Cancer Research

AACR

Copy Number Alterations that Predict Metastatic Capability of Human Breast Cancer

Yi Zhang, John W.M. Martens, Jack X. Yu, et al.

Updated Version	Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-08-4596
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2009/04/01/0008-5472.CAN-08-4596.DC1.html

Cited Articles	This article cites 27 articles, 11 of which you can access for free at: http://cancerres.aacrjournals.org/content/69/9/3795.full.html#ref-list-1
Citing Articles	This article has been cited by 3 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/69/9/3795.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.
Permissions	To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.

Copy Number Alterations that Predict Metastatic Capability of Human Breast Cancer

Yi Zhang,[1] John W.M. Martens,[2] Jack X. Yu,[1] John Jiang,[1] Anieta M. Sieuwerts,[2] Marcel Smid,[2] Jan G.M. Klijn,[2] Yixin Wang,[1] and John A. Foekens[2]

[1]Veridex LLC, a Johnson & Johnson Company, San Diego, California and [2]Department of Medical Oncology, Erasmus MC, Josephine Nefkens Institute and Cancer Genomics Centre, Rotterdam, the Netherlands

Abstract

We have analyzed the DNA copy numbers for over 100,000 single-nucleotide polymorphism loci across the human genome in genomic DNA from 313 lymph node–negative primary breast tumors for which genome-wide gene expression data were also available. Combining these two data sets allowed us to identify the genomic loci and their mapped genes, having high correlation with distant metastasis. An estimation of the likely response based on published predictive signatures was performed in the identified prognostic subgroups defined by gene expression and DNA copy number data. In the training set of 200 patients, we constructed an 81-gene prognostic copy number signature (CNS) that identified a subgroup of patients with increased probability of distant metastasis in the independent validation set of 113 patients [hazard ratio (HR), 2.8; 95% confidence interval (95% CI), 1.4–5.6] and in an external data set of 116 patients (HR, 3.7; 95% CI, 1.3–10.6). These high-risk patients constituted a subset of the high-risk patients predicted by our previously established 76-gene gene expression signature (GES). This very poor prognostic group identified by CNS and GES was putatively more resistant to preoperative paclitaxel and 5-fluorouracil-doxorubicin-cyclophosphamide combination chemotherapy ($P = 0.0048$), particularly against the doxorubicin compound, while potentially benefiting from etoposide. Our study shows the feasibility of using copy number alterations to predict patient prognostic outcome. When combined with gene expression–based signatures for prognosis, the CNS refines risk classification and can help identify those breast cancer patients who have a significantly worse outlook in prognosis and a potential differential response to chemotherapeutic drugs. [Cancer Res 2009;69(9):3795–801]

Introduction

Specific DNA copy number alterations (CNA), such as deletions and amplifications, are major genomic alterations that contribute to carcinogenesis and tumor progression through reduced apoptosis, unchecked proliferation, increased motility, and angiogenesis (1–3). Because a significant proportion of genomic aberrations are unrelated to cancer biology and merely due to random neutral events (4), it is a challenge to identify those causative gene CNAs that are responsible for gene expression regulation, which ultimately leads to malignant transformation and progression. Both fluorescence *in situ* hybridization and comparative genomic hybridizations have revealed chromosomal regions that showed CNAs in breast tumors. In a recent study, including 51 breast tumors, a high-resolution single-nucleotide polymorphism (SNP) array was used together with gene expression profiling to refine breast cancer amplicon boundaries and narrow the list of potential driver genes (5). However, only a limited number of studies (1, 6–13) investigated the CNAs in relation to their prognostic significance, whereas the sample sizes of these studies were too small to draw firm conclusions. In addition, fewer studies (1, 6, 12–14) investigated breast cancer prognosis using combined analysis of CNAs and gene expression profiling with sufficient sample size and a technology that had appropriate coverage and mapping resolution of the human genome.

In the present study, we used a high-throughput and high-resolution oligonucleotide-based SNP array technology to analyze the CNAs for >100,000 SNP loci in the breast cancer genome. In a large cohort of 200 lymph node–negative breast cancer patients, we identified copy number alterations that were correlated with the time in developing distant metastasis. The prognostic power of the CNAs was validated in two independent patient cohorts. In addition, using published predictive gene signatures, the identified patient subgroups with different prognosis were tested for putative drug efficacy. The results from our study suggest that combining DNA copy number analysis and gene expression analysis provides an additional and better means for risk assessment for breast cancer patients.

Materials and Methods

Patient and tumor material. Frozen tumor specimens of 313 lymph node–negative breast cancer patients selected from the tumor bank at the Erasmus Medical Center were used in this study. These patients were treated during the period 1980 to 1995, but none of these patients did receive any systemic (neo)adjuvant therapy. Patients who first developed a local recurrence before distant metastasis were not included to avoid the possibility that the distant metastasis later on could have originated from the local recurrence. The guidelines for local primary treatment were the same. Among these specimens, 273 were used to develop a 76-gene signature for the prediction of distant metastasis using Affymetrix U133A chips (15). The remaining 40 patients were used to study prognostic biological pathways (16). The study was approved by the medical ethics committee of Erasmus MC (MEC 02.953) and conducted in accordance to the code of conduct of Federation of Medical Scientific Societies in the Netherlands[3], and whenever possible, we adhered to the Reporting Recommendations for Tumor Marker Prognostic Studies REMARK (17).

