[CANCER RESEARCH 64, 6453–6460, September 15, 2004]

Genomic and Expression Profiling of Chromosome 17 in Breast Cancer Reveals Complex Patterns of Alterations and Novel Candidate Genes

Béatrice Orsetti,[1] Mélanie Nugoli,[1] Nathalie Cervera,[1] Laurence Lasorsa,[1] Paul Chuchana,[1] Lisa Ursule,[1] Catherine Nguyen,[2] Richard Redon,[3] Stanislas du Manoir,[3] Carmen Rodriguez,[1] and Charles Theillet[1]

[1]Génotypes et Phénotypes Tumoraux, EMI229 INSERM/Université Montpellier I, Montpellier, France; [2]ERM 206 INSERM/Université Aix-Marseille 2, Parc Scientifique de Luminy, Marseille cedex, France; and [3]IGBMC, U596 INSERM/Université Louis Pasteur, Parc d'Innovation, Illkirch cedex, France

ABSTRACT

Chromosome 17 is severely rearranged in breast cancer. Whereas the short arm undergoes frequent losses, the long arm harbors complex combinations of gains and losses. In this work we present a comprehensive study of quantitative anomalies at chromosome 17 by genomic array-comparative genomic hybridization and of associated RNA expression changes by cDNA arrays. We built a genomic array covering the entire chromosome at an average density of 1 clone per 0.5 Mb, and patterns of gains and losses were characterized in 30 breast cancer cell lines and 22 primary tumors. Genomic profiles indicated severe rearrangements. Compiling data from all samples, we subdivided chromosome 17 into 13 consensus segments: 4 regions showing mainly losses, 6 regions showing mainly gains, and 3 regions showing either gains or losses. Within these segments, smallest regions of overlap were defined (17 for gains and 16 for losses). Expression profiles were analyzed by means of cDNA arrays comprising 358 known genes at 17q. Comparison of expression changes with quantitative anomalies revealed that about half of the genes were consistently affected by copy number changes. We identified 85 genes overexpressed when gained (39 of which mapped within the smallest regions of overlap), 67 genes underexpressed when lost (32 of which mapped to minimal intervals of losses), and, interestingly, 32 genes showing reduced expression when gained. Candidate genes identified in this study belong to very diverse functional groups, and a number of them are novel candidates.

INTRODUCTION

Chromosome 17 is one of the smallest and most densely gene-loaded human chromosomes. It is frequently rearranged in human tumors and presents a number of rearrangement breakpoints mapping to either its short or long arm (1). Furthermore, comparative genomic hybridization (CGH) studies have shown it to harbor multiple regions of gains or losses in a variety of human cancers (2).

CGH, loss of heterozygosity, and molecular genetics data, taken together, show that chromosome 17 is rearranged in at least 30% of breast tumors (3, 4). Short and long arms differ in the type of events they harbor. Chromosome 17p is principally involved in losses, some of them possibly focal, whereas CGH on 17q shows complex combinations of overlapping gains and losses. Most recent efforts have focused on two regions of gains considered to be the principal events: 17q12-q21 corresponding to the amplification of *ERBB2* and collinear genes, and a large region at 17q23 (5, 6). A number of new candidate oncogenes have been identified, among which *GRB7* and *TOP2A* at 17q21 or *RP6SKB1*, *TBX2*, *PPM1D*, and *MUL* at 17q23 have drawn most attention (6–10). Furthermore, DNA microarray studies have revealed additional candidates, with some located outside current regions of gains, thus suggesting the existence of additional amplicons on 17q (8, 9).

Our previous loss of heterozygosity mapping data pointed to the existence at 17q of at least five regions of imbalance (of which two corresponded to DNA amplification; ref. 11). This is likely to be a minimal estimate, when taking into account similar data from the literature. This view was reinforced by fluorescence *in situ* hybridization studies performed in our laboratory[4] and confirmed by array-CGH (8, 9). Moreover, the observation of complex combinations of gains and losses within 40 to 50 Mb at 17q in individual breast tumors prompted us to further investigate these extensive rearrangements.

Our goal was to define with greater accuracy regions of copy number losses and/or gains on chromosome 17 and determine their boundaries. To do this, we applied the recently developed CGH on genomic arrays approach. We also sought to gain better insight on the genes involved and wanted to verify the existence of recurrent sites of rearrangements on chromosome 17. We built a genomic array covering chromosome 17 at a mean density of 1 clone per 500 Kb and used it to characterize patterns of gains and losses in 30 breast cancer cell lines and 22 primary breast tumors. Expression profiles of genomically typed tumors or cell lines were established using custom-made cDNA arrays comprising 376 expressed sequence tag sequences corresponding to 358 known genes mapping at 17q. This enabled the definition of regions of recurrent gains and losses. These were correlated with recurrent changes in expression levels that confirmed previously proposed candidates and identified novel genes. Furthermore, it appeared that individual tumors or cell lines could bear highly complex patterns of anomalies, cumulating in several amplification peaks and concomitant interstitial losses. Finally, because studied tumors and cell lines recurrently showed abrupt ruptures at the boundaries of some amplicons, we propose the existence of recurrent breakpoint sites.

MATERIALS AND METHODS

Cell Lines and Tumors. Breast cancer cell lines used in this study included BRCAMZ01, BRCAMZ02, MDAMB175, and MDAMB453 (D. Birnbaum; INSERM U119, Marseilles, France); CAL51PE, MDAMB435, SKBR7, and ZR7530 (P. Edwards; Department of Pathology, University of Cambridge, Cambridge, United Kingdom); BT474 and MCF7Rich (F. Vignon; INSERM U540, Montpellier, France); HS578T, MDAMB436, and HBL100 (A. Puisieux; INSERM U590, Lyon, France); SUM149, SUM185, and SUM52 (S. Ethier; University of Michigan, Ann Arbor, MI); EFM19, COLO824, EFM19, and EFM192A (DSMZ, Braunschweig, Germany); and BT20, BT483, CAMA1, HCC38, HCC1187, HCC1395, HCC1428, HCC1569, HCC1806, HCC1937, HCC1954, HCC2218, MCF7, MCF10F, MDAMB134,

[4] B. Orsetti, unpublished observations.

Received 3/2/04; revised 5/26/04; accepted 7/19/04.

Grant support: Funds from the CNRS, INSERM, and the Association de Recherche sur le Cancer and grant 5102, the Ligue Nationale de Lutte Contre le Cancer, as part of the Carte d'Identité des Tumeurs Program and the joint program Développement d'Outils de Diagnostic Moléculaire en Cancérologie: Applications aux Cancers du Sein Ministère de l'Enseignement Supérieur, de la Recherche et de la Technologie and Fédération Nationale des Centres de Lutte Contre le Cancer. M. Nugoli was supported by a doctoral fellowship from the Ligue Nationale Contre le Cancer. Array printing was done with the help of the Genopole platform.

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 can be found at Cancer Research Online (http://cancerres.aacrjournals.org).

