

# Clinical Cancer Research



## Search for a Gene Expression Signature of Breast Cancer Local Recurrence in Young Women

Nicolas Servant, Marc A. Bollet, Hans Halfwerk, et al.

*Clin Cancer Res* 2012;18:1704-1715. Published OnlineFirst January 23, 2012.

**Updated version** Access the most recent version of this article at:  
[doi:10.1158/1078-0432.CCR-11-1954](https://doi.org/10.1158/1078-0432.CCR-11-1954)

**Supplementary Material** Access the most recent supplemental material at:  
<http://clincancerres.aacrjournals.org/content/suppl/2012/01/23/1078-0432.CCR-11-1954.DC1.html>

**Cited Articles** This article cites by 47 articles, 14 of which you can access for free at:  
<http://clincancerres.aacrjournals.org/content/18/6/1704.full.html#ref-list-1>

**Citing articles** This article has been cited by 2 HighWire-hosted articles. Access the articles at:  
<http://clincancerres.aacrjournals.org/content/18/6/1704.full.html#related-urls>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, contact the AACR Publications Department at [permissions@aacr.org](mailto:permissions@aacr.org).

## Search for a Gene Expression Signature of Breast Cancer Local Recurrence in Young Women

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

### Abstract

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

**Experimental Design:** Gene expression 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 gene expression profile of the 343 breast cancers and to identify subgroups that differ for their risk of local recurrence.

**Results:** The previous local recurrence 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 breast cancers did not show any grouping reflecting local recurrence status. Genes related to proliferation were found differentially expressed between patients with or without local recurrence only in triple-negative tumors. Supervised classification revealed no significant gene set predictive for local recurrence or able to outperform classification based on clinical variables.

**Conclusions:** Although the previously identified local recurrence signature was robust on another platform, we were neither able to validate it on an independent data set, nor able to define a strong gene expression classifier for local recurrence using a larger data set. We conclude that there are no significant differences in gene expression pattern in tumors from patients with and without local recurrence after breast-conserving treatment. *Clin Cancer Res*; 18(6); 1704–15. ©2012 AACR.

### 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 with patients treated with modified radical mastec-

tomy (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 BCT 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 biologic characteristics of the primary breast cancer may be associated with a higher risk of developing a true local recurrence, that is, regrowth of clonogenic cells that were not removed by surgery or killed by radiotherapy (6). Gene expression profiling of primary breast carcinomas using microarray technology has already been used 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 and colleagues (10), is for a series of 165 young ( $\leq 50$  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 Niméus-Malmström and colleagues (9). Kreike and colleagues were able to construct a local recurrence classifier

**Authors' Affiliations:** <sup>1</sup>Institut Curie; <sup>2</sup>INSERM U900; <sup>3</sup>CNRS, UMR 144, Paris; <sup>4</sup>Ecole des Mines de Paris, Fontainebleau; <sup>5</sup>INRIA, Saclay, France; <sup>6</sup>Netherlands Cancer Institute; <sup>7</sup>Amsterdam Medical Centrum, Amsterdam; <sup>8</sup>Institute for Radiation Oncology Arnhem, Arnhem, The Netherlands; <sup>9</sup>Department of Statistics, University of California, Berkeley, California; and <sup>10</sup>Hôpital Fattouma Bourguiba, Monastir, Tunisia

**Note:** Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

N. Servant and M.A. Bollet contributed equally to this work.

**Corresponding Authors:** Nicolas Servant, Institut Curie, Centre de Recherche, 26 rue d'Ulm, Paris F-75248, France. Phone: 331-5624-6985; Fax: 331-5624-6946; E-mail: Nicolas.Servant@curie.fr and Marc A. Bollet. E-mail: marcbollet@gmail.com

doi: 10.1158/1078-0432.CCR-11-1954

©2012 American Association for Cancer Research.

### 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 biologic 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 biologic mechanisms involved in the relapse event and to identify novel risk factors for local recurrence after breast-conserving treatment to improve the tailoring of the local treatment in patients with breast cancer.

based on the expression of 111 genes and validate this profile on an independent data set of 161 consecutive patients with breast cancer treated by BCT who 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 local recurrence in multivariate analysis.

## Materials and Methods

### 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, the Netherlands) and one French cancer center (Institut Curie, Paris, France). The 148 Dutch patients belonged to a larger series of 165 patients with breast cancer 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 premenopausal; and no previous history of cancer (except for one nonmelanoma 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 post-operative 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. One hundred and forty-one patients received adjuvant systemic treatment with chemotherapy only (110 patients), hormone therapy only (8 patients), or a combination of both (23 patients). Chemotherapy regimens (4 missing data) were either anthracy-

cline-based (79 patients) or contained cyclophosphamide, methotrexate, and fluorouracil (CMF, 50 patients). Patients were selected if they had either experienced a local recurrence (ipsilateral breast recurrence with clinical, histologic, 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. To ensure relevant data, we conducted gene expression analyses in tumors with at least 50% of cancer cells as assessed by hematoxylin and eosin staining of histologic 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.

