Rev Endocr Metab Disord (2007) 8:185–198
DOI 10.1007/s11154-007-9035-3

Molecular profiling in breast cancer

Shannon R. Morris · Lisa A. Carey

Published online: 27 April 2007

Abstract Molecular profiling has provided biological evidence for the heterogeneity of breast cancer through the identification of intrinsic subtypes like Luminal A, Luminal B, HER2+/ER− and basal-like. It has also led to the development of clinically applicable gene expression-based prognostic panels like the Mammaprint® and Oncotype Dx™. The increasingly sophisticated understanding allowed by this and similar technology promises future individualized therapy.

Keywords Molecular profiling · Breast cancer · Microarray · Prognostic · Subtypes

1 Introduction

Over the last 10–15 years, thanks to the efforts of a variety of patient advocate groups calling for greater information and research into breast cancer, many women in the United States have a heightened awareness of the dangers of this disease. But to put their fears into a clinical perspective, the most recent data from the American Cancer Society indicates that over 200,000 new cases of invasive breast cancer were diagnosed in 2006 and that greater than 40,000 women died of the disease during that same time period. Thankfully, during the last decade and a half we've seen a decrease in the mortality rate attributed to breast cancer largely due to improvements in early detection and adjuvant treatment. While the improvement in the mortality rate is welcomed, when the patient is sitting across from the oncologist, all she wants to know is "what's going to happen to me, and what do these advances mean for my future?"

Inherent within this question are two concerns: what will my treatment involve, and what is my prognosis? Although many patients don't verbalize these concerns so concisely, this is essentially what they're asking. And as a physician, we want to answer those questions as accurately as possible. Treatment recommendations and mortality estimates largely derive from pathologic analysis of the breast tumor and axillary lymph nodes as well as the tumor's hormone receptor and HER2 status. While useful, these criteria are essentially surrogate markers that reflect our crude understanding of the underlying biology of breast cancer. One unifying observation of the last decade is that breast cancer is biologically a heterogeneous disease. Response to treatment and the prognosis of two patients with the same stage of breast cancer can be vastly different. At least a partial explanation for this disparity in behavior is found in studies using modern techniques including molecular profiling to examine the biologic underpinnings of breast cancer. Because they dovetail with and expand our understanding of breast cancer behavior, the details of molecular profiling have quickly entered the vocabulary of the breast cancer research scientist and are making their way into the clinical arena.

S. R. Morris
GlaxoSmithKline, Research Triangle Park, NC, USA

S. R. Morris · L. A. Carey (✉)
Division of Hematology and Oncology,
Lineberger Comprehensive Cancer Center,
University of North Carolina,
3009 Old Clinic Building, CB # 7305,
Chapel Hill, NC 27599-7305, USA
e-mail: Lisa_Carey@med.unc.edu

2 What is molecular profiling?

Molecular profiling is a scientific approach whereby different types of tissues are compared at the molecular level (DNA, mRNA or protein) on a global scale, i.e. thousands of comparisons made at the same time. As this technology has evolved, a variety of different methods have been developed to accomplish this goal, which makes interpretation of the final data difficult and continues to hinder progression of the field into clinical utility. In general, the majority of the scientific efforts published to date have focused on the manipulation and interpretation of cDNA arrays generated by converting mRNAs isolated from a variety of tissue types to cDNAs, which are then fixed to a solid substrate that allows quantitation of these cDNAs. By comparing the quantitative expression of a cDNA isolated from one tissue type to another, the researcher can draw conclusions regarding the biology of each tissue type.

The underlying assumption that mRNA expression adequately explains the underlying biology of the tissue under study is an assumption that ignores many of the post-translational modifications that further modulate protein expression levels, which are the ultimate determining factor of cell biology. Analysis of cDNA also ignores the underlying structural changes in the tissue's genome that may be contributing to its biology. For example, analysis of HER2 cDNA may reveal that HER2 is overexpressed in select breast tumors but does not provide insight into the mechanism of overexpression, e.g. gene amplification, an observation only made by analyzing the actual chromosomal DNA. Although slow progress is being made in the application of molecular profiling techniques to both genomic DNA and protein expression, the bulk of this review will focus on data generated by analysis of cDNA expression.

While application of existing cDNA microarray technology to patient samples carries its own technical difficulties, interpretation and statistical manipulation of the data generated from these microarrays is even more complex. In general, methods of data interpretation can be divided into two categories: unsupervised methods and supervised methods, which approach the problem from opposite directions. Unsupervised methods, as exemplified by hierarchical clustering analysis, take gene expression data generated from a cohesive set of samples and attempt to identify subclasses of samples based on differences in gene expression. In contrast, supervised methods approach the problem from the other direction—they take gene expression data generated from sample groups with distinct known clinical outcomes (like treatment response or survival) and attempt to identify differences in gene expression that correlate with the known outcome.

Despite great enthusiasm about the promise of molecular profiling to clarify biologic processes important in prediction or prognosis, technical and analytic issues still remain [1, 2]. One of the most widely expressed criticisms of any of the molecular profiling techniques (supervised or unsupervised) is whether the statistical manipulation of the gene expression data truly reflects what's happening in the tumor or whether the gene sets identified as significant are simply artifacts of data manipulation. One way in which scientists have addressed those concerns is to use a training set of patient data to generate an experimental signature gene set and function as the data source for internal validation of the gene set. External validation of the experimental gene set is then provided by using the experimental signature gene set to analyze a validation set of patient data independent of the original training set.

In general, molecular profiling of breast cancer has been used to address three major questions: (1) does the biology of breast tumors differ amongst each other and in comparison to normal tissue? (2) based on these biological differences, can we more accurately predict the clinical outcome for patients with seemingly identical tumors thereby giving them a chance to make better informed decisions regarding their future? and (3) can we more accurately predict which tumor will respond to a specific type of treatment so as to improve the risk/benefit ratio for treating each patient?

3 Breast cancer intrinsic subtypes

Perou et al. first identified distinct molecular subtypes of breast cancer using unsupervised hierarchal clustering analysis of gene expression pattern differences identified in 65 surgical breast specimens [3]. The breast cancers clustered into groups differentiated by expression patterns in several groups of coexpressed genes. An "intrinsic" set of 456 genes more prone to variability between different tumors than paired samples from the same tumor (pre- and post-chemotherapy, primary and lymph node) was established and used to further classify the samples. Subsequent work by the same group refined and expanded these classifications in a total of 78 carcinomas [4]. Reassuringly, these tumors segregated into categories differentiated by expression of estrogen receptor-related genes, supporting earlier epidemiologic and marker studies that suggested that ER-positive and ER-negative disease is biologically different. The ER-positive subtypes included Luminal A and B. The ER-negative subtypes included the HER2+/ER- subtype, characterized by expression of a HER2-related cluster of genes, and the basal-like subtype, characterized by low expression of HER2-related genes, but high expression of a group of genes characteristic of normal basal epithelial breast

tissue [3, 4]. There is a fifth subtype (the normal breast-like subgroup) originally described by Perou et al., however it is not clear that it represents a true subtype [3–6]. Rather, it is possible that this fifth subtype represents breast cancer samples in which the percentage of normal cells are overrepresented in the tumor sample thereby skewing the gene expression results.

The subclassification of these breast cancer tumors into luminal and basal-like types comes from comparison of their gene expression patterns with those of epithelial cells normally found in nonmalignant human mammary gland tissues: luminal cells stain with antibodies against keratins 8 and 18 while basal cells stain with antibodies to keratin 5 and 17. Correspondingly, luminal subtypes of breast cancer express increased levels of keratins 8 and 18 in addition to those genes associated with ER expression while basal-like subtypes of breast cancer express increased levels of keratin 5 and 17 and low to absent levels of ER and genes whose expression is linked to ER [3]. Similar classifications of breast cancers into basal and luminal types using different unsupervised clustering analyses have been seen by others [8–10]. From these observations alone, it is easy to speculate that the luminal subtypes and basal-like subtype of breast cancer either arose from different progenitor cells or differentiated along different paths, resulting in basic differences in their biology and potential differences in their response to treatment. The same type of analysis applied to breast-tissue-derived cell lines reveals that the subtype classification into luminal and basal-like is preserved when patient samples are immortalized as cell lines and provides examples of cell lines that could serve as model systems for further research [7].

