A Prognostic DNA Signature for T1T2 Node-Negative Breast Cancer Patients

Eléonore Gravier,[1,2,3,4†] Gaëlle Pierron,[5†] Anne Vincent-Salomon,[5,6†] Nadège Gruel,[1,6] Virginie Raynal,[6] Alexia Savignoni,[2] Yann De Rycke,[2] Jean-Yves Pierga,[7] Carlo Lucchesi,[6] Fabien Reyal,[8] Alain Fourquet,[9] Sergio Roman-Roman,[1] François Radvanyi,[10] Xavier Sastre-Garau,[5†] Bernard Asselain,[2,3,4†] and Olivier Delattre[5,6*†]

[1]Department of Translational Research, Institut Curie, Paris, France
[2]Department of Biostatistics, Institut Curie, Paris, France
[3]INSERM U900, Institut Curie, Paris, France
[4]Ecole des Mines de Paris, Paris, France
[5]Department of Tumor Biology, Institut Curie, Paris, France
[6]INSERM U830, Institut Curie, Paris, France
[7]Department of Medical Oncology, Institut Curie, Paris V Université Paris Descartes, Paris, France
[8]Department of Surgical Oncology, Institut Curie, Paris, France
[9]Department of Radiation Oncology, Institut Curie, Paris, France
[10]UMR CNRS/IC 144, Institut Curie, Paris, France

Predicting evolution of small node-negative breast carcinoma is a real challenge in clinical practice. The aim of this study was to search whether qualitative or quantitative DNA changes may help to predict metastasis of small node-negative breast carcinoma. Small invasive ductal carcinomas without axillary lymph node involvement (T1T2N0) from 168 patients with either good (111 patients with no event at 5 years after diagnosis) or poor (57 patients with early metastasis) outcome were analyzed with comparative genomic hybridization (CGH) array. A CGH classifier, identifying low- and high-risk groups of metastatic recurrence, was established in a training set of 78 patients, then validated, and compared with clinicopathological parameters in a distinct set of 90 patients. The genomic status of regions located on 2p22.2, 3p23, and 8q21-24 and the number of segmental alterations were defined in the training set to classify tumors into low- or high-risk groups. In the validation set, in addition to estrogen receptors and grade, this CGH classifier provided significant prognostic information in multivariate analysis (odds ratio, 3.34; 95% confidence interval 1.01–11.02; $P = 4.78 \times 10^{-2}$, Wald test). This study shows that tumor DNA contains important prognostic information that may help to predict metastasis in T1T2N0 tumors of the breast.

INTRODUCTION

Breast carcinoma patients with identical clinicopathological presentations can experience different outcomes ranging from rapid metastatic relapse to long-term disease-free survival. Patient age, tumor size, vascular invasion, grade, axillary lymph node status, hormonal receptor, and HER2 status are currently used to identify patients with a high risk of poor outcome (Cinieri et al., 2007).

Cytogenetic analyses have demonstrated that the pattern of genetic alterations may have a prognostic significance in breast cancer. Complex karyotypes or the presence of homogeneously staining regions were correlated with poor outcome, whereas detection of the t(1;16) unbalanced translocation or of 16q losses were associated with better survival (Bernardino et al., 1998; Roylance et al., 2006; Tsuda et al., 1998). More recent molecular analyses have confirmed these pioneer studies and further unraveled other genetic lesions that harbor prognostic information such as regions of amplification (Al-Kuraya et al., 2004; Cuny et al., 2000; Slamon et al., 1989).

Additional Supporting Information may be found in the online version of this article.

Supported by: Ligue Contre le Cancer that allowed the BAC array construction in the context of the "Carte d'Identité des Tumeurs" program, by a grant from the Institut Curie "PIC Sein" incentive and collaborative program, by the INCa (projet 2008 GepiG "Génétique et épigénétique fonctionnelle: Une approche intégrée pour l'étude du cancer du sein, apport des systèmes modèles à la valorisation des ressources biologiques et nouvelles générations de séquençage"); INTERFACE INSERM (to A.V.-S.).

†Eléonore Gravier, Gaëlle Pierron, Anne Vincent-Salomon, Xavier Sastre-Garau, Bernard Asselain, and Olivier Delattre contributed equally to this work.

*Correspondence to: Dr. Olivier Delattre, INSERM U830, Institut Curie, 26 rue d'Ulm, Paris 75248, Cedex 05, France.
E-mail: Olivier.delattre@curie.fr

Received 11 May 2010; Accepted 1 August 2010

DOI 10.1002/gcc.20820

Published online 14 September 2010 in Wiley Online Library (wileyonlinelibrary.com).

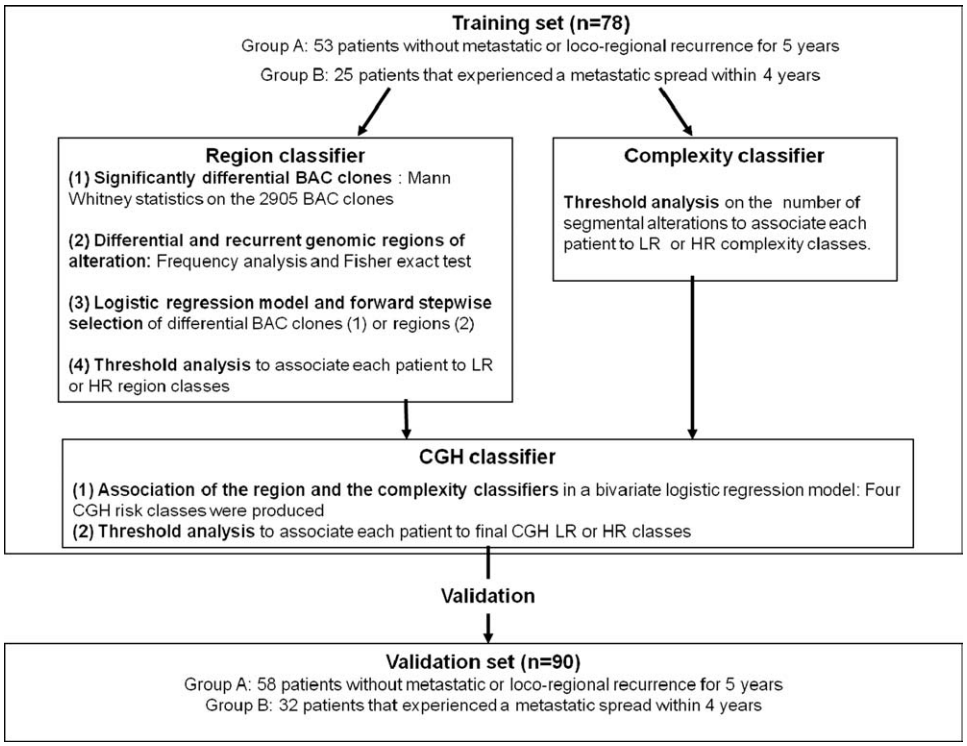

Figure 1. Diagram of the strategy used to define the CGH classifier. LR, low risk; HR, high risk.

