# High-Resolution Genomic Profiles of Breast Cancer Cell Lines Assessed by Tiling BAC Array Comparative Genomic Hybridization

**Göran Jönsson,[1†] Johan Staaf,[1†] Eleonor Olsson,[1] Markus Heidenblad,[2] Johan Vallon-Christersson,[1] Kazutoyo Osoegawa,[3] Pieter de Jong,[3] Stina Oredsson,[4] Markus Ringnér,[5] Mattias Höglund,[2] and Åke Borg[1,6*]**

[1]Department of Oncology, University Hospital, Lund, Sweden
[2]Department of Clinical Genetics, University Hospital, Lund, Sweden
[3]BACPAC Resources, Children's Hospital Oakland Research Institute, Oakland, CA
[4]Department of Cell and Organism Biology, Lund University, Lund, Sweden
[5]Department of Theoretical Physics, Lund University, Lund, Sweden
[6]Lund Strategic Research Center for Stem Cell Biology and Cell Therapy, Lund University, Lund, Sweden

A BAC-array platform for comparative genomic hybridization was constructed from a library of 32,433 clones providing complete genome coverage, and evaluated by screening for DNA copy number changes in 10 breast cancer cell lines (BT474, MCF7, HCC1937, SK-BR-3, L56Br-C1, ZR-75-1, JIMT1, MDA-MB-231, MDA-MB-361, and HCC2218) and one cell line derived from fibrocystic disease of the breast (MCF10A). These were also characterized by gene expression analysis and found to represent all five recently described breast cancer subtypes using the "intrinsic gene set" and centroid correlation. Three cell lines, HCC1937 and L56BrC1 derived from *BRCA1* mutation carriers and MDA-MB-231, were of basal-like subtype and characterized by a high frequency of low-level gains and losses of typical pattern, including limited deletions on 5q. Four estrogen receptor positive cell lines were of luminal A subtype and characterized by a different pattern of aberrations and high-level amplifications, including *ERBB2* and other 17q amplicons in BT474 and MDA-MB-361. SK-BR-3 cells, characterized by a complex genome including *ERBB2* amplification, massive high-level amplifications on 8q and a homozygous deletion of *CDH1* at 16q22, had an expression signature closest to luminal B subtype. The effects of gene amplifications were verified by gene expression analysis to distinguish targeted genes from silent amplicon passengers. JIMT1, derived from an *ERBB2* amplified trastuzumab resistant tumor, was of the ERBB2 subtype. Homozygous deletions included other known targets such as *PTEN* (HCC1937) and *CDKN2A* (MDA-MB-231, MCF10A), but also new candidate suppressor genes such as *FUSSEL18* (HCC1937) and *WDR11* (L56Br-C1) as well as regions without known genes. The tiling BAC-arrays constitute a powerful tool for high-resolution genomic profiling suitable for cancer research and clinical diagnostics. This article contains Supplementary Material available at http://www.interscience.wiley.com/jpages/1045-2257/suppmat. 

## INTRODUCTION

Cancer cells acquire various types of genomic aberrations ranging from single nucleotide substitutions to structural and quantitative alterations at the chromosomal level (Weber, 2002). DNA copy number alterations (CNA) such as somatic chromosomal gains or losses, interstitial deletions or gene amplifications are invariably found in cancer cells, but may also occur in the germline, underlying disease predisposition and congenital defects, or representing copy number polymorphisms (Sebat et al., 2004; Vissers et al., 2004). Recent improvements in resolution and sensitivity in comparative genomic hybridization (CGH) have been possible through implementation of microarray-based CGH (array CGH) (Albertson and Pinkel, 2003). The array format for CGH provides several advantages over the use of metaphase spreads (Kallioniemi et al., 1992), including higher resolution and dynamic range, direct mapping of altered clones/oligos to the genome sequence, as well as better throughput, automation and standardization.

Studies using either metaphase or array CGH have shown that breast cancer is a heterogeneous disease with regards to the number and pattern of CNA, gains commonly affect chromosomal arms

Supported by: The Swedish Cancer Society, The Swedish Research Council, The Mrs. Berta Kamprad Foundation, The Gunnar Nilsson Cancer Foundation, The Franke & Margareta Bergqvist Foundation, The Lund University Hospital Foundations, The King Gustav V:s Jubilee Foundation, The Ingabritt and Arne Lundberg Foundation, The Swedish Foundation for Strategic Research, The Marianne and Marcus Wallenberg Foundation, The Knut and Alice Wallenberg Foundation (SWEGENE program), The American Cancer Society.

*Correspondence to: Åke Borg, Department of Oncology, Lund University, SE-221 85 Lund, Sweden. E-mail: ake.borg@med.lu.se

†These authors contributed equally to this work.

Received 18 November 2006; Accepted 22 January 2007

DOI 10.1002/gcc.20438

Published online 2 March 2007 in Wiley InterScience (www.interscience.wiley.com).

1q, 8q, 11q, 17q, and 20q, whereas losses are most frequent on 6q, 8p, 9p, 13q, and 16q (Rennstam et al., 2003; Naylor et al., 2005). Genes and regions with earlier known DNA amplification such as *ERBB2* at 17q12 and *CCND1* at 11q13 have been confirmed and new amplicons identified. Clearly different CNA patterns are evident in estrogen receptor (ER) negative and ER positive tumors, the former frequently displaying gains on 3q and 10p and losses on 4p, 4q, 5q, 12q, and 15q (Loo et al., 2004). Different genomic profiles are also observed in the various tumor subtypes recently defined by expression-profiling, where frequent gains/losses are associated with the basal-like subtype, while high-level amplification is more frequent in luminal-B tumors (Sorlie et al., 2003; Bergamaschi et al., 2006). This has also been shown by genome-wide single nucleotide polymorphism (SNP) detection arrays where loss of heterozygosity (LOH) at 4p and 5q were correlated to the basal-like phenotype (Wang et al., 2004). Coherent CGH findings in tumors from *BRCA1* mutation carriers, which are typically of ER negative and basal-type (Tirkkonen et al., 1997; Jönsson et al., 2005b; van Beers et al., 2005), suggest that distinct mechanisms of genomic instability may be underlying their pathogenesis. Breast cancer cell lines commonly used as models have recently been classified into the gene expression subtypes (Charafe-Jauffret et al., 2006). However, little is known about subtype-specific DNA CNAs in these cell lines.

Here, we present the construction of a tiling BAC array platform, comprising 32,433 clones, and evaluate its performance in high-resolution genomic profiling of 10 breast cancer cell lines and one cell line derived from a breast fibrocystic disease.* Furthermore, global gene expression analysis is used to classify these cell lines into subtypes based on the "intrinsic gene set" (Sorlie et al., 2003), and to elucidate candidate genes affected by novel amplifications as well as hemizygous or homozygous deletions.

## MATERIALS AND METHODS

### Samples

Eleven human breast cell lines (BT474, ZR-75-1, SK-BR-3, MCF7, MDA-MB-231, MDA-MB-361, JIMT1, MCF10A, L56Br-C1, HCC1937, and HCC-2218) were used in this study. All, except for L56Br-C1 and JIMT1, were obtained from American type culture collection (ATCC, http://www.atcc.org/), for HCC-2218 only as extracted DNA. L56Br-C1 was established at Lund University (Johannsson, et al., 2003). JIMT1 cells were established at Tampere University (Tanner et al., 2004) and obtained from DSMZ (German collection of microorganisms and cell cultures, http://www.dsmz.de/). Cells were cultured according to the suppliers' recommendations, basically using RPMI 1640 medium supplemented with 10% heat-inactivated FCS, nonessential amino acids, insulin (10 μg/ml), penicillin (50 U/ml), and streptomycin (50 ng/ml), but with the following exceptions. SK-BR-3 cells were cultured without the addition of insulin. Cholera toxin (50 ng/ml) and hydrocortisone (500 ng/ml) were added to the medium of MCF10A cells. HCC1937 cells (Tomlinson, et al., 1998) were cultured in MEM a-medium supplemented with 10% heat-inactivated FCS, nonessential amino acids, 1entamicin (0.1 mg/ml), epidermal growth factor (20 ng/ml), and insulin (10 μg/ml). Control DNA samples from male 46, XY and female 46, XX were obtained from healthy blood donors. DNA from chromosome X aberrant cell lines with karyotype 47, XXX (repository no.GM04626) and 48, XXXX (repository no.GM01416) were obtained from the Coriell Institute for Medical Research.

### Construction of 32k-Arrays

High-resolution tiling BAC arrays were produced at the Swegene DNA Microarray Resource Center, Department of Oncology, Lund University, Sweden (http://swegene.onk.lu.se) using the BAC Re-Array set Ver. 1.0 (32,433 BAC clones), previously described by Krzywinski et al. (2004), obtained from the BACPAC Resource Center at Children's Hospital Oakland Research Institute, Oakland (CA) as prepared DNA aliquots. BAC clones were mapped to the hg17 genome build. For each BAC clone, 6 ng of DNA were amplified using degenerate oligonucleotide primed (DOP) PCR, using an adopted protocol obtained from Eric Schoenmakers, Nijmegen, The Netherlands. In short, DNA template was added in prealiquoted 96 well PCR Master Mix plates PCR plates (AB gene, Cat. No. SA-081) containing the 6MW primer, 5′-CCG ACT CGA GNN NNN NAT GTG G-3′. PCR conditions were: an initial 3 min denaturation , followed by 35 cycles of denaturation at 94°C for 30 sec, a linear ramp (37–72°C) over 10 min and 1 min extension at 72°C, and finally a 10 min step at 72°C. All PCR products were subsequently purified using PALL AcroPrep 96 Omega 10 K filter plates (Pall Corporation, Ann

*Microarray data from genomic profiling will be submitted to GEO and will be available from the Lund University, Department of Oncology website http://swegene.onk.lu.se. Scripts used for data analysis will be made available on request.

Arbor, Michigan) and verified on agarose gel (E-Gel, Invitrogen, Carlsbad, CA). Purified PCR product was dried and dissolved in 50% DMSO to a concentration of 500–1000 ng/μl to produce a print-ready probe set. All clone preparation steps were carried out on a Biomek FX automated liquid handling system (Beckman). The entire probe set was printed on a single aminosilane coated glass slide (UltraGAPS; Corning, Acton, MA) using a MicroGrid2 robot (BioRobotics, Cambridge, UK) equipped with MicroSpot 10K pins (BioRobotics).

