CAN-04-0756, Version 2

Genomic and expression profiling of chromosome 17 in breast cancer reveal 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], Charles Theillet[1]#

1. Génotypes et Phénotypes Tumoraux, EMI229 INSERM/Université Montpellier I, CRLC Val D'Aurelle-Paul Lamarque, 34298 Montpellier
2. ERM 206 INSERM/Université Aix-Marseille2, Parc Scientifique de Luminy 13288 Marseille cedex 09,
3. IGBMC, U 596 INSERM/Université Louis Pasteur, Parc d'Innovation, 1 rue Laurent Fries, BP 10142, 67404 Illkirch cedex, France

Running title: profiling of chromosome 17 alterations in breast cancer

Keywords: array-CGH, amplicon, oncogene, profiling

Abbreviations : CGH ; comparative genomic hybridization. CCGH ; chromosome CGH. CNC ; copy number change. LOH ; loss of heterozygosity. TSG ; tumor suppressor gene. cDNA ; complementary DNA. FISH ; fluorescent in situ hybridization. ER ; estrogen receptor. PR ; progesterone receptor.

#corresponding author : Charles Theillet, EMI 229 INSERM, Centre de Recherche, CRLC Val d'Aurelle 34298 Montpellier cedex 5, France

Tel : 33 (0)467 613 766

Fax : 33 (0)467 613 041

Email :theillet@valdorel.fnclcc.fr

Abstract

Chromosome 17 is severely rearranged in breast cancer. While 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-CGH 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/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 in 13 consensus segments, distributing in 4 regions showing mainly losses, 6 gains and 3 either gains or losses. Within these segments smallest regions of overlap (SRO) 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 underexpressed when lost (of which 32 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 chromosome. It is frequently rearranged in human tumors and presents a number of rearrangement breakpoints mapping to either its short or long arm (1). Furthermore, CGH studies have shown it to harbor multiple regions of gains or losses in a variety of human cancers (2).

CGH, LOH and molecular genetics data altogether 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 LOH mapping data pointed to the existence at 17q of at least 5 regions of imbalance (of which two corresponded to DNA amplification) (11). This likely to be a minimal estimate when taking into account similar data from the literature. This view was reinforced by FISH studies performed in our laboratory (B.O. unpublished) 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 genes involved and wanted to verify for the existence of recurrent sites of rearrangements on chromosome 17. We built a genomic array covering chromosome 17 at a mean density of one 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 EST 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 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.

Material and methods

Cell lines and tumors

Breast cancer cell lines used in this study included BRCAMZ01, BRCAMZ02, MDAMB175, MDAMB453 (D. Birnbaum, Inserm U119, Marseille, France), CAL51PE, MDAMB435, SKBR7, ZR7530 (P. Edwards, Department of Pathology, Cambridge, UK), BT474, MCF7Rich (F. Vignon, Inserm U540, Montpellier, France), HS578T, MDAMB436, HBL100 (A. Puisieux, Inserm U590, Lyon, France), SUM149, SUM185, SUM52 (S. Ethier, University of Michigan, Ann Arbor, MI, USA), EFM19, COLO824, EFM19, EFM192A (DSMZ, Braunschweig - Germany), BT20, BT483, CAMA1, HCC38, HCC1187, HCC1395, HCC1428, HCC1569, HCC1806, HCC1937, HCC1954, HCC2218, MCF7, MCF10F, MDAMB134, MDAMB157, MDAMB231, MDAMB330, MDAMB361, MDAMB415, MDAMB468, SKBR3, T47D, UACC812, ZR751 (ATCC, American Type Culture Collection, Manassas, VA, USA). All cell lines were maintained in DMEM or RPMI media containing 10% FBS supplemented with L-Glutamine (200 mM, 100X) and Antibiotic-Antimycotic (100X) (GibcoBRL, Life Technologies, 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 histological subtypes. The Scarff and Bloom grade distribution was 3.6 % of grade 1, 34.5% grade 2, 50.9% grade 3 and 10.9 % non determined, 75% ER+ and 67% PR+.