Several tumor suppressor genes such as *TP53* have been identified to be frequently mutated, particularly in poor prognosis carcinomas, and are associated with chromosome 17p losses. They are preferentially observed in estrogen receptor (ER)-negative patients (Langerod et al., 2007; Olivier et al., 2006). Molecular subtypes of breast cancers have been identified as distinct entities associated with specific genomic profiles and demonstrate different outcomes (Adelaide et al., 2007; Chin et al., 2006, 2007b).

Despite growing evidence that quantitative and qualitative analyses of the DNA alterations that characterize breast cancer cells may be highly informative to predict aggressiveness of the disease, few controlled studies have systematically addressed the issue of the prognostic significance of DNA lesions (Zhang et al., 2009).

Our study was performed to identify a DNA signature that could help to predict the risk of metastatic progression in T1T2N0 breast carcinoma. This prognostic DNA signature comprised two components: three different chromosomal regions and the number of DNA genomic alterations, i.e., genomic complexity. It was defined in a training set of 78 small carcinomas node-negative (pT1, tumor <2 cm; pT2, tumors from 2 to 5 cm; and pN0, pathological nonmetastatic axillary lymph nodes) invasive ductal breast carcinomas and was validated in an independent set of 90 tumors.

MATERIALS AND METHODS

Study Design

A case–control design was used to determine the DNA signature (Fig. 1). Two groups of patients with opposite outcomes were defined: group A (controls) comprised patients with no metastatic or locoregional recurrence for 5 years and group B patients (cases) who had experienced metastatic relapse within 4 years. A first set of 78 tumors (training set) was analyzed to define the signature. The accuracy of the signature was then validated in an independent validation set (90 tumors).

Patients and Tumors

All analyzed tumors were selected among the 7,469 patients treated between 1989 and 1999 with a conservative surgery followed by radiation therapy for pT1T2N0 breast carcinoma at the Institut Curie. The selection criteria were size <5 cm, node-negative, invasive ductal breast carcinomas with a follow-up of at least 5 years for

patients of group A. Patients older than 75 years, with history of cancer, with familial history of breast cancer, with multifocal or bilateral tumors, or with initial metastatic disease were excluded from the study. The final selection was based on the availability and quality of frozen tumor samples in the tumor bank and of a representative formalin-fixed tumor sample. Because early metastatic events are rare in T1T2N0 breast carcinomas, all B patients with frozen material (57 among a total of 114) were included in the study. Group A patients were randomly selected among 3,100 patients who passed the selection criteria described above and distributed among the training and validation sets. For training analyses, approximately half of the B (25 tumors) and 53 A tumors were used. The validation set contained the remaining 32 B tumors together with 58 A tumors. The use of twice the number of controls with respect to the number of cases was chosen as a compromise between the need for statistical power and the experimental costs. All experiments were performed in agreement with the French Bioethics Law 2004-800, the French National Institute of Cancer (INCa) Ethics Charter, and with the permission of institutional review boards.

Estrogens, progesterone, and HER2 receptor status was assessed by immunohistochemistry on the representative formalin-fixed tumor blocks, according to previously published protocols (Vincent-Salomon et al., 2008). The expression of ER (clone 6F11, 1/200, Novocastra, Menarini, Rungis, France; cutoff: >10% of positive cells), progesterone receptor (PR, clone 1A6, 1/200, Novocastra; cutoff: >10% of positive cells), and HER2 (clone CB11, 1/1,000, Novocastra; cutoff: >30% of positive cells (Wolff et al., 2007)) was evaluated. Internal and external controls were included in all experiments. Histoprognostic grade was assessed retrospectively according to the recommendations of Elston and Ellis (1991).

Array Comparative Genomic Hybridization

The bacterial artificial chromosome (BAC) array (2,905 BACs, 1 megabase resolution) together with experimental procedures were used according to previously published protocols (Neuvial et al., 2006; Vincent-Salomon et al., 2008). All experiments were performed blinded to the outcome and with the same male reference DNA for all cases.

The number of segmental alterations occurring in tumors, hereafter called genomic complexity, was determined on the array comparative genomic hybridization (CGH) profiles visualized using the visualization and analysis of array CGH, transcriptome, and other molecular profiles (VAMP) and gain and loss analysis of DNA (GLAD) software (Hupe et al., 2004; La Rosa et al., 2006). Segmental alterations were defined by a breakpoint between two clones harboring a different status being normal versus gained, normal versus lost, or gained versus lost. Entire chromosome copy number alterations were not taken into account because they did not demonstrate any statistical significance in analysis of the prognostic impact of the number of genomic alterations in this population (data not shown). The single outliers were not taken into account. Based on the identification of copy number variants described in the Database of Genomic Variant, known germline copy number variants were tagged and disregarded. All CGH data are available as GSE19159 study at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi.

Statistical Methods

The R software (v2.7.2) was used for all statistical analyses. The homogeneity between groups was evaluated by the Pearson's χ^2 test (or its modifications when conditions for application were not validated). The Wald test was used to compute significance in logistic regressions.

Kaplan–Meier survival plots and log-rank tests were used to assess the differences in metastasis-free survival curves. The hazard ratio and its 95% confidence interval (CI) were derived from a Cox proportional-hazards regression model.

Definition of the CGH classifier

This classifier was composed by the region and the complexity classifiers defined in the training set and then tested in the validation set (Fig. 1).

