available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Gene expression predictors in breast cancer: Current status, limitations and perspectives

C. Desmedt[a], E. Ruíz-García[b], F. André[b,*]

[a]Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium
[b]Translational Research Unit, UPRES EA03535 and Department of Medical Oncology Comite 050, Université Paris XI, Institut Gustave Roussy, 39 rue C. Desmoulins, 94805 Villejuif, France

ARTICLE INFO

Article history:
Received 16 April 2008
Accepted 23 September 2008
Available online 1 November 2008

Keywords:
Breast cancer
Gene expression profiles
Prognosis
Prediction

ABSTRACT

Breast Cancer is characterised by a wide heterogeneity regarding outcome and drug sensitivity. A better prediction of these two parameters at the individual level should improve patient management and therefore also improve both the quality of life and the overall survival of the patient. Several molecular predictors for prognosis (MammaPrint® or Oncotype DX) and drug prediction (DLD30, SET index) have been generated using DNA-based arrays or RT-PCR, some of these being tested in phase III trials. Although they exhibit good metric performance and should improve the quality of care in the next decade, these predictors are considered suboptimal regarding the potential of the technology. New study design and arrays should generate more powerful second generation gene signatures.

1. Introduction

Breast cancer is paradoxically both the leading cause of cancer deaths in women in the western countries and one of the most frequently cured cancer. The recent epidemiology of breast cancer shows an increased rate of good prognosis breast cancer, together with an increment of use in adjuvant medical treatments. As a result of screening mammograms and education, the rate of node negative breast cancers <2 cm has increased in the recent years. In these patients, according to Adjuvant! Online,[1] the 10-year relapse rate ranges between 14% and 29% after surgery alone. In the same population, the absolute benefit from adjuvant chemotherapy ranges between 2% and 16%. These data point out the fact that most of the patients treated with this treatment modality do not get any benefit, albeit presenting toxic effects. This consideration has led to the hypothesis that identifying predictors for prognosis could identify patients who could be spared from adjuvant chemotherapy. The expected benefit from such predictors would be to decrease acute and late toxic effects and to reduce the cost associated with the treatment of the disease. If not developed in line with the rules of evidence-based medicine, such predictors could lead to the under-treatment of thousands of women, and therefore be more harmful than helpful.

During the last decades, several drugs have shown efficacy in breast cancer. As an illustration, a node positive breast cancer patient is currently treated with an anthracycline and taxane containing regimen, followed by endocrine therapy in case of ER expression, and trastuzumab in case of HER2 overexpression. Although currently proven as being the most effective for the whole population, this 'one fits all' approach presents some limitations: (i) each single drug included in chemotherapy regimen is administered at suboptimal doses for sensitive cases, (ii) some highly sensitive patients to a given drug will receive unnecessary additional treatments, (iii) accumulation of treatments would not be cost-effective if highly sensitive patients could be identified. These considerations have led to the hypothesis that the identification of molecular predictors for drug efficacy could (i) improve out-

* *Corresponding author:* Tel./fax: +33 142114211.
E-mail address: fabrice.andre@igr.fr (F. André).
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2008.09.011

come by allowing an optimal drug delivery in a given patient and (ii) decrease cost and toxic effects by sparing patients from additional unnecessary medical treatments.

This introduction highlights the need for two different sets of molecular predictors: (i) predictors for prognosis and (ii) predictors for drug sensitivity. In the first part of this review, we will discuss the feasibility of a molecular diagnosis by DNA arrays (or RT-PCR). In the second part, we will present several illustrations of the so-called-'first generation' of molecular predictors, with an emphasis on the limitations that are usually pointed out. Finally, in the last part, we will discuss the perspectives of the DNA array-based gene expression signatures, including the clinical implementation of existing ones and the design of the 'second generation' of DNA array-based molecular predictors.

2. DNA array-based molecular predictors: principle and clinical applicability

2.1. *Principle*

DNA chips or microarrays allow the quantification of the expression of several thousands of genes in a single experiment. The concept, the different approaches and technologies of DNA microarrays have already been described extensively (see [2] for review). In brief, the technique relies on accurate hybridisation of strands of DNA with its corresponding mRNA derived from the tissue or cell line sample being studied. A fluorescent probe is then measured by a laser scanner which will allow the researcher to determine if the expression of the gene is up or down-regulated, unchanged or absent compared with a control level.

