

## Gene Expression Profiles of Multiple Breast Cancer Phenotypes and Response to Neoadjuvant Chemotherapy

Holly K. Dressman,<sup>1,2</sup> Christopher Hans,<sup>6</sup> Andrea Bild,<sup>2,6</sup> John A. Olson,<sup>1,3</sup> Eric Rosen,<sup>4</sup> P. Kelly Marcom,<sup>5</sup> Vlayka B. Liotcheva,<sup>4</sup> Ellen L. Jones,<sup>4</sup> Zeljko Vujaskovic,<sup>5</sup> Jeffrey Marks,<sup>3</sup> Mark W. Dewhirst,<sup>4</sup> Mike West,<sup>1,6</sup> Joseph R. Nevins,<sup>1,2</sup> and Kimberly Blackwell<sup>4,5</sup>

**Abstract Purpose:** Breast cancer is a heterogeneous disease, and markers for disease subtypes and therapy response remain poorly defined. For that reason, we employed a prospective neoadjuvant study in locally advanced breast cancer to identify molecular signatures of gene expression correlating with known prognostic clinical phenotypes, such as inflammatory breast cancer or the presence of hypoxia. In addition, we defined molecular signatures that correlate with response to neoadjuvant chemotherapy.

**Experimental Design:** Tissue was collected under ultrasound guidance from patients with stage IIB/III breast cancer before four cycles of neoadjuvant liposomal doxorubicin paclitaxel chemotherapy combined with local whole breast hyperthermia. Gene expression analysis was done using Affymetrix U133 Plus 2.0 GeneChip arrays.

**Results:** Gene expression patterns were identified that defined the phenotypes of inflammatory breast cancer as well as tumor hypoxia. In addition, molecular signatures were identified that predicted the persistence of malignancy in the axillary lymph nodes after neoadjuvant chemotherapy. This persistent lymph node signature significantly correlated with disease-free survival in two separate large populations of breast cancer patients.

**Conclusions:** Gene expression signatures have the capacity to identify clinically significant features of breast cancer and can predict which individual patients are likely to be resistant to neoadjuvant therapy, thus providing the opportunity to guide treatment decisions.

Breast cancer is an enormously complex and heterogeneous disease, with tumors reflecting the acquisition of multiple genetic alterations that disrupt normal cellular regulatory processes. The complexity is manifest in a variety of often overlapping observable phenotypes. One phenotype defines inflammatory breast cancer (IBC), a rare subtype of breast cancer seen in only about 1% of cases. IBC carries a poor prognosis with only 50% survival over a 5-year period (1). The clinical definition of IBC is somewhat imprecise and correlates with clinical features of the disease, such as breast erythema,

warmth, or s.c. edema (peau de orange), or from pathologic features of the disease, such as presence of carcinoma within the dermal lymphatics. The association of IBC with poor survival mandates a focus on improving our understanding of the molecular characteristics underlying the phenotype.

Hypoxia results in cellular responses and is another risk-related phenotype that plays roles in tumor development, progression, and therapy responsiveness (2, 3). Tumor oxygenation plays an important role in altered gene expression, multidrug resistance, tumor cell invasiveness, angiogenesis, and metastasis (4). Tumor hypoxia has been shown to be of prognostic and predictive value in several clinical trials involving radiation, chemotherapy, and surgery for various tumor types (5–8).

The heterogeneity of breast cancer presents an enormous challenge to the goal of customizing therapy for the individual patient. Therapy customization is essential to improving efficacy, limiting unnecessary treatment-related morbidity, and ultimately, eliminating unnecessary treatment. A woman diagnosed with early-stage breast cancer will undergo surgery for removal of the tumor and then typically will be treated with adjuvant chemotherapy. Nevertheless, a number of such women then unnecessarily receive potentially toxic chemotherapy.

The use of genomic data offers the potential to guide treatment options by improving risk assessments and identifying patients likely resistant to standard therapies. To address the latter, we developed gene expression data from prospectively

**Authors' Affiliations:** <sup>1</sup>Duke Institute for Genome Sciences and Policy, Departments of <sup>2</sup>Molecular Genetics and Microbiology, <sup>3</sup>Surgery, <sup>4</sup>Radiation Oncology, and <sup>5</sup>Medicine, Duke University Medical Center, <sup>6</sup>Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina  
Received 7/5/05; revised 11/1/05; accepted 11/16/05.

**Grant support:** NIH grants CA42745 (M.W. Dewhirst) and CA112952 (J.R. Nevins).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

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

H. Dressman and C. Hans contributed equally to this work.

**Requests for reprints:** Holly Dressman, Duke Inst. for Genome Sciences & Policy, Box 3382, Durham, NC 27710. Phone: 919-668-1583; Fax: 919-681-8973; E-mail: [dress002@mc.duke.edu](mailto:dress002@mc.duke.edu).

© 2006 American Association for Cancer Research.  
doi:10.1158/1078-0432.CCR-05-1447

collected pretreatment breast cancer biopsies from patients in a neoadjuvant chemotherapy trial. The resulting data were evaluated and generated molecular signatures predictive of response to chemotherapy, and, in parallel, characterizing both IBC features and the presence of tumor hypoxia, as assessed independently using polarographic electrodes with ultrasound guidance for probe placement (9). Further analysis evaluated the treatment response signature on microarray data from primary breast tumors arising from two separate, distinct and large retrospective studies. The ultimate goal in examining these molecular signatures (quantitative prognostic phenotypes and predictors of chemotherapy sensitivity) is to establish gene expression variables that will ultimately lead to improved outcomes and avoidance of unnecessary therapy in breast cancer.

