

Chapter 2

Heterogeneity of Breast Cancer: Gene Signatures and Beyond

Gaia Schiavon, Marcel Smid, Gaorav P. Gupta, Stefania Redana, Daniele Santini, and John W.M. Martens

Introduction

The initial steps into a better understanding of the heterogeneity and biology of breast cancer were made at the onset of 2000, with the first identification of distinct molecular subtypes of human breast tumors possessing different outcome [1–3]. Gene expression profiling and microarray analysis opened a road leading to the new molecular classification of breast cancer, recognizing at least five reproducible subtypes: luminal A, luminal B, ERBB2, basal, and normal-like [3–6]. This revolutionary concept was triggered from an intense research driven by the evidence that 60–70% of all breast cancers are classified as “not otherwise specified” infiltrating ductal carcinomas (IDC NOS) [4].

The methodology of microarray has been soon supported by other tools like array comparative genomic hybridization (array-CGH), single-nucleotide polymorphism (SNP), high-throughput screening (HTS) techniques and the increasing availability of multiple tools for pathways analysis. The combination of these advanced technologies is constantly applied to in vitro and in vivo research in order to improve our knowledge of breast cancer biology and our understanding of the complex process of metastasis. The ultimate goal is to create strategies and algorithms guiding a tailored management of patients with both early and advanced breast cancer.

The focus of this chapter is to provide a general summary of the genomic signatures available for breast cancer and to function as a tool for clinicians for the interpretation of gene signatures. Moreover, we also give a brief overview of the emerging tools designed to capture and study the heterogeneity of breast cancer (next-generation sequencing).

G. Schiavon (✉) • M. Smid • J.W.M. Martens
Department of Medical Oncology, Daniel den Hoed Cancer Center,
Erasmus Medical Center, Rotterdam, The Netherlands
e-mail: gaiaschiavon@gmail.com

G.P. Gupta
Department of Radiation Oncology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

S. Redana
Department of Medical Oncology I, Institute for Cancer Research and Treatment,
Fondazione Piemontese per L'Oncologia, Candiolo, Italy

D. Santini
Department of Medical Oncology, University Campus Bio-Medico, Rome, Italy

Genomic Signatures and Microarray Analysis

From Binary to Bedside

In the last decade or so, there has been a slow switch from analyzing a gene-at-a-time, to higher throughput procedures. To establish which genes were active in a biological context, traditional methods like Northern and Southern Blotting were gradually replaced by the so-called microarrays. These started out as large membrane sheets spotted with cDNA, but these have evolved substantially to much smaller chips which can contain millions of oligonucleotides. mRNA of a biological sample can be hybridized to these chips, which yields the expression levels of thousands of genes in one single experiment.

This in turn necessitates a specialized field to measure, collect, transform, analyze, and evaluate these data using statistically sound methods. Enter bioinformatics. Although part of the bioinformatics field is concerned with the (pre)processing steps to extract reliable data from the microarrays, a big part focuses on applicable analyses. Demanding as it can be, dealing with millions of data-points, it does provide a very rich platform to build solid conclusions.

If one considers that the expression of a few markers like estrogen receptor (ER) or HER2/neu (together with tumor size, grading, nodal involvement, and other few other prognostic marker) will determine the best treatment regimen, it is easy to envision that by measuring thousands of genes in hundreds of samples, valuable markers or combinations of markers can be identified, linked with many, if not all, clinical aspects of the patients involved. However, teasing out statistically sound differentially expressed genes or reliable signatures hiding in the ranks of thousands of genes does entail a considerable challenge, one that has admittedly met with its pitfalls, but more so with considerable successes.

To start on a cautionary note, it is still a very good idea to uphold the scientific principle to validate one's results thoroughly. The most notorious example to date is the paper describing expression signatures guiding the choice of chemotherapy in cancer patients which was heralded in 2006 [7], but ultimately proved unreliable and was retracted in 2011. Indeed, bioinformatics is a powerful tool, which obliges researchers to wield it in a correct manner.

One of the most abundantly used bioinformatics analysis is “hierarchical clustering,” in which order is created out of the chaos of gene expression patterns of tumor cells. By grouping tumors according to the similarity of their gene expression levels, Perou et al. were able to reflect the notorious heterogeneity which is found in the clinical outcome of breast cancer patients, in five distinct molecular subtypes [1]. This landmark paper was among the first which put the power of bioinformatics analysis on the map. The five molecular subtypes have since then been studied extensively, and many clinical relevant observations are ascribed to the subtypes, among which prognosis [1, 3] response to therapy [8], and site-specific relapse [9]. Thus, the overall gene expression patterns can be quite distinct in patients suffering from breast cancer. By linking the expression of specific genes to the clinical parameters of patients, any range of clinically relevant questions can be addressed. Milestone examples are gene expression signatures able to predict prognosis [10, 11], of which Food and Drug Administration (FDA) approval has been granted for the first multigene model, signatures for response to therapy [8, 12], and signatures for breast cancers relapsing to bone [13–15], lung [16, 17], and brain [18].

One dimension higher is the analysis of interacting genes. Similar gene functions or signaling cascades – pathways – can be identified from the expression data. Although more suited to increase understanding in the biological processes about breast cancer, clinical associations with specific pathways have been described [19]. Finally, similar bioinformatics analyses can be applied to miRNA, DNA copy-number, SNPs, sequence, and methylation data. Gathering all these data from the same tumor and integrating all this knowledge is the upcoming challenge. When available, a detailed tumor blueprint for each individual breast cancer patient could be constructed, which will guide the physician to treat future patients with the most tailored strategy.

