Concordance among Gene-Expression–Based Predictors for Breast Cancer

Cheng Fan, M.S., Daniel S. Oh, Ph.D., Lodewyk Wessels, Ph.D., Britta Weigelt, Ph.D., Dimitry S.A. Nuyten, M.D., Andrew B. Nobel, Ph.D., Laura J. van’t Veer, Ph.D., and Charles M. Perou, Ph.D.

From the Departments of Genetics (C.F., D.S.O., C.M.P.), Statistics and Operations Research (A.B.N.), and Pathology and Laboratory Medicine (C.M.P.), University of North Carolina at Chapel Hill and Lineberger Comprehensive Cancer Center, Chapel Hill; and the Divisions of Diagnostic Oncology (L.W., B.W., L.J.V.) and Radiotherapy (D.S.A.N.), the Netherlands Cancer Institute, Amsterdam. Address reprint requests to Dr. Perou at Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Campus Box 7295, Chapel Hill, NC 27599, or at cperou@med.unc.edu.

Drs. Fan and Oh contributed equally to this article.

Copyright © 2006 Massachusetts Medical Society.

ABSTRACT

BACKGROUND
Gene-expression–profiling studies of primary breast tumors performed by different laboratories have resulted in the identification of a number of distinct prognostic profiles, or gene sets, with little overlap in terms of gene identity.

METHODS
To compare the predictions derived from these gene sets for individual samples, we obtained a single data set of 295 samples and applied five gene-expression–based models: intrinsic subtypes, 70-gene profile, wound response, recurrence score, and the two-gene ratio (for patients who had been treated with tamoxifen).

RESULTS
We found that most models had high rates of concordance in their outcome predictions for the individual samples. In particular, almost all tumors identified as having an intrinsic subtype of basal-like, HER2-positive and estrogen-receptor–negative, or luminal B (associated with a poor prognosis) were also classified as having a poor 70-gene profile, activated wound response, and high recurrence score. The 70-gene and recurrence-score models, which are beginning to be used in the clinical setting, showed 77 to 81 percent agreement in outcome classification.

CONCLUSIONS
Even though different gene sets were used for prognostication in patients with breast cancer, four of the five tested showed significant agreement in the outcome predictions for individual patients and are probably tracking a common set of biologic phenotypes.
Many studies of gene expression have identified expression profiles and gene sets that are prognostic, predictive, or both for patients with breast cancer. Comparisons of the lists of genes derived from some of these apparently similar studies show that they overlap only slightly, if at all. The reasons for this lower-than-expected overlap are not completely known, but they probably include differences in the patient cohorts, microarray platforms, and mathematical methods of analysis. An important and unanswered question, however, is whether these predictors are actually concordant with respect to their predictions for individual patients. Here, we describe our analysis of a single data set on which five prognostic or predictive gene-expression–based models were simultaneously compared.

**Methods**

**Patients**

We used a single data set of breast-cancer samples from 295 women. The gene-expression data set was derived by researchers from the Netherlands Cancer Institute and Rosetta Inpharmatics–Merck using oligonucleotide microarrays (Agilent). Data on relapse-free survival (defined as the time to a first event) and overall survival were available for all patients. Most of the patients had stage I or II breast cancer; 165 had received local therapy alone, 20 had received tamoxifen only, 20 had received tamoxifen plus chemotherapy, and 90 had received chemotherapy only.

**Statistical Analysis**

**Gene Sets**

We used five prognostic or predictive gene sets (and methods) to evaluate the data set. The resulting classifications for each patient were recorded for each model (Table 1 in the Supplementary Appendix, available with the full text of this article at www.nejm.org). The gene-expression–based profiles used were the 70-gene good-versus-poor outcome model developed by van de Vijver et al. and van’t Veer et al., the wound-response model developed by Chang et al., the recurrence-score model developed by Paik et al., the intrinsic-subtype model (luminal A, luminal B, basal-like, HER2-positive and estrogen-receptor–negative (HER2+ and ER–), and normal breast-like) developed by Perou and colleagues, and the two-gene–ratio model (the ratio of the levels of expression of homeobox 13 (HOXB13) and interleukin 17B receptor (IL17BR)). (The predictions for each model are presented in the Supplementary Appendix.) The recurrence-score and two-gene–ratio models were originally designed to predict the outcomes among patients with ER+ disease who were receiving tamoxifen. We therefore performed separate analyses for the subgroup of ER+ samples and for the complete set of ER+ and ER– samples combined. A detailed description of how these methods were applied to the 295-sample data set is provided in the Supplementary Appendix.

