Genomics-Based Prognosis and Therapeutic Prediction in Breast Cancer

Stefanie S. Jeffrey, MD,* Per Eystein Lønning, MD, PhD,† and Bruce E. Hillner, MD,‡

*Stanford, California, †Bergen, Norway, and ‡Richmond, Virginia

Key Words
Gene expression profiling, DNA microarrays, genomics, breast cancer, cancer markers, prognostic markers, predictors of therapeutic response

Abstract
Breast cancer is a heterogeneous disease. DNA microarray technology is being applied to breast cancer to identify new prognostic biomarkers, to predict response to therapy, and to discover targets for the development of novel therapies. New diagnostic assays based on global gene expression are being introduced into clinical practice or tested in large-scale clinical trials. This review focuses on translational studies using microarray analyses and discusses best practice features and pitfalls. We note that factors that predict metastatic disease are not necessarily the same factors that predict therapeutic response. We believe that the characterization and discernment of different systems among breast cancers is crucial for understanding drug sensitivity and resistance mechanisms and for guiding therapy. (JNCCN 2005;3;291–300)

Prognostic and predictive factors in breast cancer have traditionally been single markers used with tumor-node-metastases (TNM) staging criteria, other features of tumor histology (tumor type, grade, and presence of lymphovascular invasion), and patient age. These are used to appraise the likelihood of recurrence of specific tumors and their response to systemic therapies. Markers in standard use in the United States include estrogen receptor (ER), progesterone receptor (PR), and the HER2/neu proto-oncogene. In many institutions, Ki-67 is used to measure tumor cell proliferation. Ongoing clinical trials prospectively evaluate newer markers for invasion and metastases, such as urokinase plasminogen activator (uPA) and plasminogen activator inhibitor-1 (PAI-1), or direct measures of tumor cells in the bone marrow or circulation. Clinical trials of new targeted therapies may require testing of a specific marker such as epidermal growth factor receptor or its mutational status,1 genes involved in apoptosis,2 proliferation,3 and multiple other signaling pathways.4

Over the past 6 years, high-throughput methods for studying global gene expression have been applied to breast cancer.5–49 From DNA microarray studies, we have learned that collections of markers may prove more useful than single markers for prognostication. Large-scale prospective corroboration is now underway. We describe the progress and pitfalls of these techniques, as well as the more limited data on the use of global-based strategies for predicting therapeutic sensitivity.

Reason for Developing Prognostic and Predictive Factors
A validated prognostic factor may serve two functions. First, it assists in the identification of a group of patients with low risk of relapse who may avoid toxic adjuvant therapy. This assumption, however, cannot be considered independent of contemporary therapeutic possibilities. Most patients are willing to accept substantial side effects to achieve a minor improvement in outcome.10 Thus, if we could offer a guarantee of cure to everybody, the great majority of patients would likely accept undergoing months of toxic chemotherapy to eliminate even a 10-year mortality risk of a few percent.

The second benefit—that a prognostic factor could be used to select “high-risk” patients in need of extensive...
therapy—includes an even more serious pitfall. Although receiving little attention, this concept actually implies that the factor associated with a poor prognosis should not predict an inferior response to therapy. Historically, the great mistakes have been avoided by chance; thus, expression of the estrogen receptor was found to be a predictive factor for endocrine treatment at the same time it was proposed as a prognostic factor. Thus, we avoided the pitfall of using ER status “prognostically,” which could have resulted in the selection of patients harboring receptor negative tumors for adjuvant endocrine treatment.

Up to the genomic era, lymph node status has been the strongest of the standard prognostic factors. The Oxford overviews have now confirmed that adjuvant therapy may reduce the hazard of relapse and death to a similar extent among node-negative and node-positive patients. This conclusion is not self-evident. Although microarray data on the identification of a “node metastatic factor” are conflicting, the fact that lymph microarray data on the identification of a “node metastatic factor.” Thus, we avoided the pitfall of using ER status “prognostically,” which could have resulted in the selection of patients harboring receptor negative tumors for adjuvant endocrine treatment.