External validation of the intrinsic subtypes came from application of the intrinsic gene set to available independent datasets [6]. Using expression centroids, which are profiles made up of the average expression of each relevant gene for each of the five main molecular subtypes, the investigators separately examined intrinsic genes that were included in arrays performed on tumors used to develop the Amsterdam 70-gene prognostic signature (461 genes) [11], a second set of 49 tumors of mixed hormone receptor and nodal status (242 genes) [12], as well as an extended set of tumors from their previous work (534 genes). Using this method between 6–36% of tumors could not be classified into a subtype. Otherwise the subtypes were represented with similar distributions in all the datasets despite differences in the populations; the original gene expression study was based upon high risk, locally-advanced tumors treated with chemotherapy or chemoendocrine therapy, the Amsterdam 70-gene prognostic signature dataset included women under 55 with lymph node-negative tumors that largely did not receive adjuvant systemic therapy [11], and the West dataset included tumors representing a mixture of stages, nodal status, and hormone receptor status [12]. A revised intrinsic gene list was also applied to the combined dataset from several of these and other sources [4, 6, 8, 11], revealing persistence of the subtype signatures across different microarray platforms [5]. Since the clustering methodology for identifying intrinsic subtypes is suboptimal for reproducible classifications, Hu et al. developed the Single Sample Predictor (SSP) tool to serve as a prognostic indicator for individual patient samples. The SSP compares the gene expression profile of an unknown sample to the prototypical profile of each intrinsic subtype and classifies the unknown according to the profile it most closely matches. A potential alternative to the SSP has also been recently described; in these experiments gene expression profiling using several different microarray platforms was used to group a set of patient tumor samples into the intrinsic subtype groups and identify signature genes associated with the luminal and basal-like subtypes [10]. Fifty-four of these signature genes were identified as the minimal set needed to distinguish between luminal and basal-like subtypes. While evaluation of the signature 54 gene set itself provokes interesting questions regarding the underlying biology of these two breast cancer subtypes, future directions may include its evaluation for clinical and clinical research applications.

4 Breast cancer subtypes, biology, and histology

Molecular profiling analyses identify differences in cDNA gene expression profiles, and have recently provided information about the differences between the two most common histologic types of breast cancer: infiltrating ductal and infiltrating lobular. In two separate sets of experiments, unsupervised analysis of cDNA microarrays generated from mRNAs was unable to consistently distinguish between infiltrating ductal carcinoma (IDC) and infiltrating lobular carcinoma (ILC), although supervised methods were able to identify specific gene signatures that appear to segregate with the two histologic types [13, 14]. Genomic profiling studies also suggest a differential imbalance between ILC and IDC in several chromosomal regions [15]. A population-based study of the intrinsic breast cancer subtypes identified by surrogate immunohistochemical (IHC) markers revealed that basal-like tumors were virtually all IDC or similarly poor prognosis histologies, while ILC was generally found among luminal subtypes [16].

While the intrinsic subtype classification system was developed from analysis of locally advanced breast carcinomas, primarily represented by infiltrating ductal carcinoma (86% of samples) [3], there are several less common histopathologic categories of breast cancer including in-

flammatory breast cancer (IBC) and medullary breast cancer (MBC) which can also be categorized by this system. IBC is a clinically distinct form of breast cancer characterized by inflammation (often involving the whole breast) and associated with a poor prognosis. The biology of IBC is poorly understood which is reflected in the fact that it is diagnosed clinically. The application of gene expression profiling and unsupervised hierarchal clustering to data generated from IBC and non-IBC samples reveals that the heterogeneity of breast cancer subtypes is also seen in the distinct clinical entity of IBC [17]; although some therapeutically targetable differences such as a higher rate of HER2-positivity [18] and upregulation of NF-κB-related genes were noted in the IBC samples [19].

In contrast to IBC, MBC is a rare breast cancer that is diagnosed pathologically and whose prognosis is better than expected based upon its associated clinical characteristics, which include high grade and ER-negativity. MBC is associated with BRCA1 mutations [20–22], which correlates with unsupervised gene expression profiling experiments suggesting that most MBC are basal-like [23, 24]. Unique gene expression profile characteristics of MBC compared to non-MBC basal-like breast cancer include underexpression of genes involved in cytoskeletal remodeling and cell invasiveness and overexpression of genes involved in apoptosis, immune response and antigen processing/presentation [24].

Unsupervised analysis of gene expression data from locally advanced breast cancers using PCA (principal component analysis), a technique different from hierarchal clustering, reveals that there may be an additional breast cancer tumor subtype characterized by expression of androgen receptor (AR) related genes [25, 26]. The authors designate this subtype as molecular apocrine and hypothesize that its cell of origin is the apocrine gland, an androgen-dependent sweat gland found in the axilla. In this analysis, three breast cancer subtypes were identified: luminal (ER+, AR+), basal (ER−, AR−) and molecular apocrine (ER−, AR+), where the molecular apocrine classification replaces the HER2+/ ER− subtype described by Perou and colleagues [3, 25]. The authors argue that the molecular apocrine gene signature more truly represents the underlying biology of the tumors since all ER− tumors outside the basal-like subtype fit in the molecular apocrine subclass. This is in contrast to the HER2+/ER− category defined by Perou et al. where some HER2+ tumors are classified outside the group and some members of the HER2+/ER− group lack HER2+ overexpression/gene amplification. The androgen-receptor classification system requires further investigation but suggests another molecule that may be a future therapeutic target.

While the identification of individual breast cancer subtypes within a clinically indistinct population of tumor samples is an interesting observation, what does it say about the biology of the underlying disease? The correlation of each subtype with a specific mammary cell of origin (luminal versus basal) as well as identification of these subtypes in preinvasive breast cancers, e.g. ductal carcinoma in situ, suggest that the changes in gene expression patterns associated with carcinogenesis occur early in the process [9, 27, 28]. Therefore, identifying the changes in gene expression that occur as a cell transforms from non-malignant to malignant may provide future therapeutic targets for cancer prevention.

Molecular profiling has also provided insight into the process of metastasis. Interestingly, while gene sets identifying tumors with a high risk of lung metastasis, bone marrow micrometastases or lymph node involvement may have been identified [29–31], these gene sets appear to differ from those predicting recurrence, suggesting that genes involved in metastasis are not always those determining prognosis [29]. Moreover, although there may be individual genes expressed in the primary tumor that predict relapse and tropism for specific metastatic sites, they are not easily detected using current molecular profiling methods. Either the alterations in gene expression determining metastatic potential are acquired early in the carcinogenesis process such that they are present in both the primary and metastatic tumor [31, 32], or they are so subtle as to escape detection [33, 34].

5 Clinical characteristics and breast cancer subtypes

The heterogeneity of breast cancer clinical outcomes and treatment response is not only related to the underlying biology of the tumor itself but is also a reflection of the genetic and biologic variability of the patient population suffering from the disease. Population-based studies demonstrate that even within the U.S., different populations vary with regards to breast cancer incidence and mortality. For example, breast cancer in African-American women is less common than in Caucasian women, but is diagnosed at a later stage and leads to worse survival even after controlling for stage of diagnosis. Breast cancers in other racial and ethnic groups also differ in terms of incidence rates and risk factors [9, 35]. In the context of molecular profiling and breast cancer subtypes, these observations raise the questions of how and why the subtypes differ within different racial and ethnic groups.

