A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen

Xiao-Jun Ma,[1] Zuncai Wang,[2] Paula D. Ryan,[3] Steven J. Isakoff,[4,5] Anne Barmettler,[2] Andrew Fuller,[2] Beth Muir,[2] Gayatry Mohapatra,[2] Ranelle Salunga,[1] J. Todd Tuggle,[1] Yen Tran,[1] Diem Tran,[1] Ana Tassin,[1] Paul Amon,[1] Wilson Wang,[1] Wei Wang,[1] Edward Enright,[1] Kimberly Stecker,[1] Eden Estepa-Sabal,[1] Barbara Smith,[3] Jerry Younger,[3] Ulysses Balis,[2] James Michaelson,[2] Atul Bhan,[2] Karleen Habin,[3] Thomas M. Baer,[1] Joan Brugge,[4] Daniel A. Haber,[3] Mark G. Erlander,[1] and Dennis C. Sgroi[2]

[1]Arcturus Bioscience, Inc., 2715 Loker Avenue West, Carlsbad, California 92008
[2]Department of Pathology, Harvard Medical School, Molecular Pathology Research Unit, Massachusetts General Hospital, Boston, Massachusetts 02129
[3]Massachusetts General Hospital Cancer Center, Harvard Medical School, Boston, Massachusetts 02129
[4]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115
[5]Department of Adult Oncology, Dana Farber Cancer Institute, Boston, Massachusetts 02115
*Correspondence: dsgroi@partners.org (D.C.S.), merlander@arctur.com (M.G.E.)

Summary

Tamoxifen significantly reduces tumor recurrence in certain patients with early-stage estrogen receptor-positive breast cancer, but markers predictive of treatment failure have not been identified. Here, we generated gene expression profiles of hormone receptor-positive primary breast cancers in a set of 60 patients treated with adjuvant tamoxifen monotherapy. An expression signature predictive of disease-free survival was reduced to a two-gene ratio, HOXB13 versus IL17BR, which outperformed existing biomarkers. Ectopic expression of HOXB13 in MCF10A breast epithelial cells enhances motility and invasion in vitro, and its expression is increased in both preinvasive and invasive primary breast cancer. The HOXB13:IL17BR expression ratio may be useful for identifying patients appropriate for alternative therapeutic regimens in early-stage breast cancer.

Introduction

Tamoxifen is the antiestrogen agent most frequently prescribed in women with both early-stage and metastatic hormone receptor-positive breast cancer (reviewed in Clarke et al., 2003; Jordan, 2002). In the adjuvant setting, tamoxifen therapy results in a 40%–50% reduction in the annual risk of recurrence, leading to a 5.6% improvement in 10-year survival in lymph node-negative patients, and a corresponding 10.9% improvement in node-positive patients (EBCTCG, 2001). Tamoxifen is thought to act primarily as a competitive inhibitor of estrogen binding to estrogen receptor (ER), and it has no effect on tumor cells lacking ER expression. The absolute levels of ER expression, as well as those of the progesterone receptor (PR, an indicator of a functional ER pathway), are currently the best predictors of tamoxifen response in the clinical setting (Bardou et al., 2003; EBCTCG, 2001). However, 25% of ER+/PR+ tumors, 66% of ER+/PR− cases, and 55% of ER−/PR+ cases fail to respond, or develop early resistance to tamoxifen, through mechanisms that remain largely unclear (Clarke et al., 2003; Nicholson et al., 2003; Osborne and Schiff, 2003). Currently, no reliable means exist that allow the identification of patients who are at risk of tumor recurrence in the setting of adjuvant tamoxifen monotherapy. In such patients, the use of alternative hormonal therapies, such as the aromatase inhibitors letrozole and anastrozole (Buzdar, 1998; Ellis et al., 2001; Goss et al., 2003), chemotherapeutic agents (Colleoni et al., 2003a), or inhibitors of other signaling pathways, such as trastuzmab and gefitinib (Atalay et al., 2003; Lichtner, 2003; Morris, 2002; Nicholson et al., 2003; Osborne and Schiff, 2003), might improve clinical outcome. Therefore, the ability to accurately predict tamoxifen treatment outcome should significantly advance the management of early-stage breast cancer.

SIGNIFICANCE

The majority of breast cancers express estrogen receptor (ER), and most patients with ER-positive tumors are treated with adjuvant tamoxifen therapy. Approximately 40% of ER+ breast cancers fail to respond or eventually develop resistance to tamoxifen, leading to disease progression. Current clinicopathological features including tumor stage and grade, and ERBB2 and EGFR expression fail to accurately identify individuals who are at risk for tumor recurrence. Our results demonstrate that a simple two-gene expression ratio of HOXB13 to IL17BR accurately predicts tumor recurrence in the setting of adjuvant tamoxifen monotherapy. Furthermore, the association of HOXB13 overexpression with poor clinical outcome and with enhanced cell invasiveness in vitro raises the possibility that it may play a role in breast cancer progression.

Table 1. Patient and tumor characteristics in this study

		Cohort 1 (frozen)		Cohort 2 (FFPE)	
		Recurrence	Nonrecurrence	Recurrence	Nonrecurrence
Size (cm)	Total	28	32	10	10
	Mean	2.7	2.1	1.9	1.7
	Range	0.9–4.7	0.8–5.5	1.1–4.0	0.8–4.0
Grade	1	2	1	1	1
	2	15	24	6	8
	3	11	7	3	1
Nodes*	0	13	15	8	10
	1–3	6	11	1	0
	>3	6	2	0	0
Age	Mean	65.1	69.1	65.5	65.2
	Range	48–84	54–85	54–93	57–74
DFS (Months)	Mean	54.8	115.6	51.4	95.8
	Range	5–137	61–169	15–117	25–123
Receptor status	ER+	27/28	32/32	10/10	10/10
	PR+	23/28	27/32	8/9*	10/10

*Cases with missing data omitted. DFS, disease-free survival; FFPE, formalin-fixed and paraffin-embedded.

Here, we performed a genome-wide microarray analysis of hormone receptor-positive invasive breast tumors from 60 patients treated with adjuvant tamoxifen alone, leading to the identification of a two-gene expression ratio that is highly predictive of clinical outcome. This expression ratio can be readily adapted to PCR-based analysis of standard formalin-fixed paraffin-embedded (FFPE) clinical specimens, which we demonstrated in an independent set of 20 patients with FFPE tissues. Further, we demonstrate that the homeobox gene HOXB13 (Zeltser et al., 1996), whose overexpression is associated with a poor outcome, confers invasive properties following ectopic expression in a nontransformed mammary epithelial cell line. Our results provide a simple molecular marker of clinical outcome in ER-positive early-stage breast cancer and may provide insight into molecular mechanisms underlying tamoxifen resistance.

Results

Identification of differentially expressed genes

To identify gene expression patterns in hormone receptor-positive, early-stage invasive breast cancers that might predict response to hormonal therapy, we performed microarray gene expression analysis of tumors from 60 women uniformly treated with adjuvant tamoxifen alone. These patients were identified from a total of 103 ER-positive early-stage cases presenting to Massachusetts General Hospital between 1987 and 1997, from whom tumor specimens were snap-frozen and for whom minimal 5-year follow-up was available (see Table 1 and Supplemental Table S1 at http://www.cancercell.org/cgi/content/full/5/6/607/DC1 for details). Within this cohort, 28 (46%) women developed distant metastasis with a median time to recurrence of 4 years ("tamoxifen recurrences") and 32 (54%) women remained disease-free with median follow-up of 10 years ("tamoxifen nonrecurrences"). Recurrence cases were matched with nonrecurrence cases with respect to TNM staging (Singletary et al., 2002) and tumor grade (Dalton et al., 2000).

