Patterns of Chromosomal Imbalances Defines Subgroups of Breast Cancer with Distinct Clinical Features and Prognosis. A Study of 305 Tumors by Comparative Genomic Hybridization

Karin Rennstam,1 Minna Ahlstedt-Sööni,2 Bo Baldetorp,1 Pär-Ola Bendahl,1 Åke Borg,1 Ritva Karhu,2 Minna Tanner,2 Mika Tirkkonen,2 and Jorma Isola2

1Jubilee Institute, Department of Oncology, University Hospital, Lund, Sweden, and 2Institute of Medical Technology, Tampere University and University Hospital, Tampere, Finland

ABSTRACT

Chromosomal copy number aberrations (CNAs) are common in breast cancer and involve genomic regions in a frequency and combination, suggesting distinct routes of tumor development. We studied chromosomal gains (+) and losses (−) by comparative genomic hybridization from a series of 305 unsolicited primary invasive breast cancers. CNAs were observed in >90% of the tumors and involved all chromosomal arms in various frequencies, the most common being +1q (55%), +8q (41%), +16p (40%), +17q (28%), −13q (27%), −16q (22%), +20q (19%), −8p (18%), and +11q (16%). Eighteen pairs of CNAs were revealed as significantly associated using Fisher’s exact test with Bonferroni correction, the most common pairs being −8p/+8q, +17q/+20q, and −4q/−13q. To study more complex relationships between individual CNAs, principal component analysis and distance-based tree modeling were performed independently. Three distinct patterns of CNAs were observed. Group A was defined by +1q, +16p, and −16q, group B by +11q, +20q, +17q, and −13q, and group C by −8p and +8q. Group A was correlated to positive estrogen receptor and progesterone receptor (PgR) status (P < 0.001 and P < 0.05, respectively). Groups B and C were correlated to DNA nondiploidy (P < 0.001 and P < 0.05), high histological grade and lymph node positivity (P < 0.05), and group B also to high proliferation rate, large primary tumor size (P < 0.001), and negative PgR status (P < 0.05). Patients with aberrations in group A only had a significantly higher breast cancer survival rate than all other patients. The worst survival was seen for patients with aberrations in group C only along with the patients displaying aberrations from all CNA pattern groups (ABC). The 5-year survival rates vary from 96% in group A to 56% in group C. These correlations were independent of node status, tumor size, and PgR status in a multivariate analysis. We conclude that patterns of copy number gains and losses define breast tumors with distinct clinicopathological features and patient prognosis.

INTRODUCTION

Breast cancer is presently the most frequently diagnosed cancer form among women in the Western world. One of nine women will develop the disease during their lifetime. Breast cancer arises due to a series of alterations in cell physiology that collectively contribute to malignant growth. Malignant cell growth is a sum of self-sufficiency in growth signals (or activation of proto-oncogenes), insensitivity to growth inhibitory signals (tumor suppressor gene inactivation), evasion of programmed cell death (apoptosis), limitless replicative potential (immortalization), sustained angiogenesis (vascularization), tissue invasion, and metastasis (1). These pathophysiological changes, in turn, arise because of multiple changes in the genome of normal epithelial cells. The genomic aberrations include single nucleotide point mutations, amplifications or deletions of single genes, insertions and translocations, gains and losses of entire or parts of chromosomes or chromosome arms, and eventually gross changes in chromosome number, i.e., aneuploidization (2–4).

Genetic abnormalities in breast cancer are known to be related to clinical outcome. HER-2/Neu gene amplification and its concomitant protein overexpression are associated with inferior outcome (5–9), as is mutation of the TP53 tumor suppressor gene and the p53 protein (10, 11). However, much less is known about how chromosomal aberrations, i.e., gains and losses of entire or parts of chromosome arms, are related to each other and what their role is in defining the clinical outcome of the patient.

Comparative genomic hybridization (CGH) provides a genomewide insight into the molecular pathogenesis of human cancers. CGH is a fluorescent in situ hybridization-based screening technique that detects, in a quantitative manner, gains and losses of genomic material, enabling a genome-wide survey of chromosomal imbalances in a single hybridization (12, 13). Thus far, only relatively small materials of CGH in breast tumors have been published (5, 14–16). Here, we studied chromosomal gains and losses by CGH from a series of 305 primary invasive breast cancer specimens. We studied how aberrations are related to each other in the carcinogenesis as well as their clinicopathological and prognostic correlations.

