The early-stage triple-negative breast cancer landscape derives a novel prognostic signature and therapeutic target

Yun-Song Yang1,2,4 · Yi-Xing Ren1,2 · Cheng-Lin Liu1,2 · Shuang Hao1,2 · Xiao-En Xu1,2 · Xi Jin1,2 · Yi-Zhou Jiang1,2 · Zhi-Ming Shao1,2,3

Received: 3 November 2021 / Accepted: 24 January 2022 / Published online: 25 March 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Purpose Triple-negative breast cancer (TNBC) is a highly heterogeneous disease. Patients with early-stage TNBCs have distinct likelihood of distant recurrence. This study aimed to develop a prognostic signature of early-stage TNBC patients to improve risk stratification.

Methods Using RNA-sequencing data, we analyzed 189 pathologically confirmed pT1-2N0M0 TNBC patients and identified 21 mRNAs that were highly expressed in tumor and related to relapse-free survival. All-subset regression program was used for constructing a 7-mRNA signature in the training set (n = 159); the accuracy and prognostic value were then validated using an independent validation set (n = 158).

Results Here, we profiled the transcriptome data from 189 early-stage TNBC patients along with 50 paired normal tissues. Early-stage TNBCs mainly consisted of basal-like immune-suppressed subtype and had higher homologous recombination deficiency scores. We developed a prognostic signature including seven mRNAs (ACAN, KRT5, TMEM101, LCA5, RPP40, LAGE3, CDKL2). In both the training (n = 159) and validation set (n = 158), this signature could identify patients with relatively high recurrence risks and served as an independent prognostic factor. Time-dependent receiver operating curve showed that the signature had better prognostic value than traditional clinicopathological features in both sets. Functionally, we showed that TMEM101 promoted cell proliferation and migration in vitro, which represented a potential therapeutic target.

Conclusions Our 7-mRNA signature could accurately predict recurrence risks of early-stage TNBCs. This model may facilitate personalized therapy decision-making for early-stage TNBCs individuals.

Keywords Triple-negative breast cancer · Transcriptome analysis · Gene expression signature · Recurrence-free survival · TMEM101

Abbreviations

AIC Akaike information criterion
AUC Area under curve
BLIS Basal-like and immune-suppressed

ER Estrogen receptor
HER2 Human epidermal growth factor receptor 2
HR Hormone receptor
HRD Homologous recombination deficiency
IM Immunomodulatory
LAR Luminal androgen receptor

Yun-Song Yang, Yi-Xing Ren and Cheng-Lin Liu have contributed equally to this work.

Xi Jin
15111230037@fudan.edu.cn

Yi-Zhou Jiang
yizhoujiang@fudan.edu.cn

Zhi-Ming Shao
zhimingshao@yahoo.com

1 Key Laboratory of Breast Cancer in Shanghai, Department of Breast Surgery, Fudan University Shanghai Cancer Center, 270 Dong-An Road, Shanghai 200032, People’s Republic of China

2 Department of Oncology, Shanghai Medical College, Fudan University, 270 Dong-An Road, Shanghai 200032, People’s Republic of China

3 Institutes of Biomedical Sciences, Fudan University, 131 Dong-An Road, Shanghai 200032, People’s Republic of China

4 Human Phenome Institute, Fudan University, 825 Zhangheng Road, Shanghai 201203, People’s Republic of China
LASSO Least absolute shrinkage and selection operator
MES Mesenchymal-like
RFS Relapse-free survival
ROC Receiver operating characteristic
PR Progestogen receptor
RT-qPCR Real-time quantitative-polymerase chain reaction
siRNA Small interfering RNAs
TNBC Triple-negative breast cancer

Introduction

Triple-negative breast cancer is the most aggressive type of breast cancer, characterized by low/no expression of estrogen receptor (ER), progestogen receptor (PR) and human epidermal growth factor receptor 2 (HER2) [1]. Recent advances in breast cancer screening have significantly improved the diagnosis of early-stage breast cancer [2]. According to breast cancer registry, stage I breast cancer accounted for 25.8% of breast cancer patients in China [3]. However, outcomes for these patients may vary by tumor subtype and biological behavior [4]. For early-stage breast cancer, TNBC patients had a higher risk of mortality compared with other molecular subtypes [4–6]. Currently, adjuvant chemotherapy remains the only systematic treatment option for early-stage TNBCs. Chemotherapy is often considered even in pT1a-bN0M0 patients, but few benefits of survival could be achieved from chemotherapy [7]. In addition, due to the intratumor heterogeneity of TNBCs, similar chemotherapy may evoke different treatment response. Therefore, there is an urgent need to develop prognostic signatures to stratify early-stage TNBCs according to prognosis.

Previous studies have tried to use mRNA, micro-RNA, long noncoding RNA, or proteome profile to build prognostic signatures to assess the risk of recurrence and magnitude of chemotherapy benefit of node-negative breast cancer [8–10]. Nonetheless, these studies had two main drawbacks: first, most of the studies used Cox proportional hazards regression analysis to build prognostic models, which is not suitable for high-dimensional data (datasets with thousands of genes) [11]; Second, these studies enrolled both node-negative and -positive TNBC patients. With a relatively small sample size of node-negative TNBC patients, it is challenging to generate a precise model for node-negative TNBCs.

In this study, we developed a 7-mRNA signature by LASSO (Least Absolute Shrinkage and Selection Operator) Cox regression model following all-subsets regression program to predict relapse-free survival (RFS) in a large cohort of early-stage TNBCs. We hypothesized that this signature could identify the high-risk subpopulation among early-stage TNBC and provide instructions for individual treatments.

