

Chemosensitivity and Stratification by a Five Monoclonal Antibody Immunohistochemistry Test in the NSABP B14 and B20 Trials

Douglas T. Ross,¹ Chung-yeul Kim,² Gong Tang,^{3,5} Olga L. Bohn,² Rodney A. Beck,⁷ Brian Z. Ring,¹ Robert S. Seitz,⁷ Soonmyung Paik,² Joseph P. Costantino,^{3,5} and Norman Wolmark^{4,6}

Abstract Purpose: To test the association between risk stratification and outcome in a prospectively designed, blinded retrospective study using tissue arrays of available paraffin blocks from the estrogen receptor – expressing, node-negative samples from the National Surgical Adjuvant Breast and Bowel Project B14 and B20 tamoxifen and chemotherapy trials.

Experimental Design: Tissue arrays were stained by immunohistochemistry targeting p53, NDRG1, SLC7A5, CEACAM5, and HTF9C. Risk stratification was done using predefined scoring rules, algorithm for combining scores, and cutoff points for low-risk, moderate-risk, and high-risk patient strata.

Results: In a univariate Cox model, this test was significantly associated with recurrence-free interval [HR, 1.3 (95% confidence interval, 1.1-1.6); $P = 0.006$]. In a multivariate model it contributed information independent of age, tumor size, and menopausal status ($P = 0.007$). The Kaplan-Meier estimates of the proportion of recurrence-free after 10 years were 73%, 86%, and 85% for the high-risk, moderate-risk, and low-risk groups ($P = 0.001$). The Kaplan-Meier estimates of the breast-cancer-specific-death rate were 23%, 10%, and 9% ($P < 0.0001$). Exploratory analysis in patients ≥ 60 years old showed Kaplan-Meier estimates of the proportion of recurrence-free of 78%, 89%, and 92%. Both high-risk and low-risk groups showed significant improvement on treatment with cytotoxic chemotherapy.

Conclusions: Immunohistochemistry using five monoclonal antibodies assigns breast cancer patients to a risk index that was significantly associated with clinical outcome among the estrogen receptor – expressing, node-negative tamoxifen-treated patients. It seems that the test may be able to identify patients who have greater absolute benefit from adjuvant chemotherapy compared with unstratified patient populations. Exploratory analysis suggests that this test will be most useful in clinical decision making for postmenopausal patients.

Great progress has been made in the development and clinical testing of treatments for early-stage, estrogen receptor – expressing breast cancer. The significant clinical benefit of adjuvant hormonal therapy has been clearly shown and has become an accepted part of standard treatment strategies. In contrast, adjuvant cytotoxic therapy has been shown to offer more moderate absolute improvement in clinical outcome, which creates uncertainty about its utility in an individual patient (1). The use of adjuvant cytotoxic therapy is therefore reliant on clinical judgment, leading to a less consistent clinical practice. Current prognostic algorithms use primarily clinical factors

including tumor size, stage, tumor grade, patient age at diagnosis, and overall comorbidity to help with stratifying risk to identify patients who might preferentially benefit from chemotherapy (2). It is widely accepted that the introduction of diagnostic tests that better stratify chemotherapy benefit based on intrinsic properties of each patient's tumor could help allow more informed choices about therapeutic options (3, 4).

Over the past several years, numerous multivariate index assays have been developed that measure gene and protein expression in breast cancer and distinguish clinically distinct patient populations. The varied approaches used to discover prognostic signatures have resulted in the creation of distinct assays measuring different targets. Nevertheless, the strength and reproducibility of these different tests and the finding that they largely classify patients similarly show that the measured biological differences between patients are stable and able to be reliably measured using different technologies (5, 6). Several clinical assays have been introduced, and one of them, based on a panel of 21 genes, has been validated as prognostic of clinical outcome using clinical trial patient populations (7, 8).

We previously reported the translation of gene expression patterns in breast cancer into a five immunohistochemistry reagent assay trained to predict recurrence in an estrogen receptor – expressing, node-negative breast cancer population, and validated using two independent institutional cohorts (9).

Authors' Affiliations: ¹Applied Genomics, Inc., Burlingame, California; ²Division of Pathology and ³Biostatistical Center, ⁴National Surgical Adjuvant Breast and Bowel Project; ⁵Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh; ⁶Allegheny General Hospital, Pittsburgh, Pennsylvania; and ⁷Applied Genomics, Inc., Huntsville, Alabama

Received 3/10/08; revised 7/8/08; accepted 7/9/08.

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.

Requests for reprints: Douglas T. Ross, Applied Genomics, Inc., 863 Mitten Avenue, Suite 103, Burlingame, CA 94010. Phone: 650-697-8900; Fax: 256-533-1474; E-mail: dross@applied-genomics.com.

©2008 American Association for Cancer Research.

doi:10.1158/1078-0432.CCR-08-0647

Translational Relevance

This article describes a validation study of Mammostrat, a five-antibody immunohistochemistry test for estimating the prognosis of tamoxifen-treated, estrogen receptor–expressing, node-negative breast cancer. These patients have a relatively good prognosis when treated with hormonal therapy alone. However, chemotherapy has been shown to provide clear benefit. Clinical tests that identify the subset of patients with higher risk of relapse and who derive the greatest benefit from chemotherapy are needed. Although there are molecular-based prognostic tests that fulfill such needs, they are expensive. The Mammostrat test can identify patients at higher risk of tumor progression and who derive robust benefit from cytotoxic chemotherapy. Conversely, patients predicted to have low risk of tumor progression by this test may elect to forgo the morbidity associated with cytotoxic therapy because their baseline prognosis is very good although they may still derive some benefit from chemotherapy. Because immunohistochemistry technology is widely distributed and relatively inexpensive, Mammostrat may be a cost-effective alternative to molecular-based tests.

