A Genomic Predictor of Response and Survival Following Taxane-Anthracycline Chemotherapy for Invasive Breast Cancer

Christos Hatzis, PhD
Lajos Pusztai, MD, DPhil
Vicente Valero, MD
Daniel J. Booser, MD
Laura Esserman, MD, MBA
Ana Lluch, MD
Tatiana Vidaurre, MD
Frankie Holmes, MD
Eduardo Souchon, MD
Hongkun Wang, PhD
Miguel Martin, MD
José Cotrina, MD
Henry Gomez, MD
Rebekah Hubbard, BS
J. Ignacio Chacón, MD
Jaime Ferrer-Lozano, MD
Richard Dyer, MD
Meredith Buxton, MS
Yun Gong, MD
Yun Wu, MD, PhD
Nuhad Ibrahim, MD
Eleni Andreopoulou, MD
Naoto T. Ueno, MD, PhD
Kelly Hunt, MD
Wei Yang, MD
Arlene Nazario, MD
Angela DeMichele, MD
Joyce O'Shaughnessy, MD
Gabriel N. Hortobagyi, MD
W. Fraser Symmans, MD

Context Prediction of high probability of survival from standard cancer treatments is fundamental for individualized cancer treatment strategies.

Objective To develop a predictor of response and survival from chemotherapy for newly diagnosed invasive breast cancer.

Design, Setting, and Patients Prospective multicenter study conducted from June 2000 to March 2010 at the M. D. Anderson Cancer Center to develop and test genomic predictors for neoadjuvant chemotherapy. Patients were those with newly diagnosed ERBB2 (HER2 or HER2/neu)–negative breast cancer treated with chemotherapy containing sequential taxane and anthracycline–based regimens (then endocrine therapy if estrogen receptor [ER]–positive). Different predictive signatures for resistance and response to preoperative (neoadjuvant) chemotherapy (stratified according to ER status) were developed from gene expression microarrays of newly diagnosed breast cancer (310 patients). Breast cancer treatment sensitivity was then predicted using the combination of signatures for (1) sensitivity to endocrine therapy, (2) chemoresistance, and (3) chemosensitivity, with independent validation (198 patients) and comparison with other reported genomic predictors of chemotherapy response.

Main Outcome Measures Distant relapse–free survival (DRFS) if predicted treatment sensitive and absolute risk reduction ([ARR], difference in DRFS between 2 predicted groups) at median follow-up (3 years).

Results Patients in the independent validation cohort (99% clinical stage II-III) who were predicted to be treatment sensitive (28%) had 56% (95% CI, 31%-78%) probability of excellent pathologic response and DRFS of 92% (95% CI, 85%-100%), with an ARR of 18% (95% CI, 6%-28%). Survival was predicted in ER-positive (30% predicted sensitive; DRFS, 97% [95% CI, 91%-100%]; ARR, 11% [95% CI, 0.1%-21%]) and ER-negative (26% predicted sensitive; DRFS, 83% [95% CI, 68%-100%]; ARR, 26% [95% CI, 4%-48%]) subsets and was significant in multivariate analysis. Other genomic predictors showed paradoxically worse survival for patients predicted to be responsive to chemotherapy.

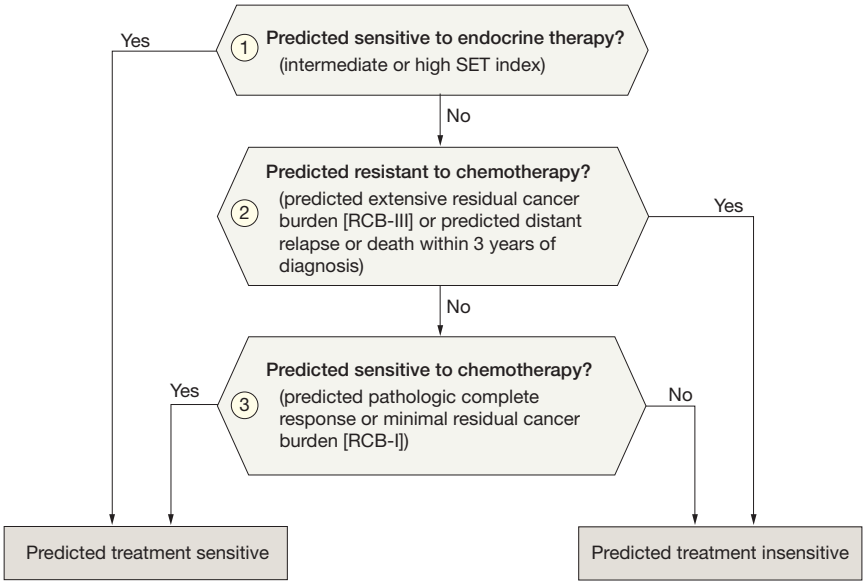
Conclusion A genomic predictor combining ER status, predicted chemoresistance, predicted chemosensitivity, and predicted endocrine sensitivity identified patients with high probability of survival following taxane and anthracycline chemotherapy.

JAMA. 2011;305(18):1873-1881 www.jama.com

THERE IS CLINICAL NEED FOR predictive tests for patients with newly diagnosed ERBB2 (HER2 or HER2/neu)–negative breast cancer whose clinical-pathologic risk at presentation favors

Author Affiliations: Nuvera Biosciences Inc, Woburn, Massachusetts (Dr Hatzis); University of Texas M. D. Anderson Cancer Center, Houston (Drs Pusztai, Valero, Booser, Gong, Wu, Ibrahim, Andreopoulou, Ueno, Hunt, Yang, Nazario, Hortobagyi, and Symmans and Ms Hubbard); I-SPY Clinical Trial Investigators for the Cancer and Leukemia Group B, University of California at San Francisco (Dr Esserman and Ms Buxton) and University of Pennsylvania (Dr DeMichelle); Grupo Español de Investigación en Cáncer de Mama (GEICAM), Madrid, Spain (Drs Martin, Lluch, Chacón, and Ferrer-Lozano); Instituto Nacional de Enfermedades Neoplásicas, Lima, Peru (Drs Vidaurre, Cotrina, Gomez, and Dyer); US Oncology The Woodlands, Texas (Dr Holmes and O'Shaughnessy); University of Texas Health Sciences Center, Lyndon B. Johnson Hospital, Houston (Dr Souchon); and Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC (Dr Wang).
Corresponding Author: W. Fraser Symmans, MD, Department of Pathology, Box 85, University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030 (fsymmans@mdanderson.org).

