

Predictive Biomarkers and Personalized Medicine**Evaluation of a 30-Gene Paclitaxel, Fluorouracil, Doxorubicin, and Cyclophosphamide Chemotherapy Response Predictor in a Multicenter Randomized Trial in Breast Cancer**

Adel Tabchy¹, Vicente Valero¹, Tatiana Vidaurre⁵, Ana Lluch⁶, Henry Gomez⁵, Miguel Martin⁶, Yuan Qi², Luis Javier Barajas-Figueroa⁷, Eduardo Souchon⁴, Charles Coutant¹, Franco D. Doimi⁵, Nuhad K. Ibrahim¹, Yun Gong³, Gabriel N. Hortobagyi¹, Kenneth R. Hess², W. Fraser Symmans³, and Lajos Pusztai¹

Abstract

Purpose: We examined in a prospective, randomized, international clinical trial the performance of a previously defined 30-gene predictor (DLDA-30) of pathologic complete response (pCR) to preoperative weekly paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide (T/FAC) chemotherapy, and assessed if DLDA-30 also predicts increased sensitivity to FAC-only chemotherapy. We compared the pCR rates after T/FAC versus FACx6 preoperative chemotherapy. We also did an exploratory analysis to identify novel candidate genes that differentially predict response in the two treatment arms.

Experimental Design: Two hundred and seventy-three patients were randomly assigned to receive either weekly paclitaxel × 12 followed by FAC × 4 (T/FAC, $n = 138$), or FAC × 6 ($n = 135$) neoadjuvant chemotherapy. All patients underwent a pretreatment fine-needle aspiration biopsy of the tumor for gene expression profiling and treatment response prediction.

Results: The pCR rates were 19% and 9% in the T/FAC and FAC arms, respectively ($P < 0.05$). In the T/FAC arm, the positive predictive value (PPV) of the genomic predictor was 38% [95% confidence interval (95% CI), 21-56%], the negative predictive value was 88% (95% CI, 77-95%), and the area under the receiver operating characteristic curve (AUC) was 0.711. In the FAC arm, the PPV was 9% (95% CI, 1-29%) and the AUC was 0.584. This suggests that the genomic predictor may have regimen specificity. Its performance was similar to a clinical variable-based predictor nomogram.

Conclusions: Gene expression profiling for prospective response prediction was feasible in this international trial. The 30-gene predictor can identify patients with greater than average sensitivity to T/FAC chemotherapy. However, it captured molecular equivalents of clinical phenotype. Next-generation predictive markers will need to be developed separately for different molecular subsets of breast cancers. *Clin Cancer Res*; 16(21); 5351-61. ©2010 AACR.

There are several clinical and molecular features that can identify generally more or less chemotherapy-sensitive subsets of breast cancers, but there are no clinically useful predictive biomarkers to select one chemotherapy regimen over another. Preoperative (neoadjuvant) chemotherapy provides a direct opportunity to assess treatment sensitivity in early-stage breast cancers, and pathologic complete response (pCR) is a powerful early surrogate of long-term

survival (1, 2). Recent studies have established that basal-like or triple receptor-negative breast cancers include a greater proportion of highly chemotherapy-sensitive tumors, reflected by the significantly higher pCR rates, compared with estrogen receptor (ER)-positive breast cancers (3, 4). Among the ER-positive cancers, high Oncotype DX recurrence score (Genomic Health, Inc.), luminal B molecular class, HER2 overexpression, and high histologic or

Authors' Affiliations: Departments of ¹Breast Medical Oncology, ²Bioinformatics and Computational Biology, and ³Pathology, The University of Texas M.D. Anderson Cancer Center; ⁴Lyndon B Johnson General Hospital, and The University of Texas Health Science Center, Houston, Texas; ⁵Instituto Nacional de Enfermedades Neoplásicas, Lima, Peru; ⁶Grupo Español de Investigación en Cáncer de Mama, Madrid, Spain; and ⁷Centro Médico Nacional de Occidente, Guadalajara, Mexico

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

Presented in part as an abstract at the 32nd Annual AACR San Antonio

Breast Cancer Symposium, December 9, 2009, San Antonio, Texas. For this, the first author (A. Tabchy) received an AACR Translational Research Scholar Award for Breast Cancer Research, supported by a grant from Susan G. Komen for the Cure.

Corresponding Author: Lajos Pusztai, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 424, Houston, TX 77230-1439. Phone: 713-792-2817; Fax: 713-794-4385; E-mail: lpusztai@mdanderson.org.

doi: 10.1158/1078-0432.CCR-10-1265

©2010 American Association for Cancer Research.

Translational Relevance

There are several clinical and molecular features that can identify generally more or less chemotherapy-sensitive subsets of breast cancers, but there are no clinically useful predictive biomarkers that can guide the selection of one chemotherapy regimen over another in individual patients. We tested a 30-gene predictor of pathologic complete response to neoadjuvant sequential weekly paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide (T/FAC) chemotherapy from fine-needle biopsies of breast cancers in a prospective randomized international trial. We also compared the treatment efficacy of T/FAC versus FACx6 chemotherapies. The results confirm the ability of the genomic predictor to identify patients with higher than average sensitivity to T/FAC chemotherapy but not to FAC treatment. Prospective gene expression biomarker analysis was technically and logistically feasible in this multicenter international trial. Most current molecular response predictors in breast cancer including this one suffer from capturing molecular equivalents of clinical phenotype. To improve their clinical utility, second-generation genomic predictors will need to be developed separately for the different molecular and phenotypic subsets of breast cancers.

genomic grade are associated with greater chemotherapy sensitivity (5–7). However, these molecular and clinicopathologic variables seem to predict general chemotherapy sensitivity, and have limited value in guiding the choice of a specific treatment regimen. Among several other markers, high expression/amplification of *topoisomerase II α* and low expression of microtubule binding protein Tau have recently been suggested as potential predictors of sensitivity to anthracyclines and taxanes, respectively (8, 9). However, neither these nor other proposed markers showed consistent clinically useful predictive value (10–12).

