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.