---

[3] http://www.fmwv.nl/

Table 1. Clinical and pathologic characteristics of patients and their tumors

Characteristics	All patients (n = 313)	Training set (n = 200)	Validation set (n = 113)	External validation set (n = 116)
Age at surgery, y				
Mean (SD)	54 (12)	54 (12)	54 (12)	57 (10)
≤40	45 (14%)	30 (15%)	15 (13%)	6 (5%)
41–55	134 (43%)	84 (42%)	50 (44%)	41 (35%)
56–70	98 (31%)	62 (31%)	36 (32%)	68 (59%)
>70	36 (12%)	24 (12%)	12 (11%)	1 (1%)
Menopausal status				
Premenopausal	152 (49%)	96 (48%)	56 (50%)	38 (33%)
Postmenopausal	161 (51%)	104 (52%)	57 (50%)	78 (67%)
T stage				
T1	153 (49%)	97 (49%)	56 (49%)	90 (78%)
T2	148 (47%)	95 (47%)	53 (47%)	26 (22%)
T3/4	11 (4%)	8 (4%)	3 (3%)	0
Unknown	1 (0%)	0	1 (1%)	0
Grade				
Poor	165 (53%)	111 (56%)	54 (48%)	48 (42%)
Moderate	45 (14%)	29 (14%)	16 (14%)	34 (29%)
Good	6 (2%)	3 (2%)	3 (3%)	34 (29%)
Unknown	97 (31%)	57 (28%)	40 (35%)	0
ER status				
Positive	199 (64%)	133 (67%)	66 (58%)	79 (68%)
Negative	114 (36%)	67 (33%)	47 (42%)	37 (32%)
Progesterone receptor status				
Positive	156 (50%)	100 (50%)	56 (50%)	NA
Negative	148 (47%)	92 (46%)	56 (50%)	NA
Unknown	9 (3%)	8 (4%)	1 (1%)	NA
Metastasis within 5 y				
Yes	99 (32%)	64 (32%)	35 (31%)	8 (7%)
No	204 (65%)	127 (64%)	77 (68%)	104 (90%)
Censored	10 (3%)	9 (4%)	1 (1%)	4 (3%)
Adjuvant systemic therapy				
Yes	0	0	0	43 (37%)
No	313 (100%)	200 (100%)	113 (100%)	71 (61%)
Unknown	0	0	0	2 (2%)

NOTE: Grade was assessed by regional pathologists and reflects the current practice during the years the tumors were collected. ER positive and progesterone receptor positive indicate >10 fmol/mg protein or >10% positive tumor cells.
Abbreviation: NA, not available.

One hundred ninety-nine tumors were classified as estrogen receptor (ER) positive and 114 as ER negative, using previously described ER (and progesterone receptor) cutoffs (15). Median age of patients at the time of surgery (breast conserving surgery, 230 patients; modified radical mastectomy, 83 patients) was 54 y (range, 26–83 y). The median follow-up time for surviving patients (n = 220) was 99 mo (range, 20–169 mo). A total of 114 patients (36%) developed distant metastasis and were counted as failures in the analysis of distant metastasis-free survival (MFS). Of the 93 patients who died, seven died without evidence of disease and were censored at last follow-up in the analysis of distant MFS; 86 patients died after a previous relapse. The clinicopathologic characteristics of the patients are given in Table 1. The data set containing the clinical and SNP data has been submitted to Gene Expression Omnibus database with accession number 10099.[4]

The external array comparative genomic hybridization data set of 116 lymph node–negative patients used in this study (6) as an independent validation was downloaded.[5] The clinical data (Table 1) related to this data set were kindly provided by Dr. Teschendorff of University of Cambridge.

DNA isolation, hybridization, and DNA copy number analysis. The methods used to isolate DNA from breast tumor samples and for hybridization of DNA to the Affymetrix GeneChip Human Mapping 100K Array are described in detail in Supplementary Materials and Methods (online only).

Identification of prognostic chromosome regions, construction, and validation of copy number signature. We designed an integrated analytic method to identify the chromosome regions and the mapped candidate genes whose CNAs were correlated with distant metastasis by taking advantage of the availability of the genomic data on both RNA gene expression, which were generated from our previous studies (15, 16), and DNA copy number from the same cohort of patients that became available in this study (Fig. 1). Our method is very similar in principle to the approach that Adler and colleagues (14) took and described as stepwise linkage analysis of microarray signatures to identify genetic regulators of expression signatures by intersecting genome-wide DNA copy number and gene expression data. We analyzed ER-positive and ER-negative patients separately and randomly split the patients in an approximate of 2:1 ratio into a training set of 200 patients and a validation set of 113 patients (Fig. 1) while balancing on the clinical and pathologic parameters, including T stage, grade, menopausal status, and recurrences. The training set was used to identify prognostic chromosome regions and mapped genes and construct a copy number signature (CNS) to predict distant metastasis;

---

the validation set was set aside solely for validation purpose. The analytic details for the identification of chromosome regions with prognostic CNAs and construction and validation of CNS are described in Supplementary Materials and Methods (online only).