### Materials and Methods

**Patient samples and tissue acquisition.** Tissue samples were collected from patients enrolled in a phase I-II, open-label study of liposomal doxorubicin (Evacet, Elan Corp., Stevenage, United Kingdom) and paclitaxel (Bristol Myers Squibb, Princeton, NJ) in combination with whole breast hyperthermia for the neoadjuvant treatment of locally advanced breast cancer (stage IIB or III). This trial required informed consent and was conducted under the approval of the Duke University Institutional Review Board. Protocol-eligible patients were treated with the combination of Evacet, paclitaxel, and hyperthermia every 3 weeks. Details of the hyperthermia procedure have been published elsewhere (9) After neoadjuvant therapy, patients received appropriate surgical removal of their primary breast tumor as well as axillary lymph node dissection. Immediately after surgery, patients underwent radiation therapy followed by an additional eight cycles of every 21-day standard dose cyclophosphamide (600 mg/m<sup>2</sup>), methotrexate (40 mg/m<sup>2</sup>), 5- and fluorouracil (600 mg/m<sup>2</sup>) and appropriate hormonal therapy. The trial has completed accrual with a total of 47 patients. Three patients were deemed nonevaluable because of failure to complete all four cycles of the neoadjuvant portion of the trial. The median follow-up for the study population from the time of completion of the neoadjuvant therapy is 18.5 months (range, 3-44 months). The median tumor size on this trial was 5.65 cm (range, 2.5-11 cm). Tissue collection was conducted at the following time points: before initiation of therapy (enrollment), before cycle 3 of neoadjuvant therapy, and at the time of definitive surgery. Under ultrasound guidance, using techniques previously described (10), polarographic oxygenation measurements were conducted followed by four to six 10-gauge cores collected through the area of invasive disease. Tissue was flash frozen over liquid nitrogen and stored at -80°C. Thirty-seven evaluable patients consented for research tissue collection and completed all four courses of neoadjuvant therapy. Clinical details of patients enrolled in this study are listed in Table 1.

**Samples used for evaluation of prognosis.** Microarray expression data was also available on 158 previously described breast cancer samples from the Koo Foundation Sun Yat Sen Cancer Center in Taipei, Taiwan (KF-SYSCC; refs. 11, 12) and on a collection of 101 breast cancer samples from Duke (<http://data.cgt.duke.edu/blackwell.php>). Data from these two separate retrospective studies were used to explore and evaluate the prognostic significance of the neoadjuvant response predictor defined in the current analysis.

**Microarray analysis.** This study involved a total of 37 evaluable patients from which pretreatment biopsies were obtained to allow gene expression analysis. All 37 tissue samples were sectioned and determined to contain at least 60% invasive disease throughout the core sample before RNA harvesting. RNA was prepared, probe generated, and used for hybridization to Affymetrix U133 Plus 2.0 GeneChip arrays (<http://www.affymetrix.com/products/arrays/specific/>

**Table 1.** Clinical characteristics of study patients

Patient characteristics	No.
Patients in study*	
Total enrolled	47
Evaluable	37
Race	
Caucasian	29
African American	6
American Indian	1
Clinical stage pretreatment	
Stage IIB	6
Stage III	30
HER-2 status (fluorescence <i>in situ</i> hybridization)	
Amplified	12
Nonamplified	24
Clinical response	
Complete	10
Partial	15
None	10
Pathologic response	
Complete	3
Partial	20
None	12
Lymph node involvement (after therapy)	
Positive	27
Negative	9
Inflammatory disease	
Positive	14
Negative	23
Hypoxia (mean $pO_2 < 10$ mm Hg)	
Present	20
Negative	14

\*The total number of patients in the study was 47. Three failed to complete the planned treatment. Of the remaining 44 patients that completed the study, biopsies were obtained from 37 patients for RNA analysis.

[hgu133plus.affx](http://hgu133plus.affx)). Expression was calculated using the robust multiarray average algorithm (13) implemented in the Bioconductor (<http://www.bioconductor.org>) extensions to the R statistical programming environment (14). Robust multiarray average generates a background-corrected and quantile-normalized measure of expression (15) on the log 2 scale of measurement. Expression estimates from the arrays of 37 tumor samples were then screened to identify genes (in reality, probe sets) showing some evidence of more than trivial variation across samples above noise levels. Specifically, we removed probe sets showing normalized expression levels varying at least 1.5-fold across the 37 samples, and whose maximum level among the 37 exceeded the 60% percentile of all data values. Thus, the reduced probe set consisted of 24,134 transcripts that were candidate predictors used in the regression model analysis.

**Determination of tumor oxygenation status.** Tumor oxygenation measurements were done under sterile conditions using local anesthetic immediately before the planned tumor core biopsies using a polarographic device (Eppendorf Netheler Hinz, GmbH, Hamburg, Germany). This technique has been described previously (10). Briefly, an anode is placed on patient's skin and polarized with a constant voltage of -700 mV. The polarographic needle electrode (cathode) consists of a 12- $\mu$ m-diameter gold filament, which is embedded within a 300- $\mu$ m-diameter flexible stainless steel housing. The opening is

covered by an oxygen-permeable membrane. Electrical current is generated that is proportional to the tissue oxygen pressure at the tip of the electrode. Polarographic electrodes were calibrated before and after the measurements in phosphate buffered normal saline with 100% nitrogen and room temperature. Location of tumor, assessment of tumor size, and the depth from the skin surface to the peripheral edge of the tumor was determined using ultrasonography by a board certified mammographic radiologist (E.R.). After determining and marking the insertion site, skin overlying the site was cleansed with betadine and anesthetized with 2% lidocaine. Under direct visual control, a 16-gauge needle was inserted and placed at the edge of tumor. Insertion of the  $pO_2$  needle electrode was done under direct ultrasound guidance. The total measurement path was adjusted according to the tumor size; thus, the measurements were only done in tumor tissue. The probe was automatically advanced forward in steps of 0.7 mm and subsequent backward step of 0.3 mm with the net increments of 0.4 mm. A mean of 180 measurements was made per tumor. At the end of the measurement path, the probe was automatically withdrawn from the tissue. Hypoxia was defined as a median  $pO_2$  of <10 mm Hg.

