

High-Resolution Genomic Profiling Reveals Association of Chromosomal Aberrations on 1q and 16p with Histologic and Genetic Subgroups of Invasive Breast Cancer

Daniel E. Stange,¹ Bernhard Radlwimmer,¹ Falk Schubert,² Frank Traub,³ Andreas Pich,³ Grischa Toedt,¹ Frank Mendrzyk,¹ Ulrich Lehmann,³ Roland Eils,² Hans Kreipe,³ and Peter Lichter¹

Abstract **Purpose:** Invasive ductal carcinoma and invasive lobular carcinoma (ILC) represent the major histologic subtypes of invasive breast cancer. They differ with regard to presentation, metastatic spread, and epidemiologic features. To elucidate the genetic basis of these differences, we analyzed copy number imbalances that differentiate the histologic subtypes. **Experimental Design:** High-resolution genomic profiling of 40 invasive breast cancers using matrix-comparative genomic hybridization with an average resolution of 0.5 Mb was conducted on bacterial artificial chromosome microarrays. The data were subjected to classification and unsupervised hierarchical cluster analyses. Expression of candidate genes was analyzed in tumor samples. **Results:** The highest discriminating power was achieved when combining the aberration patterns of chromosome arms 1q and 16p, which were significantly more often gained in ILC. These regions were further narrowed down to subregions 1q24.2-25.1, 1q25.3-q31.3, and 16p11.2. Located within the candidate gains on 1q are two genes, *FMO2* and *PTGS2*, known to be overexpressed in ILC relative to invasive ductal carcinoma. Assessment of four candidate genes on 16p11.2 by real-time quantitative PCR revealed significant overexpression of *FUS* and *ITGAX* in ILC with 16p copy number gain. Unsupervised hierarchical cluster analysis identified three molecular subgroups that are characterized by different aberration patterns, in particular concerning gain of *MYC* (8q24) and the identified candidate regions on 1q24.2-25.1, 1q25.3-q31.3, and 16p11.2. These genetic subgroups differed with regard to histology, tumor grading, frequency of alterations, and estrogen receptor expression. **Conclusions:** Molecular profiling using bacterial artificial chromosome arrays identified DNA copy number imbalances on 1q and 16p as significant classifiers of histologic and molecular subgroups.

Invasive ductal carcinoma (IDC) represents the predominant histologic subtype of breast cancer, constituting 40% to 75% of mammary carcinomas, whereas invasive lobular carcinoma (ILC) ranges second in frequency and accounts for 5% to 15% of cases. Besides differences in the histopathologic morphology,

ILC and IDC differ with regard to clinical and epidemiologic features. There has been a steady and disproportionate increase in the incidence of ILC compared with IDC in women over 50 years during the last 20 years, which has been attributed to the increased use of hormone replacement therapy (1, 2). The molecular basis for this development as well as for differences in phenotype and clinical behavior between ILC and IDC is not yet understood.

The molecular portrait of breast cancer was previously explored by a number of groups (3, 4), with two recent studies specifically comparing the gene expression profiles of IDC and ILC (5, 6). Using unsupervised cluster analysis, Zhao et al. (5) were able to subdivide ILC into a "typical" and a "ductal-like" subgroup. The typical ILC showed expression patterns similar to those of tumors that previously were grouped into a normal-like subgroup (4) because their expression profiles are more similar to normal than to cancerous breast tissue. Korkola et al. (6) were unable to distinguish ILC and IDC by unsupervised hierarchical clustering of gene expression data; however, they also identified a group of tumors, which was similar to the normal-like subgroup (4) and was characterized by low expression of proliferation associated genes. Using supervised statistical methods, sets of genes distinguishing ILC and IDC

Authors' Affiliations: Divisions of ¹Molecular Genetics and ²Theoretical Bioinformatics, German Cancer Research Center, Heidelberg, Germany; and ³Institute of Pathology, Medizinische Hochschule Hannover, Hannover, Germany Received 7/27/05; revised 10/24/05; accepted 11/7/05.

Grant support: Deutsche Forschungsgemeinschaft, Graduiertenkolleg 886 (D.E. Stange), and Bundesministerium für Bildung und Forschung grants 01GR0417, 01GS0460 (B. Radlwimmer and P. Lichter), and 01GR0417 (P. Lichter).

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.

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

Requests for reprints: Peter Lichter, Division of Molecular Genetics (B060), German Cancer Research Center, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany. Phone: 49-6221-42-4619; Fax: 49-6221-42-4639; E-mail: m.macleod@dkfz.de.

© 2006 American Association for Cancer Research.
doi:10.1158/1078-0432.CCR-05-1633

were defined in both studies. Only three genes (*CDH1*, *FHL1*, and *ADH1C*) overlap between both gene signatures (33 and 378 genes in size).

Although gene expression profiling has not allowed reliable classification of IDC and ILC, independent evidence for genetic differences between these tumor types comes from analyses of DNA copy number using chromosomal comparative genomic hybridization (CGH). These studies revealed correlation of gain of 8q and 20q with IDC and loss of 16q and loss of 22q with ILC (7–9). Additionally, preferential loss of 17q and 5p, both of which occur at low frequency in breast cancer, were reported for ILC by one study each (7, 9).

In recent years, CGH was increasingly replaced by CGH to bacterial artificial chromosome (BAC) microarrays (matrix-CGH; refs. 10, 11), which uses large genomic DNA fragments that are arrayed onto glass slides instead of metaphase chromosomes. This method offers superior resolution, which is only limited by the length and the spacing of the genomic fragments used (12, 13). Recently, matrix-CGH was applied for the first time to compare IDC and ILC (14). Differences in the frequency of copy number imbalances were detected on 1q and 11q; however, neither of these imbalances reached statistical significance. Here, we report the high-resolution genomic profiling of 40 invasive breast cancers using matrix-CGH with an average resolution of 0.5 Mb. The data were subjected to classification and unsupervised hierarchical clustering analyses to improve our understanding of the molecular differences between IDC and ILC.

Materials and Methods

Tumor material. Fresh frozen material from 46 breast tumors was obtained after informed consent and stored in an anonymous fashion according to an approval of the local ethics committee of the Medizinische Hochschule Hannover. The tumors were classified histologically as IDC (18 cases), ILC (21 cases), and one case with overlapping features of ILC and IDC. Six further cases were excluded from the analysis after repeated hybridizations because the tumor materials were consistently not analyzable. Sampling period encompassed the years 1998 to 2003. Histopathologic variables are summarized in Supplementary Table S1.