### DNA Isolation, Labeling, and Hybridization

DNA was extracted from cells using the Wizard Genomic DNA extraction kit (Promega), except for HCC-2218 (obtained from ATCC). DNA was labeled as previously described (Jönsson et al., 2005). Normal male genomic DNA was used as reference. Prior to hybridization, arrays were UV-cross-linked at 500 mJ/cm$^2$ and pretreated using the Universal Microarray Hybridization Kit (Corning) according to manufacturers' instructions. Labeled DNA was resuspended in 50 μl hybridization solution (50% formamide, 10% dextran sulfate, 2×SSC, 2% SDS, 10 μg/μl yeast tRNA) and heated to 70°C for 15 min followed by a 30 min incubation at 37°C for 30 min. Hybridization reactions were applied to arrays and incubated under cover slips for 72 hr at 37°C. Arrays were washed as previously described (Snijders et al., 2001) and fluorescence was recorded using an Agilent G2565AA microarray scanner (Agilent Technologies).

### Genome Wide Gene Expression Analysis

Oligonucleotide microarrays were produced at the Swegene DNA Microarray Resource Center, using a set of ~35,000 human oligonucleotide probes (Operon, Ver. 3.0). Probes were dissolved in Corning Universal Spotting solution to 24 μM and printed as described earlier. RNA was extracted from all cells lines (except HCC-2218) using Trizol (Invitrogen) reagent followed by RNeasy Midi purification kit (Qiagen). RNA was also isolated from normal breast tissues ($n = 7$) to use as a comparison to the cell lines. RNA concentration was determined using a spectrophotometer (NanoDrop) and integrity was confirmed with the Bioanalyzer 2100 system (Agilent technologies, Palo Alto, CA). Fluorescently labeled cDNA targets for hybridization were prepared according to manufacturers' instructions using the Corning Pronto Plus system 6. Samples were labeled with Cy3-dCTP (Amersham) and reference (Universal Human Reference RNA; Stratagene, La Jolla, CA) was labeled with Cy5-dCTP.

### Image and Data Analysis

TIFF images were analyzed using the Gene Pix Pro 4.0 software (Axon Instruments, Foster City, CA), and the quantified data matrix was loaded into a local installation of BioArray Software Environment (BASE) (Saal et al., 2002). For copy number data, positive and nonsaturated spots were background corrected using the median foreground minus the median background signal intensity for each channel and log(2)ratios were calculated from the background corrected intensities. Data were filtered for flagged features and signal to noise ratio >5 for each spot in both intensity channels. Log(2)ratios on each array were normalized and corrected for intensity-based location adjustment (Yang et al., 2001) excluding the X and Y BAC clones when estimating the normalization function. A BASE adapted R (http://www.r-project.org/) implementation of CGH-Plotter (Autio et al., 2003) was used for automatic break point analysis in which a constant parameter value of 15, was applied. Cutoff ratios for gains and losses were set to 1.23 and 0.81, respectively, corresponding to log(2)ratio of ±0.3. For gene expression data, positive and nonsaturated spots were background corrected using the median foreground minus the median background signal intensity for each channel and the log(2)ratio were calculated from the background corrected intensities. Data were filtered for flagged features for each spot in both intensity channels. The log(2)ratios on each array were normalized and corrected for intensity-based location adjustment (Yang et al., 2001) and each oligonucleotide probe (henceforth referred to as reporter) was given a SNR based uncertainty value used for merging reporters with identical locus link number. Data were then transformed using an error model with 80 percent presence required (Andersson et al., 2005).

### Comparing Genomic Alterations and Gene Expression

Before correlating genomic alterations with gene expression, relative log(2)ratio expression values from the cell lines, measured against the common Stratagene reference, were related to normal breast tissue. This approach renders an estimate of gene expression per breast cancer cell line versus an imaginable more suitable biological reference than the Stratagene reference, highlighting changes in gene expression not observed in the normal breast.

This was performed by first calculating the difference in expression for each reporter against a normal breast tissue expression centroid. In short, seven normal breast tissue samples were each hybridized against the common Stratagene reference. Data were filtered for flagged features for each spot in both intensity channels. The log(2)ratios on each array were normalized and corrected for intensity-based location adjustment (Yang et al., 2001) and each reporter was given a SNR based uncertainty value used for merging. A minimum presence in at least four out of the seven hybridizations was required before averaging log(2)ratio values on reporters across hybridizations. A normal breast expression centroid was then constructed by calculating intensity based *Z*-scores for each averaged reporter, using a sliding window of 20%, as described (Yang et al., 2002).

To map reporters to their genomic bp location, each reporter sequence was blasted against the hg17 genome build with a cutoff of >70 percent aligned length and no cross matches. A total of 10,818 reporters equaling the same number of genes remained as the final expression data set, which was then mapped to the corresponding BAC clone harboring the oligonucleotide probe. Hence, two data matrices were created, one for gene expression data and one for copy number data, with identical dimensions where each row corresponded to a reporter-BAC pair (allowing redundancy) and each column to a breast cancer cell line. Standard Pearson correlation (henceforth referred to as correlation) for each reporter-BAC pair was then calculated to find pairs where gene expression and genomic copy number behaved concordantly. Next, sample assignments for the copy number profiles were randomly permuted and correlations recalculated keeping the expression data and reporter—BAC pair mapping intact. This random permutation procedure was repeated 10,000 times and allowed us to calculate $P$ values for the different correlation bins (each bin represents a correlation range of 0.05). A $P$ value cutoff of 0.01 was selected corresponding to a correlation cutoff of 0.8. Analysis was performed in the open source statistical computing environment "R."

### Verification of Homozygous Deletions

Homozygous deletions were defined as two consecutive BAC clones with a log(2)ratio $< -2$, and were verified using STS markers within the consecutive BAC clones or vicinity. The primer sequences were obtained from UniSTS at NCBI. All cell lines were screened for the selected markers and control STS markers mapping to unaffected regions.

## RESULTS

### Quality Control of the Tiling Arrays

To analyze the reproducibility in microarray production and hybridization procedures, DNA from MCF7 cells and control male DNA was labeled in single reactions and hybridized at four different occasions using four arrays from one slide batch. Average Pearson correlation for log(2)ratios from 28,036 BAC clones present in all four replicates was 0.92. Furthermore, two self-versus-self hybridizations were conducted, resulting in an average SD of 0.135 in log(2)ratio. To assess the sensitivity and linearity in detection of copy number gains, five experiments were performed with labeled DNA from cells containing different numbers of chromosome X on an otherwise diploid autosomal background, using control male or female DNA as reference. A total of 28,620 autosomal and 1,409 X-chromosome BAC clones were used to assess the reliability of the experiments. Mean ($\pm$1 SD) log(2)ratios of X-chromosomal BAC clones were as follows: XX versus XX $-0.04$ ($-0.20$ to 0.12), expected log(2)ratio 0.0; XXX versus XX 0.42 (0.25–0.59), expected 0.5; XXXX versus XX 0.74 (0.55–0.92), expected 1.0; XXX versus XY 1.1 (0.81–1.4), expected 1.5; XXXX versus XY 1.46 (1.12–1.80), expected 2.0. Increase in copy number ratio was linearly correlated to X-chromosome number, from one to four copies, with a slope of 0.37 ($R^2 = 0.997$) (Fig. 1). Average SD in log(2)ratio of the autosomal chromosomes for all five hybridizations was 0.166.

### Genomic Profiles and Gene-Expression Subtypes of Individual Breast Cancer Cell Lines

Cutoff ratios for gains and losses were set to 1.23 and 0.81, respectively, corresponding to log(2)ratio of $\pm$0.3. Regions with log(2)ratio $> 1.5$ were considered as amplified and log(2)ratios $> 2.0$ as high-level amplification, and are listed in Table 1. By extrapolation of the linear regression curve in Figure 1, these levels would correspond to 4.0 and 5.5 times amplification (8 and 11 alleles), respectively. Homozygous deletions are defined as log(2)ratio $< -1.0$, were confirmed by STS markers and are listed in Table 2. Using the "intrinsic 500 gene set" and nearest centroid analysis (Sorlie et al., 2003), 10 of the 11 cell lines were classified into gene expression subtypes (Fig. 2). The genomic profiles of all 11 cell lines (Fig. 3), depicted in

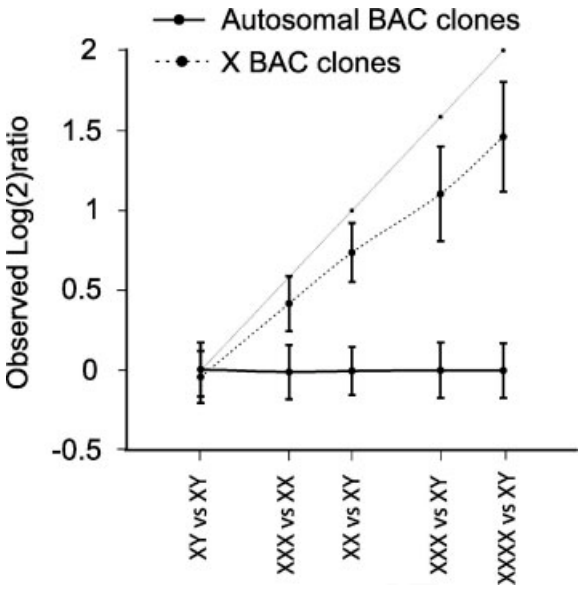


Figure 1. Assessment of signal linearity in copy number gains using DNA from cells with varying number of X chromosomes on a diploid autosomal background: Plot of mean (±1 SD) log(2)ratios for autosomal BAC clones (solid line) and X chromosome BAC clones (dashed line) for five different hybridizations with different amount of X chromosome copies. An expected standard curve of log(2)ratios on X chromosome BAC clones is plotted with solid gray line as a reference.

high-resolution for individual chromosomes, can be viewed at http://swegene.onk.lu.se.