Classical CGH

Normal metaphase chromosomes were prepared from umbilical cord blood according to standard cytogenetic protocols. Hybridizations were done on Vysis (Downers Grove, IL, USA) normal human metaphases. Genomic DNA labeling and CGH reaction were performed as described in Courjal and Theillet (1997). CGH images were captured on a Zeiss (Le Pecq, France) epifluorescence microscope equipped with a JAI (Glostrup, Denmark) CCD camera run by the Metasystems (Altlussheim, Germany) image analysis software. CGH analysis was done using the ISIS 4.4 software (Metasystems).

Genomic arrays

The chromosome 17 genomic array is consisted of 107 RPCI-BAC and PAC clones from the set of cytogenetically mapped clones reported previously http://www.ncbi.nlm.nih.gov/genome/cyto/hbrc.shtml, 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, California, USA). Nine clones CTD-2251J22, RP11-455O6, RP11-300G13, RP11-319A23, RP11-379P18, RP11-387C17, RP11-399J11, RP11-469C13, RP11-489G5 were obtained from Research Genetics (Huntsville, AL, USA). Clones corresponding to genetic markers were isolated from the Down to The Well human BAC library of GenomeSystems Inc. (St Louis, Missouri, USA). 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 UK HGMP Resource Centre, Cambridge, UK. Cosmid clones Neu1 and Neu 4, P1 clone 610 were provided by Dr. Kallioniemi (Bethesda, Maryland). Clone P1.9 was from Dr. Viskochil (Salt Lake City, Utah, USA). See list of clones in Table S1 (supplementary data).

Array-CGH conditions

We isolated BAC, PAC and Cosmid DNA using Nucleobond® BAC100 from Macherey-Nagel (Hoerdt, France). We carried out DOP-PCR amplification on 10 ng of prepared DNA in a final reaction volume of 100 µl. Primer sequences and

DOP-PCR protocol used are available on the Sanger Center web site (http://www.sanger.ac.uk/HGP/methods/cytogenetics/DOPPCR.shtml) (12). We performed it with slight modifications : second round DOP-PCR primer was not aminolinked in our experiments. Purification of PCR products was done using Nucleofast® 96 PCR plates (Macherey-Nagel, Hoerdt, France). Purified PCR products were resuspended in dd H2O at 2µg/µl. An aliquot was run on an agarose gel in order 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, Orsay, France).

Hybridization to microarrays, Image and data analysis.

Genomic DNA was digested by NdeII according to the supplier's recommendations (Roche Diagnostics, Meylan, France). 300 ng of digested genomic DNA was labelled by random-priming in a 50µl reaction containing: 0.02mM dATP, 0.02mM dGTP, 0.02mM dTTP; 0.05mM dCTP; 0.04mM Cy3-dCTP or Cy5-dCTP; 25 Units of Klenow Fragment (50U/µl, New England Biolabs, Ozyme, Saint Quentin Yvelines, France), 10 mM □-mercaptoethanol, 5mM MgCl2, 50mM Tris-HCL pH 6.8 and 300µg/ml random octamers. The reaction was incubated at 37°C for 20 hours and stopped by adding 2.5µl EDTA 0.5M pH8. 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 is checked using a spectrophotometer and incorporation of dyes is calculated using Molecular Probes software (http://www.probes.com/resources/calc/basedyeratio.html). A mix of 700 pmol Cy5 and 700 pmol Cy3 labeled probes was ethanol precipitated in the presence of 250-300µg of human Cot-1 DNA (Roche Diagnostics, Meylan, France) and 100µg herring sperm DNA (Promega, Charbonnières, France). The pellet was dried and re-suspended in 280µl Hybrisol VII (Appligene Oncor, Qbiogen, Illkirch, France). The probes were denatured at 80°C for 10 min, and repetitive sequences were blocked by pre-annealing at 37°C for 90 min. Slides were blocked for 20 min at 42°C in saturation buffer (1% BSA, 0.2% SDS, 5X SSC), washed in 2X SSC, 0.2% SDS then in 2XSSC and dehydrated in ethanol series. A 8.8 cm^2 open hybridization chamber (Gene Frame □, Abgene, Courtaboeuf,

