

# Transcriptome Analysis Derives a Novel Prognostic 7-mRNA Signature in Early-Stage Triple-Negative Breast Cancer

Yun-Song Yang (✉ [17211230025@fudan.edu.cn](mailto:17211230025@fudan.edu.cn))

Fudan University Huashan Hospital Jing'an Branch: Jing'an District Centre Hospital of Shanghai  
<https://orcid.org/0000-0001-8067-9567>

Yi-Xing Ren

Fudan University Shanghai Cancer Center

Shuang Hao

Fudan University Shanghai Cancer Center

Xiao-En Xu

Fudan University Shanghai Cancer Center

Xi Jin

Fudan University Shanghai Cancer Center

Yi-Zhou Jiang

Fudan University Shanghai Cancer Center

Zhi-Ming Shao

Fudan University Shanghai Cancer Center

---

## Research

**Keywords:** Triple-negative breast cancer, transcriptome analysis, gene-expression signature, recurrence-free survival, TMEM101

**Posted Date:** April 7th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-389386/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background:** Triple-negative breast cancer (TNBC) is a highly heterogeneous disease and patients with early-stage TNBCs have distinct likelihood of distant recurrence.

**Methods:** In this study, We extracted transcriptome data for 189 pathologically confirmed pT1-2 node-negative TNBC patients at Fudan University Shanghai Cancer Center. Candidate mRNAs were filtered, which was followed by differential expressed mRNAs analysis, survival analysis, and LASSO Cox regression model. All-subsets regression program was used for constructing a multi-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, and developed a prognostic signature based on seven mRNAs (ACAN, KRT5, TMEM101, LCA5, RPP40, LAGE3, CDKL2). In both the training (n=159) and validation cohorts (n=158), the signature could identify patients with relatively high recurrence risks and serve as an independent prognostic factor. Furthermore, the signature had better prognostic value than traditional clinicopathological features in both sets. Among the seven mRNAs, TMEM101 was identified as a prognostic biomarker of early-stage TNBC. Additional cell experiments suggested that TMEM101 could facilitate migration and proliferation of TNBC cells.

**Conclusions:** Our 7-mRNA signature could accurately predict recurrence risks of early-stage TNBCs. Clinical and genomic low risk TNBC patients may safely avoid adjuvant chemotherapy.

## Background

Triple-negative breast cancer (TNBC), which is a breast cancer subtype that lacks estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2), is the most aggressive type of breast cancer [1]. Recent advances in breast cancer screening have significantly improved the diagnosis of early breast cancer, such as pT1-2 and node-negative breast cancers [2]. However, outcomes for these patients may vary by tumor subtype and biological behavior [3]. Early-stage TNBC represents a clinically high-risk group associated with a more aggressive clinical course compared with other breast cancer subtypes [3–5]. Due to the absence of effective tools, genomic low-risk early-stage TNBC has not been identified who might not be benefited from chemotherapy.

So far, many efforts have been dedicated to identifying low clinical risk and high genomic risk subtypes based on genomic profiling. The multi-gene signature assay, Oncotype DX, is a novel genomic tumor profiling tool that determines the expression of 21 tumor-associated genes which has a high value in guiding node-negative hormone-receptor (HR)-positive breast cancer patients' individualized treatments [6]. However, there has no precise approaches to tailoring therapy of node-negative TNBCs [7]. Retrospective researches have reported the opportunity of de-escalating chemotherapy in pT1N0M0 TNBC, such as offering chemotherapy regimen with low toxicity [8], eliminating taxane [9], or even eliminating chemotherapy altogether [10]. For pT2N0M0, HR- patients have worse survival than HR+ patients, which

implied that some high-risk patients need escalating chemotherapy or frequent follow-up [4]. Thus, developing an accurate and stable prognostic signature could add significant value to individual treatments of node-negative TNBCs.

Previous studies have tried to use mRNA, micro-RNA, long-noncoding RNA, or proteome profile building prognostic signatures to assess the risk of recurrence and magnitude of chemotherapy benefit of node-negative breast cancer [11–13]. 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 data with the high variable to sample size ratios [14]. Second, these studies enrolled both node-negative and -positive TNBC patients. With a relatively small sample size of node-negative patients, it is challenging to generate a precise model for node-negative TNBCs.

In this study, we developed a new gene-expression signature by LASSO (Least Absolute Shrinkage and Selection Operator) Cox regression model to predict relapse-free survival (RFS) in a large cohort of node-negative TNBCs. We hypothesized that this signature could identify the high-risk subpopulation among node-negative TNBC and provide instructions for individual treatments.

## Methods

### Patients and samples

A total of 317 consecutive patients treated in the Department of Breast Surgery at Fudan University Shanghai Cancer Center (FUSCC) between January 1, 2008 and December 31, 2015 were enrolled in this prospective observational study, which was initiated in 2017. 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-, PR- and HER2-); (2) tumor size < 5cm and no lymph-node metastasis or distant metastasis confirmed by pathology; (3) patients who did not receive any type of treatment prior to surgery; (4) sufficient frozen tissues for RNA purification.

The ER, PR and HER2 status 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 [15]. The pathologic stage was assigned based on the 8<sup>th</sup> edition of the AJCC anatomic staging system [16]. All of the breast cancer specimens were confirmed to contain more than 80% of tumor cells. A total of 330 frozen samples (including 13 paired normal tissues) from 317 TNBC patients preserved by the Department of Pathology in FUSCC (Shanghai, P.R. China) were examined.

