ORIGINAL ARTICLE

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.

ABSTRACT

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.

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.

M ANY STUDIES OF GENE EXPRESSION have identified expression profiles and gene sets that are prognostic, predictive, or both for patients with breast cancer.[1-12] 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.[2-4] The clinical information was obtained from Chang et al.[5] 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.,[2,3] the wound-response model developed by Chang et al.,[4,5] the recurrence-score model developed by Paik et al.,[6] 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,[1,9,12,13] and the two-gene–ratio model (the ratio of the levels of expression of homeobox 13 [HOXB13] and interleukin 17B receptor [IL17BR]).[7] (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.[6,7] 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 intrinsic-subtype profiles 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
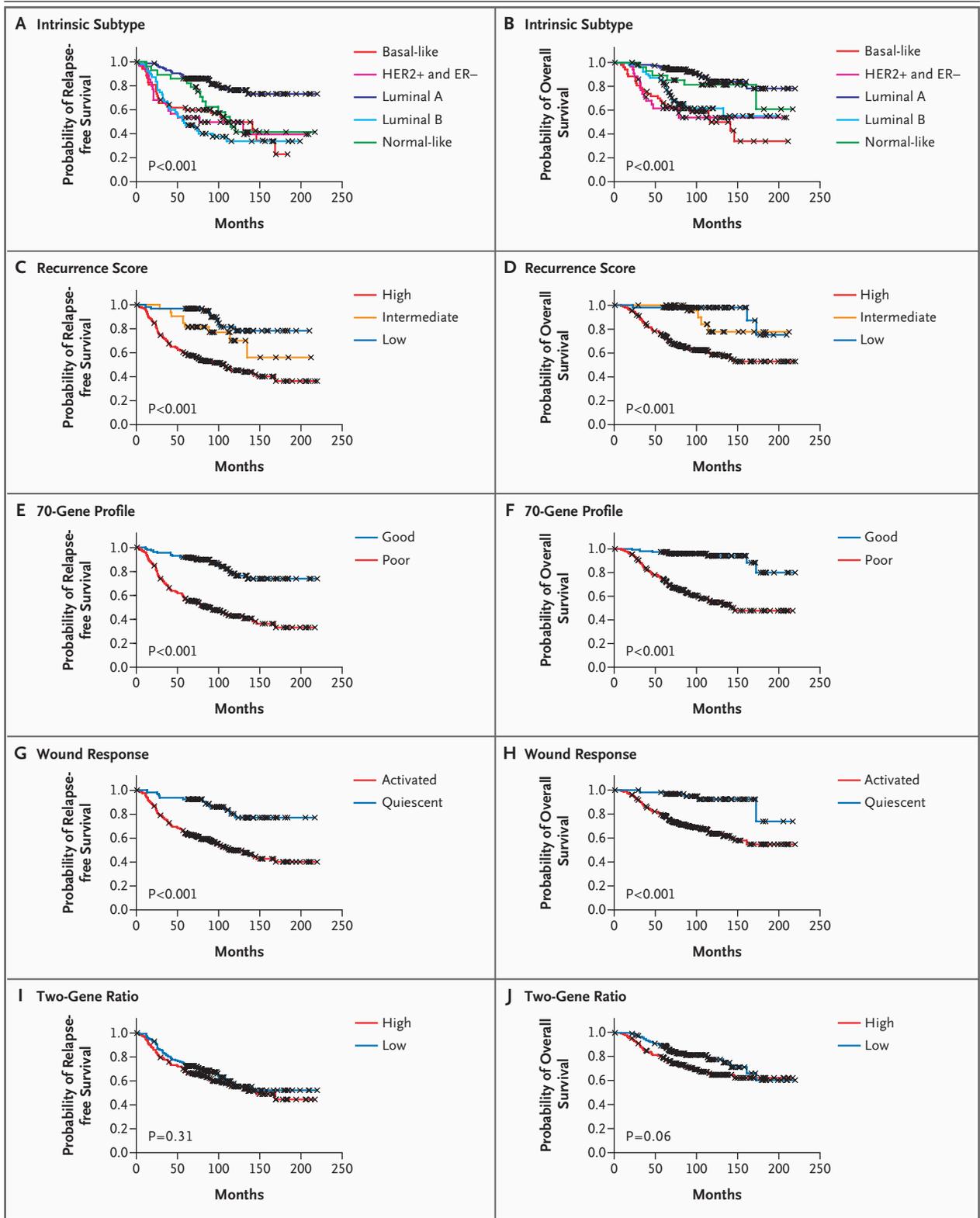
Figure 1 (facing page). Kaplan–Meier Survival Estimates of Relapse-free Survival and Overall Survival among the 295 Patients, According to the Intrinsic Subtype (Panels A and B), Recurrence Score (Panels C and D), 70-Gene Profile (Panels E and F), Wound Response (Panels G and H), and Two-Gene Ratio (Panels I and J).