MATERIALS AND METHODS

Tumor Material. Freshly frozen tumor samples were obtained from 305 patients, operated for primary invasive breast cancer at the departments of surgery in the South Swedish Health Care Region (Sweden; n = 164) and in the Tampere University Hospital District (Tampere, Finland; n = 141). The Swedish samples were taken from the frozen tumor tissue bank at the Department of Oncology, Lund University Hospital (Lund, Sweden), where they were originally collected for the routine analysis of steroid hormone receptors and DNA flow cytometric parameters (17, 18). Presence of malignant cells was verified in all tumor samples by evaluation of touch preparation imprints. To determine the histological representativeness of the tumors, frozen sections were cut (adjacent to the sections used for DNA extraction) and H&E stained. Only highly representative tumors (>60% tumor cells) were included in the analysis. The CGH profiles from 27 of the Swedish and 55 of the Finnish tumors have been reported previously (16, 19).

Patient Background Data. The median age of patients in the entire material was 62 years (range, 32–92 years). Ninety-two of the 305 patients died because of breast cancer during the follow-up (median follow-up of nondeceased patients after surgery was 98 months). Eighty-four of the patients received adjuvant therapy. One of nine women will develop the disease during their lifetime. Breast cancer arises due to a series of alterations in cell physiology that collectively contribute to malignant growth. Patients received adjuvant tamoxifen therapy, 17q (28%), and a statistically significant difference.

The costs of publication of this article were defrayed in part by the payment of page charges. The costs of publication of this article were defrayed in part by the payment of page charges. The costs of publication of this article were defrayed in part by the payment of page charges. The costs of publication of this article were defrayed in part by the payment of page charges. The costs of publication of this article were defrayed in part by the payment of page charges. The costs of publication of this article were defrayed in part by the payment of page charges.

Received 7/9/03; revised 9/9/03; accepted 10/16/03.

Grant support: The Swedish Cancer Foundation Grant 4237-099-01XAB, Berta Kamprad Cancer Foundation, Unni Nilsson Cancer Foundation, John and Augusta Persson Foundation, IngaBritt and Arne Lundberg Foundation, Hospital of Lund Foundation, Franke and Margaretha Bergqvist Foundation, King Gustaf V Jubilee Foundation Grant 99:525, the European Union Research Contract, the Tampere University Hospital Research Foundation, the Academy of Finland, Sigrid Juselius Foundation, Emil Aaltonen Foundation, and the Finnish Cancer Foundation. Prof. Jorma Isola received Swedish Cancer Society Fellowships 4047-B98-01VAA and 4047-B99-02VAA.

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.

Requests for reprints: Karin Rennstam, Jubilee Institute, Department of Oncology, Lund University, SE-221 85 Lund, Sweden. Phone: 46-46-177567; Fax: 46-46-147327; E-mail: Karin.Rennstam@onk.lu.SE.

8861

Downloaded from cancerres.aacrjournals.org
available for 8 patients). Nodal status was confirmed in 291 patients, and 128 of these (44%) were found to have lymphatic spread of their disease. Nine of 291 patients (5%) had a confirmed distant metastasis already at the time of primary surgery. These patients were included in the routine CGH profile analyses but excluded from all survival analyses. The status for estrogen receptor (ER) and progesterone receptor (PgR) was confirmed in 299 and 289 cases, respectively. One-hundred ninety-eight of 299 patients were ER positive (66%), and 151 of 289 patients were PgR positive (52%). DNA flow cytometric evaluation was made for 281 of the patients. One-hundred fifteen of these patients had a DNA diploid tumor (41%). Flow cytometric S-phase fraction (SPF) was available for 258 of the patients, and the median SPF was 8% (range, 0–36%). A high-risk group with 71 patients was defined as Swedish patients with SPF ≥ 12% and Finnish patients with SPF ≥ 15%. The cutoff used in the Swedish material has been used in other studies (17, 18), whereas the slightly higher cutoff in the Finnish material was chosen to get the same fraction of cases with high SPF in the two materials. Methods for receptor analysis and DNA flow cytometry have been described previously (17, 18).

There were no significant differences in distributions between the Swedish and Finnish tumors with respect to ER status, DNA ploidy, lymph node status, patient age at time of surgery (neither in median age nor fraction of patients ages ≤ 50 years), and categorized SPF (by definition because different cutoff values were used). However, the Swedish patients had a significantly higher proportion of PgR-positive tumors (58 versus 46%, P = 0.05) and large (≥ 20 mm) primary tumors (76 versus 63%, P = 0.02). Differences were also seen in the proportion of patients that received adjuvant therapy. In the Swedish material, 37% received adjuvant tamoxifen therapy, compared with 19% in the Finnish material (P = 0.001). Adjuvant cyclophosphamide, methotrexate, and 5-fluorouracil polichemotherapy was given to 10% of the patients in the Swedish material compared with 4% of the Finnish patients.