Figure 1. Decision Algorithm Used in the Genomic Test to Predict a Patient's Sensitivity to Adjuvant Chemotherapy or Chemoendocrine Therapy From a Biopsy of Newly Diagnosed Invasive Breast Cancer



RCB indicates residual cancer burden[12]; SET indicates sensitivity to endocrine therapy.[13]

the use of chemotherapy.[1] Identification of patients with high likelihood of survival following a current standard chemotherapy regimen (and then endocrine therapy, if estrogen receptor [ER]–positive) would reaffirm that treatment decision. Conversely, identification of those with significant risk of relapse despite standard chemotherapy could be used to advise participation in an appropriate clinical trial of potentially more effective treatment. Also, because neoadjuvant (preoperative) and adjuvant (postoperative) chemotherapy are equally effective,[2] the former provides a clinical model for development of chemopredictive tests.

Inherent chemosensitivity of breast cancers differs according to phenotype, as defined by combined ER and ERBB2 receptor status.[3] However, patients with breast cancer of any phenotype that achieves pathologic complete response (ie, no invasive or metastatic breast cancer identified) following neoadjuvant chemotherapy have excellent probability of long-term survival.[2,4,5] Unfortunately, molecular tests specifically developed to predict pathologic complete response have not demonstrated any predictive superiority over the combination of standard clinicopathologic parameters (ER status, grade, and age) and have not been compared with a survival end point.[6-8] Similarly, tests designed for molecular classification (including phenotype) or prognosis without chemotherapy have failed to predict a sufficiently high probability of survival in the patients they classify as chemosensitive.[9-11] Additionally, there currently is no clinically useful test for prognosis without chemotherapy or for prediction of response or survival following chemotherapy for patients with ER-negative/ERBB2-negative breast cancer.[7,8,11]

This study addresses our hypothesis that a predictive test for response and survival following sequential taxane and anthracycline chemotherapy for ERBB2-negative breast cancer would account for each of the following biological characteristics (FIGURE 1): tumor phenotype (ER-positive or ER-negative), sensitivity to adjuvant endocrine therapy (if ER-positive), chemoresistance (extensive residual cancer burden [RCB] or early relapse), and chemosensitivity (pathologic complete response or minimal RCB).[12,13]

METHODS

Patients and Samples

Patients prospectively provided written informed consent to participate in an institutional review board–approved research protocol (LAB99-402, USO-02-103, 2003-0321, I-SPY-1) to obtain a tumor biopsy sample by fine-needle aspiration or core biopsy prior to any systemic therapy for genomic studies to develop and test predictors of treatment outcome.[12,14] Clinical nodal status was determined before treatment from physical examination, with or without axillary ultrasound, with diagnostic fine-needle aspiration as required. Pathologic ERBB2 status was defined as negative according to American Society of Clinical Oncology/College of American Pathologists guidelines.[15] Patients with any nuclear immunostaining of ER in the tumor cells were considered eligible for adjuvant endocrine therapy.

All gene expression microarrays were profiled in the Department of Pathology at the M. D. Anderson Cancer Center (MDACC), Houston, Texas. Biopsy samples were either collected in 1.5 mL of RNA*later* (Qiagen, Valencia, California) and stored locally at −70°C and transported to the laboratory on dry ice (MDACC, Instituto Nacional de Enfermedades Neoplásicas, Lyndon B. Johnson Hospital, and Grupo Español de Investigación en Cáncer de Mama) or couriered overnight in a cooler pack from clinics to the laboratory (US Oncology), or were frozen and cryosectioned and an aliquot of RNA sent to the laboratory on dry ice (I-SPY [Investigation of Serial Studies to Predict Your Therapeutic Response With Imaging and Molecular Analysis]).

Details of our methods for RNA purification and microarray hybridization have been reported previously.[6,13,16-19] Briefly, a single-round T7 amplification was used to generate biotin-labeled complementary RNA for hybridization to oligonucleotide microarrays (U133A GeneChip; Affymetrix,

Santa Clara, California). Gene expression levels were derived from multiple oligonucleotide probes on the microarray that hybridize to different sequence sites of a gene transcript (probe sets). Details of the microarray data processing are provided in the supplementary eMethods available at http://www.jama.com.

Identification of Predictive Signature for Early Relapse Events

To identify the signature of chemoresistance, we included higher-risk patients who were clinically lymph node–positive at presentation and also predicted to have low sensitivity to adjuvant endocrine therapy[13] (Figure 1 and eFigure 1). Probe sets were evaluated in univariate Cox regression analyses under bootstrap, separately in ER-positive and ER-negative training cases to assess their association with distant relapse or death. Minimal nonredundant signatures were obtained through a univariate shrinkage approach,[20] with optimal penalization determined under cross-validation to yield the shortest probe set list that resulted in the biggest incremental improvement in the area under the receiver operating characteristic curve for predicting 3-year distant relapse–free survival (DRFS) outcome. The final predictors used 33 probe sets for the ER-positive subset and 27 probe sets for the ER-negative subset (eMethods and eFigure 2).

Identification of Predictive Signatures for Excellent Pathologic Response and for Extensive Residual Disease

Differentially expressed genes (probe sets) in 2 responder groups (pathologic complete response or minimal residual cancer burden [RCB-I] defining excellent response, vs moderate or extensive residual cancer burden [RCB-II/III] defining lesser response[12]) were identified separately in ER-positive and ER-negative training cases using a robust unequal-variance *t* statistic under a bootstrap scheme. Subsequently, a multivariate penalized optimization algorithm—gradient-directed regularization—was used with maximum penalization to select a minimal signature that maximized the area under the receiver operating characteristic curve under complete cross-validation.[21] The final response predictors used 39 probe sets for the ER-positive subset and 55 probe sets for the ER-negative subset (eMethods and eFigure 2).

A similar procedure was followed to develop the predictor for resistance by comparing patients with extensive residual disease (RCB-III) after neoadjuvant chemotherapy treatment vs remaining patients. The final predictor of extensive residual disease used 73 probe sets for the ER-positive subset and 54 probe sets for the ER-negative subset (eMethods and eFigure 2).

Development of the Predictive Testing Algorithm

We combined the individual predictions into a testing algorithm for predicted sensitivity to adjuvant treatment of ERBB2-negative breast cancer with taxane and anthracycline chemotherapy: (1) predicted moderate or high sensitivity to endocrine therapy (SET) assessed based on the published 165-gene SET index of strongly ER-correlated genes that independently predicts survival following adjuvant endocrine or chemoendocrine therapy[13]; (2) predicted resistance to chemotherapy either by early distant relapse events or by extensive residual disease after neoadjuvant chemotherapy; and (3) predicted sensitivity (pathologic response) to chemotherapy (Figure 1).

