



Research Paper

Prognostic Value of a BCSC-associated MicroRNA Signature in Hormone Receptor-Positive HER2-Negative Breast Cancer



Chang Gong^{a,1}, Weige Tan^{a,2}, Kai Chen^{a,3}, Na You^b, Shan Zhu^b, Gehao Liang^a, Xinhua Xie^c, Qian Li^a, Yunjie Zeng^d, Nengtai Ouyang^d, Zhihua Li^e, Musheng Zeng^c, ShiMei Zhuang^f, Wan-Yee Lau^{g,h}, Qiang Liu^a, Dong Yin^a, Xueqin Wang^b, Fengxi Su^{a,**}, Erwei Song^{a,i,*}

^a Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Breast Tumor Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China

^b Department of Statistical Science, School of Mathematics and Computational Science & Southern China Research Center of Statistical Science, Sun Yat-sen University, Guangzhou 510275 China

^c Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, 510060, China

^d Department of Pathology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China

^e Prevention and Cure Center of Breast Disease, Key Laboratory of Breast Disease, the Third Hospital of Nanchang City, Nanchang, Jiangxi 330009, China

^f Key Laboratory of Gene Engineering of Ministry of Education, Collaborative Innovation Center for Cancer Medicine, School of Life Science, Sun Yat-sen University, Guangzhou 510275, China

^g Faculty of Medicine, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong, China

^h Department of Hepatobiliary Surgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China

ⁱ Collaborative Innovation Center for Cancer Medicine, China

ARTICLE INFO

Article history:

Received 4 April 2016

Received in revised form 10 August 2016

Accepted 10 August 2016

Available online 17 August 2016

Keywords:

Breast cancer stem cell

miRNA

Biology-driven approach

Classifier

Prognosis

ABSTRACT

Purpose: Breast cancer patients with high proportion of cancer stem cells (BCSCs) have unfavorable clinical outcomes. MicroRNAs (miRNAs) regulate key features of BCSCs. We hypothesized that a biology-driven model based on BCSC-associated miRNAs could predict prognosis for the most common subtype, hormone receptor (HR)-positive, HER2-negative breast cancer patients.

Patients and Methods: After screening candidate miRNAs based on literature review and a pilot study, we built a miRNA-based classifier using LASSO Cox regression method in the training group ($n = 202$) and validated its prognostic accuracy in an internal ($n = 101$) and two external validation groups ($n = 308$).

Results: In this multicenter study, a 10-miRNA classifier incorporating miR-21, miR-30c, miR-181a, miR-181c, miR-125b, miR-7, miR-200a, miR-135b, miR-22 and miR-200c was developed to predict distant relapse free survival (DRFS). With this classifier, HR + HER2 – patients were scored and classified into high-risk and low-risk disease recurrence, which was significantly associated with 5-year DRFS of the patients. Moreover, this classifier outperformed traditional clinicopathological risk factors, IHC4 scoring and 21-gene Recurrence Score (RS). The patients with high-risk recurrence determined by this classifier benefit more from chemotherapy.

Conclusions: Our 10-miRNA-based classifier provides a reliable prognostic model for disease recurrence in HR + HER2 – breast cancer patients. This model may facilitate personalized therapy-decision making for HR + HER2 – individuals.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Abbreviations: BCSCs, Breast cancer stem cells; miRNAs, MicroRNAs; DRFS, Distant relapse free survival; IHC, Immunohistochemistry; ER, Estrogen receptor; HR, Hormone receptor; CSCs, Cancer stem cells; RS, Recurrence score; FFPE, Formalin-fixed paraffin-embedded; SYSMH, Sun Yat-sen Memorial Hospital; IRB, Institutional review board; LASSO, Least Absolute Shrinkage and Selection Operator; ROC, Receiver operating characteristic; AUC, Area under curve; EMT, Epithelial-mesenchymal transition; PR, Progesterone receptor; HER2, Human epidermal growth factor receptor 2; BCS, Breast conserving surgery; ET, Endocrine therapy; TAM → AI, Tamoxifen followed by aromatase inhibitor.

* Correspondence to: E. Song, Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Breast Tumor Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China, Collaborative Innovation Center for Cancer Medicine, China.

** Correspondence to: F. Su, Breast Tumor Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China.

E-mail addresses: fengxisu@vip.163.com (F. Su), songerwei02@aliyun.com (E. Song).

¹ Chang Gong, Weige Tan and Kai Chen contributed equally to this work.

² Chang Gong, Weige Tan and Kai Chen contributed equally to this work.

³ Chang Gong, Weige Tan and Kai Chen contributed equally to this work.

1. Introduction

Hormone receptor (HR)-positive, HER2-negative breast cancer accounts for approximately 60% of all primary breast cancer cases (Sorlie et al., 2003), and around 20% of early-stage estrogen receptor (ER)-positive patients may develop local or distant recurrences after treatment. HR-positive, HER2-negative breast cancer patients usually have a lower risk of tumor recurrence, and thus some of them cannot benefit from cytotoxic chemotherapies (Ring et al., 2004) leading to overtreatment of the patients. Therefore, a prognostic and predictive model for HR-positive, HER2-negative breast cancer is needed in clinical practice. To meet this end, gene assays, such as Oncotype Dx and PAM50, have been developed and validated in multiple clinical trial. However, these models employed computer-based algorithms to select candidate genes for investigation without considering their biological rationales. A hypothesis-driven approach involving factors in DNA repair pathways has been used in a scoring system for ovarian cancer patients treated with platinum-based chemotherapies, which effectively predicted their clinical outcomes (Kang et al., 2012; Paik et al., 2004). Thus, selecting target genes for

molecular signatures of prognosis and therapeutic prediction by weighing their biological rationales emerges as an effective and economic approach.

Cancer stem cells (CSCs) are a subpopulation of tumor cells with stem cell like features in solid malignancies, which play an important role in cancer recurrence and metastasis. Breast cancer patients with an elevated proportion of cancer stem cells identified by immunostaining for markers of breast cancer stem cell (BCSC), including ALDH1 and CD44^{high}CD24^{low}, reveal unfavorable clinical outcome and poor survival (Dai et al., 2012; Ginestier et al., 2007; Gong et al., 2010). However, the lack of specificity and limited number of BCSC markers restrict their application as effective biomarkers in clinical practice. Moreover, most of these markers do not actually reflect the functional features of BCSCs. On the other hand, our previous study demonstrated that BCSCs express a unique profile of microRNAs (miRNAs), and the deregulated miRNAs play a crucial role in governing BCSC biology (Dalerba and Clarke, 2013; Dirks, 2009). These BCSC-associated miRNAs may function as oncogenes or tumor suppressor genes to regulate self-renewal, anti-apoptosis, invasiveness, transdifferentiation into vascular endothelial cells and chemotherapeutic resistance of BCSCs, and thus contribute to progression and recurrence of