The array can either be a genome-wide array or a dedicated array, specifically set up for a given purpose (MammaPrint®).[3,4] The DNA array technology has allowed to build multigene-based molecular predictors, also called 'gene signatures'. These signatures are usually based on the differentially expressed genes between the two conditions they are aimed at predicting prognosis or response to a given treatment. Several bioinformatics approaches that will not be described here have been used to generate these signatures (reviewed in [5]).

During the last years, some concerns have arisen regarding the way DNA arrays are used to generate optimal predictors.[5,6] The two most frequent criticisms are related to (i) the number of events included in training set and stability of the predictive value of gene signature over series[7] and (ii) the added value of a molecular signature as compared to an optimal clinico-pathological score.[6] These two limitations as well as other more technical concerns regarding this technology will be discussed further.

2.2. *Clinical applicability*

Since the technology is highly complex and requires several steps of specific technical expertise, some criticisms have arisen regarding the reliability and clinical applicability of DNA array-based molecular predictors. Most of these technical issues have been addressed in the recent years, mainly within the Microarray Quality Control (MAQC) project.[8]

The first usual criticism was related to the reliability of the quantification of gene expression. The MAQC project has shown a high degree of correlation between quantitative RT-PCR and gene expression determined by DNA array, when the array is performed in well-trained laboratories and using commercially available arrays. Also, in a study dedicated to breast cancer, we observed a high level of correlation between ESR1 (probeset: 205225_at, Affymetrix U133A) and ERBB2/HER2 (probeset: 216836_s_at, Affymetrix U133A) gene expression levels with ER and HER2 immunostainings for both proteins.[9] These data suggest that DNA arrays are a reliable technology to evaluate gene expression levels.

The other common criticism is related to the inter-laboratory reproducibility. The MAQC project compared gene expression measurements of two RNA samples using a number of microarray platforms, as well as alternative technologies, and demonstrated intra-platform consistency and inter-platform concordance in terms of genes differentially expressed. As an illustration of gene signatures applied to breast cancer, Ach et al. showed a high intra-laboratory and inter-laboratory reproducibility regarding the 70-gene signature (MammaPrint®).[10] Interestingly, they also reported that two hybridisations of a given sample several months apart provided similar results.

Another issue relates to pre-analytical process. To be qualified for hybridisation after a single round of amplification, it is recommended to have high quality RNA, which is commonly measured by the Agilent Bioanalyser, and an RNA amount of at least 1 μg. In addition, it is usually recommended to have a minimal percentage of tumour cells in the samples to be analysed. Although the criteria for quality control are matter of controversies, it is a fact that some tumours cannot be qualified for hybridisation on DNA arrays, putting forward the fact that such technology could not be applied for a minority of patients. Several options are being developed to increase the percentage of eligible tumours for DNA array-based diagnosis. Double amplification could decrease the amount of RNA required for diagnosis, while random priming would allow samples with some degree of RNA fragmentation to be considered as eligible. Finally, some have suggested that fine needle aspiration could enrich sample in malignant cells and therefore abrogate the need for a high percentage of malignant cells in the tumour. Thus, there exist some options which allow proposing a DNA array-based diagnosis in most of the patients eligible for such approach.

Altogether, these data show that DNA microarrays are reliable to measure gene expression levels, in a reproducible manner and that this technology can be considered 'ready for use' in prospective clinical trials.[11]

3. Prognostic gene signatures

As stated in the introduction, there is a need to more accurately determine which patients are at risk for metastatic relapse. Several gene signatures have been developed in this setting, using either the top-down or the hypothesis-driven approach (reviewed in [12]). We will briefly describe the development of the 70-gene signature[2] and the Genomic Grade Index[13] in order to provide an example of each approach.

