

## GENE EXPRESSION PROFILING AS A GUIDE FOR THE MANAGEMENT OF EARLY STAGE BREAST CANCER

### *A Technology Assessment*

#### INTRODUCTION

The California Technology Assessment Forum has been asked to review the scientific literature on the safety and efficacy of gene expression profiling for the management of early stage breast cancer.

#### BACKGROUND

##### Breast Cancer

Cancer of the breast is the most common form of cancer in women. Every American woman is estimated to have a one in eight chance of developing breast cancer at some time during her life. In 2006, there will be an estimated 212,920 new cases of invasive breast cancer (>100,000 without lymph node involvement) in the United States, and an estimated 41,000 deaths from this cancer. This represents approximately 31% of all new cancer cases and 15% of all cancer deaths in women.<sup>1</sup>

The staging system of the American Joint Committee on Cancer<sup>2</sup> defines early stage (Stage I, II, and IIIA) invasive breast cancer as tumors  $\leq 5$  cm in largest dimension, with up to three positive axillary nodes and without distant metastasis or involvement of the fixed axillary or internal mammary lymph nodes (T1-2, N0-1, M0). In the TNM (Tumor Node Metastasis) staging system for breast cancer, stage T1 refers to carcinomas 2.0 cm or less in greatest dimension; stage T2 refers to tumors more than 2.0 cm, but not more than 5.0 cm in greatest dimension; stage T3, to tumors more than 5.0 cm in greatest dimension; and stage T4, to tumors with direct extension to chest wall or skin. Stage N0 refers to tumors without regional lymph node metastasis and N1 refers to tumors with ipsilateral metastases to one to three axillary lymph nodes. N2 refers to involvement of four or more axillary nodes. M0 refers to tumors without distant metastases.

- *Stage 0* - Carcinoma in situ
- *Stage I* - Tumor (T) does not exceed 2 cm, no axillary lymph nodes (N) involved.
- *Stage IIA* – T 2-5 cm, N negative, or T <2 cm and N positive.
- *Stage IIB* – T > 5 cm, N negative, or T 2-5 cm and N positive (< 4 axillary nodes).
- *Stage IIIA* – T > 5 cm, N positive, or T 2-5 cm with 4 or more axillary nodes
- *Stage IIIB* – T has penetrated chest wall or skin, and may have spread to < 10 axillary N

- *Stage III C* – T has > 10 axillary N, 1 or more supraclavicular or infraclavicular N, or internal mammary N.
- *Stage IV* – Distant metastasis (M)

### Chemotherapy in early stage breast cancer

The most recent St. Gallen consensus conference on the primary therapy of early breast cancer met in 2005. They recommend that chemotherapy be offered to patients based on risk status and hormone receptor status. They do not recommend that chemotherapy be given as part of primary therapy to low risk patients with estrogen receptor (ER) positive tumors.<sup>3</sup> Endocrine therapy is sufficient in these patients. Low risk is defined as a patient 35 years and older who has node negative (N0) disease, with a tumor less than two cm (T1) that is low grade (Grade 1), does not over express the HER2neu oncoprotein (genes are amplified, proteins overexpressed), and has no peritumoral vascular invasion. For all other women with invasive breast cancer, chemotherapy should be considered. The National Comprehensive Cancer Network (NCCN) offers similar guidelines, but recommends that women with ER-positive tumors greater than one cm in diameter consider chemotherapy.<sup>4</sup> Finally, the National Cancer Institute (NCI) treatment guidelines recommend that adjuvant chemotherapy be offered to all women with Stage I breast cancers > 1 cm in diameter, and that it is optional in women with tumors < 1 cm. A different approach is used by Adjuvant!, a free computer program used by many oncologists when counseling patients.<sup>5</sup> This program uses the patient's age, morbidity, receptor status, tumor size and grade, and number of positive lymph nodes to calculate her ten year risk of relapse and death based on data from large randomized trials and the SEER registry. The numbers can be recalculated to assess the effect of hormone therapy, chemotherapy, and combined therapy. Decisions are then made based on the patient's personal values of the absolute magnitude of the benefit compared with the harms of chemotherapy.

It has long been recognized that staging and risk assessment have important limitations. Patients with the same stage breast cancer can have markedly different responses to therapy and long-term outcomes.<sup>3, 4</sup> The majority of women with hormone receptor positive, node negative breast cancer have no evidence of breast cancer recurrence after ten years of follow-up, even if they are classified in a higher risk group using standard clinical measurements. In the NSABP-20 trial, 83% of women treated with tamoxifen alone were disease free at ten years.<sup>6</sup> However, many patients with early stage breast cancer are treated with chemotherapy in the United States. Thus, it has long been hypothesized that many of the patients with early stage breast cancer who are treated with chemotherapy derive no benefit from the treatment, yet receive all of the treatment related adverse effects. The definition of low risk described above represents years of work identifying histologic and clinical characteristics of patients who are unlikely to benefit from chemotherapy. One of the driving forces behind the characterization of tumors with gene expression profiling is to better

define which patients will benefit from chemotherapy and which will not. Additionally, some of the patients currently identified as low risk may benefit from chemotherapy. Tools such as Adjuvant! assess risk for populations, rather than for the individual. It is the hope that evaluating individual tumor biology will lead to individualization of cancer therapy to maximize efficacy and minimize toxicity.

### Gene expression profiling

Gene expression profiling refers to a number of different technologies that attempt to quantify the relative levels of messenger RNA (mRNA) for large numbers of genes in specific cells or tissues. The goal is to measure differences in the level of translation (expression) of different genes and utilize patterns of differential gene expression in order to characterize different biological states of the tissue. This allows for the simultaneous evaluation of thousands of markers and their associated patterns, rather than evaluating them individually, as has traditionally been done. In addition, gene expression is a much more accurate and powerful way to evaluate differences in markers known to correlate with response and outcome, such as the estrogen and progesterone receptors. Another potential value of this approach is the identification of genes and gene products associated with a disease process that were not previously known. In cancer biology, the technology has been used to try to differentiate between different subtypes of cancers<sup>7-13</sup>, to identify tumors with good and bad prognoses<sup>7, 12, 14-20</sup>, and to identify subgroups of tumors with a high likelihood of responding to one therapeutic regimen compared with another.<sup>21, 22</sup>

RNA is rapidly broken down in tissue samples. Thus, gene expression profiling generally requires fresh tissue that has been immediately processed to isolate and stabilize the RNA content or the tissue must be immediately frozen for later processing. The most common approach to gene expression profiling utilizes arrays of DNA sequences bound to a surface like a glass slide. Often tens of thousands of DNA sequences are organized on an individual microarray in an attempt to profile all of the 20-30,000 genes in the human genome. RNA is isolated from a test sample (tumor, white blood cells, normal tissue), amplified, and labeled with a fluorescent dye. Then it is exposed to the surface of the microarray to allow hybridization with the DNA spots bound to the microarray surface. Usually the hybridization occurs with a mixture of RNA from a control sample labeled with a second fluorescent dye. Any sample RNA that matches DNA on the microarray (complementary sequences) is bound to the microarray at a specific location. The remaining sample is then washed away. The amount of DNA binding at each site is measured by the intensity of the fluorescent signal. Since the identity of the DNA at each site on the microarray is known, the degree of fluorescence can be correlated with the relative amount of RNA in the original sample. Many genes ("housekeeping" genes: genes that tend to be transcribed continuously at a relatively constant level and usually function to maintain the cell) are expressed at the same level as the control DNA. The genes of

interest are those that are over- or underexpressed relative to the control DNA.

Another approach to the measurement of gene expression is known as real-time, reverse-transcriptase polymerase chain reaction (RT-PCR). This approach uses the reverse transcriptase enzyme to generate complementary DNA (cDNA) from the mRNA in a sample. The cDNA is then amplified using PCR. RT-PCR is more reproducible and quantitative than gene profiling with expression arrays.<sup>23, 24</sup> In particular, the precision and dynamic range of RT-PCR is greater than that of gene expression arrays. Initial discovery is often done with expression arrays, but once the discriminatory genes have been identified, RT-PCR is often used to quantify the relative amounts of a smaller set of genes. This test has the additional advantage that it can be performed on paraffin embedded tumor tissue, when a limited number of genes are analyzed.

Gene expression experiments usually start with microarrays containing many thousands of genes and compare the profiles of tissue with and without certain characteristics in order to identify a smaller subset of genes that differentiate between the two states (rejection/no rejection; metastases/no metastases). This smaller subset of genes is then validated using new patient samples. Additional candidate genes based on known biological associations may also be included.

