Search for a gene expression signature of breast cancer local recurrence in young women.

Nicolas Servant, Marc Bollet, Hans Halfwerk, Kevin Bleakley, Bas Kreike, Laurent Jacob, Daoud Sie, Ron Kerkhoven, Philippe Hupé, Rim Hadhri, et al.

Search for a gene-expression signature of breast cancer local recurrence in young women

Nicolas Servant[1,2,3,‡,*], Marc A. Bollet[1,‡,*], Hans Halfwerk[4,5], Kevin Bleakley[1,2,3,6], Bas Kreike[5,7], Laurent Jacob[1,2,3,8], Daoud Sie[4], Ron M. Kerkhoven[4], Philippe Hupé[1,2,3,9], Rim Hadhri[10], Alain Fourquet[1], Harry Bartelink[4], Emmanuel Barillot[1,2,3], Brigitte Sigal-Zafrani[1], Marc J. van de Vijver[4,5]

(1) Institut Curie, Paris, F-75248 France
(2) INSERM, U900, Paris, F-75248 France
(3) Ecole des Mines de Paris, Fontainebleau, F-77300 France
(4) Netherlands Cancer Institute, Amsterdam, The Netherlands
(5) Amsterdam Medical Centrum, Amsterdam, The Netherlands
(6) INRIA, Saclay, France
(7) Institute for radiation oncology Arnhem, Arnhem, The Netherlands
(8) Department of Statistics, University of California, Berkeley, USA
(9) CNRS, UMR 144, Paris, F-75248 France
(10) Hôpital Fattouma Bourguiba, Monastir, Tunisia

‡ These authors contributed equally to this work.

Email : Nicolas.Servant@curie.fr*, marcbollet@gmail.com*, J.B.Halfwerk@amc.uva.nl, kevbleakley@gmail.com, B.Kreike@arnhemrti.nl, laurent.jacob@gmail.com, d.sie@vumc.nl, r.kerkhoven@nki.nl, Philippe.Hupe@curie.fr, rym_hadhri@yahoo.fr, alain.fourquet@curie.net, h.bartelink@nki.nl, Emmanuel.Barillot@curie.fr, brigitte.sigal@curie.net, m.j.vandevijver@amc.uva.nl

* Corresponding authors

Running title: Gene-expression, Breast cancer, Local recurrence

Keywords: breast cancer, local recurrence, gene-expression array

Translational relevance : Several risk factors for developing an ipsilateral breast cancer after breast conserving treatment with combined surgery and radiotherapy have been identified. The fact that (young) age is the single most important risk factor, independent of all others, suggests that some unrevealed biological characteristics of the primary breast cancer may be associated with a higher risk of developing a true local recurrence. We have used gene expression microarray technology to highlight the biological mechanisms involved in the relapse event, and to identify novel risk factors for local recurrence after breast conserving treatment in order to improve the tailoring of the local treatment in breast cancer patients.

Abstract

Purpose: A gene-expression (GE) signature, predictive for local recurrence (LR) after breast-conserving treatment (BCT) has previously been identified from a series of 165 young breast cancer (BC) patients. We evaluated this signature on both another platform and an independent series, compared its performance with other published gene-sets and investigated the GE profile of a larger dataset.

Experimental Design: GE tumor profiles were obtained on 148 of the initial 165 Dutch patients and on an independent validation series of 195 French patients. Both unsupervised and supervised classifications were used to study the GE profile of the 343 BCs, and to identify subgroups that differ for their risk of LR.

Results: The previous LR signature was validated across platforms. However, when applied to the French patients, the signature did not reproduce its reported performance and did not better classify the patients than other published gene-sets. Hierarchical clustering of all 343 BCs did not show any grouping reflecting LR status. Genes related to proliferation were found differentially expressed between patients with or without LR only in triple-negative tumors. Supervised classification revealed no significant gene-set predictive for LR or able to outperform classification based on clinical variables.

Conclusions: Although the previously identified LR signature was robust on another platform, we were neither able to validate it on an independent dataset, nor to define a strong GE classifier for LR using a larger dataset. We conclude that there are no significant differences in GE pattern in tumors from patients with and without LR after BCT.

1 Introduction

Breast-conserving treatment (BCT) is established as the preferred treatment for patients with early-stage breast cancer [1]. It offers both local control and improved survival [2], as well as superior psychosocial outcomes compared to patients treated with modified radical mastectomy [3, 4]. However, the occurrence of an ipsilateral breast cancer can be traumatic and responsible for cancer-specific death [2]. Several risk factors for developing an ipsilateral breast cancer after breast conserving treatment with combined surgery and radiotherapy have been identified. The fact that (young) age is the single most important risk factor, independent of all others [5], suggests that some unrevealed biological characteristics of the primary breast cancer may be associated with a higher risk of developing a true local recurrence (LR), i.e., regrowth of clonogenic cells that were not removed by surgery or killed by radiotherapy [6]. Gene expression (GE) profiling of primary breast carcinomas using microarray technology has already been used in order to try and discover a signature associated with a higher risk of developing true recurrence after BCT [7-10]. Results have been contradictory. The most recent and largest study, by Kreike et al. [10], is for a series of 165 young (≤50 years old) premenopausal Dutch patients using the Human Genome Oligo Set version 3.0 array (Operon), and did not confirm previous classifiers, namely those based on the wound response signature [8] and that defined by Nimeus et al. [9]. Kreike et al. were able to construct a LR classifier based on the expression of 111 genes and validate this profile on an independent data set of 161 consecutive breast cancer patients treated by BCT that used a different microarray platform [10]. This classifier was mostly characterized by proliferation but did not bring a significant independent additive value, as young age remained the sole predictive factor of LR in multivariate analysis.