The Netherlands Cancer Institute in Amsterdam and Rosetta were the first to conduct comprehensive genome-wide assessments of gene expression profiling to identify broadly applicable prognostic markers using the 'top-down' approach, i.e. they derived a prognostic model from global gene expression data simply by seeking gene expression profiles that are associated with clinical outcome without any prior biological assumption.[2] Using the Agilent microarray platform, they generated this molecular predictor considering a leave-one-out approach from a list of the 231 genes differentially expressed between tumours from patients free-of-metastasis and those from patients who presented a metastatic relapse. The discovery phase presented some important features. First, the metastatic relapse was selected to have been observed within the 5 year after diagnosis. Patients who relapsed after 5 years were not considered as events, thereby avoiding heterogeneity in the biological process of metastases. Second, clinical characteristics were not taken into account when the molecular predictor was generated. This implies that the genes were not selected to provide additional information to clinical characteristics. Third, this discovery phase included a high number of cases ($n = 78$) and events ($n = 44$). After a validation phase with samples from the same centre,[14] the performance of the molecular predictor has been evaluated in a consortium of six European centres. In this study, 113 of 207 patients presented a good prognosis signature.[15] Such validation showed that MammaPrint® exhibits a significant prognostic value (hazard ratio for distant metastasis: 2.32, 95% confidence interval (CI): 1.35–4). The sensitivity and specificity to predict 10-year breast cancer death were 0.84 (0.73–0.92) and 0.42 (0.36–0.48). While Adjuvant! Online[1] exhibited a similar sensitivity 0.82 (0.71–0.90), its specificity was lower 0.29 (0.23–0.35). These findings suggested that MammaPrint® could potentially increase the detection rate of patients with good prognosis, and thereby allow a decrease in the use of chemotherapy in these patients. This has been the basis for a randomised trial that will address whether the use of MammaPrint® could allow a decrease in the use of adjuvant chemotherapy (MINDACT trial).[11] If the main hypothesis of this trial is validated, this study will be the first one to provide a level I evidence for a decrease in the indications of adjuvant chemotherapy. It must be pointed that no clinical or pathological parameter has been reported with such level of evidence to decrease the indications of adjuvant chemotherapy.

An example of deriving a prognostic gene expression signature using a hypothesis-driven approach was the study reported by our group that focused on histological grade, a well-established pathological parameter rooted in the cell biology of breast cancer.[13] Indeed, clinicians face a huge problem with respect to patients who have intermediate-grade tumours (grade 2), as these tumours, which represent 30–60% of cases, are the major source of inter-observer discrepancy and may display intermediate phenotype and survival, making treatment decisions for these patients a great challenge, with subsequent under- or over-treatment. Performing a supervised analysis, we developed a Gene expression Grade Index (GGI) score based on 97 genes. These genes were mainly involved in cell cycle regulation and proliferation and were consistently differentially expressed between low and high grade breast carcinomas. Interestingly, the GGI, which essentially quantifies the degree of similarity between the tumour expression pattern of these 97 genes and tumour grade, was able to reclassify patients with histological grade 2 tumours into two groups with distinct clinical outcomes similar to those with histological grades 1 and 3, respectively.

Additionally, several signatures have been generated to predict prognosis in patients treated with endocrine therapy. The recurrence score (Oncotype DX) was generated as a continuous variable to predict prognosis of patients treated with tamoxifen in the NSABP-B14 trial.[16] This signature is a combined score of 16 genes of interest, including ER, PGR, HER2 and Ki67. This predictor has allowed to identify a population (recurrence score < 18) that presents a very good prognosis (6.8% of 10-year distant metastasis), and could be spared from adjuvant chemotherapy. As for other molecular predictors, the added value of recurrence score as compared to optimal clinico-pathological score has not yet been formally proven. Of note, this score is a gene expression signature derived not from DNA microarrays but using quantitative RT-PCR, and an interesting feature of this test is the fact that frozen tissues are not needed since the score is generated using paraffin-embedded tissues. Since it became commercially available, Recurrence score has been ordered for 40,000 for patients.[17] As for the 70-gene signature, the process of clinical validation to achieve level I evidence is ongoing for this signature in a clinical trial named TAILORx—Trial Assigning IndividuaLized Options for Treatment (Rx).

Several criticisms, questions and perspectives have been raised regarding the prognostic gene signatures reported so far. First, some argue that most of the signatures only add little information compared to an optimal clinico-pathological score that would include ER, Her2 and Ki67 in addition to the conventional clinical parameters. It must also be pointed out that most of the genes included in the various published prognostic gene signatures are related to cell proliferation[18], and the question then arises as whether a simpler biomarker for such parameter like Ki67, which has been measured routinely for decades, could have provided similar results.[19] However, gene expression profiling studies suggest that measuring proliferation with a more objective, automated and quantitative assay may be more robust than less quantitative assays such as immunohistochemistry.

