

The *HOXB13:IL17BR* Expression Index Is a Prognostic Factor in Early-Stage Breast Cancer

Xiao-Jun Ma, Susan G. Hilsenbeck, Wilson Wang, Li Ding, Dennis C. Sgroi, Richard A. Bender, C. Kent Osborne, D. Craig Allred, and Mark G. Erlander

A B S T R A C T

Purpose

We previously identified three genes, *HOXB13*, *IL17BR* and *CHDH*, and the *HOXB13:IL17BR* ratio index in particular, that strongly predicted clinical outcome in breast cancer patients receiving tamoxifen monotherapy. Confirmation in larger independent patient cohorts was needed to fully validate their clinical utility.

Patients and Methods

Expression of *HOXB13*, *IL17BR*, *CHDH*, estrogen receptor (ER) and progesterone receptor (PR) were quantified by real-time polymerase chain reaction in 852 formalin-fixed, paraffin-embedded primary breast cancers from 566 untreated and 286 tamoxifen-treated breast cancer patients. Gene expression and clinical variables were analyzed for association with relapse-free survival (RFS) by Cox proportional hazards regression models.

Results

ER and PR mRNA measurements were in close agreement with immunohistochemistry. In the entire cohort, expression of *HOXB13* was associated with shorter RFS ($P = .008$), and expression of *IL17BR* and *CHDH* was associated with longer RFS ($P < .0001$ for *IL17BR* and $P = .0002$ for *CHDH*). In ER+ patients, the *HOXB13:IL17BR* index predicted clinical outcome independently of treatment, but more strongly in node-negative patients. In multivariate analysis of the ER+ node-negative subgroup including age, PR status, tumor size, S phase fraction, and tamoxifen treatment, the two-gene index remained a significant predictor of RFS (hazard ratio = 3.9; 95% CI, 1.5 to 10.3; $P = .007$).

Conclusion

This tumor bank study demonstrated *HOXB13:IL17BR* index is a strong independent prognostic factor for ER+ node-negative patients irrespective of tamoxifen therapy.

J Clin Oncol 24:4611-4619. © 2006 by American Society of Clinical Oncology

INTRODUCTION

Breast cancer is a heterogeneous disease with a highly variable clinical course, presenting a great challenge to prognosis and therapeutic decisions. To help guide treatment decision making, several guidelines have been established.^{1,2} The most recent (ninth) edition of the St Gallen guidelines considers both endocrine responsiveness and prognostic risk assessment in forming treatment decisions.² Historically, hormone receptor status, a target for endocrine therapies, has been considered the standard for predicting response to treatment.^{3,4} However, positive receptor status is not sufficient to ensure a therapeutic response because additional molecular alterations such as *HER2* amplification and *EGFR* expression are thought to modify a tumor's endocrine responsiveness.⁵⁻⁷ Similarly, prognosis has largely been based on clinical (eg, age and meno-

pausal status) and pathologic parameters (eg, tumor size, grade and lymph node status). However, a subset of patients with a "good" prognosis (eg, estrogen receptor-positive [ER+] and node negative) may still develop recurrence after curative surgery and adjuvant therapy.³ An improved understanding of the underlying molecular pathways that drive breast cancer development offers new opportunities for both predicting a tumor's responsiveness to treatment and assessing a tumor's intrinsic aggressiveness. The development of microarray technology has facilitated novel translational research promising significant progress in these areas.⁸⁻¹⁶

To discover novel biomarkers predicting tamoxifen response in the adjuvant setting, we have previously conducted a microarray-based survey of gene expression patterns that correlate with clinical outcome.¹⁵ In the initial cohort of 60 tamoxifen-treated patients, we identified three genes, *HOXB13*

From AviraDx Inc, Carlsbad, CA; The Breast Center, Baylor College of Medicine, Houston, TX; Department of Pathology, Harvard Medical School, Molecular Pathology Research Unit, Massachusetts General Hospital, Boston, MA; and the Department of Hematology/Oncology, Quest Diagnostics, Nichols Institute, San Juan Capistrano, CA.

Submitted March 20, 2006; accepted July 31, 2006.

Supported by Grants No. 5P01CA030195 and 5P50CA058183 (S.G.H.); National Cancer Institute Grant No. RO1-1CA112021-01 (D.C.S.); Department of Defense Grant No. W81XWH-04-1-0606 (D.C.S.); Susan G. Komen Breast Cancer Foundation Grant No. BCTR0402932 (D.C.S.); and a grant from the Avon Foundation (D.C.S.).

Presented in part at the 28th San Antonio Breast Cancer Symposium, December 8-11, 2005, San Antonio, TX.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Address reprint requests to Mark G. Erlander, PhD, 2715 Loker Ave W, Carlsbad, CA 92008; e-mail: merlander@aviaradx.com.

© 2006 by American Society of Clinical Oncology

0732-183X/06/2428-4611/\$20.00

DOI: 10.1200/JCO.2006.06.6944

(a homeo domain-containing protein), *IL17BR* (interleukin 17 receptor B) and *CHDH* (choline dehydrogenase, GenBank accession number AI240933), which were significantly associated with clinical outcome. We hypothesized that a two-gene expression index (*HOXB13:IL17BR*) might be a novel biomarker for predicting treatment outcome in tamoxifen monotherapy. Recently, Goetz et al¹⁷ analyzed *HOXB13:IL17BR* expression ratio in a carefully followed cohort of 206 postmenopausal women with ER+ breast cancer from a randomized adjuvant tamoxifen trial, and found that the two-gene index was predictive of both early relapse and death in node-negative patients, but not in node-positive patients.

To clarify the potential clinical utility of the *HOXB13:IL17BR* index, additional studies are required. Herein, we report the results of a study of 852 patients demonstrating that the two-gene index is a strong independent prognostic factor in untreated ER+ node-negative patients.