Gene expression profiling was performed using a 22,000-gene oligonucleotide microarray. In the initial analysis, we isolated RNA from frozen tumor-tissue sections taken from the archived primary biopsies. The resulting expression dataset was first filtered based on overall variance of each gene, with the top 5,475 high-variance genes (75th percentile) selected for further analysis. Using this reduced dataset, a t test was performed on each gene comparing tamoxifen nonrecurrences and recurrences, leading to the identification of 19 differentially expressed genes at the p value cutoff of 0.001 (Figure 1A; Supplemental Table S2). The probability of selecting this many or more differentially expressed genes by chance was estimated to be 0.04 by randomly permuting the patient class with respect to treatment outcome and repeating the t test procedure 2,000 times. This analysis thus demonstrated the existence of statistically significant differences in gene expression between the primary breast cancers of tamoxifen nonrecurrences and recurrences.

To refine our analysis to the tumor cells and circumvent potential variability due to stromal cell contamination, we reanalyzed the same cohort following laser-capture microdissection (LCM) of tumor cells within each tissue section (Bonner et al., 1997; Emmert-Buck et al., 1996; Ma et al., 2003). Using variance-based gene filtering and t test screening identical to that utilized for the whole tissue section dataset, we identified 9 differentially expressed genes with $p < 0.001$ (Figure 1B; Supplemental Table S3). Three genes were identified as differentially expressed in both the LCM and whole tissue section analyses: the homeobox gene HOXB13 (identified twice as AI700363 and BC007092), the interleukin 17B receptor IL17BR (AF208111), and EST AI240933 (Figure 1). HOXB13 was overexpressed in tamoxifen recurrence cases, whereas IL17BR and AI240933 were overexpressed in tamoxifen nonrecurrence cases.

Based on their identification as tumor-derived markers significantly associated with clinical outcome in two independent analyses, we evaluated the prognostic utility of each of these three genes by itself and in combination with the others. To define the sensitivity and specificity of HOXB13, IL17BR, and AI240933 expression as biomarkers of clinical outcome, we used Receiver Operating Characteristic (ROC) analysis (Pepe, 2000). For data derived from whole tissue sections, the Area Under the Curve (AUC) values for IL17BR, HOXB13, and AI240933 were 0.79, 0.67, and 0.81, respectively (Table 2; Supplemental Figure S1). ROC analysis of the data generated from

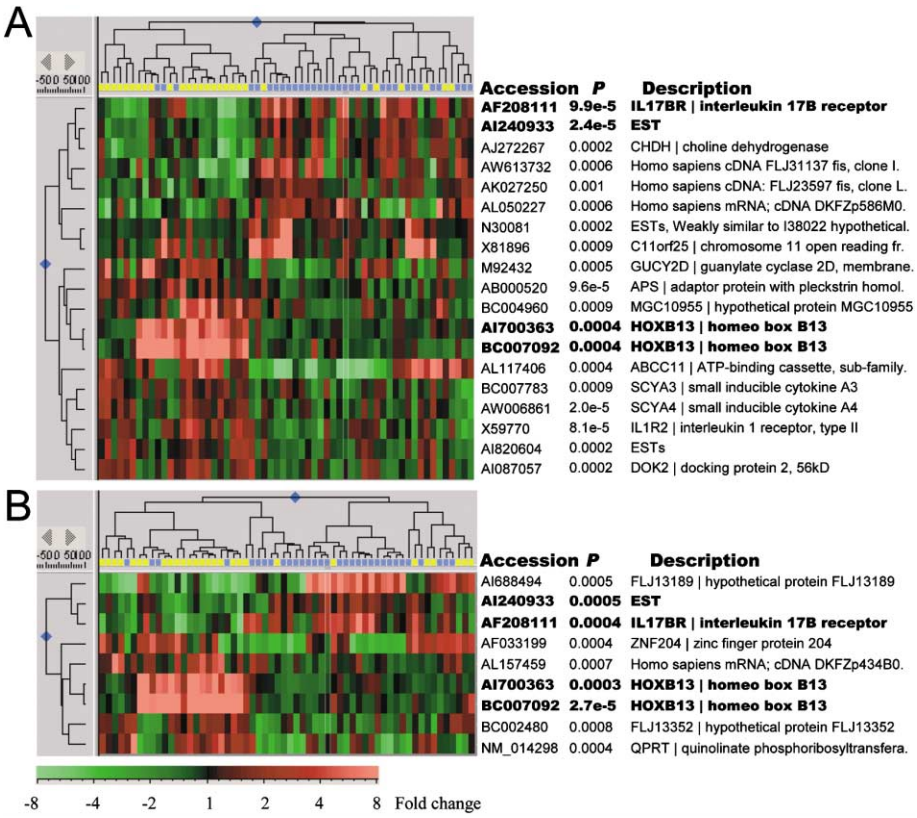

Figure 1. Heatmap of differential gene expression between tamoxifen responders and nonresponders

A: 19 genes identified in the tissue sections dataset.

B: 9 genes identified in the LCM dataset. Genes identified in bold are present in both lists. The row of yellow and blue bars below each dendrogram indicates recurrences and nonrecurrences, respectively. Bottom, color scale bar. p values are from a two-sample t test and unadjusted for multiple testing. See Supplemental Tables S1 and S2 for details.

the microdissected tumor cells yielded AUC values of 0.76, 0.8, and 0.76 for these genes (Table 2; Supplemental Figure S1). A statistical test of significance indicated that these AUC values are all significantly greater than 0.5, the expected value from the null model that predicts clinical outcome randomly (Table 2). Therefore, these three genes have potential utility for predicting clinical outcome of patients treated with adjuvant tamoxifen.

As comparison, we analyzed markers that are currently useful in evaluating the likelihood of response to tamoxifen. All but one of the 60 cases were ER+ according to clinical tests performed at diagnosis (Supplemental Table S1), but this binary receptor status is not informative for potential differences in response to tamoxifen when essentially all patients are ER+. The quantitative levels of ER (gene symbol ESR1) and progesterone receptor (PR, gene symbol PGR) are known to be positively correlated with tamoxifen response (Bardou et al., 2003; Fernandez et al., 1983; Ferno et al., 2000; Nardelli et al., 1986; Osborne et al., 1980). Therefore, mRNA measurements for ER and PR based on the microarray datasets were used to assess their predictive capacity for tamoxifen treatment outcome. In addition, growth factor signaling pathways (EGFR, ERBB2) are thought to negatively regulate estrogen-dependent signaling, and hence contribute to loss of responsiveness to tamoxifen (Dowsett, 2001; Lichtner, 2003; Nicholson et al., 2003; Osborne and Schiff, 2003). ROC analysis of the gene expression levels of these genes confirmed their correlation with clinical outcome, but with AUC values ranging only from 0.55 to 0.69, reaching statistical significance for PGR and ERBB2 (Table 2). The LCM dataset is particularly relevant, since EGFR, ERBB2, ESR1, and PGR are currently measured at the tumor cell level using either immunohistochemistry or fluorescence in situ hybridization. As individual markers of clinical outcome, HOXB13, IL17BR, and AI240933 outperformed ESR1, PGR, EGFR, and ERBB2 (Table 2).

Table 2. ROC analysis of using IL17BR, AI240933, and HOXB13 expression to predict tamoxifen treatment outcome

	Tissue sections		LCM	
	AUC	p value	AUC	p value
IL17BR	0.79	1.58E-06	0.76	2.73E-05
AI240933	0.81	3.02E-08	0.76	1.59E-05
HOXB13	0.67	0.012	0.79	9.94E-07
ESR1	0.55	0.277	0.63	0.038
PGR	0.63	0.036	0.63	0.033
ERBB2	0.69	0.004	0.64	0.027
EGFR	0.56	0.200	0.61	0.068

AUC, area under the curve; p values are AUC > 0.5.