P values were obtained from the log-rank test. X denotes observations that were censored owing to loss to follow-up or on the date of the last contact.

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

Classification	295-Sample Data Set	ER+ 225-Sample Data Set
	number (percent)	
Intrinsic subtype		
Luminal A	123 (41.7)	121 (53.8)
Luminal B	55 (18.6)	55 (24.4)
Normal-like	29 (9.8)	24 (10.7)
HER2+ and ER–	35 (11.9)	18 (8.0)
Basal-like	53 (18.0)	7 (3.1)
Recurrence score		
Low	70 (23.7)	87 (38.7)
Intermediate	33 (11.2)	18 (8.0)
High	192 (65.1)	120 (53.3)
70-Gene profile		
Good	115 (39.0)	113 (50.2)
Poor	180 (61.0)	112 (49.8)
Wound response		
Quiescent	67 (22.7)	60 (26.7)
Activated	228 (77.3)	165 (73.3)
Two-gene ratio		
Low	137 (46.4)	122 (54.2)
High	158 (53.6)	103 (45.8)

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 2. Classification of Tumor Samples from All 295 Patients, According to the Model Used.

Intrinsic Subtype	No. of Patients	Recurrence Score		70-Gene Profile		Wound Response		Two-Gene Ratio	
		Classification	No. of Patients	Classification	No. of Patients	Classification	No. of Patients	Classification	No. of Patients
Basal-like	53	Low	0	Good	0	Quiescent	3	Low	11
		Intermediate	0						
		High	53	Poor	53	Activated	50	High	42
Luminal A	123	Low	62	Good	87	Quiescent	45	Low	78
		Intermediate	25						
		High	36	Poor	36	Activated	78	High	45
Luminal B	55	Low	1	Good	9	Quiescent	4	Low	30
		Intermediate	4						
		High	50	Poor	46	Activated	51	High	25
HER2+ and ER–	35	Low	0	Good	3	Quiescent	0	Low	7
		Intermediate	0						
		High	35	Poor	32	Activated	35	High	28
Normal-like	29	Low	7	Good	16	Quiescent	15	Low	11
		Intermediate	4						
		High	18	Poor	13	Activated	14	High	18

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.[4] 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,

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.[16] However, Goetz et al. showed that in women with node-negative disease from the North Central Cancer Treatment Group Study

Figure 2 (facing page). Kaplan–Meier Survival Estimates of Relapse-free Survival and Overall Survival among the 225 Patients with ER+ Disease, According to the Intrinsic Subtype (Panels A and B), Recurrence Score (Panels C and D), 70-Gene Profile (Panels E and F), Wound Response (Panels G and H), and Two-Gene Ratio (Panels I and J).

P values were obtained from the log-rank test. X denotes observations that were censored owing to loss to follow-up or on the date of the last contact.