Preparation of BAC microarrays and labeling. A set of ~3,200 sequence-verified DNA genomic fragments (15, 16) covering the genome at ~1 Mb resolution was kindly provided to us by the Mapping Core and Map Finishing groups of the Wellcome Trust Sanger Institute (Nigel Carter and Heike Fiegler). Additionally, 3,000 gene- and region-specific genomic fragments were ordered from the RZPD (Berlin, Germany) and CalTech BAC libraries (Invitrogen, Karlsruhe, Germany). A complete list of genomic fragments spotted on the array will be available upon publication on Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under the platform accession number GPL1432. The chromosomal mapping information was based on the Ensembl (version 17) or the University of California at Santa Cruz genome database (Freeze, July 2003) and for regions of interest a more recent version of Ensembl was used (v31.35d). Detailed description of target DNA preparation and labeling procedures is described in detail in Zielinski et al. (17). Briefly, genomic DNA from tumor tissue and blood of healthy donors was isolated using the Blood and Cell Culture kit (Qiagen, Hilden, Germany) following the instructions of the suppliers. Tumor DNA and sex-matched reference DNA (pooled DNA from four healthy individuals) were labeled differentially using the Bioprime Labeling kit (Invitrogen).

Hybridization to microarrays. Ten to 15 µg of labeled test and sex-matched reference DNA each, plus 100 to 150 µg of human Cot1-DNA

(Roche, Mannheim, Germany), were precipitated and resuspended in 130 µL of hybridization buffer (ULTRAhyb, Ambion, TX). Samples were denatured for 10 minutes at 75°C and reannealed 1 hour at 42°C. Rubber cement was applied around the array to enclose an area of 2 × 3 cm. Arrays were prehybridized as follows: 750 µg of salmon sperm (Invitrogen) was precipitated, resuspended in 180 µL of hybridization buffer, denatured for 10 minutes at 75°C, and then added to the array. The array was placed in a humidity chamber on a rocking table at 5 rpm at 37°C for 60 minutes. The prehybridization solution was then removed and replaced by the solution containing reannealed genomic DNA and the array was transferred in a humidity chamber on a rocking table at 5 rpm and incubated at 37°C for 48 hours. The hybridization solution was washed away with 2× SSC/0.05% Tween 20; slides were then incubated for 2× 15 minutes with 50% formamide, 2× SSC, 0.1% Tween 20 at 43°C, followed by 15 minutes wash in 2× SSC, 0.05% Tween 20 at 43°C, and 10 minutes in 1× PBS, 0.05% Tween 20 at room temperature. All washing buffers were adjusted to pH 7.0. The slides were dried by centrifugation.

Image and data analysis. Arrays were scanned with an Axon 4000B scanner (Axon Instruments, Burlingame, CA) and images were analyzed using GenePix Pro 4.0 software (Axon Instruments). Fluorescence intensities of all spots were filtered (intensity/local background >3; mean/median intensity <1.3; SD of genomic fragment log ratios <0.25) and normalized block-wise according to Loess. Chromosomal breakpoints delimiting regions were then detected by Gain and Loss Analysis of DNA, a method developed by Hupé et al. (18) based on the Adaptive Weight Smoothing procedure. The variables of Gain and Loss Analysis of DNA were adjusted through several hybridizations of normal proband-DNA against pool-DNA from five normal probands as negative controls and cell line experiments with well-known genomic aberrations as positive controls. The threshold differentiating balanced from altered regions were 1.12 for gains and 0.88 for losses and were set in such a way that no false-positive region was found in the control hybridizations.

Significance analysis of microarrays. Significance analysis of microarrays (SAM) was done for discretized copy number ratios using an implementation of SAM for categorical variables with a false discovery rate of <5% (kindly provided by Holger Schwender, Department of Statistics, Collaborative Research Centre 475, University of Dortmund, Germany). All copy number values were discretized and encoded as -1 for a deletion (ratio <0.88), 0 for a balanced genomic fragment, and +1 for a gain (ratio >1.12). This procedure differs from the original SAM in the use of the χ^2 instead of *t* test statistics (19). SAM was done on all genomic fragments, which showed aberrations in at least 20% of all cases. This preselection of genomic fragments was done independent from the class label and a simulation showed that the number of false-positive genomic fragments is still controlled.

Support vector machines. The support vector machine (SVM) implementation LIBSVM was used as classifier in a nested leave-one-out cross-validation design with a grid variable search within the inner leave-one-out cross-validation loop (20–22). The SVM was used on discretized copy number ratios using an RBF kernel. The threshold differentiating balanced from aberrated genomic fragments was chosen for each run inside the cross-validation automatically (by a maximally selected statistics) to avoid overfitting. Inside the cross-validation were only genomic fragments that showed aberrations for at least 10% of all cases. The separating chromosomal regions were calculated from the trained SVM classifier according to the absolute value of each component of the hyperplane direction vector. The most important features were calculated by a recursive feature elimination (23).

Unsupervised hierarchical clustering. The clustering of all ILC and IDC was done for the Manhattan distance and Ward's linkage with discretized copy number ratios (24). Associations of histopathologic variables and number of aberrations with the clustering were tested by Fisher's exact test and Kruskal-Wallis test. An association of each clustering subgroup with histopathologic variables and number of aberrations was tested by using an exact Wilcoxon signed rank test

(corrected by Hochberg). All analyses were done in the open source statistical environment R, version 1.91 (<http://www.r-project.org>).

RNA extraction and quantitative real-time PCR. Extraction of RNA from fresh-frozen tissue using TRIZOL (Invitrogen) and quantitative real-time reverse transcription-PCR analysis were done as described in Lehmann et al. (25). For quantification of mRNA levels, primers were designed for *RNF40* (5'-TGGCCACAAGAAGACTCCCAC-3' and 5'-ACAGAGGTCATTACGCTGGAGG-3'), *BCL7C* (5'-GGCCAAGGATGACATCAAG-3' and 5'-GACTGTGGGCGACACTCC-3'), *FUS* (5'-GCCAGAGCTCCAATCG-3' and 5'-ACCTCGGAAGTTACGGTA-3'), and *ITGAX* (5'-CTGAGGAGAAGGAAAGCCATG-3' and 5'-GACCTGCTGTGACATCAA-3'). Expression levels of mRNA were normalized to the housekeeping genes coding for β_2 -microglobulin and TATA box binding protein essentially as described (26, 27).