We previously developed a genomic predictor of pCR to preoperative sequential weekly paclitaxel followed by fluorouracil, doxorubicin, and cyclophosphamide (T/FAC) chemotherapy from a single-arm retrospective study that included 82 patients in the discovery and 51 in the validation phase (13). The genomic predictor uses information from 30 different probe sets (i.e., genes) and uses diagonal linear discriminant analysis for prediction rules and is therefore referred to as the DLDA-30 predictor. The goal of the current study was to evaluate the predictive performance and assess the regimen specificity of this multigene predictor in a prospective, two-arm, randomized, multicenter, international, neoadjuvant clinical trial. Two commonly used standard chemotherapy regimens, including T/FAC and FAC alone, were compared for pCR rates. The predictive performance of the genomic signature was assessed independently in each treatment arm, and a

marker-treatment-outcome interaction test was done. We also examined the performance of our previously reported clinical-pathologic variable-based nomogram to predict pCR (14). It remains unknown if the sequential paclitaxel/FAC regimen is superior to six courses of FAC or FEC (the control arm of this study) in terms of pathologic response rates or survival. Therefore, we also compared the pCR rates for patients who received T/FAC versus FACx6 preoperative chemotherapy in this trial.

Materials and Methods

Patient eligibility

Patients with clinical stage I to III breast cancer were eligible. Histologic diagnosis of invasive cancer and ER, progesterone receptor (PR), and HER2 receptor status were determined from a diagnostic core needle or incisional biopsy before therapy. All patients had to agree to a separate, pretreatment research fine-needle aspiration (FNA) of the cancer for gene expression analysis. Patients were accrued at six international sites including The University of Texas M.D. Anderson Cancer Center (MDACC; $n = 96$) and the Lyndon B Johnson General Hospital ($n = 19$) in Houston, Texas; the Instituto Nacional de Enfermedades Neoplásicas in Lima, Peru ($n = 79$); the Centro Medico Nacional de Occidente in Guadalajara, Mexico ($n = 19$); and the clinical trial group Grupo Español de Investigación en Cáncer de Mama in Spain ($n = 60$). This study was approved by the institutional review boards of each participating institution, and all patients signed an informed consent for voluntary participation. The study was conducted between October 2003 and October 2006.

Treatment

Treatment was not selected based on gene expression results, and patients were centrally randomized with blocked randomization at MDACC into one of two treatment arms: arm A: (T/FAC) weekly paclitaxel ($80 \text{ mg/m}^2/\text{wk}$) \times 12 courses followed by 5-fluorouracil (500 mg/m^2), doxorubicin (50 mg/m^2), and cyclophosphamide (500 mg/m^2) all on day 1 repeated in 21-day cycles \times 4 courses (100 mg/m^2 epirubicin could be substituted for doxorubicin at the discretion of local investigators); arm B: FAC (or FEC if epirubicin was used) \times 6 courses at the same doses and schedule as above. Toxicity information was not collected during this trial because the two treatment arms were considered standard community-based therapy. Any patient developing unacceptable grade 3 or grade 4 toxicity was removed from the study. Patients with clinical or radiological disease progression were considered as having residual disease (RD) in the final analysis. After completion of neoadjuvant chemotherapy, all patients underwent modified radical mastectomy or lumpectomy and sentinel lymph node biopsy or axillary node dissection as determined by the surgeon; pCR was defined as the complete absence of invasive cancer cells in the breast and lymph nodes (15). Pathologic response was centrally reviewed by a breast pathologist (W.F.S.).

Gene expression analysis and response prediction

Two to three FNA passes obtained with 23- or 25-gauge needles were collected into vials containing 1 mL of RNA-later solution (Ambion) and stored at 4°C until mailed to MDACC in a cooler pack or dry ice. At MDACC, specimens were stored at -80°C until gene expression profiling on Affymetrix U133A gene chips. The same array platform, standard operating procedure, and normalization method (dCHIP) was used as previously reported (13). The reference chip and normalization procedure are available online (<http://www.bioinformatics.mdanderson.org/pubdata.html>). FNA samples contain on average 80% neoplastic cells and little or no normal breast epithelium or stromal cells (16). Gene expression information generated from FNAs represents the molecular characteristics of the invasive cancer, including the molecular class (3). RNA was extracted from FNA samples using the RNeasy kit (Qiagen). The amount and quality of RNA were assessed with DU-640 UV Spectrophotometer (Beckman Coulter), and they were considered adequate for further analysis if the optical density_{260/280} ratio was ≥ 1.8 and the total RNA yield was ≥ 1 μ g. Seventy-five percent of all aspirations yielded at least 1 μ g total RNA required for the gene expression profiling. Previously, 31 total RNA specimens were split, labeled, and hybridized in duplicates several months apart in the same and in a different laboratory to assess technical reproducibility of gene expression-based predictions, and showed 97% concordance in these replicate experiments (13). Gene expression profiling was done on Affymetrix U133A gene chips in batches over a 3-year period at MDACC. Genomic prediction of response was done using a standardized computer code (13, 17). Expression results from the 30 selected predictor genes were entered into the class prediction algorithm, and each case was assigned a response status of "pCR" or "RD" prospectively before actual pathologic outcome data became available. The complete microarray data of this trial are available at the Gene Expression Omnibus database (accession number GSE20271).

We also evaluated the predictive performance of a previously established multivariate clinical nomogram that is freely available online (www.mdanderson.org/pcr; ref. 14). This nomogram combines information from patient age, tumor size, histologic grade, and ER status to predict probability of pCR after preoperative T/FAC or FAC chemotherapies. None of the current cases were included in the development of the genomic or clinical response predictors.

Predictive performances are presented as sensitivity, specificity, positive (PPV) and negative predictive values (NPV), and area under the receiver operating characteristic (ROC) curve (AUC). For the clinical nomogram, we also report calibration (i.e., agreement between observed outcome frequencies and predicted probabilities) and discrimination (i.e., whether the relative ranking of individual predictions is in the correct order).