Another criticism relates to the fact that most of the predictors were generated using a mix of molecularly heterogeneous tumours. It has been reported that breast cancer population is a mix of at least four different molecular classes, i.e. basal-like tumours, Her2+/ER– tumours, luminal A and B tumours (summarised in [12]). Since oncogenic events are different across molecular classes, some have suggested that optimal predictors should be set up in each molecular class. [20] This was applied, for example, by Wang et al. who developed a 76-gene signature to identify patients at a high risk of distant recurrence based on the prognostic genes separately identified in ER– and ER+ tumours.[21] We also showed in a recent meta-analysis of publicly available gene expression breast cancer data that proliferation is the strongest parameter predicting clinical outcome in the ER+/HER2– subgroup of patients only, whereas immune response and tumour invasion appear to be the main biological processes

associated with prognosis in the ER–/HER2– and HER2+ subgroups, respectively.[18] This implies that the molecular background of the tumour should be taken into consideration to make prediction regarding prognosis.

While sensitivity of gene signatures for prognostic purpose looks good, there is still around 5–10% of patients who will present a metastatic relapse in the group predicted as low risk of relapse. If we considered that third generation chemotherapy decreases by ~30% the risk of breast cancer death in the whole population of breast cancer[22], it cannot be excluded that an optimal chemotherapy could provide benefit even in this good prognostic population. Nevertheless, it must be emphasised that most of the tumours classified as good prognosis are actually predicted to be resistant to chemotherapy.[23]

Some gene signatures have been reported to present a strong time-dependency.[15,24] This finding makes sense, since these signatures were built to predict the occurrence of metastases within the first 5 years, and are enriched in genes involved in cell proliferation. This consideration points out the need for a predictor of late metastatic relapse. This predictor could be set up based on the patients with ER+ disease who are free-of-disease at 5 years, and could help determining which patients are eligible for prolonged adjuvant endocrine therapy.

To conclude, several gene signatures for prognostic purpose have been generated at this time. At least two of them are being validated in prospective trials.[2,16] Although they allow an increase in the rate of patients who could be spared adjuvant chemotherapy while still correctly identifying the high-risk patients, they present some limitations that will have to be taken into account to generate more accurate 'second generation' gene signatures.

4. Gene expression predictors for drug sensitivity

In addition to prognostic signatures, some gene signatures have been developed to identify patients who are highly sensitive to specific drugs. The goal of such approach is to deliver the appropriate drug to the right patients, thereby decreasing the use of other unnecessary drugs. This approach is named pharmacogenomics.

Drug-specific predictors are usually developed in the preoperative setting when the drug is a chemotherapeutic agent. In a study that included 133 patients treated with a preoperative paclitaxel (T)/FAC regimen, Hess et al. have identified a 30-probe predictor for T/FAC efficacy.[25] The predictor presented a 92% sensitivity and 71% specificity to predict a pathological complete response (PCR). A prospective validation of such predictor is being started. When compared to an optimal clinico-pathological score, the predictor presented a better sensitivity (0.92 versus 0.61), but the overall accuracies of both approaches were similar (0.76 and 0.78). Further studies have shown that some probes (microtubule-associated protein-tau (MAPT)) included in the score correlate with ER and HER2 statuses.[26,27] This highlights the fact that, as for predictors for prognostic purpose, the first generation of gene signature usually includes genes related to clinical characteristics. In addition to the limitations previously discussed for prognostic signatures, a potential limitation to this predictor is related to the fact that using single arm treatment in the preoperative setting does not allow to identify pure drug-specific predictors. With this design, the predictor is indeed expected to be a mixture of non-specific cytotoxic predictors and drug-specific predictors. Using randomisedrandomised trials comparing two different drugs could be one solution to improve the drug specificity of the predictor.

Others groups have developed different approaches to generate predictor for drug sensitivity. Lee et al. have for instance used 60 panel cell lines from NIH to generate predictors for drug sensitivity (paclitaxel).[28] They showed that when applied to a small clinical dataset these predictors accurately predicted sensitivity to docetaxel ($p = 0.03$).

