

Identification of Prognosis-Relevant Subgroups in Patients with Chemoresistant Triple-Negative Breast Cancer

Ke-Da Yu¹, Rui Zhu², Ming Zhan², Angel A. Rodriguez³, Wei Yang⁴, Stephen Wong², Andreas Makris⁶, Brian D. Lehmann⁵, Xi Chen⁵, Ingrid Mayer⁵, Jennifer A. Pietenpol⁵, Zhi-Ming Shao¹, W. Fraser Symmans⁴, and Jenny C. Chang³

Abstract

Purpose: Patients with triple-negative breast cancer (TNBC) and residual disease after neoadjuvant chemotherapy generally have worse outcome; however, some patients with residual tumor after neoadjuvant chemotherapy do not relapse. We hypothesize that there are subgroups of patients with chemoresistant TNBC with different prognosis.

Experimental Design: Forty-nine chemoresistant cases from 111 patients with TNBC treated with neoadjuvant chemotherapy (M.D. Anderson Cancer Center, Houston, TX) constituted the discovery cohort, and 25 chemoresistant samples from 47 neoadjuvant chemotherapy-treated TNBC (The Methodist Hospital, Houston, TX) were chosen for validation. Extended validation was carried out in 269 operable TNBC predicted to be chemoresistant by expression pattern from published datasets.

Results: We established a seven-gene prognostic signature using dChip and gene set enrichment analyses. In the independent validation cohort, the classifier predicted correctly with positive predictive value of 75.0% and negative predictive value (i.e., relapse-free survival; RFS) of 76.9% at 3 years. Those predicted to relapse had a HR of 4.67 [95% confidence interval (CI): 1.27–17.15] for relapse in 3 years. In extended validation, patients predicted not to relapse exhibited 3-year RFS of 78.9%, whereas the 3-year RFS was 48.5% for patients predicted to relapse, with HR of 2.61 (95% CI: 1.52–4.49). The TNBC subgroup that predicted to have relatively favorable prognosis was characterized by high expression of "luminal-like" genes [androgen-receptor (AR) and GATA3], whereas the subgroup with worse prognosis was characterized by expression of cancer stem-cell markers.

Conclusion: We developed a clinically relevant signature for patients with chemoresistant TNBC. For these women, new therapeutic strategies like targeting AR activation or cancer stem cells may need to be developed. *Clin Cancer Res*; 19(10); 2723–33. ©2013 AACR.

Introduction

Triple-negative breast cancer (TNBC) is clinically defined by the lack of expression of estrogen receptor (ER), progesterone receptor (PgR), and the absence of amplification or

overexpression of HER2, and accounts for 15% to 20% of newly diagnosed breast cancer cases. In general, patients with TNBC present with larger tumors, higher grade, increased number of involved nodes, and poorer survival compared with other subtypes (1, 2). Increasing evidence indicates that TNBC is a highly heterogeneous disease (1) on a molecular (3) and genetic level (4). Treatment of patients with TNBC has been challenging due to this heterogeneity, as well as the absence of well-defined molecular targets.

Despite having higher rates of pathologic complete response (pCR) to neoadjuvant chemotherapy, patients with TNBC have a higher rate of distant recurrence and worse prognosis. Among patients with TNBC receiving neoadjuvant chemotherapy, only those with pCR have improved survival. In contrast, more than 70% of patients with TNBC have residual invasive disease after neoadjuvant chemotherapy and are at high risk of disease relapse, with significantly worse survival, particularly in the first 3 years (5, 6). Paradoxically, not all patients with TNBC and residual disease after neoadjuvant chemotherapy relapse.

Authors' Affiliations: ¹Department of Breast Surgery, Shanghai Cancer Center and Cancer Institute of Fudan University, Shanghai, China; ²Department of Systems Medicine and Bioengineering, The Methodist Hospital Research Institute; ³The Methodist Hospital Cancer Center; ⁴The University of Texas MD Anderson Cancer Center, Houston, Texas; ⁵Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, Tennessee; and ⁶Mount Vernon Hospital, London, United Kingdom

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

J. A. Pietenpol, Z. Shao, W.F. Symmans, and J.C. Chang are co-senior authors.

Corresponding Author: Jenny C. Chang, The Methodist Hospital Cancer Center, Houston, TX 77050. Phone: 713-441-0681; Fax: 713-793-1642; E-mail: jchang@tmhs.org

doi: 10.1158/1078-0432.CCR-12-2986

©2013 American Association for Cancer Research.

Translational Relevance

Although patients with triple-negative breast cancer (TNBC) and residual disease after neoadjuvant chemotherapy have worse survival than those with luminal subtypes, some of them do not relapse for a long time. In this study, we used gene expression data of patients with TNBC with residual disease and different prognosis to molecularly define the clinically relevant subgroups, and developed a seven-gene prognostic signature for chemoresistant TNBCs. A favorable prognosis was observed in patients with TNBC tumors displaying high expression of "luminal-like" genes (AR, GATA3), whereas decreased survival was observed in patients with TNBC tumors expressing cancer stem cell-like (WNT11) or EMT-associated genes (MMP28). The signature not only predicted 3-year RFS, but also showed a clinically meaningful survival difference between patients predicted to relapse versus no relapse. This signature outperforms the classical clinicopathologic features in predicting the prognosis of chemoresistant TNBC. More importantly, biologically relevant genes included in the signature might provide new potential therapeutic targets.

Identifying patients with chemoresistant TNBC who relapse versus those with relatively favorable prognosis would serve to distinguish clinically relevant subgroups for whom the targeting of different molecular pathways may be important. This study was designed to test our hypothesis that there are clinical prognosis-relevant subgroups within patients with chemoresistant TNBC. Understanding the molecular pathways, distinguishing prognostically significant subgroups will aid in the rationale design of future clinical trials.

Materials and Methods

Patients and samples from M.D. Anderson Cancer Center

To investigate the difference in genetic expression between patients with chemoresistant TNBC who relapse versus those without relapse, we chose patients treated with neoadjuvant chemotherapy (with residual cancer) and investigated survival outcomes, as our discovery and validation cohorts.

