



RESEARCH ARTICLE

Integrated analysis identifies a novel lncRNA prognostic signature associated with aerobic glycolysis and hub pathways in breast cancer

Zheng Li¹  | Juan Zheng² | Yang Feng¹ | Yaming Li¹ | Yiran Liang¹ | Ying Liu¹ | Xiaolong Wang¹ | Qifeng Yang^{1,3,4} 

¹Department of Breast Surgery, General Surgery, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, China

²Department of Ultrasound, Qilu Children's Hospital of Shandong University, Jinan, Shandong, China

³Pathology Tissue Bank, Qilu Hospital of Shandong University, Jinan, Shandong, China

⁴Research Institute of Breast Cancer, Shandong University, Jinan, Shandong, China

Correspondence

Qifeng Yang, Department of Breast Surgery, General Surgery, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, China.
Email: qifengy_sdu@163.com

Funding information

National Key Research and Development Program, Grant/Award Number: 2020YFA0712400 and 2018YFC0114705; Special Foundation for Taishan Scholars, Grant/Award Number: ts20190971; National Natural Science Foundation of China, Grant/Award Number: 81874119 and 82072912; National High Level Talents, Grant/Award Number: W01020103; Clinical Research Center of Shandong University, Grant/Award Number: 2020SDUCRCA015; Qilu Hospital Clinical New Technology Developing Foundation, Grant/Award Number: 2018-7 and 2019-3

Abstract

Long noncoding RNAs (lncRNAs) play a crucial role in cancer aerobic glycolysis. However, glycolysis-related lncRNAs are still underexplored in breast cancer. In this study, we identified the five most glycolysis-related lncRNAs in breast cancer to construct a prognostic signature, which could distinguish between patients with unfavorable and favorable prognoses. To investigate the role of signature lncRNAs in breast cancer, we profiled their expression levels in breast cancer progression cell line model. Real-time PCR revealed that the five lncRNAs could contribute to breast cancer initiation or progression. Furthermore, we observed that the levels of four lncRNAs expression had a significant trend of gradient upregulation with the addition of glycolysis inhibitor in breast cancer cells. Afterward, random forest and logistic regression were conducted to assess the model's performance in stratifying glycolysis status. Finally, a nomogram including the lncRNA signature and clinical features was developed, and its efficacy in predicting the survival time and clinical utility was evaluated using a calibration curve, concordance index, and decision curve analysis. In this study, gene set enrichment analysis showed that the mTOR pathway, a central pathway in tumor initiation and progression, was significantly enriched in the high-risk group. In addition, gene set variation analysis was performed to validate our findings in two independent datasets. Subsequent weighted gene co-expression network analysis, followed by enrichment analysis, indicated that downstream cell growth-related signaling was strikingly activated in the high-risk group, and may directly promote tumor progression and escalate mortality risk in patients with high-risk scores. Overall, our findings may provide novel insight into lncRNA-related metabolic regulation, and help to develop promising prognostic indicators and therapeutic targets for breast cancer patients.

KEYWORDS

bioinformatics, biomarkers, breast cancer, lncRNA, Prognosis

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

© 2021 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Breast cancer is the most common cancer and the leading cause of cancer-related death in women worldwide.¹ Although conventional treatment strategies have been well applied, many patients with breast cancer still have unfavorable prognosis.² Consequently, it is essential to further investigate novel prognostic indicators, diagnostic biomarkers, and therapeutic targets for improved clinical outcomes.

Altered energy metabolism is one of the pivotal fingerprints associated with cancer biological behaviors.³ Aerobic glycolysis, known as the 'Warburg effect', is a preferential metabolic phenotype for cancer cells.⁴ Although aerobic glycolysis has poor ATP production compared to mitochondrial oxidative phosphorylation, cancer cells accelerate the ATP production rate and increase glucose uptake via metabolic reprogramming.⁵ Meanwhile, glycolysis intermediates not only contribute to macromolecule formation in various biosynthetic pathways,⁶ but also induce resistance to chemotherapy and radiotherapy.^{7,8} In addition, glycolysis provides a favorable tumor microenvironment for cancer cells to thrive.⁹ Due to the crucial role of tumor aerobic glycolysis in breast cancer initiation and progression, further exploration could help to improve clinical outcomes for patients with breast cancer.

To date, long noncoding RNAs (lncRNAs), which are RNAs longer than 200 nucleotides, have been shown to play an important role in transcription, post-transcription, and epigenetic modification, and influence genes associated with glucose metabolism in several cancer types.¹⁰⁻¹⁴ In addition, lncRNAs could contribute to metabolism reprogramming, which could regulate carcinogenesis and progression by providing adequate nutrition for cancer cells to circumvent energy stress.¹⁵ Malakar et al. reported that lncRNA *MALAT1* may induce glucose metabolism reprogramming to promote tumor malignant progression by upregulating SRSF1 and activating the mTORC1-4EBP1 axis in hepatocellular carcinoma.¹⁶ Li et al. demonstrated that lncRNA *UCA1* plays a positive role in glycolysis by upregulating hexokinase two through the mTOR-STAT3/microRNA143 pathway in bladder cancer.¹⁷ Liu et al. revealed that downregulation of lncRNA *NBR2* could attenuate AMPK activation and promote mTORC1-mediated protein synthesis and cancer cell growth under glucose-starved stress.¹⁸ Additionally, Hung et al. suggested that lncRNA *PCGEM1* may function as a crucial transcription regulator in central metabolic pathways, and promote cancer cell proliferation by regulating tumor metabolism via co-activation of both c-Myc and androgen receptor (AR).¹⁹ Hence, glycolysis-related lncRNAs could provide novel

insights for further exploration of metabolic strategies in breast cancer prognosis and treatment.

In this study, we applied integrated bioinformatics analysis to identify a prognostic signature of five glycolysis-related lncRNAs, which could predict the survival time and glycolysis status in breast cancer patients. Moreover, we further investigated the potential biological roles underlying the lncRNA signature via systematic bioinformatics analysis and in vitro experiments. Thus, our findings provide a novel insight into lncRNA-related metabolic regulation and help to develop promising prognostic indicators and therapeutic targets for breast cancer patients.

2 | MATERIALS AND METHODS

2.1 | Sample datasets and data processing

The Cancer Genome Atlas (TCGA) data portal (<https://portal.gdc.cancer.gov/>) was used to obtain TCGA RNA-Seq dataset. The raw count data were transformed through the variance-stabilizing transformation method using the DESeq2²⁰ package and then were quantile-normalized using the preprocessCore package. The 888 cases of breast cancer obtained from the TCGA were screened based on the following inclusion criteria: availability of complete data on overall survival time, survival status, age, subtype, and AJCC stage. The molecular subtypes were classified by the PAM50 subtype predictor, including luminal A, luminal B, HER2-enriched, basal-like, and normal-like. The METABRIC dataset with normalized data sourced from Molecular Taxonomy of Breast Cancer International Consortium (<https://www.mbcproject.org/>) contained 1903 breast cancer cases with overall survival time and survival status. The GSE20685 dataset was downloaded from the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/gds/>), which contained 327 breast cancer cases with overall survival time, survival status, age, and clinical stage. As previously described,²¹ SeqMap was used to reannotate the probe sets of the Affymetrix Hg-U133 Plus 2.0 array. The microarray data were background corrected and normalized via the Limma package.²² The log₂-transformed normalized data were used for downstream analysis. In this study, TCGA and METABRIC datasets were used to assess the efficacy of the glycolysis score. The TCGA dataset served as the training set to select the five most glycolysis-related lncRNAs to construct a prognostic signature. The GSE20685 dataset served as the validation set to validate our findings from the training set. The clinical information for the included patients is summarized in Table S1.

2.2 | Cell lines and cell culture

All cell lines used in this study were purchased from American Type Culture Collection (ATCC), and maintained in a humidified incubator at 37°C with 5% CO₂. MDA-MB-231 and MDA-MB-468 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen), supplemented with 10% fetal bovine serum (Hyclone), 100 U/ml of penicillin, and 100 µg/ml of streptomycin. MCF-10A, MCF-10AT, and MCF-10CA1A cells were cultured in DMEM/F12 (Invitrogen, Carlsbad, California) with 5% horse serum (Invitrogen), 500 ng/ml of hydrocortisone (Sigma-Aldrich), 100 ng/ml of cholera toxin (Sigma-Aldrich), 10 µg/ml of insulin (Invitrogen), and 20 ng/ml of epidermal growth factor (EGF, Sigma-Aldrich).

2.3 | RNA extraction, reverse transcription, and real-time PCR analysis

Total RNA was extracted from breast cancer cells using the TRIzol reagent (Invitrogen). After RNA was reversely transcribed into complementary DNA (cDNA) by PrimeScript reverse transcriptase (RT) reagent kit (TaKaRa), Biosystems StepOne plus System was used to perform real-time PCR assay. Primers used for real-time PCR are listed in Table S2.

2.4 | Lactate assay

The lactate level of cell supernatant was measured by the Lactate Assay Kit (Eton Bioscience). Briefly, breast cancer cells were seeded at a density of 20,000 cells per well in a 96-well plate. The next day, the cell supernatant from each well was collected, mixed with L-Lactate assay solution, and then incubated at 37°C for 30 min. Lastly, the absorbance at 490 nm was read to measure the concentration of L-Lactate.

2.5 | 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) assay

Cell growth was measured using MTT assay. Briefly, breast cancer cells were seeded in 96-well plates at a density of 2×10^3 cells per well, treated with vehicle or 10 nM rapamycin for the indicated time, and then incubated with 20 µl MTT (5 mg/ml) for additional 4–6 h at 37°C with 5% CO₂. MTT reagent was aspirated and 100 µl of dimethyl sulfoxide (DMSO) was added into each well. Absorbance values were measured at 490 nm on a microplate reader (Bio-Rad).

