Clinical Cancer Research

Comprehensive Transcriptome Profiling Reveals Multigene Signatures in Triple-Negative Breast Cancer

Yi-Rong Liu[1,2,3], Yi-Zhou Jiang[1,2,3], Xiao-En Xu[1,2,3], Xin Hu[1,2,3], Ke-Da Yu[1,2,3], and Zhi-Ming Shao[1,2,3,4]

Abstract

Purpose: By integrating expression profiles of mRNAs and long noncoding RNAs (lncRNA), we tried to develop and validate novel multigene signatures to facilitate individualized treatment of triple-negative breast cancer (TNBC) patients.

Experimental Design: We analyzed 165 TNBC samples and 33 paired normal breast tissues using transcriptome microarrays. Tumor-specific mRNAs and lncRNAs were identified and correlated with patients' recurrence-free survival (RFS). Using Cox regression model, we built two multigene signatures incorporating mRNAs and lncRNAs. The prognostic and predictive accuracy of the signatures were tested in a training set of 165 TNBC patients and validated in other 101 TNBC patients.

Results: We successfully developed an mRNA and an integrated mRNA–lncRNA signature based on eight mRNAs and two lncRNAs. In the training set, patients in the high-risk group were more likely to suffer from recurrent disease than patients in the low-risk group in both signatures [HR, 10.00; 95% confidence interval (CI), 2.53–39.47, $P = 0.001$; HR = 4.46, 95% CI, 1.34–14.91, $P = 0.015$ for integrated signature and mRNA signature, respectively). Results were validated in the validation set ($P = 0.019$ and 0.030, respectively). In addition, time-dependent receiver operating curve showed that the integrated mRNA–lncRNA signature had a better prognostic value than both the eight-mRNA-only signature and the clinicopathologic risk factors in both sets. We also found through interaction analysis that patients classified into the low-risk group by the integrated mRNA–lncRNA signature had a more favorable response to adjuvant taxane chemotherapy.

Conclusions: The multigene signature we developed can accurately predict clinical outcome and benefit of taxane chemotherapy in TNBC patients. *Clin Cancer Res; 22(7); 1653–62. ©2016 AACR.*

Introduction

Triple-negative breast cancer (TNBC) is one of the most aggressive subtypes of breast cancer (1). Chemotherapy is the current mainstay of treatment. Despite having higher rates of clinical response to neoadjuvant chemotherapy, TNBC patients have higher rates of distant recurrence and worse prognosis compared to women with other subtypes of breast cancer. The treatment of patients with TNBC has been challenging due to the heterogeneity of the disease and the absence of well-defined molecular targets (2). Even within TNBC patients, similar chemotherapeutic strategies evoke diverse responses. At present, there is an urgent need to categorize such differences within TNBC at the time of diagnosis. Therefore, highly sensitive and specific prognostic signatures would be of great value in the individualized treatment of TNBC patients.

With the development of high-throughput technologies, several multigene signatures have been developed to predict the prognosis of breast cancer patients (3–5). Compared with traditional clinicopathologic factors, multigene signatures have higher sensitivity and specificity (3–5). However, these signatures all have limited applicable population, and only few signatures are specified for TNBC patients hitherto. The well-known multigene signature, Oncotype DX, can only help to predict the potential benefit of chemotherapy and likelihood of distant breast cancer recurrence in women with estrogen receptor (ER)-positive and human epidermal growth factor receptor 2 (HER2)-negative invasive breast cancer (4). Other available genomic prognostic signature, such as Mammoprint and Genomic Grading Index, would merely classify all TNBC samples into the poor prognosis group without further distinction (3, 5). By assembling gene data of 579 TNBC patients from the Gene Expression Omnibus database, Rody and colleagues found that the B-cell/IL8 metagene ratio could be a powerful prognostic marker for TNBC; however, intertumor heterogeneity and technical differences existing in different datasets may impair the application of this marker (6). Considering the emerging important role of long noncoding RNAs (lncRNA) in gene regulation and other cellular processes (7–10), a novel gene signature based on the transcriptome profiles of both mRNAs and lncRNAs would help better predict outcome of TNBC patients and treat them accordingly.

In the current study, we aimed to develop and validate multigene signatures by analyzing the transcriptome profiles including both mRNAs and lncRNAs. We hypothesised that signatures

[1]Department of Breast Surgery, Fudan University Shanghai Cancer Center, Shanghai, P.R. China. [2]Cancer Institute, Fudan University Shanghai Cancer Center, Shanghai, P.R. China. [3]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, P.R. China. [4]Institutes of Biomedical Sciences, Fudan University, Shanghai, P.R. China.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Y.-R. Liu, Y.-Z. Jiang, and X.-E. Xu contributed equally to this article; K.-D. Yu and Z.-M. Shao contributed equally to this article.

Corresponding Authors: Zhi-Ming Shao, Department of Breast Surgery, Fudan University Shanghai Cancer Center, Fudan University, 270 Dong-An Road, Shanghai, 200032, P.R. China. Phone: +86-21-64175590; Fax: +86-21-64438653; E-mail: zhimingshao@yahoo.com; and Ke-Da Yu, yukeda@163.com

doi: 10.1158/1078-0432.CCR-15-1555

Translational Relevance

Triple-negative breast cancer (TNBC) is a highly heterogeneous disease. By integrating the expression of messenger RNAs (mRNA) and long noncoding RNAs (lncRNA), we developed multigene signatures to facilitate individualized treatment of TNBC. In this prospective observational study, we identified tumor-specific mRNAs and lncRNAs associated with recurrence-free survival using transcriptome microarrays. An mRNA-only signature and an integrated mRNA–lncRNA signature was developed on the basis of eight mRNAs and two additional lncRNAs. The prognostic and predictive accuracy of the signatures were tested in a training set of 165 TNBC patients and further validated successfully in an independent validation set of 101 TNBC patients. Furthermore, our data revealed that the novel lncRNAs HIST2H2BC and SNRPEP4 incorporated in the integrated signature promoted cell proliferation and invasion and contributed to paclitaxel resistance in TNBC cells. The multigene signatures developed in the current study could facilitate patient counseling and individualized treatment of TNBC patients.

integrating more transcript information would improve risk stratification of TNBC patients and provide a more accurate assessment of individual treatment.

Materials and Methods

Patients and samples

All analyses were performed according to the reporting recommendations for tumor marker prognostic studies (REMARK) for prognostic and tumor marker studies, and the respective guidelines of microarray-based studies for clinical outcomes. Diagram of the study design, flow of patients, and analytic strategy is shown in Supplementary Fig. S3.