The samples of discovery cohort were from M.D. Anderson Cancer Center (MDACC, Houston, TX). Patients prospectively provided written informed consent to participate in an institutional review board-approved research protocol. As previously described, 313 HER2-negative samples from patients (45% of them were with operable stage I and II disease) treated with taxane and anthracycline-based neoadjuvant chemotherapy were obtained from June 2000 to December 2006 (7). Among them, 111 patients were identified to have TNBC, of whom 49 patients fulfilled the following criteria and were included in the discovery cohort: (i) having residual

invasive disease either in the breast or in regional lymph nodes after neoadjuvant chemotherapy (i.e., non-pCR); (ii) having grade II/III residual cancer burden (RCB; ref. 8); (iii) followed-up for longer than 20 months. The information of cohorts are provided in Table 1. Chemoresistant tumors were defined in MDACC as "non-pCR and RCB-II/III after neoadjuvant chemotherapy."

All gene expression microarrays were profiled in the Department of Pathology at the MDACC and the details of the methods for RNA purification and microarray hybridization have been reported previously (9).

Patients and samples from The Methodist Hospital, Baylor College of Medicine

The study protocol was approved by the institutional review board, and signed informed consent was obtained from all patients. From January 2002 to December 2006, 116 patients with locally advanced breast cancer presenting to the Breast Center at The Methodist Hospital, Baylor College of Medicine (TMH-BCM) were recruited into a taxane/anthracycline-based neoadjuvant chemotherapy trial. The inclusion criteria were described in the Methods in Appendix. Among them, 47 were identified as TNBC cases, and 25 were recognized as chemoresistant cases after neoadjuvant chemotherapy. The definition of chemoresistant was "pathologic grade 3B to 3D and grade 4 of modified Chevallier classification after neoadjuvant chemotherapy" (10). The 25 patients from TMH-BCM constituted the validation cohort (Table 1). Patients without relapse or death events were followed-up more than 48 months. The processes of RNA treatment and gene expression profiling have been described elsewhere (11).

Assessment of pathologic response and statuses of ER, PgR, and HER2 was described in the Methods in Appendix.

Identification and validation of prognosis signature for chemoresistant TNBC

Because of word limit, the details of the finalization and validation of a 7-gene signature for chemoresistant TNBC are described in the Methods in Appendix.

Molecular classification of chemoresistant TNBC

To investigate the relationship between our 7-gene signature and the recently described TNBC subtype molecular classification (3), we used 587 cases with TNBC in that study. Gene expression profiles of individual case were read and subtyped by Lehmann and colleagues (briefly described in the Methods in Appendix; ref. 3).

Extended validation from published adjuvant TNBC microarray data

The chemoresistant, prognosis-relevant TNBC signature was then further validated in publically available datasets. A total of 579 adjuvant TNBC from 3,488 primary breast cancer gene expression profiles representing 28 individual datasets were identified (12, 13). We predicted the sensitivity to chemotherapy using a method described in the Methods in Appendix, which is an approximation of the

Table 1. Pretreatment characteristics of the discovery and validation cohorts

	MDACC		TMH-BCM	
	Original TNBC source (n = 111) n (%)	Discovery cohort (n = 49) n (%)	Original TNBC source (n = 47) n (%)	Validation cohort (n = 25) n (%)
Follow-up time, mo				
Median (range)		25 (1–79)		36 (4–88)
Age, y				
≤50	59 (53.2)	23 (46.9)	27 (57.4)	12 (48.0)
>50	52 (46.8)	26 (53.1)	20 (42.6)	13 (52.0)
Nodal status				
Negative	26 (23.4)	9 (18.4)	10 (21.3)	8 (32.0)
Positive	85 (76.6)	40 (81.6)	37 (78.7)	17 (68.0)
Tumor size stage				
T0–2	71 (64.0)	24 (49.0)	17 (36.2)	7 (28.0)
T3–4	40 (36.0)	25 (51.0)	30 (63.8)	18 (72.0)
Grade				
I	0 (0.0)	1 (0.0)	3 (6.4)	3 (12.0)
II	13 (11.7)	9 (18.4)	7 (14.9)	3 (12.0)
III	98 (88.3)	40 (81.6)	37 (78.7)	19 (76.0)
pCR				
No	69 (65.7)	49 (100.0)	35 (74.5)	25 (100.0)
Yes	36 (34.3)	0 (0.0)	12 (25.5)	0 (0.0)
Unknown	6	0	0	0
RCB				
I	48 (46.6)	0 (0.0)	N.A.	N.A.
II	30 (29.1)	24 (51.1)	N.A.	N.A.
III	25 (24.3)	23 (48.9)	N.A.	N.A.
Unknown	8	2	N.A.	N.A.
Neoadjuvant and adjuvant CT				
P×12→FAC×4→Surgery→no CT	111 (100.0)	49 (100.0)	0 (0.0)	0 (0.0)
T×4→Surgery→AC×4	0 (0.0)	0 (0.0)	24 (51.1)	11 (44.0)
AC×4→Surgery→T×4	0 (0.0)	0 (0.0)	23 (48.9)	14 (56.0)

Abbreviations: AC, doxorubicin and cyclophosphamide; CT, chemotherapy; FAC, fluorouracil, doxorubicin and cyclophosphamide; N.A., not assessed; P, paclitaxel; T, docetaxel.

JAMA published signature (7). Finally, 269 adjuvant cases predicted to be chemoresistant with at least 3 years follow-up, and available survival outcome data were included. They could be grouped into 4 main sets according to patient sample size and patients' characteristics (Table 2).

The normalization and rescaling of 269 samples to our discovery and validation cohorts were based on a median rank score-based method (14) using ArrayMining online tools (15). Predictions were generated by applying the exact SVM model that has been learned and validated from discovery and validation cohorts, respectively.

Statistical analysis

In MDACC set, the study end point was distant relapse-free survival (DRFS), which was calculated from initial diagnostic biopsy of breast cancer to the occurrence of

distant metastasis or nonbreast cancer death. In TMH-BCM set, the study end point was relapse-free survival (RFS), calculated from initial diagnosis to the occurrence of local and regional recurrence, distant metastasis, or nonbreast cancer death. As distant metastasis is the major component of breast cancer early-relapse events (16–18), the DRFS and RFS are comparable in the first 3 years. As the relapse peak in patients with TNBC occurs within the first 3 years after surgery, the 3-year DRFS/RFS was calculated and evaluated. The log-rank test was used for comparison of differences between survival curves derived by the Kaplan–Meier method.