Statistical Analysis

Distant relapse–free survival was defined as the interval from initial diagnostic biopsy until diagnosis of distant metastasis or death from breast cancer, nonbreast cancer, or unknown causes.[22] The primary prediction end point was DRFS at 3 years (median follow-up for the validation cohort). Predictive performance was assessed by the positive predictive value (PPV), defined as the probability of distant relapse or death for patients predicted to be treatment insensitive; the negative predictive value (NPV), defined as the DRFS for patients predicted to be treatment sensitive; and the absolute risk reduction (ARR), defined as the absolute difference in DRFS between the 2 predicted groups. These were calculated from the Kaplan-Meier estimators of the survival function based on cumulative events following the interval notion for cases and controls.[23] Confidence intervals (CIs) for NPV and PPV were based on the Greenwood variance estimate and for ARR were estimated by bootstrap using 999 replicates.[24] The independent prognostic value of the genomic predictor compared with the full clinical model was assessed in multivariate Cox regression analyses using the likelihood ratio test. Pathologic response to neoadjuvant chemotherapy was defined as pathologic complete response or RCB-I for evaluation of response prediction.[6,11,25]

Statistical computations were performed in R version 2.10.1 (R Development Core Team, Vienna, Austria) by an independent statistician (H.W.). Significance was based on $P<.05$ and 95% CI estimates.

RESULTS

In the discovery cohort, biopsy samples were obtained from June 2000 to December 2006; 227 were obtained by fine-needle aspiration (MDACC) and 83 by core biopsy (I-SPY), and all chemotherapy was administered as neoadjuvant treatment. In the validation cohort, biopsy samples were obtained from April 2002 to January 2009; 157 were obtained by fine-needle aspiration (MDACC, Peru, US Oncology) and 41 by core biopsy (MDACC, Lyndon B. Johnson Hospital, Spain), and 165 of 198 patients received all chemotherapy as neoadjuvant treatment. Detailed characteristics of the patient and biospecimen cohorts used for test development and validation are provided in TABLE 1 and eFigure 1, respectively. Details of the type of biopsies collected for genomic analysis and the sequential taxane and anthracycline–

based chemotherapy treatments administered are provided in eTable 1.

Performance of the Predictive Test in the Independent Validation Cohort

The predictive test (algorithm) was applied to the discovery cohort of 310 samples (FIGURE 2) and then evaluated in the independent validation cohort of 198 patients (99% clinical stage II-III) who received sequential taxane and anthracycline chemotherapy (and then endocrine therapy if ER-positive). The validation cohort had pathologic response rates of 25% (pathologic complete response) and 30% (pathologic complete response or RCB-I), median follow-up of 3 years, and overall 3-year DRFS of 79% (95% CI, 74%-85%).

The chemopredictive test algorithm had a PPV of 56% (95% CI, 31%-78%) for prediction of pathologic response (pathologic complete response or RCB-I) after excluding patients with predicted endocrine sensitivity (TABLE 2).

In 28% of patients predicted to be treatment sensitive, the 3-year DRFS (NPV) was 92% (95% CI, 85%-100%), and there was a significant ARR of 18% (95% CI, 6%-28%). Conversely, the 3-year point estimate of DRFS for those predicted to be treatment insensitive was 75% (95% CI, 67%-82%), corresponding to a PPV of 25% (Table 2) and an odds ratio for relapse of 4.01 (95% CI, 1.60-20.4) relative to those predicted to be treatment sensitive (eTable 2). Overall, there was a significant association between predicted sensitivity to treatment and improved DRFS (P=.002) (Figure 2). Also, the diagnostic likelihood ratio for occurrence vs absence of 3-year distant relapse or death, if patients were predicted to be treatment sensitive, was 0.33 (95% CI, 0.07-0.72) (eTable 2).

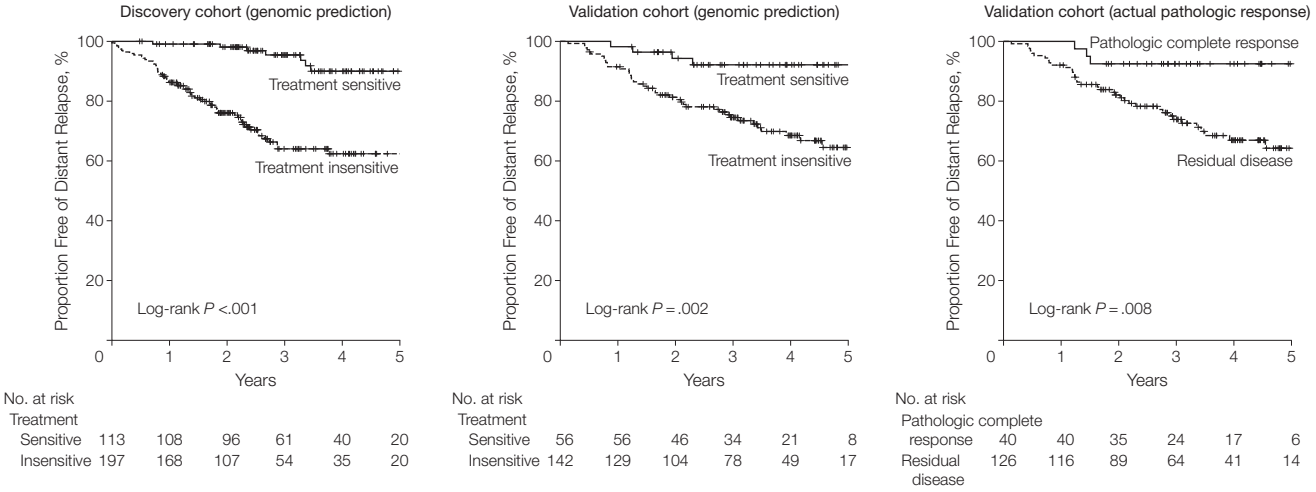
Of note, 3-year DRFS in patients predicted to be treatment sensitive at the time of diagnosis was similar to the 3-year DRFS of 93% (95% CI, 85%-100%) in the 21% of patients who