2.6 | Development and evaluation of glycolysis score

First, glycolysis-related genes were obtained from the Molecular Signatures Database (MSigDB) containing REACTOME_GLYCOLYSIS, HALLMARK_GLYCOLYSIS, and KEGG_GLYCOLYSIS_GLUONEOGENESIS.²³ Next, we calculated the glycolysis scores for each patient via single sample Gene Set Enrichment Analysis (ssGSEA). Based on the median glycolysis score, breast cancer patients were classified into two subgroups, high- and low-glycolysis groups. Lastly, GSEA and Kaplan–Meier survival analysis were used to evaluate the efficacy of glycolysis scores in two independent datasets.

2.7 | Construction of the lncRNA signature

Based on the criteria ($|r| > 0.35$, p value < 0.001), a cohort of lncRNAs significantly associated with the glycolysis score was selected in the training set by Spearman's correlation analysis. Subsequently, univariate followed by stepwise multivariate Cox regressions were performed to identify the five most promising lncRNAs and develop a prognostic signature.

2.8 | Prognostic lncRNA signature-based risk score

Multivariate Cox model analysis using bidirectional stepwise selection was performed to calculate the regression coefficients for determining each lncRNA expression level. The risk score formula is: Risk score = $(-0.249 \times \text{expression of AC007686.3}) + (-0.253 \times \text{expression of BAIAP2-DT}) + (-0.210 \times \text{expression of LINC00926}) + (0.056 \times \text{expression of LINC01016}) + (-0.107 \times \text{expression of MAPT-AS1})$.

2.9 | Construction and evaluation of a nomogram

The RMS package was used to generate a nomogram, in which predictive accuracy and discrimination ability were evaluated by calibration curve and concordance index (C-index), respectively. Moreover, a decision curve analysis (DCA) was performed to evaluate the clinical utility of the nomogram by quantifying the net benefits against a range of threshold probabilities.²⁴

2.10 | Gene set enrichment analysis and gene set variation analysis

Gene set enrichment analysis was performed by the JAVA program using the Hallmark gene sets sourced from MSigDB. All genes were ranked based on differential significance between the high- and low-risk subgroups stratified by the median risk score. After performing 1000 permutations, gene set enrichment with nominal $p < 0.05$ and FDR < 0.25 was considered as a significant difference. In GSEA,²⁵ Spearman's correlation analysis was carried out to assess the relationship between the risk score and specific hallmark gene sets in the training and validation sets.

2.11 | Weighted gene correlation network analysis

Weighted Gene Correlation Network Analysis procedure was carried out as described previously.²⁶ Briefly, a soft thresholding power of six was selected to generate a scale-free topology from adjacency matrix. DeepSplit of 2 and minModuleSize of 30 were set as the parameters of the Dynamic Tree Cut method to avoid generating too many modules. The height cut-off value was set to 0.25 to merge modules with similarity > 0.75 . Finally, the enrolled genes generated 17 modules (except the gray module) by cluster analysis. We evaluated the association between the risk score and module eigengenes (MEs) to achieve the module most closely related to the risk score. The hub genes were selected according to module membership (MM) greater than 0.8 and gene significance (GS) greater than 0.4. Biological process enrichment analysis of hub genes from highly related modules was performed using Metascape (<http://metascape.org/>).

2.12 | Statistical analysis

Multivariate survival analysis for the lncRNA signature and clinicopathological features was performed using Cox proportional hazards regression models to determine which factors could act as an independent prognostic indicator. A time-dependent receiver operating characteristic (ROC) analysis was conducted to investigate the model's predictive performance at 1, 3, 5, and 10 years. The Kaplan–Meier method combined with the log-rank test was carried out to assess the overall survival time between the two subgroups. Two well-established machine learning algorithms (random forest [RF] and logistic regression [LR]) were used to confirm the efficacy of the lncRNA

signature for stratifying glycolysis status on the basis of the area under the ROC curve (AUC) score through five-fold cross validations.²⁷ Logistic regression analysis was performed to evaluate whether the lncRNA signature had a better performance for stratifying the glycolysis status than individual lncRNA. Spearman's correlation analysis was used to assess the relationship among risk scores, glycolysis scores, and each lncRNA. The Chi-squared test was used to examine the association between the lncRNA signature and clinicopathological phenotype based on the median risk score as a cutoff threshold. For continuous data, difference between two groups was assessed using the Student's *t*-test or Wilcoxon test, and multiple groups comparison was made using the Kruskal–Wallis test. Experimental data are presented as mean (\pm SD). In this study, we used the R project (version 3.6.1) and GraphPad Prism 8 to perform the main statistical analysis. Differences with $p < 0.05$ were considered statistically significant.

3 | RESULTS

3.1 | Development and evaluation of the glycolysis score

A flow diagram illustrating our analysis procedure is shown in Figure 1. Here, TCGA and METABRIC datasets were employed to assess the efficacy of the glycolysis score (ssGSEA score). GSEA identified that three glycolysis-related gene sets were significantly enriched in the high-glycolysis group, which indicated that the glycolysis score could directly represent glycolysis status (Figure 2A,B). In addition, survival analysis revealed that patients with high-glycolysis scores had shorter survival times than those with low-glycolysis scores (Figure 2C,D).

3.2 | Construction of a five glycolysis-related lncRNA signature in the training set

Based on the criteria ($|r| > 0.35$, p value < 0.001), the 121 most glycolysis-related lncRNAs were obtained from the training set using Spearman's correlation analysis (Figure 3A; Table S3). Univariate followed by stepwise multivariate Cox regression analyses were performed, and then a five glycolysis-related lncRNA signature was constructed (Table S4). According to univariate Cox regression analysis, AC007686.3, BAIAP2-DT, LINC00926, LINC01016, and MAPT-AS1 were defined as protective

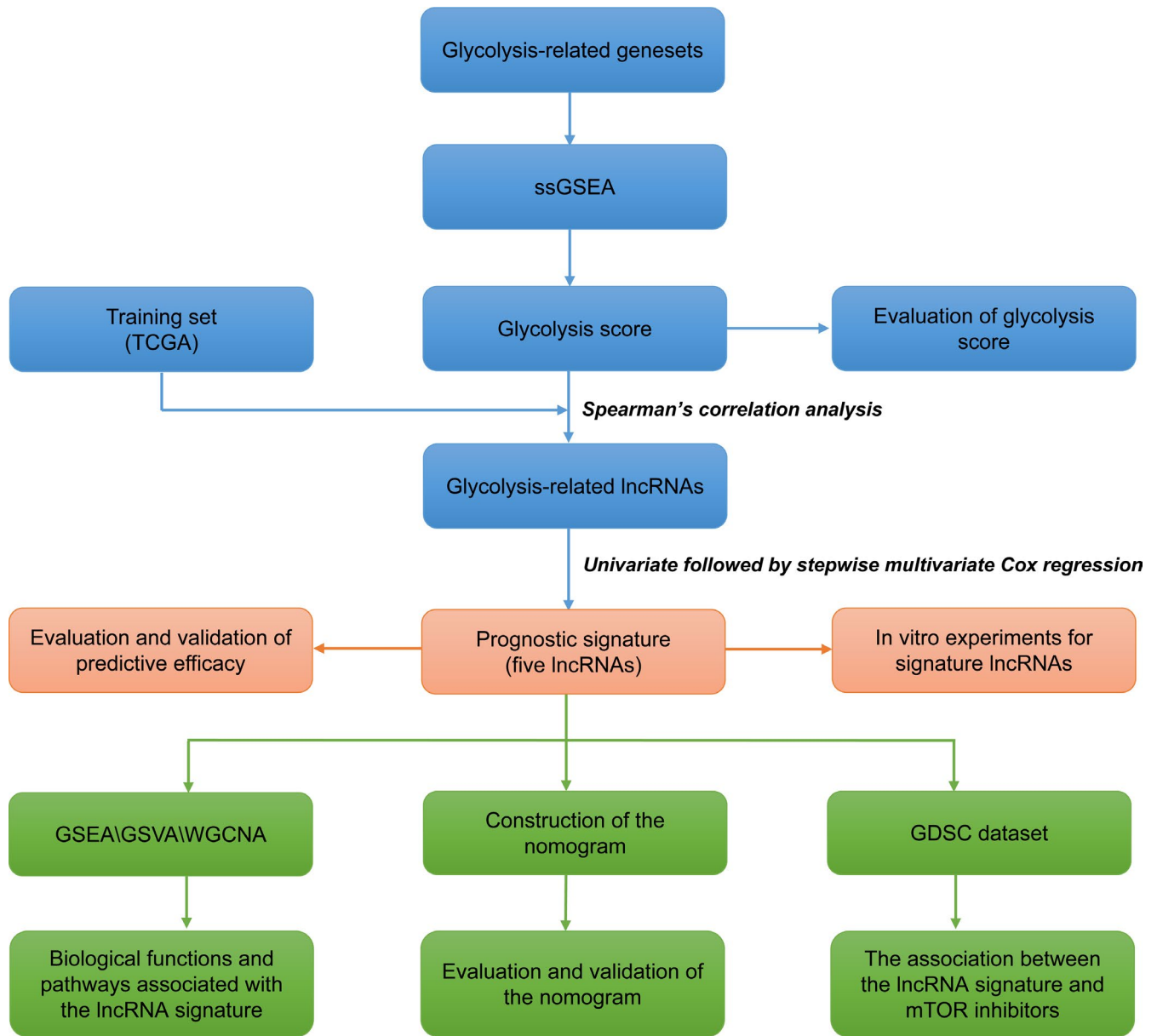


FIGURE 1 The flow diagram of our analysis procedure

factors (HR value <1) in the prognostic model (Table 1). As shown in Figure 3B, univariate Cox regression analysis was used to examine the effect of clinicopathologic features and lncRNA signature on overall survival in the TCGA cohort. Subsequent multivariate Cox regression analysis indicated that age, cancer status, and lncRNA signature had a significant association with duration of patient survival independent of other variables (Figure 3C). In addition, we found that the malignant grade of the AJCC stage was evidently associated with a high-risk value (Figure 3D). Meanwhile, the molecular subtype of breast cancer was strikingly related to the risk score, which showed that basal-like or HER2 patients had higher risk values than patients with other subtypes (Figure 3E).