Predictive performance was assessed by the positive predictive value (PPV), defined as the cumulative relapse and death rate for patients predicted to relapse or death in 3 years; the negative predictive value (NPV), defined as

Table 2. Characteristics of the extended validation cohort

	Extended validation cohort in adjuvant TNBC predicted to be chemoinensitive and using no neoadjuvant chemotherapy				
	Total (<i>n</i> = 269)	Set I, Rotterdam (<i>n</i> = 87)	Set II, Frankfurt (<i>n</i> = 53)	Set III, New York/ San Francisco/ Stockholm/ Uppsala (<i>n</i> = 78)	Set IV, Mainz/ TransBIG/Oxford/ others ^a (<i>n</i> = 51)
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Follow-up time, mo					
Median (range)	63 (0–120)	69 (0–120)	34 (6–120)	82 (2–120)	93 (6–120)
Relapse during follow-up					
No	162 (60.2)	49 (56.3)	32 (60.4)	51 (65.4)	30 (58.8)
Yes	107 (39.8)	38 (43.7)	21 (39.6)	27 (34.6)	21 (41.2)
Age, y					
≤50	118 (47.8)	45 (52.9)	23 (43.4)	25 (42.4)	25 (50.0)
>50	129 (52.2)	40 (47.1)	30 (56.6)	34 (57.6)	25 (50.0)
Unknown	22	2	0	19	1
Nodal status					
Negative	183 (79.6)	69 (100.0)	36 (69.2)	29 (50.0)	49 (96.1)
Positive	47 (20.4)	0 (0.0)	16 (30.8)	29 (50.0)	2 (3.9)
Unknown	39	18	1	20	0
Tumor size stage					
T1	78 (31.6)	30 (35.3)	14 (26.4)	19 (32.2)	15 (30.0)
T2–3	169 (68.4)	55 (64.7)	39 (73.6)	40 (67.8)	35 (70.0)
Unknown	22	2	0	19	1
Grade					
I–II	78 (33.1)	24 (28.2)	16 (30.8)	24 (46.2)	14 (29.8)
III	158 (66.9)	61 (71.8)	36 (69.2)	28 (53.8)	33 (70.2)
Unknown	33	2	1	26	4
Adjuvant CT					
No	179 (73.1)	83 (98.8)	6 (15.4)	39 (54.9)	51 (100.0)
Yes	66 (26.9)	1 (1.2)	33 (84.6)	32 (45.1)	0 (0.0)

NOTE: We grouped the 269 eligible patients into 4 main sets according to patient sample size and patients' characteristics: set I, Rotterdam (www.ncbi.nlm.nih.gov/gds, GSE2034, GSE5327, GSE12276; *n* = 87, almost all were node-negative and not treated with chemotherapy); set II, Frankfurt (GSE31519; *n* = 53, mixed node status); set III, New York (GSE2603)/San Francisco (available at www.ebi.ac.uk/arrayexpress with accession number E-TABM-158)/Stockholm (GSE1456)/Uppsala (GSE3494, GSE6232, GSE4922, GSE2990; *n* = 78, mixed nodes status in each subset); and set IV, Mainz (GSE11121)/TransBIG (GSE7390)/Oxford (GSE2990, GSE6532)/others (GSE12093, GSE9195, GSE6532; *n* = 51, almost all were node-negative and not treated with chemotherapy).

^aOthers included 3 TNBC cases from London and Veridex studies.

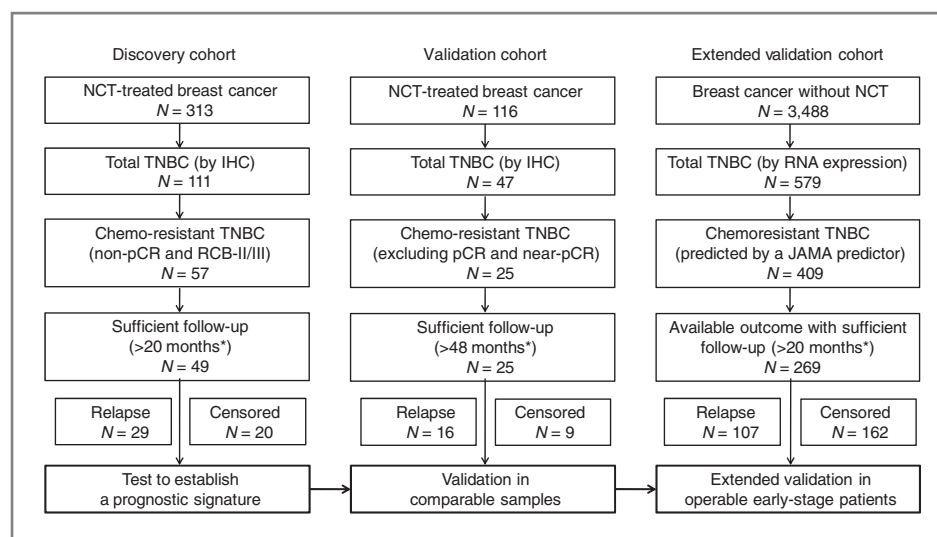
the DRFS or RFS for patients predicted to be free of relapse or death within the first 3 years. The hazard or survival was calculated from the Kaplan–Meier estimators of the survival function based on cumulative events. Confidence intervals (CI) for NPV and PPV were based on the Greenwood variance estimate. The independent prognostic value of signature was assessed in multivariate Cox regression analysis using the likelihood ratio test. The corresponding HR was calculated by the Cox model. Statistical analyses were conducted in Stata 12.0 (Stata-Corp LP.). Two-sided *P* < 0.05 was considered statistically significant.

Results

Establishment and performance of the prognostic signature in discovery cohort

To determine whether chemoprognotic predictors exist, we first examined the MDACC cohort. The study flow chart is shown in Fig. 1. For the discovery cohort, 49 of 111 TNBC samples from breast cancer women treated with neoadjuvant chemotherapy were used. This cohort had a median follow-up of 25 months, with overall 3-year DRFS of 34.4% (95% CI: 20.1%–49.2%). We compared the relapsed cases (*n* = 29) with nonrelapsed cases (*n* = 20) by dChip (19) and identified 246 genes significantly

Figure 1. Flow chart of decision algorithm used in the establishment and validation of prognostic signature in patients with chemoresistant TNBC. Abbreviations: NCT, neoadjuvant chemotherapy; RCB residual cancer burden. *, sufficient follow-up is needed for the cases without relapse event.



differentially expressed between the 2 groups, with at least a 2.14 ($2^{1.1}$)-fold difference for the ratio, with $P < 0.01$. The gene set enrichment analysis was also used to find the differentially expressed genes (see details in the Methods in Appendix). A final 7-gene signature with a minimal number and maximal prediction ability was determined.