### Clinical and histologic studies

Clinical and tumor characteristics of the 343 patients are reported in Table 1. Three pathologists (M.J. van de Vijver, B. Sigal-Zafrani, and R. Hadhri) reviewed histopathologic characteristics of all tumors. Immunohistochemical (IHC) analysis was used to determine the status of estrogen receptor (ER; clone 6F11, 1:200 dilution; Novocastra) 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 overexpression was determined according to American Society of Clinical Oncology (ASCO) guidelines.

### Isolation of RNA and gene expression profiling

Tumor material was snap-frozen in liquid nitrogen within 1 hour of surgery. Thirty 30- $\mu$ m sections were used for isolation of RNA. The first and last sections (5  $\mu$ m) 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) 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 miRNeasy 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 Bioanalyzer (Agilent Technologies). The gene expression microarrays used in this study were Illumina Human Whole Genome V3.0 arrays (Illumina, Inc.) containing 48,803 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 Gene Expression

**Table 1.** Patient and tumor characteristics with respect to the occurrence of local relapse

	Total, N = 343 (%)	Patients with LR, N = 224 (%)	Patients with no LR, N = 119 (%)
Origin			
French	195 (57)	128 (57)	67 (56)
Dutch	148 (43)	96 (43)	52 (44)
Age, <sup>a</sup> y			
≤40	78 (23)	43 (19)	35 (29)
>40	265 (77)	181 (81)	84 (71)
pT			
pT1	216 (63)	144 (64)	72 (62)
pT2–3	125 (37)	80 (36)	45 (38)
Surgical margin			
Marg–	294 (89)	197 (90)	97 (86)
Marg+	37 (11)	21 (10)	16 (14)
Surgical margin (inv.)			
Marg– inv	318 (94)	211 (94)	107 (92)
Marg+ inv	22 (6)	13 (6)	9 (8)
Surgical margin (DCIS)			
Marg– DCIS	191 (93)	125 (94)	66 (90)
Marg+ DCIS	15 (7)	8 (6)	7 (10)
pN			
Negative	215 (64)	145 (65)	70 (62)
Positive	122 (36)	79 (35)	43 (38)
RT boost			
No boost	94 (27)	65 (29)	29 (25)
Boost	248 (73)	159 (71)	89 (75)
RT total dose, Gy			
≤66	242 (71)	152 (68)	90 (76)
>66	100 (29)	72 (32)	28 (24)
Adjuvant CT or HT			
No	201 (59)	133 (60)	68 (57)
Yes	141 (41)	90 (40)	51 (43)
Adjuvant CT			
No CT	209 (61)	136 (61)	73 (61)
CT	133 (39)	87 (39)	46 (39)
Adjuvant HT			
No HT	311 (91)	207 (93)	104 (87)
HT	31 (9)	16 (7)	15 (13)
Histologic shape			
Sharply demarcated	86 (26)	55 (26)	31 (27)
Stellate	223 (69)	145 (69)	78 (68)
Multinodular	16 (5)	11 (5)	5 (4)
Histologic type			
Ductal	292 (85)	191 (85)	101 (85)
Lobular	26 (8)	18 (8)	8 (7)
Other type	25 (7)	15 (7)	10 (8)
Histologic grade (EE) <sup>a</sup>			
EEI/EEII	202 (60)	143 (65)	59 (50)
EEIII	137 (40)	78 (35)	59 (50)
Tubule formation, %			
>90	20 (6)	14 (6)	6 (5)
50–90	79 (24)	55 (25)	24 (21)
<10%	235 (70)	150 (68)	85 (74)