3.3 | Investigation of the association between signature lncRNAs and glycolysis

First, we evaluated the role of signature lncRNAs in breast cancer progression via breast cancer progression cell line model (MCF10A\MCF10AT\MCF10CA1A). Real-time PCR suggested that the levels of five lncRNAs expression were reduced in premalignant MCF10AT and malignant MCF10CA1A cells compared to parental MCF10A cells (Figure 4A). Notably, LINC00926, AC007686.3, and BAIAP2-DT appeared to be a significant trend of gradient downregulation from MCF10A to MCF10AT and MCF10CA1A cells, indicating that these three lncRNAs could play an important role in breast cancer initiation and progression. To further investigate the association

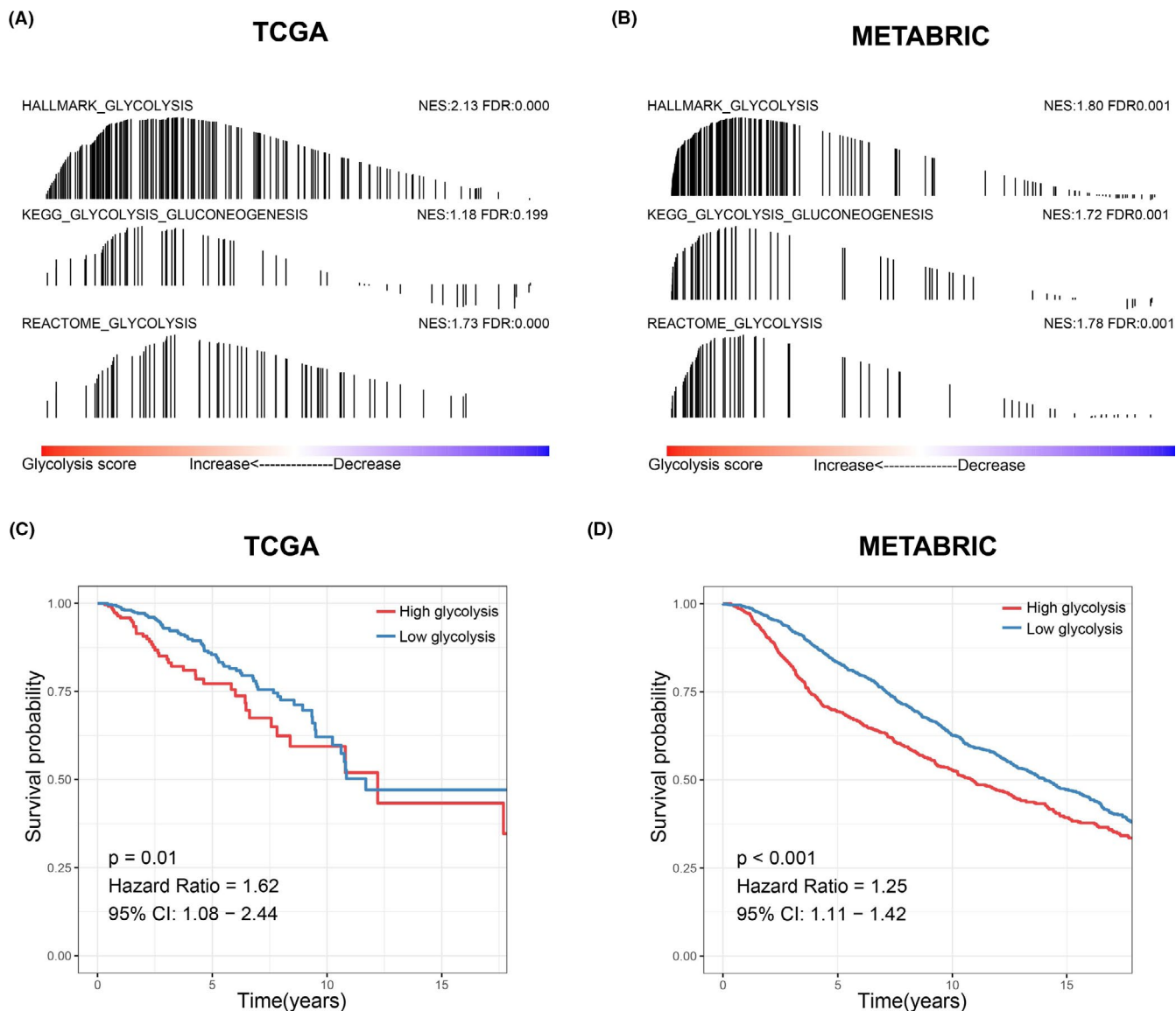


FIGURE 2 Development and evaluation of the glycolysis score. (A, B) Gene set enrichment analysis on the basis of glycolysis score. (C, D) Kaplan–Meier survival analysis on the basis of glycolysis score

of signature lncRNAs with glycolysis, we treated breast cancer cells with 2-Deoxy-D-glucose (2DG), a glycolysis inhibitor. Lactate production was used to examine for levels of the glycolytic in breast cancer cells treated with 5 or 10 mM 2DG for 12–24 h. As presented in Figure 4B, gradient descent lactate production was observed with an increase in 2DG concentration. On the contrary, the use of 2DG enhanced expression levels of LINC00926, LINC01016, AC007686.3, and MAPT-AS1 in MDA-MB-231 and MDA-MB-468 cells (Figure 4C,D). However, we did not observe significant changes in BAIAP2-DT expression levels (data not shown). Afterward, we conducted Spearman's correlation analysis to further identify the relationship between aerobic glycolysis-related factors and lncRNA signature in the training set. Our results showed that the risk score was positively correlated with

hub glycolysis-related genes, and signature lncRNAs had a negative correlation with them (Figure 4E). We next examined these correlations in the validation set (Figure S1A). A similar result was yielded to further support our findings from the training set.

3.4 | Validation and further evaluation of the lncRNA signature in the training and validation sets

Here, ROC analysis was applied to assess the predictive accuracy of signature at 1, 3, 5, and 10 years. The area under the curve (AUC) scores in the training and validation sets are shown in Figure 5A,B, respectively. Using the median risk score as the cutoff threshold, the distribution

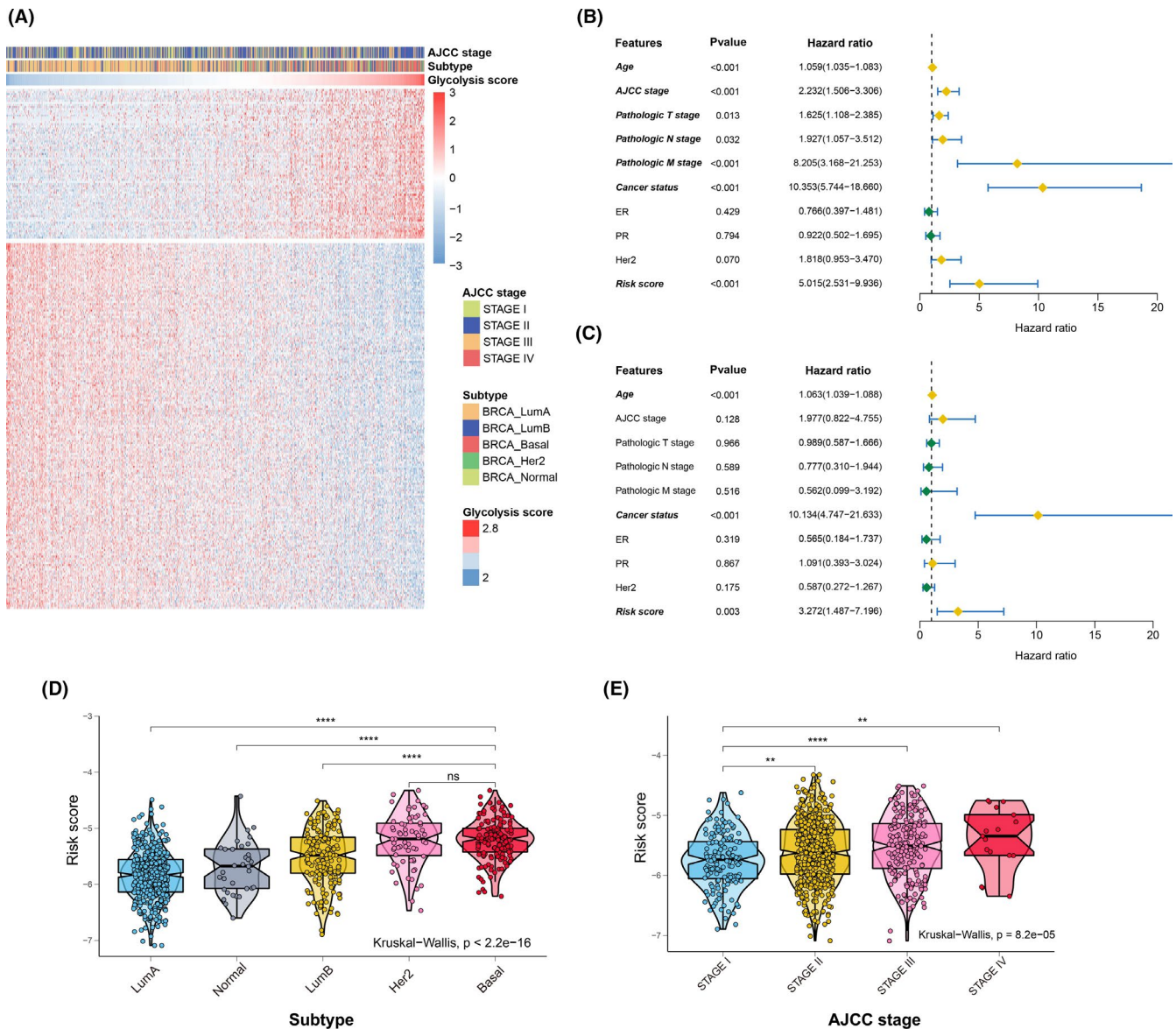


FIGURE 3 Construction of a five glycolysis-related lncRNA signature in the training set. (A) A cohort of glycolysis-related lncRNAs was identified by Spearman's correlation analysis. (B) Univariate Cox regression analysis. (C) Multivariate Cox regression analysis. Bold italics indicate statistically significant variables. (D) Comparison of risk scores associated with different clinical stages. (E) Comparison of risk scores associated with different molecular subtypes (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$; Wilcoxon test)