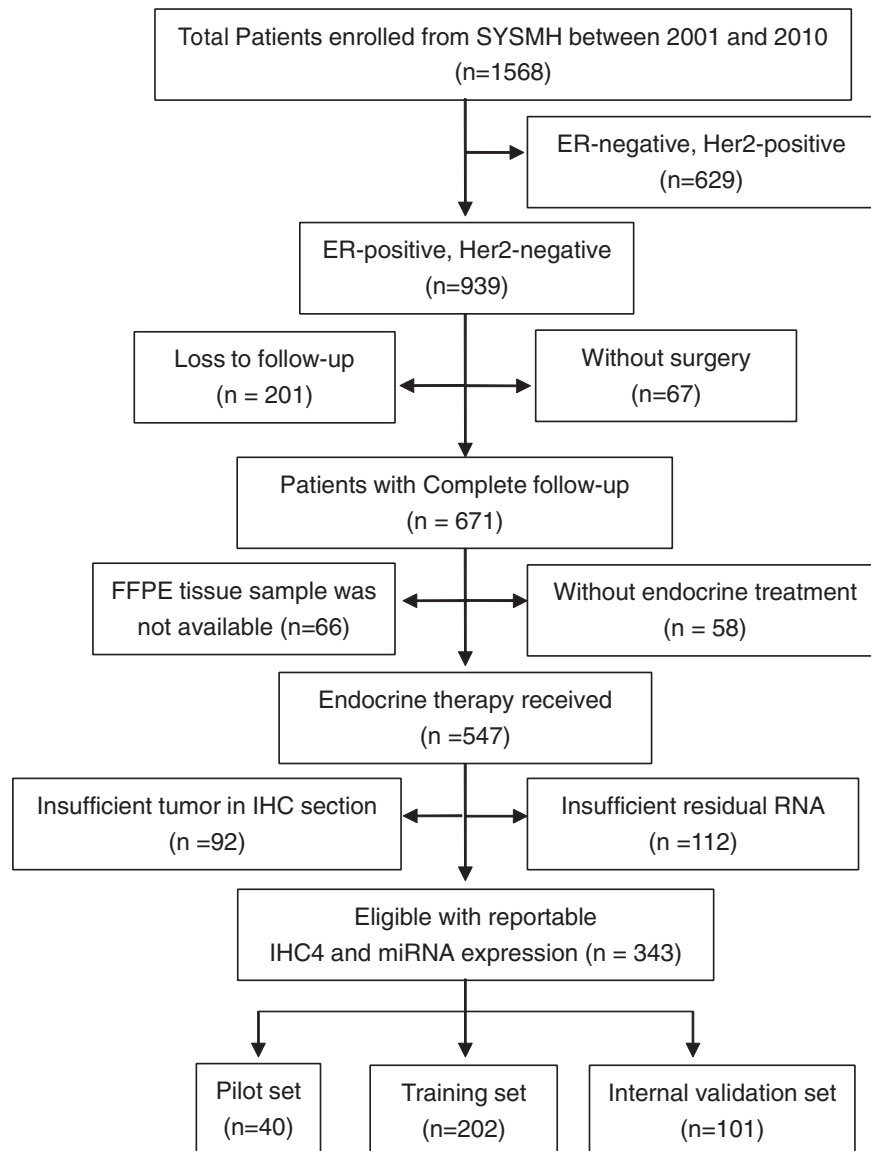


Fig. 1. Consort diagram for the availability of samples for analysis. SYSMH = Sun Yat-sen Memorial Hospital; ER = estrogen receptor; PR = progesterone receptor; HER2 = human epidermal growth factor receptor 2; and IHC = immunohistochemistry.

Table 1
Clinical characteristics of patients according to the 10-miRNA classifier in the training, internal and two external validation groups.

Variables		Training group ^a				Internal validation group ^a				External validation group-1 ^b				External validation group-2 ^c			
		Total	Low risk (%)	High risk (%)	P value	Total	Low risk (%)	High risk (%)	P value	Total	Low risk (%)	High risk (%)	P value	Total	Low risk (%)	High risk (%)	P value
Age	≤40 years	17	12 (70.6%)	5 (29.4%)	0.299	16	10 (62.5%)	6 (37.5%)	0.003	26	18 (69.2%)	8 (30.8%)	0.571	27	24 (88.9%)	3 (11.1%)	0.139
	>40 years	185	150 (81.1%)	35 (18.9%)		85	77 (90.6%)	8 (9.4%)		126	94 (74.6%)	32 (25.4%)		129	98 (76.0%)	31 (24.0%)	
Menopause	Yes	127	98 (77.2%)	29 (22.8%)	0.201	62	53 (85.5%)	9 (14.5%)	0.81	83	54 (65.1%)	29 (34.9%)	0.01	101	81 (80.2%)	20 (19.8%)	0.424
	No	75	64 (85.3%)	11 (14.7%)		39	34 (87.2%)	5 (12.8%)		69	58 (84.1%)	11 (15.9%)		55	41 (74.5%)	14 (25.5%)	
Tumor size	≤2 cm	72	61 (84.7%)	11 (15.3%)	0.23	35	34 (97.1%)	1 (2.9%)	0.02	33	26 (78.8%)	7 (21.2%)	0.452	39	34 (87.2%)	5 (12.8%)	0.117
	>2 cm	130	101 (77.7%)	29 (22.3%)		66	53 (80.3%)	13 (19.7%)		119	86 (72.3%)	33 (27.7%)		117	88 (75.2%)	29 (24.8%)	
Lymph node	Negative	97	88 (90.7%)	9 (9.3%)	<0.0001	57	57 (100%)	0 (0.0%)	<0.0001	85	71 (83.5%)	14 (16.5%)	0.002	46	41 (89.1%)	5 (10.9%)	0.033
	Positive	105	74 (70.5%)	31 (29.5%)		44	30 (68.2%)	14 (31.8%)		67	41 (61.2%)	26 (38.8%)		110	81 (73.6%)	29 (26.4%)	
TNM stage	I	45	43 (95.6%)	2 (4.4%)	<0.0001	26	26 (100%)	0 (0.0%)	<0.0001	22	20 (90.9%)	2 (9.1%)	0.05	35	29 (82.9%)	6 (17.1%)	0.006
	II	100	87 (87.0%)	13 (13.0%)		53	52 (98.1%)	1 (1.9%)		90	67 (74.4%)	23 (25.6%)		69	60 (87.0%)	9 (13.0%)	
	III	57	32 (56.1%)	25 (43.9%)		22	9 (40.9%)	13 (59.1%)		40	25 (62.5%)	15 (37.5%)		52	33 (63.5%)	19 (36.5%)	
Tumor grade	I	25	23 (92.0%)	2 (8.0%)	0.114	16	16 (100%)	0 (0.0%)	0.08	13	9 (69.2%)	4 (30.8%)	0.703	10	9 (90.0%)	1 (10.0%)	0.35
	II-III	177	139 (78.5%)	38 (21.5%)		85	71 (83.5%)	14 (16.5%)		139	103 (74.1%)	36 (25.9%)		146	113 (77.4%)	33 (22.6%)	
PR	Negative	75	63 (84.0%)	12 (16.0%)	0.362	43	36 (83.7%)	7 (16.3%)	0.572	59	41 (69.5%)	18 (30.5%)	0.45	59	46 (78.0%)	13 (22.0%)	0.955
	Positive	127	99 (78.0%)	28 (22.0%)		58	51 (87.9%)	7 (12.1%)		93	71 (76.3%)	22 (23.7%)		97	76 (78.4%)	21 (21.6%)	
Ki67	≤14%	74	63 (85.1%)	11 (14.9%)	0.181	40	38 (95%)	2 (5.0%)	0.037	84	67 (79.8%)	17 (20.2%)	0.059	73	63 (86.3%)	10 (13.7%)	0.022
	>14%	128	99 (77.3%)	29 (22.7%)		61	49 (80.3%)	12 (19.7%)		68	45 (66.2%)	23 (33.8%)		83	59 (71.1%)	24 (28.9%)	
Surgery	Mastectomy	112	89 (79.5%)	23 (20.5%)	0.77	60	49 (81.7%)	11 (18.3%)	0.116	142	106 (74.6%)	36 (25.4%)	0.309	147	114 (77.6%)	33 (22.4%)	0.508
	BCS	90	73 (81.1%)	17 (18.9%)		41	38 (92.7%)	3 (7.3%)		10	6 (60.0%)	4 (40%)		8	7 (87.5%)	1 (12.5%)	
Chemotherapy	No	43	32 (74.4%)	11 (25.6%)	0.284	17	13 (76.5%)	4 (23.5%)	0.206	44	30 (68.2%)	14 (31.8%)	0.325	38	29 (76.3%)	9 (23.7%)	0.746
	Yes	159	130 (81.8%)	29 (18.2%)		84	74 (88.1%)	10 (11.9%)		108	82 (75.9%)	26 (24.1%)		118	93 (78.8%)	25 (21.2%)	
ET	TAM	101	79 (78.2%)	22 (21.8%)	0.775	65	54 (83.1%)	11 (16.9%)	0.396	81	63 (77.8%)	18 (22.2%)	0.214	87	67 (77.0%)	20 (23.0%)	0.764
	AI	89	73 (82.0%)	16 (18.0%)		30	28 (93.3%)	2 (6.7%)		63	42 (66.7%)	21 (33.3%)		58	47 (81.0%)	11 (19.0%)	
	TAM → AI	12	10 (83.3%)	2 (16.7%)		6	5 (83.3%)	1 (16.7%)		8	7 (87.5%)	1 (12.5%)		11	8 (72.7%)	3 (27.3%)	
IHC4 score	Low risk	36	31 (86.1%)	5 (13.9%)	0.367	25	23 (92.0%)	2 (8%)	0.358	53	45 (84.9%)	8 (15.1%)	0.015	43	38 (88.4%)	5 (11.6%)	0.124
	Median risk	127	98 (77.2%)	29 (22.8%)		59	51 (86.4%)	8 (13.6%)		49	37 (75.5%)	12 (24.5%)		61	47 (77.0%)	14 (23.0%)	
	High risk	39	33 (84.6%)	6 (15.4%)		17	13 (76.5%)	4 (23.5%)		50	30 (60.0%)	20 (40.0%)		52	37 (71.2%)	15 (28.8%)	
Total		202	162 (80.2%)	40 (19.8%)		101	87 (86.1%)	14 (13.9%)		152	112 (73.7%)	40 (26.3%)		156	122 (78.2%)	34 (21.8%)	

Data are mean (SD) or n (%), unless otherwise stated; P value is calculated by χ^2 test or Fisher's exact test; a: Cohort 1 includes training group and internal validation group; b: external validation group-1 = Cohort 2; c: external validation group-2 = Cohort 3; PR: progesterone receptor; BCS: breast conserving surgery; ET: endocrine therapy; TAM → AI: tamoxifen followed by aromatase inhibitor.

the tumor. Indeed, an aberrant expression of BCSC-associated miRNAs has been related to patient prognosis and therapeutic (Arigoni et al., 2013; Yu et al., 2007). Therefore, studying the expression profiling of BCSC-associated miRNAs may provide a hypothesis-driven approach to construct a molecular signature with clear functional rationale to predict the prognosis and therapeutic benefit of breast cancer patients.