France) was fixed on the slide and the 280µl pre-annealed 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 2X SSC, 0.1% SDS pH7 at 55°C for 5min, and in 1X SSC, 0.1% SDS pH7 at 55°C for 5 min, followed by 3 times in 0.1X SSC 30 s at RT and briefly rinsed in water. Slides were dried by spinning for 5min at 1000 rpm and stored at RT until scanned. Arrays were scanned by a GenIII Array Scanner (Amersham Biosciences, Orsay, France). Images were analyzed by ARRAY-VISION 6.0 software (Amersham Biosciences, Orsay, France). 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, 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 less than 15% of deviation from the median were kept), representation of profiles with log2 ratios in Y-axis and Mb position of clones (http://genome.ucsc.edu, June 2002 freeze) along the chromosome in 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.

cDNA arrays construction and analysis

Preparation and hybridization of cDNA arrays were as described (13). Of the 720 cDNAs spotted, 376 corresponded to 358 known genes positioned on chromosome 17 (http://genome.ucsc.edu, June 2002 freeze) (Table S2, supplementary data). Hybridization signals were quantified using the HDG Analyzer software (Genomic Solutions, Ann Arbo, MI, USA) 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 non

significant variation, was defined. Its bandwidth was adjusted to fit the standard deviation 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 values < -1 (spots below the first diagonal) underexpressed.

Results

Genomic profiling of breast cancer cell lines and tumors

In order 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 (cCGH), 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 cCGH. CGH profiles showed that, while chromosome 17p suffers only losses, eventually extending into the long arm, more complex combinations of gains and losses can affect 17q (Figure S3A and S3B, supplementary data). Another distinctive feature was the existence of 8 transition sites bordering regions of either gains or losses, suggesting intense structural rearrangements. However, resolution of cCGH 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/0.5 Mb. Coverage was not even throughout the chromosome, with a higher density at 17q12-q21, 17q23-q25 and lower on 17p with 1 target/1Mb (Figure S1 supplementary data). Clones selected on the array contained 191 genes identified according to the june 2002 human genome sequence freeze (http://genome.ucsc.edu/). In order to determine the threshold for gains and losses and test for variability, 4 normal/normal hybridizations were performed and standard deviation (SD) determined (Figure S2, supplementary data).

Array-CGH data of both cell lines and tumors showed complex profiles on chromosome 17 (complete dataset in Figures S4 and S5, supplementary data). Especially for the long arm that showed combinations of gains and intervening

losses (Figure 1A and B). This elevated complexity prompted a two level analysis of genomic profiles. First we wanted to define consensus regions which 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 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 (Figure 2A and 2B). Thirteen segments were defined. These were distributed as 4 segments of losses (17p, 17q11.2, 17q21 and 17q24), 6 of gains (one at 17q12, 5 in the 17q22-q25 interval) and 3 involved in either gains or losses (17q21.3, 17q22 and 17q25). Second, we searched for the smallest regions of overlap (SROs). To be considered they had to occur in at least three tumors or cell lines. Accordingly, 18 SROs of gains and 16 of losses were defined (Figure 2C). Finally, we noted the existence of sharp transitions bordering gains or losses of elevated amplitude (Figure 1). We identified 14 transition sites, which interestingly 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 (Figure 1 and Table S3).