Results

Chromosomal imbalances in IDC and ILC. Forty invasive breast cancer samples were hybridized on a DNA microarray, composed of 6,212 genomic fragments, resulting in genomic profiles with an average resolution of ~ 0.5 Mb. We found chromosomal imbalances in all of the 40 invasive breast cancers under study. The mean number of genomic fragments showing DNA copy number gain in IDC and ILC was 8.5% (range 2.7-15.7%) and 6.9% (1-15%), the mean number of those showing losses 4.4% (0-20%) and 7.1% (0-18%), respectively. For every genomic DNA fragment, the percentage of tumors affected by either a gain or a loss was plotted separately for the two histologic subtypes (Fig. 1A and B). The overall concordance of recurring aberrations in these subtype-specific frequency plots was high. In both subtypes, the most frequently observed gains were seen on 1q, 8q, 11q, and 17q, and losses were commonly

seen on 8p, 11q, 13q, 16q, and 22q. For a more detailed analysis, a frequency difference plot was generated by calculating the absolute difference between the frequency plots (Fig. 1C). The subtype-specific frequencies and the approximate sizes (based on Ensembl v31.35d) of aberrations showing a difference of $\sim 20\%$ or more between IDC and ILC are shown in Table 1.

Statistical analysis of microarrays and tumor classification. To identify regions that are important for the discrimination of IDC and ILC by independent and rigorous biostatistical methods, we did SVM analysis to identify the genomic fragments best suited to classifying tumors as either IDC or ILC, and SAM to identify the fragments that are imbalanced with significantly different frequencies in the histologic subtypes. The 128 top-ranked fragments selected by SVM map to chromosomal regions on 1q and 16p, identifying them as the most significant discriminators of IDC and ILC. An optimal classification accuracy of 65% was achieved using a classifier consisting of 733 genomic fragments. These fragments map to 11 chromosomal regions (Ensembl v31.35d), which are listed in Table 2A in order of average fragment rank assigned by SVM. Using an implementation of SAM specifically adapted to the analysis of DNA copy number data, 116 genomic fragments were identified that cluster in two regions, one on chromosome 1q24.2-q31.3 and the other on 16p11.2 (Table 2B).

Candidate genes on 1q and 16p. The regions of interest on chromosome arms 1q and 16p, as defined by SAM and SVM analysis, were still quite large. To narrow these regions and facilitate the identification of candidate genes, we analyzed the matrix-CGH data of these regions using empirical methods. Specifically, we could identify local peaks in the frequency difference plot (Fig. 1C) on 1q and define minimal overlapping

Fig. 1. Frequency of copy number changes in IDC and ILC. The frequency of copy number gain (green) and loss (red) was calculated for all of the 6,200 genomic fragments on the array and plotted against their genomic position on the chromosomes for all tumors. This was done separately for the two histologic subtypes IDC (A, $n = 18$) and ILC (B, $n = 21$). The absolute difference of the subtype-specific frequencies was calculated by subtracting the frequency of one subtype from the other (C). Gains were assigned positive values and losses were assigned negative values. Arrows, locations of discussed genes.