To develop candidate mRNAs from high-throughput transcriptome sequencing data, we enrolled another 386 consecutive patients who underwent surgical treatment in the Department of Breast Surgery at Fudan University Shanghai Cancer Center (FUSCC) between 1 January 2007 and 31 December 2014 with Affymetrix GeneChip Human Transcriptome Array 2.0 (HTA 2.0) (n=141) and RNA-seq (n=245) data from our previous researches [17,18]. Using the same criteria as described above (except for sufficient frozen tissues), 189 patients (50 patients had paired adjacent normal tissues) were enrolled as developing set in

this study. The follow-up was completed on 30 June, 2017, and the median follow-up length was 47.6 months (interquartile range, 38.5-59.3 months). There were 79 patients who overlapped between the developing set and the above 317 patients.

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  $CT^{mRNA} - (CT^{U6} + CT^{GAPDH}) / 2$  (CT, cycle threshold).

### **Identification of candidate mRNAs**

The detailed filtration process is illustrated in **Figure 1**. 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). The differentially expressed mRNA screening 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 down-regulation, 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 [19].

### **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 [20]. 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 Yoden Index [21]. To validate this signature, the risk score of each patient in the validation set was calculated using the formula generated in training set.

### **Cell cultures and siRNA transfection**

Human breast cancer cell lines MDA-MB-231 (RRID:CVCL\_0062), BT-549 (RRID:CVCL\_1092) and human embryonic kidney cells HEK293T (RRID: CVCL\_0063) 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 were 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 hours and transfected with siRNAs. After 24 hours, 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 hours. The cell confluence was measured by IncuCyte ZOOM software, and the assays were performed in triplicate.

### **Migration assay**

For the migration assay,  $5 \times 10^4$  cells in serum-free medium were seeded on the top of the transwell chamber (pore size, 8 $\mu$ m, BD Biosciences) and 600 $\mu$ l medium containing 10% FBS was added to the bottom chamber. After incubation for 8-10 hours, cells that migrated to the opposite side of the membrane were fixed in 4% paraformaldehyde for 30 minutes and stained with crystal violet for additional 30 minutes. 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 minutes. The proteins were separated by SDS-PAGE and transferred to polyvinylidene difluoride membranes (Millipore). The primary antibody anti-TMEM101 (Abcitech, 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).

### **Statistical Analysis**

The t-test was used to compare the differences between the two groups for continuous variables and  $\chi^2$  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. A  $P$  value  $< 0.05$  was considered to be statistically significant.

## Results

### Construction of a multi-gene prognostic signature for early-stage TNBCs

We identified 996 mRNAs differently expressed between 50 paired tumor and normal breast samples that were significantly related to survival in developing set ( $n=189$ ) (**Figure 1**). LASSO Cox regression model was used to select candidate prognostic mRNAs with 200 bootstrap replications. Twenty-one candidate mRNAs were eligible for developing a prognostic signature after the filtering procedure (**Figure S1a**). Next, the expressions of the 21 mRNAs were further evaluated in training ( $n=159$ ) and validation set ( $n=158$ ) by RT-qPCR.

In order to further simplify the prognostic signature, the genetic algorithm was used to perform automated screening for the best signature based on the AIC value (Methods). A total of 542,358 candidate signatures were identified, and finally, a 7-mRNA signature with the lowest AIC was returned as the best signature to calculate the risk score of each patient based on the expression level of the seven mRNAs in the training set (**Figure S1b, c**). The formula used to calculate the risk score was the following: Risk score =  $1.108 * ACAN - 0.213 * KRT5 - 0.315 * TMEM101 - 0.464 * LCA5 + 0.446 * RPP40 - 0.373 * LAGE3 - 0.257 * CDKL2$ .

### Prognostic value of the 7-mRNA signature

With the risk score formula, patients in the training set were divided into high- (11.3%,  $N=18$ ) or low-risk (88.7%,  $N=141$ ) groups according to the optimum cut-off score (-0.87) determined by the time-dependent ROC analysis. The distribution of the risk score and RFS status showed that patients with higher risk scores had a higher recurrence rate and worse survival (**Figure 2a, b**). Yet, there was no significant difference in traditional clinicopathological factors between the two groups (**Table 1**). The time-dependent ROC analysis demonstrated that the prognostic accuracy of this signature was similar at different follow-up times (**Figure 2c**). Patients in the validation set were then stratified into the low-risk group (84.2%,  $N=133$ ) and high-risk group (15.8%,  $N=25$ ) using the same procedure as in the training cohort. Similarly, the high-risk group had worse survival compared to the low-risk group in the validation set (**Figure 2d-f**).

### Multivariate analysis and subgroup analysis

Multivariate Cox proportional hazards regression analysis showed that after adjustment of traditional clinicopathological features, the 7-mRNA signature was still an independent prognostic factor associated with RFS (**Table 2**; HR=15.68; 95% CI, 5.50-44.70;  $P<0.001$ ). In addition, none of the traditional clinicopathological features was associated with RFS in the multivariate analysis except for lymphovascular invasion in the training set. Similar results were observed in the validation set. Patients with high-risk had a higher hazard 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 (**Figure S2**).

### Comparison with other prognostic factors

To compare the performance of the 7-mRNA signature and other traditional clinicopathological features, the area under curve (AUC) was calculated on signature and traditional clinicopathological features. The results showed that the 7-mRNA signature had better prognostic value in predicting RFS (AUC=0.766; **Figure 3a**). However, other 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 (**Figure 3b**).