Region classifier

First, all BAC clones that exhibited significantly different Cy5/Cy3 values between group A and B tumors were identified (Mann–Whitney test, $P \leq 0.05$ after Benjamini and Hochberg corrections to adjust for multiple testing) (Benjamini and Hochberg, 1995). The median Cy5/Cy3 ratio was calculated over the genomic regions that contained clones with significantly different values between A and B tumors. Boundaries of these regions were defined by at least four contiguous and neighboring BAC clones not exhibiting a different frequency of alteration between groups A and B.

TABLE 1. Clinicopathological Characteristics of Tumors and Patients in Training and Validation Sets

	Training set ($n = 78$)		Validation set ($n = 90$)	
Characteristics	Group A ($n = 53$)	Group B ($n = 25$)	Group A ($n = 58$)	Group B ($n = 32$)
Age (years)				
Mean (SD)	57 (9)	57 (10)	53 (8)	51 (10)
>50	38 (72%)	18 (72%)	33 (57%)	17 (53%)
≤50	15 (28%)	7 (28%)	25 (43%)	15 (47%)
Size				
T0	4 (7%)	0 (0%)	7 (12%)	1 (3%)
T1	37 (70%)	9 (36%)	37 (64%)	16 (50%)
T2	12 (23%)	16 (64%)	14 (24%)	15 (47%)
Elston–Ellis grade				
I	37 (70%)	4 (16%)	32 (58%)	4 (13%)
II	14 (26%)	12 (48%)	18 (33%)	10 (32%)
III	2 (4%)	9 (36%)	5 (9%)	17 (55%)
Missing values	0	0	3	1
ER status (IHC)				
Positive	43 (86%)	20 (80%)	51 (93%)	18 (56%)
Negative	7 (14%)	5 (20%)	4 (7%)	14 (44%)
Missing values	3	0	3	0
PR status (IHC)				
Positive	33 (62%)	14 (56%)	43 (72%)	14 (44%)
Negative	20 (38%)	11 (44%)	12 (22%)	18 (56%)
Missing values	0	0	3	0
HER2 status				
Negative	50 (96%)	23 (92%)	50 (91%)	27 (84%)
Positive	2 (4%)	2 (8%)	5 (9%)	5 (16%)
Missing values	1	0	3	0
Vascular invasion				
No	49 (92%)	20 (80%)	50 (89%)	20 (71%)
Yes	4 (8%)	5 (20%)	6 (11%)	8 (29%)
Missing values	0	0	2	4

SD, standard deviation; ER, estrogen receptors; PR, progesterone receptor; IHC, immunohistochemistry, Group A, patients with no metastatic or locoregional recurrence for 5 years; Group B, patients with early onset metastasis (median time to metastasis, 2.1 years).

Finally, the region classifier was established on the training set using multivariate logistic regression with forward stepwise selection based on the Akaike information criterion. This multivariate analysis was based on the median Cy5/Cy3 values computed over the regions defined above. Such logistic models provide an estimation of each patient's risk corresponding to a theoretical probability, also called score, to experience a metastatic event in less than 48 months (i.e., group B patient). Each patient was assigned to a low risk (LR_{region}) or high risk (HR_{region}) class when her score was less than or greater than an optimal threshold defined to maximize the sum of sensitivity and specificity (Youden index). Sensitivity and specificity are the proportion of group B and group A patients correctly classified by the model, respectively.

Complexity classifier

The complexity classifier was built by implementing a univariate logistic model including the total number of segmental alterations. It assigned each patient to either the low risk ($LR_{complexity}$) or high risk ($HR_{complexity}$) class using the threshold described above (Youden index).

Comprehensive CGH classifier

Finally, the CGH classifier was built by associating the information provided by the region and complexity classifiers, hence defining four possible CGH risk categories: $LR_{region}/LR_{complexity}$, $LR_{region}/HR_{complexity}$, $HR_{region}/LR_{complexity}$, and $HR_{region}/HR_{complexity}$. Each patient was finally assigned to one of the two LR or HR CGH classes according to the threshold optimization scheme.

Validation of classifiers in the validation set

To assess the contribution of the CGH classifier on the definition of prognosis compared with clinicopathological parameters, univariate and multivariate analyses were performed using a logistic regression model on the clinicopathological variables (age, grade, ER and PR status,