Table 1. Pretreatment Characteristics of the Discovery and Validation Cohorts

	No. (%)						
	Discovery Cohort			Validation Cohort			
Characteristic	MDACC (n = 227)	I-SPY-1 (n = 83)	Total (N = 310)	MDACC (n = 86)	LBJ/INEN/GEICAM (n = 58)	US Oncology (n = 54)	Total (N = 198)
Age, y							
≤50	112 (49)	30 (36)	142 (46)	48 (56)	30 (52)	31 (57)	109 (55)
>50	115 (51)	53 (64)	168 (54)	38 (44)	28 (48)	23 (43)	89 (45)
Mean (SD)	51 (11)	47 (8)	50 (10)	49 (11)	51 (11)	48 (9)	49 (11)
Nodal status							
Positive	165 (73)	58 (70)	223 (72)	52 (60)	42 (72)	34 (63)	128 (65)
Negative	62 (27)	25 (30)	87 (28)	34 (40)	16 (28)	20 (37)	70 (35)
T stage							
0	2 (1)	0	2 (1)	1 (1)	0	0	1 (1)
1	19 (8)	1 (1)	20 (6)	8 (9)	1 (1)	1 (2)	10 (5)
2	131 (58)	34 (41)	165 (53)	52 (61)	19 (33)	19 (35)	90 (45)
3	35 (15)	39 (47)	74 (24)	18 (21)	19 (33)	34 (63)	71 (36)
4	40 (18)	9 (11)	49 (16)	7 (8)	19 (33)	0	26 (13)
Grade							
1	13 (6)	6 (7)	19 (6)	7 (8)	5 (8)	1 (2)	13 (7)
2	92 (40)	25 (30)	117 (38)	28 (33)	19 (33)	16 (30)	63 (32)
3	122 (54)	29 (35)	151 (49)	51 (59)	23 (40)	34 (63)	108 (54)
Unknown	0	23 (28)	23 (7)	0	11 (19)	3 (5)	14 (7)
AJCC stage							
I	6 (3)	0	6 (2)	2 (2)	0	0	2 (1)
II	126 (55)	39 (47)	165 (53)	57 (66)	18 (31)	32 (59)	107 (54)
III	95 (42)	44 (53)	139 (45)	27 (32)	40 (69)	22 (41)	89 (45)
ER status							
Positive	131 (58)	43 (52)	174 (56)	60 (70)	37 (64)	26 (48)	123 (62)
Negative	96 (42)	35 (42)	131 (42)	26 (30)	21 (36)	27 (50)	74 (37)
Indeterminate	0	5 (6)	5 (2)	0	0	1 (2)	1 (1)
PR status							
Positive	102 (45)	40 (48)	142 (46)	43 (50)	31 (53)	27 (50)	101 (51)
Negative	125 (55)	37 (45)	162 (52)	43 (50)	27 (47)	26 (48)	96 (48)
Indeterminate	0	6 (7)	6 (2)	0	0	1 (2)	1 (1)

Abbreviations: AJCC, American Joint Committee on Cancer; ER, estrogen receptor; GEICAM, Grupo Español de Investigación en Cáncer de Mama; INEN, Instituto Nacional de Enfermedades Neoplásicas; I-SPY-1, Investigation of Serial Studies to Predict Your Therapeutic Response With Imaging and Molecular Analysis; MDACC, M. D. Anderson Cancer Center; LBJ, Lyndon B. Johnson Hospital; PR, progesterone receptor.

Figure 2. Kaplan-Meier Estimates of Distant Relapse–Free Survival According to Genomic Predictions (Before Treatment) as Treatment Sensitive or Treatment Insensitive and to Pathologic Response (After Treatment) as Pathologic Complete Response or Residual Disease



Vertical ticks on the curves indicate censored observations.

Table 2. Performance of Genomic Signatures for Predicting Pathologic Response and 3-Year DRFS

	% (95% CI)											
	Prediction of Pathologic Response						Prediction of Distant Relapse or Death Within 3 Years					
	Discovery Cohort (n = 310)			Validation Cohort (n = 198)			Discovery Cohort (n = 310)			Validation Cohort (n = 198)		
Predictor	No. (%)[a]	PPV[b]	NPV[b]	No. (%)[a]	PPV[b]	NPV[b]	PPV	NPV (DRFS)[c]	ARR[d]	PPV	NPV (DRFS)[c]	ARR[d]
Genomic Grade Index, high	301 (29)	36 (30 to 43)	88 (79 to 93)	101 (30)	40 (28 to 54)	84 (70 to 93)	14 (6 to 22)	72 (65 to 79)	–14 (–25 to –3)	7 (1 to 13)	72 (64 to 80)	–21 (–30 to –10)
Genomic subtype classifier, luminal B or basal-like	301 (29)	40 (32 to 48)	85 (78 to 90)	101 (30)	40 (25 to 56)	78 (65 to 87)	13 (7 to 19)	66 (58 to 76)	–20 (–31 to –10)	12 (6 to 20)	72 (62 to 81)	–16 (–27 to –5)
Genomic predictor of pathologic complete response	301 (29)	46 (37 to 55)	83 (77 to 88)	101 (30)	40 (24 to 58)	75 (63 to 85)	15 (9 to 20)	62 (52 to 73)	–24 (–36 to –12)	10 (4 to 16)	62 (50 to 73)	–28 (–41 to –16)
ER-stratified genomic predictor of pathologic complete response/ RCB-I[e]	301 (29)	69 (60 to 77)	100 (98 to 100)	101 (30)	42 (28 to 57)	81 (68 to 91)	30 (22 to 37)	85 (78 to 93)	15 (4 to 25)	24 (14 to 32)	82 (74 to 90)	5 (–7 to 16)
Predictive test, treatment sensitive[e,f,g]	256 (31)	78 (66 to 88)	84 (78 to 89)	91 (33)	56 (31 to 78)	73 (61 to 82)	36 (27 to 44)	95 (91 to 100)	31 (22 to 41)	25 (18 to 33)	92 (85 to 100)	18 (6 to 28)

Abbreviations: ARR, absolute risk reduction; DRFS, distant relapse–free survival; NPV, negative predictive value; PPV, positive predictive value; RCB-I, minimal residual cancer burden.
[a] Percentages represent pathologic response rates.
[b] Confidence intervals based on binomial approximation.
[c] Distant relapse–free survival estimate at 3 years.
[d] For event within 3 years if predicted to be treatment sensitive (negative values indicate that any negative risk reduction was in favor of predicted treatment insensitive).
[e] Performance of the pathologic complete response predictor on the discovery cohort is optimistically biased because the predictor was trained on a subset of these samples. Performance of the pathologic complete response or RCB-I predictor and of the overall genomic prediction test on the discovery cohort represents resubstitution performance, because the predictors were trained on the same cohort.
[f] Genomic prediction of pathologic response was evaluated in the subset in both cohorts with low sensitivity to endocrine therapy to evaluate the chemopredictive component only (Figure 1).
[g] Performance of the predictive test is optimistically biased in the discovery cohort because a component of the test was trained on DRFS events to define resistance.