PATIENTS AND METHODS

Tumor Samples and Patient Clinical Data

The patients in this study derive from a prospectively assembled tumor bank (Tumor Bank and Data Network Core¹⁸) at the Breast Center of Baylor College of Medicine (Houston, TX). Tumor samples were archived in the form of formalin-fixed, paraffin-embedded (FFPE) tissue microarrays (12 samples/array 5 mm in diameter) as described previously.¹⁸ Of note, all samples were originally stored as fresh frozen tissues, and were fixed and arrayed relatively recently (2001). The quality of these FFPE specimens was comparable with those without the intervening snap-freeze step (data not shown). At the time of RNA extraction, tissue microarrays were approximately 4 years old. Patients were diagnosed between 1973 and 1993 with stage I or II primary breast cancer with no distant metastasis, treated with mastectomy or lumpectomy plus axillary dissection, with or without postoperative radiation therapy and with or without adjuvant tamoxifen monotherapy. From an initial 1,002 tumor specimens, 870 were selected on the basis of having more than 10% tumor content for RNA extraction. RNA from these small samples was insufficient in 18 cases, yielding 852 assessable cases (98%), whose patient and tumor characteristics are summarized in Table 1. The median follow-up for nonrelapse cases was 6.8 years. Receptor status had been determined by immunohistochemistry (IHC) as described previously.^{19,20} Allred scores of 3 to 8 were considered positive for ER-alpha or progesterone receptor (PR) expression.²¹ *HER2* amplification was determined by chromogenic in situ hybridization,²² and *HER2*-amplified cases had at least four copies/nucleus in the cancer cell. S-phase fraction was determined by flow cytometry at the time of original tissue collection. The study was approved by local institutional review boards according to National Institutes of Health (NIH; Bethesda, MD) guidelines.

Real-Time Polymerase Chain Reaction Analysis of Gene Expression

The TaqMan (Applied Biosystems, Foster City, CA) real-time polymerase chain reaction (RT-PCR) primers/probes for *HOXB13* and *IL17BR* used in this study were different from those published previously.¹⁵ A redesign of the assays was necessary to accommodate archival FFPE specimens. TaqMan primers and MGB probes were designed using Primer Express (Applied Biosystem) for nine genes (Table 2). The four reference genes (*ACTB*, *HMBS*, *SDHA*, and *UBC*) were selected by assessing 10 commonly used housekeeping genes in an independent breast cancer cohort, as described previously.²³ Total RNA was isolated from two 7- μ m tissue sections for each sample, and reverse transcribed into cDNA using a pool of gene-specific primers using the Paradise Reagent System (Arcturus BioScience, Mountain View, CA). Genes were quantitated by TaqMan RT-PCR in duplicate in a 384-well plate. The maximum cycling threshold (CT) value was set to 38. For each sample, CTs for the four reference genes (*ACTB*, *HMBS*, *SDHA*, and *UBC*) were averaged to obtain CT_{ref} . The relative expression level of each target gene was

expressed as $\Delta CT = CT_{ref} - CT_{target}$. The *HOXB13* and *IL17BR* ΔCT values were used to build a composite index by first *z*-transforming ΔCT s for each gene and then taking the difference, as described previously.¹⁵ Because of the *z*-transformation step, the resulting values were not simple ratios, and thus are referred to herein as the two-gene index.

Statistical Analysis

Determination of cut points for ER and PR mRNA levels was by one-dimensional Gaussian model-based clustering.²⁴ Spearman rank correlation was used to assess association between the *HOXB13:IL17BR* index and other prognostic factors. Cox proportional hazards regression and Kaplan-Meier analysis were used to examine the associations between gene expression indices and relapse-free survival (RFS). RFS was defined as the time from initial diagnosis to any recurrence (local, regional, or distant) of breast cancer. Patients who died as a result of other causes (ie, in the absence of a recurrence) were censored at the time of death because it is not thought that the tumor biology relates to other causes of death (ie, in the absence of a recurrence), and patients who remained disease-free were censored at last follow-up. Martingale residuals²⁵ from fitting a null Cox regression model were calculated to assess linearity and functional form for *HOXB13*, *IL17BR*, and the *HOXB13:IL17BR* index in a proportional hazards model and were found to be close to linearity. The proportional hazards assumption for all variables in Cox regression models was tested by correlating scaled Schoenfeld residuals with time²⁶ and no violation was detected. The cut point for the two-gene index in the initial discovery cohort of 60 tamoxifen-treated patients was determined by logistic regression, as previously described.¹⁵ For untreated patients, cut-point selection for the two-gene index was carried out by searching for a cut point yielding the smallest log-rank *P* value using values between the 10th and 90th percentile.²⁷ To avoid bias, the untreated patients were split into a training set and a test set, and 500 bootstrap samples from the training set were used for the cut-point search. The performance of the chosen cut point was assessed in the designated test set only. For plotting 5-year recurrence rate as a continuous function of the *HOXB13:IL17BR* index, a univariate Cox proportional hazards regression model was fitted first, which was then used to estimate the survival curve and confidence intervals using the *survfit* function in the survival package in R²⁸ (version 2.1.0; <http://www.R-project.org>). All *P* values are two sided. All statistical procedures were performed in R.

RESULTS

Study Design

In our previous study, microarray analysis of 60 tamoxifen-treated patients led to the identification of three genes, *HOXB13*, *IL17BR*, and *CHDH* (annotated as an expressed sequence tag with GenBank accession number AI240933), whose expression levels predicted clinical outcome.¹⁵ In this 852-patient cohort (Table 1), 566 (66%) were untreated, and 286 (34%) were treated with tamoxifen monotherapy. Univariate Cox proportional hazards regression analysis demonstrated that age, tumor size, lymph node status, S-phase fraction, and PR status were all significant factors for predicting relapse-free survival (Table 1), thus demonstrating that this cohort is consistent with the general breast cancer patient population with respect to known prognostic factors. Accordingly, we performed RT-PCR assays in this patient cohort for nine genes: five target genes (*HOXB13*, *IL17BR*, *CHDH*, *ER*, *PR*) and four reference genes (*ACTB*, *HMBS*, *SDHA*, and *UBC*).