### MCF7

MCF7 cells are derived from an ER positive epithelial adenocarcinoma metastasis (pleural effusion). Cytogenetics revealed a hypertriploid to hypertetraploid stemline with a modal chromosome number of 66–87 (ATCC). Here, CNAs affecting whole chromosomes, chromosomal arms, or major parts thereof, include gain of 1q, 3q, 7p, 8q, 14q, 16q, 17q, 20p, and 20q and loss of 1p, 4p, 8p, 9p, 11q, 13q, 15q, 16q, 18q, 21q, and 22q. A complex amplification pattern was detected on 20q13 with four narrow high-level amplification peaks, each containing only a few genes. Increased gene expression (as compared to normal breast tissue) was observed for *NCOA3*, *SULF2*, *ZNF217*, *PFDN4*, *STK6*, *VAPB*, and *BMP7*, representing possible target genes in the 20q amplicons. The region on 17q23.2 was divided into three high-level amplification peaks, where six genes (*RPS6KB1*, *TRIM37*, *USP32*, *PPM1D*, *TBX2*) showed an elevated expression. High-level amplification on chromosome 3p14.2 was split into two distinct peaks including five and six genes, respectively, two (*NIF3L1BP1*, *PSDM6*) of which have an elevated expression. Finally, high-level amplification were detected in chromosome segment 1p21.1-p13.3, encompassing ~1.2 Mbp and only two known genes (*NTNG1* and *HRMT1L6*), whereas an amplicon at 1p13.2 spans ~1 Mbp and includes 12 genes, four (*TRIM33*, *BCAS2*, *NRAS*, *UNR*) of which showed an elevated expression. One homozygous deletion was detected, located in 4q34.3, comprising ~240 kbp but no known coding or noncoding genes. As expected, MCF7 has correlation to the luminal A subtype and inverse correlations to the ERBB2+ and basal-like subtypes.

### BT474

The BT474 cell line is derived from an invasive ductal carcinoma, and has a chromosome count in the hypertetraploid range (ATCC). Here, larger chromosomal regions with CNA comprise 1q, 3q, 5p, 7, 8q, 11q, 12, 14, 17q, 18, 19q, and 20 (gains) and 3p, 6q, 9p, and 10q (losses). Four chromosomes have local high-level amplification, including 9p13.3, 15q12 as well as multiple amplicons on 17q and 20q. Moreover, lower level amplification was found on three loci on 1q, two loci on chromosome 4, and two loci on 11q. The region on 9p13.3 harbors a narrow peak spanning ~1 Mbp including 25 known genes, two (*FANCG* and *STOML2*) of which have increased gene expression. The 15q11.2 amplicon spans ~1.4 Mbp and contains five genes (*LOC283755*, *POTE15*, *LOC651769*, *OR4M2*, and *OR4N4*), although none of them show increased expression. The ~3Mbp high-level amplification on 17q12-q21.2 contains a considerable number of genes, with *ERBB2* as the obvious target. However, 24 additional genes in the amplicon have an increased expression. A ~2.7 Mbp amplicon that maps to 17q21.32-q21.33 also includes a large number of candidate genes, such as *HOXB7* (Hyman et al., 2002). Moreover, a high-level amplification region on 17q22-23.2 spans ~2.2 Mbp and includes eight known genes (*COX11*, *TOM1L1*, *STXBP4*, *HLF*, *MMD*, *TMEM100*, *PCTP*, *ANKFN1*, and *NOG*) of which one (*TOML1*) shows elevated expression. A 17q23.2 amplicon includes 22 genes whereof five (*RAD51C*, *FLJ10587*, *TRIM37*, *CLTC*, and *PTRH2*) have increased transcript levels. Like MCF7, BT474 cells harbor a highly complex amplification pattern on chromosome arm 20q with at least three distinct amplified regions. One of these overlaps with an amplicon found in MCF7 cells, and span ~11.7 Mbp on 20q13.13-q13.32 including a large number of candidate target genes. Eleven of these genes display increased expression (*NCOA3*, *KCNG1*, *CSE1L*, *PREX1*, *PFDN4*, *STK6*, *BMP7*, *RAE1*, *VAPB*, *RAB22A*, and *RNPC1*). Two other 20q regions amplified in BT474 did not

TABLE 1. Regions Found Amplified in 10 Breast Cancer Cell Lines

| Cell line[a] | Cytoband | Start Clone | Start position (bp) | Size (kbp) | Peak log(2)R | N genes | Candidate target genes |
|---|---|---|---|---|---|---|---|
| MCF7 | 1p21.1-p13.3 | RP11-441C19 | 106,373,928 | 1287 | 3.2 | 2 | |
| | 1p13.2-p13.1 | RP11-541A20 | 113,961,795 | 1220 | 2.8 | 12 | *TRIM33, BCAS2, NRAS, CSDE1* |
| | 3p14.2 | RP11-401G18 | 61,500,834 | 1579 | 2.7 | 5 | *PTPRG* |
| | 3p14.2-p14.1 | RP11-177C11 | 63,270,810 | 1684 | 3.6 | 6 | *PSMD6, NIF3L1BP1* |
| | 17q23.2 | RP11-795C13 | 53,871,587 | 839 | 3.2 | 11 | *RAD51C, TRIM37, PPM1E* |
| | 17q23.2 | RP11-168J8 | 55,205,410 | 1007 | 3.7 | 11 | *TUBD1, RPS6KB1, USP32* |
| | 17q23.2 | RP11-113J9 | 56,558,187 | 900 | 4.4 | 8 | *TBX2* |
| | 20q13.12-q13.13 | RP11-702E3 | 45,194,473 | 1903 | 4.4 | 6 | *SULF2, NCOA3* |
| | 20q13.13 | RP13-625L11 | 48,515,164 | 522 | 3.3 | 3 | *BCAS4* |
| | 20q13.2 | RP11-694L10 | 51,487,854 | 1649 | 5.0 | 6 | *ZNF217, PFDN4* |
| | 20q13.31-q13.32 | RP11-46O3 | 54,786,446 | 2069 | 4.7 | 9 | *BMP7, STK6, RAB22A, VAPB* |
| BT474 | 1q24.2 | RP11-745I5 | 166,303,401 | 578 | 1.8 | 7 | *KIFAP3* |
| | 1q44 | RP11-706E22 | 242,748,587 | 1716 | 2.0 | 29 | *TFB2M, ZNF124, ZNF695* |
| | 4p16.1-p15.33 | RP11-270I3 | 10,716,786 | 930 | 1.7 | 1 | *HS3ST1* |
| | 4q21.1 | RP11-200G12 | 76,740,121 | 2033 | 1.6 | 21 | *CCNG2, SEPT11* |
| | 9p13.3 | RP11-752M5 | 34,421,289 | 1076 | 2.4 | 25 | *FANCG, STOML2* |
| | 11q13.4 | CTD-2165B14 | 72,376,209 | 1203 | 1.7 | 15 | |
| | 11q22.1 | RP11-239C4 | 98,998,299 | 2689 | 1.8 | 8 | *PGR* |
| | 14q32.11-q32.12 | RP11-661E19 | 89,918,395 | 545 | 1.6 | 4 | *CALM1* |
| | 15q11.2 | RP11-638O1 | 19,068,207 | 978 | 2.0 | 4 | |
| | 15q11.2 | RP11-716E17 | 21,543,989 | 1390 | 2.6 | 3 | |
| | 17q12 | RP11-8D3 | 31,132,119 | 943 | 2.3 | 29 | *MYOHD1* |
| | 17q12-q21.2 | RP11-722B4 | 32,333,405 | 3319 | 4.0 | 51 | *STARD3, PERLD1, ERBB2, GRB7* |
| | 17q21.32-q21.33 | RP13-495A21 | 43,597,619 | 2783 | 3.6 | 57 | *HOXB7, FLJ13855, EAP30, PHB, PPP1R9B* |
| | 17q22 | RP11-734K17 | 48,575,697 | 999 | 2.0 | 1 | *KIF2B* |
| | 17q22-q23.2 | RP11-515J20 | 49,994,527 | 2212 | 2.6 | 9 | *TOM1L1, COX11* |
| | 17q23.2 | RP11-639P5 | 54,011,739 | 2731 | 3.1 | 23 | *RAD51C, TRIM37, FLJ10587, CLTC, Bit1, TUBD1, BCAS3* |
| | 20q11.22 | RP11-601G7 | 32,388,522 | 1161 | 3.0 | 18 | *NCOA6* |
| | 20q13.12 | RP11-770M1 | 42,271,027 | 1082 | 2.2 | 19 | *STK4, C20orf121* |
| | 20q13.12-q13.32 | RP11-702E3 | 45,194,473 | 11661 | 4.1 | >50 | *NCOA3, STK6, BMP7, RAB22A, ZNF217* |
| | 20q13.32-q13.33 | RP11-648D7 | 57,579,717 | 314 | 2.8 | 2 | |
| HCC1937 | 3q27.2-27.3 | RP11-238G24 | 186,795,912 | 2556 | 2.0 | 24 | |
| | 3q29 | RP11-272C21 | 194,809,788 | 895 | 2.3 | 7 | |
| | 16p11.2-p11.1 | RP11-258P17 | 34,135,769 | 578 | 1.9 | – | |
| HCC2218 | 1q42.12-q42.2 | CTD-2185P6 | 222,699,955 | 7054 | 1.9 | >50 | |
| | 8q11.21 | RP11-163E15 | 51,161,135 | 560 | 1.8 | 1 | *SNTG1* |
| | 17q12-q21.1 | RP11-85O18 | 34,781,214 | 579 | 3.3 | 15 | *PERLD1, ERBB2, GRB7* |
| | 17q21.32-q21.33 | RP11-759D3 | 43,752,243 | 2808 | 3.6 | >50 | *HOXB7, FLJ13855, EAP30, PHB, PPP1R9B* |
| | 17q23.2-q23.3 | RP11-758H9 | 54,973,342 | 2548 | 3.5 | 21 | *RPS6KB1, TBX2* |
| | 17q25.1 | RP11-751O16 | 68,898,787 | 2406 | 2.7 | >50 | |

(Continued)