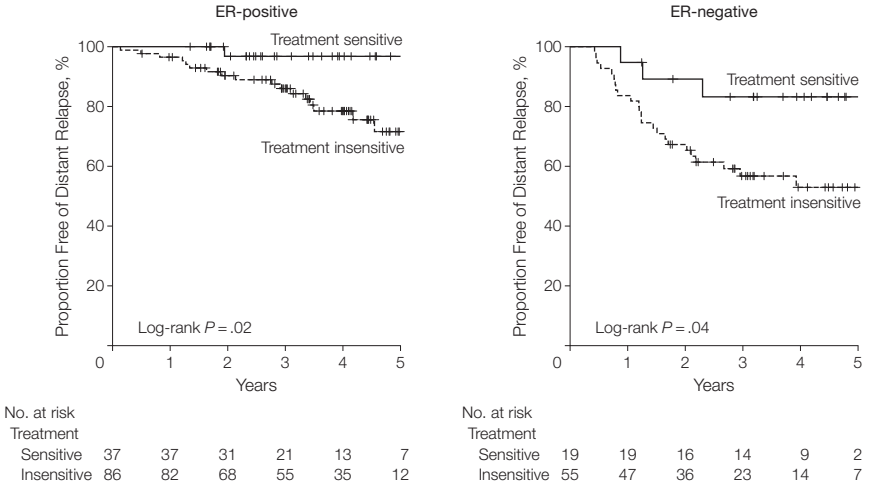
achieved pathologic complete response after completion of neoadjuvant chemotherapy (Figure 2). Similarly, 3-year DRFS for those predicted to be treatment insensitive was identical to the 3-year DRFS of 75% (95% CI, 68%-83%) in those with residual disease (Figure 2). Furthermore, DRFS estimates for the predicted treatment sensitive and the actual pathologic complete response groups were unchanged at 5 years and were identical at 65% (95% CI, 56%-75%) for the predicted treatment insensitive and the actual residual disease groups (Figure 2).

Predicted Treatment Sensitivity According to ER Status

Treatment sensitivity was predicted in 37 of 123 patients (30%) in the ER-positive phenotypic subgroup and in 19 of 74 (26%) in the ER-negative subgroup. In the ER-positive subgroup, these patients had excellent DRFS (NPV) of 97% (95% CI, 91%-100%) and a significant ARR of 11% (95% CI, 0.1%-21%) at 3 years of follow-up (FIGURE 3). In the subset of ER-positive patients predicted to be insensitive to endocrine therapy (low SET index), the PPV for pathologic response (pathologic complete response or RCB-I) was 42% (95% CI, 15%-72%) in 20% of patients predicted to be treatment sensitive. Conversely, if patients were ER-positive and predicted to be treatment insensitive, the PPV for 3-year DRFS event was 14% (95% CI, 6%-21%).

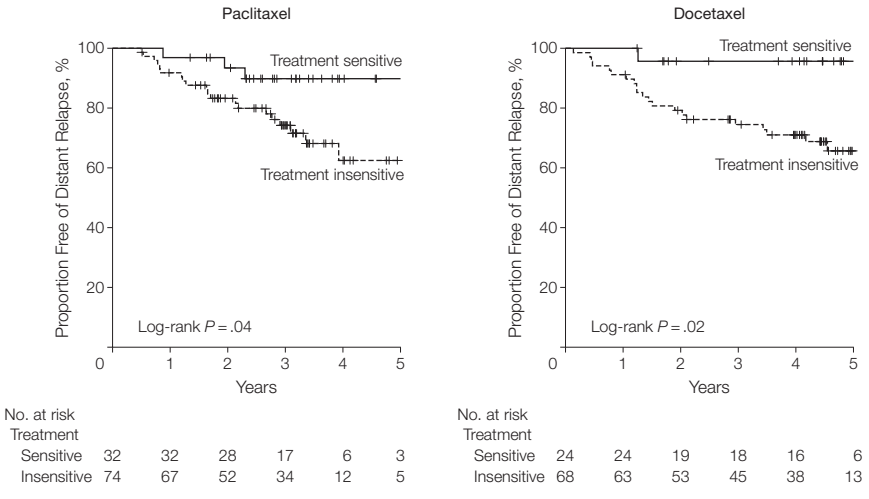
Patients with ER-negative cancer predicted to be treatment sensitive had significantly improved 3-year DRFS (NPV) of 83% (95% CI, 68%-100%) (Figure 3), with an ARR of 26% (95% CI, 4%-48%) and a PPV for pathologic response (pathologic complete response or RCB-I) of 83% (95% CI, 36%-100%). Conversely, the PPV for 3-year relapse was 43% (95% CI, 28%-55%) for patients predicted to be treatment insensitive. Lastly, this test had a significant diagnostic likelihood ratio for predicted occurrence vs absence of 3-year distant relapse or death, if patients were predicted to be treatment sensitive, of 0.27 (95% CI, 0.01-0.94) for ER-positive and 0.35 (95% CI, 0.04-0.91) for ER-negative breast cancer (eTable 2).

Figure 3. Phenotypic Subset Analysis of Genomic Predictions in the Validation Cohort



Vertical ticks on the curves indicate censored observations. ER indicates estrogen receptor.

Figure 4. Taxane Chemotherapy Regimen Subset Analysis of Genomic Predictions in the Validation Cohort



Taxane chemotherapy administered as 12 cycles of weekly paclitaxel or as 4 cycles of 3-times-weekly docetaxel with capecitabine, each sequentially administered before or after 4 cycles of anthracycline-based therapy. Vertical ticks on the curves indicate censored observations.

Performance of the Predictive Test in Other Relevant Subsets

The association between predicted treatment sensitivity and DRFS appears to be unrelated to the type of taxane therapy administered (FIGURE 4). For patients predicted to be treatment sensitive, the 3-year DRFS was 90% (95% CI, 80%-100%) in the subset who received 12 cycles of weekly paclitaxel and 96% (95% CI, 88%-100%) in the subset who received 4 cycles of 3-times-weekly docetaxel with capecitabine. Similarly, the 3-year DRFS was 93% (95% CI, 84%-100%) with significantly improved DRFS compared with prediction of treatment insensi-

tivity ($P = .003$) in 128 patients with clinically positive nodes. The 3-year DRFS was 91% (95% CI, 81%-100%) in 70 patients with clinically negative nodes but was not significantly different from predicted insensitivity.

