Genomic Differences Between Pure Ductal Carcinoma In Situ of the Breast and that Associated with Invasive Disease: a Calibrated aCGH Study

Vladimir V. Iakovlev,1 Nona C.R. Arneson,1 Vietty Wong,1 Chunjie Wang,1 Stephanie Leung,3,5 Gaiane Iakovleva,6 Keisha Warren,1 Melania Pintilie,2 and Susan J. Done1,3,4,5

Abstract

Purpose: In the quest for new targets, genomes of ductal carcinoma in situ (DCIS) and infiltrating duct carcinoma (IDC) have been compared previously; however, genomic alterations associated with cancer progression were difficult to identify. We hypothesized that significant events can be detected by comparing lesions with a broader range of behavior: from pure DCIS to IDC associated with lymph node metastasis.

Experimental Design: Array comparative genomewide hybridization, calibrated by self-self hybridization tests, was used to study 6 cases of pure DCIS and 17 cases of DCIS paired with IDC where 8 tumors had spread to the local lymph nodes.

Results: Pure DCIS exhibited a marginally higher degree of genomic complexity than DCIS and IDC components of invasive tumors. The latter two showed similarity between tumors and between components of the same tumor with several regions detected preferentially compared with pure DCIS. IDC associated with lymph node metastases showed similarity of genomic profiles as a group. Gain on 17q22-24.2 was associated with higher histologic grade, large IDC size, lymphatic/vascular invasion, and lymph node metastasis (P < 0.05).

Conclusions: Our findings suggest that DCIS and IDC are associated with specific genomic events. DCIS associated with IDC is genomically similar to the invasive component and therefore may represent either a clone with high invasive potential or invasive cancer spreading through the ducts. Specifically, gain on 17q22-24.2 is a candidate region for further testing as a predictor of invasion when detected in DCIS and predictor of nodal metastasis when detected in DCIS or IDC.

The role of DNA copy number changes is well documented in the neoplastic process. Discovery of MYC oncogene amplification in neuroblastoma serves as an example of gene amplification as a mechanism to promote gene function. It is known that breast cancers contain numerous genomic alterations (1), of which amplification of ERBB2 has been successfully targeted therapeutically. Recent expression profiling studies of frozen samples of invasive breast cancer have resulted in an understanding that infiltrating duct carcinoma (IDC) encompasses several different molecular types (e.g., luminal, basal, or HER-2 amplified). Unfortunately, our understanding of the molecular events associated with initial disease development and subsequent progression lags behind. Array comparative genomic hybridization (aCGH) has been used mainly to study IDC (2–4), and only a few studies have aimed to compare ductal carcinoma in situ (DCIS) and IDC searching for genomic alterations associated with cancer progression or genomic differences between these lesions. This work showed that DCIS and IDC foci of the same tumor have similar genomic profiles (5, 6).

We hypothesized that important genomic changes can be revealed by comparing samples with a broader range of biological behavior: from pure DCIS without invasion to invasive cancers associated with local lymph node metastasis. To our knowledge, no previous studies have covered this spectrum. Screening for DNA copy number changes associated with cancer subtypes or stages of progression can serve a dual purpose: selection of candidate genes for anticancer drug targeting as well as identification of diagnostic markers for prognosis and conventional treatment stratification. It is known that DCIS without detected invasion is a heterogeneous group of lesions with unpredictable prognosis; therefore, tests stratifying these patients according to their risk of progression would prevent undertreatment or overtreatment (7).
Pure DCIS is an early lesion limited to the duct system. Although it may be scattered within a large area, individual foci are often small and only found in formalin-fixed, paraffin-embedded tissue samples. We have developed and refined techniques to allow aCGH analysis to be done on histologically defined, microdissected, archival, formalin-fixed, paraffin-embedded samples (8). In this study, we describe aCGH analysis of pure DCIS and compare it with DCIS associated with invasive carcinoma and IDC with and without lymph node metastasis. The aCGH data were analyzed as nonsegmented for correlation between samples as well as segmented by circular binary segmentation (9) to define minimal common regions (MCR) of recurrent alterations. The amount of background noise was measured by self-self hybridization tests of DNA from normal tissues to determine cutoffs for significant segments. The samples were further analyzed by continuous moving average to identify peaks of genomic alterations within the MCRs. We used quantitative PCR (Q-PCR) as an independent technique to validate detected regions of gain. This analysis showed that specific genomic events are likely to occur at different breast cancer stages and need to be studied for development of novel targeted therapies and clinical prognostic markers.