This prospective observational study was initiated in 2011. In the training cohort, a total of 198 frozen tissues from 165 consecutive TNBC patients (including 33 pairs of tumor and adjacent normal tissues) were collected in the Department of Breast Surgery at Fudan University Shanghai Cancer Center (FUSCC, Shanghai, P.R. China) between January 1, 2011 and December 31, 2012. The percentage of tumor cells was over 80% in all breast cancer specimens. Two individual pathologists evaluated ER, progesterone receptor (PR), and HER2 expression levels by immunohistochemistry and FISH. Statuses of the three receptors were assessed using the American Society of Clinical Oncology/College of American Pathologists guidelines of that time (11, 12). Patients selected for the study fulfilled the following inclusion criteria: (i) female patients diagnosed with unilateral histologically confirmed invasive ductal carcinoma with phenotype ER− PR−, and HER2−; *in situ* breast carcinomas (with or without microinvasions) were excluded, (ii) had pathologic examination of tumor specimens carried out by the Department of Pathology in FUSCC (Shanghai, P.R. China), and (iii) no evidence of metastasis at diagnosis (13). Recurrence-free survival was defined as the time from the date of surgery to the date of confirmed tumor recurrence and censored at the date of death from other causes, or the date of the last follow-up visit for recurrence-free patients. Using the same inclusion criteria as above, we recruited another 101 consecutive TNBC patients between January 1, 2010 and December 31, 2010 to validate the signature developed from the training set.

Procedures

RNA was isolated from 266 frozen TNBC samples and 33 adjacent normal breast tissue using the RNeasy Plus Mini Kit (Qiagen). The Affymetrix GeneChip Human Transcriptome Array 2.0 (HTA 2.0) was used to examine the expression profiles of RNAs in the training set of 165 patients according to the standard protocol, which covers more than 285,000 full-length transcripts. The details of the infiltration procedures are listed in Supplementary Fig. S4. We used a random variance model to identify the differentially expressed RNAs between the 33 paired tumor and normal breast tissues. Differences in RNA expression were regarded as significant if values for false discovery rate were less than 0.001 (10, 14). Considering the low expression level of lncRNAs in tissue, only upregulated lncRNAs in tumor samples were included for further analysis (15). We analyzed the association between each of the RNAs and patient recurrence-free survival (RFS) using univariate Cox proportional hazards regression model with BRB-Array Tools. RNAs significantly correlated with RFS were selected as candidates for further analysis. Duplicated mRNAs were excluded as well as the nonintergenic lncRNAs. Using the expression of GAPDH as a reference, we examined the tumor-specific and RFS-related RNAs using qRT-PCR in the training set of 165 TNBCs and 33 paired normal breast tissues. A maximum of six pairs of primers were designed for each RNA and validated in the paired samples. If all six primers failed, we deemed the RNA as technical error from microarray analysis. Further correlation analyses were conducted to examine the association between data obtained from microarray platform and qRT-PCR platform in the training set. For mRNAs, we further examined their expression pattern in the Oncomine database (16). Selected RNAs were used to construct signatures based on their coefficients in the multivariate Cox proportional hazards regression model (17). To calculate the risk score of each patient using the formula, the expression level of each RNA was recorded as high or low based on cohort median expression, and, respectively, given the values 1 or 0. We selected the optimum cut-off scores for the signatures using time-dependent ROC analysis. To test the efficacy of the signatures, we conducted a time-dependent ROC analysis and used the AUC to measure the prognostic accuracy in the training set (17–19).

We further tested the signature's performance in another independent cohort of 101 consecutive TNBC patients. The expression levels of all RNAs included in the signature were assessed using qRT-PCR with GAPDH as a reference and also coded as high or low expression level based on the cohort median expression. Using the multivariate Cox proportional hazards regression model, formulas calculating each patient's recurrence risk score were developed, in which high and low expression equaled to 1 and 0, respectively. As in the training set, time-dependent ROC analyses were applied to decide the optimum cut-off values for the signature and the performance of the signatures were evaluated.

Cell cultures

Breast cancer cell lines (MDA-MB-468 and MDA-MB-231) and 293T cells were obtained from the ATCC and maintained in complete growth medium as described previously (20). Liquid

nitrogen stocks were created upon receipt, and cells were maintained in liquid nitrogen until the start of each study. Cell morphology and doubling times were also regularly recorded to ensure the maintenance of phenotypes. Cells were used for no more than 6 months after being thawed.

RNA interference

siRNAs against two candidate lncRNAs were designed using BLOCK-iT RNAi Designer (Life Technologies). The siRNA oligonucleotides were synthesized by GenePharma Co. Ltd. For reverse siRNA transfection, the procedure was performed as follows: Briefly, 25 μL Opti-MEM medium dissolved in 0.3 μL Lipofectamine RNAiMAX was added to 25 μL Opti-MEM medium containing 7.5 pmol siRNA duplex (final concentration 50 nmol/L), and the mixture was dripped into each well of a 96-well plate. Approximately 5×10^3 cells suspended in 100 μL antibiotic-free growth medium were added to each well. To test for cell viability, the proliferation rate was measured with Cell Counting Kit-8 (Dojindo) after 72 hours of transfection. A scrambled sequence served as negative control. All raw data were collected at 450 nm.

Measurement of cell proliferation

MDA-MB-468 and MDA-MB-231 cells transfected with mock, HIST2H2BC, or SNRPEP4 siRNA (5×10^3 per well) were seeded in 96-well plates. After 6 hours, the cells were treated with indicated concentrations of paclitaxel. In parallel, 5×10^3 cells per well were seeded in 96-well plates and treated with PBS. Cell proliferation was determined from the metabolic reduction of WST-8 (Cell Counting Kit-8 cell proliferation assay) as described previously (21). Relative cell viability was calculated using the formula: each siRNA: OD of paclitaxel group/OD of non-paclitaxel group)/ negative control (NC): OD of paclitaxel group/OD of non-paclitaxel group.

Cell invasion assay

For the Boyden chamber invasion assay, cells were added to the top compartment of the chamber, and 800 μL of medium (containing 0.1% BSA) was added into the bottom chamber. Cells were incubated and allowed to migrate through Matrigel (BD Biosciences) for 24 hours. After removal of nonmigrated cells, cells that had migrated through the filter were counted.

Cell apoptosis and cycle arrest assay

For cell apoptosis and cycle arrest assay, 2×10^5 cells per well were seeded in 6-well plates and transfected with siRNA. After 48-hour transfection, cells were treated with 5 nmol/L paclitaxel for 16 hours. Cell apoptosis was evaluated using Alexa Fluor 488 Annexin V/Dead Cell Apoptosis Kit (Life Technologies) followed by flow cytometry according to the standard protocol. For cell-cycle arrest assay, cells were stained with propidium iodide and tested using cytometry (22).

Statistical analysis

All experiments were repeated at least three times. All numerical data were expressed as median ± IQR or mean ± SD. The data were analyzed using a two-sided Student *t* test or a one-way ANOVA test. We used a random variance model to pick out the differentially expressed RNAs between tumor samples and paired normal breast cancer samples. To develop the prognostic and predictive RNA signatures, univariate Cox proportional hazards regression model was performed, associating the RNA expression with the RFS time of patients in the training set. Cox regression coefficients and corresponding *P* values were determined for all tested RNAs. An mRNA-only and an integrated mRNA–lncRNA signature were developed on the basis of the coefficients of the candidate RNAs in the multivariate Cox proportional hazards regression model. To test whether the signature was an independent prognostic factor, the clinicopathologic factors, which were significantly associated with RFS in the univariate Cox proportional hazards regression model, were included in the multivariate analysis with each signature. All statistical analyses were performed with R software version 3.0.3 with two-tailed tests, and significance was defined with *P* values less than 0.05.