Putative response to chemotherapy. To estimate the putative responses of the validation set of patients to chemotherapeutic compounds, gene expression signatures (GES) in two published studies were used (18, 19). The original gene expression data set and the *R* function for the prediction algorithm of diagonal linear discriminant analysis for the 30-gene preoperative paclitaxel and 5-fluorouracil-doxorubicin-cyclophosphamide (T/FAC) response signature was downloaded[6] (19). Because the original authors did not provide the necessary model parameters to use the algorithm directly, the model was trained from the original data set using the provided *R* function and then validated in our gene expression data set. For each of the seven GES that predict sensitivity to individual chemotherapeutic drugs, the predicted probability of sensitivity to each compound using the Bayesian fitting of binary probit regression models was calculated with the help of Drs. Anil Potti and Joseph Nevins (for details, see ref. 18).

Statistical analysis. Unsupervised analysis using principal component analysis was performed on the copy number data set with all SNPs to examine the potential subclasses of the tumors. Kaplan-Meier survival plots (20) and log-rank tests were used to assess the differences in MFS of the predicted high-risk and low-risk groups. Cox proportional hazard regression was performed to compute the hazard ratio (HR) and its 95% confidence interval (95% CI). Due to missing data on grade, multivariate Cox regression analysis was done by multiple imputations using the Markov Chain Monte Carlo method under the general location model (21). Dunnett's tests in the context of ANOVA were performed to assess the significance of differential therapeutic responses between the very poor prognostic group and each of the good/poor prognostic groups while controlling the type I error. All tests of statistical significance were two sided. All statistical analyses were performed using *R* version 2.6.2.

Results

Identification of prognostic chromosomal regions. The median copy number estimate of the copy number data set, calculated as the median of the means of each SNP copy number estimate across all SNPs, which was computed as the average of the SNP interquartile (middle, 50%) copy number estimates, was 2.1, consistent with the general assumption that the majority of the genome is diploid. Unsupervised analysis using principle component analysis on all 313 tumors showed that chromosomal copy number variations were clearly different for ER-positive and ER-negative tumors (Supplementary Fig. S1). Therefore, these two types of breast tumors not only differ on global gene expression profiles, as indicated in many studies before (15, 22–24), but also have distinct chromosomal variations on the DNA level. Therefore, it is necessary that subsequent analysis be performed separately for ER-positive and ER-negative tumors. Furthermore, we randomly divided the patients into a training set of 200 patients (133 for ER-positive and 67 for ER-negative tumors) and a validation set of 113 patients (66 for ER-positive and 47 for ER-negative tumors; Table 1; Fig. 1) in an approximate of 2:1 ratio. The training set was used to identify prognostic chromosome regions and the mapped genes and construct a CNS to predict distant metastasis; the validation set was set aside solely for validation purposes.

First, we identified chromosome regions whose CNAs were correlated with the MFS of patients. For ER-positive tumors, 45 chromosomal regions distributed over 17 chromosomes were identified as having CNAs that correlated with MFS; for ER-negative

tumors, there were 56 regions distributed over 19 chromosomes (Fig. 2). The total of these region sizes for ER-positive and ER-negative tumors were 521 (Supplementary Table S1) and 496 Mb (Supplementary Table S2), respectively. The prognostic chromosomal regions identified from the ER-positive tumors share limited similarities with those from the ER-negative tumors (Fig. 2).

Search for prognostic candidate genes to construct CNS. The gene expression profiling data from our previous studies of the same tumors were used (15, 16) to screen for genes that had consistent change patterns between gene expression profiles and copy number variations. We reasoned that the change in copy numbers has to be reflected in the corresponding change in gene expression levels to have a phenotypic effect. Within these prognostic regions, a total of 2,833 and 3,656 genes were mapped for ER-positive tumors (Supplementary Table S1) and ER-negative tumors (Supplementary Table S2), respectively. For the ER-positive tumors, 122 genes had significant Cox regressions (*P* < 0.05) in both gene expression data and copy number data and showed the same direction for the changes in DNA copy number and gene expression. For the ER-negative tumors, 78 genes had significant *P* values in both data sets and showed the same direction of alterations (Supplementary Fig. S2). Of these, 53 genes (43%) for ER-positive and 28 genes (36%) for ER-negative tumors, respectively, had correlation coefficients between gene expression and copy number of >0.5. Thus, in total, 81 prognostic candidate genes were identified, which were then used as a CNS for prognosis (Table 2 and Supplementary Table S3).