HOXB13:IL17BR expression ratio is a robust composite predictor of outcome

Since HOXB13 and IL17BR have opposing patterns of expression (Figure 1), we examined whether the expression ratio of HOXB13 over IL17BR provides a better composite predictor of tamoxifen response. Indeed, both t test and ROC analyses demonstrated that the two-gene ratio had a stronger correlation with treatment outcome than either gene alone, both in the whole tissue sections and LCM datasets (Table 3). AUC values for HOXB13:IL17BR reached 0.81 for the tissue sections dataset and 0.84 for the LCM dataset. Pairing HOXB13 with AI240933 or

Table 3. HOXB13:IL17BR ratio is a stronger predictor of treatment outcome

		t test		ROC	
		t statistic	p value	AUC	p value
Tissue section	IL17BR	4.15	1.15E-04	0.79	1.58E-06
	HOXB13	−3.57	1.03E-03	0.67	0.01
	HOXB13:IL17BR	−4.91	1.48E-05	0.81	1.08E-07
LCM	IL17BR	3.70	5.44E-04	0.76	2.73E-05
	HOXB13	−4.39	8.00E-05	0.79	9.94E-07
	HOXB13:IL17BR	−5.42	2.47E-06	0.84	4.40E-11

AUC, area under the curve; p values are AUC > 0.5.

analysis of all three markers together did not provide additional power (data not shown).

We compared the HOXB13:IL17BR ratio to well-established prognostic factors for breast cancer, such as patient age, tumor size, grade, and lymph node status (Fitzgibbons et al., 2000). Univariate logistic regression analysis indicated that only tumor size was marginally significant in this cohort ($p = 0.04$); this was not surprising given that the nonrecurrence group was closely matched to the recurrence group with respect to tumor size, tumor grade, and lymph node status during patient selection. Among the known positive (ESR1 and PGR) and negative (ERBB2 and EGFR) predictors of tamoxifen response, ROC analysis of the tissue sections data indicated that only PGR and ERBB2 were significant (Table 2). Therefore, we compared logistic regression models containing the HOXB13:IL17BR ratio either by itself or in combination with tumor size, and expression levels of PGR and ERBB2 (Table 4). The HOXB13:IL17BR ratio alone was a highly significant predictor ($p = 0.0003$) and had an odds ratio of 10.2 (95% CI 2.9–35.6). In the multivariate model, HOXB13:IL17BR ratio is the only significant variable ($p = 0.002$) with an odds ratio of 7.3 (95% CI 2.1–26). Controlling for tumor size, tumor grade, and nodal status in patient selection is an effective strategy to reveal potential gene expression patterns predicting clinical outcome independent of known prognostic factors (Gruvberger et al., 2003). However, the relative predictive power of HOXB13:IL17BR in the context of these known prognostic factors will need to be assessed in larger population-based patient cohorts. The inclusion of only 10 PR-negative patients may also underestimate the predictive utility of PR in this cohort. Nonetheless, these results demonstrated that the expression ratio of HOXB13:IL17BR is a strong independent predictor of treatment outcome in the setting of adjuvant tamoxifen therapy.

Table 4. Logistic regression analysis

Univariate model			
Predictor	Odds ratio	95% CI	p value
HOXB13:IL17BR	10.17	2.9–35.6	0.0003
Multivariate model			
Predictors	Odds ratio	95% CI	p value
Tumor size	1.5	0.7–3.5	0.3289
PGR	0.8	0.3–1.8	0.5600
ERBB2	1.7	0.8–3.8	0.1620
HOXB13:IL17BR	7.3	2.1–26.3	0.0022

All predictors are continuous variables. Gene expression values were from microarray measurements. Odds ratio is the interquartile odds ratio, based on the difference of a predictor from its lower quartile (0.25) to its upper quartile (0.75); CI, confidence interval.

Independent validation of HOXB13:IL17BR expression ratio to predict tamoxifen treatment outcome

The reduction of a complex microarray signature to a two-gene expression ratio allows the use of simpler detection strategies, such as real-time quantitative PCR (RT-QPCR) analysis. We therefore analyzed the HOXB13:IL17BR expression ratio by RT-QPCR using tissue sections that were available from 59 of the 60 frozen specimens (Figure 2A). RT-QPCR data were highly concordant with the microarray data of frozen tumor specimens (Pearson correlation coefficient $r = 0.83$ for HOXB13, 0.93 for IL17BR). In addition, the PCR-derived HOXB13:IL17BR ratios, represented as ΔC_Ts, where C_T is the PCR amplification cycles to reach a predetermined threshold amount (e.g., Figures 2A and 2B) and ΔC_T is the C_T difference between HOXB13 and IL17BR, were highly correlated with the microarray-derived data ($r = 0.83$) and with treatment outcome (t test $p = 5.8e\text{-}06$, Figure 2C). Thus, conventional RT-QPCR analysis for the expression ratio of HOXB13 to IL17BR appears to be equivalent to microarray-based analysis of frozen tumor specimens.

One significant benefit of using a simple two-gene ratio test is the possibility of applying the test in routine clinical specimens, which are mostly formalin-fixed and paraffin-embedded (FFPE). To demonstrate the predictive utility of the HOXB13:IL17BR expression ratio in an independent patient cohort of FFPE specimens, we identified 20 additional ER-positive early-stage primary breast tumors from women treated with adjuvant tamoxifen monotherapy between 1991 and 2000, for whom both medical records and FFPE blocks were available. Of the 20 archival cases, 10 had recurred with a median time to recurrence of 5 years, and 10 had remained disease-free with a median follow up of 9 years (see Table 1 and Supplemental Table S4 for details). RNA was extracted from FFPE tissue sections, linearly amplified, and used as template for RT-QPCR analysis. Consistent with the results of the frozen cohort, the HOXB13:IL17BR expression ratio in this independent patient cohort was highly correlated with clinical outcome (t test $p = 0.024$) with higher HOXB13 expression (lower ΔC_Ts) correlating with poor outcome (Figure 2D).

To test the predictive accuracy of the HOXB13:IL17BR ratio, we used the RT-QPCR data from the frozen tissue sections ($n = 59$) to build a logistic regression model. The model fitted well to this training set, with an overall accuracy of 81%, and the positive and negative predictive values were 81% and 82%,

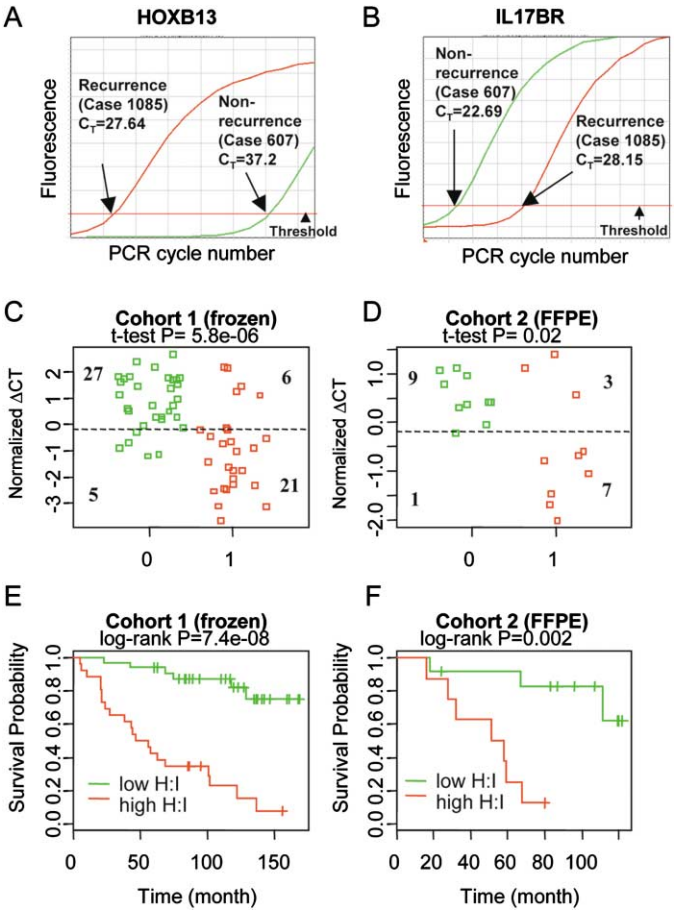

Figure 2. Prediction of clinical outcome by HOXB13:IL17BR ratio

A and B: HOXB13 (**A**) and IL17BR (**B**) real-time quantitative PCR for recurrence case 1085 and nonrecurrence case 607.