TABLE 1 Top five glycolysis-related lncRNAs identified from Spearman's correlation analysis and Cox regression analysis

lncRNA_symbol	Ensemble_ID	Spearman's correlation analysis		Univariate cox regression analysis			
		R	p value	HR	HR.95L	HR.95H	p value
AC007686.3	ENSG00000273729	-0.384	1.43E-32	0.660	0.499	0.872	0.0034
BAIAP2-DT	ENSG00000226137	-0.375	5.35E-31	0.709	0.562	0.893	0.0035
LINC00926	ENSG00000247982	-0.403	4.69E-36	0.740	0.625	0.876	0.0005
LINC01016	ENSG00000249346	-0.420	3.64E-39	0.949	0.900	1.000	0.0492
MAPT-AS1	ENSG00000264589	-0.364	3.69E-29	0.889	0.834	0.947	0.0003

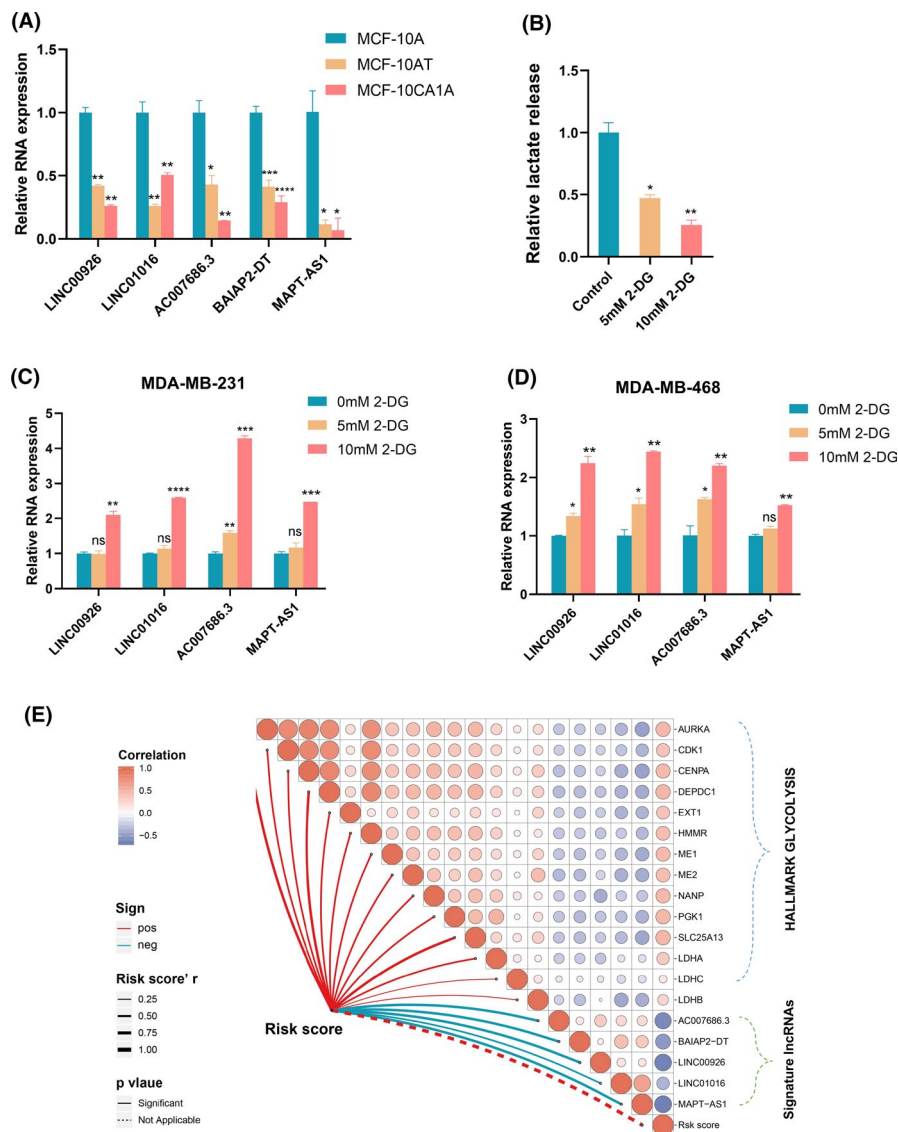


FIGURE 4 Investigation of the association between signature lncRNAs and glycolysis. (A) Expression profiles of signature lncRNAs in breast cancer progression cell line model. (B) The effect of 2DG on lactate generation was estimated in MDA-MB-468 cells. (C, D) Breast cancer cells treated with 2DG-containing medium were subjected to real-time PCR analysis to measure signature lncRNAs expression. (E) The relationship among aerobic glycolysis-related factors, lncRNA signature, and each lncRNA (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$; Student's t -test)

of survival status, overall survival time, and lncRNA expression in the training and validation sets is presented separately in Figure 5C,D. Survival analysis showed that patients with high-risk scores had poor survival time compared to those with low-risk scores (Figure 5E,F).

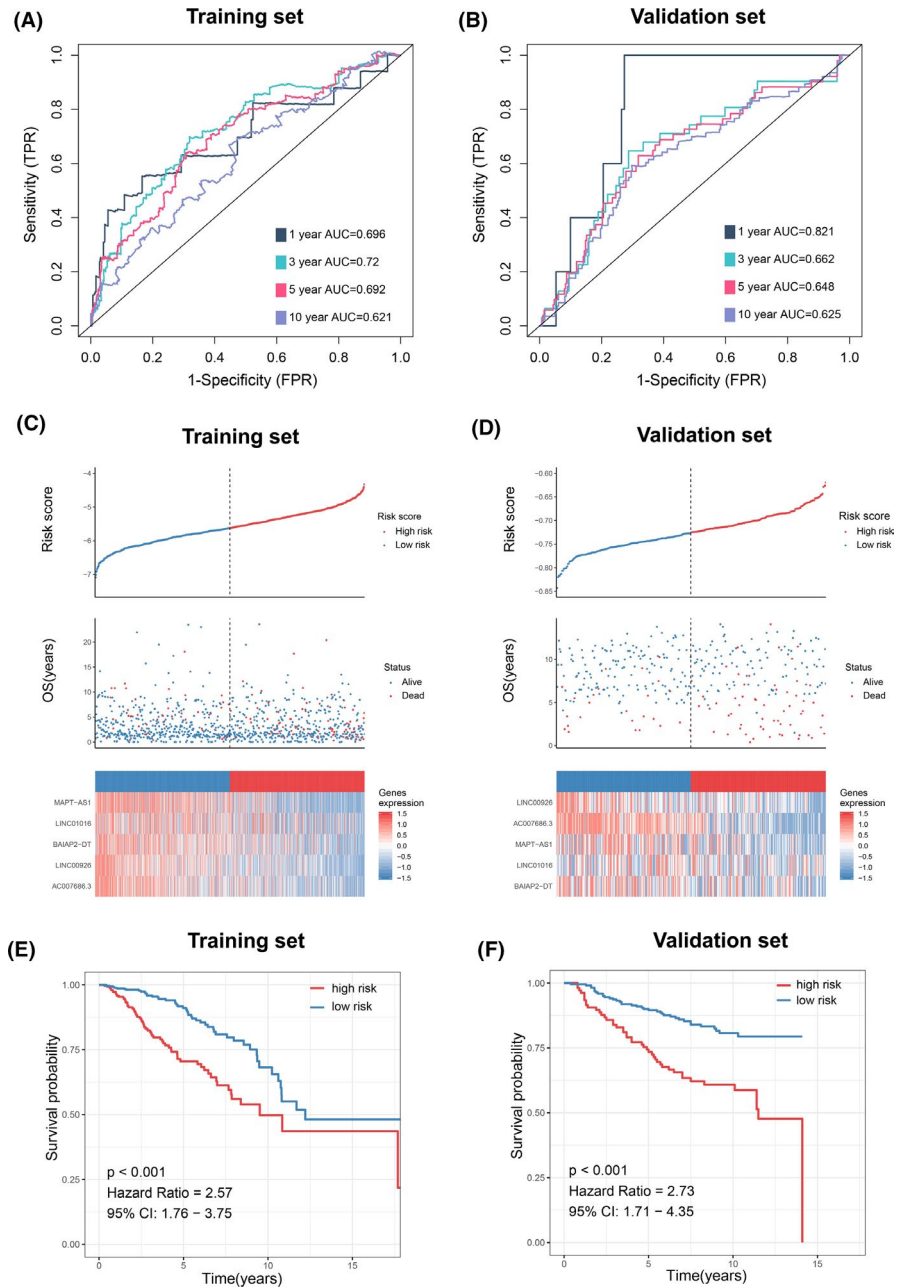
Subsequently, the signature's predictive capability for glycolysis status was further assessed by RF and LR analyses. Importantly, moderate predictive performance for glycolysis status was observed in the training and validation sets (Figure 6A,B). Additionally, LR analysis identified that the lncRNA signature predicted glycolysis status more efficiently than individual lncRNA (Figure 6C,D). The interplay among risk scores, glycolysis scores, and five lncRNAs was further confirmed by Spearman's correlation analysis. It revealed that the risk score was positively associated with the glycolysis score, whereas five lncRNAs were negatively associated with the risk and glycolysis scores (Figure 6E,F). Additionally, the interactions among the five lncRNAs are shown in Figure 6E,F.