CGH. CGH was performed according to a previously published protocol (20, 21). Briefly, tumor DNA was extracted from freshly frozen tumor tissue (QIagen, Hilden, Germany) and labeled with FITC-dUTP and FITC-dCTP (DuPont, Boston, MA) using standard nick translation. Labeled DNAs (400–800 ng each, labeled reference DNA; Vysis, Inc.) and 10 μg of unlabeled Cot-1 DNA (Life Technologies, Inc., Gaithersburg, MD) were hybridized onto commercially available normal metaphase chromosomes (Vysis, Inc., Downers Grove, IL). Included in each hybridization batch were one positive control (MPE-600 cell line) and one negative control (normal female DNA). Some of the control hybridizations suggested false aberration-like patterns on chromosomes 19 and 22, and these regions were therefore excluded from all additional analyses (21, 22). The hybridizations were evaluated using the QUIPS digital image analysis system (Vysis Inc.). At least five metaphases from each tumor were analyzed. Limits for gain and loss were set to 1.15 and 0.85, respectively. Gains and losses were detected on an arm-specific basis, except for regions on 8q, 17q, and 20q, which already at the start of this study were known to be individually correlated to patient outcome (23–25). Therefore, these regions were studied in greater detail.

Statistical Methods. The CGH data analyzed in the present article were coded as 82 binary variables, one gain and one loss variable for each of the 41 chromosome arms analyzed (chromosomes 13, 14, 15, 21, and 22 lack the p-arm.) CGH profiles showing both gain and loss on the same arm were coded as gains if the gains were dominating but otherwise as losses. No case was thus coded as having both gain and loss on a specific chromosome arm. Some of the analyses were based on the 81 binary variables with nonzero SD, excluding the loss variable for 1q, an arm without losses among the 305 patients in this material.

Associations between individual copy number aberrations (CNAs) were analyzed in various ways. The first approach was to search for pairs of CNAs occurring more often together than expected under the null hypothesis of no association. In this analysis, loss or gain of both arms on the same chromosome was disregarded because this might indicate monosomies or polysomies of the entire chromosome and not rearrangements of specific genomic regions that are essential to cancer development. Fisher’s exact test was used, and associations with P < 0.000016 were considered significant. This significance level corresponds to a Bonferroni correction of the 5% level adjusting for 3200 tests.

Higher order interactions between CNAs were evaluated using PCA and distance-based tree modeling. The nine most common genetic aberrations, each occurring in >15% of the patients, were included in these analyses.

Inclusion of the nine most common CNAs is a compromise between large nonrobust models and those including too few events to be biologically interesting. Simon et al. (26) suggest a method by Brodeur et al. (27) for identification of a subset of aberrations that occur more often than would be expected by chance. When applied to our material, the method suggests the inclusion of 16 CNAs. However, interpretation of the results from PCA of 16 CNAs became uncertain and was therefore reduced to the nine most common CNAs. The structure of the distance-based trees converged to three distinct groups of aberrations as the number of CNAs was reduced from 16 to 9, excluding the least prevalent aberration in each step. Software for distance-based tree modeling has been developed by Desper et al. (28) and is available for download online.3

The PCA analysis was based on the correlation matrix and three components were needed to explain >50% of the variability.

Death because of breast cancer was chosen as the end point in survival analyses. Differences in breast cancer survival (BCS) were evaluated using log-rank tests, Kaplan-Meier estimates, and Cox regression. Schoenfeld’s test (29) was used to check proportional hazards assumptions. The prognostic value of each of the 81 binary CNAs was evaluated using both standard asymptotic and exact log-rank tests. The number of significant findings was then related to the expected number of significant findings under the null hypothesis of no association between the CNAs and BCS.

A simulation procedure was used to estimate the distribution of number of significant associations under the null hypothesis. Briefly, in each simulation, the survival data were shuffled randomly, log-rank tests for each CNA (n = 81) were performed, and the number of significant associations was recorded. After 1000 simulations, we obtained an estimate of number of significant associations at the 5% level under the null hypothesis of no association between the CNAs and BCS. This distribution was used to assign P-values to the null hypothesis that the observed number of significant associations is equal to the expected number assuming that the null hypothesis is true.

Fisher’s exact test was used to test for associations in two-by-two tables, χ2 tests for trend for two-by-n tables where the n categories are ordered, and linear regression when testing for association between two ordered variables, both of which have more than two categories. The nonparametric Mann-Whitney U test was used to compare median levels of continuous variables in two groups. All tests were two sided, and the significance level was set to 5%. The statistical packages Stata 7.0 (Stata Corporation, College Station, TX) and StatXact 4.0 (Cytel Software Corporation, Cambridge, MA) were used.

RESULTS

Overview of Chromosomal Aberrations. We found 270 breast tumors (89%) with gains and 216 tumors (71%) with losses. Twenty-seven tumors (8.9%) showed no CNAs at all. The median number of losses/tumor was 2.6 (range, 0–18), the median number of gains 3.5 (range, 0–15), and the median number of CNAs was 5.0 (range, 0–28).