Efforts have also been made to identify predictors for endocrine sensitivity. The identification of patients highly sensitive to endocrine therapy could lead to a decrease in the use of chemotherapeutics, by identifying the patients displaying a good clinical outcome because of endocrine therapy who would therefore not need additional treatment.

In two studies, investigators used the concept that the activation of ER-regulated genes is associated with a high sensitivity to endocrine therapy. Symmans et al. have generated a predictor that includes 200 probesets selected to be correlated with ER expression (SET index).[29] In the first validation cohorts, the SET index, as a continuous variable, significantly predicted prognosis in patients treated with tamoxifen ($n = 211$, HR: 0.71, $p = 0.005$), but not in those who did not receive endocrine therapy ($n = 174$, HR: 0.92, $p = 0.19$). In the second validation cohorts, SET index was associated with a 32% reduction in the risk of metastatic relapse for tamoxifen-treated patients (HR: 0.68, $p = 0.01$), as compared to a 2% reduction in untreated patients (HR: 0.98, $p = 0.81$). A similar approach was used by Oh et al.[30] In this study, the authors used a set of genes induced by an in vitro exposure of MCF7 cell line to estradiol to further generate a gene signature of 822 genes that aim at predicting endocrine sensitivity.

5. Perspectives

5.1. *Development of second generation predictors*

As discussed in the previous sections, while gene signatures clearly represent a major step forward in the molecular prediction of patient outcome and drug sensitivity, they are still showing some limitations that nowadays are the basis for the development of the second generation gene predictors.

These limitations are summarised in Table 1, together with some solutions which we will describe here. First, considering that most of the first generation gene signatures capture genes that are correlated to clinico-pathological characteristics, some are developing molecular predictors that contain genes selected to provide additional information to the clinico-pathological characteristics. Second, since it is now clear that breast cancer can be considered as four distinct molecular groups, next generation of molecular predictors should focus on homogenous classes, as already done by several groups.[18,31] Third, predictors for metastatic relapse should be designed to predict both early and late relapses. Fourth, drug-specific predictors should be optimally defined either in randomised trials that compare two drugs, or based on the two series of similar patients treated with different

Table 1 – Strengths and weaknesses of the first generation gene expression predictors

Strengths	Weaknesses	Solutions (second generation molecular predictors)
Feasible and reproducible across platforms Validated in retrospective studies Medical usefulness under evaluation in phase III trials (will provide level I evidence for clinical use)	*Predictors for prognosis and drug sensitivity:* • Added value to optimal clinico-pathological characteristics not proven • Metric performances not optimal • Generated regardless of breast cancer molecular classes • Generated using the 1st generation of arrays *Predictors for prognosis* • Poor performance to predict metastases over 5 years *Predictors for drug sensitivity* • Not generated based on the direct comparison between two drugs: drug specificity to be determined	• Predictors generated to provide additional prediction to an optimal clinico-pathological score: select probes that increase performance of clinical characteristics, etc. • Predictors generated for each molecular class • Use of last generation of arrays to generate predictors: exon arrays and splice arrays • Predictor for late events • Generation of drug-specific predictor based on the randomised trials (interaction test)

treatment schedules. This latter approach was used to generate a predictor that is specific to ixabepilone B as opposed to paclitaxel.[25]

DNA microarrays used to generate first generation predictors contained between 22 and 44 K probesets. Although some genes presented multiple probesets, the vast majority of such probes were located in 3'. This implies that gene profiling actually detects the transcriptional activity, but is unable to determine which of the transcripts are expressed for a single gene. Several new arrays[32], including Splice Arrays, are detecting both gene expression and splicing events, thereby providing a more functional picture of the genomic programme in every single patient.

5.2. *Long term perspectives*

While most of the predictors previously discussed will be determined either in a central laboratory (Oncotype DX) and/or using dedicated array (MammaPrint®), it must be emphasised that DNA arrays offer the major advantage of being able to provide genome-wide gene expression measurements.