Statistical design

The primary objective of this study was to establish that patients with DLDA-30-positive tumors are significantly

more likely to experience pCR to T/FAC chemotherapy than patients who are predicted to have residual cancer by the genomic predictor. The secondary objectives of the study were to compare the pCR rates between the sequential T/FAC and FACx6 treatment arms, and to assess interaction between prediction status assigned by the DLDA-30 predictor and pCR to two different neoadjuvant chemotherapies. The primary end point of this study was to assess the rate of pCR after completion of preoperative chemotherapy. Sample size was calculated based on the primary objective using computer simulations. Computer simulations were carried out to estimate the power to detect gene expression profile effects and profile-by-treatment interaction effects at different sample sizes. The following assumptions were used based on the original discovery and validation results (13): (a) the prevalence of DLDA-30 marker-positive patients is 30% (because the pCR rate after T/FAC is expected to be approximately 25-30% in unselected patients), (b) pCR rate to T/FAC treatment in the marker-positive group is $\geq 60\%$ (based on the PPV observed in the previous single arm study; ref. 13), and (c) pCR rate to FAC treatment in the marker-positive group is between 20% and 40% (higher than 10-15% seen historically in unselected patients but lower than observed for T/FAC). Repeated fitting of a logistic regression model with 10,000 iterations for each case was done with pCR as the dependent variable and included terms for treatment, microarray profile group, and treatment-profile interaction. With these assumptions, a study with 210 patient (105 in each arm) would have a $\geq 95\%$ power to detect a significant marker effect for T/FAC therapy (i.e., significantly higher pCR rates in marker-positive compared with marker-negative cases). However, with this sample size, the power varies substantially to detect significant marker-treatment arm interaction effect depending on differences in pCR rates between the treatment arms (Supplementary Table S1). We assumed a 25% to 30% loss of samples, and therefore, the maximum sample size was set to 273. In the final analysis, we used multivariate logistic regression models with terms for age, treatment, tumor size, grade, ER, and HER2 status and genomic prediction to calculate odds ratios for pCR.

In a hypothesis-generating exploratory analysis, we searched for novel differentially predictive genes by fitting logistic regression models with response (pCR versus RD) as the outcome and with treatment type (FAC versus T/FAC) and gene expression values of gene *i* (for each probe set on the U133A chip) as covariates. We included a test for interaction between treatment and gene expression, and calculated the *P* value for the gene-treatment interaction term. To adjust for multiple testing, we used a β -uniform mixture model to estimate the false discovery rate (FDR; ref. 18). To explore the power of the interaction test in the completed study, we generated four random normal distributions representing particular gene expression values—one for each treatment-response group. The values for the means and SDs were taken from the observed normalized and \log_2 -transformed microarray data. Each

distribution had SD = 0.3 and sample size equal to that from the data (pCR/FAC: $n = 7$, RD/FAC: $n = 80$, pCR/T/FAC: $n = 19$, RD/T/FAC: $n = 72$), we set the means for pCR/FAC and RD/FAC to 2.5 (i.e., no effect on FAC response by gene expression), and we also set the mean for RD/T/FAC to 2.5 and varied the mean for pCR/T/FAC from 2.5 to 3.2. We fit the logistic regression model described above. Over 50 iterations, we tracked how often the interaction test P value was <0.05 for a given gene and took this value as the measure of the power of the test. When the pCR/T/FAC mean was 2.6, the power was 14%. For 2.7, the power was 30%; for 2.8, the power was 51%; for 2.9, the power was 72%; for 3.0, the power was 86%; for 3.1, the power was 92%; and for 3.2, the power of the interaction test was 98%.

Results

Patient characteristics and response to chemotherapy

Two hundred and seventy-three patients were enrolled: 138 were randomized to T/FAC and 135 to FAC chemotherapy (intent-to-treat population). Twenty (7%) and 16 (6%) patients were excluded from genomic response analysis in each treatment arm, respectively, due to eligibility violations including nonstudy treatment regimen, patient withdrawal, or lack of pathologic assessment of response. Of the 118 patients who received T/FAC, 9 patients progressed clinically, and these were considered as RD for response prediction analysis. Of the 119 patients who were assigned to receive FAC chemotherapy, 11 received T/FAC treatment (to maximize response or due to progression on FAC), and these cases were assigned to the T/FAC treatment group for genomic response prediction analysis. The remaining 108 cases, including 5 cases that progressed, comprised the FAC treatment cohort for the final response prediction analysis (Supplementary Table S2). Figure 1 illustrates the flow and assignment of specimens.

Based on treatment received, the pCR rates were significantly higher (19%, $n = 24$ of 129) in the T/FAC arm compared with 9% ($n = 10$ of 113) in the FAC arm ($P < 0.05$). For the calculation of FAC efficacy, the five cases that were resistant to FAC and were crossed over to the T/FAC arm were counted as FAC failures; thus, $n = 108 + 5 = 113$. In the intent-to-treat population, the pCR rates were the same as above, 19.6% ($n = 27$ of 138) and 9.6% ($n = 13$ of 135) in the T/FAC and FAC arms ($P < 0.05$), respectively.

Two hundred and four FNA samples (75%) yielded sufficient quality and quantity of RNA to do gene expression analysis. The main reasons for failure were acellular aspirates and low RNA yield; five profiles (2.5%) failed array QC after hybridization. After excluding the patients who had no response information available (Fig. 1), 178 cases remained with complete pathologic response and genomic prediction results for final analysis. Of these, 91 received T/FAC and 87 received FAC chemotherapy. Clinical characteristics of these patients are presented in Table 1. Both treatment groups were well balanced for age, race, histo-

logic type and grade, tumor size, clinical nodal status, and hormone receptor and HER2 status. The pCR rates were 21% [95% confidence interval (95% CI), 0.13-0.29] in the T/FAC group ($n = 19$) and 8% (95% CI, 0.02-0.14) in the FAC group ($n = 7$; $P = 0.019$).

Performance of the DLDA-30 genomic predictor and the clinical nomogram to predict pCR

In the T/FAC arm, the PPV of the genomic predictor was 38% (95% CI, 21-56%), the NPV was 88% (95% CI, 77-95%), and sensitivity and specificity were 63% (95% CI, 38-84%) and 72% (95% CI, 60-82%), respectively. The observed pCR rate was 38% in the cohort that was predicted to achieve pCR (marker-positive patients), compared with 19% in unselected patients ($P = 0.032$), and 12% in the marker-negative patients (patients predicted to have RD; $P = 0.006$). The AUC was 0.711 (95% CI, 0.570-0.852). In the FAC-only arm, the PPV and NPV were 9% (95% CI, 1-29%), and 92% (95% CI, 83-97%), respectively. The sensitivity and specificity were 29% (95% CI, 4-71%) and 75% (95% CI, 64-84%), and the AUC was 0.584 (95% CI, 0.353-0.815; Fig. 2A; Table 2). The observed pCR rates were identical (9%) in the overall population and in the marker-positive and marker-negative patient subsets.