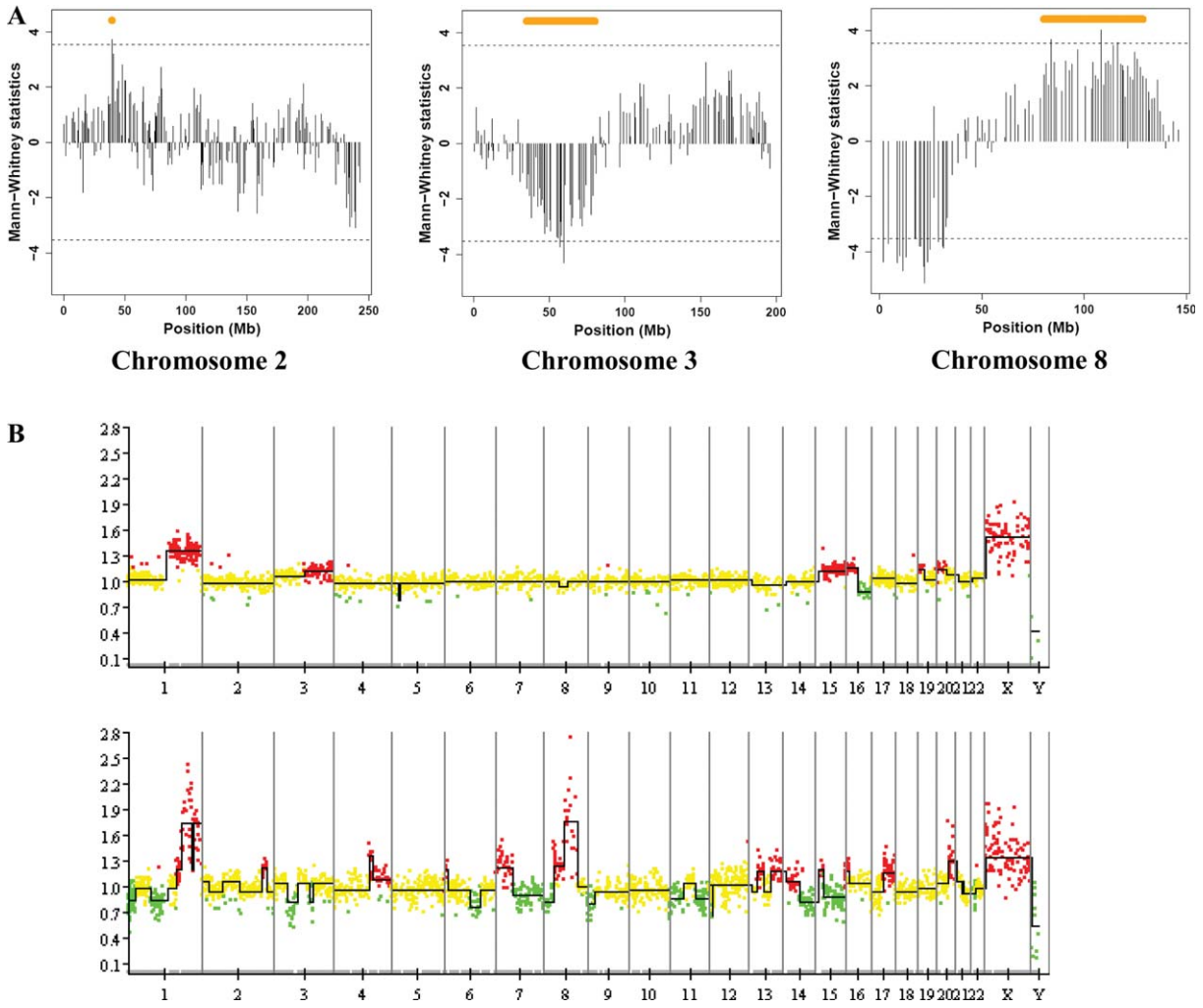

Figure 2. Analysis of the differential regions and genomic complexity in the training set. (A) Mann–Whitney statistics for clones on chromosomes 2, 3, and 8. The X-axis represents the position (Mb) of the clones along the chromosomes and the Y-axis the values of Mann–Whitney statistics. Negative statistics mean that the median Cy5/Cy3 values of group B tumors were lower than the median Cy5/Cy3 values of group A tumors. Horizontal dashed lines correspond to thresholds of significance after Benjamini Hochberg correction. The three prognostic regions located on 2p22.2, 3p23, and 8q21-24 that define the region classifier are represented by orange lines at the top of each of the three graphs. (B) Two-array CGH profiles illustrating the various levels of genomic complexity. Each graph shows the Cy5/Cy3 values (Y-axis) of a tumor sample along the chromosomes (X-axis). Chromosome boundaries are indicated by vertical gray lines. The smoothing line of the Cy5/Cy3 values generated by the GLAD algorithm (Hupe et al., 2004) is drawn in black. Normal statuses of Cy5/Cy3 values are shown in yellow, gains in red, and losses in green. (Top) The tumor harbors a simple profile with few segmental alterations (low genomic complexity). (Bottom) The tumor harbors a more complex profile with distal or interstitial (segmental) gains and losses (high genomic complexity).

vascular invasion, tumor size, and HER2 status) with and without the CGH classifier in the validation set.

RESULTS

Clinicopathological Characteristics of Tumors

Median follow-up for group A training and validation sets was 11.4 and 9.4 years, respectively. Time to metastasis for group B training and validation sets was 2.1 years. The main difference between training and validation sets regarding treatment protocols was the increased indication of chemotherapy. Indeed, one B patient from the training set received hormonal therapy contrasting with 25 patients from the validation set who received anthracycline-based adjuvant chemotherapy or hormonal therapy or both. Most (16 of 25) of the treated patients were indeed B patients. The clinicopathological characteristics of tumors and patients are summarized in Table 1 and detailed in Supporting Information Table S1. No statistically significant difference as regards to clinicopathological characteristics was observed between the A groups from training and validation sets or between the B groups from training and validation sets.

CGH Analysis and Definition of the Classifiers in the Training Set

DNAs from tumors of the training set were analyzed on BAC arrays as previously described (Vincent-Salomon et al., 2008). Comparison of the Cy5/Cy3 values between groups A and B at each of the 2,905 BAC loci identified 24 BACs with statistically significant differences. Twenty-two of these BACs were located within four recurrently altered genomic regions on 3p, 8p, 8q, and 11p. Two clones, harboring very significant differences between the two groups of tumors on 2p and 6q, were solitary and apparently not located in recurrent regions of alteration (see the chromosome 2 clone in Fig. 2A). Median values across the four regions or individual clone values (for the two clones on the 2p and 6q) were used for multivariate logistic regression analyses that resulted in the selection of the 2p22.2, 3p23, and 8q21-24 loci to build the region classifier (Fig. 2A and Supporting Information Table S2). The optimal threshold analysis maximizing the sum of sensitivity and specificity provided a value of 0.45 to assign each patient to either low-risk or high-risk region classes (Supporting Information Table S1 and Fig. S3).

The genomic complexity observed in tumors of the training set was also evaluated (Fig. 2B). Strikingly, univariate analysis of genomic complexity showed that it was very significantly associated with metastasis (logistic regression model, $P = 0.00056$, Wald test). An optimal threshold of 11 chromosomal segmental alterations was defined (Youden index) to assign each patient to a risk class. A tumor with more than 11 segmental alterations was classified in the $HR_{complexity}$ group; otherwise, it was assigned in the $LR_{complexity}$ group (Supporting Information Table S1 and Fig. S3).

The region and complexity classifiers were then included in a bivariate regression logistic model and were both shown to remain highly significant ($P = 0.0000052$ and $P = 0.0053$, Wald test, respectively). The threshold analysis performed on the CGH classifier classified a tumor in the LR group when it was predicted in the LR group by both the region and complexity classifiers. When at least one component of the classifier (region or complexity) assigned the patient to the HR group, the

TABLE 2. Distribution of Patients in the Four CGH Risk Classes Defined by the Region and Complexity Classifiers in the Validation Set

CGH risk classes	Group A ($n = 58$)	Group B ($n = 32$)	Total ($n = 90$)
LRr/LRc	49 (82%)	11 (18%)	60
LRr/HRc	6 (40%)	9 (60%)	15
HRr/LRc	1 (20%)	4 (80%)	5
HRr/HRc	2 (20%)	8 (80%)	10

LR and HR, low-risk and high-risk classes, respectively; r, region classifier; c, complexity classifier; Group A, patients with no metastatic or locoregional recurrence for 5 years; Group B, patients with early onset metastasis (median time to metastasis, 2.1 years); CGH, comparison genomic hybridization.

