Prognostic & Predictive Assays for Breast Cancer: A Practical Update

Frederick L. Baehner, MD

Assistant Professor of Clinical Pathology
University of California, San Francisco
Director of Pathology
Genomic Health, Inc.
Redwood City, CA
“No classification is perfect nor is it likely that it ever will be ... All classifications depend on our knowledge of the pathology and histogenesis of the tumors being classified and, since this knowledge is far from perfect or complete, no classification can be other than a reasonable working compromise”

Dr. John Azzopardi

Introduction

The Philadelphia Chromosome

Using standard cytogenetic preparations, the molecular era began in 1960 when Nowell et al described a minute chromosome in human chronic myeloid leukemia, in 1973 recognized as the t(9;22) translocation. With continued technologic advancement, basic science further defined CML molecular pathogenesis as due to unregulated signal transduction by the BCR-ABL tyrosine kinase. This constitutively activated kinase results from the fusion of the BCR (breakpoint cluster region) gene on chromosome 22 to the ABL (Ableson leukemia virus) gene on chromosome 9. This is the sine qua non of CML. It also provides prognostic and predictive information. In regards to the latter, it predicts response to Imatinib, an orally available ABL kinase inhibitor that can induce hematologic and cytogenetic remission in all stages of CML.

Breast Carcinoma

Similar advancements in prognostic and predictive factors as well as tailored treatments have occurred in breast cancer. This began at the onset of the 20th century with recognition of the prognostic significance of tumor grading and histologic subtypes and it culminated at the end of the century with the identification of hormone (estrogen and progesterone receptors) and growth (HER2) receptors as therapeutic targets. It is interesting to note, that methods for the semiquantitative assessment of breast tumor grade were described before a full classification of histologic types was established. Today, the prognostic factors required by College of American Pathologists include, lymph node status, tumor size, tumor grade, histologic type, hormone receptor status, skin/chest wall involvement and lymphovascular invasion (optional).

Traditional Prognostic and Predictive Factors

Introduction

Breast carcinoma is the second leading cause of cancer death in women. The pathologist’s role in patient management is to correctly diagnose and classify the tumor and then provide key prognostic information to the clinician and patient: tumor type, tumor size, tumor grade, margin status, presence of lymphovascular invasion and axillary lymph node status. This information helps determine whether the tumor will recur, either locally or distantly, and whether the patient may benefit from hormonal and/or adjuvant chemotherapy. Unfortunately, these metrics do not specify the clinical outcome for an individual patient. Breast carcinoma is a heterogeneous disease, with equally heterogeneous outcomes, even among “matched” patients with similar histologic tumors. Invasive duct carcinoma, not otherwise specified, is the largest group of malignant
mammary tumors, constituting 65 to 80% of mammary carcinomas, and within this cohort, patient outcome is extremely varied. Heimann recently articulated the dilemma as follows:

“…Required of tumors is the development of critical phenotypic attributes: growth, invasion, metastagenicity, and angiogenesis…Recognizing tumor heterogeneity emphasizes the need to determine an individual tumor’s place in the evolutionary spectrum. This may be accomplished using clinical feature such as size, nuclear grade, patient age, as well as by examining angiogenesis, metastatic capacity, and proliferation. Identification of the extent of tumor progression with regard to these major tumor phenotypes should allow individual therapy to be fashioned for each patient.”

The future of clinical prognostication and therapy prediction will require, in addition to the “tried and true” traditional clinicopathologic metrics, the discovery of the genes and signaling pathways that are responsible for tumor heterogeneity and differences in patient outcome.

**Tumor Stage: Axillary lymph node status**
The presence of metastatic disease within the axillary lymph nodes is still the single most powerful prognostic factor for primary breast cancer. It has been noted in numerous studies that there is a direct relationship between the number of positive lymph nodes and clinical outcome (Figure 1).

![Figure 1. Relapse-free survival by number of positive axillary lymph nodes (n = 2,873 patients, median follow-up of 37 months) from Chang JC and Hilsenbeck, SG as presented in the AJCC staging manual, 6th Ed.](image-url)

To date, there have been no genes or other factors that have been able to accurately identify *a priori* patients at risk of being node positive; as such, sentinel lymph node biopsy (SLN) continues to be the standard of care.

**Isolated tumor cells: pN0i+**
By definition isolated tumor cells (ITCs) are tumor cell clusters that do not measure larger than 0.2 mm or a single section containing less than 200 tumor cells. The current diagnostic guidelines have been published as well as additional diagnostic clarification by CAP regarding how to make the diagnosis of ITCs versus micrometastases. The CAP
diagrams of possible interpretive scenarios are helpful. The clinical significance of ITCs and micrometastases has been evaluated in several recent manuscripts. Connolly recently pointed out the obvious diagnostic questions that pathologists face when confronted with ITCs: what about spaces between clusters of malignant cells, how are these small clusters with intervening normal lymphoid tissue measured, how close do the deposits need to be in order to sum their measurements and what about diffuse involvement of the lymph node by ITCs, as is often seen with infiltrating lobular carcinoma. In this latter case, the individual cells may be counted and it is recommended to classify such an involved lymph node as positive (pN1a), based on the number of tumor cells. The clinical significance of ITCs was recently demonstrated by de Boer and colleagues and is noted below in the discussion on micrometastases.

**Micrometastases: pN1mi**

Tumor deposits greater than 0.2 mm but not greater than 2.0 mm in largest dimension are termed micrometastases (pN1mi). Cases in which only micrometastases are detected (none greater than 2.0 mm) are classified pN1mi. The prognostic significance of micrometastases is not clear. Retrospective studies have reported decreases in disease-free survival ranging from 10% to 22% in some subgroups of patients where micrometastatic axillary disease was detected by immunohistochemical studies. Most recently, at the Breast Cancer: Current Controversies and New Horizons Meeting, it was reported that patients with micrometastases after chemotherapy fared as poorly as patients with macrometastases; whereas, in patients with micrometastases prior to treatment, the prognostic significance has yet to be determined. Recently the Dutch presented their data from before 2006 and had breast cancer with favorable primary-tumor characteristics and isolated tumor cells or micrometastases in the regional lymph nodes. The primary end point was disease-free survival. 856 patients were node-negative disease and had not received systemic adjuvant therapy (the node-negative, no-adjuvant-therapy cohort), 856 patients with isolated tumor cells or micrometastases and had not received systemic adjuvant therapy (the node-positive, no-adjuvant-therapy cohort), and 995 patients with isolated tumor cells or micrometastases who had received such treatment (the nodepositive, adjuvant-therapy cohort) were studied. The median follow-up was 5.1 years. The adjusted hazard ratio for disease events among patients with isolated tumor cells who did not receive systemic therapy, as compared with women with node-negative disease, was 1.50 (95% confidence interval [CI], 1.15 to 1.94); among patients with micrometastases, the adjusted hazard ratio was 1.56 (95% CI, 1.15 to 2.13). Among patients with isolated tumor cells or micrometastases, the adjusted hazard ratio was 0.57 (95% CI, 0.45 to 0.73) in the node-positive, adjuvant-therapy cohort, as compared with the node-positive, no-adjuvant-therapy cohort.

**AJCC Cancer Staging Manual**

The seventh edition of the AJCC staging manual updates the definitions of pN0(i+) and pN0(i-). The updates state that i+ or i- refers to the presence or absence of ITCs detected by any morphologic technique, including hematoxylin-eosin staining and IHC.

- pN0(i-) is: no regional lymph node metastasis histologically, including negative morphologic findings for ITCs (by any morphologic technique, including
hematoxylin-eosin and IHC).
- pN0(i+) is: no regional lymph node metastasis histologically, but positive morphologic findings for ITCs (any morphologic technique, including hematoxylin-eosin and IHC), and no ITC cluster greater than 0.2 mm.

**Tumor Size**

After nodal status, tumor size is one of the most consistently prognostic factors for predicting risk of distant relapse, especially in lymph node negative patients as illustrated from one data set below (Figure 2). Given increased mammographic surveillance, many tumors are noted earlier and this impact on outcome is unclear. Are more malignant tumors being caught earlier, at a smaller size? Does this “stage migration” bias the historical size staging data? The varied outcomes of similarly staged patients by size is most consistent with breast cancer not being a homogeneous disease, but rather a spectrum of disease states, with varying capacities for growth and metastasis.

