A Novel Seven Gene Signature-Based Prognostic Model to Predict Distant Metastasis of Lymph Node-Negative Triple-Negative Breast Cancer

Wenting Peng1,2,3,4†, Caijin Lin1,2,3†, Shanshan Jing3,5†, Guanhua Su1,2,3, Xi Jin1,2,3*, Genhong Di1,2,3* and Zhiming Shao1,2,3*

1 Department of Breast Surgery, Fudan University Shanghai Cancer Center, Shanghai, China, 2 Key Laboratory of Breast Cancer in Shanghai, Fudan University Shanghai Cancer Center, Shanghai, China, 3 Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China, 4 Department of Breast Surgery, The Affiliated Changzhou No. 2 People’s Hospital of Nanjing Medical University, Changzhou, China, 5 Department of Nursing Administration, Fudan University Shanghai Cancer Center, Shanghai, China

Background: The prognosis of lymph node-negative triple-negative breast cancer (TNBC) is still worse than that of other subtypes despite adjuvant chemotherapy. Reliable prognostic biomarkers are required to identify lymph node-negative TNBC patients at a high risk of distant metastasis and optimize individual treatment.

Methods: We analyzed the RNA sequencing data of primary tumor tissue and the clinicopathological data of 202 lymph node-negative TNBC patients. The cohort was randomly divided into training and validation sets. Least absolute shrinkage and selection operator Cox regression and multivariate Cox regression were used to construct the prognostic model.

Results: A clinical prognostic model, seven-gene signature, and combined model were constructed using the training set and validated using the validation set. The seven-gene signature was established based on the genomic variables associated with distant metastasis after shrinkage correction. The difference in the risk of distant metastasis between the low- and high-risk groups was statistically significant using the seven-gene signature (training set: \( P < 0.001 \); validation set: \( P = 0.039 \)). The combined model showed significance in the training set \( (P < 0.001) \) and trended toward significance in the validation set \( (P = 0.071) \). The seven-gene signature showed improved prognostic accuracy relative to the clinical signature (training set: \( P = 0.001 \); validation set: \( P = 0.039 \)). The combined model showed improved prognostic accuracy relative to the clinical signature (AUC value of 4-year ROC: 0.879 vs. 0.699, \( P = 0.046 \)). Moreover, the composite clinical and gene signature also showed improved prognostic accuracy relative to the clinical signature (AUC value of 4-year ROC: 0.882 vs. 0.693, \( P = 0.038 \)). A nomogram model was constructed with the seven-gene signature, patient age, and tumor size.
INTRODUCTION

Breast cancer is estimated to be the most common cancer diagnosed in women and the second leading cause of cancer-related death in the United States in 2021 (1). Triple-negative breast cancer (TNBC) is characterized by a lack of expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), representing 10%-20% of all breast cancers (2, 3). TNBC is more likely to show lymph node involvement at diagnosis and exhibit invasive and metastatic tendencies (2, 4). Nonetheless, the incidence of lymph node-negative TNBC has markedly increased owing to early detection and initiated screening programs (5–8).

To date, lymph node-negative TNBC is generally considered at moderate risk of disease recurrence and is often recommended for adjuvant chemotherapy (9). Small lymph node-negative tumors tend to have an excellent prognosis without chemotherapy (10). However, the risk of metastasis and death of partial lymph node-negative TNBC patients is still high despite the high proportion of adjuvant chemotherapy (2, 11, 12). A more quantitative approach is required to inform the risk of distant metastasis and individualized treatment in lymph node-negative TNBC.

Several multigene assays have been developed to facilitate prognosis prediction and treatment planning in early-stage breast cancer, but most of the enrolled patients are hormone receptor-positive (13–15). Although many publications have attempted to identify gene signatures that predict the prognosis of TNBC patients, several limitations need to be considered due to the limited sample size and incomplete follow-up information (16–20). Above all, most previous studies include all TNBC patients as a cohort. Because lymph node status is a well-known prognostic value, there is an urgent need to identify a robust risk stratification tool for lymph node-negative TNBC patients (21, 22). Based on detailed clinicopathological information, well-documented follow-up, and complete RNA-sequencing data, we constructed a gene expression-based prognostic signature combined with clinicopathological factors to provide quantitative predictions of short- and long-term disease outcomes for Chinese lymph node-negative TNBC patients.