The assay measures SLC7A5, involved in nutrient transport; p53, involved in cell cycle checkpoint control; HTF9C, a gene whose expression oscillates during the cell cycle; NDRG1, a stress- and hypoxia-inducible gene; and CEACAM5, a carcinoembryonic differentiation antigen. Our results showed that the assay was independent of clinical predictors and allowed a superior stratification of patients compared with a widely used measure of standard clinical parameters, the Nottingham prognostic index. To further validate and define the clinical utility of this assay in early-stage patients, we stained archived tumor samples from the National Surgical Adjuvant Breast and Bowel Project (NSABP) trials B14 and B20. These seminal clinical studies helped to establish the clinical benefit of adjuvant tamoxifen therapy and the addition of cytotoxic chemotherapy in nonmetastatic estrogen receptor–expressing breast cancer (1, 10–13). The prospectively designed retrospective studies reported herein were done to further test the association between the five-antibody test and clinical outcomes in estrogen receptor–expressing, node-negative tamoxifen-treated breast cancer patients and to determine whether the test identified patients who would have selectively benefited from adjuvant chemotherapy treatment.

Materials and Methods

Patients

Patients were included for study if there was clinical follow-up information and a paraffin block available for construction of tissue microarrays. Triplicate sections derived from the replicate tissue microarray blocks were stained by the five monoclonal antibodies and scored manually; inclusion of a patient in analysis required that at least one of the three blocks yielded an interpretable stain for each of the five studied biomarkers. For the work done at Applied Genomics, Inc., this study was classified as exempt by the New England Institutional Review Board because no identifiable patient clinical

information was made available to Applied Genomics at any time. The study was approved by the Institutional Review Board of Allegheny General Hospital.

Tissue arrays and immunohistochemistry

Tissue arrays and staining. Each tissue microarray block contained single 0.6-mm cores sampled from a maximum of 100 cases arranged in a 10 × 10 array format with three position orientation cores on one corner. Tissue microarrays are protocol specific. The Division of Pathology of the NSABP Foundation has an Institutional Review Board–approved protocol in place for assignment of a pathology serial number for all patients. Tissue array sections were deparaffinized and dehydrated by submerging in xylene thrice for 10 min each, followed by rinsing thrice in 100% ethanol, twice in 95% ethanol, and then treated by microwaving, boiling for 11 min in 10 μmol/L buffered citrate (pH 6.0). Slides were allowed to cool to room temperature and rinsed in distilled water followed by PBS. Slides were dipped in 0.03% hydrogen peroxide followed by rinsing with PBS and then stained with dilutions of antibodies in Dako diluent (DakoCytomation, Inc.) for 1 h at room temperature. Secondary antibody was applied for 1 h, and staining visualized with the DakoCytomation Envision staining kit according to the manufacturer's instructions. For each antibody, dilutions were first tested on a small "titer" tissue array that had breast cancer cases with positive and negative cases for all antibodies in the panel, in addition to a set of tumor-derived cell lines suspended in paraffin.

Tissue array scoring. Manual scoring was first done on a semiquantitative scale wherein the invasive breast cancer epithelium present in each tissue core was scored by microscopic evaluation as negative, weak, or strong. Scores were converted to a binary scale of positive or negative or no information when instructive tumor was absent. Triplicate scores were reviewed by evaluation of images of the cores using a custom database of scores and images. Staining on any one of the replicate tissue array cores was interpreted as positive staining for the patient as a whole.

Scoring rules.

SLC7A5 (rabbit monoclonal s0720): Score "positive" when membrane staining is present on >10% of invasive tumor cells.

p53 (mouse monoclonal D0-7): Score positive when nuclear staining is present on >10% of invasive tumor cells.

NDRG1 (rabbit monoclonal s0721): Score positive when a confluent area of uniform cytoplasmic or membrane positive staining is present. NDRG1 staining is present proximal to hypoxic areas near necrotic centers. This nonhomogeneous staining pattern is interpreted as negative.

HTF9C (rabbit monoclonal antibody s0722): Score positive only when a cytoplasmic staining pattern is present on >10% of invasive tumor cells. Occasional nuclear staining is ignored for scoring purposes.

CEACAM5: Score positive when cytoplasmic and/or membrane staining is present on >10% of invasive tumor cells.

Measurement of estrogen receptor expression by the ligand binding assay has previously been described (14).

Study design and analysis

Concordance between individual tissue array replicates was determined by pairwise comparison of all available duplicate stains on independent stained tissue array cores scored by a single scorer or by single stained cores scored by two different scorers. Concordance results reflect scores obtained by primary evaluation by microscope before subsequent review of discordant scores.

A prospective analysis plan including scoring rules and predefined study end points was established as a written document before staining of the tissue array slides. Staining and scoring was done in Applied Genomics laboratories and scores were linked to an anonymous case identifier. Scores and a calculated risk index for each patient were sent

Table 1. Patient numbers in the B14 and B20 clinical trial randomization arms used for development and validation of the 21 gene test; for validation the 5 antibody test described herein; and total trial participant populations

	Enrolled patients		
	21 gene	5 Ab	Clinical trial
Placebo	N/A	287	1,414
B14 Tam	668	550	2,615
B20 Tam	233	161	771
B20 Chemo	440	296	1,535

Abbreviations: Ab, antibody; N/A, not available; Tam, tamoxifen; Chemo, chemotherapy.

by Applied Genomics to an honest broker who merged the staining data with clinical data from the NSABP Biostatistical Center according to a link file from the Pathology Division of NSABP. All statistical analyses were done by a statistician from the NSABP Biostatistical Center, and exploratory analyses were done at the discretion of the statistician. Prognostic score was calculated using the prespecified algorithm: Prognostic index = (1.542 × SLC7A5) + (1.124 × p53) + (1.058 × NDRG1) + (0.712 × HTF9C) + (0.504 × CEACAM) - 0.864. Patients were classified into the three risk categories such that prognostic index ≤0 defines the “low-risk” group; prognostic index >0 and ≤0.7 define the “moderate-risk” group; prognostic index >0.7 defines the “high-risk” group. The algorithm and end points are unchanged relative to that specified in previously published studies (9).