Requests for reprints: Charles Theillet, EMI 229 INSERM, Centre de Recherche, CRLC Val d'Aurelle 34298 Montpellier cedex 5, France. Phone: 33-467-613-766; Fax: 33-467-613-041; E-mail: theillet@valdorel.fnclcc.fr.

MDAMB157, MDAMB231, MDAMB330, MDAMB361, MDAMB415, MDAMB468, SKBR3, T47D, UACC812, and ZR751 (American Type Culture Collection, Manassas, VA). All cell lines were maintained in Dulbecco's modified Eagle's medium or RPMI 1640 containing 10% fetal bovine serum supplemented with L-glutamine (200 mmol/L, 100×) and antibiotic-antimycotic (100×) (Life Technologies, Inc., Cergy Pontoise). A total of 55 primary breast cancers were collected at the Pathology Department of Val d'Aurelle Cancer Center (Montpellier, France). The present collection included 54.5% ductal carcinomas, 21.8% lobular carcinomas, 18.2% invasive carcinoma of undetermined type, and 5.5% of rare histologic subtypes. The Scarff and Bloom grade distribution was 3.6% grade 1, 34.5% grade 2, 50.9% grade 3, 10.9% nondetermined, 75% estrogen receptor positive, and 67% progesterone receptor positive.

Classical Comparative Genomic Hybridization. Normal metaphase chromosomes were prepared from umbilical cord blood according to standard cytogenetic protocols. Hybridizations were done on Vysis (Downers Grove, IL) normal human metaphases. Genomic DNA labeling and CGH reaction were performed as described by Courjal and Theillet (3). CGH images were captured on a Zeiss (Le Pecq, France) epifluorescence microscope equipped with a JAI (Glostrup, Denmark) charge-coupled device camera run by Metasystems (Altlussheim, Germany) image analysis software. CGH analysis was done using ISIS 4.4 software (Metasystems).

Genomic Arrays. The chromosome 17 genomic array consisted of 107 Roswell Park Cancer Institute (RPCI)-bacterial artificial chromosome (BAC) and P1 artificial chromosome (PAC) clones from the set of cytogenetically mapped clones reported previously,[5] 20 BACs selected using sequence data, and 46 BAC and PAC clones corresponding to genetic markers and known genes. A large majority of RPCI-1,3, 5 PAC clones and RPCI-11 BAC clones were obtained from the Children's Hospital Oakland Research Institute (Oakland, CA). Nine clones (CTD-2251J22, RP11-455O6, RP11-300G13, RP11-319A23, RP11-379P18, RP11-387C17, RP11-399J11, RP11-469C13, and RP11-489G5) were obtained from Research Genetics (Huntsville, AL). Clones corresponding to genetic markers were isolated from the Down to The Well human BAC library of GenomeSystems Inc. (St. Louis, MO). Clones D152 and PO135 were isolated from the RZPD Human Chromosome-sorted Cosmid Library of chromosome 17 (Berlin, Germany). Clones 56K13 and 201L4 were obtained by screening the HGMP Human PAC Library of the United Kingdom HGMP Resource Centre (Cambridge, United Kingdom). Cosmid clones Neu1 and Neu4 and P1 clone 610 were provided by Dr. A. Kallioniemi (Bethesda, MD). Clone P1.9 was from Dr. D. Viskochil (Salt Lake City, UT). See the list of clones in Supplementary Table S1.

Array-Comparative Genomic Hybridization Conditions. We isolated BAC, PAC, and cosmid DNA using Nucleobond BAC100 from Macherey-Nagel (Hoerdt, France). We carried out degenerated nucleotide primer (DOP)-polymerase chain reaction (PCR) amplification on 10 ng of prepared DNA in a final reaction volume of 100 μL. Primer sequences and the DOP-PCR protocol used are available on the Sanger Center web site (12).[6] We performed it with slight modifications: second-round DOP-PCR primer was not amino-linked in our experiments. Purification of PCR products was done using Nucleofast 96 PCR plates (Macherey-Nagel). Purified PCR products were resuspended in double-distilled H_2O at 2 μg/μL. An aliquot was run on an agarose gel to ascertain even distribution of product in all wells. Prior spotting products were diluted 1:1 in spotting solution (Amersham Biosciences, Orsay, France) and spotted in quadriplicate onto Corning GapsII slides (Schiphol-Rijk, the Netherlands) using a Lucidea array spotter IV (Amersham Biosciences).

Hybridization to Microarrays and Image and Data Analysis. Genomic DNA was digested by *Nde*II according to the supplier's recommendations (Roche Diagnostics, Meylan, France). Three hundred nanograms of digested genomic DNA were labeled by random priming in a 50-μL reaction containing 0.02 mmol/L dATP, 0.02 mmol/L dGTP, 0.02 mmol/L dTTP, 0.05 mmol/L dCTP, 0.04 mmol/L Cy3-dCTP or Cy5-dCTP; 25 units of Klenow fragment (50 units/μL; New England Biolabs, Ozyme, Saint Quentin Yvelines, France), 10 mmol/L β-mercaptoethanol, 5 mmol/L $MgCl_2$, 50 mmol/L Tris-HCl (pH 6.8), and 300 μg/mL random octamers. The reaction was incubated at 37°C for 20 hours and stopped by adding 2.5 μL of 0.5 mol/L EDTA (pH 8). The reaction product size was about 100 bp. We purified labeled products using microcon 30 filters (Amicon, Millipore, Molsheim, France). Abundance of the labeled DNA was checked using a spectrophotometer, and incorporation of dyes was calculated using Molecular Probes software.[7] A mixture of 700 pmol of Cy5-labeled probes and 700 pmol of Cy3-labeled probes was ethanol precipitated in the presence of 250 to 300 μg of human Cot-1 DNA (Roche Diagnostics) and 100 μg of herring sperm DNA (Promega, Charbonnières, France). The pellet was dried and resuspended in 280 μL of Hybrisol VII (Appligene Oncor, Qbiogen, Illkirch, France). The probes were denatured at 80°C for 10 minutes, and repetitive sequences were blocked by preannealing at 37°C for 90 minutes. Slides were blocked for 20 minutes at 42°C in saturation buffer (1% bovine serum albumin, 0.2% SDS, and 5× SSC), washed in 2× SSC and 0.2% SDS and then in 2× SSC, and dehydrated in an ethanol series. A 8.8-cm^2 open hybridization chamber (Gene Frame, Abgene, Courtaboeuf, France) was fixed on the slide, and the 280-μL preannealed mix was applied and hybridized in a humid chamber at 37°C on a rocking table for 16 hours. After hybridization, slides were washed in 2× SSC and 0.1% SDS (pH 7) at 55°C for 5 minutes and in 1× SSC and 0.1% SDS (pH 7) at 55°C for 5 minutes, followed by three washes in 0.1× SSC for 30 seconds at room temperature, and briefly rinsed in water. Slides were dried by spinning for 5 minutes at 1,000 rpm and stored at room temperature until scanned. Arrays were scanned by a GenIII Array Scanner (Amersham Biosciences). Images were analyzed by ARRAY-VISION 6.0 software (Amersham Biosciences). Spots were defined by use of the automatic grid feature of the software and manually adjusted when necessary. Fluorescence intensities of all spots were then calculated after subtraction of local background. These data were then analyzed using a custom-made MS-Excel VBA script. Cy3 and Cy5 global intensities were normalized with the entire set of spots on the array, the Cy3/Cy5 ratios were calculated, the median values of replicate spots were calculated, and these values were used to define the selection threshold for individual spots (only replicates showing <15% of deviation from the median were kept), with representation of profiles with log 2 ratios on the *Y* axis and Mb position of clones[8] along the chromosome on the *X* axis. For each sample, at least two experiments were performed (Cy3/Cy5 and Cy5/Cy3), and the final profile corresponds to the mean of two experiments.