**Determination of IBC status.** Patients had their initial presentation recorded as either having IBC or not. IBC was defined as erythema involving at least 30% of the breast and the presence of subcutaneous edema. In addition, pretreatment pathologic specimens (skin must have been present on the original tumor biopsy or a separate skin punch biopsy was done) were required to have evidence of dermal lymphatic tumor cell involvement.

**Determination of lymph node involvement.** All patients enrolled on the study had a formal pathologic review of their diagnostic biopsy as well as tissue obtained at the time of definitive surgery. Twenty of 37 evaluable patients had pathologic confirmation (FNA/core biopsy), and 10 of 37 evaluable patients had clinical confirmation of axillary involvement before the initiation of therapy. No patients underwent prechemotherapy axillary lymph node dissection. At the time of the diagnostic surgical procedure, all patients underwent a standard lymph node dissection with a median number of 12 (range, 9-25) lymph nodes removed. Each lymph node removed was examined using both H&E staining as well as cytokeratin staining (AE1/AE3; Zymed Laboratories, South San Francisco, CA; Cam 5.2; Becton Dickinson, Franklin Lakes, NJ).

**Statistical analysis.** Statistical analysis of the gene expression data evaluated binary logistic regression models to predict, in three separate analyses, the clinical states: IBC (yes/no), hypoxia (yes/no), and treatment response (yes/no). In each of the three analyses, very many individual binary regression models were generated and evaluated, each based on a set of genes selected from the filtered set of 24,134. The analysis evaluated large numbers of such individual regression models using stochastic search methods implemented on a cluster computer to rapidly search the space of such subsets. This analysis method has been

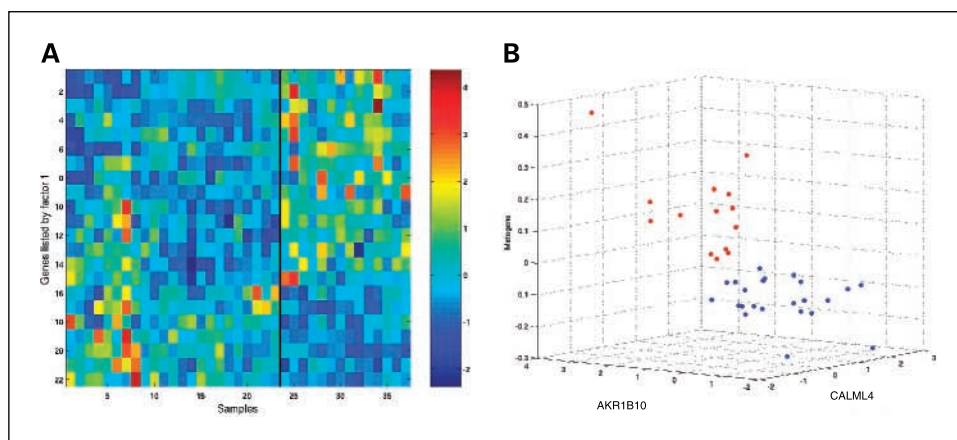
previously used in a similar study in brain cancer genomics in exploring subsets of gene expression predictors in a linear regression format (16). See also ref. (17) for statistical details. In each of the three analyses here, a number of regression models involving a small number of genes were identified. The Bayesian statistical analysis penalizes larger numbers heavily, to address both the need for parsimonious models when dealing with limited sample sizes, but also critically, to automatically and appropriately avoid the false discovery propensity when searching across so many potential predictive models due to the large number of genes available as candidate predictors. In each of the three studies, a resulting set of binary regression models was generated this way, each model having an associated approximate probability based on its fit to the data. The practical relevance of the analysis was evaluated by cross-validation prediction, repeatedly performing the analysis with each tumor held out, to define leave-one-out predictions of the outcomes. For each leave-one-out case, overall predictions are based on averaging across the set of regression models identified and weighed. Further analysis explored small sets of genes identified as relevant in the more highly scoring regression models for each of the three clinical outcomes. The resulting three sets of genes (reported in results) were also analyzed to define principal component (metagene) summaries of expression useful for visual presentation.

## Results

We have made use of tissue samples collected from patients enrolled in a phase I-II, open-label study of liposomal doxorubicin and paclitaxel, in combination with whole breast hyperthermia for the neoadjuvant treatment of locally advanced breast cancer (stage IIB or III), as an opportunity to identify gene expression profiles that reflect and predict response to therapy. The study provided the additional opportunity to explore two significant breast cancer phenotypes: tumor hypoxia and IBC.

**Gene expression profiles that characterize IBC.** Our initial focus was on the identification of genes that have the capacity to discriminate tumors with the IBC phenotype. Within the study population, a total of 37 patients were ascertained, and 14 were positive for the phenotype of IBC. Two of the 14 patients (14%) had expression of the estrogen receptor, and 8 of 14 (57%) had amplification of the *Her-2* gene. An expression profile that was selected to discriminate the IBC phenotype is depicted in Fig. 1A. Although there is considerable heterogeneity in the profiles, with noise and apparently only weak signatures gene by gene, the aggregate pattern that is

**Fig. 1.** Gene expression profiles that identify an inflammatory breast cancer phenotype. **A**, image intensity display of genes identified by top models. The vertical line demarks the non-IBC from IBC tumors. **B**, scatter plot depicting the classification of samples based on the expression patterns. The 37 samples were plotted according to the expression level of the two most highly weighted genes (*AKR1B10* and *CALML4*) in regression models predicting IBC. The metagene, represented on the vertical axis, represents the key predictive metagene underlying the set of 22 genes appearing in a thresholded selection of the top-scoring 100 regression models. IBC cases (red), non-IBC cases (blue).