### 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 (**Figure S3a**). Furthermore, Kaplan-Meier analysis demonstrated that high expression of *TMEM101* was significantly associated with RFS (**Figure 4a**;  $P=0.019$ ), while *LAGE3* was not related to patients survival (**Figure S3b**). Therefore, *TMEM101* resulted as a good candidate target or prognostic biomarker of TNBC.

The exact function of *TMEM101* in human still remains unknown. Some studies have suggested an association between the NF- $\kappa$ B pathway and *TMEM101* [22,23]. In this study, we found higher expression of *TMEM101* in tumor samples than in paired normal samples, which further suggested that *TMEM101* might be a prognostic biomarker of early-stage TNBCs (**Figure 4b**;  $P=0.003$ ). We also designed three siRNAs targeting *TMEM101*, and two siRNAs with the highest interference efficient inhibited the migration ability of MDA-MB-231 and BT-549 cell lines (**Figure 4c, d, f**). Besides, the downregulation of *TMEM101* significantly impaired the proliferation of tumor cells *in vitro* (**Figure 4e, g**). 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 status, and lymphovascular invasion status. Furthermore, we identified that *TMEM101* was related to poor prognosis and promoted of TNBC progression.

The prognosis of node-negative TNBC patients is rarely associated with clinical factors, including age, histological grade, and tumor size [24]. Previous retrospective research showed that lymphovascular invasion was the only traditional parameter associated with the survival of pT1a-b node-negative TNBC [9]. 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 [11]. According to this signature, 60% early-stage TNBC were with 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 [11] 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 a 286 cohort and a 180 cohort; 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 [13,25]. Also, in this study, ER-negative patients included both HER2-positive and TNBC subtype, which had entirely different biological features [26]. 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. High-throughput data was used to construct prognostic signatures, but it might not be widely used in clinical practice because of the relatively high costs of transcriptome sequencing. 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 [27]. 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 [28]. A two-gene epigenetic signature revealed that higher methylation of *CDKL2* was related to non-responders to neoadjuvant chemotherapy in triple-negative breast cancer [29]. *KRT5* encoded 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 [26,30–33]. Aggrecan encoded by *ACAN* is an important component for cartilage structure, which may facilitate ECM remodeling [34,35]. Although the function of aggrecan in cancer is unknown, in breast cancer, ECM components are significantly dysregulated and associated with tumor progression [36]. *TMEM101* has been rarely reported in cancer research. Moreover, using a cDNA library, Matsuda and colleagues revealed that 83 genes could activate the NF- $\kappa$ B pathway, including *TMEM101* [22]. Recently, Khan and colleagues demonstrated that *TMEM101* could increase the expression of NF- $\kappa$ B in zebrafish ovarian samples [23]. 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. Yet, according to our previous research that anthracycline-based taxane-free regimen might be sufficient for pT1N0M0 TNBCs so that patients with low recurrence risk had the opportunity to omit taxane [9]. 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 [37,38], 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, since about 5–15% of node-negative TNBC patients developed tumor recurrence, our 7-mRNA prognostic signature could identify patients at high recurrence risks. The 7-mRNA signature could be a powerful tool for oncologists to escalate systemic therapy for high-risk node-negative TNBC. Subsequent clinical trials are needed to further validate the signature.

## Abbreviations

AIC: akaike information criterion; AUC: area under curve; ER: estrogen receptor; HER2: epidermal growth factor receptor 2; HR: hormone-receptor; LASSO: least absolute shrinkage and selection operator; RFS: relapse-free survival; ROC: receiver operating characteristic; PR: progesterone receptor; RT-qPCR: real-time quantitative-polymerase chain reaction; siRNA: small interfering RNAs; TNBC: triple-negative breast cancer.

## Declarations

### Ethics approval and consent to participate

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

### Consent for publication

All authors read and approved the manuscript as submitted.

### Availability of data and materials

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

### Competing interests

The authors declare that they have no competing interests.

### Funding

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).

### Author contributions

Conception and design: Yun-Song Yang, Yi-Xing Ren, Xi Jin, Xiao-En Xu, Yi-Zhou Jiang, Zhi-Ming Shao; Development of methodology: Yun-Song Yang, Yi-Xing Ren, Xiao-En Xu, Yi-Zhou Jiang; Acquisition of data: Yun-Song Yang, Yi-Xing Ren, Shuang Hao, Yi-Zhou Jiang; Analysis and interpretation of data: Yun-Song Yang, Yi-Xing Ren, Xi Jin, Yi-Zhou Jiang; Writing, review, and/or revision of the manuscript: Yun-Song Yang, Yi-Xing Ren, Shuang Hao, Xi Jin, Yi-Zhou Jiang, Zhi-Ming Shao; Administrative, technical, or material support: Yun-Song Yang, Yi-Xing Ren, Shuang Hao, Xi Jin, Xiao-En Xu, Yi-Zhou Jiang; Study supervision: Yi-Zhou Jiang, Zhi-Ming Shao