Clinical end points were defined as follows: recurrence-free interval, time from enrollment to any breast cancer recurrence or death with evidence of breast cancer; distant recurrence-free interval, time from enrollment to breast cancer recurrence in any area of the body excluding invasive or *in situ* breast cancer recurrence (except lobular carcinoma *in situ*) in the ipsilateral breast or chest wall or in the ipsilateral internal mammary, supraclavicular, infraclavicular, and/or axillary nodes, as well as the soft tissue of the ipsilateral axilla; breast cancer-specific death, time from enrollment to death due to breast cancer. Pathologic grading was not available in this study.

Prospectively defined study aims

Prognosis in tamoxifen-treated patients was evaluated by using patients from both the B14 and B20 trials. The primary aim was to test the hypothesis that the risk category, coded as an interval variable with values 0, 1, and 2, is significantly associated with recurrence-free interval in a univariate Cox model. A two-sided *P* value for the likelihood ratio test of <0.05 was considered significant. Secondary aims were to test the hypothesis that the distributions of distant recurrence-free interval and breast cancer-specific death were significantly distinct among the high-risk, moderate-risk, and low-risk groups and to test the hypothesis that the five-antibody model contributed independently from age and clinical tumor size in predicting recurrence-free interval, distant recurrence-free interval, and breast cancer-specific death. A two-sided *P* value of <0.05 for the effect associated with the prognostic categories as an interval variable after adjusting age and clinical tumor size in a Cox proportional hazards model was considered significant. Exploratory analysis was done to determine recurrence events (recurrence-free interval) in patients separated into age and quantitative estrogen receptor expression groups measured by the ligand binding. The explored age groups were <40, 40 to 50, 50 to 60, and ≥60 y old. The estrogen receptor-expressing groups were 10 to 49, 50 to 99, 100 to 199, and ≥200 fmol/mg.

Prediction of chemotherapy benefit was carried out by comparing patients treated with tamoxifen only to those treated with tamoxifen plus chemotherapy using patients from the B20 trial. The primary aim was to test the hypothesis that the distribution of recurrence events in comparison between tamoxifen-treated and tamoxifen plus chemotherapy-treated patients during the 10-year study period was significantly distinct within any of the (Applied Genomics)-defined patient subclasses. The interaction between the risk categories and treatment options (tamoxifen plus chemotherapy versus tamoxifen only) was tested in a multivariate Cox model.

Prognosis in patients who did not receive adjuvant treatment was evaluated based on patients from the B14 placebo arm. The primary aim was to test the hypothesis that the distribution of the recurrence events (recurrence-free interval) was significantly distinct among the high-risk, moderate-risk, and low-risk groups.

Cohort Characteristics

The NSABP B14 and B20 studies were randomized prospective studies of the role of adjuvant tamoxifen and adjuvant chemotherapy in treatment of node-negative, estrogen receptor-expressing breast cancer, respectively. The B14 trial was a randomized study comparing

Table 2. Summary of clinical characteristics of the populations used for the twenty-one gene and five antibody tests

	Cohort characteristics (% of cohort)				
	21-gene B14	5-Ab B14 and B20 Tam only	Clinical B14 Tam only	Clinical B20 Tam only	Clinical B20 Chemo
Age (y)					
<50	29	32	31	45	46
50-59	26	27	29	28	28
>60	45	41	40	27	26
Clinical tumor size (cm)					
<2.0	62	59	63	70	69
2.1-4.0	33	35	33	27	27
>4.1	5	5	4	3	4
ER (fmol/mg)					
10-49	36	37	43	44	45
50-99	21	22	23	22	22
>100	42	41	34	34	33

Abbreviation: ER, estrogen receptor.