Unsupervised hierarchal clustering analysis of gene expression patterns in 98 invasive breast cancers from a predominantly Chinese patient population revealed that the Luminal, basal-like and HER2+/ER− subtypes are relatively well conserved; although the HER2+/ER− subtype tumors in this patient population also exhibited low level expression of

ER and ER-related genes [9]. Using IHC surrogates for the breast cancer subtypes described by Perou et al., a population-based study evaluating tumors from 196 African-American women and 300 non-African-American women in the United States revealed that while all subtypes were represented in both African-American and Caucasian women, there was an interaction between subtype, race, and age. Basal-like tumors comprised 39% of breast cancers occurring in premenopausal African-American women, as compared to postmenopausal African-American women or non-African-American women regardless of their menopausal status, in whom this subtype made up only 15% [16]. The higher prevalence of basal-like tumors in premenopausal African-American women corresponded to a lower prevalence of Luminal A type tumors which may explain, in part, the epidemiologic observation that young African-American women with breast cancer have a poorer prognosis, although decreased access to care and lower socioeconomic status also play a role.

Ethnicity is not the only clinical characteristic that can be associated with a specific intrinsic subtype [16]. While there is little variation by subtype in stage at presentation, both the basal-like and HER2+/ER− subtypes are primarily invasive ductal histologies, invasive lobular cancers are more likely to be luminal, and metaplastic, anaplastic and undifferentiated carcinomas are more likely to be basal-like than other subtypes. There are marked differences in grade among subtypes, with the majority of basal-like and HER2+/ER− cancers exhibiting high nuclear and histologic grade and mitotic index. This is in contrast to the luminal subtype where only approximately one-third of tumors exhibit high grade histology. Adjusted for other variables, the basal-like subtype is approximately 10-fold more likely to have a high mitotic index or overall grade relative to the Luminal A subtype. While tumors arising in BRCA1 carriers had previously been noted to have a characteristic gene expression signature [11], more recent analysis indicated that these tumors are generally basal-like [6, 36–38]. Although most basal-like tumors do not arise in BRCA1 carriers [16], the relevance of the BRCA1 pathway in the pathogenesis and behavior of sporadic basal-like breast cancer is a topic of great interest.

6 Prognostic profiles

Prognostic indicators based on currently available clinical and histopathologic variables already exist and are used in clinical practice. Examples of such indicators include the Nottingham Prognostic Indicator (NPI), the St Gallen criteria, the NIH consensus guidelines, and Adjuvant! Online which use criteria like tumor size, tumor grade, lymph node status, and hormone receptor status to predict a patient's clinical outcome in certain situations [39–42]. However, these indicators are still inadequate in that within a given patient population with a specific predicted risk of recurrence, there are always patients whose actual clinical outcome doesn't match that predicted by the indicator. Even well-validated tools like Adjuvant! Online, which are used to predict recurrence, mortality risks and the benefit of adjuvant systemic therapy, can still lead to patients being unnecessarily treated with toxic therapies or not treated when their outcomes out to be poor. Therefore, scientists have attempted to use molecular profiling via either unsupervised or supervised methods to create more accurate prognostic indicators to address these issues [4, 6, 8, 11, 29, 43–48]. For example, oncologists already know that the prognosis for breast cancer patients with lymph node positive disease is poorer and that adjuvant systemic therapy decreases their risk of recurrence, but for patients with lymph node negative disease (LNN), the benefit of adjuvant systemic therapy is not so clear. Therefore the ability to risk stratify LNN patients according to prognosis could provide important information for the patient and the treating oncologist when discussing treatment options and could keep many women from suffering the side effects of adjuvant systemic therapy in the absence of benefit. Although many different prognostic indicators are in development, there are six that are relatively well characterized, four of which have been specifically developed to address this question of prognosis in LNN patients: the Amsterdam 70-gene profile, the Recurrence Score, the Rotterdam 76-gene signature, and the wound response signature (Table 1). The fifth prognostic indicator (the invasiveness gene set) was developed by comparing the expression levels of genes expressed in tumorigenic breast cancer cells versus normal breast epithelial cells, while the sixth prognostic indicator (the intrinsic subtype described above) was originally developed from analysis of patients with locally advanced breast cancer and was not specifically designed to risk stratify patients with LNN disease (Table 1).

The Amsterdam 70-gene profile (Mammaprint®) was first developed from supervised gene expression profiling analysis of frozen tumor samples from two distinct patient populations; all were <55 years of age and had lymph node negative disease but 34 of 78 (44%) of the patients had distant metastasis within 5 years of completing treatment and 44 of 78 (56%) of the patients did not [11]. By comparing the gene expression profile of these two groups, a signature 70-gene set was identified that correlated with clinical outcome. Internal validation of the set indicated that it could accurately predict disease outcome for 65 of 78 (83%) of the patients used to generate the 70-gene signature [11]. External validation of the Amsterdam 70-gene prognostic indicator came from a retrospective analysis of 295 young patients (age <53 years) with both lymph node

Table 1 Prognostic profiles

Profile	Developed from	External validation population	Adjusted hazard ratio	Clinical use
Amsterdam 70 gene profile (Mammaprint®) [11, 47, 49]	78 LNN pts, age <55 years, followed for >5 years [11]	295 pts with LNP and LNN disease (61 from training set) age <53 years, T1 or T2, heterogeneous Rx, followed for >5 years [47]	4.6 (2.3–9.2); likelihood of distant metastasis as first event [47]	Predictor of distant metastasis in Stage I–II. Requires frozen tissue.
		302 pts, T1-2 with LNN dis., age <60 years, no adjuvant systemic therapy, median followup >10 years [49]	2.13 (1.19–3.82) time to distant metastasis [49] 2.63 (1.45–4.79) overall survival [49] 1.36 (0.91–2.03) disease-free survival [49] 3.21 (2.23–4.61); likelihood of distant recurrence at 10 years	
Recurrence Score (Oncotype Dx™) [46]	Candidate list of 250 genes applied to 447 pts with LNN and LNP disease, ER+ and ER−, heterogeneous treatment.	668 (of 2,617) pts with ER+, LNN disease Rx with tamoxifen on NSABP B-14, median followup >10 years		Predictor of distant relapse in pts with ER+, LNN disease Can be performed in fixed archival tissue.
Rotterdam 76 gene signature [44, 53]	115 pts w/ LNN disease, no systemic neoadjuvant or adjuvant Rx, followed for >5 years	171 pts w/ LNN disease, 75 % ER+, no systemic neoadjuvant or adjuvant Rx, followed for >5 years [53]	5.55 (2.46–12.5); distant metastasis-free survival [53]	Predictor of distant metastasis-free survival in pts with LNN disease not treated with systemic therapy. Validated primarily in ER+. Requires frozen tissue.
		180 pts w/ LNN disease, >90% ER+, heterogeneous Rx, followed for >5 years [44]	11.36 (2.67–48.4) likelihood of distant metastasis-free survival [44]	
Wound response signature [55]	Identification of core serum response genes expressed in serum-stimulated fibroblasts	295 pts with LNP and LNN disease, age <53 years, T1 or T2, heterogeneous Rx, followed for >5 years (Amsterdam validation study)	7.25 (1.75–30.0) metastasis as first event 11.18 (2.52–49.6) overall survival	None at this time
Invasiveness gene set [60]	Identification of 186 genes that differentiate tumorigenic CD44+/CD24− cells from normal breast epithelium	295 pts with LNP and LNN disease, age <53 years, T1 or T2, heterogeneous Rx, followed for >5 years (Amsterdam validation study)	1.2 (1.1–1.4) metastasis-free survival 1.2 (1.0–1.4) overall survival	None at this time
Intrinsic Subtype [4–6]	Gene list from unsupervised analysis, 49 pts with locally advanced disease, Rx neoadjuvant doxorubicin [4]	97 pts, mostly LNN, followup >5 years [6] (from Amsterdam training set [11])	Not available	None at this time
		311 pts with heterogeneous disease and Rx from multiple datasets and microarray platforms (includes tumors from training set and first validation set) [5]	Relapse-free survival compared with Luminal A: 2.02 (1.1–3.9) Basal-like; 3.47 (1.8–6.8) HER2+/ER−; 1.92 (1.1–3.5) Luminal B [5]	