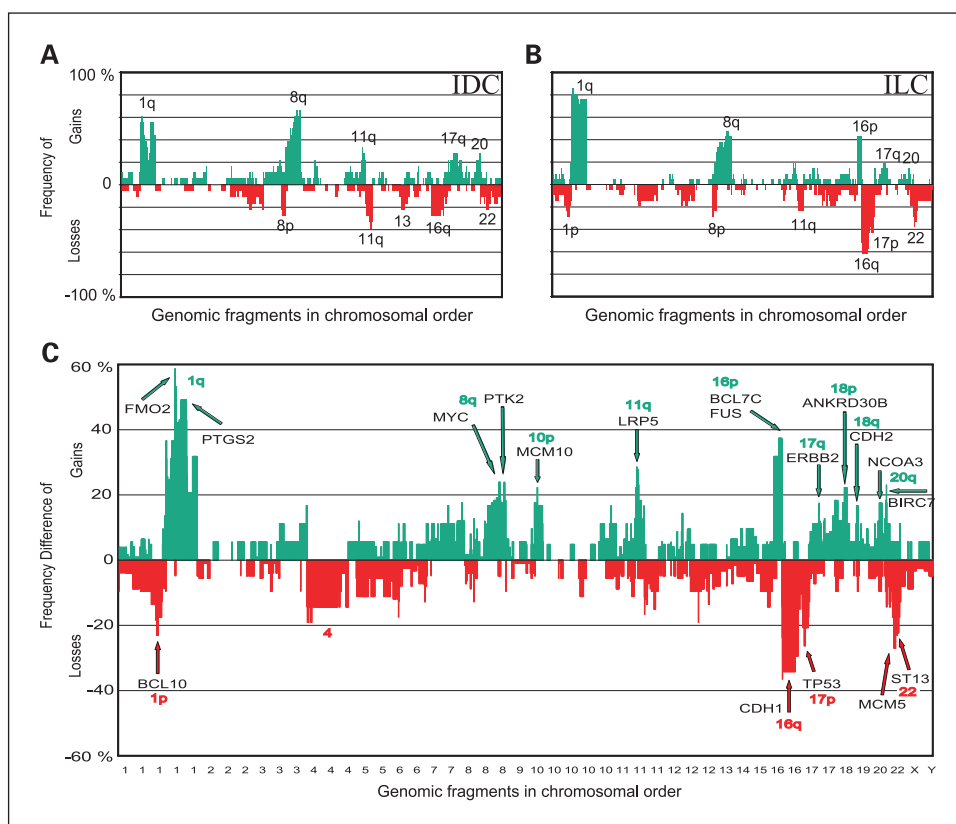


Table 1. Chromosomal regions with frequency differences between IDC and ILC

Chromosome	Start (Mb)	End (Mb)	Size (Mb)	BAC start	BAC end	IDC (%)	ILC (%)	Average difference (%)
-1p22.3-p22.2	84.8	89.2	4.4	RP11-484D4	RP11-82K18	5.6	28.6	23.0
+1q24.2-q25.1	166.6	170	3.4	RP1-190I16	RP3-395P12	25.9	76.2	50.3
+1q25.3-q31.3	180.4	196.4	16	RP11-293B7	RP11-152M20	23.3	72.2	48.8
+8q24.13-q24.21	124.4	128.5	4.1	RP11-468O2	RP11-382A18	66.7	42.9	23.8
+8q24.3	141.6	142.4	0.8	RP11-642A1	RP11-10J21	66.7	42.9	23.8
+10p13	0.3	13.4	13.1	CTC-306F7	RP11-425A6	22.2	4.8	17.4
+11q13.2-q13.3	67.7	68.7	1	RP11-856M9	RP11-554A11	23.6	0.6	23.0
+11q13.4	7	70.8	0.8	RP11-826F13	RP11-512I24	28.6	5.4	23.1
+16p11.2-p13.12	12	33.7	21.7	RP11-276H1	RP11-274A17	6.5	42.0	35.5
-16q	45.1	88.6	43.5	RP11-5L1	CTB-121I4	27.3	60.2	32.9
-17p13.1-p12	6.7	14.6	7.9	RP11-530N7	RP11-64B12	21.1	42.9	21.8
+17q12	34.5	35.1	0.6	RP5-906A24	RP11-62N23	22.2	4.8	17.5
+17q25.1-q25.3	69.8	78.4	8.6	RP11-478P5	RP11-567O16	27.8	9.5	18.3
+18p11.21	14.5	16.9	2.4	RP11-19M12	RP11-666N19	22.2	0.0	22.2
+18q11.2-q12.1	21.4	24.2	2.7	RP11-5G23	RP11-430E17	22.2	0.0	22.2
+20q13.12-q13.13	42.8	46.1	3.3	RP4-781B1	RP11-347D21	22.2	4.8	17.5
+20q13.33	60.1	62.4	2.2	RP5-1005F21	CTB-81F12	26.4	7.7	18.7
-22q12.3	30.8	36	5.2	RP1-127L4	RP1-151B14	11.1	37.7	26.6
-22q13.2	39.5	41.6	2.1	RP3-362J20	RP3-437M21	9.8	31.9	22.0

regions of copy number gain by overlaying the profiles of all tumors displaying a gain on 16p (data not shown). Applying these approaches, intervals of ~3.4 and 16 Mb (Table 1), and ~790 kb (30.5-31.3 Mb) could be delineated on 1q and 16p, respectively (Fig. 2). The regions on 1q contain two genes (*FMO2* and *PTGS2*) that previously had been identified as overexpressed in ILC compared with IDC (5). On 16p, no potential oncogenes were described in the identified region thus far.

Based on gene localization and annotated gene function, we therefore selected four candidate genes (*RNF40*, *BCL7C*, *FUS*, and *ITGAX*) and measured their expression level in three different tumor groups by real-time quantitative PCR: ILC with

16p gain ($n = 10$), ILC with balanced 16p ($n = 11$), and IDC ($n = 18$; Fig. 3). One gene, *ITGAX*, showed moderate overexpression in ILC with 16p gain and IDC compared with ILC with balanced 16p. The other three genes (*RNF40*, *BCL7C*, and *FUS*) showed overexpression in ILC with 16p gain compared with IDC and ILC with balanced 16p. Expression levels in the groups with balanced 16p were comparable. *RNF40* was moderately overexpressed in ILC with 16p gain. *BCL7C* and *FUS* were significantly overexpressed ($P < 0.001$) by a factor of $5\times$ to $10\times$ comparing the median expression level of ILC with 16p gain to tumors with balanced 16p. Of the 10 ILC with 16p gain under study, eight showed a $>2.5\times$ and six showed a $>4.5\times$ overexpression of *BCL7c* compared with ILC and IDC

Table 2. Regions ranked by SVM (A) and identified by SAM (B) analyses

Chromosome	Start (Mb)	End (Mb)	Size (Mb)	BAC start	BAC end
A					
+1q12-q44	141.4	245.4	104	RP3-365I19	CTB-160H23
+16p11.2-p13.3	0	33.7	33.7	RP11-344L6	RP11-274A17
+11q13.3-q13.4	68.5	70.4	1.9	RP11-554A11	RP11-21D20
+17q12-q21.2	35	35.9	0.9	RP11-62N23	RP11-58O9
+17q22-q25.3	49	78.4	29.4	RP11-312B18	RP11-567O16
+11q13.2	66.5	67	0.5	RP11-699D4	RP11-678D20
+11q13.4-q13.5	70.8	76.2	5.4	RP11-660L16	RP11-30J7
+8p11.21-q24.3	40.5	146.2	105.7	RP11-51K12	RP5-1056B24
+20q13.13-q13.33	47.1	62.4	15.3	RP1-155G6	CTB-81F12
+1p13.1	116.5	117.3	0.8	RP4-787H6	RP11-27K13
+14q32.33-q32.33	104.1	105.1	1	RP11-982M15	RP11-521B24
B					
+1q24.2-q32.1	166.3	196.9	30.6	RP4-780M13	RP11-47A17
+16p11.2-p13.12	12	33.7	21.7	RP11-276H1	RP11-274A17

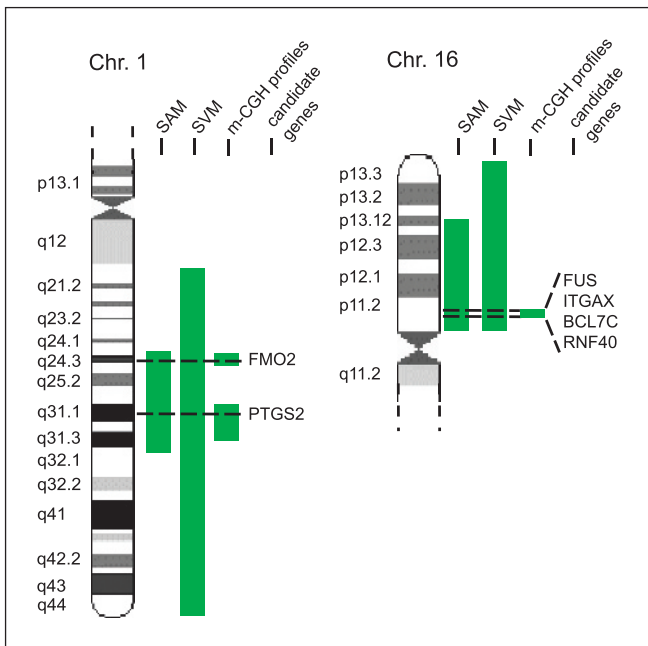


Fig. 2. Regions on 1q and 16p identified by different analysis methods. The different regions identified by SVM and SAM were compared with matrix-CGH profiles and plotted against the ideograms of chromosomes 1 and 16. The localization of candidate genes is indicated.

without 16p gain. For *FUS*, eight showed a $>2.5\times$ and seven showed a $>3.75\times$ overexpression.