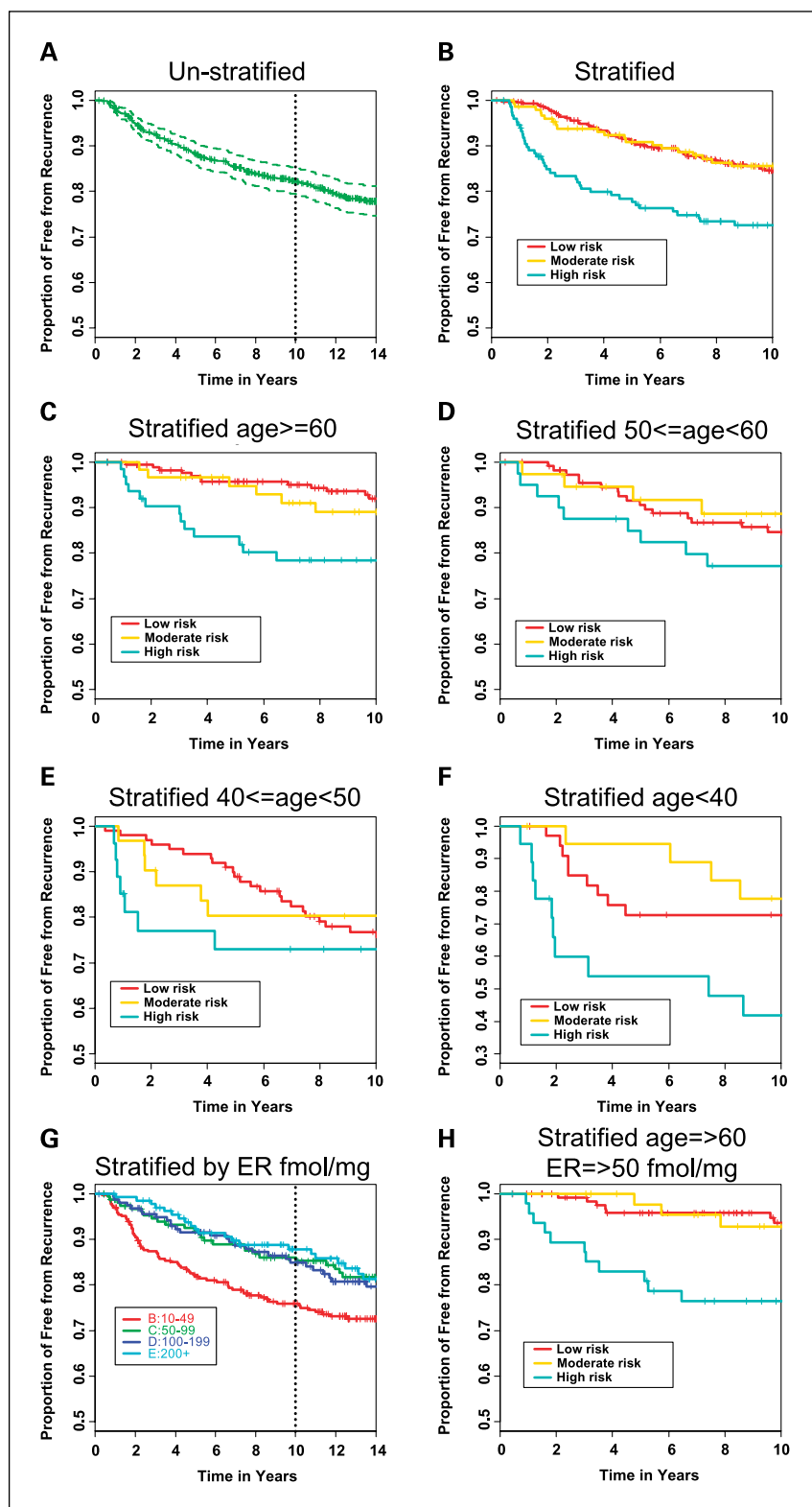


Fig. 1. Tamoxifen-treated patients. Recurrence-free interval (RFI) in B14 and B20 combined tamoxifen-only – treated patients. *A*, all patients unstratified. *B*, all patients stratified into low-risk, moderate-risk, and high-risk groups by the prognostic index. *C*, patients ages ≥ 60 y old. *D*, patients $50 \leq \text{age} < 60$ y. *E*, patients $40 \leq \text{age} < 50$ y. *F*, patients ages < 40 y. *G*, all patients stratified by quantitative estrogen receptor status measured by the ligand binding assay (14). *H*, patients with relatively high estrogen receptor staining (≥ 50 fmol/mg) and ages ≥ 60 y stratified into low-risk, moderate-risk, and high-risk groups by the prognostic index.

placebo-treated patients with tamoxifen-treated patients and added a tamoxifen-only registration arm after early results showed a clear and striking benefit for adjuvant tamoxifen therapy. The B20 trial randomized patients either to the arm receiving tamoxifen only or to one of two arms receiving either tamoxifen plus cyclophosphamide and

methotrexate (CM) or tamoxifen plus CM and fluorouracil (CMF; ref. 10). Paraffin blocks were archived for a subset of trial patients and available for construction of tissue arrays. These same samples were previously used for the development and validation of the 21-gene predictor, a validated prognostic reverse transcription-PCR (RT-PCR)

Table 3. Cox proportional hazard models used to estimate the hazard ratio associated with assignment to risk strata by the five antibody model in tamoxifen-only treated patients

Cox proportional hazard models				
Variable	P	HR (95% CI)	P	
RFI				
Low, moderate, high	0.006	1.3 (1.1-1.6)	0.003	
Moderate vs low	0.51	0.9 (0.5-1.3)		
High vs low	0.002	1.8 (1.2-2.6)		
DRFI				
Low, moderate, high	0.001	1.4 (1.1-1.7)	0.002	
Moderate vs low	0.85	1.0 (0.6-1.7)		
High vs low	0.0004	2.1 (1.4-3.1)		
BCSD				
Low, moderate, high	0.0003	1.5 (1.2-1.9)	0.0004	
Moderate vs low	0.96	1.0 (0.6-1.7)		
High vs low	<0.0001	2.3 (1.5-3.5)		

Abbreviations: RFI, recurrence-free interval; DRFI, distant recurrence-free interval; BCSD, breast cancer-specific death.

assay for early-stage, tamoxifen-treated, estrogen receptor-expressing breast cancer patients (7).