The most commonly observed gains were seen at 1q (55%), 8q (41%), 16p (40%), 17q (28%), 20q (19%), and 11q (16%), and the most common losses were found at 13q (27%), 16q (22%), and 8p (18%; Table 1).

Gains and losses of regions on 8q, 17q, and 20q were studied in detail because of prior knowledge of the regional CNAs (namely amplifications) located in these chromosome arms. On 8q, 64 (52%) proximal (8q11-q13) and 120 (97%) distal (8q21-q24) regional gains were found (n = 124). Gain of the entire arm was seen in 61 (49%) of the cases. For chromosome arm 17q, we found 42 (48%) proximal (17q11-q12) and 84 (97%) distal regional gains (17q21-q25), and the entire arm was gained in 34 (39%) of the cases (n = 87). On 20q, we found 35 (60%) proximal (20q11), 55 (95%) distal (20q12-13), and 35 (60%) entire-arm gains (n = 58). Losses on these chromosomal regions were only found on rare occasions, typically when the entire arm was lost.


18862

Downloaded from cancerres.aacrjournals.org
Common aberrations, with a prevalence of more than 15%, are shown in boldface. These aberrations were later included in further statistical analyses. The most common aberrations, a distance matrix was created, describing how closely events are related to each other. Two aberrations often occurring in the same tumor will appear close together, and aberrations that rarely occur together will appear far apart. The distance matrix was then transformed into a distance-based tree, consisting of three very distinct branches. As in the PCA, one group of tumors carries the aberrations +1q, +16p, and −16q (group A); the second group −13q, +11q, +20q, and +17q (group B); and the third group carrying −8p and +8q (group C; Fig. 2).

Taking together, the two methods here used identify exactly the same groups of tumors.

Next, we studied whether the CNA patterns are different in DNA diploid and nondiploid tumors separately. No major changes were found with respect to the structure of the trees or which were the nine most common aberrations of the three groups (for diploid tumors +1q was substituted for −6q as the eighth most frequent aberration; data not shown).

**CNA and Tumor Phenotype.** Total number of gains, losses, CNAs, and the most common individual aberrations were correlated to established prognostic factors such as histological grade (data only available from the Finnish part of the material), ER status, PgR status, DNA ploidy status, SPF, lymph node status, tumor size, and patient age at time of diagnosis (Table 3). The strongest associations were seen between number of gains, number of losses, and total number of CNAs versus DNA nondiploidy. Of the individual aberrations, loss of 16q (P < 0.001), gain of 16p (P < 0.001), and gain of 1q (P < 0.01) were significantly correlated to ER positivity, loss of 16q and gain of 16p was also correlated to PgR positivity (P < 0.01 and P < 0.05, respectively), and gain of 16p was correlated to low histological grade (P < 0.01). Gain of 17q was significantly correlated to DNA nondiploidy (P < 0.001) and a high SPF (P < 0.001), as well as low patient age (P < 0.01), lymph node positivity (P < 0.05), and large primary tumor (P < 0.05). Loss of 13q was strongly correlated to a large primary tumor (P < 0.001) and DNA nondiploidy (P < 0.01). Gain of 11q was correlated to PgR negativity (P < 0.01) and large primary tumor (P < 0.05). Gain of 20q was correlated to DNA nondiploidy (P < 0.01), and gain of 8q was correlated to high histological grade (P < 0.01), DNA nondiploidy (P < 0.05), and lymph node positivity (P < 0.05).

Comparing the three CNA pattern groups to the prognostic factors mentioned above shows that group A is strongly correlated to ER and PgR positivity (P < 0.001 and P < 0.05, respectively). Group B is strongly correlated to DNA nondiploidy, high SPF, and large primary tumors.

### Table 1. All copy number aberrations found in 305 invasive breast carcinomas

Copy number aberrations are listed in the order of their prevalence. The nine most common aberrations, with a prevalence of more than 15%, are shown in boldface. These aberrations were later included in further statistical analyses.

<table>
<thead>
<tr>
<th>Chromosome arm</th>
<th>No. of cases</th>
<th>Frequency (%)</th>
<th>Chromosome arm</th>
<th>No. of cases</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>04</td>
<td>0</td>
<td>0</td>
<td>09</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>01</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Scoring coefficients for the first three principal components, together describing ~50% of the variance in the binary variables corresponding to absence/presence of the nine most common aberrations

The first two principal components, PC-1 and PC-2, separate group A from groups B and C, whereas the third component, PC-3, identifies group C.