We applied the clinical nomogram to the same 178 patients for which microarray-based prediction was available (Fig. 2B; Table 2). In the T/FAC arm, the discrimination was high with an AUC of 0.89 (95% CI, 0.85-0.93), which was not statistically significantly different from the AUC of the DLDA-30 predictor. The model was also well calibrated with no significant difference between the predicted and the observed probability ($P = 0.21$). When the nomogram was applied to patients who received FAC, the discrimination remained high with an AUC of 0.82 (95% CI, 0.75-0.89), but the calibration was less good [the nomogram significantly ($P = 0.03$) overpredicted pCR]. Thus, the clinical predictor was accurate in predicting pCR after T/FAC, and less accurate but still effective in predicting pCR after FACx6.

Multivariate logistic regression model

In a multivariate analysis using all samples ($n = 178$) and including treatment arm, age, clinical tumor size, clinical nodal status, grade, ER and HER2 status, and DLDA-30 score as variables, only ER status ($P = 0.008$), tumor size ($P = 0.018$), and treatment arm ($P = 0.015$) were significant independent predictors of pCR. In the T/FAC treatment cohort, only ER status ($P = 0.022$) and tumor size ($P = 0.046$) were independent predictors of pCR. In the FAC treatment cohort, none of the variables was a significant predictor for pCR. The small number of events (seven pCR) and lack of power may have prevented the identification of any significant predictors with confidence within this subset.

New candidate biomarker identification using marker-treatment interaction test

In an exploratory analysis, we examined what genes were differentially predictive of response (pCR versus RD) by treatment arm by testing for gene-treatment interaction.

When all probe sets on the U133A microarray were considered individually, logistic regression models identified 206 probe sets with interaction P values of ≤ 0.05 (Supplementary Table S3). The gene with the most significant gene expression–treatment interaction P value was the *inositol polyphosphate-5-phosphatase (INPP5A)*; Fig. 3). After adjustment for multiple comparisons using the β -uniform mixture model of P values, no probe sets remained significant at a reasonable FDR. The P value distribution showed a paucity of low P values, indicating a lack of power for the individual comparisons.

Supplementary Table S4 contains the gene-treatment interaction results for the individual 30 probe sets included in the DLDA-30 predictor. The overall DLDA-30 score showed no significant interaction with treatment ($P = 0.443$). Three of the individual probe sets including Na^+/K^+ transporting ATPase-interacting protein 1 (*NKAIN1*), meteorin (*METRN*), and delta 2 catenin (*CTNND2*) showed significant interaction with treatment ($P \leq 0.05$), and one probe set [G protein–coupled receptor activity modifying protein 1 (*RAMP1*)] had borderline significance ($P = 0.051$). Figure 4 shows the plots of the fitted logistic regression models for the four genes and for the overall score. The

plots show how the probability of pCR varies by treatment arm as a function of gene expression level or DLDA-30 score values.

Retrospective power calculations for interaction test based on mean gene expression values for the DLDA-30 probe sets and observed response rates using logistic regression model indicated that the current study had a power between 14% and 50% to detect significant interaction effects.

Discussion

We tested a multigene predictor of pCR to preoperative sequential weekly paclitaxel and FAC chemotherapy from fine-needle biopsies of breast cancers in a prospective randomized international trial. pCR is important as a clinical end point because these patients experience excellent cancer-free long-term survival.

The 30-gene predictor was predictive of response to T/FAC chemotherapy. Patients who were predicted to achieve pCR to T/FAC (marker-positive patients) had significantly higher pCR rates (38%; 95% CI, 21-56%) than unselected patients (19%; $P = 0.032$) or patients predicted

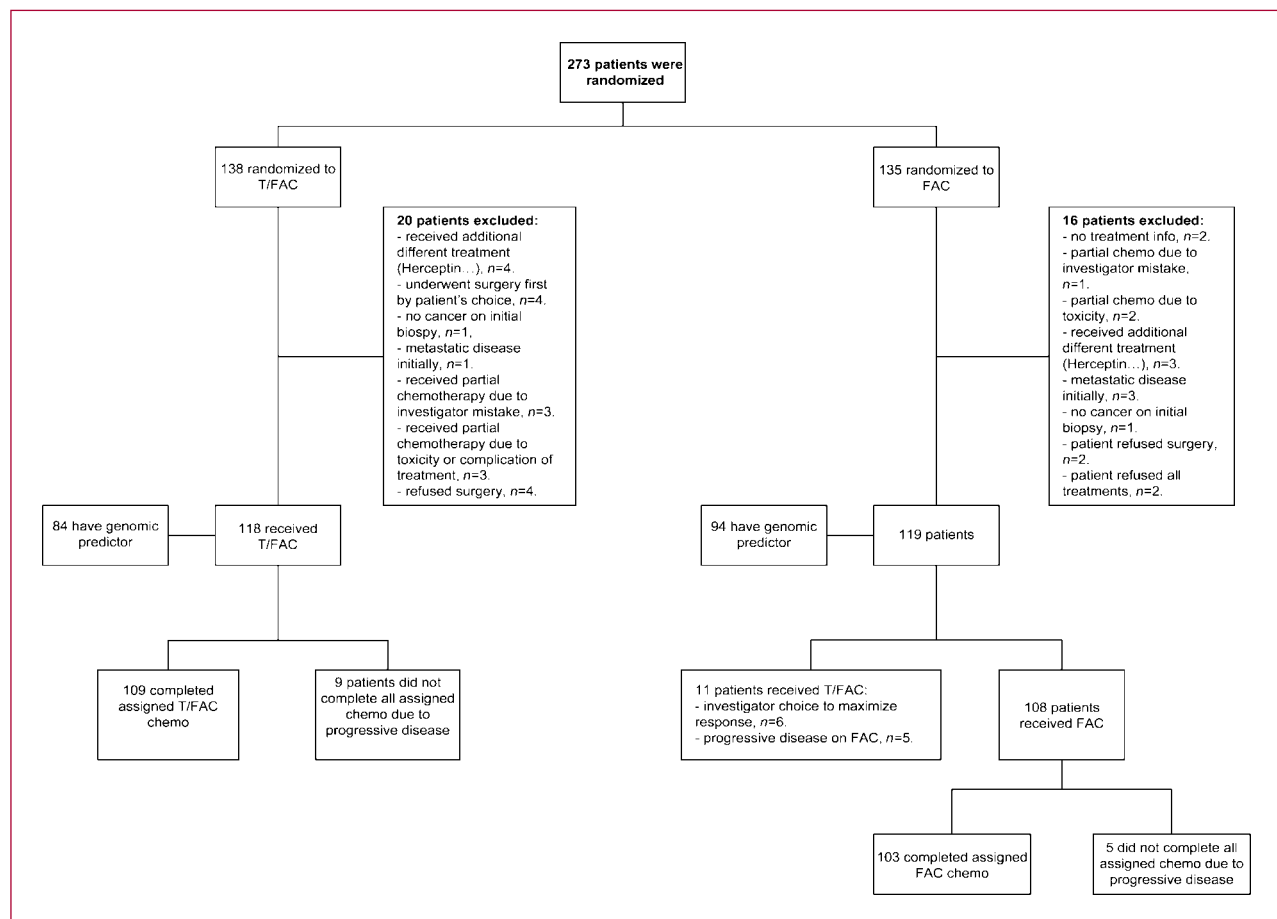


Fig. 1. Flow chart of patients and samples used in the analysis.