MATERIALS AND METHODS

Patient Samples and Study Design

We included 202 eligible patients from our previously published cohort of 465 primary TNBC patients treated at Fudan University Shanghai Cancer Center (FUSCCTNBC) (23). Patients were included based on the following criteria: histologic diagnosis of lymph node-negative TNBC with RNA-sequencing data and follow-up information for recurrence and metastasis. The RNA-sequencing data are available in the Sequence Read Archive (RNA-seq: SRP157974). Patients with contralateral breast cancer, lymph node recurrence, and unknown sites of recurrence were excluded. Lymph node status was independently confirmed by two experienced pathologists. The date of diagnosis of metastasis was defined when metastasis was either confirmed by biopsy or clinically diagnosed. The follow-up of this cohort was completed on June 11, 2019. Distant metastasis-free survival (DMFS) was defined as the interval between diagnosis and the first distant metastasis (viscera/bone/brain). Patients without events were censored from the time point of the last follow-up.

Ethics Statement

The present study was reviewed and approved by the Ethics Committee of Fudan University Shanghai Cancer Center (Ethics number: 050432-4-1212B). The patients provided written informed consent to participate in this study.

Gene Selection and Risk-Score Algorithm

To identify mRNAs of prognostic value, analysis for differentially expressed mRNAs between two groups was performed using the limma package (version 3.48.0) in R software. We also performed Gene Set Enrichment Analysis (GSEA) of differentially expressed genes between the two groups with or without distant metastasis using the RNA-sequencing data and GSEA software (GSEA_4.1.0) (24, 25).

The cohort was randomly divided into the training set (n=142) and validation set (n=60) at a ratio of 7 to 3 by the caret package (version 6.0-88) in R software. Pearson chi-square test or Fisher’s exact test was used to ensure that there was no significant difference and that no bias was introduced in clinicopathological characteristics between the two sets. Least absolute shrinkage and selection operator (LASSO) Cox regression analysis was performed to further filter the differentially expressed mRNAs. A multivariate Cox regression model was used to determine the coefficient of each factor. The risk score of each model was used to estimate the probability of distant metastasis. The genomic risk score was calculated from individual gene expression measurements as follows: Genomic risk score = (β_B3GALT5-AS1 × B3GALT5-AS1) + (β_DNER × DNER) + (β_CPS1 × CPS1) + (β_KIF5A × KIF5A) + (β_SIX3 × SIX3) + (β_NOTUM × NOTUM) + (β_CPS1 × CPS1).

Conclusions: The proposed signature may improve the risk stratification of lymph node-negative TNBC patients. High-risk lymph node-negative TNBC patients may benefit from treatment escalation.

Keywords: triple-negative breast cancer, distant metastasis, prognostic biomarker, modeling, transcriptomics
Validation of Different Prognostic Models
Patients were stratified into high- and low-risk groups based on optimum cutoff risk scores determined by the “surv_cutpoint” function in the survminer package (version 0.4.9) in R software. Kaplan–Meier analyses and log-rank tests were performed to assess the differences in DMFS between the high- and low-risk groups. The time-dependent receiver operating characteristic (ROC) curve was used to measure the prognostic performance by comparing the area under the ROC curve (AUC) values.

Construction and Validation of a Nomogram Model
Based on data availability and clinical evidence (9, 26, 27), a nomogram was constructed integrating the seven-gene risk score, age of the patients at surgery, and pathological tumor size. We measured the predictive accuracy of the nomogram via Harrell’s concordance index (C-index) in the training and validation sets. In addition, the predictive capacity of the nomogram was also evaluated using calibration curve and decision curve analysis (DCA).

Statistical Analysis
Pearson’s chi-square test or Fisher’s exact test was used to compare the clinical and pathological characteristics between the training set and validation set. All statistical analyses were performed using the SPSS 22.0 (SPSS Inc.) or R software (version 4.1.0, www.r-project.com). A value of $P < 0.05$ was considered statistically significant.

RESULTS
Patient Characteristics
The clinical and pathological characteristics of 202 patients and their primary tumors are summarized in Table 1. Of 202 lymph node-negative TNBC patients, the median follow-up was 68.2 months (interquartile range, 57.6-80.6 months). Overall, 12 (5.9%) cases with distant metastasis were observed. Of the 12 patients, 4 (33.3%) patients had multisite metastasis, and 7 (58.3%) patients died due to breast cancer during follow-up. The median tumor size and age of the patients at surgery in this study cohort were 2.5 centimeters (range 0.8-12.0) and 53 years (range 25-82), respectively.

