REVIEW ARTICLE

MOLECULAR ORIGINS OF CANCER

Gene-Expression Signatures in Breast Cancer

Christos Sotiriou, M.D., D.Phil., and Lajos Pusztai, M.D., D.Phil.

From the Medical Oncology Department, Translational Research Unit, Jules Bordet Institute, Université Libre de Bruxelles, Brussels (C.S.); and the Department of Breast Medical Oncology, the University of Texas M.D. Anderson Cancer Center, Houston (L.P.). Address reprint requests to Dr. Sotiriou at the Translational Research Unit, Jules Bordet Institute, 121 Blvd. de Waterloo, Brussels 1000, Belgium, or at christos.sotiriou@bordet.be.

N Engl J Med 2009;360:790-800.

Gene-expression profiling with the use of DNA microarrays allows measurement of thousands of messenger RNA (mRNA) transcripts in a single experiment. Results of such studies have confirmed that breast cancer is not a single disease with variable morphologic features and biomarkers but, rather, a group of molecularly distinct neoplastic disorders. Profiling results also support the hypothesis that estrogen-receptor (ER)–negative and ER-positive breast cancers originate from distinct cell types and point to biologic processes that govern metastatic progression. Moreover, such profiling has uncovered molecular signatures that could influence clinical care. In this review, we summarize the results of gene-expression studies that hold the most promise to accelerate the transition between empirical and molecular medicine.

Molecular Classification of Breast Cancer

Four main molecular classes of breast cancer have been distinguished by gene-expression profiling.[1-5] The "intrinsic" classification by Perou et al.[1] proposes that these four classes be called basal-like breast cancers, which mostly correspond to ER-negative, progesterone-receptor (PR)–negative, and HER2-negative tumors (hence, "triple-negative" tumors); luminal-A cancers, which are mostly ER-positive and histologically low-grade; luminal-B cancers, which are also mostly ER-positive but may express low levels of hormone receptors and are often high-grade; and HER2-positive cancers, which show amplification and high expression of the *ERBB2* gene and several other genes of the *ERBB2* amplicon. These subgroups correspond reasonably well to clinical characterization on the basis of ER and HER2 status, as well as proliferation markers or histologic grade.

Microarray studies have shown that luminal types of tumors express high amounts of luminal cytokeratins and genetic markers of luminal epithelial cells of normal breast tissue.[6] In contrast, basal-like breast cancers do not express ER, PR, and ER-related genes and do not overexpress several genes that typify myoepithelial cells of normal breast tissue: luminal cytokeratins, smooth-muscle–specific markers, and certain integrins. In some basal-like cancers, there is high expression of "basal" cytokeratins such as CK5 and a variety of growth factor receptors, including high levels of epidermal growth factor receptor, c-kit (a tyrosine kinase in breast epithelium), and growth factors such as hepatocyte growth factor and insulin growth factor.[3,4] Immunohistochemical methods for defining basal-like cancers[7] have not gained wide acceptance, partly because correspondence with molecular classification is less than perfect and also because logistic complexities limit the feasibility of combining five or more immunohistochemical markers in routine clinical practice.

Another feature that differentiates sporadic basal-like tumors from luminal-like tumors is dysfunction of the BRCA1 pathway caused by *BRCA1* gene promoter methyl-

ation, *BRCA1* transcriptional inactivation, or both.[8-11] *BRCA1* expression is important in DNA repair, activation of cell-cycle checkpoints, maintenance of chromosomal stability, and perhaps differentiation of ER-negative stem or progenitor cells into ER-positive luminal cells.[12] These findings are in line with suggestions of a link between the basal-like phenotype and germ-line mutation of *BRCA1*.[13,14] Indeed, almost all breast cancers that are associated with a *BRCA1* mutation, whether sporadic or hereditary, have a basal-like triple-negative phenotype.[6,15] Tumors associated with the *BRCA2* mutation have the distribution of phenotypes encountered in the general population.

Tumor grade can discriminate luminal A from luminal B tumors. This distinction can be further refined by the application of a genomic grade, a gene-expression signature of tumor differentiation.[4,16] Luminal B tumors typically have a high genomic grade, similar to basal-like and HER2-positive tumors, whereas luminal A tumors have a genomic grade similar to that of normal breast tissue.

Microarray-based comparative genomic hybridization has revealed differences in copy numbers of particular genes in different subtypes of breast cancer. The increased copy-number variation in basal-like tumors indicates more genetic complexity than in the other subtypes, suggesting a greater degree of genetic instability in these tumors.[17-19] Basal-like cancers are relatively enriched for low-level copy-number gains involving several chromosomal regions, whereas high-level amplification at any locus is infrequent. In contrast, high-level amplifications are seen more frequently in HER2-positive and luminal B tumors. Similar aberrant genomic patterns occur in familial breast cancers that are not associated with *BRCA1* or *BRCA2*.[18,20-22] Both hereditary *BRCA1*-associated tumors and sporadic basal-like tumors do not have markers of X-chromosome inactivation (Xi); duplication of the active X chromosome and loss of Xi suggest that X-chromosome abnormalities contribute to the pathogenesis of basal-like cancers.[23,24]