Up to the genomic era, lymph node status has been the strongest of the standard prognostic factors. The Oxford overviews have now confirmed that adjuvant therapy may reduce the hazard of relapse and death to a similar extent among node-negative and node-positive patients. This conclusion is not self-evident. Although microarray data on the identification of a “node metastatic factor” are conflicting, the fact that lymph node status predicts prognosis independent of local therapy indicates that similar genetic changes in tumor cells are associated with lymph node and distant metastasis. However, the data indirectly suggest that the genetic disturbances associated with risk of metastases may not play a critical role in treatment sensitivity. In contrast, TP53 mutations predict a poor prognosis as well as lack of responsiveness to certain chemotherapeutics. For most prognostic factors in the literature, the fact is that we do not know whether they influence sensitivity to different treatment options.

What is important to remember is that these considerations relate to prognostication in general, whether outlined by a single factor or through multigene expression profiles.

**Use of Gene Expression Profiles (DNA Microarrays) to Determine Prognosis**

DNA microarrays are available in multiple formats. They can be based on glass slides or flexible nylon membranes. They can be generated from cDNA clones or oligonucleotides of different sizes (25-mer, 60-mer, or 70-mer). cDNA or oligonucleotide spots can be printed using stainless steel print tips, inkjet printers, or photolithographic techniques. Most currently used microarrays measure 25,000 to 42,000 features (spots), which represent the majority of expressed genes. Breast tumor tissue is snap-frozen to prevent RNA degradation; unamplified or amplified RNA from the tumor is reverse transcribed, labeled with fluorescent dye (or, less commonly, radioactivity), and hybridized to the microarray, applying the sensitivity and specificity of base pairing to identify expressed genes on a global scale. Some microarray formats use comparative hybridization of tumor and reference RNA to generate a relative abundance ratio that cancels out differences in hybridization kinetics; other formats directly measure target abundance of nucleotide sequences along a gene.

After hybridization, arrays are scanned and fluorescence signal intensity data are downloaded and analyzed using multiple bioinformatics tools. The data are “mined” and visualized to reveal collections of genes that classify tumors by their molecular profile, outline prognosis, or predict response to specific therapies. One method of analysis uses hierarchical clustering to group similar sets of tumors and genes. Clustering can be “unsupervised” (when no specific gene lists are applied) or “supervised” (when only specific gene sets or tumor/patient characteristics are analyzed). (For further review, see Jeffrey et al., or http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html.)

The gene expression profile represents an overall pattern of relatively up- or down-regulated genes expressed by a given tumor. The relative level of gene expression is more analogous to a light dimmer switch, rather than a gene being completely “turned on” or “turned off.” A breast cancer’s expression profile is unique, reproducible, and more similar to the profile of any associated lymph node or distant metastasis than to other breast cancers. Importantly, the expression profiles of breast cancers can be analyzed to identify tumor types associated with different prognostic outcomes. Expression profiling has also identified diverse prognostic groups in other malignancies such as cancers of the lung, central nervous system, head and neck, gastrointestinal system, Wilms tumor, lymphomas, and leukemias. However, the need for clinical validation trials of expression profiles and the evaluation of potential association to drug sensitivity is similar to that required for individual parameters.

**Tumor Classification by Expression Profile**

Breast cancer was the first solid tumor to be studied using microarray analysis. A collaborative research
Figure 1  (A) Gene expression patterns of 85 experimental samples analyzed by hierarchical-clustering, identifying different molecular subtypes. (B) Overall and relapse-free survival analysis of 51 locally advanced breast cancer patients, uniformly treated in a prospective study, based on five expression-based tumor subtypes. From Sorlie et al.11 Proc Natl Acad Sci U S A 2001;98:10869–10874; reprinted with permission.
group from Stanford University and Norway used microarray expression data to classify breast cancers into five or six distinct molecular subtypes. These include at least two or three ER-positive subgroups that co-express markers associated with luminal cellular characteristics; at least one ER-negative group of tumors that express basal (myoepithelial) cell characteristics; ER negative/ERBB2 (HER2/neu)-overexpressing tumors; and normal-like tumors that have some characteristics in common with normal breast tissue (Fig. 1). These molecular subtypes are associated with distinct prognoses and have been identified in diverse tumor sets from different countries and across different racial populations. Based on the subtype concept, it is expected that different therapies would be tested for their effect on tumors of a specific subtype, analogous to how we currently apply therapeutics like selective estrogen receptor modulators (SERMs) for ER-positive tumors or trastuzumab for HER2-overexpressing tumors. Although studies addressing these issues have yet to be implemented, indirect evidence suggests better tamoxifen sensitivity for “luminal A” compared with non-A luminal tumors, even though both subtypes express the estrogen receptor. The 70-gene classifier outperformed St. Gallen and NIH consensus criteria. In a follow-up study, the 70-gene profile was tested on 295 node-negative (including members of the original trial) and node-positive patients younger than 53 years of age, with median follow-up of 6.7 years (range, 0.05–18.3). Multivariable Cox regression analysis showed that this multi-gene expression profile was an independent and stronger prognostic factor than currently-used criteria.