These experiments generate tens of thousands of data points, but microarrays are expensive. Appropriate tissue from patients with outcomes of interest is limited, so the number of patients evaluated in microarray experiments is often quite low. Much has been written about the statistical dangers of evaluating thousands of predictor variables in small datasets (multiple hypothesis testing, overfitting).<sup>25-27</sup> It is essential that any pattern identified by such experiments be independently validated. Unfortunately, excitement about the results from initial experiments has often overwhelmed statistical caution. One recent paper re-evaluated the data from seven gene expression profiles of cancer prognosis and showed that five of them were likely to predict outcome no better than chance.<sup>28</sup> Ideally, results from a gene expression profile should, at a minimum, be validated in a new set of patients by a group of investigators independent from those initially developing the test. Validation means re-evaluating the test characteristics exact same assay in a new set of patients.

Gene expression profiles of breast tumors were initially analyzed with unsupervised clustering, which groups sets of cancers together based on similar patterns of DNA expression.<sup>29</sup> These intrinsic gene patterns identified five subtypes of breast cancer that apparently were subsequently found to have prognostic, as well as biological, implications.<sup>10, 30, 31</sup> That spawned a series of gene expression array experiments focused on improving risk prediction based on RNA expression levels in the primary breast tumors. Two commercial tests based on these experiments, MammaPrint and Oncotype DX, are currently marketed to physicians treating breast cancer. This is likely the tip of the iceberg: as preservation of tumor RNA from tumors

becomes standard and technology improves, novel profiles and refinements of the existing profiles will be developed and commercialized.

#### MammaPrint / The Amsterdam Profile / The 70-gene prognostic signature

MammaPrint is a custom microarray chip designed to assay the mRNA expression of 70 genes in triplicate. It requires either fresh or snap-frozen tissue. These 70 genes were identified in one of the first gene expression profiling experiments performed explicitly to find gene patterns that differentiated tumors likely to metastasize from tumors likely to be cured with local therapy alone.<sup>20</sup> It is sometimes known as the Amsterdam profile because it was developed by scientists at the Netherlands Cancer Institute in Amsterdam. The pattern was developed in patients with lymph node negative breast cancers (N0) and the commercial test is marketed for use in this population irrespective of ER status. Overexpressed genes in the profile are involved with cell cycle regulation, angiogenesis, invasion, cell migration, and signal transduction. The test result is binary, either good prognosis or poor prognosis, based on the correlation of the 70 genes with the profile identified in the initial study.

#### Oncotype DX / Recurrence Score

Oncotype DX uses RT-PCR to measure the expression levels of 21 genes in tissue samples from breast cancer. The key advantage of this test compared with usual expression profiling is that it can be performed on slides prepared from fixed, paraffin -embedded tumor tissue<sup>32</sup>; it does not require fresh or snap-frozen tumor samples. Sixteen of the genes are related to breast cancer (proliferation set: Ki67, STK15, Survivin, CCNB1 (cyclin B1), MYBL2; HER2 set: GRB7, HER2; Estrogen set: ER, PGR, BCL2, SCURBE2; invasion set: MMP11 [stromolysin 3], CTSL2 [cathepsin L2]; and three other genes: GSTM1, CD68, BAG1). The remaining five (ACTB, GAPDH, GUS, RPLPO, and TFRC) are reference genes used to standardize the test results. The reference set represent genes with minimal variation in expression level across tumors. Thus, any decrease in RNA signal intensity that occurs over time in archived, paraffin embedded tissue<sup>32</sup> can be mathematically corrected using the levels measured for the reference genes. The reference normalized expression levels for each gene ranging from zero to 15, with each one unit increase corresponding to approximately a doubling of RNA level. A proprietary algorithm uses these data to calculate a number between zero and 100, known as the recurrence score; higher scores indicate a higher probability of distant site breast cancer recurrence at ten years despite five years of adjuvant tamoxifen. Scores less than 18 are considered low risk, scores between 18 and 30 are considered intermediate risk, and scores greater than 30 are considered high risk. A high score correlates with high gene expression of proliferation genes and/or HER2 related genes; whereas a low score correlates with high gene expression for estrogen receptor related genes and low expression for the proliferation/HER2 genes. Details of the methods used to select

the final genes included in the test have not been published except in outline form. The only published information indicated that an initial list of approximately 250 genes were evaluated in three independent studies that included a total of 447 patients.<sup>33</sup> Currently, the test is directed at women with hormone-receptor positive, lymph node negative breast cancer. Among these women, tumors with low recurrence scores appear to receive no benefit from chemotherapy but significant benefit from tamoxifen, while those with high scores appear to receive significant benefit from chemotherapy, and potentially less benefit from tamoxifen. Patients with intermediate scores (representing the smallest data set in the publications to date) may receive some benefit, but the degree of benefit is less clear. An ongoing randomized clinical trial will try to answer this question.

### Technology Assessment (TA)

TA Criterion 1:           The technology must have the appropriate regulatory approval.

Until recently Oncotype DX, the Amsterdam signature, and other Gene Expression Profiles were considered Home Brew tests and exempted from FDA oversight. However, on September 7, 2006, the FDA published draft guidance on planned regulation of In Vitro Diagnostic Multivariate Assays (IVMIA). Complex tests combining data from multiple laboratory tests using a complex algorithm, like those derived from gene expression profiles, will be subject to FDA review in the future. In particular, those with direct implications for medical therapy will be considered Class III devices and will be subject to the Pre-Market Approval (PMA) process.

Genomic Health Inc. (Redwood City, CA) received a Clinical Laboratory Improvement Amendments (CLIA) Certificate of Accreditation on June 30, 2005 to conduct the Oncotype DX test.

TA Criterion 1 is met.

TA Criterion 2: The scientific evidence must permit conclusions concerning the effectiveness of the technology regarding health outcomes. For diagnostic tests, there is evidence that use of the test would result in improved medical management in a way that will benefit the patient.

The Medline database, Cochrane clinical trials database, Cochrane reviews database and the Database of Abstracts of Reviews of Effects (DARE) were searched using the key words 'breast neoplasms' and 'gene expression profiling'. These were cross-referenced with the keyword 'human'. The search was performed for the period from 1966 through September 2006 and identified 789 articles. The bibliographies of systematic

reviews and key articles were manually searched for additional references. The abstracts of citations were reviewed for relevance and all potentially relevant articles (n=68) were reviewed in full. In order to be included in this systematic review, articles had to compare the five year outcomes for women with breast cancer for high and low risk groups identified by the gene expression profile. Ideally, the study would evaluate the incremental benefit of gene expression profiling over standard risk assessment tools or guidelines such as Adjuvant!<sup>5</sup> Online or the St. Gallen Consensus Guidelines.

This review will focus on the two commercially available expression profiles: the 70-gene prognostic signature developed by Dr. van't Veer and colleagues<sup>19, 20, 34-37</sup> and the 21-gene profile developed by Dr. Paik and colleagues.<sup>33, 38-43</sup>

A number of other prognostic patterns have been developed based on gene expression profiling, but are not available for commercial use.<sup>44-56</sup> Studies predicting response to specific chemotherapeutic agents have been published, but the results are not yet mature enough to affect clinical practice.<sup>21, 22, 57-60</sup>

None of the studies used the gene expression profiles to guide patient management. Ideally, randomized clinical trials would compare the clinical outcomes of patients with early stage breast cancer managed using standard risk assessment to those of patients managed using information from gene expression profiling. Two randomized trials are currently accruing patients or about to open, but will require at least five to eight years to report preliminary outcomes. Both trials are assessing the value of chemotherapy at intermediate levels of risk, defined by either the Oncotype DX recurrence score or the 70-gene prognostic score, rather than directly comparing care guided by a gene expression profile to usual care.

Level of evidence: 1-5

TA Criterion 2 is met.

**TA Criterion 3: The technology must improve the net health outcomes.**

The primary outcomes of interest should be overall survival and disease-free survival over a minimum of five to ten years. The stated goal of many of the gene expression profiles is to reduce the use of chemotherapy in patients who do not benefit from it, while potentially identifying patients who are low risk by clinical criteria who would still benefit from chemotherapy. Ideally, trials using expression arrays would demonstrate equivalent long-term outcomes as current approaches to targeting adjuvant therapy, while reducing the number of patients receiving chemotherapy. In the absence of studies evaluating these outcomes, the incremental predictive accuracy, as measured by improvements in the concordance index<sup>61</sup> or area under the receiver operating characteristic curve, would be a reasonable surrogate outcome.