Materials and methods

Patients and samples

Previously we have constructed a multiomics TNBC cohort, containing 386 consecutive TNBC patients who underwent surgeries at the Department of Breast Surgery, Fudan University, Shanghai Cancer Center (FUSCC), from 1 January 2007 to 31 December 2014 [12]. To identify candidate mRNAs for prognostic signature development, we reviewed the 386 TNBC patients with high-throughput transcriptome data (GSE76250 and SRP157974) and 189 patients (50 patients had paired adjacent normal tissues) meet the following inclusion criteria were included as developing set in this prospective observational study, which was initiated in 2017 (Fig. S1). The inclusion criteria were the following: (1) Female patients with unilateral breast tumors and histologically confirmed invasive ductal carcinoma (IDC) and triple-negative subtype (ER-negative, PR-negative, and HER2-negative); (2) Pathology confirmed stage pT2 or below (tumor size ≤5 cm); (3) Pathology confirmed pN0 (no cancer cells in the lymph nodes). Patients with isolated tumor cells (pN0i+) and micro metastases (pN1mi) in lymph nodes were excluded. (4) Patients who did not receive any type of treatment prior to surgery; (5) Follow-up time > 1 year. The follow-up of patients in developing set was completed on 30 June, 2017, and the median follow-up length was 47.6 months (interquartile range 38.5–59.3 months). To construct the training and validation set, a total of 317 consecutive early-stage TNBC patients treated in FUSCC between January 1, 2008 and December 31, 2015 were enrolled according to the inclusion criteria described above. A total of 330 frozen samples (317 tumor tissues and 13 paired normal tissues) from 317 TNBC patients preserved by the Department of Pathology in FUSCC were collected. There were 79 patients who were overlapped between the developing set and the above 317 patients.

The ER, PR and HER2 status were measured by immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) were confirmed by two experienced pathologists (Ruo-Hong Shui and Wen-Tao Yang), and the cut-off points were set as the guideline from the American Society of Clinical Oncology and the College of American Pathologists [13]. The pathologic stage was assigned based on the 8th edition of the AJCC anatomic staging system [14].

To train and validate a multi-gene signature, we randomly assigned the 317 patients into a training set (n = 159) and validation set (n = 158). The follow-up was completed on 31 December, 2017 and the median follow-up length...
was 39.3 months (interquartile range 25.6–55.2 months) in training set and 38.5 months (interquartile range 26.8–52.7 months) in the validation set. RFS events were defined as the first recurrence of invasive disease at a local, regional, or distant site; contralateral breast cancer; and death from any cause. Patients without events were censored at the last follow-up.

The independent ethics committee/institutional review board of FUSCC (Shanghai Cancer Center Ethics Committee) approved our study and written informed consent from patients in our study was obtained before enrollment.

RNA purification, reverse transcription and RT-qPCR

The RNeasy Plus Mini Kit (QIAGEN) was used to isolate total RNA from frozen tissues and cells and reverse transcribed to cDNA using GoScript Reverse Transcription Kit (Promega) according to the manufacturer’s protocol. Real-time quantitative-polymerase chain reaction (RT-qPCR) was performed using SYBR Premix Ex Taq (TAKARA) and amplified on QuantStudio 7 Flex System (Applied Biosystems). The primers used for RT-qPCR were listed in Table S1. The relative expression value of each mRNAs was calculated as CTmRNA – (CTU6 + CTGAPDH)/2 (CT, cycle threshold).

Identification of candidate mRNAs

The detailed filtration process is illustrated in Fig. S1. The mRNAs that were differentially expressed between tumor and paired normal tissues and related to RFS were selected for further research using transcriptome data in the developing set (N = 189 containing 50 patients with paired adjacent normal tissues). In detail, the paired differentially expressed mRNA analysis was processed using the Bioconductor package ‘Limma’ (Linear models for microarray analysis); significant differentially expressed genes were defined as false discovery rate < 0.05 and fold change > 1.5 or < 0.33 for up or downregulation, respectively. The univariate Cox proportional hazards regression model was used to select mRNAs that were significantly correlated with RFS. Next, the LASSO Cox regression model analysis was processed to select the most useful prognostic mRNAs as candidate mRNAs with 200 bootstrap replicates [15].

Development and validation of the prognostic signature

The expression of the candidate mRNAs was measured using RT-qPCR in the training set (N = 159) and internal validation set (N = 158). In order to simplify the prognostic signatures, we used R package ‘glmulti’ to perform automated screening for the best combination of mRNAs based on Akaike information criterion (AIC) value and calculate the coefficients of each mRNAs in the training set [16]. The coefficients of each mRNAs were used to construct a risk score formula. The accuracy of this signature was assessed by time-dependent receiver operating characteristic (ROC) analysis, and the best cut-off score was selected according to Youden Index [17]. To validate this signature, the risk score of each patient in the validation set was calculated using the formula generated in training set. In addition, we used transcriptome data from METABRIC database as a preliminary external validation. As the gene expression level in METABRIC database was profiled using RNA-sequencing (RNA-seq), we could not use the same coefficients derived from training set to calculate the risk score of patients from METABRIC database. Therefore we used Cox regression model to recalculate the coefficients of 7 mRNAs and the cut-off of risk score was selected according to Youden Index.

Gene set enrichment analysis (GSEA)

GSEA (version 2.2.0) was performed using the hallmark gene set and a preranked differential expression gene list [18]. The differential gene analysis between early-stage and late-stage was conducted using DEseq2 package.