**Materials and Methods**

**Samples.** After University Health Network Research Ethics Board approval, paraffin blocks and H&E slides of 17 breast specimens containing DCIS and IDC and 6 specimens containing pure DCIS from patients without a history of IDC were acquired. In our institution, excision specimens thought to contain pure DCIS are sampled extensively to exclude an occult invasive lesion. Synchronous lobular carcinoma in situ was seen in 2 cases of pure DCIS (03 and 06), which was avoided during microdissection. Columnar cell hyperplasia was also documented with a higher frequency in pure DCIS (5 of 6, 83%) than in cases with an invasive component (6 of 17, 35%); however, this may be due to more extensive sampling of the pure DCIS cases. Median patient age for the pure DCIS group was 51 years (range, 45-64) and 53.5 years (range, 41-86) for the invasive group. In total, 38 blocks were used to obtain 5-μm-thick sections, which were deparaffinized, stained in hematoxylin for 30 s, and microdissected in a stereoscopic dissecting microscope using 18-gauge needles to selectively collect DCIS and IDC samples [H&E slides as a reference, under the guidance of a breast pathologist (S.J.D.)]. Of 17 tumors with both DCIS and IDC, 15 tumors had sufficient amounts of both components for dissection and DNA extraction (small amount of either DCIS or IDC in two tumors, 16 and 23). DNA was extracted from 16 DCIS and 16 IDC components of invasive tumors and 6 pure DCIS tumors.

**DNA extraction and whole-genome amplification.** DNA was extracted using a Qiagen DNA Mini Kit according to the manufacturer's
<table>
<thead>
<tr>
<th>Pure DCIS</th>
<th>DCIS associated with IDC</th>
<th>IDC associated with DCIS</th>
<th>MCR cytoband (Mb)</th>
<th>Detection rate(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pure DCIS</td>
</tr>
<tr>
<td>1q25.3-pter (179-pter)</td>
<td>1q25.3-pter (179-pter)</td>
<td></td>
<td></td>
<td>67 50 25 43</td>
</tr>
<tr>
<td>2q24.1-q33.1 (155-200)</td>
<td>2q24.1-q33.1 (155-200)</td>
<td></td>
<td></td>
<td>67 31 25 35</td>
</tr>
<tr>
<td>21q11.1-q21.3 (12-29)</td>
<td>21q11.1-q21.3 (12-29)</td>
<td></td>
<td></td>
<td>67 38 19 35</td>
</tr>
<tr>
<td>12q14.3-q23.1 (64-99)</td>
<td>12q14.3-q23.1 (64-99)</td>
<td></td>
<td></td>
<td>50 31 13 27</td>
</tr>
<tr>
<td>Xpter-q13.1 (70-pter)</td>
<td>Xpter-q13.1 (70-pter)</td>
<td></td>
<td></td>
<td>50 13 13 19</td>
</tr>
<tr>
<td>20p13-p11.23 (4-18)</td>
<td>20p13-p11.23 (4-18)</td>
<td></td>
<td></td>
<td>33 25 6 19</td>
</tr>
<tr>
<td>15q23-qter (65-pter)</td>
<td>15q23-qter (65-pter)</td>
<td></td>
<td></td>
<td>17 13 6 11</td>
</tr>
<tr>
<td>21q22.2-pter (40-pter)</td>
<td>21q22.2-pter (40-pter)</td>
<td></td>
<td></td>
<td>67 25 6 24</td>
</tr>
<tr>
<td>Xp11.3-cen (45-60)</td>
<td>Xp11.3-cen (45-60)</td>
<td></td>
<td></td>
<td>67 25 6 24</td>
</tr>
<tr>
<td>9q33.2-qter (121-pter)</td>
<td>9q33.2-qter (121-pter)</td>
<td></td>
<td></td>
<td>83 38 25 41</td>
</tr>
</tbody>
</table>

7p22.1-p15.2 (6-26) *,**
17q21.1-q21.2 (32-36) *,**
19q12-q13.2 (30-44)
10pter-q11.21 (pter-42)
5q11.2-q23.3 (55-131)
9q24.22-qter (134-pter)
14pter-cen (pter-15)
3p24.1-p14.2 (31-61)
20cen-q11.23 (30-36)
1pter-p34.3 (pter-34)
17pter-cen (pter-23) * 1-2 times Preference for pure DCIS