2 Materials and Methods

2.1 Selection of patients

This study reports on a series of 343 patients with primary breast cancers selected from the fresh-frozen–tissue bank of 4 Dutch hospitals (Netherlands Cancer Institute, Amsterdam; Radboud University Nijmegen Medical Centre, Nijmegen; Erasmus Medical Center, Rotterdam; and Ziekenhuis Amstelland, Amstelveen) and one French cancer center (Institut Curie, Paris). The 148 Dutch patients belonged to a larger series of 165 breast cancer patients already reported [10], were combined with the 195 French patients. Patients were included in the study if they met the following criteria: invasive primary breast carcinoma diagnosed at or before the age of 50; all patients were pre-menopausal; no previous history of cancer (except for one non-melanoma skin cancer). All patients were treated between January 1984 and November 2002 by initial breast-conserving surgery (216 pT1, 124 pT2, 1pT3, 2 missing data), including either axillary lymph node dissection or sentinel node procedures (215 pN0, 99 pN1, 21 pN2, 3pN3, 6 missing data). All patients underwent postoperative radiotherapy to the breast (median dose of 50 Gy, range 45-55) with, in 248 patients, a boost to the tumor bed (external beam radiotherapy or brachytherapy with a median dose of 15 Gy, range 6-26) and/or the regional lymph node bearing areas. 141 patients received adjuvant systemic treatment with either chemotherapy only (110 patients), hormone-therapy only (8 patients) or a combination of both (23 patients). Chemotherapy regimens (4 missing data) were either anthracycline-based (79 patients) or contained cyclophosphamide, methotrexate and fluorouracile (CMF, 50 patients). Patients were selected if they had either experienced a local recurrence (ipsilateral breast recurrence with clinical, histological and/or genomic features consistent with true recurrences) within the first 10 years after primary treatment, or remained free from local recurrence (any type of ipsilateral breast cancer) for at least 10 years after primary treatment. In order to ensure relevant data, we performed gene-expression analyses in tumors with at least 50% of cancer cells as assessed by Hematoxylin, and Eosin staining of histological sections of the snap-frozen samples.

This research was approved by the institutional review boards of both the Netherlands Cancer Institute and the Institut Curie.

2.2 Clinical and histological studies

Clinical and tumor characteristics of the 343 patients are reported in Table 1. Three pathologists (M.J. v.d.V, B. S-Z, R.H.) reviewed histopathologic characteristics of all tumors. Immunohistochemical analysis was used to determine the status of Estrogen receptor (ER, clone 6F11, 1:200 dilution; Novocastra, Newcastle Upon Tyne, England) and progesterone receptor (PR, clone 1A6, 1:200 dilution; Novocastra). Tumors were considered to be positive for these receptors if at least 10% of the invasive tumor cells in a section showed nuclear staining [11]. HER2 over-expression was determined according to American Society of Clinical Oncology (ASCO) guidelines.

2.3 Isolation of RNA and gene expression profiling

Tumor material was snap-frozen in liquid nitrogen within one hour of surgery. Thirty 30-µm sections were used for isolation of RNA. The first and last sections (5 µm) were stained with hematoxylin and eosin; only samples containing an average of at least 50% tumor cells were used in this analysis. Two protocols of RNA isolation were used. For Dutch samples, total RNA was isolated with RNA-Bee (Tel-Test, Friendswood, United States) and dissolved in RNase-free water. Then, total RNA was treated with DNase with use of the Qiagen® RNase-free DNase kit and RNeasy spin columns and dissolved in RNase-free water. Detailed information on protocols can be found at the Central Microarray Facility website of the Netherlands Cancer Institute (http://microarrays.nki.nl/). For French samples, RNA was extracted from frozen samples using the Trizol method (Invitrogen) according to the manufacturer's instructions and purified using mirRNeasy kit (Qiagen). The concentration, integrity and purity of each RNA sample (260/280, 260/230, 28S/18S, RIN) were measured using the RNA 6000 LabChip kit with the Agilent 2100 bioanalyser (Agilent technologies, Palo Alto, CA). The gene-expression microarrays used in this study were Illumina Human Whole Genome V3.0 arrays (Illumina, Inc. San Diego, CA) containing 48803 reporters. Details of RNA amplification, labeling and hybridization are available on the Illumina website (http://www.illumina.com). The arrays were processed in the Central Microarray Facility of the Netherlands Cancer Institute and are publicly available at GEO, series record GSE30682 (http://www.ncbi.nlm.nih.gov/geo/).