Cell cultures and siRNA transfection

Human breast cancer cell lines MDA-MB-231, BT-549 and human embryonic kidney cells HEK293T were obtained from American Type Culture Collection. These human cell lines have been authenticated using STR profiling and monitoring mycoplasma contamination. The Lipofectamine RNAiMAX Transfection Reagent (Thermo Fisher Scientific) was used to transfect small interfering RNAs (siRNAs). The target sequences are listed in Table S2. The transfection procedure was performed according to the manufacturer’s instructions.

Proliferation assay

For the proliferation assay, cells were cultured in 6-well plates for 24 h and transfected with siRNAs. After 24 h, 1000 cells per well were seeded in 96-well plates. The cell confluence was monitored using an IncuCyte Live-Cell Analysis Systems (IncuCyte ZOOM System, ESSEN Bioscience) and imaged every 12 h. The cell confluence was measured by IncuCyte ZOOM software, and the assays were performed in triplicate.

Migration assay

For the migration assay, 5 × 10^4 cells in serum-free medium were seeded on the top of the transwell chamber (pore size,
8μm, BD Biosciences) and 600 μl medium containing 10% FBS was added to the bottom chamber. After incubation for 8–10 h, cells that migrated to the opposite side of the membrane were fixed in 4% paraformaldehyde for 30 min and stained with crystal violet for additional 30 min. ImageJ was used to quantify the number of cells migrated to the opposite side per field.

**Western blot**

Cells were lysed in lysis buffer (50 mM tris [pH 8.1], 1 mM EDTA, 1% SDS, 1 mM fresh dithiothreitol, sodium fluoride, and leupeptin). The cell lysates were boiled in SDS-PAGE loading buffer for 15 min. The proteins were separated by SDS-PAGE and transferred to polyvinylidene difluoride membranes (Millipore). The primary antibody anti-TMEM101 (Abscitech, AB-35251-2, 1:2000) and secondary antibody HRP-conjugated goat anti-rabbit antibody (Jackson Immuno-Research; 1:5000) were blotted. Signals were detected with an enhanced chemiluminescence substrate (Pierce Biotechnology), and images were collected by Molecular Imager ChemiDoc XRS + (Bio-Rad) with Image Lab Software (Bio-Rad).

**Immunohistochemistry (IHC)**

A total of 56 primary TNBC specimens were obtained from the Department of Pathology, Fudan University Shanghai Cancer Center. IHC staining was performed as previously described [19]. The anti-TMEM101 (Abscitech, AB-35251-2, 1:100) primary antibody was used. Slides were evaluated using light microscopy and scored with a microscope by pathologists with the H-score, which evaluated staining intensity (0 to 3) and the percentage of positively stained cells (0 to 1), with a final score ranging from 0 to 3.

**Statistical analysis**

The t-test was used to compare the differences between the two groups for continuous variables and χ² test for categorical variables. We used the Kaplan–Meier method to perform univariate survival analysis and log-rank test to calculate the hazard ratio and compare two survival curves. For multivariate survival analysis, the multivariate Cox proportional hazards regression model was used to test whether the signature was an independent prognostic factor associated with RFS. All statistical analyses were performed with R software version 3.5.3. or GraphPad Prism 8. A P value < 0.05 was considered to be statistically significant.

**Results**

**Transcriptome characteristic of early-stage TNBCs**

We reviewed 17,877 breast cancer patients diagnosed in our center from January 1, 2008 and December 31, 2015 and 663 of them were at pT1-2N0M0 TNBC which accounted for 28.8% of all TNBC patients. Survival analysis showed that pT1-2N0M0 TNBCs (early-stage) showed great survival benefits compared to other TNM stage (late-stage) (Fig. 1A). According to our previous research, TNBC could be classified into 4 mRNA-based TNBC intrinsic subtype, including basal-like and immune-suppressed (BLIS), luminal androgen receptor (LAR), immunomodulatory (IM) and mesenchymal stem-like (MSL) subtype [20]. We found that more early-stage TNBCs were classified as BLIS (Fig. 1B, Fig. 1C, Fig. 1D).

![Fig. 1](image-url) The transcriptome features of early-stage TNBCs. A. Kaplan–Meier curves of relapse-free survival in early-stage and late-stage TNBCs. B. Distribution of TNBC molecular subtypes (B) and HRD scores (C) between early-stage and late-stage TNBCs. D. Representative gene set enrichment analysis plot showing gene set unregulated in the early-stage versus the late-stage TNBCs. RFS relapse-free survival; BLIS basal-like and immune-suppressed; IM immunomodulatory; LAR luminal androgen receptor; MSL mesenchymal-like; HRD homologous recombination deficiency; EMT epithelial-mesenchymal transition.
Patients with higher HRD score had a higher recurrence rate and worse survival (Fig. 2A, B). Yet, there was no significant difference in lymphovascular invasion in the training set. Similar results were observed in the validation set. Patients with high-risk had a higher risk of recurrence (Table 2, HR 5.87; 95% CI 2.07–16.66; \( P = 0.001 \)), and lymphovascular invasion was not a significant prognostic factor in the validation set.

Subgroup analysis was proposed to verify the performance of this signature in different subpopulations. The forest plot illustrated that the 7-mRNA signature had good prognostic value in predicting RFS independent of tumor size, Ki-67, tumor grade, and lymphovascular invasion (Fig. S3C).