Concordance of ER and PR mRNA Measurements With Immunohistochemistry

Expression profiling of FFPE samples represents a significant technical challenge because of RNA fragmentation and chemical modifications that occur during fixation and storage. We therefore

Table 1. Patient and Tumor Characteristics

Description	Tamoxifen Treated (n = 286)		Untreated (n = 566)		All (N = 852)		P Value*	Hazard Ratio*
	No.	%	No.	%	No.	%		
Age, years								
< 50	26	9	129	23	155	18	.2	
> 50	260	91	435	77	695	82		0.7
Unknown			2	0.2	2	0.2		
Nodes								
0	138	48	475	84	613	72	< .001	
1-3	79	28	55	10	134	16		1.4
4+	69	24	36	6	105	12		3.7
Tumor size, cm								
< 2	102	36	207	37	309	36	< .001	
2-5	157	55	303	54	460	54		1.2
> 5	26	9	49	9	75	9		3.1
Unknown	1	0.1	7	1	8	1		
ER (IHC)								
Negative	29	10	185	33	214	25	.11	
Positive	255	89	370	65	625	73		0.8
Unknown	2	1	11	2	13	2		
PR (IHC)								
Negative	97	34	285	50	382	45	.45	
Positive	188	66	273	48	461	54		0.8
Unknown	1	0.1	8	1	9	1		
Her2 (CISH)								
0-3	204	71	385	68	589	69	.48	
4-12	42	15	91	16	133	16		1.1
Unknown	40	14	90	16	130	15		
S-phase fraction, %								
< 6	69	24	146	26	215	25	.005	
6-10	89	31	124	22	213	25		1.7
> 10	89	31	198	35	287	34		2.0
Unknown	39	14	98	17	137	16		
Ploidy								
Diploid	95	33	186	33	281	33	.19	
Aneuploid	166	58	299	53	465	55		1.2
Unknown	25	9	81	14	106	12		
RFS, months								
Mean	68		71		70			
Range	0-231		0-285		0-285			
Recurrence status								
Nonrecurrence	203	71	383	68	586	69		
Recurrence	83	29	183	32	266	31		
Cause of death								
Cancer	69	24	137	24	206	24		
Other causes	49	17	106	19	155	18		

Abbreviations: CISH, chromogenic in situ hybridization; IHC, immunohistochemistry; ER, estrogen receptor; PR, progesterone receptor; RFS, relapse-free survival. *P values and hazard ratios are from univariate Cox proportional hazards regression models using the entire cohort. All explanatory variables were treated as categorical, omitting unknown.

assessed concordance between mRNA measurements and IHC results for ER and PR. Using a Gaussian model-based clustering technique,^{24,29} both ER and PR mRNA measurements were found to be bimodal, which was most pronounced for ER (Fig 1). Using the midpoint between the two natural clusters of ER mRNA levels as cut point (-2.5), ER status determinations between mRNA and IHC were highly concordant (91% concordance, kappa = 0.83; $P < .0001$). Using a similarly determined cut point (-5.9) for PR, mRNA and IHC measurements were again highly concordant (85%; kappa = 0.70; $P < .0001$). This level of agreement between mRNA and IHC results is similar to those reported by others.³⁰ These results confirmed the significant correlations between

mRNA and protein levels for ER and PR,³¹ and provided validation of our FFPE gene expression assay platform.

HOXB13, IL17BR, and CHDH Were Significantly Associated With Disease-Free Survival in Node-Negative Patients

Using the entire cohort, univariate Cox regression analysis indicated that gene expression levels of ER, PR, HOXB13, IL17BR, and CHDH, treated as continuous explanatory variables, were all significantly associated with RFS (Fig 2). Specifically, higher expression of HOXB13 and lower expression of IL17BR or CHDH, and a

Table 2. Real-Time Polymerase Chain Reaction Primer and Probe Design

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')	TaqMan Probe (5'-3')
<i>ACTB</i>	CTTCCTGGGCATGGAGTCC	ACGTACACATTCATGATGGAGTT	ATCCACGAAACTAC
<i>CHDH</i>	GCATCGGGAATGCTGATGA	GGCCAACCCAGGTAGGT	CAAGAACTGGGCATCC
<i>ESR1</i>	ATGATCAACTGGGCGAAGA	GGTGGACCTGATCATGGA	TGCCAGGCTTTGTGGA
<i>HMBS</i>	CTGCCACTGTGCTTCCT	TTTTCCCGCTTGCAGATG	CTGGCTTACCATCGG
<i>HOXB13</i>	TGTTGCCAGGAGAACAGAAC	CGCTGGAGTCTGCAAATGCT	ACCAGGTCCTTTTG
<i>IL17BR</i>	GAGCCGACCGTTCAATGT	AGATCATGTTGTAGCATCCACTCT	CTGAAACTGGGCCATC
<i>PGR</i>	CATTGCCAGGTTTTCGAA	CCAAGAATACTGAATGAGAGTTATC	ACTTACATATTGATGACCA
<i>SDHA</i>	CGCCGTGGTCGAGCTAGA	ACGCTGATAAATCTCCCATCTTC	CATGCCGTTTAGCAGAA
<i>UBC</i>	GATTTGGGTCGCGTTCT	TCTGCATTGTCAAGTGACGAT	TTGTGGATCGCTG

higher *HOXB13:IL17BR* index were all associated with a higher risk of relapse, in a manner similar to that in our original study.¹⁵ As positive controls, both high ER and PR mRNA levels correlated with lower risk of relapse as expected. Using the cutoff values established herein, mRNA-based ER and PR status were stronger predictors of RFS than their IHC counterparts: hazard ratio and *P* values for ER were 0.73 and .02 for mRNA versus 0.80 and .11 for IHC, and for PR were 0.68 and .0017 for mRNA versus 0.79 and .045 for IHC, respectively. This suggests that quantitative mRNA measurements may be superior to conventional IHC for determining hormone receptor status.

We next examined the association of these genes with RFS as a function of lymph node status. In subset analysis of the entire cohort, univariate Cox regression indicated that *IL17BR* and *CHDH* and the

HOXB13:IL17BR index were only significant in node-negative patients; in contrast, both ER and PR were significant factors regardless of nodal status (Fig 3). Additional analysis indicated that *HOXB13*, *IL17BR*, *CHDH*, and *HOXB13:IL17BR* were all significantly associated with RFS only in ER+ patients (data not shown). The interaction with node status was statistically significant for *IL17BR* (likelihood ratio test *P* = .0037), *CHDH* (*P* = .010), and the *HOXB13:IL17BR* index (*P* = .018) in ER+ patients.