<table>
<thead>
<tr>
<th>Copy number aberrations</th>
<th>PC-1</th>
<th>PC-2</th>
<th>PC-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1q</td>
<td>0.12</td>
<td>0.59</td>
<td>−0.04</td>
</tr>
<tr>
<td>+16p</td>
<td>0.14</td>
<td>0.40</td>
<td>0.47</td>
</tr>
<tr>
<td>−16q</td>
<td>−0.12</td>
<td>0.67</td>
<td>−0.12</td>
</tr>
<tr>
<td>+13q</td>
<td>0.34</td>
<td>−0.01</td>
<td>0.34</td>
</tr>
<tr>
<td>+11q</td>
<td>0.35</td>
<td>0.05</td>
<td>0.18</td>
</tr>
<tr>
<td>+20q</td>
<td>0.45</td>
<td>−0.07</td>
<td>0.25</td>
</tr>
<tr>
<td>+17q</td>
<td>0.46</td>
<td>−0.16</td>
<td>0.08</td>
</tr>
<tr>
<td>−8q</td>
<td>0.38</td>
<td>0.08</td>
<td>−0.48</td>
</tr>
<tr>
<td>+8q</td>
<td>0.39</td>
<td>0.03</td>
<td>−0.56</td>
</tr>
</tbody>
</table>
tumor (P < 0.001), high histological grade, PgR negativity, and lymph node positivity (P < 0.05). Group C shows a statistical correlation to high histological grade, DNA nondiploidy, and lymph node positivity (P < 0.05). These analyses were made dichotomously according to the CNA pattern groups (i.e., A versus non-A, B versus non-B, and C versus non-C), regardless of what other aberrations the tumor may present. When analyzing the group of tumors with aberrations from group A only, we find that this group is associated with PgR positivity (P < 0.001), low histological grade, and DNA diploidy (P < 0.001), lymph node negativity, small primary tumor, and high patient age (P < 0.05). Groups B only and C only are too small to examine in this way.

**CGH Aberrations and Clinical Outcome.** To verify that our material is representative of an invasive breast cancer material, a univariate statistical analysis was performed, correlating the most common prognostic factors to BCS (Table 4). A multivariate analysis (data not shown) confirmed that the most informative prognostic factors in this material were lymph node status, tumor size, and PgR.

When studying the correlation between number of CGH gains, losses, total CNAs, and patient survival, we found that each added gain increases the hazard of dying from breast cancer by 11% (P < 0.001), which each added loss increases the hazard by 6% (P = 0.06). Each added CNA increases the hazard of dying from breast cancer by 5% (P = 0.03). However, none of the quantitative variables of chromosomal aberrations provided prognostic information after adjustment for lymph node status, tumor size, and PgR status in a multivariate analysis (data not shown).

In a univariate survival analysis, 20 of the 81 CNAs turned out to be significantly associated with BCS (P < 0.05). This number is significantly higher than the four significant associations expected under the null hypothesis of no associations (P = 0.003). Gains of 17q (P < 0.0001), 8q (P = 0.0026), and 7p (P = 0.0029) and losses of 9q (P = 0.0021) and 11p (P = 0.0038) have the strongest prognostic value. Four of the nine most common aberrations (included in PCA and distance-based tree modeling) proved to be significantly associated with patient outcome. All aberrations, significant at the 5% level, are shown in Table 5 (aberrations from groups A, B, and C highlighted).

When studying the banding analysis of gains on chromosome 8q, 17q, and 20q in correlation to breast cancer related death, we found that amplification of certain regions increases the hazard more than others. On chromosome 8, gain of 8q21-q24, but not those at q11 and q13, was significantly correlated to prognosis, yielding a hazard ratio (HR) of 1.8 (P < 0.05). For chromosome 17, gain of all bands...
analyzed is correlated to an increased risk of breast cancer related death. The strongest correlation is seen for band 17q23-q24, corresponding to a 2.5-fold increased hazard \( (P < 0.0003; \text{Fig. 3}) \). In this material, none of the bands on 20q proved to be statistically correlated to patient survival, nor did 20q as a whole.

**Survival Analysis On the Basis of the CNA Pattern Groups A, B, and C.** To study the prognostic impact of the CNA patterns as defined above, we first analyzed the 294 nonmetastasized patients to study the prognostic impact of the CNA patterns as A, B, C, and 0 to established prognostic factors connected to poor patient outcome. We found that aberrations from group B were associated with worse BCS (HR = 2.0; \( P = 0.002 \)). The same is true for aberrations from group C (HR = 2.0; \( P = 0.001 \)), whereas aberration from group A seems to be of no prognostic value (HR = 1.1; \( P = 0.7 \)). Because the same tumor can be classified into more than one group, a bivariate analysis, including indicators for B, C, and their interaction, was fitted. This model shows that B and C are independent prognostic factors and that they interact in a roughly multiplicative way, corresponding to a HR of 3.9 for BC-carriers compared with patients without aberrations in the B and C groups.