## References

1. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. *N Engl J Med*. 2010;363:1938–48.
2. Fan L, Strasser-Weippl K, Li J-J, St Louis J, Finkelstein DM, Yu K-D, et al. Breast cancer in China. *Lancet Oncol*. 2014;15:e279–89.
3. Bao J, Donovan C, Amersi F, Zhang X, Giuliano AE, Chung A. Outcomes in patients with small node-negative invasive breast cancer. *Breast J*. 2019;25:638–43.
4. Parise CA, Caggiano V. Risk of mortality of node-negative, ER/PR/HER2 breast cancer subtypes in T1, T2, and T3 tumors. *Breast Cancer Res Treat*. 2017;165:743–50.
5. Amar S, McCullough AE, Tan W, Geiger XJ, Boughey JC, McNeil RB, et al. Prognosis and outcome of small ( $\leq 1$  cm), node-negative breast cancer on the basis of hormonal and HER-2 status. *Oncologist*. 2010;15:1043–9.
6. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med*. 2018;379:111–21.
7. Carey LA. De-escalating and escalating systemic therapy in triple negative breast cancer. *Breast Edinb Scotl*. 2017;34(Suppl 1):112–5.
8. Kim H-A, Seong M-K, Kim E-K, Kang E, Park S, Hur MH, et al. Evaluation of the Survival Benefit of Different Chemotherapy Regimens in Patients with T1-2N0 Triple-Negative Breast Cancer. *J Breast*

- Cancer. 2015;18:271–8.
9. Ren Y-X, Hao S, Jin X, Ye F-G, Gong Y, Jiang Y-Z, et al. Effects of adjuvant chemotherapy in T1N0M0 triple-negative breast cancer. *The Breast*. 2019;43:97–104.
  10. de Nonneville A, Gonçalves A, Zemmour C, Cohen M, Classe JM, Reyal F, et al. Adjuvant chemotherapy in pT1ab node-negative triple-negative breast carcinomas: Results of a national multi-institutional retrospective study. *Eur J Cancer*. 2017;84:34–43.
  11. Liu NQ, Stingl C, Look MP, Smid M, Braakman RBH, De Marchi T, et al. Comparative proteome analysis revealing an 11-protein signature for aggressive triple-negative breast cancer. *J Natl Cancer Inst*. 2014;106:djt376.
  12. Liu Y-R, Jiang Y-Z, Xu X-E, Yu K-D, Jin X, Hu X, et al. Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer. *Breast Cancer Res*. 2016;18:33.
  13. Wang Y, Klijn JGM, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. 2005;365:9.
  14. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer*. 1994;69:979–85.
  15. Hammond MEH, Hayes DF, Wolff AC, Mangu PB, Temin S. 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*. 2010;6:195–7.
  16. Amin MB, Edge SB. *AJCC cancer staging manual*. Springer; 2017.
  17. Jiang Y-Z, 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.
  18. Xiao Y, Ma D, Zhao S, Suo C, Shi J, Xue M-Z, et al. Multi-Omics Profiling Reveals Distinct Microenvironment Characterization and Suggests Immune Escape Mechanisms of Triple-Negative Breast Cancer. *Clin Cancer Res*. 2019;25:5002–14.
  19. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997;16:385–95.
  20. Calcagno V, de Mazancourt C. glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models. *J Stat Softw*. 2010;34.
  21. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32–5.
  22. Matsuda A, Suzuki Y, Honda G, Muramatsu S, Matsuzaki O, Nagano Y, et al. Large-scale identification and characterization of human genes that activate NF-kappaB and MAPK signaling pathways. *Oncogene*. 2003;22:3307–18.
  23. Khan ZA, Labala RK, Yumnamcha T, Devi SD, Mondal G, Sanjita Devi H, et al. Artificial Light at Night (ALAN), an alarm to ovarian physiology: A study of possible chronodisruption on zebrafish (*Danio rerio*). *Sci Total Environ*. 2018;628–629:1407–21.
  24. Rakha EA, El-Sayed ME, Green AR, Lee AHS, Robertson JF, Ellis IO. Prognostic markers in triple-negative breast cancer. *Cancer*. 2007;109:25–32.