We next investigated correlations of the *HOXB13:IL17BR* index with standard prognostic factors in ER+ patients. *HOXB13:IL17BR* correlated significantly with predictors of poor prognosis (ie, HER2 amplification, S-phase fraction, and number of positive lymph nodes) and correlated inversely with ER and PR expression

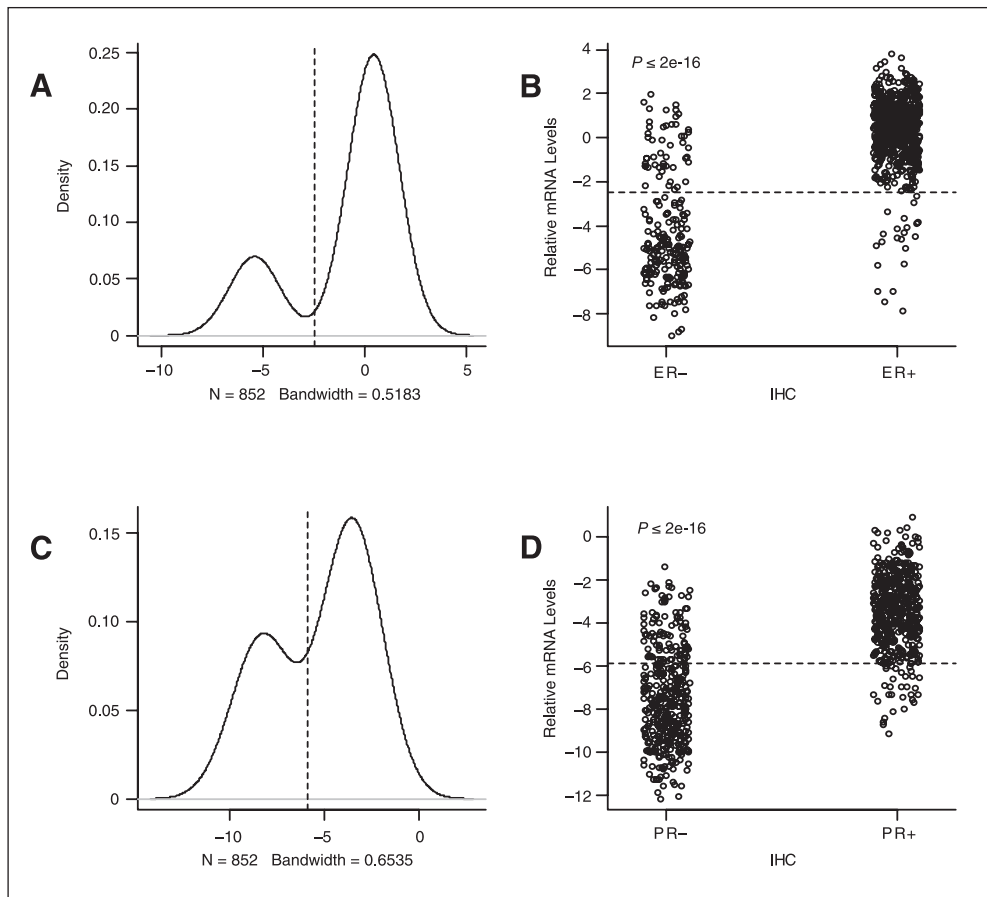


Fig 1. Concordance between estrogen receptor (ER) and progesterone receptor (PR) mRNA with immunohistochemistry (IHC). Histogram of (A) ER and (C) PR mRNA levels. Stripcharts of quantitative mRNA levels (y-axis) versus receptor status (x-axis, negative = 0; positive = 1 for (B) ER and (D) PR. *P* values are from Wilcoxon two-sample test.

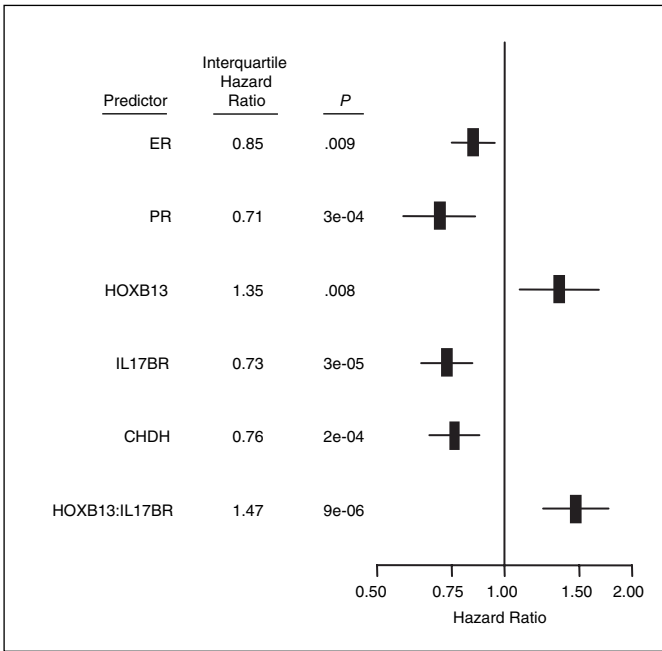


Fig 2. Forest plot of univariate Cox regression analysis of gene expression. ER, estrogen receptor; PR, progesterone receptor.

(Table 3), although the proportion of variation explained in all cases is small (<10%).

Cut-Point Selection for HOXB13:IL17BR Index

Our analysis of the combined set of untreated and tamoxifen-treated patients thus far demonstrated that the HOXB13:IL17BR index is a prognostic factor in breast cancer, particularly in ER+ node-negative patients (n = 430 in this cohort). We next wished to define an optimal cut point for stratifying untreated patients (n = 308) into low- and high-risk groups. To allow an unbiased estimate of performance, we randomly partitioned the untreated ER+ node-negative group into a training set (two thirds, n = 205) and a test set (one third, n = 103). The training set was used to generate 500

Table 3. Spearman Rank Correlation Between HOXB13:IL17BR and Standard Prognostic Factors in ER+ Patients

	No.	r _{sp}	P
Age	625	-0.06	.17
ER (mRNA)	625	-0.23	< .0001
PR (mRNA)	625	-0.26	< .0001
HER2 (CISH)	546	0.17	< .0001
Nodes	625	0.13	.0015
S-phase	521	0.15	.0004
Tumor size	619	0.06	.15

Abbreviations: r_{sp}, Spearman rank correlation; CISH, chromogenic in situ hybridization; ER, estrogen receptor; PR, progesterone receptor.

bootstrap data sets. In each bootstrap sample, we searched for a cut point yielding minimal log-rank P value by dichotomizing the HOXB13:IL17BR index using potential cut points between the 10th and 90th percentile of the index values. The distribution of the resulting cut points indicated that the selected cut points were strongly nonrandom and clustered around the median value of about 1.0, which was at approximately the 75th quantile (Fig 4). We then applied the 1.0 cut-point value to the reserved test set from the untreated group and also the tamoxifen-treated group (n = 122). In both independent test sets, Kaplan-Meier curves and univariate Cox regression analysis indicated that this cut point stratified patients into significantly different risk groups (Fig 5). Comparing the two Kaplan-Meier plots suggests that the prognostic power of the two-gene index was independent of tamoxifen therapy.