In this study, we aimed to assess the association between clinical outcomes in early-stage, estrogen receptor-expressing breast cancer and stratification by a novel five-antibody immunohistochemistry test. For the study of tamoxifen-treated patients, we combined all available tamoxifen-only-treated patients from both the B14 and B20 studies. For the study of chemotherapy responsiveness, the CM and CMF arms were combined because they were similar, and they were compared with the B20 tamoxifen-only-treated patients. All available B14 placebo-treated patients were also analyzed. Paraffin blocks were available for ~20% of the original combined B14 and B20 tamoxifen-only arms, the B20 chemotherapy trial, and the B14 placebo treatment arm (Table 1). Compared with samples previously used for development and validation of the OncoType DX assay, there were 20% fewer patients available in the tissue array for the B14 tamoxifen arm and 32% fewer patients available from the B20 trial. In the B14 study population herein, 69% of patients were >50 years of age, whereas in the B20 clinical study only 55% were in this age group. The latter differs significantly from the population demographics in the United States where 73% of estrogen receptor-expressing, node-negative patients are between 50 and 70 years of age⁸ (Table 2).

Tamoxifen-Treated Patients

Among the 711 combined B14 and B20 tamoxifen-treated patients who are eligible for this study, ~58% were in the low-risk category and 21% each in moderate-risk and high-risk categories. There was a statistically significant association between patients stratified by the immunohistochemistry test and the recurrence-free interval in the combined B14 and B20 tamoxifen-treated patient cohorts [HR, 1.3; 95% confidence interval (95% CI), 1.1-1.6; *P* = 0.006; Fig. 1A and B]. Low-risk and moderate-risk groups were not significantly different (log-rank test, *P* = 0.5), whereas low-risk was significantly distinct from high-risk [HR, 1.8 (95% CI, 1.2-2.6); log-rank test, *P* = 0.001]. In the unstratified cohort, the Kaplan-Meier estimate of the proportion of patients who were recurrence-free after 10 years was 82% (95% CI, 79-85%) overall. When comparing the three categories of risk level, the Kaplan-Meier estimates were 85% (95% CI, 81-88%) for patients

classified as low-risk, 85% (95% CI, 80-91%) with moderate-risk, and 73% (95% CI, 65-80%) for high-risk patients. There was also a statistically significant association with two secondary predefined end points: distant recurrence-free interval and breast cancer-specific death (Table 3).

In a multivariate Cox model, the immunohistochemistry model provided significant prognostic power that was independent of age and tumor size [HR, 1.3 (95% CI, 1.1-1.6); *P* = 0.007]. Age ≥60 years contributed independent information to the model [HR, 0.5 (95% CI, 0.36-0.74); *P* = 0.003], whereas the effect of tumor size was not significant [HR, 1.3 (95% CI, 1.0-1.8); *P* = 0.07; Table 4; Fig. 1C-F]. The immunohistochemistry model was able to identify a group of patients with significantly worse outcomes than cohort averages regardless of age, and there was no significant interaction between age ≥60 years and the model (*P* = 0.6). Among patients ≥60 years old, the low-risk group had a Kaplan-Meier estimate of 92% (88-96%), the moderate-risk group had 89% (95% CI, 81-97%), and the high-risk group had 78% (95% CI, 68-88%). In the same age group, the low-risk patient stratum had a Kaplan-Meier estimate of breast cancer-specific death of 6% (95% CI, 2-9%) at 10 years, whereas the high-risk patients had a breast cancer-specific death rate of 22% (95% CI, 11-32%; Table 5). In the unstratified cohort, low expression of estrogen receptor measured by the ligand binding assay (<50 fmol/mg) was associated with significantly worse outcome (ref. 14; Fig. 1G). In patients ≥60 years of age and limited to the subset that showed moderately high level expression of estrogen receptor and higher (≥50 fmol/mg), the five-antibody algorithm identified a high-risk group of patients with a Kaplan-Meier estimate of the recurrence-free interval of 24% (95% CI, 14-38%) and a low-risk group with a 6% (95% CI, 3-13%) 10-year recurrence rate (Fig. 1H).

Three of the five individual biomarkers were associated with the recurrence-free interval by univariate analysis (SLC7A5, HTF9C, and p53) whereas two (NDRG1 and CEACAM5) were not significant in this cohort (data not shown). Pairwise reproducibility between stains done on independent triplicate arrays and read by a single scorer ranged from 85% (p53) to 92% (CEACAM5) concordance. Reproducibility of the interpretation of a single stained cohort array read by two different scorers ranged from 86% (p53) to 96% (SLC7A5 and NDRG1; Table 6).

Tamoxifen- and Cytotoxic Chemotherapy-Treated Patients

In the unstratified B20 patient population used in this study, the chemotherapy-treated patients showed a 5% absolute decrease in recurrences with the addition of chemotherapy. The immunohistochemistry test identified high-risk and low-risk groups that both showed significant improvement on treatment with cytotoxic chemotherapy. The Kaplan-Meier estimate of the recurrence-free interval in low-risk patients improved by 5%, from 86% to 91% [HR, 0.4 (95% CI, 0.2-0.8); *P* = 0.01], whereas in high-risk patients the recurrence-free interval improved by 21%, from 64% to 85% [HR, 0.4 (95% CI, 0.2-0.9); *P* = 0.02; Fig. 2A-D]. The moderate-risk group showed no significant difference between chemotherapy-treated and

Table 4. Estimated hazard ratios and significance of individual components of a multivariable model, which includes the five antibody model, age, and clinical tumor size

Variables	P	HR (95% CI)
Age ≥60 y	0.0003	0.5 (0.4-0.7)
CTS >2 cm	0.07	1.3 (1.0-1.8)
Low, moderate, high	0.007	1.3 (1.1-1.6)

Abbreviation: CTS, clinical tumor size.

⁸ Surveillance, Epidemiology, and End Results (SEER) program. (<http://www.seer.cancer.gov>) released April 2007, based on the November 2006 submission.