Based on these results, the 70-gene prognosis profile will be tested in a large prospective clinical trial developed by TRANSBIG, a translational research network of the Breast International Group, which involves 39 researchers from 21 countries in Europe and Latin America (http://www.breastinternationalgroup.org). The trial is called Microarray for Node-Negative Disease may Avoid ChemoTherapy (MINDACT) and is supervised by the EORTC. The study is expected to enroll 5,000 patients over 3 years. Microarray analysis will be performed on frozen tumor samples.

A different 76-gene prognostic signature was recently published by the Rotterdam and Veridex LLC group, with 60 genes for ER-positive and 16 genes for ER-negative tumors, identifying lymph node-negative patients who developed distant metastases within five years. The profile was developed in 115 patients and validated in an independent set of 171 lymph node-negative patients (total group 286 patients, the largest breast cancer microarray study to date). The profile was found applicable to pre- and post-menopausal patients as well as patients with tumors.
10 to 20 mm, an especially common but not well-studied group.

**Paraffin-Based Prognosis Profile**

Genomic Health, a California-based company, developed a multi-gene assay using reverse transcription polymerase chain reaction (RT-PCR) on RNA isolated from paraffin-embedded tissue. Two-hundred and fifty candidate genes were chosen from an analysis of three independent microarray data sets involving 447 patients: the Stanford-Norwegian data, the NKI data, and a study on tumor class prediction from the Whitehead Institute at MIT and Dana-Farber Cancer Institute. The candidate genes were then tested on tumor samples from three separate trials. A final panel of 21 genes (16 cancer-related and 5 reference genes) was selected for entry into an algorithm that generates a recurrence score that assigns patients into a recurrence risk group (Fig. 2). This prognostic assay, marketed as Oncotype DX, is intended for use in stage I or II node-negative, ER-positive breast cancer patients who will be treated with tamoxifen. A validation study based on 675 patients from NSABP B-14 was recently published.

**Predictive Markers for Response to Primary Chemotherapy**

Two groups, one from Baylor College of Medicine and the other from Millennium Pharmaceuticals and the University of Texas M. D. Anderson Cancer Center, have used microarrays to study response of locally advanced breast cancer to primary chemotherapy in 30 and 42 patients, respectively. Multi-gene predictors of response were generated. These included a 92-gene list predicting at least 75% tumor regression in response to four cycles of docetaxol in 24 patients that was validated in an independent set of six patients and a 74-gene list predicting pathologic complete response to sequential weekly paclitaxel and fluorouracil + doxorubicin + cyclophosphamide (T/FAC) neoadjuvant chemotherapy in 24 patients that was independently validated with an additional 18 patients.

A Japanese group analyzed 44 primary or locally recurrent breast cancers treated with docetaxel using a 2,453-gene high-throughput reverse transcriptase polymerase chain reaction technique. They developed an 85-gene classifier for partial or complete response which was validated in an additional 26 patients. Although all groups reported an association between gene expression profile and treatment outcome, the predictive power was too low for current clinical use and larger validation studies are underway or planned. In the NKI study, 120 of the 144 node-positive patients and 10 of the 151 node-negative patients received any adjuvant chemotherapy or hormone therapy. This indirectly suggests that their prognostic profile may work independent of treatment and may not be associated with drug sensitivity. This, however, may also represent an oversimplification. The fact that we observe lack of cross-resistance to different chemotherapeutics administered in the advanced setting means that the mechanisms of resistance must be (at least partly) different between different compounds. Therefore, the potential for a predictive power needs to be addressed for each regimen separately.

