exact alignment). All analyses were implemented using the rms package (Regression Modeling Strategies) in the R statistical environment (v4.2.1).

Time-dependent ROC analysis was used to assess the accuracy of each feature in predicting prognosis, including the multi-lncRNA classifier, and the nomogram [17].

We employed the survivalROC package in R platform to compute the area under ROC curve (AUC) for assessing the discrimination performance of the prediction model.

Results

Clinicopathological features of patients

A series of 200 patients of early-stage breast cancer were recruited. Patients were randomly divided into distinct training and validation groups described in the methodology section. The presenting clinical and pathological characteristics of participants were well balanced in each cohort inluding age, stage, molecularly defined subgroup, grade surgery type and lymph vascular invasion, which are summarized in Table 1. The study design is shown in Fig. 1A.

Construction of a tumor tissue LncRNA signature

In the discovery cohort, through hypothesis-free transcriptome profiling, we uncovered progression-relevant lncRNAs of breast cancer. RNA-seq was performed to identify lncRNAs differently expressed in 3 paired breast cancer primary tumor and liver metastasis tissues. From the pool of lncRNA candidates detected by RNA-seq, we identified 12 that were differentially expressed ($P < 0.05$). From among these, 6 lncRNAs were upregulated (CBR3-AS1, HNF4A-AS1, LINC02499, LINC00261, LINC00844 and LINC01554), and 6 lncRNAs (LINC00993, LINC01578, LINC00342, LINC02381, LBX2-AS1 and LINC00622) were downregulated in the liver metastases tissues (Fig. 1C). Hierarchical clustering analysis, guided by differentially expressed lncRNAs, successfully segregated primary tumor and metastatic liver tissue samples into two separate clusters (Fig. 1B).

Furthermore, we observed collinearity among those lncRNAs we mentioned above, which might distort the outcomes of ordinary Cox regression analysis. To address this, we utilized a LASSO Cox regression mothed to filter lncRNAs for predicting PFS in patients. The model incorporated 200 bootstrap replicates and relied on penalized maximum likelihood (Fig. 1D). We computed the regularization path over a range of lambda values, resulting in the selection of five lncRNAs (CBR3-AS1, HNF4A-AS1, LINC00622, LINC00993, and LINC00342) from the original 12. These findings were utilized in order to establish the optimal weighting coefficients that would enable the construction of a prognostic signature (Fig. 1E). Building upon the RNA quantification data of these five lncRNAs, we derived a formula to calculate prognostic risk score (RS)), were RS = (2.286 × expression of HNF4A-AS1) − (0.690 × expression of LINC00622) + (2.813 × expression of CBR3-AS1) − (3.329 × expression of LINC00993) − (0.136 × expression of LINC00342). Expression of each gene was calculated using the formular: expression value = $2-\log_{10}(\Delta Ct)$.

Evaluation of the RS formula

Using the RS formula, the training cohort patients were categorized into high- and low-risk subgroups. X-tile plots calculated the optimal cutoff score was 0.91 [18] (Supplementary Fig. 1A), was used based on its correlation with PFS. Clinical-pathological parameters showed adequate equilibrium between the study groups, as indicated by non-significant statistical differences (all $P > 0.05$). (Table 1).

Between patients in training cohort with high RS and low RS, the 5-year DFS was 61.1% vs. 92.2% (HR 6.329, 95% confidence interval (CI) 3.463–11.573; $P < 0.001$) (Fig. 2A), and 5-year OS was 88.9% vs. 99.2% (HR 3.315, 95% CI 1.550-11.573; $P = 0.002$) (Fig. 2B).

For additional validation, the effectiveness of these five lncRNA-based signatures was examined. The formula and cutoff setting were validated in the validation cohort. In the validation cohort, low score patients exhibited better prognosis outcomes than higher score patients (Fig. 2C, D). The 5-year DFS was 72.9% for high-score patients compared to 85.4% for low-score patients (HR 2.565, 95% CI 1.448–4.534; $P = 0.0012$) (Fig. 2C). The 5-year OS was 87.6% for high-score patients versus 94.2% for low-score patients (HR 2.070, 95% CI 0.913–4.696; $P = 0.0816$) (Fig. 2D).

Univariable and multivariable regression analyses in validation sets

Single-factor test revealed that each single lncRNA was a prognostic factor for EBC patients (Table 2). In multivariable analyses incorporating clinicopathological confounders, the 5-lncRNA signature retained independent prognostic significance across the full cohort ($n = 400$; HR = 3.584, 95% CI 2.227–5.769, $P < 0.001$). (Table 2). Furthermore, pN stage was an independent prognostic factor (Table 2).

Stratified COX analysis based on individual clinical and pathological features was performed when 5-lncRNA signature was considered as a continuous variable. The signature continued to be a significant model for predicting DFS, independent of clinicopathological variables, including age, ER, PR and HER2 status. (Fig. 2E).

Nomogram combined LncRNAs signature and clinical-related factors

To develop a quantitative method for predicting cancer progression probability, a nomogram were constructed. This nomogram included both clinical pathological factors and the 5-lncRNA signature (Fig. 3A). Calibration plots demonstrated the nomogram did well when compare with an ideal model (Fig. 3B).