After >15 years, untreated lymph node negative, ER positive patients with tumors less than 2.0 cm still have a risk of distant recurrence of 25%. The Memorial Sloan Kettering Cancer Center group has advocated a 12% risk of relapse for tumors less than 1 centimeter\(^\text{14}\). Even tumors less than 1.0 cm have a 10% chance of distant metastasis and such patients may benefit from adjuvant chemotherapy\(^\text{15}\). Thus systemic therapy is now considered even for patients with small node-negative breast cancer.

![Figure 2. Relapse free survival by tumor size for node-negative patients (n = 1,613 node-negative patients, median follow-up of 43 months) from Chang JC and Hilsenbeck, SG as presented in the AJCC staging manual, 6th ed)\(^\text{16}\).](image)

**Tumor Grade**

Semiquantitative assessment of breast tumor grade was described before a full histologic system of classification was established. Tumor grading was contentious over 50 years ago when Willis, an influential figure in pathology, stated that attempts at precise numerical histological grading of tumors were arbitrary, unscientific and wasted effort\(^\text{17}\). Today, the standard method is the Nottingham Combined Histologic Grade (NCHG) or Elston-Ellis/modified SBR (mSBR) and it is well known to the reader (Table 1)\(^\text{18-19}\). Although some studies show poor reproducibility and lack of agreement among different observers\(^\text{20}\), when performed by experienced pathologists histologic grade has been
shown to correlate with clinical outcome as is illustrated from one data set below (Figure 3)\textsuperscript{21-25}. According to the College of American Pathologists, all invasive breast cancers, except medullary carcinoma, should be graded.

Table 1. Summary of semi-quantitative method for assessing histological grade in breast carcinoma\textsuperscript{26}

<table>
<thead>
<tr>
<th>Feature</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tubule formation</td>
<td></td>
</tr>
<tr>
<td>Majority of tumor (&gt;75%)</td>
<td>1</td>
</tr>
<tr>
<td>Moderate degree (10-75%)</td>
<td>2</td>
</tr>
<tr>
<td>Little or none (&lt;10%)</td>
<td>3</td>
</tr>
<tr>
<td>Nuclear pleomorphism</td>
<td></td>
</tr>
<tr>
<td>Small, regular nuclei</td>
<td>1</td>
</tr>
<tr>
<td>Moderate increase in size &amp; variability</td>
<td>2</td>
</tr>
<tr>
<td>Marked variation</td>
<td>3</td>
</tr>
<tr>
<td>Mitotic counts</td>
<td></td>
</tr>
<tr>
<td>Dependent on microscopic field</td>
<td>1-3</td>
</tr>
</tbody>
</table>

Perhaps the most poorly reproducible component is the assessment of nuclear pleomorphism and as such deserves a word. According to Elston and Ellis:

“…to introduce a degree of objectivity we have suggested that the size and shape of normal epithelial cells present in breast tissue adjacent to the tumor should be used as a reference point….When the tumor nuclei are small, with little increase or variation in size compared with normal nuclei and have regular outlines and uniformity of nuclear chromatin, one point is appropriate. A score of 2 points is given when the nuclei are larger than normal, have more open vesicular nuclei with visible, usually single, nucleoli and there is a moderate variation in size and shape. A marked variation in size and shape, especially when very large and bizarre nuclei are present, scores 3 points; nuclei are vesicular with prominent enlarged and often multiple nucleoli\textsuperscript{26}.”

Figure 3. Correlation between histologic grade and overall survival in 2005 patients with primary operable carcinoma of the breast from Elston & Ellis’s The Breast\textsuperscript{26}.

**Histologic Subtypes**

Pure papillary, tubular and mucinous morphologies identify a small subset of patients with a better prognosis than infiltrating ductal carcinoma. However, the majority of
tumors (80-90%) are ductal carcinomas of no special type (NOS) and the risk of recurrence with ductal carcinoma, NOS is quite variable.26

**Hormone Receptors: An Introduction**

More than one hundred years ago Beatson noted that an oophorectomy increased survival in some patients with advanced breast carcinoma.27 Almost 40 years ago researchers first noted that radiolabeled estrogens were preferentially concentrated in human breast cancers and that many breast cancers are dependent on estrogen and/or progesterone for growth.28 The receptors for estrogen and progesterone (ER/PR) are highly expressed in the majority malignant breast carcinomas (~75%). ER exerts effects through:

1. Classical ligand-dependent pathway in which the ER complex regulates gene transcription through its interaction with ERE consensus DNA sequences.
2. Ligand-independent pathway in which growth factors and their tyrosine kinase receptors may activate ER and increase the expression of ER target genes in the absence of estrogen.
3. DNA binding-independent pathway in which induction of gene regulation by ER complexes is through interactions with no ERE-like promoter elements such as AP1, SP1, and CREs.
4. Cell surface (nongenomic) signaling in which estrogen activates a putative membrane–associated binding site that generates rapid tissue responses.

However, the details of the estrogen effect on downstream gene targets, the role of cofactors, and crosstalk between other signaling pathways are still largely unknown. What is known is that all four mechanisms contribute to ER proliferative and survival effects. As the estrogen receptor has become the target of hormonal therapy, estrogen receptor status is now recognized as a strong predictive factor. Progesterone receptor status has been variously reported as prognostic and/or predictive.29 Of great significance is that patients with hormone receptor-positive disease do not have a uniform response to endocrine therapy.

**The estrogen and progesterone receptors**

There are two estrogen receptors, ERα (chromosome 6) and ERβ (chromosome 14) and two progesterone receptors, PRA and PRB. They share a common structural and functional organization: their functional domains have been designated A-F (ERα, figure 4). The ERα gene is complex, 140 kb with 8 exons and it has 595 amino acids with a central DNA-binding domain (DBD) and a carboxyl-terminal, hormone binding domain (HBD). Binding of estrogen to ER activates the receptor with coincident dissociation of chaperonin proteins, i.e., heat shock protein 90. Hormone-bound ER then dimerizes and binds to estrogen response elements (ERE’s) in the promoter of estrogen-responsive genes. Through binding the DNA, ER influences expression of estrogen-responsive genes such as PR. There are two distinct activation domains: (1) a hormone independent domain in exon 1 and 2 (AF-1); and, (2) a hormone dependent, carboxyl-terminal activation domain is contained in portions of exon 4 through 8 (AF-2). AF-1 and AF-2 activate transcription independently and/or synergistically. Both are needed for maximal ER transcriptional activity. Tamoxifen inhibits AF-2 activation and thereby ER activity but tamoxifen does not prevent activation of AF-1. Additionally, when AF-2 is not
required and AF-1 is sufficient for ERα activity, then tamoxifen can function as a partial agonist\(^30\).

![ERα Functional Domains Diagram](image)

**Figure 4:** A diagram of ERα functional domains labeled A through F. The central DNA-binding domain (DBD) and hormone-binding domain (HBD). Regions important for dimerization (dimer) and transactivation functions (AF-1, AF-2a, AF-2), respectively. Region A/B (AF-1) is key for hormone-independent ER transcription, region F is important for modulation of ER activity (AF-2).

**ER Negativity**
Tumors lacking the estrogen receptor have been postulated to be negative due to: genomic deletion or rearrangement, gene methylation and lack of gene transcription. Most ER negative tumors show little ER mRNAs\(^31\). Variant ERα isoforms due to exon deletions have been described. Exon 5 deletion, involves the hormone-binding domain and results in constitutive, hormone-independent activity. Missense mutations have been described but appear rare. Those reportedly of significance include tyrosine 537 asparagine (causing constitutive activation).

**Estrogen receptor β**
On the one hand ERβ has been associated with clinical tamoxifen resistance and is thought to enhance tumor aggressiveness; while, on the other hand, most studies have found it to be a marker of good prognosis. Critics contend that the studies have not been powered appropriately and larger studies appear to be required. ERβ is generally coexpressed with ERα (76% ERα+/β+, 14% ERα+/ERβ-, 15% ERα-/ERβ+)\(^32\).