*(Continued on the following page)*

**Table 1.** Patient and tumor characteristics with respect to the occurrence of local relapse (Cont'd)

	Total, N = 343 (%)	Patients with LR, N = 224 (%)	Patients with no LR, N = 119 (%)
Polymorphism			
P1	15 (5)	8 (4)	7 (6)
P2	131 (39)	92 (42)	39 (34)
P3	187 (56)	119 (54)	68 (60)
Mitotic index <sup>a</sup>			
MI1	174 (53)	126 (59)	48 (42)
MI2	59 (18)	33 (15)	26 (23)
MI3	96 (29)	56 (26)	40 (35)
DCIS inside			
Absent	152 (62)	99 (61)	53 (65)
Present	93 (38)	64 (39)	29 (35)
Extensive DCIS inside			
Absent	296 (91)	199 (93)	97 (87)
Present	29 (9)	14 (7)	15 (13)
DCIS outside			
Absent	160 (48)	112 (52)	48 (42)
Present	171 (52)	104 (48)	67 (58)
Extensive DCIS outside <sup>a</sup>			
Absent	291 (88)	196 (91)	95 (83)
Present	40 (12)	20 (9)	20 (17)
DCIS			
Absent	118 (36)	85 (39)	33 (29)
Present	211 (64)	131 (61)	80 (71)
Extensive DCIS <sup>a</sup>			
Absent	273 (84)	187 (87)	86 (77)
Present	53 (16)	28 (13)	25 (23)
DCIS grade			
Low	20 (10)	16 (12)	4 (5)
Intermediate	69 (33)	42 (33)	27 (34)
High	120 (57)	71 (55)	49 (61)
LCIS			
Absent	301 (91)	197 (90)	104 (91)
Present	31 (9)	21 (10)	10 (9)
LVI			
Absent	254 (96)	175 (97)	79 (94)
Present	11 (4)	6 (3)	5 (6)
LVI quantity			
1	11 (14)	6 (14)	5 (14)
2-5	31 (39)	16 (37)	15 (41)
>5	38 (47)	21 (49)	17 (46)
p53 mutation			
Absent	73 (75)	56 (75)	17 (77)
Present	24 (25)	19 (25)	5 (23)
Side			
Right	86 (44)	52 (41)	34 (51)
Left	109 (56)	76 (59)	33 (49)
Familial history of BC			
Absent	148 (81)	101 (84)	47 (75)
Present	35 (19)	19 (16)	16 (25)
ER expression by GE			
Negative	76 (22)	47 (21)	29 (24)
Positive	267 (78)	177 (79)	90 (76)

*(Continued on the following page)*

**Table 1.** Patient and tumor characteristics with respect to the occurrence of local relapse (Cont'd)

	Total, N = 343 (%)	Patients with LR, N = 224 (%)	Patients with no LR, N = 119 (%)
PR expression by GE			
Negative	139 (41)	88 (39)	51 (43)
Positive	204 (59)	136 (61)	68 (57)
HR by GE			
Negative	73 (21)	45 (20)	28 (24)
Positive	270 (79)	179 (80)	91 (76)
HER2 by GE			
Negative	291 (85)	196 (88)	95 (80)
Positive	52 (15)	28 (12)	24 (20)
TNBC by GE			
No	284 (83)	185 (83)	99 (83)
Yes	59 (17)	39 (17)	20 (17)

Abbreviations: BC, breast cancer; CT, chemotherapy; EE, Ellis and Elston; GE, gene expression; HT, hormone therapy; Inv, invasion; LCIS, lobular carcinoma *in situ*; LR, local recurrence; LVI, lobular vascular *in situ*; MI, mitotic index; pN, regional nodal stage according to American Joint Committee on Cancer (AJCC); pT, classification of the pathological tumor size according to AJCC; RT, radiotherapy; TNBC, triple-negative breast cancer.

<sup>a</sup>Significant ( $P < 0.05$ )  $\chi^2$  tests, for age:  $P = 0.044$ , histologic grade:  $P = 0.012$ , extensive DCIS outside:  $P = 0.047$ , extensive DCIS:  $P = 0.041$ . Significant ( $P < 0.05$ ) ordinal  $\chi^2$  test, for mitotic index:  $P = 0.011$ .

Omnibus (GEO), series record GSE30682 (<http://www.ncbi.nlm.nih.gov/geo/>).

### Cross-platform validation of the local recurrence signature

The population of 148 Dutch patients was first studied to validate the signature developed by Kreike and colleagues (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; ref. 12) was applied to the 111-gene signature reported by Kreike and colleagues using the R-package FactoMineR (v1.14; ref. 13). MFA aims to integrate the 2 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 data sets share a common geometric structure, thus indicating that the same underlying biologic signal can be extracted from the same samples and the same set of genes, with different technology. The local recurrence status was added as an illustrative variable.

### 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), radioresistance 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), 3 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; ref. 33). The area under the receiver operating characteristic (ROC) curve (AUC) was used as the prediction quality criterion.