AUC curve was calculated to evaluate the 5-lncRNA signature cumulative impact on risk prediction for both of the training and validation cohorts. DFS predictive

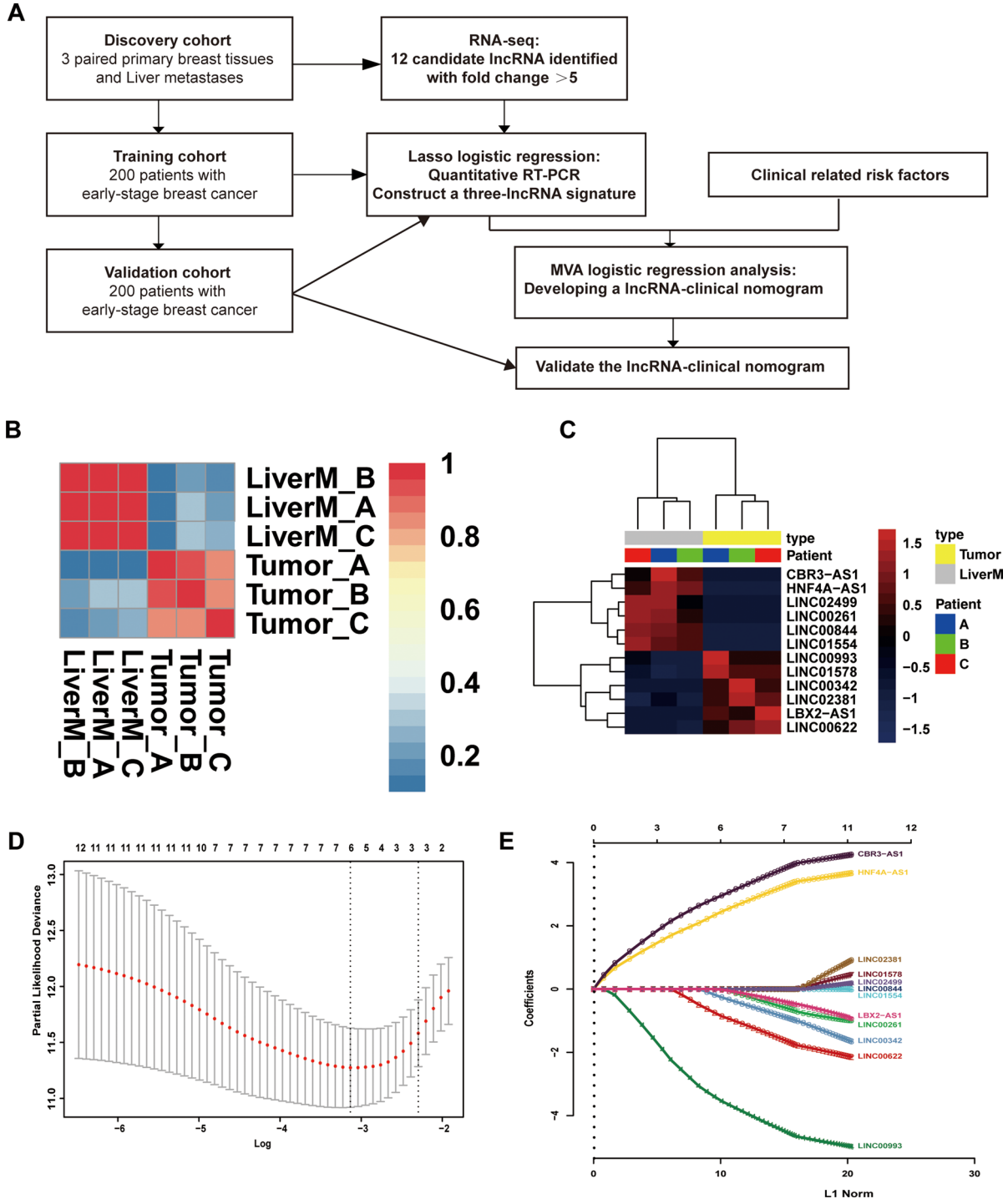

Fig. 1 Establishing a molecular signature comprising five long non-coding RNAs (**A**) Procedures diagram for building predictive signature and validation in early-stage breast cancer patients. (**B**) Hierarchical clustering investigation demonstrated the collinearity of three paired primary breast cancer tissues and liver metastasis tissue. A correlation matrix heatmap was generated for each tissue type, with each cell in the matrix depicted Pearson correlation coefficient of the corresponding row and column tissues. The color gradient in the legend corresponds to the range of correlation coefficients. (**C**) Hierarchical clustering of 12 lncRNAs expression level in three paired primary breast cancer tissues (in yellow) and liver metastasis tissues (in grey). In this representation, rows correspond to distinct lncRNA entries while columns reflect individual biological samples. Relative expression intensities are visualized through pseudocolor gradients, scaled from – 1.5 (lowest) to 1.5 (highest). (**D**) Determine the optimal parameter selection for LASSO regression through cross validation. Select the optimal value based on the minimum criterion and 1-SE, two dotted vertivcal lines were drawn. (**E**) LASSO-based coefficient profiles of the 12 EBC-associated lncRNAs in the discovery cohort