**Methods of measuring hormone receptors**
Most of the data on the clinical utility of ER content have been generated using biochemical ligand binding assays (LBAs), such as the dextran-coated charcoal assay (DCCA). Since the first report of ER’s independent prognostic significance almost two decades ago, the assessment of ER status by DCCA has been validated repeatedly and is generally regarded as the standard by which other methods are assessed\(^33\). Currently the validated method of choice is immunohistochemistry, performed on fixed, paraffin embedded tissue\(^34\). An emerging method is by quantitative RT-PCR to measure mRNA levels of ER and PR\(^35\).

**Methods of measuring hormone receptors: The ligand binding method**
The ligand binding method (DCC) measures radiolabeled steroid (ligand) bound to the homogenized breast tissue. Using a standardized Scatchard plot, total concentration of receptor protein in the cytosol is obtained and expressed as femtomoles of receptor protein per milligram of total cytosol protein. The most common cut off for considering
a tumor positive was >3.0 fmol/mg (others have used >10 fmol/mg). Specimens needed to be immediately frozen and shipped on dry ice to prevent protein degradation. The tumor needed to be reviewed so that representative portions, from 0.5 to 1.0 cm in diameter could be chosen. Variability in results was caused by inclusion of benign tissues, tumor receptor heterogeneity, pregnancy, delayed time to tissue freezing or tamoxifen usage. DCC was complicated, required central laboratories and used low-level radioactive materials.

Methods of measuring hormone receptors: Immunohistochemistry
Immunohistochemistry (IHC) uses antibodies specifically directed against epitopes unique to each receptor protein and can be used on all fixed specimens as well as frozen sections. Correlation between DCC and IHC is high, from 80–90%36. PR has been compared, though less exhaustively and concordance is 70–80%37. Problematic is variability of result and as many as 10% of cases may be incorrectly classified. This may occur due to fixation differences (fixative, duration of fixation), differences in assay protocol (particularly antigen retrieval, widely thought to be the most important factor in assay variability38) and antibody clone. Additional problems include that the method is semiquantitative and interpretation is subjective. Attempts to automate staining and the use of image analysis systems, such as the ACIS II (DakoCytomation), are steps toward greater objectivity in quantitation of hormone receptor expression.

Protocol: Estrogen Receptor
Consistent ER results have been repeatedly obtained and reported from tissue sections fixed for 6 to 8 hours in 10% neutral buffered formalin39. The debate over the best monoclonal antibody continues. The most widely used antibodies are 6F11 and 1D5, both mouse monoclonal antibodies and SP1, a rabbit monoclonal with reported increased affinity. 1D5 is perhaps the most widely used and clinically validated antibody40. Recently there has been concern that ER assessment using the 1D5 antibody misclassifies ~5-10% of breast cancers as ER negative (Cheang, M et al; manuscript submitted to JCO). Cheang and colleagues contend that SP1 is more sensitive, better predicts tamoxifen response and is better correlated with the ligand binding assay (figure 5). The new ER/PR pharmDX assay is a FDA 510(k) cleared assay and includes both 1D5 and a second ER antibody, ER-2-123.
Figure 5. Comparison of ER antibodies SP1 and 1D5. The SP1+/1D5- cases have cumulative survival at 10 years similar to those of SP1+/1D5+ tumors (Cheang M, J Clin Oncol. 2006 Dec 20;24(36):5637-44).

Clinical Validation: Estrogen receptor as a continuous variable
To identify a clinically meaningful cut point for defining ER-positive tumors, Allred et al developed and used an eponymous scoring system that measures both the proportion of cells staining (5 pts, 100%; 4 pts, 66%; 3 pts, 33%, 2 pts, 10%, 1pt, 1%; 0 pts, 0) and their intensity (3 pts, strong; 2 pts, intermediate; 1 pt, weak)40. The proportion of cells staining was added to the intensity of staining and results in a score from 0-8. Using disease free survival curves, tumors were defined as ER-positive if their total IHC score was greater than 2 and ER-negative if their score was 0 to 2. Note that an Allred score of 3, the lowest possible positive score, corresponds to as few as 1% intermediately staining tumor cells to 10% weakly staining tumor cells. For patients receiving no systemic adjuvant therapy (n = 701), ER status was only a weakly prognostic factor. For patients who received adjuvant endocrine therapy, either alone (n = 517) or in combination with chemotherapy (n = 260), ER status was a highly significant predictive marker of disease free survival (DFS). For these latter two groups combined (n = 777), the best cut point (IHC score > 2) was highly significant (P > 0.0001) (Figure 6). It is not yet known whether other laboratories performing this assay would obtain similar results, or a similar cut-off point.

Figure 6. Disease free survival (DFS) curves for all possible IHC scores within the different treatment groups. For patients who received adjuvant endocrine therapy, either alone (n = 517) or in combination with chemotherapy (n = 260), ER status was a highly significant predictor of DFS. For these latter two groups combined (n = 777), the best cut point (IHC score > 2) was highly significant (P > 0.0001) (Figure 6). It is not yet known whether other laboratories performing this assay would obtain similar results, or a similar cut-off point.

Clinical Validation: Progesterone Receptor
Progesterone receptor (PR) is a surrogate marker of estrogen receptor activity in breast cancer37. Its utility in predicting clinical outcome has been established using biochemical assays. Allred, et al validated an immunohistochemical assay for PR in breast cancer using the monoclonal antibody 1294, the slides were scored microscopically using the Allred score on a scale of 0–8. The assay was compared to ligand-binding assay in 1235 breast cancers, and a subset (n = 362) that received only hormonal therapy was used to
define a cutoff for progesterone receptor positive. Clinical utility was validated in an independent set of samples (n = 423) from a clinical trial randomizing premenopausal breast cancer patients to tamoxifen + oophorectomy vs observation following surgery. A cutoff of Allred score > 2 (corresponding to > 1% positive cells) dichotomized patients with significantly better or worse clinical outcome (P = 0.0014). Progesterone receptor by immunohistochemistry provided significantly better results than progesterone receptor by ligand-binding assay in predicting clinical outcome. In the clinical trial, a positive result in univariate analyses was associated with significantly improved disease-free and overall survival both in untreated and hormonally treated patients. Positive progesterone receptor remained significant for improved disease-free and overall survival in multivariate analyses including the standard variables of tumor size, nodal status, treatment, histological grade, and HER2 status (PR remained a strongly prognostic factor). Estrogen and progesterone receptors were noted to be codependent variables and progesterone receptor was a weaker predictor of response to endocrine therapy than estrogen receptor when both were included in multivariate analysis (PR is only weakly predictive of hormone therapy).

Debates: Hormone Receptor Testing a Continuous Variable or Dichotomous?

Recently Collins et al analyzed 825 cases using the Allred score (figure 7). They noted that in 817 cases (99%), tumor cells showed either complete absence (score 0) or strong immunostaining of 70% or more of cells (score 7 or 8). They concluded that with the immunohistochemical method used in their laboratory, ER staining is essentially bimodal. The overwhelming majority of breast cancers were either completely ER– or unambiguously ER+, and cases with weak ER immunostaining were rare. The debate over whether ER expression is continuous or dichotomous continues.

![Figure 7. Frequency distribution of the percentage of cells showing “bimodal” or dichotomous nuclear staining for estrogen receptor among 825 primary breast cancers.](image)

Endocrine Therapy

Although the levels of ER are weakly prognostic with regards to breast cancer recurrence and overall survival survival, the true clinical utility of ER is that it is a as powerful, predictive marker for benefit of endocrine therapy. The Early Breast Cancer Trialists’ Collaborative Group data have confirmed the survival benefit for the average patient with early-stage disease (especially younger women) by tamoxifen for any patient with ER-positive disease regardless of age. Aromatase inhibitors have been show in the Arimidex Versus Tamoxifen Alone or in Combination (ATAC) trial and the Breast International Groups (BIG) 1-98 trial to improve disease-free survival (DFS) over
tamoxifen, but thus far none have demonstrated a significant improvement in overall survival.