Molecular subclassification by unsupervised hierarchical clustering. Unsupervised hierarchical clustering was applied to detect molecular subgroups with similar patterns of chromosomal aberrations. The tumors clustered into three main groups (Fig. 4). Alterations characterizing the different clusters were located on chromosome 1q, 8p, 8q, 11q, 16p, 16q, and 22q. Group I tumors ($n = 11$) were characterized by a gain of chromosome 8q24.13-q24.21 in $>90\%$ of cases. Group II tumors ($n = 15$) showed a gain of 1q21.1-q23.1 and 1q32.1-q44 in $>90\%$ and a loss of complete 16q in 40% of cases. Group III tumors ($n = 14$) showed a combination of the patterns found in group I and II with gain on 1q in 70%, gain on 8q in 85%, and loss of 16q in 90%. Additionally, group III showed high frequencies of gain of 16p (65%) and 17q (43%) as well as losses on 1p (34%), 6q (35%), 8p (57%), 11q (64%), 13q (36%), 17p (60%), and 22q (64%).

We further analyzed associations between groups I, II, and III and histopathologic as well as immunohistochemical variables (Table 3). Lobular cancers mainly clustered in groups II and III, which together contained $>90\%$ of the ILC cases. In contrast, IDC was overrepresented in group I ($n = 9$ of 11, $P = 0.02$) compared with the other two groups. Besides correlations to histologic subtypes, we found that the cluster groups exhibited significant differences with regard to histopathologic grade ($P = 0.004$), estrogen receptor (ER) expression ($P = 0.005$), and frequency of aberrations ($P < 0.001$; Table 3). Group II cases revealed a significantly lower grading ($P = 0.02$) compared with the other two groups. In group III, a significantly higher number of aberrations was encountered ($P < 0.01$) and in group I a higher frequency of ER negative tumors ($P = 0.008$) was observed when compared with the other two groups.

Discussion

Our analysis of 40 invasive breast cancer specimens revealed frequent gains on 1q, 8p, 8q, 11q, 16p, 17q, and 20q, as well as frequent losses on 1p, 8p, 11q, 13q, 16q, 17p, and 22q (Fig. 1), in good agreement with earlier chromosomal CGH analyses (28–31). Of these aberrations, 16q and 22q loss and 8q and 20q gain were previously reported to occur at different frequencies in ILC and IDC (7–9). Loss of 16q, in particular, is a frequent event in invasive breast cancer (32). In ILC, *CDH1* (*E-cadherin*) was identified as the target gene of the 16q loss by mutation analysis (33). In our data, no significant difference concerning loss of 16q was found between IDC and ILC, although ILC had a higher frequency of loss (60% compared with 27% in IDC).

Due to the superior resolution of matrix-CGH, several chromosomal regions were newly identified that are differentially imbalanced in IDC and ILC. Furthermore, a number of known candidate regions were considerably refined. Several of these regions contain oncogenes and tumor suppressor genes that are important in breast cancer (34, 35). Our data indicate that these genes might at least partially be responsible for the difference between the histologic subtypes. For example, *MYC* (8q24.21) and *ERBB2* (17q12) were found to be more often gained in IDC, whereas *TP53* (17q22.3) was more frequently lost in ILC. New genes distinguishing IDC and ILC, like the oncogenes *PTK2* (*FAK*) on 8q24.3, *NCOA3* (*AIB1*) on 20q13.12, and *BIRC7* on 20q13.33, were more commonly gained in IDC, and tumor suppressor genes like *MCM5* (22q12.3) and *ST13* (22q13.2) were more frequently lost in ILC (Figure 1 and Table 1).

Genomic fragments on chromosome arms 1q and 16p were identified as the most important discriminators between IDC

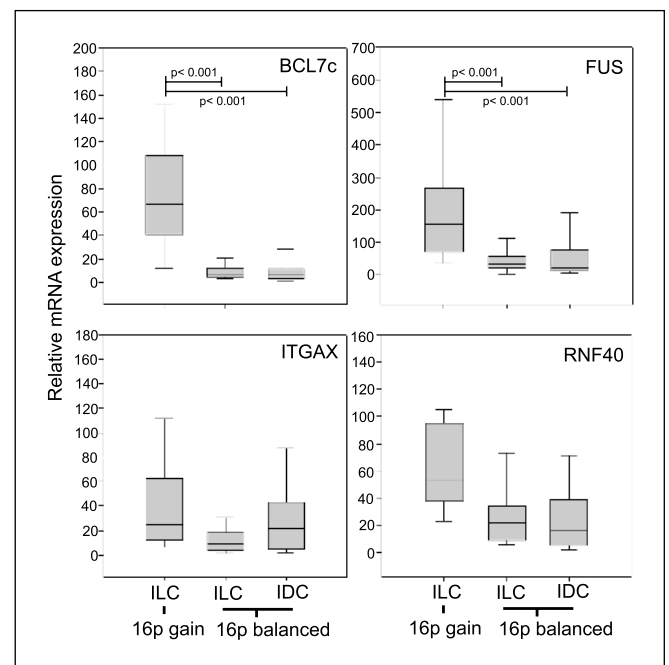


Fig. 3. Expression analysis of genes located in the candidate region 16p11.2. Expression levels of mRNA of four candidate genes were determined by quantitative reverse transcription-PCR. The expression was measured in three groups: ILC with 16p gain ($n = 10$), ILC ($n = 11$), and IDC ($n = 18$) with balanced 16p. $P < 0.001$ are indicated.

and ILC in independent bioinformatics analyses. These fragments not only occurred at significantly different frequencies in the two histologic subtypes but also proved to best distinguish IDC and ILC in SVM classification analysis. A subtype-specific frequency difference of >40% of gain of 1q also was detected by Loo et al. (14) in an recent array CGH study but the difference did not prove to be significant in their data. Classification accuracy was further improved by taking into account aberration patterns on chromosome 11q.13, 17q12-q21.2, 17q22-q25.3, 8p11.21-8q24.3, 20q13.13-q13.33, 1p13.1, and 14q32.33. Both the *ERBB2* and *CCND1* loci are included in the classifier, together with the 8q gain containing the *MYC* oncogene. These are well-characterized loci of amplification in breast cancer (36) and seem to be genomic markers for IDC rather than ILC.