Complementary DNA Array Construction and Analysis. Preparation and hybridization of cDNA arrays were as described previously (13). Of the 720 cDNAs spotted, 376 corresponded to 358 known genes positioned on chromosome 17 (Supplementary Table S2).[8] Hybridization signals were quantified using HDG Analyzer software (Genomic Solutions, Ann Arbor, MI) by integrating all spot pixel signal intensities and removing spot background values determined in the neighboring area.

Expression values for each sample were normalized according to the median expression levels in all samples (tumors and cell lines). This was done to favor the selection of expression differences related to quantitative genomic anomalies. Using an adaptation of the Spline function proposed by Cole (14), the variance was adjusted to be constant in the whole dataset (for low and high expression levels). Then a confidence interval determining genes that showed nonsignificant variation was defined. Its bandwidth was adjusted to fit the SD in the dataset. It encompassed 68.3% of the spots on the array. The distance separating the limit of the confidence interval from its orthogonal projection on the first diagonal was defined as the basic unit of expression variation. Thus, within the confidence interval, all values equaled $|1|$. This defined the baseline, and genes with values > 1 (spots above the first diagonal) were considered overexpressed, and genes with values < -1 (spots below the first diagonal) were considered underexpressed.

RESULTS

Genomic Profiling of Breast Cancer Cell Lines and Tumors. To produce a comprehensive survey of genetic anomalies affecting chromosome 17 in breast cancer, we selected 30 of 51 breast cancer cell lines we had analyzed by classical CGH, on the basis of their patterns of gains and/or losses on chromosome 17. We also studied 22 primary tumors, of which 4 had previously been typed by classical CGH. CGH

[5] http://www.ncbi.nlm.nih.gov/genome/cyto/hbrc.shtml.
[6] http://www.sanger.ac.uk/HGP/methods/cytogenetics/DOPPCR.shtml.
[7] http://www.probes.com/resources/calc/basedyeratio.html.
[8] http://genome.ucsc.edu, June 2002 freeze.

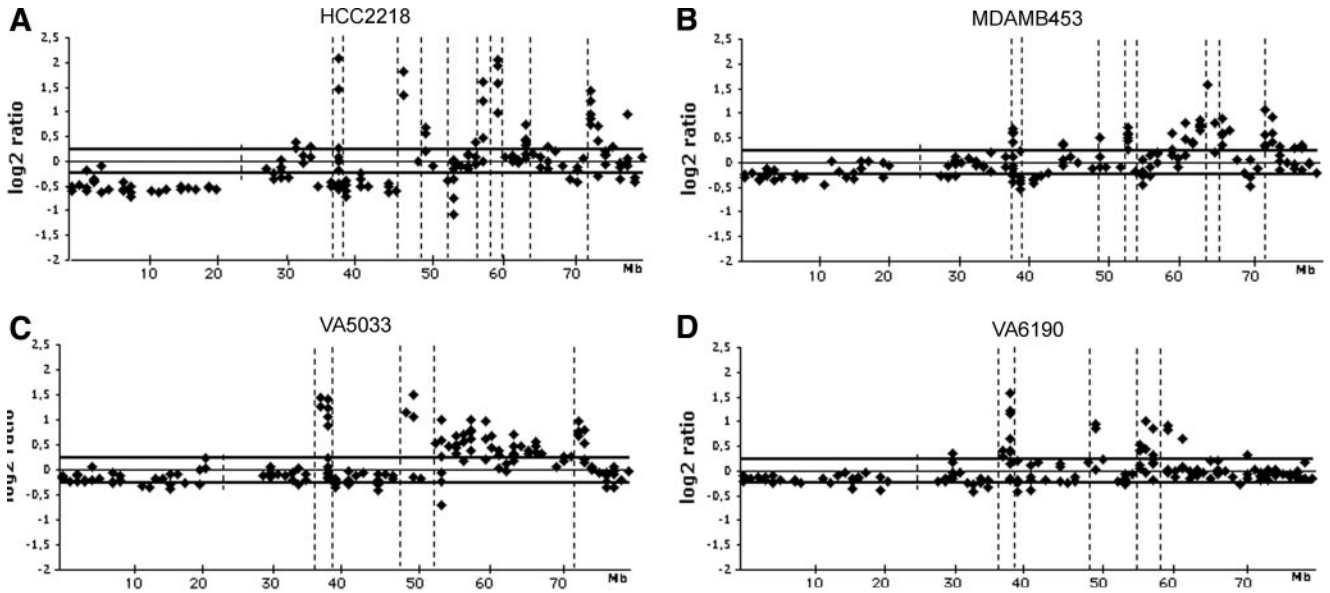

Fig. 1. Chromosome 17 high-resolution array-CGH profiles in breast cancer. Log 2 ratios were plotted according to the Mb positions of the clones on the University of California Southern California June 2002 freeze of the genome sequence (http://genome.ucsc.edu). *Horizontal bars* indicate log 2 ratio thresholds for gains (0.25) and losses (−0.25). Centromere position is depicted by a *dotted vertical line*. *Dotted vertical lines across the graphs* indicate sites of recurrent abrupt transitions flanking peaks of amplification or losses of elevated amplitude. These correspond to sites where log 2 ratio [clone(x) − clone(x +1)] = 2 SD.log 2 ratio(array). To be selected, these abrupt transitions had to occur in at least four different tumors or cell lines.

profiles showed that whereas chromosome 17p suffers only losses, eventually extending into the long arm, more complex combinations of gains and losses can affect 17q (Supplementary Figs. S3*A* and *B*). Another distinctive feature was the existence of eight transition sites bordering regions of either gains or losses, suggesting intense structural rearrangements. However, resolution of classical CGH was insufficient to draw firm conclusions.

To address this in greater detail, we built a genomic array covering chromosome 17 with 173 genomic clones (BAC, PAC, and cosmids). The average density was 1 target per 0.5 Mb. Coverage was not even throughout the chromosome, with a higher density at 17q12-q21 and 17q23-q25 and lower density on 17p with 1 target per 1 Mb (Supplementary Fig. S1). Clones selected on the array contained 191 genes identified according to the June 2002 human genome sequence freeze.[8] To determine the threshold for gains and losses and test for variability, four normal/normal hybridizations were performed, and SD was determined (Supplementary Fig. S2).