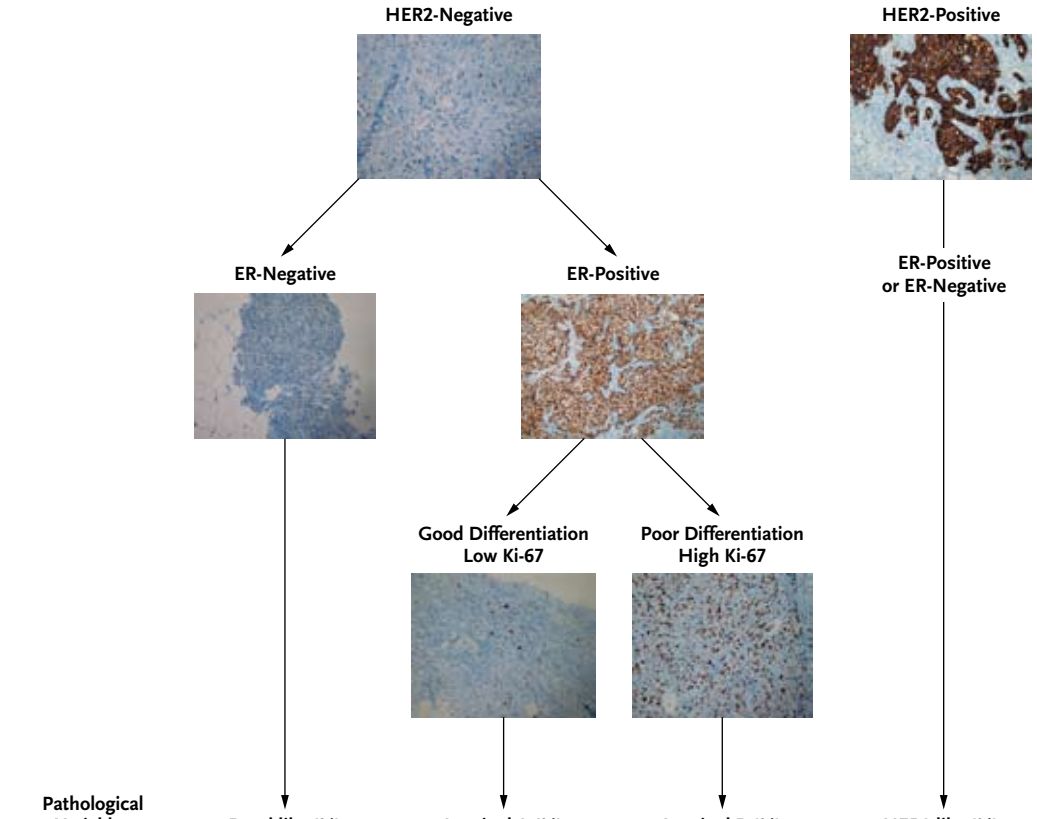
These distinct transcriptional and genomic aberrations that differentiate the four subtypes of breast cancer indicate that these variants may arise from different transformed stem or progenitor cells, each with distinct biologic properties.[25-27] Moreover, these subgroups track with prognosis and responses to therapy. The low-grade luminal A tumors are indolent and sensitive to antiestrogens. Luminal B tumors and tumors that are HER2-positive and ER-positive have incomplete sensitivity to endocrine therapy, and HER2-positive tumors, which have an aggressive natural history, are sensitive to trastuzumab, an anti-HER2 antibody. Basal-like tumors also have a more aggressive natural history, though they can be especially sensitive to chemotherapy.[28]

The additional clinical value of molecular classification is limited by its close correspondence with the status of ER, PR, and HER2, along with tumor grade (Fig. 1). However, molecular classification is changing the design of clinical trials. Moreover, the molecular differences that underlie the phenotypes of breast cancer could reveal new therapeutic targets. Examples are the identification of a functional androgen-receptor pathway in a subgroup of ER-negative and PR-negative breast tumors and defects in DNA-repair pathways in *BRCA1* and *BRCA2* carriers and probably in many basal-like cancers.[30-32]

GENE-EXPRESSION SIGNATURES AND CLINICAL OUTCOME

Gene-expression profiling has been used to develop genomic tests that may provide better predictions of clinical outcome than the traditional clinical and pathological standards[33-44] (Table 1 in the Supplementary Appendix, available with the full text of this article at NEJM.org). Three different strategies have been explored for this purpose (Fig. 2).

Using the supervised top-down approach, investigators from the Netherlands Cancer Institute developed a gene signature (MammaPrint, Agendia) from a selected retrospective series of 78 patients with node-negative breast cancer who had received no systemic adjuvant therapy.[33] The assay, which measures the expression of 70 genes and calculates a prognostic score that categorizes patients into "good" or "poor" risk groups, was recently cleared by the Food and Drug Administration (FDA) to aid in formulating a prognosis for patients with breast cancer who are under 61 years of age and who have node-negative, stage I or II disease with a tumor size of 5 cm or less.[33] However, the assay has not been tested in a prospective study. The Dutch researchers also reported a validation study from a retrospectively collected consecutive series of breast tumors, including both node-negative and node-positive cancers.[29] In this study, however, 130 patients had received systemic adjuvant chemotherapy or hor-



Pathological Variables	Basal-like (%)	Luminal A (%)	Luminal B (%)	HER2-like (%)
HER2-positive (IHC)	10	12	20	100
ER-positive (IHC)	12	96	97	46
Grade III	84	19	53	74
Tumor size >2 cm	75	53	69	74
Node-positive	40	52	65	66

Figure 1. Correspondence between Molecular Class and Clinicopathological Features of Breast Cancer.
Data are from Sorlie et al.,[3] Hu et al.,[5] Rouzier et al.,[28] and van de Vijver et al.[29] HER2 status, as determined by immunohistochemical (IHC) analysis, was available for only 82 of 535 patients in the combined studies. ER denotes estrogen receptor, and Ki-67 nuclear antigen Ki-67.

monal therapy, and 61 had also been included in the original study. Therefore, the results could have been biased. A second, more appropriate validation study that included 307 patients who had received no systemic therapy confirmed the Dutch findings.[45]