Copy number changes vs. RNA expression

Ahving established the boundaries of the different segments we sought to identify the genes showing expression changes in conjunction with copy number changes (CNC). We produced a custom made cDNA chip comprising 376 ESTs 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, loss) in each of the 13 previously defined segments on chromosome 17 (Figure 2B). Mean expression levels were calculated for each gene within each group (no CNC, gained, 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 (Figure 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 (Table S4 supplementary data). Of these 85 genes, 39 were located within the SROs 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, BIRC5*), as well as a number of novel candidates (*TOM1L1, COX11, ZNF161, FLJ20062, SMARCD2, LLGL2, SMT3H2, CDK3, SECTM1*). In addition to overexpressed genes, we noted 32 genes showing reduced expression levels in segments of gains. This pattern unexpected and it will be worthwhile exploring if this apparent paradox is related to any function detrimental to cancer growth. Remarkably, the retinoid receptor homolog *NR1D1* and *CBX1* (human ortholog of the chromo-domain protein HP1), which act both as transcriptional repressors, were members of this group (Table S6, supplementary data).

Finally, we searched for genes with reduced expression with conjunction with genomic losses and identified 67 genes (Table S5, supplementary data). 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, 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 (Figure 2C), 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 TSG 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 aberrations 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 smallest regions of overlap among the consensus regions limited by transition sites. These regions represented events that could occur independently and may be thus a more accurate representation of core events. In total, 17 SROs of gains and 16 SROs of losses were defined. Interestingly, 10/17 SROs of gains could be involved in high level amplification and 7/16 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 to 2 Mb). We identified 14 such sites and aCGH profiles suggested the occurrence of multiple breaks within a single tumor. This was supported by FISH data showing multiple clusters of amplified chromosome 17 sequences dispersed at several chromosomal locations (11). These breakpoints could correspond to chromosomal fragile sites and play an active part in the occurrence of copy number changes 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 like 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 which arises 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 copy number changes, such as those resulting from aneuploidy (24). From the studies of Virtaneva and colleagues, a trisomy of chromosome 8 in AML apparently results in a global expression increase of genes on this chromosome (25) 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, since

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 copy number changes, as gains or losses, are associated with important modifications in RNA expression levels. Five to 50% of genes comprised in an amplified segment showed increased expression, while 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 downregulation 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. This respect, we were motivated to determine whether genes mapping in different amplification cores presented coordinated expression profiles, thus suggesting co-selection processes. Therefore, we analyzed expression profiling data by hierarchical clustering and searched for groups of recurrently co-clustering 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, TAF2N*), 17q23 (*TOM1L1*) and 17q25 (*SECTM1, TBCD*).

The diversity of functions among these genes was striking. It covered almost every area of cell physiology and metabolism. These included transcription (*ZNF161, SMARCD2*), DNA replication (*CDC6*), recombination (*RAD51, TOP2A*), chromatin remodeling (*CBX1, HBOA*), protein catabolism (*PSMB1, SMT3H2*),

vesicular trafficking (*TOM1L1*), RNA translation (*RPL19, RPS6KB1*) or 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, as it is either presented as a nuclear protein interacting with pRB (29), or as a chaperone stabilizing respiratory complexes at the mitochondrial inner membrane. Interestingly, *PHB* is upregulated in case of mitochondrial stress (30).

Chromosome 17 is commonly and intensely rearranged in a number of human malignancies. Our work as well as 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 will affect chromatin conformation, will have profound consequences on genomic integrity and, thus, worsen the cancer phenotype.

Acknowledgements: This study was supported by funds from the CNRS, INSERM, the Association de Recherche sur le Cancer (ARC), grant 5102, the Ligue Nationale de Lutte Contre le Cancer (LNCC), 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.N. was supported by a doctoral fellowsh from the Ligue Nationale Contre le Cancer. Authors wish to thank Prof. Philippe Jeanteur for proofreading the manuscript and Prof. Jean-Bernard Dubois for constant support and Mrs Annick Causse for excellent technical support.