TABLE 3. Comparison Between the CGH Classifier and the Clinicopathological Parameters Prognostic Values in the Validation Set

	Univariate analysis		Multivariate analysis: Clinicopathological model[a]		Multivariate analysis: Confrontation model[b]	
	Odds ratio (95% CI)	*P*	Odds ratio (95% CI)	*P*	Odds ratio (95% CI)	*P*
Age $\leq$ 50 vs. >50 years	1.16 (0.49–2.77)	7.30×10^{-1}				
Size T2 vs. T0/T1	2.77 (1.11–6.95)	2.95×10^{-2}				
Grade II/III vs. I (Elston–Ellis)	9.39 (2.89–30.53)	1.96×10^{-4}	4.17 (1.14–15.27)	3.12×10^{-2}	3.35 (0.88–12.8)	7.71×10^{-2}
ER status negative vs. positive (IHC)	9.92 (2.89–34.07)	2.69×10^{-4}	5.85 (1.5–22.83)	1.10×10^{-2}	4.12 (0.99–17.24)	5.24×10^{-2}
PR status negative vs. positive (IHC)	4.61 (1.79–11.88)	1.57×10^{-3}				
Vascular invasion yes vs. no	3.33 (1.03–10.83)	4.53×10^{-2}				
HER2 positive vs. negative	1.85 (0.49–6.97)	3.62×10^{-1}				
CGH classifier HR vs. LR	10.39 (3.75–28.78)	6.63×10^{-6}	–	–	3.34 (1.01–11.02)	4.78×10^{-2}

[a]Multivariate logistic regression with forward stepwise selection of clinicopathological parameters significant in univariate analysis ($P \leq 0.05$, Wald test).
[b]Multivariate logistic regression including both clinicopathological parameters selected in the clinicopathological model and the CGH classifier.
95% CI, 95% confidence interval; IHC, immunohistochemistry.

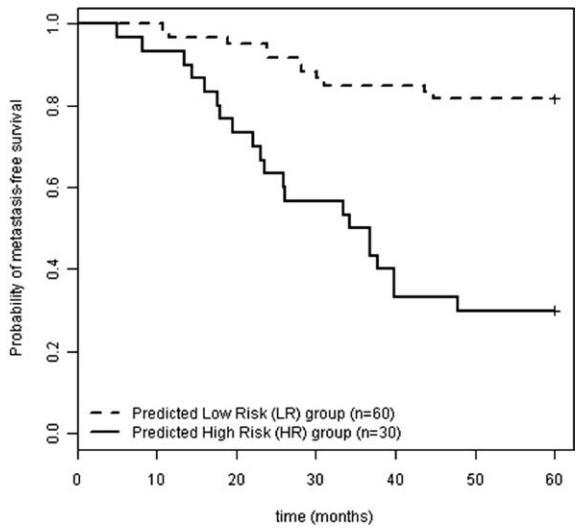

Figure 3. Kaplan–Meier curves of the low-risk (LR) and high-risk (HR) groups predicted by the CGH classifier in the validation set. The difference between the two predicted groups is highly significant (hazard ratio = 5.7, 95% CI: 2.73–11.8, $P = 1.82 \times 10^{-7}$, log-rank test).

patient was classified in the HR group (Supporting Information Table S1).

Analysis of the CGH Classifier in the Validation Set

Analysis of the region and complexity classifiers was performed in an entirely distinct set of tumors and produced receiving operating characteristics curves with significant areas of 0.76 (95% CI: 0.66–0.87) and 0.83 (95% CI: 0.75–0.92), respectively (Supporting Information Fig. S4). Highly significant odds ratio of 11 (95% CI: 2.81–43.06, $P = 0.00057$, Wald test) and 7.08 (95% CI: 2.56–19.63, $P = 0.00017$, Wald test) were also observed in univariate logistic analyses for region and complexity classifiers, respectively.

The four CGH risk categories generated by the association of the region and the complexity classifiers in the validation set and the metastatic rates according to the risk groups are displayed in Table 2. Eighteen percent of predicted low risk presented a metastatic relapse within 4 years, whereas 65% of the high risk prediction group did present a metastatic relapse.

Validation of the CGH classifier produced a highly significant odds ratio of 10.39 (95% CI: 3.75–28.78, $P = 0.0000066$, Wald test) in univariate analysis (Table 3). A sensitivity of 66% (21 of 32 patients), a specificity of 84% (49 of 58 patients), and an accuracy rate of 78% (70 of 90 patients) were also observed. The 5-year metastasis-free survival analyzed with the log-rank test showed a highly significant difference between the two predicted groups (hazard ratio = 5.7, 95% CI: 2.73–11.88, $P = 1.82 \times 10^{-7}$, log-rank test; Fig. 3). Given the small differences between treatment in the training and validation sets, it has to be pointed out that our classifier still identifies metastatic patients even though some of them received chemotherapy that may have improved their expected survival.

When tumors were classified into basal, luminal A or B, and HER2, the predicted high-risk group contained approximately equal number of basal, luminal B, and luminal A tumors (10 of 27 = 37%, 9 of 27 = 33%, and 8 of 27 = 30%, respectively), whereas in the predicted low-risk group, the majority but not all cases were luminal A (49 of 56 = 88%). A few low-risk tumors belonged to the basal and luminal B subtypes (4 of 56 = 7% and 2 of 56 = 3%, respectively).

Prognostic Value of the CGH Classifier in ER-Positive Breast Carcinoma Patients

The prognostic value of the CGH classifier on the subgroup of ER-positive tumors (51 group A and 18 group B patients, respectively) was then tested. In univariate analysis, the CGH classifier was associated with a highly significant odds ratio of 11.8 to develop metastasis (95% CI: 3.3–42.2, $P = 1.48 \times 10^{-4}$, Wald test). The 5-year metastasis-free survival analyzed by the log-rank test showed a highly significant difference between the two predicted groups (hazard ratio = 6.82, 95% CI: 2.63–17.72, $P = 0.0000052$, log-rank test).