The 7 genes were *AR* (androgen receptor), *ESR2*, *GATA3* (GATA-binding protein 3), *GBX2* (gastrulation brain homeobox 2), *KRT16* (keratin 16), *MMP28* (matrix metalloproteinase 28), and *WNT11* (wingless-type MMTV integration site family, member 11; Table A1). The basal marker *KRT16*, stem cell maker *WNT11*, and epithelial-to-mesenchymal transition (EMT) marker *MMP28*, integrally defined a subset of TNBC with unfavorable prognosis.

In contrast, luminal hormone receptor *AR* and luminal marker *GATA3* were relatively highly expressed in TNBC tumors with favorable prognosis (Fig. 2A). *GATA3* is recognized as a marker of luminal ER-positive breast tumor and there is a strong relationship between coexpression of ER α and *GATA3*, as reported in the literature (20, 21). Here, we observed that *GATA3* was moderately-to-high expressed in approximately 30% of 49 TNBCs. Similarly, among the 313 HER2-negative patients (all available cases from MDACC; Fig. 2B), some ER-negative tumors expressed *GATA3* as high as that in ER-positive tumors. ER-negative tumors had a wide range of *GATA3* expression, compared with ER-positive ones. We further plotted the *GATA3* expression in 313 patients according to PAM50-predicted subtypes (22), and the results reconfirmed a wide range of *GATA3* expression in these non-luminal tumors (Fig. 2C). No obvious association between *GATA3* and *ESR1* expression was observed in these basal-like cases (Fig. 2D). Thus, *GATA3* expression is present in a subset of TNBC.

In the discovery set, the 7-gene prognostic signature had PPV of 95.4% (95% CI: 81.7%–99.6%) and NPV (DRFS) of 100% (95% CI: 80%–100%) for the first 3

years after diagnosis (Table 3). Compared with other clinicopathologic factors available, the 7-gene signature was the only factor that could effectively predict the outcome of patients with TNBC and residual disease after neoadjuvant chemotherapy (log-rank P for 7-gene signature, <0.001 ; for age, 0.301; for tumor size, 0.114; for nodes status, 0.810; and for grade, 0.737).

Association between chemoresistant prognosis-relevant subgroups and the Pietenpol's molecular subtypes

As mentioned above, the 2 prognosis-relevant subgroups (early relapse vs. nonrelapse) could be molecularly defined by the 7-gene signature. The subgroup expressing high luminal-like genes (*AR*, *GATA3*) was associated with good prognosis, whereas the subgroup expressing high cancer stem cell-like genes (*WNT11*, *MMP28*) was related to early metastasis. Recently, Lehmann and colleagues (3) identified 6 TNBC subtypes including 2 basal-like (BL1/2), an immunomodulatory (IM), a mesenchymal (M), a mesenchymal stem-like (MSL), and a luminal AR (LAR) subtype. We examined the expression of our 7 genes in their 587 TNBC samples (Fig. 2E). The values for *AR*, *GATA3*, and *KRT16* were higher than the rest of the genes. There was a clear absence of *GATA3* in the M and MSL subtypes. *AR* and *GATA3* were enriched in LAR subtype, whereas *WNT11* and *MMP28* were commonly expressed in M and MSL subtypes.

Performance of the prognostic signature in validation cohort

Independent validation was conducted in a second cohort from TMH-BCM which included 25 TNBC. This cohort was followed-up for median 36 months, with 3-year RFS of 48.0% (95% CI: 27.8%–65.6%).

In this independent cohort, this 7-gene signature predicted correctly prognosis for 9 of 12 patients predicted not to relapse in 3 years [NPV (RFS), 76.9%], and for 10 of 13 patients predicted to relapse in 3 years (PPV, 75.0%; Table 3;