Study approval

Tissue samples of TNBC were obtained with approval of an independent ethical committee/institutional review board at FUSCC, Shanghai Cancer Center Ethical Committee (Shanghai, P.R. China), and informed consent from patients undergoing treatment in our cancer center.

Results

Patients with pathologically confirmed TNBC were included in the study according to the selection criteria. The baseline clinicopathologic characteristics are shown in Table 1. A total of 165 and 101 patients were recruited in the training and the validation set, respectively. The median follow-up time for each set was 13.9 [interquartile range (IQR) 8.6–21.1] and 18.5 (IQR 15–26.2) months, respectively.

Development of an mRNA-only and an integrated mRNA–lncRNA signature for TNBC patients

Using transcriptome microarray analysis of 33 paired tumor and normal breast samples, we identified 183 mRNAs and 231 lncRNAs that were differently expressed after adjustment using random variance model. Association between the expression of every mRNA or lncRNA and RFS of the patients was assessed in 165 TNBCs (training set; Supplementary Fig. S1). A total of eight mRNAs and two lncRNAs (Supplementary Table S1) were eligible for developing signatures after the stringent filtering procedure. Next, we examined the expression of the eight mRNAs and two lncRNAs using qRT-PCR in 165 TNBCs and 33 paired normal breast tissues. Assessing correlation between microarray and qRT-PCR data, we found that expression levels of all the RNAs were tightly associated between the two platforms (Supplementary Fig. S2). Expression of these mRNAs and lncRNAs measured by qRT-PCR was notably different between tumor and paired normal breast tissues, which were significantly correlated with their microarray data.

Using the coefficients from multivariate Cox proportional hazards model, we derived an mRNA-only signature based on the expression levels of the eight mRNAs determined by qRT-PCR in the training set of 165 TNBCs (23, 24). The formula is as follows: recurrence risk score (mRNA signature) = 0.877*ABCA8-2.553*CHRDL1+0.531*ADH1B-0.238*CDK1+0.086*CDC6+0.219*SQLE-1.14*FCGR1A+1.38*RSAD2. Adding two upregulated lncRNAs (HIST2H2BC and SNRPEP4) into the signature, we constructed an integrated mRNA–lncRNA signature

Table 1. Clinicopathologic characteristics of TNBC patients according to the integrated RNA signatures in two sets

		Training set					Validation set			
		mRNA signature		mRNA–lncRNA signature			mRNA signature		mRNA–lncRNA signature	
Characteristics	N	High risk (%)	Low risk (%)	High risk (%)	Low risk (%)	N	High risk (%)	Low risk (%)	High risk (%)	Low risk (%)
Age, y										
Median	55	50	56	50	56	53.5	53.5	53.5	54	53
IQR	46–61	43–58	49.8–62	43–59	49–61.3	44.8–59	45.5–58.3	44.8–60.3	46.5–59	44–59
≤50	68	38 (55.9)	30 (44.1)	34 (50.0)	34 (50.0)	39	19 (48.7)	20 (51.3)	20 (51.3)	19 (48.7)
>50	97	31 (32.3)	66 (68.8)	31 (32.3)	66 (68.8)	62	26 (41.9)	36 (58.1)	32 (51.6)	30 (48.4)
Menopause										
Yes	101	37 (36.6)	64 (63.4)	35 (34.7)	66 (65.3)	64	26 (40.6)	38 (59.4)	31 (48.4)	33 (51.6)
No	64	32 (50.0)	32 (50.0)	30 (46.9)	34 (53.1)	32	14 (43.8)	18 (56.3)	16 (50.0)	16 (50.0)
Unknown	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	5	5 (100.0)	0 (0.0)	5 (100.0)	0 (0.0)
Tumor size, cm										
≤2	58	23 (39.7)	35 (60.3)	19 (32.8)	39 (67.2)	36	17 (47.2)	19 (52.8)	19 (52.8)	17 (47.2)
>2	104	44 (42.3)	60 (57.7)	45 (43.3)	59 (56.7)	65	28 (43.1)	37 (56.9)	33 (50.8)	32 (49.2)
Unknown	3	2 (66.7)	1 (33.3)	1 (33.3)	2 (66.7)	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Tumor grade										
I–II	32	9 (28.1)	23 (71.9)	10 (31.3)	22 (68.8)	31	15 (48.4)	16 (51.6)	18 (58.1)	13 (41.9)
III	104	48 (46.2)	56 (53.8)	45 (43.3)	59 (56.7)	70	30 (42.9)	40 (57.1)	34 (48.6)	36 (51.4)
Unknown	29	12 (41.4)	17 (58.6)	10 (34.5)	19 (65.5)	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Positive LNs										
≤3	115	47 (40.9)	68 (59.1)	44 (38.3)	71 (61.7)	82	38 (46.3)	44 (53.7)	44 (53.7)	38 (46.3)
>3	50	22 (44.0)	28 (56.0)	21 (42.0)	29 (58.0)	19	7 (36.8)	12 (63.2)	8 (42.1)	11 (57.9)
Chemotherapy										
Taxane	124	52 (41.9)	72 (58.1)	48 (38.7)	76 (61.3)	66	29 (43.9)	37 (56.1)	34 (51.5)	32 (48.5)
Non-taxane	27	12 (44.4)	15 (55.6)	12 (44.4)	15 (55.6)	35	16 (45.7)	19 (54.3)	18 (51.4)	17 (48.6)
Unknown	14	5 (35.7)	9 (64.3)	5 (35.7)	9 (64.3)	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Radiotherapy										
Yes	50	23 (46.0)	27 (54.0)	21 (42.0)	29 (58.0)	39	19 (48.7)	20 (51.3)	21 (53.8)	18 (46.2)
No	103	41 (39.8)	62 (60.2)	41 (39.8)	62 (60.2)	42	18 (42.9)	24 (57.1)	21 (50.0)	21 (50.0)
Unknown	12	5 (41.7)	7 (58.3)	3 (25.0)	9 (75.0)	20	8 (40.0)	12 (60.0)	10 (50.0)	10 (50.0)
Follow-up time, mo										
Median	13.9	12.6	14.2	11.7	14.3	18.5	18.5	18.3	18	19.2
IQR	8.6–21.1	8.4–19.3	9.3–22.2	8.2–18.4	9.5–22.6	15–26.2	13.6–26.1	15.1–26.5	13.7–25.7	15.2–27.5
RFS event	22	17	5	17	5	9	7	2	8	1

Abbreviations: IQR, interquartile range; LN, lymph node; mo, months; RFS, recurrence-free survival; y, year.

using the same method, in which the recurrence risk score (integrated mRNA–lncRNA signature) is 0.939*ABCA8-2.593*CHRDL1+0.517*ADH1B-0.329*CDK1-0.071*CDC6+0.02*SQLE-1.146*FCGR1A+1.366*RSAD2+0.361*SNRPEP4+0.277*HIST2H2BC. We used cohort median expression levels to classify the expression level of the 10 RNAs included in the signatures: low expression status equalled 0 and high expression status equalled 1 in the formulas (21). According to the formulas, every patient in the training set received a score and was classified into high- or low-risk group based on the optimum cut-off scores from time-dependent ROC analysis.