Of note, the validation set had an acceptable performance with the training set.

3.5 | Construction and evaluation of the nomogram

Before the signature was sent to construct a nomogram, we employed the Chi-squared test to explore the association between the lncRNA signature and clinicopathological features by stratifying TCGA-derived patients into high- and low-risk groups based on the median risk score as a cutoff threshold. As shown in Table 2, the risk score was significantly associated with AJCC stage, cancer status, and subtype. To enhance the predictive efficacy of the lncRNA signature, we developed a nomogram that incorporated age, AJCC stage, subtype, cancer status, and the lncRNA signature (Figure 7A). DCA was performed to estimate net benefit and clinical utility for this nomogram

FIGURE 5 Validation of the lncRNA signature in the training and validation sets. (A, B) Time-dependent ROC curves at 1, 3, 5, and 10 years. (C, D) From top to bottom are the risk scores, patients' survival status distribution, and the expression heatmap of five glycolysis-related lncRNAs in the low- and high-risk groups. (E, F) Kaplan–Meier survival analysis on the basis of risk score



at 1, 3, and 5 years (Figure 7B,D). It revealed that the nomogram displayed consistent positive and larger net benefit across a broad range of threshold probabilities (more than 70%) at 3 and 5 years compared to either the none-treat scheme or all-treat scheme. However, the 1-year DCA showed that patients could only acquire net benefits within nearly 30% of the threshold probability. The calibration curve indicated that the nomogram survival prediction for breast cancer patients had an excellent agreement with actual observations at 1, 3, and 5 years, with a C-index of 0.855 (95% CI, 0.812–0.898) (Figure 7E). Importantly, a good performance for predicting survival was also observed in the validation set, with a moderate discrimination (C-index of 0.725 [95% CI, 0.671–0.779]) (Figure 7F).

3.6 | The cancer-related hallmark gene sets associated with the lncRNA signature in breast cancer

As presented in Figure S1B, ssGSEA followed by Spearman's correlation analysis was performed, and it was found that the hallmark gene sets significantly associated with the lncRNA signature included mTORC1 signaling ($r = 0.607$, $p < 0.001$), G2M checkpoints ($r = 0.516$, $p < 0.001$), E2F targets ($r = 0.476$, $p < 0.001$), unfold protein response ($r = 0.430$, $p < 0.001$), mitotic spindle ($r = 0.420$, $p < 0.001$), glycolysis ($r = 0.409$, $p < 0.001$), and MYC targets V1 ($r = 0.408$, $p < 0.001$).

Based on the median risk score, TCGA-derived patients were stratified into two subgroups, high- and low-risk

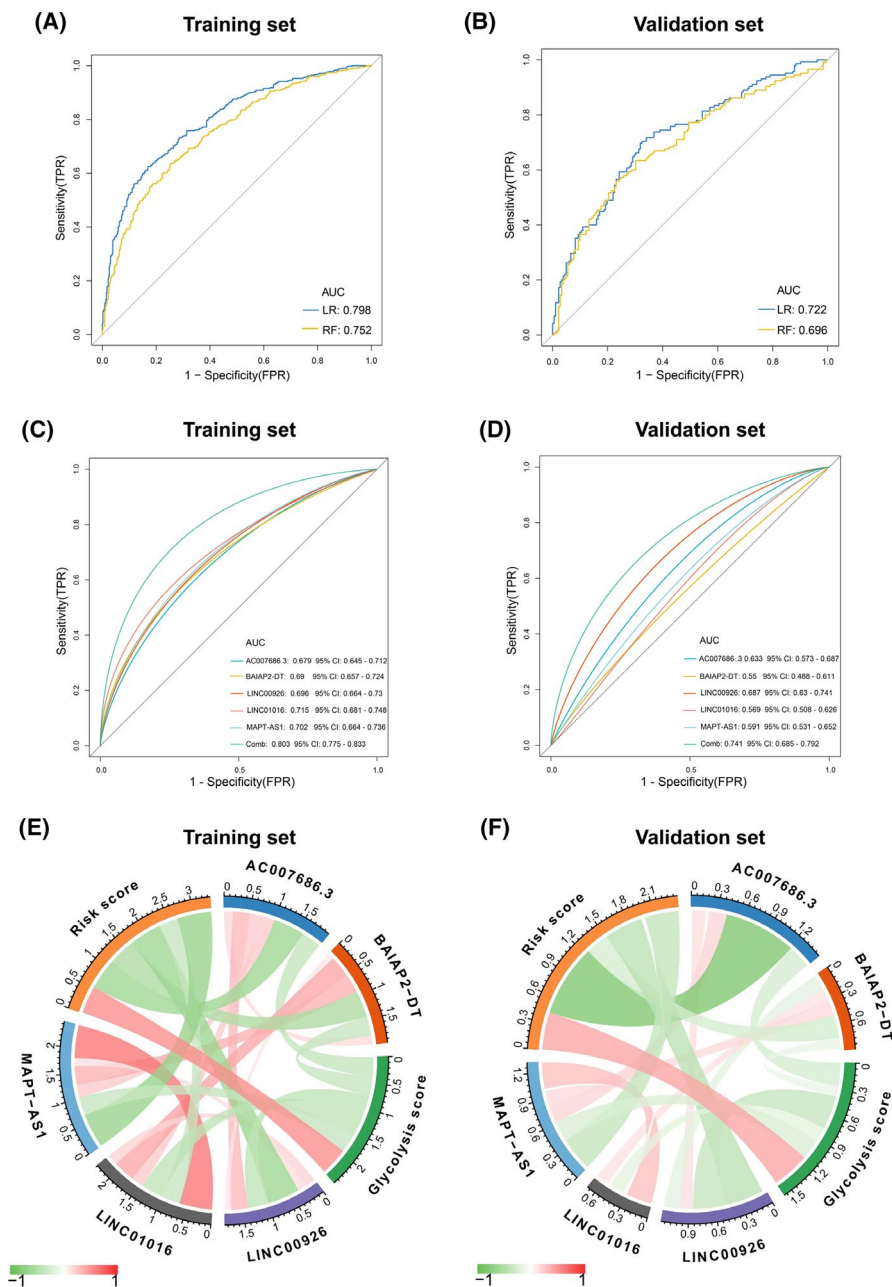


FIGURE 6 Further evaluation of the lncRNA signature in the training and validation sets. (A, B) Logistic regression (LR) and random forest (RF) were used to evaluate the signature's performance in stratifying glycolysis status. (C, D) Logistic regression analysis was used to identify that the lncRNA signature predicted glycolysis status more efficiently than individual lncRNA. (E, F) The interplay among risk scores, glycolysis scores, and five lncRNAs

groups. GSEA showed that the mTORC1 signaling pathway was most significantly enriched in the high-risk group (NES = 2.07, FDR = 0.015) (Figure 8A), which suggested that the lncRNA signature may contribute to the regulation of the mTORC1 signaling pathway. Moreover, the interplay between mTORC1 signaling, glycolysis signaling, and the prognostic signature is shown in Figure 8B, which revealed that mTORC1 signaling was significantly positively correlated with glycolysis signaling, and the high-risk group displayed higher levels of enrichment for mTORC1 and glycolysis signaling compared to the low-risk group.

To validate and further clarify the association between the lncRNA signature and hub hallmark gene sets, we performed GSA as described in Figure 8C,D. Accordant

with the above results, we observed that several hallmark gene sets related to cell growth were significantly upregulated in the training and validation sets; these gene sets included mTORC1 signaling, G2M checkpoints, E2F targets, unfold protein response, mitotic spindle, glycolysis, and MYC targets V1.

To estimate whether the lncRNA signature can predict clinical response to mTOR inhibitors, we extracted the data of related drugs from the Genomics of Drug Sensitivity in Cancer (GDSC),²⁸ including four mTOR inhibitors (Rapamycin, AZD8055, NVP-BEZ235, and Temozolimus). Our results indicated that the high-risk group appeared as a higher IC50 value for mTOR inhibitors than the low-risk group (Figure 8E). We next treated MDA-MB-231 cells with 10 nM rapamycin, an mTORC1 inhibitor. MTT assay

TABLE 2 The chi-squared test of the association between the lncRNA signature and clinicopathological features in TCGA breast cancer dataset

Features	Alive (768)	Dead with high risk (79)	Dead with low risk (41)	Total (888)	p value
Age					
Mean (SD)	57.5 (12.6)	60.2(15.4)	63.2(14.3)	58(13)	
Median (Min, Max)	57.5 (26, 89)	60 (26, 90)	66 (34, 88)	58 (26,90)	
Subtype					
BRCA_Basal	135 (17.6\%)	19 (24.1\%)	2 (4.9\%)	156 (17.6\%)	
BRCA_Her2	53 (6.9\%)	14 (17.7\%)	NA	67 (7.5\%)	
BRCA_LumA	406 (52.9\%)	22 (27.8\%)	29 (70.7\%)	457 (51.5\%)	
BRCA_LumB	146 (19.0\%)	20 (25.3\%)	8 (19.5\%)	174 (19.6\%)	
BRCA_Normal	28 (3.6\%)	4 (5.1\%)	2 (4.9\%)	34 (3.8\%)	0.004*
AJCC stage					
Stage I	142 (18.5\%)	8 (10.1\%)	7 (17.1\%)	157 (17.7\%)	
Stage II	459 (59.8\%)	36 (45.6\%)	17 (41.5\%)	512 (57.7\%)	
Stage III	163 (21.2\%)	26 (32.9\%)	13 (31.7\%)	202 (22.7\%)	
Stage IV	4 (0.5\%)	9 (11.4\%)	4 (9.8\%)	17 (1.9\%)	2.4e-13*
Cancer status					
Tumor free	668 (87.0\%)	32 (40.5\%)	17 (41.5\%)	717 (80.7\%)	
With tumor	20 (2.6\%)	34 (43.0\%)	15 (36.6\%)	69 (7.8\%)	1.9e-52*
Glycolysis status					
High	375 (48.8\%)	56 (70.9\%)	13 (31.7\%)	444 (50.0\%)	
Low	393 (51.2\%)	23 (29.1\%)	28 (68.3\%)	444 (50.0\%)	5.3e-05*

*Significant.

showed that rapamycin indeed resulted in cell growth inhibition compared to vehicle-treated group (Figure 8F). On the other hand, expression levels of LINC00926, LINC01016, AC007686.3, and MAPT-AS1 also significantly increased in response to rapamycin treatment (Figure 8G), which indicated that the mTORC1 signaling pathway was negatively associated with the expression of signature lncRNAs.