Validation of CNS. Validation was done in the independent validation set of 66 ER-positive and 47 ER-negative tumors separately using 53 and 28 genes from CNS, respectively. The HR and 95% CI for time to distant metastasis of patients with a poor CNS compared with a good CNS were 2.8 (1.3–6.3; *P* = 0.0088) for ER-positive and 8.7 (1.1–74.4; *P* = 0.0166) for ER-negative tumors, respectively. The Kaplan-Meier analyses of the combined two patient groups stratified by the 81-gene CNS showed a statistically significant difference in time to distant metastasis (Fig. 3*A*) with a HR of 2.8 (*P* = 0.0036). The estimated rate of distant metastasis at 5 years for the two groups was 27% (95% CI, 17–35%) and 67% (95% CI, 32–84%), respectively. We chose not to further stratify the patients by other clinical variables because the subgroups would become too small to allow statistically justifiable conclusions. When used in conjunction with our previously identified (15) and independently validated 76-gene GES (25–27), the patient group with worse prognosis outcome defined by the 81-gene CNS remained the same with 67% of estimated distant metastasis at 5 years. The 76-gene GES stratified the other patient group with better prognosis further to good and poor prognosis groups with the 5-year estimated rate of recurrence at 11% and 37%, respectively (Fig. 3*B*). This result led to three prognostic groups, which we defined as good, poor, and very poor groups for GES good/CNS good, GES poor/CNS good, and GES poor/CNS poor groups, respectively. Multivariate Cox regression analysis of both signatures, together with traditional clinical and pathologic factors, showed that the combination of the two signatures was the only significant (likelihood ratio test, *P* = 0.0003) prognostic factor for MFS, with HR of 8.86 comparing the very poor versus good prognostic groups, and 3.59 for comparison of the poor versus the good prognostic groups (Table 3). The patients in the very poor prognosis group are not significantly different from the good and poor prognosis groups with respect to the traditional clinical variables: age, T stage, grade, menopause status, and progesterone

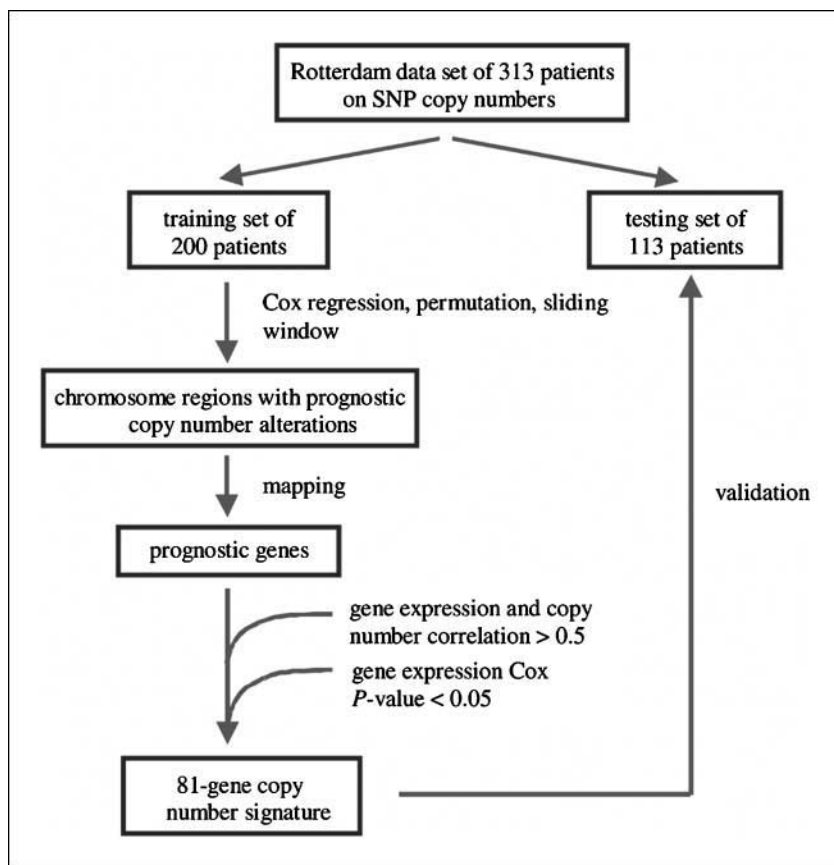**Figure 1.** The analysis workflow to identify the genes (SNPs) with prognostic copy number alterations and the establishment of CNS.

receptor status, except ER status. ER status, however, could not be used alone to identify the patients in the very poor prognostic group. For example, 14 of the 66 ER-positive patients were in the very poor prognostic group whereas 52 ER-positive patients were in good or poor prognostic groups. In the analysis for the 10-year overall survival, 84% (95% CI, 70–99%) of the patients in the good prognosis group were alive after 10 years compared with 55% (95% CI, 41–74%) and 27% (95% CI, 10–78%) in the poor and very poor prognosis groups, respectively ($P = 0.0005$).

Next, we validated the CNS in a completely independent external data set of 116 lymph node–negative patients (79 ER-positive and 37 ER-negative tumors) derived from a lower resolution array comparative genomic hybridization technology (6). The 81-gene CNS significantly stratified this patient cohort (Fig. 3*C*) into two prognostic groups with an HR of 3.7 ($P = 0.0102$) and remained to be the only significant prognosticator in a multivariate Cox regression analysis, including age, tumor size, grade, and ER status ($P = 0.0150$). The lower rate of distant metastasis at 5 years (19%) for the poor prognostic group compared with that of our own data set was likely due to the smaller tumor sizes (78% smaller than 2 cm) and the fact that over one third of the patients had received adjuvant hormone and/or chemotherapy in this cohort (Table 1).