visually apparent is predictive of phenotype as detailed below. In complex biological phenotyping problems, typified by a wide range of outcomes and states in human breast cancer, there is often little or no opportunity or relevance for single-gene or simple “fold-change” evaluations. Rather, the power of gene expression is in the aggregate patterns and the derivation of relevant, predictive summaries underlying these patterns, based on regression methods or other forms of analysis of sets of genes together. Indeed, the ability of these patterns to discriminate IBC tumors is illustrated by the scatter plot shown in Fig. 1B. This shows a scatter plot of the 37 breast tumor cases according to expression levels of the two most highly weighed genes (*AKR1B10* and *CALML4* are those two genes receiving the highest probability of inclusion in regression models predicting IBC) together with a metagene (the first principal component) underlying the set of 22 genes appearing in a thresholded selection of the top-scoring regression models (Table 2). IBC cases appear as red, and non-IBC cases appear as blue. Importantly, note how the metagene separates IBC from non-IBC cases. The overall validity of the set of regressions was evaluated by cross-validation prediction, where the analysis is repeatedly done in a leave-one-out context, with the tumor left out then being predicted based on the set of models defined

and weight by the analysis of the remaining samples. These validations show the capacity of the gene expression patterns to identify samples with the IBC phenotype (Supplementary Fig. 1).

An examination of the genes that constitute the profile classifying an IBC phenotype are enriched for those that encode stromal proteins including a variety of proteoglycans. This is clearly evident in the analysis of Gene Ontology categories enriched in the IBC signature (Supplementary Fig. 3). This enrichment in the classifier of the IBC phenotype would suggest a significant role for the expression of these stromal activities in eliciting the inflammatory response associated with this breast cancer phenotype.

**Gene expression profiles that characterize tumor hypoxia.** We next evaluated the gene expression data for evidence of patterns that might reflect the state of tumor hypoxia. Of the 34 samples for which O<sub>2</sub> measures were taken, 14 exhibited a hypoxia phenotype (defined as having a median pO<sub>2</sub> < 10 mm Hg). There was no significant relationship between the presence of hypoxia and estrogen receptor or *HER-2* status.

Genes that were identified in the analysis, which have the capacity to discriminate tumors with evidence of hypoxia, are depicted as an expression profile in Fig. 2A, and listed in Table 3. Similar to the IBC characterization, there is heterogeneity in the profiles, but there was also evidence of a pattern that distinguished the samples. Indeed, the ability of these patterns to discriminate hypoxic tumors is illustrated by the scatter plot shown in Fig. 2B. Scatter of the 34 breast tumor cases according to expression levels of the two most highly weighed genes (*SLIC1* and *EROL1* are those two genes receiving the highest probability of inclusion in regression models predicting hypoxia) together with the key predictive metagene shows clear separation of the samples. Once again, we evaluated the extent to which these patterns truly reflected the underlying distinction of hypoxia by performing leave-one-out cross-validations (Supplementary Fig. 1).

The genes that constitute the profile classifying the hypoxic phenotype include several that have previously been identified in hypoxia conditions: *ERO1L*, *CDC34*, *TRGV9*, *TOP1*, and *TRA*. Furthermore, an evaluation of Gene Ontology designations revealed an enrichment for genes involved in DNA replication (Supplementary Fig. 3), perhaps indicating the effect of hypoxia in the arrest of cell cycle progression.

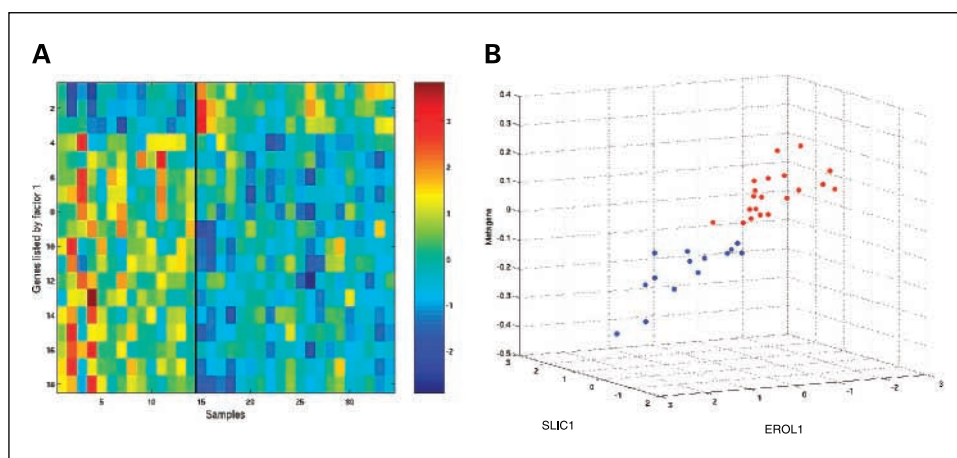
**Gene expression profiles that predict clinical response to neoadjuvant chemotherapy.** Although the identification of gene expression profiles reflecting inflammation or hypoxia provides an opportunity to explore the underlying biology of breast cancer, an ability to predict the ultimate clinical response to neoadjuvant therapy will be of most immediate significance. Response to therapy in breast cancer can be defined in many different ways. A clinical response is usually determined by either physical exam or radiological exam (mammogram, ultrasound, or magnetic resonance imaging). Either outcome measure is somewhat imprecise, allowing for interobserver variability as well inter-technique and intra-technique inconsistencies. A pathologic response can also be defined in a number of ways. Whether defined as the absence of microscopically detected invasive disease in either the breast alone or the breast and the axillary lymph nodes, a pathologic response has been shown in a number of studies to correlate with a favorable outcome (18–20). Pathologic response measures are