3.7 | Cell growth-related signaling significantly activated in the high-risk group

Given that the hallmark gene sets related to cell growth were upregulated in the high-risk group, we continued to perform WGCNA to identify the biological processes involved in the lncRNA signature. As presented in Figure 9A, the genes enrolled in the training set were clustered into 18 modules using cluster analysis. Subsequently, the brown module was found to be highly associated with the lncRNA signature (Figure 9B). In the brown module, 93 hub genes were selected based on the criteria of MM values greater than 0.8, and a GS value greater than 0.4 (Figure 9C; Table S5). Finally, biological process

enrichment analysis of hub genes from the brown module was performed using Metascape. As expected, we found that cell growth-related signaling, including cell cycle, cell division, and regulation of cell cycle process, were significantly enriched in the high-risk group (Figure 9D,E). Taken together, the lncRNA signature may be associated with tumor malignant progression and higher mortality risk by promoting tumor cell proliferation.

4 | DISCUSSION

For decades, great advances have been made in breast cancer treatment; however, several mechanisms associated with breast cancer progression remain elusive.^{29–32} Reprogrammed energy metabolism is currently identified as an emerging hallmark of cancer cells.⁶ This alteration is characterized by preferential dependence on glycolysis for energy production even in the presence of adequate oxygen and fully functioning mitochondria, namely ‘aerobic glycolysis’ or ‘Warburg effect’.^{3,9,33} Furthermore, previous studies have shown that tumor aerobic glycolysis frequently contributes to poor clinical outcomes in patients with breast cancer.^{34–36} Thus, continued investigation of aerobic glycolysis could help to gain insight into

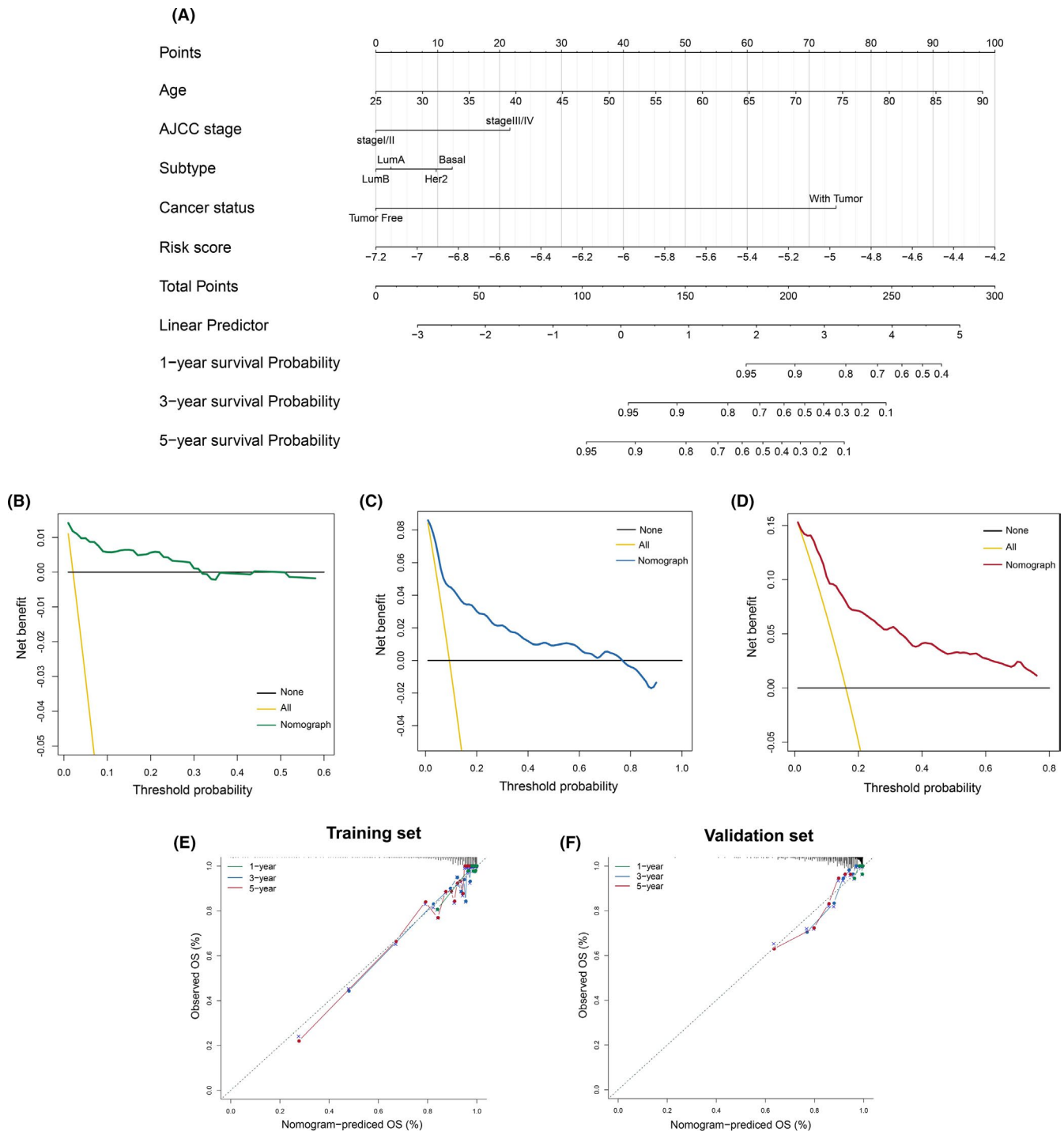


FIGURE 7 Construction and evaluation of the nomogram. (A) Construction of the nomogram. (B–D) Decision curve analysis associated with the nomogram at 1, 3, and 5 years. Note that yellow line: net benefit of all-treat scheme; black line: net benefit of none-treat scheme; green line: net benefit of the nomogram for predicting 1-year survival time; blue line: net benefit of the nomogram for predicting 3-year survival time; red line: net benefit of the nomogram for predicting 5-year survival time. E–F, Calibration curve of the nomogram in the training and validation sets. The prognostic model's performance at 1, 3, and 5 years is presented by green, blue, and red lines, respectively. And the gray dashed line of 45° represents the ideal performance

the crucial mechanism of breast cancer initiation and progression and develop better prognostic indicators, diagnostic biomarkers, and therapeutic targets for breast cancer patients.

LncRNAs were previously reported to be involved in tumor metabolism reprogramming.^{37–41} In this study, we developed a glycolysis score to further construct a five glycolysis-related lncRNA signature, which was associated