25. Foekens JA, Atkins D, Zhang Y, Sweep FCGJ, Harbeck N, Paradiso A, et al. Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 2006;24:1665–71.
26. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406:747–52.
27. Liu Y-R, Jiang Y-Z, Xu X-E, Hu X, Yu K-D, Shao Z-M. Comprehensive Transcriptome Profiling Reveals Multigene Signatures in Triple-Negative Breast Cancer. *Clin Cancer Res*. 2016;22:1653–62.
28. Li L, Liu C, Amato RJ, Chang JT, Du G, Li W. CDKL2 promotes epithelial-mesenchymal transition and breast cancer progression. *Oncotarget*. 2014;5:10840–53.
29. Pineda B, Diaz-Lagares A, Pérez-Fidalgo JA, Burgués O, González-Barrallo I, Crujeiras AB, et al. A two-gene epigenetic signature for the prediction of response to neoadjuvant chemotherapy in triple-negative breast cancer patients. *Clin Epigenetics*. 2019;11:33.
30. Zhao SG, Chen WS, Das R, Chang SL, Tomlins SA, Chou J, et al. Clinical and Genomic Implications of Luminal and Basal Subtypes Across Carcinomas. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2019;25:2450–7.
31. Dadhania V, Zhang M, Zhang L, Bondaruk J, Majewski T, Siefker-Radtke A, et al. Meta-Analysis of the Luminal and Basal Subtypes of Bladder Cancer and the Identification of Signature Immunohistochemical Markers for Clinical Use. *EBioMedicine*. 2016;12:105–17.
32. Rodriguez-Pinilla SM. Prognostic Significance of Basal-Like Phenotype and Fascin Expression in Node-Negative Invasive Breast Carcinomas. *Clin Cancer Res*. 2006;12:1533–9.
33. van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, et al. Expression of Cytokeratins 17 and 5 Identifies a Group of Breast Carcinomas with Poor Clinical Outcome. *Am J Pathol*. 2002;161:1991–6.
34. Rowlands D, Lensjø KK, Dinh T, Yang S, Andrews MR, Hafting T, et al. AggreCAN Directs Extracellular Matrix-Mediated Neuronal Plasticity. *J Neurosci Off J Soc Neurosci*. 2018;38:10102–13.
35. Suna G, Wojakowski W, Lynch M, Barallobre-Barreiro J, Yin X, Mayr U, et al. Extracellular Matrix Proteomics Reveals Interplay of AggreCAN and AggreCANases in Vascular Remodeling of Stented Coronary Arteries. *Circulation*. 2018;137:166–83.
36. Oskarsson T. Extracellular matrix components in breast cancer progression and metastasis. *Breast Edinb Scotl*. 2013;22(Suppl 2):66–72.
37. Joensuu H, Kellokumpu-Lehtinen P-L, Huovinen R, Jukkola-Vuorinen A, Tanner M, Kokko R, et al. Adjuvant Capecitabine in Combination With Docetaxel, Epirubicin, and Cyclophosphamide for Early Breast Cancer: The Randomized Clinical FinXX Trial. *JAMA Oncol*. 2017;3:793–800.
38. Masuda N, Lee S-J, Ohtani S, Im Y-H, Lee E-S, Yokota I, et al. Adjuvant Capecitabine for Breast Cancer after Preoperative Chemotherapy. *N Engl J Med*. 2017;376:2147–59.

## Tables

**Table 1.** Clinicopathological features of three sets of pT1-2N0M0 TNBC patients stratified according to the 7-mRNA signature

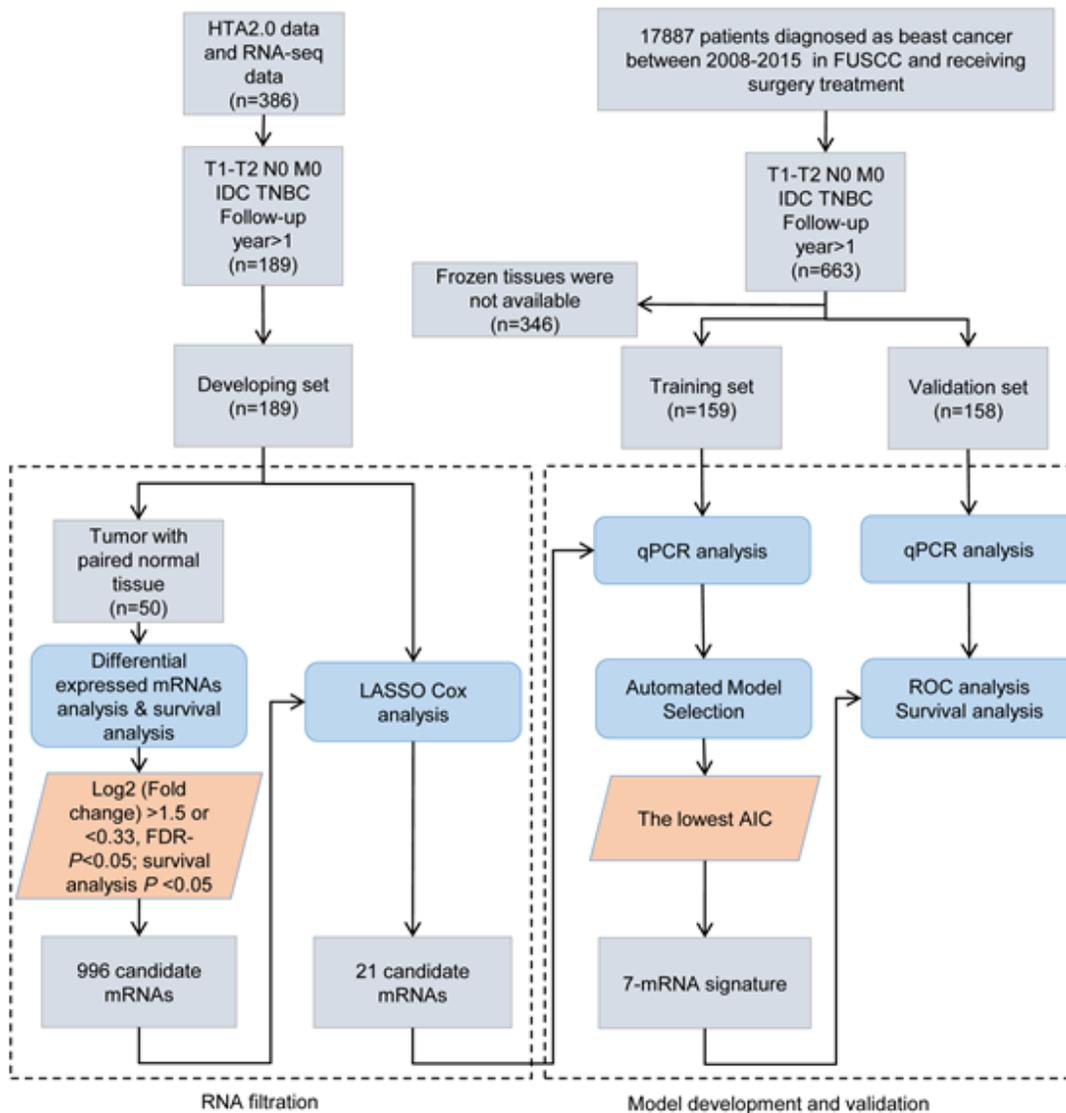
Clinical features	Developing set	Training set			<i>P</i> value <sup>a</sup>	Validation set			<i>P</i> value <sup>a</sup>
		N	Low risk (%)	High risk (%)		N	Low risk (%)	High risk (%)	
	189	159	141 (88.7)	18 (11.3)		158	133 (84.2)	25 (15.8)	
Age (years)					0.469				0.873
<55	74 (39.2)	98	85 (60.3)	13 (72.2)		83	69 (51.9)	14 (56.0)	
>=55	115 (60.8)	61	56 (39.7)	5 (27.7)		75	64 (48.1)	11 (44.0)	
Menopause					0.315				0.384
Yes	124 (65.6)	101	92 (65.2)	9 (50.0)		105	86 (64.7)	19 (76.0)	
No	65 (34.4)	58	49 (34.8)	9 (50.0)		53	47 (35.3)	6 (24.0)	
Tumor size (cm)					0.58				0.99
<=2	81 (42.9)	76	69 (48.9)	7 (38.9)		71	60 (45.1)	11 (44.0)	
>2	108 (57.1)	83	72 (51.1)	11 (61.1)		87	73 (54.9)	14 (56.0)	
Tumor grade					0.557				0.369
II	35 (18.5)	28	26 (18.4)	2 (11.1)		41	37 (27.8)	4 (16.0)	
II-III	17 (9.0)	24	20 (14.2)	4 (22.2)		17	13 (9.8)	4 (16.0)	
III	137 (72.5)	107	95 (67.4)	12 (66.7)		100	83 (62.4)	17 (68.0)	
Ki-67					0.861				0.99
<=50%	86 (45.5)	87	78 (55.3)	9 (50.0)		80	67 (50.4)	13 (52.0)	
>50%	103 (54.5)	72	63 (44.7)	9 (50.0)		78	66 (49.6)	12 (48.0)	
Lymphovascular invasion					0.715				0.08
Yes	NA	18	15 (10.6)	3 (16.7)		28	20 (15.0)	8 (32.0)	