### MammaPrint / The Amsterdam Profile / The 70-gene prognostic signature

The first study explicitly using gene expression profiling to identify a set of genes that strongly predict a poor prognosis was published in January 2002.<sup>20</sup> Investigators at the Netherlands Cancer Institute selected 98 patients with breast cancer, including 34 who developed distant metastases within five years, 44 who were disease free at five years, 18 who carried the BRCA1 mutation, and two who carried the BRCA2 mutation. The 78 patients who did not carry a BRCA mutation were all lymph node negative. All patients were less than 55 years old and the tumors were less than 5 cm in diameter. The investigators isolated RNA from snap-frozen tumor tissue. Each sample was hybridized in duplicate to microarrays containing approximately 25,000 human genes. The relative RNA expression level for each tumor was compared to that of RNA from a pool of equal amounts of RNA from each patient sample. The investigators initially identified a group of 5000 genes with at least a two-fold variation in expression across the tumors. Limiting the analysis to patients without a germline BRCA mutation, the investigators first identified 231 genes among the 5000 that were significantly correlated with the development of distant metastasis (correlation coefficient  $>0.3$  or  $< -0.3$ ). Next, they ranked the genes by the magnitude of the correlation coefficient. Finally, they evaluated the ability of the genes to classify the tumors by sequentially adding groups of five genes until the model's accuracy stopped improving (70 genes). Many of the genes upregulated in the final model were functionally linked to cell cycle regulation, invasion, metastasis, angiogenesis, and signal transduction. In the derivation cohort, the model correctly classified 29 out of 34 of the patients developing metastases (sensitivity 85%) and 36 out of 44 of the patients who remained disease free (specificity 82%). Since the investigators were interested in ensuring that most patients with a poor prognosis were identified as candidates for chemotherapy, they chose a lower threshold to ensure that the sensitivity was  $> 90\%$ . Using the new threshold, only three out of 34 patients with poor prognosis were misclassified (sensitivity 91%), but 12 out of 44 patients with a good prognosis were misclassified (specificity 73%). The poor prognosis signature had an odds ratio of 28 (95% CI 7-107) for distant metastases, though this dropped to 15 (95% CI 4-56) with leave-one-out cross validation. Finally, the investigators validated the model in an additional 19 young patients with lymph node negative breast cancer. The model correctly classified 11 out of 12 patients who developed distant metastases (sensitivity 92%) and six out of seven patients who remained disease free for five years (specificity 86%)

Given the large number of potential gene predictors assessed (25,000), the small number of outcomes (34) and the post-hoc shift in the threshold used to define a poor prognosis, the model had strong potential for overfitting. One group had demonstrated that the 70 genes identified by the methods described by van't Veer et al. were highly dependent on the set of patients used to derive the model.<sup>62</sup> Many other sets of 70 genes can be found using the same data and methods that do not overlap with the 70 genes in the

prognostic signature, but predict outcome equally well in the derivation and validation sets. The fact that the results of the small validation study were better than those obtained in the derivation cohort was surprising. Clearly, larger independent validation studies are needed before the 70-gene profile is used to guide patient management.

A second study by the same group validated the 70-gene prognosis profile in a series of 295 patients with stage I or II breast cancer who were younger than 53 years.<sup>36</sup> Approximately half of the patients had lymph node negative disease (51%). All patients were followed annually for at least five years. The median follow-up was 6.7 years. RNA expression levels were assessed using the same methods as in the prior study.<sup>20</sup> The ten year survival was 54.6% among the 180 patients with a poor prognosis profile, and 94.5% among the 115 patients with a good prognosis profile. In this study, the poor prognosis signature had a hazard ratio of 5.1 (95% CI 2.9-9.0) for distant metastases. The prognosis signature was strongly associated with clinical characteristics known to be associated with prognosis, such as patient age, histologic grade of the tumor, and estrogen receptor status. In a multivariable analysis adjusting for age, lymph node status, tumor size, tumor grade, vascular invasion, ER expression, and types of treatment, the poor prognosis signature remained strongly associated with the occurrence of distant metastases (hazard ratio 4.6, 95% CI 2.3-9.2). The prognosis profile predicted being metastasis-free and overall survival at both five and ten years in this study (see Table 3 in the original paper<sup>36</sup>). The absolute differences in mortality are large enough to be clinically significant. For example, the ten-year overall survival was 94.5% in the good prognosis group and 54.6% in the poor prognosis group. Similarly, in patients with lymph node negative disease, the proportion of patients free from distant metastases was 93.4% in the good prognosis group and 56.2% in the poor prognosis group.

The investigators introduced a significant methodologic flaw by including 61 of the 78 patients used to develop the 70-gene prognosis profile in this validation study. The investigators addressed this by performing subgroup analyses of previously excluded patients from the earlier study – these results were similar to those reported for the larger set of patients. An additional confusing decision described by the investigators was the use of different thresholds to define poor prognosis (0.55 for the 61 patients from the prior study and 0.40 for all other patients). Finally, the analysis was complicated by the fact that 130 of the 295 patients received adjuvant therapy (90 chemotherapy, 20 hormonal therapy, 20 both therapies) in a non-randomized fashion. However, the magnitude of the results is compelling: the 70-gene prognosis profile clearly gives prognostic information beyond that available from standard clinical data. The appropriate patient population whose outcomes would be improved through use of the prognosis profile remains to be defined.

Espinoza et al<sup>35</sup> evaluated the feasibility of using quantitative RT-PCR as the basis for calculating the 70 gene prognostic signature because RT-PCR is thought to be more reproducible and accurate than expression array technology. They isolated tumor RNA from 96 patients with stage I or II breast cancer. Half of the patients had lymph node positive disease. Median follow-up was 70 months. The investigators were unable to perform RT-PCR for ten of the 70 genes in the original prognostic signature, but constructed an analog to the signature described by van't Veer et al. using the data from the 60 available genes.

A small study investigated the 70-gene expression profile in seven pairs of primary breast cancers and distant metastases.<sup>63, 64</sup> All five tumors with a poor prognosis signature in the primary tumor also had poor prognosis signatures in their matched metastatic tumor. However, one of the two tumors that initially had a good prognosis signature had a poor prognosis signature in the corresponding metastatic focus. All but one of the tumor pairs had a worse prognosis score in the metastasis than it did in the primary tumor. The authors also evaluated the classification of five of the pairs into the intrinsic breast cancer subtypes initially described by Perou et al.<sup>10, 29</sup> They found that all five metastatic tumors expressed genes mapping to the same breast cancer subtype as their matched primary tumor. The authors conclude that their findings support the hypothesis that treatment based on the expression profile of the original tumor should be effective at preventing future metastatic disease.

Buyse et al performed the first truly independent validation of the 70-gene prognostic signature.<sup>34</sup> Their goal was to assess whether the 70-gene signature had prognostic value that was independent of the best clinical risk classifications. They selected patients from five European centers who were younger than 61 years, were diagnosed prior to 1999 with node negative breast cancers < 5 cm in diameter, and who had not been treated with adjuvant therapy. The investigators identified 403 eligible patients with frozen tumor samples. They were able to extract useable RNA from 326 of these samples (81%). An additional 19 patients were ineligible for other reasons, leaving a final validation series of 307 patients. Agendia performed the microarray analysis using MammaPrint, a custom microarray chip designed to assay the mRNA expression of the 70 genes in triplicate. The investigators defined a tumor signature as low risk if the Pearson correlation coefficient for the 70-gene profile was above 0.4. All other profile signatures were considered high risk. During a median follow-up of 13.6 years, there were 68 recurrences, 77 distant metastases and 82 deaths. The investigators used the Adjuvant! Software<sup>5</sup>, which uses the patient's age, tumor size and grade, ER status and nodal status, to calculate the patient's ten year probability of survival. Due to missing data, they were not able to calculate the survival probability for five patients in the validation series because of missing ER status. For dichotomous analysis, they considered ER-positive patients with estimated survival  $\geq 88\%$  and ER-negative patients with estimated survival  $\geq 92\%$  to be clinically at low risk of recurrence.

The ten year overall survival was 69% for patients with a high risk gene signature in both the clinical low risk and high risk groups. Similarly, the ten year survival for patients with a low risk gene signature was 88% and 89% for the clinical low risk and high risk groups respectively. The hazard ratios of the high risk gene signature for time to distant metastases (2.32, 95% CI 1.35-4.00), overall survival (2.79, 95% CI 1.60-4.87), and disease-free survival (1.50, 95% CI 1.04-2.16) were greater than the comparable hazard ratios for age  $\leq 50$ , T2 tumor size, poorly differentiated tumor grade, ER-negative status, high risk by Adjuvant! Software, and high risk by the Nottingham Prognostic Index. Those considered high risk by the St.Gallen criteria had a greater hazard ratio (HR) for disease-free survival (HR 2.18, 95% CI 0.96-4.96), but not for time to distant metastases (HR 2.22) and overall survival (HR 1.69). The sensitivity of the 70-gene signature for predicting metastases within five years was slightly higher than that of Adjuvant! (90% vs. 87%), and it had a much higher specificity (42% vs. 29%). The area under the receiver operating characteristic curves, a measure of the discriminatory power of a prognostic tool, was 0.681 for the gene signature score and 0.659 for the Adjuvant! Score. The difference between the prognostic models on this measure was not large. Both of these numbers reflect the relatively modest discriminatory power. The results were similar for the prediction of ten year overall survival.