C and D: Dot plot of HOXB13:IL17BR ratios according to clinical outcome (0 = nonrecurrence, 1 = recurrence) in the 59 frozen sections (**C**) and 20 FFPE sections (**D**). In each plot, the dashed line denotes the cutoff point determined by logistic regression analysis; numerals represent the number of nonrecurrences and recurrences.

E and F: Kaplan-Meier disease-free survival curves for the frozen (**E**) and FFPE (**F**) cohorts comparing patients with HOXB13:IL17BR ratios above (high H:I) or below (low H:I) the cutoff point defined in the logistic regression model.

respectively (the cutoff point derived from the model was indicated by the dashed line in Figure 2C). Because we identified the two genes and built the model in the same patient cohort, these performance estimates may be biased. Therefore, we applied the model directly to the 20 independent patients in the FFPE cohort. In the resulting "honest" prediction, treatment outcome for 16 of the 20 patients was correctly predicted with an overall accuracy of 80% and positive and negative predictive values of 87% and 75%, respectively (Figure 2D). The probability of correctly predicting 16 of 20 cases by chance was 0.01 (exact binomial test; 95% confidence interval 56%–94%), demonstrating the statistical significance of the two-gene ratio test in the independent FFPE cohort. To further illustrate the predictive ability of HOXB13:IL17BR ratio, the two patient groups stratified by the HOXB13:IL17BR ratio cutoff point exhibited significantly different disease-free survival curves in both the frozen and the FFPE cohorts (Figures 2E and 2F). Therefore, the HOXB13:IL17BR ratio model defined in the frozen cohort had comparable performance in an independent set of 20 FFPE cases, suggesting the potential of this two-gene ratio as an accurate predictive biomarker of clinical outcome for tamoxifen monotherapy.

HOXB13 stimulates mammary epithelial cell migration and invasion

The reduction of our tamoxifen response signature to two genes raised the possibility that these genes may not only be markers of clinical outcome but may be involved in breast tumorigenesis. Little information exists in the literature linking IL17BR to breast cancer. In contrast, HOXB13 has been shown to be overexpressed in a small number of breast cancer tissue samples (n = 11) (Cantile et al., 2003). Indeed, RT-QPCR analysis of LCM-procured normal and malignant breast epithelial cells from our expanded, previously published cohort (n = 45) (Ma et al., 2003) demonstrated that compared to normal breast epithelial cells, the mean expression levels of HOXB13 were significantly higher in both ductal carcinoma in situ (DCIS, $p = 0.002$) and invasive ductal carcinoma (IDC, $p = 0.006$) (Figure 3A). Compared to patient-matched normals, 56% DCIS or IDC cases overexpressed HOXB13 by >2-fold. Furthermore, RNA in situ hybridization confirmed the tumor cell-specific expression of HOXB13 (Figure 3B). Interestingly, a subset of normal breast specimens demonstrated expression of HOXB13 in terminal duct lobular unit, raising the possibility that it may play a role in normal mammary physiology (Figure 3B).

To investigate the potential biological function of HOXB13, we expressed a retroviral construct in MCF-10A cells, a non-transformed human mammary epithelial cell line spontaneously derived from normal breast tissue (Soule et al., 1990). Ectopic expression of HOXB13 in MCF10A was confirmed by RT-QPCR (Figure 3C, insert). Cells expressing HOXB13 displayed distinct morphological changes, characterized by a reduction in epithelial-type junctions (data not shown). Compared to cells infected with the empty vector, MCF10A cells expressing HOXB13 had a 5-fold increase in cell motility in trans-well migration assays in the presence of EGF (Figure 3C). Invasion through a Matrigel-coated modified Boyden chamber, a well-established assay correlated with metastatic potential in vivo (Albini et al., 1987; Repesh, 1989), was also enhanced 5-fold by HOXB13 expression in the presence of EGF (Figure 3D). These observations suggest that HOXB13 may regulate a pathway that functions synergistically with EGF-dependent signaling to stimulate cell motility and invasion in vitro.

Discussion

Previous studies linking gene expression profiles to clinical outcome in breast cancer have demonstrated that the potential for distant metastasis and overall survival probability may be attributable to biological characteristics of the primary tumor at the time of diagnosis (Huang et al., 2003; Sorlie et al., 2001, 2003; Sotiriou et al., 2003; van de Vijver et al., 2002; van 't Veer et al., 2002). In particular, a 70-gene expression signature has proven to be a strong prognostic factor, outperforming all known clinicopathological parameters. However, in those studies, patients either received no adjuvant therapy (van 't Veer et al., 2002) or were treated nonuniformly with hormonal and chemo-