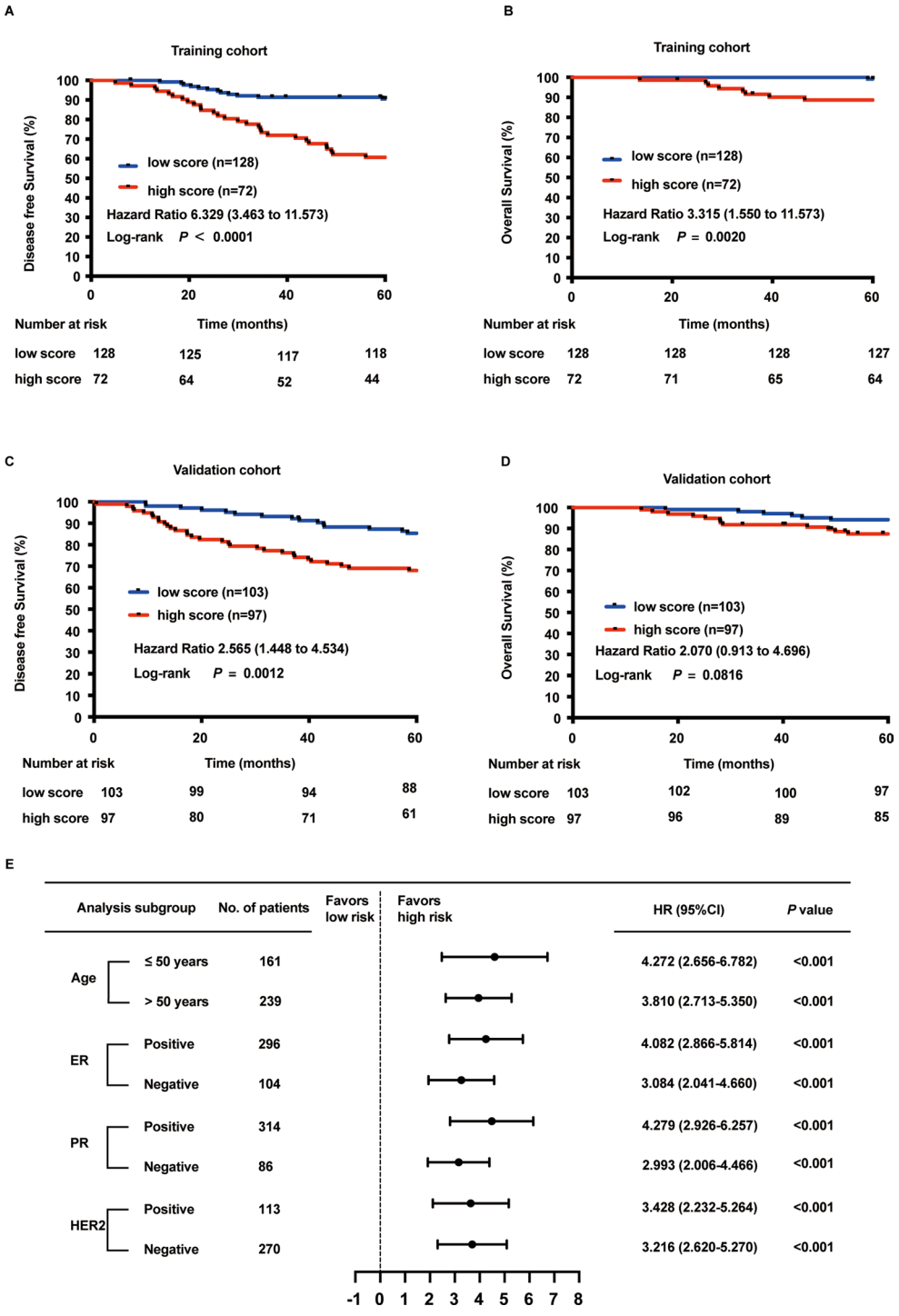

Fig. 2 Survival outcomes were analyzed using the five-lncRNA classifier, with stratification according to clinicopathological risk factors. (**A**-**B**) Kaplan-Meier analysis of DFS and OS between high- and low- risk for EBC patients in training cohort ($n = 200$). (**C**-**D**) Kaplan-Meier analysis of DFS and OS between high- and low- risk EBC patients in validation cohort ($n = 200$). HR were calculated through Cox proportional hazards regression, with P-values determined by two-sided log-rank testing in combined training and validation cohorts. The 95% CI associated with HR estimates are graphically represented using square markers with error bars. (**E**) The forest plot illustrates DFS performance of the predictive risk score with stratification based on clinicopathological features, encompassing all 400 patients from training and validation cohorts. Two-sided log-rank test was used, with HR and 95% CI depicted through square markers and error bars

Table 2 Cox regression analysis of 5-lncRNA signature and clinicopathological characteristics with DFS and OS