The three group indicators A, B, and C can be combined in eight ways leading to the subgroups A, B, C, AB, AC, BC, ABC, and 0, as explained above. Fig. 4 shows the BCS for each of these eight groups. Interestingly, group A has clearly the best prognosis of the patients, even better than group 0. The worst prognosis was seen for patients from group C along with the group of patients that have aberrations from all CNA pattern groups (ABC). The 5-year survival rates (95% confidence interval) vary from 96% (95% confidence interval, 85–99%) in group A to 56% (95% confidence interval, 20–80%) in group C.

A multivariate Cox regression analysis, including lymph node status, tumor size, PgR status, and group indicators for A, B, C and their interactions, shows that the patient prognostic information found in the CNA pattern groups was independent of all three of the above-mentioned prognostic indicators. The significant effects after adjustment for lymph node status, tumor size, and PgR status were A \( \text{versus } 0 \) (HR = 0.18, \( P = 0.028 \)) and C \( \text{versus } 0 \) (HR = 4.2, \( P = 0.013 \); Table 6).

To investigate whether the individual CNAs showing prognostic importance in a univariate analysis (shown in Table 5) give any added information about patient survival to the groups of the tree model, each of the aberrations in Table 5 were added, one at a time, to a Cox regression model, including the three CNA pattern groups and all interactions between them. The Ps from these analyses are given in the last column of Table 5. We found that the aberrations +17q, -9q,
Table 5. Aberrations showing prognostic importance in a univariate analysis correlating individual copy number aberrations to patient survival ($n = 294$)

Four of the 9 most common aberrations (Table 1) proved to be significant for patient outcome. These are shown in boldface. The last column shows the $P$ values for individual aberrations giving additional information about patient survival when copy number aberration pattern group membership is known.

<table>
<thead>
<tr>
<th>Aberration</th>
<th>No. of observed cases</th>
<th>No. of deceased patients</th>
<th>Expected no. of deceased patients</th>
<th>$P$ (log rank)</th>
<th>$P$ (added survival information)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1q</td>
<td>81</td>
<td>40</td>
<td>23</td>
<td>$&lt;0.0001$</td>
<td>0.027</td>
</tr>
<tr>
<td>−9q</td>
<td>15</td>
<td>9</td>
<td>3.4</td>
<td>0.0021</td>
<td>0.018</td>
</tr>
<tr>
<td>+8q</td>
<td>119</td>
<td>46</td>
<td>32</td>
<td>$0.0026$</td>
<td>0.76</td>
</tr>
<tr>
<td>+7p</td>
<td>24</td>
<td>14</td>
<td>6.6</td>
<td>0.0029</td>
<td>0.058</td>
</tr>
<tr>
<td>−11p</td>
<td>9</td>
<td>6</td>
<td>2.0</td>
<td>0.0038</td>
<td>0.012</td>
</tr>
<tr>
<td>+2q</td>
<td>4</td>
<td>3</td>
<td>0.66</td>
<td>0.0039</td>
<td>0.074</td>
</tr>
<tr>
<td>+4p</td>
<td>3</td>
<td>2</td>
<td>0.34</td>
<td>0.0046</td>
<td>0.082</td>
</tr>
<tr>
<td>−7q</td>
<td>4</td>
<td>3</td>
<td>0.75</td>
<td>0.0090</td>
<td>0.12</td>
</tr>
<tr>
<td>−6q</td>
<td>39</td>
<td>5</td>
<td>14</td>
<td>0.0094</td>
<td>0.006</td>
</tr>
<tr>
<td>−1p</td>
<td>27</td>
<td>14</td>
<td>7.7</td>
<td>0.012</td>
<td>0.11</td>
</tr>
<tr>
<td>−15q</td>
<td>7</td>
<td>4</td>
<td>1.2</td>
<td>0.013</td>
<td>0.070</td>
</tr>
<tr>
<td>−8q</td>
<td>53</td>
<td>22</td>
<td>14</td>
<td>0.018</td>
<td>0.67</td>
</tr>
<tr>
<td>+16q</td>
<td>163</td>
<td>57</td>
<td>47</td>
<td>0.035</td>
<td>0.012</td>
</tr>
<tr>
<td>−Xq</td>
<td>8</td>
<td>5</td>
<td>2.0</td>
<td>0.037</td>
<td>0.33</td>
</tr>
<tr>
<td>+4q</td>
<td>38</td>
<td>18</td>
<td>12</td>
<td>0.042</td>
<td>0.51</td>
</tr>
<tr>
<td>+7q</td>
<td>28</td>
<td>13</td>
<td>7.7</td>
<td>0.044</td>
<td>0.33</td>
</tr>
<tr>
<td>−3p</td>
<td>11</td>
<td>6</td>
<td>2.7</td>
<td>0.046</td>
<td>0.38</td>
</tr>
<tr>
<td>−20p</td>
<td>3</td>
<td>2</td>
<td>0.55</td>
<td>0.048</td>
<td>0.28</td>
</tr>
</tbody>
</table>