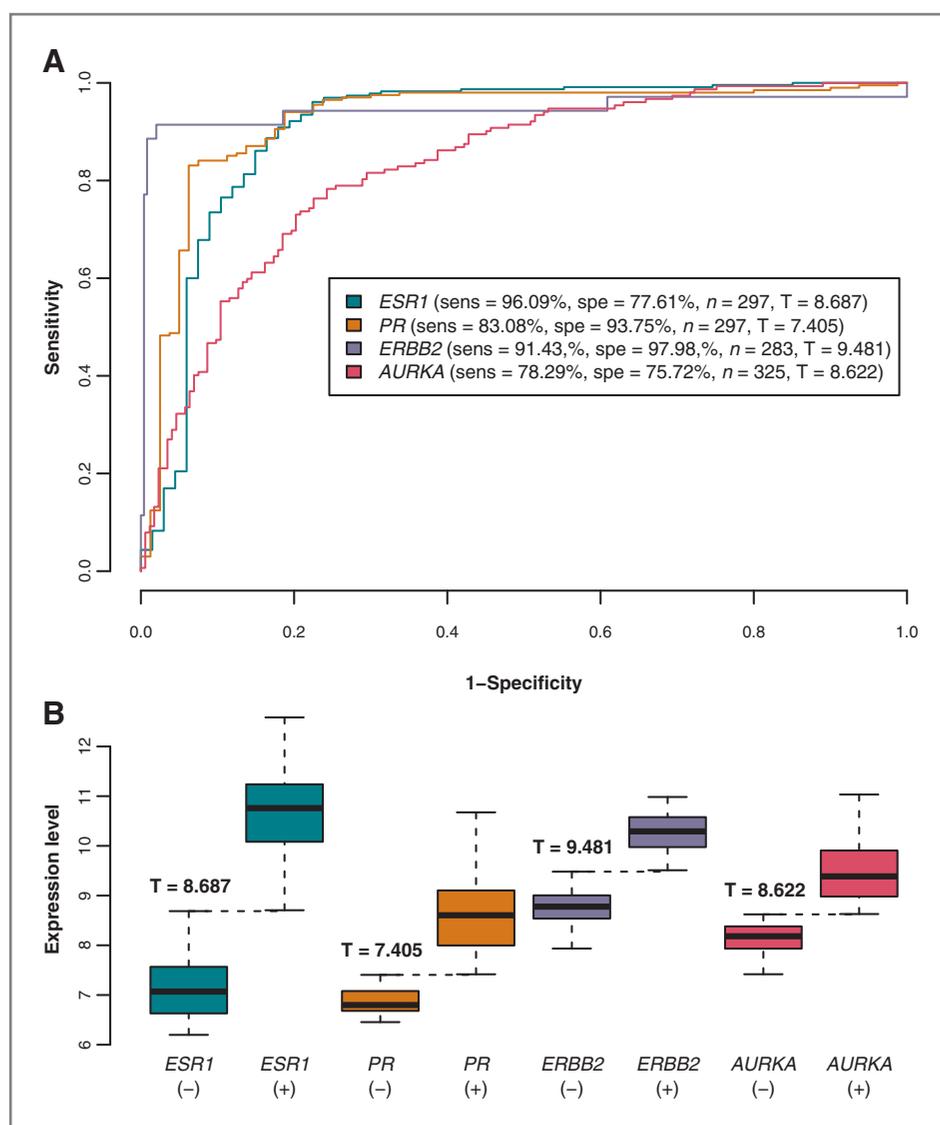
### Data analysis of 343 primary tumors

The 343 arrays were normalized using the variance stabilization transformation and the robust spline algorithm suggested by Du and colleagues (lumi R-package v2.2.0; ref. 34). Probes were then filtered to discard those called as unexpressed (or undetectable) in all samples. This left 30,198 probes to be used for analyses. To ensure comparability between the Dutch and French samples, the data set was corrected for a population effect using the following one-way ANOVA model applied to each probe  $i$ ,  $i \in [1, 30198]$ :

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

where  $Y_{ij}$  is the expression of probe  $i$  in sample  $j \in [1, 343]$ ,  $O_p$  is the effect of the population,  $p \in \{\text{Dutch, French}\}$ ,  $\mu_i$  is the expression of gene  $i$  when  $p = \text{Dutch}$ ,  $\varepsilon_{ijp} \sim N(0, \sigma_i^2)$  is a residual term. The corrected expression level of probe  $i$  was computed as follows:

$$Y_{ij}^c = Y_{ij} - O_{\text{French}}$$



**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 IHC data. The microarray-derived ER, PR, and HER2 statuses were determined according to the optimal sensitivity–specificity cutoff and applied to the 343 samples.  $n$ , number of tumors with known IHC data; sens, sensitivity; spe, specificity;  $T$ , optimal sensitivity–specificity cutoff. B, expression level of *ESR1*, *PR*, *ERBB2*, and *AURKA* for each defined immunohistochemical class. The 343 tumors were then grouped into molecular subtypes according to their gene expression, as defined by Voduc and colleagues (37).