References

1. Mitelman F, Mertens F, and Johansson B A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nat Genet*1997;15 Spec No:417-74.
2. Struski S, Doco-Fenzy M, and Cornillet-Lefebvre P Compilation of published comparative genomic hybridization studies. *Cancer Genet Cytogenet*2002;135:63-90.
3. Courjal F and 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 U S A*2001;98:5711-6.
6. Kauraniemi P, Barlund M, Monni O, and 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 U S A*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, and 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. *Journal of the Royal Statistical Society*1988;A: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, and 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, and 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 Cell*1998;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 Sur*1986;5:1-23.
24. Li R, Sonik A, Stindl R, Rasnick D, and Duesberg P Aneuploidy vs. gene mutation hypothesis of cancer: recent study claims mutation but is found to support aneuploidy. *Proc Natl Acad Sci U S A*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 U S A*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, and 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 U S A*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, and 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.

Figure legends

Figure 1

Chromosome 17 high-resolution array-CGH profiles in breast cancer. Log2 ratios were plotted according to the Mb positions of the clones on the UCSC June 2002 freeze of the genome sequence (http://genome.ucsc.edu). Horizontal bars indicate log2 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 log2.ratio[clone(x)−clone(x+1)]=2 SD.log2.ratio(array). To be selected these abrupt transitions had to occur at least in 4 different tumors or cell lines.

Figure 2

Regions of recurrent gains and losses on chromosome 17 in breast cancer. Panel 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 (log2.ratio >0.25), while the bottom curve shows the frequency of losses (log2.ratio <-0.25). Plots are with respect to the Mb positioning of the clones on the array. Hence clones positioned close to each other may appear as merged. Panel B: Consensus segments of gains and losses on chromosome 17 defined according to frequencies of events. Black segments correspond to losses, grey to gains and white to segments showing both events. Panel C: definition of smallest regions of overlap and events of elevated amplitude. Top of the ideogram: regions of gains in each tumor or cell line are represented as grey horizontal bars. The smallest regions of gains, indicated as bold grey bars at the bottom of the graph, correspond to minimal overlap in at least three tumors or cell lines. Bottom of the ideogram: regions of losses are represented as black bars. Smallest regions of overlap were defined as for gains. Black arrows indicate SROs that could show events of elevated amplitude. Cell lines 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 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.

Figure 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|x(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 differences d for each gene in the corresponding segment. The threshold for significant expression difference was $d \geq 1.5$ for overexpression and ≤ -1.5 for reduced expression in at least 20% of the tumors or cell lines showing copy number changes. Significant expression differences are represented as black bars. Grey bars correspond to non-significant 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). SROs are depicted as grey (gains) or black (losses) horizontal bars at the bottom of the graph. Panel **A**: expression differences in conjunction with gains in segment 3. Panel **B**: expression differences in conjunction with gains in segment 6. Panel **C**: expression differences in conjunction with losses in segment 4.

Legends to Tables

Table 1: 39 overexpressed genes mapping in the smallest regions of gains.

Table 2: 32 genes with reduced RNA expression levels mapping in the smallest regions of losses. Asterisk indicates genes which were also selected as overexpressed when gained.