	Univariate analysis			Meltivariate analysis		
	HR	**95%CI**	***P* value**	**HR**	**95%CI**	***P* value**
Disease-free survival						
CBR3-AS1			< 0.001			——
HNF4A-AS1			0.01			——
LINC00622			0.035			——
LINC00993			< 0.001			——
LINC00342			0.043			——
Signature (high vs. low)	3.757	2.432–5.806	< 0.001	3.584	2.227–5.769	< 0.001
Age	0.989	0.972–1.007	0.239	0.995	0.975–1.016	0.662
ER status (positive vs. negative)	0.465	0.269–0.804	0.006	1.091	0.603–1.974	0.774
PR status (positive vs. negative)	0.547	0.307–0.974	0.04	0.881	0.479–1.623	0.685
HER2 status (positive vs. negative)	1.300	0.722–2.342	0.381	1.290	0.800-2.078	0.296
Grade	1.607	1.081–2.389	0.019	1.299	0.824–2.046	0.260
pT stage	1.743	1.213–2.505	0.003	1.227	0.809–1.816	0.336
pN stage	1.638	1.384–1.938	< 0.001	1.630	1.290–2.060	< 0.001
Stage	2.224	1.626–3.040	< 0.001			——
LVI (yes vs. no)	1.575	1.048–2.367	0.028	0.802	0.466–1.380	0.426
Overall survival						
CBR3-AS1			< 0.001			——
HNF4A-AS1			< 0.001			——
LINC00622			0.027			——
LINC00993			0.043			——
LINC00342			0.356			——
Signature (high vs. low)	2.437	1.398–4.249	0.002	1.912	1.049–3.485	0.034
Age	1.017	0.995–1.041	0.156	1.024	0.997–1.052	0.077
ER status (positive vs. negative)	0.416	0.241–0.720	0.002	0.515	0.236–1.124	0.095
PR status (positive vs. negative)	0.531	0.298–0.945	0.031	1.000	0.448–2.233	1.000
HER2 status (positive vs. negative)	1.408	0.781–2.537	0.255	1.037	0.548–1.962	0.911
Grade	2.410	1.408–4.123	0.001	1.384	0.784–2.559	0.301
pT stage	2.045	1.251–3.344	0.004	1.492	0.818–2.720	0.192
pN stage	1.804	1.439–2.263	< 0.001	1.838	1.360–2.484	< 0.001
Stage	2.591	1.681–3.993	< 0.001			——
LVI (yes vs. no)	2.004	1.160–3.464	0.013	0.868	0.421–1.791	0.702

ER, estrogen recepter; PR, progestrone recepter; Her2, epidermal growth factor receptor-2; BCS, breast conserving surgery; LVI, lymphovascular invasion, HR hazard ratio, CI confidence interval

P-values were calculated with the two-sided log-rank test

performance metrics attained an AUC value of 0.72 (95% CI: 0.63–0.79) in the training cohort via ROC curve analysis. (Fig. 4A). Similarly, the AUC for DFS in validation cohort was 0.71 (95% CI 0.62–0.78) (Fig. 4C). These results suggest that the signature is a stable predictor. The accuracy of the 5-lncRNA signature in predicting prognosis was further compared to that of each individual lncRNA. The combined signature demonstrated a higher AUC than any of the single lncRNAs. (Fig. 4A, C).

The construction of prognostic nomogram combining both 5-lncRNA signature and clinical factors further improved the sensitivity and specificity of model. The AUC for DFS of nomogram were 0.79 (95% CI 0.71–0.87) in training cohort (Fig. 4A), 0.81 (95% CI 0.75–0.87) in validation cohort (Fig. 4C). The AUC of nomogram was significantly higher than 5-lncRNA signature (Fig. 4A, C). Compared to individual clinicopathological risk factors, the nomogram exhibited significantly higher specificity and sensitivity ($P < 0.05$) (Fig. 4B, D).

Discussion

To optimize clinical decision-making and enhance prognosis for EBC patients, a novel molecular-based prognostic model incorporating both molecular and clinical data has been established. This comprehensive approach enables accurate prediction of DFS and OS post-surgery in EBC patients. The model successfully distinguishes between high- and low-risk patients. The results revealed significant disparities in DFS and OS for these patients.

Furthermore, a LASSO Cox regression model was developed to integrate multiple lncRNAs within a single prognostic predicting model, demonstrating greater