**Survival**

To evaluate the prognostic value of each gene-expression–based model, we performed univariate Kaplan–Meier analysis using the Cox–Mantel log-rank test in WinStat for Excel (R. Fitch Software). We also used SAS software to perform a multivariate Cox proportional-hazards analysis of each model individually in a model that included estrogen-receptor status (positive vs. negative), tumor grade (1 vs. 2 and 1 vs. 3), nodal status (no positive nodes vs. one to three positive nodes and no positive nodes vs. more than three positive nodes), age (as a continuous variable), tumor diameter (2 cm or less vs. more than 2 cm), and treatment received (no adjuvant therapy vs. chemotherapy, hormonal therapy, or both). Relapse-free survival (defined as the time to a first event) and overall survival were the end points. (For multivariate analysis of the intrinsic subtypes and recurrence score, estrogen-receptor status was not included as a variable because it was based on the same microarray data that were used in the gene-expression models).

Two-way contingency-table analyses and the calculation of Cramer’s V statistic were performed with WinStat for Excel. Cramer’s V statistic provides a quantitative measure of the strength of the association between the two variables in a contingency table (information that cannot be obtained from the P value). The values range from 0 to 1, with 0 indicating no relation and 1 indicating a perfect association. Traditionally, values of 0.36 to 0.49 indicate a substantial relation, and values of 0.50 or more indicate a strong relation. The V statistic is a generalization of the more
familiar phi statistic for non–two-by-two contingency tables, and for two-by-two tables, the V statistic is equal to the phi statistic.14

RESULTS

ANALYSIS OF ALL TUMORS

For all 295 tumors, all gene-expression–based models except the two-gene–ratio model, estrogen-receptor status, tumor grade, tumor diameter, and nodal status were significant predictors of relapse-free survival and overall survival, according to univariate Kaplan–Meier survival analyses (Fig. 1 and Table 1). For the four significant models, the groups with a poor outcome were as expected: those with a poor 70-gene profile, an activated wound response, a high recurrence score, and the basal-like, luminal B, and HER2+ and ER− intrinsic subtypes.

To evaluate the prognostic value of each gene-expression–based model, we next performed a multivariate Cox proportional-hazards analysis — that included estrogen-receptor status, tumor grade, nodal status, age, tumor diameter, and treatment status — of each model individually (Table 2 in the Supplementary Appendix). The models based on intrinsic subtype, 70-gene profile, wound response, and recurrence score were significant predictors of both relapse-free survival and overall survival. Thus, each gene-expression profile (except for the two-gene ratio) added new and important prognostic information beyond that provided by the standard clinical predictors. In fact, the 70-gene, recurrence-score and wound-response models were the most predictive variables in each analysis, as reflected by their having the lowest nominal P value.

As a point of reference, we next analyzed each model relative to the intrinsic-subtype assignments, which were largely based on an unsupervised analysis of breast-tumor gene-expression profiles (Table 2). All 53 basal-like tumors were classified as having a high recurrence score and a poor 70-gene profile, and 50 were classified as having an activated wound-response signature. A nearly identical finding was observed for the HER2+ and ER− subtype, as well as for the poor-outcome luminal B subtype that is defined clinically as ER+. Conversely, the normal-like and luminal A tumors showed heterogeneity in terms of how they were classified by the other models; however, 62 of 70 samples with low recurrence scores were of the luminal A subtype. These data suggest that if a sample is classified as basal-like, HER2+ and ER−, or luminal B, then it most likely would be in the poor-prognosis groups of the 70-gene, wound-response, and recurrence-score models.

We next compared the results of the 70-gene, wound-response, recurrence-score, and two-gene models with one another, using two-way contingency-table analyses. For these analyses, we combined the low and intermediate recurrence-score categories into a single group, because their survival curves were not significantly different (Table 2E in the Supplementary Appendix). All the comparisons yielded significant correlations, with the two-gene model having the lowest level of correlation. The results of the recurrence score, 70-gene, and wound-response models were all highly correlated (Table 3 in the Supplementary Appendix) (P<0.001 by the chi-square test).