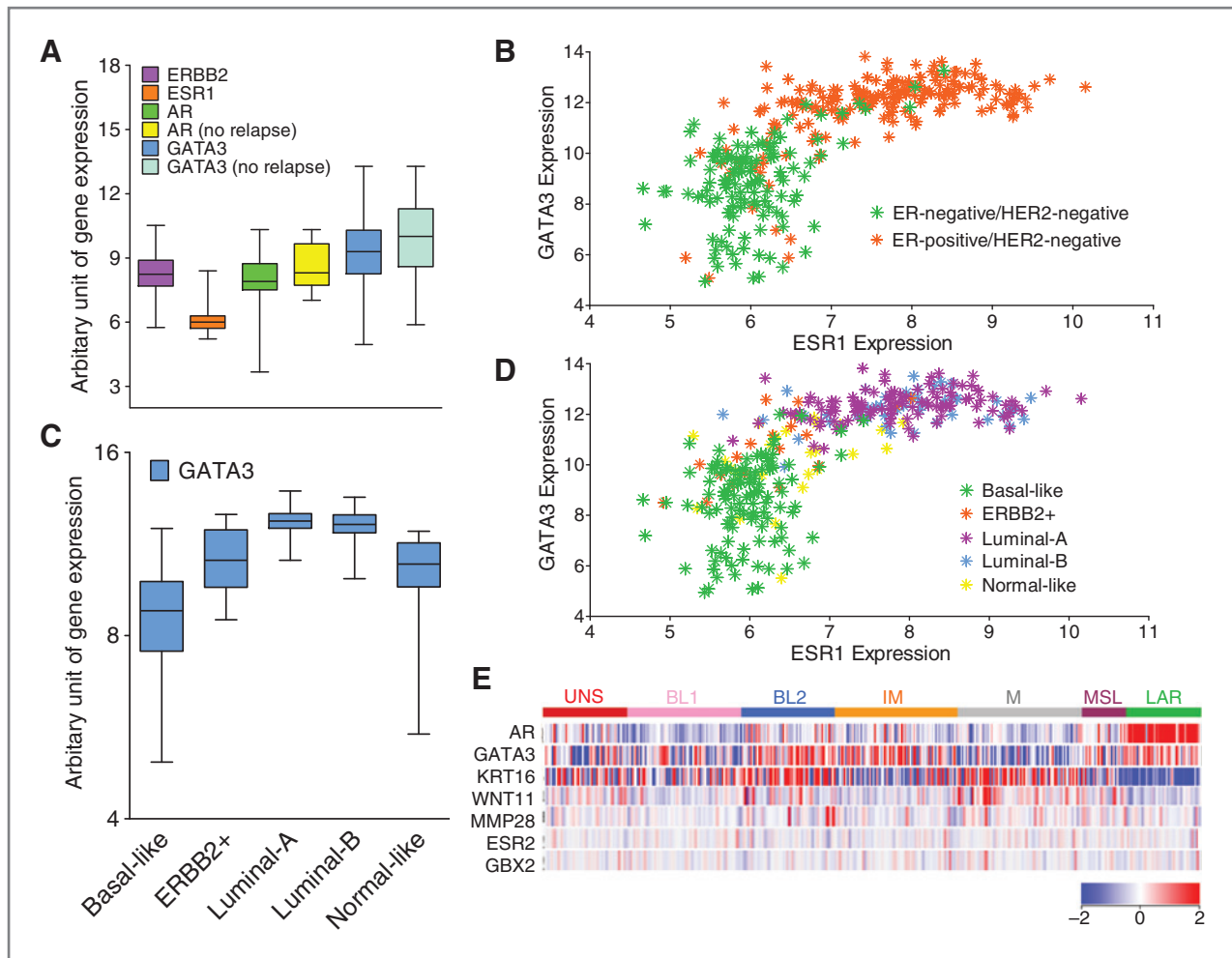


Figure 2. Expression of featured genes in the signature. **A**, expression of 2 luminal-like genes (AR and GATA3) in 49 chemoresistant TNBC in the MDACC cohort. Expressions of ESR1 and ERBB2 are shown as reference for "low-expression" and "nonoverexpression," respectively. **B**, the association between ESR1 and GATA3 in 313 HER2-negative breast cancer (all available cases from MDACC). **C**, box plot of expression of GATA3 in 313 patients according to PAM50-predicted subtypes. **D**, association between ESR1 and GATA3 in 313 patients according to PAM50-predicted subtypes. **E**, association between the 7-gene defined subgroups and Pietenpol's TNBC subtypes classification in 587 TNBC in Pietenpol's study. BL1 and BL2, basal-like 1 and 2; IM, immunomodulatory; M, mesenchymal; MSL, mesenchymal stem-like; LAR, luminal AR. The color scale is also shown: the red color represents expression level above mean expression of a gene across all samples, the white color represents mean expression, and the blue color represents expression lower than the mean.

Kaplan–Meier plots in Fig. 3A). Thus, the 3-year RFS estimate for the patients predicted to have good prognosis was 76.9%, compared with those predicted to relapse within 3 years was only 25.0%. Similarly, the likelihood ratio for relapse versus absence of 3-year relapse was 4.67 (95% CI: 1.27–17.15), after adjustment for other clinicopathologic factors (Table 3).

Extended validation in patients with operable TNBC treated with adjuvant chemotherapy

The 7-gene signature was useful in predicting the prognosis of patients with TNBC with known resistance to neoadjuvant chemotherapy. Its use in the adjuvant TNBC was unclear. To validate the use of this signature in patients with adjuvant TNBC, chemoresistance to treatment was first determined. The previously established signature (7) which

could discriminate between chemoresistant (RCB-II/III) and chemosensitive (pCR or RCB-I) in ER-negative and HER2-negative patients was used. As expected, the 7-gene signature could not accurately predict the prognosis in patients predicted to be chemosensitive (log rank $P = 0.172$; data not shown). In contrast, the 7-gene signature discriminated well in women predicted to be chemoresistant, either in the overall cohort (Fig. 3B) or in each subset (log rank P significant in set I and II; borderline in set III and IV; Fig. 3C–F).

Regarding the degree of accuracy, patients predicted not to relapse exhibited high 3-year RFS (NPV) of 78.9% (95% CI: 72.4%–84.1%) compared with only 48.5% (95% CI: 37.0%–59.0%) for those predicted to relapse (calculated by "1-PPV"; Table 3). The results were concordant in each set, indicating the robustness of prediction. Moreover, the

Table 3. Performance of signature for predicting prognosis of patients with chemoresistant TNBC

	Discovery cohort	Validation cohort		Extended validation cohort			
	MDACC	TMH-BCM	Set I, Rotterdam	Set II, Frankfurt	Set III, New York/San Francisco/Stockholm/Uppsala	Set IV, Mainz/TransBIG/Oxford/others	
PPV (95% CI) at 3-year (CDRR or CRR)	95.4 (81.7–99.6)	75.0 (40.8–91.2)	66.7 (49.1–83.2)	56.3 (34.4–80.2)	34.8 (19.2–57.7)	41.7 (19.1–73.0)	
NPV (95% CI) at 3-year (DRFS or RFS)	100 (80.0–100.0)	76.9 (52.5–94.4)	73.3 (60.2–82.7)	77.0 (59.0–87.9)	83.6 (70.9–91.1)	82.1 (66.0–91.0)	
Univariate log rank P value at 3-year	<0.0001	0.009	<0.0001	0.006	0.063	0.066	
Multivariate hazard ratio at 3-year ^a	/	4.67 (1.27–17.15)	4.00 (1.21–13.20)	3.55 (1.77–7.14)	2.40 (0.93–6.23)	2.46 (0.77–7.85)	
Univariate log rank P value during follow-up	<0.0001	0.017	0.001	0.013	0.053	0.105	
Multivariate hazard ratio during follow-up ^a	/	3.39 (1.14–10.10)	2.92 (1.51–5.65)	3.08 (1.10–8.63)	1.64 (0.66–4.09)	1.88 (0.75–4.75)	

Abbreviations: CDRR, cumulative distant relapse rate; CI, confidence interval; CRR, cumulative relapse rate; DRFS, distant relapse-free survival; RFS, relapse-free survival; NPV, negative predictive value; PPV, positive predictive value

^aMultivariate hazard ratio was adjusted for age, lymph nodes status, tumor size stage, grade, and neoadjuvant chemotherapy regimen in TMH-BCM cohort; for age, lymph nodes status, tumor size stage, grade, adjuvant chemotherapy, and subset group in the overall extended validation cohort; for age, tumor size stage, and grade in set I and IV; for age, lymph nodes status, tumor size stage, grade, and adjuvant chemotherapy in set II and III.

