Treatment of pT1N0 breast cancer: multigene predictors to assess risk of relapse

D. Fumagalli & C. Sotiriou*

Jules Bordet Institute, Translational Research Unit, Universite Libre de Bruxelles, Bruxelles, Belgium

Breast cancer is a complex disease and even at a favourable stage, such as the pT1N0 one, it is unlikely to be understood and cured by focusing only on single gene or protein alterations determined with suboptimal technologies, as the standard clinico-pathological predictors are. Improving breast cancer treatment will require a more systematic, structured and multidimensional approach able to integrate tumour biology, disease burden and host-related factors. In this scenario, multigene predictors capturing gene-expression profiling or other molecular measurements have great potential for improving breast cancer management. Nevertheless, even if the gene signatures generated so far clearly represent a step forward in the prediction of patient outcome, they are still showing some limitations that nowadays are the basis for the development of a second generation of multigene predictors. Their strength will stand in the investigation of the tumour-surrounding stroma and tumour microenvironment, in the interrogation of different molecular subtypes of breast cancer as distinct entities and in the ability to predict both early and late relapse. It is prospected that the greater accuracy of this new wave of predictors will provide substantial support to the existing decision tools and will significantly ameliorate our current ability to define breast cancer prognosis.

Key words: Breast cancer, molecular signatures, prognosis

background

The clinical and pathological parameters that are currently used to determine the risk of relapse of patients affected by breast cancer, such as tumour size, lymph node involvement, histological grade, oestrogen and ERBB2 receptor status, have shown limited ability to predict distinct patient outcomes and individuals with the same clinical assessment can have markedly different courses [1]. It is likely that considering one or a few predictive markers is not enough to capture the heterogeneity and complexity of breast cancer disease; in contrast, the simultaneous measurement of the expression of thousands of genes or gene products, such as that provided by microarray technologies, should be able to offer a more accurate and precise evaluation of a patient’s clinical course. So far, using these technologies, various prognostic signatures have been produced by our and other groups [2–6]; however, despite their unquestionable performance, some enhancements still need to be made in order to improve their ability to accurately predict patients’ risk of relapse and to become of valuable help in the clinical decision making process.

first-generation prognostic signatures

Our group, in collaboration with the Swiss Institute of Bioinformatics, has recently performed a comprehensive meta-analysis of publicly available gene expression and clinical data from almost 3000 breast tumours [7]; several interesting conclusions were drawn from this collaborative effort:

- All reported prognostics signatures, despite the disparity in their gene lists, carry similar information with regard to prognostication and the disparity of the gene lists produced by several investigators can be attributed to heterogeneity in patient characteristics, expression-profiling methodologies and sampling variation owing to small sample size relative to the number of genes examined;
- All these prognostic signatures are very useful for determining the risk of recurrence in the ER+ subgroup and are less informative for ER− and ERBB2+ diseases, which are assigned to the high-risk category in almost all cases. The power of these signatures resides then in their high accuracy in identifying low-risk patients who could be spared aggressive adjuvant chemotherapy, while the identification of high-risk patients could still be improved;
- Proliferation genes appear to be the common driving force of the prognostic signatures being considered; clinical variables measuring the extent of tumour progression, like nodal status and tumour size, still add independent prognostic information.

From what has been stated so far, it emerges that these first-generation signatures could provide important additional prognostic information for patients with pT1N0 breast disease. These tumours are in fact generally considered at good prognosis, and the genomic signatures could identify among
them a higher number of patients who do not need treatment compared with standard clinical and pathological parameters, thus sparing more women from unnecessary cures and undesirable side-effects [8, 9]. For this reason, two of the guidelines currently used by oncologists [10, 11] recommend the use of genomic tools when these could provide additional information beyond anatomical staging and determination of estrogen receptor (ER)/progesterone receptor (PR) and HER2 status.

The characteristics of the most relevant first-generation gene expression signatures are reported in Table 1. However, although the potential of these predictors is huge, they remain suboptimal for several reasons. First of all, even if they outperform the currently used clinical–pathological criteria in identifying low-risk patients, a level I evidence of their superiority is still missing and it will be available only at the completion of two large prospective clinical trials evaluating two of these genomic signatures [12, 13], namely MammaPrint® and OncotypeDX®. Another important consideration is that although these predictors perform well in identifying early relapses, they fail to predict late relapses. This finding is actually not entirely unexpected, as signatures like MammaPrint® for instance [2] were originally developed to identify patients with distant metastases within 5 years; it also indicates that different molecular mechanisms are likely to be involved during the development of early and late distant metastases. Finally, it is likely that the complexity of breast cancer disease could be better understood not only by focusing on the tumour itself, but also by considering the role of the tumour-surrounding stroma, or host-related factors, such as the immune system.