We then assessed the strength of the correlation between the models using Cramer’s V statistic. Comparison of the 70-gene and recurrence-score models yielded a Cramer’s V statistic of 0.60 (indicating a strong relation), comparisons of the recurrence-score and wound-response models yielded a V statistic of 0.42 (indicating a substantial relation), and comparison of the 70-gene and wound-response models yielded a V statistic of 0.36 (indicating a substantial relation). Thus, most tumors classified as resulting in a poor outcome according to one of these three models were also classified as such by the other two. With regard to the Cramer’s V values, the model showing the best agreement with the other two was the recurrence score (i.e., of the three, recurrence score came the closest to functioning as a consensus predictor). To determine whether the use of the
Probability of Relapse-free Survival

- **A** Intrinsic Subtype
  - Basal-like
  - HER2+ and ER-
  - Luminal A
  - Luminal B
  - Normal-like

- **B** Intrinsic Subtype
  - Basal-like
  - HER2+ and ER-
  - Luminal A
  - Luminal B
  - Normal-like

Probability of Overall Survival

- **C** Recurrence Score
  - High
  - Intermediate
  - Low

- **D** Recurrence Score
  - High
  - Intermediate
  - Low

70-Gene Profile

- **E** Good
  - Poor

- **F** Good
  - Poor

Wound Response

- **G** Activated
  - Quiescent

- **H** Activated
  - Quiescent

Two-Gene Ratio

- **I** High
  - Low

- **J** High
  - Low
three models together would result in a better model than the use of any one alone, we derived a single model based on the most common findings of the three models. The performance of this model according to the Kaplan–Meier analysis was similar to that of each of the three models but was not noticeably better.

Histologic grade is an important clinical and biologic feature of tumors, especially in a comparison of the clinical characteristics of grade 1 and grade 3 breast tumors. An often-asked question regarding these gene-expression–based models is whether the predicted prognosis correlates with tumor grade. We therefore performed two-way contingency-table analyses comparing tumor grade and the results of each of four models (70-gene, wound-response, two-gene ratio, and recurrence score [low plus intermediate vs. high]). All four models showed significant correlations with grade (P<0.001). The 70-gene model was the most highly correlated with grade (Cramer’s V statistic, 0.52), followed by recurrence score (V statistic, 0.48), wound response (V statistic, 0.35), and the two-gene ratio (V statistic, 0.25).

Thus, to varying degrees, all the models correlated with grade, but the 70-gene, recurrence-score, intrinsic-subtype, and wound-response models added prognostic information beyond that provided by the tumor grade. Moreover, the use of these four models involved an assay that is objective and quantitative and could be automated and easily standardized across institutions.

Of the five models, the 70-gene\(^2,3\) and recurrence-score\(^6,15\) models are the most well validated and are beginning to be used in the clinical setting to assist in treatment decisions. We therefore specifically compared these two models in a group of 295 patients with cancer, using a simple method. We considered low and intermediate recurrence scores to be equivalent to a good score on the 70-gene model and a high recurrence score to be equivalent to a poor score on the 70-gene model and then determined how many scores agreed between the two models. We observed agreement in 239 of 295 samples (81 percent). In particular, 81 of the 103 samples with a recurrence score of low or intermediate were classified as having a good 70-gene profile.

### Table 1. Classification of the Netherlands Cancer Institute Patient Data Set According to Five Gene-Expression–Based Models.