Array-CGH data of both cell lines and tumors showed complex profiles on chromosome 17 (complete dataset is in Supplementary Figs. S4 and S5), especially for the long arm, which showed combinations of gains and intervening losses (Fig. 1*A* and *B*). This elevated complexity prompted a two-level analysis of genomic profiles. First we wanted to define consensus regions that we would subsequently use as a basis for a comparison of genomic and expression profiles. Compilation of data from primary tumors and cell lines allowed us to define segments according to the main type of event observed (gain or loss). To do this, losses or gains were scored for each target clone along the chromosome, and ruptures in their frequency curve defined the boundaries of different segments (Fig. 2*A* and *B*). Thirteen segments were defined. These were distributed as four segments showing mainly losses (17p, 17q11.2, 17q21, and 17q24), six segments showing mainly gains (one at 17q12 and five in the 17q22-q25 interval), and three segments involved in either gains or losses (17q21.3, 17q22, and 17q25). Second, we searched for the smallest regions of overlap (SRO). To be considered, they had to occur in at least three tumors or cell lines. Accordingly, 18 SRO of gains and 16 SRO of losses were defined (Fig. 2*C*). Finally, we noted the existence of sharp transitions bordering gains or losses of elevated amplitude (Fig. 1). We identified 14 transition sites; interestingly, these tended to cluster within narrow intervals. One striking example is a transition downstream of *ERBB2-GRB7* observed 15 times within an interval of 0.2 Mb (Fig. 1; Supplementary Table S3).

Copy Number Changes *versus* RNA Expression. Having established the boundaries of the different segments, we sought to identify the genes showing expression changes in conjunction with copy number changes (CNCs). We produced a custom-made cDNA chip comprising 376 expressed sequence tags corresponding to 358 known genes on 17q. We compared array-CGH and expression array data in 18 primary tumors and 29 cell lines studied by both approaches. Primary tumors and cell lines were grouped according to their genomic status (gain, no CNC, or loss) in each of the 13 previously defined segments on chromosome 17 (Fig. 2*B*). Mean expression levels were calculated for each gene within each group (no CNC, gained, and lost) for each segment. Next, expression for each gene of the gained or lost group was normalized according to that of the no CNC group, and the expression difference d was calculated (Fig. 3).

We first searched for genes with modified expression in segments of gain. Overall, 85 genes showed significantly increased expression in conjunction with genomic gains (Supplementary Table S4). Of these 85 genes, 39 were located within the SRO of gains (Table 1). Genes retained in this restricted screen included a number of previously identified genes (*LASP1*, *RPL19*, *ERBB2*, *GRB7*, *HOXB7*, *NDP52*, *RPS6KB1*, *GRB2*, and *BIRC5*), as well as a number of novel candidates (*TOM1L1*, *COX11*, *ZNF161*, *FLJ20062*, *SMARCD2*, *LLGL2*, *SMT3H2*, *CDK3*, and *SECTM1*). In addition to overexpressed genes, we noted 32 genes showing reduced expression levels in segments of gains. This unexpected pattern will be worth exploring in order to determine whether this apparent paradox is related to any function detrimental to cancer growth. Remarkably, the retinoid receptor homologue *NR1D1* and *CBX1* (human orthologue of the chromo-domain protein HP1), which both act as transcriptional repressors, were members of this group (Supplementary Table S6).

Finally, we searched for genes with reduced expression in con-

junction with genomic losses and identified 67 genes (Supplementary Table S5). Remarkably, of these 67 genes, 19 had been previously selected as consistently overexpressed when gained. These included proven or strong candidate oncogenes, such as *MLLT6*, *GRB7*, or *TOP2A*, as well as novel candidates that we have identified (*TOM1L1* and *ZNF161*). These data suggested that expression levels of these genes are highly dependent on genomic dosage. Searching for genes located within minimal intervals of losses (Fig. 2*C*), we selected a subset of 32 genes (Table 2). Disregarding the 9 genes alternatively overexpressed when gained or underexpressed when lost, there remained 23 genes whose expression was consistently reduced as a consequence of a genomic loss. Whether this is related to the inactivation of a tumor suppressor gene remains to be determined.

DISCUSSION

Despite its relatively small size, chromosome 17 is a prevalent target of genetic anomalies in human cancer (3, 4, 15, 16). It harbors a number of *bona fide* cancer genes and contributes to a sizeable fraction of newly identified candidate cancer genes (5–9, 17, 18).