**second-generation signatures**

Most of the first-generation signatures were developed focusing on the epithelial cancer cells, and the contribution of stromal cells to them is unknown. A recent report has shown that varying proportions of stromal components may influence the accuracy of gene predictors [14] and comprehensive gene expression profiling of each cell type has demonstrated that at the transcriptome level, changes occur in epithelial, myoepithelial and stromal cells that are already evident at the carcinoma in situ stage [15]. Lately, various investigators have tried to derive prognostic signatures from the profile of stromal cells [16–18]. Other equally significant signatures have been derived by analysing the immune system [16, 19, 20] or cancer-related pathways [21–23].

Finak and colleagues [17], for example, have isolated tumour stroma and matched normal stroma from breast tumours and have derived a 26-gene signature strongly associated with outcome called the stroma-derived prognostic predictor (SDPP). SDPP predicted prognosis with greater accuracy than MammaPrint®, stratified several published whole-tumour-derived data sets into clinically meaningful subgroups, and was independent of grade, age, lymph node involvement, chemotherapy, hormonal therapy, and ER and HER2 status. Interestingly, the poor-outcome cluster showed markers of hypoxic and angiogenic responses, whereas the good-outcome cluster overexpressed immune-response genes.

Or again, Loi and collaborators [21], by analysing gene expression and protein data from nearly 1800 human breast cancers, defined a PIK3CA mutation-associated signature (PIK3CA-GS) that was able to predict PIK3CA mutation status in two independent data sets and to identify better clinical outcome in a subset of ER+/HER2– disease. Moreover, in ER+ breast cancer cell lines, PIK3CA-GS seemed to be a better indicator of mTORC1 functional output than mutational status alone, indicating that it was a good predictive marker of PI3K/mTOR inhibitor response.

Table 2 reports and describes other examples of second-generation signatures. This second generation of signatures should try to overcome some other limitations of the first-generation ones. For instance, since it is now clear that breast cancer can be considered as four distinct molecular subgroups, namely basal-like, ERBB2-like, luminal A and luminal B [24], the new molecular predictors should focus on homogeneous classes, as our group has already tried to do [16, 25, 26].

Or again, predictors for metastatic relapse should be designed to predict both early and late relapse; this would be of value especially for tumours with a favorable clinical presentation, such as the pT1N0 ones.

Not less important are the improvements that could be made at the technical level, for both the definition of these signatures and their clinical applicability. Although the cost of conducting microarray experiments is decreasing, the requirement for fresh or snap-frozen tissue may still limit their clinical use. High-throughput, real-time reverse transcriptase–polymerase chain reaction (RT–PCR) can now be performed on sections of...
The second-generation gene expression signatures have been developed to provide even more precise and strong molecular tools. They are based on the literature, genes to act as “prototypes” for different biological processes (ER for ER signaling, HER2 for HER2 signaling, AURKA for proliferation, CASP3 for apoptosis, STAT1 for immune response, VEGF for angiogenesis and in this case PLAU for tumor invasion/metastasis) were selected. A comparison of linear models was then applied to generate modules of genes specifically associated with each of the prototype genes but not with the other prototypes. Genes whose expression varied most between tumor tissue and normal stroma for some tissue-matched pairs were identified and were used to cluster the complete data set of tumor stroma samples. Three clusters with different clinical behavior were generated. Unique genes with the greatest differential expression pattern between the three clusters were selected and a stroma-derived predictor was built after ranking the genes by their independent prognostic ability using a multivariate logistic regression with clinical and biological variables as covariates. A subclass of ER- tumors that over-express immune response genes and that has a good prognosis compared with the rest of ER- breast tumors independently of lymph node status or lymphocytic infiltration was identified. Subsequently, an associated module of complement and immune response genes that define prognostic markers was identified and validated in over 240 ER- samples.

Unsupervised hierarchical clustering of genes in 12 primary invasive breast cancer datasets as well as combined datasets revealed a large cluster of genes with functions in immune cells. Among it, clusters that contained a minimum number of elements and a minimal average correlation were selected and 7 metagenes were derived. Each metagene was then associated with a cell type and/or immunological state. A PIK3CA mutation-associated gene signature was developed starting from probes differentially expressed between PIK3CA mutant and wild type primary ER+/HER2-breast cancers. Gene expression signatures that reflect the activity of a given pathway were identified using supervised classification methods that select set of genes for which the expression levels are mostly highly correlated with the classification of human primary mammary epithelial cell cultures samples into oncogene-activated/deregulated versus control. Two multigene signatures were selected and a stroma-derived predictor was built after ranking the genes by their independent prognostic ability using a multivariate logistic regression with clinical and biological variables as covariates.