2.4 Cross-platform validation of the local recurrence signature

The population of 148 Dutch patients was first studied to validate the signature developed by Kreike et al. [10] and to assess cross-platform reproducibility. For these 148 patients, both

Operon (Human Genome Oligo Set version 3.0) and Illumina (Human Whole Genome V3.0) data were available. A Multiple Factor Analysis (MFA [12]) was applied to the 111-gene signature reported by Kreike et al. using the R-package FactoMineR (v1.14) [13]. MFA aims to integrate the two groups of variables (111 genes expression profiles from both Operon and Illumina data) which are expected to describe the same observations (148 tumor samples). The goal is to check whether both datasets share a common geometric structure, thus indicating that the same underlying biological signal can be extracted from the same samples and the same set of genes, with different technology. The LR status was added as an illustrative variable.

2.5 Validation of previous local recurrence signatures

The previous local recurrence signature developed at the NKI was tested in the French population along with 21 other signatures: mutation status [14], chromosomal instability [15], fibroblast serum response [16], wound serum response [17], hypoxia signature [18], radio-resistance signature [19], invasiveness signature [20], local recurrence signature [9], 70 gene NKI signature [21], P53 signature [22], recurrence score [23], Rotterdam signature [24], PTEN-loss signature [25], genomic grade [26], proliferation signature [27], molecular portraits [28], prognostic gene-expression classifier for ER positive breast cancer [29], pooled signature [30], markers of proliferation [31], three modules [32] and the wound signature local relapse signature [8].

A Diagonal Linear Discriminant Analysis (DLDA) was applied to classify the samples and the performances were evaluated using Monte Carlo cross-validation (10 fold - 100 permutations) [33]. The area under the Receiver Operating Characteristic (ROC) curve (AUC) was used as the prediction-quality criterion.

2.6 Data analysis of 343 primary tumors

The 343 arrays were normalized using the variance stabilization transformation and the robust spline algorithm suggested by Du et al. (lumi R-package v2.2.0) [34]. Probes were then filtered to discard those called as unexpressed (or undetectable) in all samples. This left 30198 probes to be used for analyses. To ensure comparability between the Dutch and French samples, the dataset was corrected for a population effect using the following one-way ANOVA model applied to each probe *i, i* $\in$ [1, 30198]:

$$Y_{ij} = \mu_i + O_p + \varepsilon_{ijp} \quad (1)$$

where:

Y_{ij} is the expression of probe *i* in sample $j \in [1, 343]$,

O_p is the effect of the population, $p \in \{\text{Dutch, French}\}$,

μ_i is the expression of gene *i* when *p* = Dutch,

$\varepsilon_{ijp} \sim N(0, \sigma_i^2)$ is a residual term.

The corrected expression level of probe *i* was computed as follows:

$$Y^{c}_{ij} = Y_{ij} - O_{French}$$

Ward-linkage hierarchical clustering of Pearson correlation similarity matrices was then applied on the 5000 most variant probes (highest interquartile range) of the 343 samples, using the R-package EMA (v1.2) [35].

Based on the expression level of genes ESR1, PR, ERBB2 and AURKA, each sample was assigned to a molecular subtype. All tumors were assessed using the average expression ratio of the probes for each gene, respectively ESR1 (1731003), PGR (1710082), ERBB2 (1729826/1753944) and the AURKA (1710733/1755528) proliferation marker. ROC curves were computed for all tumors with known immunohistochemical data. The microarray-derived ER, PR and HER2 statuses were determined from these tumors according to the optimal sensitivity-specificity cut-off (Figure 1) [36] and then applied to all tumors without immunohistochemical data. Tumors were grouped into molecular subtypes according to their gene-expression, as defined by Voduc et al. [37]: luminal A (ER+ and/or PR+, HER2-, low AURKA), luminal B (ER+ and/or PR+, HER2-, high AURKA), luminHER (ER+ and/or PR+, HER2+), triple-negative (ER-, PR-, HER2-) and HER2 enriched (ER-, PR-, HER2+).

In addition, the single sample predictor (SSP) provided by Hu et al. [28] was mapped onto our microarray platform and tested for class assignment. In order to do this, the expression data were centered on genes and the Pearson correlation of each sample with the centroids of the 5 molecular subtypes (luminal A, luminal B, HER2+, Basal-like, Normal-like) was calculated. Each sample was assigned to a subtype based on the largest correlation. Samples with largest correlation less than 0.1 were considered unclassified.

A one-way ANOVA model, taking into account the subtype information and its interaction with relapse, was applied to all probes *i*, $i \in [1, 30308]$ in order to identify differentially expressed genes (DEGs) between LR and no-LR:

$$Y^{c}_{ij} = \mu_i + T_t + LR_r + (T*LR)_{tr} + \varepsilon_{ijtr} \quad (2)\ ,$$

where:

Y^{c}_{ij} is the expression of gene *i* in sample $j \in [1, 343]$ corrected from the population effect (1),

T_t is the effect of the molecular subtype, $t \in$ {HER2+, luminal A, luminal B, luminal-HER, triple-negative},

LR_r is the effect of local recurrence, $r \in \{0, 1\}$,

$(T*LR)_{tr}$ is the interaction term between molecular subtype and recurrence event,

μ_i is the expression of gene i when t = HER2+ and $r = 0$,

$\varepsilon_{ijtr} \sim N(0, \sigma_i^2)$ is a residual term.