Ward linkage hierarchical clustering of Pearson correlation similarity matrices was then applied on the 5,000 most variant probes (highest interquartile range) of the 343 samples, using the R-package EMA (v1.2; ref. 35).

On the basis of 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), *PR* (1710082), *ERBB2* (1729826/1753944), and the *AURKA* (1710733/1755528) proliferation marker. ROC curves were computed for all tumors with known IHC data. The microarray-derived ER, PR, and HER2 statuses were determined from these tumors according to the optimal sensitivity–specificity cutoff point (Fig. 1;

ref. 36) and then applied to all tumors without IHC data. Tumors were grouped into molecular subtypes according to their gene expression, as defined by Voduc and colleagues (37): luminal A ( $ER^+$  and/or  $PR^+$ ,  $HER2^-$ , low *AURKA*), luminal B ( $ER^+$  and/or  $PR^+$ ,  $HER2^-$ , high *AURKA*), luminal HER ( $ER^+$  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 and colleagues (28) was mapped onto our microarray platform and tested for class assignment. 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 on the basis of 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  $i$ ,  $i \in [1, 30198]$  to identify differentially expressed genes (DEG) between local recurrence and no local recurrence:

$$Y_{ij}^c = \mu_i + T_t + LR_r + (T \times LR)_{tr} + \varepsilon_{ijtr}(B)$$

where  $Y_{ij}^c$  is the expression of gene  $i$  in sample  $j \in [1, 343]$  corrected from the population effect ( $A$ ),  $T_t$  is the effect of the molecular subtype,  $t \in \{\text{HER2}^+, \text{luminal A, luminal B, luminal HER, triple-negative}\}$ ,  $LR_r$  is the effect of local recurrence,  $r \in \{0, 1\}$ ,  $(T \times LR)_{tr}$  is the interaction term between molecular subtype and recurrence event,  $\mu_i$  is the expression of gene  $i$  when  $t = \text{HER2}^+$  and  $r = 0$ ,  $\varepsilon_{ijtr} \sim N(0, \sigma_i^2)$  is a residual term.

The analysis was conducted 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.

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

#### 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 and colleagues have already assessed this question on the Dutch samples, using the overall gene expression pattern of 15 pairs of primary tumors and local recurrences. Regarding the French series, we analyzed 14 pairs of tumors using DNA copy number alteration (Affymetrix GeneChips Human Mapping 50K array Xba) profiles and the partial identity score proposed by Bollet and colleagues (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.

#### Supervised classification and variable selection

After appropriate scaling of the set of clinical variables (Table 1) and expression data, we used Monte Carlo cross-validation (5-folds–100 permutations) and conducted binary classification using the support vector machines (SVM) algorithm (39) to predict outcome (local recurrence or no local recurrence). A radial basis function (rbf) kernel was used, with variance fixed at 10 for the clinical data and 1,000 for expression data; internal cross-validation was carried out inside each fold to select the SVM penalty parameter  $\lambda$ . We applied the algorithm separately for clinical and expression data, as well as combining the 2 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, that is, 0.5 (sensitivity + specificity). Analyses were first conducted on all patients, then for patients with each clinical subtype to see if this led to improvement in accuracy for certain subtype.

Furthermore, we applied the least angle regression (LARS) algorithm (40), which does both supervised classification and variable selection, and can help to suggest a ranking of important predictive variables.

#### Results

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

#### Clinical and histologic features

The clinical and pathologic characteristics of patients and tumors are displayed in Table 1. Patients who developed local recurrences were significantly younger at diagnosis than those who did not ( $P = 0.044$ ;  $\chi^2$  test). Their tumors had significantly higher histologic grade according to Ellis and Elston ( $P = 0.012$ ;  $\chi^2$  test), higher mitotic indexes ( $P = 0.011$ ; ordinal  $\chi^2$  test), and more frequently contained an extensive ductal carcinoma *in situ* (DCIS) component ( $P = 0.041$ ;  $\chi^2$  test).

#### Gene expression studies

**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 cutoff from the ROC curves produced using IHC data (Fig. 1). These statuses were respectively defined from the probes ESR1 (1731003, cutoff = 8.687), PR (1710082, cutoff = 7.405), ERBB2 (1729826/1753944, cutoff = 9.481), and AURKA (1710733/1755528, cutoff = 8.622).

**Cross-platform comparison.** Of the complete data set of 343 patients, 148 were also used by Kreike and colleagues to define the 111-gene signature and studied using the Operon microarray platform.