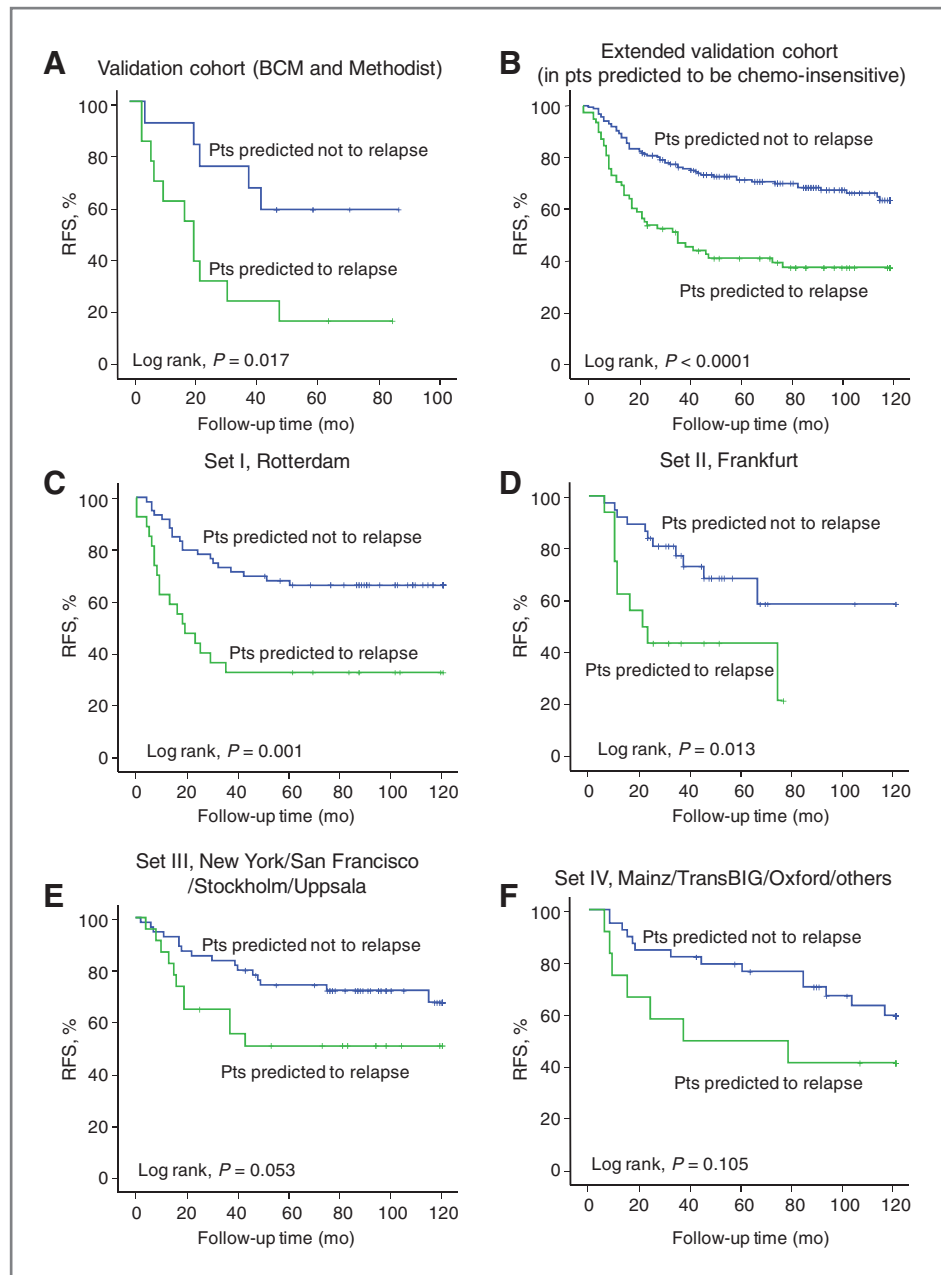


Figure 3. Kaplan–Meier estimates of RFS according to 7-gene chemoresistant prognostic signature in validation cohorts. A, validation in the TMH-BCM cohort; B, extended validation in the whole cohort of operative TNBC without neoadjuvant chemotherapy but predicted to be chemoinensitive by an approximation of a JAMA predictor. C–F, four subsets within the whole cohort of operative TNBC. Vertical ticks on the curves indicate censored observations.

prediction of relapse by our signature was independent to clinicopathologic factors such as nodal status, tumor size, age, etc. (Fig. A1). After adjustment, the 7-gene signature was independently and significantly associated with risk of relapse in 3 years among adjuvant TNBC predicted to be chemoresistant (HR = 2.61; 95% CI: 1.52–4.49). Of note, the 7-gene signature had limits in predicting long-term relapse beyond 3 years and the predicted results were less reliable beyond this time frame. Relapse in TNBC after median follow-up of 3 years is rare, and the loss of prognostic accuracy of this signature most likely reflects small sample sizes.

Discussion

Although patients with TNBC with residual disease after neoadjuvant chemotherapy have worse survival than those with luminal subtypes (5), some of them do not relapse for a long time. In this study, we used gene expression data of patients with TNBC and residual disease and different prognosis to molecularly define the clinically relevant subgroups, and developed a 7-gene prognostic signature for chemoresistant TNBCs. A favorable prognosis was observed in patients with TNBC tumors displaying high expression of "luminal-like" genes (AR, GATA3), whereas decreased survival was

observed in patients with TNBC tumors expressing cancer stem cell-like (WNT11) or EMT-associated genes (MMP28). The signature not only predicted 3-year RFS, but also showed a clinically meaningful survival difference between patients predicted to relapse versus no relapse. Furthermore, the signature is the only significant marker that can effectively predict prognosis of chemoresistant TNBC in a multivariate clinicopathologic model (including age, tumor size, nodal status, grade, and adjuvant chemotherapy).