In the present study, we developed a 10 BCSC-associated miRNA-based classifier, to divide the HR-positive, HER2-negative breast cancer patients into high- and low-risk groups. We performed a multi-institutional study to both internally and externally validate this classifier as prognostic biomarkers of distant relapse-free survival (DRFS) for HR + HER – patients. Additionally, its prognostic efficacy was further compared with each single miRNA, clinicopathological risk factors and IHC4 scoring.

2. Patients and Methods

2.1. Patients and Clinical Database

We reviewed the 1568 breast cancer patients who were admitted in Sun Yat-sen Memorial Hospital (SYSMH) between June 1, 2000 and May 31, 2010, and enrolled 343 eligible patients according to the inclusion and exclusion criteria in this study. Please refer to the Supplementary material for the details of eligibility criteria. Sample availability is shown in the consort diagram in Fig. 1. In addition to 40 patients for pilot study, the remaining 303 eligible samples (Cohort 1) are randomly divided into two groups for the model training ($n = 202$) and internal validation ($n = 101$), stratified by the distant recurrence events. The patients who enrolled in the Cancer Center of Sun Yat-sen University (Cohort 2, $n = 152$) and the Third Hospital of Nanchang City (Cohort 3, $n = 156$) during the same period are selected to match the clinicopathological characteristics of Cohort 1, and used for external validation. Among all the HR-positive, lymph node-negative breast cancer patients enrolled in this study, 21-gene assay (Surexam®, Guangzhou, China) (Paik et al., 2004; Zhang et al., 2015) was performed in 153 cases to generate a 21-gene Recurrence Score (21-gene RS) in the paraffin-embedded tumor tissue samples and the results were compared with those generated with our 10 BCSC-associated miRNA classifier. This was a retrospective study with de-identified information for all participants. Therefore, informed consents from patients were not needed and the institutional review board (IRB) reviews were waived based on the institutional policy.

2.2. Identification of BCSC-associated miRNAs

A total of 35 miRNAs were identified BCSC associated and constantly detectable for the further study. The detailed identification procedure of BCSC-associated miRNAs was referred in Fig. S1 and Supplementary materials.

2.3. Construction of a BCSC-associated miRNA classifier

qRT-PCR was used to examine the expression level of 35 BCSC-associated miRNAs according to a standard protocol as described in the Supplementary Method Section. LASSO (Least Absolute Shrinkage and Selection Operator) Cox regression method (Simon et al., 2011) was employed to construct a BCSC-associated miRNA classifier for clinical prognosis. The detailed methods were described in the

Supplementary Method Section. We divided the patients into high- and low-risk groups by K-means (Hartigan and Wong, 1979), and validated the sensitivity and specificity of this classifier in predicting DRFS in all groups. Furthermore, we compared the prognostic value of the classifier with IHC4 scoring and 21-gene RS, and investigated the association between chemotherapy and risk scores defined by this classifier.

2.4. Statistical Analysis

Clinical follow-up data was recorded by retrospective case-note review with data from surviving patients censored every 3 months or from telephone counseling every month. Outcome measures were distant relapse free survival (DRFS) defined as the time from the date of surgery to the time of the first distant relapse or the date of the last follow-up visit for patients (Gourgou-Bourgade et al., 2015). *t*-test and the χ^2 test were used to compare continuous and categorical variables of two groups. The Cox regression model was performed for multivariate survival analysis. We used Kaplan-Meier method to display the survival curves and log-rank test to compare the difference between high- and low-risk groups in survival. The receiver operating characteristic (ROC) curves were employed to test the sensitivity and specificity of variables in predicting DRFS. All statistical analyses were performed at Sun Yat-sen Memorial Hospital, Sun Yat-sen University (Supervised by E.S.) and blind to group assignment and outcome assessment.

3. Results

3.1. Baseline Clinicopathological Status

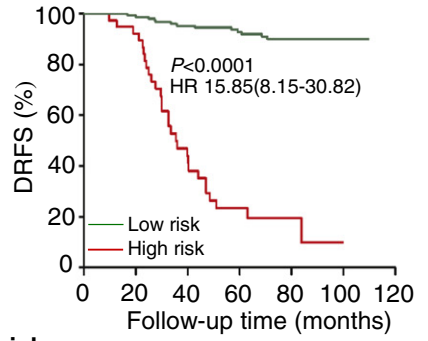
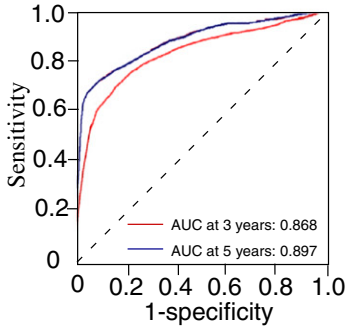
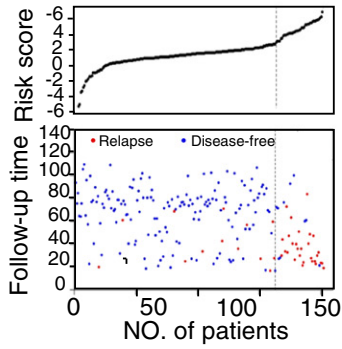
The baseline clinicopathological features and treatments of three cohorts were comparable (Table 1). All the 611 patients received surgical resection and adjuvant endocrine therapy after surgery, and 469 (76.8%) of them received adjuvant chemotherapy with anthracycline or paclitaxel-based regimens or both, 334 (54.7%) of them received upfront tamoxifen, 240 (39.3%) of them received upfront aromatase inhibitor (AI), 37 (6.0%) of them received tamoxifen followed by AI. There were 33.0% (123 out of 373) postmenopausal patients received upfront tamoxifen. The median follow-up was 69.4 months, ranging from 8.1 to 120.4 months, and 130/611 (21.3%) patients developed distant recurrence during the follow-up period. The 5-year and 7-year DRFS were 81.5% and 78.9% in Cohort 1, 74.3% and 73.0% in Cohort 2, and 84.6% and 83.3% in Cohort 3, respectively.

3.2. Construction of the 10 BCSC-associated miRNA-Based Prognostic Model

We screened 35 BCSC-associated miRNAs as candidates based on literature review and a pilot assay (Fig. S1). The Kaplan-Meier analysis with Log-rank test suggested that 13 out of the 35 BCSC-associated miRNAs were associated with DRFS in these training and internal validation groups (Table S1). We used LASSO Cox regression method and identified 10 miRNAs with the highest frequency of being selected by this method among 200 bootstrap replicates, which were as follow: miR-30c, miR-21, miR-181a, miR-181c, miR-125b, miR-7, miR-200a, miR-135b, miR-22 and miR-200c. These 10 miRNAs were incorporated into a Cox proportional hazards regression model and the coefficients

Fig. 2. Construction of the ten-miRNA-based classifier. Risk score by the 10 miRNA-based classifier (left), time-dependent ROC curves (middle) and Kaplan-Meier survival (right) in the training group ($n = 202$) (a), internal testing group ($n = 101$) (b), independent validation group-1 ($n = 152$) (c) and independent validation group-2 ($n = 158$) (d). Data are AUC or hazard ratio (95% CI). ROC = receiver operator characteristic. AUC = area under the curve. We used K-means to generate the optimum cutoff score for the prognostic model and patients were classified as high- and low- risk subgroups. We used AUCs at 3 and 5 years to assess prognostic accuracy, and calculated *p* values using the log-rank test.