Prognostic value of the multigene signatures in the training set

The prognostic value of the signatures was tested by using the multivariate Cox proportional hazards regression analyses (Table 2). According to the mRNA-only signature, patients in the high-risk group were more likely to suffer from recurrence than the low-risk group (HR = 4.46; 95% CI, 1.34–14.91, $P = 0.015$). However, other factors were not significantly associated with RFS in the multivariate analysis. Similar results were observed in the analysis for the integrated mRNA–lncRNA signature. Patients deemed as high-risk in the integrated signature had higher hazard of recurrence (HR = 10.00; 95% CI, 2.53–39.47; $P = 0.001$).

To test the performance of the signatures developed, we conducted time-dependent ROC analyses and calculated the AUC on both the signatures and traditional prognostic factors. For better comparison, we treated all factors as categorical variables. The time used in the analyses was set as 24 months for the limitation of the follow-up. Only tumor grade, number of positive lymph nodes, and the signatures could be regarded as significant prognostic factors with AUCs larger than 0.5 (Fig. 1). Our analysis showed that the integrated signature might have better prognostic value than the mRNA-only signature and the combined clinicopathologic factors in predicting 2-year RFS (AUC, 0.826 vs. 0.767 and 0.712).

Validation of the multigene signatures in the validation set

We validated the signatures in another independent cohort consisting of 101 TNBC patients. The mRNA–lncRNA expression profiles in this set were only examined by using qRT-PCR with GAPDH expression as reference. Each RNA was coded as high or low expression based on the median expression level, and the risk scores for each patient were calculated using the signatures developed in the training set. Then, patients with scores higher than the optimum cut-off scores, as determined by time-dependent ROC analysis, were allotted to the high-risk group, and the others to the low-risk group.

In the multivariate Cox proportional hazards regression model, the mRNA-only and the integrated signature were also significantly correlated with RFS (Table 2; HR = 6.31; 95% CI, 1.20–33.26, $P = 0.030$; HR = 14.04; 95% CI, 1.56–126.71, $P = 0.019$; for the mRNA-only and integrated mRNA–lncRNA signature, respectively). Two-year time-dependent ROC curves were also

Table 2. Multivariate Cox proportional hazards regression analysis of the derived RNA signatures and traditional characteristics with RFS

	mRNA-only signature				Integrated mRNA–lncRNA signature			
	Training set		Validation set		Training set		Validation set	
Variable[a]	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*
Age (≤50 y as reference)	1.06 (0.16–6.95)	0.949	1.29 (0.10–16.09)	0.845	1.01 (0.09–11.11)	0.995	1.18 (0.11–12.90)	0.890
Menopause (no as reference)	0.67 (0.12–3.73)	0.646	0.43 (0.03–5.77)	0.521	0.63 (0.07–5.38)	0.672	0.56 (0.05–6.18)	0.635
Tumor grade (≤II as reference)	1.61 (0.40–6.45)	0.501	0.71 (0.10–5.00)	0.729	1.79 (0.44–7.31)	0.416	0.91 (0.14–6.09)	0.920
Tumor size (≤2 cm as reference)	1.13 (0.34–3.70)	0.846	4.45 (0.48–41.47)	0.189	0.78 (0.20–2.99)	0.712	4.72 (0.51–43.68)	0.171
Positive LNs (≤3 as reference)	1.95 (0.53–7.24)	0.319	2.48 (0.53–11.53)	0.247	2.37 (0.61–9.16)	0.211	3.38 (0.70–16.31)	0.130
Radiotherapy (no as reference)	4.03 (0.93–17.39)	0.062	3.87 (0.64–23.49)	0.141	3.79 (0.83–17.31)	0.086	3.58 (0.63–20.33)	0.151
Chemotherapy (non-taxane as reference)	0.49 (0.15–1.57)	0.228	1.22 (0.25–5.95)	0.804	0.69 (0.17–2.74)	0.593	1.41 (0.27–7.26)	0.683
Signature (low-risk as reference)	4.46 (1.34–14.91)	0.015	6.31 (1.20–33.26)	0.030	10.00 (2.53–39.47)	0.001	14.04 (1.56–126.71)	0.019

Abbreviation: LN, lymph node.

[a]Adjusted by Cox proportional hazards models including age, menopausal status, tumor grade, tumor size, positive lymph nodes, radiotherapy, chemotherapy, and integrated RNA signature.

calculated. The signatures showed better performance in predicting 2-year RFS compared with traditional prognostic factors, and the integrated signature seemed to be the most superior (AUC = 0.714). Adding lncRNAs into the mRNA-only signature was shown to improve the accuracy of the signature in the validation set.

Predictive value of the signatures to taxane-based chemotherapy

We hypothesized that the signatures might also have predictive value in patient sensitivity to taxane-based chemotherapy. To validate this, we conducted interaction analysis in multivariate Cox regression model (Table 3). We assessed the interactions