Multiple Gene Signatures: Do Not Get Lost

As already mentioned, the identification of breast cancer molecular subtypes was the first insight into the biologic heterogeneity of breast cancer [20]. Gene-expression profiling resulted to be – and still is – a very appealing approach, but the clinical utility of classifying breast cancer into molecular subtypes using unsupervised cluster analysis has limitations. For example, with the addition of a new case to the data, the dendrograms of hierarchical cluster analysis are reorganized, and therefore it is not possible to prospectively classify new cases using this methodology [21]. If larger sample sets are used, more clusters and molecular subtypes of breast cancer could become evident [22]. In this perspective, several large studies gave the major clinical contribution of gene-expression profiling in predicting prognosis and response to therapies. Different approaches have been used and different questions have been addressed. For example, some signatures were derived from the gene profiling of human breast cancer cell lines with particular propensity to metastasize to one or another organ and/or mouse models of breast cancer and then applied to breast cancer patients. In other cases, the signature was directly derived from the profiling of fresh-frozen tissues from breast cancer patients (test set) and then validated in one or more independent datasets. Samples are selected on the basis of a specific aim: identification of molecular subtype, prediction of prognosis, resistance to chemo/hormonal-therapy, risk of relapse, etc. Table 2.1 provides a large overview of the most relevant signatures produced by breast cancer research. Of note, a source of concern has been the little or absent overlap between the different gene sets, when compared to each other. For example, the 70-gene signature and the 21-gene RS have only the SCUBE2 gene in common [10, 23, 32]. The reasons for this lower-than-expected overlap are not completely known, but they probably include differences in the patient cohorts (e.g., 70-genes and 76-genes prognostic signatures), microarray platforms, the large number of genes associated with prognosis, and bioinformatics–mathematical methods used for analysis [10, 11, 23]. To answer the question whether these predictors are concordant with respect to their predictions for individual patients, Fan et al. analyzed a single dataset on which five prognostic or predictive gene-expression-based models were simultaneously compared (the 70-gene signature model, the wound-response model, the 21-gene RS model, the intrinsic-subtype model, and the two-gene-ratio model) [1–3, 5, 10, 11, 23–26, 32, 33]. With this analysis, all gene-expression-based models with the exception of the two-gene ratio model, significantly predicted relapse-free survival and overall survival. A limitation of this study was that the 21-gene RS and two-gene ratio models were developed to be used in different clinical scenarios than was represented in this patient cohort [20]. In fact, the two-gene ratio and the 21-gene RS models were designed to predict outcomes in patients with ER-positive disease receiving tamoxifen as an adjuvant treatment. However, the Fan dataset included only 40 such patients and a substantial portion of it was used as the training set for the development of the intrinsic-type, 70-gene signature, and wound-response models.

Despite the lack of gene overlap, four of the five models showed significant agreement in predicting outcome for individual patients leading to the hypothesis that different gene sets may predict a biologically similar breast cancer phenotype. The concordance among the models in the identification of patients with a genomic high-risk for recurrence was excellent, but it remains unknown how much clinical utility these predictive models provide over standard clinicopathologic features. While in multivariate analysis, three of the gene-based models proved to be more predictive for outcome than standard clinicopathologic criteria, features routinely used by physicians, especially in clinically intermediate-risk cancers (e.g., progesterone receptor (PR) status, HER2 status, lymphovascular invasion, and mitotic rate) were not included in the analysis. On-going clinical trials will help to clarify the potential benefits of these genomic tools over standard clinicopathologic assessment. In fact, what we have learned from this field is the importance of clinical validation of gene signatures in independent datasets with long follow-up available and a subsequent prospective validation (see section “Clinical Application of Gene Signatures: Ongoing Trials”).

Table 2.1 Gene-expression profiling models in breast cancer

Models	Aim of the signature	Tissue	Validation sets in the same paper	Assay	Number of genes in panel	Used in prospective studies
Intrinsic subtype [1–5]	Subtypes/prognosis	Fresh frozen	–	Microarray	534	None
Netherlands signature (MammaPrint TM) [10, 22]	Prognosis	Fresh frozen	–	Microarray	70	MINDACT (on-going)
Recurrence score (OncotypeDX TM) [23]	Relapse	Paraffin-embedded	–	RT-PCR	16 (+ 5 reference genes)	TAILORx (on-going)
Rotterdam Signature [11]	Prognosis	Fresh frozen	–	Microarray	76	None
Wound response [24, 25]	Progression/prognosis	Fresh frozen	–	Microarray	512	None
Two-gene ratio [26]	Prognosis	Paraffin-embedded	–	RT-PCR	2	None
Gene-expression grade index [27, 28]	Grading/prognosis	Fresh frozen	–	Microarray	97	None
p53 Signature [29]	p53 Status/prognosis	Fresh frozen	GSE:4382,1379 [10]	Microarray	32	None
p53 Signature [30]	Subtypes/prognosis/chemo-response	Developed in [29]	GSE:3494,2034,7390,9195,16716	Microarray	39 (ER+) 30 (ER-)	None
Smid et al. [9]	Subtype/site of relapse	Developed in EMC344 (GSE:2034,5327)	–	Microarray	529 (bone), 67 (lung), 149 (brain), 18 (liver), 39 (pleura)	None
Jansen et al. [12]	Tamoxifen resistance	Fresh frozen (test+validation sets)	–	Microarray	44	None
Kang (bone) [13]	Bone relapse	human breast cancer cell line (highly metastatic to bone)	–	Microarray	102	None
Src responsive signature (SRS) [14]	Bone relapse	Developed in [31]	Fresh frozen EMC344, EMC189, MSK82 (GEO2603, GSE:5327,2034,12276)	Microarray	159	None
Predictor of Tumor Relapse to Bone [15]	Bone relapse	Fresh frozen	–	Microarray	31	None
Lung metastasis signature (LMS) [16]	Lung relapse	human breast cancer cell line (metastatic to lung) in a xenograft model	MSK-82 (fresh frozen)	Microarray	95 (54 for functional validation)	None
Lung metastasis signature (LMS) [17]	Lung relapse	Developed in [16]	MSK99, NK1295, EMC344	Microarray	18	None
Brain metastasis signature [18]	Brain relapse	brain-metastatic-derived (BrM) cells + MSK82/EMC286 (training set)	EMC192, NK1295 (independent sets)	Microarray	17	None