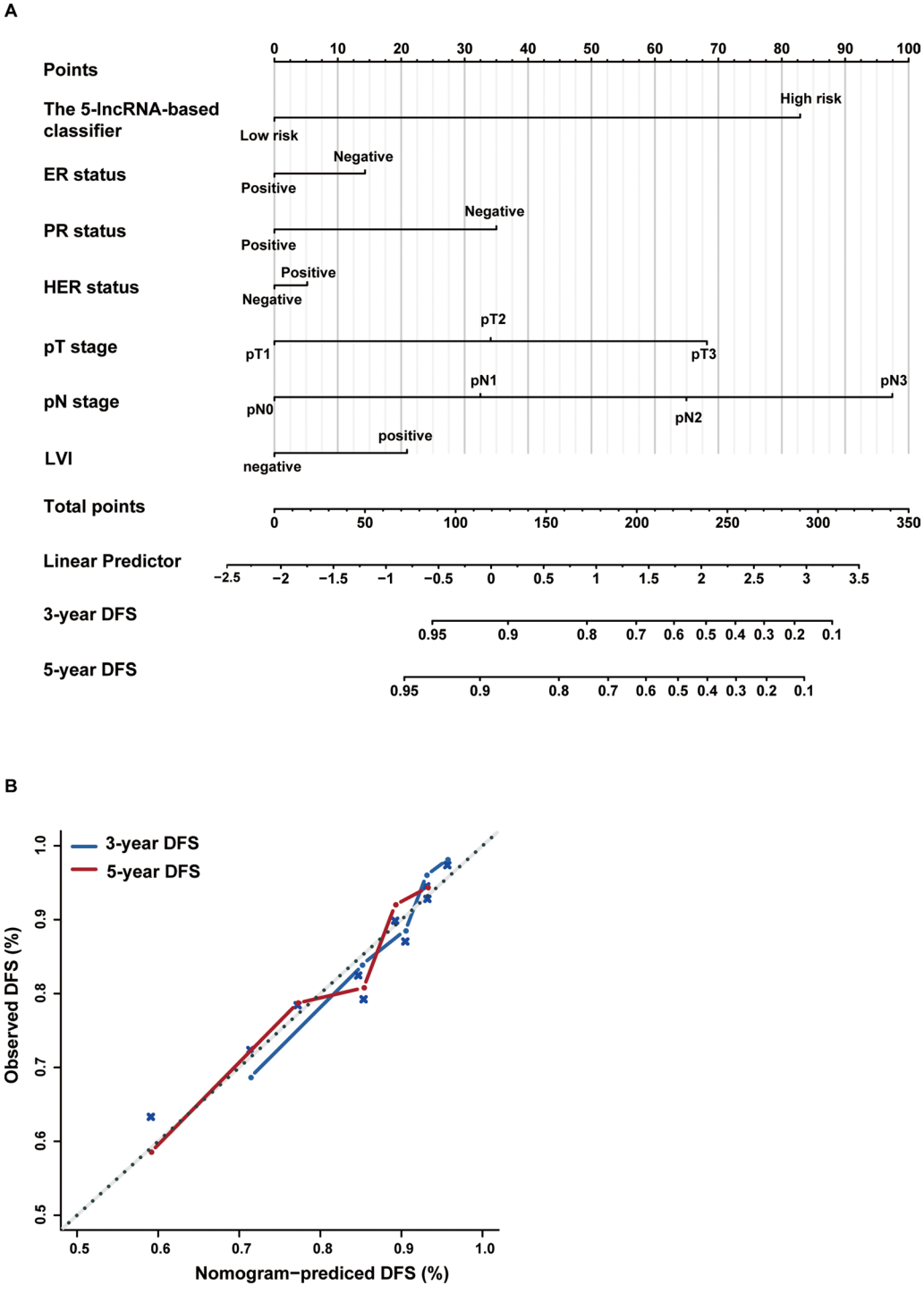

Fig. 3 Construction of nomogram including 5-lncRNAs signature and clinicopathological features. (**A**) Constructed a nomogram integrated 5-lncRNAs signature as well as clinical pathological factors to predict 3-year and 5-year DFS for EBC patients, including the 5-lncRNAs-based classifier, ER status, PR status, HER2 status, pT stage, pN stage and lymph-vascular invasion (LVI). (**B**) Plots illustrate the calibration curve of the nomogram by demonstrating agreement between the predicted and the observed 3-year and 5-year prognostic outcomes. The plot illustrates model performance, where the 45-degree line signifies ideal predictive accuracy