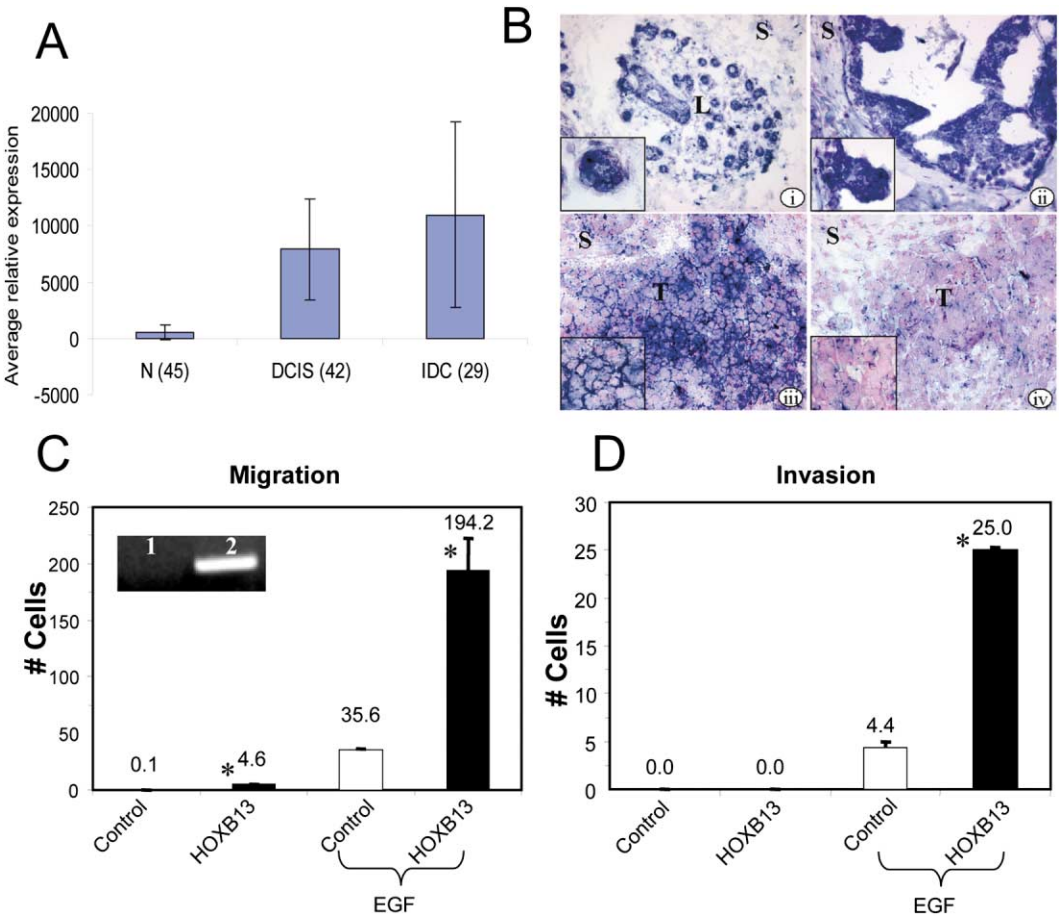

Figure 3. HOXB13 expression in breast cancer and its ectopic expression enhances cell motility and invasion

A: Relative quantitative HOXB13 gene expression values in normal (N, n = 45), DCIS (n = 42), and IDC (n = 29) cases. Error bars denote 95% confidence intervals.

B: In situ hybridization of HOXB13 mRNA. DIG11UTP-labeled RNA probes with anti-sense hybridization to human breast epithelium of (i) the normal terminal duct lobular unit (200× magnification), (ii) ductal carcinoma in situ (400× magnification), and (iii) invasive ductal carcinoma (400× magnification), and sense probe hybridization to (iv) invasive ductal carcinoma (400× magnification). Inserts represent select regions of each field at 1000× magnification. L, S, and T denote lobule, stroma, and tumor, respectively.

C: Migration assay. The mean numbers of cells that migrated through the transwell filter per 20× field are shown ± standard deviation of triplicate wells; insert represents ectopic expression of HOXB13 in MCF10A cells; reverse transcription PCR analysis of HOXB13 from expression constructs with pBABE vector alone (lane 1) or HOXB13 (lane 2).

D: Invasion assay. The mean number of cells that invaded is shown. Error bars indicate one standard deviation. *, $p < 0.05$ compared to control cells (MCF10A cells infected with pBABE).

therapeutic regimens (Huang et al., 2003; Sorlie et al., 2001, 2003; Sotiriou et al., 2003; van de Vijver et al., 2002). Patients with ER-positive early-stage breast cancer treated with tamoxifen alone, such as the cohorts studied here, represent only a subset of the population tested with the 70-gene signature. Of note, 61 of the genes in the 70-gene signature were present on the microarray used in our study, but no significant association with clinical outcome was observed in our patient cohorts (data not shown). Conversely, the HOXB13:IL17BR ratio identified in our tamoxifen-treated cohorts is not effective in predicting clinical outcome of patients receiving no adjuvant therapy in the van 't Veer study (data not shown). Hence, the genes identified here as correlated with clinical outcome in the setting of adjuvant tamoxifen monotherapy may differ from those associated with inherent biological properties of a more diverse collection of breast cancers. We also attempted to evaluate our two-gene ratio test in a microarray data set containing 45 patients with ER+ breast cancer treated with tamoxifen monotherapy (Sotiriou et al., 2003). However, the published cDNA microarray dataset did not include HOXB13 or IL17BR. Even when all 25 genes identified in this study (Figures 1A and 1B) were considered, only two genes (IL1R2 and APS) were present on their cDNA microarray, making it impossible to evaluate our predictive signature.

The observation that a simple expression ratio of two genes, HOXB13:IL17BR, accurately predicts tumor recurrence in adjuvant tamoxifen-treated patients with early-stage ER-positive breast cancer is limited by the size of patient cohorts in this study and will require confirmation in a larger population-based cohort. Furthermore, it remains to be determined whether this two-gene ratio predicts a tumor's response to tamoxifen or its intrinsic aggressiveness, or both. A similarly case-matched

cohort of untreated patients will be required to address this issue. If confirmed, this simple assay may prove helpful in identifying patients who may benefit from more complete ablation of estrogen signaling using aromatase inhibitors (Goss et al., 2003) or from the addition of chemotherapeutic agents (Colleoni et al., 2003b). Whereas most reported prognostic signatures in human cancer have involved a large number of markers and required microarray-based analysis, we were able to reduce the tamoxifen outcome signature to a two-gene ratio with no loss of accuracy. The application of a simple two-gene PCR-based assay to standard FFPE clinical tissues will allow for broader access and utilization of this technology in the clinical setting.

Little is known about the relevance of HOXB13 in breast cancer biology. Homeobox genes encode transcriptional regulators implicated in normal organ development, whose disruption has been most clearly linked with hematopoietic malignancies (reviewed in Abate-Shen, 2002; Maulbecker and Gruss, 1993). However, expression of HOXC8 was recently shown to be upregulated in prostate cancer and to suppress androgen receptor signaling (Miller et al., 2003), raising the possibility that some homeobox genes may interact with hormone receptor signaling. Consistent with a previous report (Cantile et al., 2003), we demonstrated that HOXB13 was frequently upregulated in breast cancer cells relative to normal breast epithelial cells. Of note, we detected HOXB13 expression in normal cells of the terminal duct lobular unit, the anatomic substructure of the human breast from which breast cancer arises, suggesting that this homeobox protein may play a role in breast development and physiology. Importantly, ectopic expression of HOXB13 in MCF10A cells potentiates EGF-induced cell migration and invasion, suggesting that HOXB13 may directly contribute to tumor invasion and metastasis. Functional cooperation between HOXB13 and EGFR signaling pathways may be relevant in the context of tamoxifen resistance, since activation of growth factor signaling pathways can cause tamoxifen-resistant tumor growth (Clarke et al., 2003; Nicholson et al., 1994; Osborne and Schiff, 2003). Given the known role of ERBB2 overexpression in human breast cancer, the apparent in vitro interaction between HOXB13 expression and EGF signaling pathways may also point to possible therapeutic options in tumors with high expression of HOXB13. In fact, targeting the ERBB2 pathway through blocking antibodies (Herceptin) has been suggested in the context of tamoxifen resistance based on the link between activation of growth factor signaling pathways and estrogen-independent tumor growth (Clarke et al., 2003; Nicholson et al., 2003; Osborne and Schiff, 2003). HOXB13 may also have a direct effect on ER signaling, since homeobox proteins have been shown to inhibit the histone acetyltransferase activity of CBP/p300 (Shen et al., 2001), a key coactivator for ER-dependent transcriptional regulation (Chakravarti et al., 1996; Hanstein et al., 1996). Therefore, HOXB13 may be involved directly and indirectly in the modulation of ER signaling pathways, a possibility that is of particular interest given its clinical correlation with tamoxifen resistance.

In summary, our findings demonstrate the utility of a two-gene expression biomarker in identifying a subset of patients with ER-positive breast cancer who are at risk for tumor recurrence in the setting of tamoxifen therapy, who may benefit from alternative therapeutic options. Furthermore, our results have the potential to open new avenues of research into the molecular mechanisms of resistance to tamoxifen.

Experimental procedures

Patient and tumor selection

Patient inclusion criteria for this study were: women diagnosed at the Massachusetts General Hospital (MGH) between 1987 and 2000 with hormone receptor-positive breast cancer, treatment with standard breast surgery (modified radical mastectomy or lumpectomy), and radiation followed by five years of systemic adjuvant tamoxifen; no patient received chemotherapy prior to recurrence. Clinical and follow-up data were derived from the MGH tumor registry. There were no missing registry data and all available medical records were reviewed as a second tier of data confirmation. This study was approved by the institutional review board of the MGH in accordance with federal human research study guidelines.