Tamoxifen resistance in ER-positive breast cancer
The expression of ER is the most important factor in predicting endocrine response and classification of ER-positive tumors is made by detecting ER-α. However, ER expression alone does not accurately predict response to endocrine therapy as response to tamoxifen in ER positive patients is less than 50%. ER status/expression does remain positive in most tumors which initially are tamoxifen responsive but then develop resistance. Interestingly, over two thirds of patients who develop resistance will respond upon changing endocrine therapies.

There are many theories about the cause(s) of tamoxifen resistance. Perhaps most likely is activation of AP-1-response elements that regulate genes involved in cell proliferation, motility, and apoptosis. Enhanced AP-1 activation has been associated with tamoxifen resistance. Recently, nongenomic mechanisms for ER action have been postulated as alternative mechanisms for ER-positive endocrine resistance. ER can cause rapid tyrosine phosphorylation of SHC. A transient activation of extracellular-signal-regulated kinases ERK1 and ERK2 is mediated by ER. This may result in different cell survival and proliferative signals via the AKT and mitogen-activated protein kinase (MAPK) pathways (figure 8). Additionally, tamoxifen may exert agonist effects via the interaction of membrane ER with growth factor receptor signaling (epidermal growth factor receptor or HER2). Finally, “crosstalk” may occur in ER-positive tumors with HER2 amplification as activation of mitogen-activated protein kinases (MAPK), ERK 1 & 2 by overexpressed HER2 may result in phosphorylation of ER, resulting in tumor stimulation rather than inhibition.
Acquired endocrine resistance
EGFR and HER2 seem to become selectively upregulated in breast cancer cells that acquire resistance to tamoxifen during prolonged exposure. Overexpression of HER2, amplified in about 10% of ER-positive breast cancers, results in loss of tamoxifen’s ER antagonist activity and the acquisition of tamoxifen-stimulated growth. Tamoxifen, like estrogen, may activate HER2 via the membrane functions of ER, which, in turn, phosphorylate both ER and AIB1 (SRC3), an important ER coactivator.

The cumulative data from clinical studies supports the hypothesis that overexpression of HER2, EGFR, and high levels of phosphorylated AKT or ERK, appear to contribute to tamoxifen resistance in some patients. Aromatase inhibitors (estrogen deprivation), on the other hand, may be more effective in such tumors since ligand deprivation would shut off both nuclear and membrane ER activity, thereby eliminating the cross talk generated in the presence of estrogen or tamoxifen.

Other markers
Other markers of prognostic significance have been p53 mutated forms, patient age, patient menstrual/menopausal status, race, epidermal growth factor receptor (EGFR), epidermal like growth factor B-2 (HER-2/neu), Bcl-2 and plasminogen activators. These are outside this discussion.

21st Century Prognostic and Predictive Factors
Roadmap
The ensuing discussions will examine new molecular methodologies and their role in defining the current breast carcinoma taxonomy and new and/or improved prognostic and predictive markers. Technology is the driving force in many of these developments and it includes quantitative real time polymerase chain reaction (QRT-PCR), gene expression microarrays and comparative genomic hybridization (CGH). These techniques will be introduced below. CGH involves examining tumor DNA, looking for large genomic gains and deletions. Expression microarrays were first described in the mid-1990s as a way to examine the expression (mRNA levels) of thousands of genes simultaneously. QRT-PCR is a robust, precise, sensitive and reproducible way to quantify tumor mRNA. The study of DNA is referred to as genomics while the study of mRNA is referred to as functional genomics.

Technology Introduction:
*Microarray analysis: What is it and why is it valuable?
As morphologists, pathologists are experts at examining fixed paraffin sections that have been stained with hematoxylin and eosin. By identifying mitotic figures the pathologist is able to assess the proliferative activity of a cell. Immunohistochemical stains for the Ki-67 protein allow for semiquantitative monitoring of the growth fraction of normal and neoplastic cells. DNA microarray analyses measure quantities of individual messenger RNA, that is, gene expression. Using this method, a snapshot of all the proliferation genes that a cell is transcribing can be examined and quantified at once. Indeed, the entire genome maybe surveyed at once. DNA microarray analysis was first used to compare different tumors in order to identify genes and functional gene groups that differ in relative expression, i.e., the estrogen receptor and associated genes, as they differ between ER+ and ER- tumors. Such comparisons in gene expression have enabled identification of genes relevant to cancer progression, to new tumor subclasses, and to the identification of biomarkers associated with disease.

**DNA Expression Microarrays: “Expression Arrays”**

**Functional Genomics**

Gene microarray analysis allows for the quantification of thousands of unique messenger RNA (mRNA), obtained from fresh frozen tissue samples, on a single slide or chip. From an individual tumor, levels of mRNA can be detected, quantitated and compared among multiple tumor samples. This method is excellent for assessing correlations in gene expression, i.e., expression patterns, obtained from thousands of genes which can be assessed on a single array.

DNA microarrays are available commercially and all are based on hybridization of nucleic acid strands but there are differences between platforms. An important distinction is the length of the probe. Probes are gene-specific and represent part of a gene. Microarrays may be classified as either: cDNA arrays with probes up to a thousand base pairs (mer) or oligonucleotide arrays using short 25-30mers or long oligonucleotide (60-70mer) probes. The probes can be either contact-spotted, ink-jet deposited or directly synthesized on the substrate. An array approximates the size of a glass microscope slide and it typically has thousands of individual genes represented on its surface, each gene represented multiple times (different segments of the gene at each site respectively). Thousands of copies of the probe are printed or synthesized (“arrayed”) at each site on an inert substrate (such as a glass slide or cartridge) in spatially specific locations.

**Gene Arrays: Technical Comments**

For the most part, DNA expression arrays require fresh frozen tissue (new assay, DASL: cDNA-mediated annealing, selection, extension, and ligation appears to be a promising exception). The sensitivities of the various platforms range from being able to detect from 1 to 10 copies per cell. They measure relative or absolute transcript concentrations of genes above the sensitivity level of the microarray (rendering ~1/2 of the transcriptome beyond reach of arrays). They have modest reproducibilities (the best reported for Affymetrix, Agilent and Codelink are 0.9). Cross-platform correlations can be difficult due to only moderate correlation coefficients.
**Gene Expression Arrays: Methodology**

- Typically RNA is extracted from fresh frozen tissue and reverse transcribed into cDNA.
  - Some newer assays are available for research purposes using fixed paraffin embedded tissues.
- cDNA is labeled with a detectable fluorescent dye, the efficiency of this labeling step is crucial for assay precision and reproducibility.
- The cDNA is placed on the array and allowed to hybridize to the arrays.
  - Individual molecules hybridize to complementary *gene-specific* probes on the array.
- Images are captured with the use of confocal laser scanning.
- The relative fluorescence intensity of each gene-specific probe is a measure of the level of expression of the particular gene.
  - The greater the degree of hybridization, the more intense the signal, implying a higher relative level of expression (more copies binding to the array).
- Comparisons from different laboratories and across various types of microarray may differ significantly.
- After collection, the data are normalized for comparison between the different assays.
  - Normalization compensates for differences in labeling, hybridization, and detection methods.
  - Normalization and filtering transformations must be carefully applied due to effects on the results.
- The data are then filtered by objective criteria or statistical analyses to select expression levels that correlate with particular groups of samples.
- Different methods of statistical analysis applied to the same data set can produce different sets of significant genes.
  - Validation is important (confirmatory testing in a second series of patients/samples).
- Caution should be exercised in comparing data sets from different laboratories. Comparisons of published lists of genes can produce discordant results, because they rarely take into account the differences in the methods of analysis of the data.

**Comparative Genomic Hybridization: Methodology**

Comparative genomic hybridization (CGH) provides an overview of DNA sequence copy number changes (losses, deletions, gains, amplifications) in a tumor specimen and maps these changes on normal chromosomes (figure 9). CGH is a powerful method for molecular cytogenetic analysis of tumors. CGH is based on the *in situ* hybridization of differentially labeled total genomic tumor DNA (red) and normal reference DNA (green) to normal human metaphase chromosomes. It may be used with fixed or fresh tissue and the fact that it works with fixed tissues explains its popularity. After hybridization and differential fluorescent staining of the bound DNAs, copy number variations among the different sequences in the tumor DNA are detected by measuring and comparing the tumor to normal fluorescence intensity ratio for each locus in the target metaphase
chromosomes. Each chromosome is represented by a software package as an *ideogram* (figure 9). Areas of *increased* copy hybridization of tumor DNA are represented in red. Areas of *decreased* copy hybridization are represented in green.