Table 1 : 39 overexpressed genes mapping in minimal regions of gain

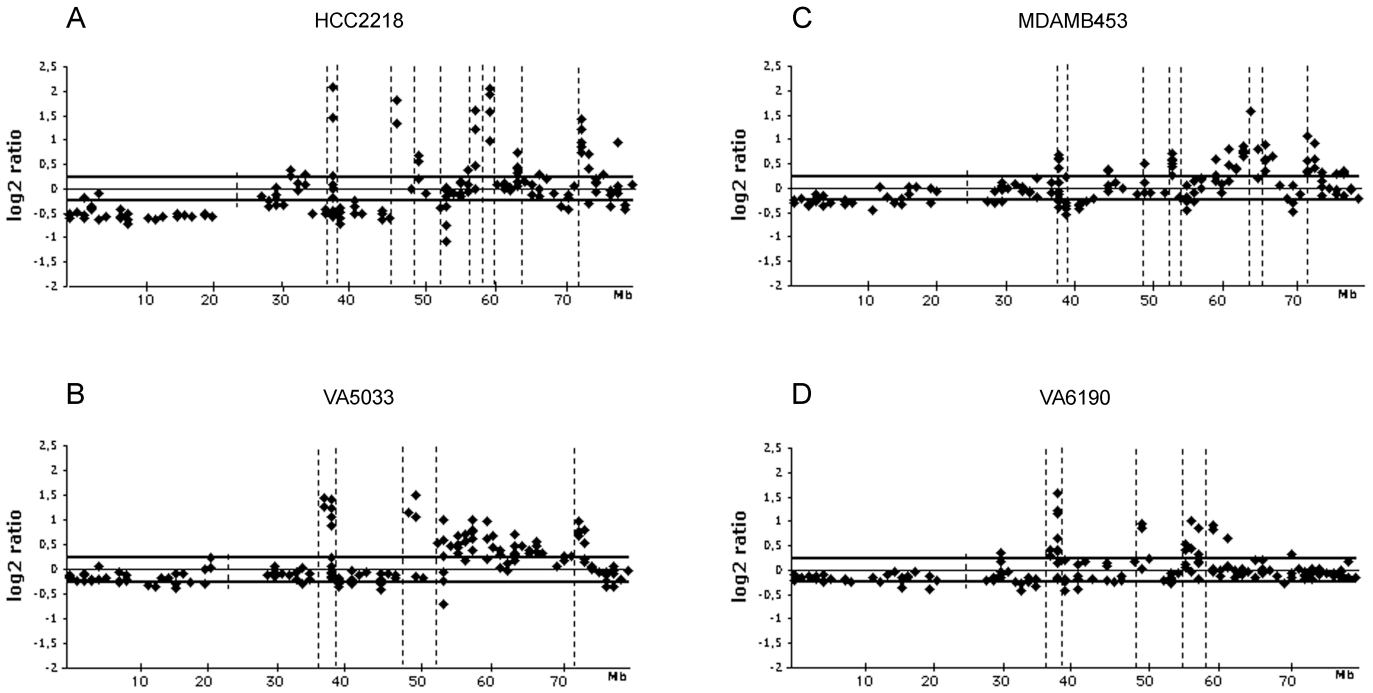
Segment number	Gene Symbol	Location (Mb)	Cytoband	Gene name
3	LASP1	36529854-36581558	17q12	LIM and SH3 protein 1
3	RPL19	36860105-36864515	17q12	ribosomal protein L19
3	PPARBP	37066388-37111066	17q12	PPAR binding protein
3	PPP1R1B (DARPP32)	37283120-37289786	17q12	protein phosphatase 1, regulatory rotein 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 homolog 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 homolog, 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 homolog C (S. cerevisiae)
8	RPS6KB1	57819125-57873488	17q23.2	ribosomal protein S6 kinase, 70kDa, polypeptide 1
8	FLJ20062	61547301-61553465	17q23.3	FTSJ3 FtsJ homolog 3 (E. coli)
8	SMARCD2	61560621-61570860	17q23.3	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, me
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, gamma subunit 4
9	PRKAR1A	66304550-66324342	17q24.2	protein kinase, cAMP-dependent, regulatory, type I, alpha (tissue specific extinguisher 1)
12	FLJ20721	71117987-71318577	17q25.1	hypothetical protein FLJ20721
12	KIAA0176	73068800-73087366	17q25.1	KCTD2: potassium channel tetramerisation domain containing 2
12	SMT3H2	73161097-73176005	17q25.1	SMT3 suppressor of mif two 3 homolog 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 homolog 2 (Drosophila)
12	WBP2	73818122-73827758	17q25.1	WW domain binding protein 2
12	SRP68	73853819-73888378	17q25.1	signal recognition particle 68kDa
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, alpha; 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