**Comparison with other prognostic factors**

To compare the performance of the 7-mRNA signature and other traditional clinicopathological features, we conducted time-dependent ROC analyses (the time was set as 3 years) and calculated the area under curve (AUC) on both the signature and traditional clinicopathological features. The results showed that the 7-mRNA signature had better prognostic value in predicting RFS (Fig. 3A). In line with the current statistical consensus, the predictor model with AUC over 0.7 could be considered as adequate for clinical implications [22, 23]. Traditional clinicopathological features including tumor size, tumor grade, Ki-67 status and lymphovascular invasion had poor prognostic value in early-stage TNBCs with AUC lower than 0.7. This finding was further validated in the validation set, as the 7-mRNA signature had higher AUC; however, when compared to tumor size and lymphovascular invasion, the difference was not statistically significant (Fig. 3B). Therefore, the 7-mRNA signature could be a better prognostic model than clinicopathological prognostic features.

**The biological function of TMEM101 in the 7-mRNA signature**

To further study the biological function of 7 mRNAs in the prognostic signature, we explored their role in cell migration. Knockdown of each mRNAs was conducted...
by a pool of three target-specific siRNAs. The transwell assay revealed that the downregulation of TMEM101 or LAGE3 restrained the migration of the MDA-MB-231 cell line (Fig. S4A). Furthermore, Kaplan–Meier analysis demonstrated that high expression of TMEM101 was significantly associated with shorter RFS (Fig. 4A; \( P = 0.019 \)), while LAGE3 was not related to patients survival (Fig. S4B, C). Therefore, TMEM101 could be a good candidate target or prognostic biomarker of TNBC.
Table 1  Clinicopathological features of three sets of pT1-2N0M0 TNBC patients stratified according to the 7-mRNA signature

| Clinical features | Developing set | Training set | Validation set |
|-------------------|----------------|--------------|----------------|
|                   | N  | Low risk (%) | High risk (%) | P valuea | N  | Low risk (%) | High risk (%) | P valuea |
| Age (years)       |    |              |              |          |    |              |              |          |
| < 55              | 189 | 159 (88.7)   | 18 (11.3)    | 0.469   | 158 | 133 (84.2)  | 25 (15.8)    | 0.873     |
| ≥ 55              | 115 | 61 (56.7)    | 56 (43.3)    | 0.119   | 75  | 14 (48.1)   | 11 (51.9)    | 0.548     |
| Menopause         |    |              |              |          |    |              |              |          |
| Yes               | 124 | 92 (65.6)    | 28 (34.4)    | 0.393   | 101 | 92 (65.2)   | 9 (34.8)     | 0.205     |
| No                | 65  | 56 (85.8)    | 9 (14.2)     | 0.119   | 58  | 49 (85.6)   | 9 (14.4)     | 0.205     |
| Tumor size (cm)   |    |              |              |          |    |              |              |          |
| < 2               | 108 | 72 (67.2)    | 36 (32.8)    | 0.035   | 73  | 54 (54.9)   | 19 (45.1)    | 0.088     |
| > 2               | 100 | 95 (95.0)    | 5 (5.0)      | 0.284   | 100 | 95 (97.0)   | 5 (3.0)      | 0.284     |
| Tumor grade       |    |              |              |          |    |              |              |          |
| II                | 52  | 52 (100.0)   | 0 (0.0)      | 1.000   | 58  | 52 (96.6)   | 4 (3.4)      | 0.697     |
| II–III            | 137 | 107 (79.2)   | 30 (20.8)    | 0.861   | 100 | 83 (83.0)   | 17 (17.0)    | 0.861     |
| Ki-67             |    |              |              |          |    |              |              |          |
| ≤ 50%             | 103 | 72 (70.2)    | 31 (29.8)    | 0.058   | 78  | 66 (85.5)   | 12 (14.5)    | 0.128     |
| > 50%             | 86  | 87 (99.1)    | 1 (0.9)      | 0.001   | 80  | 67 (84.0)   | 13 (16.0)    | 0.001     |
| Lymphovascular invasion | NA | 18 (100.0) | 0 (0.0) | 1.000 | 28 | 20 (92.0) | 4 (8.0) | 0.084 |
| Yes               |    |              |              |          |    |              |              |          |
| No                | 92  | 76 (82.4)    | 16 (17.6)    | 0.009   | 51  | 39 (76.5)   | 12 (23.5)    | 0.009     |
| Chemotherapy      |    |              |              |          |    |              |              |          |
| Yes               | 177 | 145 (82.3)   | 32 (17.7)    | 0.545   | 129 | 107 (85.0) | 22 (15.0)    | 0.212     |
| No                | 4   | 3 (75.0)     | 1 (25.0)     | 0.001   | 7   | 5 (71.4)   | 2 (28.6)     | 0.001     |
| UNK               | 8   | 8 (100.0)    | 0 (0.0)      | 0.001   | 22  | 21 (95.5)  | 1 (4.5)      | 0.001     |
| Radiotherapy      |    |              |              |          |    |              |              |          |
| Yes               | 19  | 13 (68.4)    | 6 (31.6)     | 0.441   | 4   | 4 (100.0)  | 0 (0.0)      | 0.679     |
| No                | 169 | 136 (80.0)   | 33 (20.0)    | 0.001   | 142 | 119 (83.9) | 23 (16.1)    | 0.001     |
| UNK               | 1   | 10 (100.0)   | 0 (0.0)      | 0.001   | 12  | 10 (83.3)  | 2 (16.7)     | 0.001     |
| Follow-up time (Median, Months) | 47.6 | 41 | 32.8 | 0.009 | 47.6 | 41 | 32.8 | 0.009 |