Response to chemotherapy. We subsequently investigated the chemotherapy response profiles of the three prognostic groups determined by the GES and CNS prognostic assays using well-validated gene signatures derived from two studies (18, 19) for which follow-up validation studies in human clinical samples were also available (28, 29). Firstly, using a previously published 30-gene signature that predicted a pathologic complete response to

preoperative T/FAC chemotherapy (19), we assigned each patient in the different prognostic subgroups into two response groups, either as having pathologic complete response or still with residual disease. Only 2 of the 15 patients (13%) in the very poor prognostic group were predicted as having pathologic complete response, whereas 34 of the 60 patients (57%) and 14 of the 38 patients (37%) in the poor and good prognostic groups, respectively, were predicted as having pathologic complete response. The chemo-response score for the very poor prognostic group was significantly lower than those of the poor prognostic group ($P = 0.0048$), indicating that these patients would be much more resistant to preoperative T/FAC chemotherapy in case these patients would have received this preoperative combination chemotherapy (Supplementary Fig. S3). Secondly, we determined the response profiles of the three prognostic groups against seven individual chemotherapeutic compounds using expression signatures established on cell lines (18). For each compound, we calculated the predicted probability of sensitivity to the compound (Supplementary Fig. S3) using the Bayesian fitting of binary probit regression models (18). Compared with the poor prognostic group, the patients in the very poor prognostic group had significantly lower mean sensitivity score, i.e., they were more resistant to doxorubicin ($P = 0.0037$). On the other hand, the very poor prognosis group seemed to be more sensitive to etoposide ($P = 0.0359$) and, although not statistically significant, to topotecan ($P = 0.0542$). Thus, when combined with gene expression–based signatures for prognosis and therapy prediction, CNAs measured by SNP arrays improve risk classification and can identify those breast cancer patients who have a significantly worse outlook in prognosis and a potential differential response to chemotherapeutic drugs.

Discussion

In this study, we have performed a combined analysis of DNA copy number and gene expression on a large cohort of 313 lymph node–negative breast cancer patients who received no adjuvant systemic therapy. To our knowledge, this is the largest study to analyze CNAs for breast cancer prognosis using the high-density SNP array technology that has much higher resolution than array comparative genomic hybridization. We identified, from a training set of 200 lymph node–negative patients, a signature of 81 genes that showed CNAs and concordant gene expression regulation and validated this CNS in the independent 113 lymph node–negative patients, as well as in an external array comparative genomic hybridization data set of 116 lymph node–negative patients. We also showed that applying CNS, in addition to GES, in risk classification for the prognosis of breast cancer patients is clearly

improved, particularly in the poor prognostic patients predicted by the 76-gene GES alone. Although the very poor prognosis group, defined as patients in the poor prognosis group according to GES and CNS, constitutes only 13% of all patients, we consider this group of patients clinically relevant because of very poor overall survival. Furthermore, in the concept of personalized medicine, other patient groups similar in size (basal, triple negative, and HER2 breast cancer subtypes) have been and are attracting major attention. Our view is that the clinical utility of the combination of GES and CNS may be better assessed with positive predictive value (PPV) and negative predictive value (NPV) using 5-year distant metastasis as the defining end point. Because PPV and NPV can only be calculated for binary classifications, we calculated two separate PPVs for the patients predicted as having either very poor or very poor plus poor prognosis by either CNS or GES,