The analysis was performed using the R-package limma (v3.6.9) and the *p*-values were adjusted for multiple testing using the Benjamini-Hochberg correction. All adjusted *p*-values smaller than 5% were considered significant.

Over-representation of Gene Ontology categories in the list of DEGs was identified using a hypergeometric test (R-package GOstats v2.16.0). Categories represented by less than 10 genes were discarded.

2.7 Pairs of primary breast carcinoma and ipsilateral breast cancers

The distinction of secondary primary cancers and true recurrences is an important issue, which can be used to validate patient selection. In the previous study, Kreike et al. have already assessed this question on the Dutch samples, using the overall gene expression pattern of 15 pairs of primary tumors and LRs. Regarding the French series, we analyzed 14 pairs of tumors using DNA copy number alterations (Affymetrix GeneChips Human Mapping 50K array Xba) profiles and the partial identity score proposed by Bollet, Servant et al. [38]. Briefly, this score aims to estimate the clonal relatedness of the ipsilateral breast cancer and the primary tumor using the number of common DNA breakpoints among both profiles, weighted by their occurrence frequency in a control population.

2.8 Supervised classification and variable selection

After appropriate scaling of the set of clinical variables (see Table 1) and expression data, we used Monte Carlo cross-validation (5 folds - 100 permutations) and performed binary classification using the SVM algorithm [39] in order to predict outcome (LR or no LR). A radial basis function (rbf) kernel was used, with variance fixed at 10 for the clinical data and 1000 for expression data; internal cross-validation was performed inside each fold in order to select the SVM penalty parameter λ. We applied the algorithm separately for clinical and expression data, as well as combining the two using either the mean or the product of the individual rbf kernels to create a new valid kernel [39]. In experiments, we calculated the balanced accuracy, i.e., ½ (sensitivity + specificity). Analyses were first performed on all

patients, then for patients with each clinical subtype in order to see if this led to improvement in accuracy for certain subtype.

Furthermore, we applied the Least Angle Regression (LARS) algorithm [40], which performs both supervised classification and variable selection, and can help to suggest a ranking of important predictive variables.

3 Results

We investigated the relationship between LR and gene expression profiles of primary breast cancers of patients treated with BCT in a series of 343 patients. Of these, 119 patients had LR (median time-lapse of 3.3 years, range 0.3 -9.4 years) and 224 patients remained free from LR for at least 10 years.

3.1 Clinical and histological features

The clinical and pathological characteristics of patients and tumors are displayed in Table 1. Patients who developed LRs were significantly younger at diagnosis than those who did not (χ^2 test p=0.044). Their tumors had significantly higher histological grade according to Ellis and Elston (χ^2 test, p=0.012), higher mitotic indexes (ordinal χ^2 test, p=0.011) and more frequently contained an extensive DCIS component (χ^2 test, p=0.041).

3.2 Gene-expression studies

3.2.1 *ER, PR, HER2 and proliferation status from gene expression*

The ER, PR, HER2 and proliferation statuses were determined from the gene expression data according to the optimal sensitivity-specificity cut-off from the ROC curves produced using IHC data (Figure 1). These statuses were respectively defined from the probes ESR1 (1731003, cut-off=8.687), PR (1710082, cut-off=7.405), ERBB2 (1729826/1753944, cut-off=9.481) and AURKA (1710733/1755528, cut-off=8.622).

3.2.2 *Cross-platform comparison*

Of the complete dataset of 343 patients, 148 were also used by Kreike et al. to define the 111-gene signature, and studied using the Operon microarray platform.

MFA analysis (Figure 2) aims to integrate the two groups of variables (111 genes expression profiles from both Operon and Illumina data) which are expected to describe the same observations (148 tumor samples) in order to assess the robustness of this signature on

another microarray platform, as well as its association with LR. The two groups have coordinates that are close in the first dimension, which means that their contribution to the first principal component is similar. It also means that the first principal component of the MFA (the strongest signal) is common to the two datasets. The LR event is well represented on the first component of the MFA. This observation is consistent with the performance of the 111 gene signature previously described.

3.2.3 *Validation of the previous local recurrence signature*

Another way to assess the predictive power of the signature from Kreike et al. is to consider the French samples as an independent validation series. The class prediction (DLDA) on this population showed an AUC of 0.56, a specificity of 0.53 for a sensitivity of 0.63, and a misclassification rate of 43%. These results are worse than the initial performances published by Kreike et al. (sensitivity 0.77 and specificity 0.43). We assessed how other GE signatures performed in pairing tumors to LR status. The 111-gene classifier did not show convincingly better performance on the French samples. None of the tested signatures were able to classify the samples with an average balanced accuracy higher than 0.585 (Table S1).

3.2.4 *Overall gene expression profiles of 343 primary tumors*

Hierarchical clustering of the 343 tumors and the 5000 most variant probes resulted in the formation of two main clusters that could be further subdivided into four sub-clusters (Figure 3). None of the tumor characteristic related to LR was associated with these clusters.