An interesting point of view is that gene expression signatures can reflect the activation status of several oncogenic pathways. Yu et al. suggested that it might be more appropriate to interrogate the gene lists for biological themes, rather than individual genes [19]. Moreover, identification of the distinct biological processes between subtypes of cancer patients is more relevant to understand the mechanism of the tumorigenesis and metastatic capability and for targeted drug development. They resampled their dataset numerous times to get multiple gene lists whose expression correlated with patients' outcome. For example, based on these gene lists, they identified overrepresented pathways defined in gene ontology biological process (GOBP) for ER-positive or ER-negative breast cancer patients, separately. Then, they compared the pathways represented by different published prognostic gene signatures with the overrepresented pathways associated with metastatic capability. This study also demonstrated that it is feasible to construct a gene signature from the key pathways to predict clinical outcomes. Clustering tumors based on pathway signatures defines prognosis in different patient subsets, demonstrating that patterns of oncogenic pathway deregulation underlie the development of the oncogenic phenotype and reflect the biology and outcome of specific cancers. According to Bild et al., prediction of pathway deregulation in cancer cell lines is also able to predict the sensitivity to therapeutic agents that target components of the pathway. Linking pathway deregulation with sensitivity to therapeutics that target members of the pathway likely allows using these oncogenic pathway signatures to guide the choice of targeted therapeutics [31].

Triple Negative Breast Cancer: Any Signature Available?

Among the 4–5 molecular subgroups, basal-like tumors present the worst outcome. According to current estimates, triple negative breast cancers (TNBCs) account for 10–17% of all breast carcinomas, depending on thresholds used to define ER and PR positivity and HER2 overexpression [34]. In different series and patient populations TNBC may range 6–28% of breast cancers, but higher incidence rates are reported for some ethnical groups, such as African-Americans and for younger patients, as well as for BRCA-mutation carriers [35]. Despite its relatively small proportion among all breast cancers, TNBC is responsible for a large fraction of breast cancer deaths, because of the aggressive tumor phenotype(s), only partial response to chemotherapy and present lack of clinically established targeted therapies. It should be emphasized that TNBC currently includes a heterogeneous group of tumors. By simple morphology, a group of TNBC patients with a more favorable outcome can be identified, for example patients with invasive adenoid cystic, apocrine, and typical medullary tumors. Even within the relatively homogeneous group of patients with triple-negative invasive ductal carcinoma (IDC), patients with higher or lower risk may be identified, based on specific molecular markers. For example, Viale et al. found that epidermal growth factor receptor (EGFR) immunoreactivity significantly correlates with worse prognosis in 284 patients with triple-negative IDC [35]. This underscores the importance of defining underlying risk factors for TNBC as a crucial step toward its prevention. While several potential therapeutic targets have recently surfaced from the gene expression profiling of the triple-negative tumors, the search is still onto unravel the modifiable and nonmodifiable risk factors associated with this aggressive disease. Also, additional tumor markers might allow identification of patients at higher risk of relapse [36].

A TNBC metastasis-associated signature, currently nonavailable, would be extremely useful in clinical setting, to select patients that probably have benefit from treatment and patients that can avoid toxicity of not necessary treatments. Recent discoveries in this field have been presented from three groups at San Antonio Breast Cancer Symposium 2010.

Lehmann and colleagues analyzed 386 TNBC gene-expression profile training sets from 21 independent breast cancer studies and identified six stable clusters that display unique

gene-expression patterns and gene ontologies [37]. These clusters are: two basal-like subtypes characterized by cell cycle and DNA damage response genes; two mesenchymal-like subtypes enriched in cell differentiation, epithelial–mesenchymal transition and growth factor pathways; an immunomodulatory subgroup defined by immune cell surface antigens, receptors, and signal transduction genes; and a luminal subgroup driven by androgen-receptor signaling. Lehmann et al. after the identification of representative cell lines to model each of the subgroups, treated xenografts of these TNBC subtypes and found that basal-like triple-negative disease is sensitive to cisplatin, mesenchymal-like TNBC may preferentially respond to Src and PI3K/mTOR inhibitors, and the luminal subtype is sensitive to the androgen-receptor antagonist bicalutamide and to HSP90 inhibitors. These data for the target selection in drug discovery, clinical trial design, and selection of biomarkers represent a potential approach to assign a personalized treatment to patients with TNBC.