It must also be noticed that DNA arrays can be used for many other purposes than to predict prognosis and treatment efficacy. Indeed, this technology has been used to generate molecular predictors for the diagnosis of malignancy (Andre F, data not shown), organ-specific metastases[33], lymph node

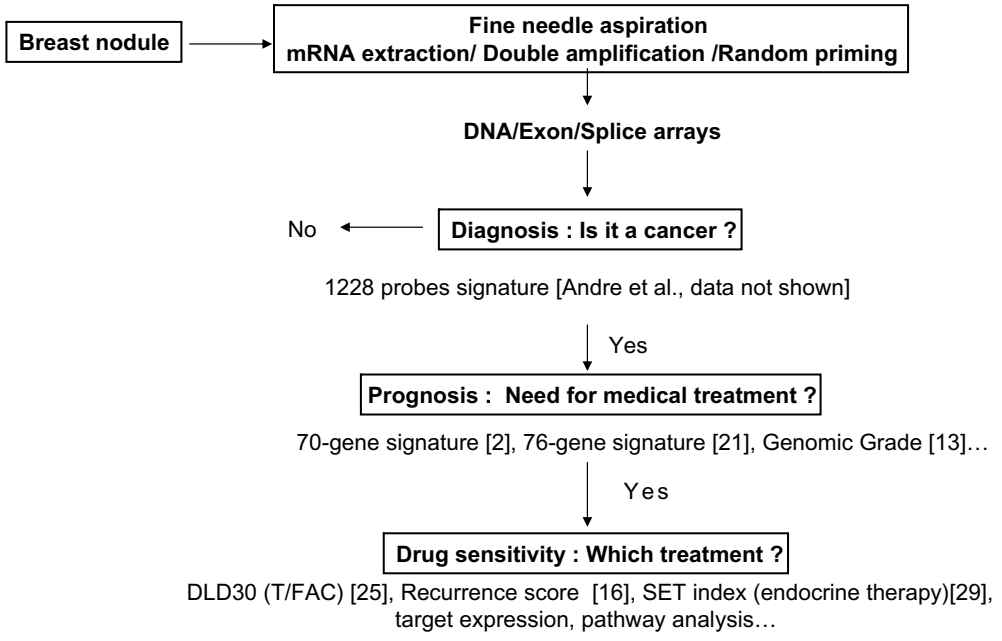

Fig. 1 – Potential of molecular predictors for breast cancer in the next decades. This figure reports the author's view of how gene signature will be integrated in the clinical practice in the next decades. This view highlights the point that array-based predictors could be used as an 'all-in-one' bioassay to get information about diagnosis, prognosis and drug sensitivity within a very short timeframe.

involvement[34], as well as the identification of activated pathways[35] and the expression of therapeutic targets.[9] Also the fact that recent technological advances have made possible the measurement of gene expression in single cells[36,37] opens new avenues to evaluate gene expression profile of circulating tumour cells, cancer stem cells or some other cells of interest including mesenchymal stem cells and endothelial progenitor cells.

This suggests that besides prognostic and predictive information related to chemotherapy and endocrine therapy, DNA arrays could allow in the future to determine all the information needed for optimal patient's care. Since the current trend is to increase the number of independent bioassays or analyses to be done in a single sample (conventional pathology, immunohistochemistry, FISH, RT-PCR,...), the use of DNA arrays could change this tendency by doing all analyses in a single experiment in certified laboratories. In addition to facilitate logistics, decrease time before treatment decision, this approach would probably save costs by decreasing the rate of bioassays to be done. Fig. 1 illustrates how genome-wide DNA arrays could affect patient's management in the next decades.

6. Conclusion

Pioneer studies have shown that gene signatures provide powerful information that should allow a better tailoring of treatment for breast cancer patients in the next few years. Based on the limitations exhibited by the first generation predictors, the second generation of gene signatures is being developed. This new wave of predictors should present a higher accuracy and be complementary to the pre-existing decision tools.

Conflict of interest statement

None declared.

Acknowledgement

This work was supported by the Belgian National Foundation for Cancer Research FNRS (C.D.).