**Table 2.** Genes that characterize the phenotype of inflammatory breast cancer

Gene symbol	Gene title
<i>CALML4</i>	Calmodulin-like 4
<i>AKR1B10</i>	Aldo-keto reductase family 1, member B10
<i>RAB3D</i>	RAB3D, member RAS oncogene family
<i>CLGN</i>	Calmeglin
<i>AQP3</i>	Aquaporin 3
<i>PITPNC1</i>	Phosphatidylinositol transfer protein, cytoplasmic 1
<i>CKB</i>	Creatine kinase, brain
<i>CHST5</i>	Carbohydrate ( <i>N</i> -acetylglucosamine 6- <i>O</i> ) sulfotransferase 5
—	Similar to KIAA0563 gene product
<i>TFCP2L1</i>	Transcription factor CP2-like 1
<i>G1P2</i>	IFN, $\alpha$ -inducible protein (clone IFI-15K)
<i>PEX14</i>	Peroxisomal biogenesis factor 14
<i>GCHFR</i>	GTP cyclohydrolase I feedback regulator
<i>TAF6</i>	TAF6 RNA polymerase II, TATA box binding protein – associated factor
—	Ribosomal protein S4-like (RPS4L)
<i>RNF24</i>	Ring finger protein 24
<i>NCOR1</i>	Nuclear receptor co-repressor 1
<i>SULT1E1</i>	Sulfotransferase family 1E, estrogen-preferring, member 1
<i>SIAT7E</i>	Sialyltransferase 7
<i>ZNF496</i>	Zinc finger protein 496
—	EST
<i>CHST12</i>	Carbohydrate (chondroitin 4) sulfotransferase 12

NOTE: Genes are listed in order of their data-based posterior probabilities of inclusion in the regression models. The image in Fig. 1 is based on the metagene (first principal component or singular factor) of mean-corrected expression of the genes listed in the table.

**Fig. 2.** Gene expression profiles that identify a hypoxia phenotype. *A*, image intensity display of genes identified by top models predicting the hypoxic phenotype. *B*, scatter plot depicting the classification of samples based on the expression patterns. The 34 samples were plotted according to the expression level of the two most highly weighted genes (*SLIC1* and *EROL1*) in regression models predicting hypoxia. The metagene, represented on the vertical axis, represents the key predictive metagene underlying the set of 18 genes appearing in a thresholded selection of the top-scoring 100 regression models. Hypoxic cases (*red*), nonhypoxic case (*blue*).



useful as they do not suffer from the variability variables seen in determining a clinical response; however, to use pathologic response, a large number of patients are needed as complete pathologic responses with current treatments are uncommon.

We have sought to use the gene expression data to identify profiles predictive of a relevant pathologic response. Although we have also attempted to derive a signature predicting clinical response, it has not been possible to identify a clear pattern reflecting of this measure. In contrast, we have identified a pattern predictive of the persistence of positive axillary lymph nodes. Persistence of tumor in the lymph nodes has been shown in previous large studies to be the single most significant

prognostic factor in disease-free survival for breast cancer (15). A total of 36 samples were available from the study to allow the development of a predictor of lymph node persistence; of these, 27 were positive for lymph node involvement and nine were negative. There was no significant relationship between persistent lymph node involvement and estrogen receptor or *HER-2* status.

Similar to the analysis of IBC and hypoxia, a gene expression pattern was identified that could discriminate patient samples based on lymph node persistence (Fig. 3A). Again, there is heterogeneity in the profiles, but there was also evidence of a pattern that distinguished the samples (Fig. 3B). We also evaluated the extent to which these patterns truly reflected an ability to predict clinical outcome by performing leave-one-out cross-validations as shown in Fig. 4. The results are presented as the estimated probability that a given sample exhibits a pattern characteristic of a positive clinical response as measured by the absence of positive lymph nodes at the time of surgery. Although the analysis is limited in terms of numbers because there are only nine patients that failed to show a clinical response as seen by persistence of positive lymph nodes, the results do nevertheless indicate that gene expression data can be used to predict the clinical response.

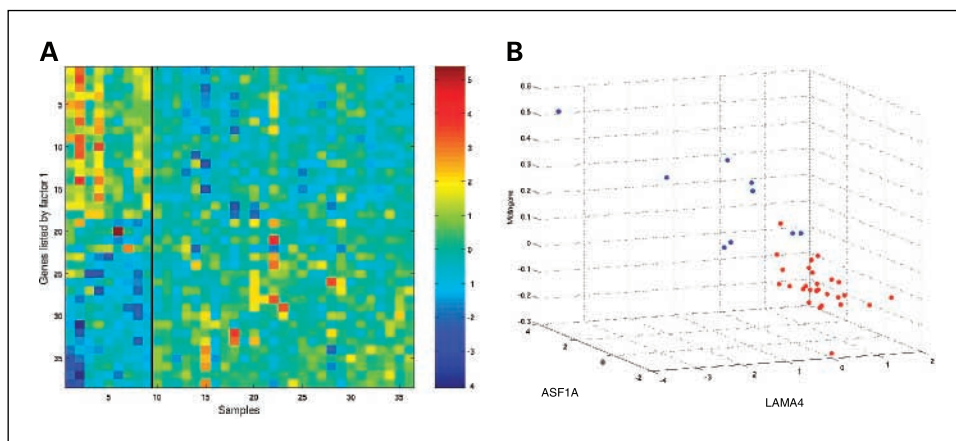
The genes that predict clinical response include some that have been previously linked to breast cancer outcome and metastasis, including *LAMA4*, *YY1*, and *RARA*. We would also expect that the other genes in the profile would likely make a contribution to the process and thus worth further study (Table 4).