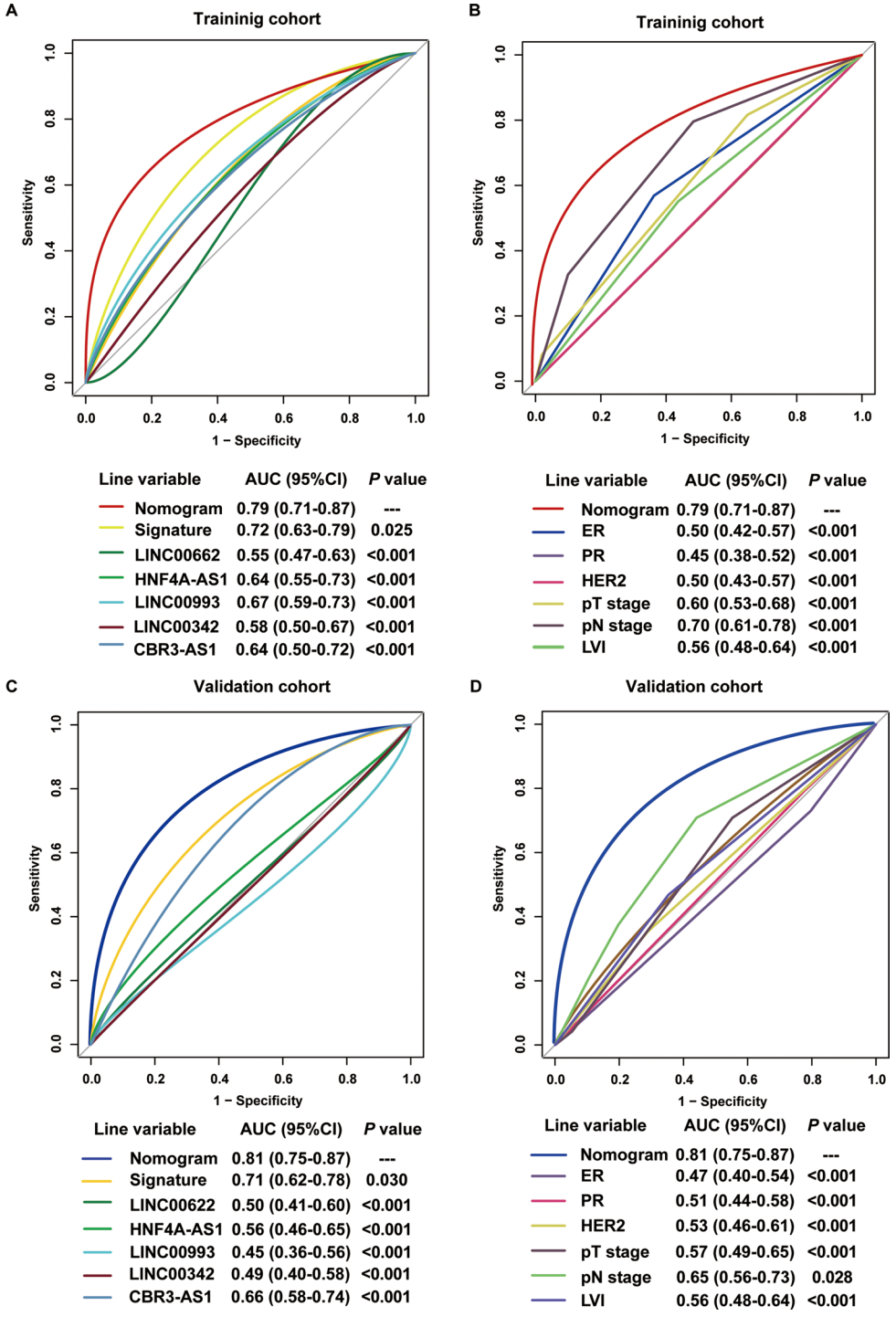

Fig. 4 Receiver operating characteristic (ROC) curves and area under curve (AUC) values. (**A**) ROC curves of prognostic accuracy in training cohort by the nomogram, 5-lncRNA signature and five single lncRNAs. (**B**) ROC curves of prognostic accuracy in training cohort by the nomogram and ER, PR, HER2 status (positive vs. negative), pT stage (T1-3), pN stage (N0-3) and LVI (yes vs. no). (**C**) ROC curves of prognostic accuracy in validation cohort by the nomogram, 5-lncRNA signature and five single lncRNAs. (**D**) ROC curves of prognostic accuracy in validation cohort by the nomogram and ER, PR, HER2 status (positive vs. negative), pT stage (T1-3), pN stage (N0-3) and LVI (yes vs. no). The AUC was computed, with its estimated 95% CI via Bootstrap resampling. P values were determined using two-sided Bootstrap tests

accuracy in forecasting early breast cancer prognosis compared to individual lncRNA analysis. Using multivariable analysis, we demonstrated the 5-lncRNAs classifier is an independent prognostic factor. When partitioned according to clinicopathological characteristics, the 5-lncRNAs signature remains a strong prognostic prodicting model.

Earlier studies have detected multiple lncRNAs that have association witn the progress of breast cancer [19, 20]. For example, MALAT1 [21], HOTAIR [22] and LINC-A [23] demonstrate established clinical relevance, correlating with survival trajectories and treatment responsiveness across breast cancer cohorts. The regulation network of lncRNAs is complicated in breast cancer with large number of upregulated and down regulated lncRNAs. Thus, use of multi-lncRNAs model may further improve the clinical value of lncRNAs.

In our study, we identified and validated five lncRNAs: CBR3-AS1, HNF4A-AS1, LINC00622, LINC00993, and LINC00342. These lncRNAs are critical for the prognosis of patients with EBC. The biological functions of these five lncRNAs have been reported in previous studies. CBR3-AS1 was identified accelerate the malignant progression of osteosarcoma, lung cancer and gestational choriocarcinoma [24–27]. Research has found that HNF4A-AS1 can promote a series of processes in neuroblastoma cells, including lactate production, glycolysis, glucose uptake, and ATP levels [28]. It might be a therapeutic target for the progression of aerobic glycolysis and neurobldownoma [29]. Additionally, HNF4A-AS1 has been shown to inhibit the hepatocellular carcinoma progression by facillitate the PCBP2 degradation through ubiquitin [30]. A previous study found that adipose-derived stem cell-derived extracellular vesicles carried LINC00622, which inhibited neuroblastoma cell proliferation, invasion, and migration [31]. Our study initially reported the prognostic value of LINC00622 and LINC00342 in breast cancer. LINC00993 was found to exhibit tumor-suppressive functionality in triple-negative breast cancer, as evidenced by its inhibition of malignant proliferation and metastasis [32]. This conclusion is similar as we demonstrated in this study, that we showed a high expression of LINC00993 is protective factor in breast cancer. LINC00342 was extensively explored in lung cancer [33], colon adenocarcinoma [34] and oral cancer [35].

Meanwhile, we constructed a nomogram including both 5-lncRNAs signature and clinicopathological risk factors. This nomogram can predict survival of EBC patients, which is better than 5-lncRNAs signature and other clinicopathological risk factors. This model enables precise stratification of patients with differential responsiveness to adjuvant therapy, suggesting current therapeutic regimens may require intensification for the identified high-risk molecular subset.

This study has same limitations. The absence of an external validation cohort. The clinical characteristics may differ across hospitals and provinces, potentially introducing selection bias. Further validations are required for further confirming the study's findings through external cohorts or prospective clinical research.

In conclusion, The five prognostic tools based on lncRNA and the clinical pathological column chart of lncRNA can accurately stratify EBC patients into low-risk and high-risk categories, with significant differences in prognosis. The findings highlight the potential of lncRNAs in predicting recurrence risk, offering valuable insights for therapeutic strategies and patient management.