8q13.3-q23.1 (71-110) *
13q21.33-q33.2 (71-105) *
14q12-q23.1 (24-59) *
6p21.1-q24.3 (44-147)
7q11.23-q22.3 (75-106)
17q11.2 (24-28)
1p3.11-p21.3 (69-94)
16p13.13-p12.1 (11-23)
2p13.1-cen (75-90)
8pter-p11.21 (pter-42)
14q24.1-qter (68-pter)
19q13.2-pter (44-qter)
20pter-p13 (pter-5)
6p22.2-p21.1 (26-44)
chrom 22
4pter-p15.33 (pter-13)
14q12-q23.2 (28-61)
2q37.1-qter (232-pter)

4q12-q26 (57-115) *
3q13.31-qter (118-pter) *
9pter-q31.2 (pter-106) *
2pter-p13.3 (pter-70)
12q13.11-q15 (46-66)
17q25.2-qter (73-pter)

17q22-q24.2 (53-63) *,***
3pter-p24.1 (pter-29)*
15q11.1-q15 (22-37)
11p15.2-p13 (13-34)
13pter-q21.32 (pter-67)
11p23.2-qter (133-pter)
11pter-p15.4 (pter-8)
16cen-qter (34-pter)

Detection rate(%) Preference for IDC 1-2 times
44 43 31 43 67 50 25 43 67 31 25 35 67 38 19 35 50 31 13 27 50 13 13 19 33 25 6 19 17 19 6 14 17 13 6 11 67 25 6 24 83 38 25 41
67 44 31 43 50 25 25 30 33 25 19 24 50 12 31 27 50 25 25 30 50 19 31 30 50 38 31 38 50 44 31 41 83 50 44 54
67 75 81 78 50 75 75 65 50 44 63 54 50 69 38 54 67 31 50 46 33 38 38 39 33 25 31 30 33 6 25 19 33 19 25 24 33 19 25 24 50 44 38 43 67 44 50 51 67 69 56 65 17 6 19 14 17 6 19 14 17 19 19 19
33 63 65 59 33 44 50 46 17 44 31 35 17 31 31 30 17 19 31 24 17 19 25 22 17 44 31 35
17 38 38 35 17 31 39 32 0 19 19 16 0 13 19 14 0 6 19 11 0 6 19 11 0 19 25 19 17 31 38 32
17 38 38 35 17 31 39 32 0 19 19 16 0 13 19 14 0 6 19 11 0 6 19 11 0 19 25 19 17 31 38 32
instructions after 72 h in lysis buffer at 56°C. DNA from 12 normal lymph nodes was pooled for the reference sample. The minimal yield of DNA from DCIS foci was 0.2 µg and the median for all samples was 1.3 µg. Because 1.0 µg DNA was required for microarray hybridization, whole-genome amplification was required to perform aCGH. We used the single-cell comparative genomic hybridization protocol for genome amplification, which was originally introduced for analysis of single cells and then validated for use in archival tissue (10). To minimize the effect of possible artifacts introduced by the amplification, all samples including normal reference DNA were amplified, whereas the protocol remained the same regardless of initial DNA yield. Self-set tests were done on single-cell comparative genomic hybridization protocol amplified DNA as well to account for amplification artifacts in determining cutoffs for significant segments.

**aCGH.** The data with detailed protocols are available at the National Center for Biotechnology Information Gene Expression Omnibus Depository, accession no. GSE10197. Repeat cDNA spots, clones with incomplete genomic data, or spots flagged by GenePix software for inadequate quality were removed. Filtering was done in all samples simultaneously to maintain the same set of cDNA clones in all samples. Normalized log₂ ratios of each remaining spot were averaged between duplicates, arranged in genomic sequence (National Center for Biotechnology Information Build 201 and University of California-Santa Cruz Build hg18), centered by the median (median value brought to zero), and saved in a data frame suitable for R.8 The data of breast cancer samples and self-self tests of normal DNA were segmented by a circular binary segmentation algorithm (“DNAcopy” package for “R”; ref. 9). Segmentation of self-set tests revealed that shorter segments containing <20 data points had a significantly larger spread (Fig. 1); therefore, our search for MCRs was based on the segments longer than 20 data points (median of 20 data points 14.7 Mb). The maximum and minimum values of segments longer than 20 data points of the self-set tests were used as cutoffs (0.086 and -0.083 log₂ ratios). When two or more consecutive segments crossed the cutoffs in the same direction (subsegments), they were counted as one segment. MCR were defined as an overlapping region between significant long (>20 data points) segments when it was present in more than two of all DCIS and IDC samples. The ends of significant segments limiting the overlapping region were used as MCR boundaries, whereas MCR length was not limited. A sample was counted as positive for a MCR when it had a significant segment of the same sign equal or larger than the MCR. Few samples were counted as positive when they had segments of borderline significant length (15-20 data points) within short (<20 data points) MCR. This exception was made because the probability of these segments to be significant was higher than that of random borderline-long segments. All other samples were recorded as negative for a MCR.