Interestingly, another group of investigators evaluated 28 breast cancer datasets with gene-expression data and identified 12 different molecular phenotypes among 579 TNBC samples [38]. The analysis showed that 73% of TNBC are basal-like tumors, with the rest classified into phenotypes according to gene function (e.g., immune activity, angiogenesis, proliferation, apocrine activity, inflammation). Notably, there were no outcome differences between basal-like tumors and nonbasal-like tumors. However, high B-cell (immune system) and low IL-8 (inflammation) metagene expression were able to identify a subset of patients (32% of all tumors) with a favorable prognosis and a 5-year metastasis-free survival of 84%. The inhibition of the related pathways might provide new therapeutic approaches. Only the metagene ratio and lymph node status significantly predicted of prognosis in the multivariate analysis.

Goga A, on the basis of gene-expression arrays of 149 patients from the I-SPY trial, demonstrated that TNBC with high expression of MYC have worse clinical outcomes [39]. The oncogene MYC, whose genomic locus is amplified in 20–50% of all breast tumors, is associated with poorly differentiated tumors and abundant in TNBC. In the trial, patients with tumors expressing high MYC signatures had worse outcome. Disease-free survival at 5 years was approximately 95, 80, and 55% for patients with low, intermediate, and high MYC expression, respectively. In a multivariate analysis considering receptor status and MYC pathway activation as a continuous variable, triple-negative status, and MYC pathway activation had a hazard ratio of 1.5 and 16.7, respectively. The CDK1 inhibitor SCH-727965, currently in phase II trials, is a promising agent targeting MYC pathway via upregulation of the proapoptotic Bcl2 family member BIM. The cooperation between MYC overexpression and CDK1 inhibition induces cell death in triple-negative breast cancer cells and regression of tumor xenografts.

In conclusion, on the basis of transcriptomic analyses, TNBCs seem to represent a molecularly and clinically heterogeneous disease and not all of them have an unfavorable prognosis [37].

The results of these mentioned and other studies are enthusiastically awaited to discover driving pathways involved in TNBC progression and to better individualize therapy for these patients.

Clinical Application of Gene Signatures: Ongoing Trials

Adjuvant treatment for early-stage breast cancer is offered to the majority of patients after definitive surgery, assuming the presence of residual microscopic disease. Clinical–pathological prognostic factors (tumor stage, hormone receptors and HER2 expression, tumor grade, proliferative rate) are useful tools to estimate the risk of disease recurrence and thus, to decide whether to use adjuvant chemotherapy (CT), hormone therapy (HT), or biologic agents. However, during the last decade, a deeper insight into breast cancer biology led to the classification of breast cancer into different subtypes, characterized by different biological features and prognosis [1]. Several gene signatures proved to