Comparison of the Predictive Test With Clinical-Pathologic Parameters

Genomic predictions were independently and significantly associated with risk of distant relapse or death (sensitive vs insensitive; hazard ratio, 0.19 [95% CI, 0.07-0.55]; $P = .002$) after adjusting for standard clinical-pathologic parameters (eTable 3). Addition of the genomic prediction to a multivariate Cox model of the clinical-pathologic factors significantly increased the predictive utility of the model (likelihood ratio of complete model vs clinical model, 13.8; $P < .001$). In this model, higher clinical tumor stage (tumor stage T3 or T4 vs T1 or T2: hazard ratio, 2.13 [95% CI, 1.13-4.02]; $P = .02$) and ER-negative status (ER-positive vs ER-negative: hazard ratio, 0.34 [95% CI, 0.18-0.65]; $P = .001$) were associated with statistically significant greater risk of distant relapse or death.

Comparison With Other Predictive Genomic Signatures

We also evaluated other phenotypic predictors that have published association with higher probability of pathologic complete response to neoadjuvant chemotherapy, have a predefined threshold for prediction of pathologic complete response based on Affymetrix microarray data, and that we have confirmed to be correctly calculated: the 96-gene genomic grade index (GGI) to define high vs low grade (high GGI predicted pathologic complete response),[11] a 52-gene signature (PAM50) to assign intrinsic subtype (basal-like, ERBB2, and luminal B subtypes predicted pathologic complete response),[25] and a 30-gene signature (DLDA30) developed to predict pathologic complete response vs residual disease.[6] These tests were significantly predictive of pathologic response in the discovery cohort (lower 95% confidence limit of the PPV greater than the baseline response [pathologic complete response or RCB-I] rate of 29%), and the tests had NPVs of 83% or greater (Table 2). Performance in the validation cohort was similar (Table 2), but none of these tests had a PPV significantly greater than the baseline response (pathologic complete response or RCB-I) rate of 30%.

The entire prediction algorithm (Figures 1 and 2) and its component to predict pathologic complete response or RCB-I (Figure 1 and eFigure 3D and H) demonstrated significantly improved DRFS for patients predicted to be treatment sensitive (Table 2). The other tests (GGI, PAM50, DLDA30) demonstrated worse DRFS for patients predicted to have chemosensitive breast cancer (eFigure 3) and within the ER-positive (eFigure 4) and ER-negative (eFigure 5) subsets, as indicated by their negative ARRs (Table 2) and paradoxical diagnostic likelihood ratios (eTable 2).

COMMENT

Any test based on predicted sensitivity, resistance, or both to guide the selection of a standard adjuvant treatment regimen should predict a high probability of survival for patients predicted to be treatment sensitive (NPV, no relapse if predicted to be treatment sensitive) and a clinically meaningful survival difference between patients predicted to be treatment sensitive and insensitive (ARR) as well as improve on predictions using existing clinical-pathological information. The performance of our predictive test meets these criteria in an independent validation cohort. The 3-year DRFS of 92% for patients predicted to be treatment sensitive (NPV) was higher than in the unselected cohort (79%), with a statistically significant ARR of 18%. Furthermore, the predictive test added significantly to a multivariate clinical-pathologic model (age, tumor size, nodal status, grade, ER status, and type of taxane administered), wherein patients predicted to be treatment sensitive had a 5-fold reduction in the risk of distant relapse (eTable 3). It also should be noted that in the validation cohort the a priori test results (predicted treatment sensitive or insensitive) from a tumor sample obtained before treatment were as predictive of DRFS as the pathologic response assessed after the completion of chemotherapy (Figure 2).

Strictly, an unequivocal estimate of the extent of chemopredictive effect of this test on DRFS (rather than prognosis from natural history) would require comparison with a randomized control population who did not receive chemotherapy. However, this approach is not justifiable for patients with stage II-III disease without strong evidence of safety. Therefore, we separately demonstrated that this test was not prognostic using available data from 2 retrospective cohorts that did not receive chemotherapy for earlier-stage cancer (eFigure 6).

We observed similar performance of this test in patients who received equivalent chemotherapy regimens containing 12 cycles of weekly paclitaxel or 4 cycles of 3-times-weekly docetaxel with capecitabine (Figure 4), in each case sequentially administered before or after 4 cycles of anthracycline-based chemotherapy.[26] Of course, additional studies must address the generalizability of these results to other study cohorts and other chemotherapy regimens that combine taxanes and anthracyclines in sequence or concurrently or that do not include a taxane or anthracycline component.

A predictive test with this performance could potentially assist medical decision-making. It could identify patients with stage II-III, ER-positive and ERBB2-negative breast cancer with excellent 3-year and 5-year DRFS (97%) following a standard adjuvant treatment (Figure 3). This group included the subset of patients predicted to be treatment sensitive who were also predicted to have low endocrine sensitivity and so were specifically predicted to be chemosensitive (eFigure 7). The

subset of ER-negative cancers predicted to be treatment sensitive had a DRFS of 83% and a significant ARR of 26% (Figure 3).

At issue is whether an 83% 3-year DRFS for ER-negative breast cancers is sufficiently high to justify use of this test to choose a currently standard taxane and anthracycline chemotherapy regimen. One clinical strategy might be to perform the test on needle biopsy samples obtained before treatment and select patients predicted to be treatment sensitive for neoadjuvant taxane and anthracycline chemotherapy. Using this approach, additional postoperative adjuvant therapy, preferably in a clinical trial, could be considered if the patient proved to have significant residual disease (RCB-II or RCB-III) at the time of surgery (17% of evaluable cases in our study). Conversely, the 3-year probability of relapse or death if a patient were predicted to be treatment insensitive was 25% overall (14% if ER-positive; 43% if ER-negative). This probability should not be misinterpreted as futility of current standard treatment (there was no randomized control) but for some patients may be sufficiently high to encourage consideration of other therapeutic options, including participation in a clinical trial.