Of note, there was no significant heterogeneity in the hazard ratio estimates for any outcome between the five clinical sites in this study. However, the initial estimates from the first validation reported from the Netherlands Cancer Institute<sup>36</sup> were significantly higher than those in this validation series for time to distant metastases (HR 6.1 vs. 2.1), overall survival (HR 17.5 vs. 2.6), and disease free survival (HR 4.8 vs. 1.4). There are likely many factors that contribute to these large differences between these two “validation” studies of the 70-gene prognostic signature. First, there were methodologic problems with the initial validation study of van de Vijver because they included patients that were used to develop the gene signature in their validation series. There is always concern about overfitting in models developed from tens of thousands of predictors – one re-analysis suggested that this played a role in the overly optimistic estimates seen in the first validation study.<sup>28</sup> Of perhaps greater importance, the more recent validation study of Buyse et al used a different technology platform to estimate the relative expression levels of the 70 genes. They used the MammaPrint array that forms the basis of the commercially available assay and thus, is probably a more realistic estimate of how the signature will perform in the real world. Studies have demonstrated that a prediction model developed on one expression profiling platform may perform with much lower accuracy when evaluated on a different platform.<sup>65</sup> Finally, the patient populations were different: the second validation study included older women, followed them almost twice as long, and excluded patients who received adjuvant therapy.

The Microarray In Node negative Disease may Avoid ChemoTherapy (MINDACT) trial is a randomized clinical study that will assess the value of the 70-gene prognostic signature in predicting the response to chemotherapy for patients with lymph node negative breast cancer. Women 18 to 70 years old with T1, T2, or operable T3 invasive breast cancers are eligible for enrollment. This prospective, randomized phase III study will compare risk assessment using gene expression with risk assessment using common clinical-pathological criteria (Adjuvant!)<sup>5</sup> in selecting patients for adjuvant chemotherapy in node-negative breast cancer. An initial pilot trial enrolling 800 patients is currently accruing patients. The eventual goal is to study 6000 women with the prognostic signature. If both the gene signature and the clinical assessment are high risk (n=3300), patients will be randomized to one of two chemotherapy regimens. If both are low risk (n=780), then no chemotherapy will be administered. If the two forms of risk assessment are discordant (n=1920), then patients will be randomized to therapy based either on the clinical assessment or the gene expression signature. Patients with hormone receptor positive disease will be randomized to one of two hormonal regimens.

In summary, the 70-gene prognostic signature was one of the first gene expression profiles developed specifically to identify patients with a poor prognosis. It focuses on all lymph node negative patients and is most effective at predicting outcomes in untreated patients. It has been evaluated in approximately 700 patients, although the commercially available test using the MammaPrint chip has only been evaluated in one study reporting outcomes on 307 patients: a recently published multicenter validation study demonstrating that the low risk prognostic signature predicted disease-free survival independent of usual clinical and pathologic predictors of recurrence. However, the improvements in predictive accuracy were modest (from .66 for Adjuvant! to .68 for the prognostic signature) and the association between the prognostic signature and recurrent disease were much weaker than reported in the prior studies. More importantly, it is unclear what subgroup of patients would derive the greatest benefit from use of the 70-gene prognostic signature. The authors of the validation study suggest that patients with lymph node negative breast cancer and a prognostic signature that is opposite from the clinical risk assessment using current tools may benefit from therapy guided by the prognostic signature. This is being tested in a clinical trial that began recruiting patients in 2006. It is not yet clear whether the use of the prognostic signature would improve patient outcomes through increases in disease-free and overall survival or from a decrease in the number of patients unnecessarily treated with chemotherapy. Thus, TA criterion 3 is not met for the 70-gene prognostic signature.

## Oncotype DX / The Recurrence Score

The first paper describing outcomes based on the results of the Oncotype DX Recurrence Score was published in late 2004.<sup>33</sup> The study used stored specimens collected in the 1980's from the NSABP B-14 trial, a randomized clinical trial of tamoxifen in women with estrogen receptor positive, lymph node negative breast cancer. None of these samples had been used in the initial development and validation of the recurrence score. The investigators included participants in the tamoxifen-only arm of the study if tumor blocks were available in the NSABP Tissue Bank. They excluded participants with insufficient tumor tissue, insufficient RNA, or a weak RT-PCR signal. The investigators at Genomic Health extracted RNA from 10- $\mu$ m sections of the paraffin blocks and quantified the total RNA content. They then used reverse transcription to convert the RNA to DNA, followed by quantitative TaqMan RT-PCR to amplify the DNA. RT-PCR was successful in 668 of 675 tissue blocks available from the 2617 tamoxifen treated patients in the study. They measured expression of each gene in triplicate and normalized the values relative to the set of five reference genes. The investigators assessed the reproducibility of the assay using five serial sections from six blocks in two patients. The within block standard deviation for the recurrence score was .72 recurrence score units and the within patient standard deviation was 2.2 units. The pre-specified primary outcome was to compare the ten year distant recurrence rates between the high and low risk groups defined by the recurrence score.

The Oncotype DX assay classified the majority of women in the study as low risk (51% with recurrence score <18; 22% 18-30; and 27% >30). The rate of distant recurrence was significantly lower in the low risk group (6.8%, 95% CI 4.0-9.6%) than in the high-risk group (30.5%, 95% CI 23.6-37.4%,  $p < 0.001$ ). The intermediate risk group fell between the two (14.3%, 95% CI 8.3-20.3%). Age and tumor size significantly predicted distant recurrence, but neither was significant after adjusting for the recurrence score in a multivariate model. Similarly, in models including estrogen receptor status, progesterone receptor status, and HER2 status, only the recurrence score significantly predicted distant recurrence. Finally, in a model that included age, tumor size, tumor grade, ER protein level, PR (progesterone receptor) protein level, HER2 amplification, and the recurrence score, only tumor grade and the recurrence score were independent predictors of distant recurrence.

This study validated the ability of the recurrence score to identify high and low risk subgroups among tamoxifen-treated patients with ER-positive, lymph node negative breast cancers. Strengths of the study include the large sample size, independence from earlier studies used to develop the recurrence score, the centralized high quality pathology data available for all tumors, and centrally adjudicated ten year outcomes. The recurrence score was shown to predict recurrence independent from standard measures used to

assess risk of recurrence such as patient age, tumor size and grade, estrogen and progesterone receptor status, and HER2 status. However, there was not a formal assessment of the amount of independent information that the recurrence score added to traditional markers. Ideally, the investigators should have calculated the concordance index, or area under the receiver operating statistic curve, for the risk of recurrence using their model that included age, tumor size, tumor grade, ER and PR levels, HER2 levels or a standard risk assessment tool like Adjuvant! Online. Then, they should have assessed the same measure using the same model plus the recurrence score. That way the investigators and readers could evaluate the additional prognostic information provided by the recurrence score beyond what would be available from standard measurements.

Furthermore, the results of this study can only be generalized to similar patients: estrogen receptor positive, lymph node negative patients treated with tamoxifen. It is not clear how this information should be used to guide patient care in this subgroup. Almost half of women diagnosed with breast cancer in the United States have ER+ LN- breast cancer and receive hormone therapy. Thus, the recurrence score has the potential to influence the care of a large portion of women with breast cancer. The oncology community makes a clear distinction between prognostic tests (those that predict outcome) and predictive tests (those that predict which patients will respond to a particular therapy).<sup>66, 67</sup> For example, estrogen receptor status is both prognostic (ER-positive tumors have a better prognosis than ER-negative tumors) and predicts response to hormonal therapy and HER2-positive tumors have a poor prognosis, but HER-2 status predicts response to trastuzumab (Herceptin). It appears that the recurrence score is prognostic, but it is not clear that it is predictive and thus, it is not clear from this study that it can be used to improve patient outcomes.

A much smaller study evaluated the recurrence score among 149 patients who received no adjuvant therapy for their breast cancer.<sup>38</sup> The investigators from a single institution studied all patients with stage I or IIA breast cancer who had adequate tumor for evaluation, no evidence of lymph node involvement, had undergone definitive surgical treatment without adjuvant treatment and had at least five years of follow-up. Both ER positive and ER negative tumors were studied. The recurrence score assay was performed by Genomic Health. The primary outcome was distant recurrence free survival. The patients were primarily post-menopausal (82%) with a mean tumor size of 2.3 cm. The tumors were predominantly ER+ (69%) and 17% were HER-2 positive. The investigators found no association between the recurrence score and distant recurrence free survival over a median of 18 years of follow-up. The authors hypothesized that the lack of an association was due to the development of the recurrence score in cohorts of women primarily treated with tamoxifen – thus, the recurrence score, through its incorporation of estrogen receptor related genes, may partially incorporate a predisposition to respond to hormonal therapy. The authors also highlight another unusual finding in the study: there was an association of high nuclear grade with improved survival,

a pattern contrary to that observed in most studies of breast cancer prognosis. As the authors concluded, this study offers no support for the use of the recurrence score as a prognostic test.