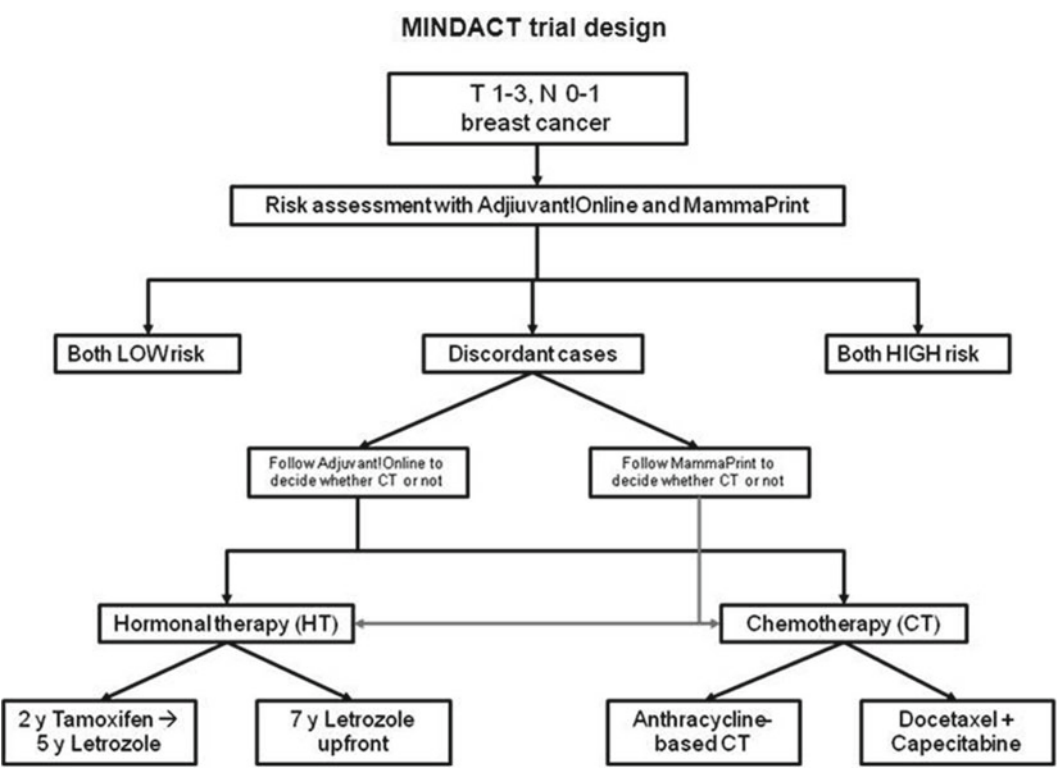


Fig. 2.1 MINDACT trial design

better estimate the risk of disease recurrence (Table 2.1), when compared to classic prognostic factors, and some of them are recommended from the American Society of Clinical Oncology (ASCO) and the NCCN Guidelines both as a prognostic and predictive tool in patients with node negative, endocrine positive disease. One of the priorities in breast cancer management is to identify patients with good-prognosis early-stage disease who could be spared adjuvant CT. Hence, the application of gene signatures in predicting the benefit of CT and identifying patients who will mostly benefit from a specific cytotoxic agent is an extremely active research field. Promising data suggesting a role in predicting benefit from adjuvant CT over endocrine therapy alone prompted the development of two large randomized phase III trials [MINDACT (Fig. 2.1) and TAILORx (Fig. 2.2)], actually ongoing.

MammaPrint™

The 70-gene expression profile was developed in the Netherlands, applying DNA-microarray technology to 78 frozen tumor samples from untreated node-negative breast cancer patients. Patients relapsing within 5 years from definitive surgery were defined as “poor prognosis,” those who remained disease-free after 5 years were considered as “good prognosis.” Researchers selected 70 genes that demonstrated to accurately classify tumors in either the poor or good-prognosis group [10]. The prognostic value of the 70-genes signature was validated in retrospective case series, showing that the use of MammaPrint could reduce misclassification of patients’ risk, hence the overtreatment of the low-risk group [40].

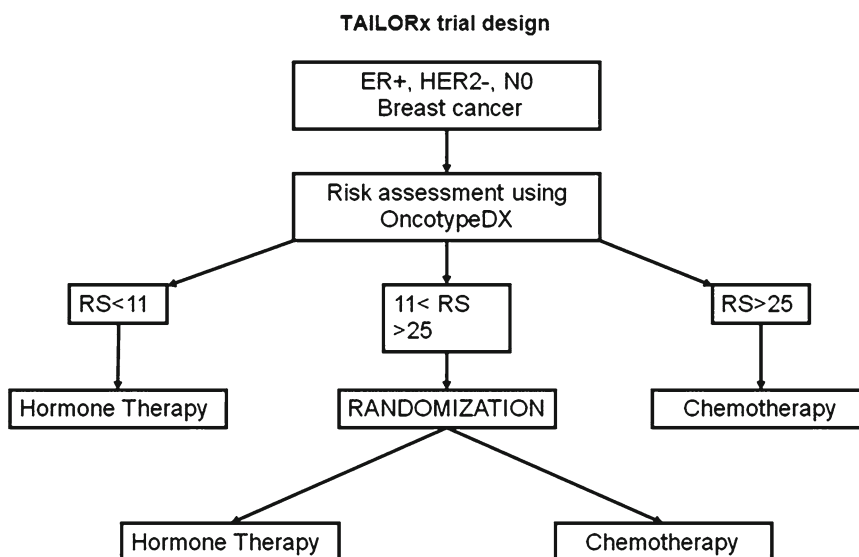


Fig. 2.2 TAILORx trial design

MammaPrint identifies a low-risk group of patients having a 10-year breast cancer survival probability $\geq 88\%$ if their tumor had ER expression $> 1\%$, and of at least 92% if ER negative [41]. Recently, the role of MammaPrint to predict CT benefit in addition to HT was assessed: patient in the high-risk group derived a greater benefit from CT in terms of distant disease-free survival (DDFS) and breast cancer specific survival (BCSS), compared to high-risk patients treated with HT alone [42]. Moreover, the application of the 70-gene profile in the neoadjuvant setting further provided evidences of the predictive value of the signature [43].

A prospective validation of MammaPrint is ongoing in a large, multicentric, randomized, controlled, phase III trial: the microarray in node negative disease may avoid chemotherapy (MINDACT) trial. Primary objective of the trial is to confirm that patients with molecular low risk can safely be spared adjuvant CT even if they have clinical high-risk tumor.

The risk of relapse of 6,000 breast cancer patients (0–3 lymph nodes involved) will be assessed using both traditional clinical–pathological criteria (Adjuvant!Online) and the MammaPrint signature. Patients estimated low risk with both methods will be spared CT; patients estimated high risk with both methods will be proposed CT; if the methods are discordant, patients are randomized to be treated according to the clinical–pathological or the genomic result. Estimating a 35% rate of discordance between Adjuvant!Online and MammaPrint, it is expected that a third of them will not be treated with CT, while it would be recommended using the conventional risk assessment criteria. Patients defined as high clinical and low genomic risk will not receive CT, and will be closely followed. The role of MammaPrint in predicting benefit from a specific chemotherapeutic agent will be assessed in a second randomization: patients will be randomly treated with an anthracycline-based CT or the combination of docetaxel and capecitabine. A third randomization, offered to all ER positive patients, will compare 2 years of T followed by 5 years of letrozole to 7 years of upfront letrozole (see Fig. 2.1 for trial design). The whole genome will be analyzed for all the 6,000 patients, aiming at discovering new signatures with prognostic and predictive value.

Accrual started in February 2007 and is still ongoing. Results of this ambitious trial are eagerly awaited.