<table>
<thead>
<tr>
<th>Classification</th>
<th>295-Sample Data Set</th>
<th>ER+ 225-Sample Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>number (percent)</td>
<td>number (percent)</td>
</tr>
<tr>
<td><strong>Intrinsic subtype</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luminal A</td>
<td>123 (41.7)</td>
<td>121 (53.8)</td>
</tr>
<tr>
<td>Luminal B</td>
<td>55 (18.6)</td>
<td>55 (24.4)</td>
</tr>
<tr>
<td>Normal-like</td>
<td>29 (9.8)</td>
<td>24 (10.7)</td>
</tr>
<tr>
<td>HER2+ and ER−</td>
<td>35 (11.9)</td>
<td>18 (8.0)</td>
</tr>
<tr>
<td>Basal-like</td>
<td>53 (18.0)</td>
<td>7 (3.1)</td>
</tr>
<tr>
<td><strong>Recurrence score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>70 (23.7)</td>
<td>87 (38.7)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>33 (11.2)</td>
<td>18 (8.0)</td>
</tr>
<tr>
<td>High</td>
<td>192 (65.1)</td>
<td>120 (53.3)</td>
</tr>
<tr>
<td><strong>70-Gene profile</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>115 (39.0)</td>
<td>113 (50.2)</td>
</tr>
<tr>
<td>Poor</td>
<td>180 (61.0)</td>
<td>112 (49.8)</td>
</tr>
<tr>
<td><strong>Wound response</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quiescent</td>
<td>67 (22.7)</td>
<td>60 (26.7)</td>
</tr>
<tr>
<td>Activated</td>
<td>228 (77.3)</td>
<td>165 (73.3)</td>
</tr>
<tr>
<td><strong>Two-gene ratio</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>137 (46.4)</td>
<td>122 (54.2)</td>
</tr>
<tr>
<td>High</td>
<td>158 (53.6)</td>
<td>103 (45.8)</td>
</tr>
</tbody>
</table>
In this analysis, we compared the capacity of each model to predict recurrence in a group of patients with either node-negative or node-positive tumors and with or without adjuvant chemotherapy. However, the profiles were developed to predict the distant metastasis-free survival among patients with node-negative disease only, and they are meant to be used either to predict prognosis without adjuvant treatment (70-gene predictor) or with the use of tamoxifen (recurrence score).

### Analysis of Estrogen-Receptor–Positive Tumors

Two of the five models (recurrence score and two-gene ratio) were specifically designed to evaluate outcomes in patients with ER+ tumors who were treated with tamoxifen. We therefore performed the same analyses described above (Table 1) on the 225 samples in the 295-sample data set that were classified as ER+ on the basis of the level of expression of the estrogen-receptor gene. Again, all the gene-expression–based models (except for the two-gene ratio) were significant predictors of relapse-free survival and overall survival in univariate Kaplan–Meier analyses (Fig. 2). In multivariate Cox proportional-hazards analyses (in which each model was evaluated individually in a model that included the standard clinical variables), the 70-gene, wound-response, and recurrence-score models and the luminal A and B intrinsic subtypes added considerable prognostic information regarding relapse-free survival and overall survival; each gene-expression–based model typically had the lowest P value as compared with the traditional clinical variables (Table 4 in the Supplementary Appendix). The ER+ samples were also classified according to intrinsic subtype (Table 3); 7 were classified as basal-like and 18 as HER2+ and ER−, suggesting that approximately 10 percent of the ER+ tumors could be considered ER−, according to hierarchical clustering analysis.

As for the 295-sample data set, we performed a pairwise comparison of the 70-gene, wound-response, recurrence-score, and two-gene ratio assignments for the 225 ER+ samples, using two-way contingency-table analyses. All comparisons yielded significant correlations except for the two-gene model (Table 5 in the Supplementary Appendix). The recurrence-score, 70-gene, and wound-response profiles were highly correlated ($P<0.001$); the Cramer’s V values were 0.54 for the 70-gene model as compared with the recurrence-score model, 0.38 for the recurrence-score model as compared with the wound-response model,

<table>
<thead>
<tr>
<th>Intrinsic Subtype</th>
<th>No. of Patients</th>
<th>Recurrence Score</th>
<th>70-Gene Profile</th>
<th>Wound Response</th>
<th>Two-Gene Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Classification</td>
<td>No. of Patients</td>
<td>Classification</td>
<td>No. of Patients</td>
</tr>
<tr>
<td>Basal-like</td>
<td>53</td>
<td>Low</td>
<td>0</td>
<td>Good</td>
<td>0</td>
</tr>
<tr>
<td>Intermediate</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>53</td>
<td>Poor</td>
<td>53</td>
<td>Activated</td>
<td>50</td>
</tr>
<tr>
<td>Luminal A</td>
<td>123</td>
<td>Low</td>
<td>62</td>
<td>Good</td>
<td>87</td>
</tr>
<tr>
<td>Intermediate</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>36</td>
<td>Poor</td>
<td>36</td>
<td>Activated</td>
<td>78</td>
</tr>
<tr>
<td>Luminal B</td>
<td>55</td>
<td>Low</td>
<td>1</td>
<td>Good</td>
<td>9</td>
</tr>
<tr>
<td>Intermediate</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>50</td>
<td>Poor</td>
<td>46</td>
<td>Activated</td>
<td>51</td>
</tr>
<tr>
<td>HER2+ and ER−</td>
<td>35</td>
<td>Low</td>
<td>0</td>
<td>Good</td>
<td>3</td>
</tr>
<tr>
<td>Intermediate</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>35</td>
<td>Poor</td>
<td>32</td>
<td>Activated</td>
<td>35</td>
</tr>
<tr>
<td>Normal-like</td>
<td>29</td>
<td>Low</td>
<td>7</td>
<td>Good</td>
<td>16</td>
</tr>
<tr>
<td>Intermediate</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>18</td>
<td>Poor</td>
<td>13</td>
<td>Activated</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 2. Classification of Tumor Samples from All 295 Patients, According to the Model Used.
and 0.34 for the 70-gene model as compared with the wound-response model. Thus, recurrence score showed the best agreement with the other two models. We again derived a model based on the most common results for the three models, and its performance in Kaplan–Meier analysis was similar to that of the three individual models.