Habel et al studied the utility of the recurrence score to predict breast cancer mortality in a population-based study using stored specimens and data from the Kaiser Permanente tumor registry.<sup>40</sup> They conducted a nested case-control study of death from breast cancer within a larger tumor registry. All women diagnosed with lymph node negative invasive breast cancer from 1985 to 1994 were eligible for inclusion in the study. Patients with prior breast cancer, bilateral disease, or metastatic disease at diagnosis were excluded. Cases were defined as women dying from their breast cancer. The investigators chose up to three controls for each case matched with respect to age, race, and year of diagnosis, pathology department, and use of tamoxifen. The pre-specified analysis plan was to stratify the analyses by ER status and treatment with tamoxifen. After chart review to confirm eligibility, there were 234 cases and 631 controls with tumor for pathology studies. Insufficient tumor (7.9%) and failed RT-PCR (1%) left a total of 220 cases and 570 controls. The recurrence score was evaluated in the standard fashion at Genomic Health. The authors cite unpublished data that evaluated the impact of tumor heterogeneity on the reproducibility of the recurrence score. The standard deviation of the recurrence score measured in 60 blocks from 20 patients was 3.0 units. The median time to death in this study was 4.9 years. As expected, the cases had tumors that were larger, less differentiated, more likely to be ER-negative, and had higher recurrence scores. As expected, tumors with low risk recurrence scores were more common in the controls (56%) than in the cases (26%). Among the ER+ patients, there were 55 cases and 150 controls treated with Tamoxifen. Using patients with low risk scores (recurrence score <18) as the referent group, the relative risk of death was 4.0 (95% CI 1.8-8.8) for patients with intermediate risk scores (18-30) and 6.2 (95% CI 2.4-15.8) for patients with high risk scores (>30). Among patients not treated with tamoxifen (110 cases and 251 controls), the relative risks were somewhat lower (2.7 and 3.3 for intermediate and high risk respectively), though still highly significant ( $p < 0.0001$ ). When considered as a continuous variable in 50 unit increments, the recurrence score remained statistically significant after adjustment for tumor size and grade among tamoxifen treated (RR 5.3, 95% CI 1.6-17.2) and untreated patients (RR 2.4, 95% CI 1.1-5.2). Among the ER-negative patients, the recurrence score was also positively associated with breast cancer death in multivariate analyses adjusting for tumor size and grade (RR 6.2, 95% CI 1.2-31.8). Overall, the recurrence score identified a larger subset of low risk patients than was possible with tumor size, grade, and ER status. The study confirms the early finding of Paik et al.<sup>33</sup> that the recurrence score is particularly useful at risk stratifying ER-positive, lymph node negative patients who receive tamoxifen, although it offered independent prognostic information for patients who did not receive tamoxifen and those with ER-negative breast cancers.

A small study in Italy studied the utility of the recurrence score derived from core biopsy tissue for the prediction of pathologic complete response in patients receiving neoadjuvant chemotherapy for locally advanced breast cancer.<sup>39</sup> This is of interest because pathologic complete response in breast to neoadjuvant therapy predicts long term survival in breast cancer.<sup>68</sup> Core biopsy specimens were available for 95 consecutive patients. Two of the specimens had insufficient tumor for the assay and four yielded insufficient RNA, leaving 89 patients with study results. The recurrence score assay was performed by Genomic Health in California according to their standard protocol. All patients were treated with doxorubicin and paclitaxel prior to surgery. The primary endpoint, pathological complete response, was defined as the complete absence of invasive breast cancer on pathologic evaluation of the surgical specimen. The mean tumor size at diagnosis was 6.4 cm. Pathologic complete response was observed in 11 patients. Complete response to neoadjuvant therapy correlated with a high recurrence score ( $p=0.005$ ), suggesting that the recurrence score may not only give prognostic information about the probability of recurrence, but may also predict which patients are most likely to respond to chemotherapy. Because the study is small and uses a surrogate marker (pathologic complete response) as the primary outcome, the results should be considered preliminary. Larger studies with clinical outcomes are required to validate the utility of the recurrence score for predicting which patients will have the best long term outcomes following chemotherapy.

In a more recent study, Paik et al evaluated the ability of the recurrence score to predict response to chemotherapy.<sup>43</sup> The study used stored specimens from the NSABP B-20 trial, a prospective, randomized clinical trial of chemotherapy plus tamoxifen versus tamoxifen alone in women with estrogen receptor positive, lymph node negative breast cancer < 5 cm in diameter and, in the larger trial, received one of two chemotherapy regimens studied (cyclophosphamide, methotrexate, and fluorouracil [CMF] or methotrexate and fluorouracil [MF]). There were no differences between the two regimens in the B-20 trial, so the investigators combined them in this evaluation of the recurrence score. The study had sufficient tumor to evaluate the recurrence score on 670 of 2,299 patients in the trial. The RT-PCR assay was successful on 651 of the 670 patients (97%). The distribution of patient age, tumor size, tumor grade, estrogen and progesterone protein levels in the 651 patients included in this study were similar to those in the remaining 1648 patients in the larger trial. The investigators used samples from patients in the tamoxifen-alone arm in the initial development of the assay, but not samples from the chemotherapy arm. The primary outcome of the study was distant recurrence of breast cancer. The research question for this study was to evaluate whether there was a significant interaction between the recurrence score and treatment with chemotherapy.

Once again Oncotype DX assay classified the majority of women in the study as low risk (54% with recurrence score <18; 21% 18-30; and 25% >30). There was a significant interaction between treatment with chemotherapy and the recurrence score ( $p=.038$ ). Patients in the low-risk group had similar outcomes

regardless of the receipt of chemotherapy. Patients in the high-risk group received significant benefits from chemotherapy (see Table 1).

Table 1: Effect of chemotherapy on ten-year distant recurrence rates by recurrence score subgroups

Recurrence Score Group	<u>Tam</u> 10-year DRR (%)	Tam+CRX 10-year DRR (%)	RR (95% CI)
Low risk (<18)	3.2	4.4	1.31 (0.46-3.78)
Intermediate risk (18-30)	9.1	10.9	0.61 (0.24-1.59)
High risk (>30)	29.5	11.9	0.26 (0.13-0.53)

Tam = Tamoxifen; CRX = chemotherapy; DRR = distant recurrence rate, RR = relative risk

None of the clinical variables showed a clinically significant interaction with chemotherapy treatment in this study, but in the larger study younger age predicted a better response to chemotherapy ( $p=0.029$ ) and poor tumor grade was almost a significant predictor of response to chemotherapy ( $p=0.057$ ). No combination of clinical variables was assessed. Within the recurrence score itself, the proliferation gene group and invasion gene group were positively associated with response to chemotherapy, while the estrogen receptor group was inversely associated with response to chemotherapy. The HER2 group did not appear to predict response to chemotherapy.

The investigators were able to use stored specimens from a prior study to effectively perform a retrospective study nested within a prospective, randomized clinical trial in order to evaluate the relative benefits of chemotherapy by recurrence score level in a large group of patients with mature ten year outcome data. However, less than half of the patients in the original study were included in this study, which raises the issue of selection bias. The authors presented data suggesting that patients in this substudy were not significantly different from the remaining patients in the NSABP B-20 trial (Appendix Table A1). Ideally, they should have presented an additional “Table 1” comparing the clinical characteristics of patients randomized to tamoxifen alone to those of patients randomized to tamoxifen plus chemotherapy in their substudy so that readers could assess whether the key benefit of randomization (equal distribution of potential confounders) was preserved within this substudy. Patients in the tamoxifen only arm of the study were used in the original development of the recurrence score. The Oncotype DX assay was re-run for all of these patients, but the non-independence of these samples decreases the value of the study as an independent validation of the recurrence score and is a significant flaw in the study. Their results clearly demonstrated a clinically and statistically significant difference in the response to chemotherapy by recurrence score risk group. Low risk patients received no benefit from chemotherapy, but high-risk patients reduced their risk for distant recurrence by 74% with the use of chemotherapy. It would have been instructive to know what percentage

of patients in the recurrence score substudy would be considered high enough risk to receive chemotherapy in addition to hormonal therapy by today's standards, and how those proportions would change with the use of the recurrence score. Other weaknesses of the study include the fact that it was a retrospective study and that current standards for therapy are more likely to include an aromatase inhibitor for hormonal therapy, a chemotherapy regimen that includes an anthracycline, and trastuzumab (Herceptin) for patients with HER2 overexpressing breast cancer. Thus, there remains some uncertainty about whether the recurrence score would predict response to therapy as effectively with today's medical care. The fact that subsets of the genes that make up the recurrence score predicted response to chemotherapy in the direction that current understanding of tumor biology would predict enhances the likelihood that the findings of this study will still be useful today.

The National Cancer Institute is currently supporting a large randomized trial, the Trial Assigning Individualized Options for Treatment (Rx), or TAILORx, to further delineate the value of the recurrence score. The study plans to enroll over 10,000 women ages 18 to 75 years old with hormone receptor positive, HER2 negative, lymph node negative breast cancer. Patients with low recurrence scores (<11) will receive hormone therapy alone. Patients with high scores (>25) will all receive hormone therapy plus chemotherapy. Patients with intermediate scores (11-25) will be randomized to either hormonal therapy alone or chemotherapy plus hormone therapy. This study is currently enrolling patients, but given the long time to recurrence expected in this generally good risk population, results are not expected for many years.