TABLE I. Regions Found Amplified in 10 Breast Cancer Cell Lines (Continued)

| Cell line[a] | Cytoband | Start Clone | Start position (bp) | Size (kbp) | Peak log(2)R | N genes | Candidate target genes |
|---|---|---|---|---|---|---|---|
| SK-BR-3 | 8q13.3-q21.13 | RP11-746L20 | 71,312,586 | 10798 | 3.2 | 33 | *LACTB2, TCEB1, MRPS28* |
| | 8q21.2 | RP11-509F16 | 86,620,595 | 278 | 2.0 | 1 | |
| | 8q21.3 | RP11-778C19 | 87,744,061 | 652 | 2.1 | 2 | |
| | 8q21.3 | RP11-502M10 | 88,591,743 | 1366 | 3.1 | 3 | |
| | 8q21.3 | RP11-627A6 | 90,790,023 | 598 | 3.3 | 5 | *DECR1* |
| | 8q21.3 | RP11-662E23 | 91,717,298 | 836 | 2.1 | 6 | *CGI-77* |
| | 8q23.3-q24.13 | RP11-500K1 | 112,613,664 | 14089 | 3.8 | 47 | *EIF3S3, MAL2* |
| | 8q24.21 | RP11-664D24 | 127,851,608 | 1403 | 2.9 | 2 | *MYC* |
| | 17q12-q21.2 | RP11-25P3 | 34,574,333 | 2344 | 3.5 | >50 | *STARD3,PERLD1, ERBB2, GRB7* |
| | 17q25.3 | RP11-467J3 | 74,388,427 | 1367 | 1.7 | 13 | *CBX4, CBX2* |
| | 20q13.2 | RP11-359J9 | 50,602,592 | 2385 | 1.9 | 6 | *ZNF217* |
| MDA-MB-361 | 17q12 | RP11-747F4 | 30,835,528 | 631 | 1.9 | 17 | |
| | 17q12-q21.1 | RP11-689B15 | 34,946,406 | 414 | 2.3 | 12 | *STARD3,PERLD1, ERBB2, GRB7* |
| | 17q23.2 | RP11-720I5 | 54,168,604 | 1043 | 1.9 | 9 | *TRIM37, PPM1E, CLTC, Bit1* |
| | 17q23.3 | RP11-282E10 | 58,167,577 | 631 | 2.5 | 1 | *RFP190* |
| | 17q23.3-q24.1 | RP11-630H24 | 59,198,385 | 1112 | 2.6 | 21 | *CCDC47, SMARCD2, ERN1* |
| | 17q24.1 | RP11-484A6 | 60,383,462 | 895 | 2.1 | 4 | *GNA13* |
| | 17q25.1 | RP11-713D20 | 69,101,794 | 1079 | 1.8 | 12 | *GPRC5C, RAB37, FLJ40319* |
| JIMT1 | 1p32.3 | RP11-92H3 | 54,460,877 | 1104 | 2.2 | 12 | |
| | 1p32.2-p31.3 | RP11-636K17 | 58,363,970 | 4397 | 2.0 | 14 | *USP1* |
| | 1p22.2 | RP11-592L19 | 88,603,537 | 515 | 1.8 | 2 | *PKN2* |
| | 1q23.3 | RP11-418C6 | 157,394,241 | 700 | 2.3 | 25 | *PFDN2* |
| | 1q24.1-q24.2 | CTD-2017D9 | 163,853,475 | 545 | 2.3 | 4 | |
| | 1q24.3-q25.1 | RP11-220H11 | 168,116,428 | 2510 | 1.9 | 19 | |
| | 1q25.2 | RP11-604C16 | 173,599,415 | 1208 | 1.7 | 3 | |
| | 3q22.3 | RP11-394K22 | 139,420,142 | 600 | 1.9 | 7 | *PIK3CB, FAIM* |
| | 8p12-p11.23 | RP11-621B1 | 37,365,131 | 1870 | 2.2 | 24 | *FGFR1, BLP1* |
| | 11p15.3-p15.2 | RP11-204J9 | 12,314,946 | 3227 | 2.0 | 15 | |
| | 11p15.1 | RP11-14D9 | 17,443,334 | 2049 | 2.0 | 27 | *LOC11317, PSMA1, FLJ23311* |
| | 11p13 | RP11-64P1 | 34,699,699 | 1867 | 1.7 | 12 | |
| | 12q13.3-q14.1 | RP11-799H16 | 56,206,303 | 1474 | 3.0 | 18 | *CDK4, MARCH-IX* |
| | 17q12-q21.2 | RP11-62P3 | 34,442,984 | 1210 | 3.4 | 27 | *STARD3,PERLD1, ERBB2, GRB7* |
| | 17q21.31 | RP11-21I21 | 38,864,523 | 727 | 3.1 | 14 | *ETV4, MPP3, DUSP3* |
| L56Br | 6q14.1 | RP11-467K7 | 78,789,403 | 327 | 2.0 | – | |
| | 6q14.1 | RP11-14A19 | 79,370,442 | 1155 | 2.0 | 5 | *PHIP, HMGN3* |
| | 6q14.1 | RP11-316P15 | 81,083,056 | 530 | 2.8 | 1 | |
| | 6q14.1 | RP11-185M7 | 81,784,421 | 720 | 3.1 | – | |
| | 6q22.31 | RP11-14D7 | 124,548,194 | 634 | 2.1 | 1 | |
| | 6q23.2 | RP11-203B4 | 133,490,152 | 231 | 2.3 | 1 | *EYA4* |
| | 6q23.2-q23.3 | CTD-2130M21 | 134,789,587 | 3415 | 3.1 | 15 | *MYB, FAM54A, AHI1* |
| | 6q24.1 | RP11-649D20 | 139,835,754 | 1749 | 2.8 | – | |
| | 6q24.1-q24.2 | RP11-89C11 | 142,505,116 | 3000 | 3.4 | 12 | *AIG1, C6orf93, PHACTR2, SF3B5, UTRN* |
| | 7q21.11-q21.12 | RP11-287O4 | 85,757,083 | 807 | 2.0 | 4 | *C7orf23* |

(Continued)

TABLE 1. Regions Found Amplified in 10 Breast Cancer Cell Lines (Continued)

| Cell line[a] | Cytoband | Start Clone | Start position (bp) | Size (kbp) | Peak log(2)R | N genes | Candidate target genes |
|---|---|---|---|---|---|---|---|
| | 11p13 | RP11-92D15 | 32,037,530 | 330 | 1.7 | 1 | |
| | 11p11.2 | RP11-784M21 | 44,665,246 | 1022 | 2.6 | 5 | *TP53I11* |
| | 11p11.2 | RP11-318N19 | 48,092,602 | 242 | 2.1 | 6 | |
| ZR-75-1 | 11q13.2 | RP11-157K17 | 66,670,511 | 504 | 1.9 | 24 | *RPS6KB2* |
| | 11q13.3-q13.4 | CTD-2009H2 | 68,979,527 | 1195 | 1.9 | 11 | *CCND1, FADD, CTTN* |
| MDA-MB-231 | 6p21.31-p21.2 | RP11-479F12 | 35,048,432 | 3515 | 1.6 | 42 | *STK38, MAPK13, FANCE, MTCH1* |
| | 6p21.1 | RP11-279E6 | 41,040,461 | 2230 | 1.6 | 40 | *RPL7L1, TRERF1, MRPL2, BYSL, PPP2R5D, MEA1* |

MCF10A cells had no amplification.

overlap with the MCF7 20q amplicons. The most centromeric region maps to 20q11.22, spans ~1 Mbp including 17 known genes, six of which have an elevated expression (*ITGB4BP*, *C20orf44*, *CEP2*, *GSS*, *NCOA6*, and *CDC91L1*). The third amplicon located in 20q13.12 is small (~965 kbp) and includes 19 known genes, but none of these show increased expression. Moreover, BT474 cells harbored increased gene copy number and expression of *CCND1* and *EMSY* on 11q13. BT474 cells are both ER and ERBB2 positive, representing a rather unusual subtype in breast cancer since ERBB2 positive breast cancers are predominantly ER negative. BT474 shows strongest correlation to the Luminal A subtype centroid, and inverse correlation to the basal-like and, surprisingly, ERBB2+ subtypes.

### HCC1937

HCC1937 is derived from a *BRCA1* mutation (5382insC) carrier and has no wildtype *BRCA1* allele. These cells have a high degree of aneuploidy, an acquired *TP53* mutation and homozygous deletion of *PTEN*, as well as LOH at multiple loci known to be involved in the pathogenesis of breast cancer (Tomlinson et al., 1998). Here, we confirmed the high frequency of CNA in HCC1937 cells, with gain on 3q13.11-qter and losses on 4pter-p16.1, 4q13.1-q21.23 and 5q11.2-q14.3. Gene amplifications were limited to three regions, two residing on the 3q arm, which has an overall copy number gain, and one on 16p. In addition to the known homozygous deletion on 10q23.31 including *PTEN*, a second more centromeric homozygous deletion on 10q21.3 was discovered, but contains no known genes. A homozygous deletion was also found on 18q21.1, including the *FUSSEL18* gene. HCC1937 cells are ER negative and show an expected correlation to the basal-like phenotype.