Table 1. Clinical characteristics of the 178 patients by treatment received for whom an observed pathologic tumor response and molecular prediction of tumor response are available

Clinical and pathologic characteristics	Patients received T/FAC		Patients received FAC		P
	No. patients	(%)	No. patients	(%)	
No. patients	91	51.1	87	48.9	
pCR	19	20.9	7	8.0	0.02
RD	72	79.1	80	92.0	0.02
Race					0.42
White	40	44.0	41	47.1	
Black	9	9.9	4	4.6	
Hispanic	41	45.1	42	48.3	
Asian	1	1.1	0	0.0	
Mean age, y (range)	51.5 (26-73)		50.3 (31-74)		0.41
Menopausal status					0.63
Premenopausal	46	50.5	48	55.2	
Postmenopausal	44	48.4	38	43.7	
Unknown	1	1.1	1	1.1	
Histology					0.11
Ductal	81	89	83	96.7	
Lobular	6	6.6	1	1.1	
Mixed	4	4.4	2	2.2	
Clinical T size					0.89
T0-T1	8	8.8	5	5.7	
T2	39	42.8	37	42.6	
T3	19	20.9	18	20.7	
T4	25	27.5	26	29.9	
Unknown	0	0.0	1	1.1	
Clinical N stage					0.76
N0	31	34.1	28	32.3	
N1	38	41.7	33	37.9	
N2-3	22	24.2	25	28.7	
Unknown	0	0.0	1	1.1	
Grade*					0.45
1	10	11	5	5.7	
2	30	33	31	35.6	
3	36	39.5	36	41.5	
Unknown	15	16.5	15	17.2	
ER status†					0.85
Negative	42	46.2	38	43.7	
Positive	49	53.8	49	55.7	
PR status†					0.98
Negative	49	53.8	46	52.9	
Positive	42	46.2	41	47.1	
HER2 status‡					0.35
Not overexpressed	75	82.4	77	88.5	
Overexpressed	16	17.6	10	11.5	
Patients underwent ALND or SLNB	84	92.3	81	93	0.93
Mean no. LN removed (range)	15.8 (1-38)		15.8 (1-43)		0.97
Patients with ALN involvement	46/84		46/81		0.9

(Continued on the following page)

Table 1. Clinical characteristics of the 178 patients by treatment received for whom an observed pathologic tumor response and molecular prediction of tumor response are available (Cont'd)

Clinical and pathologic characteristics	Patients received T/FAC		Patients received FAC		P
	No. patients	(%)	No. patients	(%)	
Type of surgery					0.99
Breast conservation	11	12.1	11	12.6	
Mastectomy	66	72.5	65	74.8	
Surgery not done due to PD	6	6.6	6	6.9	
Unknown	8	8.8	5	5.7	

Abbreviations: LN, lymph node; ALN, axillary lymph node; ALND, axillary lymph node dissection; SLNB, sentinel lymph node biopsy; PD, progressive disease.

*Histologic grade according to the modified Black's nuclear grade.

†Status for ER and PR was determined by immunohistochemistry.

*Status for HER2 was determined by immunohistochemistry or fluorescence *in situ* hybridization.

to have RD (12%; $P = 0.006$; marker-negative patients) when treated with this regimen. These results confirm that the multigene predictor can identify patients with greater than average sensitivity to T/FAC chemotherapy; they are consistent with two previous small studies (13, 17).

A test that could be used to select one therapy over another needs to have a high PPV and high sensitivity. Our test achieved a PPV of 38%, a sensitivity of 63%, and the NPV was 88%. These performance measures were less than what we have observed in the earlier validation studies but were still within the 95% CIs of the earlier performance estimates. Therefore, these results are consistent with the previous reports (13, 17).

It is increasingly clear that general chemotherapy sensitivity can also be gauged by considering routine clinical

variables such as proliferative activity, ER and HER2 status, molecular subtype, and histologic grade. Basal-like (or triple-negative) breast cancers have higher likelihood of pCR after preoperative chemotherapy than other molecular subtypes, and among the ER-positive cancers, high Oncotype DX recurrence score, luminal B molecular class, HER2 overexpression, and high genomic grade are each associated with greater chemotherapy sensitivity (3–7). All these assays tend to capture molecular characteristics of similar patients with clinically ER-negative and/or high-grade and high-proliferation tumors. When we compared the predictive performance of the DLDA-30 genomic test in T/FAC-treated patients with a multivariate clinical prediction model including grade and ER status, the overall predictive performance of the genomic test was similar to the performance of