These four clusters were closely related to the subtype definitions. We clearly identified the triple-negative tumors and two other clusters mainly composed of luminal A and luminal B-like tumors. The luminal A, luminal B and the triple-negative clusters are clearly identified with respectively 72%, 76% and 96% (Table S2, hyper-geometric tests, p=2.86e-36, p=1.60e-18 and p=1.46e-53) of the tumors in the three groups. The last cluster is enriched in HER2 profiles and can be further divided in the HER2+ profiles and a mixture between luminal-HER2 and luminal B profiles. The class assignments predicted from the Hu et al. [28] SSP do not outperform the clustering based on the ESR1, PR, HER2, and proliferation subtype prediction (Supplementary Figure 1).

Information on breast cancer subtypes is crucial and has to be taken into account in any supervised analysis in order to avoid confounding effects. The association between breast cancer subtype and relapse events (Table S3, χ^2 test, p = 0.033) has already been reported in several previous studies [10, 37]. In order to identify DEGs between patients developing a LR

and those that do not, we used the linear model defined in (2). *P*-values lower than 5% after Benjamini-Hochberg correction were reported as significant.

No DEGs in the groups luminal A, luminal B and luminal-HER were found. A list of 182 probes was generated for the triple-negative tumors and a list of 115 probes for the HER2+ tumors. When the subtype factor was removed from the linear model, no genes were found to be significant.

To identify if these two lists of probes were enriched for genes representing particular Gene Ontology categories compared with the entire list of probes, a hyper-geometric test was used and the Gene Ontology categories with *p*-values lower than 0.001 were reported. Over-representation of genes related to proliferation, immune response and cell death were found in the triple-negative differentially expressed probes (DEGs). No interesting category was found in the list of DEGs in the HER2 over-expressed population (Table S4).

3.3 Partial Identity Score of primary tumors and LRs

We applied the partial identity score proposed by Bollet, Servant et al. [38] on 14 pairs of tumors. Among the 14 pairs, all but one had a significant score and thus were classified as a true recurrence (Figure 4). This indicates that the DNA copy number profile of the LRs and the primary tumors are highly similar. It also validates the inclusion of the French samples and the relevance of their recurrence status.

3.4 Supervised classification and variable selection

Using the SVM algorithm with an rbf kernel, we obtained an average balanced accuracy of 62±2% using clinical variables and 59±2% using expression data. Using a mean kernel to combine the two sets of variables gave 64±2% average balanced accuracy, whereas a product kernel gave 63±2%. This indicates that combining the two types of variables in this way may give a small prediction accuracy improvement compared with treating them separately.

The same analyses with the SVM algorithm and rbf kernel were performed by subtype (Table 2). There were too few HER2 patients to give meaningful results. Balanced accuracy results did not significantly improve with respect to treating all subtypes together, and in some cases, decreased.

Using LARS on the clinical data with 5-fold cross-validation and averaged over 1000 trials gave a maximum best balanced accuracy of 63% obtained when including 32 variables in the predictor. Running LARS on all clinical data for variable selection purposes, we obtained a

list of pertinent variables along with their weight in the prediction function at the maximum where 32 variables had been included. The first 10 selected variables are shown in Table 3.
Including the 5 subtypes as additional binary variables did not change the above ranking. Using LARS on the expression data with 5-fold cross-validation and averaged over 100 trials gave a maximum balanced accuracy of 58% obtained when including 99 variables in the predictor. Using LARS on the combined clinical and expression data with 5-fold cross-validation and averaged over 100 trials gave a maximum balanced accuracy of 61% obtained when including 73 variables in the predictor.

4 Discussion

We have performed gene expression profiling of 343 primary breast cancers that were treated with BCT. We chose to base our study on a series of young (≤50 years old), pre-menopausal women, not only because (young) age is recognized as one of the most important independent prognostic factors for ipsilateral breast recurrence [5, 41-46], but also in order to ensure a high level of homogeneity. Additionally, all patients had undergone breast conserving surgery followed by whole-breast radiotherapy for their initial breast cancers.

In a previous study, Kreike et al. [10] found that there were no significant differences in the overall gene expression profiles between primary tumors with or without LR. With an increased number of patients and the use of a new microarrays platform (Operon), they were able to construct and to validate a predictive classifier for LR with a specificity of 43% and sensibility of 77%. We have significantly extended the number of patients to include 343 primary breast cancers, using a different microarray platform (Illumina). From these 343 samples, 148 were common with the previous study of Kreike et al. Among the 165 new French samples, the single-nucleotide polymorphism profiles of 14 pairs of primary tumors and ipsilateral breast cancers were analyzed. The relapse status of all but one patient was validated.

The first part of this study consisted in confirming the previous 111-genes signature. As expected, the MFA analysis on the 148 Dutch samples common to both studies and platforms (Operon and Illumina) on the 111 genes resulted in good correlation between the two datasets, and further, shows an association between LR and the expression profiles of these 111 genes. However, when we applied this signature to an independent dataset (French samples), we were not able to reproduce the reported performance, finding a sensitivity of 63% and a specificity of 53%. We also assessed a large number of alternative gene expression signatures and found that the 111-gene classifier did not show convincingly better performance than the others. With a large number of false positives (specificity=53%) and many false negatives (sensitivity=63%), this signature has limited value in the prediction of LR.