Although the majority of TNBCs classified as basal-like (1, 3, 23), the clinically diagnosed TNBC is a heterogeneous collection of distinct phenotypes (3). Our study, unlike previous reports, focuses on only chemoresistant TNBC and subdivides these cancers according to relapse outcomes. A simple combination of luminal-like genes and cancer stem cell-like genes defines the subgroup of TNBC with relatively favorable or unfavorable survival. Our discovery also challenge the value of non-pCR in TNBC and the universal applicability of the concept that non-pCR in TNBC equals to recurrence or poor survival.

AR and its ligand androgens may have some essential role in breast cancer (24). AR expression was found in 20% to 30% of the cases with TNBC (25, 26). Most studies confirm a significantly positive correlation between AR expression and favorable survival in patients with TNBC (3, 26, 27). Our study suggests a relatively favorable prognosis in patients with chemoresistant TNBC with higher expression of AR. Several novel and druggable pathways, including AR, are being studied in patients with TNBC (3). Another marker defining the favorable prognosis is GATA3. Previously, studies have shown that GATA3 expression is highly correlated with ER α (encoded by ESR1; refs. 20, 21). We confirmed a high coincidence between GATA3 and ESR1 at mRNA level; however, when ESR1 is lowly expressed, the range of GATA3 expression is wide and approximately 30% to 40% of TNBCs have moderately-to-high expression of GATA3. Low GATA3 expression is associated with aggressive phenotype, and in most studies, worse RFS (28–31). Increasing evidence indicates that the role of GATA3 is not ER-dependent and that GATA3 is functional in TNBC cells (29, 31, 32). Expression of GATA3 reprograms TNBCs to a less aggressive phenotype (30).

The subgroup with unfavorable prognosis is characterized by the stem cell-like and EMT-associated genes. Inhibitors of WNT/ β -catenin are of great interest for such a subtype and they currently are in preclinical development (33).

Central to this study is whether chemotherapy is still needed for the TNBC subgroup with relatively favorable prognosis. With only a 78% 3-year RFS, chemotherapy as yet cannot be avoided. Generally, by consensus, low-clinical risk group is defined as patients with 10-year overall survival probabilities of at least 92% for ER-negative tumors (34). Thus, even for patients with TNBC predicted not to relapse, some alternate therapy to further decrease the risk of relapse is needed. However, the nature

of these chemoresistant cases implies limited benefits from standard chemotherapy. Novel treatment strategies based on the biologic features of chemoresistant TNBC need to be developed. According to our study, there are 2 main entities in chemoresistant TNBCs: one is AR-related luminal-like tumors and the other is stem cell-like tumors. For the former, an AR antagonist might be more effective than traditional chemotherapy (3); for the latter, targeting proteins involved in cell-renewal or EMT may provide a more reasonable therapeutic strategy (3), as chemotherapy may not effectively eliminate tumor-initiating cells (35).

Our observation is important because most currently available genomic prognostic signatures (e.g., 70-gene profile, Recurrence Score, Genomic Grading Index) assign poor prognostic risk status to all TNBC samples despite their variable outcomes. A few signatures have been developed to allow prognostic stratification of TNBC cancers with consideration of the chemosensitivity of the tumors (12, 36). The implication of our study is that, ER/PgR/HER2 biomarkers have some limitation in defining a subtype with similar biologic behavior, a patient with TNBC could have received an inadequate or untargeted treatment, and a more accurate evaluation of TNBC biology before planning neoadjuvant/adjuvant treatment is needed. Our signature, for the first time, considers chemosensitivity and excludes chemosensitive cases who achieve pCR and have excellent prognosis (5), and focuses primarily on the chemoresistant tumors. Our 7-gene signature has the potential to assist treatment decision-making (e.g., guide to participate appropriate clinical trials) and predict clinical outcomes for chemoresistant TNBC. Of note, our signature should be used only in patients proven or predicted to be chemoresistant. There is a need for studies introducing molecularly targeted therapies in the adjuvant management of patients with TNBC and the strategies to prospectively validate the signature as well as the novel therapeutic approach.

This study has several limitations. First, the sample size in the discovery cohort and in the homogeneous validation cohort is limited. Although our signature is successfully validated in the extended validation, further optimization is needed. Second, we used the normalized gene expression data as provided in public databases (12); no attempts to renormalize the microarray data were made, although a robust rescaling procedure ensured that the gene expressions were similarly distributed across datasets.

In conclusion, we have developed a prognostic classifier specific to chemoresistant TNBC. It is derived from patients with TNBC receiving neoadjuvant chemotherapy, and is further validated in patients with either locally advanced disease or operable tumors. This signature outperforms the classical clinicopathologic features in predicting the prognosis of chemoresistant TNBC. More importantly, biologically relevant genes included in the signature might provide new potential therapeutic targets. Further validation in a large prospective series and additional research on new therapeutic strategy for chemoresistant TNBC is warranted.

Information of Microarray Data

Gene expression data have been deposited into the GEO database (<http://www.ncbi.nlm.nih.gov>) under accession identification numbers GSE25066 (discovery cohort from MDACC), GSE43502 (validation cohort from TMH-BCM), and GSE31519 (extended validation cohort).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: K. Yu, F. Symmans, J.C. Chang

Development of methodology: K. Yu, R. Zhu, J.C. Chang

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): K. Yu, A. Makris, I.A. Mayer, F. Symmans, J.C. Chang

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): K. Yu, R. Zhu, M. Zhan, S.T.C. Wong, B.D. Lehmann, X. Chen, J.A. Pietenpol, J.C. Chang

Writing, review, and/or revision of the manuscript: K. Yu, R. Zhu, M. Zhan, W. Yang, A. Makris, B.D. Lehmann, J.A. Pietenpol, J.C. Chang

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.T.C. Wong, F. Symmans
Study supervision: A.A. Rodriguez, Z. Shao

Grant Support

This research is supported by grants from the Breast Cancer Research Foundation (to J. Chang); Komen Promise for the Cure KG081694 (to J. Chang); NIH U54 CA 149196 (to S. Wong, J. Chang, R. Zhu, and M. Zhan); NIH R01 CA138197 (to J. Chang); National Natural Science Foundation of China (30971143, 30972936, 81001169; to K. Yu, Z. Shao); the Zhuo-Xue Project of Fudan University; and the Shanghai Committee of Science and Technology Fund for 2011 Qimingxing Project (to K. Yu).