### SK-BR-3

SK-BR-3 originates from a breast adenocarcinoma pleural effusion metastasis. It is hypertriploid with the modal chromosome number of 80–84, and is known to express ERBB2 (ATCC). Here, we confirm the high-level amplification at the *ERBB2* locus, spanning ~1.3 Mbp. As expected, a large number of other genes have an elevated expression. Moreover, a complex amplification pattern was observed on 8q displaying both losses and high-level amplifications. Five amplicons are found in the 8q13.3-q21.3 region, three of them reaching log(2)ratios > 2. A large region (~13 Mbp) on 8q23.3-q24.13 displayed high-level

TABLE 2. Regions with Homozygous Deletions in 11 Breast Tumor Cell Lines

| Cell line | Cytoband | Start clone | Start position (bp) | Size (kbp) | Genes |
|---|---|---|---|---|---|
| MCF7 | 4q34.3 | RP11-400H15 | 182202439 | 759 | No genes |
| HCC1937 | 10q21.3 | RP11-757B19 | 65916163 | 1391 | No genes |
| | 10q23.31 | RP11-210E13 | 89400386 | 579 | *PTEN* |
| | 18q21.1 | RP11-391C19 | 43010108 | 264 | *FUSSEL18* |
| SK-BR-3 | 16q22.1 | RP11-604O10 | 67167809 | 263 | *CDH1* |
| | 19p12 | RP11-1318H24 | 20254829 | 4018 | *ZNF493, ZNF43, ZNF253* |
| MDA-MB-361 | 20p12.1 | RP11-224A21 | 14802718 | 232 | No genes |
| MDA-MB-231 | 9p22.3-p22.2 | RP11-554H2 | 16354864 | 1621 | *C9orf39, BNC2* |
| | 9p21.3 | RP11-66P3 | 20784479 | 3556 | *CDKN2A, CDKN2B* |
| | 9p21.3 | RP11-20A20 | 24485736 | 467 | |
| MCF10A | 9p21.3 | RP11-615P15 | 21754402 | 397 | *CDKN2A, CDKN2B* |
| L56Br-C1 | 10q26.12 | RP11-499E6 | 122574466 | 435 | *WDR11* |
| JIMT1 | 17q24.3 | RP11-74N19 | 67077096 | 310 | No genes |
| | 17q24.3 | RP11-95C11 | 67746245 | 313 | No genes |

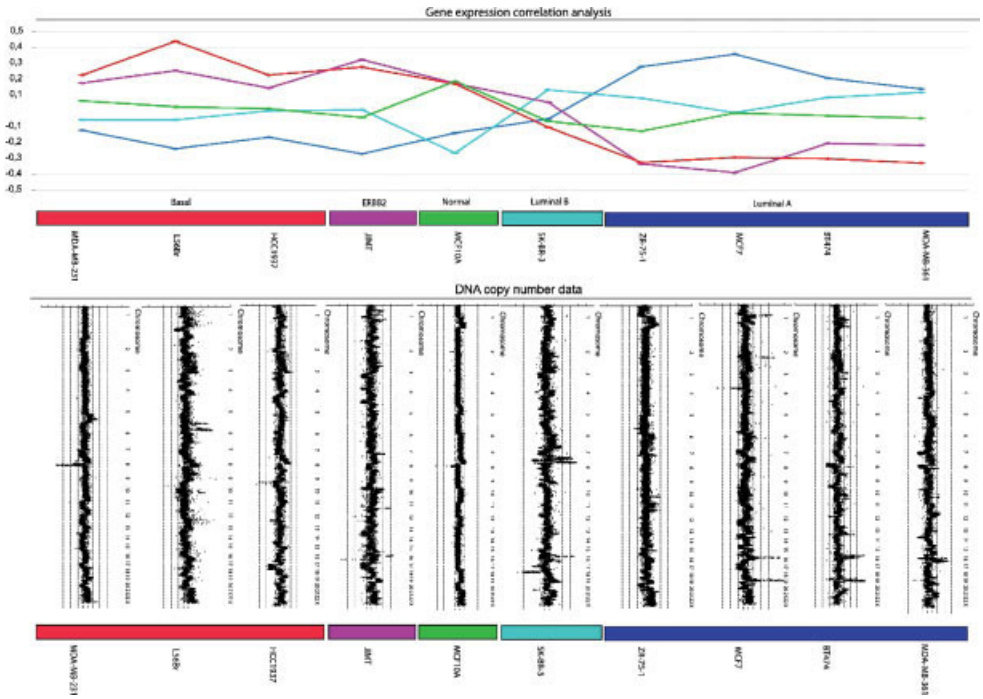


Figure 2. Breast cancer cell lines were classified into molecular breast cancer subtypes. Top panel displays gene expression classification results based on the "intrinsic 500 gene set." Correlation coefficients for each subtype is calculated and plotted for each cell line. Three cell lines (L56Br-C1, HCC1937, and MDA-MB-231) had highest correlation coefficient for the basal-like subtype, one cell line (JIMT1) had highest for the ERBB2 subgroup, one cell line (MCF10A) had highest for the normal-like subtype, one cell line (SK-BR-3) had highest for the luminal B subtype and four cell lines (MCF7, BT474, MDA-MB-361, and ZR-75-1) had highest correlation for the luminal A subtype. Bottom panel displays genome wide DNA copy number profiles for each breast cancer subtype. High-resolution profiles for individual chromosomes can be viewed at http://swegene.onk.lu.se.

amplification and included 30 genes of which four show an elevated expression (*EIF3S3*, *MAL2*, *ATAD2*, *MGC2165*). Finally, a narrow amplification peak maps to 8q24.21 only comprise *MYC*, which intriguingly did not show an elevated expression. Two previously reported homozygous deletions were also detected, one at 16q22.1 including *CDH1*, and the second at 19p12 including a number of zinc finger genes (Table 2) (Hiraguri et al., 1998). Interestingly, SK-BR-3 shows no strong correlation or inverse correlation to any of the five pheno-subtypes, but is closest to luminal B followed by the ERBB2+ type.

## MDA-MB-361

MDA-MB-361 is derived from a breast adenocarcinoma brain metastasis. It is hyperdiploid with a

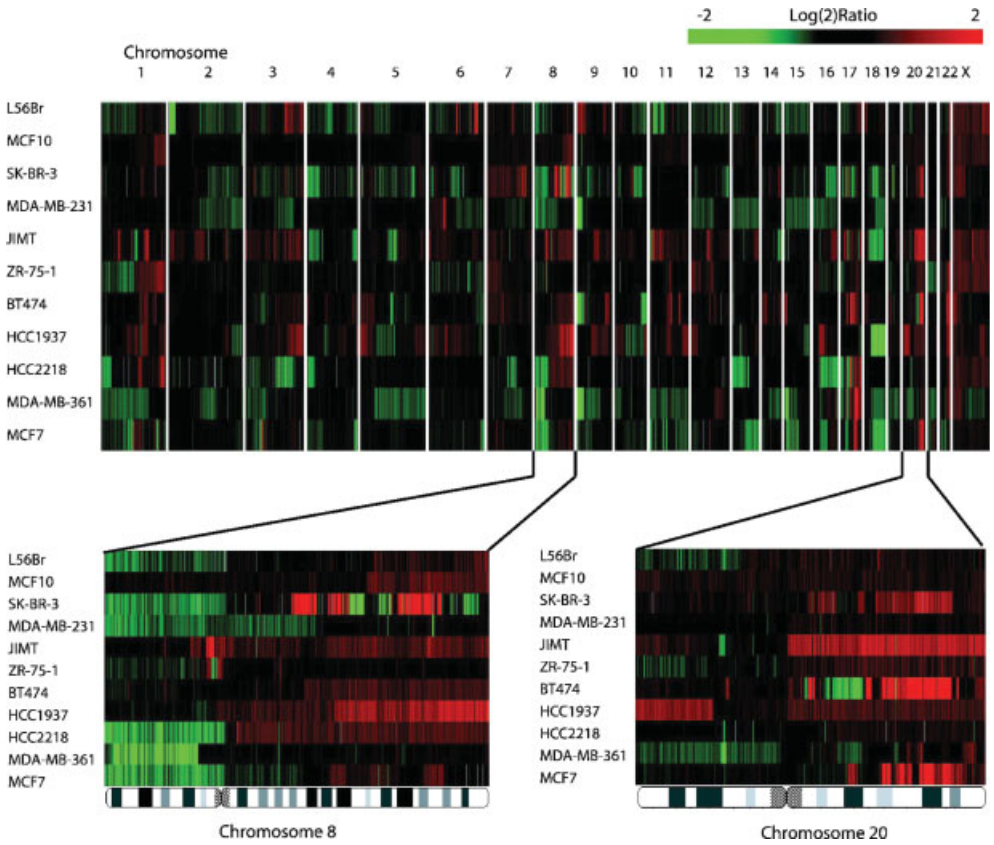


Figure 3. Genomic profiling of breast cancer cells demonstrating the precise delineation and resolving power of tiling arrays. BAC clones are ordered according to their genomic localization on the vertical axis. Cell lines are indicated on the horizontal axis. Red represents increased copy number and green represents decreased copy number. Zoom-in of chromosome 8 is showed in the lower left panel where a complex amplification pattern is seen for SK-BR-3, but it is also apparent that increased copy number of 8q is common in breast cancer. Lower right panel displays a zoom-in of chromosome 20 where interesting 20q amplification patterns are shown.

modal chromosome number of 54-61 (ATCC). Here, we found large regions of CNA affecting 5p, 8q, 12q, 16, 17q, 20q (gains) and 1p, 2q, 3p, 5q, 7q, 8p, 9, 10p, 11, 14q, 15q, 17p, 18q, 19, 20p, 21 (losses). High-level amplifications were few and confined to chromosome 17. One narrow peak was observed at the *ERBB2* locus on 17q12-q21.1, and another at 17q23.3 including only one gene, encoding a ring finger protein 190. The amplified region on 17q23.3-q24.1 includes seven known genes, of which only *POLG2* had an elevated expression. Additional amplification was detected on 17q24.1 including three known genes (*RGS9*, *AXIN2*, and *CCDC46*); however, none of these have an increased expression. A copy number increase was also seen at *CCND1* on 11q13.3, but not including *EMSY* at 11q13.5. A homozygous deletion was found on 20p12.1, but again no genes reside in this region. Despite *ERBB2* amplification and overexpression, MDA-MB-361 cells do not have an ERBB2+ gene expression phenotype, but are rather of luminal A or B subtype, probably related to the expression of ER.

## MDA-MB-231

MDA-MB-231 is derived from adencarcinoma pleural effusion. It is near-triploid with a chromosome count of 52-68 (ATCC). Here, we found large regions with CNA at 6p, 20q (gains) and 2q, 3q, 6q, 8p, 9p, 12, 13, 15, 16, 18q, 22 (losses). No high-level amplification peaks were detected; however, multiple noncontiguous homozygous deletions were found on 9p. One (~1.4 Mbp) maps to 9p22.2 and includes three known genes (*BNC2*, *C9orf39*, and *SH3GL2*), a second deletion targets the 9p21.3 region with *CDKN2A* and *CDKN2B* as well as *MTAP*, *DMRTA1*, and *ELAVL2*. The third deletion target also maps to 9p21.3, spans ~400 kbp and includes no known genes. MDA-MB-231 cells are ER negative but express the EGFR, and show strongest correlation to the basal-like subtype.