Construction and Validation of the Novel Seven-Gene Signature
An overview of the study design is shown in Figure 1. Using log$_2$ (fold change) $> 1$ or $< -1$ and $P < 0.05$, we identified 71 differentially expressed mRNAs between the two groups with or without distant metastasis. We also performed Gene Set Enrichment Analysis of differentially expressed genes between the two groups with or without distant metastasis using the RNA-sequencing data. In patients with distant metastasis, 25 gene sets were significantly enriched at nominal $P$ value $< 0.05$. The top ten gene sets enriched in 12 lymph node-negative TNBC patients with distant metastasis compared to 190 patients without distant metastasis were illustrated in Figure S1. In patients with distant metastasis, 56 mRNAs were upregulated, whereas 15 mRNAs were downregulated (Figure 2).

We constructed a matrix integrating RNA-sequencing data of 71 differentially expressed mRNAs and clinicopathological data of all 202 patients. Next, patients were randomly classified into the training set (n = 142) and validation set (n = 60). There was no difference in all characteristics between the training and internal validation sets (Table 1). Seven genes, including B3GALT5-AS1, DNER, CSN1S1, KIF5A, SIX3, NOTUM, and CPS1, were selected using the LASSO Cox regression model in the training set. The summary of log$_2$(fold change), multivariable Cox regression coefficient, hazard ratio, 95% confidence interval, and $P$ value for selected genes are presented in Table 2. Time-dependent ROCs and Kaplan–Meier curves were used to evaluate the prognostic potential of the seven-gene signature for DMFS (Figures 3A, B). The AUC values for 3-, 4-, and 5-year DMFS were 0.823, 0.879, and 0.870 in the training set and 0.727, 0.705, and 0.689 in the validation set, respectively (Figure 3A). The formula of genomic risk score is as follows: genomic risk score = 0.18801037 × CPS1 - 0.06761698 × NOTUM + 0.3601127 × KIF5A + 0.57677377 × SIX3 + 0.70105693 × CSN1S1 + 0.30611227 × DNER + 0.28358112 × B3GALT5-AS1.

In patients with distant metastasis, the seven genes were statistically significantly different in the testing set (Table 1). The formula of genomic risk score was used to calculate the genomic risk score, and the results were compared between the high-risk group (genomic risk score $> 0.14$) and the low-risk group (genomic risk score $< 0.14$) in the training set and validation set. The AUC values for the genomic risk score were 0.865, 0.861, and 0.860 in the training set and 0.727, 0.705, and 0.689 in the validation set.

Since the genomic risk score is a continuous variable, we determined the optimal cutoff value using the Youden index. The genomic risk score was divided into high- and low-risk groups based on the optimal cutoff value (1.78). The patients in the high-risk group had a significantly lower survival rate than those in the low-risk group (Figure 3B). The formula of genomic risk score was used to calculate the genomic risk score, and the results were compared between the high-risk group (genomic risk score $> 1.78$) and the low-risk group (genomic risk score $< 1.78$) in the training set and validation set. The AUC values for the genomic risk score were 0.865, 0.861, and 0.860 in the training set and 0.727, 0.705, and 0.689 in the validation set.

*P values were calculated using Pearson’s chi-square test or Fisher’s exact test to compare the clinical and pathological characteristics between the training set and validation set.

Table 1 | Clinicopathological characteristics of patients and their tumors.

| Characteristics | Number of patients (%) | $P^*$ |
|-----------------|------------------------|------|
| Age, years      |                        |      |
| ≤50             | 86 (42.6%)             | 0.865|
| >50             | 116 (57.4%)            |      |
| Menopausal status|                        | 0.468|
| Premenopausal   | 75 (37.1%)             |      |
| Postmenopausal  | 127 (62.9%)            |      |
| Histological grade|                      | 0.183|
| I               | 35 (17.3%)             |      |
| II              | 13 (6.4%)              |      |
| III             | 134 (66.3%)            |      |
| Unknown         | 20 (9.9%)              |      |
| Tumor size      |                        | 0.239|
| ≤2cm            | 85 (42.1%)             |      |
| >2-5cm          | 111 (55.0%)            |      |
| >5cm            | 6 (3.0%)               |      |
| Ki-67           |                        | 0.820|
| ≤20%            | 28 (13.9%)             |      |
| >20%            | 169 (83.7%)            |      |
| Unknown         | 5 (2.5%)               |      |
| Chemotherapy    |                        | 0.644|
| No              | 6 (3.0%)               |      |
| Yes             | 186 (93.1%)            |      |
| Unknown         | 8 (4.0%)               |      |
| Radiotherapy    |                        | 0.861|
| No              | 180 (89.1%)            |      |
| Yes             | 21 (10.4%)             |      |
| Unknown         | 1 (0.5%)               |      |
| Metastasis      |                        | 0.345|
| No              | 190 (94.1%)            |      |
| Yes             | 12 (5.9%)              |      |