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 September 17, 2012; revised February 8, 2013; accepted March 5, 2013; published OnlineFirst April 2, 2013.

References

- Metzger-Filho O, Tutt A, de Azambuja E, Saini KS, Viale G, Loi S, et al. Dissecting the heterogeneity of triple-negative breast cancer. *J Clin Oncol* 2012;30:1879–87.
- Blows FM, Driver KE, Schmidt MK, Broeks A, van Leeuwen FE, Wesseling J, et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med* 2010;7:e1000279.
- 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.
- Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 2012;486:395–9.
- Liedtke C, Mazouni C, Hess KR, Andre F, Tordai A, Mejia JA, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J Clin Oncol* 2008;26:1275–81.
- Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin Cancer Res* 2007;13:2329–34.
- Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A, et al. A genomic predictor of response and survival following taxane–anthracycline chemotherapy for invasive breast cancer. *JAMA* 2011;305:1873–81.
- Symmans WF, Peintinger F, Hatzis C, Rajan R, Kuerer H, Valero V, et al. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. *J Clin Oncol* 2007;25:4414–22.
- Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, Mejia JA, et al. Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* 2006;24:4236–44.
- Chevallier B, Chollet P, Merrouche Y, Roche H, Fumoleau P, Kerbrat P, et al. Lenograstim prevents morbidity from intensive induction chemotherapy in the treatment of inflammatory breast cancer. *J Clin Oncol* 1995;13:1564–71.
- Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Tham YL, et al. Patterns of resistance and incomplete response to docetaxel by gene expression profiling in breast cancer patients. *J Clin Oncol* 2005;23:1169–77.
- Rody A, Karn T, Liedtke C, Pusztai L, Ruckhaeberle E, Hankaer L, et al. A clinically relevant gene signature in triple negative and basal-like breast cancer. *Breast Cancer Res* 2011;13:R97.
- Karn T, Metzler D, Ruckhaeberle E, Hankaer L, Gätje R, Solbach C, et al. Data driven derivation of cutoffs from a pool of 3,030 Affymetrix arrays to stratify distinct clinical types of breast cancer. *Breast Cancer Res Treat* 2010;120:567–79.
- Warnat P, Eils R, Brors B. Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes. *BMC Bioinformatics* 2005;6:265.
- Glaab E, Garibaldi JM, Krasnogor N. ArrayMining: a modular web-application for microarray analysis combining ensemble and consensus methods with cross-study normalization. *BMC Bioinformatics* 2009;10:358.
- Yu KD, Li S, Shao ZM. Different annual recurrence pattern between lumpectomy and mastectomy: implication for breast cancer surveillance after breast-conserving surgery. *Oncologist* 2011;16:1101–10.
- Voogd AC, Nielsen M, Peterse JL, Blichert-Toft M, Bartelink H, Overgaard M, et al. Differences in risk factors for local and distant recurrence after breast-conserving therapy or mastectomy for stage I and II breast cancer: pooled results of two large European randomized trials. *J Clin Oncol* 2001;19:1688–97.
- Mansell J, Monypenny IJ, Skene AI, Abram P, Carpenter R, Gattuso JM, et al. Patterns and predictors of early recurrence in postmenopausal women with estrogen receptor-positive early breast cancer. *Breast Cancer Res Treat* 2009;117:91–8.
- Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* 2001;98:31–6.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869–74.
- Voduc D, Cheang M, Nielsen T. GATA-3 expression in breast cancer has a strong association with estrogen receptor but lacks independent prognostic value. *Cancer Epidemiol Biomarkers Prev* 2008;17:365–73.
- Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009;27:1160–7.
- Kreike B, van Kouwenhove M, Horlings H, Weigelt B, Peterse H, Bartelink H, et al. Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. *Breast Cancer Res* 2007;9:R65.
- Liao DJ, Dickson RB. Roles of androgens in the development, growth, and carcinogenesis of the mammary gland. *J Steroid Biochem Mol Biol* 2002;80:175–89.
- Loibl S, Muller BM, von Minckwitz G, Schwabe M, Roller M, Darb-Esfahani S, et al. Androgen receptor expression in primary breast cancer and its predictive and prognostic value in patients treated with neoadjuvant chemotherapy. *Breast Cancer Res Treat* 2011;130:477–87.
- He J, Peng R, Yuan Z, Wang S, Peng J, Lin G, et al. Prognostic value of androgen receptor expression in operable triple-negative breast

- cancer: a retrospective analysis based on a tissue microarray. *Med Oncol* 2012;29:406–10.
27. Gonzalez-Angulo AM, Stemke-Hale K, Palla SL, Carey M, Agarwal R, Meric-Bertram F, et al. Androgen receptor levels and association with PIK3CA mutations and prognosis in breast cancer. *Clin Cancer Res* 2009;15:2472–8.
 28. Mehra R, Varambally S, Ding L, Shen R, Sabel MS, Ghosh D, et al. Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res* 2005;65:11259–64.
 29. Dydensborg AB, Rose AA, Wilson BJ, Grote D, Paquet M, Giguere V, et al. GATA3 inhibits breast cancer growth and pulmonary breast cancer metastasis. *Oncogene* 2009;28:2634–42.
 30. Chu IM, Michalowski AM, Hoenerhoff M, Szauter KM, Luger D, Sato M, et al. GATA3 inhibits lysyl oxidase-mediated metastases of human basal triple-negative breast cancer cells. *Oncogene* 2012;31:2017–27.
 31. Yan W, Cao QJ, Arenas RB, Bentley B, Shao R. GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition. *J Biol Chem* 2010;285:14042–51.
 32. Usary J, Llaca V, Karaca G, Presswala S, Karaca M, He X, et al. Mutation of GATA3 in human breast tumors. *Oncogene* 2004;23:7669–78.
 33. MacDonald BT, Tamai K, He X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell* 2009;17:9–26.
 34. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 2006;98:1183–92.
 35. Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM, Sjolund A, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc Natl Acad Sci U S A* 2009;106:13820–5.
 36. Sabatier R, Finetti P, Cervera N, Lambaudie E, Esterni B, Mamessier E, et al. A gene expression signature identifies two prognostic subgroups of basal breast cancer. *Breast Cancer Res Treat* 2011;126:407–20.