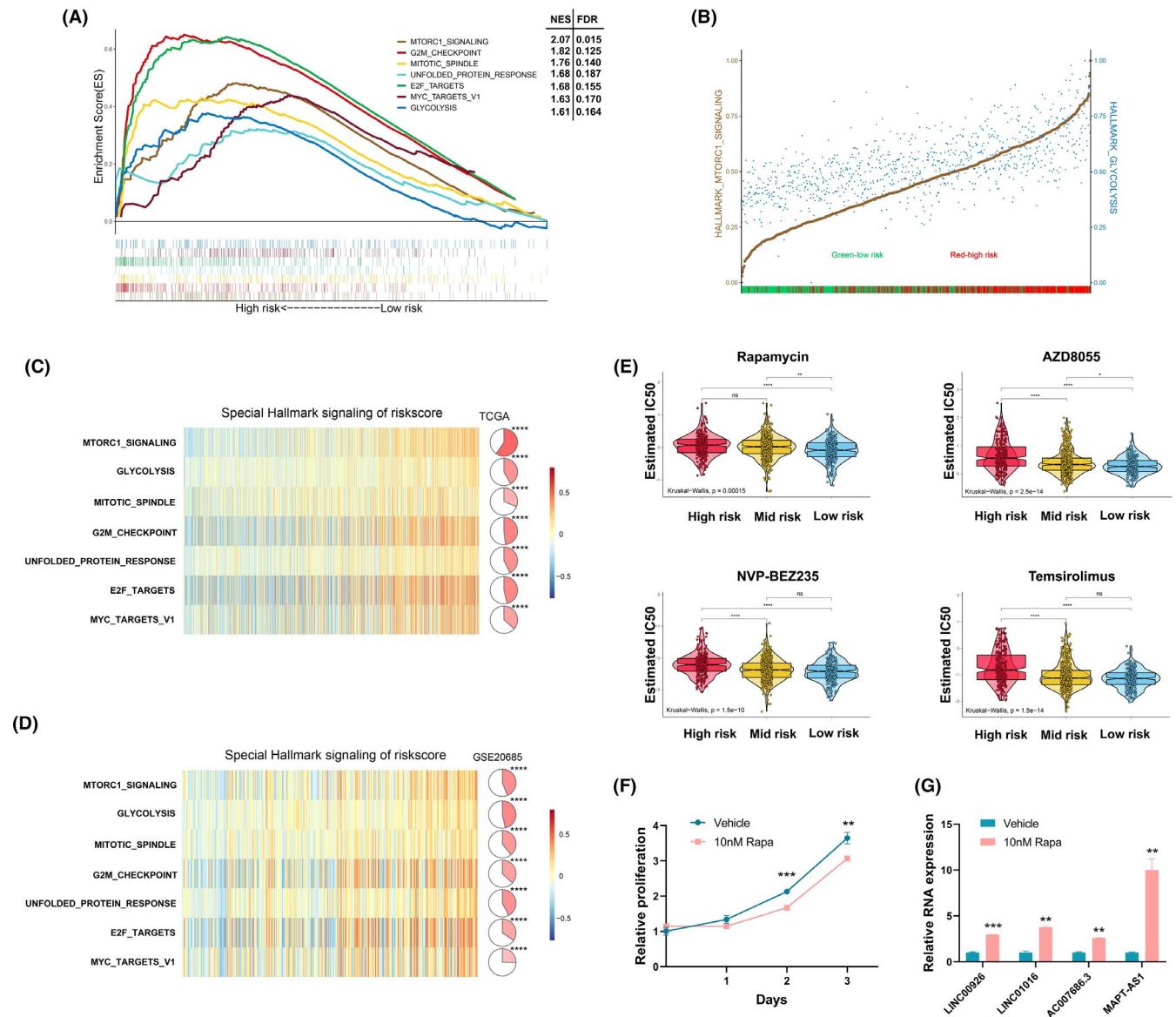


FIGURE 8 The cancer-related hallmark gene sets associated with the lncRNA signature in breast cancer. (A) Gene set enrichment analysis. (B) The interplay between the lncRNA signature, mTORC1 signaling, and glycolysis signaling. Brown: the ssGSEA score of mTORC1 signaling; Blue: the ssGSEA score of glycolysis signaling; Red: high-risk patients; Green: low-risk patients. The ssGSEA score was scaled to a range between 0 and 1 in the plot. (C, D) Gene set variation analysis. (E) The GDSC drug response data were used to estimate the association between the lncRNA signature and mTOR inhibitors in TCGA breast cancer patients. (F) Proliferation of MDA-MB-231 cells treated with vehicle control or 10 nM rapamycin. (G) Signature lncRNAs with significant expression change in response to rapamycin treatment ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$)

with malignant progression of breast cancer and acted as an independent prognostic factor in breast cancer patients. Subsequent *in vitro* experiments also supported these findings. Moreover, the lncRNA signature could well distinguish patients with unfavorable prognosis from those with favorable prognosis. Further analyses demonstrated that the lncRNA signature had moderate discrimination for glycolysis status, and the combination of five lncRNAs possessed better predictive efficacy for glycolysis status compared with each lncRNA from the prognostic model. Importantly, a consistent performance was observed in the validation set. In an

effort to enhance the predictive efficacy of the lncRNA signature, we further integrated age, AJCC stage, subtype, cancer status, and the lncRNA signature to develop a nomogram which predicts the efficacy for survival and clinical utility, and was validated by calibration curve, C-index, and DCA, respectively. Lastly, our findings suggest that the nomogram based on the lncRNA signature could contribute to predicting survival probability and help to guide personalized therapeutic strategies for breast cancer patients.

lncRNAs play pivotal roles in metabolism reprogramming of breast cancer by regulating important cancer-related

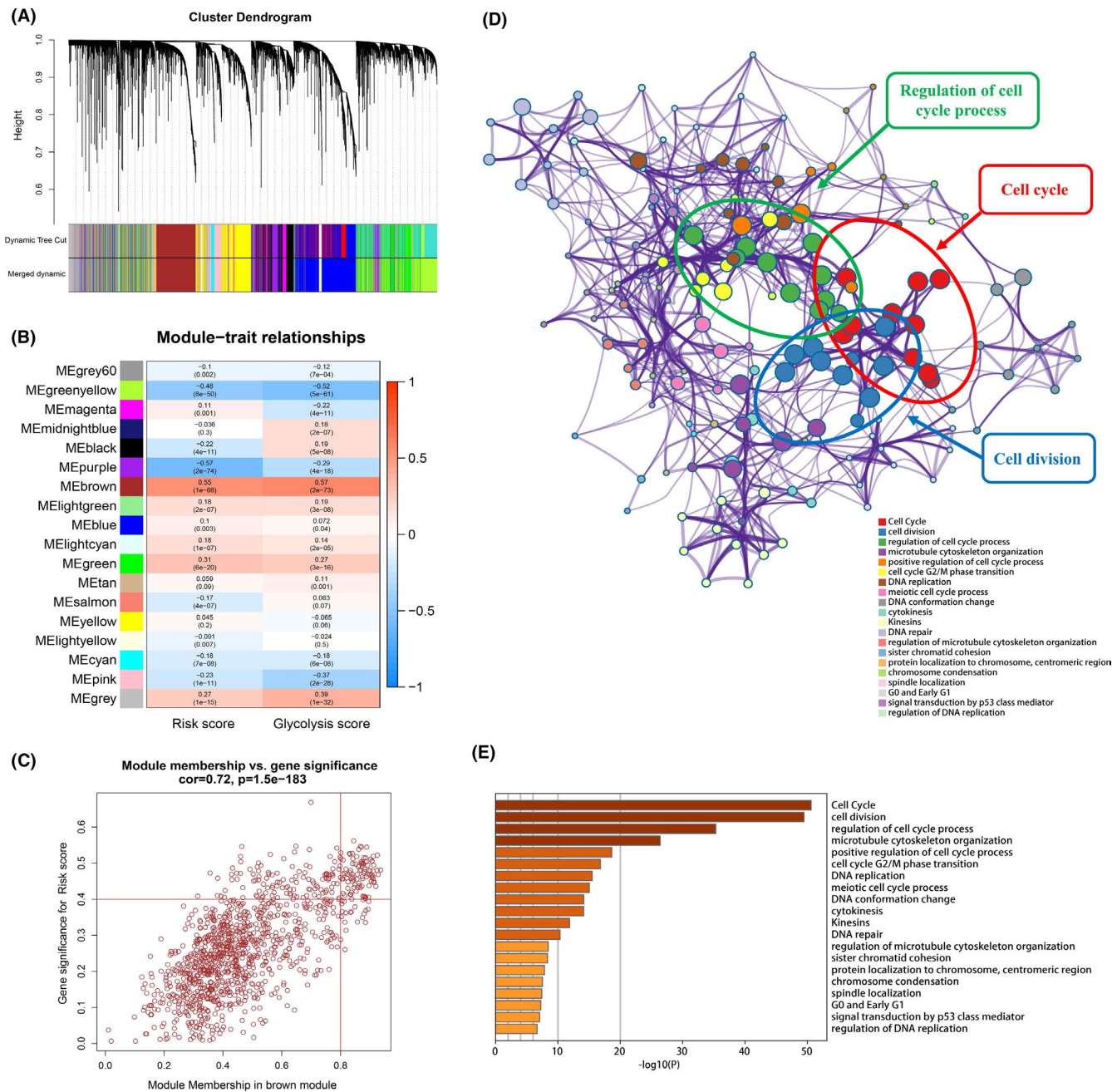


FIGURE 9 Cell growth-related signaling significantly activated in the high-risk group. (A) Clustering dendrogram of mRNAs. The two colored rows represent the original modules and merged modules, respectively. (B) The relationship between modules and traits. (C) A scatter plot of GS for risk scores versus MM for brown module. Red line represents the screening criteria: MM value greater than 0.8 and GS value greater than 0.4. (D, E) Biological process enrichment analysis of 93 hub genes from the brown module

pathways.^{42–44} GSEA of hallmark gene sets was performed and identified that mTORC1 signaling was significantly enriched in the high-risk group. In addition, we observed that mTORC1 signaling was positively correlated with glycolysis signaling. According to previous reports, the mTOR signaling pathway could integrate both intracellular and extracellular signals and function as a central pathway in tumor initiation and progression.^{45–47} Accumulating evidence also demonstrated that the mTORC1 signaling pathway may act as a mediator of aerobic glycolysis to promote

cell proliferation.^{45,48,49} Subsequent GSVA further identified that the risk score was positively correlated with the mTORC1 signaling pathway as well as other hallmark gene sets associated with cell growth. Notably, the results were mutually validated in two independent datasets.

To further investigate the biological processes related to the lncRNA signature, we applied WGCNA and identified that the brown module was highly associated with the risk score and glycolysis score. Furthermore, Metascape was conducted and demonstrated that hub genes sourced

from the brown module were significantly enriched in cell growth-related signaling, which could promote tumor cell proliferation and contribute to a higher mortality risk in breast cancer patients.

When glycolysis inhibitors are employed, mTORC1 could be involved in metabolism reprogramming to escape from glycolytic dependency.⁵⁰ Currently, mTOR inhibitors have been used in clinical practice. Therefore, we tried to estimate the association between the lncRNA signature and drug response via the GDSC drug response data and in vitro experiments. Finally, our data suggested that the lncRNA signature can serve as a promising indicator for measuring response to mTOR inhibitors in breast cancer patients. Moreover, previous studies have shown that tumor cell proliferation could be inhibited by co-targeting glycolytic enzyme and mTORC1 signaling.⁵⁰ Given that the lncRNA signature had a significantly positive association with mTORC1 and glycolysis signaling, it may help in developing novel therapeutic strategies for combination therapy, and achieving desirable clinical benefits for breast cancer patients.

Our data presented here provided a basis for further exploration of metabolic strategies in breast cancer prognosis and treatment. However, this study was mainly based on the publicly available datasets. Because the public sample size is limited, further exploring these findings will be a crucial direction for our future work.

In conclusion, we identified five glycolysis-related lncRNAs to construct an lncRNA signature on the basis of the glycolysis score, which could predict the survival probability and glycolysis status. Moreover, hallmark gene sets associated with cell growth were significantly activated in the high-risk breast cancer patient subgroup. Overall, the lncRNA signature could function as a robust prognostic indicator and help to develop novel therapeutic strategies for breast cancer patients.