Table 2  Multivariate Cox regression analysis of 7-mRNA signature and traditional clinical features associated with RFS

| Clinical features | Training set | Validation set |
|-------------------|--------------|----------------|
|                   | HRa (95% CI) | P value | HRa (95% CI) | P value |
| 7-mRNA signature (low risk as reference) | 15.68 (5.50–44.70) | <0.001 | 5.87 (2.07–16.66) | 0.001 |
| Tumor size (tumor size < 2 cm as reference) | 1.97 (0.66–5.86) | 0.221 | 1.57 (0.51–4.80) | 0.434 |
| Tumor grade II–III (grade II as reference) | 1.01 (0.32–3.21) | 0.982 | 0.56 (0.19–1.62) | 0.284 |
| Ki-67 (Ki-67 < 50% as reference) | 0.58 (0.19–1.80) | 0.347 | 2.43 (0.78–7.54) | 0.125 |
| Lymphovascular invasion (no as reference) | 4.57 (1.45–14.39) | 0.009 | 2.53 (0.83–7.71) | 0.104 |

*UNK unknown; RFS relapse-free survival

*P value was calculated using χ² test

Table 2  Multivariate Cox regression analysis of 7-mRNA signature and traditional clinical features associated with RFS

| Clinical features | Training set | Validation set |
|-------------------|--------------|----------------|
|                   | HRa (95% CI) | P value | HRa (95% CI) | P value |
| 7-mRNA signature (low risk as reference) | 15.68 (5.50–44.70) | <0.001 | 5.87 (2.07–16.66) | 0.001 |
| Tumor size (tumor size < 2 cm as reference) | 1.97 (0.66–5.86) | 0.221 | 1.57 (0.51–4.80) | 0.434 |
| Tumor grade II–III (grade II as reference) | 1.01 (0.32–3.21) | 0.982 | 0.56 (0.19–1.62) | 0.284 |
| Ki-67 (Ki-67 < 50% as reference) | 0.58 (0.19–1.80) | 0.347 | 2.43 (0.78–7.54) | 0.125 |
| Lymphovascular invasion (no as reference) | 4.57 (1.45–14.39) | 0.009 | 2.53 (0.83–7.71) | 0.104 |

*RFS relapse-free survival; HR hazard ratio

*Adjusted by Cox proportional hazards models including age, tumor size, tumor grade, Ki-67, lymphovascular invasion, chemotherapy, radiotherapy, and the 7-mRNA signature
The exact function of TMEM101 in human still remains unknown. Some studies have suggested an association between the NF-κB pathway and TMEM101 [24, 25]. In this study, we found higher expression of TMEM101 in tumor samples than paired normal samples (Fig. 4B; \(P = 0.003\)) and TMEM101 showed great mRNA-protein correlations in TNBC tumor samples (Fig. S4D). We also designed three siRNAs targeting TMEM101, and two siRNAs with the highest interference efficiently inhibited the migration ability of MDA-MB-231 and BT-549 cell lines (Fig. 4C, D, F). Similarly, the downregulation of TMEM101 significantly impaired the proliferation of tumor cells in vitro (Fig. 4E, G). In addition, TMEM101 depletion could also decrease the mRNA level of NF-κB downstream genes [26] (Fig. S4E). Collectively, these data suggest that knocking down TMEM101 could restrain migration and proliferation of TNBC cell lines in vitro, thus revealing it as a potential therapeutic target.

**Discussion**

In this study, we developed and validated a 7-mRNA signature based on transcriptome analysis to predict prognosis for node-negative TNBC. The prognostic value of this signature was more efficient than traditional clinicopathological features, including tumor size, tumor grade, Ki-67, and LVI in training (A) and validation set (B). \(P\) value was generated by bootstrap at 200 replicates. AUC area under curve; LVI lymphovascular invasion.

The prognosis of node-negative TNBC patients is rarely associated with clinical factors, including age, histological grade, and tumor size [27]. Previous retrospective research showed that lymphovascular invasion was the only traditional parameter associated with the survival of pT1a-b node-negative TNBC [28]. Our results also revealed that lymphovascular invasion could reflect worse survival of
node-negative TNBCs, but the 7-mRNA signature had a better prognostic value of RFS.

Several studies constructed prognostic signatures to stratify node-negative TNBC into different risk groups, to provide de-escalating or escalating treatments for early-stage TNBCs. Liu and colleagues constructed an 11-protein signature from 126 node-negative and adjuvant chemotherapy-naive TNBCs [8]. According to this signature, 60% early-stage TNBC were with relatively low recurrence risks, while over 80% of patients in our cohort were identified to have low recurrence risks (Table 1). However, it is difficult to compare our results with those of Liu and colleagues [8] because of the different inclusion criteria. Most of the patients enrolled in our study received adjuvant chemotherapy. Furthermore, Wang and colleagues constructed a 76-gene signature (16 genes for ER-negative; 60 genes for ER-positive) to predict distant metastasis for node-negative breast cancer patients in two cohorts.