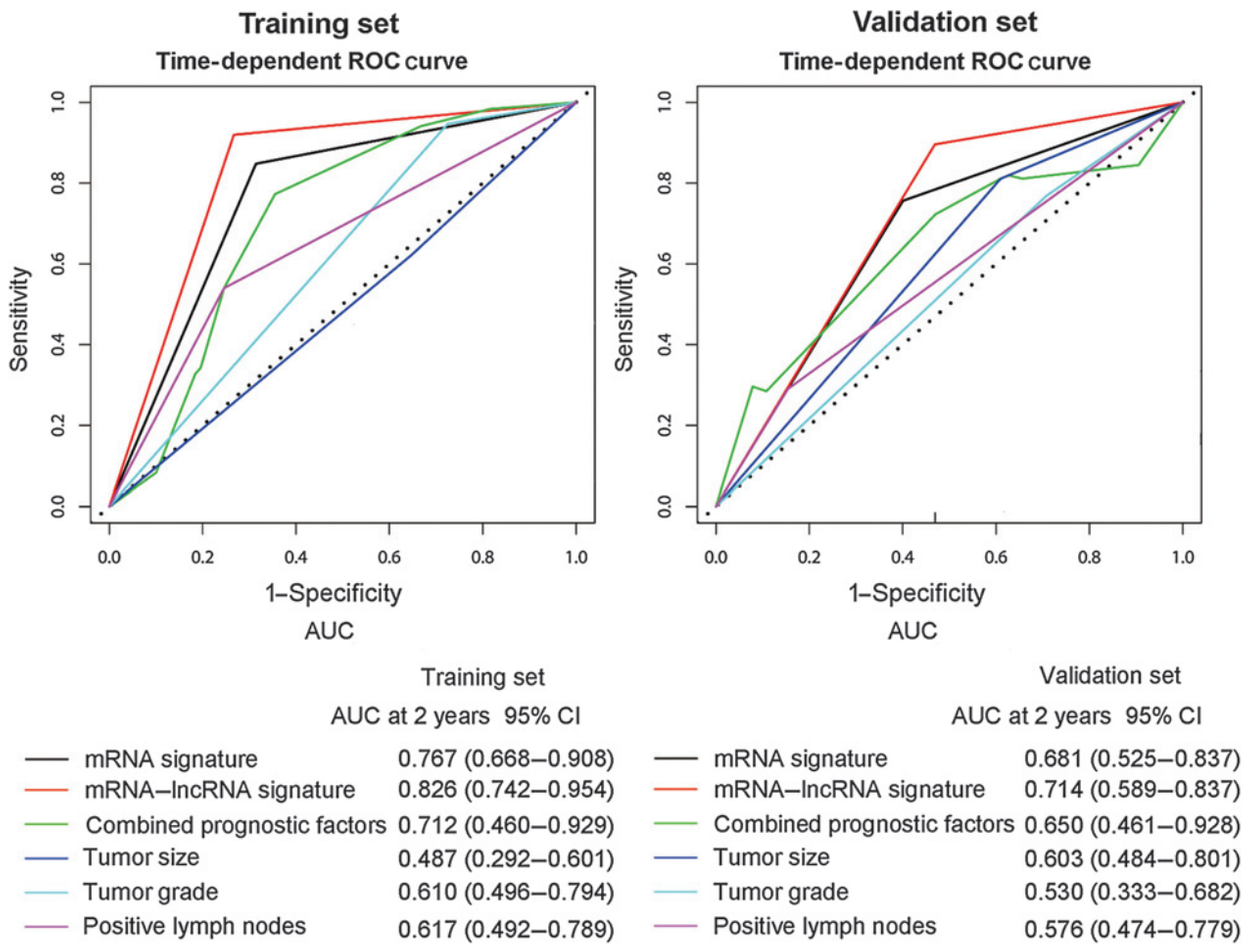

Figure 1.
Comparison of the sensitivity and specificity of prognosis by the integrated mRNA–lncRNA signature, the mRNA-only signature, and the traditional clinicopathologic factors in the training ($n = 165$) and validation ($n = 101$) sets. Time-dependent ROC curves were plotted to assess the efficacy of the signatures, with AUCs reported. We also calculated the performances of the combined traditional prognostic factors, including tumor size, grade, and number of involved lymph nodes. All factors were coded as categorical variables. Limited by the follow-up time, we could only perform analysis up to two years.

between each risk group and the taxane-based chemotherapy after adjusting the traditional clinicopathologic factors. For the mRNA-only signature, the interactions were not statistically significant in the training and validation sets (HR = 1.82; 95% CI, 0.62–5.37, $P = 0.277$; HR = 4.14; 95% CI, 0.93–18.48, $P = 0.063$, for the training and validation sets, respectively). For the integrated mRNA–lncRNA signature, the interaction was significantly associated with RFS in both the training and the validation sets (HR = 5.74; 95% CI, 1.54–21.33, $P = 0.009$; HR = 4.46; 95% CI, 1.00–19.88, $P = 0.050$, for the training and validation sets, respectively), and was further validated using the multivariate Cox proportional hazards regression analysis after stratifying according to the receipt of the taxane-based chemotherapy (Fig. 2). Collectively, these data implied that patients in the high-risk group, according to the integrated signature, benefited less from the taxane-based chemotherapy than patients in the low-risk group.

Biologic function of the lncRNAs incorporated in the integrated signature

We further explored the effect of the lncRNAs HIST2H2BC and SNRPEP4 on cell invasion, proliferation, and paclitaxel resistance (Fig. 3). For each lncRNA included in the signature, we designed three small double-strand interfering RNAs (siRNA), then selected the two with highest transfection efficiency (validated by qRT-PCR) for further study (data not shown). The downregulation of either of the lncRNA was significantly associated with decreased cell proliferation (MCF-7 cell line: HIST2H2BC; $P = 0.005$ and 0.003 for siRNA-1, siRNA-2, respectively; SNRPEP4: $P = 0.035$ and 0.043 for siRNA-1, siRNA-2, respectively; MDA-MB-231 cell line: $P < 0.001$ for both lncRNAs and siRNAs). Transwell Matrigel invasion assay revealed significant effect of both lncRNAs on cell invasion (HIST2H2BC: $P = 0.009$ and 0.001 for siRNA-1, siRNA-2, respectively; SNRPEP4: $P < 0.001$ for both siRNAs). The proliferation of MDA-MB-468 cells treated with paclitaxel was determined from the metabolic reduction of WST-8 (CCK-8 cell proliferation assay). After transfection with siRNA, cells were cultured with or without 5 nmol/L paclitaxel for 48 hours. Cells transfected with siRNAs were more sensitive to paclitaxel (HIST2H2BC: $P < 0.001$ for both siRNA-1, $P = 0.002$ for siRNA-2; SNRPEP4: $P = 0.086$ and 0.024 for siRNA-1 and siRNA-2, respectively). This may be partially explained by the effect of two lncRNAs on cell apoptosis and cell-cycle arrest (Supplementary Figs. S5 and S6). Taken together, these data suggested that the lncRNAs HIST2H2BC and SNRPEP4 promote cell proliferation and invasion and contribute to paclitaxel resistance in TNBC cells.

Discussion

The prognosis of TNBC patients is extremely heterogeneous and is rarely associated with the conventional prognostic parameters (patient age, tumor size, tumor grade, and lymph node status; ref. 25), a conclusion which was concordant with the results of this study. Approximately 30% of TNBC patients eventually experience relapse (1), while a substantial proportion of patients are overtreated with systemic adjuvant therapy. In this prospective observational study, we identified and independently validated prognostic and predictive RNA signatures for TNBC, which could be used to classify TNBC patients into high- or low-risk groups of recurrence. At the same time, patients predicted to have low recurrence risk would likely benefit more from the taxane-based adjuvant chemotherapy.