In summary, the Oncotype DX recurrence score has been studied in over 2500 patients. The assay failed in only one percent to three percent of patients in the validation studies. Among ER-positive, lymph node negative patients it strongly predicted long-term outcome, particularly among patients treated with tamoxifen. In multivariable analyses that include patient and tumor characteristics known to be associated with long-term outcomes, the recurrence score remains a strong, independent risk factor predicting the occurrence of distant metastases. However, the predictive accuracy of the recurrence score was never carefully compared to that of standard risk assessment tools. The data from the randomized NSABP B-14 trial provided the strongest evidence supporting benefits derived from use of the recurrence score to guide the use of adjuvant chemotherapy. More than half of patients with evaluable tissue had low risk recurrence scores. Over ten years of follow-up, the low risk patients randomized to chemotherapy did not have better outcomes than those patients not receiving chemotherapy. Thus, the recurrence score identified a large group of patients who could safely not suffer the short and long term consequences of chemotherapy. Conversely, patients with a high recurrence score derived a large benefit from chemotherapy (rate of distant metastases 12% vs. 30% in the untreated group). The value of chemotherapy in the intermediate risk group was negligible, but there was a trend towards benefit when follow-up was extended beyond ten years. Thus,

the recurrence score has been shown to be predictive as well as prognostic. If used to guide decision making regarding chemotherapy for patients with ER-positive LN-negative invasive breast cancers, the recurrence score should improve outcomes through improved targeting of therapy.

TA Criterion 3 is not met for the 70-gene signature

TA Criterion 3 is met for Oncotype DX

**TA Criterion 4:** The technology must be as beneficial as any established alternatives.

There is no single established alternative to gene expression profiling for risk stratification of patients with breast cancer. Guidelines from the NCI, the NCCN, St Gallen, the Nottingham Prognostic Index, and Adjuvant! Software all rely to varying degrees on patient age, tumor size and grade, lymph node status, and hormone receptor status to risk stratify patients. The risk status is then used to guide recommendations for chemotherapy. Recommendations for chemotherapy vary significantly for early stage, lymph node negative breast cancers. No studies directly compare outcomes of chemotherapy guided by one of the guidelines noted above with chemotherapy guided by risk classification by gene expression profiling.

As noted above, Buyse et al.<sup>34</sup> directly compared the risk classification of the MammaPrint 70-gene prognostic index to risk classification by several of the guidelines (St Gallen, Nottingham Prognostic Index, Adjuvant!) commonly used to help make decisions about whether to use chemotherapy. Time dependent ROC curves were used to compare MammaPrint to Adjuvant! for distant metastases within five years and for death within ten years. In both cases, the area under the ROC curve was greater for MammaPrint than for Adjuvant! (.681 vs. .659 for distant metastases, .648 vs. .576 for death, no p values reported). The authors also report the sensitivity and specificity for the different risk methods to classify risk. The sensitivity of the 70-gene signature was 90% and the specificity was 42%; 37% of the women were considered low risk. The St Gallen criteria were the most sensitive (93%), but very few patients were considered low risk (9%); hence the specificity was low (10%). Using St Gallen criteria, most patients (91%) would receive chemotherapy. The Nottingham Prognostic Index had the highest specificity (48%), but failed to identify 21% of patients who would develop metastases within five years (sensitivity 79%). Adjuvant! performed reasonably well on both sensitivity (87%) and specificity (29%), but the gene signature was better on both measures. The study demonstrated that the gene signature is at least as good as standard risk classification tools at predicting outcomes in lymph node negative women who do not receive adjuvant hormonal or chemotherapy. It also classifies a significant proportion of women as low risk. However, it is unclear what effect use of the gene signature would have on overall patient outcomes when hormonal treatments are

given to hormone receptor positive patients and chemotherapy is given to the higher risk patients. There are no published studies that directly evaluate the ability of the 70-gene prognostic signature to predict response to therapy.

No published analyses have directly compared risk classification by the Oncotype DX recurrence score with established approaches to risk classification (Adjuvant!, Nottingham Prognostic Index, St Gallen classification). In the initial validation study<sup>33</sup>, the recurrence score was a strong, statistically significant risk factor for metastases independent of patient age, tumor size and grade, ER and PR levels, and HER2 expression. However, a direct comparison of the predictive accuracy is more clinically relevant when assessing prognostic models.<sup>69, 70</sup> According to one analysis<sup>71</sup>, less than 10% of patients in the NSABP B-14 trial would have been classified as low risk by the St Gallen<sup>3</sup> or National Comprehensive Cancer Network criteria.<sup>4</sup> In contrast, the recurrence score classified 51% of the patients as low risk. Thus, approximately 40% of all patients with ER-positive, lymph node negative breast cancer who might have received chemotherapy, could potentially avoid chemotherapy through the use of the recurrence score without affecting their long-term outcome. Furthermore, in the NSABP-B20 trial, the recurrence score was shown to predict response to chemotherapy.<sup>43</sup> In this study, more than half of ER-positive, LN-negative patients were classified as low risk and in the randomized comparison, there was a trend towards harm (more distant recurrences), not benefit, from chemotherapy. In an editorial published in the same issue of the Journal of Clinical Oncology, Sandra Swain of the National Cancer Institute estimated that widespread use of the recurrence score to guide therapy could reduce the number of patients receiving recommendations for chemotherapy by 50,000 annually in the United States alone.<sup>72</sup> Thus, the evidence from a prospective, randomized clinical trial strongly supports the conclusion that use of the Oncotype DX recurrence score can impact medical decision making about the appropriate use of chemotherapy in ways that should improve patient outcomes, primarily by avoiding the use of chemotherapy in patients who are unlikely to derive any benefit from chemotherapy.

One study directly compared risk prediction using the 70-gene prognostic score to the recurrence score and several other gene profiles.<sup>73</sup> The Recurrence Score was not calculated using the RT-PCR based commercial Oncotype DX assay, but was estimated using data on the 21 genes from the expression arrays results used to develop the 70-gene prognostic score. Similarly, a special algorithm was used to approximate the Perou / Sorlie intrinsic subtype classification. The study was performed using the 295 tumor samples that contributed to the initial derivation and validation of the 70-gene profile and the wound response model.<sup>36, 37</sup> The study sample included a mix of ER-positive (n=225) and ER-negative (n=70) breast cancers. Treatment was also heterogeneous: 20 received tamoxifen alone, 20 received tamoxifen plus chemotherapy, and 90 received chemotherapy alone. Even though there was very little overlap

between the genes included in the four profiles based on multiple genes, there was a high degree of concordance in outcome predictions. This was particularly true for the 70-gene prognostic signature and the recurrence score. The two classifiers agreed on 81% of the samples (239/295). The authors suggest that the four classifiers based on gene expression profiling may be representing the same underlying biological phenotype. If true, they may be able to be used interchangeably. Ideally, subsets of patients in the two ongoing randomized trials will be assessed with both tests to better delineate their respective strengths and weaknesses.

TA Criterion 4 is not met for the 70-gene prognostic signature.

TA Criterion 4 is met for the Oncotype DX recurrence score.

TA Criterion 5:           The improvement must be attainable outside the investigational setting.

No studies have been published on the reproducibility of the 70-gene prognostic signature. The development of a chip custom-designed to assess the signature in triplicate may enhance the notoriously finicky reproducibility of expression arrays. Additionally, the tumor tissue must be handled carefully to avoid degradation of mRNA. Unlike DNA, RNA is unstable, so the length of time from excision to freezing or fixation, prolonged storage, and other factors related to specimen processing can lead to tremendous variability in the quality of mRNA available for expression profiling. The multi-center validation study suggests that it is possible to attain useful data from specimens obtained at multiple sites, although part of the explanation for the weaker association between the prognostic signature and breast cancer recurrence in that study may be variability in specimen handling and measurement error. Further evaluation of the reproducibility of the prognostic signature is warranted.

The same issues surrounding the influence of specimen processing and storage on mRNA quality apply to the Oncotype DX recurrence score. In order to partially account for expected RNA degradation, the Oncotype DX assay is designed specifically to work with mRNA fragments. Several small reproducibility studies were reported in the validation studies of the Oncotype DX recurrence score. The score appears to have reasonable precision, although the number of samples evaluated was small and the standard deviation was reported rather than the coefficient of variation. More importantly, the prevalence of low recurrence scores in the target population appears to be consistent across studies and the recurrence score has been validated in specimens obtained at many sites across the country including community based clinics in the Kaiser system.

The field would benefit from standard protocols and new reagents specifically designed to preserve mRNA for gene expression profiling. Reduction in the noise introduced by RNA degradation during sample processing, storage, and preparation should increase the precision of the measurements of mRNA levels and increase the power to predict important clinical outcomes using gene expression profiling.

TA Criterion 5 is not met for the 70-gene prognostic signature.

TA Criterion 5 is met for the Oncotype DX recurrence score.