Table 5. Kaplan-Meier estimates of the event-free rates after 10 years in tamoxifen-only treated patients age greater than 60 years old

K-M estimates for the event-free rates after 10 y, age >60 y (%)			
Strata	RFI	DRFI	BCSD
Risk			
Low	92.0 (87.3-95.9)	92.8 (87.3-95.9)	94.30 (89.3-97)
Moderate	89.0 (77.1-94.9)	88.7 (76.4-94.8)	92.6 (81.4-97.2)
High	78.5 (65.8-86.9)	79.9 (67.2-88)	78.3 (65.5-86.8)

Abbreviation: K-M, Kaplan-Meier.

tamoxifen-only-treated patients (Table 7). A test for an interaction between chemotherapy and the risk group stratification was not significant ($P = 0.13$).

Placebo-Treated Patients

There was no significant association between the recurrence-free interval at 10 years and stratification by the immunohistochemistry test for placebo-treated patients (Fig. 3A and B).

Discussion

The current goal of prognostic testing in early-stage, estrogen receptor-expressing breast cancer is to provide information to aid in the choice between using hormonal therapy alone or hormonal therapy in conjunction with cytotoxic chemotherapy. Prior clinical studies with unstratified patient populations have not established whether there is a clinical benefit by adding adjuvant chemotherapy to tamoxifen in estrogen receptor-expressing patients >60 years old (1). In general, there is a bias in younger patients, who have overall worse prognosis relative to older patients, to treat with cytotoxic chemotherapy. Cytotoxic chemotherapy is used more cautiously in elderly or frail patients with otherwise limited expected life expectancy.

This work shows that a prognostic model that uses immunohistochemistry staining using five monoclonal antibodies combined in a multivariate index assay is able to identify a group of estrogen receptor-expressing, tamoxifen-treated, node-negative patients who have a relatively poor prognosis and a greater absolute benefit from chemotherapy compared with unstratified patient populations. In the study using samples from B20 adjuvant cytotoxic chemotherapy trial, high-risk patients and low-risk patients both showed an absolute benefit from chemotherapy; however, the magnitude of benefit in high-risk patients was four times greater than in low-risk patients. The intermediate-risk group was not well separated from the low-risk group in these studies and therefore, with current data, is uninformative for managing patients classified as intermediate risk. These studies were blinded, prospectively designed retrospective trials in which no specimens from the test cohorts were used in the development of the test.

The strong contribution of age to multivariate models containing the immunohistochemistry prognosticator prompted us to explore in more detail the strength of the immunohistochemistry model in different age strata. The training cohort for the five-antibody test was composed of

83% of patients >50 years old, whereas both the B14 and B20 clinical trial cohorts were significantly younger (69% and 55%, respectively). Younger patients in the low-risk group identified by this immunohistochemistry test had a 20% risk of disease progression that warrants consideration of aggressive treatment strategies, consistent with current clinical practice. In contrast, in patients of ages ≥ 60 years with whom cytotoxic chemotherapy is currently used much more cautiously, the test identified high-risk patients with a 22% risk of breast cancer-specific death compared with 6% in low-risk patients. This, combined with the absolute 21% decrease in recurrence rate associated with the administration of adjuvant chemotherapy identified in the high-risk patient strata in the B20 study, suggests that elderly high-risk patients may gain more clinical benefit relative to low-risk patients from accepting the risk and morbidity of cytotoxic chemotherapy. It should be noted that stratification into age groups was not a prespecified analysis in the trial design and therefore should be confirmed in additional studies.

The B14 and B20 paraffin-block trial samples were also used in the development and validation of the 21-gene predictor, OncoType DX (7, 8). This test uses preparation of RNA from macrodissected paraffin-block sections and RT-PCR of prepared RNA to detect levels of gene expression. Because 20% and 30%, respectively, of blocks were depleted by the RT-PCR analysis before this immunohistochemistry study, a direct comparison of results is not possible. However, a number of interesting trends are worth noting. The two overlapping studied cohorts had comparable 10-year recurrence rates overall, but the RT-PCR test identified a low-risk group with modestly better outcomes compared with the immunohistochemistry test (7% for the molecular test compared with 11% for the immunohistochemistry test). The difference in outcomes seems to be biased toward differences in performance in younger patients. For example, in patients >60 years of age, the RT-PCR and immunohistochemistry cohorts had very similar recurrence rates before risk stratification, and the tests done comparably in assigning patients to low-risk strata (low-risk distant recurrence-free interval of 7.2% for the immunohistochemistry test compared with 7.5% for the 21-gene test), whereas in younger patients, the RT-PCR cohort had lower rates of tumor progression in its low-risk group. Thus, a difference in performance in younger patients seems to account for much of the overall difference in performance of the two tests. Because epidemiologic work suggests that breast tumors that occur in premenopausal as opposed to postmenopausal patients are biologically distinct and the immunohistochemistry test was trained predominantly on postmenopausal patients, it is possible that the immunohistochemistry test is best suited to postmenopausal patients (15). Whether the difference in

Table 6. Concordance between replicate stains interpreted by a single scorer or between duplicate scores of single stains read by two independent scorers

	Pairwise stain reproducibility				
	SLC7A5	NDRG1	HTF9C	CEACAM5	p53
Replicate stains	0.89	0.91	0.89	0.92	0.85
Duplicate scores	0.96	0.96	0.92	0.94	0.86

performance in younger patients is due to such intrinsic biological differences between the two tests or to biased depletion of cohort samples between the molecular work and tissue array construction will require additional trials wherein more comparable cohorts are studied.

