Integrated Profiling of Basal and Luminal Breast Cancers

José Adélaïde,1 Pascal Finetti,1 Ismahane Bekhouche,1 Laetitia Repellini,1 Jeannine Geneix,1 Fabrice Sircoulomb,1 Emmanuelle Charafe-Jauffret,1,2L Nathalie Cervera,1 Jérôme Desplas,3 Daniel Parzy, Eric Schoenmakers,5 Patrice Viens,3,4 Jocelyn Jacquemier,1,2 Daniel Birnbaum, François Bertucci,1,4 and Max Chaffanet1

1Marseille Cancer Research Center, Institut Paoli-Calmettes and UMR599 Institut National de la Santé et de la Recherche Médicale, Department of Molecular Oncology, "Department of Biopathology, and "Department of Medical Oncology, Institut Paoli-Calmettes; 2Université de la Méditerranée; 3Institut de Médecine Tropicale du Service de Santé des Armées, Unité de Recherche en Pharmacologie et Physiopathologie Parasitaire, Marseille, France; and 4Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands

Abstract
Basal and luminal are two molecular subtypes of breast cancer with opposite histoclinical features. We report a combined, high-resolution analysis of genome copy number and gene expression in primary basal and luminal breast cancers. First, we identified and compared genomic alterations in 45 basal and 48 luminal tumors by using 244K oligonucleotide array comparative genomic hybridization (aCGH). We found various genome gains and losses and rare high-level gene amplifications that may provide therapeutic targets. We show that gain of 10p is a new alteration in basal breast cancer that and that a subregion of the 8p12 amplification is specific of luminal tumors. Rare high-level amplifications contained BCL2L2, CCNE, EGFR, FGFR2, IGF1R, NOTCH2, and PIK3CA. Potential gene breaks involved ETV6 and FLT3. Second, we analyzed both aCGH and gene expression profiles for 42 basal and 32 luminal breast cancers. The results support the existence of specific oncogenic pathways in basal and luminal breast cancers, involving several potential oncogenes and tumor suppressor genes (TSG). In basal tumors, 73 candidate oncogenes were identified in chromosome regions 1q21-23, 10p14, and 12p13 and 28 candidate TSG in regions 4q32-34 and 5q11-23. In luminal breast cancers, 33 potential oncogenes were identified in 1q21-23, 8p12-q21, 11q13, and 16p12-13 and 61 candidate TSG in 16q12-13, 16q22-24, and 17p13. HORMAD1 (P = 6.5 × 10−4) and ZNF703 (P = 7 × 10−4) were the most significant basal and luminal potential oncogenes, respectively. Finally, among 10 candidate oncogenes associated with basal subtype, we validated CDC123/C10orf7 protein as a basal marker. [Cancer Res 2007;67(24):11565–75]

Introduction
Breast cancer is a heterogeneous disease whose evolution is difficult to predict. Consequently, treatment is not as adapted as it should be. Breast cancer can be defined at the clinical, histologic, cellular, and molecular levels. Progress and efforts to integrate these definitions allow a better management of the disease (1). Five major breast cancer molecular subtypes (luminal A and B, basal, ERBB2 overexpressing, and normal-like) have been identified (2–5). They are associated with different prognosis. Luminal A breast cancers express hormone receptors, have an overall good prognosis, and can be treated by hormone therapy. ERBB2-overexpressing breast cancers have a poor prognosis and are treated by targeted therapy using trastuzumab or lapatinib (see ref. 6 for review). No specific therapy is available for the other subtypes, although the prognosis of basal and luminal B tumors is poor.

Progress can be made in several directions. First, by identifying, among good prognosis tumors such as luminal A breast cancers, the cases that will relapse and metastasize. Second, by developing new drugs that will allow a better management of poor-prognosis breast cancers, such as basal tumors. The identification of new markers and therapeutic targets is necessary to meet these objectives. This relies on the study of tumor samples at the genomic (7–14), gene expression (2, 4, 15, 16), or proteomic (17–19) levels.

We have established here the genomic and gene expression profiles of a series of breast cancer samples by using high-resolution array comparative genomic hybridization (aCGH) and DNA microarrays, respectively. We have focused our analysis on two breast cancer subtypes with overall opposite features: luminal and basal.

Materials and Methods
Breast cancer samples. Tumor tissues were collected from 202 patients with invasive adenocarcinoma who underwent initial surgery at the Institut Paoli-Calmettes (Marseille, France) between 1992 and 2004 (from a cohort of 1,185 patients with frozen tumor sample). Each patient gave written informed consent and the study was approved by our institutional review board. Samples were macrodissected and frozen in liquid nitrogen within 30 min of removal. Several sets of tumors were studied. Gene expression was analyzed on a set of 202 tumors. Gene expression and genome profilings were established on a subset of 74 tumors (2 basal and 32 luminal A; as defined in ref. 4) for which RNA and DNA samples were both available. Genome profiles were established on 93 tumors, including these 74 tumors and an additional set of 19 cases selected from immunohistochemical data on estrogen receptor (ER), progesterone receptor (PR), ERBB2, and other basal and luminal markers (20): 2 ER+/PR−/ERBB2− ("triple negative") and 17 ER+/PR+ (21) tumors associated with basal-like and luminal-like immunohistochemical phenotypes, respectively. Although a third of the luminal tumors were not profiled on expression microarrays, immunohistochemical-based phenotypes were established for all studied tumors: 98% and 88% of the luminal samples used in this study were ER and PR positive, respectively (see Supplementary Table S1). The set studied by aCGH thus comprised 44 basal and 49 luminal tumors (mainly luminal A tumors). All tumor sections were reviewed by two pathologists...
DNA and RNA extraction. DNA and RNA were extracted from frozen samples by using guanidine isothiocyanate and cesium chloride gradient as described previously (23). RNA integrity was controlled on Agilent Bioanalyzer (Agilent Technologies).