No	NA	141	126 (89.4)	15 (83.3)		130	113 (85.0)	17 (68.0)
Chemotherapy					0.545	0.212		
Yes	177 (93.7)	145	128 (90.8)	17 (94.4)		129	107 (80.5)	22 (88.0)
No	4 (2.1)	6	5 (3.5)	1 (5.6)		7	5 (3.8)	2 (8.0)
UNK	8 (4.2)	8	8 (5.7)	0 (0.0)		22	21 (15.8)	1 (4.0)
Radiotherapy					0.441	0.679		
Yes	19 (10.1)	13	12 (8.5)	1 (5.6)		4	4 (3.0)	0 (0.0)
No	169 (89.4)	136	119 (84.4)	17 (94.4)		142	119 (89.5)	23 (92.0)
UNK	1 (0.5)	10	10 (7.1)	0 (0.0)		12	10 (7.5)	2 (8.0)
RFS status					<0.001	<0.001		
Yes	11 (5.8)	17	7 (5.0)	10 (55.6)		15	7 (5.3)	8 (32.0)
No	178 (94.2)	142	134 (95.0)	8 (44.4)		143	126 (94.7)	17 (68.0)
Follow-up time (Median, Months)	47.6		41	32.8			39.1	36.7
Abbreviations: UNK, unknown; RFS, relapse-free survival								
a <i>P</i> -value was calculated using $\chi^2$ test								