The future ability to effectively prioritize and complete prospective clinical trials in breast oncology may be challenged by an increasing number of new treatments to test, questions of synergy in combined treatments, questions of efficacy in biological subsets, and generally low rates of participation in prospective clinical trials of adjuvant treatments.[27] It is relevant to consider whether more patients would choose to participate in a clinical trial from which they might benefit if they knew in advance that they had significant probability of early relapse following a current and intensive adjuvant treatment strategy. Furthermore, any increase in the rate of participation in prospective clinical trials, especially those that target therapeutic strategies for patients predicted to be insensitive to current standard treatments, could profoundly accelerate clinical investigation of new adjuvant treatments.[27]

It is essential to realize that prediction of pathologic response from neoadjuvant chemotherapy does not necessarily predict improved survival. This apparent paradox is attributable to the relationship between the biologic information captured by the predictor and the frequency and prognosis of false-positive response predictions. Tumors that are less differentiated often have high proliferation and are generally more likely to have poor prognosis, respond to cytotoxic chemotherapy, and have poor prognosis if they do not respond to chemotherapy.[1,28] Therefore, tests that are prognostic in the absence of chemotherapy may predict response in patients at higher risk but without an associated high probability of survival. This is likely to be most apparent in stage II-III breast cancers, as illustrated for GGI and intrinsic subtype (Table 2, eFigures 3-5),[11] and in reports concerning the commercially available 70-gene prognostic signature (that uses a different microarray platform)[10] and the recurrence score.[9,29,30] The latter tests could not be directly compared in our study because each uses a different technology lacking direct correlation with our Affymetrix microarray platform. A similar paradox was observed for DLDA30, a predictor trained on pathologic complete response in unselected patients,[6] and is probably related to different frequency of pathologic complete response according to ER status and grade, reflecting differentiation and proliferation (Table 2, eFigure 3C and G).[28] We conclude that prediction of pathologic response alone is not sufficient to demonstrate clinical validity of a test. There also must be a survival advantage, with an appropriately high survival estimate, for patients predicted to be treatment sensitive.

In this study, overcoming the prediction paradox involved the incorporation of several elements in the prediction algorithm (Figure 1) potentially relevant to other technologies or other types of cancer. One element was to identify patients whose excellent survival after chemoendocrine therapy was likely attributable to the endocrine sensitivity of their breast cancer.[13] That convinced us to evaluate predictors of chemotherapy response in breast cancers with predicted low endocrine sensitivity. Another element was to separately develop predictors within relevant phenotypic subsets using an improved measure of pathologic response (pathologic complete response or RCB-I).[12] That improvement is sufficient to reverse the prediction paradox, but not to a level of clinical utility (Table 2, eFigure 3D and H, and eTable 2). Yet another element was to develop predictors based on clinically relevant definitions of resistance, so that resistant disease could be identified first to avoid misclassification as responsive.[12] That further improved predictions to a level of potential clinical utility (Figures 2-4, Table 2, and eTable 2).

In its current format, this predictive test would be performed on fresh primary tumor sample obtained from clinical core needle biopsy (2 cores), fine-needle aspiration (2 passes), or surgical resection (eg, tumoral punch biopsy) and placed into a vial containing 1.5 mL of a standard RNA preservative (RNA*later*) at room temperature. Similar methods could become feasibly implemented into diagnostic practice if procurement of a tumor sample of optimal quality became a priority for molecular diagnostic tests of proven accuracy and medical utility. Meanwhile, it is also imperative to continue to evaluate the predictive accuracy of this test in additional validation studies.

Author Contributions: Dr Symmans had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Hatzis, Symmans.
Acquisition of data: Pusztai, Valero, Booser, Esserman, Lluch, Vidaurre, Holmes, Souchon, Martin, Cotrina, Gomez, Hubbard, Chacón, Ferrer-Lozano, Dyer, Buxton, Gong, Wu, Ibrahim, Andreopoulou, Ueno, Hunt, Yang, Nazario, DeMichele, Symmans.
Analysis and interpretation of data: Hatzis, Pusztai, Esserman, Wang, Ibrahim, Ueno, Hunt, O'Shaughnessy, Hortobagyi, Symmans.
Drafting of the manuscript: Hatzis, Symmans.
Critical revision of the manuscript for important intellectual content: Hatzis, Pusztai, Valero, Booser,

Esserman, Lluch, Vidaurre, Holmes, Souchon, Wang, Martin, Cotrina, Gomez, Hubbard, Chacón, Ferrer-Lozano, Dyer, Buxton, Gong, Wu, Ibrahim, Andreopoulou, Ueno, Hunt, Yang, Nazario, DeMichele, O'Shaughnessy, Hortobagyi, Symmans.
Statistical analysis: Hatzis, Wang.
Obtained funding: Pusztai, Esserman, Symmans.
Administrative, technical, or material support: Pusztai, Valero, Esserman, Lluch, Vidaurre, Holmes, Souchon, Hubbard, Hortobagyi, Symmans.
Study supervision: Pusztai, Valero, Holmes, Souchon, Martin, Gomez, Hortobagyi, Symmans.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Hatzis reported holding equity in his employer, Nuvera Biosciences, and holding joint patents with Nuera Biosciences. Dr Pusztai reported serving as an unpaid scientific advisor for Nuvera Biosciences and holding joint patents with Nuera Biosciences. Dr Holmes reported receiving honoraria from Novartis, Genentech, and Philips (with relevance outside the submitted work). Dr Gomez reporting receiving honoraria from GlaxoSmithKline and Bristol-Meyers Squibb (with relevance outside the submitted work). Dr Hortobagyi reported serving as a paid consultant to Allergan, Genentech, Novartis, and sanofi-aventis and receiving research funding from Novartis (with relevance outside the submitted work). Dr Symmans reported serving as an unpaid scientific advisor to Nuvera Biosciences; holding equity in Nuvera Biosciences; holding joint patents with Nuera Biosciences with royalty; and receiving honoraria from Agendia BV.

Funding/Support: This study was supported by grants from Susan G. Komen for The Cure (KG081680) to Dr Symmans; the National Cancer Institute (1 R01 CA106290) to Dr Pusztai; the Breast Cancer Research Foundation to Dr Pusztai and Dr Symmans; the Safeway Foundation to Dr Pusztai and Dr Symmans; faculty funds from Dr Gong, Dr Wu, and Dr Symmans; and generally by the Nellie B. Connally Breast Cancer Research Fund at the M. D. Anderson Cancer Center (MDACC). Additionally, the I-SPY-1 (Investigation of Serial Studies to Predict Your Therapeutic Response With Imaging and Molecular Analysis) trial was supported in part by grants from the National Cancer Institute to the Cancer and Leukemia Group B (Monica M. Bertagnolli, MD, chair [CA31946]) and to the Cancer and Leukemia Group B Statistical Center (Daniel J. Sargent, PhD [CA33601]) and by grants to participating sites: Georgetown University Medical Center, Washington, DC (Minetta C. Liu, MD [CA77597]), Memorial Sloan-Kettering Cancer Center, New York, NY (Clifford A. Hudis, MD [CA77651]), University of California at San Francisco (Charles J. Ryan, MD [CA60138]), University of Chicago, Chicago, Illinois (Hedy L. Kindler, MD [CA41287]), University of Minnesota, Minneapolis (Bruce A. Peterson, MD [CA16450]), University of North Carolina at Chapel Hill (Thomas C. Shea, MD [CA47559]), and University of Texas Southwestern Medical Center, Dallas (Debasish Tripathy, MD [CA37347]).