---

**Figure 2** Twenty-one gene panel of the Oncotype DX™ assay from Genomic Health, Inc. Genes are grouped on the basis of function, correlated expression, or both. The GRB7, ER, proliferation, and invasion group scores are calculated from individual gene-expression measurements, as follows: GRB7 group score = 0.9 x GRB7 + 0.1 x HER2 (if the result is less than 8, then the GRB7 group score is considered 8); ER group score = (0.8 x ER + 1.2 x PGR + BCL2 + SCUBE2) ÷ 4; proliferation group score = [(Survivin + Ki67 + MYBL2 + CCNB1 [the gene encoding cyclin B1] + STK15] ÷ 5 (if the result is less than 6.5, then the proliferation group score is considered 6.5); and invasion group score = (CTSL2 [the gene encoding cathepsin L2] + MMP11 [the gene encoding stromolysin 3]) ÷ 2. The unscaled recurrence score (RSU) is calculated as RSU = + 0.47 x GRB7 group score – 0.08 x CD68 – 0.07 x GSTM1 – 0.05 x CXCL12 – 0.08 x CD31 + 1.2 x Survivin + 1.04 x proliferation group score + 1.0 x invasion group score + 0.05 x CXCL12 – 0.08 x GSTM1 – 0.07 x BAG1. The recurrence score (RS) is rescaled from the unscaled recurrence score, as follows: RS = 0 if RSU < 0; RS = 20 x (RSU – 6.7) if 0 ≤ RSU ≤ 100; and RS = 100 if RSU > 100. Cutoff points classify patients into the following categories: low risk (recurrence score, less than 18), intermediate risk (recurrence score, 18 or higher but less than 31), and high risk (recurrence score, 31 or higher). From Paik et al.®; with permission. © 2004 Massachusetts Medical Society; all rights reserved.
Two-Gene Expression Ratio for Predicting Tamoxifen Sensitivity

A group from Massachusetts General Hospital and Arcturus Bioscience performed microarray analysis of whole and laser microdissected frozen tumor from 60 early-stage hormone receptor-positive patients uniformly treated with 5 years of adjuvant tamoxifen alone and with median follow-up of 10 years for the patients without recurrence.83 A two-gene expression biomarker (HOXB13:IL17BR ratio) that predicted tamoxifen treatment outcome was identified. This marker was also predictive when tested using real-time quantitative PCR on formalin-fixed paraffin-embedded tissue in an additional 20 patients. Functional studies indicated that HOXB13 may be involved in cell motility and invasion. Interestingly, they found that 61 genes of the NKI 70-gene profile did not predict outcome in their patient cohort. A follow-up study in an independent ER-positive patient cohort revealed that the two-gene expression ratio was more robust in lymph node-negative patients.84

Newer Gene Lists

A recent study showed that a “wound-response” signature predicted for diminished overall survival and distant metastasis-free survival in a consecutive series of 295 early breast cancer patients.85 The wound-response signature performed particularly well as an adjunct to the 70-gene prognosis profile originally tested in this patient cohort, because there was minimal overlap between the two gene lists. Importantly, basal-like and ERBB2-overexpressing tumor subtypes generally expressed both the 70-gene poor-prognosis signature and the poor-prognosis wound-response signature, suggesting a different underlying biology from other breast cancers. This, however, raises the important issue of co-variance (well known from studies of conventional prognostic factors). Even if a gene profile can be shown reproducible to outline prognosis, the possibility exists that this may be due to covariance with other biologic factors and may not necessarily implicate a direct biological role of the genes involved in the “signature.”

Design and Analysis Concerns

In all areas of medicine, clinicians seek better tools for identifying the risk of disease, the prognosis of a disease, and for predicting response to therapy. Over the past 30 years, decision rules have been reported for numerous common clinical situations.86,87 These decision rules and most of the recent initial genetic association reports share a common concern or flaw: lack of subsequent validation or reproducibility. For the clinical oncologist, this pattern should be familiar when one recalls the hundreds of single-institution phase II studies reported at professional meetings or published that either are never subsequently assessed or found not to be beneficial in comparative phase III trials. As in most elements of clinical oncology, the practitioner should be cautiously optimistic about the power of microarray studies.