89-30-52, the two-gene ratio was a significant predictor of relapse-free survival and disease-free survival.[17] 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.,[7] 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-
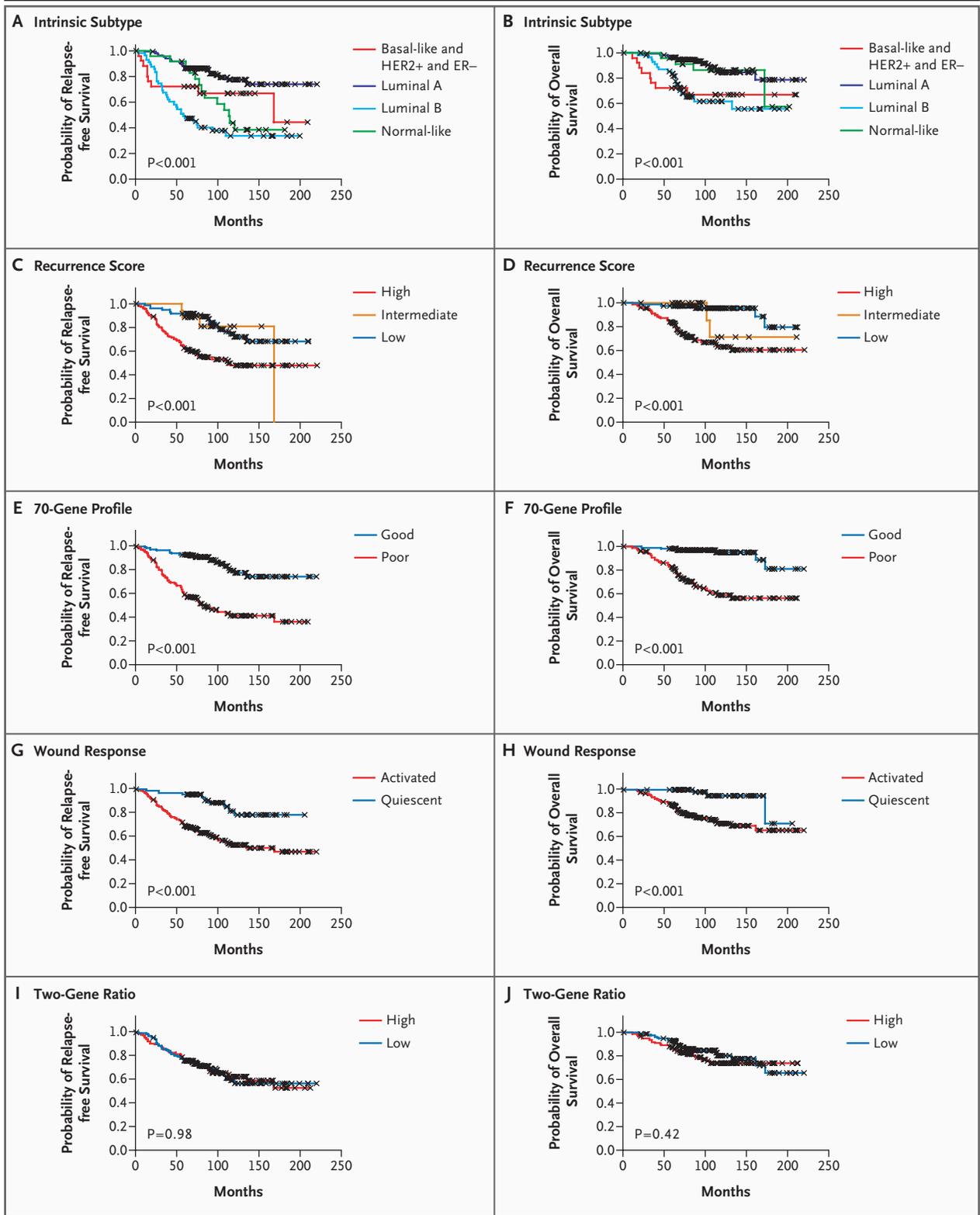
Table 3. Classification of Tumor Samples from the 225 Patients with ER+ Disease, According to the Model Used.