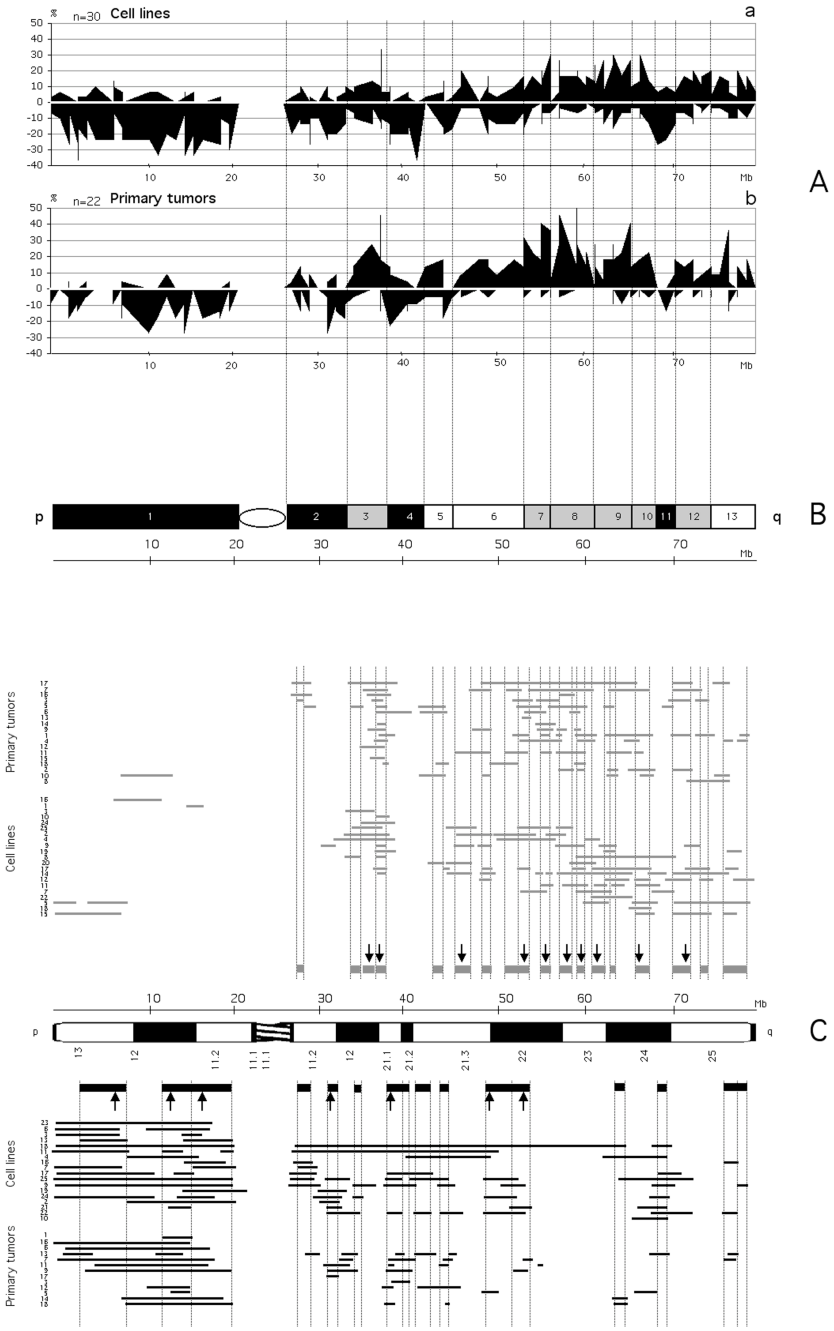
Table 2 : 32 genes with reduced RNA expression levels mapping in reduced intervals of losses. Asterisk indicates genes wihich were also selected as overexpressed when gained

Segment number		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	non-metastatic 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 alpha
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, gamma 1
4		TUBG2	40390227-40397943	17q21.2	tubulin, gamma 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 kD
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)

Orsetti et al.
Figure 1



Orsetti et al.
Figure 2



Orsetti et al.
Figure 3