$P^*$ values were calculated using Pearson’s chi-square test or Fisher’s exact test to compare the clinical and pathological characteristics between the training set and validation set.
were also divided into high-risk (n = 8) and low-risk (n = 52) groups in the validation set (Figures 3B, C). The Kaplan-Meier analyses for DMFS as a function of the seven-gene signature showed highly significant differences between the high- and low-risk groups (Figure 3B, \( P < 0.001 \) in the training set; \( P = 0.039 \) in the validation set).

Construction and Validation of the Combined Gene and Clinical Model

We also created a clinical prognostic model using the following clinically significant predictors: age and tumor size. The summary of multivariable Cox regression coefficient, hazard ratio, 95% confidence interval, and \( P \) value for age and tumor size are presented in Table S1. The formula of clinical risk score is as follows: clinical risk score = 0.21532 × Tumor size (cm) - 0.04466 × Age (years). The AUC values of the clinical model for 3-, 4-, and 5-year DMFS were 0.755, 0.699, and 0.693 in the training set and 0.574, 0.651, and 0.631 in the validation set, respectively (Figure 4A). The genomic risk score remained an independent prognostic factor in the multivariate Cox analysis after adjusting for patient age and tumor size in both the training set (hazard ratio = 2.64, 95% CI: 1.76-3.96, \( P < 0.001 \)) and validation set (hazard ratio = 1.63, 95% CI: 1.07-2.49, \( P = 0.02 \)). The combined risk score was derived from the genomic and clinical risk score as follows: combined risk score = 0.9702 × Genomic risk score + 1.0854 × Clinical risk score. After integrating the clinical model with the genomic risk score, the AUC values for 3-, 4-, and 5-year DMFS were 0.836, 0.888, and 0.882 in the training set, respectively (Figure 4B). The AUC values of the combined model remained high in the validation set with values of 0.801, 0.793, and 0.768 for 3-, 4-, and 5-year DMFS, respectively (Figure 4B). Patients were stratified into high- (n = 15 or 9) and low-risk groups (n = 127 or 51) in the training set or validation set (Figure 4C). The Kaplan-Meier analyses for DMFS as a function of the combined model showed a significant difference between the high- and low-risk groups in the training set (Figure 4C, \( P < 0.001 \)). Likewise, the trend was also observed in the validation set (Figure 4C, \( P = 0.071 \)).

Construction and Validation of a Predictive Nomogram

We integrated the seven-gene signature with age and tumor size to construct a prognostic nomogram in the training set (Figure 5A). The C-index value for the combined models was

| Gene symbol | Log2 FCa | Coefficientb | HR (95% CI)b | \( P^b \) |
|-------------|---------|--------------|-------------|------|
| B3GALT5-AS1 | 1.18    | -0.06761697  | 0.93 (0.41-2.15) | 0.87 |
| DNER        | 1.60    | 0.18801037   | 1.21 (0.39-3.73) | 0.74 |
| CSN1S1      | 1.61    | 0.28358112   | 1.33 (1.03-4.30) | 0.11 |
| KIF5A       | 1.10    | 0.36011127   | 1.43 (0.79-2.61) | 0.24 |
| SIX3        | 1.28    | 0.57677377   | 1.78 (0.92-3.44) | 0.09 |
| NOTUM       | 1.81    | 0.70105693   | 2.02 (1.22-3.33) | 0.01 |
| CPS1        | 1.51    | 0.74508978   | 2.11 (1.08-4.30) | 0.04 |

FC, fold change; HR, hazard ratio; CI, confidence interval.

aThe difference in the expression of seven genes between the group with and without distant metastasis was calculated using the limma package in R software.