CGH is excellent for low resolution maps of genomic gains and losses. It is a poor method if a high resolution map is required (sensitivity of ~50 kb).

Figure 9. This is an example of the hybridization of reference DNA (green) and tumor DNA (red) to an interphase spread of chromosomes (ready for mitosis and frozen by adding chlokicine to stop spindle apparatus). Above the red line shows amplification of tumor gene copy and below the green line shows loss of gene copy number.

CGH is particularly useful for analysis of DNA sequence copy number changes in common solid tumors where high-quality metaphase preparations are often difficult to grow, and where complex karyotypes with numerous markers, double minutes, and homogeneously stained chromosomal regions are common. CGH only detects changes that are present in a substantial proportion of tumor cells (i.e., clonal aberrations). It does not reveal translocations, inversions, and other aberrations that do not change copy number. Gains or losses less than 50 kb will not be identified. At present, CGH is a research tool that complements previous methods for genetic analysis. CGH will advance understanding of the genetic progression of cancer and highlight important genomic regions for further study.

**Array CGH**

Array-based CGH (array-CGH or a-CGH) is a modification of standard chromosomal CGH using cloned fragments of DNA spotted onto glass slide arrays as the target of hybridization, rather than normal metaphase spreads. As in conventional CGH, differentially labeled tumor and normal reference DNA are hybridized to this array, and a normalized ratio of tumor to reference intensity is calculated for every clone. This application takes advantage of the mapping information and the cloned DNA fragments generated by the human genome project. The DNA, which is spotted onto glass slides to create CGH arrays, is prepared from bacterial artificial chromosomes (BACs), which contain ~100,000 base pairs of cloned normal DNA derived from carefully mapped regions of the human genome. Current CGH arrays include BACs coding for 100 kilobases of DNA distributed on average every million base pairs (10% coverage, 1 megabase resolution). The next generation of CGH arrays currently being developed and validated will contain ~30,000 BACs of 100 kb each, covering 100% of the genomic sequence. Again, this is useful for fixed tissues.
Quantitative Real Time Polymerase Chain Reaction
This method allows for the quantitation of mRNA transcripts. This characterizes the cell’s gene expression at the moment the cell was fixed. It is highly sensitive, specific, precise and reproducible. It is generally considered to be the “gold-standard” for measurement of gene expression. Figure 10 describes one widely used QRT-PCR method, the TaqMan assay (Applied Biosystems; Foster City, CA). In brief:
1. mRNA is extracted from a fixed or fresh tissue.
2. Complementary DNA (cDNA) is made from the total RNA using oligo-dT priming, random priming, or gene-specific primers.
3. The cDNA for the specific target is amplified by PCR using both a gene specific probe and pair of gene specific primers.
4. The probe has both a fluorescent reporter and a quencher attached to either end so there is no fluorescence signal when the probe is intact.
5. The specific primers and probe anneal to the specific target gene.
6. Due to the 5’-exonuclease activity of the Taq polymerase, the bound fluorescent-labeled probe is degraded while new PCR product is being synthesized. This process releases the fluorescent tag from the quencher and generates a fluorescence signal that is directly proportional to the amount of PCR product in the tube.
7. Real-time polymerase chain reaction thermocyclers can detect the abundance of fluorescence and thus determine the relative amount of mRNA present in a given sample.

Introduction: 21st Century Prognostic and Predictive Factors
Gene expression profiling defines, at the molecular level, the unique gene expression inherent to many kinds of tumors. One of the common features of these studies has been the emergence, through unsupervised, hierarchical clustering analysis, of breast cancer subtypes with distinct gene expression patterns for each subtype. The differences in thousands of individual gene expression patterns among these subtypes likely reflect basic differences in the cell biology of the tumors; and therefore, these molecular subtypes may be considered as separable diseases. The molecular differences between the tumor subtypes are often accompanied by differences in clinical features, such as statistically robust differences in relapse-free and overall survival. This certainly is not a new idea. As early as 40 years ago, Fox noted in JAMA that the survival curves of
women with breast cancer suggests that two or more populations exist, with about 40% suffering fatal outcome unaffected by treatment. The remaining 60% exhibit a relative mortality only modestly different from that of women of similar ages without evidence of disease. He speculated that increasing detection of an entity that is histologically defined as malignant but biologically relatively benign could account for the observed increase in incidence\textsuperscript{56}.

When an alternative approach, analysis \textit{supervised} by outcome data, is used many groups identify smaller numbers of individual genes whose expression is associated with prognosis and treatment response\textsuperscript{57-58}. These genes define potential prognostic and predictive molecular markers without respect to the biological diversity represented by the subtypes.

**A New Breast Taxonomy: Molecular Subtypes using Unsupervised Classification**

\textit{Unsupervised Molecular Taxonomy of Breast: Luminal A/B, HER2 & Basal Subtypes of Breast Carcinoma}

Perou and Sorlie first reported a molecular taxonomy of breast cancer based on variation in global gene expression patterns measured by cDNA microarrays, using \textit{unsupervised} statistical correlations in expression patterns. This demonstrated that tumors could be grouped into molecular types distinguished by unique gene expression patterns. The types include: luminal A and B (ER positive), HER2 (Erb-B2 positive) and basal type (ER/PR/HER2 negative or “triple negative”). These types differ in disease outcome and therapeutic response. In further studies using a directed or \textit{supervised} approach, they identified and attempted to reconcile the definition of those subtypes and the accompanying differences in disease outcome (figure 11). Perhaps most exciting about these elegant studies was the identification of two types of ER positive breast carcinoma with differing outcomes. The luminal A type breast cancers patients were noted to have significantly better clinical outcomes with regards to recurrence and overall survival than those ER positive tumors classified as luminal B.

![Figure 11. Gene expression patterns of 85 experimental samples representing 78 carcinomas, three benign tumors, and four normal tissues, analyzed by hierarchical clustering. The closer samples are together, the more similar are their expression profiles. The tumor specimens were divided into six subtypes based on](image)
differences in gene expression. Note differences in ER- and ER+. The cluster dendrogram shows the six subtypes of tumors from left to right (colored as): basal-like, red; ERBB21, pink; normal breast-like, green; luminal subtype C, light blue; luminal subtype B, yellow; luminal subtype A, dark blue54.

Confirmation of Taxonomy: Evaluation of Breast Cancer Data From van 't Veer et al
Van 't Veer et al used gene expression data from 24,480 genes in a set of 117 tumors from young breast cancer patients. Again a supervised method of hierarchical clustering was used, exactly as described for the Perou/Sorlie data, to display the expression patterns of 461 intrinsic genes in the 97 tumor samples that were obtained from patients diagnosed with sporadic cancer (figure 12). Individual dendrogram branches are colored according to the strongest correlation of the corresponding tumor with the subtype centroid as defined for the Perou/Sorlie samples. As in the Perou/Sorlie data, the best discrimination was between tumors that expressed genes in the luminal A cluster at high levels (many related to ER and ER associated genes) and the tumors that were negative for these genes and exhibited expression profiles characteristic of either the basal or the HER2 types. All samples that showed the strongest correlation with the basal subtype (red branches) are all contained within the left branch of the dendrogram in a tight cluster. The luminal A/luminal B distinction, though much less clear than the basal versus luminal distinction, is also seen, with many of the luminal B tumors clustering near each other on the right branch of the dendrogram.

![Figure 12](image-url)
expression profiles characteristic of the basal, HER2 or luminal B subtypes. The strongest correlation was with the basal subtype (red branches) – all of which are contained within the left branch of the dendrogram in a tight cluster.