We present here a comprehensive study on copy number aberra-

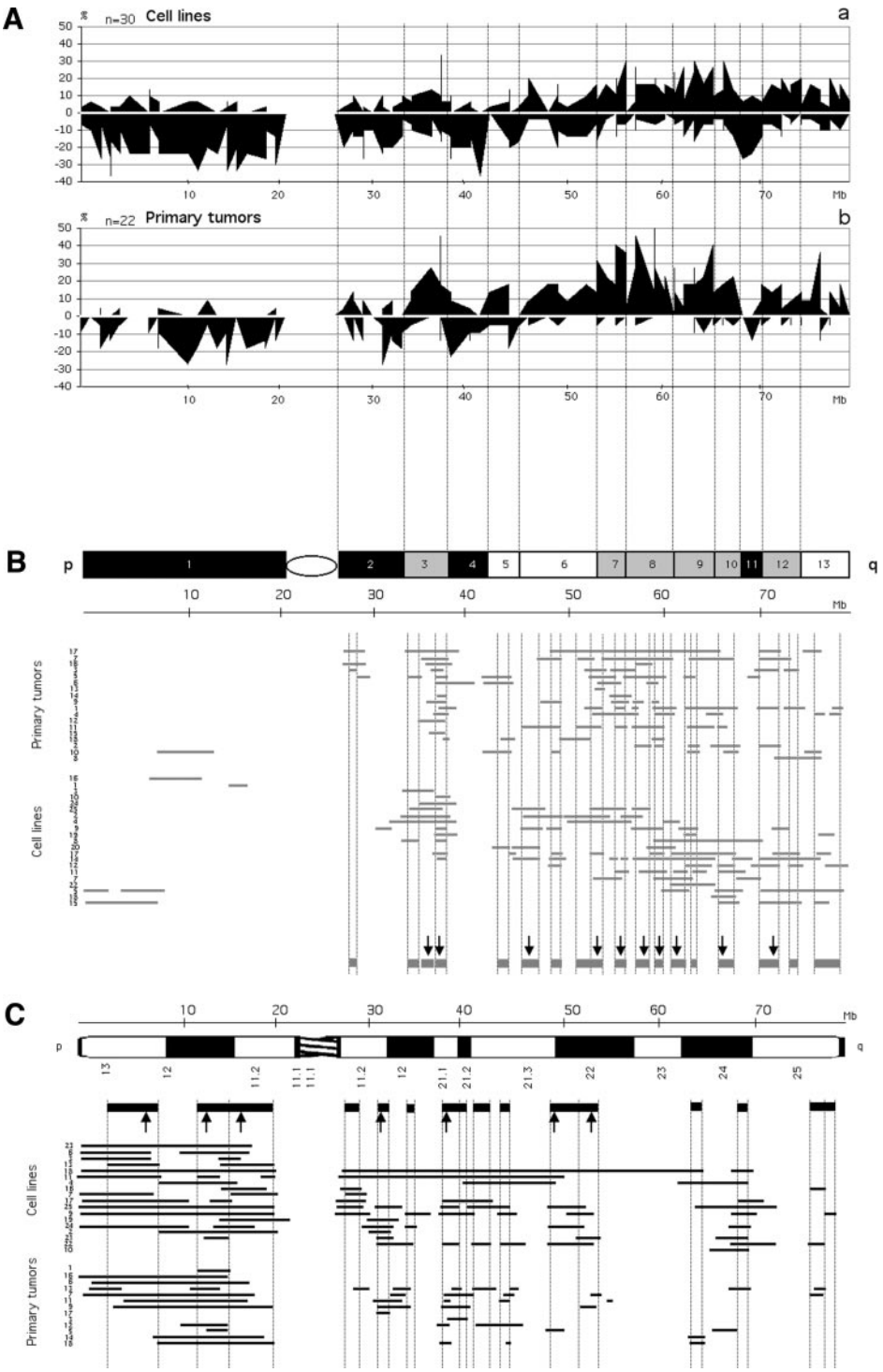

Fig. 2. Regions of recurrent gains and losses on chromosome 17 in breast cancer. *A*, frequencies of gains or losses along chromosome 17 in 30 cell lines (*a*) and 22 primary tumors (*b*). In each panel, the *top curve* indicates the frequency of gains (log 2 ratio $>$ 0.25), whereas the *bottom curve* shows the frequency of losses (log 2 ratio $<$ -0.25). Plots are shown with respect to the Mb positioning of the clones on the array, hence clones positioned close to each other may appear as merged. *B*, consensus segments of gains and losses on chromosome 17 defined according to frequencies of events. *Black segments* correspond to losses, *gray segments* correspond to gains, and *white segments* correspond to segments showing both events. *C*, definition of SRO and events of elevated amplitude. *Top*, regions of gains in each tumor or cell line are represented as *gray horizontal bars*. The smallest regions of gains, indicated as *bold gray bars* at the *bottom* of the graph, correspond to minimal overlap in at least three tumors or cell lines. *Bottom*, regions of losses are represented as *black bars*. SRO were defined as *gray horizontal bars*. *Black arrows* indicate SRO that could show events of elevated amplitude. Cell lines were as follows: *1*, BT20; *2*, BT474; *3*, BT483; *4*, EFM19; *5*, HCC1395; *6*, HCC1187; *7*, HCC1428; *8*, HCC1954; *9*, HCC2218; *10*, Hs578T; *11*, MCF7Rich; *12*, MDAMB157; *13*, MDAMB175; *14*, MDAMB361; *15*, MDAMB435; *16*, MDAMB436; *17*, MDAMB453; *18*, MDAMB468; *19*, SKBR3; *20*, SUM52; *21*, SUM149; *22*, SUM185; *23*, T47D; *24*, UACC812; *25*, ZR7530. Primary tumors were as follows: *1*, VA1593; *2*, VA4055; *3*, VA4380; *4*, VA4390; *5*, VA4435; *6*, VA4956; *7*, VA5033; *8*, VA4956; *9*, VA5450; *10*, VA6204; *11*, VA6219; *12*, VA6277; *13*, VA6582; *14*, VA6586; *15*, VA6660; *16*, VA7106; *17*, VA7079; *18*, VA7417.