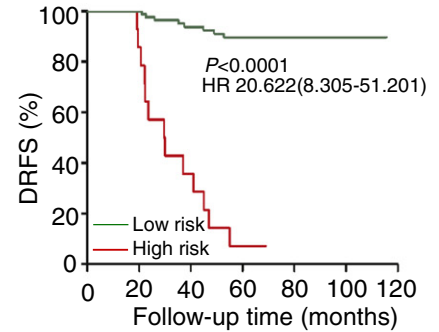
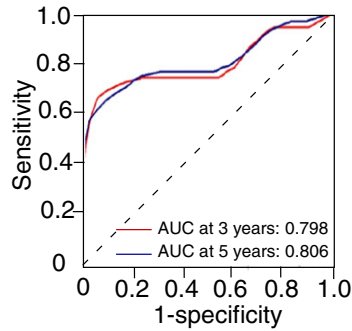
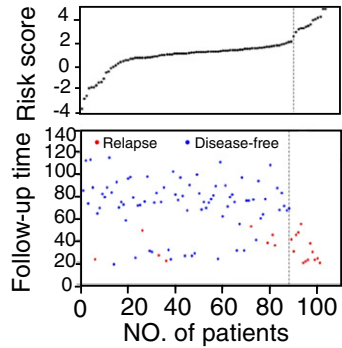
a. Training group



NO. at risk

High risk	40	38	17	10	3	0	0
Low risk	162	158	134	111	47	9	0

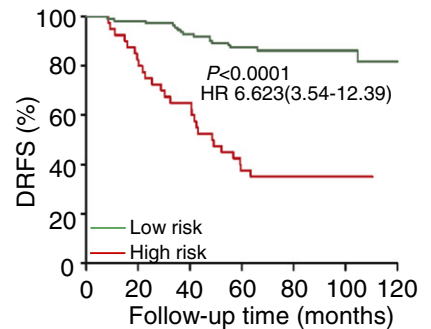
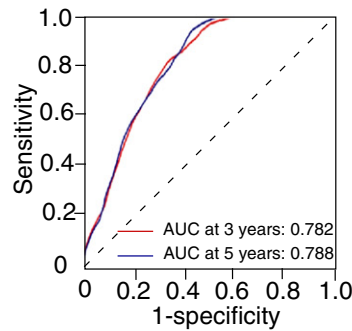
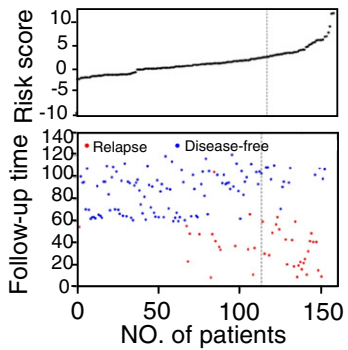
b. Internal validation group



NO. at risk

High risk	14	12	5	1	0	0	0
Low risk	87	86	70	63	31	7	0

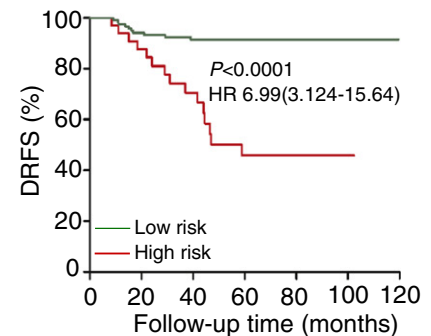
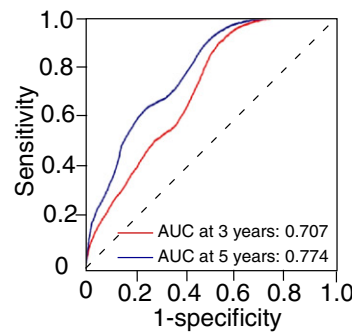
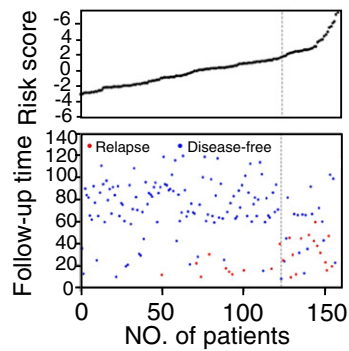
c. External validation group-1



NO. at risk

High risk	40	33	26	15	13	8	0
Low risk	112	110	104	96	60	25	1

d. External validation group-2



NO. at risk

High risk	38	27	19	10	4	1	0
Low risk	118	111	100	89	52	12	0

were estimated using the data of patients from the training group. The risk score which were used to predict disease recurrence was given as follows,

$$\begin{aligned} \text{Risk score} = & (0.165 \times \text{expression}_{J_{\text{miR}-21}}) \\ & - (0.812 \times \text{expression}_{J_{\text{miR}-30c}}) \\ & + (1.053 \times \text{expression}_{J_{\text{miR}-181a}}) \\ & + (0.179 \times \text{expression}_{J_{\text{miR}-181c}}) \\ & + (0.672 \times \text{expression}_{J_{\text{miR}-125b}}) \\ & - (0.588 \times \text{expression}_{J_{\text{miR}-7}}) \\ & - (0.469 \times \text{expression}_{J_{\text{miR}-200a}}) \\ & + (1.065 \times \text{expression}_{J_{\text{miR}-135b}}) \\ & - (0.986 \times \text{expression}_{J_{\text{miR}-22}}) \\ & - (0.820 \times \text{expression}_{J_{\text{miR}-200c}}). \end{aligned}$$

In this formula, $\text{expression}_{J_{\text{miR}-i}}$ indicates the log10-scaled expression value of miRNA-*i*.

Patients were divided into high- and low-risk groups with the optimal cut-off of risk scores (1.324) (Fig. 2 left panel). As shown in Table 1, the traditional clinicopathological status (e.g. LN status and TNM-stage) was associated with the predictive risk scores.

In the training group, a time-dependent ROC analysis demonstrated that the AUC (Area under curve) for 3-year and 5-year DRFS of this classifier were 0.868 and 0.897 (Fig. 2a middle panel), respectively. Patients in the high-risk subgroup had a significantly lower 5-year DRFS than those in the low-risk subgroup (28.3% vs. 92.1%, HR = 15.85, 95% CI 8.15–30.82; $p < 0.0001$) (Fig. 2a right panel).

3.3. Internal and External Validation

In the internal validation group, patients defined as high-risk by the classifier had significantly lower 5-year DRFS than the low-risk patients (7.1% vs. 89.7%, HR 20.62, 95% CI 8.31–51.20; $p < 0.0001$; Fig. 2b). In Cohort 2, 40 (26.3%) and 112 (73.7%) of the patients were classified into the high- and low-risk subgroups, respectively, and the 5-year DRFS in the high-risk subgroup was significantly lower than that of the low-risk one (37.5% vs. 87.5%, HR 6.62, 95% CI 3.54–12.39; $p < 0.0001$; Fig. 2c). Similarly, in Cohort 3, the 5-year DRFS were 46.1% and 91.5% in the high-risk ($n = 34$, 21.8%) and the low-risk ($n = 122$, 78.2%) subgroup, respectively (HR 6.99, 95% CI 3.12–15.64; $p < 0.0001$; Fig. 2d).

3.4. Multivariate Analysis and Stratification Analysis

The prognostic risk score were significantly associated with the classical prognostic factors, including menopausal status, tumor size, lymph node status, TNM, grade and ki67 (Fig. S2). After adjustment with the individual clinicopathological features, the prognostic risk score was still significantly associated with the 5-year DRFS (Table 2 and S2). Stratification analysis showed that our miRNA signature is superior in predicting DRFS of HR + HER2 – breast cancer patients, independent of clinic pathological features including TNM staging, tumor size, LN

status and histological grade (Fig. 3). In addition, the prognostic risk score was constantly accurate when the analysis was stratified by different regimens of endocrine therapies including Tamoxifen upfront and AI upfront, ki67 expression, PR status and menopausal status (Fig. S3).

3.5. Comparison with Other Prognostic Factors

The prognostic accuracy of the 10 BCSC-associated miRNAs based classifier for the DRFS was much better than any of the single BCSC-associated miRNA (Fig. 4a). Additionally, the classifier was also significantly more accurate than any other single clinicopathological risk factor evaluated in our study (Fig. 4b). Specifically, we compared the AUC between the miRNA panel and TNM stage, and found that there was statistically significant difference between these two variables ($P = 0.0366$), which was consistent with the stratification analysis shown in Fig. 3.

We quantified the IHC4 score for each patient and stratified the patients into low-, intermediate- or high- risk group (Fig. S4). We observed much closer association between the miRNA risk scoring with DRFS of the HR + HER2 – breast cancer patients than the IHC4 risk scoring (miRNA RS high vs. low, HR 10.313, 95% CI 7.182–14.809; $p < 0.0001$; IHC4 RS high vs. low, HR 1.454, 95% CI 1.133–1.866; $p = 0.003$; IHC4 RS high vs. intermediate, HR 1.370, 95% CI 0.931–2.061; $p = 0.11$; IHC4 RS intermediate vs. low, HR 1.508, 95% CI 0.941–2.418; $p = 0.088$; Fig. 5a). In patients with intermediate-risk scores as defined by IHC4 scoring, our classifier could further distinguish a high-risk subgroup of patients with significantly lower 5-year DRFS (Fig. 5b). This classifier had better prognostic accuracy than IHC4 score in 611 HR + HER2 – patients (AUC 0.772 vs AUC 0.692) (Fig. 4b), probably due to the exclusion of HER2 positive patients. In addition, compared to IHC4 score, stratification analysis showed that the miRNA RS was superior in predicting DRFS of HR + HER2 – breast cancer patients with high tumor burden including advanced TNM staging, larger tumor size, higher histological grade, lymph nodes metastasis and higher proliferation index as well as in the patient subgroup aged over 40 (Fig. S5). Therefore, these data suggest that the 10 BCSC-associated miRNAs based prognostic scoring may serve as a better prognostic model than IHC4 scoring in the HR + HER2 – subtype of Asian breast cancer patients.