Within the subtype-discriminating region on 1q, smaller intervals were delineated using empirical methods. These contain two genes, *FMO2* and *PTGS2* (*COX-2*), which in a recent gene expression profiling study were found to be over-expressed in IDC, identifying them as potential new candidates of pathogenic relevance (5, 6). *FMO2* thus far has not been implicated in cancer biology by any other studies. The expression of *PTGS2* (*COX2*), on the other hand, is deregulated in

epithelial tumors and has been associated with variables of aggressive breast cancer, including large tumor size, positive axillary lymph node metastases, and HER2-positive tumor status (37, 38). Prostaglandin E₂, the protein product of *PTGS2* (*COX-2*), has been suggested to be important in stimulating estrogen synthesis and thus in promoting carcinogenesis of breast tumors (39). DNA copy number gain and overexpression of *PTGS2* might thus confer a greater growth advantage in the ER-positive ILC than in the mostly ER-negative IDC.

For 16p, we could define a small minimally gained region on band 11.2, which is affected in all cases with 16p gain. This region contains four genes of interest with respect to tumor pathomechanism (*RNF40*, *BCL7C*, *FUS*, and *ITGAX*), which were further analyzed by real-time quantitative PCR. Only *BCL7C* and *FUS* are significantly overexpressed in ILC with 16p gain, however, and remain as candidates for possible pathogenic relevance in this tumor subgroup. *BCL7C* was identified by its similarity to *BCL7A*, which has been shown to be part of a three-way gene translocation in a Burkitt lymphoma cell line (40). The RNA-binding protein *FUS* was shown to participate in a fusion protein playing an important role in malignant liposarcomas (41). Further studies will be needed to elucidate their role in lobular breast cancer with 16p gain in general.

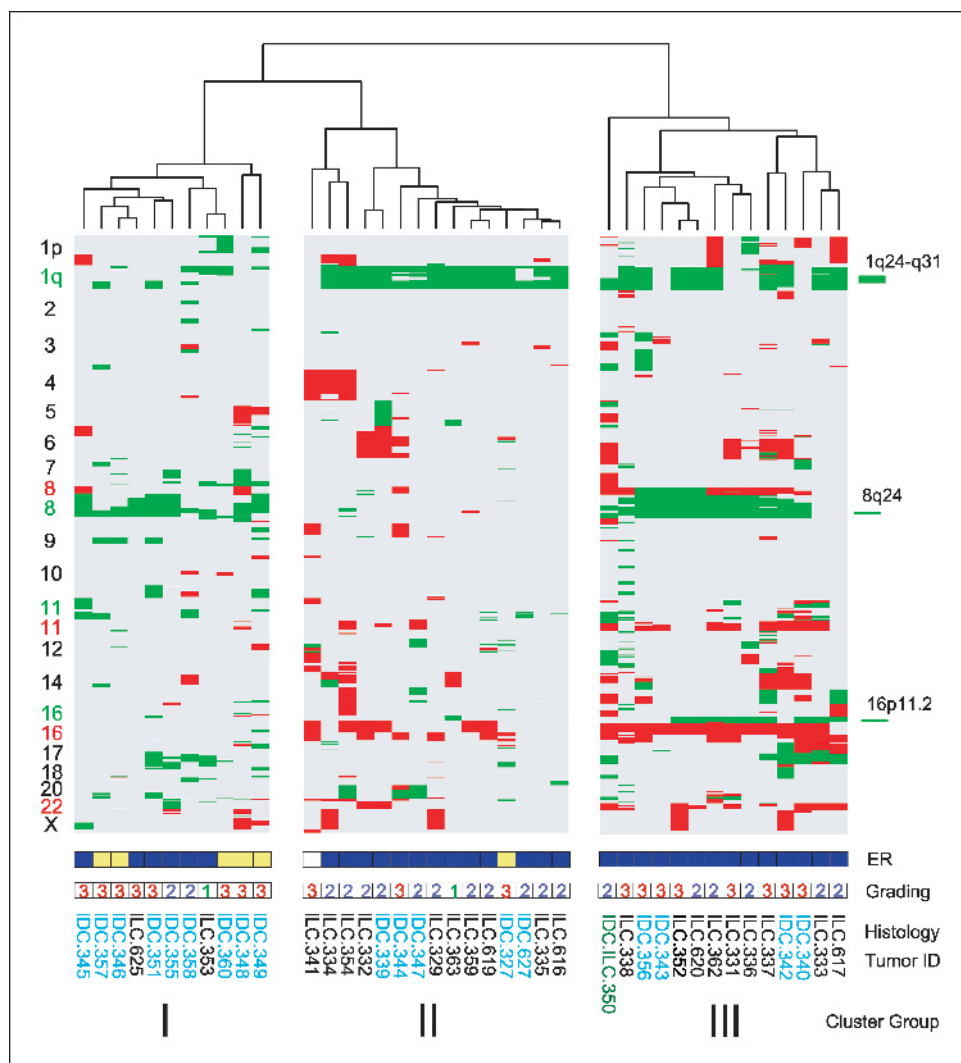


Fig. 4. Unsupervised hierarchical cluster analysis was done with all 40 invasive breast cancers. The genomic fragments on the Y axis are ordered according to their genomic position on the chromosomes from top to bottom. Three distinct cluster groups were identified (I, II, and III). The chromosomal arms of frequently altered regions in one or more cluster groups are indicated on the Y axis for gains (green) and losses (red). Right, the discussed candidate regions. ER status (yellow, negative; blue, positive) as well as the grading and histology are stated for every tumor.

Downloaded from http://aacrjournals.org/clinccancerres/article-pdf/12/2/345/1967055/345.pdf by guest on 04 February 2023

Table 3. Association of histopathologic variables and number of aberrations with the cluster groups identified by the unsupervised hierarchical cluster analysis

	Group I (n = 11)	Group II (n = 15)*	Group III (n = 14)*	P
Histology				
ILC	2	10	9	0.02
IDC	9	5	4	
ER				
Positive	6	13	14	0.005
Negative	5	1	0	
Grading				
1	1	1	—	0.004
2	2	11	6	
3	8	3	8	
Aberrations	10.2%	11.3%	21.3%	<0.001
Gains	7.9%	5.1%	10.7%	<0.001
Losses	2.3%	6.2%	10.6%	<0.001

*Group II contains tumor ILC.341 with unknown ER and Her2neu status, and tumor IDC.327 with unknown Her2neu status. Group III contains tumor 350, which exhibits both histologies.

The importance of 16p in ILC is independently supported by two recently published studies, one showing a high frequency of gain (54% compared with 42% in our study) of the whole chromosome arm 16p (42) and the other describing more frequent 16p gains in ER-positive than in ER-negative IDC (14). As ILCs normally are ER positive, the difference between ILC and IDC on 16p could be a result of the association of lobular histology and ER expression.