The nomogram modeling has the characteristics of intuitiveness and ease of use, which can display prognostic value independent of traditional clinical pathological features in multivariate analysis. It can enhance the clinical relevance of the model and facilitate the screening of patients with different risks, which is of great significance for personalized treatment and risk management [36–37].With these tools, we can better classify patients for personalized therapeutic regimen selection, so as to avoid overtreatment and insufficient treatment.

Abbreviations

EBC	Early-stage breast cancer
lncRNAs	Long non-coding RNAs
DFS	Disease-free survival
OS	Overall survival
LASSO	Least absolute shrinkage and selection operator method
ECOG	Eastern Cooperative Oncology Cohort
ROC	Receiver operating characteristic
AUC	Area under the curve

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12935-025-03832-9.

Supplementary Material 1: **Supplementary Table 1**. The primers used in the current study.

Supplementary Material 2: **Supplementary Figure S1**. X-tile plots of the five studied lncRNAs and LASSO risk value in the training cohort based on patients' DFS and OS. The colour of the plot (left panel) reflects the degree of association strength with each division, ranging from low (black) to high (bright, red or green). Red represents the correlation between high and poor prognosis of this indicator, while green represents the correlation between high and good prognosis of this indicator. The frequency histograms (second panel) display optimized cutoff values for distinguishing high score from low score in terms of expression level. Comparison of PFS (third panel) and OS (right panel) between five lncRNAs-based signature high score and low score (A), or high expression and low expression of indicated lncRNAs (B-F)

Author contributions

R.G. designed and wrote the original manuscript, X.T. W. and Y.J.Y. analyzed the results and verified the research, J.Y. Z. organized the data and plotted it, Z.Y. L and M.L. collected and organized the data, Y. Y. and N.L. helped with the

reference, J.Y.N and Y.Y. T. and G.J. Z. approved the final manuscript.All authors have read and agreed to the published version of the manuscript.

Funding
The present study was supported by the National Natural Science Foundation of China (grant no. 82360614), First-Class Discipline Team of Kunming Medical University (Kunming Medical University Breast Cancer Precision & Translational Medicine Research Team (2024XKTDYS08), the Science and Technology Project of Yunnan Provincial Science and Technology Department (grant no. 202401AT070005) and the Yunnan Health Training Project of High Level Talents (H-2024087).

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
The manuscript is approved by all authors for publication.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Breast Surgery, Yunnan Cancer Hospital, the Third Affiliated Hospital of Kunming Medical University, Kunming 650000, P. R. China
[2]First-Class Discipline Team of Kunming Medical University, Kunming 650000, P. R. China