Because the amount of subsegmentation was different between samples and the median segment value was partially dependent on segment length (Fig. 1), we did not make further distinction by degree of alteration (e.g., amplification versus gain). Instead, peaks of alterations within MCR were measured by continuous moving average of alteration (e.g., amplification versus gain). Instead, peaks of long segments. All other samples were recorded as negative for a MCR.

**Statistical analysis.** Pearson correlation between paired DCIS-IDC samples was calculated using nonsegmented log₂ ratios (filtered, normalized, and centered). Unsupervised cluster analysis of non-segmented data was done by squared Euclidean distance and the weighted pair-group average agglomeration method (XLSTAT, Addinsoft).7 Concordance of segmented data in the paired samples was assessed by counting the number of times MCR detection had the same result in the DCIS-IDC pairs. The exact Mann-Whitney test was used for Q-PCR validation and to assess the association between tumor size, patient age, and detection of 17q22-24 MCR (z = 0.05). The associations between 17q12 gain and HER-2 overexpression, 17q22-24 MCR detection in DCIS and presence of IDC, 17q22-24 MCR detection in DCIS/IDC components and vascular invasion, and tumor grade and number of involved lymph nodes were investigated using the Fisher’s exact test (z = 0.05).

**Q-PCR.** Nonamplified (single-cell comparative genomic hybridization protocol omitted for both test and control samples) DNA was used for real-time PCR (Q-PCR). Tissue remaining in the paraffin blocks used for aCGH was assessed as sufficient to allow validation of 10 genes in 24 samples (5 pure DCIS, 7 DCIS, and 12 IDC). Actual DNA yield allowed testing of all 10 genes in 21 samples and 7 to 8 genes in 3 samples (1 sample from each group). In total, 232 duplicate Q-PCR (40 ng DNA/well) were done using the ABI Prism 7900HT Sequence Detection system and 2× Quantitect SYBR Green PCR Master Mix (Applied Biosystems) according to the manufacturer’s instructions. Pooled DNA of 12 lymph nodes served as a normal control. Genes located within selected MCR (see supplementary material for primer sequences) were assessed by Primer sequences (Operon) and primers designed by Primer Express version 2.0 (Applied Biosystems) using National Center for Biotechnology Information published sequences. Primers were tested for specificity by PCR and running products on a 2% agarose gel. The PCR efficiencies were determined from the standard curve plots using the control DNA. Cycle threshold values were averaged between duplicates and analyzed by the ΔΔCt method. Copy number of the genes was assessed relative to the pooled normal lymph node DNA using β₂-microglobulin as a housekeeping gene located outside of all identified MCR.

**Results**

**aCGH.** 19K human cDNA arrays from the University Health Network Microarray Centre10 contain 19,200 spots with map positions identified for ~11,000 cDNA clones. After filtering, our data set contained 9,820 cDNA clones with the median distance between mapped positions 73.4 kb where 93% of the clones were spaced <1 Mb and 99% of the clones were spaced <3 Mb. We analyzed the data as nonsegmented as well as categorical segmented by the circular binary segmentation algorithm. The latter identifies segments of adjacent statistically uniform data points and was found superior compared with two other aCGH analysis methods (11) and successfully used for analysis validated by other techniques in solid tumors and hematologic malignancies (12–14). To define the extent of nonspecific data variation, we used self-self hybridization of DNA from normal tissues and successfully used for analysis validated by other techniques in solid tumors and hematologic malignancies (12–14). To define the extent of nonspecific data variation, we used self-self hybridization of DNA from normal tissues and measured the level of data variation due to noise of our methods. The data from these tests induced segmentation where segments longer than 20 data points had a more uniform pattern of deviation from the zero line. Segments containing <20 cDNA clones had a distinctly greater degree of deviation inversely proportional to their length (Fig. 1). The criteria to identify MCR were based on these findings (see Materials and Methods).
Analysis of nonsegmented microarray profiles showed similarity between DCIS and IDC foci of the same tumor (median Pearson correlation $r = 0.7$). Unsupervised cluster analysis of nonsegmented data revealed similarity of microarray profiles within the following groups: pure DCIS samples, DCIS and IDC samples from the same tumor, and IDC samples of tumors associated with ipsilateral axillary lymph node metastases. Figure 2 shows that 5 of 6 (83%) pure DCIS samples clustered in one class, 10 of 15 (66%) of DCIS-IDC paired samples clustered in the same class, and 6 of 8 (75%) IDC associated with lymph node metastases clustered in two small classes.