bThe coefficients, hazard ratios, 95% confidence intervals, and \( P \) values of seven genes were calculated using a multivariate Cox proportional hazards regression model.
0.874 in the training set and 0.805 in the validation set. The 4- and 5-year time-dependent ROC curves for the seven-gene, clinical, and combined models are illustrated in Figure 5B. Both the seven-gene model and combined model showed better prognostic performance than the clinical model for predicting 4-year DMFS ($P = 0.046$ for the gene model; $P = 0.029$ for the combined model). The combined model showed significantly better prognostic performance than the clinical model for predicting 5-year DMFS ($P = 0.038$), and the seven-gene model also trended toward significance ($P = 0.065$). The calibration analysis of the 4-year DMFS prediction is shown in Figure 5C. The solid blue line has a closer fit to the dotted gray line, indicating great predictive accuracy of the nomogram. Decision curve analysis (DCA) revealed that compared to the clinical model, the seven-gene model and combined model were superior in predicting 4-year DMFS (Figure 5D).

**DISCUSSION**

We constructed a novel seven-gene signature (B3GALT5-AS1, DNER, CSN1S1, KIF5A, SIX3, NOTUM, and CPS1) and a combined prognostic model integrating a seven-gene signature with patient age and tumor size to quantify the likelihood of
distant metastasis in lymph node-negative TNBC. Both the seven-gene signature and the combined prognostic model had higher AUC values for 4- and 5-year survival than the clinical model. Patients were divided into low- and high-risk groups based on optimal cutoff values. Compared to the low-risk group, patients in the high-risk group had significantly poorer DMFS in both the training set and validation set. Finally, we constructed a prognostic nomogram and validated it in an internal validation set.

Several multigene assays have been employed in breast cancer, including the 76-gene signature, MammaPrint® (70-gene profile), Breast Cancer Index (BCI) test, Oncotype® DX Breast Recurrence Score (RS), EndoPredict® (EP), and Prosigna® (Risk Of Recurrence, ROR) (13, 28–32). None of the above is specifically designed and validated for TNBC patients. Most previous prognostic evaluation studies have focused on all TNBC patients (20, 33–37). One publication has reported the first validated proteomic signature of lymph node-negative TNBC patients (38), but all patients involved in this study were adjuvant treatment-naive, differing from actual clinical practice. The present study focused only on lymph node-negative TNBC patients with more than 90% of patients receiving adjuvant treatment. Apart from the study cohort, the flowchart to construct the gene signature in our study differed from previous studies. The seven differentially expressed mRNAs between the two groups with or without distant metastasis were utilized in our study, while we constructed our previous integrated mRNA-lncRNA signature after comparing the

![FIGURE 4](image-url)
tumor tissues with the paired normal tissues as in most previous studies (39, 40). Therefore, genes selected for model development in the present study correlated more closely to prognosis based on well-documented follow-up information. Although more than 90% of patients received adjuvant chemotherapy in our cohort, the high-risk groups classified by the seven-gene signature and combined model presented poor DMFS within four years after surgery. Chemotherapy escalation may be required for these patients.

Among the seven genes, B3GALT5-AS1 was the only RNA gene. A previous study has revealed the suppressive roles of the B3GALT5-AS1/miR-203/epithelial-mesenchymal transition (EMT) regulation axis in colon cancer liver metastasis (41). Similarly, B3GALT5-AS1 was the only gene with a negative