−11p, −6q, and +1q are significant for patient survival even after adjustment for respective CNA pattern group membership. All of these aberrations, except loss of chromosome arm 6q, are correlated to an increased hazard of dying from breast cancer. Gain of 1q and 17q are already components of the tree model. However, the prognosis of the patients in groups A and B, respectively, is significantly different if the 1q/17q aberration is present or not.

**DISCUSSION**

A CGH analysis of 305 breast cancer tumors gave an enormous amount of cancer biological data, as revealed by complex biostatistical analyses. Previous studies with smaller tumor materials have shown how individual CNAs correlate with tumor biological features or patient survival (5, 19, 23, 30–32), but the results are difficult to interpret, on the one hand because of the possibility of obtaining chance correlations in tens or even hundreds of repeated statistical tests, and on the other hand because of the lack of information on the common coexisting patterns of CNAs. By using PCA and distance-based tree analysis, we were able to define three distinct groups of tumors based on their CNA composition. The fact that two different biostatistical approaches identified the same groups suggests that the CNA pattern groups A, B, and C are true biological entities rather than reflection of the statistical modeling used.

Group A, containing gains of 1q and 16p and loss of 16q, may reflect the cytogenetically described translocation t(1q;16p; Ref. 33) or the well-differentiated ductal carcinoma in situ amplification on chromosome 1q and deletion on 16q (34). Gains of 1q and losses of 16q have been reported to be involved in early steps of cancer progression (19, 35) and were in our material correlated with favorable prognostic factors, indicating that patients belonging to this group might have a better prognosis.

Group B, containing −13q, +11q, +20q, and +17q consists of common individual aberrations (Table 1 and Refs. 19, 34, 36, 37) and might be characterized as a group of tumors that are prone to onco-generation. These aberrations have, to our knowledge, not previously been reported as linked to each other. Chromosome arms 11q, 20q, and 17q gains and 13q losses have individually been associated with aggressive tumor phenotype (19, 35), and in our material, this CNA pattern was linked to high histological grade, loss of −11p, −6q, and +1q are significant for patient survival even after adjustment for respective CNA pattern group membership. All of these aberrations, except loss of chromosome arm 6q, are correlated to an increased hazard of dying from breast cancer. Gain of 1q and 17q are already components of the tree model. However, the prognosis of the patients in groups A and B, respectively, is significantly different if the 1q/17q aberration is present or not.

![Chromosome 8](image1.png)  ![Chromosome 17](image2.png)

![Fig. 3. A detailed analysis of gains on chromosome 8q and 17q in relation to breast cancer related survival. $P$s are calculated for each band separately and grouped together only to simplify the figure.](image3.png)

![Fig. 4. A Kaplan-Meier Survival analysis based on the three CNA pattern groups A, B, and C, as well as the subgroups AB, AC, BC, ABC, and 0 ($n = 294$). Curves are terminated when ≤5 patients remain in the group.](image4.png)

Table 6. Multivariate Cox regression analysis including lymph node status, tumor size, progesterone receptor status, and group indicators for A, B, C, and their interactions

<table>
<thead>
<tr>
<th>Hazard ratio</th>
<th>95% confidence interval</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A versus 0</td>
<td>0.18</td>
<td>0.38–0.83</td>
</tr>
<tr>
<td>B versus 0</td>
<td>0.93</td>
<td>0.31–2.8</td>
</tr>
<tr>
<td>C versus 0</td>
<td>4.2</td>
<td>1.4–13</td>
</tr>
<tr>
<td>AB versus 0</td>
<td>1.3</td>
<td>0.55–3.3</td>
</tr>
<tr>
<td>AC versus 0</td>
<td>0.90</td>
<td>0.28–2.9</td>
</tr>
<tr>
<td>ABC versus 0</td>
<td>1.5</td>
<td>0.70–3.2</td>
</tr>
</tbody>
</table>

The significant effects after adjustment for lymph node status, tumor size, and progesterone receptor status are shown in boldface.
of PgR, DNA nondiploidy, high SPF, positive lymph nodes, and a large primary tumor size.