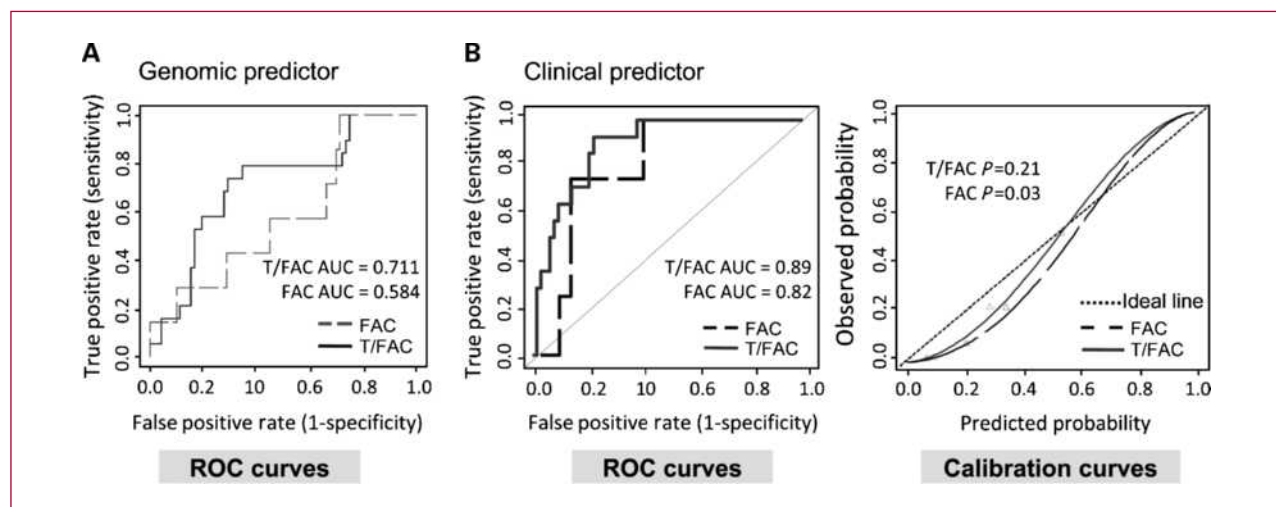


Fig. 2. ROC and calibration curves measure the performance of the genomic and clinical predictors. A, ROC curve for the predictive performance of the DLDA-30 genomic pCR predictor on the 178 cases for whom an observed pathologic tumor response and molecular prediction of tumor response were available. B, ROC and calibration curves for the predictive performance of the clinical nomogram to predict pCR on the same cases used in the genomic analysis. The corresponding performance metrics of the genomic and clinical predictors are found in Table 2.

Table 2. Performance metrics of the genomic and clinical predictors

	T/FAC (n = 91)	FACx6 (n = 87)
Genomic predictor		
ROC, AUC	0.711 (95% CI, 0.570-0.852)	0.584 (95% CI, 0.353-0.815)
PPV	38% (95% CI, 21-56)	9% (95% CI, 1-29)
NPV	88% (95% CI, 77-95)	92% (95% CI, 83-97)
Sensitivity	63% (95% CI, 38-84)	29% (95% CI, 4-71)
Specificity	72% (95% CI, 60-82)	75% (95% CI, 64-84)
Clinical predictor		
Discrimination (ROC), AUC	0.89 (95% CI, 0.85-0.93)	0.82 (95% CI, 0.75-0.89)
Calibration		
<i>P</i>	0.21	0.03
<i>E</i> _{max}	10.5%	15.5%
<i>E</i> _{average}	4.8%	9.2%

NOTE: The corresponding ROC and calibration curves are represented graphically in Fig. 2.

Abbreviations: *E*, difference in predicted probabilities and observed frequencies; *E*_{max}, maximal error; *E*_{average}, average error.

the clinical nomogram. In a multivariate logistic regression analysis that included all patients, only ER status, tumor size, and treatment arm but not the genomic test results were independent predictors of pCR. This indicates that this first-generation genomic chemotherapy response predictor mostly captures gene expression information associated with clinical phenotype, particularly ER status and proliferative activity that is reflected in histologic grade. This study illustrates an inherent pitfall in developing predictive markers from patient cohorts that include different breast cancer subtypes. ER-positive and ER-negative cancers have large-scale gene expression differences, and they also have substantially different sensitivities to preoperative chemotherapy, and this can lead to confounding of ER-status related genes with treatment response genes in studies that consider all breast cancers together. During the development of the DLDA-30, predictor cases with pCR were compared with cases with residual cancer and both ER-positive and ER-negative

cancers were included in the analysis. Inevitably, the predictor included many genes that reflected the phenotypic differences between responders (pCR) and nonresponders (RD), responders being predominantly triple-negative cancers and highly proliferative ER-positive cancers. To illustrate this point, we tested the DLDA-30 pCR predictor as a predictor of ER status: in the current study population, it had a PPV of 0.7 and AUC of 0.766 as predictor of ER-negative versus ER-positive phenotype. Similar phenotype-associated predictive value limits the use of the 70-gene prognostic signature (MammaPrint) or Oncotype DX in ER-negative breast cancers (19). The next-generation predictive (and prognostic) markers will need to be developed separately for different molecular subsets of breast cancers to increase their clinical utility (20).

A secondary objective of this randomized study was to evaluate the regimen specificity of the DLDA-30 predictor by assessing its performance on the FAC-treated patients. Excellent survival in patients who achieve pCR most likely

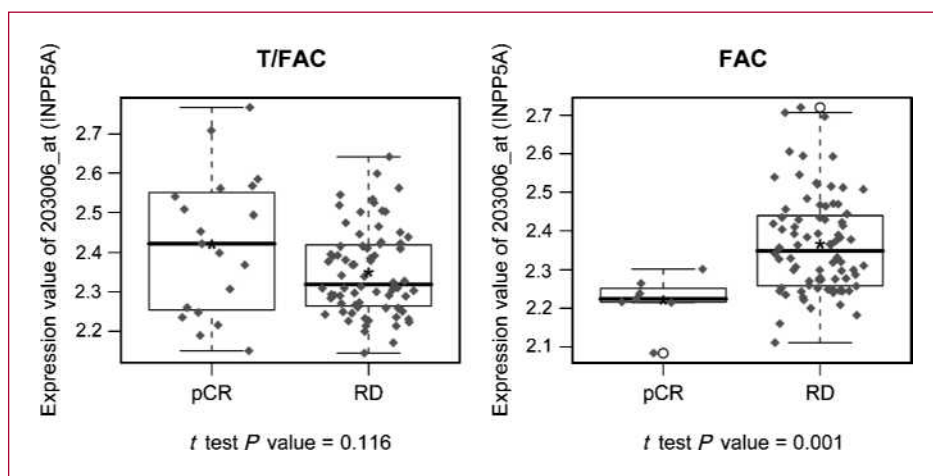


Fig. 3. Expression of the *INPP5A* gene in the two response cohorts within the two treatment arms. *INPP5A* had the most significant gene expression–treatment arm interaction *P* value (*P* = 0.00193, unadjusted for multiple comparisons). Asterisk indicates the mean value.