To address this shortcoming, we investigated the gene expression profile of a significantly larger set of 343 tumors. Hierarchical clustering analysis of the 343 tumors showed clusters of specific breast cancer subtypes. The triple-negative, luminal A and luminal B samples were well separated. The HER2+ samples were also well separated from the luminal-HER2 which are grouped with some luminal B samples. The subtype prediction is a potential confounding

factor that may have strong impact on the results of the study. Recently, Weigelt et al. [47] discussed the robustness of several microarray-based SSP and the need to establish a standard methodology for breast cancer subtype definition. Here, the combination of IHC status and gene-expression data offers effective subtype discrimination. Moreover, the IHC status gives robust information that is mainly used in routine clinical practice.

We did not observe significant grouping of the tumors with respect to their relapse status. Previous studies [10, 37] have already described the association between molecular breast cancer subtypes and LR. This observation is critical for further analysis, and particularly for the search of DEGs. Based on the linear model (2) which takes into account the molecular subtype and its association with LR, we found significant DEGs only for the HER2+ and triple-negative subtypes. The DEGs found in the triple-negative cancers were associated with Gene Ontology categories such as proliferation, cell death and immune response. The enrichment in genes involved in proliferation was also reported by Kreike et al. [10] and agrees with previously published classifiers. The fact that no interesting categories were found in the HER2+ over-expressed subtype might be explained by the small number of patients in this class (8 LR and 6 no-LR). Further investigations and a larger number of patients with HER2+ tumors are necessary to optimize the search for DEGs in this tumor subtype.

More surprisingly, no DEGs were found if we did not take into account subtype information. This observation is remarkable and contradicts the previous results published by Kreike et al. [10]. The two different microarrays platforms, the heterogeneity of the data and the different subtype proportions might have some influence on this finding.

Based on the 343 tumors profiles, we attempted to define a new classifier using clinical data and microarray gene expression profiling. Combining both variable types gave a small balanced accuracy improvement compared with treating them separately.

We performed supervised classification (SVM and LARS) on both the clinical and expression data, in order to create classifiers to predict LR or no LR. Using SVM, balanced accuracy of between 60-64% was achieved and remains unsatisfactorily low. Separation of the data into subtypes gave similar balanced accuracy, though in some case worse. This may be partially due to a loss of predictive power in smaller data sets when using supervised classification.

Using LARS on the clinical data confirmed that the total dose of radiotherapy to the tumor bed and age at diagnosis were the two most important prognostic factors associated with decreased risk of local relapse [5]. The third and fourth factors were the surgical margin involvement by invasive component and the tumor proliferation, as already described by Kreike et al. [10, 48]. Adding subtype information to the clinical variables did not improve

significantly the algorithm's performances. Using LARS on expression data or on combined clinical and expression data did not give improved classification performance.

In conclusion, although we were unable to validate the 111 gene classifier proposed by Kreike et al. [10] and to define a strong classifier for local relapse after BCT from this large and well-characterized dataset, it cannot be concluded that such a classifier does not exist. New studies, such as one that is currently under progress at the Netherlands Cancer Institute, the Institut Gustave Roussy and the Karolinska Institutet on preoperative radiotherapy (« image-guided preoperative accelerated partial breast irradiation (PAPBI) : defining radiotherapy sensitivity » RCB 2010-A00573-36) should help in finding new factors associated with radiosensitivity in breast cancer. As previously reported, our study also demonstrated that proliferation, age and tumor invasion are important criteria in predicting LR. Lastly, this study also highlights the difficulty of reproducing and validating gene expression signatures between competing microarray platforms and from one population to another.

5 Acknowledgments

We thank F. Reyal for helpful discussion about this study and the manuscript.

6 Funding and conflict of interest

This work was financed by the Institut Curie, the 'Courir pour la vie, Courir pour Curie' association, the 'Odyssea' association, and PHRC 2006. No conflict of interest is declared by any of the authors.

LR: Local relapse, pT: classification of the pathological tumor size according to AJCC, DCIS: ductal carcinoma in-situ, pN: regional nodal stage according to AJCC, CT: chemotherapy, HT: hormone-therapy, LCIS: lobular carcinoma in-situ, LVI: lobular vascular in-situ, BC: breast cancer, ER: estrogen receptor, GE: gene expression, PR: progesterone receptor, HR: hormone receptor, HER2: Human Epidermal Receptor 2, TNBC: triple-negative (ER-, PR-, HER2-) breast cancer, DEGs: differentially expressed genes.