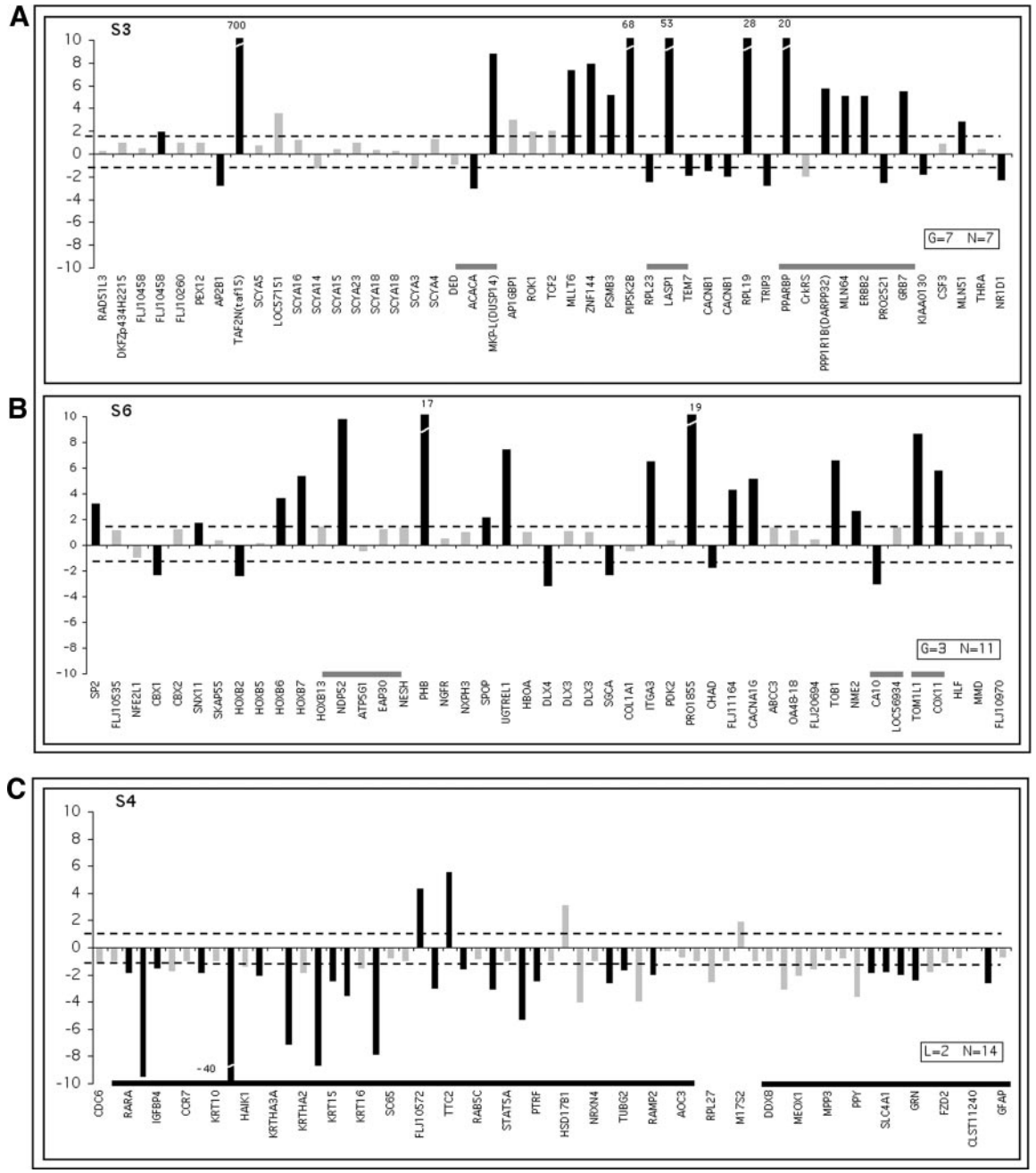

Fig. 3. Expression differences of genes in conjunction with CNC in consensus segments. The mean expression value of each gene was calculated in each genomic segment for gained, lost, or no CNC samples. Expression difference d was calculated as follows: $d = [|a| \times (b - a)]/a$, where a is the mean expression in no CNC samples, and b is the mean expression in gained or lost samples. *Graphs* represent levels of expression difference d for each gene in the corresponding segment. The threshold for significant expression difference was $d \geq 1.5$ for overexpression and $d \leq -1.5$ for reduced expression in at least 20% of the tumors or cell lines showing CNCs. Significant expression differences are represented as *black bars*. *Gray bars* correspond to nonsignificant differences. The number of samples used to calculate the mean expression value in a segment is indicated on each graph for normal and altered samples (*G*, gains; *L*, loss, *N*, no CNC). SRO are depicted as *gray* (gains) or *black* (losses) *horizontal bars* at the *bottom* of the graph. *A*, expression differences in conjunction with gains in segment 3. *B*, expression differences in conjunction with gains in segment 6. *C*, expression differences in conjunction with losses in segment 4.

tions on chromosome 17 and their consequences at the RNA expression level in breast cancer. At the genomic level, our data clearly showed that this chromosome was severely rearranged in breast cancer, and anomalies were found throughout the entire length of chromosome 17. We applied a two-level definition for regions of anomalies. Compiling data obtained on 22 tumors and 30 cell lines, we defined 13 consensus segments according to the main anomaly observed. We then searched for the SRO among the consensus regions limited by transition sites. These regions represented events that could occur independently and may thus be a more accurate representation of core events. In total, 17 SRO of gains and 16 SRO of losses were defined. Interestingly, 10 of 17 SRO of gains could be involved in high-level amplification, and 7 of 16 regions of losses showed events of high amplitude. This strongly suggested that these events resulted from a positive selection. Moreover, a number of these events of elevated amplitude were bordered by sharp transitions, and these breakpoints tended to cluster in narrow intervals (0.2–2 Mb). We identified 14 such sites, and array-CGH profiles suggested the occurrence of multiple breaks within a single tumor. This was supported by fluorescence *in situ* hybridization data showing multiple clusters of amplified chromosome 17 sequences dispersed at several chromosomal locations (11). These breakpoints could correspond to chromo-

somal fragile sites and play an active part in the occurrence of CNCs at 17q in breast tumors. Indeed, it is well established that unrepaired double-strand breaks are initiating events for DNA amplification (19). Sites of rupture related to regions of chromosomal fragility are apparently essential for DNA amplification to occur (20). It is noteworthy that the rupture site we mapped at 17q21.2 (38 Mb) colocalized with the t(15;17)(q22;q12-q21) translocation breakpoint cluster stereotypical of acute promyelocytic leukemia (21). It would be interesting to verify whether some of these rupture sites on 17q correspond to recurrent breast cancer-specific translocation breakpoints such as the recently characterized breakpoint at 8p12-p21 (22).

Copy number variations are expected to affect RNA expression levels. This is well accepted for DNA amplification, which was shown to arise as a selective mechanism for increased expression of one or more target genes (23). Some authors have proposed to extend this model to lower level CNCs, such as those resulting from aneuploidy (24). From the studies of Virtaneva *et al.* (25), a trisomy of chromosome 8 in acute myeloid leukemia apparently results in a global expression increase of genes on this chromosome, and other reports on tumors with more deeply affected karyotypes also suggest global modifications in expression concordant with chromosomal dosage (26). However, the selective advantage of such unstable events may be questionable because aneuploidy is a byproduct of mitotic instability in tumor cells (27) and is therefore prone to undergo rapid changes, as recently shown by us (13).

Overall, our data clearly indicate that CNCs, as gains or losses, are associated with important modifications in RNA expression levels. Five to fifty percent of genes in an amplified segment showed increased expression, whereas up to 30% of genes in a region of loss presented reduced RNA levels. A search for the most consistent expression changes led to the selection of 85 genes gained and overexpressed and 67 genes underexpressed in conjunction with a genomic loss. We observed 19 genes that showed both overexpression when gained and underexpression when lost. A number of these genes were either proven oncogenes or strong candidates. This finding emphasizes the strong influence of genomic dosage on expression levels. Transcription levels appeared to be almost mechanically adjusted according to copy numbers, and for such genes, DNA amplification could be the most efficient mechanism to select for increased RNA expression.

In contrast, 32 genes showed reduced expression in conjunction with genomic gains. This suggests a down-regulation of these genes when amplified, at variance with collinear genes, which were selected for increased expression. It will be interesting to see whether this effectively corresponds to transcriptional repression, thus favoring an interpretation that these genes could act as tumor suppressors.

Further work based on the analysis of a large set of breast tumors will be needed to validate the relative significance of the different candidates and, eventually, to evaluate their interplay. In this respect, we were motivated to determine whether genes mapping in different amplification cores presented coordinated expression profiles, thus suggesting coselection processes. Therefore, we analyzed expression profiling data by hierarchical clustering and searched for groups of recurrently coclustering genes. Three clusters grouping genes located in different regions of gains at 17q were identified. Cluster 2, for example, grouped genes at 17q12 (*PSMD11*, *PSMB3*, *RPL19*, and *TAF2N*), 17q23 (*TOM1L1*), and 17q25 (*SECTM1* and *TBCD*).

The diversity of functions among these genes was striking and

Table 1 *Thirty-nine overexpressed genes mapping in the smallest regions of gains*