Fig. 4 In vitro biologic function of TMEM101 incorporated in the 7-mRNA signature. A Kaplan–Meier curves of relapse-free survival of TMEM101 in node-negative TNBC cohort (n = 317). B The expression of TMEM101 between tumor and paired adjacent samples measured by RT-qPCR. C Western blot analysis of TMEM101 expression of MDA-MB-231 interfered with TMEM101 siRNAs. Representative images and the number of migrated MDA-MB-231 (D) and BT-549 cells (F) interfered with TMEM101 siRNAs. In vitro growth curves of MDA-MB-231 (E) and BT-549 (G) interfered with TMEM101 siRNAs for 5 days. Scale bar, 100 µm. NC negative control; TM TMEM101. *P < 0.05; **P < 0.01; ***P < 0.001
(contained 286 and 180 lymph node-negative breast cancer patients, respectively); however, the number of patients with ER-negative tumors (77 and 16 patients, respectively) in their study was too small to perform a subgroup analysis [10, 29]. Also, in this study, ER-negative patients included both HER2-positive and TNBC subtype, which had entirely different biological features [30]. Therefore, the first strength of our prognostic signature compared with previous studies is the large cohort of node-negative TNBC patients. Second, we used RT-qPCR to construct the signature in order to apply it in clinical practice. Profiling of multiple gene expression by RT-qPCR or microarray, such as PAM50 assay, 21-gene Oncotype DX and 70-gene MammaPrint signature, has been widely used in clinical practice [31]. High-throughput data are a time-consuming and costly assay which might not be easily used in clinical practice. Liu and colleagues constructed an mRNA-lncRNA signature by RT-qPCR to identify high-risk TNBC patients, and a clinical trial (NCT02641847) based on this signature was conducted in our center [32]. Therefore, from technical aspects, our signature could be widely used in our center.

The literature review showed that the seven mRNAs incorporated in our signature and their potential roles in cancer have already been reported. For example, CDKL2 was reported to promote breast cancer epithelial-mesenchymal transition through the up-regulation of ZEB1 expression in breast cancer [33]. A two-gene epigenetic signature revealed that higher methylation of CDKL2 was related to non-response to neoadjuvant chemotherapy in triple-negative breast cancer [34]. KRT5 encodes a member of the keratin gene family, which was identified as a basal marker in cancer research and related to worse survival, especially in node-negative breast cancer [30, 35–38]. In our study KRT5 served as a predictor of worse survival in early-stage TNBC, but the in vitro migration assay illustrated that KRT5 had no impact on the migration of TNBC cell lines. We speculated that it may regulate other biological functions in TNBC, which need to be further investigated. Aggrecan encoded by ACAN is an important component for cartilage structure, which may facilitate extracellular matrix (ECM) remodeling [39, 40]. Although the function of aggrecan in cancer is unknown, in breast cancer, ECM components are significantly dysregulated and associated with tumor progression [41]. TMEM101 has been rarely reported in cancer research. Moreover, using a cDNA library, Matsuda and colleagues revealed that 83 genes could activate the NF-κB pathway, including TMEM101 [24]. Recently, Khan and colleagues demonstrated that TMEM101 could increase the expression of NF-κB in zebrafish ovarian samples [25]. In our study, we first demonstrated that the expression of TMEM101 in TNBC was higher in tumors than in adjacent normal tissues, which suggests that TMEM101 could be used as a prognostic biomarker at early stage of TNBCs. Furthermore, interference expression of TMEM101 could inhibit the migration and proliferation abilities of TNBC cell lines.

This study has a few limitations. First, most of the patients enrolled in our study received adjuvant chemotherapy containing taxane and (or) anthracycline. Therefore, it was not accurate to identify the subpopulation of early-stage TNBCs in whom chemotherapy might be entirely omitted. However, according to our previous research, addition of taxane to an anthracycline-based adjuvant chemotherapy cannot further improve the RFS of pT1N0M0 TNBC patients. Therefore early-stage TNBC patients with low recurrence risk identified by this signature had the opportunity to omit taxane in the adjuvant chemotherapy [28]. Further randomized prospective studies should verify our hypotheses that patients with a high risk of relapse could benefit from adding chemotherapy such as capecitabine, which has been shown to bring additional survival benefits for TNBCs in clinical trials [42, 43], whereas patients in low-risk group could avoid taxane-contained chemotherapy. Second, as for using the RT-qPCR platform, it is inappropriate to use public high-throughput data sets as independent external validation. Therefore, to further validate the accuracy and stability of our signature, we have already started collecting patients’ samples and follow-up information at our center since 2018.

Conclusion

In summary, our 7-mRNA prognostic signature could identify early-stage TNBC patients at high recurrence risks. The 7-mRNA signature could be a powerful tool for oncologists to individualize treatment to patients’ early-stage TNBCs. Subsequent clinical trials are needed to further validate the signature.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10549-022-06537-z.

Acknowledgements This work was supported by the Program of Shanghai Academic/Technology Research Leader (20XD1421100), the Fok Ying-Tong Education Foundation for College Young Teachers (171034), the Shanghai Sailing Program (19YF1409000), the Innovation Team of Ministry of Education (IRT1223), and the Shanghai Key Laboratory of Breast Cancer (12DZ2260100). The funders had no role in the study design, data collection and analysis, or manuscript preparation.

Author contributions Conception and design: Y-SY, Y-XR, C-LL, XJ, X-EX, Y-ZJ, and Z-MS; Development of methodology: Y-SY, Y-XR, X-EX, and Y-ZJ; Acquisition of data: Y-SY, Y-XR, SH, and Y-ZJ; Analysis and interpretation of data: Y-SY, Y-XR, C-LL, XJ, and Y-ZJ; Writing, review, and/or revision of the manuscript: Y-SY, Y-XR, C-LL, SH, XJ, Y-ZJ, and Z-MS; Administrative, technical, or material
support: Y-SY, Y-XR, SH, XJ, X-EX, and Y-ZJ; Study supervision: Y-ZJ, Z-MS.

Data availability The datasets analyzed during the current study are available in the Gene Expression Omnibus (GSE76250, HTA2.0) and Sequence Read Archive (SRP157974, RNA-seq).