Comparing the two signatures, there are some pros and cons. For the integrated signature, it has better prognostic and predictive value, while the mRNA signature might be more applicable in clinical practice. Like the ER, PR, and HER2 markers currently used in the clinic, the mRNA signature could be easily applied via immunohistochemical method. Among the 8 mRNAs included in the signature, 5 mRNAs were significantly associated with RFS ($P < 0.05$). This implies that each of the 5 mRNA might be an individual prognostic factor for TNBC with less efficacy. These results were validated in both array and qRT-PCR platform. Of these 8 mRNAs, three have been previously studied regarding their potential role in cancer (*CDK1*, *CDC6*, and *SQLE*). Cdk1, a protein encoded by *CDK1*, is a catalytic subunit of the highly conserved protein kinase complex known as M-phase–promoting factor (MPF), which is essential for the G_1–S and G_2–M phase transitions of the eukaryotic cell cycle. Cdk1 plays an important role in multiple processes during mitosis. In breast cancer, previous studies reported that high Cdk1 activity predicts poor survival, suppresses DNA damage response, promotes tumorigenesis, and controls Fas-mediated apoptosis (14, 26–29). Also Cdc6, coded by *CDC6*, is an essential regulator of DNA replication and maintenance of the checkpoint mechanism in the cell cycle. The expression of *CDC6* has been proved to be associated with the survival of breast cancer patients, grade of breast cancer, and response to methionine stress (30–33). Another mRNA is *SQLE*, which catalyses the first oxygenation step in sterol biosynthesis. Helms and colleagues found out that *SQLE* mRNA expression might indicate high-risk ER^+ stage I/II breast cancers (34). The other five mRNAs' relationships with cancer, especially breast cancer, have not been reported until now, and future research will be needed to clarify their potential function in breast cancer. Collectively, after rigorous selection and comprehensive validation process, the signatures we developed could successfully classify TNBC patients into high- and low-risk groups, indicating that they may serve as potential prognostic markers.

In a study by Su and colleagues, breast cancer was classified into four subtypes based on lncRNA profile using The Cancer Genome Atlas data, and the first lncRNA subtype of breast cancer was

Table 3. Multivariate Cox proportional hazards regression analysis of RFS, including interaction of signatures with adjuvant chemotherapy

	mRNA signature				mRNA–lncRNA signature			
	Training set		Validation set		Training set		Validation set	
Variable[a]	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*
Chemotherapy (non-taxane as reference)	0.49 (0.15–1.57)	0.228	1.22 (0.25–5.95)	0.804	0.69 (0.17–2.74)	0.593	1.41 (0.27–7.26)	0.683
Signature (low-risk as reference)	4.46 (1.34–14.91)	0.015	6.31 (1.20–33.26)	0.030	10.00 (2.53–39.47)	0.001	14.04 (1.56–126.71)	0.019
Interaction	1.82 (0.62–5.37)	0.277	4.14 (0.93–18.48)	0.063	5.74 (1.54–21.33)	0.009	4.46 (1.00–19.88)	0.05

[a]Adjusted by Cox proportional hazards models including clinical variables as categorized in Table 2. Here, we present only three items: chemotherapy, signature, and the interaction between them. Other parameters (age, menopausal status, tumor grade, tumor size, positive lymph nodes and radiotherapy) are not shown.

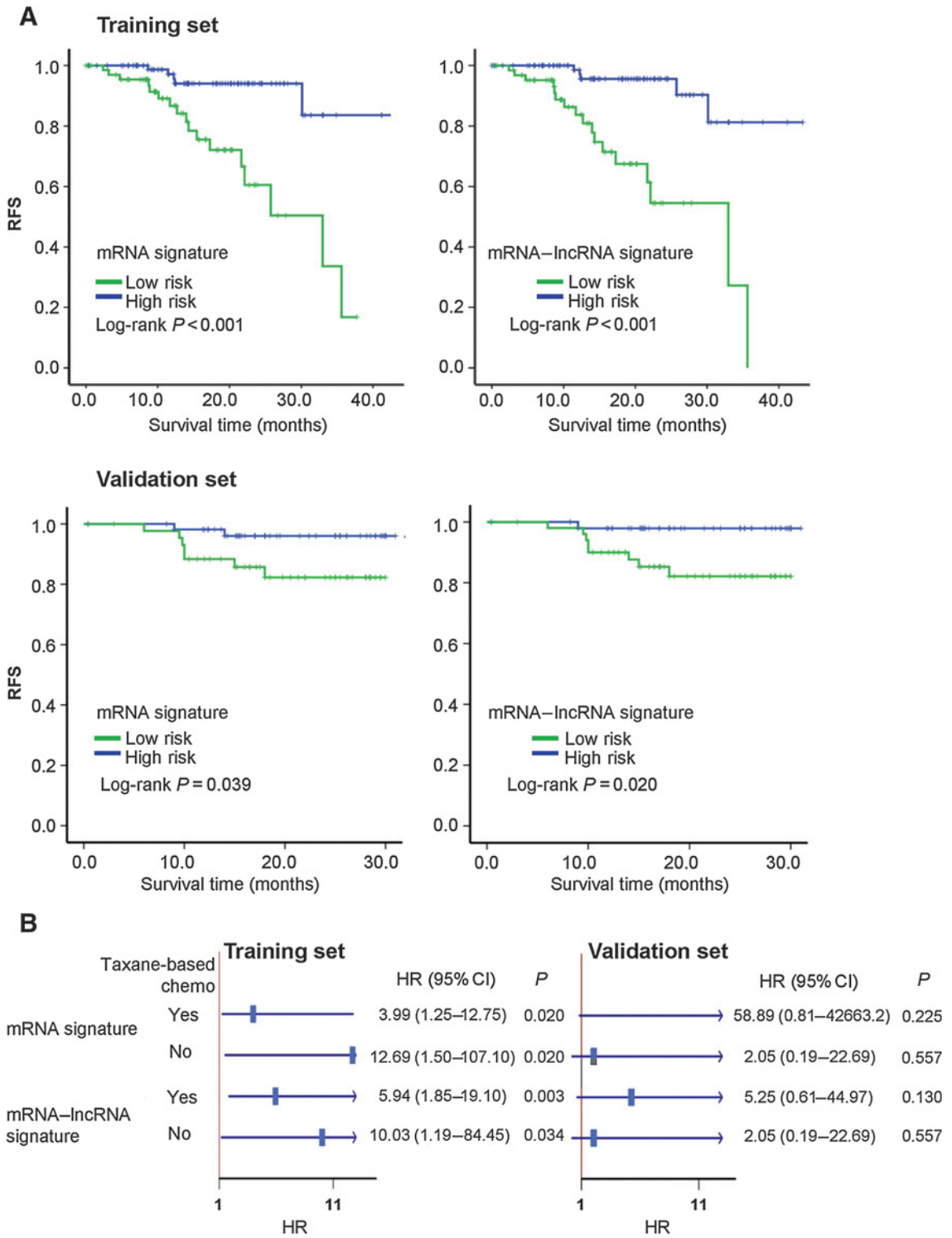

Figure 2.
Estimates of RFS according to the scores calculated by the integrated mRNA–lncRNA signature and the mRNA-only signature in the training ($n = 165$) and the validation ($n = 101$) sets. A, Kaplan–Meier analysis of RFS according to the scores calculated by the integrated mRNA–lncRNA signature and the mRNA-only signature. B, patients were stratified according to the receipt of the taxane-based chemotherapy to validate the interaction between each risk group and the taxane-based chemotherapy using multivariate Cox proportional hazards regression analysis adjusted for clinicopathologic factors in Table 2. The low-risk group was used as reference. The bars represent the HRs in different set and the lines represent the 95% CI. If the limit of 95% CI outranges the scale on the *x*-axis, the data were shown as arrows.