Figure 2. The chromosomal regions with prognostic copy number alterations. *A*, the smoothed Cox regression *P*-value (log$_{10}$ scale) curves along the 23 chromosomes for ER-positive (*top*) and ER-negative (*bottom*) tumors. The log$_{10}$ smoothed *P* values were signed by the corresponding signs of the log scores *S*. The red color curve indicates that the copy number was higher in the relapsed tumors than in the relapse-free tumors; the blue color curve indicates the opposite. The horizontal dotted line indicates nominal significance at *P* value cutoffs (*P* = 0.05). *B*, the ideogram of the human genome labeled by the chromosomal regions for ER-positive (*top edge of each chromosome*) and ER-negative (*bottom edge*) tumors of which copy number alterations were prognostic. *Red and blue*, regions of gains and losses, respectively, in patients who developed distant metastasis.
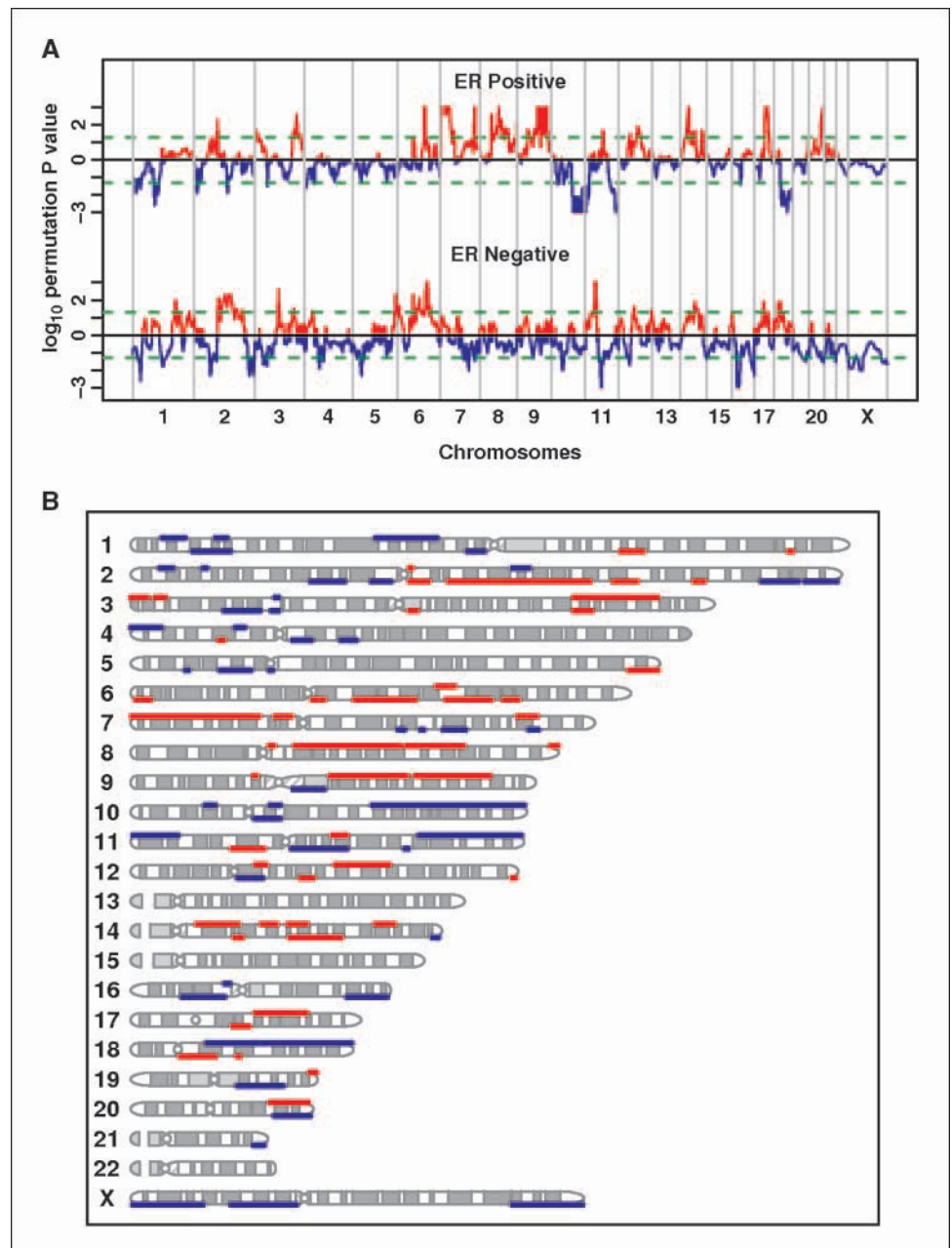
Table 2. Description of the 81 genes used as CNS

	Prognostic genes with copy number alteration
Gain in ER+ tumors	SMC4, PDCD10, PREP, CBX3, NUP205, TCEB1, TERF1, TPD52, GGH, TRAM1, ZBTB10, YTHDF3, EIF3E, POLR2K, RPL30, CCNE2, RAD54B, MTERFD1, ENY2, DPY19L4, ZNF623, SCRIB, SLC39A4, ATP6V1G1, PSMA6, STRN3, CLTC, TRIM37, NME1, NME2, RPS6KB1, PPM1D, MED13, SLC35B1, APPBP2, MKS1, C17orf71, HEATR6, TMEM49, USP32, ANKRD40, NME1-NME2, ZNF264, ZNF304, ATP5E, CSTF1, PPP1R3D, AURKA, RAE1, STX16, C20orf43, RAB22A
Loss in ER+ tumors	TCTN3
Gain in ER− tumors	C1orf9, COX5B, EIF5B, DDX18, TSN, p20, METTL5, MGAT1, TUBB2A, RWDD1, PGM3, FOXO3, CDC40, REV3L, HDAC2, TSPYL4, C6orf60, ASF1A, MED23, TSPYL1, ACTR10, KIAA0247, RARA, KRT10, RIOK3, IMPACT
Loss in ER− tumors	HDAC1, BSDC1

respectively, because these patients should be treated with adjuvant systemic therapies; on the other hand, we only calculated one NPV for the good prognostic group by the combination of GES and CNS because we believe that the patients in this group had such a good prognosis that adjuvant systemic therapies might be withheld. The PPV for the very poor prognostic group stratified by CNS alone (i.e., the very poor prognostic group predicted by the combination of GES and CNS) is 67% compared with 42% PPV for the poor prognostic group stratified by GES alone and 30% PPV when no signature was used at all as the standard practice today. Therefore, this could lead to a potential 37% (70−33%) reduction of unnecessary treatment upon further validation. The NPV for the good prognostic group predicted that the GES and CNS combination is 90%, indicating that ∼10% of the patients in this good prognostic group may have their treatment erroneously withheld if this is the only information used for decision making. Further improvements on NPV are desired.