High-resolution molecular profiling approaches provide a means of identifying tumor subtypes, which correlate with clinical behavior. Because histologic classification of invasive breast cancer does not allow the prediction of the clinical course, substantial efforts have been made to further subclassify this type of breast cancer. Several studies attempted to group invasive breast cancer according to their profile obtained by chromosomal CGH (14, 31, 43–46). In this study, unsupervised cluster analysis revealed three molecular subgroups associated with tumor histology, ER status, grading as well as pattern of genetic aberrations. Groups I and II apparently overlap with previously reported subclassification schemes. In a study of primary invasive breast cancer by Rennstam et al. (31), which does not distinguish between IDC and ILC, chromosomal CGH data were used to define three distinct groups of tumors. Group A is described by whole chromosome arm gains of 1q and 16p as well as loss of 16q and was strongly correlated to ER and prostaglandin receptor positivity. 1q gain and 16q loss are frequently found in well-differentiated *in situ* ductal carcinomas (47). The existence of a distinct low-grade tumor group characterized by frequent 16q loss and ER/prostaglandin receptor positivity was also suggested in a multipathway model of breast cancer recently proposed by Simpson et al. (43). Group A of Rennstam et al. and the low-grade group of Simpson et al. are very similar to group II of the present study with 1q gain, 16q loss,

ER receptor positivity, and a significantly lower grading ($P = 0.02$) than the tumors in the other two groups (Table 3). Group B by Rennstam et al. is characterized by 8p gain, 8q loss, and adverse prognosis, closely matching our group I with the highest tumor grades (Table 3). An association of chromosome 8 alterations and poor clinical outcome was also observed by Tirkkonen et al. (30). The distinction of a high- and a low-grade tumor group by 16q copy number status (groups I and II, respectively) is consistent with a model of breast cancer development proposing two different genetic pathways, characterized by an early loss of 16q in the low-grade group (44, 46). Group III in our cluster analysis exhibits a combination of aberrations from groups I and II in addition to frequent aberrations of 8p, 11q, 17p, and 22q loss. This cluster group shows a significantly higher frequency of gains and losses than the two other groups ($P < 0.01$). Whereas group C of Rennstam et al. also exhibits higher numbers of genomic aberrations, it differs considerably with respect to histopathology and clinical variables. Whether group III of our study presents an independent third group of breast cancer or develops out of one of the other two groups remains to be elucidated. In the context of a progression model, in which breast cancer develops from low- to high-grade tumors (48, 49), it is tempting to speculate that group III tumors develop from group II tumors acquiring genomic imbalances observed in the high-grade group I tumors.

Our unsupervised hierarchical cluster analysis reflects the existence of considerable genetic overlap between IDC and ILC. However, significant correlations between histologic and molecular classifications do exist. Group I is significantly correlated with IDC ($P = 0.02$), whereas >90% of ILC cluster in groups II and III. A major difference between ILC in these groups is gain of chromosome 16p. Eight of nine ILC in group III have a gain on chromosome arm 16p, whereas 9 of 10 ILC in group II have a balanced chromosome arm 16p. These findings might explain why the classification of the histologic subtypes, which is primarily dependent on 16p and 1q, did not exceed an overall accuracy of 65%. Importantly, however, the newly identified regions on 1q and 16p (and by implication the candidate genes they contain) that best distinguish the histologic subtypes are important for the characterization of the molecular subgroups in our unsupervised cluster analysis. The regions on 1q, including the two candidate genes, are gained in none of the group I tumors but in 11 of 15 and 10 of 15 of group II and 8 of 14 and 10 of 14 of group III tumors for 1q24.2–25.1 and 1q25.3–q31.3, respectively. The latter two groups can be distinguished by gain of the candidate region on 16p11.2, present in 9 of 14 of group III tumors but in only one tumor of group II.

In conclusion, high-resolution molecular signatures allowed to further subdivide and refine current diagnostic subtypes of invasive breast cancer. In the future, integration of genomic profiles with comprehensive transcriptomic and proteomic analyses in the same series of tumors will likely further improve our understanding of invasive breast cancer.

Acknowledgments

We thank Holger Schwender for providing an implementation of SAM for categorical values and Felix Kokocinski for bioinformatics support as well as Stefanie Hofman and Andrea Wittman for excellent technical assistance.