3.6. Comparison with 21-Gene Recurrence Score

Since the 21-gene RS was validated to predict distant recurrence in patients with node-negative, estrogen-receptor-positive breast cancer (Paik et al., 2004), we selected node-negative patients from our cohort for the comparison between the miRNA and 21 gene RS. In our entire cohort, there were 33.0% (285 out of 611) patients with negative lymph node. Among the 285 patients without lymph node metastasis, we performed 21-gene assay in 153 cases with qualified FFPE samples for the assay and generated a 21-gene recurrence score (21-gene RS) for each individual. The median follow-up period of this cohort was 70.7 (8.1–

Table 2
Multivariate association of the 10-miRNA classifier, clinicopathological characteristics with DRFS.

Variable	Cohort 1 ($n = 303$)		Cohort 2 ($n = 152$)		Cohort 3 ($n = 156$)	
	HR (95% CI)	<i>P</i> value	HR (95% CI)	<i>P</i> value	HR (95% CI)	<i>P</i> value
10-miRNA classifier (High vs. Low risk)	11.04(6.26–19.49)	<0.0001	4.95(2.51–9.74)	<0.0001	5.65(2.38–13.42)	<0.0001
Age (>40 years vs. ≤40 years)	0.86(0.44–1.65)	0.642	1.04(0.47–2.28)	0.926	0.65(0.20–2.13)	0.474
Menopause (Yes vs. No)	0.85(0.46–1.56)	0.588	0.76(0.37–1.56)	0.455	0.41(0.16–1.04)	0.061
T stage (≤2 cm vs. >2 cm)	1.80(0.82–3.99)	0.145	2.29(0.74–7.06)	0.15	1.37(0.07–26.02)	0.836
LN (positive vs. negative)	2.14(1.02–4.48)	0.044	1.96(0.98–3.91)	0.058	1.51(0.22–10.51)	0.678
TNM stage (III–II vs. I)	1.93(0.35–10.75)	0.452	2.82(0.34–23.19)	0.336	2.65(0.11–63.06)	0.546
Grade (III–II vs. I)	2.18(0.50–9.55)	0.303	3.06(0.70–13.43)	0.139	1.10(0.45–9.02)	0.927
Ki67(>14% vs. ≤14%)	1.62(0.85–3.10)	0.143	1.09(0.56–2.13)	0.801	2.90(1.04–8.09)	0.042
Chemotherapy (yes vs. no)	0.56(0.32–0.98)	0.042	0.52(0.26–1.02)	0.058	0.44(0.19–1.06)	0.068

DRFS = disease relapse free survival; and LN = lymph node.

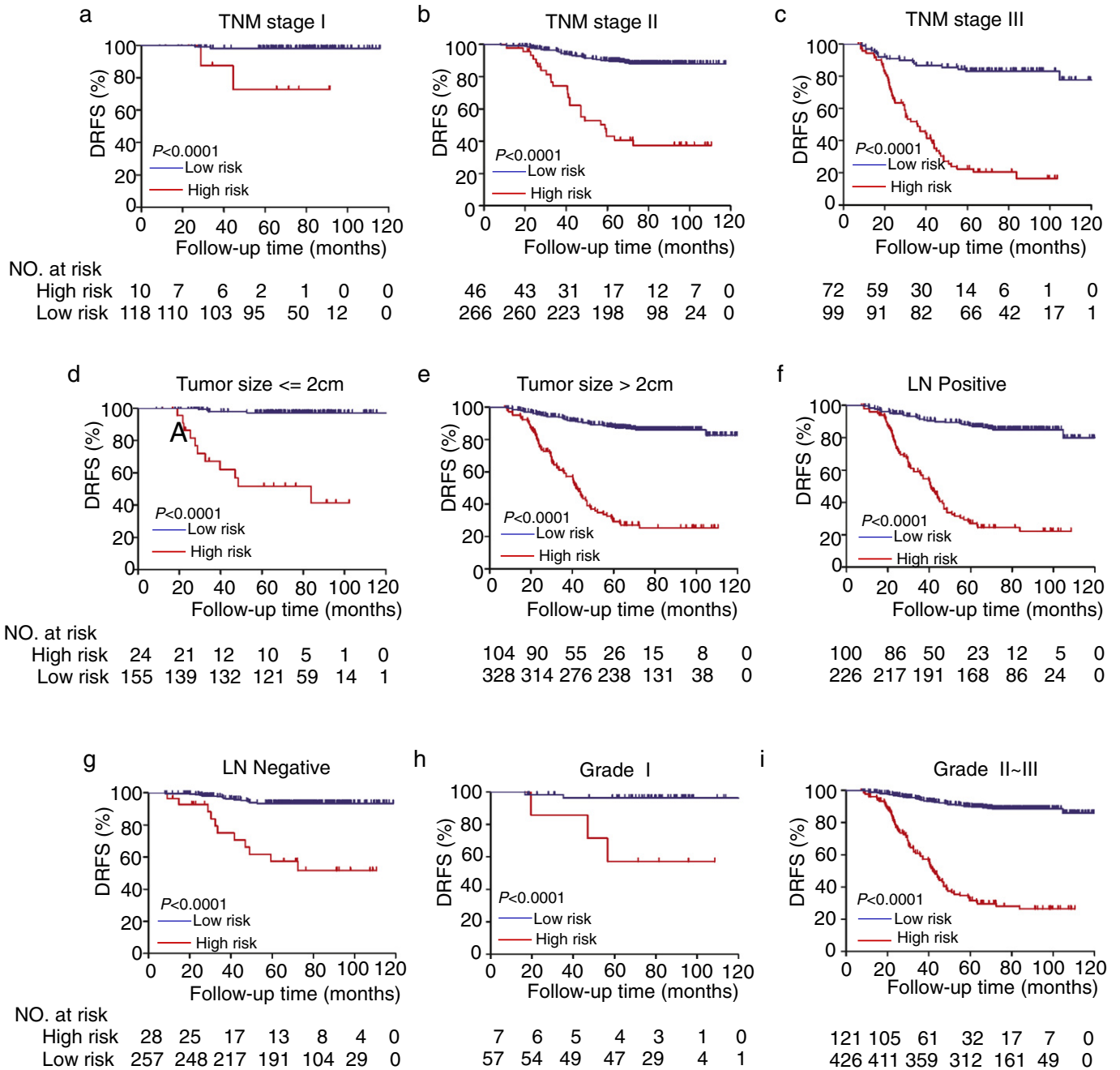


Fig. 3. Kaplan-Meier survival analysis for 611 HR + HER2 – patients according to the 10 miRNA-based classifier stratified by clinical risk factors. (a–c) TNM stage (d–e) Tumor size. (f–g) Lymph node (LN) status. (h–i) Grade.

118.8) months. These LN negative cases ($n = 153$) were enrolled in three independent centers, with 85 cases from SYSMH, 43 cases from Cancer Center of Sun Yat-sen University, and 25 cases from the Third Hospital of Nanchang City. The baseline clinicopathological features of the 153 patients were comparable with the total 285 cases of LN-negative patients (Table S4). A direct comparison of the prognostic value between 10-miRNA classifier and 21-gene RS was further performed in this cohort. As shown in Fig. 5c, the miRNA RS demonstrated a much closer association with the DRFS of HR + HER2 – breast cancer patients than the 21-gene RS (miRNA RS high vs. low, HR 9.748, 95% CI 3.415–27.842, $p < 0.0001$; 21-gene RS high vs. low, HR 2.181, 95% CI 1.109–4.291, $p = 0.004$; 21-gene RS high vs. intermediate, HR 1.894, 95% CI 0.554–6.474, $p = 0.108$; 21-gene RS intermediate vs. low, HR 1.012, 95% CI 0.674–1.520, $p = 0.953$; Fig. 5c). Additionally, among the 40

patients with intermediate 21-gene RS, the 10 BCSC-associated miRNA classifier could further identify a high-risk subgroup of patients with significantly lower 5-year DRFS (miRNA RS high vs. low HR 6.59, 95% CI 2.20–19.70, $P = 0.001$, HR 12.55, 95% CI 1.720–91.544, $P = 0.013$, Fig. 5d). More importantly, the 10 BCSC-associated miRNA classifier showed a trend of better prognostic accuracy than the 21-gene RS in the 153 HR + HER2 – patients tested, although it did not reach statistical difference (AUC 0.710 vs. 0.685, $P = 0.36$) (Fig. 5e). Stratification analysis showed that the prognostic accuracy between miRNA RS and 21-gene RS is comparable (Fig. S6), probably due to limited population in each subgroup. Therefore, the 10 miRNAs-based classifier seems to outperform the 21-gene RS in predicting the DRFS for HR + HER2 – breast cancer patients.