Comparison Between the CGH Classifier and the Clinicopathological Parameters' Prognostic Values in the Validation Set

Among the various clinicopathological parameters investigated in this study, tumor size, grade, ER and PR status, and the presence of vascular invasion were significantly associated with metastatic outcome (groups A and B) at a significance level of 5% in univariate analysis (Table 3).

In multivariate logistic regression analysis of these parameters, only ER status and grade were retained in the model summarizing clinicopathological parameters. Interestingly, the CGH classifier remained significant when tested with ER status and grade in multivariate analysis (odds ratio of 3.34, 95% CI: 1.01–11.02, $P = 0.048$, Wald test)

DISCUSSION

Because metastatic events are rare in small node-negative breast carcinoma, a case–control

design was used to compare tumors from patients with rapid metastatic onset and tumors from patients with longer event-free survivals. Defined on a training set, the proposed DNA signature includes both qualitative criteria, particularly copy number alterations at three specific genomic regions (2p22.2 gain, 3p23 loss, and 8q21-24 gains) that summed up the prognostic information on genomic regions and quantitative criteria evaluating genomic complexity (number of chromosomal segmental alterations). Finally, the combined information of genomic regions and genetic complexity accurately predicted early metastatic recurrence in a completely independent validation set of T1T2 node-negative breast cancers. Logistic multivariate analysis performed on this validation set taking into account clinical parameters and CGH data showed that the CGH classifier had the ability to predict metastasis besides grade and ER status, two major clinical prognostic variables.

Some regions with very different ratios between group A and B tumors were previously shown to be of potential prognostic value in breast cancers, as 11p which has been associated with poor prognosis (Dellas et al., 2002). Chromosome arm 3p loss is observed relatively more frequently in basal-like and luminal B tumors than in other molecular types of breast cancers (Chin et al., 2006). It includes a fragile site region and encodes the *FHIT* gene suspected to play a tumor suppressor role (Pekarsky et al., 2002). Chromosome arm 8p loss and 8q gain are frequently associated and characterize particularly aggressive breast cancers (Adelaide et al., 2007; Chin et al., 2006; Marchio et al., 2008; Natrajan et al., 2009; Rennstam et al., 2003; Thor et al., 2002; Vincent-Salomon et al., 2007). In addition, the 8q region encompasses the *MYC* gene, overexpression or amplification of which has also been associated with poor outcome (Chandriani et al., 2009; Chin et al., 2007a; Schlotter et al., 2003). It is also noteworthy that the 2p region described here encompasses 2.6 Mb and encodes several genes involved in RNA metabolism, a process that is frequently represented in prognostic expression signatures (Reyal et al., 2008).

Conversely, some regions previously shown to be associated with prognosis, such as 17q12 encompassing *HER2*, were not present in our CGH signature. HER2 overexpression was found in less than 10% of cases and was not a prognostic parameter in this study. Similarly, because the majority of the studied tumors are ER-positive, it could explain why regions more preferentially associated with basal-like tumors, such as 5p loss and 10p and 9p gains were not detected. Chromosome arm 16q loss, a genomic alteration that has been associated with good prognosis, low-grade, ER-positive tumors (Cervantes and Glassman, 1996), was observed in more than 20% of all cases, regardless of the prognosis, and therefore did not contribute to the present signature.

Previous reports have also highlighted the potential prognostic significance of a number of genetic alterations that characterize the tumors using different approaches (Al-Kuraya et al., 2004; Bergamaschi et al., 2006; Carter et al., 2006; Chin et al., 2006; Courjal et al., 1997). Copy number abnormalities (high level copy number alterations) have been shown to be more frequent in large, high-grade, node-positive tumors (Chin et al., 2006, 2007b; Courjal et al., 1997), and, more recently, the presence of amplicons has been associated with the luminal B molecular subtype (Bergamaschi et al., 2006; Chin et al., 2006).

This CGH signature predicts short-term metastatic risk in small ER-positive breast cancers corresponding to the majority of cases in this study. Importantly, with the advent of mammographic screening, these tumors now constitute the most frequent initial presentation. Accurate prognostic tools in addition to grade, hormonal, and HER2 status must, therefore, be urgently defined. Gene expression prognostic signatures are mainly based on proliferation, RNA metabolism genes (Reyal et al., 2008; Wang et al., 2005), and stromal/immune genes (Finak et al., 2008). Some of these signatures have been specifically designed in ER-positive breast cancer patients (Paik et al., 2004), and clinical trials are ongoing to evaluate the efficacy and accuracy of such signatures on large cohorts of patients. We determined whether some of the 21 genes from the Oncotype Dx (Paik et al., 2004) or the 70 genes from the Mammaprint (van 't Veer et al., 2002) signatures were localized within the regions of our classifier. Five genes of the Mammaprint signature are encoded by the 8q region (8q21–8q24 from 73 to 146 Mb: *CCNE2*, *TSYPL5*, *EXT1*, *AA555029*, and *WISP1*), and none of the Oncotype Dx genes belong to the described regions (data not shown). It is noteworthy that both DNA and RNA profiles provide complementary prognostic information in breast cancer. It has to be emphasized that DNA is more robust than RNA and less sensitive to stromal and inflammatory component. Moreover,

DNA can be amplified from fixed and paraffin-embedded material that may represent a clear advantage for a prognosis tool to be used in clinical practice. However, combined studies of the two types of nucleic acids in a clinical perspective and subsequent multivariate analyses have yet to be performed to evaluate the overlap and complementarities between these two types of investigations. The DNA signature proposed shows that, as in other cancers such as chronic lymphoid leukemia (Kujawski et al., 2008), glioma (Idbaih et al., 2005), or neuroblastoma (Janoueix-Lerosey et al., 2009), specific genomic profiles provide prognostic information in breast cancer. In addition, assessment of genomic profiles also provide information about the presence of high level copy number alterations that may encompass known drug targets (Chin et al., 2006, 2007a) such as *HER2* and *FGFR1* (Reis-Filho et al., 2006) or new ones (Andre et al., 2009). In practice, it would be interesting to settle a clinical trial for small breast cancers to evaluate the feasibility and the reliability of CGH array. The CGH could be performed in a central laboratory to stratify patients in different therapeutic schemes according to their DNA signature of poor or good prognosis as already performed for example in the Mammaprint gene expression signature for the MINDACT trial.

