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

Ke-Da Yu, Rui Zhu, Ming Zhan, Angel A. Rodriguez, Wei Yang, Andreas Makris, Brian D. Lehmann, Xi Chen, Ingrid Mayer, Jennifer A. Pietenpol, Zhi-Ming Shao, W. Fraser Symmans, and Jenny C. Chang

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.

Clinical Cancer Research; 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.
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
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 the cumulative relapse or death rate for patients predicted not to relapse or death in 3 years.

### Table 1. Pretreatment characteristics of the discovery and validation cohorts

<table>
<thead>
<tr>
<th></th>
<th>MDACC</th>
<th>Discovery cohort</th>
<th>TMH-BCM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original TNBC source (n = 111)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Follow-up time, mo</td>
<td>Median (range)</td>
<td>25 (1–79)</td>
<td>36 (4–88)</td>
</tr>
<tr>
<td>Age, y</td>
<td>≤50</td>
<td>59 (53.2)</td>
<td>23 (46.9)</td>
</tr>
<tr>
<td></td>
<td>&gt;50</td>
<td>52 (46.8)</td>
<td>26 (53.1)</td>
</tr>
<tr>
<td>Nodal status</td>
<td>Negative</td>
<td>26 (23.4)</td>
<td>9 (18.4)</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>85 (76.6)</td>
<td>40 (81.6)</td>
</tr>
<tr>
<td>Tumor size stage</td>
<td>T0–2</td>
<td>71 (64.0)</td>
<td>24 (49.0)</td>
</tr>
<tr>
<td></td>
<td>T3–4</td>
<td>40 (36.0)</td>
<td>25 (51.0)</td>
</tr>
<tr>
<td>Grade</td>
<td>I</td>
<td>0 (0.0)</td>
<td>1 (0.0)</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>13 (11.7)</td>
<td>9 (18.4)</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>98 (88.3)</td>
<td>40 (81.6)</td>
</tr>
<tr>
<td>pCR</td>
<td>No</td>
<td>69 (65.7)</td>
<td>49 (100.0)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>36 (34.3)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>RCB</td>
<td>I</td>
<td>48 (46.6)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>30 (29.1)</td>
<td>24 (51.1)</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>25 (24.3)</td>
<td>23 (48.9)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Neoadjuvant and adjuvant CT</td>
<td>P × 12—FAC × 4—Surgery—no CT</td>
<td>111 (100.0)</td>
<td>49 (100.0)</td>
</tr>
<tr>
<td></td>
<td>T × 4—Surgery—AC × 4</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>AC × 4—Surgery—T × 4</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

Abbreviations: AC, doxorubicin and cyclophosphamide; CT, chemotherapy; FAC, fluorouracil, doxorubicin and cyclophosphamide; N.A., not assessed; P, paclitaxel; T, docetaxel.
Table 2. Characteristics of the extended validation cohort

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

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 (GSE4394, 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.

Yu et al.

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 (StataCorp LP.). Two-sided P < 0.05 was considered statistically significant.

Results

Establishment and performance of the prognostic signature in discovery cohort

To determine whether chemoprognostic 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
differentially expressed between the 2 groups, with at least a 2.14 (2.14)-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 marker 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\( \alpha \) and GATA3, as reported in the literature (20, 21). Here, we observed that GATA3 was moderately-to-highly 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 confirmed 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, GATA 3 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;
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

<table>
<thead>
<tr>
<th></th>
<th>Discovery cohort</th>
<th>Validation cohort</th>
<th>Extended validation cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MDACC</td>
<td>TMH-BCM</td>
<td>Overall</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Set I, Rotterdam</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Set II, Frankfurt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Set III, New York/ San Francisco/Uppsala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Set IV, Mainz/ TransBIG/ Oxford/others</td>
</tr>
<tr>
<td>PPV (95% CI) at 3-year CDRR or CRR</td>
<td>95.4 (81.7–99.6)</td>
<td>75.0 (40.8–91.2)</td>
<td>51.5 (41.0–63.0)</td>
</tr>
<tr>
<td>NPV (95% CI) at 3-year DRFS or RFS</td>
<td>100 (80.0–100.0)</td>
<td>76.9 (52.5–94.4)</td>
<td>78.9 (72.4–84.1)</td>
</tr>
<tr>
<td>Univariate log rank P value at 3-year</td>
<td>&lt;0.0001</td>
<td>0.009</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Multivariate hazard ratio at 3-year</td>
<td>/</td>
<td>4.67 (1.27–17.15)</td>
<td>2.61 (1.52–4.49)</td>
</tr>
<tr>
<td>Univariate log rank P value during follow-up</td>
<td>&lt;0.0001</td>
<td>0.017</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Multivariate hazard ratio during follow-up</td>
<td>/</td>
<td>3.39 (1.14–10.10)</td>
<td>2.11 (1.33–3.34)</td>
</tr>
</tbody>
</table>

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.

*multivariate 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, and neoadjuvant 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.
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 ERα (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, 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/PrR/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 microarray 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 GSE15119 (extended validation cohort).

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors' Contributions
Conceptualization and design: K. Yu, F. Symmans, 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, bioinformatics, computational analysis): K. Yu, R. Zha, M. Zhan, S.T.C. Wong, B.D. Lehmann, X. Chen, I.A. Pietinenpol, J.C. Chang
Writing, review, and/or revision of the manuscript: K. Yu, R. Zha, M. Zhan, W. Yang, A. Makris, B.D. Lehmann, I.A. Pietinenpol, J.C. Chang

References


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

Ke-Da Yu, Rui Zhu, Ming Zhan, et al.