Tumor Subtypes Are Associated with Significant Difference in Clinical Outcome.
In the previous work of Sorlie et al, the expression-based tumor subtypes were associated with a significant difference in overall survival as well as disease-free survival for the patients suffering from locally advanced breast cancer and belonging to the same treatment protocol. To investigate whether these subtypes were also associated with a significant difference in outcome in other patient cohorts, they performed a univariate Kaplan–Meier analysis with time to development of distant metastasis as a variable in the data set comprising the 97 sporadic tumors taken from van’t Veer et al. As shown in figure 13, the probability of remaining disease-free was significantly different between the subtypes; patients with luminal A type tumors lived considerably longer before they developed distant disease, whereas the basal and HER2 groups showed much shorter disease-free time intervals. The methodological differences prevent a definitive interpretation, but it is notable that the order of severity of clinical outcome associated with the several subtypes is similar in the two dissimilar cohorts. This implies that these differences are real and are due to difference in tumor biology.

Figure 13. Kaplan-Meier analysis of disease outcome in two patient cohorts. A. time to development of distant metastasis in the 97 cases from van ’t Veer et al. Patients were stratified according to the subtypes as shown in Figure 14. B. Overall survival for 72 patients with locally advanced breast cancer in the Norway cohort. The normal-like tumor subgroup was omitted in both analyses.

Breast Tumor Subtypes Represent Reproducible Distinct Biological Entities
Gene expression studies show that there is considerable diversity among breast tumors, both biologically as well as clinically. This is not a new idea, as epidemiological studies previously had inferred the existence of two or more subpopulations of breast cancer. A parsimonious interpretation of the reproducibility of several different patterns of gene expression is to regard each as representing a different biological entity. One exciting possible basis for the differences in these patterns between tumor subtypes may be that they originate from different mammary epithelial cell types. The findings support this interpretation, as breast tumor subtypes with patterns of gene expression similar to those
of luminal epithelial cells (the cells that line the duct and give rise to the majority of breast cancers) and patterns of at least one other subtype (basal) seem to resemble the pattern found in basal epithelial cells of the normal mammary gland (characterized by expression of cytokeratins 18 & 19 for luminal cells and cytokeratins 5/6 & 17 for basal cells).

Conclusion: New Breast Taxonomy
Luminal and basal tumor subtypes appear to be distinct biological entities, as the expression patterns have been show to be detectable in other genome-scale studies of breast cancer. Sorlie et al have found strong evidence for the universality of a distinction between basal-like and luminal-like subtypes in three independent data sets comprising different patient populations whose gene expression profiles had been determined by using different microarray technology platforms. They also found considerable evidence, in one of the studies, for the distinction between the luminal A and B subtypes. The fact that these distinctions for the basal and luminal subtypes are reproducible (less so for the luminal B subtype versus luminal A) means that the substantial differences in the characteristics of the patients (e.g., age and tumor stage) are less important determinants of tumor expression phenotypes than intrinsic biology.

Genomic Classifiers: Prognostic and Predictive Signatures

Introduction
The first published breast cancer classifier was that of van ‘t Veer et al from the Netherlands. This was followed by an independent validation study and a subsequent validation study. This work resulted in a commercially available clinical product, Agendia’s MammoPrint assay. Similarly, the Genomic Health Oncotype DX assay was developed, refined and independently validated. It too resulted in a commercially available clinical product. There are other classifiers in various stages of development and validation. These include a two gene ratio from Sgroi et al, a 76 gene classifier from Veridex as well as a classifier being developed by Celera. For the purposes of this discussion, we will limit it to the first two classifiers, which have been independently validated in more than one study, and which are both being used in two large prospective trials, in Europe (the MammoPrint assay in MINDACT: microarray in node negative disease may avoid chemotherapy) and the United States (the Oncotype DX assay in TAILORx: trial assigning individualized options for treatment). The discussion will provide the basis for evaluating these and other emerging classifiers.

Classifier Development
There is a large body of literature on prognostic factors for cancer patients unfortunately only a few are used in clinical practice. They are unlikely to be used unless they are therapeutically relevant and most publications do not establish such therapeutic relevance. Most prognostic factor studies are conducted by use of a convenient sample of patients for whom tissue is available, and the cohort is often far too heterogeneous with regard to age, ER status, stage and treatment to support therapeutically relevant conclusions.
As there are many new classifiers in development, it is useful to divide genomic classifier studies into developmental studies and validation studies. Developmental studies define the genes and the algorithms used in multi-gene classifiers and are analogous to phase I-II clinical trials while validation studies that use prespecified genes, algorithms and analysis plans are analogous to phase III clinical trials.

In development studies, one significant concern is that the number of candidate genes studied by DNA expression arrays for use in the classifier is typically much larger than the number of cases available for analysis. In these studies it is always possible to find genes or sets of genes that perfectly classify the data on which they were developed. This apparently perfect classification can be achieved even if there is no relationship between expression of any of the genes and outcome. Therefore, even in developmental studies, both control for false discovery of genes and validation of the selected genes and classifier is necessary. “Internal validation” may be accomplished by: (1) splitting the data into two portions, one used for training the model and the other used for testing the model; or (2) cross-validation that is based on repeated model development and testing on random data partitions. However, neither of these statistical methods provides true “independent” validation. Thus, these methods for internal validation do not constitute external validation of the classifier in a setting simulating broad clinical application.

In an editorial to JNCI, Dr. Simon, Chief, Biometric Research Branch, NIH recommends the following roadmap for developing a genomic classifier:

- Developmental and validation studies should be based on cohorts of patients that are sufficiently homogeneous for therapeutically relevant classifiers to be developed...this is best achieved by studying patients who were included in a single, large, randomized clinical trial.
- Developmental studies should be sufficiently large so that they can incorporate either cross-validation or split sample validation and demonstrate that the internally validated prediction error is statistically significantly less than would be expected by chance.
- Independent validation studies are essential before results are accepted into medical practice.
- Independent validation studies should apply the classifier completely specified, including cut-offs, by the developmental study and measure prediction accuracy.
- The size of the validation study should be sufficient so that meaningful confidence intervals on predictive accuracy and positive and negative predictive values can be reported.
- The size of the validation study should be sufficient so that the extent to which the classifier adds predictive accuracy to established prognostic factors can be meaningfully evaluated.

The Mammoprint Breast Cancer Assay: 70 Gene Assay

Breast Cancer Classifier: van’t Veer et al.61
The Netherlands group identified a 70 gene classifier developed using expression microarrays with ~25,000 human genes from fresh tissues obtained from primary breast
tumors of 117 young female patients. This group used supervised classification to identify a 70 gene expression signature strongly predictive of a short interval to distant metastases (poor prognosis signature) in 78 of the lymph node negative patients.

To do this they selected 98 primary breast cancers and tested ~25,000 genes which clustered the tumors on the basis of gene expression. All sporadic patients were node negative and under 55 years of age. RNA was isolated from frozen tissues and was used to derive complementary RNA (cRNA). A reference pool was made from a pool of all sporadic carcinomas. Two hybridizations were made for each tumor using a fluorescent dye reversal technique on microarrays synthesized by inkjet technology on arrays containing approximately 25,000 human genes. Approximately 5,000 genes were significantly regulated across the group of samples (P-value of less than 0.01 in more than 5 tumors). Using unsupervised, hierarchical clustering, they clustered the 98 tumors on the basis of their similarities. They found 2 distinct groups, one with a poor prognosis with distant metastases and another good prognosis group without progressive disease. To gain insight into the genes of the dominant expression signatures they found that the ER negative cases clustered together and there was a second group associated lymphocytic infiltrate, including several genes expressed primarily by B and T lymphocytes.

They then used 78 sporadic lymph node negative tumors to search for a prognostic signature in their gene expression profiles. 44 remained disease free for an interval of at least 5 years (good prognosis group) and 34 patients that developed metastases within 5 years (poor prognosis group) (figure 14).
The 70 gene classifier was identified by using a three-step supervised classification method. The accuracy was improved until the optimal number of genes, 70, were identified. This classifier predicted correctly the actual outcome of disease for 65 out of the 78 patients (83%), with respectively 5 poor prognosis and 8 good prognosis patients assigned to the opposite category. An optimized threshold was established by optimizing the algorithm so that no more than 10% of poor prognosis patients were misclassified. This optimized sensitivity threshold resulted in 15 misclassifications: 3 poor prognosis were classified as good and 12 good prognosis tumors were classified as poor.