Role of the Sponsors: The sponsors had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

Statistical Analysis: Dr Hatzis conducted all analyses for signature development for the discovery cohort. Dr Wang had full access to all data from the validation cohort and conducted independent statistical analyses. Data sets for this study and gene expression weights for signatures are accessible via the GEO repository (http://www.ncbi.nlm.nih.gov/geo/) under accession identification numbers GSE25066 (combined series), GSE25055 (discovery cohort), and GSE25065 (validation cohort).

Disclaimer: The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute.

Online-Only Material: The eMethods, eTables 1-3, and eFigures 1-7 are available at http://www.jama.com.

Additional Contributions: We thank the patients who participated in this research and the following colleagues (uncompensated for their contributions) for technical laboratory work in the Department of Pathology: Feng Lin, MS, Bin Zeng, MS, Hongxia Sun, PhD, and Chunxiao Fu, PhD (MDACC Breast Cancer Pharmacogenomics Laboratory); obtaining biopsy samples: Nour Sneige, MD (cytopathology); obtaining clinical data used in this work: Eva Carrasco (Grupo Español de Investigación en Cáncer de Mama), Patricia de los Rios (Lyndon B. Johnson Hospital), Marc Lenburg (Investigation of Serial Studies to Predict Your Therapeutic Response With Imaging and Molecular Analysis), and Chingyi Young, MS, and Anna Kratz (MDACC); and the Departments of Breast Imaging, Pathology, Surgery, and Breast Medical Oncology at MDACC.

REFERENCES

1. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. *N Engl J Med*. 2009;360(8):790-800.
2. Wolmark N, Wang J, Mamounas E, Bryant J, Fisher B. Preoperative chemotherapy in patients with operable breast cancer: nine-year results from National Surgical Adjuvant Breast and Bowel Project B-18. *J Natl Cancer Inst Monogr*. 2001;(30):96-102.
3. Hayes DF, Thor AD, Dressler LG, et al; Cancer and Leukemia Group B (CALGB) Investigators. HER2 and response to paclitaxel in node-positive breast cancer. *N Engl J Med*. 2007;357(15):1496-1506.
4. Bear HD, Anderson S, Smith RE, et al. Sequential preoperative or postoperative docetaxel added to preoperative doxorubicin plus cyclophosphamide for operable breast cancer: National Surgical Adjuvant Breast and Bowel Project Protocol B-27. *J Clin Oncol*. 2006; 24(13):2019-2027.
5. 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(8):1275-1281.
6. 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(26):4236-4244.
7. Lee JK, Coutant C, Kim YC, et al. Prospective comparison of clinical and genomic multivariate predictors of response to neoadjuvant chemotherapy in breast cancer. *Clin Cancer Res*. 2010;16(2):711-718.
8. Popovici V, Chen W, Gallas BG, et al. Effect of training-sample size and classification difficulty on the accuracy of genomic predictors. *Breast Cancer Res*. 2010; 12(1):R5.
9. Albain KS, Barlow WE, Shak S, et al; Breast Cancer Intergroup of North America. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor–positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol*. 2010;11(1):55-65.
10. Straver ME, Glas AM, Hannemann J, et al. The 70-gene signature as a response predictor for neoadjuvant chemotherapy in breast cancer. *Breast Cancer Res Treat*. 2010;119(3):551-558.
11. 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(19):3185-3191.
12. Symmans WF, Peintinger F, Hatzis C, et al. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. *J Clin Oncol*. 2007;25(28):4414-4422.
13. Symmans WF, Hatzis C, Sotiriou C, et al. Genomic index of sensitivity to endocrine therapy for breast cancer. *J Clin Oncol*. 2010;28(27):4111-4119.
14. Symmans WF, Ayers M, Clark EA, et al. Total RNA yield and microarray gene expression profiles from fine-needle aspiration biopsy and core-needle biopsy samples of breast carcinoma. *Cancer*. 2003;97 (12):2960-2971.
15. Wolff AC, Hammond ME, Schwartz JN, et al; American Society of Clinical Oncology; College of American Pathologists. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol*. 2007; 25(1):118-145.
16. Gong Y, Yan K, Lin F, et al. Determination of oestrogen-receptor status and ERBB2 status of breast carcinoma: a gene-expression profiling study. *Lancet Oncol*. 2007;8(3):203-211.
17. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor–positive breast carcinomas through genomic grade. *J Clin Oncol*. 2007;25(10):1239-1246.
18. Desmedt C, Piette F, Loi S, et al; TRANSBIG Consortium. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res*. 2007; 13(11):3207-3214.
19. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node–negative primary breast cancer. *Lancet*. 2005;365 (9460):671-679.
20. Tibshirani RJ. Univariate shrinkage in the Cox model for high dimensional data. *Stat Appl Genet Mol Biol*. 2009;8(1):e21.
21. Friedman JH, Popescu BE. Gradient directed regularization for linear regression and classification. Stanford University Web site. http://www-stat.stanford.edu/~jhf/ftp/path.pdf. March 29, 2004. Accessed April 13, 2011.
22. Hudis CA, Barlow WE, Costantino JP, et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol*. 2007;25(15):2127-2132.
23. Moskowitz CS, Pepe MS. Quantifying and comparing the accuracy of binary biomarkers when predicting a failure time outcome. *Stat Med*. 2004; 23(10):1555-1570.
24. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer-Verlag; 2000.
25. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27(8):1160-1167.
26. Sparano JA, Wang M, Martino S, et al. Weekly paclitaxel in the adjuvant treatment of breast cancer. *N Engl J Med*. 2008;358(16):1663-1671.
27. Committee on Cancer Clinical Trials and the Cooperative Group Program. *A National Cancer Clinical Trials System for the 21st Century: Reinvigorating the NCI Cooperative Group Program*. Washington, DC: National Academies Press; 2010.
28. Symmans WF. A pathologist's perspective on emerging genomic tests for breast cancer. *Semin Oncol*. 2007;34(2)(suppl 3):S4-S9.
29. 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(29):7265-7277.
30. Goldstein LJ, Gray R, Badve S, et al. Prognostic utility of the 21-gene assay in hormone receptor–positive operable breast cancer compared with classical clinicopathologic features. *J Clin Oncol*. 2008; 26(25):4063-4071.