MFA (Fig. 2) aims to integrate the 2 groups of variables (111-gene expression profiles from both Operon and Illumina data) which are expected to describe the same observations (148 tumor samples) to assess the robustness of this signature on another microarray platform, as well as its association with local recurrence. The 2 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 2 data sets. The local recurrence 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.

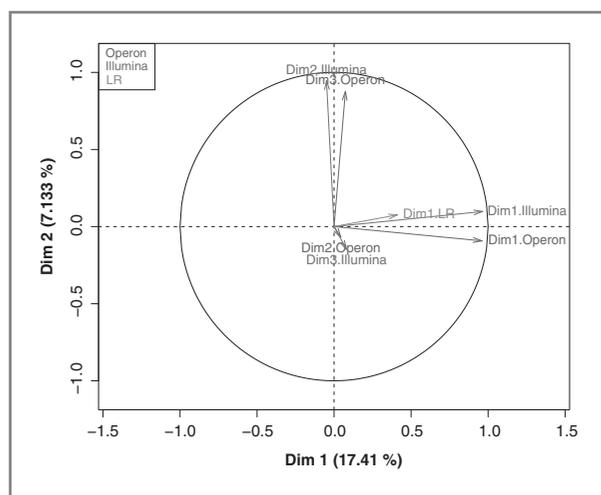


Figure 2. Multiple factor analysis. The partial axis representation shows the link between the 2 first principal components of the MFA and those of each individual group (Illumina and Operon). Local relapse information is added as an illustrative variable. LR, local recurrence.

#### Validation of the previous local recurrence signature.

Another way to assess the predictive power of the signature from Kreike and colleagues 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 and colleagues (sensitivity, 0.77 and specificity, 0.43). We assessed how other gene expression signatures conducted in pairing tumors to local recurrence 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 (Supplementary Table S1).

#### Overall gene expression profiles of 343 primary tumors.

Hierarchical clustering of the 343 tumors and the 5,000 most variant probes resulted in the formation of 2 main clusters that could be further subdivided into 4 subclusters (Fig. 3). None of the tumor characteristic related to local recurrence was associated with these clusters.

These 4 clusters were closely related to the subtype definitions. We clearly identified the triple-negative tumors and 2 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% (Supplementary Table S2; hypergeometric tests,  $P = 2.86 \times 10^{-36}$ ,  $P = 1.60 \times 10^{-18}$ , and  $P = 1.46 \times 10^{-53}$ ) of the tumors in the 3 groups. The last cluster is enriched in HER2 profiles and can be further divided in the HER2<sup>+</sup> profiles and a mixture between luminal HER2 and luminal B profiles. The class assignments predicted from the study of Hu and colleagues (28), SSP, do not outperform the clustering based on the ESR1, PR, HER2, and proliferation subtype prediction (Supplementary Fig. S1).

Information on breast cancer subtypes is crucial and has to be taken into account in any supervised analysis to avoid confounding effects. The association between breast cancer subtype and relapse events (Supplementary Table S3;  $P = 0.033$ ;  $\chi^2$  test) has already been reported in several previous studies (10, 37). To identify DEGs between patients developing a local recurrence and those who do not, we used the linear model defined in the work of Clarke and colleagues (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<sup>+</sup> tumors. When the subtype factor was removed from the linear model, no genes were found to be significant.

To identify whether these 2 lists of probes were enriched for genes representing particular Gene Ontology categories compared with the entire list of probes, a hypergeometric test was used and the Gene Ontology categories with  $P$  values lower than 0.001 were reported. Overrepresentation of genes related to proliferation, immune response, and cell death were found in the triple-negative differentially expressed probes (DEG). No interesting category was found in the list of DEGs in the HER2 overexpressed population (Supplementary Table S4).

#### Partial identity score of primary tumors and local recurrences

We applied the partial identity score proposed by Bollet and colleagues (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 (Fig. 4). This indicates that the DNA copy number profile of the local recurrences and the