Furthermore, by using previously reported gene signature profiles for sensitivity to chemotherapeutic compounds, it was shown that this very poor prognostic group might be much more resistant to preoperative T/FAC combination chemotherapy, particularly against the doxorubicin compound while benefiting from etoposide. If confirmed in independent studies, this may suggest that patients belonging to this category might benefit from different chemotherapy regimens compared with other patient groups and that the 81 genes of the CNS might be used to determine chemosensitivity.

Previous studies investigating the association between gene amplification and breast cancer prognosis considered different breast cancer subtypes, such as ER positive and ER negative, as a single homogenous cohort. However, it is well known that these tumors are pathologically and biologically very different, evidenced by tremendous distinct global gene expression profiles (15, 22–24). In this study, we showed that this dichotomy also extended to the global pattern of the DNA copy numbers. Therefore, the analysis needed to be performed separately for ER-positive and

ER-negative tumors. Indeed, the prognostic chromosomal regions identified from the ER-positive tumors share limited similarities with those from the ER-negative tumors. For example, chromosome region 8q is a widely known site of DNA amplification associated with poor prognosis in breast cancer (7, 9–11). Our results showed that 8q was indeed a hotspot for amplification in ER-positive tumors but contained no significant amplified areas for ER-negative tumors. Because ER-negative tumors constitute only a small percentage (∼25%) of the lymph node–negative breast cancers, it is reasonable to speculate that those studies that did not separate the two types of breast tumors in their analysis may have had their conclusions overwhelmed by the results from the
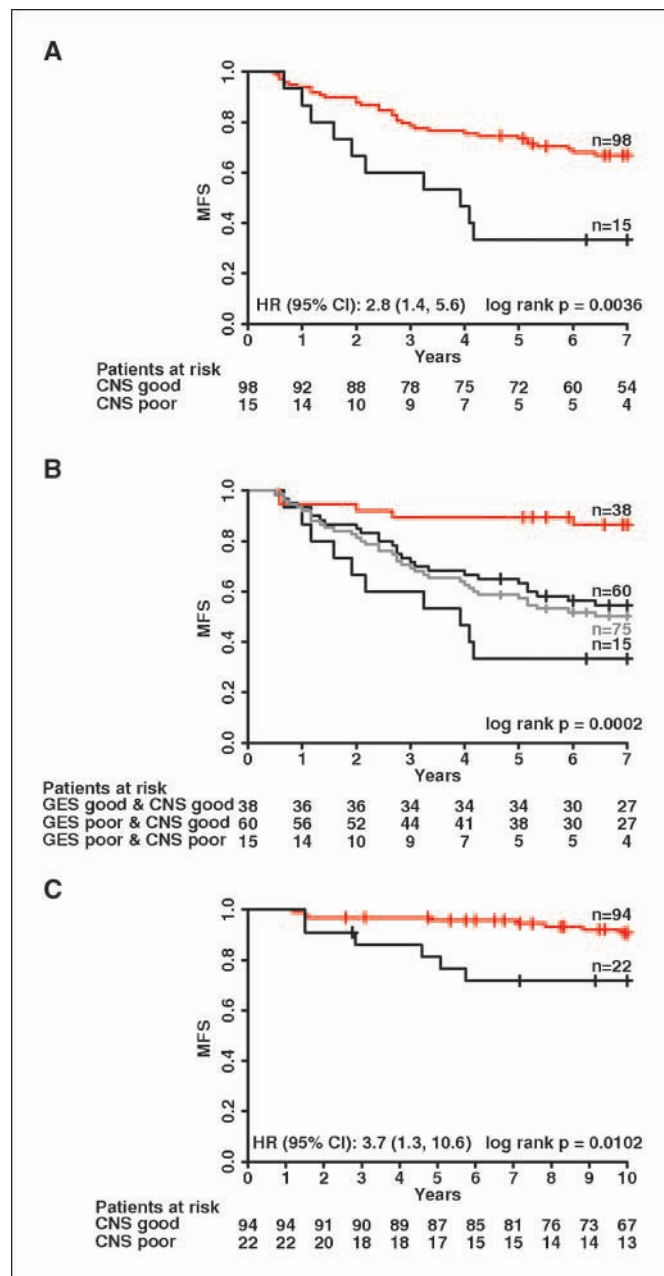

Figure 3. Distant MFS as a function of CNS. *A,* lymph node–negative validation set patients stratified by the 81-gene CNS. *B,* lymph node–negative validation set patients stratified by the 76-gene GES and CNS. *Gray line,* poor prognostic group defined by GES alone. *C,* lymph node–negative patients of the external validation set (6) stratified by CNS.

Table 3. Multivariate Cox regression analysis of GES and CNS combination

	Multivariate analysis		
	HR	95% CI	*P*
Age (per 10-y increment)	0.77	0.48–1.22	0.2573
Post versus premenopausal	1.34	0.45–3.97	0.5920
Grade 1 and grade 2 versus grade 3	0.45	0.17–1.19	0.1060
Tumor size of >20 mm versus ≤20 mm	1.02	0.54–1.92	0.9583
ER negative versus ER positive	1.07	0.52–2.19	0.8590
Gene expression signature and CNS combination			
Poor versus good	3.59	1.35–9.49	0.0102
Very poor versus good	8.86	2.76–28.4	0.0002

majority of the samples of ER-positive tumors. Another apparent difference between the two types of tumors observed from our analysis was at chromosome region 20q13.2-13.3. A gain in copy number of this region in ER-positive tumors, but by contrast a loss in copy number of this region in ER-negative tumors, was related to an early recurrence. Taken together, these results reemphasize that ER-positive and ER-negative tumors follow different biological pathways for cancer development and progression.