ACKNOWLEDGMENTS

The authors thank Fabien Valet for his contribution to statistical analysis, Ingrid Lebigot and her team from the Institut Curie's tumor bank, and the Institut Curie Breast Cancer Study Group headed by Brigitte Sigal-Zafrani.

REFERENCES

Adelaide J, Finetti P, Bekhouche I, Repellini L, Geneix J, Sircoulomb F, Charafe-Jauffret E, Cervera N, Desplans J, Parzy D, Schoenmakers E, Viens P, Jacquemier J, Birnbaum D, Bertucci F, Chaffanet M. 2007. Integrated profiling of basal and luminal breast cancers. Cancer Res 67:11565–11575.

Al-Kuraya K, Schraml P, Torhorst J, Tapia C, Zaharieva B, Novotny H, Spichtin H, Maurer R, Mirlacher M, Kochli O, Zuber M, Dieterich H, Mross F, Wilber K, Simon R, Sauter G. 2004. Prognostic relevance of gene amplifications and coamplifications in breast cancer. Cancer Res 64:8534–8540.

Andre F, Job B, Dessen P, Tordai A, Michiels S, Liedtke C, Richon C, Yan K, Wang B, Vassal G, Delaloge S, Hortobagyi G, Symmans WF, Lazar V, Pusztai L. 2009. Molecular characterization of breast cancer with high-resolution oligonucleotide comparative genomic hybridization array. Clin Cancer Res 15:441–451.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J Roy Stat Soc B 57:289–300.

Bergamaschi A, Kim YH, Wang P, Sorlie T, Hernandez-Boussard T, Lonning PE, Tibshirani R, Borresen-Dale AL, Pollack JR. 2006. Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. Genes Chromosomes Cancer 45:1033–1040.

Bernardino J, Gerbault-Seureau M, Zafrani B, Dericke Y, Boudou E, Magdelenat H, Dutrillaux B. 1998. Homogeneously staining regions in 223 breast carcinomas: Cytogenetic and clinicopathological correlations. Br J Cancer 78:1214–1218.

Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. 2006. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat Genet 38:1043–1048.

Cervantes M, Glassman AB. 1996. Breast cancer cytogenetics: A review and proposal for clinical application. Ann Clin Lab Sci 26:208–214.

Chandriani S, Frengen E, Cowling VH, Pendergrass SA, Perou CM, Whitfield ML, Cole MD. 2009. A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response. PLoS One 4:e6693.

Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW. 2006. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 10:529–541.

Chin SF, Teschendorff AE, Marioni JC, Wang Y, Barbosa-Morais NL, Thorne NP, Costa JL, Pinder SE, van de Wiel MA, Green AR, Ellis IO, Porter PL, Tavare S, Brenton JD, Ylstra B, Caldas C. 2007a. High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. Genome Biol 8:R215.

Chin SF, Wang Y, Thorne NP, Teschendorff AE, Pinder SE, Vias M, Naderi A, Roberts I, Barbosa-Morais NL, Garcia MJ, Iyer NG, Kranjac T, Robertson JF, Aparicio S, Tavaré S, Ellis I, Brenton JD, Caldas C. 2007b. Using array-comparative genomic hybridization to define molecular portraits of primary breast cancers. Oncogene 26:1959–1970.

Cinieri S, Orlando L, Fedele P, Cusmai A, D'Amico M, Rizzo P, Chetri MC. 2007. Adjuvant strategies in breast cancer: New prospectives, questions and reflections at the end of 2007 St Gallen International Expert Consensus Conference. Ann Oncol 18(Suppl 6):vi63–vi65.

Courjal F, Cuny M, Simony-Lafontaine J, Louason G, Speiser P, Zeillinger R, Rodriguez C, Theillet C. 1997. Mapping of DNA amplifications at 15 chromosomal localizations in 1875 breast tumors: Definition of phenotypic groups. Cancer Res 57:4360–4367.

Cuny M, Kramar A, Courjal F, Johannsdottir V, Iacopetta B, Fontaine H, Grenier J, Culine S, Theillet C. 2000. Relating genotype and phenotype in breast cancer: An analysis of the prognostic significance of amplification at eight different genes or loci and of p53 mutations. Cancer Res 60:1077–1083.

Dellas A, Torhorst J, Schultheiss E, Mihatsch MJ, Moch H. 2002. DNA sequence losses on chromosomes 11p and 18q are associated with clinical outcome in lymph node-negative ductal breast cancer. Clin Cancer Res 8:1210–1216.

Elston C, Ellis I. 1991. Pathological prognostic factors in breast cancer. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. Histopathology 19:403–410.

Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H, Chen H, Omeroglu G, Meterissian S, Omeroglu A, Meterissian S, Omeroglu A, Hallett M, Park M. 2008. Stromal gene expression predicts clinical outcome in breast cancer. Nat Med 14:518–527.

Hupe P, Stransky N, Thiery JP, Radvanyi F, Barillot E. 2004. Analysis of array CGH data: From signal ratio to gain and loss of DNA regions. Bioinformatics 20:3413–3422.

Idbaih A, Marie Y, Pierron G, Brennetot C, Hoang-Xuan K, Kujas M, Mokhtari K, Sanson M, Lejeune J, Aurias A, Delattre O, Delattre JY. 2005. Two types of chromosome 1p losses with opposite significance in gliomas. Ann Neurol 58:483–487.

Janoueix-Lerosey I, Schleiermacher G, Michels E, Mosseri V, Ribeiro A, Lequin D, Vermeulen J, Couturier J, Peuchmaur M, Valent A, Plantaz D, Rubie H, Valteau-Couanet D, Thomas C, Combaret V, Rousseau R, Eggert A, Michon J, Speleman F, Delattre O. 2009. Overall genomic pattern is a predictor of outcome in neuroblastoma. J Clin Oncol 27:1026–1033.

Kujawski L, Ouillette P, Erba H, Saddler C, Jakubowiak A, Kaminski M, Shedden K, Malek SN. 2008. Genomic

complexity identifies patients with aggressive chronic lymphocytic leukemia. Blood 112:1993–2003.