Multivariate Analysis

To demonstrate that the HOXB13:IL17BR index provides prognostic information beyond standard clinical and molecular factors, we performed multivariate Cox regression analysis incorporating the two-gene index dichotomized at the 1.0 cut point. To allow an unbiased estimate of the performance of the two-gene index, we combined the designated test set from the untreated group (Fig 4) and from the tamoxifen-treated group into a true validation set (n = 225). In this data set of ER+ node-negative patients, the multivariate Cox

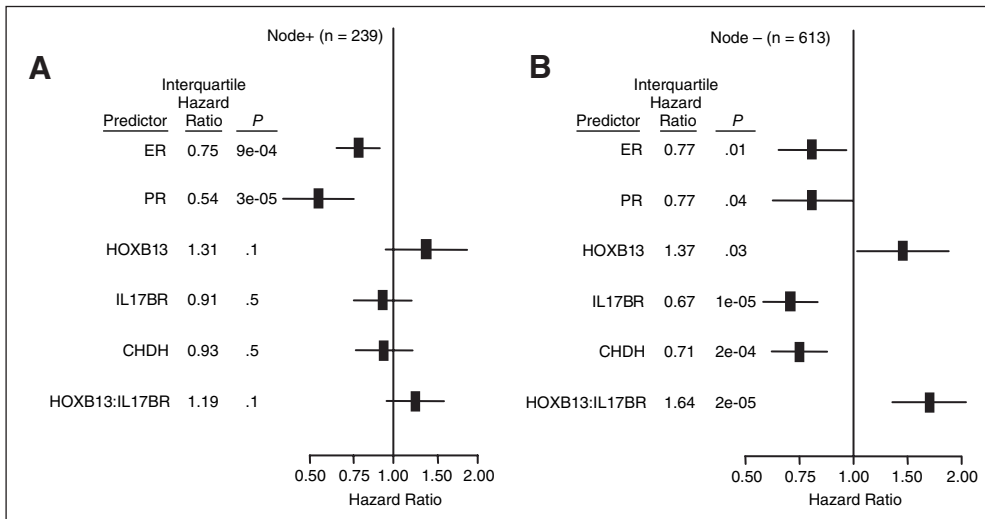


Fig 3. Forest plot of univariate Cox regression analysis of gene expression for (A) node-negative and (B) node-positive patients. Node+, node positive; Node-, node negative; ER, estrogen receptor; PR, progesterone receptor.

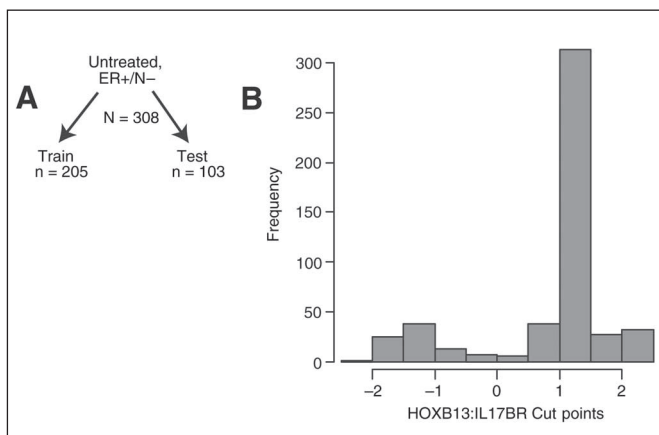


Fig 4. Cut-point selection for the *HOXB13:IL17BR* index. (A) Schematic of data partitions. (B) Distribution of cut points from 500 bootstrap samples from the training set. ER, estrogen receptor; N-, node-negative.

regression model including age, tumor size, S-phase fraction, PR status, and tamoxifen therapy (Table 4), the two-gene index remained a highly significant factor for predicting RFS with a hazard ratio of 3.9 (95% CI, 1.5 to 10.3; $P = .007$).

***HOXB13:IL17BR* As a Continuous Predictor of Prognosis**

Although a single cut point may be useful for patient stratification, Martingale residual analysis²⁵ indicated that two-gene index predicted risk of recurrence on a continuous scale. To demonstrate the prognostic value of the two-gene index, we used the untreated group of ER+ node-negative patients ($n = 308$) to estimate the 5-year recurrence rate as a function of the two-gene index (Fig 6). On the continuous scale, an untreated patient with a two-gene index of -2.0 has a 5-year recurrence risk of 15% (95% CI, 9.8% to 20.5%), whereas a patient with an index of $+2.0$ has a significantly higher 5-year recurrence risk of 36% (95% CI, 26.5% to 45.2%).

Validation of *HOXB13:IL17BR* Index in Tamoxifen-Treated Patients

We have thus far demonstrated a prognostic role for the two-gene index irrespective of tamoxifen therapy. We next examined the tamoxifen-treated subgroup separately. First, we reanalyzed the initial tamoxifen-treated 60-patient cohort ($n = 59$, material for one case unavailable) with the RT-PCR assays used in this study. Using this data set, we derived an optimal cut point of 0.06 for the two-gene index separating the recurrence cases from the nonrecurrence cases (Fig 7). Applying this cut point to the tamoxifen-treated ER+ node-negative subgroup ($n = 122$), Kaplan-Meier curves of the resulting patient stratification demonstrated significantly different RFS (Fig 7). In fact, the cut point of 0.06 performed slightly better for the treated group than did the 1.0 cut point derived from untreated patients (Fig 5). In contrast, this cut point had no predictive value in the treated ER+ node-positive subgroup ($n = 133$, Fig 7). Therefore, these results confirmed previous work demonstrating that the two-gene index was a significant predictor of clinical outcome in ER+ node-negative tamoxifen-treated patients.¹⁷