A comparison of this gene signature with the Adjuvant! Online program (www.adjuvantonline.com), which assigns risk according to conventional criteria of tumor size, nodal status, grade, and ER status, showed that 87 of 302 patients had discordant results (29%). Of these 87 patients, 59 (68%) had tumors that were rated as clinically high-risk according to the conventional criteria but low-risk according to their gene signature, and 28 (32%) had tumors that were rated as clinically low-risk but high-risk according to their gene signature. In these discordant cases, the genomic test appeared to predict the outcome more accurately. Patients with clinically low-risk tumors that were rated as high-risk on genomic assay had a 10-year overall survival rate of 69%, whereas patients with clinically high-risk tumors that were low-risk on genomic assay had a 10-year overall survival rate of 89%. Whether these data would be relevant in patients treated with adjuvant therapy remains unclear. It is interesting that in a prospectively conducted multicenter study

that included 427 patients in various Dutch hospitals, the use of MammaPrint in combination with clinical guidelines led to altered adjuvant treatment recommendations in 26% of patients.[46]

The genomic-grade signature exemplifies the "bottom-up" discovery strategy.[39] It seeks to define molecular features of tumor differentiation and tumor grade, both of which influence tumor progression and metastatic spread.[47-49] A 97-gene signature consistently discriminated between low-grade and high-grade tumors. This signature, which is driven by proliferation and cell-cycle genes, separates the intermediate-grade tumors that are problematic for making decisions about treatment into two subgroups of low genomic grade and high genomic grade, with outcomes similar to those of low and high histologic grade, respectively. These results were observed across multiple independent data sets that were generated on different microarray platforms. In more than 650 patients with ER-positive breast cancer who were untreated or who had received only tamoxifen, genomic grade was associated with outcome more than clinical variables.[16] This result highlights the importance of tumor-differentiation and tumor-proliferation genes in the ER-positive subgroup, as reported previously.[16,37,50-52]

MammaPrint, the genomic-grade signature, and a 76-gene outcome signature (developed by researchers from Rotterdam, the Netherlands, in collaboration with Veridex) appear to quantify mainly tumor grade and proliferation. When the three assays were analyzed in the same population of patients who had received no systemic adjuvant therapy, they had similar performance, suggesting that genes controlling tumor differentiation and proliferation account for a large proportion of these classifiers.[53]

Another molecular assay, Oncotype DX (Genomic Health), exemplifies the candidate-gene approach to estimating outcome (Table 1).[43] It measures the expression of ER and HER2, as well as that of ER-regulated transcripts and several proliferation-related genes, with the use of the quantitative reverse-transcriptase–polymerase-chain-reaction (RT-PCR) assay. Most of these genes are associated with outcome, and several can be assessed with the use of conventional methods. The Oncotype DX system combines these measurements into a quantitative "recurrence score,"

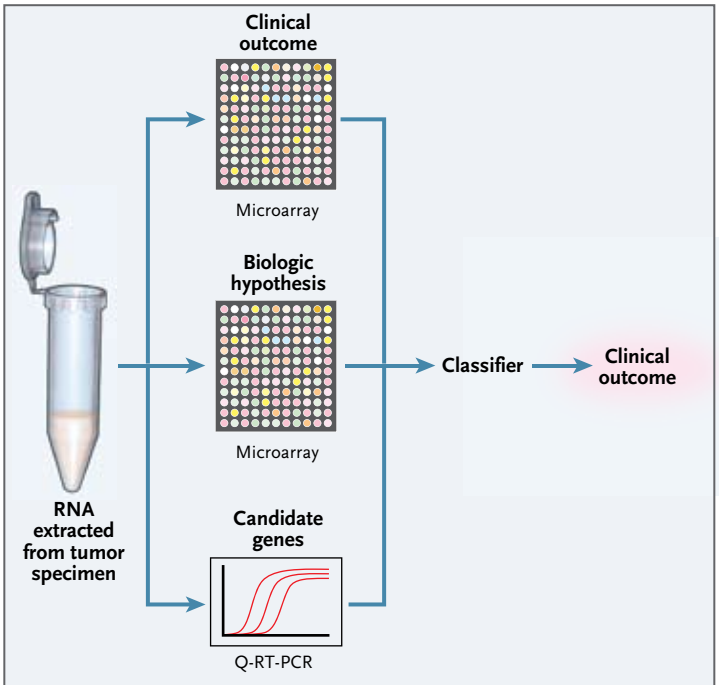


Figure 2. Three Strategies for the Development of a Gene-Expression Prognostic Signature.

In the "top-down" approach, gene-expression data from cohorts of patients with known clinical outcomes are compared to identify genes that are associated with prognosis without any a priori biologic assumption. In the "bottom-up" approach, gene-expression patterns that are associated with a specific biologic phenotype or a deregulated molecular pathway are first identified and then subsequently correlated with the clinical outcome. In the candidate-gene approach, selected genes of interest on the basis of existing biologic knowledge are combined into a multivariate predictive model. Q-RT-PCR denotes quantitative reverse-transcriptase–polymerase chain reaction.

which can be used as a continuous variable to estimate the probability of recurrence at 10 years or to group patients into low-risk, intermediate-risk, and high-risk categories.