### MCF10A

MCF10A is a nontumorigenic epithelial cell line derived from a 36-year-old female with fibrocystic disease of the breast (ATCC). Here, a number of CNA were detected, including gains at 1q, 5q22.3-qter as well as of whole chromosomes 7, 8, 11, 13, 19, and 20. Loss was observed at a ~3 Mbp region on 9p21.3, which includes a homozygous deletion spanning one BAC clone and *CDKN2A* and *CDKN2B*. MCF10A cells have a gene expression signature that correlates to the normal-like phenotype, but also to the basal-like and ERBB2+ subtypes, and that shows inverse correlation especially to the luminal B subtype.

### L56Br-C1

L56Br-C1 originates from a *BRCA1* germ line nonsense mutation (Q563X) carrier. It has no wild-type *BRCA1* allele, and has acquired a *TP53* mutation (Johannsson et al., 2003). As expected, a high frequency of CNAs was detected, including losses at 4q13.3-q24, 4q31.23-qter, 5q11.2-q14.2, and 5q35.1-qter. Copy number gain was detected at 3q22.1-qter. A high-level amplification has previously been mapped to 6q22-q24 (Kauraniemi et al., 2000). Here, this is verified and delineated into several narrow peaks, the first one maps to 6q23.2-q23.3 and includes six genes of which three (*HBS1L*, *MYB*, and *AHI1*) show clear overexpression. A second peak locates to 6q24.1 and contains no known genes, while the third amplicon at 6q24.2 includes 10 genes of which four show an increased expression (*FUCA2*, *C6orf93*, *SF3B5*, *UTRN*). Furthermore, chromosome 11 harbors two narrow amplifications, one at 11p11.2 includes five genes of which only *TP53I11* shows some increased activity. The second peak spans only ~240 kbp and includes a cluster of olfactory receptor genes (not included in the expression array) and *PTPRJ*, which did not show an elevated expression. A homozygous deletion was detected in chromosome band 10q26.12 spanning ~433kbp and covering the *WDR11* gene and the 5′ part of *FGFR2* gene. L56Br-C1 cells are ER negative and show strong correlation to the basal-like subtype and inverse correlation to luminal A.

### ZR-75-1

ZR-75-1 is derived from a ductal breast carcinoma ascites metastasis. It is hypertriploid with a modal chromosome number was 72 (ATCC). Here, CNA were found at 1q, 7p, 12p, 16p, 17q, 18, 19q, 20q, 22q (gains) and 1p, 17p, 21 (losses). No high-level amplification peak was observed; however, a narrow amplicon with a log(2)ratio of 1.8–2.0 was seen at 11q13.3 including *CCND1*, which was also found to be overexpressed. ZR-75-1 has a luminal A phenotype with inverse correlation to the basal-like and ERBB2+ subtypes.

### JIMT1

JIMT1 is an ER negative, *ERBB2* amplified cell line established from a ductal carcinoma pleural metastasis of a 62-year-old patient, who did not respond to Herceptin treatment (Tanner, et al., 2004). Here, the high-level *ERBB2* amplicon was confirmed to span ~950 kbp and include more than 20 genes. A narrow amplification peak was also present at 17q21.31 including 17 genes of which three, *ETV4*, *MPP3*, and *DUSP3*, display overexpression. Also, a 1.4 Mbp high-level amplification peak was observed at 12q13.3-q14.1 including 17 genes. The obvious target is *CDK4*, but also *MARCH-IX* shows increased expression. A homozygous deletion was identified on 17q24.3; however, no genes map to this region. JIMT1 cells show a strong correlation to the ERBB2+ and basal-like subtypes, and inverse correlation to the luminal A subtype.

### HCC2218

HCC2218 was derived from a primary invasive ductal carcinoma. These cells are ER negative, cytokeratin 19 positive, and express p53 protein (ATCC, Stephens et al., 2005). Here, several regions on chromosome 17 harboring high-level amplification were discovered. A distinct peak spanning 355 kbp at 17q12-q21 *ERBB2* locus includes only nine genes (*PPARBP*, *CRKS*, *NEUROD2*, *PPP1R1B*, *STARD3*, *TCAP*, *PNMT*, *PERLD1*, and *ERBB2*). A second amplicon of ~2.8 Mbp maps to 17q21.32-q21.33 and includes the *HOX* gene cluster among others. A third amplicon of ~2.6 Mbp maps to 17q23.2 and includes candidate target genes such as *RPS6KB1*, *BCAS3*, and *TBX2*. A fourth amplicon spans ~2.2 Mbp on 17q25.1 and contains a large number of genes. A homozygous deletion of *CDH1* at 16q22.1 was also discovered. Additional CNAs include losses at 1p36.33-p35.2, 3pter-p14.3, 3q13.3-q26.1, 4pter-p14, 7q11.22-q21.3, 8p, 13qter-q21.33, 16q, 17p as well as smaller regions on 17q. Regions of CN gains were found at 1p11.2-qter, 5p, 5q13.3-qter, 7q21.3-qter, 8q, and 14q32.12-qter, but also at 17q in a complex pattern. No RNA was obtained from HCC2218.

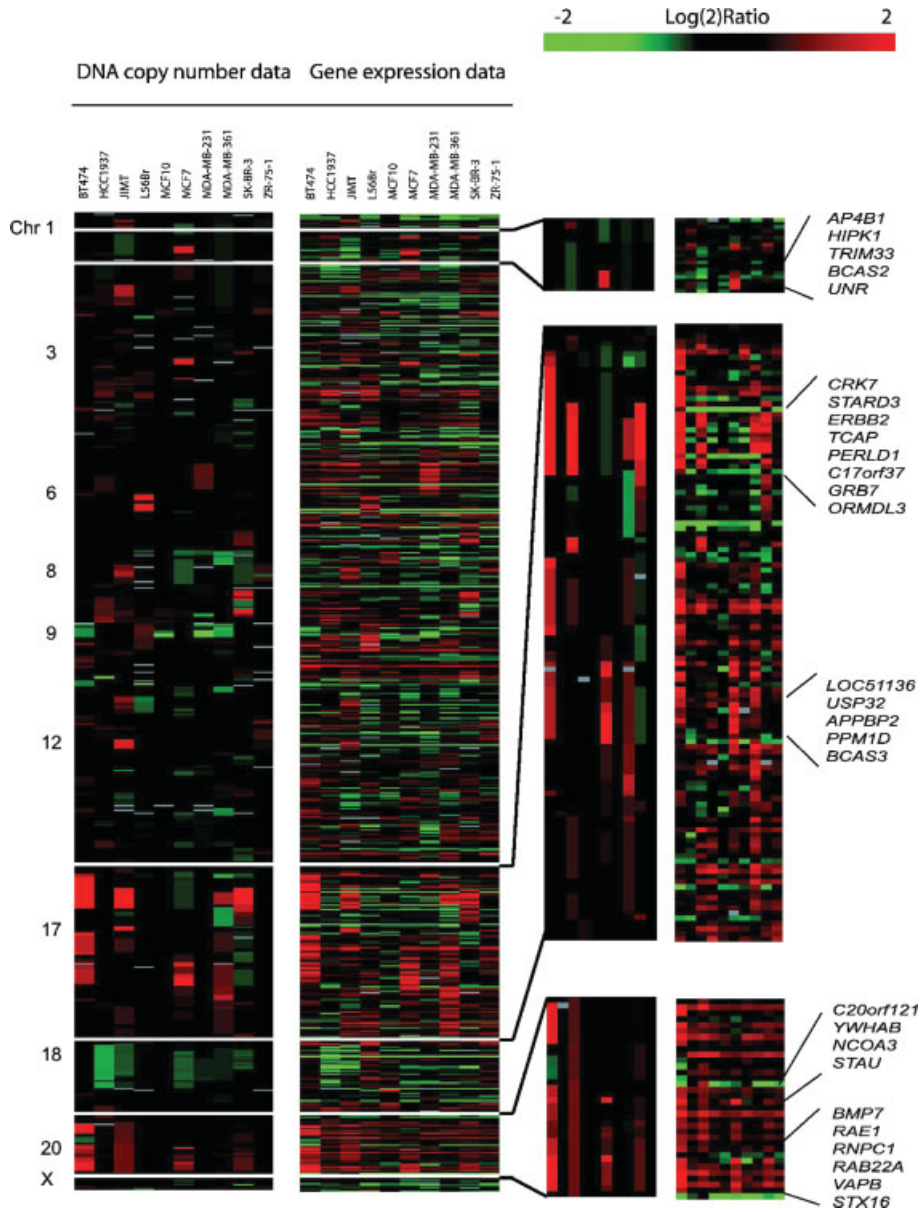


Figure 4. Correlation of gene expression and DNA copy number data. BAC clones and oligonucleotides are ordered vertically according to their genomic localization. Left column represents DNA copy number data and right column represents gene expression data. Red and green represents amplification/overexpression and deletion/low expression, respectively. Moreover, top right panel is a zoom-in of chromosome 1p13 amplified in MCF7 cells where *AP4B1*, *HIPK1*, *TRIM33*, *BCAS2* and *UNR* showed overexpression. Below, zoom-in of chromosome 17q including the ERBB2 amplicon as well as additional amplicons more distal on 17q. Bottom panel represents a zoom-in of chromosome 20q identifying several candidate oncogenes such as *NCOA3*, *BMP7*, and *RAB22A*.