La Rosa P, Viara E, Hupe P, Pierron G, Liva S, Neuvial P, Brito I, Lair S, Servant N, Robine N, Manie E, Brennetot C, Janoueix-Lerosey I, Raynal V, Gruel N, Rouveirol C, Stransky N, Stern MH, Delattre O, Aurias A, Radvanyi F, Barillot E. 2006. VAMP: Visualization and analysis of array-CGH, transcriptome and other molecular profiles. Bioinformatics *22*: 2066–2073.

Langerod A, Zhao H, Borgan O, Nesland JM, Bukholm IR, Ikdahl T, Karesen R, Borresen-Dale AL, Jeffrey SS. 2007. TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer. Breast Cancer Res 9:R30.

Marchio C, Iravani M, Natrajan R, Lambros MB, Savage K, Tamber N, Fenwick K, Mackay A, Senetta R, Di Palma S, Schmitt FC, Bussolati G, Ellis LO, Ashworth A, Sapino A, Reis-Filho JS. 2008. Genomic and immunophenotypical characterization of pure micropapillary carcinomas of the breast. J Pathol 215:398–410.

Natrajan R, Lambros MB, Rodriguez-Pinilla SM, Moreno-Bueno G, Tan DS, Marchio C, Vatcheva R, Rayter S, Mahler-Araujo B, Fulford LG, Hungermann D, Mackay A, Grigoriadis A, Fenwick K, Tamber N, Hardisson D, Tutt A, Palacios J, Lord CJ, Buerger H, Ashworth A, Reis-Filho JS. 2009. Tiling path genomic profiling of grade 3 invasive ductal breast cancers.2009.Clin Cancer Res 15:2711–2722.

Neuvial P, Hupe P, Brito I, Liva S, Manie E, Brennetot C, Radvanyi F, Aurias A, Barillot E. 2006. Spatial normalization of array-CGH data. BMC Bioinformatics 7:264.

Olivier M, Langerod A, Carrieri P, Bergh J, Klaar S, Eyfjord J, Theillet C, Rodriguez C, Lidereau R, Bieche I, Varley J, Bignon Y, Uhrhammer N, Winqvist R, Jukkola-Vuorinen A, Niederacher D, Kato S, Ishioka C, Hainaut P, Borresen-Dale AL. 2006. The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. Clin Cancer Res 12:1157–1167.

Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher E, Wickerham D, Bryant J, Wolmark N. 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351:2817–2826.

Pekarsky Y, Zanesi N, Palamarchuk A, Huebner K, Croce CM. 2002. FHIT: From gene discovery to cancer treatment and prevention. Lancet Oncol 3:748–754.

Reis-Filho JS, Simpson PT, Turner NC, Lambros MB, Jones C, Mackay A, Grigoriadis A, Sarrio D, Savage K, Dexter T, Iravani M, Fenwick K, Weber B, Hardisson D, Schmitt FC, Palacios J, Lakhani SR, Ashworth A. 2006. FGFR1 emerges as a potential therapeutic target for lobular breast carcinomas. Clin Cancer Res 12:6652–6662.

Rennstam K, Ahlstedt-Soini M, Baldetorp B, Bendahl PO, Borg A, Karhu R, Tanner M, Tirkkonen M, Isola J. 2003. Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis. A study of 305 tumors by comparative genomic hybridization. Cancer Res 63:8861–8868.

Reyal F, van Vliet MH, Armstrong NJ, Horlings HM, de Visser KE, Kok M, Teschendorff AE, Mook S, van 't Veer L, Caldas C, Salmon RJ, van de Vijver MJ, Wessels LF. 2008. A comprehensive analysis of prognostic signatures reveals the high predictive capacity of the proliferation, immune response and RNA splicing modules in breast cancer. Breast Cancer Res 10:R93.

Roylance R, Gorman P, Papior T, Wan YL, Ives M, Watson JE, Collins C, Wortham N, Langford C, Fiegler H, Carter N, Gillett C, Sasieni P, Pinder S, Hanby A, Tomlinson I. 2006. A comprehensive study of chromosome 16q in invasive ductal and lobular breast carcinoma using array CGH. Oncogene 25:6544–6553.

Schlotter CM, Vogt U, Bosse U, Mersch B, Wassmann K. 2003. C-myc, not HER-2/neu, can predict recurrence and mortality of patients with node-negative breast cancer. Breast Cancer Res 5:R30–R36.

Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, Levin WJ, Stuart SG, Udove J, Ullrich A. 1989. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science 244:707–712.

Thor AD, Eng C, Devries S, Paterakos M, Watkin WG, Edgerton S, Moore DH 2nd, Etzell J, Waldman FM. 2002. Invasive micropapillary carcinoma of the breast is associated with chromosome 8 abnormalities detected by comparative genomic hybridization. Hum Pathol 33:628–631.

Tsuda H, Sakamaki C, Tsugane S, Fukutomi T, Hirohashi S. 1998. Prognostic significance of accumulation of gene and chromosome alterations and histological grade in node-negative breast carcinoma. Jpn J Clin Oncol 28:5–11.

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530–536.

Vincent-Salomon A, Gruel N, Lucchesi C, MacGrogan G, Dendale R, Sigal-Zafrani B, Longy M, Raynal V, Pierron G, de Mascarel I, Taris C, Stoppa-Lyonnet D, Pierga JY, Salmon R, Sastre-Garau X, Fourquet A, Delattre O, de Cremoux P, Aurias A. 2007. Identification of typical medullary breast carcinoma as a genomic sub-group of basal-like carcinomas, a heterogeneous new molecular entity. Breast Cancer Res 9:R24.

Vincent-Salomon A, Raynal V, Lucchesi C, Gruel N, Delattre O. 2008. ESR1 gene amplification in breast cancer: A common phenomenon? Nat Genet 40:809.

Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EM, Atkins D, Foekens JA. 2005. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 365:671–679.

Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, Dowsett M, Fitzgibbons PL, Hanna WM, Langer A, McShane LM, Paik S, Pegram MD, Perez EA, Press MF, Rhodes A, Sturgeon C, Taube SE, Tubbs R, Vance GH, van de Vijver M, Wheeler TM, Hayes, DF. 2007. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. Arch Pathol Lab Med 131:18.

Zhang Y, Martens JW, Yu JX, Jiang J, Sieuwerts AM, Smid M, Klijn JG, Wang Y, Foekens JA. 2009. Copy number alterations that predict metastatic capability of human breast cancer. Cancer Res 69:3795–37801.