The association between the recurrence score and distant relapse was examined retrospectively in 668 patients with ER-positive, node-negative cancers treated with tamoxifen who were enrolled in the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14 clinical trial. The 10-year distant recurrence rates were 7%, 14%, and 30% for the low-risk, intermediate-risk, and high-risk categories, respectively. Similar results were found in a community-based population of patients.[55] The Oncotype DX assay appears to identify tumors that are likely to respond to adjuvant chemotherapy in addition to tamoxifen

Table 1. Commercially Available Genomic Assays for the Prediction of Clinical Outcome in Patients with Breast Cancer.*

Variable	MammaPrint	Oncotype DX	Theros	MapQuant Dx
Provider	Agendia	Genomic Health	Biotheranostics	Ipsogen
Type of assay	70-Gene assay	21-Gene recurrence score	2-Gene ratio of HOXB13 to IL17R (H/I) and molecular-grade index	Genomic grade
Type of tissue sample	Fresh or frozen	Formalin-fixed, paraffin-embedded	Formalin-fixed, paraffin-embedded	Fresh or frozen
Technique	DNA microarrays	Q-RT-PCR	Q-RT-PCR	DNA microarrays
Centrally certified laboratory†	Yes	Yes	Yes	Yes
Indication	To aid in prognostic prediction in patients <61 yr of age with stage I or II, node-negative disease with a tumor size of ≤5 cm	To predict the risk of recurrence in patients with ER-positive, node-negative disease treated with tamoxifen; to identify patients with a low risk of recurrence who may not need adjuvant chemotherapy	To stratify ER-positive patients into groups with a predicted low risk or high risk of recurrence and a predicted good or poor response to endocrine therapy	To restratify grade 2 tumors into low-risk grade 1 or high-risk grade 3 tumors, specifically for invasive, primary, ER-positive grade 2 tumors
Level of evidence (I–V)‡	III	II	III	III
FDA clearance	Yes	No	No	No
Availability	Europe and United States	Europe and United States	United States	Europe

* ER denotes estrogen receptor, FDA Food and Drug Administration, and Q-RT-PCR quantitative reverse-transcriptase–polymerase chain reaction.
† Laboratories were certified according to the criteria of the Clinical Laboratory Improvement Amendments or by the International Organization for Standardization.
‡ Levels of evidence are measured on a scale ranging from I (strongest) to V (weakest).[54]

therapy. The association of the recurrence score with benefit from adjuvant cyclophosphamide, methotrexate, and fluorouracil chemotherapy in ER-positive, node-negative, tamoxifen-treated patients was examined in 651 patients who were enrolled in the NSABP B-20 randomized clinical trial.[56] Higher recurrence scores were associated with greater benefit from adjuvant chemotherapy, and more critically, lower recurrence scores were associated with a lack of even marginal benefit from chemotherapy. Similar results were found in a subgroup analysis of the Southwest Oncology Group (SWOG) Intergroup 0100 trial, a randomized study of tamoxifen with or without anthracycline-based chemotherapy for postmenopausal women with node-positive breast cancer.[57]

In limited feasibility studies, published only in abstract form, it was reported that molecular information may change treatment recommendations for approximately 30% of patients, usually prompting less use of chemotherapy.[58,59] The Oncotype DX assay has been endorsed as a tumor marker by the American Society of Clinical Oncology (ASCO) and as an aid to decision making regarding adjuvant chemotherapy in patients with ER-positive, node-negative breast cancer by the breast cancer panel of the National Comprehensive Cancer Network (NCCN) (www.nccn.org).[54]

Other gene signatures that may predict the risk of recurrence in ER-positive patients treated with tamoxifen have been developed through a "bottom-up" discovery strategy (Table 2 in the Supplementary Appendix).[44] Among these signatures, a high mRNA-expression ratio of HOXB13 to IL17R (H/I) was associated with a high risk of recurrence in patients treated with tamoxifen.[44] The H/I ratio was also confirmed in independent retrospective series with the use of standard formalin-fixed, paraffin-embedded tissue samples from both untreated and tamoxifen-treated patients.[60-62] Recently, the accuracy of the H/I assay was improved by including the molecular-grade index, which mainly measures cell proliferation.[63]

The fact that different gene signatures have very few genes in common may be surprising at first, but it is a common feature of complex gene-

expression data that contain large numbers of highly correlated variables (i.e., gene-expression measurements). Several different combinations of the correlated variables can be selected to build similarly accurate prediction models. Indeed, in a study in which five different signatures were tested on the same data, four of the five had similar performance.[64] This study also showed that the different signatures identified a largely overlapping population of patients as high-risk. Among the ER-positive, luminal B tumors, the rates of high-risk designation were 93% with the use of Oncotype DX and 84% with the use of MammaPrint. This concordant risk assignment occurred even though only a single gene (*SCUBE2*) is common to MammaPrint and Oncotype DX. An important limitation of these classifiers is that they assign the high-risk category to almost all ER-negative patients.

A meta-analysis of publicly available gene-expression and clinical data from almost 3000 breast tumors[65-67] supported the relationship between the risk of recurrence and molecular subtype, including several different signatures, as well as routine clinical and pathological variables. An encouraging finding is that all evaluated signatures showed similar performance despite the limited overlap of genes. Several common features also emerged from this analysis: basal-like or triple-negative tumors and HER2-positive tumors had high expression of tumor-differentiation genes, including several cell-cycle and proliferation genes. In contrast, the ER-positive combined luminal A and luminal B subtypes were more heterogeneous. The low-risk luminal A tumors were associated with a low expression of proliferation-related genes. The expression of several cell-cycle and proliferation-related genes drove the performance of several signatures. All the signatures were most useful in ER-positive tumors as a consequence of identifying the low-proliferation luminal A tumors at low risk of recurrence, whereas they were less informative for the basal-like and HER2-positive tumors, since most of these tumors were classified as high-risk by all signatures. Testing with more than one signature did not improve the performance, and lymph-node status and tumor size, which essentially capture the clinical tumor stage, had an independent prognostic value. These results suggest that both genomic and clinical variables should be included in a common algorithm to yield the most accurate prediction model (Fig. 3).

Other gene signatures that assess the role of tumor microenvironment,[68] chromosomal instability,[69] stem-cell biology,[41,42,70] and the process of metastatic spread and colonization[71-73] have been reported. Although they yield new insights into the biology of stem cells and the metastatic process, their clinical use beyond that of other classifiers is unknown.