## CONCLUSION

The majority of breast cancers in the United States are diagnosed at an early stage. Significant improvements in long-term outcomes for women with breast cancer have been achieved by targeting therapy based on the results of tests that predict response to therapy (hormone receptor status for tamoxifen and aromatase inhibitors; HER2 status for trastuzumab). Many women with early stage tumors receive no benefit from chemotherapy, but accrue all of the associated risks and side effects. Current guidelines and risk assessment tools recommend that 80 to 90% of these women be offered chemotherapy, but fewer than 50% will benefit. The primary clinical goal of the two commercially available gene expression profile tools is to improve risk stratification of women with early stage breast cancers in order to more precisely individualize use of chemotherapy.

Gene expression profiling describes several technologies that quantify the relative expression of mRNA levels for many genes. Patterns of gene expression can be used to differentiate one tumor type from another and to separate tumors likely to be associated with a good prognosis from those with a poor prognosis. The MammaPrint test is the commercial version of the 70-gene prognostic signature developed at the Netherlands Cancer Institute in Amsterdam. It was designed to predict five year rates of distant metastases in younger women with lymph node negative breast cancer. The interpretation of the results from the initial validation study in 295 patients was clouded by the inclusion of patients used to develop the signature in the group of patients used to validate the signature, the inclusion of a large number of women with lymph node positive breast cancer (49%), and the use of a different array platform than the commercial test. A larger validation study of 307 patients with lymph node negative disease < 5 cm in diameter was the first to use the commercial microarray. The prognostic signature was good for 37% of the patients. The investigators demonstrated that the prognostic signature performed better than standard guidelines and risk assessment tools at predicting recurrent disease, but the magnitude of the association with the poor prognostic signature was much lower than in the earlier validation study (relative risk for distant metastases of 2.1 vs. 6.1 in the earlier study). In summary, a relatively small number of patients have been evaluated

with the commercial version of the 70-gene prognostic signature and no studies have directly evaluated the benefits of chemotherapy in the good and poor prognosis groups and the results of that study were much weaker than prior published results. Further studies are needed to clarify the role of the 70-gene prognostic in managing patients with early stage breast cancer. A large, randomized clinical trial (MINDACT) will enroll 6000 patients in Europe in part to test that hypothesis.

The Oncotype DX test uses relative expression levels of 16 cancer-related genes and five house-keeping genes to calculate a recurrence score between 0 and 100. This test focused on predicting the ten year risk of metastases in ER-positive, LN-negative women treated with tamoxifen. The initial validation study identified the tumors of 51% of 668 ER-positive, LN-negative women as low risk. Distant recurrence was much more common in the high risk group than in the low risk group (31% vs. 7%,  $p < 0.001$ ). A second large validation study in a population based sample of 790 patients reported that the relative risk of breast cancer death for patients with high risk recurrence scores compared to those with low recurrence scores was 6.2 for patients treated with tamoxifen and 3.3 for untreated patients ( $p < 0.001$ ). Finally, a third validation study demonstrated that the benefits of chemotherapy varied by recurrence score ( $p = .038$  for the interaction) in a randomized trial of chemotherapy for women with ER-positive, LN-negative breast cancer treated with tamoxifen. Again, more than half of women (54%) were classified as low risk. The key result from this study is that the recurrence score not only offers information on prognosis (who is at high risk of recurrence vs. low risk of recurrence), but also predicts which patients are likely to benefit from chemotherapy. Low risk patients derived no benefit from chemotherapy (3.2% ten year recurrence rate for tamoxifen alone vs. 4.4% for tamoxifen plus chemotherapy). High risk patients, on the other hand, derived significant benefits chemotherapy (29.5% ten year recurrence rate for tamoxifen alone vs. 11.9% for tamoxifen plus chemotherapy). The data were equivocal for intermediate risk patients. The benefits of chemotherapy in the intermediate risk group are being evaluated in a clinical trial of 10,000 women in the United States and Canada (TAILORx). In summary, several large validation studies reported that more than half of ER-positive, LN-negative women were low risk using the recurrence score. In a retrospective study nested within a high quality, prospective randomized clinical trial, women with a low risk recurrence score received no benefit from chemotherapy. This provides strong evidence that women with low risk recurrence scores may safely not include chemotherapy as part of their treatment for ER-positive, LN-negative breast cancer.

## RECOMMENDATION

It is recommended that the use of gene expression profiling using the Oncotype DX recurrence score meets Technology Assessment Criterion 1 through 5 for safety, effectiveness and

improvement in health outcomes when used to inform the decision to use chemotherapy in patients recently diagnosed with invasive breast cancer meeting the following criteria:

- Tumor size less than 5 cm
- Estrogen or progesterone receptor positive
- Lymph node negative

It is recommended that the use of other forms of gene expression profiling, including the 70-gene prognostic signature, do not meet Technology Assessment Criterion 3 through 5 for safety, effectiveness and improvement in health outcomes.

October 18, 2006

*The California Technology Assessment Forum voted to accept the alternate recommendation as follows:*

It is recommended that the use of gene expression profiling using the Oncotype DX recurrence score meets CTAF Technology Assessment Criteria 1 through 5 for safety, effectiveness and improvement in health outcomes when used with other tools to inform the decision to use chemotherapy in patients recently diagnosed with invasive breast cancer meeting the following criteria:

- Tumor size less than 5 cm.
- Estrogen or progesterone receptor positive
- Lymph node negative

It is recommended that the use of other forms of gene expression profiling, including the 70-gene prognostic signature, do not meet CTAF Technology Assessment Criteria 3 through 5 for safety, effectiveness and improvement in health outcomes.

## RECOMMENDATIONS OF OTHERS

### Blue Cross Blue Shield Association (BCBSA)

In February 2005, the BCBSA Technology Evaluation Center Medical Advisory Panel determined that gene-expression profiling for managing breast cancer treatment did not meet TEC criteria.

### Centers for Medicare and Medicaid Services (CMS)

National Heritage Insurance Corporation (NHIC), the local CMS contractor for CA has determined that “the Oncotype DX test is considered safe and effective and reasonable and necessary to contribute to breast cancer diagnosis and major treatment decisions” with specific limitations.

### American College of Obstetricians and Gynecologists (ACOG)

ACOG does not have an opinion on the use of this technology at this time.

### American Society of Breast Surgeons (ASBS)

The ASBS has provided an opinion in support of the use of this technology. A representative was not available to attend the meeting.

### Association of Northern California Oncologists (ANCO)

ANCO provided opinion in support of the use of this technology.

### Medical Oncology Association of Southern California (MOASC)

MOASC provided opinion in support of the use of this technology.

### American Cancer Society (ACS)

ACS does not have a formal opinion or position on the use of this technology.

**ABBREVIATIONS USED IN THIS ASSESSMENT:**

TNM: Tumor Node Metastasis

NCCN: National Comprehensive Cancer Network

mRNA: messenger RNA

RT-PCR: Reverse-transcriptase polymerase chain reaction

cDNA: Complementary DNA

NO: Lymph node negative breast cancers

IVMIA: In Vitro Diagnostic Multivariate Assays

PMA: Pre-Market Approval

DARE: Database of Abstracts of Reviews of Effects

MINDACT: Microarray In Node negative Disease may Avoid ChemoTherapy

TAILORx: The Trial Assigning Individualized Options for Treatment (Rx)

ER: Estrogen Receptor

NSABP: National Surgical Adjuvant Breast and Bowel Project

CI: Confidence Interval

HR: Hazard Ratio

PR: Progesterone Receptor

LN: Lymph node

HER2: Human Epidermal Growth Factor Receptor

RR: Relative risk

## REFERENCES

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2006. *CA Cancer J Clin.* Mar-Apr 2006;56(2):106-130.
2. Breast. *American Joint Committee on Cancer: AJCC Cancer Staging Manual.* 5th ed. Philadelphia, Pa: Lippincott-Raven Publishers; 1997:171-180.
3. Goldhirsch A, Glick JH, Gelber RD, Coates AS, Thurlimann B, Senn HJ. Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005. *Ann Oncol.* Oct 2005;16(10):1569-1583.
4. Carlson RW, Brown E, Burstein HJ, et al. NCCN Task Force Report: Adjuvant Therapy for Breast Cancer. *J Natl Compr Canc Netw.* Mar 2006;4 Suppl 1:S1-26.
5. Ravdin PM, Siminoff LA, Davis GJ, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol.* Feb 15 2001;19(4):980-991.
6. Fisher B, Jeong JH, Bryant J, et al. Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from National Surgical Adjuvant Breast and Bowel Project randomised clinical trials. *Lancet.* Sep 4-10 2004;364(9437):858-868.
7. Bhattacharjee A, Richards WG, Staunton J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A.* Nov 20 2001;98(24):13790-13795.
8. Blaveri E, Simko JP, Korkola JE, et al. Bladder cancer outcome and subtype classification by gene expression. *Clin Cancer Res.* Jun 1 2005;11(11):4044-4055.
9. Maxwell GL, Chandramouli GV, Dainty L, et al. Microarray analysis of endometrial carcinomas and mixed mullerian tumors reveals distinct gene expression profiles associated with different histologic types of uterine cancer. *Clin Cancer Res.* Jun 1 2005;11(11):4056-4066.
10. 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.* Sep 11 2001;98(19):10869-10874.