OncotypeDx

OncotypeDx is the most widely used gene signature in everyday practice. It is an RT-PCR assay performed on formalin-fixed, paraffin embedded tissues that evaluate expression of 21 genes (16 cancer related and 5 reference genes). Levels of gene expression are combined to provide a continuous variable: the recurrence score (RS). The assay was developed in a population of patients with node-negative ER positive disease, treated with tamoxifen (T). Thus, the RS quantifies the likelihood of 10-year distant recurrence (10 yDR) in patients with the aforementioned characteristics [32]. Patients with an RS <18 are considered low risk with a 10 yDR of 6.8%; an RS ≥ 18 and <31 defines the intermediate risk group, with 10 yDR of 14.3%; finally, an RS ≥ 31 is associated with a 10 yDR rate of 30.5%. The prognostic value of the assay was validated in several retrospective trials [44]. OncotypeDx to predict the benefit of CT over T in ER positive breast cancer patients has been assessed using samples from the NSABP B20 and the SWOG S8814 trials. In both cases, patients with a high RS derived the greater benefit from the addition of CT to T, while it did not affect outcome of low-risk patients, suggesting that the later ones could safely be spared C [45, 46]. The later study was run in node-positive patients and proved an increase of both disease-free survival (DFS) and BCSS when CT was added to T only in the high-risk group (10-years BCSS in the high-risk group 73% vs. 54% for CAF+T vs. T alone, respectively) [46]. The predictive value of OncotypeDx has been confirmed in the neoadjuvant setting, proving that low RS tumors are less likely to obtain pathologic complete response than high RS [47].

Trial assigning individualized options for treatment (Rx) (TAILORx) trial is a prospective, randomized, controlled, phase III trial that aims at identifying the best individual treatment for patients with ER positive, HER2 negative, node negative breast cancer (see Fig. 2.2 for trial design). Primary objective of the trial is to compare DFS and distant recurrence-free interval (DRFI) of patients in the intermediate risk group treated with CH + HT vs. HT alone, other than create a tissue bank repository. Other objectives are to determine DFS, DRFI, and overall survival of patients with low RS and determine whether HT alone can be considered the optimal treatment for these patients.

The definitions of low, intermediate, and high risk utilized in this trial are slightly different than those previously defined:

- Low-risk patients (RS <11) will receive HT alone. A RS <11 is associated with a 10 yDR rate <5%, on average, if treated with T alone, and no benefit from CT has been demonstrated in this group.
- Intermediate-risk patients (RS 11–25) will be randomly assigned to receive CT and HT vs. HT alone. The risk of 10 yDR is approximately 10% if treated with T alone, and is considered sufficiently high to recommend CT even if CT benefit has not been clearly established for this group. The treating physician can chose the CT regimen among all those recommended from ASCO and NCCN guidelines.
- High-risk patients (RS > 26) will be assigned to receive CT. An RS >26 was selected because it is associated with a 10 yDR >20%, on average, and because CT has been shown to be beneficial in this group.

Accrual of the 11,248 was started in early 2007 and was very rapid. Actually, the study is ongoing but not accruing (clinicalTrials.gov – last assessed 4/1/2011).

Next-Generation Sequencing: An Emerging Tool to Characterize the Molecular Heterogeneity of Breast Cancer

Mutation Signatures: Mountains and Hills

Advances in whole genome sequencing technology have led us into an era of cancer genome discovery [48]. Systematic Sanger sequencing of over 13,000 genes in 11 human breast cancers has revealed preliminary insights into the breast cancer genome [49]. According to this analysis, the

average tumor harbors approximately 90 point mutations in gene coding regions. Only small minorities of these mutations were recurrent, and were in genes already known to be important in breast cancer, including *p53* and *PIK3CA*. These relatively common mutations figuratively reflect “mountains” in the mutation landscape of breast cancer [50]. In contrast, much more numerous in this landscape are the “hills,” which represent infrequent mutations in a diversity of genes that are much more numerous in any given tumor, and may or may not be contributing to the oncogenesis. Indeed, clarifying the drivers from the passengers in this complex landscape will likely be necessary to understand the heterogeneity of breast cancer, and remains one of the biggest challenges of cancer genomics.

Contrasting Paths of Genomic Evolution

Luminal and basal-like breast cancers exhibit distinct gene expression signatures, biology, and clinical behavior. Next-generation resequencing now suggests that genomic evolution during tumor progression might also follow divergent patterns between these biologic subtypes.

Comprehensive paired-end deep sequencing and transcriptome sequencing was performed of a metastatic lobular ER-positive breast cancer specimen [51]. Thirty-two nonsynonymous coding sequence mutations were identified. Significantly, 19 of these mutations were not present in the patient’s primary tumor, which had been surgically excised 9 years prior to emergence of the metastasis. Notably, no significant genome rearrangements were identified in the tumor samples, when compared to normal tissue DNA. These findings suggested that a relatively small number of point mutations might be responsible for tumorigenesis in this breast cancer, and that additional mutations arose after intervening therapy and several years of dormancy. The contribution of these mutations to metastatic recurrence remains unclear, although it seems plausible that some of these might have biologic importance.