GENE-EXPRESSION SIGNATURES AND RESPONSE TO CHEMOTHERAPY

The development of tests to predict responses to chemotherapy poses several practical challenges. There are theoretical limits to the accuracy of any response predictor that measures the characteristics of only the cancer. Host characteristics, including the rate of drug metabolism, and the association between benefit from tamoxifen and the genetic variants of CYP2D6, a cytochrome involved in the metabolism of tamoxifen, can also affect the response to therapy.[74] Moreover, there is considerable uncertainty as to what level of predictive accuracy would be clinically useful, since different levels of accuracy may be required for different clinical situations, depending on the availability of alternative treatments, the frequency and severity of adverse effects, and the risks of disease progression in the absence of therapy. Given these complexities, many of the studies of genomic markers have focused on preoperative (neoadjuvant) treatment in breast cancer. Pathologic complete response to chemotherapy indicates that the cancer is extremely sensitive to chemotherapy. Most of the clinical trials examining the correlation between pathologic complete response and long-term cancer-free survival have reported a strong association between these two outcomes.[75]

Several small studies have shown that the gene-expression profiles of cancers that are highly sensitive to chemotherapy differ from those of less responsive tumors. The largest study to date included prospectively collected needle-biopsy samples from 133 patients with stage I, II, or III breast cancer who received preoperative weekly paclitaxel and a combination of fluorouracil, doxorubicin, and cyclophosphamide.[76,77] Data from the first 82 patients were used to develop a multigene signature predictive of pathologic complete response, and data from the remaining 51 patients were used to test the accuracy of the predictor. This 30-gene predictor showed higher sensitivity than a clinical predictor that included age, nuclear grade, and

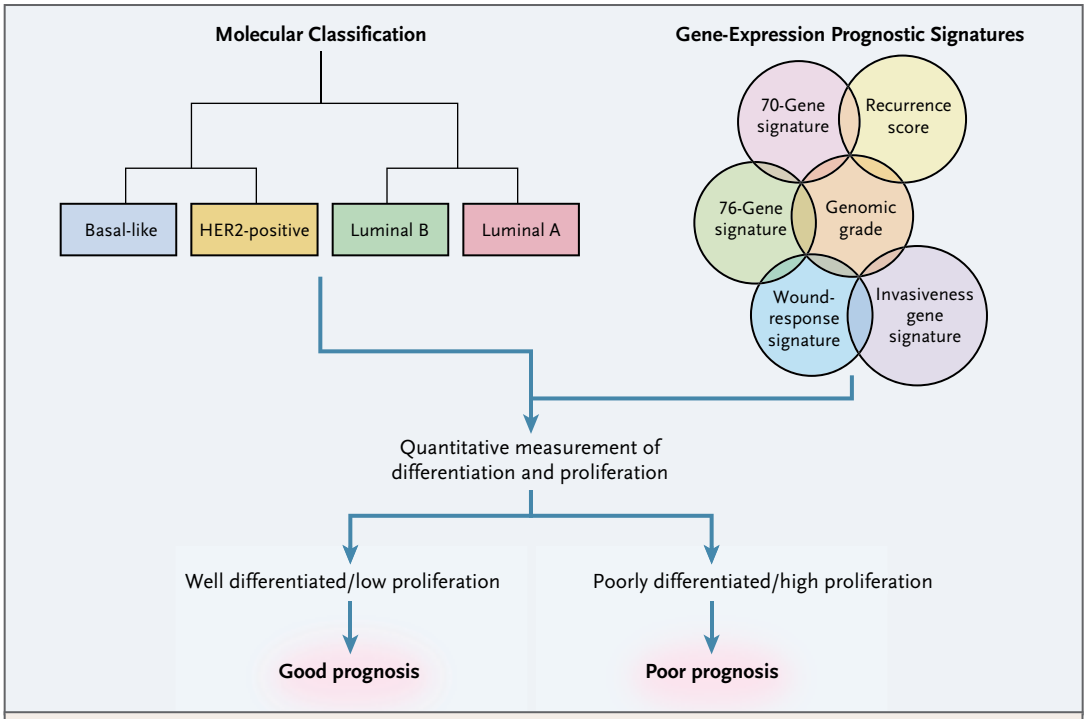


Figure 3. Molecular Classification, Gene-Expression Signatures, and Clinical Outcome.
Genes that are associated with tumor differentiation and cell cycle drive the prognostic power of the intrinsic molecular classification and several gene-expression signatures.

ER status (92% vs. 61%). It also correctly identified 92% of the patients who achieved a pathologic complete response. The positive predictive value of the pharmacogenomic signature was a modest 52%, but its negative predictive value was 96%. Similar results were reported in several other small pilot studies (Table 3 in the Supplementary Appendix). None of these predictors of chemotherapy response are commercially available, and much larger studies are needed to validate these observations, assess the specificity of the treatment regimen, and determine the true performance characteristics of these tests.

An alternative approach is the use of experimental cancer models to define gene signatures that correlate with the response to particular drugs and to test the performance of these signatures in humans. A few groups have reported gene-expression signatures associated with response or resistance to chemotherapy in vitro.[78-80] Initial efforts to validate genomic predictors that are derived from cell lines with the use of data from humans have been reported,[81] but the results remain unconfirmed.

READINESS OF GENOMIC SIGNATURES FOR ROUTINE CLINICAL USE

Several signatures are under clinical development, and some that are commercially available have been cleared by the FDA for clinical use (i.e., MammaPrint) or endorsed by ASCO and NCCN guidelines (i.e., Oncotype DX) to assist clinicians in making decisions about treatment (Table 1).[82,83] However, appropriate treatment recommendations can often be made without using these tests. The genetic tests add modest prognostic information for patients with HER2-positive and triple-negative tumors, but when measures of clinical risk are equivocal (e.g., intermediate expression of ER and intermediate histologic grade), these assays could guide clinical decisions.