When the recurrence scores were compared with the 70-gene profile scores for the 225-sample subgroup as they were for the complete data set, 173 of the 225 samples (76.9 percent) showed agreement. In particular, of the 105 samples with low or intermediate recurrence scores, 83 were classified as having a good 70-gene profile.

We did not perform any multivariate Cox proportional-hazards analyses using all predictors simultaneously to identify the optimal model for either the 225-patient group or the 295-patient group. We believed that doing so would not be a fair test for any model for which this group was a true test set (recurrence score and two-gene ratio) or for those that were developed with the use of a different platform (recurrence score, two-gene ratio, and intrinsic subtype).

**Discussion**

We analyzed a single data set for which enough genes had been assayed to allow the simultaneous analysis of five gene-expression–based models. Four of these models resulted in similar predictions — for example, each model assigned the same samples to the poor-outcome groups. Tumors classified as basal-like, HER2+ and ER−, and luminal B by the intrinsic-subtype model were almost all classified as having a poor outcome (regardless of estrogen-receptor status) by the 70-gene, recurrence-score, and wound-response models. Only within the luminal A and normal-like intrinsic subtypes was variability in the outcome predictions found.

Of the five models analyzed in our study, only the two-gene ratio failed to identify significant differences in outcome within the data set. In an independent data set of patients with ER+ disease who were receiving tamoxifen, Reid et al. reported that the two-gene model failed to detect differences in outcome. However, Goetz et al. showed that in women with node-negative disease from the North Central Cancer Treatment Group Study 89-30-52, the two-gene ratio was a significant predictor of relapse-free survival and disease-free survival. A model based on the analysis of only two genes is much more likely to be sensitive to technical differences in analysis platforms than one based on many genes, and it is possible that one of the features representing HOXB13 or IL17BR in the Netherlands Cancer Institute data set may not faithfully reflect the values seen by Ma et al., owing to alternative splicing or differences in probe-hybridization conditions.

Pairwise comparisons of the 70-gene, wound-response, recurrence-score, and two-gene models showed that the results of all but the two-gene model were highly concordant. Comparison of the 70-gene and recurrence-score models showed that their sample predictions agreed in 77 percent of patients with ER+ cancer and 81 percent of all patients. These analyses suggest that even though there was very little gene overlap (the 70-gene and recurrence-score profiles overlapped by only 1 gene: SCUBE2) and different algorithms were used, the outcome predictions for the majority of patients with breast cancer would be similar. It is also likely that the recurrence-score model, originally developed for patients with ER+ disease, is accurate for all patients with breast cancer, because almost all (69 of 70) patients with ER− tumors were classified as having a high recurrence score.

The outcome predictions derived from the various models largely overlapped, according to multivariate Cox proportional-hazards analyses (the 95 percent confidence intervals of the hazard ratios for relapse-free and overall survival are given in Table 2 in the Supplementary Appendix). The discordance rate of up to 20 percent among the patients in different categories led to slight differences in outcome prediction and emphasiz-
CONCORDANCE AMONG GENE-EXPRESSION–BASED PREDICTORS FOR BREAST CANCER

A Intrinsic Subtype

B Intrinsic Subtype

C Recurrence Score

D Recurrence Score

E 70-Gene Profile

F 70-Gene Profile

G Wound Response

H Wound Response

I Two-Gene Ratio

J Two-Gene Ratio
The need for further validation of this approach. The National Cancer Institute and the European Union have designed randomized clinical trials (Trial Assigning Individualized Options for Treatment (Rx) [TAILORx] and Translating Molecular Knowledge into Early Breast Cancer Management Building on the Breast International Group network for Improved Treatment Tailoring [TRANSBIG]-Microarray in Node-Negative Disease May Avoid Chemotherapy [MINDACT], respectively) that will prospectively address the prognostic and predictive powers of the recurrence-score and 70-gene models, respectively.