All tumor specimens collected at the time of initial diagnosis were obtained from frozen and formalin-fixed paraffin-embedded (FFPE) tissue repositories at the Massachusetts General Hospital. Tumor samples with greater than 20% tumor cells were selected with a median of greater than 75% for all samples. Each sample was evaluated for the following features: tumor type (ductal versus lobular), tumor size, and Nottingham combined histological grade (Dalton et al., 2000). Estrogen and progesterone receptor expression were determined by biochemical hormone binding analysis and/or by immunohistochemical staining as described (Harvey et al., 1999). More specifically, ER positivity was defined as greater than 10 fmol/mg tumor tissue and greater than 1% nuclear staining or immunohistochemistry (IHC) score of at least 3 for the biochemical and immunohistochemical assays (Harvey et al., 1999), respectively; all but three tumors (2 recurrences and 1 nonrecurrence) demonstrated greater than 10% ER staining by IHC (see Supplemental Table S1). Immunohistochemistry for Her-2/neu protein expression was performed on paraffin-embedded or frozen tissue sections using anti-HER-2/neu polyclonal antibody (Dako Corp, Carpinteria CA, dilution 1:300) or anti-HER2/neu monoclonal antibody NCL-CB11 (Novocastra Laboratories, Newcastle upon Tyne, UK, dilution 1:50) and employing standard immunoperoxidase techniques (Jacobs et al., 1999). The membrane staining of cells was scored as no or weak staining (0–1+ or negative), moderate staining (2+ or indeterminate) and strong staining (3+ or positive); in our experience, most tumors showing 2+ staining do not demonstrate HER-2/neu amplification by FISH. Tumors from recurrence cases were matched to the nonrecurrences with respect to TNM staging and tumor grade (Dalton et al., 2000; Singletary et al., 2002).

LCM and RNA isolation and amplification

With each frozen tumor sample within the 60-case cohort, RNA was isolated from both a whole tissue section of 8 μm in thickness and a highly enriched population of 4,000–5,000 malignant epithelial cells procured by laser capture microdissection using a PixCell IIe LCM system (Arcturus, Mountain View, CA). From each tumor sample within the 20-case test set, RNA was isolated from four 8 μm-thick FFPE tissue sections using the Paradise Reagent System (Arcturus, Mountain View, CA). Isolated RNA was subjected to DNase treatment followed by one round of T7 polymerase in vitro transcription using the RiboAmp kit (frozen samples) or the Paradise system (FFPE samples) according to manufacturer's instructions (Arcturus, Mountain View, CA). Labeled cRNA was generated by a second round of T7-based RNA in vitro transcription in the presence of 5-[3-Aminoallyl]uridine 5′-triphosphate (Sigma-Aldrich, St. Louis, MO). Universal Human Reference RNA (Stratagene, San Diego, CA) was amplified in the same manner. The purified aRNA was later conjugated to Cy5 (experimental samples) or Cy3 (reference sample) dye (Amersham Biosciences).

Data analysis

Raw data are available in GEO (http://www.ncbi.nlm.nih.gov/geo/, accession numbers GSE1378 and GSE1379). A custom-designed 22,000-gene oligonucleotide (60mer) microarray was fabricated using ink-jet in-situ synthesis technology (Agilent Technologies, Palo Alto, CA). Cy5-labeled sample RNA and Cy3-labeled reference RNA were cohybridized at 65°C, 1× hybridization buffer (Agilent Technologies). Slides were washed at 37°C with 0.1× SSC/0.005% Triton X-102. Image analysis was performed using Agilent's image analysis software.

Raw Cy5/Cy3 ratios per array were normalized using nonlinear local regression (loess, Yang et al., 2002); no background adjustment was made due to generally low local background values. Between-array normalization

was achieved by subtracting the median log2 ratios gene-wise across all arrays. The variance of each gene over all samples was calculated and the top 25% high variance genes (5,475) selected for further analysis. Identification and permutation testing for significance of differential gene expression were performed using BRB ArrayTools, developed by Richard Simon and Amy Peng (http://linus.nci.nih.gov/BRB-ArrayTools.html). Hierarchical cluster analysis was performed with GeneMaths software (Applied-Maths, Belgium) using cosine correlation as distance and complete linkage clustering. All other statistical procedures (two-sample t test, receiver operating characteristic analysis, logistic regression, and survival analysis) were performed in the open-source R statistical environment (http://www.r-project.org). Statistical test of significance of ROC curves was by the method of DeLong et al. (1988) as implemented by Doug Mahoney and Beth Atkinson in their S-Plus roc package (http://www.mayo.edu/hsr/Sfunc.html). Disease-free survival was calculated from the date of diagnosis. Events were scored as the first distant metastasis, and patients remaining disease-free at the last follow-up were censored. Survival curves were calculated by the Kaplan-Meier estimates and compared by log-rank tests. To build a predictive model using HOXB13 and IL17BR, raw C_T values from RT-QPCR were normalized across the sample set (subtracting the mean C_T and dividing by the standard deviation), and the two-gene ratio was the difference of normalized C_T values (HOXB13 − IL17BR); normalization was done separately for the frozen and FFPE cohorts. The two-gene ratio was then used in a univariate logistic regression model, and the cutoff point for predicting recurrences = −intercept/β coefficient for the ratio, both of which were derived from the model.

In the analysis of the data published previously, data were downloaded from the authors' website (van 't Veer et al., 2002) or the publisher's website (Sotiriou et al., 2003). Genes common between the microarray in this study and the published microarrays were mapped using the UniGene database (http://www.ncbi.nlm.nih.gov/) by UniGene ID. In the evaluation of the 70-gene prognostic signature (van 't Veer et al., 2002) in our study cohort, we used the compound covariate predictor algorithm (Hedenfalk et al., 2001) as implemented in BRB ArrayTools to predict recurrences.

Real-time quantitative PCR analysis

RT-QPCR was performed on 59 of the 60 training samples (one case was excluded due to insufficient materials) and the 20 validation samples. All LCM and tissue-section derived RNA samples were subjected to DNase treatment prior to cDNA synthesis and RNA amplification. Briefly, 2 μg of amplified RNA was converted into double stranded cDNA. For each case, 12 ng of cDNA in triplicates was used for real-time PCR with an ABI 7900HT (Applied Biosystems) as described (Gelmini et al., 1997). The sequences of the PCR primer pairs and fluorogenic MGB probe (5′ to 3′), respectively, that were used for each gene are as follows: HoxB13, TTCATCCTGACAGTG GCAATAATC, CTAGATAGAAAATATGAGGCTA-ACGATCAT, VIC-CGATAA CCAGTACTAGCTG; IL17BR, GCATTAACTAACGATTGGAAACTACATT, GGA AGATGCTTTATTGTTGCATTATC, VIC-ACAACTTCAAAGCTGTTTTA. Relative expression levels of HOXB13 in normal, DCIS, and IDC samples were calculated as follows. First, all C_T values were adjusted by subtracting the highest C_T (40) among all samples, then relative expression = 1/2^C_T.

In situ hybridization

Dig-labeled RNA probes were prepared using DIG RNA labeling kit (SP6/T7) from Roche Applied Science, following the protocol provided with the kit. In situ hybridization was performed on frozen tissue sections as described (Long et al., 1992). The probe covered the entire coding region of the HOXB13 transcript (GenBank Accession BC007092).