Contrasting insights have been gleaned from deep resequencing of a basal-like primary breast cancer and its associated brain metastasis that arose in an African–American woman who succumbed to a particularly rapid and aggressive clinical course of disease [52]. The primary breast cancer in this case was ER-negative, lacked ERBB2 amplification, and was classified as an inflammatory breast cancer. Clinically, this tumor exhibited resistance to chemotherapy, and gave rise to brain metastases within 8 months, resulting in fatality shortly thereafter. In this pair of primary and metastatic specimens, only two de novo mutations distinguished the metastasis from the primary tumor. Nonetheless, the relative abundance of shared mutations varied between the metastasis and primary tumor, suggesting that selective processes during metastasis emergence were operative on preexisting dominant clones within the primary tumor. Somatic genomic rearrangements and copy number alterations were also abundant in this primary tumor, and most of these were also shared in the metastasis. However, there was evidence of increased copy number alterations and structural variants during the course of tumor progression, which might contribute to metastasis emergence, or alternatively might be a byproduct of chemotherapy or ionizing radiation exposure.

Indeed, a comparison of genomic profiles from primary tumor to metastasis in luminal and basal-like breast cancer reveals more differences than similarities. Further analysis of primary tumor–metastasis–normal tissue triads will determine the generality of these observations. An improved understanding of the differing rates of genome mutation and evolution among breast cancers would likely have implications for prognosis, and may possibly also reveal novel approaches to therapeutic intervention.

Genome Rearrangement Profiles: Diversity and Complexity

Paired-end sequencing and structural genomic rearrangement analysis has been performed on a cohort of 24 primary breast tumors and immortalized breast cancer cell lines [53]. This high-resolution analysis revealed, for the first time, the landscape of chromosomal rearrangements in breast cancer. A striking discovery was the abundance of intrachromosomal rearrangements, which could not be readily characterized by previous techniques. Also apparent were distinct patterns of rearrangements in the various biologic subtypes of breast cancer. Hormone receptor-positive breast cancers exhibited either minimal genomic instability or rearrangements that almost exclusively involved amplified segments of the genome. In contrast, TNBCs exhibited rampant instability, frequently predominated by intrachromosomal tandem duplications. Sequence analysis of these break-points generally revealed short segments of microhomology at the junction of these rearranged segments, implicating DNA repair processes in their etiology. Of note, *BRCA1* and *BRCA2* null tumors did not exhibit this preponderance of tandem duplications, suggesting that a distinct pathway may be deranged in this subgroup of tumors.

Do these distinct patterns of genomic instability have implications for patient prognosis? An analysis by the Borreson-Dale group using multiple cohorts of array-CGH data suggests that it does. In this study, the authors characterized two quantitative indexes of genomic instability – a whole arm aberration index (WAAI) and a complex arm aberration index (CAAI) – that can be applied across array CGH platforms [54]. These two instability indices were used to segregate 595 breast tumors into distinct subgroups. Significantly, both indices correlated with poorer patient prognosis in univariate and multivariate analyses. These findings suggest that patterns of genomic instability may reflect an underlying biology of breast cancer that has implications for the likelihood of cancer recurrence, as well as the clinical responsiveness to DNA-damaging therapies.

Summary

Interestingly, particular gene classifiers able to predict the aggressiveness of a tumor and the tumor's ability to home and proliferate in an organ-specific manner have been identified. Gene-expression profiling has already provided important contribution into the biologic heterogeneity of breast cancer. The combination of several advanced genomic tools with new approaches described in this chapter has to be optimally incorporated. Whether this is feasible and effective will depend on trials, such as TAILORx and MINDACT. Moreover, technical and economic issues need to be examined, and optimization of the eventual use of these tools in decision-making process is necessary.

References

1. Perou CM, Sørli T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406:747–52.
2. Perou CM, Jeffrey SS, van de Rijn M, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci USA*. 1999;96:9212–7.
3. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA*. 2001;98:10869–74.
4. Stingl J, Caldas C. Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nat Rev Cancer*. 2007;7(10):791–9.
5. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA*. 2003;100:8418–23.

6. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci USA*. 2003;100:10393–8.
7. Potti A, Dressman HK, Bild A, et al. Genomic signatures to guide the use of chemotherapeutics. *Nat Med*. 2006;12:1294–300.
8. Rouzier R, Perou CM, Symmans WF, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res*. 2005;11:5678–85.
9. Smid M, Wang Y, Zhang Y, et al. Subtypes of breast cancer show preferential site of relapse. *Cancer Res*. 2008;68(9):3108–14.
10. Van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415:530–6.
11. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*. 2005;365:671–9.
12. Jansen MP, Foekens JA, van Staveren IL, et al. Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. *J Clin Oncol*. 2005;23(4):732–40.
13. Kang Y, Siegel PM, Shu W, et al. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell*. 2003;3(6):537–49.
14. Zhang XH, Wang Q, Gerald W, et al. Latent bone metastasis in breast cancer tied to Src-dependent survival signals. *Cancer Cell*. 2009;16(1):67–78.
15. Smid M, Wang Y, Klijn JG, et al. Genes associated with breast cancer metastatic to bone. *J Clin Oncol*. 2006;24(15):2261–7.
16. Minn AJ, Gupta GP, Siegel PM, et al. Genes that mediate breast cancer metastasis to lung. *Nature*. 2005;436(7050):518–24.
17. Minn AJ, Gupta GP, Padua D, et al. Lung metastasis genes couple breast tumor size and metastatic spread. *Proc Natl Acad Sci U S A*. 2007;104(16):6740–5.
18. Bos PD, Zhang XH, Nadal C, et al. Genes that mediate breast cancer metastasis to the brain. *Nature*. 2009;459(7249):1005–9.
19. Yu JX, Sieuwerts AM, Zhang Y, et al. Pathway analysis of gene signatures predicting metastasis of node-negative primary breast cancer. *BMC Cancer*. 2007;7:182.
20. Stadler ZK, Come SE. Review of gene-expression profiling and its clinical use in breast cancer. *Crit Rev Oncol Hematol*. 2009;69(1):1–11.
21. Chang JC, Hilsenbeck SG, Fuqua SA. Genomic approaches in the management and treatment of breast cancer. *Br J Cancer*. 2005;92:618–24.
22. McShane LM, Radmacher MD, Freidlin B, et al. Methods for assessing reproducibility of clustering patterns observed in analyses of microarray data. *Bioinformatics*. 2002;18:1462–9.
23. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347:1999–2009.
24. Chang HY, Sneddon JB, Alizadeh AA, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol*. 2004;2:E7.
25. Chang HY, Nuyten DS, Sneddon JB, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci USA*. 2005;102:3738–43.
26. Ma XJ, Wang Z, Ryan PD, et al. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell*. 2004;5:607–16.
27. Sotiriou C, Wirapati P, Loi S, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst*. 2006;98:262–72.
28. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol*. 2007;25:1239–46.
29. Miller LD, Smeds J, George J, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA*. 2005;102:13550–5.
30. Coutant C, Rouzier R, Qi Y, et al. Distinct p53 gene signatures are needed to predict prognosis and response to chemotherapy in ER-positive and ER-negative breast cancers. *Clin Cancer Res*. 2011. doi:10.1158/1078-0432.CCR-10-1045.
31. Bild AH, Yao G, Chang JT, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature*. 2006;439(7074):353–7.
32. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351:2817–26.
33. Fan C, Oh DS, Wessels L, et al. Concordance among gene expression-based predictors for breast cancer. *N Engl J Med*. 2006;355:560–9.
34. Podo F, Buydens LM, Degani H, et al. Triple-negative breast cancer: present challenges and new perspectives. *Mol Oncol*. 2010;4(3):209–29.

35. Viale G, Rotmensz N, Maisonneuve P, et al. Invasive ductal carcinoma of the breast with the “triple-negative” phenotype: prognostic implications of EGFR immunoreactivity. *Breast Cancer Res Treat.* 2009;116(2):317–28.
36. Maiti B, Kundranda MN, Spiro TP, et al. The association of metabolic syndrome with triple-negative breast cancer. *Breast Cancer Res Treat.* 2010;121(2):479–83.
37. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest.* 2011;121(7):45014. doi:10.1172/JCI45014.
38. Rody A, Karn T, Liedtke C, et al. Identification of a clinically relevant gene signature in triple negative and basal-like breast cancer. 33rd Annual San Antonio Breast Cancer Symposium. Abstract S5-5. Presented December 11, 2010.
39. Goga A, Horiuchi D, Kusdra L, et al. Synthetic-lethality of triple-negative breast cancers via the MYC oncogene pathway. 33rd Annual San Antonio Breast Cancer Symposium. Abstract S5-4. Presented December 11, 2010.
40. Mook S, Van’t Veer LJ, Rutgers EJ, et al. Individualization of therapy using MammaPrint: from development to the MINDACT Trial. *Cancer Genomics Proteomics.* 2007;4(3):147–55.
41. Bogaerts J, Cardoso F, Buyse M, et al. Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat Clin Pract Oncol.* 2006;3(10):540–51.
42. Knauer M, Mook S, Rutgers EJ, et al. The predictive value of the 70-gene signature for adjuvant chemotherapy in early breast cancer. *Breast Cancer Res Treat.* 2010;120(3):655–61.
43. Straver ME, Glas AM, Hannemann J, et al. The 70-gene signature as a response predictor for neoadjuvant chemotherapy in breast cancer. *Breast Cancer Res Treat.* 2010;119(3):551–8.
44. Habel LA, Shak S, Jacobs MK, et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res.* 2006;8(3):R25.
45. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol.* 2006;24(23):3726–34.
46. Albain KS, Barlow WE, Shak S, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol.* 2010;11(1):55–65.
47. Gianni L, Zambetti M, Clark K, et al. Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. *J Clin Oncol.* 2005;23(29):7265–77.
48. Stratton MR. Exploring the genomes of cancer cells: progress and promise. *Science.* 2011;331:1553–8.
49. Sjöblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science.* 2006;314:268–74.
50. Wood LD, Parsons DW, Jones S, et al. The genomic landscapes of human breast and colorectal cancers. *Science.* 2007;318:1108–13.
51. Shah SP, Morin RD, Khattra J, et al. Mutational evolution in a lobular breast tumor profiled at single nucleotide resolution. *Nature.* 2009;461:809–13.
52. Ding L, Ellis MJ, Li S, et al. Genome remodeling in a basal-like breast cancer metastasis and xenograft. *Nature.* 2010;464:999–1005.
53. Stephens PJ, McBride DJ, Lin ML, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature.* 2009;462:1005–10.
54. Russnes HG, Vollen HK, Lingjaerde OC, et al. Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci Transl Med.* 2010;2:38–47.

Diagnostic, Prognostic and Therapeutic Value of Gene
Signatures

Russo, A.; Iacobelli, S.; Iovanna, J. (Eds.)

2012, XI, 182 p. 28 illus., 24 illus. in color., Hardcover

ISBN: 978-1-61779-357-8

A product of Humana Press