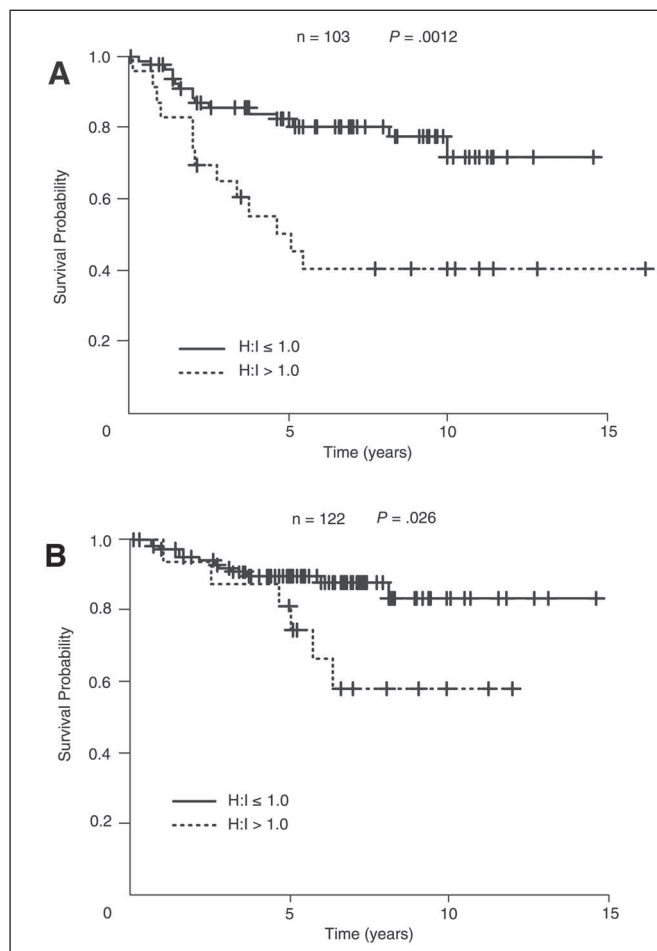


Fig 5. Validation of *HOXB13:IL17BR* cut point in two test sets of estrogen receptor-positive node-negative patients. Kaplan-Meier plots of the (A) untreated test set and (B) tamoxifen-treated cohort.

DISCUSSION

We have demonstrated previously that the *HOXB13*, *IL17BR*, and *CHDH* genes, particularly the *HOXB13:IL17BR* index, predict distant metastasis in tamoxifen-treated patients with breast cancer.¹⁵ However, in that study it was not possible to address the question whether these genes are prognostic factors for a tumor's natural history or predictors of tamoxifen response, or both. In this cohort consisting of both untreated and tamoxifen-treated patients, we show that these genes predicted relapse in untreated patients as well. However, a direct comparison of the untreated and tamoxifen-treated patients in this cohort was difficult because of both the limited sample sizes and the fact the patients were not randomly assigned for treatment. It thus remains to be determined whether the two-gene index is also a predictive factor for tamoxifen response, as suggested by a study of first-line tamoxifen therapy in metastatic breast cancer.³² Nevertheless, a reanalysis of our initial tamoxifen-treated cohort with the RT-PCR assays used in this study resulted in a different cut point (0.06), which performed well in the independent tamoxifen-treated ER+ node-negative patients. The existence of different cut points for untreated and tamoxifen-treated patients warrants further studies.

Table 4. Multivariate Cox Proportional Hazards Regression Analysis

Predictor	ER+ Node Negative (untreated test set and tamoxifen treated, n = 225)		
	Hazard Ratio	95% CI	P
Age, years			
> 50 v < 50	0.8	0.3 to 2.0	.62
Tumor size, cm			.71
2-5 v < 2	1.2	0.5 to 2.5	
> 5 v < 2	2.4	0.3 to 21.7	
S to phase fraction, %			.74
6-10 v < 6	0.8	0.3 to 1.9	
> 10 v < 6	0.7	0.3 to 1.9	
PR (IHC)			
Positive v negative	1.4	0.6 to 3.5	.48
Tamoxifen treatment			
Treated v untreated	0.6	0.3 to 1.3	.20
HOXB13:IL17BR index			
High v low	3.9	1.5 to 10.3	.007

Abbreviations: IHC, immunohistochemistry; ER, estrogen receptor; PR, progesterone receptor.

A surprising feature of the *HOXB13:IL17BR* index is that it is a much better predictor in lymph node-negative patients than in lymph node-positive patients, both in this and a previous study.¹⁷ Consistent with these results, Reid et al³³ failed to demonstrate a predictive value for the two-gene index in a mostly node-positive cohort (n = 58). The mechanism for this index-nodal status interaction is unclear; however, we note that tumors from node-positive patients tend to have a higher *HOXB13:IL17BR* index.

At present, three other expression-based prognostic signatures for breast cancer have been published.^{9,14,34} Perhaps surprisingly, these gene sets, including our two-gene index, are largely unique

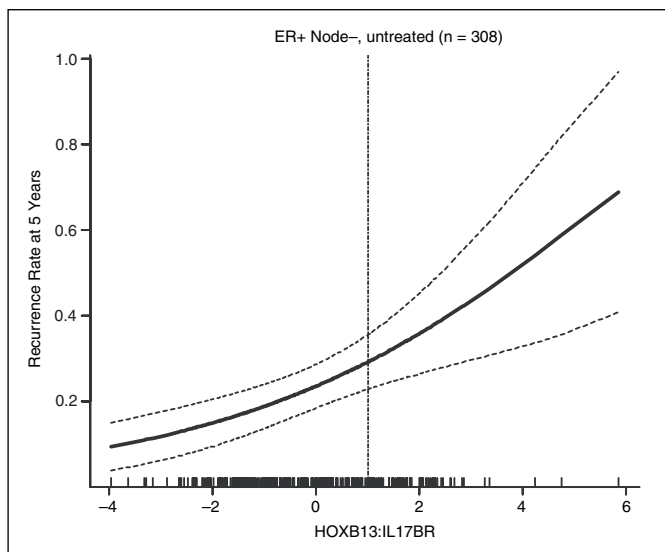


Fig 6. The *HOXB13:IL17BR* index as a continuous predictor of recurrence at 5 years in untreated estrogen receptor-positive (ER+) node-negative (Node-) patients. Vertical line indicates optimized cut point (1.0). The rug plot along the x-axis shows the *HOXB13:IL17BR* index values. Solid line shows the estimated recurrence rate and dashed lines show the 95% CI.