All the signatures have been developed starting from fresh frozen tissue using microarray technology. Abbreviations: PLAU, plasminogen activator, urokinase; SDPP, stroma-derived prognostic predictor; STAT1, signal transducer and activator of transcription 1; IR, immune response; IgG, immunoglobulin G; HCK, hemopoietic cell kinase; MHC, major histocompatibility complex; LCK, lymphocyte-specific kinase; PIK3CA-GS, phosphatidylinositol 3-kinase gene signature; IGF-I, insulin-like growth factor 1; MYC, v-myc myelocytomatosis viral oncogene homolog; RAS, Kirsten rat sarcoma viral oncogene homolog; E2F3, E2F transcription factor 3; SRC, Rous sarcoma oncogene; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; BC, breast cancer.

Formalin-fixed paraffin-embedded tissue [3, 27] and it could be the RNA quantitative method of choice in the application of molecular predictors in the clinical setting.

conclusions

Results obtained by our and other groups show that genomic tools have the potential to provide an enormous contribution to prognostication of individual breast cancer patients. The results of two prospective genomic trials will provide level I evidence of the prognostic performance of two multigene predictors [12, 13] and the new wave of signatures is expected to provide even more precise and strong molecular tools. It is likely that in the near future these expression signatures will become part of an integrative decision-making model based on

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>GENE SIGNATURE</th>
<th>SIGNATURE'S DEVELOPMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>STROMA SIGNATURES</td>
<td>PLAU MODULE [16]</td>
<td>Based on the literature, genes to act as “prototypes” for different biological processes (ER for ER signaling, HER2 for HER2 signaling, AURKA for proliferation, CASP3 for apoptosis, STAT1 for immune response, VEGF for angiogenesis and in this case PLAU for tumor invasion/metastasis) were selected. A comparison of linear models was then applied to generate modules of genes specifically associated with each of the prototype genes but not with the other prototypes</td>
</tr>
<tr>
<td></td>
<td>SDPP [17]</td>
<td>Genes whose expression varied most between tumor tissue and normal stroma for some tissue-matched pairs were identified and were used to cluster the complete data set of tumor stroma samples. Three clusters with different clinical behavior were generated. Unique genes with the greatest differential expression pattern between the three clusters were selected and a stroma-derived predictor was built after ranking the genes by their independent prognostic ability using a multivariate logistic regression with clinical and biological variables as covariates.</td>
</tr>
<tr>
<td>IMMUNE-RELATED SIGNATURES</td>
<td>STAT1 MODULE [16]</td>
<td>Based on the literature, genes to act as “prototypes” for different biological processes (ER for ER signaling, HER2 for HER2 signaling, AURKA for proliferation, CASP3 for apoptosis, VEGF for angiogenesis, PLAU for tumor invasion/metastasis and in this case STAT1 for immune-response) were selected. A comparison of linear models was then applied to generate modules of genes specifically associated with each of the prototype genes but not with the other prototypes.</td>
</tr>
<tr>
<td></td>
<td>IR MODULE [19]</td>
<td>A subclass of ER- tumors that over-express immune response genes and that has a good prognosis compared with the rest of ER- breast tumors independently of lymph node status or lymphocytic infiltration was identified. Subsequently, an associated module of complement and immune response genes that define prognostic markers was identified and validated in over 240 ER- samples.</td>
</tr>
<tr>
<td>DCN MODULE [18]</td>
<td>A representative gene for a specific biological process (in this case Decorin for stroma) was selected and included as explanatory variable in a multivariate regression model in order to identify groups of genes related to each process.</td>
<td></td>
</tr>
<tr>
<td>PATHWAYS SIGNATURES</td>
<td>PIK3CA-GS [21]</td>
<td>A PIK3CA mutation-associated gene signature was developed starting from probes</td>
</tr>
<tr>
<td></td>
<td>IGF-I Gene Signature [22]</td>
<td>After stimulation of ER+ and IGF-responsive MCF-7 human breast cancer cells with IGF-I for 3 or 24 hours, an IGF-I gene expression signatures was developed starting from genes that were up or down regulated by IGF-I at both 3 and 24 hours.</td>
</tr>
<tr>
<td></td>
<td>MYC, RAS, E2F3, SRC and β-CATENIN</td>
<td>Gene expression signatures that reflect the activity of a given pathway were identified using supervised classification methods that select set of genes for which the expression levels are mostly highly correlated with the classification of human primary mammary epithelial cell cultures samples into oncogene-activated/deregulated versus control.</td>
</tr>
</tbody>
</table>

Table 2. Second-generation gene expression signatures
multiple-level sources of prognostic data able to support oncologists in the increasingly complex decision making process.

**disclosure**
The authors declare no conflict of interest.

**references**