*Lymph node persistence signature is a prognostic factor for overall survival.* Finally, as an additional measure of the prognostic significance of the gene expression patterns trained to predict chemotherapy response, we used the signature to predict the status of a large series of breast cancer samples for which Affymetrix data were available. This includes two cohorts of patients: a group of 158 breast cancer samples from Taiwanese patients (KF-SYSCC set; Supplementary Table 1), described in previous studies focused on developing predictors of recurrence (11, 12) and 101 samples from a group of Duke patients (Supplementary Table 2). For each tumor, we evaluated the probability that the clinical response signature is evident, thus generating groups of patients that were classified as positive or negative response. The survival characteristics of

**Table 3.** Genes that characterize the phenotype of tumor hypoxia

Gene symbol	Gene title
—	EST
<i>ERO1L</i>	ERO1-like ( <i>Saccharomyces cerevisiae</i> )
<i>PGRMC1</i>	Progesterone receptor membrane component 1
<i>HN1</i>	Hematologic and neurologic expressed 1
<i>HN1</i>	Hematologic and neurologic expressed 1
<i>CDC34</i>	Cell division cycle 34
<i>C14orf151</i>	Chromosome 14 open reading frame 151
—	EST
<i>ZNF135</i>	Zinc finger protein 135 (clone pHZ-17)
<i>RHOBTB3</i>	Rho-related BTB domain containing 3
<i>TRGV9</i>	T-cell receptor gamma variable 9
<i>FCRH3</i>	Fc receptor-like protein 3
<i>CST7</i>	Cystatin F (leukocystatin)
<i>TOP1</i>	Topoisomerase (DNA) I
—	EST
<i>TRA<math>\alpha</math></i>	Tcell receptor $\alpha$ locus
—	Olfactory receptor, family 7, subfamily A, member 126 pseudogene
<i>CAMTA1</i>	Calmodulin-binding transcription activator 1

NOTE: Genes are listed in order of their data-based posterior probabilities of inclusion in the regression models. The image in Fig. 2 is based on the metagene (first principal component or singular factor) of mean-corrected expression of the genes listed in the table.



**Fig. 3.** Gene expression profiles reflecting clinical response to neoadjuvant chemotherapy. *A*, image intensity display of genes identified by top models predicting clinical response. *B*, scatter plot depicting the classification of samples based on the expression patterns. The 36 samples were plotted according to the expression level of the two most highly weighted genes (*ASF1A* and *LAMA4*) in regression models predicting clinical response. The metagene, represented on the vertical axis, represents the key predictive metagene underlying the set of 38 genes appearing in a thresholded selection of the top-scoring 100 regression models. Positive response, red symbols; negative response, blue symbols.

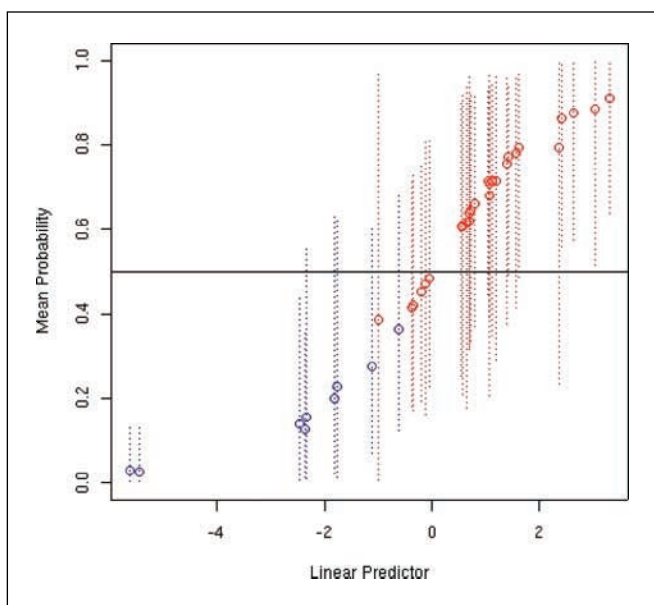
these patients were then examined by Kaplan-Meier analysis. As shown in Fig. 5, the chemotherapy response signature did identify a population of patients in both cohorts with reduced survival compared with those lacking the signature, consistent with previous studies showing that lymph node persistence following neoadjuvant chemotherapy is a significant prognostic factor. We conclude that the gene expression pattern that predicts response to neoadjuvant therapy does identify an aspect of underlying biology that is significant for clinical outcome.

### Discussion

The complexity and heterogeneity of breast cancer is substantial, challenging the capacity to develop an under-

standing of the unique attributes of a given tumor or patient that can guide the development of more effective therapeutic strategies. The ability to predict whether an individual patient will respond to a specific therapy can be of considerable value by stratifying patients to receive the most effective treatment based on their individual characteristics. Perhaps the best example can be seen from the analysis of *HER-2* amplification that can guide the selection of patients most likely to benefit from trastuzumab (Herceptin); likewise, the use of estrogen receptor status to guide the use of hormonal therapy is another example. In contrast, the ability to predict response to chemotherapy, whether in the adjuvant or neoadjuvant setting, is obviously more challenging than these examples of targeted therapeutic agents. Nevertheless, the principle of predicting response and selection of patients most likely to benefit from a given therapy can be extended to any circumstance where a given therapeutic shows effectiveness in only a fraction of the patients, even when the mechanism of action of the drug is not understood. The logic is to make use of the power of genomic data that can reflect subtle differences in tumors that uniquely define a positive or negative response to the drug. The data presented illustrate this approach, making use of complex gene expression data to identify a profile that can predict the response to neoadjuvant chemotherapy in breast cancer patients. Although the study was limited by the small cohort of patients available for analysis, the ability to show the prognostic significance for long-term survival based on the neoadjuvant response predictor, using two independent patient populations, does provide strong evidence for the value of this predictor to eventually stratify patients to the most effective treatment.

The value in this genomic approach is the ability to identify those patients most likely to benefit from a particular therapeutic strategy, in this case, the use of neoadjuvant chemotherapy. Other studies also point to the potential use of gene expression profiles as a mechanism to predict response to neoadjuvant chemotherapy (21, 22). Success in this strategy does raise the important additional question of how then to treat those patients predicted to be resistant to a given therapy. Clearly, the identification of characteristics unique to the resistant patient's tumors, which can drive the development of new therapeutics specific to these tumors, will be a critical part of the overall strategy that uses genomic information to guide therapeutic decisions.