Despite the absence of gene overlap, the different gene models yielded similar predictions largely because they reflected common cellular phenotypes, which encompass the consistent differences in ER+ (i.e., luminal) breast cancer and ER− (basal-like and HER2+ and ER−) breast cancers. Although these differences are correlated with histologic grade, it is clear that these profiles provided additional information beyond that provided by grade. Our findings also show that outcomes can readily be predicted by a large number of genes and that any randomly selected subgroup that is sufficiently large (approximately 100 genes) reproduces the hierarchical clustering obtained with the use of the full gene set.18

We conclude that overlap in gene identity among gene-expression profiles is not a good measure of reproducibility and that the classification of individual samples is the relevant measure of concordance. Our results are encouraging and can be interpreted to mean that although different gene sets are being used as predictors, they each track a common set of biologic characteristics that are present in different groups of patients with breast cancer, resulting in similar predictions of outcome.

Supported by grants from the National Cancer Institute (RO1-CA-101227-01, to Dr. Perou), the National Cancer Institute Breast Specialized Programs of Research Excellence program (P50-CA58223-09A1, to the University of North Carolina at Chapel Hill), the Breast Cancer Research Foundation, the Dutch Cancer Society (DCS-NKI02-2575, to Dr. van’t Veer), the Dutch National Genomics Initiative (NGI02-01, to the Cancer Genomics Center), and the National Science Foundation (DMS 0406361, to Dr. Nobel).


Dr. van’t Veer reports holding equity in Agendia BV. No other potential conflict of interest relevant to this article was reported.

We are indebted to Lisa Carey and Melissa A. Troester for reading and commenting on the manuscript.

Table 3. Classification of Tumor Samples from the 225 Patients with ER+ Disease, According to the Model Used.

<table>
<thead>
<tr>
<th>Intrinsic Subtype</th>
<th>No. of Patients</th>
<th>Recurrence Score</th>
<th>70-Gene Profile</th>
<th>Wound Response</th>
<th>Two-Gene Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Classification</td>
<td>Classification</td>
<td>Classification</td>
<td>Classification</td>
</tr>
<tr>
<td>Basal-like</td>
<td>7</td>
<td>Low</td>
<td>Good</td>
<td>Quiescent</td>
<td>Low</td>
</tr>
<tr>
<td>Intermediate</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luminal A</td>
<td>121</td>
<td>Low</td>
<td>Good</td>
<td>Quiescent</td>
<td>Low</td>
</tr>
<tr>
<td>Intermediate</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luminal B</td>
<td>55</td>
<td>Low</td>
<td>Good</td>
<td>Quiescent</td>
<td>Low</td>
</tr>
<tr>
<td>Intermediate</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HER2+ and ER−</td>
<td>18</td>
<td>Low</td>
<td>Good</td>
<td>Quiescent</td>
<td>Low</td>
</tr>
<tr>
<td>Intermediate</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal-like</td>
<td>24</td>
<td>Low</td>
<td>Good</td>
<td>Quiescent</td>
<td>Low</td>
</tr>
<tr>
<td>Intermediate</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

expressed in various tissues and that any randomly selected subgroup that is sufficiently large (approximately 100 genes) reproduces the hierarchical clustering obtained with the use of the full gene set.18

We conclude that overlap in gene identity among gene-expression profiles is not a good measure of reproducibility and that the classification of individual samples is the relevant measure of concordance. Our results are encouraging and can be interpreted to mean that although different gene sets are being used as predictors, they each track a common set of biologic characteristics that are present in different groups of patients with breast cancer, resulting in similar predictions of outcome.

Supported by grants from the National Cancer Institute (RO1-CA-101227-01, to Dr. Perou), the National Cancer Institute Breast Specialized Programs of Research Excellence program (P50-CA58223-09A1, to the University of North Carolina at Chapel Hill), the Breast Cancer Research Foundation, the Dutch Cancer Society (DCS-NKI02-2575, to Dr. van’t Veer), the Dutch National Genomics Initiative (NGI02-01, to the Cancer Genomics Center), and the National Science Foundation (DMS 0406361, to Dr. Nobel).