REFERENCES

1. <http://www.adjuvantonline.com/index.jsp>.
2. Loi S, Desmedt C, Cardoso F, Piccart M, Sotiriou C. Breast cancer gene expression profiling: clinical trial and practice implication. *Pharmacogenomics* 2005;**6**:49–58.
3. 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.
4. <http://www.agendia.com/en/Professional/About-MammaPrint/About-MammaPrint>.
5. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 2007;**99**:147–57.
6. Koscielny S. Critical review of microarray-based prognostic tests and trials in breast cancer. *Curr Opin Obstet Gynecol* 2008;**20**:47–50.
7. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2005;**365**:488–92.
8. MAQC Consortium. The microarray quality control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 2006;**24**:1151–61.
9. Gong Y, Yan K, Lin F, et al. Determination of oestrogen-receptor status and ERBB2 status of breast carcinoma: a gene-expression profiling study. *Lancet Oncol* 2007;**8**:111–203.
10. Ach RA, Floore A, Curry B, et al. Robust interlaboratory reproducibility of a gene expression signature measurement consistent with the needs of a new generation of diagnostic tools. *BMC Genomics* 2007;**8**:148.
11. Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ. Clinical application of the 70-gene profile: the MINDACT trial. *J Clin Oncol*. 2008;**26**:729–35.
12. Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat Rev Cancer* 2007;**7**(7):545–53.
13. 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.
14. 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.
15. Buyse M, Loi S, van't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 2006;**98**:1183–92.
16. 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.
17. Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. *J Clin Oncol* 2008;**26**:721–8.
18. Desmedt C, Haibe-Kains B, Wirapati P, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res* 2008;**14**:5158–65.
19. de Azambuja E, Cardoso F, de Castro G, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. *Br J Cancer* 2007;**96**:1504–13.
20. Andre F, Pusztai L. Molecular classification of breast cancer: implications for selection of adjuvant chemotherapy. *Nat Clin Pract Oncol* 2006;**3**:621–32.
21. 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.
22. Martin M, Pienkowski T, Mackey J, et al. Adjuvant docetaxel for node-positive breast cancer. *N Engl J Med* 2005;**352**:2302–13.
23. 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**:3726–34.
24. Desmedt C, Piette F, Loi S, et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* 2007;**13**:3207–14.
25. Hess KR, Anderson K, Symmans WF, et al. Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* 2006;**24**:4236–44.

26. Andre F, Hatzis C, Anderson K, et al. Microtubule-associated protein-tau is a bifunctional predictor of endocrine sensitivity and chemotherapy resistance in estrogen receptor-positive breast cancer. *Clin Cancer Res* 2007;**13**:2061–7.
27. Andre F, Mazouni C, Liedtke C, et al. HER2 expression and efficacy of preoperative paclitaxel/FAC chemotherapy in breast cancer. *Breast Cancer Res Treat* 2007; [Epub ahead of print].
28. Lee JK, Havaleshko DM, Cho H, et al. A strategy for predicting the chemosensitivity of human cancers and its application to drug discovery. *Proc Natl Acad Sci U S A* 2007;**104**:13086–91.
29. Symmans WF, Hatzis C, Sotiriou C, et al. The Microarray-based Endocrine Prediction Collaborat. Ability of a 200-gene endocrine sensitivity index (SET) to predict survival for patients who receive adjuvant endocrine therapy or for untreated patients. In: ASCO 2007, breast cancer symposium.
30. Oh DS, Troester MA, Usary J, et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *J Clin Oncol* 2006;**24**:1656–64.
31. Loi S, Sotiriou C, Buyse M, et al. Molecular forecasting of breast cancer: time to move forward with clinical testing. *J Clin Oncol* 2006;**24**:721–2.
32. Pando MP, Kotraiah V, McGowan K, Bracco L, Einstein R. Alternative isoform discrimination by the next generation of expression profiling microarrays. *Expert Opin Ther Targets* 2006;**10**:613–25.
33. Smid M, Wang Y, Klijn JG, et al. Genes associated with breast cancer metastatic to bone. *J Clin Oncol* 2006;**24**:2261–7.
34. Sauer GW, Schneiderhann-Marra C, Kurzeder C, et al. Predicting nonsentinel node status in breast cancer preoperatively by molecular characterization of core needle biopsies with multiparametric protein analyses. *J Clin Oncol* 2007;**25**:21068.
35. Bild AH, Potti A, Nevins JR. Linking oncogenic pathways with therapeutic opportunities. *Nat Rev Cancer* 2006;**6**:735–41.
36. Schardt JA, Meyer M, Hartmann CH, et al. Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer. *Cancer Cell* 2005;**8**:227–39.
37. Nagrath S, Sequist LV, Maheswaran S, et al. Isolation of rare circulating tumour cells in cancer patients by microchip technology. *Nature* 2007;**450**:1235–9.