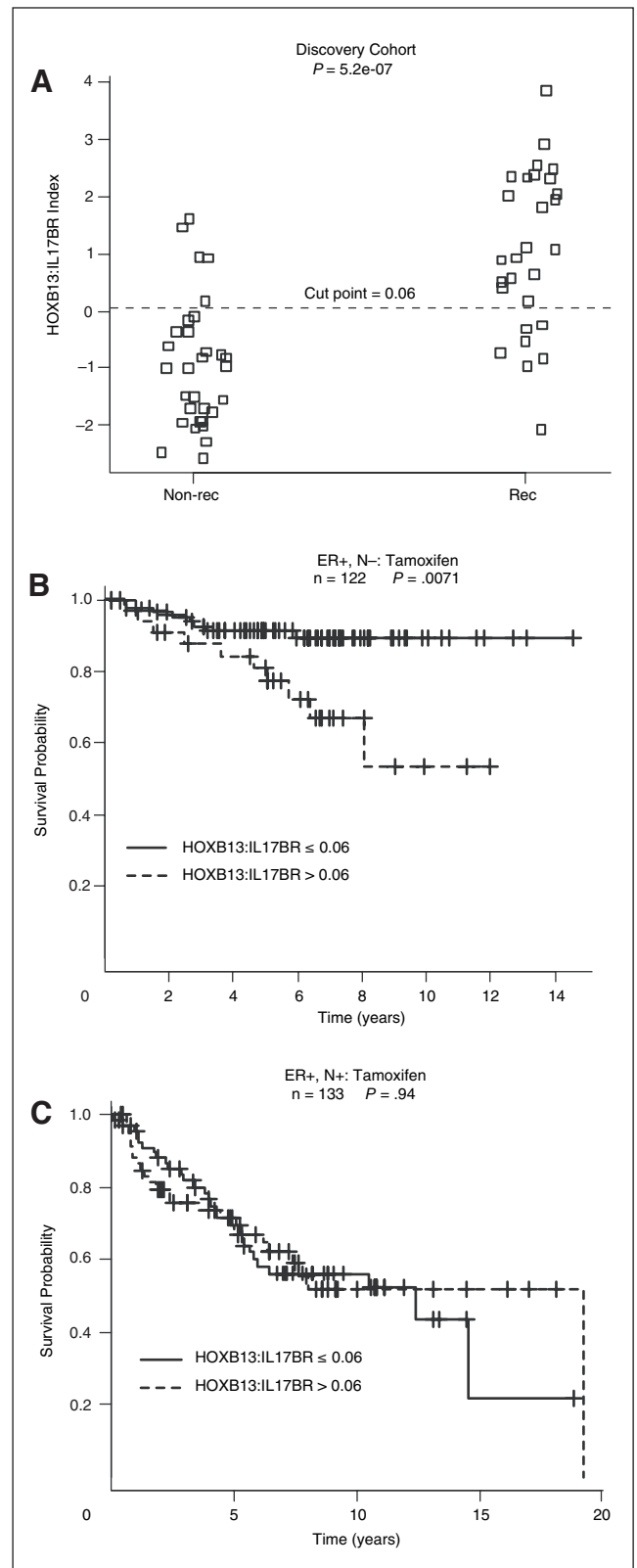


Fig 7. *HOXB13:IL17BR* cut point in tamoxifen-treated patients. (A) Determination of optimal cut point in the initial discovery cohort of tamoxifen-treated patients (n = 59). The dashed line indicates the 0.06 cut point derived from logistic regression. This cut point was then applied to the estrogen receptor-positive (ER+) tamoxifen-treated subset in the current cohort for Kaplan-Meier analysis. (B) Node-negative patients. (C) Node-positive patients. Nonrec, nonrecurrence cases; rec, recurrence cases.

from one another. However, it should be noted that these gene sets were derived using different platforms. For example, the Affymetrix GeneChip U133A microarray detects little signal from *HOXB13* (unpublished data), and *IL17BR* mRNA has multiple variants,³⁵ making it difficult to compare results across platforms. Technical differences notwithstanding, an important question remains to be addressed: do these gene signatures provide the same or unique prognostic information? The prognostic utility for *HOXB13* is supported by evidence indicating that its overexpression promotes tumor growth and invasion in multiple tumors.^{15,36-39}

As part of our study, we demonstrated that determining hormonal receptor status by mRNA expression from FFPE tissues provided excellent concordance with IHC, as reported by others.³⁰ However, the concordance between mRNA expression and IHC for

PR is less than that for ER (91% v 85%; $P < .001$), and mRNA-derived receptor status is more strongly associated with clinical outcome, suggesting that mRNA quantitation by RT-PCR may be a more reliable method for assessing receptor status.

In summary, two principal findings have emerged from this study of 852 breast cancer patients. First, we have extended our previous finding that *HOXB13*, *IL17BR*, and *CHDH* are predictive of RFS in tamoxifen-treated patients to untreated patients, suggesting a prognostic role. Second, the two-gene index performs the best in ER+ node-negative patients. If corroborated in further studies, the two-gene index may provide a new prognostic biomarker for identifying a subset of high-risk ER+ node-negative patients for alternative treatment strategies (eg, *EGFR* inhibitors or chemotherapies with or without tamoxifen).