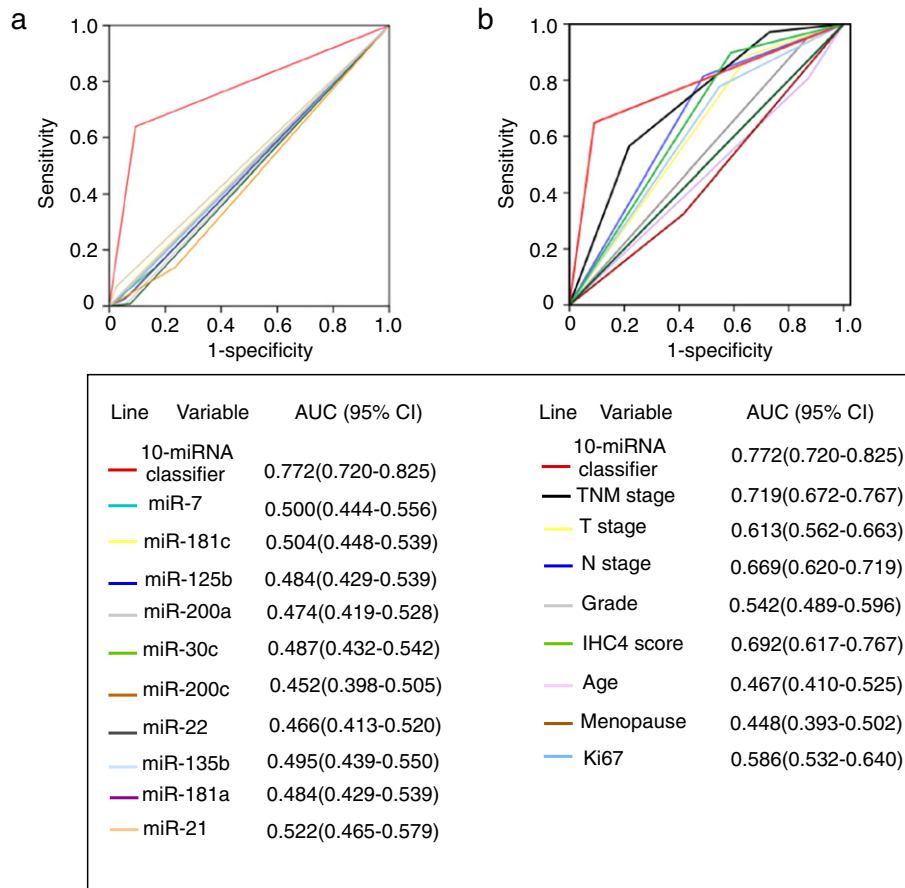


Fig. 4. ROC curves compare the prognostic accuracy of the 10 based-miRNA classifier with clinicopathological risk factors and single miRNA in the 611 patients with HR+ Her2-. (a) Comparisons of the prognostic accuracy by the 10 miRNA-based classifier (high vs. low risk), and any single BCSC-associated miRNA in the prediction of DRFS. miR-7 (high vs. low expression), miR-181c (high vs. low expression), miR-125b (high vs. low expression), miR-200a (high vs. low expression), miR-30c (low vs. high expression), miR-200c (low vs. high expression), miR-22 (low vs. high expression), miR-135b (low vs. high expression), miR-181a (low vs. high expression) and miR-21 (low vs. high expression). (b) Comparisons of prognostic accuracy by the 10 miRNA-based classifier (high risk vs. low risk), TNM stage (II-III vs. I), tumor size (>2 cm vs. ≤2 cm), pathological grade (II-III vs. I), lymph node (positive vs. negative), IHC4 score (high vs. low-intermediate risk), age (≤40 vs. >40), menopause (yes vs. no), Ki67 (≥14% vs. <14%). AUC = area under curve. ROC = receiver operator characteristic.

3.7. Association with Chemotherapy and Clinical Outcome

In the validation population ($n = 409$), adjuvant chemotherapy was associated with a slight increase in 5-year DRFS (83.2% with chemotherapy vs. 66.7% without chemotherapy; HR 1.723, 95% CI 1.117–2.658; $P = 0.014$; Fig. 6a). We observed a 30.7% improvement of 5-year DRFS by chemotherapy in the high-risk subgroup (HR 1.765, 95% CI 1.024–3.043; $P = 0.041$; Fig. 6c), whereas chemotherapy was not associated with improvement of the 5-year DRFS in the low-risk subgroup (HR 1.652, 95% CI 0.805–3.390; $P = 0.171$; Fig. 6b).

4. Discussion

In our present study, we developed and validated a novel prognostic model based on 10 BCSC-associated miRNAs to improve the prediction of disease recurrence in patients with non-metastatic HR+HER2- breast cancer. The prognostic value of this proposed classifier is better than the well-established IHC4 scoring, 21-gene RS and other clinicopathological risk factors. As shown in our retrospective analysis, while patients in the low risk group gained little benefit from chemotherapy, patients with higher recurrent risk defined by our classifier seem to benefit from chemotherapy. Therefore, the 10-miRNA classifier may be useful to predict chemotherapy sensitivity, but this needs further prospective studies.

The 10 miRNAs incorporated into our predictive model have been shown to regulate the biology of BCSCs and non-stem cell like breast cancer cells (non-BCSCs) by various target genes, suggesting these miRNAs play crucial roles during the development of tumor. Among them, reduction in miR-30 family maintains self-renewal and inhibits apoptosis of BCSCs by relieving their suppression to UBC9 (Yu et al., 2010), AVEN and FOXD1 (Ouzounova et al., 2013), while higher microRNA-30c expression was significantly associated with benefit of tamoxifen treatment in advanced ER+ breast cancer. Increased expression of miR-125b is associated with chemotherapy resistance and metastasis in breast cancers (Okuda et al., 2013) whereas increased miR-7 suppresses brain metastasis of BCSCs by modulating KLF4 expression. These findings implied that the selected miRNAs in our classifier may have biological importance. Although in our study we did not have direct evidence to support that these miRNAs influenced the clinical outcomes via their target genes, this has no direct impact on their usefulness as biomarker.

The prognostic and predictive roles of miRNAs have been investigated in several studies. For example, a MAPK-microRNA signature has been shown to predict endocrine resistance and poor survival in ER-positive breast cancers (Miller et al., 2015). Nonetheless, we are still lack of well validated miRNA signatures that are applicable in clinical practice for diagnosis, prognosis and personalized therapy in breast cancer. As distinct from previous miRNA prognostic models, the 10 BCSC-associated miRNAs classifier was set up on the basis of the aggressive cancer biology, and its prognostic value has been validated both