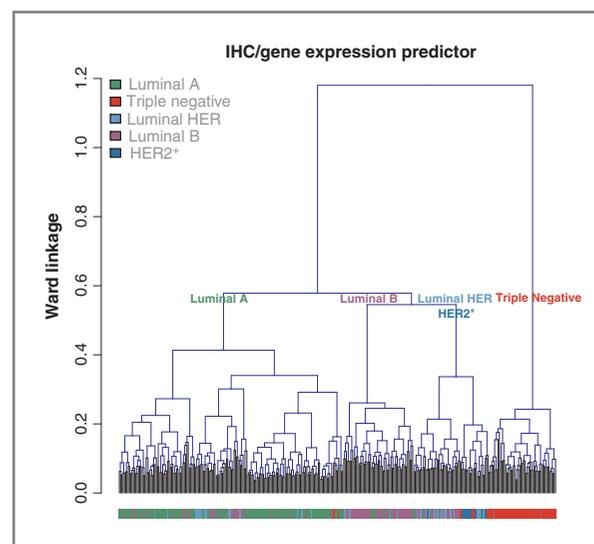
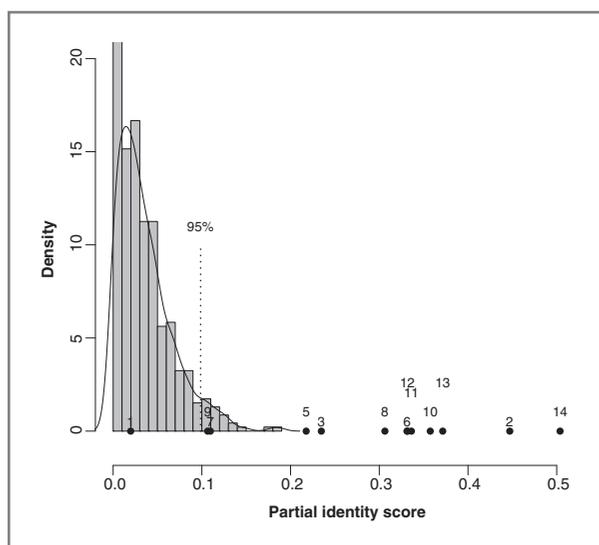


Figure 3. Hierarchical clustering of the 343 primary tumors. The samples can be divided into 4 groups according to their subtype. We clearly identify the triple-negative tumors. Two other clusters are mainly composed with luminal A and luminal B tumors. Finally, the third group is composed with most of the HER2<sup>+</sup> profiles. Group composition is given in more detail in Supplementary Table S2.



**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 and colleagues (38). The histogram was constructed using pairs from different patients. The vertical dashed line represents the upper fifth percentile of the null hypothesis and the threshold above which true recurrences were defined. Each dot represents one of the 14 natural pairs.

primary tumors are highly similar. It also validates the inclusion of the French samples and the relevance of their recurrence status.

#### Supervised classification and variable selection

Using the SVM algorithm with an rbf kernel, we obtained an average balanced accuracy of  $62\% \pm 2\%$  using clinical variables and  $59\% \pm 2\%$  using expression data. Using a mean kernel to combine the 2 sets of variables gave  $64\% \pm 2\%$  average balanced accuracy whereas a product kernel gave  $63\% \pm 2\%$ . This indicates that combining the 2 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 conducted 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 1,000 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.

#### Discussion

We have conducted gene expression profiling of 343 primary breast cancers that were treated with BCT. We chose to base our study on a series of young ( $\leq 50$  years old), premenopausal 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 to ensure a high level of homogeneity. In addition, all patients had undergone breast-conserving surgery followed by whole-breast radiotherapy for their initial breast cancers.

In a previous study, Kreike and colleagues (10) found that there were no significant differences in the overall gene expression profiles between primary tumors with or without local recurrence. 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 local recurrence 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 and colleagues. Among the 165 new French samples,

**Table 2.** Supervised classification (SVM algorithm; ref. 39) by subtype to predict outcome (LR or no LR)

	Clinical	Expression	Mean kernel	Product kernel
Luminal A	55 $\pm$ 4	52 $\pm$ 2	53 $\pm$ 3	53 $\pm$ 3
Luminal B	64 $\pm$ 4	45 $\pm$ 3	64 $\pm$ 3	65 $\pm$ 4
Luminal HER	51 $\pm$ 6	62 $\pm$ 5	51 $\pm$ 5	48 $\pm$ 6
Triple-negative	49 $\pm$ 6	54 $\pm$ 5	52 $\pm$ 6	50 $\pm$ 6
HER2+	—	—	—	—

NOTE: Monte Carlo cross-validation (5-folds—100 permutations) was used to assess performances. The rbf kernel had variance of 10 for clinical and 1,000 for expression data. Balanced accuracy is reported ( $\pm$ SD). Not enough HER2<sup>+</sup> patients were available to conduct internal cross-validation step.

**Table 3.** Top 10 variables selected by LARS

Ranking	Variable	Weight
1	RT whole breast dose (in Gy)	-0.33
2	Age (in y)	-0.09
3	Surgical margin positive (invasive component)	0.08
4	Mitotic index	0.06
5	RT total dose in Gy [ $<60$ , (60–70), $>70$ Gy]	-0.10
6	Micropapillary	0.06
7	Origin	0.11
8	Hormone therapy	0.04
9	Surgical margin positive (DCIS component)	0.06
10	Progesterone receptor percent	-0.08

NOTE: The respective weights are specified for the 10 first variables.