Although many studies document the repro-

Table 2. Key Features of the TAILORx and MINDACT Validation Trials.*

Variable	TAILORx	MINDACT
Eligibility	Node-negative, ER-positive or PR-positive, HER2-normal, stage I or II tumor	Node-negative and 0 to 3 node-positive, any hormone receptor status, stage I, II, or III tumor
Molecular assay	Oncotype DX (Q-RT-PCR)	Mammaprint (DNA microarray)
Tissue requirement	Formalin-fixed, paraffin-embedded	Fresh or frozen
Number of participants	10,500	6000
Number to be randomized	4,390	1920
Randomized group	Patients with a recurrence score of 11 to 25 (44%)†	Discordant risk between AdjuvantOnline and MammaPrint (32%)†
Randomization	Endocrine therapy alone or endocrine therapy plus chemotherapy	Treatment recommendation on the basis of clinical or genomic risk
Treatment of nonrandomized groups	For recurrence score of <11 (29%), endocrine therapy; for recurrence score of >25 (27%), chemotherapy plus endocrine therapy†	For low risk on both predictions (13%), endocrine therapy alone; for high risk on both predictions (55%), chemotherapy plus hormonal therapy†
Primary research question	To determine whether adjuvant endocrine therapy alone is not inferior to chemotherapy and endocrine therapy in patients with an intermediate recurrence score (11 to 25)	To determine whether chemotherapy can be safely avoided in patients who are predicted to be at low risk by MammaPrint but at high risk by AdjuvantOnline
Primary end point	Disease-free survival	Distant metastasis–free survival
Secondary objectives	To create a tissue bank (formalin-fixed, paraffin-embedded samples and blood)	To create a specimen and gene-expression data bank (frozen or formalin-fixed, paraffin-embedded samples and blood)
Risk groups	Risk of distant recurrence in patients with a recurrence score of 11 to 25 treated with hormonal therapy, about 12.5%	Risk of death at 10 yr for low-risk patients: untreated, about 12%; treated with endocrine therapy, <10%
Treatment regimens	Choice of treating physician for both endocrine and chemotherapy	For endocrine therapy for ER-positive tumors: randomization between tamoxifen for 2 yr, followed by letrozole for 5 yr or letrozole for 7 yr; for chemotherapy: randomization between anthracycline-based regimen or docetaxel–capecitabine

* The Microarray in Node Negative and 0 to 3 Positive Lymph Node Disease May Avoid Chemotherapy Trial (MINDACT) (ClinicalTrials.gov number, NCT00433589) was coordinated by the European Organization for Research and Treatment of Cancer, and the Trial Assigning Individualized Options for Treatment (TAILORx) (NCT00310180) was coordinated by the Eastern Cooperative Oncology Group. ER denotes estrogen receptor, PR progesterone receptor, and Q-RT-PCR quantitative reverse-transcriptase–polymerase chain reaction.
† Percentages denote the proportions of patients in the study.

ducibility and associative properties of these assays, it is noteworthy that no prospective, randomized comparisons between genomic testing and clinical factors in making clinical decisions have been completed. Two such studies are under way: the Microarray in Node Negative and 0 to 3 Positive Lymph Node Disease May Avoid Chemotherapy Trial (MINDACT) in Europe[84] (testing MammaPrint) and the Trial Assigning Individualized Options for Treatment (TAILORx) study in the United States[85] (testing Oncotype DX) (Table 2). The results of these trials could provide valuable information[86] about the use of gene-expression signatures in daily breast-cancer management and should address important logistic, technical,[87,88] and analytical[89] issues, such as those related to the handling, shipping, reproducibility, quality control, and standardization of these new molecular tools.

CONCLUSIONS

Results from studies of gene-expression profiling have altered our view of breast cancer and provided us with a new tool for molecular diagnosis. Technical advances are rapid in this field, and the

microarray platforms that were used to develop these signatures interrogate the "mRNA world." The next generation of DNA microarrays (e.g., tiling arrays, microRNA arrays, and direct sequencing of complementary DNA) will enable investigators to study the clinical and diagnostic potential of new RNA species, including microRNAs and RNA transcribed from noncoding DNA, pseudogenes, and antisense DNA strands.

An exciting prospect of microarray-based tests is that multiple, distinct predictions — including prognosis, ER and HER2 status, and sensitivity to various treatment approaches — could be generated from a single assay. This type of test would use information from different sets of genes from the same tissue for different predictions. This outcome is technically feasible and could substantially improve the cost-effectiveness of a multigene assay.[90] To provide treatment recommendations that are truly molecularly tailored to individual patients in the future, it will be important to measure the risk of relapse and the probability of benefit from endocrine therapy and chemotherapy separately and to consider the preferences of patients in light of these results. Another promising direction of research is to examine the hypothesis that different markers and biologic pathways may be involved in determining prognosis, response, and resistance to therapy in different molecular subgroups of breast cancers. As ever-larger clinical data sets become available for gene-expression analysis, it is conceivable that predictors of molecular class–specific prognosis and treatment response will be developed in the future.

Supported by grants from the Breast Cancer Research Foundation (to Drs. Sotiriou and Pusztai) and the Belgian National Foundation for Research (to Dr. Sotiriou).

No potential conflict of interest relevant to this article was reported.

We thank Drs. Martine Piccart-Gebhart and Gabriel N. Hortobagyi for their comments.