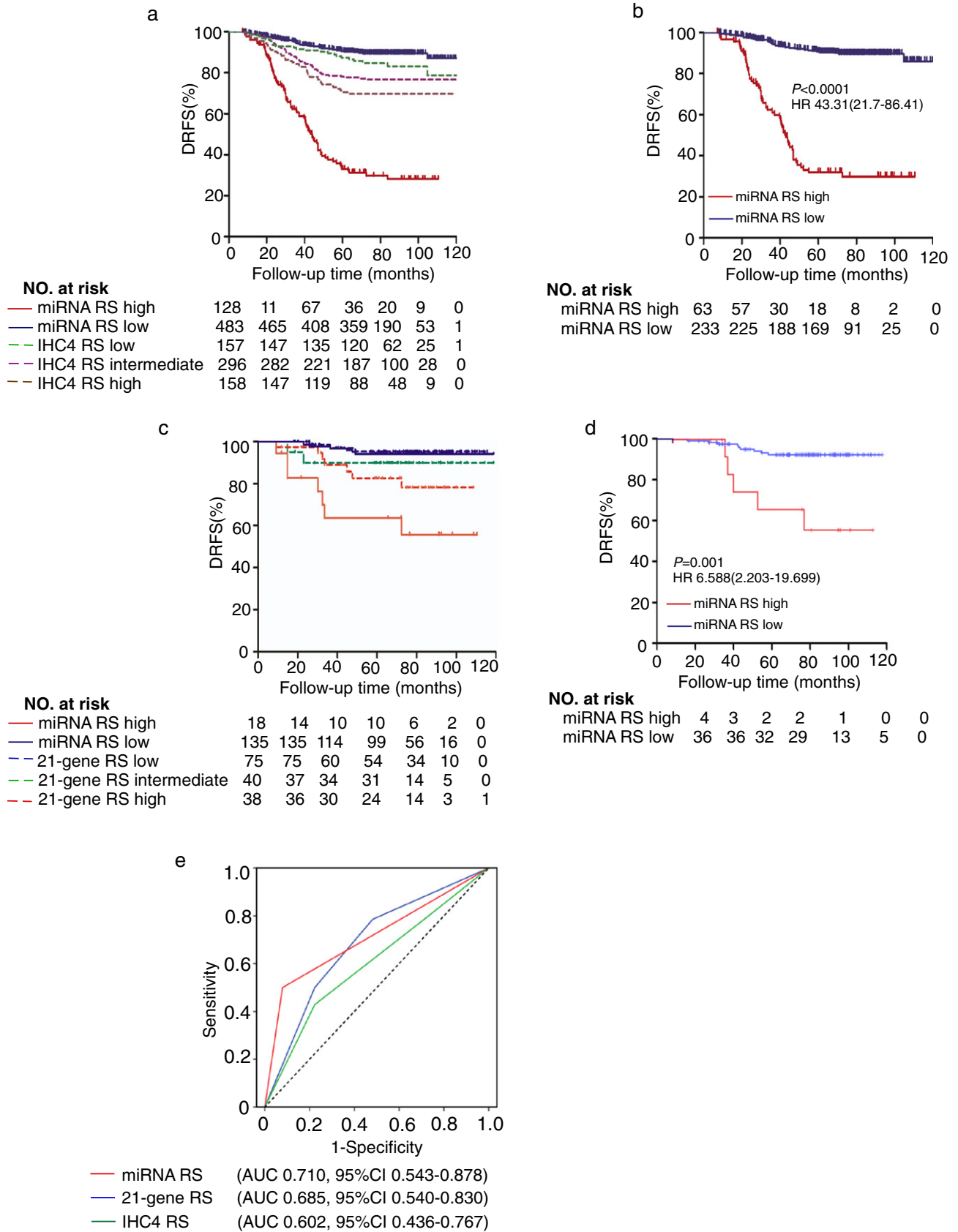


Fig. 5. Comparison of the prognostic accuracy of miRNA RS with IHC4 RS and 21-gene RS. (a) Kaplan-Meier survival curves for patients indifferent subgroups, which were stratified by IHC4 RS and miRNA RS. (b) Kaplan-Meier survival curves for the patients with intermediate-risk defined by IHC4 scoring. (c) Kaplan-Meier survival curves for patients stratified by 21-gene RS and the miRNA RS. 21-gene RS low risk, defined as a recurrence score < 18; 21-gene RS intermediate risk, defined as $18 \leq \text{score} < 31$; 21-gene RS high risk, defined as a score ≥ 31 . (d) Kaplan-Meier survival curves for the patients with 21-gene RS intermediate risk stratified by miRNA RS. (e) Comparisons of the prognostic accuracy by miRNA RS (high vs. low risk), IHC4 RS and 21-gene RS. DRFS = distant relapse-free survival; RS = recurrence/risk score; and AUC = area under curve.

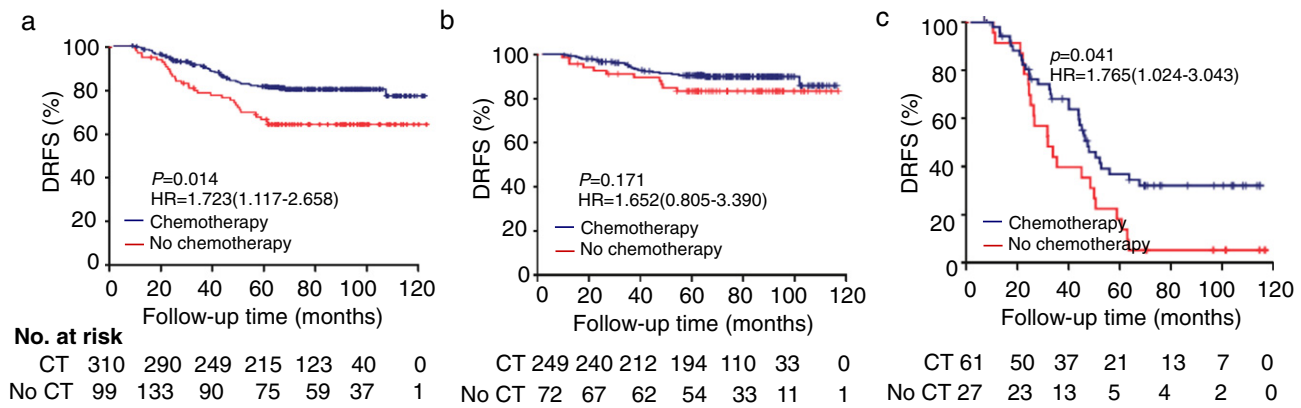


Fig. 6. Effect of chemotherapy in different subgroups. (a) Kaplan-Meier survival curves for patients in different subgroups, which were stratified by the receipt of chemotherapy. (b–c) Effect of adjuvant chemotherapy on DRFS in the low-risk (b) and high-risk (c) subgroups. DRFS = distant relapse-free survival. CT = chemotherapy.

internally and externally in multi-institutional study cohorts of HR + HER2 – breast cancer.

Profiling of multiple gene expression by qRT-PCR or microarray, such as 21-gene testing and 70-gene MammaPrint signature, has been used in clinical practice to provide prognostic information and help in clinical decision for breast cancer patients. These assays have been approved to be used in clinics for predicting recurrence risks for early-stage ER-positive breast cancer patients in western countries (Paik et al., 2004, 2006; van't Veer et al., 2005, 2002). Similarly, PAM50 assay provides a risk prediction for distant recurrence in HR-positive postmenopausal breast cancers (Wallden et al., 2015). It has been reported that IHC4 scoring based on ER, PR, HER2 and Ki67 status has similar predictive value as the 21-gene risk scoring for ER + breast cancer patients, and IHC4 scoring has also been widely used in routine clinical practice (Gourgou-Bourgade et al., 2015). When compared with the IHC4 scoring, our 10 BCSC-associated miRNA classifier demonstrated better prognostic value for DRFS in the HR-positive, HER2-negative breast cancer patients regardless of lymph node or menopausal status. Furthermore, we performed a comparison of the prognostic value of miRNA classifier and 21-gene RS in HR + HER2 – patients with negative LN status. Our retrospective data showed that the miRNAs-based classifier seemed to outperform the 21-gene RS in predicting the DRFS in this cohort. Several lines of explanation may be responsible for the different prognostic values of these gene assays. Firstly, the prognostic information of the first generation signatures stems almost exclusively from the degree of expression of proliferation-related genes in ER + cancers (Reis-Filho and Pusztai, 2011). In contrast, our classifier did not display statistically significant associations with Ki67 expression two out of the four cohorts, and the prognostic risk score was constantly accurate when the analysis was stratified by ki67 expression level and histological grade, suggesting our signature provides information more than proliferation. Secondly, meta-analyses have revealed that tumor size and lymph-node status provided prognostic information that is independent of that offered by prognostic generation signature (Wirapati et al., 2008). Also, a model combining risk score with traditional anatomical pathological factors could be more prognostic than risk alone (Pusztai, 2011). In contrast, our proposed classifier can predict the DRFS of patients with HR + HER2 – breast cancer, independent of clinicopathological features, although further prospective trials are needed to validate its prognostic value. Thirdly, previous predictive models were largely founded on the basis of computer-based algorithm, the function of some of the selected genes is not clear, and additional efforts were needed to explore their biological functions after the models were built. Therefore, hypothesis-driven approaches are alternative strategies to develop prognostic signatures for clinical outcomes (Reis-Filho and Pusztai, 2011), which had been demonstrated in ovarian cancer. In our study, combining biology-driven and empirical approaches, we have selected

a score of critical BCSC-associated miRNAs based on literature reviewing, trained them in our studying population and developed a valid prognostic model for HR + HER2 – breast cancers. This approach is less time-consuming and more economical, cost-effective than the traditional strategy. However, it should also be noted that the pre-selection process of biomarker candidates based on one biology-driven hypothesis may have inevitable bias. In addition to cancer stem cell properties, many other biological behaviors, such as epithelial-mesenchymal transition (EMT), drug resistance and anoikis, may also contribute to cancer progression and patient survival. Hence, pre-selected biomarkers focusing on one biological process may not completely reflect the whole picture. Taken together, our present findings suggest that developing prognostic gene models using the combination of biology-driven and empirical approaches is feasible, efficient and cost-effective.

Albeit the present study demonstrated a successful example in constructing a prognostic model for cancer prognosis using a biology-driven approach, the limitations of our study should be addressed. Firstly, although our present study found that the 10 BCSC-associated miRNAs classifier could provide prognostic information for HR + HER2 – breast cancer and potentially outperformed the 21-gene RS and IHC4 score, our study is limited by its retrospective design. Prospective randomized clinical trial is needed to validate the prognostic value of our current miRNA model. Also, the prospective randomized clinical trials with larger cohort for head-to-head comparison between current miRNA model and 21-gene RS are required in the future. Secondly, there are significant discrepancies in the epidemiological characteristics, diagnostic methods, biological features, and treatment strategies of breast cancer between Asian and western countries (Sivasubramaniam et al., 2015). Given these discrepancies, the clinical applicability of this miRNA classifier should be validated in breast cancer patients from western countries. Thirdly, since most patients in our study were diagnosed before 2008, the types of endocrine therapy, treatment periods and combined therapies changed over time, and may not be optimal. Thus, the association between the miRNA classifier and chemotherapeutic benefit need further evaluation. Fourthly, approximately 50% of all tumor recurrence occurs in the ER-positive breast cancer patients after 5 years, the prognostic value of this classifier for prediction of late-recurrence should be tested in a breast cancer population with a longer follow-up. Last but not the least, although our present study found that the benefit from chemotherapy in the low-risk group is small, our study is limited by its retrospective design. This limitation also applies to the development and validation of our prognostic signature as described above. Therefore, prospective randomized clinical trial is needed to further validate the predictive value of our 10 BCSC-associated miRNA model for chemotherapeutic benefits in HR-positive, HER2-negative breast cancer patients.