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethical approval The independent ethics committee/institutional review board of FUSCC (Shanghai Cancer Center Ethics Committee) approved our study.

Consent to participate Written informed consent from patients in our study was obtained before enrollment.

Consent for publication All authors read and approved the manuscript as submitted.

References

1. Foulkes WD, Smith IE, Reis-Filho JS (2010) Triple-negative breast cancer. N Engl J Med 363:1938–1948
2. Fan L, Strasser-Weippl K, Li J-J, St Louis J, Finkelstein DM, Yu K-D et al (2014) Breast cancer in China. Lancet Oncol 15:e279–289
3. Zeng H, Ran X, An L, Zheng R, Zhang S, Ji JS et al (2021) Disparities in stage at diagnosis for five common cancers in China: a multicentre, hospital-based, observational study. Lancet Public Health 6:e877–e887
4. Bao J, Donovan C, Ameris F, Zhang X, Giuliano AE, Chung A (2019) Outcomes in patients with small node-negative invasive breast cancer. Breast J 25:638–643
5. Parise CA, Caggiano V (2017) Risk of mortality of node-negative, ER/PR/HER2 breast cancer subtypes in T1, T2, and T3 tumors. Breast Cancer Res Treat 165:743–750
6. Amar S, McCullough AE, Tan W, Geiger XJ, Boughey JC, McNeil RB et al (2010) Prognosis and outcome of small (<=1 cm), node-negative breast cancer on the basis of hormonal and HER-2 status. Oncologist 15:1043–1049
7. de Nonneville A, Gonçalves A, Zemmour C, Cohen M, Classe JM, Reynal F et al (2017) Adjuvant chemotherapy in pT1ab node-negative triple-negative breast carcinomas: results of a national multi-institutional retrospective study. Eur J Cancer 84:34–43
8. Liu NQ, Stingl C, Look MP, Smid M, Braakman RBH, de Marchi T et al (2014) Comparative proteome analysis revealing an 11-protein signature for aggressive triple-negative breast cancer. J Natl Cancer Inst 106:djt576
9. Liu Y-R, Jiang Y-Z, Xu X-E, Yu K-D, Jin X, Hu X et al (2016) Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer. Breast Cancer Res 18:33
10. Wang Y, Klijn JGM, Zhang Y, Siewerts AM, Look MP, Yang F et al (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 365:9
11. Simon R, Altman DG (1994) Statistical aspects of prognostic factor studies in oncology. Br J Cancer 69:979–985
12. Xiao Y, Ma D, Zhao S, Suo C, Shi J, Xue M-Z et al (2019) Multiomics profiling reveals distinct microenvironment characterization and suggests immune escape mechanisms of triple-negative breast cancer. Clin Cancer Res 25:5002–5014
13. Hammond MEH, Hayes DF, Wolffe AC, Mangu PB, Temin S (2010) American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Oncol Pract 6:195–197
14. Amin MB, Edge SB (2017) AJCC cancer staging manual. Springer, Cham
15. Tibshirani R (1997) The lasso method for variable selection in the Cox model. Stat Med 16:385–395
16. Calcagno V, de Mazancourt C (2010) glmulti: an R package for easy automated model selection with (generalized) linear models. J Stat Softw. https://doi.org/10.18637/jss.v034.i12
17. Youden WJ (1950) Index for rating diagnostic tests. Cancer 3:32–35
18. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102:15545–15550
19. Jin X, Xu X-E, Jiang Y-Z, Liu Y-R, Sun W, Guo Y-J et al (2019) The endogenous retrovirus-derived long noncoding RNA TROJAN promotes triple-negative breast cancer progression via ZMYND8 degradation. Sci Adv 5:eaao9820
20. Jiang Y-Z, Ma D, Suo C, Shi J, Xue M, Hu X et al (2019) Genomic and transcriptomic landscape of triple-negative breast cancers: subtypes and treatment strategies. Cancer Cell 35:428–440.e5
21. Timms KM, Abkevich V, Hughes E, Neff C, Reid J, Morris B et al (2014) Association of BRCA1/2 defects with genomic scores predictive of DNA damage repair deficiency among breast cancer subtypes. Breast Cancer Res. https://doi.org/10.1186/s13058-014-0475-x
22. Sriureja W, Kyulo NL, Runyon BA, Hu K-Q (2005) MELD score is a better prognostic model than Child-Turcotte-Pugh score or Discriminant Function score in patients with alcoholic hepatitis. J Hepatol 42:700–706
23. Panczykowski DM, Puccio AM, Scruggs BJ, Bauer JS, Hricik AJ, Beers SR et al (2012) Prospective independent validation of IMPACT modeling as a prognostic tool in severe traumatic brain injury. J Neurotrauma 29:47–52
24. Matsuda A, Suzuki Y, Honda G, Muramatsu S, Matsuzaki O, Nagano Y et al (2003) Large-scale identification and characterization of human genes that activate NF-kappaB and MAPK signaling pathways. Oncogene 22:3307–3318
25. Khan ZA, Labala RK, Yamnamacha T, Devi SD, Mondal G, Sanjita Devi H et al (2018) Artificial light at night (ALAN), an alarm to ovarian physiology: a study of possible chronodisruption on zebrafish (Danio rerio). Sci Total Environ 628–629:1407–1421
26. Fusella F, Sechi L, Busso E, Krepelova A, Moiso E, Rocca S et al (2017) The IKK/NF-κB signaling pathway requires Morgana to drive breast cancer metastasis. Nat Commun 8:1636
27. Rakha EA, El-Sayed ME, Green AR, Lee AHS, Robertson JF, Ellis IO (2007) Prognostic markers in triple-negative breast cancer. Cancer Cell 35:428–440.e5
28. Panzyszewski DM, Puccio AM, Scruggs BJ, Bauer JS, Hricik AJ, Beers SR et al (2012) Prospective independent validation of IMPACT modeling as a prognostic tool in severe traumatic brain injury. J Neurotrauma 29:47–52
29. Furcell F, Sechi L, Busso E, Krepelova A, Moiso E, Rocca S et al (2017) The IKK/NF-κB signaling pathway requires Morgana to drive breast cancer metastasis. Nat Commun 8:1636
30. Khan ZA, Labala RK, Yamnamacha T, Devi SD, Mondal G, Sanjita Devi H et al (2018) Artificial light at night (ALAN), an alarm to ovarian physiology: a study of possible chronodisruption on zebrafish (Danio rerio). Sci Total Environ 628–629:1407–1421
31. Fusella F, Sechi L, Busso E, Krepelova A, Moiso E, Rocca S et al (2017) The IKK/NF-κB signaling pathway requires Morgana to drive breast cancer metastasis. Nat Commun 8:1636
32. Rakha EA, El-Sayed ME, Green AR, Lee AHS, Robertson JF, Ellis IO (2007) Prognostic markers in triple-negative breast cancer. Cancer Cell 35:428–440.e5
33. Panzyszewski DM, Puccio AM, Scruggs BJ, Bauer JS, Hricik AJ, Beers SR et al (2012) Prospective independent validation of IMPACT modeling as a prognostic tool in severe traumatic brain injury. J Neurotrauma 29:47–52
34. Matsuda A, Suzuki Y, Honda G, Muramatsu S, Matsuzaki O, Nagano Y et al (2003) Large-scale identification and characterization of human genes that activate NF-kappaB and MAPK signaling pathways. Oncogene 22:3307–3318
35. Khan ZA, Labala RK, Yamnamacha T, Devi SD, Mondal G, Sanjita Devi H et al (2018) Artificial light at night (ALAN), an alarm to ovarian physiology: a study of possible chronodisruption on zebrafish (Danio rerio). Sci Total Environ 628–629:1407–1421
36. Fusella F, Sechi L, Busso E, Krepelova A, Moiso E, Rocca S et al (2017) The IKK/NF-κB signaling pathway requires Morgana to drive breast cancer metastasis. Nat Commun 8:1636
30. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA et al (2000) Molecular portraits of human breast tumours. Nature 406:747–752