### Integrating Gene Copy Number and Expression Data

Pearson correlation on data from matching BAC and oligonucleotide probes demonstrated that expression of a large number of genes appears to be affected by CNA (Fig. 4). This includes for instance several genes in addition to *ERBB2* in the 17q12-q21 amplicon, i.e., *STARD3*, *GRB7*, *PPARBP*, and *PERLDD1*. Overall, we observed 686 unique genes, where expression data and genomic alterations displayed a Pearson correlation $\geq$0.8, as compared to an expected 140 genes under the null hypothesis of random distribution at a $P$ value of 0.01. Chromosome arm 1q gain is a common CNA in breast cancer and here we observe 42 candidate genes with increased expression. Furthermore, 27 genes with significant copy-number gain and increased expression correlation were located on 8q, 124 genes on 17q and 34 genes on 20q. These correlations included also DNA copy number

losses and decreased gene expression, for instance for the *PTEN* gene, a target for deletion or homozygous deletion (in HCC1937). *CDKN2A* and *CDKN2B* located in chromosome band 9p21 also displayed a high correlation over assays indicating that the p16 pathway is commonly disrupted in breast cancer as well. The complete list of genes with significant correlation between DNA copy number and gene expression is found in Table 3 (Supplementary material for this article can be found at http://www.interscience.wiley.com/jpages/1045-2257/suppmat).

## DISCUSSION

In this study we describe a BAC microarray-based platform for CGH and high-resolution characterization of genomic alterations and evaluate its performance in 11 breast cell lines, also characterized by gene expression analysis. Genome-wide DNA copy number aberrations are profiled along each chromosome down to single genes affected by narrow amplifications or homozygous deletions. The effect of gene amplification is further verified by the level of mRNA expression to distinguish targeted genes from silent amplicon passengers. We have included 11 cell lines that were shown to represent all five recently described gene-expression based subtypes of breast cancer; the luminal A and B, basal-like, ERBB2 positive, and normal-like subtypes (Perou et al., 2000; Sorlie et al., 2003). This allows us to delineate whether subtype specific genomic profiles exist in breast cancer cell lines, as was recently suggested in breast tumors analyzed by CGH to cDNA microarrays (Bergamaschi et al., 2006). Seven of the cell lines used in present study were also investigated by Charafe-Jauffret et al. (2006) using gene expression profiling, resulting in similar grouping into basal, luminal, and mesenchymal subtypes.

The series included two cell lines established from *BRCA1* germline mutation carriers. As expected, the gene expression signatures of both HCC1937 and L56Br-C1 show strongest similarity to the basal-like subtype. Their genomes harbor the typical high frequency of low-level CNA as has been demonstrated in *BRCA1* tumors (Jönsson et al., 2005; van Beers et al., 2005). The patterns of CNA are also similar to what are found in *BRCA1* tumors, as well as in basal-like tumors (Wang et al., 2004; Bergamaschi et al., 2006) and, to lesser extent, in ER negative tumors (Loo et al., 2004). For instance, chromosome arm 5q deletions are frequently seen in *BRCA1* tumors (Johannsdottir et al., 2006) and were here restricted to deletions on 5q11.2-q14.3 in HCC1937 cells and on 5q11.2-q14.2 and 5q35.1-qter in L56Br cells. Genes within these regions with decreased expression in HCC1937 and L56Br cells include *PELO*, *IL31RA*, *PIK3R1*, *MGC13034*, *IQGAP2*, *LHFPL2*, and *APG10L*, possibly representing targets for deletions. Intriguingly, *PIK3R1* was recently found to be homozygously deleted in a tumor derived from a *BRCA1* mutation carrier (Johannsdottir et al., 2006). Also the third cell line classified as of basal-like subtype (MDA-MB-231) harbored a high frequency of low-level CNA similar to the *BRCA1* mutated cell lines, emphasizing a common genomic program in the pathogenesis of *BRCA1*/basal-like tumors.

Four cell lines (ZR-75-1, BT474, MCF7, and MDA-MB-361) had a gene expression signature closest to the luminal A subtype, which fits well with their ER positive phenotype. ZR-75-1, BT474, and MDA-MB-361 harbored amplification or increased copy number of the 11q13 locus including *CCND1*, which is known to be preferentially amplified in ER positive breast tumors (Reis-Filho et al., 2006). In general, these luminal A type cell lines had a lower frequency of CNA, common regions affected being found at 8p, 8q, 11q13, and 20q. BT474 and MDA-MB-361 cells have also massive amplification on several regions on 17q including *ERBB2*, which is less typical of ER positive and luminal A type tumors, suggesting that the ER signature is strong and overrides the influence of *ERBB2*. However, it should be noted that the two cell lines showing strongest resemblance with luminal A subtype are the *ERBB2* negative MCF7 and ZR-75-1 cells, whereas both BT474 and MDA-MB-361 show correlation also to the luminal B subtype. Most extreme in this respect is SK-BR-3, here classified as luminal B-like but in fact not showing strong correlation or anticorrelation to any subtype. SK-BR-3 cells harbored several high-level amplification peaks including 17q12-q21 (*ERBB2*) and 20q, and massive amplification on 8q and increased activity of >20 genes, the latter which may be an attribute of luminal B tumors and a possible reason to their suggestive aggressive behavior (Bergamaschi et al., 2006). SK-BR-3 had also a homozygous deletion at 16q22.1 including *CDH1*. Loss of E-cadherin activity, by means of point mutations, promotor methylation or deletions, is typically connected to lobular breast tumors (Berx et al., 1996) and an epithelial-mesenchymal transition, which may contribute to the complex character of SK-BR-3.

JIMT1 was the only cell line in this series that had closest resemblance to the ERBB2+ subtype. It also shows correlation to the basal-like subtype, which may reflect the "epithelial progenitor cell" origin of JIMT1, expressing both basal CK5/14 and luminal CK8/18 cytokeratins (Tanner et al., 2004). JIMT1 was established from a patient clinically resistant to Herceptin, despite having a tumor with *ERBB2* amplification. JIMT1 cells were found to express *ERBB2* as well as *ERBB1*, *ERBB3*, and *ERBB4* mRNA at similar levels as trastuzumab-sensitive SK-BR-3 cells, providing no clue for the Herceptin-resistance (Tanner et al., 2004). Here we found that JIMT1 cells have a ~950 kbp amplification peak at the *ERBB2* locus, but also numerous other CNAs including hemizygous loss of 10q and *PTEN*, suggesting alternative explanations for their trastuzumab resistance (Fig. 4). The CNA profile of JIMT1 is dissimilar to HCC1937 and L56Br-C1 cells and not a typical basal-like sample harboring low-level gains and losses.

MCF10A is nontumorigenic and derived from a breast fibrocystic lesion. As expected, its gene expression signature was normal-like, although similar also to the basal-like and ERBB2+ subtypes. Intriguingly, although MCF10A have a much less complex genomic profile compared to the cancer cell lines, it harbors major alterations such as whole chromosome 7, 8, 11, 13, 19, and 20 gains, as well as a narrow homozygous deletion at 9p21.3 spanning only *CDKN2A* and *CDKN2B*. HCC2218 was included in the series because of its "mutator" phenotype, as recently shown in a comprehensive screening of the protein kinase family for somatic mutations (Stephens et al., 2005). That study suggests a novel DNA repair deficiency and mechanism in breast carcinogenesis, with accumulation of numerous point mutations, primarily tranversions at G:C bp, distinct from the mismatch repair defects and MIN phenotype in colorectal cancers. Whereas the latter typically have a near-diploid genome, we here show that HCC2218 cells have acquired many CNA, including *ERBB2* amplification and a 17q23 amplicon, suggesting that a high rate of somatic point mutations is not sufficient in tumor development.

The high resolution of the tiling BAC-array platform allowed discovery of 14 homozygous deletions in 8 cell lines, many of which have not previously been described (Table 2). These include well-known tumor suppressor genes such as *PTEN* (HCC1937), *CDKN2A* (MCF10A, MDA-MB-231), and *CDH1* (SK-BR-3), but also novel regions. A 264 kbp deleted region on 18q21.1 in HCC1937 included *FUSSEL18*, described to interact with Smad proteins and TGF-βsignaling (Arndt et al., 2005). Our data also show very low expression of *SMAD2*, *4* and 7 in HCC1937. L56Br-C1 cells displayed a homozygous deletion on 10q26.13 including a single gene, *WDR11*, also a candidate tumor suppressor gene found disrupted in glioblastoma cells (Chernova et al., 2001). Other affected regions in chromosome bands 9p22 and 19p12 included zing finger proteins of unknown role in breast cancer. It should be noted that homozygous deletions of *CDKN2A* may be restricted to cultured cells, since, to our knowledge, this has not been reported in primary breast tumors. Moreover, several regions with homozygous deletions included no known genes (Table 2). While these could target important regulatory elements, we cannot exclude that they are constitutional CNA polymorphisms.

We have constructed a 32k BAC array CGH platform for genome-wide determination of CNA in cancer genomes. It allows precise mapping of segmental gains and losses, resolution of complex amplicons into narrow peaks, and identification of homozygous deletions of sizes equivalent of single genes (<100 kbp). We have confirmed its ability to detect single copy gains and losses and linearity in quantification of gene amplifications. The same BAC clone library was earlier used to establish sub-megabase-resolution tiling (SMRT) arrays (Ishkanian et al., 2004) and subsequently applied in analysis of breast cancer cells (Shadeo and Lam, 2006), including four (MCF7, BT474, SK-BR-3, MDA-MB-231) of the cell lines used in the present study. In addition, five of the cell lines (MCF7, HCC2218, MDA-MB-231, HCC1937, and MDA-MB-361) have been analyzed using a 10K SNP array (http://www.sanger.ac.uk/cgi-bin/genetics/CGP/CGHviewer.cgi?tissue = breast). This provides opportunity for direct comparison, and data from copy number gains and amplifications are in agreement with results presented here, most discrepancies being explained by the definition of signal cutoff levels. However, while a number of homozygous deletions were discovered in the present study, none were reported by Shadeo and Lam, or obvious using the 10K SNP arrays referred to above, suggesting that differences in clone processing and array production or DNA preparation and hybridization conditions may be of importance. Microarrays constructed from BAC clones constitute excellent probes for hybridization because of multiple PCR representations and large fragment size, providing robust conditions allowing also partly fragmented DNA to be analyzed. BAC

arrays provide a superior signal-to-noise ratio as compared to short-oligo arrays, but fall short in regard to possibilities of probe design (e.g., avoiding cross-hybridization to low copy repeats), industrial standardized production and potential higher print density and resolution. SNP-based arrays have the further advantage of allowing allele-specific analysis and detection of copy number neutral alterations. Regardless of which high-density platform preferred, they offer powerful tools for cancer genome profiling, diagnostics, and prediction of clinical outcome.