The segmentation analysis identified 57 MCRs among all samples, 31 of DNA copy number gain/amplification, and 26 of loss/deletion. The median number of MCR per sample within the set was 15.5 (range, 3-41) with a median 11 MCR of gain (range, 2-22) and 7 of loss (range, 0-20). The median numbers for pure DCIS, DCIS-IDC, and IDC groups were 22 (range, 6-37), 17 (range, 5-20), and 14.5 (range, 1-17), respectively (differences between the groups not statistically significant). The median concordance rate of MCR detection between DCIS and IDC foci of the same tumor was 80% for both detection and nondetection events.

To analyze segmentation profiles of pure DCIS and IDC, we organized MCRs according to their prevalence in these groups. MCRs were separated into panels using fold difference of MCR detection rates in pure DCIS and IDC (Fig. 3). The most frequent MCRs detected >2 times in IDC were regions of gain 17q22-q24.2 and 3pter-p24.1 and loss of the entire 16q arm (Fig. 3, bottom). These MCRs were detected in 32.4% to 35.1% of all samples, which was close to the median 32.8% detection rate of all MCRs in the set. Gains at 15q11-q15 and 11p15.2-p13 and loses at 11pter-15.4, 11q23.2-qter, and 13pter-q21.32 were detected uniquely in DCIS associated with IDC and IDC although at lower overall rates.

**PCR validation.** Ten MCRs of gain were validated by Q-PCR of genes located within the 9 most frequent MCRs shown in Fig. 3 (bottom) and the MCR centered at ERBB2. The amount of DNA was sufficient to test 232 sample-gene combinations without single-cell comparative genomic hybridization protocol genome amplification. Relative gene quantities were compared between the samples with detected corresponding MCR (expected $>1$) and without detectable MCR (expected $1$). Each gene showed median values higher in the expected group with variable significance. The difference between $>1$ and $1$ groups was statistically significant when all reactions were analyzed together ($P = 0.014$). We did not observe significance for all individual genes, which was attributed to focal variations of DNA copy number and combined noise of microarray and Q-PCR methods (see supplementary material for full Q-PCR data). The median relative gene quantities assessed by Q-PCR were 1.59, 1.84, and 1.47 for all samples, expected $>1$, and expected $1$ groups, respectively. 60% of values in the $>1$ group were higher than the overall median and 57% of values in the $1$ group were lower than the median. Median value $>1$ in the

### Table 1. Clinical variables and their association with the MCR of gain at 17q22-24

<table>
<thead>
<tr>
<th>Case no.</th>
<th>17q22-q24.2+ detected in DCIS</th>
<th>17q22-q24.2+ detected in IDC</th>
<th>17q22-q24.2+ detected in tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure DCIS</td>
<td>0.16</td>
<td>0.93</td>
<td>0.20</td>
</tr>
<tr>
<td>Invasive tumors</td>
<td>0.55</td>
<td>0.68</td>
<td>0.28</td>
</tr>
</tbody>
</table>

**NOTE:** Invasive tumors were sorted by the number of lymph nodes involved by metastases. $P$ value was calculated for the MCR detection separately in DCIS or IDC components (latter when present) and anywhere in the tumor (Fisher’s exact tests for IDC grade, LVI, and number of lymph nodes and Mann-Whitney for age and IDC size). Case 10 had missing data.

**Abbreviations:** LVI, lymphovascular invasion; LN+, number of lymph nodes involved by metastases.

*Excluding two cases of micrometastasis and isolated tumor cells borderline for size criteria.

*Including a case of a 0.2-mm micrometastasis and a case of isolated tumor cells spread over a distance of 3.5 mm (49).
A region centered at 17q23 showed frequent gain with a MCR defined between 53 and 63 Mb (17q22-q24.2). Although all three groups had median values above zero within the region, the MCR was detected in 38% of DCIS associated with IDC, 38% of IDC, and 17% (one sample) of pure DCIS (Fig. 4A). As stated earlier, its detection was associated with prognostic variables. When all samples positive for this MCR were grouped together, a broad region of gain was observed between 45 and 72 Mb with two peaks at 54 to 55 and 60 to 62 Mb (Fig. 4C). When IDCs associated with lymph node metastases were grouped together, they differed from the remaining IDCs by the gain between 60 and 64 Mb (Fig. 4D).