Few previous studies have evaluated biomarkers in randomized clinical trial populations relative to their utility in prediction of chemotherapy benefit. Only patient age, tumor size, and tumor grade have gained widespread use as markers of prognosis. There is relatively strong evidence to support the use of immunohistochemistry markers of proliferation (e.g., Ki67/MIB-1) and measurements of hormone receptor status (e.g., quantitative estrogen receptor and progesterone receptor) as prognosticators for the natural history of hormonally treated breast cancer (2, 3). The utility of additional diagnostic testing hinges on both the clinical effect of the result and the technical reliability of the test. Every technology applied to diagnostic testing, whether it is morphologic, immunologic, or molecular based, requires demonstration of the analytic reproducibility and clinical effect of testing. Morphology and immunohistochemistry have mixed records as diagnostic tests, but both have shown remarkably good reproducibility, when done by skilled specialists and in experienced high-volume laboratories. Immunohistochemistry-based assays have the advantage over molecular assays in that detection of biomarker expression is done and evaluated *in situ*, which gives confidence that measurement of expression is not confounded by expression in a variable admixture of normal or benign breast tissue and/or contaminating inflammatory cells and stroma. The ~85% to 95% reproducibility of the replicate stains and stain interpre-

Table 7. Kaplan-Meier estimates of the event-free rates after 10 years in individual risk strata comparing those treated with tamoxifen-only to those treated with tamoxifen plus chemotherapy

RFI: K-M estimates of proportion of recurrence-free after 10 y			
Strata	Treatment	K-M estimate (95% CI)	P
All patients	Tam	0.83 (0.77-0.89)	0.014
	Tam + Chemo	0.88 (0.84-0.92)	
Low risk	Tam	0.86 (0.79-0.93)	0.01
	Tam + Chemo	0.91 (0.87-0.96)	
Moderate risk	Tam	0.87 (0.73-1)	0.62
	Tam + Chemo	0.82 (0.71-0.93)	
High risk	Tam	0.64 (0.45-0.83)	0.017
	Tam + Chemo	0.85 (0.77-0.93)	

tation in this study using the very limited tissue available in 600- μ m tissue array cores suggest that these immunohistochemistry biomarkers are technically robust markers. Assay performance should improve further under clinical practice conditions wherein whole sections are available for interpretation (16). In addition, standardization of laboratory practice, automated staining, and quantitative image analysis approaches promise to improve laboratory consistency and performance of immunohistochemistry (17). Information from biomarker expression *in situ* such as offered by immunohistochemistry assays may be prudent in the setting of limited tissue

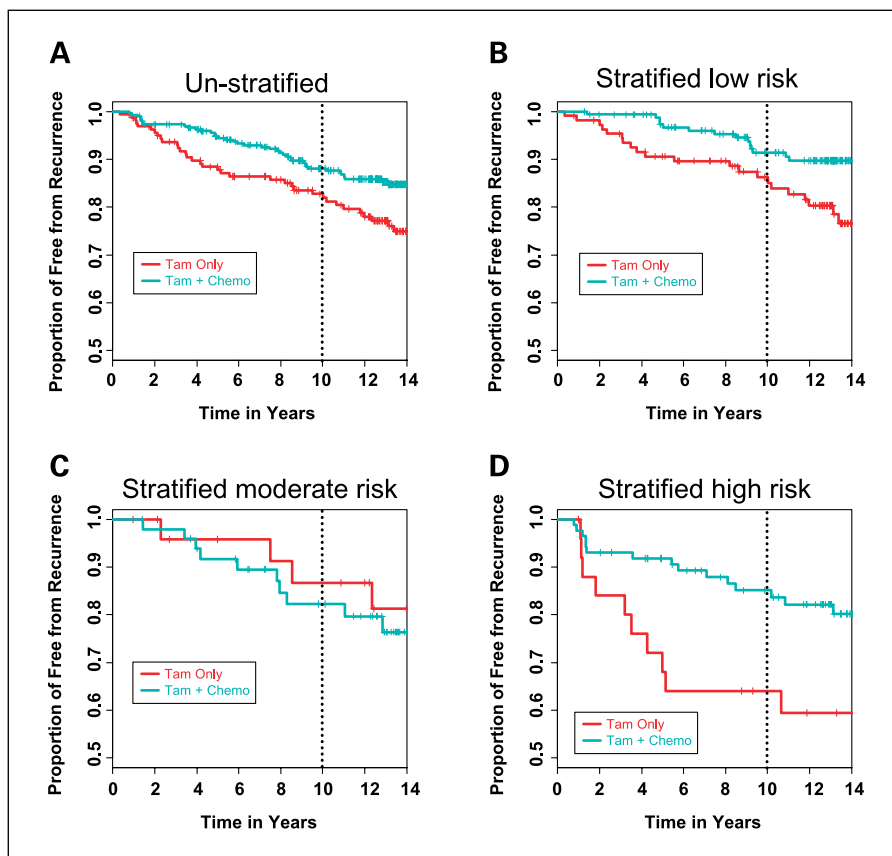
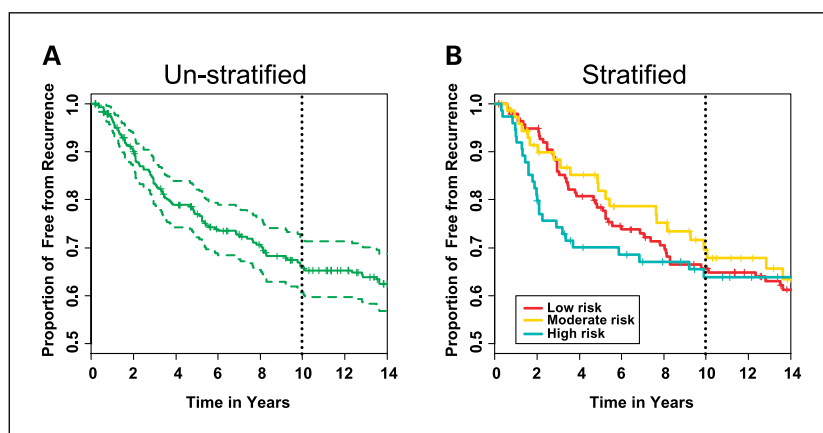