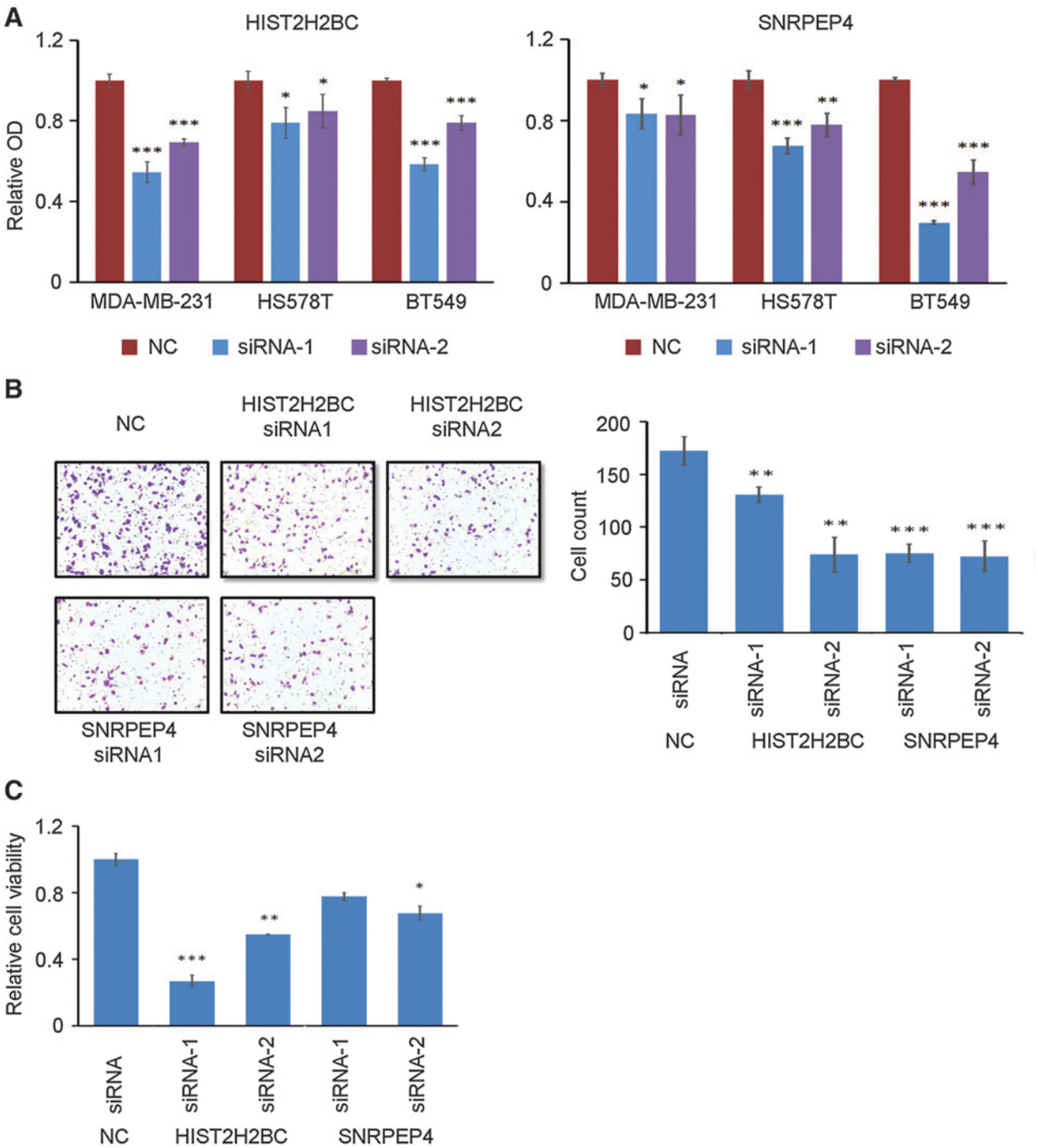

Figure 3.
Biologic function of the lncRNAs HIST2H2BC and SNRPEP4 incorporated in the integrated signature. A, cell proliferation was determined by CCK-8 assay after transfection with siRNAs for 48 hours. The results are shown as the percentage of optical density (OD) with negative control (NC) as reference. B, representative light microscopic images of migrated cells through the Transwell chamber (magnification, 100×). The number of migrated cells was calculated and compared between each siRNA with NC. C, effect of lncRNAs on the resistance to paclitaxel. The results were assessed by CCK-8 assay and the relative cell viability was calculated. All results are represented as the mean ± SD from three independent experiments. Notes: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

proposed (15). After unsupervised hierarchical consensus clustering and comparison, they found that cluster I was highly correlated with the basal-like subtype. In another study (35), also based on The Cancer Genome Atlas data, Yan and colleagues comprehensively analyzed lncRNA alterations at transcriptional, genomic, and epigenetic levels across 13 human cancer types, and found several dysregulated lncRNAs. We did not find our two lncRNAs (HIST2H2BC and SNRPEP4) in Su and colleagues' list of

basal-like enriched lncRNAs, nor in Yan and colleagues' list of breast cancer related lncRNAs. As both of the two studies were based on The Cancer Genome Atlas data, we think the reason for not finding our lncRNAs in their lists might be differences in inclusion criteria and technology platforms. In our study, we focused exclusively on the TNBC subtype, and compared expression difference between normal tissue and TNBC breast cancer, but not among different subtypes of breast cancer. Also, lncRNAs correlated with RFS were selected as candidates for further analysis. Furthermore, our preliminary data show that these two lncRNAs promoted the proliferation and invasion of TNBC cells and contributed to resistance towards paclitaxel, which partially explained their roles in TNBC progression and treatment. These results add more evidence to the predictive and prognostic value of the integrated mRNA–lncRNA signature. Further investigation into their functions may provide additional targets and strategies for treatment.

Our study has several limitations. First, the median follow-up time of the prospective observational study is relatively short, and may be not enough to reveal all patients in high-risk, thus could underestimate their recurrence risk and impair the efficacy of the signatures. We will continue to follow the patients in both cohorts and keep updating the signatures assessment. Second, the microarray/qRT-PCR–based platform is difficult to apply in routine clinical practice. Therefore, our future work will focus on developing simpler sampling strategies and high-throughput–selected reaction monitoring assay to reliably measure the integrated signature.