LNN lymph node negative, *LNP* lymph node positive, *Rx* treatment

negative and lymph node positive disease, some of whom were included in the earlier trial [47]. Based on their 70-gene expression profile, 180 of the patients were classified as poor prognosis and 115 as good prognosis. The mean 5 year overall survival for the poor prognosis group of patients was 74% as compared to 97% for the good prognosis patients (Fig. 1). The 70-gene signature was able to predict prognosis regardless of lymph node status and in

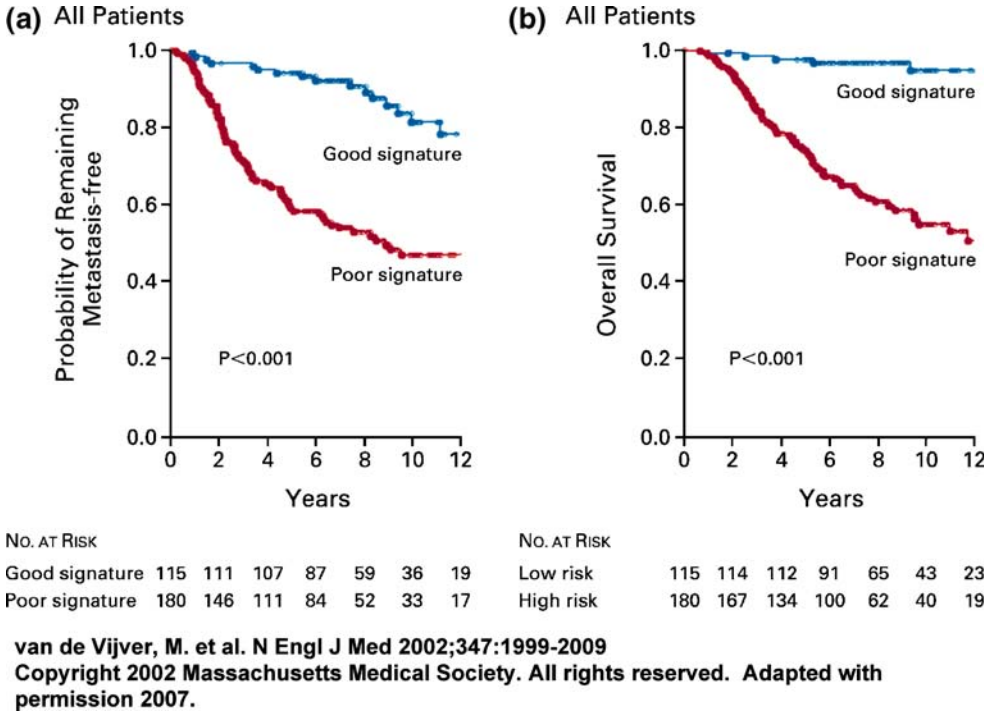

Fig. 1 Probability of time to distant metastasis (a) and overall survival (b) in poor 70 gene signature and good 70 gene signature patients

van de Vijver, M. et al. N Engl J Med 2002;347:1999-2009

multivariate analysis of the risk of distant metastasis as the first event after treatment for the primary cancer, the only independent predictive risk factors were the 70-gene prognosis signature, tumor size, nodal involvement, and use of adjuvant chemotherapy [47]. A second retrospective external validation study of the 70-gene prognostic signature was performed in 302 women age 60 and younger with node-negative T1-2 breast cancers that were not treated with adjuvant systemic therapy and were followed for over 10 years. In these patients, the 70-gene prognostic indicator was better at predicting time to distant metastasis (hazard radio (HR) 2.13) and overall survival (HR 2.63) compared to the clinical variables used by Adjuvant! Online. It did not perform as well for disease-free survival as the other endpoints (HR 1.36 adjusted for Adjuvant!) [49] and the hazard ratios were lower than in the previous validation study, which may reflect the fact that this prognosticator was developed to evaluate expression of genes related to early relapse, the untreated nature of this population, or that the earlier validation study had included some tumors from the training set. However, the external validation data still provided the evidence needed for FDA approval of the Mammaprint® assay and implementation of the MINDACT trial where the usefulness of the 70-gene prognostic signature in determining systemic therapy will be evaluated in a prospective randomized trial for patients with node negative breast cancer [49].

The 21 gene Recurrence Score prognostic indicator (Oncotype Dx™) was developed using slightly different methods than those described above [46]. In this series of experiments, 250 candidate genes were selected from the published literature, genomic databases, and gene expression profiling experiments and correlated with breast cancer recurrence in 447 patients. From these 250 genes, 16 cancer-related genes and five reference genes were selected and their expression levels used to develop the Recurrence Score assay, which is unique in that it can be performed on fixed tumor samples and does not require frozen samples. External validation of the 21 gene Recurrence Score came from the application of this prognostic indicator to patient samples collected in the large multicenter NSABP (National Surgical Adjuvant Breast and Bowel Project) B-14 trial, which examined the benefit of adjuvant tamoxifen in patients with hormone receptor-positive, lymph node-negative breast cancers. 668 of 2,617 tumors from the tamoxifen arm of the trial were assayed and their Recurrence Score compared with outcome at greater than 10 years followup [46]. In those patients classified as low risk by the Recurrence Score (RS<18) only 7% relapsed, compared to high risk patients (RS>31) among whom 31% relapsed within 10 years (Fig. 2). Multivariate analysis of the role of patient age, tumor size, tumor grade, HER2 status, hormone receptor status, and Recurrence Score in predicting distant recurrence revealed that only the Recurrence Score and poor tumor grade were significant predictors of clinical outcome [46]. Since the Recurrence Score was validated in a relatively homogeneous patient population (LNN, ER+, tamoxifen-treated), it is not clear if its ability to predict distant recurrence is related to the natural history of the disease, to the tumor's responsiveness to tamoxifen, or to

both. However, subsequent studies suggest that the Recurrence Score is independently associated with breast cancer mortality [50], chemotherapy sensitivity and tamoxifen resistance (discussed further below) [51, 52]

As mentioned above, the Rotterdam 76-gene signature was specifically devloped to address the clinical question of how to identify those patients with lymph node negative breast cancer that would benefit from adjuvant systemic therapy, regardless of hormone receptor status, since the majority of these patients are cured with locoregional treatment [53]. Two hundred eighty-six patients with LNN breast cancer that had not received adjuvant therapy were divided into ER− and ER+ groups and subjected to gene expression profiling. 115 patients served as the source of the training set data, from which a prognostic model was created by combining the 76 genes selected from the profiling experiments with ER status data. The remaining 171 mixed ER+ (75%) and ER− (25%) tumors served as the validation set. The sensitivity of the 76-gene test in predicting distant metastasis was 93%, and the specificity was 48%. In multivariate analysis of distant metastasis-free survival, the 76-gene prognostic indicator outperformed clinical variables and was the only significant variable to contribute to prognosis prediction. In a subsequent study, the Rotterdam 76-gene signature was also externally validated using a retrospective analysis of an independent data set of 180 LNN patients who did not receive adjuvant systemic therapy. The 76-gene signature was able to accurately identify poor prognosis patients (increased risk of distant metastasis within 5 years) versus good prognosis patients with a hazard ratio of 7.41 (95% CI 2.63–20.9) [44]. Only 16 patients had ER-negative disease, making generalizations to this subset difficult. In multivariate analysis of distant metastasis free survival, the Rotterdam 76-gene signature was the only factor significantly affecting prognosis.