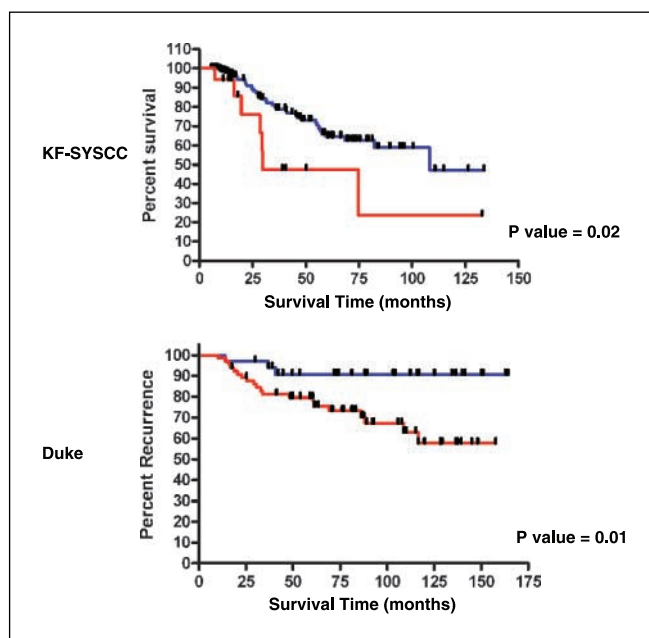
**Fig. 4.** Prediction of clinical response to neoadjuvant chemotherapy. Cross-validation predictions from the aggregate binary regression model for clinical response as measured by lymph node persistence. Each tumor is case plotted in terms of its predicted probability of clinical response (*red*) versus nonresponse (*blue*) based on the analysis of the remaining samples. These predicted cross-validation probabilities are located at the corresponding predicted metagene score from the average of multiple models (*horizontal axis*) with the vertical bar indicating an ~95% interval about the predicted probability.

Clearly, the ability to develop a more precise and detailed description of the molecular processes underlying breast cancer phenotypes will be critical to the development of more effective treatment strategies. Two relevant examples can be seen in

**Table 4.** Genes that predict the clinical response to neoadjuvant chemotherapy

Gene symbol	Gene title
<i>LAMA4</i>	Laminin, $\alpha$ 4
<i>ASF1A</i>	ASF1 anti-silencing function 1 homologue A ( <i>S. cerevisiae</i> )
<i>PPP2R5C</i>	Protein phosphatase 2, regulatory subunit B (B56), gamma isoform
<i>ANG</i>	Angiogenin, ribonuclease, RNase A family, 5
—	CDNA FLJ36638 fis
<i>ROD1</i>	ROD1 regulator of differentiation 1 ( <i>S. pombe</i> )
<i>FLJ10404</i>	Hypothetical protein FLJ10404
—	EST
<i>DKFZP434B0335</i>	DKFZP434B0335 protein
<i>MGC26717</i>	Hypothetical protein MGC26717
<i>UBE4B</i>	Ubiquitination factor E4B (UFD2 homolog, yeast)
<i>YY1</i>	YY1 transcription factor
—	Clone IMAGE:4871993, mRNA
<i>PPP2R5C</i>	Protein phosphatase 2, regulatory subunit B (B56), gamma isoform
<i>DR1</i>	Down-regulator of transcription 1, TBP-binding (negative cofactor 2)
<i>FLJ10359</i>	Protein BAP28
—	EST
<i>LYAR</i>	Hypothetical protein FLJ20425
<i>DKFZp434N2030</i>	Hypothetical protein DKFZp434N2030
<i>RARA</i>	Retinoic acid receptor, $\alpha$
<i>CELSR3</i>	Cadherin, EGF LAG seven-pass G-type receptor 3
<i>KPNA6</i>	Karyopherin $\alpha$ 6 (importin $\alpha$ 7)
<i>FLJ20516</i>	Timeless-interacting protein
—	EST
<i>HOXC6</i>	Homeobox C6
<i>LOC88523</i>	CG016
<i>EED</i>	Embryonic ectoderm development
<i>C14orf65</i>	Chromosome 14 open reading frame 65
—	EST
<i>NPDC1</i>	Neural proliferation, differentiation and control, 1
<i>C5orf14</i>	Chromosome 5 open reading frame 14
<i>RND1</i>	Rho family GTPase 1
<i>C10orf18</i>	Chromosome 10 open reading frame 18
<i>HOXD8</i>	Homeobox D8
<i>MTX3</i>	Metaxin 3
<i>C6orf66</i>	Chromosome 6 open reading frame 66
<i>TIMM22</i>	Translocase of inner mitochondrial membrane 22 homologue (yeast)
<i>FLJ39963</i>	Hypothetical protein FLJ39963

NOTE: Genes are listed in order of their data-based posterior probabilities of inclusion in the regression models. The image in Fig. 3 is based on the meta-gene (first principal component, or singular factor) of mean-corrected expression of the genes listed in the table.



**Fig. 5.** Signature for predicting response to neoadjuvant chemotherapy is a prognostic factor for breast cancer outcome. The response signature was used to predict overall survival in a set of 158 breast cancer samples from Taiwan and disease-free survival in 101 breast cancer samples from Duke University. In each instance, the predicted probability of the signature was used to identify individual patients exhibiting the phenotype. The survival characteristics of the two patient populations were then examined by Kaplan-Meier survival analysis.

the analysis of IBC and hypoxia as important breast cancer phenotypes. IBC, although rare, clearly defines a subclass of the disease with very poor prognosis. Likewise, previous studies have pointed to the hypoxic state of tumors as being a significant determinant of disease outcome. Although the understanding of the hypoxia response is now quite advanced, with components of the response pathway well defined, it remains largely unclear how this phenotype, as well as the inflammatory phenotype, contributes to disease outcome. The studies we present here are an initial attempt towards a better understanding of these phenotypes, making use of gene expression profiling as a mechanism to identify additional genes that contribute to the phenotype. An analysis of the genes that classify and predict these phenotypes reveals several that are logical components of an inflammatory or hypoxic response, but many additional genes were identified whose link to these processes are unclear. Nevertheless, it is this unbiased approach to the analysis of these phenotypes, driving the gene expression profiles to reflect these biological phenotypes, which represents the power of the genomic strategy. The principal limitation at this stage of genomic profiling is the lack of integrative biology knowledge: how different pathways and processes interact within the cell to mediate a response to therapy. Understanding these connections will be key to evaluating complex phenotypes, such as IBC and the hypoxic response. However, it is the identification of genes that associate with and define each phenotype as described here, which are essential first steps towards this understanding.