11. Wilson CS, Davidson GS, Martin SB, et al. Gene expression profiling of adult acute myeloid leukemia identifies novel biologic clusters for risk classification and outcome prediction. *Blood*. Jul 15 2006;108(2):685-696.
12. Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*. Mar 2002;1(2):133-143.
13. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. Feb 3 2000;403(6769):503-511.
14. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med*. Aug 2002;8(8):816-824.
15. Pomeroy SL, Tamayo P, Gaasenbeek M, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*. Jan 24 2002;415(6870):436-442.
16. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet*. Jan 2003;33(1):49-54.
17. Rosenwald A, Staudt LM. Clinical translation of gene expression profiling in lymphomas and leukemias. *Semin Oncol*. Jun 2002;29(3):258-263.
18. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med*. Jun 20 2002;346(25):1937-1947.
19. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Expression profiling predicts outcome in breast cancer. *Breast Cancer Res*. 2003;5(1):57-58.
20. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. Jan 31 2002;415(6871):530-536.
21. 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*. Jun 15 2004;22(12):2284-2293.

22. Chang JC, Wooten EC, Tsimelzon A, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet*. Aug 2 2003;362(9381):362-369.
23. Lossos IS, Czerwinski DK, Alizadeh AA, et al. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N Engl J Med*. Apr 29 2004;350(18):1828-1837.
24. Miller LD, Long PM, Wong L, Mukherjee S, McShane LM, Liu ET. Optimal gene expression analysis by microarrays. *Cancer Cell*. Nov 2002;2(5):353-361.
25. Halloran PF, Reeve J, Kaplan B. Lies, damn lies, and statistics: the perils of the P value. *Am J Transplant*. Jan 2006;6(1):10-11.
26. Ransohoff DF. Discovery-based research and fishing. *Gastroenterology*. Aug 2003;125(2):290.
27. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer*. Apr 2004;4(4):309-314.
28. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet*. Feb 5-11 2005;365(9458):488-492.
29. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. Aug 17 2000;406(6797):747-752.
30. 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*. Jul 8 2003;100(14):8418-8423.
31. Sorlie T, Wang Y, Xiao C, et al. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. *BMC Genomics*. 2006;7:127.
32. Cronin M, Pho M, Dutta D, et al. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *Am J Pathol*. Jan 2004;164(1):35-42.
33. 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*. Dec 30 2004;351(27):2817-2826.

34. 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.* Sep 6 2006;98(17):1183-1192.
35. Espinosa E, Vara JA, Redondo A, et al. Breast cancer prognosis determined by gene expression profiling: a quantitative reverse transcriptase polymerase chain reaction study. *J Clin Oncol.* Oct 10 2005;23(29):7278-7285.
36. 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.* Dec 19 2002;347(25):1999-2009.
37. van't Veer LJ, Paik S, Hayes DF. Gene expression profiling of breast cancer: a new tumor marker. *J Clin Oncol.* Mar 10 2005;23(8):1631-1635.
38. Esteva FJ, Sahin AA, Cristofanilli M, et al. Prognostic role of a multigene reverse transcriptase-PCR assay in patients with node-negative breast cancer not receiving adjuvant systemic therapy. *Clin Cancer Res.* May 1 2005;11(9):3315-3319.
39. Gianni L, Zambetti M, Clark K, et al. Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. *J Clin Oncol.* Oct 10 2005;23(29):7265-7277.
40. 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.* May 31 2006;8(3):R25.
41. Paik S. Methods for gene expression profiling in clinical trials of adjuvant breast cancer therapy. *Clin Cancer Res.* Feb 1 2006;12(3 Pt 2):1019s-1023s.
42. Paik S, Kim CY, Song YK, Kim WS. Technology insight: Application of molecular techniques to formalin-fixed paraffin-embedded tissues from breast cancer. *Nat Clin Pract Oncol.* May 2005;2(5):246-254.
43. 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.* May 23 2006.
44. Ahr A, Holtrich U, Solbach C, et al. Molecular classification of breast cancer patients by gene expression profiling. *J Pathol.* Oct 2001;195(3):312-320.

45. Ahr A, Karn T, Solbach C, et al. Identification of high risk breast-cancer patients by gene expression profiling. *Lancet*. Jan 12 2002;359(9301):131-132.
46. 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*. Mar 8 2005;102(10):3738-3743.
47. 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*. Feb 2004;2(2):E7.
48. 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*. May 15 2005;65(10):4059-4066.
49. 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*. Apr 10 2006;24(11):1665-1671.
50. Glinsky GV, Higashiyama T, Glinskii AB. Classification of human breast cancer using gene expression profiling as a component of the survival predictor algorithm. *Clin Cancer Res*. Apr 1 2004;10(7):2272-2283.
51. 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*. Apr 1 2006;12(7 Pt 1):2080-2087.
52. Iwao K, Matoba R, Ueno N, et al. Molecular classification of primary breast tumors possessing distinct prognostic properties. *Hum Mol Genet*. Jan 15 2002;11(2):199-206.
53. Nagahata T, Onda M, Emi M, et al. Expression profiling to predict postoperative prognosis for estrogen receptor-negative breast cancers by analysis of 25,344 genes on a cDNA microarray. *Cancer Sci*. Mar 2004;95(3):218-225.
54. Onda M, Emi M, Nagai H, et al. Gene expression patterns as marker for 5-year postoperative prognosis of primary breast cancers. *J Cancer Res Clin Oncol*. Sep 2004;130(9):537-545.

55. Tsumagari K, Chijiwa K, Nagai H, et al. Postoperative prognosis of node-negative breast cancers predicted by gene-expression profiling on a cDNA microarray of 25,344 genes. *Breast Cancer*. 2005;12(3):166-177.
56. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*. Feb 19-25 2005;365(9460):671-679.
57. Chang JC, Wooten EC, Tsimelzon A, et al. Patterns of resistance and incomplete response to docetaxel by gene expression profiling in breast cancer patients. *J Clin Oncol*. Feb 20 2005;23(6):1169-1177.
58. Dressman HK, Hans C, Bild A, et al. Gene expression profiles of multiple breast cancer phenotypes and response to neoadjuvant chemotherapy. *Clin Cancer Res*. Feb 1 2006;12(3 Pt 1):819-826.
59. Rouzier R, Perou CM, Symmans WF, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res*. Aug 15 2005;11(16):5678-5685.
60. Thuerigen O, Schneeweiss A, Toedt G, et al. Gene expression signature predicting pathologic complete response with gemcitabine, epirubicin, and docetaxel in primary breast cancer. *J Clin Oncol*. Apr 20 2006;24(12):1839-1845.
61. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. Feb 28 1996;15(4):361-387.
62. Ein-Dor L, Kela I, Getz G, Givol D, Domany E. Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics*. Jan 15 2005;21(2):171-178.
63. Weigelt B, Glas AM, Wessels LF, Witteveen AT, Peterse JL, van't Veer LJ. Gene expression profiles of primary breast tumors maintained in distant metastases. *Proc Natl Acad Sci U S A*. Dec 23 2003;100(26):15901-15905.
64. Weigelt B, Hu Z, He X, et al. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Res*. Oct 15 2005;65(20):9155-9158.

65. Stec J, Wang J, Coombes K, et al. Comparison of the predictive accuracy of DNA array-based multigene classifiers across cDNA arrays and Affymetrix GeneChips. *J Mol Diagn.* Aug 2005;7(3):357-367.
66. Hayes DF, Trock B, Harris AL. Assessing the clinical impact of prognostic factors: when is "statistically significant" clinically useful? *Breast Cancer Res Treat.* 1998;52(1-3):305-319.
67. Henderson IC, Patek AJ. The relationship between prognostic and predictive factors in the management of breast cancer. *Breast Cancer Res Treat.* 1998;52(1-3):261-288.
68. Kaufmann M, von Minckwitz G, Smith R, et al. International expert panel on the use of primary (preoperative) systemic treatment of operable breast cancer: review and recommendations. *J Clin Oncol.* Jul 1 2003;21(13):2600-2608.
69. Kattan MW. Judging new markers by their ability to improve predictive accuracy. *J Natl Cancer Inst.* May 7 2003;95(9):634-635.
70. Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol.* Oct 10 2005;23(29):7332-7341.
71. Burstein HJ, Paik S, Ravdin P, Albain KS. Adjuvant chemotherapy for patients with estrogen receptor-positive breast cancer. *ASCO Educational Book: American Society for Clinical Oncology;* 2006:49-55.
72. Swain SM. A step in the right direction. *J Clin Oncol.* Aug 10 2006;24(23):3717-3718.
73. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med.* Aug 10 2006;355(6):560-569.