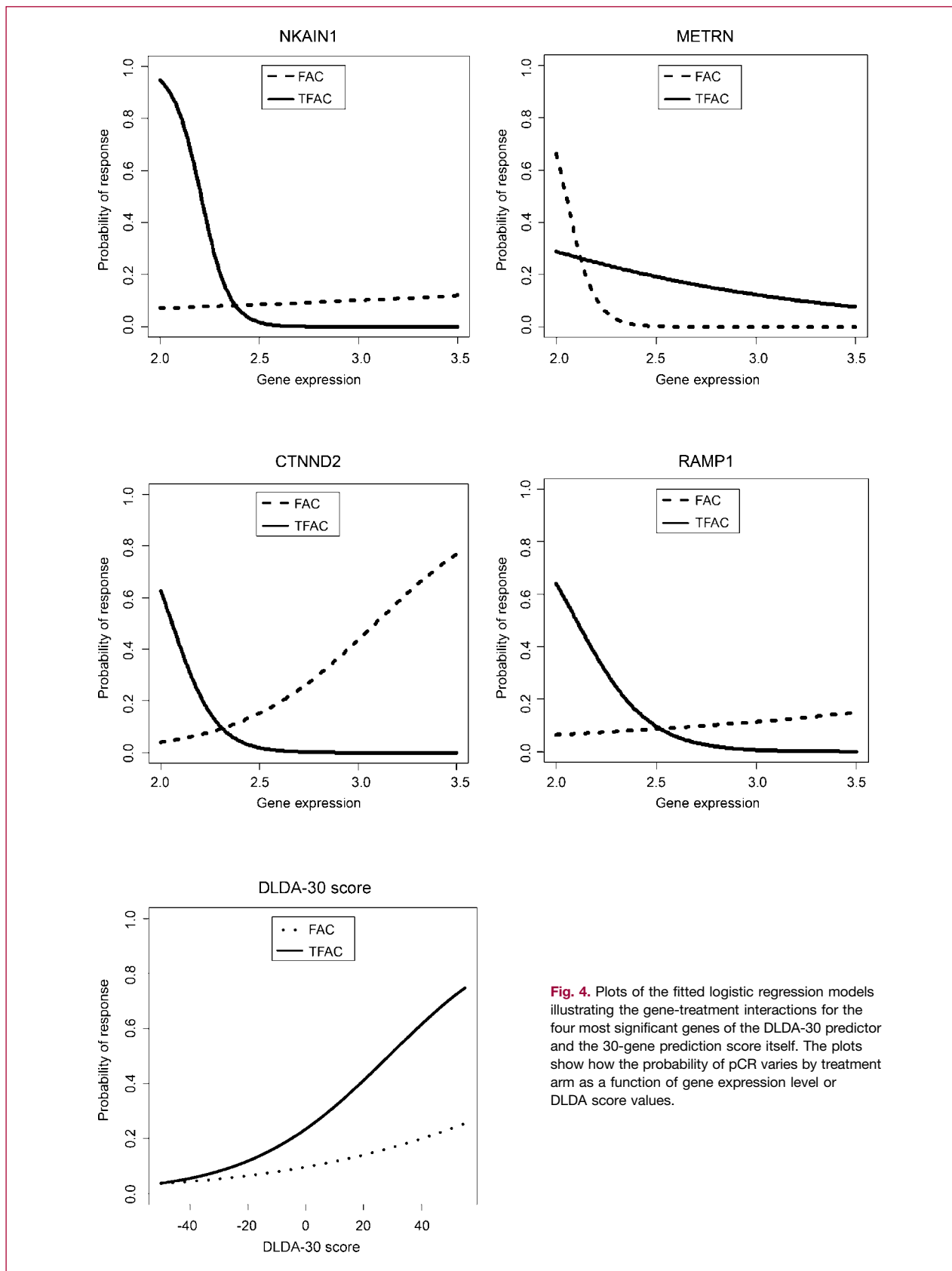


Fig. 4. Plots of the fitted logistic regression models illustrating the gene-treatment interactions for the four most significant genes of the DLDA-30 predictor and the 30-gene prediction score itself. The plots show how the probability of pCR varies by treatment arm as a function of gene expression level or DLDA score values.

reflects benefit from chemotherapy because most clinical and molecular characteristics associated with pCR (i.e., ER-negative status, high histologic or genomic grade, high Oncotype DX recurrence score, and basal-like or luminal B molecular class “intrinsic subtype”) predict worse natural history in the absence of chemotherapy (5–7). But the above variables predict general chemosensitivity of a tumor, which can be useful clinically. However, more useful will be the predictive test that can discriminate between the probability of response to different chemotherapies and thus help guide the rational selection of specific treatment in individual patients. We tested the performance of the genomic predictor in the two treatment arms. The DLDA-30 identified a patient population that is more likely to achieve pCR after T/FAC chemotherapy than the general patient population, but it did not predict pCR in the FAC-only treatment arm. In that arm, the PPV and NPV were 9% and 92%, respectively, and the sensitivity and specificity were 29% and 75% with an AUC of 0.584. This indicates that our genomic predictor may have some regimen-specificity. Alternatively, this may also represent a false-negative finding due to lack of power because of the small number of events in the FAC group (pCR, $n = 7$) and small overall sample size. The marker-treatment interaction test was also not significant, but post hoc power calculations showed 14% to 50% power to detect significant interaction effects for individual probe sets or for the combined DLDA-30 score. Interestingly, although the genomic predictor performed similarly to the clinical pCR predictor nomogram at predicting response to T/FAC, the clinical nomogram retained a comparable predictive accuracy in FAC-treated patients, whereas the genomic test did not. This further supports some degree of regimen-specific predictive value for the genomic test that the clinical predictor lacks.

The ideal data to develop treatment-specific predictors is a randomized trial that is adequately powered to identify individual genes with significant treatment-marker interaction effect. However, such data sets are rare and prospective power calculations for gene-treatment interaction tests are difficult because the magnitude of effect is unknown and it is likely to be variable for different genes. The current data set provided an opportunity to do an exploratory analysis. We examined each probe set on the array for potential marker-treatment interaction effect. This analysis was undertaken to assess if we could identify genes that have larger gene-treatment-

response interaction effect than the probes sets that were included in the DLDA-30 predictor. We identified numerous, potential treatment arm-specific predictor candidates (Supplementary Table S3). These genes will need to be tested in other independent data sets.

We also show that weekly paclitaxel \times 12 followed by FAC \times 4 regimen results in significantly higher pCR rate (19% versus 9%, $P < 0.05$) than 6 courses of FAC (or FEC). The pCR rates for T/FAC are consistent with findings from a larger study using the same preoperative therapy (21). These results support the increasing consensus that the addition of a taxane to anthracycline-based therapy improves pCR rates and long-term outcomes in breast cancer.