Intrinsic Subtype	No. of Patients	Recurrence Score		70-Gene Profile		Wound Response		Two-Gene Ratio	
		Classification	No. of Patients	Classification	No. of Patients	Classification	No. of Patients	Classification	No. of Patients
Basal-like	7	Low	1	Good	0	Quiescent	0	Low	1
		Intermediate	1						
		High	5	Poor	7	Activated	7	High	6
Luminal A	121	Low	68	Good	87	Quiescent	45	Low	77
		Intermediate	13						
		High	40	Poor	34	Activated	76	High	44
Luminal B	55	Low	2	Good	9	Quiescent	4	Low	30
		Intermediate	2						
		High	51	Poor	46	Activated	51	High	25
HER2+ and ER−	18	Low	1	Good	2	Quiescent	0	Low	5
		Intermediate	0						
		High	17	Poor	16	Activated	18	High	13
Normal-like	24	Low	15	Good	15	Quiescent	11	Low	9
		Intermediate	2						
		High	7	Poor	9	Activated	13	High	15

es 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 a model that uses a sufficiently representative subgroup of these genes should be effective. This is consistent with an observation made by Son et al., who reported that approximately 19,000 genes are differentially 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.

REFERENCES

1. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100: 8418-23.

2. van de Vijver MJ, He YD, van 't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999-2009.

3. van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530-6.

4. Chang HY, Nuyten DS, Sneddon JB, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 2005;102:3738-43.

5. Chang HY, Sneddon JB, Alizadeh AA, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS Biol 2004;2(2):E7.

6. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817-26.

7. Ma XJ, Wang Z, Ryan PD, et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell 2004;5: 607-16.

8. Bertucci F, Finetti P, Rougemont J, et al. Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. Cancer Res 2005;65:2170-8.

9. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869-74.

10. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 2003;100:10393-8.

11. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671-9.

12. Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 2006;7:96.

13. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747-52.

14. van Belle G, Fisher L. Biostatistics: a methodology for the health sciences. 2nd ed. Hoboken, N.J.: Wiley-Interscience, 2004.

15. Paik S, Shak S, Tang G, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen-receptor—positive breast cancer. J Clin Oncol (in press).

16. Reid JF, Lusa L, De Cecco L, et al. Limits of predictive models using microarray data for breast cancer clinical treatment outcome. J Natl Cancer Inst 2005;97:927-30.

17. Goetz MP, Suman VJ, Ingle JN, et al. A two-gene expression ratio of homeobox 13 and interleukin-17B receptor for prediction of recurrence and survival in women receiving adjuvant tamoxifen. Clin Cancer Res 2006;12:2080-7.

18. Son CG, Bilke S, Davis S, et al. Database of mRNA gene expression profiles of multiple human organs. Genome Res 2005;15:443-50.

Copyright © 2006 Massachusetts Medical Society.

2. Amarenco P, Labreuche J, Lavallee P, Touboul PJ. Statins in stroke prevention and carotid atherosclerosis: systematic review and up-to-date meta-analysis. Stroke 2004;35:2902-9.

3. Castelli WP, Anderson K, Wilson PW, Levy D. Lipids and risk of coronary heart disease: the Framingham Study. Ann Epidemiol 1992;2:23-8.

4. Prospective Studies Collaboration. Cholesterol, diastolic blood pressure, and stroke: 13,000 strokes in 450,000 people in 45 prospective cohorts. Lancet 1995;346:1647-53.

5. Tirschwell DL, Smith NL, Heckbert SR, Lemaitre RN, Longstreth WT Jr, Psaty BM. Association of cholesterol with stroke risk varies in stroke subtypes and patient subgroups. Neurology 2004;63:1868-75.

6. Hayward RA, Hofer TP, Vijan S. Lack of evidence for recommended LDL treatment targets: a solvable problem. Ann Intern Med (in press).

7. Segal AZ, Chiu RI, Eggleston-Sexton PM, Beiser A, Greenberg SM. Low cholesterol as a risk factor for primary intracerebral hemorrhage: a case-control study. Neuroepidemiology 1999; 18:185-93.

8. Ioannidis JP, Lau J. The impact of high-risk patients on the results of clinical trials. J Clin Epidemiol 1997;50:1089-98.

9. Hayward RA, Kent DM, Vijan S, Hofer TP. Reporting clinical trial results to inform providers, payers, and consumers. Health Aff (Millwood) 2005;24:1571-81.

10. Grundy SM, Cleeman JI, Merz CN, et al. Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. Circulation 2004; 110:227-39. [Erratum, Circulation 2004;110:763.]