In summary, our study demonstrated that the 10 BCSC-associated miRNA prognostic model can effectively distinguish HR-positive, HER2-negative breast cancer patients with low recurrence risk from those with high recurrence risk, regardless of lymph node or menopausal status and age at diagnosis. Since our 10-miRNA-based classifier can provide better prognostic value than the traditional clinicopathological predictors in HR + HER2 – breast cancer patients, it may be helpful to effectively distinguish the low risk cases from the high risk ones among HR-positive, HER2-negative breast cancer patients. In this scenario, clinicians may be able to recommend less aggressive therapy for low-risk individuals in directing personalized therapy. Therefore, this model may facilitate personalized clinical decision making for HR + HER2 – breast cancer patients.

Declaration of Interests

We declare that we have no conflicts of interest.

Author Contributions

Conception and design: Erwei Song, Chang Gong.

Financial support: Erwei Song, Fengxi Su, Chang Gong, Kai Chen, Xinhua Xie, Zhihua Li.

Administrative support: Erwei Song.

Provision of study materials or patients: Erwei Song, Fengxi Su, Musheng Zeng, Zhihua Li.

Collection and assembly of data: Weige Tan, Gehao Liang, Qian Li.

Data analysis and interpretation: Chang Gong, Weige Tan, Kai Chen, Na You, Shan Zhu, Dong Yin, Xueqin Wang.

Manuscript writing: Erwei Song, Chang Gong, Weige Tan, Kai Chen, Wan-Yee Lau, Qiang Liu.

Final approval of manuscript: All authors.

Acknowledgments

This work was funded by grants from the Natural Science Foundation of China (81490750, 81230060, 81442009, 81272893, 81472466, 81302318, 81402201, 81260389, 81572580); Program for New Century Excellent Talents in University (NCET-12-0565); National Science Foundation of Guangdong Province (2014A03036003, S2012030006287, 2014A030310378, 2016A050502018), Guangzhou Science Technology and Innovation Commission (201508020008, 201508020249), Guangdong Science and Technology Department (2015B050501004), Translational medicine public platform of Guangdong Province (4202037) and Guangdong Department of Science & Technology Translational Medicine Center grant (2011A080300002); Elite Young Scholars Program of Sun Yat-sen Memorial Hospital (Y201401); Grant KLB09001 from the Key Laboratory of Malignant Tumor Gene Regulation and Target Therapy of Guangdong Higher Education Institutes; Grant (2013) 163 from Key Laboratory of Malignant Tumor Molecular Mechanism and Translational Medicine of Guangzhou Bureau of Science and Information Technology.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ebiom.2016.08.016>.

References

Arigoni, M., Barutello, G., Riccardo, F., Ercole, E., Cantarella, D., Orso, F., Conti, L., Lanzardo, S., Taverna, D., Merighi, I., et al., 2013. miR-135b coordinates progression of ErbB2-

- driven mammary carcinomas through suppression of MID1 and MTCH2. *Am. J. Pathol.* 182, 2058–2070.
- Dai, J., Hu, Z., Jiang, Y., Shen, H., Dong, J., Ma, H., Shen, H., 2012. Breast cancer risk assessment with five independent genetic variants and two risk factors in Chinese women. *Breast Cancer Res.* 14, R17.
- Dalerba, P., Clarke, M.F., 2013. Oncogenic miRNAs and the perils of losing control of a stem cell's epigenetic identity. *Cell Stem Cell* 13, 5–6.
- Dirks, P.B., 2009. MicroRNAs and parallel stem cell lives. *Cell* 138, 423–424.
- Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S., et al., 2007. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* 1, 555–567.
- Gong, C., Yao, H., Liu, Q., Chen, J., Shi, J., Su, F., Song, E., 2010. Markers of tumor-initiating cells predict chemoresistance in breast cancer. *PLoS One* 5, e15630.
- Gourgou-Bourgade, S., Cameron, D., Poortmans, P., Asselain, B., Azria, D., Cardoso, F., A'Hern, R., Bliss, J., Bogaerts, J., Bonnefoi, H., et al., 2015. Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (definition for the assessment of time-to-event endpoints in CANcer trials). *Ann. Oncol.* 26, 2505–2506.
- Hartigan, J.A., Wong, M.A., 1979. A K-means clustering algorithm. *Appl. Stat.* 28.
- Kang, J., D'Andrea, A.D., Kozono, D., 2012. A DNA repair pathway-focused score for prediction of outcomes in ovarian cancer treated with platinum-based chemotherapy. *J. Natl. Cancer Inst.* 104, 670–681.
- Miller, P.C., Clarke, J., Koru-Sengul, T., Brinkman, J., El-Ashry, D., 2015. A novel MAPK-microRNA signature is predictive of hormone-therapy resistance and poor outcome in ER-positive breast cancer. *Clin. Cancer Res.* 21, 373–385.
- Okuda, H., Xing, F., Pandey, P.R., Sharma, S., Watabe, M., Pai, S.K., Mo, Y.Y., Iizumi-Gairani, M., Hirota, S., Liu, Y., et al., 2013. miR-7 suppresses brain metastasis of breast cancer stem-like cells by modulating KLF4. *Cancer Res.* 73, 1434–1444.
- Ouzounova, M., Vuong, T., Ancey, P.B., Ferrand, M., Durand, G., Le-Calvez Kelm, F., Croce, C., Matar, C., Hecceg, Z., Hernandez-Vargas, H., 2013. MicroRNA miR-30 family regulates non-attachment growth of breast cancer cells. *BMC Genomics* 14, 139.
- Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., et al., 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* 351, 2817–2826.
- Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., Cronin, M., Baehner, F.L., Watson, D., Bryant, J., et al., 2006. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J. Clin. Oncol.* 24, 3726–3734.
- Puztai, L., 2011. Anatomy and biology: two complementary sides of breast cancer prognostication. *J. Clin. Oncol.* 29, 4347–4348.
- Reis-Filho, J.S., Puztai, L., 2011. Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet* 378, 1812–1823.
- Ring, A.E., Smith, I.E., Ashley, S., Fulford, L.G., Lakhani, S.R., 2004. Oestrogen receptor status, pathological complete response and prognosis in patients receiving neoadjuvant chemotherapy for early breast cancer. *Br. J. Cancer* 91, 2012–2017.
- Simon, N., Friedman, J., Hastie, T., Tibshirani, R., 2011. Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* 39, 1–13.
- Sivasubramaniam, P.G., Zhang, B.L., Zhang, Q., Smith, J.S., Zhang, B., Tang, Z.H., Chen, G.J., Xie, X.M., Xu, X.Z., Yang, H.J., et al., 2015. Breast cancer disparities: a multicenter comparison of tumor diagnosis, characteristics, and surgical treatment in China and the U.S. *Oncologist* 20, 1044–1050.
- Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., et al., 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. U. S. A.* 100, 8418–8423.
- van't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al., 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530–536.
- van't Veer, L.J., Paik, S., Hayes, D.F., 2005. Gene expression profiling of breast cancer: a new tumor marker. *J. Clin. Oncol.* 23, 1631–1635.
- Wallden, B., Storhoff, J., Nielsen, T., Dowidar, N., Schaper, C., Ferree, S., Liu, S., Leung, S., Geiss, G., Snider, J., et al., 2015. Development and verification of the PAM50-based Prosigna breast cancer gene signature assay. *BMC Med. Genet.* 8, 54.
- Wirapati, P., Sotiriou, C., Kunkel, S., Farmer, P., Pradervand, S., Haibe-Kains, B., Desmedt, C., Ignatiadis, M., Sengstag, T., Schutz, F., et al., 2008. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res.* 10, R65.
- Yu, F., Deng, H., Yao, H., Liu, Q., Su, F., Song, E., 2010. Mir-30 reduction maintains self-renewal and inhibits apoptosis in breast tumor-initiating cells. *Oncogene* 29, 4194–4204.
- Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J., Song, E., 2007. let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* 131, 1109–1123.
- Zhang, Y.N., Zhou, Y.D., Mao, F., Sun, Q., 2015. Impact of the 21-gene recurrence score assay in adjuvant chemotherapy selection for node-negative, hormone receptor-positive breast cancer in the Chinese population. *Neoplasma* 62, 658–665.