Loss of the region distal to 73 Mb (7q25.2-qter) was 1.87 times more frequent in IDC than pure DCIS and had a variable proximal boundary in samples with gain on 17q22-q24.2 extending distally to 73 Mb in some samples.

**Discussion**

In our material, the genomic profiles of pure DCIS and DCIS associated with IDC showed a degree of dissimilarity between these two types of DCIS, whereas the samples showed similarity within the groups. Additionally, we found similarity of genomic alterations of paired DCIS and IDC samples from the same tumor as has been observed by other investigators (5, 6, 15, 16). In case of gain on 17q22-24 that was similarly detected in DCIS and IDC of the same tumors, it was associated with poor prognostic factors, which excludes similarity due to individual polymorphisms. These data suggest that DCIS associated with IDC is genomically different from pure DCIS and may, at least partially, represent either a preinvasive clone with high invasive potential or an invasive cancer spreading through the duct system. It is possible to hypothesize that the

**Table 1. Clinical variables and their association with the MCR of gain at 17q22-24 (Cont’d)**

<table>
<thead>
<tr>
<th>LN*</th>
<th>LN+</th>
<th>Age</th>
<th>ER</th>
<th>PR</th>
<th>HER-2</th>
<th>DCIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td>2-3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td>Solid, cribriform, clinging +</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td>2-3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td>Solid, cribriform +</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td>Solid, cribriform, micropapillary +</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>86</td>
<td>+</td>
<td>+</td>
<td>+/-</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>46</td>
<td>+</td>
<td>-</td>
<td>+/-</td>
<td>2-3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Missing data</td>
<td></td>
<td></td>
<td></td>
<td>Solid, cribriform +</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>61</td>
<td>+</td>
<td>+/-</td>
<td>-</td>
<td>2-3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>77</td>
<td>+</td>
<td>+/</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>48</td>
<td>+</td>
<td>+</td>
<td></td>
<td>2-3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>44</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>41</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>2-3</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>68</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>Solid, cribriform</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>59</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>Solid, cribriform +</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>42</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>61</td>
<td>-</td>
<td>-</td>
<td>+/-</td>
<td>2-3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>52</td>
<td>-</td>
<td>-</td>
<td>+/-</td>
<td>2-3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>44</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Solid, cribriform, micropapillary</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>65</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Insufficient amount of DCIS for aCGH</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>55</td>
<td>+</td>
<td>+</td>
<td>+/-</td>
<td>Solid, cribriform +</td>
</tr>
<tr>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

~1 group suggests a trend for false-negative rather than false-positive MCR detection during aCGH analysis. This was expected because the cutoffs of nonspecific segmentation were set at the maximum rather than a percentile of self-self segment values.

**Correlation with clinical data.** Regions 17q12-21.2 and 17q22-24.2 showed associations with relevant clinical data. Detection of the MCR at 17q12-21.2 (32-36 Mb) in IDC correlated with positive immunostaining for HER-2/neu (P = 0.048) assessed by pathologists in IDC using routine diagnostic protocols (Novocastra monoclonal CB-11 antibody). Gain at 17q22-24.2 (53-63 Mb) was associated with higher histologic grade, tumor size, lymphatic/vascular invasion, and number of lymph nodes involved by metastatic spread (Table 1). Detection of this MCR either anywhere in the tumor or separately in DCIS and IDC components was associated with lymph node metastasis. There was a trend to detect gain on 17q22-24.2 in tumors of younger, premenopausal patients: median age 44 and 57 years for the MCR-positive and MCR-negative patients, respectively (Table 1). Other clinical and morphologic variables did not show significant associations.

**Analysis of chromosome 17.** Loss of 17p was commonly detected without preference for any group. Genomic loss in the region was not detected by segmentation in the samples with gain of whole chromosome 17; however, these samples showed at least partial loss of the arm by moving average. The lowest median values were between 4 and 7.5 Mb, which includes the location of p53 (Fig. 4A).

The region of genomic gain peaking at 35 Mb, accepted as containing ERBB2, was shorter and more centered around this gene in pure DCIS. This MCR was detected in 67% of pure DCIS and produced a distinct peak (Fig. 4A, blue moving average). The IDC samples with detectable gain of the 17q12-21.2 MCR also showed a peak around 35 Mb although of lower amplitude (Fig. 4B).
latter may occur in de novo IDC where invasive cancer evolves without a DCIS stage but mimics it spreading through the ducts.