### Acknowledgments

We thank Kaye Culler for her assistance in the preparation of the article.

**References**

1. Low JA, Berman AW, Steinberg SM, Danforth DN, Lippman ME, Swain SM. Long-term follow-up for locally advanced inflammatory breast cancer patients treated with multimodality therapy. *J Clin Oncol* 2004;22:4067–74.
2. Goonewardene TI, Sowter HM, Harris AL. Hypoxia-induced pathways in breast cancer. *Microsc Res Tech* 2002;59:41–8.
3. Harris AL. Hypoxia: a key regulatory factor in tumour growth. *Nat Rev Cancer* 2002;2:38–47.
4. Brown NS, Bicknell R. Thymidine phosphorylase, 2-deoxy-D-ribose and angiogenesis. *Biochem J* 1998; 334:1–8.
5. Fyles A, Milosevic M, Hedley D, et al. Tumor hypoxia has independent predictor impact only in patients with node-negative cervix cancer. *J Clin Oncol* 2002;20: 680–7.
6. Fyles A, Voduc D, Syed A, Milosevic M, Pintilie M, Hill R. The effect of smoking on tumour oxygenation and treatment outcome in cervical cancer. *Clin Oncol* 2002;14:442–6.
7. Brizel DM, Sibley GS, Prosnitz LR, Scher RL, Dewhirst MW. Tumor hypoxia adversely affects the prognosis of carcinoma of the head and neck. *Int J Radiat Oncol Biol Phys* 1997;38:285–9.
8. Brizel DM, Rosner GL, Harrelson J, Prosnitz LR, Dewhirst MW. Pretreatment oxygenation profiles of human soft tissue sarcomas. *Int J Radiat Oncol Biol Phys* 1994;30:635–42.
9. Jones EL, Prosnitz LR, Dewhirst MW, et al. Thermochemoradiotherapy improves oxygenation in locally advanced breast cancer. *Clin Cancer Res* 2004;10: 4287–93.
10. Vujaskovic Z, Rosen EL, Blackwell KL, et al. Ultrasound guided pO<sub>2</sub> measurement of breast cancer reoxygenation after neoadjuvant chemotherapy and hyperthermia treatment. *Int J Hyperthermia* 2003;19: 498–506.
11. Huang E, Cheng SH, Dressman H, et al. Gene expression predictors of breast cancer outcomes. *Lancet* 2003;361:1590–6.
12. Pittman J, Huang E, Dressman H, et al. Models for individualized prediction of disease outcomes based on multiple gene expression patterns and clinical data. *Proc Natl Acad Sci U S A* 2004;101:8431–6.
13. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; 4:249–64.
14. Ihaka R, Gentleman R. A language for data analysis and graphics. *J Comput Graph Stat* 1996;5:299–314.
15. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003;19:185–93.
16. Rich J, Jones B, Hans C, et al. Gene expression profiling and genetic markers in glioblastoma survival. *Cancer Res* 2005;65:8869–77.
17. Hans C, Dobra A, West M. Shotgun stochastic search for regression with many candidate predictors. ISDS Discussion paper 2005.
18. Wolmark N, Wang J, Mamounas E, Bryant J, Fisher B. Preoperative chemotherapy in patients with operable breast cancer: nine-year results from National Surgical Adjuvant Breast and Bowel Project B-18. *J Natl Cancer Inst Monogr* 2001;30:96–102.
19. Cure H, Amat S, Penault-Llorca R, et al. Prognostic value of residual node involvement in operable breast cancer after induction chemotherapy. *Breast Cancer Res Treat* 2002;76:37–45.
20. Chollet P, Amat S, Cure H, et al. Prognostic significance of a complete pathological response after induction chemotherapy in operable breast cancer. *Br J Cancer* 2002;86:1041–6.
21. 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* 2003;362:362–9.
22. Ayers M, Symmans WF, Stec J, et al. Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J Clin Oncol* 2004;22:2284–93.



# Clinical Cancer Research

## Gene Expression Profiles of Multiple Breast Cancer Phenotypes and Response to Neoadjuvant Chemotherapy

Holly K. Dressman, Christopher Hans, Andrea Bild, et al.

*Clin Cancer Res* 2006;12:819-826.

<b>Updated version</b>	Access the most recent version of this article at: <a href="http://clincancerres.aacrjournals.org/content/12/3/819">http://clincancerres.aacrjournals.org/content/12/3/819</a>
<b>Supplementary Material</b>	Access the most recent supplemental material at: <a href="http://clincancerres.aacrjournals.org/content/suppl/2006/02/28/12.3.819.DC1">http://clincancerres.aacrjournals.org/content/suppl/2006/02/28/12.3.819.DC1</a>

<b>Cited articles</b>	This article cites 19 articles, 8 of which you can access for free at: <a href="http://clincancerres.aacrjournals.org/content/12/3/819.full#ref-list-1">http://clincancerres.aacrjournals.org/content/12/3/819.full#ref-list-1</a>
<b>Citing articles</b>	This article has been cited by 14 HighWire-hosted articles. Access the articles at: <a href="http://clincancerres.aacrjournals.org/content/12/3/819.full#related-urls">http://clincancerres.aacrjournals.org/content/12/3/819.full#related-urls</a>

<b>E-mail alerts</b>	<a href="#">Sign up to receive free email-alerts</a> related to this article or journal.
<b>Reprints and Subscriptions</b>	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a> .
<b>Permissions</b>	To request permission to re-use all or part of this article, contact the AACR Publications Department at <a href="mailto:permissions@aacr.org">permissions@aacr.org</a> .