11. Sacco RL, Adams R, Albers G, et al. Guidelines for prevention of stroke in patients with ischemic stroke or transient ischemic attack: a statement for healthcare professionals from the American Heart Association/American Stroke Association Council on Stroke. Stroke 2006;37:577-617.

12. Ovbiagele B, Saver JL, Bang H, et al. Statin treatment and adherence to national cholesterol guidelines after ischemic stroke. Neurology 2006;66:1164-70.

13. Ovbiagele B, Saver JL, Fredieu A, et al. In-hospital initiation of secondary stroke prevention therapies yields high rates of adherence at follow-up. Stroke 2004;35:2879-83.

Copyright © 2006 Massachusetts Medical Society.

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 subtypes of breast cancer increasingly being recognized as separate diseases that require biologically based therapies. One type of breast cancer, HER2-positive disease, was definitively identified as a separate entity when it was found that survival among patients with early-stage breast cancer is substantially improved by trastuzumab, a monoclonal antibody that interrupts HER2 signaling, when combined with standard chemohormonal therapy.[5,6] The large magnitude of the benefit seen with the addition of trastuzumab heralds the advances to come, as other biologically defined subtypes become the focus of adjuvant-therapy trials.

Gene-expression profiling has contributed to this evolving realization that the biologic heterogeneity of breast cancer has implications for treatment. There are now several predictors based on this method. One such predictor is the intrinsic-subtype classifier, which uses gene-expression profiles to distinguish among breast cancers on the basis of either their cell type of origin — the luminal cell (which is ER-positive) or the basal cell (which lacks expression of ER, the progesterone receptor, and HER2) — or whether the tumor is HER2-positive.[7]

A second microarray-based predictor, specifically based on the levels of expression of 70 genes, discriminates between a good and a poor outcome (risk of recurrence) in patients with early-stage breast cancer. The signature associated

with a poor prognosis demonstrates overexpression of genes regulating the cell cycle, invasion, metastasis, and angiogenesis.[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.[10,11]

A fourth predictor, based on a wound-response gene-expression signature derived from the transcriptional 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.[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.[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.[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 local 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 local 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.

1. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 2005;365:1687-717.
2. Carlson RW, Brown E, Burstein HJ, et al. NCCN task force report: adjuvant therapy for breast cancer. J Natl Compr Canc Netw 2006;4:Suppl 1:S1-S26.
3. International Breast Cancer Study Group. Endocrine responsiveness and tailoring adjuvant therapy for postmenopausal lymph node-negative breast cancer: a randomized trial. J Natl Cancer Inst 2002;94:1054-65. [Erratum, J Natl Cancer Inst 2002;94:1339.]
4. Albain K, Barlow W, O'Malley F, et al. Concurrent (CAFT) versus sequential (CAF-T) chemohormonal therapy (cyclophosphamide, doxorubicin, 5-fluorouracil, tamoxifen) versus T alone for postmenopausal, node-positive estrogen (ER) and/or progesterone (PgR) receptor-positive breast cancer: mature outcomes and new biologic correlates on phase III Intergroup trial 0100-(SWOG 8814). San Antonio Breast Cancer Symposium, 2004. (Accessed July 14, 2006, at http://www.SABCS.org.)
5. Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med 2005;353:1673-84.
6. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. N Engl J Med 2005;353:1659-72.
7. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747-52.
8. van de Vijver MJ, He YD, van 't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999-2009.
9. Wang Y, Klijn JGM, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671-9.
10. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol (in press).
11. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817-26.
12. Chang HY, Nuyten DS, Sneddon JB, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 2005;102:3738-43.
13. Goetz MP, Suman VJ, Ingle JN, et al. A two-gene expression ratio of homeobox 13 and interleukin-17B receptor for prediction of recurrence and survival in women receiving adjuvant tamoxifen. Clin Cancer Res 2006;12:2080-7.
14. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression–based predictors for breast cancer. N Engl J Med 2006;355:560-9.
15. Heimann R, Hellman S. Clinical progression of breast cancer malignant behavior: what to expect and when to expect it. J Clin Oncol 2000;18:591-9.