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

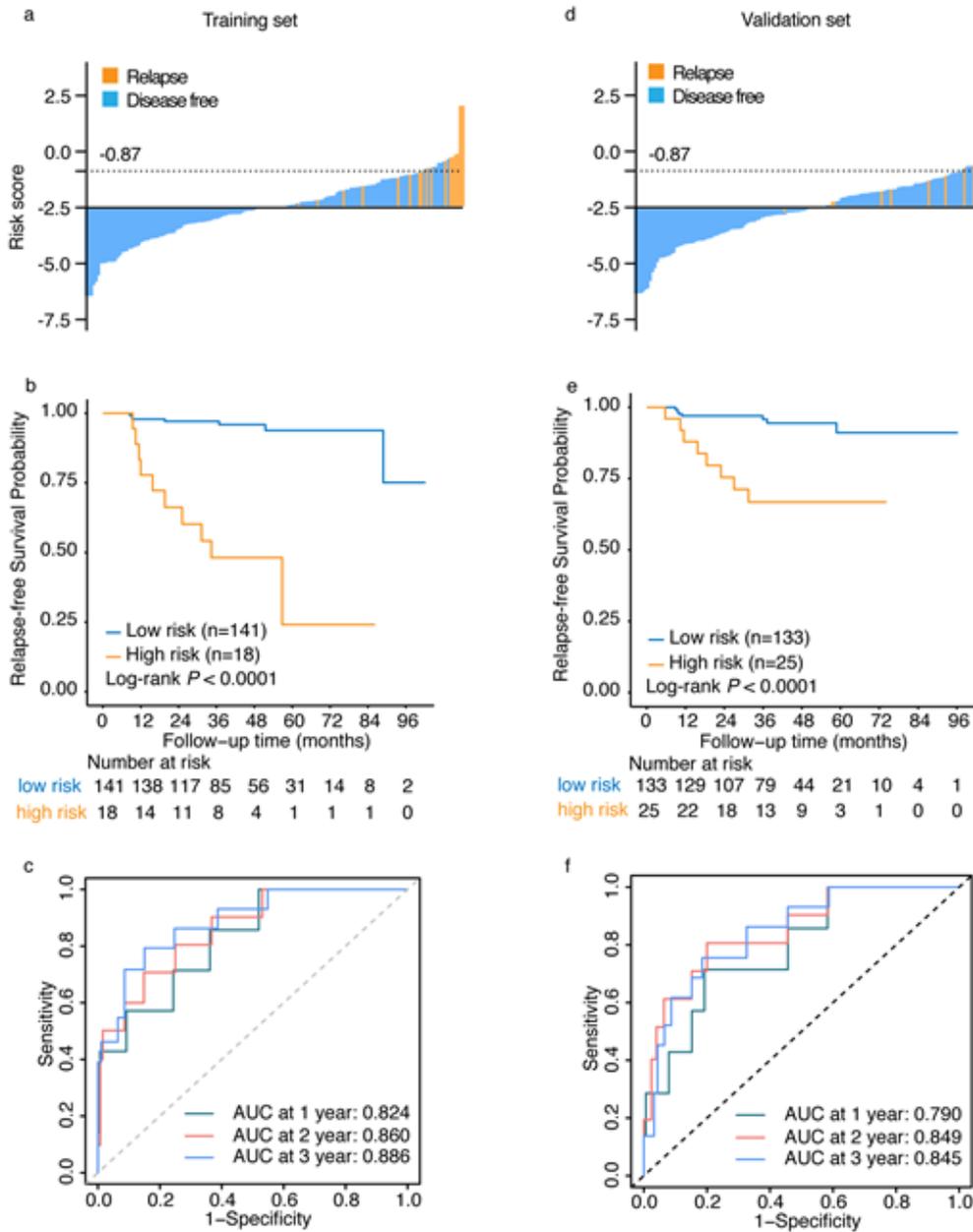
Clinical features	Training set		Validation set	
	HR <sup>a</sup> (95% CI)	<i>P</i> value	HR <sup>a</sup> (95% CI)	<i>P</i> value
7-mRNA signature (low risk as reference)	15.68 (5.50-44.70)	<b>&lt;0.001</b>	5.87 (2.07-16.66)	<b>0.001</b>
Tumor size (tumor size < 2cm as reference)	1.97 (0.66-5.86)	0.221	1.57 (0.51-4.80)	0.434
Tumor grade II-III (grade II as reference)	1.18 (0.19-7.38)	0.859	3.49 (0.74-16.48)	0.115
Tumor grade III (grade II as reference)	1.12 (0.22-5.77)	0.890	0.93 (0.24-3.63)	0.920
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)	<b>0.009</b>	2.53 (0.83-7.71)	0.104
Abbreviations: RFS, relapse-free survival; HR, hazard ratio				
a Adjusted by Cox proportional hazards models including age, tumor size, tumor grade, Ki-67, lymphovascular invasion, chemotherapy, radiotherapy, and the 7-mRNA signature				