Table 1 describes select salient features that the reader should look for in genomic reports. For more detailed guidance, the reader is directed to work of Simon, Pusztai, and the ASCO tumor marker panel.88–90

The studies on the 70-gene profile by the NKI group17,18 merit particular consideration, because they may illustrate these principles. First, they include the largest number of patients reported in a study assessing the prognostic power of a gene profile in breast cancer. Second, they were able to confirm the prognostic impact of their profile in node-negative and node-positive patients separately. Considering validation of its prognostic impact, one paper questioned whether their gene profile could improve prognostication compared with optimal use of conventional indices.91

Perhaps even more important, another study questioned the uniqueness of their “gene list” by revealing several potential combinations extracted from the same dataset to be of similar importance.92 A recent analysis on survival-related outcomes in seven of the largest cancer-related microarray studies (only one of which represented solely breast cancer) revealed “instability” of gene lists based on selection of the training set, with five of the seven studies performing no better than chance for cancer prognosis classification.93 The NKI dataset,
however, was one of the two performing better than chance. However, prospective confirmation data are still required. Breast cancers may recur after many years, so data based on 5-year follow-up may potentially identify genes based on tumors that recur relatively early, such as the basal-like, ERBB2-overexpressing, and non-A luminal subtypes, and overlook the long-term impact of slower growing luminal A tumors.

Are Microarrays Too Easy?

One of the downsides to the advent of high-throughput genomic technologies is that it is often too easy to assay hundreds to thousands of genes without any specific pre-planned knowledge of the biologic processes that could be associated. Because thousands of genes are assessed, the use of traditional thresholds for statistical significance, such as 1-in-20, will lead to having many associations being coincidental rather than biologically meaningful (multiple hypothesis testing). In addition, a range of technical aspects involved in accrual of fresh frozen tissue, microarray hybridization, and clinical follow-up often limits the number and composition of tumor specimens in a cohort. For example, the very process of excising and freezing a piece of tumor from within a surgical specimen generally requires that the tumor be palpable and of a generous enough size to not interfere with routine pathology processing. This tends to bias many datasets to include a disproportionate number of large or palpable tumors.

When interpreting microarray literature, an important distinction to consider is whether the method of analysis is unsupervised or supervised. Unsupervised analysis searches for patterns in the data without prior assumptions about the biological relevance of these patterns (the classic “fishing expedition”). Such analyses are primarily used for hypothesis generation. In contrast, a supervised approach uses a prediction model to specifically search out genes known to be associated with specific biologic processes (e.g., growth, apoptosis, or known cell-surface receptors or pathways) and typically starts with a much smaller set of candidate genes. The association of specific candidate genes with biologic processes greatly reduces the number of innocent bystander genes. Other approaches to supervision analyze gene expression with respect to a defined outcome, such as prognosis or response to therapy, or using some other phenotypic distinction, such as hormone receptor status or histologic features.

Training and Test Sets

When searching for molecular distinctions between two groups of patients (e.g., those with and without tumor recurrences), differentially expressed genes are identified and rank ordered by univariate analysis methods. Multiple hypothesis issues should be addressed in the choice of threshold for statistical significance. Ideally, the cohort should first be split into training and test sets with differentially expressed genes selected in the training set and then tested on the remaining patients (test set). This simple step of splitting the cohort reduces the false-positive rate. However, many microarray studies develop a "test model" without using a split sample. A prognostic or predictive index is subsequently built from various candidate markers using multivariate methods.

The validation step is best performed by taking the index marker set to a completely new cohort and testing its reproducibility in predicting the relevant endpoint. The sample set must be assessed blindly to primary endpoint. At a minimum, a pre-planned sample size should be reported.

The Opening Chapter

Reports of genetic biomarkers in oncology are growing exponentially. To date, few studies have been validated in large, independent datasets. There has been little overlap between different gene lists for the same cancers. If these reports are to change routine clinical practice, they must be guided by sound design and reproducibility.

Currently, we face uncertainty regarding the prognostic impact of gene signatures and their application. A major advantage of microarray technology at this stage is that it offers new and varied approaches for exploring breast and other cancer biology. In the rush toward a single end-point of prognostication, data may be noisy and confounding. Although a prognostic impact corroborates biologic relevance for a molecular profile, it must not always be a primary endpoint on its own. Although the classification of breast tumors into distinct molecular subtypes based on hierarchical clustering appears robust, reproducible, and is also associated with prognosis, the most interesting aspect may be further characterization of these subtypes to determine the biologic distinctions and genetic differences that may be of critical importance. Experimental as well as clinical studies have linked mutations in genes like TP53, of critical importance to apoptosis, and genes involved in DNA repair to involvement in
chemoresistance. We expect that future microarray data will be collected in prospective clinical studies designed for this purpose. We would hope that data will be analyzed according to “functional hypotheses” based on biologic knowledge about defined genetic pathways, particularly mechanisms of drug sensitivity/resistance to guide treatment decisions.

References