---

**FIGURE 5** | A predictive nomogram was established in the training set. AUC, area under the curve. (A) The nomogram was built by the seven-gene risk score and clinical characteristics, including age and tumor size. (B) The time-dependent receiver operating characteristic (ROC) curves of the seven-gene model, clinical model, and combined model for 4- and 5-year distant metastasis-free survival (DMFS). The combined model was better than the clinical model for predicting 4-year ($P = 0.029$) and 5-year ($P = 0.038$) DMFS. (C) Calibration plots of the nomogram for 4-year DMFS. (D) Decision curve analysis (DCA) of the seven-gene model, clinical model, and combined model for 4-year DMFS.
correlation coefficient in the present study. Delta/Notch-like EGF repeat containing (DNER) is a transmembrane protein that regulates EMT to enhance the proliferation and metastasis of breast cancer cells via the Wnt/β-catenin pathway (42). The other three genes, SIX3, NOTUM, and CPS1, have also been reported in other types of tumors. A systematic meta-analysis of non-small cell lung cancer has indicated that higher expression of SIX homeobox 3 (SIX3) is associated with a greater probability of tumorigenesis and a higher TNM stage (43). NOTUM acts as a key negative regulator of the Wnt signaling pathway, and knockdown of NOTUM genes inhibits the proliferation and migration of colorectal cancer cells (44). Previous studies have demonstrated that CPS1 expression is upregulated in glioblastoma multiforme and that overexpression of CPS1 is associated with poor therapeutic response and adverse outcomes among rectal cancer patients receiving concurrent chemoradiotherapy (45, 46). Inconsistent with our study, Mou et al. found a positive correlation between the lower expression of Csn1si1 and patients surviving with breast cancer (47). Kinesin family member 5A (KIF5A) encodes a member of the kinesin family of proteins. Previous research has confirmed that kinesin overexpression correlates with specific taxane resistance in basal-like breast cancer cell lines and tissues (48). Investigational kinesin protein inhibitors, such as GSK-923295, may be promising drugs in the future.

Our study had several limitations. First, external validation is required to ensure generalization. Second, our study did not explore the expression and prognostic effects of the seven genes at the protein level due to the incomplete protein expression information of partial genes. Finally, the reliability of our prognostic model needs further clinical validation.

In conclusion, we identified and validated a novel seven-gene signature model and constructed a nomogram combined with the patient age and tumor size for predicting DMFS in lymph node-negative TNBC patients. A higher risk score may indicate an increased likelihood of distant metastasis and vice versa. After taking the potential benefits and increased risks of distant metastasis into account, treatment escalation may be considered as an alternative strategy for lymph node-negative TNBC patients with a high-risk score. In contrast, de-escalation chemotherapy might be taken into consideration in patients with a low-risk score.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Fudan University Shanghai Cancer Center. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

WP, CL, SJ, XJ, GD, and ZS: study concept and design. WP, CL, SJ, and GS: data analysis and interpretation. WP: wrote the first draft of the manuscript. WP, CL, and GS: visualization. XJ and ZS: funding acquisition. XJ, GD, and ZS: final approval. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the National Natural Science Foundation of China (81902684) and Shanghai three year action plan for Traditional Chinese Medicine [ZY(2018-2020)-CCXX-2005-04]. The funders had no role in the design of the study and collection, analysis, interpretation of data or writing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.746763/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin (2021) 71:7–33. doi: 10.3322/caac.21654
2. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, et al. Triple-Negative Breast Cancer: Clinical Features and Patterns of Recurrence. Clin Cancer Res (2007) 13:4429–34. doi: 10.1158/1078-0432.CCR-06-3045
3. Venkitaraman R. Triple-Negative/Basal-Like Breast Cancer: Clinical, Pathologic and Molecular Features. Expert Rev Anticancer Ther (2010) 10:199–207. doi: 10.1586/era.09.189
4. Carey L, Winer E, Viale G, Cameron D, Gianni L. Triple-Negative Breast Cancer: Disease Entity or Title of Convenience? Nat Rev Clin Oncol (2010) 7:683–92. doi: 10.1038/nrclinonc.2010.154
5. Liu Q, Xing P, Dong H, Zhao T, Jin F. Preoperative Assessment of Axillary Lymph Node Status in Breast Cancer Patients by Ultrasonography Combined with Mammography: A STROBE Compliant Article. Med (Baltimore) (2018) 97:e11441. doi: 10.1097/MD.0000000000011441
6. Welch HG, Prorok PC, O’Malley AJ, Kramer BS. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. N Engl J Med (2016) 375:1438–47. doi: 10.1056/NEJMoa1600249
7. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of Screening and Adjuvant Therapy on Mortality From Breast Cancer. N Engl J Med (2005) 353:1784–92. doi: 10.1056/NEJMoa050518
8. Otto SJ, Fracheboud J, Looman CW, Boer R, Hendriks JH, et al. Initiation of Population-Based Mammography Screening in Dutch Municipalities and Effect on Breast-Cancer Mortality: A Systematic Review. Lancet (2003) 361:1411–7. doi: 10.1016/S0140-6736(03)13132-7
9. Gradishar WJ, Anderson BO, Abraham J, Aft R, Agnese D, Allison KH, et al. Breast Cancer, Version 3.2020, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw (2020) 18:452–78. doi: 10.6004/jnccn.2020.0016
22. Rakha EA, El-Sayed ME, Green AR, Lee AH, Robertson JF, Ellis IO. A Novel Prognostic Two-Gene Signature for Triple-Negative Breast Cancer. J Clin Oncol (2014) 32:3513–9. doi: 10.1200/JCO.2013.54.9931

23. Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, et al. 70-Immune-Related-Gene Signature in Patients With Early Triple-Negative Breast Cancer. J Natl Cancer Inst (2009) 101:283–9. doi: 10.1093/jnci/djp498

24. Bonastre J, Marguet S, Lueza B, Michiels S, Delalosge S, Saghatzian M. Cost Effectiveness of Molecular Profiling for Adjuvant Decision Making in Patients With Node-Negative Breast Cancer. J Clin Oncol (2014) 32:3513–9. doi: 10.1200/JCO.2013.54.9931

25. Vaz-Luis I, Ottesen RA, Hughes ME, Mamet R, Burstink HJ, Edge SB, et al. Outcomes by Tumor Subtype and Treatment Pattern in Women With Small, Node-Negative Breast Cancer: A Multinational Case-Control Study. J Clin Oncol (2014) 32:2142–50. doi: 10.1200/JCO.2013.53.1608

26. Goldhirsh A, Glick JH, Gelber RD, Coates AS, Senn HJ. Meeting Highlights: International Consensus Panel on the Treatment of Primary Breast Cancer. Seventh International Conference on Adjuvant Therapy of Primary Breast Cancer. J Clin Oncol (2001) 19:3817–27. doi: 10.1200/JCO.2001.19.18.3817

27. Iasonos A, Schrag D, Raj GV, Panagase KS. How to Build and Interpret a Nomogram for Cancer Prognosis. J Clin Oncol (2008) 26:1364–70. doi: 10.1200/JCO.2007.12.9791

28. van ’t Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer. Nature (2002) 415:530–6. doi: 10.1038/415530a

29. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A Multigene Assay to Predict Recurrence of Tumoxifen-Treated, Node-Negative Breast Cancer. N Engl J Med (2004) 351:2817–26. doi: 10.1056/NEJMoa041588

30. Fâpîts M, Rudas M, Jakesz R, Dubsb P, Fitzal F, Singer CF, et al. A New Molecular Predictor of Distant Recurrence in ER-Positive, HER2-Negative Breast Cancer Adds Independent Information to Conventional Clinical Risk Factors. Clin Cancer Res (2011) 17:6012–20. doi: 10.1158/1078-0432.CCR-11-0926

31. Jerevall PL, Ma XJ, Li H, Salunga R, Kesty NC, Erlander MG, et al. Prognostic Utility of HOXB13:IL17BR and Molecular Grade Index in Early-Stage Breast Cancer Patients From the Stockholm Trial. Br J Cancer (2011) 104:1762–9. doi: 10.1038/bjc.2011.145

32. Park JS, Mullins M, Cheung MC, Leung S, Voduc D, Vickery T, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. J Clin Oncol (2009) 27:1160–7. doi: 10.1200/JCO.2008.18.1370

33. Li CI, Zhang Y, Cieslik M, Wu YM, Xiao L, Cobain E, et al. Cancer Cell Intrinsic and Immunologic Phenotypes Determine Clinical Outcomes in Basal-Like Breast Cancer. Clin Cancer Res (2021) 27:3079–93. doi: 10.1158/1078-0432.CCR-20-3890

34. Qin Y, Deng J, Zhang L, Yuan F, Yang H, Li Q. Tumor Microenvironment Characterization in Triple-Negative Breast Cancer Identifies Prognostic Gene Signature. Aging (Albany NY) (2021) 13:5485–505. doi: 10.18632/aging.202477

35. Park YH, Jung HH, Do IG, Cho EY, Sohn I, Jung SH, et al. Triple-Negative Breast Cancers Who Receive Adjuvant Chemotherapy Following Surgery. Int J Cancer (2015) 136:1976–84. doi: 10.1002/ijc.29233

36. Kim JY, Jung HH, Sohn I, Woo SY, Cho H, Cho EY, et al. Prognostication of a 13-Immune-Related-Gene Signature in Patients With Early Triple-Negative Breast Cancer. Breast Cancer Res Treat (2020) 184:325–34. doi: 10.1007/s10549-020-05874-1