There are several prognostic signatures that are less clinically developed but are of interest. The "wound response" gene expression signature arose from the identification of core serum response (CSR) genes that changed expression levels when cultured fibroblasts were activated with serum. Evaluation of the CSR genes suggested that they represent important processes in wound healing like matrix remodeling, cell motility and angiogenesis, all of which are predicted to play a role in cancer invasion and metastasis [54]. Subsequent evaluation of the expression of these CSR genes in an external gene expression profiling data set generated from 295 patient samples used to validate the Amsterdam 70-gene profile indicated that patients with tumors that expressed an activated wound response signature had a significantly decreased survival and increased probability of distant metastasis as compared to patients

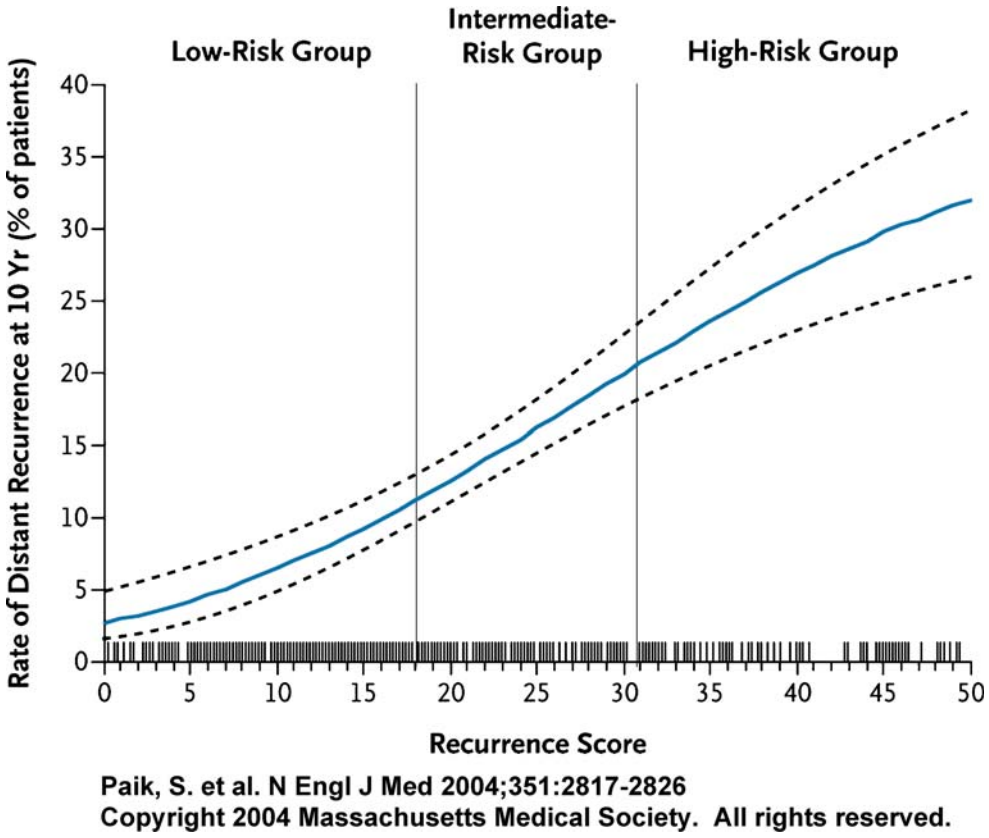

Fig. 2 Risk of distant metastasis as a function of Recurrence Score

whose tumors expressed a quiescent wound response signature [55]. In addition, multivariate analysis of metastasis and death in this patient population indicated that the wound response signature was an independent predictor of prognosis. Genes associated with proliferation alone may also provide prognostic information within a subset of patients. The proliferation gene profile was derived from the Amsterdam 70-gene dataset [11, 47], in which investigators noted that outcome heterogeneity still existed within patient populations classified as having good and poor outcome signatures. They found that after stratification by ER expression and age, the expression level of a group of 50 cell cycle-related genes predicted outcome among those patients identified as having higher than expected ER expression levels for their age [56]. The proliferation signature is an example of a prognostic indicator that may play a role in a specific patient population. Expression levels of hypoxia-induced genes are also prognostic in early stage breast cancer [57]. While the independent contribution of this signature is not yet clear, it may be therapeutically relevant since we currently have no strategies for selecting appropriate patients for antiangiogenic strategies. Recent reports have focused upon the genes associated with the putative cancer "stem cell" [58, 59], which comprise less than 10% of the cells in breast cancer and are highly tumorigenic. These cells are characterized by high expression of the cell surface marker CD44, which is implicated in cell adhesion, migration, and proliferation, and low expression of the less well-characterized CD24. Comparison of CD44+/CD24− cells with normal epithelial cells identified 186 genes associated with the tumorigenic cells, called the "invasiveness gene set" (IGS), which appeared prognostic in both breast and other tumor types. Examination of the 295-patient Amsterdam dataset revealed that the IGS is prognostic independent of clinical characteristics, and appears to be particularly so among ER-positive or intermediate grade tumors. The IGS gene set overlapped little with other prognostic gene sets, and its impact was independent of the wound response signature [60].

Although the intrinsic gene subtypes described by Perou et al. were not originally intended to function as prognostic indicators, the subtypes correlated with prognosis in the original population of 49 patients with relatively locally advanced tumors who had been treated with neoadjuvant doxorubicin on a clinical trial [4]. Patients with the Luminal A subtype had the best prognosis as evaluated by overall survival (OS) and relapse-free survival (RFS) followed by Luminal B. Both the basal-like and HER2+/ER− subtypes had the worst OS and RFS rates. Correlation of outcome with subtype in the independent Amsterdam dataset revealed a significantly longer time to development of distant metastasis among patients with Luminal A tumors compared to patients with basal-like or HER2+/ER− tumors [6]. Similarly, in both a far larger combined dataset of 311 frozen samples of heterogeneous breast cancers and a study using immunohistochemical proxies for the subtypes in a population-based study of nearly 500 tumors, the association of intrinsic subtype with prognosis remained, with the best outcome observed among patients with Luminal A tumors compared with the other subtypes [5, 16].

7 Comparision of prognostic profiles

In order for a new prognostic or predictive assay to be clinically accepted it must be accurate, reproducible, and feasible using clinical samples (a topic beyond the scope of this manuscript) and it must perform better than existing prognostic indicators, i.e. it has to provide better information for clinical decision-making. As described above, there are currently several tools, each incorporating slightly different clinical and histopathologic variables into a prognostic model, available to the practicing oncologist to guide breast cancer treatment decisions. These conventional clinical-pathologic tools are useful but sufficiently inaccurate in predicting either good or bad outcomes such that many patients are either undertreated or overtreated with adjuvant therapy. Therefore, any genomic tool that could more accurately predict those patients more or less likely to benefit from therapy would be welcomed. In addition, now that there are a number of genomic predictive models, the extent to which these are independent remains in question, as does selection of the appropriate patient population to test.