Abbreviation: RT, radiotherapy.

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-gene 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 2 data sets, and further, shows an association between local recurrence and the expression profiles of these 111 genes. However, when we applied this signature to an independent data set (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 local recurrence.

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<sup>+</sup> 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 and colleagues (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 local recurrence. This observation is critical for further analysis, and particularly for the search of DEGs. On the basis of the linear model (2) which takes into account the molecular subtype and its association with local recurrence, we found significant DEGs only for the HER2<sup>+</sup> 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 and colleagues (10) and agrees with previously published classifiers. The fact that no interesting categories were found in the HER2<sup>+</sup> overexpressed subtype might be explained by the small number of patients in this class (8 local recurrence and 6 no local recurrence). Further investigations and a larger number of patients with HER2<sup>+</sup> 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 and colleagues (10). The 2 different microarrays platforms, the heterogeneity of the data and the different subtype proportions might have some influence on this finding.

On the basis of 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 conducted supervised classification (SVM and LARS) on both the clinical and expression data, to create classifiers to predict local recurrence or no local recurrence. Using SVM, balanced accuracy between 60% and 64% was achieved and remains unsatisfactorily low. Separation of the data into subtypes gave similar balanced accuracy, although in some case worse. This may be partially because of 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 2 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 and colleagues (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 and colleagues (10) and to define a strong classifier for local relapse after BCT from this large and well-characterized data set, 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 (Villejuif, France), and the Karolinska Institutet (Stockholm, Sweden) 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 showed that proliferation, age, and tumor invasion are important criteria in predicting local recurrence. Finally, this study also highlights the difficulty of reproducing and validating gene expression signatures between competing microarray platforms and from one population to another.

## 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;24:3367–73.
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;366:2087–106.
3. Engel J, Kerr J, Schlesinger-Raab A, Sauer H, Hölzel D. Quality of life following breast-conserving therapy or mastectomy: results of a 5-year prospective study. *Breast J* 2004;10:223–31.
4. Moyer A. Psychosocial outcomes of breast-conserving surgery versus mastectomy: a meta-analytic review. *Health Psychol* 1997;16:284–98.
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;39:932–44.
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;27:575–83.
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;12:5705–12.
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;15:4181–90.
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;16:144–8.
12. Escofier B, Pagès J. Multiple factor analysis. *Comput Stat Data Anal* 1990;18:121–40.
13. Lê HF. FactoMineR: an R package for multivariate analysis. *J Stat 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;318:1108–13.
15. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 2006;38:1043–8.
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 U S A* 2005;102:3738–43.
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, Weichselbaum RR. STAT1 is overexpressed in tumors selected for radioresistance and confers protection from radiation in transduced sensitive cells. *Proc Natl Acad Sci U S A* 2004;101:1714–9.
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;356:217–26.
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;415:530–6.
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 U S A* 2005;102:13550–5.
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;351:2817–26.
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;365:671–9.
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 U S A* 2007;104:7564–9.
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;98:262–72.
27. Starmans MH, Krishnapuram B, Steck H, Horlings H, Nuyten DS, van de Vijver MJ, et al. Robust prognostic value of a knowledge-based

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interests were disclosed.

## Acknowledgments

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

## Grant Support

This work was financed by the Institut Curie, the 'Courir pour la vie, Courir pour Curie' association, the 'Odyssee' association, and PHRC 2006.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 2, 2011; revised December 22, 2011; accepted January 9, 2012; published OnlineFirst January 23, 2012.

- proliferation signature across large patient microarray studies spanning different cancer types. *Br J Cancer* 2008;99:1884–90.
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 Genom* 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, Wessels LFA. Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability. *BMC Genom* 2008;9:375.
  31. Whitfield ML, George LK, Grant GD, 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, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 2005;21:3301–7.
  34. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 2008;24:1547–8.
  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;3:32–5.
  37. Voduc KD, Cheang MCU, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. *J Clin Oncol* 2010;28:1684–91.
  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, Smola A. Learning with kernels: support vector machines, regularization, optimization, and beyond. Cambridge, MA: MIT Press; 2002.
  40. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. *Annals Stat* 2004;32:407–99.
  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;17:719–25.
  42. Borger J, Kemperman H, Hart A, Peterse H, van Dongen J, Bartelink H. Risk factors in breast-conservation therapy. *J Clin Oncol* 1994;12:653–60.
  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;40:859–67.
  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;65:1345–52.
  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;82:272–80.
  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;71:1014–21.