In summary, prospective gene expression analysis for response prediction was feasible in this randomized, international trial. Seventy-five percent of the FNA specimens mailed to a central laboratory yielded adequate RNA for genomic analysis. A 30-gene molecular test was predictive of pCR to T/FAC and not to FAC chemotherapy, but it did not do better than a freely available web-based clinical nomogram. The clinical nomogram, however, lacked regimen specificity and predicted response equally well to both T/FAC and FAC. Like most other currently in use molecular response predictors, which rely on measuring molecular equivalents of clinical phenotype, this first-generation genomic predictor derives its predictive value from detecting the large-scale gene expression differences that distinguish ER-negative from ER-positive tumors and high-grade from low-grade cancers. To improve their clinical utility, second-generation genomic predictors will need to be developed separately for the different molecular and phenotypic subsets of breast cancers.

Disclosure of Potential Conflicts of Interest

W.F. Symmans: stock owner, Nuvera Biosciences; L. Pusztai and W.F. Symmans: uncompensated consultant/advisory board, Nuvera Biosciences.

Grant Support

Grants from the Breast Cancer Research Foundation, the Commonwealth Foundation, and NCI RO1-CA106290 (L. Pusztai). There was no direct support from the pharmaceutical industry to conduct this trial.

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 05/12/2010; revised 08/10/2010; accepted 08/31/2010; published OnlineFirst 09/09/2010.

References

1. Fisher B, Bryant J, Wolmark N, et al. Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. *J Clin Oncol* 1998;16:2672–85.
2. Liedtke C, Mazouni C, Hess KR, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J Clin Oncol* 2008;26:1275–81.
3. Rouzier R, Perou CM, Symmans WF, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res* 2005;11:5678–85.
4. Carey LA, Dees EC, Sawyer L, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin Cancer Res* 2007;13:2329–34.
5. Gianni L, Zambetti M, Clark K, et al. Gene expression profiles in paraffin-embedded core biopsy tissue predict response to

- chemotherapy in women with locally advanced breast cancer. *J Clin Oncol* 2005;23:7265–77.
6. Andre F, Mazouni C, Liedtke C, et al. HER2 expression and efficacy of preoperative paclitaxel/FAC chemotherapy in breast cancer. *Breast Cancer Res Treat* 2008;108:183–90.
 7. Liedtke C, Hatzis C, Symmans WF, et al. Genomic grade index is associated with response to chemotherapy in patients with breast cancer. *J Clin Oncol* 2009;27:3185–91.
 8. Knoop AS, Knudsen H, Balslev E, et al. Retrospective analysis of topoisomerase IIa amplification and deletions as predictive markers in primary breast cancer patients randomly assigned to cyclophosphamide, methotrexate, and fluorouracil or cyclophosphamide, epirubicin or fluorouracil: Danish Breast Cancer cooperative Group. *J Clin Oncol* 2005;23:7483–90.
 9. Rouzier R, Rajan R, Hess KR, et al. Microtubule associated protein τ is a predictive marker and modulator of response to paclitaxel-containing preoperative chemotherapy in breast cancer. *Proc Natl Acad Sci U S A* 2005;102:8315–20.
 10. Esteva FJ, Hortobagyi GN. Topoisomerase II α amplification and anthracycline-based chemotherapy: the jury is still out. *J Clin Oncol* 2009;27:3416–7.
 11. Pusztai L, Jeong JH, Gong Y, et al. Evaluation of microtubule-associated protein- τ expression as a prognostic and predictive marker in NSABP-B 28 randomized clinical trial. *J Clin Oncol* 2009;27:4287–92.
 12. Pusztai L. Markers predicting clinical benefit in breast cancer from microtubule-targeting agents. *Ann Oncol* 2007;18:xii15–20.
 13. Hess KR, Anderson K, Symmans WF, 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.
 14. Rouzier R, Pusztai L, Delaloge S, et al. Nomograms to predict pathologic complete response and metastasis-free survival after preoperative chemotherapy for breast cancer. *J Clin Oncol* 2005;23:8331–9.
 15. Mazouni C, Peintinger F, Kau SW, et al. Residual ductal carcinoma *in situ* in patients with complete eradication of invasive breast cancer after neoadjuvant chemotherapy does not adversely affect patient outcome. *J Clin Oncol* 2007;25:2650–5.
 16. Symmans WF, Ayers M, Clark EA, et al. Total RNA yield and microarray gene expression profiles from fine needle aspiration and core needle biopsy samples of breast cancer. *Cancer* 2003;97:2960–71.
 17. Peintinger F, Anderson K, Mazouni C, et al. Thirty-gene pharmacogenomic test correlates with residual cancer burden after preoperative chemotherapy breast cancer. *Clin Cancer Res* 2007;13:4078–82.
 18. Pounds S, Morris SW. Estimating the occurrence of false positive and false negatives in microarray studies by approximating and partitioning the empirical distribution of *P*-values. *Bioinformatics* 2003;19:1236–42.
 19. Sotiriou C, Pusztai L. Gene expression signatures in breast cancer. *N Engl J Med* 2009;360:790–800.
 20. Hatzis C, Symmans WF, Lin F, et al. Genomic predictors of pathologic response to preoperative chemotherapy for triple-negative and ER-positive/HER2-negative breast cancers [abstract 571]. *Proc Am Soc Clin Oncol* 2008;26:23s.
 21. Green MC, Buzdar AU, Smith T, et al. Weekly paclitaxel improves pathologic complete remission in operable breast cancer when compared with paclitaxel once every 3 weeks. *J Clin Oncol* 2005;23:5983–92.

Clinical Cancer Research

Evaluation of a 30-Gene Paclitaxel, Fluorouracil, Doxorubicin, and Cyclophosphamide Chemotherapy Response Predictor in a Multicenter Randomized Trial in Breast Cancer

Adel Tabchy, Vicente Valero, Tatiana Vidaurre, et al.

Clin Cancer Res 2010;16:5351-5361. Published OnlineFirst September 9, 2010.

Updated version	Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-10-1265
Supplementary Material	Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2011/03/02/1078-0432.CCR-10-1265.DC1

Cited articles	This article cites 21 articles, 15 of which you can access for free at: http://clincancerres.aacrjournals.org/content/16/21/5351.full#ref-list-1
Citing articles	This article has been cited by 22 HighWire-hosted articles. Access the articles at: http://clincancerres.aacrjournals.org/content/16/21/5351.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, use this link http://clincancerres.aacrjournals.org/content/16/21/5351 . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.