## REFERENCES


Albertson DG, Pinkel D. 2003. Genomic microarrays in human genetic disease and cancer. Hum Mol Genet 12 Spec No 2:R145–R152.

Andersson A, Edén P, Lindgren D, Nilsson J, Lassen C, Heldrup J, Fontes M, Borg A, Mitelman F, Johansson B, Hoglund M, Fioretos T. 2005. Gene expression profiling of leukemic cell lines reveals conserved molecular signatures among subtypes with specific genetic aberrations. Leukemia 19:1042–1050.

Arndt S, Poser I, Schubert T, Moser M, Bosserhoff AK. 2005. Cloning and functional characterization of a new Ski homolog, Fussel-18, specifically expressed in neuronal tissues. Lab Invest 85:1330–1341.

Autio R, Hautaniemi S, Kauraniemi P, Yli-Harja O, Astola J, Wolf M, Kallioniemi A. 2003. CGH-Plotter: MATLAB toolbox for CGH-data analysis. Bioinformatics 19:1714–1715.

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

Berx G, Cleton-Jansen AM, Strumane K, de Leeuw WJ, Nollet F, van Roy F, Cornelisse C. 1996. E-cadherin is inactivated in a majority of invasive human lobular breast cancers by truncation mutations throughout its extracellular domain. Oncogene 13: 1919–1925.

Charafe-Jauffret E, Ginestier C, Monville F, Finetti P, Adelaide J, Cervera N, Fekairi S, Xerri L, Jacquemier J, Birnbaum D, Bertucci F. 2006. Gene expression profiling of breast cell lines identifies potential new basal markers. Oncogene 25:2273–2284.

Chernova OB, Hunyadi A, Malaj E, Pan H, Crooks C, Roe B, Cowell JK. 2001. A novel member of the WD-repeat gene family, *WDR11*, maps to the 10q26 region and is disrupted by a chromosome translocation in human glioblastoma cells. Oncogene 20:5378–5392.

Hiraguri S, Godfrey T, Nakamura H, Graff J, Collins C, Shayesteh L, Doggett N, Johnson K, Wheelock M, Herman J, Baylin S, Pinkel D, Gray J. 1998. Mechanisms of inactivation of E-cadherin in breast cancer cell lines. Cancer Res 58:1972–1977.

Hoglund M, Johansson B, Pedersen-Bjergaard J, Marynen P, Mitelman F. 1996. Molecular characterization of 12p abnormalities in hematologic malignancies: Deletion of KIP1, rearrangement of TEL, and amplification of CCND2. Blood 87:324–330.

Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, Ling V, MacAulay C, Lam WL. 2004. A tiling resolution DNA microarray with complete coverage of the human genome. Nat Genet 36:299–303.

Johannsdottir HK, Jonsson G, Johannesdottir G, Agnarsson BA, Eerola H, Arason A, Heikkila P, Egilsson V, Olsson H, Johannsson OT, Nevanlinna H, Borg A, Barkardottir RB. 2006. Chromosome 5 imbalance mapping in breast tumors from BRCA1 and BRCA2 mutation carriers and sporadic breast tumors. Int J Cancer 119:1052–1060.

Johannsson OT, Staff S, Vallon-Christersson J, Kytola S, Gudjonsson T, Rennstam K, Hedenfalk IA, Adeyinka A, Kjellen E, Wennerberg J, Baldetorp B, Petersen OW, Olsson H, Oredsson S, Isola J, Borg A. 2003. Characterization of a novel breast carcinoma xenograft and cell line derived from a BRCA1 germ-line mutation carrier. Lab Invest 83:387–396.

Jonsson G, Bendahl PO, Sandberg T, Kurbasic A, Staaf J, Sunde L, Cruger DG, Ingvar C, Olsson H, Borg A. 2005a. Mapping of a novel ocular and cutaneous malignant melanoma susceptibility locus to chromosome 9q21.32. J Natl Cancer Inst 97:290–300.

Jonsson G, Naylor TL, Vallon-Christersson J, Staaf J, Huang J, Ward MR, Greshock JD, Luts L, Olsson H, Rahman N Stratton M, Ringner M, Borg A, Weber BL. 2005b. Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. Cancer Res 65:7612–7621.

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258:818–821.

Kauraniemi P, Hedenfalk I, Persson K, Duggan DJ, Tanner M, Johannsson O, Olsson H, Trent JM, Isola J, Borg A. 2000. *MYB* oncogene amplification in hereditary BRCA1 breast cancer. Cancer Res 60:5323–5328.

Krzywinski M, Bosdet I, Smailus D, Chiu R, Mathewson C, Wye N, Barber S, Brown-John M, Chan S, Chand S, Cloutier A, Girn N, Lee D, Masson A, Mayo M, Olson T, Pandoh P, Prabhu AL, Schoenmakers E, Tsai M, Albertson D, Lam W, Choy CO, Osoegawa K, Zhao S, de Jong PJ, Schein J, Jones S, Marra MA. 2004. A set of BAC clones spanning the human genome. Nucleic Acids Res 32:3651–3660.

Loo LW, Grove DI, Williams EM, Neal CL, Cousens LA, Schubert EL, Holcomb IN, Massa HF, Glogovac J, Li CI, Malone KE, Daling JR, Delrow JJ, Trask BJ, Hsu L, Porter PL. 2004. Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. Cancer Res 64:8541–8549.

Naylor TL, Greshock J, Wang Y, Colligon T, Yu QC, Clemmer V, Zaks TZ, Weber BL. 2005. High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. Breast Cancer Res 7:R1186–R1198.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D. 2000. Molecular portraits of human breast tumours. Nature 406:747–752.

Reis-Filho JS, Savage K, Lambros MB, James M, Steele D, Jones RL, Dowsett M. 2006. Cyclin D1 protein overexpression and CCND1 amplification in breast carcinomas: an immunohistochemical and chromogenic in situ hybridisation analysis. Mod Pathol 19:999–1009.

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

Saal LH, Troein C, Vallon-Christersson J, Gruvberger S, Borg A, Peterson C. 2002. BioArray Software Environment (BASE): A platform for comprehensive management and analysis of microarray data. Genome Biol 3:SOFTWARE0003.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. 2004. Large-scale copy number polymorphism in the human genome. Science 305:525–528.

Shadeo A, Lam WL. 2006. Comprehensive copy number profiles of breast cancer cell model genomes. Breast Cancer Res 8:R9.

Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D, Albertson DG. 2001. Assembly of microarrays for genome-wide measurement of DNA copy number. Nat Genet 29:263–264.

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, Botstein D. 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 100:8418–8423.

Stephens P, Edkins S, Davies H, Greenman C, Cox C, Hunter C, Bignell G, Teague J, Smith R, Stevens C, O'Meara S, Parker A, Tarpey P, Avis T, Barthorpe A, Brackenbury L, Buck G, Butler A, Clements J, Cole J, Dicks E, Edwards K, Forbes S, Gorton M, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jones D, Kosmidou V, Laman R, Lugg R, Menzies A, Perry J, Petty R, Raine K, Shepherd R, Small A, Solomon H, Stephens Y, Tofts C, Varian J, Webb A, West S, Widaa S, Yates A, Brasseur F, Cooper CS, Flanagan AM, Green A, Knowles M, Leung SY, Looijenga LH, Malkowicz B, Pierotti MA, Teh B, Yuen ST, Nicholson AG, Lakhani

S, Easton DF, Weber BL, Stratton MR, Futreal PA, Wooster R. 2005. A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. Nat Genet 37:590–592.

Tanner M, Kapanen AI, Junttila T, Raheem O, Grenman S, Elo J, Elenius K, Isola J. 2004. Characterization of a novel cell line established from a patient with Herceptin-resistant breast cancer. Mol Cancer Ther 3:1585–1592.

Tirkkonen M, Johannsson O, Agnarsson BA, Olsson H, Ingvarsson S, Karhu R, Tanner M, Isola J, Barkardottir RB, Borg A, Kallioniemi OP. 1997. Distinct somatic genetic changes associated with tumor progression in carriers of BRCA1 and BRCA2 germ-line mutations. Cancer Res 57:1222–1227.

Tomlinson GE, Chen TT, Stastny VA, Virmani AK, Spillman MA, Tonk V, Blum JL, Schneider NR, Wistuba II, Shay JW, Minna JD, Gazdar AF. 1998. Characterization of a breast cancer cell line derived from a germ-line BRCA1 mutation carrier. Cancer Res 58:3237–3242.

van Beers EH, van Welsem T, Wessels LF, Li Y, Oldenburg RA, Devilee P, Cornelisse CJ, Verhoef S, Hogervorst FB, van't Veer LJ, Nederlof PM. 2005. Comparative genomic hybridization profiles in human BRCA1 and BRCA2 breast tumors highlight differential sets of genomic aberrations. Cancer Res 65:822–827.

Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, Janssen IM, van der Vliet WA, Huys EH, de Jong PJ, Hamel BC, Schoenmakers EF, Brunner HG, Veltman JA, van Kessel AG. 2004. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. Nat Genet 36:955–957.

Wang ZC, Lin M, Wei LJ, Li C, Miron A, Lodeiro G, Harris L, Ramaswamy S, Tanenbaum DM, Meyerson M, Iglehart JD, Richardson A. 2004. Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers. Cancer Res 64:64–71.

Weber BL. 2002. Cancer genomics. Cancer Cell 1:37–47.

Yang MC, Ruan QG, Yang JJ, Eckenrode S, Wu S, McIndoe RA, She JX. 2001. A statistical method for flagging weak spots improves normalization and ratio estimates in microarrays. Physiol Genomics 7:45–53.

Yang IV, Chen E, Hasseman JP, Liang W, Frank BC, Wang S, Sharov V, Saeed AI, White J, Li J, Lee NH, Yeatman TJ, Quackenbush J. 2002. Within the fold: Assessing differential expression measures and reproducibility in microarray assays. Genome Biol 3: research0062.