REFERENCES

1. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747-52.
2. 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.
3. 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.
4. 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.
5. Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 2006;7:96.
6. Rakha EA, El-Sayed ME, Green AR, et al. Biologic and clinical characteristics of breast cancer with single hormone receptor positive phenotype. J Clin Oncol 2007;25:4772-8.
7. Rakha EA, Tan DS, Foulkes WD, et al. Are triple-negative tumours and basal-like breast cancer synonymous? Breast Cancer Res 2007;9:404.
8. Matros E, Wang ZC, Lodeiro G, Miron A, Iglehart JD, Richardson AL. BRCA1 promoter methylation in sporadic breast tumors: relationship to gene expression profiles. Breast Cancer Res Treat 2005;91:179-86.
9. Turner N, Tutt A, Ashworth A. Hallmarks of 'BRCAness' in sporadic cancers. Nat Rev 2004;4:814-9.
10. Turner NC, Reis-Filho JS, Russell AM, et al. BRCA1 dysfunction in sporadic basal-like breast cancer. Oncogene 2007;26:2126-32.
11. Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. N Engl J Med 2001;344:539-48.
12. Liu S, Ginestier C, Charafe-Jauffret E, et al. BRCA1 regulates human mammary stem/progenitor cell fate. Proc Natl Acad Sci U S A 2008;105:1680-5.
13. Foulkes WD, Brunet JS, Stefansson IM, et al. The prognostic implication of the basal-like (cyclin E high/p27 low/p53+/glomeruloid-microvascular-proliferation+) phenotype of BRCA1-related breast cancer. Cancer Res 2004;64:830-5.
14. Foulkes WD, Stefansson IM, Chappuis PO, et al. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 2003;95:1482-5.
15. Kreike B, van Kouwenhove M, Horlings H, et al. Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. Breast Cancer Res 2007;9:R65.
16. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. J Clin Oncol 2007;25:1239-46. [Erratum, J Clin Oncol 2007;25:3790.]
17. Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 2006;10:529-41.
18. Melchor L, Honrado E, Garcia MJ, et al. Distinct genomic aberration patterns are found in familial breast cancer associated with different immunohistochemical subtypes. Oncogene 2008;27:3165-75.
19. Chin SF, Teschendorff AE, Marioni JC, et al. High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. Genome Biol 2007;8:R215.
20. Melchor L, Honrado E, Huang J, et al. Estrogen receptor status could modulate the genomic pattern in familial and sporadic breast cancer. Clin Cancer Res 2007;13:7305-13.
21. Bergamaschi A, Kim YH, Wang P, et al. Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. Genes Chromosomes Cancer 2006;45:1033-40.
22. Vincent-Salomon A, Gruel N, Lucchesi C, et al. Identification of typical medullary breast carcinoma as a genomic subgroup of basal-like carcinomas, a heterogeneous new molecular entity. Breast Cancer Res 2007;9:R24.
23. Richardson AL, Wang ZC, De Nicolo A, et al. X chromosomal abnormalities in basal-like human breast cancer. Cancer Cell 2006;9:121-32.
24. Wessels LF, van Welsem T, Hart AA, van't Veer LJ, Reinders MJ, Nederlof PM. Molecular classification of breast carcinomas by comparative genomic hybridization: a specific somatic genetic profile for BRCA1 tumors. Cancer Res 2002;62:7110-7.
25. Glinsky GV. Stem cell origin of death-from-cancer phenotypes of human prostate and breast cancers. Stem Cell Rev 2007;3:79-93.

26. *Idem*. "Stemness" genomics law governs clinical behavior of human cancer: implications for decision making in disease management. J Clin Oncol 2008; 26:2846-53.
27. Kakarala M, Wicha MS. Implications of the cancer stem-cell hypothesis for breast cancer prevention and therapy. J Clin Oncol 2008;26:2813-20.
28. Rouzier R, Perou CM, Symmans WF, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 2005;11: 5678-85.
29. 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.
30. Farmer P, Bonnefoi H, Becette V, et al. Identification of molecular apocrine breast tumours by microarray analysis. Oncogene 2005;24:4660-71.
31. Doane AS, Danso M, Lal P, et al. An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. Oncogene 2006;25: 3994-4008.
32. Ashworth A. A synthetic lethal therapeutic approach: poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. J Clin Oncol 2008;26:3785-90.
33. 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.
34. 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.
35. Foekens JA, Atkins D, Zhang Y, et al. Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. J Clin Oncol 2006;24:1665-71.
36. Naderi A, Teschendorff AE, Barbosa-Morais NL, et al. A gene-expression signature to predict survival in breast cancer across independent data sets. Oncogene 2007;26:1507-16.
37. Teschendorff AE, Naderi A, Barbosa-Morais NL, et al. A consensus prognostic gene expression classifier for ER positive breast cancer. Genome Biol 2006;7:R101.
38. 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.
39. Sotiriou C, Wirapati P, Loi S, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 2006;98:262-72.
40. Miller LD, Smeds J, George J, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci U S A 2005; 102:13550-5.
41. Glinsky GV, Berezovska O, Glinskii AB. Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. J Clin Invest 2005;115:1503-21.
42. Liu R, Wang X, Chen GY, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med 2007;356:217-26.
43. 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.
44. 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.
45. Buyse M, Loi S, van't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst 2006;98:1183-92.
46. Bueno-de-Mesquita JM, van Harten WH, Retel VP, et al. Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER). Lancet Oncol 2007;8:1079-87. [Erratum, Lancet Oncol 2008;9:10.]
47. Ma XJ, Salunga R, Tuggle JT, et al. Gene expression profiles of human breast cancer progression. Proc Natl Acad Sci U S A 2003;100:5974-9.
48. Rhodes DR, Yu J, Shanker K, et al. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc Natl Acad Sci U S A 2004;101:9309-14.
49. Weinberg RA. Mechanisms of malignant progression. Carcinogenesis 2008;29: 1092-5.
50. Dai H, van't Veer L, Lamb J, et al. A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. Cancer Res 2005;65:4059-66.
51. Ivshina AV, George J, Senko O, et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res 2006;66:10292-301.
52. Oh DS, Troester MA, Usary J, et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. J Clin Oncol 2006;24:1656-64.
53. Haibe-Kains B, Desmedt C, Piette F, et al. Comparison of prognostic gene expression signatures for breast cancer. BMC Genomics 2008;9:394.
54. Harris L, Fritsche H, Mennel R, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol 2007;25:5287-312.
55. Habel LA, Shak S, Jacobs MK, et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. Breast Cancer Res 2006;8:R25.
56. 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 2006;24:3726-34.
57. Albain K, Barlow W, Shak S, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal, node-positive, ER-positive breast cancer (S8814,INT0100). Breast Cancer Res Treat 2007;106:Suppl:S10. abstract.
58. Liang H, Burfsky AM, Lembersky BB, et al. A retrospective analysis of the impact of OncotypeDX low recurrence score results on treatment decisions in a single academic breast cancer center. Breast Cancer Res Treat 2007;106:Suppl:S105. abstract.
59. Mumby PB, Lo SS, Norton J, et al. Prospective multicenter study of the impact of the 21-gene recurrence score assay on patient satisfaction, anxiety and decisional conflict for adjuvant breast cancer treatment selection. Breast Cancer Res Treat 2007;106:Suppl:S73. abstract.
60. Ma XJ, Hilsenbeck SG, Wang W, et al. The HOXB13:IL17BR expression index is a prognostic factor in early-stage breast cancer. J Clin Oncol 2006;24:4611-9.
61. 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.
62. Jansen MP, Sieuwerts AM, Look MP, et al. HOXB13-to-IL17BR expression ratio is related with tumor aggressiveness and response to tamoxifen of recurrent breast cancer: a retrospective study. J Clin Oncol 2007;25:662-8.
63. Ma XJ, Salunga R, Dahiya S, et al. A five-gene molecular grade index and HOXB13:IL17BR are complementary prognostic factors in early stage breast cancer. Clin Cancer Res 2008;14:2601-8.
64. 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.
65. Desmedt C, Haibe-Kains B, Wirapati P, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. Clin Cancer Res 2008;14:5158-65.
66. Haibe-Kains B, Desmedt C, Sotiriou C, Bontempi G. A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? Bioinformatics 2008;24:2200-8.
67. Wirapati P, Sotiriou C, Kunkel S, et al. Meta-analysis of gene-expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res 2008;10:R65.