The upregulated genes associated with a poor prognosis included those involved in cell cycle, invasion and metastasis, angiogenesis, and signal transduction (e.g., cyclin E2, metalloproteinases MMP9 and MMP1 and the VEGF receptor FLT1). The prognosis classifier was validated in an additional, independent set of 19 young, node negative patients. This group consisted of 7 patients who remained disease free for at least 5 years and 12 patients who developed distant metastases within 5 years. The disease outcome
was predicted by the 70-gene classifier and resulted in 2 out of 19 incorrect classifications using the optimized threshold.

The prediction of the classifier presented would indicate that women under 55 years of age who are diagnosed with lymph node negative breast cancer that have a poor prognosis signature have a 28-fold odds ratio (OR) (95% confidence interval, CI 7-107, \(P=1.0 \times 10^{-8}\)). This classifier provided additional prognostic information in addition to the traditional clinical and histopathological prognostic factors: high grade (OR = 6.4, 95%CI 2.1-19, \(P = 0.0008\)), tumor size greater than 2 cm (OR = 4.4, 95%CI 1.7-11, \(P = 0.0028\)), angioinvasion (OR = 4.2, 95%CI 1.5-12, \(P = 0.01\)), age less than or equal to 40 yrs (OR = 3.7, 95%CI 0.9-6.6, \(P = 0.13\)), and ER negative (OR = 2.4, 95%CI 0.9-6.6, 95%CI 0.9-6.6, \(P = 0.13\)). Multivariate analysis that included all classical prognostic factors indicated that it was an independent factor in predicting outcome of disease (logistic regression OR = 18, 3.3-94, \(P=1.4 \times 10^{-4}\)).  


Using fresh frozen tissue and microarray analysis to evaluate the previously established 70-gene prognosis profile, the Netherlands group classified a series of 295 consecutive patients with primary breast carcinomas as having a gene-expression signature associated with either a poor prognosis or a good prognosis. All patients had stage I or II breast cancer and were younger than 53 years old; 151 had lymph-node–negative disease, 144 had lymph node–positive disease; 155 had tumors less than or equal to 2.0 cm, 140 had tumors greater than 2.0 cm; 99 were ER negative and 226 were ER positive; 185 received adjuvant chemotherapy, 110 did not; 40 received hormonal therapy and 255 did not. They evaluated the predictive power of the prognosis profile using univariate and multivariate statistical analyses.

Among the 295 patients, 180 were found to have a poor prognosis signature and 115 had a good-prognosis signature, and the mean (±SE) overall 10-year survival rates were 54.6 ± 4.4% and 94.5 ± 2.6 %, respectively. At 10 years, the probability of remaining free of distant metastases was 50.6 ± 4.5 % in the group with a poor-prognosis signature and 85.2 ± 4.3 % in the group with a good-prognosis signature (figure 15). The prognosis profile was significantly associated with tumor histologic grade (\(P<0.001\)), the estrogen-receptor status (\(P<0.001\)), and age (\(P>0.001\)) but not tumor size, extent of vascular invasion, the number of positive lymph nodes or treatment.
Figure 15. Kaplan-Meier analysis of the probability that patient would remain free of distant metastases and the probability of overall survival among all patients.

The estimated hazard ratio for distant metastases in the group with a poor-prognosis signature, as compared with the group with the good-prognosis signature, was 5.1 (95% CI 2.9 - 9.0; P < 0.001). This ratio remained significant when the groups were analyzed according to lymph-node status (figure 16). The prognosis profile was a strong independent factor in predicting disease outcome (HR 5.5 among those with poor prognosis signature compared to those with a good signature). The profile was also strongly associated with the outcome in the group of 144 patients with lymph node positive disease, HR for distant metastases = 4.5 (95% CI 2.0 – 10.2; P < 0.001)(figure 16).

Figure 16. Kaplan-Meier analysis of the probability that lymph node negative patient (left two panels) and lymph node positive patients (right two panels) would remain free of distant metastases and the probability of overall survival, respectively.

Between the two studies, the Nature study and the NEJM study, in lymph node negative patients, the odds ratios for the development of distant metastases within 5 years were similar, 15.3 and 15.0. The prognosis signature was also highly predictive of the risk of distant metastases among the subgroup of patients with lymph node positive disease, which were not present in the original Nature study. The authors conclude that the gene-expression profile is a more powerful predictor of the outcome of disease in young patients with breast cancer than standard systems based on clinical and histologic criteria.

Criticisms
The study has been criticized by many statisticians because of the inclusion of patients in
the development set in the validation cohort. In an independent multicenter validation study, presented at the 2004 San Antonio Breast Cancer Symposium, the 70-gene prognostic index performed less well than in previous studies, although it was still prognostic (note that in a Breast Cancer International Group (BIG) newsletter (volume 7, Number 3, 2005), reanalysis of the data has been claimed to show a much more robust prognostic power). BIG is launching a large clinical trial called MINDACT (n = 6,000 patients) based on upfront stratification using this 70-gene assay, now available as a commercial reference laboratory test, called MammaPrint, in Europe (by Agendia). It is currently available in the USA and has received FDA 510(k) approval63.

21 Gene Assay: Onco
type DX®

Introduction: 21 Gene Assay
The 21 gene assay breast cancer assay is a quantitative RT-PCR assay for fixed paraffin embedded samples. The assay is clinically validated to assess the risk of 10-year distant recurrence, the likelihood of patient survival within 10 years of diagnosis and the magnitude of chemotherapy benefit. This test is currently validated for women with early-stage, node negative, estrogen receptor-positive (N-, ER+) breast cancer who will be treated with hormonal therapy. The assay quantifies the expression of 21 genes using TaqMan™ (quantitative PCR). The assay was clinically validated in a large, multi-center clinical trial with prospectively defined endpoints, conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP), and the assay results were confirmed in a community-based patient population study with Northern California Kaiser Permanente.

Assay Design
There were four phases in the design the assay (1) technical feasibility studies; (2) development studies; (3) analytical validation (4) clinical validation studies.

Technical Feasibility Studies
These studies demonstrated the ability to reproducibly extract mRNA from fixed paraffin embedded tissues from archival tissues up to 30 years in age, developed normalization strategies to account for increasing RNA degradation over time, resolved issues of tumor heterogeneity and determined objective cutpoints for tumor manual microdissection.

Development Studies
In the development process, 250 candidate genes were selected from the published literature, genomic databases and experimental microarray data for breast cancer. The genes were tested in three independent studies including cases from NSABP B-20, in total, involving 447 patients. Multivariate analyses indicated that panels of multiple genes had greater predictive power than any single gene. The data from all three studies were used to select a 21-gene panel (16 cancer related genes and 5 reference genes) that strongly and consistently correlated with likelihood of distant recurrence. The Recurrence Score (RS) is used to predict patient prognosis, it ranges from 0-100 and it is divided into three risk groups: (1) a low risk score correlating with a low risk of distant recurrence, RS 0-18; an intermediate risk score correlating with an intermediate risk of distant
recurrence, $18 < RS < 31$; and, a high risk score correlating with a high risk of distant recurrence, $RS \geq 31$.

**Recurrence Score Calculation**
The Recurrence Score Calculation is as follows: $RS = + 0.47 \times$ HER2 Group Score - $0.34 \times$ ER Group Score + $1.04 \times$ Proliferation Group Score + $0.10 \times$ Invasion Group Score + $0.05 \times$ CD68 - $0.08 \times$ GSTM1 - $0.07 \times$ BAG1. Although some of the coefficients are greater than others, each of the individual genes can greatly influence the individual Recurrence Score.