Fig. 2. Tamoxifen-treated versus tamoxifen plus chemotherapy – treated patients. Patients in the B20 tamoxifen versus tamoxifen plus chemotherapy trial stratified into treatment arms (chemotherapy arms combined). A, all patients unstratified; B, patients classified as low risk; C, patients classified as moderate risk; D, patients classified as high risk.

Fig. 3. Placebo-treated patients. All patients in the B14 placebo treatment arm. *A*, all patients unstratified. *B*, patients stratified into low-risk, moderate-risk, and high-risk groups by the prognostic index.



such as that available through the use of needle or core biopsies, which are becoming more common as diagnostic specimens with the emergence of neoadjuvant approaches to treating breast cancer.

The demonstration that five immunohistochemistry antibodies distinguish breast cancer patients at high risk of recurrence and who seem to have a robust response to chemotherapy suggests that this test may be useful for helping manage early-stage, estrogen receptor-expressing breast cancer. Patients identified as high risk for recurrence may be more inclined to elect aggressive treatment approaches, whereas patients predicted to have low risk of tumor progression may elect to forgo the morbidity associated with cytotoxic therapy because their baseline prognosis is very good although they may still derive some benefit from chemotherapy. Further testing in postmenopausal patients as well as exploration of additional markers selectively on premenopausal patients should be done to continue to tailor the test to accommodate

the recognized biological and clinical diversity of breast cancer patients. Because immunohistochemistry technology is widely distributed and relatively inexpensive, this five-antibody test may be a cost-effective alternative to molecular-based tests. Given the cost advantage, the ease of testing, and the confidence engendered by *in situ* visualization of marker expression, immunohistochemistry approaches will likely continue to have a significant role in managing decision making for early-stage breast cancer.

Disclosure of Potential Conflicts of Interest

D.T. Ross, R.A. Beck, R.S. Seitz, and B.Z. Ring are employees of and holders of stock interest in Applied Genomics, Inc., which developed and sells the five-antibody test described herein. C. Kim, G. Tang, O. M. Mejia, S. Paik, J. P. Costantino, and N. Wolmark declare no competing financial conflicts.

Applied Genomics, Inc., remunerated the NSABP to in part defray costs of the collaborative work done by the NSABP.

References

- Fisher B, Jeong JH, Bryant J, et al. Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from National Surgical Adjuvant Breast and Bowel Project randomised clinical trials. *Lancet* 2004;364:858–68.
- Schnitt SJ. Traditional and newer pathologic factors. *J Natl Cancer Inst Monogr* 2001;(30):22–6.
- Hayes DF, Isaacs C, Stearns V. Prognostic factors in breast cancer: current and new predictors of metastasis. *J Mammary Gland Biol Neoplasia* 2001;6:375–92.
- Di Leo A, Cardoso F, Durbecq V, et al. Predictive molecular markers in the adjuvant therapy of breast cancer: state of the art in the year 2002. *Int J Clin Oncol* 2002;7:245–53.
- Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006;355:560–9.
- 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;25:1239–46.
- Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817–26.
- Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;24:3726–34.
- Ring BZ, Seitz RS, Beck R, et al. Novel prognostic immunohistochemical biomarker panel for estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;24:3039–47.
- Fisher B, Dignam J, Bryant J, Wolmark N. Five versus more than five years of tamoxifen for lymph node-negative breast cancer: updated findings from the National Surgical Adjuvant Breast and Bowel Project B-14 randomized trial. *J Natl Cancer Inst* 2001;93:684–90.
- Fisher B, Dignam J, Wolmark N, et al. Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. *J Natl Cancer Inst* 1997;89:1673–82.
- Fisher B, Jeong JH, Anderson S, Wolmark N. Treatment of axillary lymph node-negative, estrogen receptor-negative breast cancer: updated findings from National Surgical Adjuvant Breast and Bowel Project clinical trials. *J Natl Cancer Inst* 2004;96:1823–31.
- Fisher B, Jeong JH, Dignam J, et al. Findings from recent National Surgical Adjuvant Breast and Bowel Project adjuvant studies in stage I breast cancer. *J Natl Cancer Inst Monogr* 2001;(30):62–6.
- Bryant J, Fisher B, Gunduz N, Costantino JP, Emir B. S-phase fraction combined with other patient and tumor characteristics for the prognosis of node-negative, estrogen-receptor-positive breast cancer. *Breast Cancer Res Treat* 1998;51:239–53.
- Anderson WF, Chen BE, Brinton LA, DeVesa SS. Qualitative age interactions (or effect modification) suggest different cancer pathways for early-onset and late-onset breast cancers. *Cancer Causes Control* 2007;18:1187–98.
- Goethals L, Perneel C, Debucquoy A, et al. A new approach to the validation of tissue microarrays. *J Pathol* 2006;208:607–14.
- Goldstein NS, Hewitt SM, Taylor CR, Yaziji H, Hicks DG. Recommendations for improved standardization of immunohistochemistry. *Appl Immunohistochem Mol Morphol* 2007;15:124–33.