References

1. Temple WJ, Russell ML, Parsons LL, Huber SM, Jones CA, Bankes J, et al. Conservation surgery for breast cancer as the preferred choice: a prospective analysis. J Clin Oncol 2006; 24:3367-3373.
2. Clarke M, Collins R, Darby S, Davies C, Elphinstone P, Evans E, et al. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. Lancet 2005; 366:2087-2106.
3. Engel J, Kerr J, Schlesinger-Raab A, Sauer H, and Hölzel D. Quality of life following breast-conserving therapy or mastectomy: results of a 5-year prospective study. Breast J 2004; 10:223-231.
4. Moyer A. Psychosocial outcomes of breast-conserving surgery versus mastectomy: a meta-analytic review. Health Psychol 1997; 16:284-298.
5. Vrieling C, Collette L, Fourquet A, Hoogenraad WJ, Horiot JC, Jager JJ, et al. Can patient-, treatment- and pathology-related characteristics explain the high local recurrence rate following breast-conserving therapy in young patients? Eur J Cancer 2003; 39:932-944.
6. Haffty BG, Carter D, Flynn SD, Fischer DB, Brash DE, Simons J, et al. Local recurrence versus new primary: clinical analysis of 82 breast relapses and potential applications for genetic fingerprinting. Int J Radiat Oncol Biol Phys 1993; 27: 575-583.
7. Kreike B, Halfwerk H, Kristel P, Glas A, Peterse H, Bartelink H, et al. Gene expression profiles of primary breast carcinomas from patients at high risk for local recurrence after breast-conserving therapy. Clin Cancer Res 2006; 12:5705-5712.
8. Nuyten DSA, Kreike B, Hart AAM, Chi JA, Sneddon JB, Wessels LFA, et al. Predicting a local recurrence after breast-conserving therapy by gene expression profiling. Breast Cancer Res 2006; 8:R62.
9. Niméus-Malmström E, Krogh M, Malmström P, Strand C, Fredriksson I, Karlsson P, et al. Gene expression profiling in primary breast cancer distinguishes patients developing local recurrence after breast-conservation surgery, with or without postoperative radiotherapy. Breast Cancer Res 2008; 10:R34.
10. Kreike B, Halfwerk H, Armstrong N, Bult P, Foekens JA, Veltkamp SC, et al. Local recurrence after breast-conserving therapy in relation to gene expression patterns in a large series of patients. Clin Cancer Res 2009 ; 15:4181-4190.
11. Balaton AL, Coindre JM, Collin F, Ettore F, Fiche M, Jacquemier J, et al. [Recommendations for the immunohistochemical evaluation of hormone receptors on paraffin sections of breast cancer. Study Group on Hormone Receptors using Immunohistochemistry FNCLCC/AFAQAP. National Federation of Centres to Combat Cancer/French Association for Quality Assurance in Pathology]. Ann Pathol 1996; 16:144-148.
12. Escofier B and Pagès J. Multiple factor analysis. Computational Statistics and Data Analysis 1990; 18:121-40.
13. Lê HF. FactoMineR: an R package for multivariate analysis. Journal of statistical software 2008; 25:1-18.
14. Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, et al. The genomic landscapes of human breast and colorectal cancers. Science 2007; 318: 1108-1113.

15. Carter SL, Eklund AC, Kohane IS, Harris LN and Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat Genet 2006; 38:1043-1048.
16. Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS Biol 2004; 2:E7.
17. Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sorlie T, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci USA 2005; 102:3738-3743.
18. Chi JT, Wang Z, Nuyten DS, Rodriguez EH, Schaner ME, Salim A, et al. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. PLoS Med 2006; 3:e47.
19. Khodarev NN, Beckett M, Labay E, Darga T, Roizman B and Weichselbaum RR. STAT1 is overexpressed in tumors selected for radioresistance and confers protection from radiation in transduced sensitive cells. Proc Natl Acad Sci USA 2004; 101:1714-1719.
20. Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. N Engl J Med 2007; 356:217-226.
21. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415:530-536.
22. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, 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; 102:13550-13555.
23. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004; 351:2817-2826.
24. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005; 365:671-679.
25. Saal LH, Johansson P, Holm K, Gruvberger-Saal SK, She QB, Maurer M, et al. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. Proc Natl Acad Sci USA 2007; 104:7564-7569.
26. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 2006; 98:262-272.
27. Starmans MH, Krishnapuram B, Steck H, Horlings H, Nuyten DS, van de Vijver MJ, et al. Robust prognostic value of a knowledge-based proliferation signature across large patient microarray studies spanning different cancer types. Br J Cancer 2008; 99:1884-1890.
28. Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 2006; 7:96.
29. Teschendorff AE, Naderi A, Barbosa-Morais NL, Pinder SE, Ellis IO, Aparicio S, et al. A consensus prognostic gene expression classifier for ER positive breast cancer. Genome Biol 2006; 7:R101.
30. van Vliet MH, Reyal F, Horlings HM, van de Vijver MJ, Reinders MJ, and Wessels LFA. Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability. BMC Genomics 2008; 9:375.
31. Whitfield ML, George LK, Grant GD and Perou CM. Common markers of proliferation. Nat Rev Cancer 2006; 6:99-106.