ACKNOWLEDGMENTS

This work was supported by National Key Research and Development Program (No. 2020YFA0712400), Special Foundation for Taishan Scholars (No. ts20190971), National Natural Science Foundation of China (No. 81874119; No. 82072912), Special Support Plan for National High Level Talents (Ten Thousand Talents Program W01020103), National Key Research and Development Program (No. 2018YFC0114705), Foundation from Clinical Research Center of Shandong University (No. 2020SDUCRCA015), and Qilu Hospital Clinical New Technology Developing Foundation (No. 2018-7; No. 2019-3).

CONFLICT OF INTEREST

No potential conflict of interest was reported by the authors.

ETHICAL APPROVAL

All analyses in this study were based on the publicly available database. No ethical approval and patient consent are required.

DATA AVAILABILITY STATEMENT

The datasets analyzed in this study were sourced from the TCGA (<http://cancergenome.nih.gov/>), METABRIC (<https://www.mbcproject.org/>), GEO (<https://www.ncbi.nlm.nih.gov/geo/>), and GDSC (<https://www.cancerrxgene.org/>).

ORCID

Zheng Li  <https://orcid.org/0000-0002-9828-5475>

Qifeng Yang  <https://orcid.org/0000-0003-0576-8513>

REFERENCES

1. Wei Y, Lv J, Guo YU, et al. Soy intake and breast cancer risk: a prospective study of 300,000 Chinese women and a dose-response meta-analysis. *Eur J Epidemiol.* 2020;35:567-578.
2. Mahmood N, Arakelian A, Khan HA, Tanvir I, Mazar AP, Rabbani SA. Upar antibody (Huatn-658) and zometa reduce breast cancer growth and skeletal lesions. *Bone Res.* 2020;8:18.
3. Ganapathy-Kanniappan S, Geschwind JF. Tumor glycolysis as a target for cancer therapy: progress and prospects. *Mol Cancer.* 2013;12:152.
4. Koppenol WH, Bounds PL, Dang CV. Otto warburg's contributions to current concepts of cancer metabolism. *Nat Rev Cancer.* 2011;11:325-337.
5. Locasale JW, Cantley LC. Metabolic flux and the regulation of mammalian cell growth. *Cell Metab.* 2011;14:443-451.
6. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144:646-674.
7. Zhou Y, Tozzi F, Chen J, et al. Intracellular ATP levels are a pivotal determinant of chemoresistance in colon cancer cells. *Can Res.* 2012;72:304-314.
8. Pitroda SP, Wakim BT, Sood RF, et al. STAT1-dependent expression of energy metabolic pathways links tumour growth and radioresistance to the warburg effect. *BMC Med.* 2009;7:68.
9. Liberti MV, Locasale JW. The Warburg effect: how does it benefit cancer cells? *Trends Biochem Sci.* 2016;41:211-218.
10. Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell.* 2013;152:1298-1307.
11. Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet.* 2015;47:199-208.
12. Anastasiadou E, Jacob LS, Slack FJ. Non-coding RNA networks in cancer. *Nat Rev Cancer.* 2018;18:5-18.
13. Costa FF. Non-coding RNAs, epigenetics and complexity. *Gene.* 2008;410:9-17.
14. Xiao ZD, Zhuang L, Gan B. Long non-coding RNAs in cancer metabolism. *BioEssays.* 2016;38:991-996.
15. Liu H, Luo J, Luan S, He C, Li Z. Long non-coding RNAs involved in cancer metabolic reprogramming. *Cell Mol Life Sci.* 2019;76:495-504.
16. Malakar P, Stein I, Saragovi A, et al. Long noncoding RNA MALAT1 regulates cancer glucose metabolism by

- enhancing mTOR-mediated translation of TCF7L2. *Can Res.* 2019;79:2480-2493.
17. Li Z, Li X, Wu S, Xue M, Chen W. Long Non-coding RNA UCA1 promotes glycolysis by upregulating hexokinase 2 through the mTOR-STAT3/MICRORNA143 pathway. *Cancer Sci.* 2014;105:951-955.
 18. Liu X, Xiao Z-D, Han L, et al. LncRNA NBR2 engages a metabolic checkpoint by regulating AMPK under energy stress. *Nat Cell Biol.* 2016;18:431-442.
 19. Hung CL, Wang LY, Yu YL, et al. A long noncoding RNA connects c-Myc to tumor metabolism. *Proc Natl Acad Sci USA.* 2014;111:18697-18702.
 20. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* 2014;15:550.
 21. Li Z, Li Y, Wang X, et al. LINC01977 promotes breast cancer progression and chemoresistance to doxorubicin by targeting miR-212-3p/GOLM1 axis. *Front Oncol.* 2021;11:657094.
 22. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47.
 23. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1:417-425.
 24. Rousson V, Zumbo T. Decision curve analysis revisited: overall net benefit, relationships to ROC curve analysis, and application to case-control studies. *BMC Med Inform Decis Mak.* 2011;11:45.
 25. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics.* 2013;14:7.
 26. Li Z, Li Y, Wang X, Yang Q. PPP2R2B downregulation is associated with immune evasion and predicts poor clinical outcomes in triple-negative breast cancer. *Cancer Cell Int.* 2021;21:13.
 27. Han L, Yuan Y, Zheng S, et al. The pan-cancer analysis of pseudogene expression reveals biologically and clinically relevant tumour subtypes. *Nat Commun.* 2014;5:3963.
 28. Yang W, Soares J, Greninger P, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 2013;41:D955-D961.
 29. Lawson DA, Bhakta NR, Kessenbrock K, et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature.* 2015;526:131-135.
 30. Sosa MS, Lopez-Haber C, Yang C, et al. Identification of the Rac-GEF P-Rex1 as an essential mediator of ErbB signaling in breast cancer. *Mol Cell.* 2010;40:877-892.
 31. Kessenbrock K, Smith P, Steenbeek SC, et al. Diverse regulation of mammary epithelial growth and branching morphogenesis through noncanonical Wnt signaling. *Proc Natl Acad Sci USA.* 2017;114:3121-3126.
 32. Maitland ML, Schilsky RL. Clinical trials in the era of personalized oncology. *Cancer J Clin.* 2011;61:365-381.
 33. Li Z, Zhang H. Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. *Cell Mol Life Sci.* 2016;73:377-392.
 34. Ma F, Zhang L, Ma L, Zhang Y, Zhang J, Guo B. MiR-361-5p inhibits glycolytic metabolism, proliferation and invasion of breast cancer by targeting FGFR1 AND MMP-1. *J Exp Clin Cancer Res.* 2017;36:158.
 35. Peng F, Wang J-H, Fan W-J, et al. Glycolysis gatekeeper PDK1 reprograms breast cancer stem cells under hypoxia. *Oncogene.* 2018;37:1062-1074.
 36. Ma F, Liu X, Zhou S, et al. Long non-coding RNA FGF13-AS1 inhibits glycolysis and stemness properties of breast cancer cells through FGF13-AS1/IGF2BPs/Myc feedback loop. *Cancer Lett.* 2019;450:63-75.
 37. Shankaraiah RC, Veronese A, Sabbioni S, Negrini M. Non-coding RNAs in the reprogramming of glucose metabolism in cancer. *Cancer Lett.* 2018;419:167-174.
 38. Chen J, Yu Y, Li H, et al. Long non-coding RNA PVT1 promotes tumor progression by regulating the miR-143/HK2 axis in gallbladder cancer. *Mol Cancer.* 2019;18:33.
 39. Zheng X, Han H, Liu GP, et al. LncRNA wires up Hippo and Hedgehog signaling to reprogramme glucose metabolism. *EMBO J.* 2017;36:3325-3335.
 40. Sun LY, Li XJ, Sun YM, et al. LncRNA anril regulates AML development through modulating the glucose metabolism pathway of AdipoR1/AMPK/SIRT1. *Mol Cancer.* 2018;17:127.
 41. Xiao Z-D, Han L, Lee H, et al. Energy stress-induced lncRNA FILNC1 represses C-Myc-mediated energy metabolism and inhibits renal tumor development. *Nat Commun.* 2017;8:783.
 42. Xing Z, et al. Expression of long noncoding RNA YIYA promotes glycolysis in breast cancer. *Can Res.* 2018;78:4524-4532.
 43. Chen F, Chen J, Yang L, et al. Extracellular vesicle-packaged HIF-1 α -stabilizing lncRNA from tumour-associated macrophages regulates aerobic glycolysis of breast cancer cells. *Nat Cell Biol.* 2019;21:498-510.
 44. Lin A, Li C, Xing Z, et al. The link-A lncRNA activates normoxic HIF1 α signalling in triple-negative breast cancer. *Nat Cell Biol.* 2016;18:213-224.
 45. Perl A. mTOR activation is a biomarker and a central pathway to autoimmune disorders, cancer, obesity, and aging. *Ann NY Acad Sci.* 2015;1346:33-44.
 46. Boutouja F, Stiehm CM, Platta HW. mTOR: a cellular regulator interface in health and disease. *Cells.* 2019;8:18.
 47. Murugan AK. mTOR: role in cancer, metastasis and drug resistance. *Semin Cancer Biol.* 2019;59:92-111.
 48. Laplante M, Sabatini DM. mTOR signaling in growth control and disease. *Cell.* 2012;149:274-293.
 49. Saxton RA, Sabatini DM. mTOR signaling in growth, metabolism, and disease. *Cell.* 2017;168:960-976.
 50. Pusapati RV, Daemen A, Wilson C, et al. mTORC1-dependent metabolic reprogramming underlies escape from glycolysis addiction in cancer cells. *Cancer Cell.* 2016;29:548-562.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Li Z, Zheng J, Feng Y, et al. Integrated analysis identifies a novel lncRNA prognostic signature associated with aerobic glycolysis and hub pathways in breast cancer. *Cancer Med.* 2021;10:7877-7892. doi:[10.1002/cam4.4291](https://doi.org/10.1002/cam4.4291)