68. Finak G, Bertos N, Pepin F, et al. Stromal gene expression predicts clinical outcome in breast cancer. Nat Med 2008; 14:518-27.
69. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat Genet 2006;38:1043-8.
70. Ben-Porath I, Thomson MW, Carey VJ, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet 2008;40:499-507.
71. Kang Y, Siegel PM, Shu W, et al. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell 2003;3:537-49.
72. Minn AJ, Gupta GP, Siegel PM, et al. Genes that mediate breast cancer metastasis to lung. Nature 2005;436:518-24.
73. Minn AJ, Kang Y, Serganova I, et al. Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. J Clin Invest 2005;115:44-55.
74. Schroth W, Antoniadou L, Fritz P, et al. Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes. J Clin Oncol 2007;25:5187-93.
75. Gralow JR, Burstein HJ, Wood W, et al. Preoperative therapy in invasive breast cancer: pathologic assessment and systemic therapy issues in operable disease. J Clin Oncol 2008;26:814-9.
76. 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.
77. Hess KR, Anderson K, Symmans WF, et al. Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. J Clin Oncol 2006;24:4236-44.
78. Staunton JE, Slonim DK, Coller HA, et al. Chemosensitivity prediction by transcriptional profiling. Proc Natl Acad Sci U S A 2001;98:10787-92.
79. Potti A, Dressman HK, Bild A, et al. Genomic signatures to guide the use of chemotherapeutics. Nat Med 2006;12: 1294-300. [Errata, Nat Med 2007;13:1388, 2008;14:889.]
80. Lee JK, Havaleshko DM, Cho H, et al. A strategy for predicting the chemosensitivity of human cancers and its application to drug discovery. Proc Natl Acad Sci U S A 2007;104:13086-91.
81. Bonnefoi H, Potti A, Delorenzi M, et al. Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00-01 clinical trial. Lancet Oncol 2007;8:1071-8.
82. Ross JS, Hatzis C, Symmans WF, Pusztai L, Hortobágyi GN. Commercialized multigene predictors of clinical outcome for breast cancer. Oncologist 2008; 13:477-93.
83. Marchionni L, Wilson RF, Wolff AC, et al. Systematic review: gene expression profiling assays in early-stage breast cancer. Ann Intern Med 2008;148:358-69.
84. Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ. Clinical application of the 70-gene profile: the MINDACT trial. J Clin Oncol 2008;26:729-35.
85. Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. J Clin Oncol 2008;26:721-8.
86. Hayes DF, Bast RC, Desch CE, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. J Natl Cancer Inst 1996; 88:1456-66.
87. Pusztai L, Ayers M, Stec J, et al. Gene expression profiles obtained from fine-needle aspirations of breast cancer reliably identify routine prognostic markers and reveal large-scale molecular differences between estrogen-negative and estrogen-positive tumors. Clin Cancer Res 2003;9:2406-15.
88. Symmans WF, Ayers M, Clark EA, et al. Total RNA yield and microarray gene expression profiles from fine-needle aspiration biopsy and core-needle biopsy samples of breast carcinoma. Cancer 2003; 97:2960-71.
89. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Natl Cancer Inst 2007;99:147-57.
90. Pusztai L, Hatzis C, Cardoso F. Combined use of genomic prognostic and treatment response predictors in lymph node-negative breast cancer. J Clin Oncol 2008;26:Suppl:15S. abstract.