## Figures



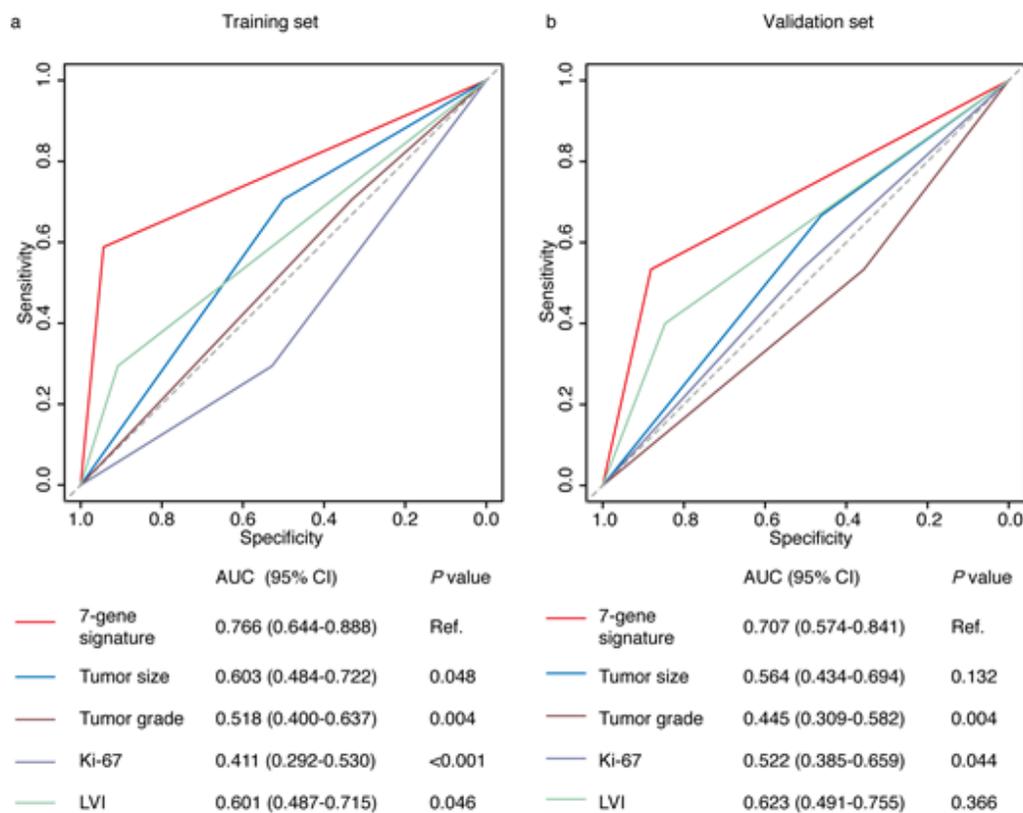
**Figure 1**

The procedure of RNA filtration, signature construction, and validation. Abbreviations: HTA, Human Transcriptome Array; IDC, invasive ductal carcinoma; TNBC, triple-negative breast cancer; FDR, false discovery rate; LASSO, Least Absolute Shrinkage and Selection Operator; qPCR, real-time quantitative polymerase chain reaction; ROC, receiver operating characteristic; AIC, Akaike information criterion;



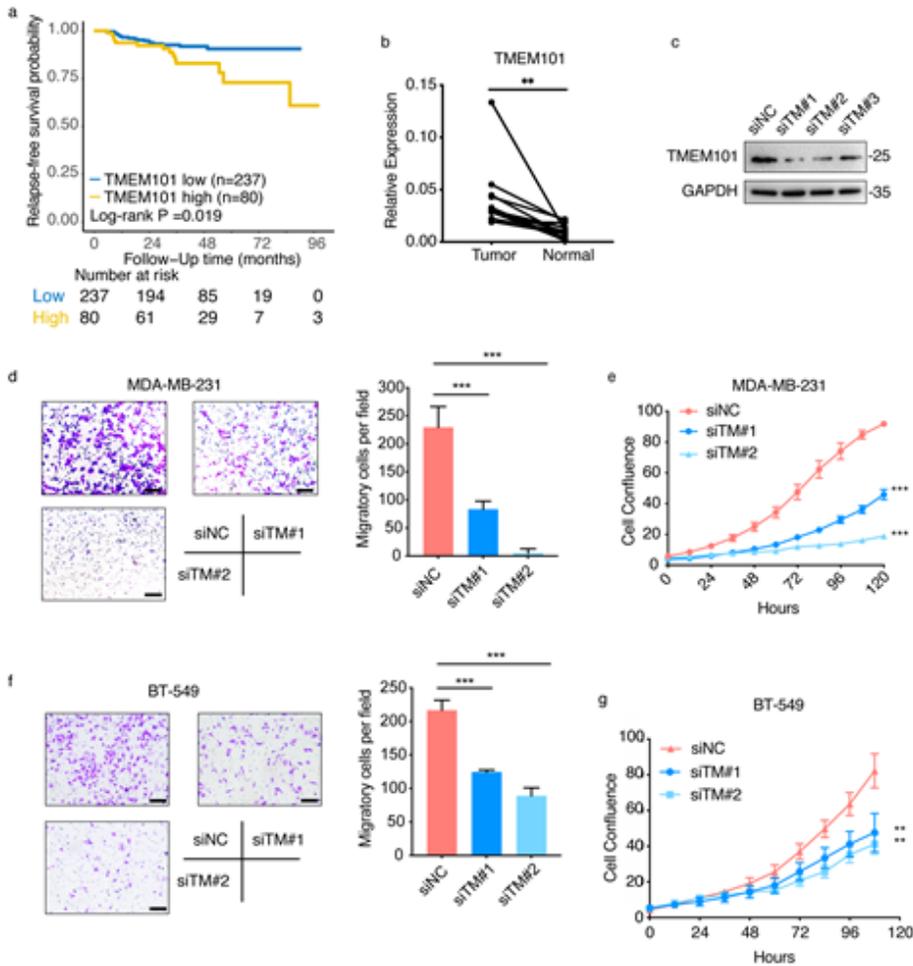
**Figure 2**

Prognostic value and accuracy of the 7-mRNA signature tested in training and validation set. Distribution of risk score by the 7-mRNA signature in training a and validation set d. Kaplan-Meier curves of relapse-free survival in training b and validation set e, and P value were calculated by log-rank test. Time-dependent ROC calculated at 1, 2, 3 years in training c and validation set f. Abbreviations: AUC, area under curve.



**Figure 3**

Comparison of the prognostic accuracy of the 7-mRNA signature with traditional clinicopathological features. a, b Comparison the sensitivity and specificity of prognosis by the 7-mRNA signature and the traditional clinicopathological features, including tumor size, tumor grade, Ki-67 and LVI in training and validation set. P value was generated by bootstrap at 200 replicates. Abbreviations: AUC, area under curve; LVI, lymphovascular invasion.



**Figure 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. d, f Representative images and the number of migrated MDA-MB-231 and BT-549 cells interfered with TMEM101 siRNAs. e, g In vitro growth curves of MDA-MB-231 and BT-549 interfered with TMEM101 siRNAs for 5 days. Scale bar, 100 $\mu$ m. Abbreviations: TM, TMEM101. \*, P < 0.05; \*\*, P < 0.01; \*\*\*, P < 0.001

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialsV5.docx](#)