We observed a trend for pure DCIS to have a degree of genomic complexity not less than that of invasive cancers. The finding of a high number of genomic alterations in pure preinvasive breast cancer supports the previous conclusions of in vitro mathematical modeling and loss of heterozygosity analyses (15, 17, 18). Despite its high genomic complexity, its microdissected multiple tumor foci were shown to be genetically more homogeneous than DCIS associated with IDC (16). Relative homogeneity of pure DCIS throughout the duct system was also shown by clonality of primary pure DCIS and matched recurrences (19). Higher clonal homogeneity of pure DCIS and heterogeneity of DCIS associated with IDC is in keeping with the idea that the latter type of DCIS is likely to be partially represented by more aggressive invasive or near-invasive clones.

Fig. 4. Analysis of chromosome 17. X axis, position (in Mb); Y axis, log2 ratio. Moving average of 20 data points (median distance 14.7 Mb) was used to visualize median nonsegmented data of specific groups (continuous waveform lines). A, segmentation and moving average in three groups: pure DCIS, DCIS associated with IDC, and IDC. Segments of gain are represented by horizontal lines above the X axis of moving averages; segments of loss are represented by horizontal lines below the X axis of moving averages. Note distinct peak around ERBB2 in pure DCIS. B, pure DCIS and IDC samples with detected gain on 17q12-21.2. Pure DCIS had a tall peak at 35 Mb, whereas it was much lower in IDC samples. C, samples with gain on 17q22-24 versus all others. The difference is pronounced between 45 and 72 Mb with two peaks at 54 to 55 and 60 to 62 Mb. D, IDC samples grouped by lymph node status. Lymph node–positive IDC as a group had gain between 60 and 64 Mb, which was not observed in tumors without metastases.
Microarray technology generates large data sets that require choices to be made to filter nonsignificant data or noise. Segmentation of aCGH data is aimed to merge statistically uniform individual data points into contiguous segments; however, segments need to be filtered by their median log2 values. We addressed this issue experimentally by self-self hybridization of normal DNA. We found that the amplitude of nonspecific segmentation is dependent on segment length measured by data points and is more pronounced for segments containing <20 data points, which we attributed to technical noise and spontaneous data clustering. This strategy may prove useful in assessing reliable resolution of aCGH platforms and the number of probes per gene in expression arrays.

We identified several regions that were either preferentially or uniquely detected in DCIS associated with IDC and IDC (Fig. 3). When we tested all MCR for correlation with clinical variables, two MCR had significant associations. Gain at 17q12-21.2 was predictive of ERBB2 overexpression (P = 0.048), which was expected and served as an additional validation test. This region was amplified in 67% of pure DCIS compared with 31% and 44% for DCIS associated with IDC and IDC. The degree of amplification in pure DCIS was also higher than that of DCIS associated with IDC and IDC (Fig. 4A). As reported by others, ERBB2 amplification with overexpression is seen in ~50% of pure DCIS with lower rate in DCIS associated with IDC and not more than 30% of IDC (20, 21). These data indicate that the invasive switch or alternatively development of IDC de novo is likely to occur on the background of a sustained degree of ERBB2 gain, which may reflect selection of slower proliferating cells by metabolic challenges (18).

We expected to find an alteration or a set of alterations associated with lymph node involvement after we observed that lymph node–positive IDC clustered closely (Fig. 2). Gain at 17q22-24.2 showed a significant association with lymph node status and other important prognostic factors (Table 1: P < 0.0001 for lymph node metastasis). This MCR was also detected in DCIS of lymph node–positive tumors (P = 0.01), which further indicates that this type of DCIS contains cells genotypically similar to adjacent IDC. Practical utility of this finding is that DCIS can be a source of information regarding IDC behavior and existence of invasion. However, prognostically useful cells may be present focally because DCIS associated with IDC was shown previously to display intratumoral heterogeneity (16), which is also supported by the lower significance of detection of this MCR in DCIS compared with IDC samples from the same tumors.