32. Wirapati P, Sotiriou C, Kunkel S, Farmer P, Pradervand S, Haibe-Kains B, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res 2008; 10:R65.
33. Molinaro AM, Simon R and Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. Bioinformatics 2005, 21, 3301-3307
34. Du P, Kibbe WA and Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008; 24:1547-1548.
35. Servant N, Gravier E, Gestraud P, Laurent C, Paccard C, Biton A, et al. EMA - A R package for Easy Microarray data analysis. BMC Res Notes 2010; 3:277.
36. Youden WJ. Index for rating diagnostic tests. Cancer 1950; 3:32-35.
37. Voduc KD, Cheang MCU, Tyldesley S, Gelmon K, Nielsen TO and Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. J Clin Oncol 2010; 28:1684-1691.
38. Bollet MA, Servant N, Neuvial P, Decraene C, Lebigot I, Meyniel JP, et al. High-Resolution Mapping of DNA Breakpoints to Define True Recurrences Among Ipsilateral Breast Cancers. J Natl Cancer Inst 2008;100:48–58
39. Schölkopf B and Smola A. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press; 2002
40. Efron B, Hastie T, Johnstone I and Tibshirani R. Least Angle Regression. Annals of Statistics. Annals of Statistics 2004; 32:407-499.
41. Fourquet A, Campana F, Zafrani B, Mosseri V, Vielh P, Durand JC, et al. Prognostic factors of breast recurrence in the conservative management of early breast cancer: a 25-year follow-up. Int J Radiat Oncol Biol Phys 1989; 17:719-725.
42. Borger J, Kemperman H, Hart A, Peterse H,van Dongen J, Bartelink H. Risk factors in breast-conservation therapy. J Clin Oncol 1994; 12:653-660.
43. Elkhuizen PH, van de Vijver MJ, Hermans J, Zonderland HM, van de Velde CJ, Leer JW. Local recurrence after breast-conserving therapy for invasive breast cancer: high incidence in young patients and association with poor survival. Int J Radiat Oncol Biol Phys 1998; 40:859-867.
44. Elkhuizen PH, Voogd AC, van den Broek LC, Tan IT, van houwelingen HC, Leer JW, et al. Risk factors for local recurrence after breast-conserving therapy for invasive carcinomas: a case-control study of histological factors and alterations in oncogene expression. Int J Radiat Oncol Biol Phys 1999; 45:73-83.
45. Oh JL, Bonnen M, Outlaw ED, Schechter NR, Perkins GH, Strom EA, et al. The impact of young age on locoregional recurrence after doxorubicin-based breast conservation therapy in patients 40 years old or younger: How young is "young"? Int J Radiat Oncol Biol Phys 2006; 65:1345-1352.
46. Bollet MA, Sigal-Zafrani B, Mazeau V, Savignoni A, de la Rochefordiere A, Vincent-Salomon A, et al. Age remains the first prognostic factor for loco-regional breast cancer recurrence in young (<40 years) women treated with breast conserving surgery first. Radiother Oncol 2007; 82:272-280.
47. Weigelt B, Mackay A, A'hern R, Natrajan R, Tan DSP, Dowsett M, et al. Breast cancer molecular profiling with single sample predictors : a retrospective analysis. Lancet Oncol 2010; 11: 339-49
48. Kreike B, Hart AA, van de Velde T, Borger J, Peterse H, Rutgers E, et al. Continuing risk of ipsilateral breast relapse after breast-conserving therapy at long-term follow-up. Int J Radiat Oncol Biol Phys 2008; 71:1014-1021

Tables and Figures

Table 1: Patient and tumor characteristics with respect to the occurrence of local relapse. *Significant (p<0.05) χ^2 tests, for age: p=0.044, histological grade: p=0.012, extensive DCIS outside: p=0.047, extensive DCIS: p=0.041. Significant (p<0.05) ordinal χ^2 test, for mitotic index: p=0.011.

Table 2: Supervised classification (SVM algorithm [39]) by subtype to predict outcome (LR or no LR). Monte Carlo cross-validation (5 folds - 100 permutations) was used to assess performances. The rbf kernel had variance of 10 for clinical and 1000 for expression data. Balanced accuracy is reported (± s.d.) Not enough HER2+ patients were available to perform internal cross-validation step.

Table 3: Top 10 variables selected by LARS. The respective weights are specified for the 10 first variables.

Figure 1: Definition of molecular subtype based on the expression level of genes ESR1, PR, ERBB2 and AURKA. A. ROC curves were computed for all tumors with known immunohistochemical data. The microarray-derived ER, PR and HER2 statuses were determined according to the optimal sensitivity-specificity cut-off and applied to the 343 samples. sens: sensibility, spe: specificity, n: number of tumors with known immunohistochemical data, T: optimal sensitivity-specificity cut-off **B.** Expression level of ESR1, PR, ERBB2 and AURKA for each defined immunohistological class. The 343 tumors were then grouped into molecular subtypes according to their gene-expression, as defined by Voduc et al. [37].

Figure 2: Multiple Factor Analysis. The partial axis representation shows the link between the two first principal components of the MFA and those of each individual group (Illumina and Operon). Local relapse information is added as an illustrative variable.

Figure 3: Hierarchical clustering of the 343 primary tumors. The samples can be divided into four groups according to their subtype. We clearly identify the triple-negative tumors. Two others clusters are mainly composed of luminal A and luminal B tumors. Lastly, the third group is composed of most of the HER2+ profiles. Group composition is given in more detail in Table S2. LumA: Luminal A, LumB: Luminal B, LumHER: Luminal-HER, TNBC: Triple-negative breast cancer.

Figure 4: Partial identity score of 14 pairs of tumors. Primary tumors and their associated ipsilateral breast cancer were analyzed using the partial identity score proposed by Bollet, Servant et al. [38]. The histogram was constructed using pairs from different patients. The vertical dashed line represents the upper 5th percentile of the null hypothesis and the threshold above which true recurrences were defined. Each dot represents one of the 14 natural pairs.