Comparison of the Amsterdam 70-gene signature with the St. Gallen or NIH criteria reveals that the 70-gene signature assigns more LNN patients to the low risk prognosis group than either of the other two clinical indicators: 40% versus 15% versus 7% respectively [47]. Those patients identified as low risk by the 70-gene profile had a higher likelihood of metastasis-free survival than those identified as low risk by the other two methods, thereby indicating that use of the Amsterdam signature could still identify those patients with high risk disease while resulting in fewer patients being inappropriately treated. Comparison of the Amsterdam 70-gene signature to the Adjuvant! Online risk assessment also confirmed the added benefit of the 70-gene profile to clinical risk assessment. The additional benefit of this and similar genomic tools over conventional clinical-pathologic criteria is still an area of some controversy [61]. The Rotterdam 76-gene signature also appears to be superior to both the St. Gallen and NIH consensus criteria with respect to being able to identify those patients with high risk disease while reducing the numbers of patients with LNN disease unnecessarily exposed to the toxicity of adjuvant systemic therapy [44, 53]. More specifically, 40% of patients

classified as average or high risk patients by St. Gallen and 41% of patients classified as average or high risk by NIH would have been reclassified accurately as low risk using the 76-gene signature [44]. As this analysis suggests, these molecular profiling prognosticators will likely provide the most impact when applied in conjunction with clinical prognostic variables rather than instead of clinical variables.

While it appears as if at least some of the gene expression profiling prognostic indicators in development may be superior to those that rely solely on clinical and histopathologic variables, it is unclear how the different gene expression profiling tools compare to one another in terms of their ability to predict clinical outcome. A direct comparison of the Amsterdam 70-gene signature, the Recurrence Score, the wound response profile and the intrinsic subtype classification across a single data set indicated that all four models were highly concordant with respect to their ability to predict prognosis (Table 2). Specifically, those tumors classified into known biologically distinct intrinsic subtypes with poor prognosis, i.e. basal-like, HER2+/ER− or Luminal B, were also classified as having a poor Amsterdam 70-gene profile, an activated wound response signature and a high Recurrence Score [5, 62]. This is in keeping with previous comparisons of the Amsterdam 70-gene signature, the wound response profile and the intrinsic subtype classifications [55]. The most interesting observation to be made from the concordance of the different gene expression profile prognostic indicators with respect to predicting clinical outcome is that there is little gene overlap between the various tools, i.e. they have few genes in common among all of them [62]. This argues that the ability to predict clinical outcome is not related to the expression of a specific and unique set of breast cancer-promoting genes, but that there are a multitude of genes or gene sets within important pathways that can serve as correlates for the biological processes driving these tumors [63].

8 Molecular profiling to predict treatment response

The holy grail of molecular profiling in cancer is the ability to provide individualized treatment plans to each patient so that all patients gain maximal therapeutic benefit with minimal toxicity. This use of genomic techniques is far less advanced than prognostic applications, but is a topic of great interest and rapid evolution. The first indication that molecular profiling could predict chemosensitivity came from gene expression profiling experiments in cell culture lines where cell lines were classified as sensitive or resistant to a specific compound. Gene expression profiles were developed and asked to predict the sensitivity or resistant of a given cell line to that specific compound. Evaluation of 60 cell lines and 232 compounds revealed that 88 of 232 (38%) of profiles could accurately predict sensitivity or resistance to a given compound while only 12 of 232 (5%) of such profiles would be predicted to do so if the profiles were created by chance [64]. These data suggested that gene expression profiles differed between cells that were sensitive or resistant to a given compound, and that evaluation of these differences might be used in a predictive way.

Table 2 Correlation of prognostic indicators [62]

Intrinsic subtype	No. of patients	Recurrence score		70 gene profile		Wound response	
		Classification	No. of patients	Classification	No. of patients	Classification	No. of patients
Basal-like	53	Low	0 (0%)	Good	0 (0%)	Quiescent	3 (6%)
		Intermediate	0 (0%)				
		High	53 (100%)	Poor	53 (100%)	Activated	50 (94%)
Luminal A	123	Low	62 (50%)	Good	87 (71%)	Quiescent	45 (37%)
		Intermediate	25 (20%)				
		High	36 (29%)	Poor	36 (29%)	Activated	78 (63%)
Luminal B	55	Low	1 (2%)	Good	9 (16%)	Quiescent	4 (7%)
		Intermediate	4 (7%)				
		High	50 (91%)	Poor	46 (84%)	Activated	51 (93%)
HER2+/ER−	35	Low	0 (0%)	Good	3 (9%)	Quiescent	0 (0%)
		Intermediate	0 (0%)				
		High	35 (100%)	Poor	32 (91%)	Activated	35 (100%)
Normal-like	29	Low	7 (24%)	Good	16 (55%)	Quiescent	15 (52%)
		Intermediate	4 (14%)				
		High	18 (62%)	Poor	13 (45%)	Activated	14 (48%)

Fan et al. [62]

To bring similar approaches into the clinical arena, gene expression profiling of tumor samples before and after treatment and correlation of those profiles with clinical outcome is critical. The best way to do this in breast cancer is to evaluate tumor samples from patients treated in either the metastatic or neoadjuvant setting, with either single agent or multi-agent therapeutic regimens. Given the toxicity of chemotherapy, much initial focus has been on individualizing this form of therapy, although similar efforts for endocrine and biologic therapy are also underway. In the first line treatment setting, there are two classes of agents that serve as the backbone of the majority of chemotherapeutic regimens: anthracyclines and taxanes. Accordingly the majority of the published scientific data has focused on predicting chemosensitivity of breast cancers to these types of agents.

Support of the observation that gene expression profiles can predict chemotherapy sensitivity comes from attempts to use the 21-gene Recurrence Score to predict chemotherapy benefit in patients treated with a multi-agent chemotherapy regimen. As might have been predicted by the influence of proliferative and HER2 genes, the RS correlates with the probability of pathologic complete response (pCR) in patients treated with an anthracycline/taxane neoadjuvant regimen [51]. In NSABP B-20, a randomized study of tamoxifen with or without MF or CMF chemotherapy in LNN, hormone receptor-positive patients, those patients with high risk Recurrence Scores (RS≥31) not only had a higher likelihood of relapse despite endocrine therapy, but also derived a larger benefit from adjuvant chemotherapy (mean absolute decrease in distant recurrence rate at 10 yrs of 28%) compared to those with low risk Recurrence Scores (RS<18) who received minimal benefit from adjuvant chemotherapy (mean absolute decrease in distant recurrence rate at 10 yrs of −1%) [52]. Like the Recurrence Score, the intrinsic subtypes have also been correlated with pCR after treatment with an anthracycline/taxane regimen in the neoadjuvant setting. Paradoxically, in spite of the poor prognosis associated with the basal-like and HER2+/ER− subtypes, both of these tumor subtypes demonstrate increased rates of pCR as compared to Luminal A tumors when treated with neoadjuvant chemotherapy [65, 66]. This fits with the finding that ER negative disease appears to benefit more from chemotherapy advances [67] and may reflect the lack of non-chemotherapy adjuvant options historically available for ER-negative disease.

A number of efforts at identifying gene sets predictive of chemotherapy response have been published [68–75]. For example, three distinct gene sets predicting tumor response to single agent docetaxel have been described. Two were developed from supervised gene expression profile analysis of patient material obtained before and after treatment with single agent neoadjuvant docetaxel, where tumors were classified as sensitive or resistant based on the percentage of residual tumor after treatment. In one series of experiments, an 85-gene signature was able to predict clinical response with a prediction accuracy of 80% [74]. Functional analysis of the 85 gene set revealed that tumors isolated from nonresponders exhibited elevated expression of genes controlling the cellular redox environment, thereby suggesting that the use of drugs inhibiting redox could convert docetaxel-resistant tumors to docetaxel-sensitive tumors. Another study identified an alternative 92-gene signature with a prediction accuracy of 88% [69]. In a subsequent study by the same group, comparison of the gene expression profiles of all tumors after docetaxel treatment (regardless of whether they were classified as sensitive or resistant), revealed that the profiles of tumors subjected to the selection pressure of docetaxel were relatively homogeneous. This observation suggested that those tumors originally sensitive to docetaxel may have developed resistance to the drug or led to selection of a resistant clone as evidenced by convergence of the gene expression profiles of sensitive and resistant tumors [70]. A third 50-gene predictor of docetaxel sensitivity was developed from gene expression profile analysis of sensitive versus resistant cancer cell lines in vitro and shown to have a prediction accuracy of 91.6% when used to identify patients who responded to docetaxel in vivo [75]. A practical approach to developing a gene profile predicting sensitivity to an entire regimen was taken by another group [68], who identified a 74 marker profile. While these molecular profiles provide interesting information about sensitivity to commonly used breast cancer drugs, all of these assays require further validation, and caution is reinforced when noting that while these appear to be similar studies, they included different tumor sizes, patient populations and different methodologies, and confounding by varying proportions of tumor and stroma can impact results [76]. The ability to identify predictive profiles in these typically small neoadjuvant studies is controversial [73]. It must also be noted that for assays used in clinical decision-making, even 10–20% inaccuracy is often unacceptable. These studies were performed in unselected breast cancers; it is also possible that different subtypes of breast cancer respond differently to cytotoxic chemotherapy [77] so predictive profiles, like prognostic profiles, may differ among populations.