Dr. van’t Veer reports holding equity in Agendia BV. No other potential conflict of interest relevant to this article was reported.

We are indebted to Lisa Carey and Melissa A. Troester for reading and commenting on the manuscript.
CONCORDANCE AMONG GENE-EXPRESSION–BASED PREDICTORS FOR BREAST CANCER

REFERENCES


Copyright © 2006 Massachusetts Medical Society.

CLINICAL TRIAL REGISTRATION

The Journal encourages investigators to register their clinical trials in a public trials registry. The members of the International Committee of Medical Journal Editors plan to consider clinical trials for publication only if they have been registered (see N Engl J Med 2004;351:1250-1). The National Library of Medicine’s www.clinicaltrials.gov is a free registry, open to all investigators, that meets the committee’s requirements.
Molecular Signatures Predict Outcomes of Breast Cancer
Joyce A. O’Shaughnessy, M.D.

Breast cancer is classified and managed largely on the basis of anatomy — in contrast with lymphoma, which has been classified and treated according to grade for more than 20 years. Tumor size and the degree of involvement of the axillary nodes are used to estimate the risk of systemic micrometastases at diagnosis and, accordingly, whether systemic adjuvant therapy, which improves overall survival in largely unselected populations, is needed.\(^1\)

A routine question faced by oncologists is, which of the two thirds of patients with hormone-receptor–positive breast cancer require systemic adjuvant chemotherapy to decrease their chance of recurrence? Although there are substantial differences in the prognosis and natural history between histologically defined low-grade and high-grade breast cancers that express hormone receptors, national consensus guidelines currently recommend the consideration of adjuvant chemotherapy for estrogen receptor (ER)-positive, node-negative tumors that are more than 1 cm in diameter.\(^2\) However, retrospective analyses suggest that adjuvant chemotherapy does not benefit patients with highly ER–positive breast cancer (regardless of nodal status), whereas it does appear to benefit patients with lower levels of ER expression.\(^3,4\) This finding suggests that biology trumps anatomy in the determination of prognosis and the benefit of chemotherapy.

Accordingly, a sea change is under way with the results of clinical trials. J Clin Epidemiol 1997;50:1089-98.


Copyright © 2006 Massachusetts Medical Society.
with a poor prognosis demonstrates overexpression of genes regulating the cell cycle, invasion, metastasis, and angiogenesis.\textsuperscript{8,9}

A third predictor calculates a recurrence score on the basis of the expression of 21 genes, with the use of reverse transcriptase–polymerase chain reaction (RT-PCR) in formalin-fixed, paraffin-embedded tissue. The predictor separates node-negative, ER-positive breast cancers into categories of high risk, intermediate risk, and low risk of recurrence.\textsuperscript{10,11}

A fourth predictor, based on a wound-response gene-expression signature derived from the transscriptional response of normal fibroblasts to serum in cell culture, has also been shown to improve the risk stratification of early breast cancer over that provided by standard clinicopathological features, in that the development of distant metastases is more likely among patients whose breast cancers have activated pathways for matrix remodeling, cell motility, and angiogenesis than among those whose cancers do not.\textsuperscript{12}

A fifth predictor uses a ratio of the levels of expression of two genes, one encoding homeobox 13 and the other encoding the interleukin-17B receptor. This predictor, which is based on assays using RT-PCR in formalin-fixed, paraffin-embedded tissue, was developed to determine the risk of recurrence in women with node-negative, ER–positive breast cancers who had received treatment with tamoxifen.\textsuperscript{13}

These diagnostic advances have galvanized the international breast-cancer research community and have led to the launch of the Microarray in Node-Negative Disease May Avoid Chemotherapy (MINDACT) and the Trial Assigning Individualized Options for Treatment (Rx) (TAILORx) studies. These trials will use the 70-gene profile and the recurrence score, respectively, to determine prospectively which patients with ER-positive, node-negative breast cancer benefit from adjuvant chemotherapy and which patients have a risk of recurrence sufficiently low that chemotherapy is unlikely to change their outcome. The results will probably alter standard medical practice such that, in the future, 30 to 50 percent fewer patients with ER-positive breast cancer will receive adjuvant chemotherapy.