References

1. Tavassoli FADP, editor. World Health Organization classification of tumours. Pathology and genetics of tumours of the breast and female genital organs. Lyon (France): WHO; 2003.
2. Li CI, Anderson BO, Daling JR, Moe RE. Trends in incidence rates of invasive lobular and ductal breast carcinoma. *JAMA* 2003;289:1421–4.
3. 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.
4. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003; 100:8418–23.
5. Zhao H, Langerod A, Ji Y, et al. Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Mol Biol Cell* 2004;15:2523–36.
6. Korkola JE, DeVries S, Fridlyand J, et al. Differentiation of lobular versus ductal breast carcinomas by expression microarray analysis. *Cancer Res* 2003;63:7167–75.
7. Loveday RL, Greenman J, Simcox DL, et al. Genetic changes in breast cancer detected by comparative genomic hybridisation. *Int J Cancer* 2000;86:494–500.
8. Nishizaki T, Chew K, Chu L, et al. Genetic alterations in lobular breast cancer by comparative genomic hybridization. *Int J Cancer* 1997;74:513–7.
9. Gunther K, Merkelbach-Bruse S, Amo-Takyi BK, Handt S, Schroder W, Tietze L. Differences in genetic alterations between primary lobular and ductal breast cancers detected by comparative genomic hybridization. *J Pathol* 2001;193:40–7.
10. Solinas-Toldo S, Lampel S, Stilgenbauer S, et al. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 1997;20:399–407.
11. Pinkel D, Seagraves R, Sudar D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 1998;20:207–11.
12. Mendrzyk F, Radlwimmer B, Joos S, et al. Genomic and protein profiling identifies CDK6 as novel independent prognostic marker in medulloblastoma. *J Clin Oncol* 2005;23:8853–62.
13. Schwaenen C, Nessling M, Wessendorf S, et al. Automated array-based genomic profiling in chronic lymphocytic leukemia: development of a clinical tool and discovery of recurrent genomic alterations. *Proc Natl Acad Sci U S A* 2004;101:1039–44.
14. Loo LW, Grove DJ, Williams EM, et al. Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. *Cancer Res* 2004; 64:8541–9.
15. Knight SJ, Lese CM, Precht KS, et al. An optimized set of human telomere clones for studying telomere integrity and architecture. *Am J Hum Genet* 2000;67: 320–32.
16. Fiegler H, Carr P, Douglas EJ, et al. DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer* 2003;36:361–74.
17. Zielinski B, Gratijs S, Toedt G, et al. Detection of chromosomal imbalances in retinoblastoma by matrix-based comparative genomic hybridization. *Genes Chromosomes Cancer* 2005;43:294–301.
18. Hupé P, Stransky N, Thiery JP, Radvanyi F, Barillot E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics* 2004;20: 3413–22.
19. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98:5116–21.
20. Ruschhaupt M, Huber W, Poustka A, Mansmann U. A compendium to ensure computational reproducibility in high-dimensional classification tasks. *Statistical Applications in Genetics and Molecular Biology* 2004; 3:Article 37. Available from <http://www.bepress.com/sagmb/vol3/iss1/art37>.
21. Aliferis CF, Hardin D, Massion PP. Machine learning models for lung cancer classification using array comparative genomic hybridization. *Proc AMIA Symp* 2002;7–11.
22. Fritz B, Schubert F, Wrobel G, et al. Microarray-based copy number and expression profiling in dedifferentiated and pleomorphic liposarcoma. *Cancer Res* 2002;62:2993–8.
23. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Machine Learning* 2002;46:389–422.
24. Quackenbush J. Computational analysis of microarray data. *Nat Rev Genet* 2001;2:418–27.
25. Lehmann U, Bock O, Langer F, Kreipe H. Demonstration of light chain restricted clonal B-lymphoid infiltrates in archival bone marrow trephines by quantitative real-time polymerase chain reaction. *Am J Pathol* 2001;159:2023–9.
26. Specht K, Kremer M, Muller U, et al. Identification of cyclin D1 mRNA overexpression in B-cell neoplasias by real-time reverse transcription-PCR of microdissected paraffin sections. *Clin Cancer Res* 2002;8:2902–11.
27. Bijwaard KE, Aguilera NS, Monczak Y, Trudel M, Taubenberger JK, Lichy JH. Quantitative real-time reverse transcription-PCR assay for cyclin D1 expression: utility in the diagnosis of mantle cell lymphoma. *Clin Chem* 2001;47:195–201.
28. Kallioniemi A, Kallioniemi OP, Piper J, et al. Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc Natl Acad Sci U S A* 1994;91:2156–60.
29. 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.
30. Tirkkonen M, Tanner M, Karhu R, Kallioniemi A, Isola J, Kallioniemi OP. Molecular cytogenetics of primary breast cancer by CGH. *Genes Chromosomes Cancer* 1998;21:177–84.
31. 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.
32. Cleton-Jansen AM. E-cadherin and loss of heterozygosity at chromosome 16 in breast carcinogenesis: different genetic pathways in ductal and lobular breast cancer? *Breast Cancer Res* 2002;4:5–8.
33. Droufakou S, Deshmane V, Roylance R, Hanby A, Tomlinson I, Hart IR. Multiple ways of silencing E-cadherin gene expression in lobular carcinoma of the breast. *Int J Cancer* 2001;92:404–8.
34. Nessling M, Richter K, Schwaenen C, et al. Candidate genes in breast cancer revealed by microarray-based comparative genomic hybridization of archived tissue. *Cancer Res* 2005;65:439–47.
35. Callagy G, Pharoah P, Chin SF, et al. Identification and validation of prognostic markers in breast cancer with the complementary use of array-CGH and tissue microarrays. *J Pathol* 2005;205:388–96.
36. Kreipe H, Feist H, Fischer L, et al. Amplification of c-myc but not of c-erbB-2 is associated with high proliferative capacity in breast cancer. *Cancer Res* 1993; 53:1956–61.
37. Saji S, Hirose M, Toi M. Novel sensitizing agents: potential contribution of COX-2 inhibitor for endocrine therapy of breast cancer. *Breast Cancer* 2004; 11:129–33.
38. Arun B, Goss P. The role of COX-2 inhibition in breast cancer treatment and prevention. *Semin Oncol* 2004;31:22–9.
39. Brodie AM, Lu Q, Long BJ, et al. Aromatase and COX-2 expression in human breast cancers. *J Steroid Biochem Mol Biol* 2001;79:41–7.
40. Zani VJ, Asou N, Jadayel D, et al. Molecular cloning of complex chromosomal translocation t(8;14;12)(q24.1;q32.3;q24.1) in a Burkitt lymphoma cell line defines a new gene (BCL7A) with homology to caldesmon. *Blood* 1996;87:3124–34.
41. Schwarzbach MH, Koesters R, Germann A, et al. Comparable transforming capacities and differential gene expression patterns of variant FUS/CHOP fusion transcripts derived from soft tissue liposarcomas. *Oncogene* 2004;23:6798–805.
42. Shelley Hwang E, Nyante SJ, Yi Chen Y, et al. Clonality of lobular carcinoma *in situ* and synchronous invasive lobular carcinoma. *Cancer* 2004;100: 2562–72.
43. Simpson PT, Reis-Filho JS, Gale T, Lakhani SR. Molecular evolution of breast cancer. *J Pathol* 2005;205: 248–54.
44. Roylance R, Gorman P, Hanby A, Tomlinson I. Allelic imbalance analysis of chromosome 16q shows that grade I and grade III invasive ductal breast cancers follow different genetic pathways. *J Pathol* 2002;196: 32–6.
45. Jones C, Ford E, Gillett C, et al. Molecular cytogenetic identification of subgroups of grade III invasive ductal breast carcinomas with different clinical outcomes. *Clin Cancer Res* 2004;10:5988–97.
46. Buerger H, Mommers EC, Littmann R, et al. Ductal invasive G2 and G3 carcinomas of the breast are the end stages of at least two different lines of genetic evolution. *J Pathol* 2001;194:165–70.
47. Buerger H, Mommers EC, Littmann R, et al. Correlation of morphologic and cytogenetic parameters of genetic instability with chromosomal alterations in *in situ* carcinomas of the breast. *Am J Clin Pathol* 2000;114:854–9.
48. Ried T, Heselmeyer-Haddad K, Blegen H, Schrock E, Auer G. Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: a phenotype/genotype correlation. *Genes Chromosomes Cancer* 1999;25:195–204.
49. Brenner AJ, Aldaz CM. The genetics of sporadic breast cancer. *Prog Clin Biol Res* 1997;396:63–82.