31. Piccart M, van’t Veer LJ, Poblet C, Lopes Cardozo JMN, Delaloge S, Pierga J-Y et al (2021) 70-gene signature as an aid for treatment decisions in early breast cancer: updated results of the phase 3 randomised MINDACT trial with an exploratory analysis by age. Lancet Oncol 22:476–488

32. Liu Y-R, Jiang Y-Z, Xu X-E, Hu X, Yu K-D, Shao Z-M (2016) Comprehensive transcriptome profiling reveals multigene signatures in triple-negative breast cancer. Clin Cancer Res 22:1653–1662

33. Li L, Liu C, Amato RJ, Chang JT, Du G, Li W (2014) CDKL2 promotes epithelial-mesenchymal transition and breast cancer progression. Oncotarget 5:10840–10853

34. Pineda B, Díaz-Lagares A, Pérez-Fidalgo JA, Burgués O, González-Barrallo I, Crujeiras AB et al (2019) A two-gene epigenetic signature for the prediction of response to neoadjuvant chemotherapy in triple-negative breast cancer patients. Clin Epigenetics 11:33

35. Zhao SG, Chen WS, Das R, Chang SL, Tomlins SA, Chou J et al (2019) Clinical and genomic implications of luminal and basal subtypes across carcinomas. Clin Cancer Res 25:2450–2457

36. Dadhania V, Zhang M, Zhang L, Bondaruk J, Majewski T, Sieker-Radtke A et al (2016) Meta-analysis of the luminal and basal subtypes of bladder cancer and the identification of signature immunohistochemical markers for clinical use. EBioMedicine 12:105–117

37. Rodriguez-Pinilla SM (2006) Prognostic significance of basal-like phenotype and fascin expression in node-negative invasive breast carcinomas. Clin Cancer Res 12:1533–1539

38. van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J et al (2002) Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. Am J Pathol 161:1991–1996

39. Rowlands D, Lensjo KK, Dinh T, Yang S, Andrews MR, Hafting T et al (2018) Aggrecan directs extracellular matrix-mediated neuronal plasticity. J Neurosci 38:10102–10113

40. Suna G, Wojakowski W, Lynch M, Barallobre-Barreiro J, Yin X, Mayr U et al (2018) Extracellular matrix proteomics reveals interplay of aggrecan and aggrecanases in vascular remodeling of stented coronary arteries. Circulation 137:166–183

41. Oskarsson T (2013) Extracellular matrix components in breast cancer progression and metastasis. Breast 22(Suppl 2):S66–72

42. Joensuu H, Kellokumpu-Lehtinen P-L, Huovinen R, Jukkola-Vuorinen A, Tanner M, Kokko R et al (2017) Adjuvant capecitabine in combination with docetaxel, epirubicin, and cyclophosphamide for early breast cancer: the randomized clinical FinXX trial. JAMA Oncol 3:793–800

43. Masuda N, Lee S-J, Ohtani S, Im Y-H, Lee E-S, Yokota I et al (2017) Adjuvant capecitabine for breast cancer after preoperative chemotherapy. N Engl J Med 376:2147–2159

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.