Cell culture and cell line construction

MCF-10A cells (ATCC) (Soule et al., 1990) were maintained in growth medium as described (Debnath et al., 2003) in DMEM/F12 (Invitrogen) with 5% horse serum (Invitrogen), 20 ng/ml EGF (Peprotech), 10 μg/ml insulin (Sigma), 100 ng/ml cholera toxin, 0.5 μg/ml hydrocortisone, 50 U/ml penicillin, 50 μg/ml streptomycin. Assay medium is identical to the growth medium except 2% horse serum was used instead of 5%. Human cDNA HOXB13 in the pDNR plasmid was generously provided by Joshua LaBaer (Harvard Medical School). HOXB13 was subcloned into the SnaB1 site of the retroviral expression vector pBabe-puro (Morgenstern and Land, 1990) and proper orientation determined by restriction mapping. Replication-incompetent virus with the vesicular stomatitis virus (VSV) envelope was generated from VSV-GPG packaging cells as described (Ory et al., 1996), and stable pools of MCF-10A cells were generated by retroviral infection as described (Debnath et al., 2003) using 2 μg/ml puromycin for selection. Control cells were MCF10 cells infected with the empty pBABE vector.

Trans-well migration and invasion assay

In vitro migration and invasion assays were performed using 24-well modified Boyden chamber trans-well with PET (polyethylene terephthalate) membranes containing 8 micron pores (BD BioCoat). Uncoated membranes were used for migration assays, and Matrigel-coated membranes were used for invasion (Repesh, 1989). Stable pools of MCF-10A cells infected with retroviral constructs were maintained in growth medium until the day of the assay. 5×10^4 cells in 100 μl of assay medium were seeded in the upper chamber, and 500 μl assay medium with or without 20 ng/ml EGF was added to the lower chamber. Cells were incubated at 37°C for 24 hr, then fixed in 70% ethanol for 20 min, rinsed with PBS, and stained with DAPI (500 ng/ml). Cells that remained on the upper surface were mechanically removed with a cotton swab. Cells remaining on the underside were counted (5 fields at 20× magnification per trans-well). Trans-wells were plated in triplicates and the results were averaged. Invasion through a Matrigel-coated modified Boyden chamber was assayed as described (Albini et al., 1987; Repesh, 1989).

Acknowledgments

We thank Leif Ellisen, Emmett Schmidt, and Michael Overholtzer for helpful discussions, and Philip McQuary, Anitha Male, Sonika Dahiya, Teresa Alvarez, Leah Gaumont, Katrina Mesina, Marc Laughlin, and Brian M. Zhang for technical assistance. This work is supported in part by grants from the Avon Foundation, the Department of Defense Grants DAMD17-03-1-0428 (to D.C.S.) and the National Cancer Institute-Specialized Programs of Research Excellence (NCI SPORE) in Breast Cancer at Massachusetts General Hospital. X.-J.M., R.S., J.T.T., Y.T., D.T., A.T., P.A., Wilson Wang, Wei Wang, E.E., K.S., E.E.-S., T.M.B., and M.G.E. are employees of Arcturus Bioscience Inc. X.-J.M., M.G.E., and D.C.S. are named inventors on a patent to use the HoxB13:IL17BR expression ratio to ascertain breast cancer prognosis.

Received: February 20, 2004
Revised: April 15, 2004
Accepted: May 10, 2004
Published online: June 3, 2004

References

Abate-Shen, C. (2002). Deregulated homeobox gene expression in cancer: cause or consequence? Nat. Rev. Cancer *2*, 777–785.

Albini, A., Iwamoto, Y., Kleinman, H.K., Martin, G.R., Aaronson, S.A., Kozlowski, J.M., and McEwan, R.N. (1987). A rapid in vitro assay for quantitating the invasive potential of tumor cells. Cancer Res. *47*, 3239–3245.

Atalay, G., Cardoso, F., Awada, A., and Piccart, M.J. (2003). Novel therapeutic strategies targeting the epidermal growth factor receptor (EGFR) family and its downstream effectors in breast cancer. Ann. Oncol. *14*, 1346–1363.

Bardou, V.J., Arpino, G., Elledge, R.M., Osborne, C.K., and Clark, G.M. (2003). 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.

Bonner, R.F., Emmert-Buck, M., Cole, K., Pohida, T., Chuaqui, R., Goldstein, S., and Liotta, L.A. (1997). Laser capture microdissection: Molecular analysis of tissue. Science *278*, 1481–1483.

Buzdar, A.U. (1998). Anastrozole: a new addition to the armamentarium against advanced breast cancer. Am. J. Clin. Oncol. *21*, 161–166.

Cantile, M., Pettinato, G., Procino, A., Feliciello, I., Cindolo, L., and Cillo, C. (2003). In vivo expression of the whole HOX gene network in human breast cancer. Eur. J. Cancer *39*, 257–264.

Chakravarti, D., LaMorte, V.J., Nelson, M.C., Nakajima, T., Schulman, I.G.,

Juguilon, H., Montminy, M., and Evans, R.M. (1996). Role of CBP/P300 in nuclear receptor signalling. Nature *383*, 99–103.

Clarke, R., Liu, M.C., Bouker, K.B., Gu, Z., Lee, R.Y., Zhu, Y., Skaar, T.C., Gomez, B., O'Brien, K., Wang, Y., and Hilakivi-Clarke, L.A. (2003). Antiestrogen resistance in breast cancer and the role of estrogen receptor signaling. Oncogene *22*, 7316–7339.

Colleoni, M., Curigliano, G., Minchella, I., Peruzzotti, G., Nole, F., Mazzarol, G., Renne, G., Orlando, L., Rocca, A., Veronesi, P., et al. (2003a). Preoperative and perioperative chemotherapy with 5-fluorouracil as continuous infusion in operable breast cancer expressing a high proliferation fraction: cytotoxic treatment during the surgical phase. Ann. Oncol. *14*, 1477–1483.

Colleoni, M., Zarieh, D., Gelber, R.D., Viale, G., Luini, A., Veronesi, P., Intra, M., Galimberti, V., Renne, G., and Goldhirsch, A. (2003b). Preoperative systemic treatment: prediction of responsiveness. Breast *12*, 538–542.

Dalton, L.W., Pinder, S.E., Elston, C.E., Ellis, I.O., Page, D.L., Dupont, W.D., and Blamey, R.W. (2000). Histologic grading of breast cancer: linkage of patient outcome with level of pathologist agreement. Mod. Pathol. *13*, 730–735.

Debnath, J., Muthuswamy, S.K., and Brugge, J.S. (2003). Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. Methods *30*, 256–268.

DeLong, E.R., DeLong, D.M., and Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics *44*, 837–845.

Dowsett, M. (2001). Overexpression of HER-2 as a resistance mechanism to hormonal therapy for breast cancer. Endocr. Relat. Cancer *8*, 191–195.

EBCTCG. (2001). Tamoxifen for early breast cancer. Cochrane Database Syst. Rev. *1*, CD000486.

Ellis, M.J., Coop, A., Singh, B., Mauriac, L., Llombert-Cussac, A., Janicke, F., Miller, W.R., Evans, D.B., Dugan, M., Brady, C., et al. (2001). 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.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A., and Liotta, L.A. (1996). Laser capture microdissection. Science *274*, 998–1001.

Fernandez, M.D., Alaghband-Zadeh, J., and Burn, J.I. (1983). Quantitative oestrogen and progesterone receptor values in primary breast cancer and predictability of response to endocrine therapy. Clin. Oncol. *9*, 245–250.

Ferno, M., Stal, O., Baldetorp, B., Hatschek, T., Kallstrom, A.C., Malmstrom, P., Nordenskjold, B., and Ryden, S. (2000). Results of two or five years of adjuvant tamoxifen correlated to steroid receptor and S-phase levels. South Sweden Breast Cancer Group, and South-East Sweden Breast Cancer Group. Breast Cancer Res. Treat. *59*, 69–76.