Segment no.	Gene symbol	Location (Mb)	Cytoband	Gene name
3	*LASP1*	36529854–36581558	17q12	LIM and SH3 protein 1
3	*RPL19*	36860105–36864515	17q12	Ribosomal protein L19
3	*PPARBP*	37066388–37111066	17q12	PPAR-binding protein
3	*PPP1R1B (DARPP32)*	37283120–37289786	17q12	Protein phosphatase 1, regulatory protein phosphatase 1, regulatory (inhibitor) subunit 1B
3	*MLN64 (STARD3)*	37292846–37319201	17q12	START domain containing 3 (STARD3)
3	*ERBB2*	37355863–37384391	17q12	v-erb-b2 erythroblastic leukemia viral oncogene homologue 2, neuro
3	*GRB7*	37569442–37578773	17q12	Growth factor receptor-bound protein 7
5	*GOSR2*	44498391–44516025	17q21.32	Golgi SNAP receptor complex member 2
6	*HOXB7*	46395738–46399523	17q21.32	Homeo box B7
6	*NDP52*	46559566–46652458	17q21.32	Nuclear domain protein 52
6	*TOM1L1*	52708811–52770009	17q22	Target of myb1-like 1 (chicken)
6	*COX11*	52759961–52776713	17q22	COX11 homologue, cytochrome *c* oxidase assembly protein (yeast)
7	*ZNF161*	55796051–55810451	17q23.2	Zinc finger protein 161
7	*SFRS1*	55825798–55829476	17q23.2	Splicing factor, arginine/serine-rich 1
7	*FLJ20315*	56175262–56237324	17q23.2	Hypothetical protein FLJ20315
7	*RAD51C*	56517886–56559615	17q23.2	RAD51 homologue C (*Saccharomyces cerevisiae*)
8	*RPS6KB1*	57819125–57873488	17q23.2	Ribosomal protein S6 kinase, 70 kDa, polypeptide 1
8	*FLJ20062*	61547301–61553465	17q23.3	FTSJ3 FtsJ homologue 3 (*Escherichia coli*)
8	*SMARCD2*	61560621–61570860	17q23.3	SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily d, member 2
9	*DKFZP586L0724*	62575195–62601293	17q23.3	DKFZP586L0724 protein
9	*KIAA0054 (HELZ)*	63140494–63341040	17q24.1	Helicase with zinc finger domain
9	*CACNG4*	63387986–63454096	17q24.1	Calcium channel, voltage-dependent, γ subunit 4
9	*PRKAR1A*	66304550–66324342	17q24.2	Protein kinase, cAMP-dependent, regulatory, type I, α (tissue-specific extinguisher 1)
12	*FLJ20721*	71117987–71318577	17q25.1	Hypothetical protein FLJ20721
12	*KIAA0176*	73068800–73087366	17q25.1	KCTD2: potassium channel tetramerization domain containing 2
12	*SMT3H2*	73161097–73176005	17q25.1	SMT3 suppressor of mif two 3 homologue 2
12	*AD023*	73259304–73264299	17q25.1	AD023 protein
12	*GRB2*	73313175–73398819	17q25.1	Growth factor receptor-bound protein 2
12	*LLGL2*	73523022–73542161	17q25.1	Lethal giant larvae homologue 2 (*Drosophila*)
12	*WBP2*	73818122–73827758	17q25.1	WW domain-binding protein 2
12	*SRP68*	73853819–73888378	17q25.1	Signal recognition particle 68 kDa
12	*CDK3*	73921942–73926181	17q25.1	Cyclin-dependent kinase 3
13	*SYNGR2*	75988409–75992834	17q25.3	Synaptogyrin 2
13	*BIRC5*	76124977–76135412	17q25.3	Baculoviral IAP repeat-containing 5 (survivin)
13	*TIMP2*	76623255–76641648	17q25.3	Tissue inhibitor of metalloproteinase 2
13	*FLJ20748*	77529010–77533990	17q25.3	Hypothetical protein FLJ20748
13	*GAA*	77620684–77638860	17q25.3	Lucosidase, α; acid (Pompe disease, glycogen storage disease type II)
13	*CD7*	79074198–79076869	17q25.3	CD7 antigen (p41)
13	*SECTM1*	79165545–79178439	17q25.3	Secreted and transmembrane 1

Abbreviation: IAP, inhibitor of apoptosis protein.

Table 2 *Thirty-two genes with reduced RNA expression levels mapping in the smallest regions of losses.*

Segment no.	Gene symbol	Location (Mb)	Cytoband	Gene name
2	*FLJ10120*	28971716–29004074	17q11.2	Hypothetical protein FLJ10120
2	*HCA66*	29084568–29122520	17q11.2	Hepatocellular carcinoma-associated antigen 66
2	*CREME9*	29215318–29840209	17q11	Cytokine receptor-like factor 3 (CRLF3)
2	*NME1*	31298019–31312916	17q11.2	Nonmetastatic cells 1, protein (NM23A) expressed in
2	*SCYA7*	32366930–32368947	17q11.2	Chemokine (C-C motif) ligand 7
4	*TOP2A**	38045208–38073667	17q21.2	Topoisomerase (DNA) II α
4	*IGFBP4*	38099553–38113665	17q21.2	Insulin-like growth factor-binding protein 4
4	*SMARCE1**	38285228–38304358	17q21.2	SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily e, member 1
4	*KRT20**	38532465–38541442	17q21.2	Keratin 20
4	*KRT14*	38859634–38864236	17q21.2	Keratin 14 (epidermolysis bullosa simplex, Dowling-Meara, Koebner)
4	*KRTHA1**	39134483–39138352	17q21.2	Keratin, hair, acidic, 1
4	*KRT13**	39241737–39246354	17q21.2	Keratin 13
4	*KRT19*	39264373–39269057	17q21.2	Keratin 19
4	*KRT15**	39254502–39259645	17q21.2	Keratin 15
4	*JUP*	39465596–39497955	17q21.2	Junction plakoglobin
4	*ACLY*	39579631–39631756	17q21.2	ATP citrate lyase
4	*GCN5L2*	39819036–39827252	17q21.2	GCN5 general control of amino-acid synthesis 5-like 2 (yeast)
4	*STAT5B**	39960502–40026777	17q21.2	Signal transducer and activator of transcription 5B
4	*STAT3*	40032144–40105859	17q21.2	Signal transducer and activator of transcription 3 (acute-phase response factor)
4	*PTRF*	40194164–40197002	17q21.2	Polymerase I and transcript release factor
4	*TUBG1**	40369221–40374780	17q21.2	Tubulin, γ 1
4	*TUBG2*	40390227–40397943	17q21.2	Tubulin, γ 2
4	*RAMP2*	40484705–40486595	17q21.2	Receptor (calcitonin) activity-modifying protein 2
4	*UBTF*	42026411–42038852	17q21.31	Upstream binding transcription factor, RNA polymerase I
4	*SLC4A1*	42068967–42087395	17q21.31	Solute carrier family 4, anion exchanger, member 1
4	*RPIP8*	42127708–42137966	17q21.31	RAP2-interacting protein 8
4	*GRN*	42164416–42172398	17q21.31	Granulin
4	*U5–116KD*	42446478–42495441	17q21.31	*Homo sapiens* U5 snRNP-specific protein, 116 kDa
5	*NMT1*	42667176–42714604	17q21.31	*N*-Myristoyltransferase 1
5	*C17orf1B*	42849376–42854356	17q21.31	Chromosome 17 open reading frame 1 (FMNL = formin-like)
6	*CA10*	49420806–49950501	17q21.33	Carbonic anhydrase X
6	*TOM1L1**	52708811–52770009	17q22	Target of myb1-like 1 (chicken)

* Genes that were also selected as overexpressed when gained.

covered almost every area of cell physiology and metabolism, including transcription (*ZNF161* and *SMARCD2*), DNA replication (*CDC6*), recombination (*RAD51* and *TOP2A*), chromatin remodeling (*CBX1* and *HBOA*), protein catabolism (*PSMB1* and *SMT3H2*), vesicular trafficking (*TOM1L1*), RNA translation (*RPL19* and *RPS6KB1*), and respiratory chain (*COX11*). *COX11* encodes for an enzyme located at the mitochondrial inner membrane (28), and its amplification/overexpression could be related to the selection of *PHB* in our list of 85 amplified/overexpressed genes. Indeed, *PHB* codes for prohibitin and was originally proposed as a tumor suppressor. However, its role is unclear because it is presented either as a nuclear protein interacting with pRB (29) or as a chaperone stabilizing respiratory complexes at the mitochondrial inner membrane. Interestingly, *PHB* is up-regulated in case of mitochondrial stress (30).