In summary, our study identified a panel of 81 genes based on two-dimensional evidence and showed that the copy number of these genes drives their transcriptional regulation, yielding a cascade of downstream genetic changes that ultimately result in breast tumor progression. Because of the high correlation between the copy number and gene expression level of the 81 genes of the CNS, our data provided initial evidence that the 81 genes might function as candidate oncogenes or tumor suppressor genes, which deserves further in-depth experimental investigation. Our study also shows the feasibility of using DNA alterations as a prognostic assay to predict patient outcome. When combined with gene expression–based signatures for prognosis and therapy prediction, CNAs measured by SNP arrays improve risk classification and can identify those patients who have a worse outlook in prognosis and a potential differential response to chemotherapeutic drugs. Regarding the latter, the limitation of our study lies in the fact that we were only able to assess the putative prediction to treatment as based on published signatures and not the actual efficacy of chemotherapy because the patients in the study did not receive adjuvant chemotherapy.

Disclosure of Potential Conflicts of Interest

Acknowledgments

References

1. Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 2006;10:529–41.
2. Albertson DG, Collins C, McCormick F, Gray JW. Chromosome aberrations in solid tumors. Nat Genet 2003;34:369–76.
3. Knuutila S, Autio K, Aalto Y. Online access to CGH data of DNA sequence copy number changes. Am J Pathol 2000;157:689.
4. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. Nature 2006;444:444–54.
5. Haverty PM, Fridlyand J, Li L, et al. High-resolution genomic and expression analyses of copy number alterations in breast tumors. Genes Chromosomes Cancer 2008;47:530–42.
6. Chin SF, Teschendorff AE, Marioni JC, et al. High-resolution array-CGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. Genome Biol 2007;8:R215.
7. Al-Kuraya K, Schraml P, Torhorst J, et al. Prognostic relevance of gene amplifications and coamplifications in breast cancer. Cancer Res 2004;64:8534–40.
8. Cheng KW, Lahad JP, Kuo WL, et al. The RAB25 small GTPase determines aggressiveness of ovarian and breast cancers. Nat Med 2004;10:1251–6.
9. Rennstam K, Ahlstedt-Soini M, Baldetorp B, et al. 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 2003;63:8861–8.
10. Isola JJ, Kallioniemi OP, Chu LW, et al. Genetic aberrations detected by comparative genomic hybridization predict outcome in node-negative breast cancer. Am J Pathol 1995;147:905–11.
11. Jain AN, Chin K, Borresen-Dale AL, et al. Quantitative analysis of chromosomal CGH in human breast tumors associates copy number abnormalities with p53 status and patient survival. Proc Natl Acad Sci U S A 2001;98:7952–7.
12. Gelsi-Boyer V, Orsetti B, Cervera N, et al. Comprehensive profiling of 8p11-12 amplification in breast cancer. Mol Cancer Res 2005;3:655–67.
13. Ginestier C, Cervera N, Finetti P, et al. Prognosis and gene expression profiling of 20q13-amplified breast cancers. Clin Cancer Res 2006;12:4533–44.
14. Adler AS, Lin M, Horlings H, Nuyten DS, van de Vijver MJ, Chang HY. Genetic regulators of large-scale transcriptional signatures in cancer. Nat Genet 2006;38:421–30.
15. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671–9.
16. Yu JX, Sieuwerts AM, Zhang Y, et al. Pathway analysis of gene signatures predicting metastasis of node-negative primary breast cancer. BMC Cancer 2007;7:182.
17. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst 2005;97:1180–4.
18. Potti A, Dressman HK, Bild A, et al. Genomic signatures to guide the use of chemotherapeutics. Nat Med 2006;12:1294–300.
19. Hess KR, Anderson K, Symmans WF, et al. Pharma-cogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. J Clin Oncol 2006;24:4236–44.
20. Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. J Am Stat Assoc 1958;53:457–81.
21. Schafer JL. Analysis of incomplete multivariate data. London: Chapman & Hall/CRC Press; 1997.
22. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.
23. Van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–6.
24. Gruvberger S, Ringner M, Chen Y, et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res 2001;61:5979–84.
25. Foekens JA, Atkins D, Zhang Y, et al. Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. J Clin Oncol 2006;24:1665–71.
26. Desmedt C, Piette F, Loi S, et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res 2007;13:3207–14.
27. Zhang Y, Sieuwerts AM, McGreevy M, et al. The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. Breast Cancer Res Treat. Epub ahead of print 2008 Sep 23.
28. Bonnefoi H, Potti A, Delorenzi M, et al. Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00–01 clinical trial. Lancet Oncol 2007;8:1071–8.
29. Peintinger F, Anderson K, Mazouni C, et al. Thirty-gene pharmacogenomic test correlates with residual cancer burden after preoperative chemotherapy for breast cancer. Clin Cancer Res 2007;13:4078–82.