Group C, characterized by loss of 8p and gain of 8q as a common denominator, might result from an isochromosome 8 formation. This, in turn, may lead to loss of tumor suppressor genes on 8p (i.e., TACC1, DLEC1, and N33) and amplification or excessive copies of oncogenes in 8q (i.e., c-MYC and EBAG9). Gain of 8q and loss of 8p have also been associated with poor patient outcome (19, 35). Our multivariate analyses showed that a chromosomal defect or defects on chromosome 8 define a distinct entity of breast tumors in which the tumor biological features, reflected in the clinicopathological and prognostic correlations, are likely to be driven by chromosome 8 and not the CNAs present in the same tumors in a highly variable manner.

CNA pattern trees from diploid and nondiploid tumors separately revealed no structural changes of the trees, and the three CNA pattern groups remained essentially the same as in the entire material. This may indicate that ploidy shift, i.e., aneuploidization of the entire genome, is a late genetic event in breast cancer, which is preceded by both oncogene amplifications and many chromosome rearrangements, as shown previously using flow cytometric sorting techniques (38).

Aberrations most commonly associated were loss of 8p and gain of 8q. This could reflect an isochromosome 8 formation, as previously shown (5). The second most common pair of aberrations was gain of 17q and gain of 20q, commonly involved in unbalanced translocations (39). Similarly, losses of 4q and 13q were commonly associated. These regions have never previously been reported as being mutually related, a fact that might be explained by the rather infrequent loss of 4p. Still, when it occurs in this material, it is often in association with loss of 13q.

In our material, as in others previously published (19, 36), gains of genetic material were found to be slightly more common than losses. In general, the frequency of the most common gains and losses found in this material correlates well with findings previously reported in the literature (5, 19, 23, 30–32). The small percentage of tumors (8.9%) presenting no aberrations may reflect tumors with CNAs too small to be detected by CGH. It is also possible that in some cases the DNA extracted from the tumor was too degraded to allow a successful CGH hybridization to be performed or that the tumor sample from which DNA was extracted contained too high a proportion of normal cells and was not sufficiently representative of carcinoma, despite all efforts being made to select histologically representative samples.

Survival analysis shows that number of gains, losses, and total number of CNAs is correlated to patient survival but neither of these factors gives any added information to lymph node status, tumor size, and PgR status.

When linking the CNA patterns to patient survival, we found that the prognosis of group A patients is much better than that of group B or C patients. Group A patients also have a much better survival rate than the patients from groups AB, AC, and ABC. This explains why group A aberrations were not correlated to BCS in the univariate analysis comparing patients having A aberrations to those not having A aberrations, regardless of the coexisting CNAs belonging to groups B and C. The prognosis of group A patients was also much better than that of the group of patients that carries none of the nine most common aberrations (group 0). Because group A patients are strongly correlated to having a positive ER and PgR status, the survival difference could be attributed to adjuvant hormonal therapy. This was ruled out, however, because a smaller proportion of group A patients, (~20%) compared with group B and C patients (~30%) as well as ABC patients (~40%), had received adjuvant hormonal therapy. It is noteworthy that there were no breast cancer related deaths among the patients from group A who received adjuvant hormonal treatment. Patients belonging to groups B or C had a much worse prognosis than those of group A, and the worst prognosis is seen for the group of patients that has aberrations from all three CNA pattern groups (ABC), again indicating that accumulation of genetic aberrations is a sign of a grossly malignant cell.

The patient survival information received from the CNA pattern groups was independent of node status, tumor size, and PgR status, indicating that CNAs provide true biological information, which is independent of clinical stage at the time of diagnosis. Although the aim of this study was not to create a tool to predict survival in breast cancer patients in the clinical setting, we found that the CNA pattern groups withheld a lot of prognostic information. Adding the five CNAs that showed the strongest correlation to patient survival in the univariate analysis actually predicted patient outcome, as well as the established prognostic factors (data not shown).

The more detailed banding analysis of chromosome 8q, 17q, and 20q revealed that for 8q and 17q distal amplifications have a higher correlation to patient prognosis. CGH identified two independent aberrations at 17q: amplification at 17q12-q21 with the possible target HER-2/Neu, as well as amplification at 17q22-q24 where several candidate proto-oncogenes have been proposed (mm23, S6K, BCA53).

Interestingly, the second region was more strongly associated with breast cancer related death, indicating that genes other than HER-2/Neu located on 17q could be more important for patient survival. Gain of chromosome 8q21-q22 correlated to a 2-fold increased hazard of dying from breast cancer. Possible target genes in this area are TP53 at 8q21 and CCNE2 at 8q22. In contrast to other studies (40), gain of 20q was not correlated to patient survival. This can be explained by the fact that the resolution of chromosomal CGH might be insufficient for banding analysis at a short chromosome arm such as 20q. Newer techniques such as microarray-based CGH may give a much higher resolution and is better suited for analysis of CNAs located in short chromosome arms.

REFERENCES