In summary, we have developed multigene signatures integrating coding and noncoding RNAs for predicting disease recurrence and the benefit of taxane chemotherapy in TNBC patients. Future prospective clinical trials are needed to further consolidate the validity of the signatures.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Disclaimer

The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

Conception and design: Y.-R. Liu, X. Hu, K.-D. Yu, Z.-M. Shao
Development of methodology: Y.-R. Liu, Y.-Z. Jiang, X.-E. Xu, K.-D. Yu, Z.-M. Shao
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): Y.-R. Liu, Y.-Z. Jiang, X.-E. Xu, X. Hu
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Y.-R. Liu, Y.-Z. Jiang, X.-E. Xu, K.-D. Yu
Writing, review, and/or revision of the manuscript: Y.-R. Liu, Y.-Z. Jiang, X.-E. Xu, X. Hu, K.-D. Yu, Z.-M. Shao
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): X.-E. Xu, X. Hu, K.-D. Yu, Z.-M. Shao
Study supervision: K.-D. Yu, Z.-M. Shao

Grant Support

This work was supported by grants from the Research Project of Fudan University Shanghai Cancer Center (YJ201401), the National Natural Science Foundation of China (81572583, 81502278, 81372848, 81370075), the Municipal Project for Developing Emerging and Frontier Technology in Shanghai Hospitals (SHDC12010116), the Cooperation Project of Conquering Major Diseases in Shanghai Municipality Health System (2013ZYJB0302), the Innovation Team of Ministry of Education (IRT1223), and the Shanghai Key Laboratory of Breast Cancer (12DZ2260100).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 30, 2015; revised November 5, 2015; accepted November 12, 2015; published OnlineFirst January 26, 2016.

References

1. 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.
2. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011;121:2750–67.
3. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 2006;98:262–72.
4. 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.
5. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.
6. Rody A, Karn T, Liedtke C, Pusztai L, Ruckhaeberle E, Hanker L, et al. A clinically relevant gene signature in triple negative and basal-like breast cancer. Breast Cancer Res 2011;13:R97.
7. Mourtada-Maarabouni M, Pickard MR, Hedge VL, Farzaneh F, Williams GT. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. Oncogene 2009;28:195–208.
8. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 2010;464:1071–6.
9. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. Cancer Discov 2011;1:391–407.
10. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell 2011;43:904–14.
11. Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. J Clin Oncol 2007;25:118–45.
12. Hammond MEH, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol 2010;28:2784–95.
13. Jiang YZ, Yu KD, Bao J, Peng WT, Shao ZM. Favorable prognostic impact in loss of TP53 and PIK3CA mutations after neoadjuvant chemotherapy in breast cancer. Cancer Res 2014;74:3399–407.
14. Zhang W, Peng G, Lin SY, Zhang P. DNA damage response is suppressed by the high cyclin-dependent kinase 1 activity in mitotic mammalian cells. J Biol Chem 2011;286:35899–905.
15. Su X, Malouf GG, Chen Y, Zhang J, Yao H, Valero V, et al. Comprehensive analysis of long non-coding RNAs in human breast cancer clinical subtypes. Oncotarget 2014;5:9864–76.
16. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia 2007;9:166–80.
17. Paralkar VR, Weiss MJ. A new 'Linc' between noncoding RNAs and blood development. Genes Dev 2011;25:2555–8.

18. Banfai B, Jia H, Khatun J, Wood E, Risk B, Gundling WE Jr, et al. Long noncoding RNAs are rarely translated in two human cell lines. Genome Res 2012;22:1646–57.
19. Wu SC, Kallin EM, Zhang Y. Role of H3K27 methylation in the regulation of lncRNA expression. Cell Res 2010;20:1109–16.
20. Jiang YZ, Yu KD, Peng WT, Di GH, Wu J, Liu GY, et al. Enriched variations in TEKT4 and breast cancer resistance to paclitaxel. Nat Commun 2014; 5:3802.
21. Zhang JX, Song W, Chen ZH, Wei JH, Liao YJ, Lei J, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. Lancet Oncol 2013;14:1295–306.
22. Krishan A. Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining. J Cell Biol 1975;66:188–93.
23. Yu SL, Chen HY, Chang GC, Chen CY, Chen HW, Singh S, et al. MicroRNA signature predicts survival and relapse in lung cancer. Cancer Cell 2008; 13:48–57.
24. Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D, et al. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med 2004;350:1828–37.
25. Rakha EA, El-Sayed ME, Green AR, Lee AH, Robertson JF, Ellis IO. Prognostic markers in triple-negative breast cancer. Cancer 2007;109:25–32.
26. Cepeda D, Ng HF, Sharifi HR, Mahmoudi S, Cerrato VS, Fredlund E, et al. CDK-mediated activation of the SCF(FBXO) (28) ubiquitin ligase promotes MYC-driven transcription and tumourigenesis and predicts poor survival in breast cancer. EMBO Mol Med 2013;5:999–1018.
27. Matthess Y, Raab M, Sanhaji M, Lavrik IN, Strebhardt K. Cdk1/cyclin B1 controls Fas-mediated apoptosis by regulating caspase-8 activity. Mol Cell Biol 2010;30:5726–40.
28. Paruthiyil S, Cvoro A, Tagliaferri M, Cohen I, Shtivelman E, Leitman DC. Estrogen receptor beta causes a G2 cell cycle arrest by inhibiting CDK1 activity through the regulation of cyclin B1, GADD45A, and BTG2. Breast Cancer Res Treat 2011;129:777–84.
29. Suryadinata R, Sadowski M, Steel R, Sarcevic B. Cyclin-dependent kinase-mediated phosphorylation of RBP1 and pRb promotes their dissociation to mediate release of the SAP30.mSin3.HDAC transcriptional repressor complex. J Biol Chem 2011;286:5108–18.
30. Booher K, Lin DW, Borrego SL, Kaiser P. Downregulation of Cdc6 and pre-replication complexes in response to methionine stress in breast cancer cells. Cell Cycle (Georgetown, Tex) 2012;11:4414–23.
31. Buechler S. Low expression of a few genes indicates good prognosis in estrogen receptor positive breast cancer. BMC Cancer 2009; 9:243.
32. Lopez FJ, Cuadros M, Cano C, Concha A, Blanco A. Biomedical application of fuzzy association rules for identifying breast cancer biomarkers. Med Biol Eng Comput 2012;50:981–90.
33. Stevens KN, Wang X, Fredericksen Z, Pankratz VS, Cerhan J, Vachon CM, et al. Evaluation of associations between common variation in mitotic regulatory pathways and risk of overall and high grade breast cancer. Breast Cancer Res Treat 2011;129:617–22.
34. Helms MW, Kemming D, Pospisil H, Vogt U, Buerger H, Korsching E, et al. Squalene epoxidase, located on chromosome 8q24.1, is upregulated in 8q+ breast cancer and indicates poor clinical outcome in stage I and II disease. Br J Cancer 2008;99:774–80.
35. Yan X, Hu Z, Feng Y, Hu X, Yuan J, Zhao SD, et al. Comprehensive genomic characterization of long non-coding RNAs across human cancers. Cancer Cell 2015;28:529–40.