Chromosome 17 is commonly and intensely rearranged in a number of human malignancies. Our work and published data show that a large number of genes can be involved. The prevalent involvement of chromosome 17 in cancer is puzzling and suggests that it harbors genes instrumental to the cancer process. Additionally, the presence of a number of chromosomal fragility sites could be a synergistic element. Chromosomal breaks will favor DNA copy number aberrations and modify expression profiles. This will in turn result in accelerated cell proliferation and bypass of cell cycle checkpoints, which will eventually end up in additional genetic aberrations. Similarly, it can easily be envisioned that deregulated expression of genes such as *RAD51*, which is instrumental for homologous recombination-mediated DNA repair, or *HBOA*, which affects chromatin conformation, will have profound consequences on genomic integrity and thus worsen the cancer phenotype.

ACKNOWLEDGMENTS

We thank Prof. Philippe Jeanteur for proofreading the manuscript, Prof. Jean-Bernard Dubois for constant support, and Annick Causse for excellent technical support.

REFERENCES

1. Mitelman F, Mertens F, Johansson B. A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. Nat Genet 1997;15:417–74.
2. Struski S, Doco-Fenzy M, Cornillet-Lefebvre P. Compilation of published comparative genomic hybridization studies. Cancer Genet Cytogenet 2002;135:63–90.
3. Courjal F, Theillet C. Comparative genomic hybridization analysis of breast tumors with predetermined profiles of DNA amplification. Cancer Res 1997;57:4368–77.
4. Forozan F, Mahlamaki EH, Monni O, et al. Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data. Cancer Res 2000;60:4519–25.
5. Monni O, Barlund M, Mousses S, et al. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. Proc Natl Acad Sci USA 2001;98:5711–6.
6. Kauraniemi P, Barlund M, Monni O, Kallioniemi A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. Cancer Res 2001;61:8235–40.
7. Clark J, Edwards S, John M, et al. Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. Genes Chromosomes Cancer 2002;34:104–14.
8. Hyman E, Kauraniemi P, Hautaniemi S, et al. Impact of DNA amplification on gene expression patterns in breast cancer. Cancer Res 2002;62:6240–5.
9. Pollack JR, Sorlie T, Perou CM, et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci USA 2002;99:12963–8.
10. Willis S, Hutchins AM, Hammet F, et al. Detailed gene copy number and RNA expression analysis of the 17q12–23 region in primary breast cancers. Genes Chromosomes Cancer 2003;36:382–92.
11. Orsetti B, Courjal F, Cuny M, Rodriguez C, Theillet C. 17q21–q25 aberrations in breast cancer: combined allelotyping and CGH analysis reveals 5 regions of allelic imbalance among which two correspond to DNA amplification. Oncogene 1999;18:6262–70.
12. 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.
13. Nugoli M, Chuchana P, Vendrell J, et al. Genetic variability in MCF-7 sublines: evidence of rapid genomic and RNA expression profile modifications. BMC Cancer 2003;3:13.
14. Cole TJ. Fitting smoothed curves to reference data. J R Stat Soc 1988;48:385–412.
15. Phelan CM, Borg A, Cuny M, et al. Consortium study on 1280 breast carcinomas: allelic loss on chromosome 17 targets subregions associated with family history and clinical parameters. Cancer Res 1998;58:1004–12.
16. Plummer SJ, Paris MJ, Myles J, et al. Four regions of allelic imbalance on 17q12-qter associated with high-grade breast tumors. Genes Chromosomes Cancer 1997;20:354–62.
17. Barlund M, Monni O, Kononen J, et al. Multiple genes at 17q23 undergo amplification and overexpression in breast cancer. Cancer Res 2000;60:5340–4.

18. Wu GJ, Sinclair CS, Paape J, et al. 17q23 amplifications in breast cancer involve the PAT1, RAD51C, PS6K, and SIGma1B genes. Cancer Res 2000;60:5371–5.
19. Paulson TG, Almasan A, Brody LL, Wahl GM. Gene amplification in a p53-deficient cell line requires cell cycle progression under conditions that generate DNA breakage. Mol Cell Biol 1998;18:3089–100.
20. Coquelle A, Toledo F, Stern S, Bieth A, Debatisse M. A new role for hypoxia in tumor progression: induction of fragile site triggering genomic rearrangements and formation of complex DMs and HSRs. Mol Cell1998;2:259–65.
21. Yoshida H, Naoe T, Fukutani H, et al. Analysis of the joining sequences of the t(15;17) translocation in human acute promyelocytic leukemia: sequence non-specific recombination between the PML and RARA genes within identical short stretches. Genes Chromosomes Cancer 1995;12:37–44.
22. Adelaide J, Huang HE, Murati A, et al. A recurrent chromosome translocation breakpoint in breast and pancreatic cancer cell lines targets the neuregulin/NRG1 gene. Genes Chromosomes Cancer 2003;37:333–45.
23. Stark GR. DNA amplification in drug resistant cells and in tumours. Cancer Surv 1986;5:1–23.
24. Li R, Sonik A, Stindl R, Rasnick D, Duesberg P. Aneuploidy vs. gene mutation hypothesis of cancer: recent study claims mutation but is found to support aneuploidy. Proc Natl Acad Sci USA 2000;97:3236–41.
25. Virtaneva K, Wright FA, Tanner SM, et al. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. Proc Natl Acad Sci USA 2001;98:1124–9.
26. Phillips JL, Hayward SW, Wang Y, et al. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. Cancer Res 2001;61:8143–9.
27. Borel F, Lohez OD, Lacroix FB, Margolis RL. Multiple centrosomes arise from tetraploidy checkpoint failure and mitotic centrosome clusters in p53 and RB pocket protein-compromised cells. Proc Natl Acad Sci USA 2002;99:9819–24.
28. Petruzzella V, Tiranti V, Fernandez P, et al. Identification and characterization of human cDNAs specific to BCS1, PET112, SCO1, COX15, and COX11, five genes involved in the formation and function of the mitochondrial respiratory chain. Genomics 1998;54:494–504.
29. Wang S, Fusaro G, Padmanabhan J, Chellappan SP. Prohibitin co-localizes with Rb in the nucleus and recruits N-CoR and HDAC1 for transcriptional repression. Oncogene 2002;21:8388–96.
30. Coates PJ, Nenutil R, McGregor A, et al. Mammalian prohibitin proteins respond to mitochondrial stress and decrease during cellular senescence. Exp Cell Res 2001; 265:262–73.

Genomic and Expression Profiling of Chromosome 17 in Breast Cancer Reveals Complex Patterns of Alterations and Novel Candidate Genes

Béatrice Orsetti, Mélanie Nugoli, Nathalie Cervera, et al.

Cancer Res 2004;64:6453-6460.