Gain at 17q23 (within 17q22-24.2 MCR in our data set) was detected previously in a range of human malignancies including medulloblastoma, rhabdomyosarcoma, hepatocellular, and carcinoma of the prostate, pancreas, and bladder (22–27). It is predictive for prognosis in neuroblastoma, ovarian clear cell carcinoma, pancreatic adenocarcinoma, and anaplastic transformation of meningiomas and differentiates dermatofibrosarcoma protuberans from its benign counterpart, dermatofibroma (23, 28–31). It is most frequent in breast tumors with percentages of detection ranging between 18% and 68% (28, 32–35) and is highly prevalent (>75%) in BRCA2 tumors (4). A study screening 3,520 specimens detected gain of 17q23 in tumors from 40 different tissue types where 78% of high-level amplifications were detected in breast tumors (36). Its recurrent nature in a wide range of neoplasms may be related to the fragile site FRA17B at 17q23.1. Our finding that gain at 17q22-24.2 is associated with metastasis supports earlier reports of its detection in 14% of primary breast tumors and 36% of breast cancer metastases (36) and detection at a lower frequency (<10%) in node-negative breast cancer (34). It was also identified as predictive of worse prognosis and lymph node metastases when detected in IDC (32). In addition to the evidence from clinical samples, experimental work shows that 17q23 is involved in cancer progression. The prostatic cancer cell line, PC3M, retained the gain of 17q22-q23 after four successive orthotopic implantations, which was thought to be due to selective pressure of growth in the lymph node (37).

To define locations of potential targets in the region, we grouped all samples positive for the 17q22-24.2 MCR (53-63 Mb) and observed a broad region of gain between 45 and 72 Mb with two peaks at 54 to 55 and 60 to 62 Mb (Fig. 4C). Other studies of primary breast tumors and cell lines report the following boundaries and number of distinct amplicons: two regions 52.5 to 54.2 and 57.9 to 59.8 Mb (13); one region 49 to 78.4 Mb (3); one region 58.1 to 60.4 Mb (38); two regions 53.8 to 56.6 and ~59.2 Mb (39); and three regions 52.47 to 55.8, 63.8 to 69.7, and 69.93 to 74.99 Mb (40). Thus, it appears that the region contains at least two distinct recurrent amplicons with the broadest boundaries estimated as 52.47 to 56.6 and 57.9 to 63.3 Mb, which is in agreement with the two peaks we observed at 54 to 55 and 60 to 62 Mb. When we grouped all lymph node–positive IDCs, they differed from the other IDCs by the gain between 60 and 64 Mb, which suggests that the distal amplicon may have a stronger association with metastatic phenotype (Fig. 4D); however, to date, it has attracted less attention in the literature than the proximal amplicon at 17q23.

Several candidate genes within 17q23 were found amplified and overexpressed, of which RPS6KB1, TRIM37, TXB2, PPP1TD, and APPBP2 are the most frequently reported (39–41). RPS6KB1, PAT1, and TXB2 were also found coamplified in 10% of tumors (41). A recent study of primary breast tumors revealed that overexpression of 11 genes in the region of 54.5 to 56.1 Mb (FAM33A, DHX40, CLTC, PTHR1, TEMEM49, TUBD1, RPS6KB1, ABC1, USP32, APPBP2, and PPM1D) is activated by high level of amplification and was not observed with low-level copy number gain (42). These data are in agreement with our finding that, although gain in the region was common (by moving average), only distinct amplification detectable by the segmentation algorithm was associated with prognostic variables.

We found that 88% of samples with gain at 17q22-24.2 showed synchronous gain on 17q12-21.1, which may provide a basis for a synergistic effect of the genes in these regions in aggressive breast cancer phenotypes (Fig. 4C). Coamplification of 17q12-21 and 17q23 was observed previously in ERBB2-positive breast cancers (28, 43). Coamplification of placental lactogens located at 17q24 and ERBB2 was observed in 12% of breast tumors, where 86% of these tumors were lymph node positive compared with 59% in the set (44). Genes other than ERBB2 may also be affected by the amplification. For example, Pip4k2β, GRB7, MLN64, and EST 48582 were coamplified and overexpressed along with ERBB2 within 17q11-12 (45, 46).

Loss of 17p that contains well-known p53 was detected in >50% of samples, which is in agreement with studies of spontaneous breast cancer (2, 38) and BRCA1/BRCA2 hereditary tumors (4). Loss of the opposite subtelomeric 17q occurred almost twice as often in DCIS associated with IDC and IDC compared with pure
node metastasis. Thus, it is a candidate region for further testing in a larger data set as a predictor of invasion when detected in DCIS as well as a potential predictor of nodal metastasis.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

We thank Dr. Jeremy Squire (Ontario Cancer Institute/Princess Margaret Hospital) for reading this article and for helpful suggestions.

**References**