Fitzgibbons, P.L., Page, D.L., Weaver, D., Thor, A.D., Allred, D.C., Clark, G.M., Ruby, S.G., O'Malley, F., Simpson, J.F., Connolly, J.L., et al. (2000). Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. Arch. Pathol. Lab. Med. *124*, 966–978.

Gelmini, S., Orlando, C., Sestini, R., Vona, G., Pinzani, P., Ruocco, L., and Pazzagli, M. (1997). Quantitative polymerase chain reaction-based homogeneous assay with fluorogenic probes to measure c-erbB-2 oncogene amplification. Clin. Chem. *43*, 752–758.

Goss, P.E., Ingle, J.N., Martino, S., Robert, N.J., Muss, H.B., Piccart, M.J., Castiglione, M., Tu, D., Shepherd, L.E., Pritchard, K.I., et al. (2003). A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer. N. Engl. J. Med. *349*, 1793–1802.

Gruvberger, S.K., Ringner, M., Eden, P., Borg, A., Ferno, M., Peterson, C., and Meltzer, P.S. (2003). Expression profiling to predict outcome in breast cancer: the influence of sample selection. Breast Cancer Res. *5*, 23–26.

Hanstein, B., Eckner, R., DiRenzo, J., Halachmi, S., Liu, H., Searcy, B., Kurokawa, R., and Brown, M. (1996). p300 is a component of an estrogen receptor coactivator complex. Proc. Natl. Acad. Sci. USA *93*, 11540–11545.

Harvey, J.M., Clark, G.M., Osborne, C.K., and Allred, D.C. (1999). 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.

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O.P., et al. (2001). Gene-expression profiles in hereditary breast cancer. N. Engl. J. Med. *344*, 539–548.

Huang, E., Cheng, S.H., Dressman, H., Pittman, J., Tsou, M.H., Horng, C.F., Bild, A., Iversen, E.S., Liao, M., Chen, C.M., et al. (2003). Gene expression predictors of breast cancer outcomes. Lancet *361*, 1590–1596.

Jacobs, T.W., Gown, A.M., Yaziji, H., Barnes, M.J., and Schnitt, S.J. (1999). Comparison of fluorescence in situ hybridization and immunohistochemistry for the evaluation of HER-2/neu in breast cancer. J. Clin. Oncol. *17*, 1974–1982.

Jordan, C. (2002). Historical perspective on hormonal therapy of advanced breast cancer. Clin. Ther. *24* (*Suppl. A*), A3–16.

Lichtner, R.B. (2003). Estrogen/EGF receptor interactions in breast cancer: rationale for new therapeutic combination strategies. Biomed. Pharmacother. *57*, 447–451.

Long, A.A., Mueller, J., Andre-Schwartz, J., Barrett, K.J., Schwartz, R., and Wolfe, H. (1992). High-specificity in-situ hybridization. Methods and application. Diagn. Mol. Pathol. *1*, 45–57.

Ma, X.J., Salunga, R., Tuggle, J.T., Gaudet, J., Enright, E., McQuary, P., Payette, T., Pistone, M., Stecker, K., Zhang, B.M., et al. (2003). Gene expression profiles of human breast cancer progression. Proc. Natl. Acad. Sci. USA *100*, 5974–5979.

Maulbecker, C.C., and Gruss, P. (1993). The oncogenic potential of deregulated homeobox genes. Cell Growth Differ. *4*, 431–441.

Miller, G.J., Miller, H.L., van Bokhoven, A., Lambert, J.R., Werahera, P.N., Schirripa, O., Lucia, M.S., and Nordeen, S.K. (2003). Aberrant HOXC expression accompanies the malignant phenotype in human prostate. Cancer Res. *63*, 5879–5888.

Morgenstern, J.P., and Land, H. (1990). Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. Nucleic Acids Res. *18*, 3587–3596.

Morris, C. (2002). The role of EGFR-directed therapy in the treatment of breast cancer. Breast Cancer Res. Treat. *75* (*Suppl. 1*), S51–S55.

Nardelli, G.B., Lamaina, V., and Siliotti, F. (1986). Estrogen and progesterone receptors status in the prediction of response of breast cancer to endocrine therapy (preliminary report). Eur. J. Gynaecol. Oncol. *7*, 151–158.

Nicholson, R.I., McClelland, R.A., Gee, J.M., Manning, D.L., Cannon, P., Robertson, J.F., Ellis, I.O., and Blamey, R.W. (1994). Epidermal growth factor receptor expression in breast cancer: association with response to endocrine therapy. Breast Cancer Res. Treat. *29*, 117–125.

Nicholson, R.I., Gee, J.M., Knowlden, J., McClelland, R., Madden, T.A., Barrow, D., and Hutcheson, I. (2003). The biology of antihormone failure in breast cancer. Breast Cancer Res. Treat. *80* (*Suppl 1*), S29–S34.

Ory, D.S., Neugeboren, B.A., and Mulligan, R.C. (1996). A stable human-derived packaging cell line for production of high titer retrovirus/vesicular stomatitis virus G pseudotypes. Proc. Natl. Acad. Sci. USA *93*, 11400–11406.

Osborne, C.K., Yochmowitz, M.G., Knight, W.A., 3rd, and McGuire, W.L. (1980). The value of estrogen and progesterone receptors in the treatment of breast cancer. Cancer *46*, 2884–2888.

Osborne, C.K., and Schiff, R. (2003). Growth factor receptor cross-talk with estrogen receptor as a mechanism for tamoxifen resistance in breast cancer. Breast *12*, 362–367.

Pepe, M.S. (2000). An interpretation for the ROC curve and inference using GLM procedures. Biometrics *56*, 352–359.

Repesh, L.A. (1989). A new in vitro assay for quantitating tumor cell invasion. Invasion Metastasis *9*, 192–208.

Shen, W.F., Krishnan, K., Lawrence, H.J., and Largman, C. (2001). The HOX

homeodomain proteins block CBP histone acetyltransferase activity. Mol. Cell. Biol. *21*, 7509–7522.

Singletary, S.E., Allred, C., Ashley, P., Bassett, L.W., Berry, D., Bland, K.I., Borgen, P.I., Clark, G., Edge, S.B., Hayes, D.F., et al. (2002). Revision of the American Joint Committee on Cancer staging system for breast cancer. J. Clin. Oncol. *20*, 3628–3636.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA *98*, 10869–10874.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., et al. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc. Natl. Acad. Sci. USA *100*, 8418–8423.

Sotiriou, C., Neo, S.Y., McShane, L.M., Korn, E.L., Long, P.M., Jazaeri, A., Martiat, P., Fox, S.B., Harris, A.L., and Liu, E.T. (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc. Natl. Acad. Sci. USA *100*, 10393–10398.

Soule, H.D., Maloney, T.M., Wolman, S.R., Peterson, W.D., Jr., Brenz, R., McGrath, C.M., Russo, J., Pauley, R.J., Jones, R.F., and Brooks, S.C. (1990). Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. Cancer Res. *50*, 6075–6086.

van de Vijver, M.J., He, Y.D., van 't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. *347*, 1999–2009.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. *30*, e15.

Zeltser, L., Desplan, C., and Heintz, N. (1996). Hoxb-13: a new Hox gene in a distant region of the HOXB cluster maintains colinearity. Development *122*, 2475–2484.

Accession numbers

The microarray datasets from this study have been deposited to GEO (http://www.ncbi.nlm.nih.gov/geo, accession numbers GSE1378 and GSE1379).