**Clinical Validation Studies**

**NSABP B-14: Introduction**
NSABP B-14 was a significant clinical trial that established the value of tamoxifen in hormone receptor positive breast cancer patients. This validation study was carefully designed to meet rigorous statistical standards and included a prospectively defined gene list and Recurrence Score calculation, prospectively defined Recurrence Score cutoffs for the risk groups, prospectively defined endpoints and a prospectively defined analysis plan. The NSABP controlled the clinical data and the data analysis. In the validation study, the assay was found to provide a better and/or more reproducible indication of prognosis for ER-positive tumors in node-negative patients than age, tumor size, or histologic grade. The quantitative data allow an individualized risk estimate to be derived (on a scale of 0 to 100), which is a significant improvement over classical prognostic indicators.

**Study Design**
The B-14 study involved 668 evaluable patients: those B-14 participants for whom both tissue and clinical follow-up data was available. All these patients were stage I or II, node negative, ER positive and received tamoxifen treatment. The 668 patients were similar in terms of age distribution and the distribution of tumor size to the entire group of 2,617 tamoxifen-treated patients. Patients from the tamoxifen-treatment arm were chosen, instead of those from the placebo arm, because hormonal therapy such as tamoxifen is the standard of care for women with ER positive breast cancer.

**Methodology**
For each patient, three 10-micron formalin-fixed, paraffin-embedded tissue specimens from the NSABP B-14’s baseline tumor samples were sent to GHI in a blinded fashion. For each specimen, the 21-gene RT-PCR assay was performed, and the Recurrence Score calculated without the knowledge of the patients’ clinical outcomes.

**Analysis**
The pre-specified endpoints were: the primary endpoint was Distant Recurrence-Free Survival (DRFS), while the secondary endpoints included Relapse-Free Survival (RFS) and Overall Survival (OS). The clinical outcomes were compared with the measurement for each individual of their Recurrence Score. The comparison was conducted in a blinded fashion.
Results

Of note is that in this “retroprospective study” paraffin blocks from NSABP trial B-14, with median follow-up of more than 14 years, were used. The technical success rate of the 21 gene assay was 99% (of the 675 patients who were eligible (pathology and clinical), insufficient RNA or RT-PCR outside of specifications occurred in 7 patients (1%). Thus 668 or (99%) of the patients were evaluable in the final analysis). The study met its prospectively defined endpoints: Recurrence Scores were independent and highly significant predictors of recurrence-free survival and Recurrence Score provided accuracy and precision in predicting likelihood of distant recurrence (p < 0.001). Assignment to the risk groups accurately predicted distant recurrence at 10 years (figure 17).

![Figure 17](image)

Figure 17. The Recurrence Score is used to predict patient prognosis, it ranges from 0-100 and it is divided into three risk groups: (1) a low risk score correlating with a low risk of distant recurrence, RS 0-18; an intermediate risk score correlating with an intermediate risk of distant recurrence, 18 < RS < 31; and, a high risk score correlating with a high risk of distant recurrence, RS ≥ 31.

The RS performance was shown to exceed standard measures, such as age, tumor size, and tumor grade either in predictive power or in reproducibility (table 2). There is a near-linear relationship between the numerical Recurrence Score and the patient’s actual risk of distant recurrence.

Table 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Analysis without RS</th>
<th>Analysis with RS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P Value</td>
<td>Hazard Ratio</td>
</tr>
<tr>
<td>Age at surgery</td>
<td>0.1</td>
<td>0.7</td>
</tr>
<tr>
<td>Clinical tumor size</td>
<td>0.13</td>
<td>1.35</td>
</tr>
<tr>
<td>Tumor Grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>0.04</td>
<td>1.87</td>
</tr>
<tr>
<td>Poor</td>
<td>&lt;0.001</td>
<td>5.14</td>
</tr>
</tbody>
</table>
### HER2 amplification

<table>
<thead>
<tr>
<th>Level</th>
<th>Value1</th>
<th>Value2</th>
<th>Value3</th>
<th>Value4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.89 1.04</td>
<td>0.06</td>
<td>0.51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### ER protein

<table>
<thead>
<tr>
<th>Level</th>
<th>Value1</th>
<th>Value2</th>
<th>Value3</th>
<th>Value4</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-99 fmol/mg</td>
<td>0.23</td>
<td>0.71</td>
<td>0.32</td>
<td>0.75</td>
</tr>
<tr>
<td>100-199 fmol/mg</td>
<td>0.38</td>
<td>0.78</td>
<td>0.72</td>
<td>0.9</td>
</tr>
<tr>
<td>&gt;200 fmol/mg</td>
<td>0.9</td>
<td>0.97</td>
<td>0.94</td>
<td>1.02</td>
</tr>
</tbody>
</table>

### Recurrence Score

<table>
<thead>
<tr>
<th>Value1</th>
<th>Value2</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0.001</td>
<td>2.81</td>
</tr>
</tbody>
</table>

---

**Additional Validation Study: Kaiser Permanente Study**

An additional validation study of in 220 cases and 570 controls, conducted by the Northern California Kaiser Permanente in its community-based patient set, confirmed the results from the NSABP B-14 study, that the RS was statistically significantly associated with breast cancer survival in tamoxifen-treated patients with node-negative, ER-positive breast cancer (p = 0.0002). The risk of breast cancer death for patients in the low-risk group as determined by the Recurrence Score (RS < 18) was 2.8% at 10 years. Moreover, over 50% of the patients were in the low risk group. It is notable that this genomic assay can identify such a large cohort of patients who retain such a low risk of breast cancer death.

**Prediction of Chemothrerapy Benefit: NSABP B-20**

NSABP B-20 also assessed the assay’s ability to predict the magnitude of benefit from chemotherapy. This study examined the performance of the 21 gene assay in tumor blocks from 651 patients enrolled from 1988 to 1993 in the tamoxifen alone and tamoxifen plus either CMF or MF chemotherapy treatment randomization arms of the NSABP Study B-20.

**Results**

Patient-specific Recurrence Scores derived from the multi-gene assay were obtained for a total of 651 eligible patients from the tamoxifen alone arm (n = 227) and the tamoxifen plus chemotherapy treatment arm (n = 424) of B-20. Cox proportional hazards models for the global test of interaction between treatment effect (either tamoxifen alone or tamoxifen plus chemotherapy) and gene expression reveal that the Recurrence Score is a significant predictor of chemotherapy benefit (p-value = 0.038).

As the Recurrence Score increases, the likelihood of chemotherapy benefit increases. Not all patients benefit equally from chemotherapy. Patients with low RS tumors (RS < 18) derived minimal, if any, benefit from chemotherapy (an estimated increase in DRFS at 10 years of -1.1 ± 2.2%, mean ± SE). Patients with high risk tumors (RS ≥ 31) had a large absolute benefit of chemotherapy (an absolute increase in DRFS at 10 years of 27.6 ± 8.0%, mean ± SE) (figure 18).
Figure 18 Not all patients benefit equally from chemotherapy. Patients with low RS tumors derived minimal benefit from chemotherapy while patients with high risk tumors had a 28% absolute benefit from chemotherapy.

RS predicted the absolute risk of breast cancer death at 10 Years [low RS = 2.8% (95%CI 1.7%, 3.9%); Intermediate RS = 10.7% (95%CI 6.3%, 14.9%); high RS = 15.5% (95%CI 7.6%, 22.8%). These risk estimates were similar to those in the NSABP B-14 Clinical Validation Study.

**Reclassification vs. Guidelines**
Analysis of the 668 patients from the B-14 study compared the patient Recurrence Scores with patient risk of distant recurrence as determined by clinical guidelines. Overall, 47% of patients were reclassified by 21 gene assay when compared to NCCN guidelines.

**Conclusion**
Significant progress has been made in the use of DNA microarray and QRT-PCR analysis in developing robust genomic classifiers that will compliment the current prognostic metrics that pathologists measure today. The hope is that these platforms and gene combinations, either as ratios, centroids or equations will provide further predictive tools to assist in personalizing chemotherapeutic decisions in the 21st century.
Bibliography

14. Rosen PP, Groshen S, Saigo PE, Kinne DW, Hellman S. Pathological prognostic factors in stage I (T1N0M0) and stage II (T1N1M0) breast carcinoma: a study of 644 patients with median follow-up of 18 years. J Clin Oncol 1989;7:1239-51.
23. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer 1957;11:359-77.