Received: 8 December 2024 / Accepted: 16 May 2025
Published online: 24 June 2025

References

1. Waks AG, Winer EP. Breast Cancer treatment: A review. JAMA. 2019;321(3):288–300.
2. Johansson ALV, Trewin CB, Hjerkind KV, Ellingjord-Dale M, Johannesen TB, Ursin G. Breast cancer-specific survival by clinical subtype after 7 years follow-up of young and elderly women in a nationwide cohort. Int J Cancer. 2019;144(6):1251–61.
3. Fei F, Zhang K, Siegal GP, Wei S. A simplified breast cancer prognostic score: comparison with the AJCC clinical prognostic staging system. Mod Pathol. 2021;34(12):2141–7.
4. Phung MT, Tin Tin S, Elwood JM. Prognostic models for breast cancer: a systematic review. BMC Cancer. 2019;19(1):230.
5. Goyal B, Yadav SRM, Awasthee N, Gupta S, Kunnumakkara AB, Gupta SC. Diagnostic, prognostic, and therapeutic significance of long non-coding RNA MALAT1 in cancer. Biochim Biophys Acta Rev Cancer. 2021;1875(2):188502.
6. Goodall GJ, Wickramasinghe VO. RNA in cancer. Nat Rev Cancer. 2021;21(1):22–36.
7. Guo R, Su Y, Si J, Xue J, Yang B, Zhang Q, Chi W, Chen J, Chi Y, Shao Z, et al. A nomogram for predicting axillary pathologic complete response in hormone receptor-positive breast cancer with cytologically proven axillary lymph node metastases. Cancer. 2020;126(Suppl):3819–29.
8. Yu Y, Tan Y, Xie C, Hu Q, Ouyang J, Chen Y, Gu Y, Li A, Lu N, He Z, et al. Development and validation of a preoperative magnetic resonance imaging Radiomics-Based signature to predict axillary lymph node metastasis and Disease-Free survival in patients with Early-Stage breast Cancer. JAMA Netw Open. 2020;3(12):e2028086.
9. Huang L, Zeng L, Chu J, Xu P, Lv M, Xu J, Wen J, Li W, Wang L, Wu X, et al. Chemoresistancerelated long noncoding RNA expression profiles in human breast cancer cells. Mol Med Rep. 2018;18(1):243–53.
10. Shi SH, Jiang J, Sun L, Zhang W, Zhuang ZG. Dynamic regulative biomarker: long noncoding RNA (lncRNA) in metastatic breast Cancer. Clin Lab. 2020;66(9). https://doi.org/10.7754/Clin.Lab.2020.191140.
11. Sun M, Liu X, Xia L, Chen Y, Kuang L, Gu X, Li T. A nine-lncRNA signature predicts distant relapse-free survival of HER2-negative breast cancer patients receiving taxane and anthracycline-based neoadjuvant chemotherapy. Biochem Pharmacol. 2021;189:114285.
12. Wu Q, Li Q, Zhu W, Zhang X, Li H. Identification of autophagy-related long non-coding RNA prognostic signature for breast cancer. J Cell Mol Med. 2021;25(8):4088–98.
13. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. Br J Cancer. 1994;69(6):979–85.
14. Li H, Liu J, Chen J, Wang H, Yang L, Chen F, Fan S, Wang J, Shao B, Yin D, et al. A serum MicroRNA signature predicts trastuzumab benefit in HER2-positive metastatic breast cancer patients. Nat Commun. 2018;9(1):1614.
15. Tibshirani R. The Lasso method for variable selection in the Cox model. Stat Med. 1997;16(4):385–95.
16. Utazirubanda JC, Leon T, Ngom P. Variable selection with group LASSO approach: application to Cox regression with frailty model. Commun Stat Simul Comput. 2021;50(3):881–901.
17. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000;56(2):337–44.
18. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res. 2004;10(21):7252–9.
19. Li X, Li Y, Yu X, Jin F. Identification and validation of stemness-related LncRNA prognostic signature for breast cancer. J Transl Med. 2020;18(1):331.
20. Liu L, Zhang Y, Lu J. The roles of long noncoding RNAs in breast cancer metastasis. Cell Death Dis. 2020;11(9):749.
21. Kim J, Piao HL, Kim BJ, Yao F, Han Z, Wang Y, Xiao Z, Siverly AN, Lawhon SE, Ton BN, et al. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. Nat Genet. 2018;50(12):1705–15.
22. Mozdarani H, Ezzatizadeh V, Rahbar Parvaneh R. The emerging role of the long non-coding RNA HOTAIR in breast cancer development and treatment. J Transl Med. 2020;18(1):152.
23. Lin A, Li C, Xing Z, Hu Q, Liang K, Han L, Wang C, Hawke DH, Wang S, Zhang Y, et al. The LINK-A LncRNA activates normoxic HIF1alpha signalling in triple-negative breast cancer. Nat Cell Biol. 2016;18(2):213–24.
24. Yao W, Hou J, Liu G, Wu F, Yan Q, Guo L, Wang C. LncRNA CBR3-AS1 promotes osteosarcoma progression through the network of miR-140-5p/DDX54-NUCKS1-mTOR signaling pathway. Mol Ther Oncolytics. 2022;25:189–200.
25. Zhang Y, Meng W, Cui H. LncRNA CBR3-AS1 predicts unfavorable prognosis and promotes tumorigenesis in osteosarcoma. Biomed Pharmacother. 2018;102:169–74.
26. Hou M, Wu N, Yao L. LncRNA CBR3-AS1 potentiates Wnt/beta-catenin signaling to regulate lung adenocarcinoma cells proliferation, migration and invasion. Cancer Cell Int. 2021;21(1):36.
27. Zhang Y, Zhang H, Zhang X, Liu B. CBR3-AS1 Accelerates the Malignant Proliferation of Gestational Choriocarcinoma Cells by Stabilizing SETD4. Dis Markers, 2022:7155525.
28. Jia W, Yu L, Xu B, et al. HNF4A-AS1 inhibits the progression of hepatocellular carcinoma by promoting the ubiquitin-modulated degradation of PCBP2 and suppressing the stability of ARG2 mRNA. Int J Biol Sci. 2024;20(13):5087–108. Published 2024 Sep 23.
29. Song H, Li D, Wang X, Fang E, Yang F, Hu A, Wang J, Guo Y, Liu Y, Li H, et al. HNF4A-AS1/hnRNPU/CTCF axis as a therapeutic target for aerobic Glycolysis and neuroblastoma progression. J Hematol Oncol. 2020;13(1):24.
30. 30 Song H, Wang J, Wang X, et al. HNF4A-AS1-encoded small peptide promotes self-renewal and aggressiveness of neuroblastoma stem cells via eEF1A1-repressed SMAD4 transactivation. Oncogene. 2022;41(17):2505–19.
31. Guo M, Li D, Feng Y, Li M, Yang B. Adipose-derived stem cell-derived extracellular vesicles inhibit neuroblastoma growth by regulating GABBR1 activity through LINC00622-mediated transcription factor AR. J Leukoc Biol. 2022;111(1):19–32.
32. Guo S, Jian L, Tao K, Chen C, Yu H, Liu S. Novel breast-Specific long Non-coding RNA LINC00993 acts as a tumor suppressor in Triple-Negative breast Cancer. Front Oncol. 2019;9:1325.
33. Su H, Yu S, Sun F, Lin D, Liu P, Zhao L. LINC00342 induces metastasis of lung adenocarcinoma by targeting miR-15b/TPBG. Acta Biochim Pol. 2022;69(2):291–7.
34. Miao Z, Liu S, Xiao X, Li D. LINC00342 regulates cell proliferation, apoptosis, migration and invasion in colon adenocarcinoma via miR-545-5p/MDM2 axis. Gene. 2020;743:144604.

35. Chen Y, Wang Y, Zhang W. LINC00342 regulates the PI3K-AKT signaling pathway via the miR-149-5p/FGF11 axis and affects the progression of oral cancer. Discov Oncol. 2024;15(1):606.
36. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than Meets the eye. Lancet Oncol. 2015;16(4):e173–80.
37. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008;26(8):1364–70.

Publisher's note