REFERENCES

- Carlson RW, McCormick B: Update: NCCN Breast Cancer Clinical Practice Guidelines. *J Natl Compr Canc Netw* 3:S7-S11, 2005
- Goldhirsch A, Glick JH, Gelber RD, et al: Meeting highlights: International expert consensus on the primary therapy of early breast cancer 2005. *Ann Oncol* 16:1569-1583, 2005
- Early Breast Cancer Trialists Collaborative Group: Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: An overview of the randomised trials. *Lancet* 365:1687-1717, 2005
- Bardou VJ, Arpino G, Elledge RM, et al: Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. *J Clin Oncol* 21:1973-1979, 2003
- Ellis MJ, Coop A, Singh B, et al: Letrozole is more effective neoadjuvant endocrine therapy than tamoxifen for ErbB-1- and/or ErbB-2-positive, estrogen receptor-positive primary breast cancer: Evidence from a phase III randomized trial. *J Clin Oncol* 19:3808-3816, 2001
- Dowsett M, Harper-Wynne C, Boeddinghaus I, et al: HER-2 amplification impedes the antiproliferative effects of hormone therapy in estrogen receptor-positive primary breast cancer. *Cancer Res* 61:8452-8458, 2001
- Osborne CK, Bardou V, Hopp TA, et al: Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. *J Natl Cancer Inst* 95:353-361, 2003
- van de Vijver MJ, He YD, van't Veer LJ, et al: A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347:1999-2009, 2002
- Wang Y, Klijn JG, Zhang Y, et al: Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365:671-679, 2005
- Chang JC, Wooten EC, Tsimelzon A, et al: Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* 362:362-369, 2003
- Huang E, Cheng SH, Dressman H, et al: Gene expression predictors of breast cancer outcomes. *Lancet* 361:1590-1596, 2003
- Sorlie T, Perou CM, Tibshirani R, et al: Gene expression patterns of breast carcinomas distinguish tumor subtypes with clinical implications. *Proc Natl Acad Sci U S A* 98:10869-10874, 2001
- Sotiriou C, Neo SY, McShane LM, et al: Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* 100:10393-10398, 2003
- 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* 351:2817-2826, 2004
- Ma XJ, Wang Z, Ryan PD, et al: A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 5:607-616, 2004
- Jansen MP, Foekens JA, van Staveren IL, et al: Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. *J Clin Oncol* 23:732-740, 2005
- Goetz M, Suman V, Ingle JN, et al: A two-gene expression ratio of *HOXB13* and *IL-17BR* for prediction of recurrence and survival in women receiving adjuvant tamoxifen. *Clinical Cancer Res* 12:2080-2087, 2006
- Martin MD, Hilsenbeck SG, Mohsin SK, et al: Breast tumors that overexpress nuclear metastasis-associated 1 (MTA1) protein have high recurrence risks but enhanced responses to systemic therapies. *Breast Cancer Res Treat* 95:7-12, 2006
- Harvey JM, Clark GM, Osborne CK, et al: Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol* 17:1474-1481, 1999
- Mohsin SK, Weiss H, Havighurst T, et al: Progesterone receptor by immunohistochemistry and clinical outcome in breast cancer: A validation study. *Mod Pathol* 17:1545-1554, 2004
- Allred DC, Harvey JM, Berardo M, et al: Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol* 11:155-168, 1998
- Tanner M, Gancberg D, Di Leo A, et al: Chromogenic in situ hybridization: A practical alternative for fluorescence in situ hybridization to detect HER-2/neu oncogene amplification in archival breast cancer samples. *Am J Pathol* 157:1467-1472, 2000
- Vandesompele J, De Preter K, Pattyn F, et al: Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3:RESEARCH0034, 2002
- Fraley C, Raftery AE: MCLUST: Software for model-based cluster analysis. *J Class* 16:297-306, 1999
- Therneau TM, Grambsch PM, Fleming TR: Martingale-based residuals for survival analysis. *Biometrika* 77:216-218, 1990
- Pettitt AN, Bin Daud I: Investigating time dependence in Cox's proportional hazards model. *Appl Stat* 39:205-217, 1990
- Faraggi D, Simon R: A simulation study of cross-validation for selecting an optimal cutpoint in univariate survival analysis. *Stat Med* 15:2203-2213, 1996
- R Development Core Team: R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing, 2005
- Yeung KY, Fraley C, Murua A, et al: Model-based clustering and data transformations for gene expression data. *Bioinformatics* 17:977-987, 2001
- Cronin M, Pho M, Dutta D, et al: Measurement of gene expression in archival paraffin-embedded tissues: Development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *Am J Pathol* 164:35-42, 2004
- Tong D, Schneeberger C, Czerwenka K, et al: Messenger RNA determination of estrogen receptor, progesterone receptor, pS2, and plasminogen activator inhibitor-1 by competitive reverse transcription-polymerase chain reaction in human breast cancer. *Clin Cancer Res* 5:1497-1502, 1999
- Jansen MP, Foekens JA, Klijn JG, et al: Re: Limits of predictive models using microarray data for breast cancer clinical treatment outcome. *J Natl Cancer Inst* 97:1851-1853, 2005
- Reid JF, Lusa L, De Cecco L, et al: Limits of predictive models using microarray data for breast cancer clinical treatment outcome. *J Natl Cancer Inst* 97:927-930, 2005
- van't Veer LJ, Dai H, van de Vijver MJ, et al: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415:530-536, 2002
- Tian E, Sawyer JR, Largaespada DA, et al: Evi27 encodes a novel membrane protein with homology to the IL17 receptor. *Oncogene* 19:2098-2109, 2000
- Chen H, Sukumar S: HOX genes: Emerging stars in cancer. *Cancer Biol Ther* 2:524-525, 2003
- Svingen T, Tonissen KF: Altered HOX gene expression in human skin and breast cancer cells. *Cancer Biol Ther* 4:518-523, 2003
- Maeda K, Hamada J, Takahashi Y, et al: Altered expressions of HOX genes in human cutaneous malignant melanoma. *Int J Cancer* 114:436-441, 2005
- Yamashita T, Tazawa S, Yawei Z, et al: Suppression of invasive characteristics by antisense introduction of overexpressed HOX genes in ovarian cancer cells. *Int J Oncol* 28:931-938, 2006

Acknowledgment

We thank JoAnn Kop, Ranelle Salunga, Rajiv Raja, Jaji Murage, Evelyn Cheung, Rajesh Patel, and Walter Lee for excellent technical assistance.

Authors' Disclosures of Potential Conflicts of Interest

Although all authors completed the disclosure declaration, the following authors or their immediate family members indicated a financial interest. No conflict exists for drugs or devices used in a study if they are not being evaluated as part of the investigation. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.

Authors	Employment	Leadership	Consultant	Stock	Honoraria	Research Funds	Testimony	Other
Xiao-Jun Ma	AviaraDx (N/R)	AviaraDx (B)		AviaraDx (B)				
Xianqun Wilson Wang	AviaraDx (N/R)			AviaraDx (B)				
Li Ding	AviaraDx (N/R)			AviaraDx (A)				
Mark G. Erlander	AviaraDx (N/R)	AviaraDx (B)		AviaraDx (B)				
Dollar Amount Codes (A) < \$10,000 (B) \$10,000-99,999 (C) ≥ \$100,000 (N/R) Not Required								

Author Contributions

Conception and design: Xiao-Jun Ma, Susan G. Hilsenbeck, Dennis C. Sgroi, C. Kent Osborne, D. Craig Allred, Mark G. Erlander
Provision of study materials or patients: C. Kent Osborne, D. Craig Allred
Collection and assembly of data: Susan G. Hilsenbeck, Li Ding, C. Kent Osborne, D. Craig Allred
Data analysis and interpretation: Xiao-Jun Ma, Susan G. Hilsenbeck, Wilson Wang, D. Craig Allred, Mark G. Erlander
Manuscript writing: Xiao-Jun Ma, Susan G. Hilsenbeck, Wilson Wang, Richard A. Bender, D. Craig Allred, Mark G. Erlander
Final approval of manuscript: Xiao-Jun Ma, Susan G. Hilsenbeck, Wilson Wang, Li Ding, Dennis C. Sgroi, Richard A. Bender, C. Kent Osborne, D. Craig Allred, Mark G. Erlander