In this issue of the Journal, Fan et al.\textsuperscript{14} report on the extent to which the five predictors are concordant in their classification of the risk of recurrence. They applied the predictors to a single data set that included both gene-expression data and clinical-outcome data for 295 patients. Four of the predictors were highly concordant in the prediction of recurrence and death. The predictor based on the two-gene ratio was not concordant with the other predictors; however, it was designed to predict the benefit from tamoxifen rather than to establish the prognosis for patients with ER–positive disease who had received loco therapy only, and only 40 patients in the data set had received tamoxifen.

The study was limited by its inclusion in the 295-patient data set of both patients who had received loco therapy only — whose prognosis can be clearly discerned — and patients who had received tamoxifen, chemotherapy, or both — whose natural history was potentially perturbed by one or more interventions. Thus, the ability of the gene-expression assays to predict prognosis was somewhat confounded. Another limitation was that a subgroup of the gene-expression data, derived from the 295 patients to test the prognostic power of the multigene assays, was first used as the training set for the intrinsic-subtype, 70-gene, and wound-response predictors (i.e., the set in which gene-expression cutoff points were selected). As the authors point out, having the training set embedded within the test set positively biases the performance of the predictive assays in their estimates of the recurrence-free survival in multivariate analyses.

To what extent are these concordant gene-expression predictors useful in the management of early-stage breast cancer? Do they add value over that provided by standard prognostic factors and factors predictive of the response to treatment? There was excellent concordance among the predictors in the identification of patients at high risk for recurrence, such that each predictor indicated a poor prognosis for almost all the patients with ER-negative, HER2–positive, or ER-positive and high-grade cancers. Most patients with these high-grade cancers associated with an elevated risk of recurrence routinely receive adjuvant chemotherapy, and multigene classifiers are therefore not needed to identify them. Moreover, the multivariate analyses conducted by Fan et al. showed that the gene-expression assays had a prognostic value independent of that of some standard prognostic factors, including grade. However, the assays did not include quantitative assessments of ER status and progesterone-receptor status, or an evaluation of the HER2 status, mitotic rate, or presence of lympho-
vascular invasion — other tumor characteristics that are available on routine histopathological assessment and that provide important prognostic information (especially for intermediate-grade breast cancers, whose natural history is the most variable within-grade). At present, therefore, it is not clear that the quantification of the level of expression of dozens or hundreds of genes provides more information about the potential of a cancer for metastasis, virulence,15 and response to therapy for an individual patient than does an optimal analysis of the standard and readily available histopathological prognostic factors.

However, the final judgment about the clinical usefulness of gene-expression profiling may ultimately be practical: if the recommendation for potentially life-saving adjuvant endocrine therapy, chemotherapy, or both is to be based on biologic factors, then the assessment of those factors must be reproducible and reliable. The literature is replete with documentation that assessments of hormone-receptor and HER2 status are highly variable, with substantial rates of false negative and false positive results, and that expert pathologists disagree in their assignments of breast-cancer grade, especially for intermediate-grade cancers. Gene-expression assays that can be performed on formalin-fixed, paraffin-embedded tissue and that provide highly reproducible prognostic information in prospective studies will be of great clinical utility, even if they do not independently predict the prognosis in multivariate analyses that include all the standard clinicopathological prognostic factors. To this end, the high degree of prognostic concordance among four of the predictors, including the recurrence score, is encouraging — especially because the recurrence-score predictor uses formalin-fixed, paraffin-embedded tissue (thus avoiding the need for the processing and storing of fresh tissue) and RT-PCR (an established, specific, and reproducible technique with a wide dynamic range).

Perhaps most important, molecular-expression profiles have the potential to identify the dominant growth and survival networks (networks of proteins in the breast-cancer cell that enable its growth and survival) assessed by the various predictive assays. There is a great, unmet need in the treatment of ER-negative and HER2-negative basal cancers and ER-positive, high-grade cancers, because the major molecular networks that sustain these cancers are not yet known. With the abundance of molecularly targeted inhibitors available either commercially or within clinical trials, the identification of these survival networks is an urgent research priority.

Dr. O’Shaughnessy reports having received consulting fees from the Molecular Profiling Institute. No other potential conflict of interest relevant to this article was reported.

From the Baylor Sammons Cancer Center, Texas Oncology, and US Oncology — all in Dallas.


Copyright © 2006 Massachusetts Medical Society.