Not only is gene expression profiling promising for predicting sensitivity to chemotherapy, it has also been used to predict sensitivity to endocrine therapies like tamoxifen. Although current histopathologic evaluation of breast cancer tumors involves determination of ER status which, in general, correlates with response to endocrine therapy, a large percentage of patients with ER+ disease will display

de novo resistance to endocrine therapy or will develop resistance over time. As mentioned above, within the appropriate patient population, the RS identifies those most likely to develop distant metastases despite adjuvant tamoxifen [46].Unsupervised gene expression profiling analysis of microarrays created from ER-positive tumors has also revealed a 44-gene signature that correlated with tamoxifen resistance in 77% of patients. Clinical ER status correctly predicts response to tamoxifen in only 50–60% of patients [78]. This tamoxifen resistance profile is undergoing independent validation. Not surprisingly, functional analysis of the gene signature revealed a large number of genes known to be regulated by estrogen, although genes involved in apoptosis and extracellular matrix remodeling were also detected.

9 Conclusions

- Breast cancer is a biologically heterogeneous entity.
- Four intrinsic subtypes of breast cancer have been identified based on differences in patterns of gene expression.
- Several prognostic indicators including the Amsterdam 70-gene profile (Mammaprint®) and Recurrence score (Oncotype Dx™) have been developed from gene signature sets and are in clinical trials.
- Both the intrinsic subtypes and prognostic indicators are being evaluated for their ability to predict metastasis and provide for individualized therapy.

10 Key unanswered questions

- How many breast cancer subtypes are there? There may be subtypes not yet identified, or subtypes within the major subtypes that have clinical implications, particularly for efficacy of targeted therapy.
- Do prognostic profiles work better in one subtype or population than another? The Recurrence Score has been validated in node-negative, hormone receptor-positive tumors treated with adjuvant endocrine therapy, the 70-gene and 76-gene prognosticators in node-negative disease. Do these work equally in more expanded populations, and can we identify node-positive breast cancers that do not benefit from chemotherapy, or node-negative breast cancers that do not require any systemic therapy at all?
- Are the site-specific metastatic signatures identified in recent studies real? If so, what does it mean for our ability to predict the risk/benefit of site-specific adjuvant therapies such as bisphosphonates?
- Can analysis of the genes identified in these molecular profiles provide new targets for future therapeutic intervention?

References

1. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. Nat Rev Cancer 2005;5:142–9.
2. Simon R, Radmacher MD, Dobbin K, McShane, LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J Natl Cancer Inst 2003;95:14–8.
3. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.
4. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74.
5. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 2006;7:96.
6. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al: Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003; 100:8418–23.
7. Ross DT, Perou CM. A comparison of gene expression signatures from breast tumors and breast tissue derived cell lines. Dis Markers 2001;17:99–109.
8. Sotiriou C, Neo SY, McShane LM, Korn L, ong PM, Jazaeri A, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 2003;100:10393–8.
9. Yu K, Lee CH, Tan PH, Tan P. Conservation of breast cancer molecular subtypes and transcriptional patterns of tumor progression across distinct ethnic populations. Clin Cancer Res 2004;10:5508–17.
10. Sorlie T, Wang Y, Xiao C, Johnsen H, Naume B, Samaha RR, et al. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. BMC Genomics 2006;7:127.
11. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–6.
12. West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, et al. Predicting the clinical status of human breast cancer by using gene expression profiles. Proc Natl Acad Sci U S A 2001;98:11462–7.
13. Zhao H, Langerod A, Ji Y, Nowels KW, Nesland JM, Tibshirani R, et al. Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. Mol Biol Cell 2004;15: 2523–36.
14. Korkola JE, DeVries S, Fridlyand J, Estep AL, Chen YY, Chew KL, et al. Differentiation of lobular versus ductal breast carcinomas by expression microarray analysis. Cancer Res 2003;63:7167–75.
15. Stange DE, Radlwimmer B, Schubert F, Traub F, Pich A, Toedt G, et al. High-resolution genomic profiling reveals association of chromosomal aberrations on 1q and 16p with histologic and genetic subgroups of invasive breast cancer. Clin Cancer Res 2006;12:345–52.
16. Carey LA, Perou CM, Livasy CA, Dressler LG, Croan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 2006;295:2492–502.