37. Alsaeima AL, Ball G, Toss MS, Rafat S, Aleskandarany M, Joseph C, et al. A Novel Prognostic Two-Gene Signature for Triple Negative Breast Cancer. Mod Pathol (2020) 33:2208–20. doi: 10.1038/s41379-020-0563-7

38. Hong HC, Chuang CH, Huang WC, Weng SL, Chen CH, Chang KH, et al. A Panel of Eight microRNAs Is a Good Predictive Parameter for Triple-Negative Breast Cancer Relapse. Theranostics (2020) 10:8771–89. doi: 10.7150/thno.46142

39. Carter CL, Allen C, Henson DE. Relation of Tumor Size, Lymph Node Status, and Survival in 24,740 Breast Cancer Cases. Cancer (1989) 63:1817–7. doi: 10.1002/1097-0142(19890701)63:1<181::aid-cncr2820630129>3.0.co;2-h

40. Rakhka EA, El-Sayed ME, Green AR, Robertson JF, Ellis IO, Prognostic Markers in Triple-Negative Breast Cancer (2007) 109:25–32. doi: 10.1002/1054-90-0204.1

41. Jiang YZ, Ma D, Suo C, Shi J, Xue M, Hu X, et al. Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. Cancer Cell (2019) 35:428–40.e5. doi: 10.1016/j.ccell.2019.02.001

42. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lerner J, et al. PGC-1alpha-Responsive Genes Involved in Oxidative Phosphorylation Are Coordinately Downregulated in Human Diabetics. Nat Genet (2003) 34:267–73. doi: 10.1038/ng1180

43. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Buetl BR, Gillette MA, et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. Proc Natl Acad Sci U S A (2005) 102:15545–50. doi: 10.1073/pnas.0506580102

44. Goldhirsh A, Glick JH, Gelber RD, Coates AS, Senn HJ. Meeting Highlights: International Consensus Panel on the Treatment of Primary Breast Cancer. Seventh International Conference on Adjuvant Therapy of Primary Breast Cancer. J Clin Oncol (2001) 19:3817–27. doi: 10.1200/JCO.2001.19.18.3817

45. Iasonos A, Schrag D, Raj GV, Panagase KS. How to Build and Interpret a Nomogram for Cancer Prognosis. J Clin Oncol (2008) 26:1364–70. doi: 10.1200/JCO.2007.12.9791

46. van ’t Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer. Nature (2002) 415:530–6. doi: 10.1038/415530a

47. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. N Engl J Med (2004) 351:2817–26. doi: 10.1056/NEJMoa041588

48. Fâpîts M, Rudas M, Jakesz R, Dubsb P, Fitzal F, Singer CF, et al. A New Molecular Predictor of Distant Recurrence in ER-Positive, HER2-Negative Breast Cancer Adds Independent Information to Conventional Clinical Risk Factors. Clin Cancer Res (2011) 17:6012–20. doi: 10.1158/1078-0432.CCR-11-0926

49. Jerevall PL, Ma XJ, Li H, Salunga R, Kesty NC, Erlander MG, et al. Prognostic Utility of HOXB13:IL17BR and Molecular Grade Index in Early-Stage Breast Cancer Patients From the Stockholm Trial. Br J Cancer (2011) 104:1762–9. doi: 10.1038/bjc.2011.145

50. Park JS, Mullins M, Cheung MC, Leung S, Voduc D, Vickery T, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. J Clin Oncol (2009) 27:1160–7. doi: 10.1200/JCO.2008.18.1370
47. Mou MA, Keya NA, Islam M, Hossain MJ, Al Habib MS, Alam R, et al. Validation of CSN1S1 Transcriptional Expression, Promoter Methylation, and Prognostic Power in Breast Cancer Using Independent Datasets. *Biochem Biophys Rep* (2020) 24:100867. doi: 10.1016/j.bbrep.2020.100867

48. Tan MH, De S, Bebek G, Orloff MS, Wesolowski R, Downs-Kelly E, et al. Specific Kinesin Expression Profiles Associated With Taxane Resistance in Basal-Like Breast Cancer. *Breast Cancer Res Treat* (2012) 131:849–58. doi: 10.1007/s10549-011-1500-8

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

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

Copyright © 2021 Peng, Lin, Jing, Su, Jin, Di and Shao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.