17. Bertucci F, Finetti P, Rougemont J, Charafe-Jauffret E, cervera N, Tarpin C, et al. Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. Cancer Res 2005; 65:2170–8.
18. Charafe-Jauffret E, Tarpin C, Bardou VJ, Bertucci F, Ginestier C, Braud AC, et al. Immunophenotypic analysis of inflammatory breast cancers: identification of an 'inflammatory signature.' J Pathol 2004;202:265–73.
19. Van Laere S, Van der Auwera I, Van den Eynden GG, Fox SB, Bianchi F, Harris AL, et al. Distinct molecular signature of inflammatory breast cancer by cDNA microarray analysis. Breast Cancer Res Treat 2005;93:237–46.
20. Pathology of familial breast cancer: differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and sporadic cases. Breast Cancer Linkage Consortium. Lancet 1997;349:1505–10.
21. Eisinger F, Jacquemier J, Charpin C, Stoppa-Lyonnet D, Bressac-de-Paillerets B, Peyrat JP, et al. Mutations at BRCA1: the medullary breast carcinoma revisited. Cancer Res 1998;58:1588–92.
22. Lakhani SR, Jacquemier J, Sloane JP, Gusterson BA, Anderson TJ, van de Vijver, et al. Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. J Natl Cancer Inst 1998;90:1138–45.
23. Jacquemier J, Padovani L, Rabayrol L, Lakhani SR, Penault-Llorca F, Denoux Y, et al. Typical medullary breast carcinomas have a basal/myoepithelial phenotype. J Pathol 2005;207:260–8.
24. Bertucci F, Finetti P, Cervera N, Charafe-Jauffret E, Mamessier E, Adelaide J, et al. Gene expression profiling shows medullary breast cancer is a subgroup of basal breast cancers. Cancer Res 2006;66:4636–44.
25. Farmer P, Bonnefoi H, Becette V, Tubiana-Hulin M, Fumolean P, Larsimont D, et al. Identification of molecular apocrine breast tumours by microarray analysis. Oncogene 2005;24: 4660–71.
26. Doane AS, Danso M, Lal P, Donaton M, Zhang L, Hudis C, et al. An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. Oncogene 2006;25:3994–4008.
27. Hannemann J, Velds A, Halfwerk JB, Kreike B, Petersen JL, van de Vijver MJ. Classification of ductal carcinoma in situ by gene expression profiling. Breast Cancer Res 2006;8:R61.
28. Schuetz CS, Bonin M, Clare SE, Nieselt K, Sotlar K, Walter M, et al. Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. Cancer Res 2006;66:5278–86.
29. Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, et al. Gene expression predictors of breast cancer outcomes. Lancet 2003;361:1590–6.
30. Woelfle U, Cloos J, Sauter G, Riethdorf L, Janicke F, van Diest P, et al. Molecular signature associated with bone marrow micrometastasis in human breast cancer. Cancer Res 2003;63:5679–84.
31. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. Nature 2005;436:518–24.
32. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. Nat Genet 2003;33:49–54.
33. Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MG, et al. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. Cancer Res 2005;65:9155–8.
34. Weigelt B, Wessels LF, Bosma AJ, Glas AM, Nuyten DS, He YD, et al. No common denominator for breast cancer lymph node metastasis. Br J Cancer 2005;93:924–32.
35. Chlebowski RT, Chen Z, Anderson GL, Rohan T, Aragaki A, Lane D, et al. Ethnicity and breast cancer: factors influencing differences in incidence and outcome. J Natl Cancer Inst 2005;97:439–48.
36. Foulkes WD, Stefansson IM, Chappuis PO, Begin LR, Goffin JR, Wong N, et al. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 2003;95:1482–5.
37. Olopade OI, Grushko T. Gene-expression profiles in hereditary breast cancer. N Engl J Med 2001;344:2028–9.
38. Foulkes WD, Brunet JS, Stefansson IM, Straume O, Chappnis PO, Begin LR, et al. The prognostic implication of the basal-like (cyclin E high/p27 low/p53+/glomeruloid-microvascular-proliferation+) phenotype of BRCA1-related breast cancer. Cancer Res 2004;64:830–5.
39. Olivotto IA, Bajdik CD, Ravdin PM, Speers CH, Coldman AJ, Norris BO, et al. Population-based validation of the prognostic model ADJUVANT! for early breast cancer. J Clin Oncol 2005;23:2716–25.
40. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham Prognostic Index in primary breast cancer. Breast Cancer Res Treat 1992;22:207–19.
41. Goldhirsch A, Wood WC, Gelber RD, Coates AS, Thurlimann B, Senn HJ. Meeting highlights: updated international expert consensus on the primary therapy of early breast cancer. J Clin Oncol 2003;21:3357–65.
42. Eifel P, Axelson JA, Costa J, Crowley J, Curran WJ, Deshler A, et al. National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1–3, 2000. J Natl Cancer Inst 2001;93:979–89.
43. Ahr A, Karn T, Solbach C, Seiter T, Strebhardt K, Holtrich U, et al. Identification of high risk breast-cancer patients by gene expression profiling. Lancet 2002;359:131–2.
44. Foekens JA, Atkins D, Zhang Y, Sweep FC, Harbeck N, Paradiso A, et al. Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. J Clin Oncol 2006;24:1665–71.
45. Ma XJ, Wang Z, Ryan PD, Isakoff SJ, Barmettler A, Fuller A, et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell 2004;5:607–16.
46. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817–26.
47. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.
48. Iwao K, Matoba R, Ueno N, Ando A, Miyoshi Y, Matsubara K, et al. Molecular classification of primary breast tumors possessing distinct prognostic properties. Hum Mol Genet 2002;11:199–206.
49. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst 2006;98:1183–92.
50. Habel LA, Quesenberry CP, Jacobs M, Greenberg D, Fehrenbacher L, Alexander C, et al. Gene expression and breast cancer mortality in Northern California Kaiser Permanente Patients: a large population-based case control study. J Clin Oncol 2005 ASCO Annual Meeting Proceedings 2005;23:603.
51. Gianni L, Zambetti M, Clark K, Baker J, Cronin M, Wu J, et al. Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. J Clin Oncol 2005;23:7265–77.
52. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol 2006;24:3726–34.
53. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671–9.

54. Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS Biol 2004;2:E7.
55. Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshitani R, Sorlie T, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 2005;102:3738–43.
56. Dai H, van't Veer L, Lamb J, He YD, Mao M, Fine BM, et al. A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. Cancer Res 2005;65:4059–66.
57. Chi JT, Wang Z, Nuyten DS, Rodriguez EH, Schaner ME, Salim H, et al. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. PLoS Med 2006;3:e47.
58. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A 2003;100:3983–8.
59. Ponti D, Costa A, Zaffaroni N, Pratesi G. Petrangolini G, Coradini D, et al. Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. Cancer Res 2005;65:5506–11.
60. Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med 2007;356:217–26.
61. Eden P, Ritz C, Rose C, Ferno M, Peterson C. "Good Old" clinical markers have similar power in breast cancer prognosis as micro-array gene expression profilers. Eur J Cancer 2004;40:1837–41.
62. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, et al. Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 2006;355:560–9.
63. Ein-Dor L, Kela I, Getz G, Givol D, Domany E. Outcome signature genes in breast cancer: is there a unique set? Bio-informatics 2005;21:171–8.
64. Staunton JE, Slonim DK, Coller HA, Tamayo P, Angelo MJ, Park J, et al. Chemosensitivity prediction by transcriptional profiling. Proc Natl Acad Sci U S A 2001;98:10787–92.
65. Carey LA, Dees EC, Sawyer LR, et al. The triple negative paradox: primary tumor chemosensitivity of the basal-like breast cancer phenotype. Clin Cancer Res in press.
66. Rouzier R, Perou CM, Symmans WF, Cristofanilli M, Anderson K, Hess KR, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res 2005;11:5678–85.
67. Berry DA, Cirrincione C, Henderson IC, Citron ML, Budman DR, Goldstein LJ, et al. Estrogen-receptor status and outcomes of modern chemotherapy for patients with node-positive breast cancer. JAMA 2006;295:1658–67.
68. Ayers M, Symmans WF, Stec J, Damokosh AI, Clark E, Hess K, et al. Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. J Clin Oncol 2004;22:2284–93.
69. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. Lancet 2003;362:362–9.
70. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Tham YL, et al. Patterns of resistance and incomplete response to docetaxel by gene expression profiling in breast cancer patients. J Clin Oncol 2005;23:1169–77.
71. Cleator S, Tsimelzon A, Ashworth A, Dowsert M, Dexter T, Powles T. Gene expression patterns for doxorubicin (Adriamycin) and cyclophosphamide (cytoxan) (AC) response and resistance. Breast Cancer Res Treat 2006;95:229–33.
72. Folgueira MA, Carraro DM, Brentani H, Patrao DF, Barbosa EM, Netto MM, et al. Gene expression profile associated with response to doxorubicin-based therapy in breast cancer. Clin Cancer Res 2005;11:7434–43.
73. Hannemann J, Oosterkamp HM, Bosch CA, Velds A, Wessels LF, Loo C, et al. Changes in gene expression associated with response to neoadjuvant chemotherapy in breast cancer. J Clin Oncol 2005;23:3331–42.
74. Iwao-Koizumi K, Matoba R, Ueno N, Kim SJ, Ando A, Miyoshi Y, et al. Prediction of docetaxel response in human breast cancer by gene expression profiling. J Clin Oncol 2005;23:422–31.
75. Potti A, Dressman HK, Bild A, Riedel RF, Chan G, Sayer R, et al. Genomic signatures to guide the use of chemotherapeutics. Nat Med 2006;12:1294–300.
76. Cleator SJ, Powles TJ, Dexter T, Fulford L, Mackay A, Smith Ie, et al. The effect of the stromal component of breast tumours on prediction of clinical outcome using gene expression microarray analysis. Breast Cancer Res 2006;8:R32.
77. Troester MA, Hoadley KA, Sorlie T, Herbert BS, Borresen-Dale AL, Lonning PE, et al. Cell-type-specific responses to chemo-therapeutics in breast cancer. Cancer Res 2004;64:4218–26.
78. Jansen MP, Foekens JA, van Staveren IL, Dirkzwager-Kiel MM, Ritstier K, Look MP, et al. Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. J Clin Oncol 2005;23:732–40.