Gene expression profiling with DNA microarrays. Expression profiles were established for 202 breast cancers and 4 normal breast samples with Affymetrix U133 Plus 2.0 human oligonucleotide microarrays following a described protocol (24). Scanning was done with Affymetrix GeneArray scanner and quantification with Affymetrix GeneChip Operating Software. Data were analyzed by the robust multichip average method in R using Bioconductor and associated packages (25). Robust multichip average did background adjustment, quantile normalization, and summarization of 11 oligonucleotides per gene. Before analysis, a filtering process removed from the data set the genes with low and poorly measured expression as defined by an expression value inferior to 100 units in all samples. All data were then log2 transformed for display and analysis. Before combined analysis, gene expression levels for each tumor sample were centered by the average expression levels of the four normal breast samples. The up-regulation or down-regulation status for each probe set was assigned using a threshold of |1| corresponding to twice the expression found in the normal breast pool. For multiple probe sets mapping to the same gene, the probe sets with an extension "-at", next "s_at", and followed by all other extensions were preferentially kept. When several probe sets with the best extension were available, the one with the highest median value was retained. Probe sets were mapped to the human genome according to the hg17 Build 35 from the National Center for Biotechnology Information (NCBI) Ensembl database and the University of California at Santa Cruz Genome Bioinformatics database Genome Browser.

Array-comparative genomic hybridization. Genomic imbalances of 93 breast tumors were analyzed by aCGH using 244K CGH Microarrays (Hu-244A, Agilent Technologies) following a described protocol (26). Scanning was done with Agilent Autofocus Dynamic Scanner (G2565BA, Agilent Technologies). Data analysis was done as described (27) and visualized with CGH Analytics 3.4 software (Agilent Technologies). Extraction of data (log, ratio) was done from CGH Analytics, whereas normalized and filtered log2 ratio was obtained from "Feature Extraction" software (Agilent Technologies). Data generated by probes mapped to X and Y chromosomes were eliminated. The final data set contained 225,388 unique probes covering 22,509 genes and intergenic part following the hg17 human genome mapping. Copy number changes were characterized as reported previously (10, 28). aCGH data were analyzed using circular binary segmentation (29) with default variables to translate intensity measurements in regions of equal copy number as implemented in the DNA copy R/Bioconductor package (25). Thus, each probe was assigned a segment value referred to as its "smoothed" value. The gain and loss status for each probe was assigned using threshold of 0.5 on at least five consecutive probes across the genome. Data were submitted to the Cluster program (30) using Pearson correlation as similarity metric and centroid linkage clustering. Results were displayed using the TreeView program (30).7

Fluorescence in situ hybridization analysis. Fluorescence in situ hybridization (FISH) analysis on tissue microarrays using bacterial artificial chromosome (BAC) probes has been described previously (34).

Immunohistochemistry on breast cancer tissues. Anti-epidermal growth factor receptor (EGFR) antibody (clone 31G7, 1:120 dilution; Zymed Laboratories) and anti-CDC123 antibody (clone 4B9, 1:300 dilution; Abnova Corp.) were used for immunohistochemistry. A pepsin treatment was done on tissue sections for 30 min before immunohistochemistry with EGFR antibody. Immunohistochemistry was done on 5-mm sections using DAKO LSAB2 kit in a DAKO Autostainer (DakoCytomation). Antibodies were incubated for 1 h in citrate buffer. After staining, slides were evaluated by two pathologists. EGFR and CDC123 (also called C10orf7) showed membrane and cytoplasmic signal, respectively. Intensity of EGFR and CDC123 staining was scored from 0 (no staining) to 3 (strong and diffuse staining). The normality of these results was expressed by the percentage of positive cells (P, from 0% to 100%) and the intensity of the staining (I, from 0 to 3) and expressed by the Quick score (Q = P × I, from 0 to 300).

Statistical analyses. Differences in clinical and pathologic features between basal and luminal cases were assessed by Fisher's exact test for qualitative variables with discrete categories and Wilcoxon test for continuous variables.

Results

Identification of molecular subtypes of breast cancer by gene expression profiling. The five molecular subtypes of breast cancer have been identified using an intrinsic set of ~ 500 genes (2–4). We looked for these subtypes in our series of 202 breast cancers profiled with Affymetrix DNA microarrays by using the 476 probe sets common to the intrinsic 500-gene set and our 29,623 filtered probe sets. We first defined the expression centroid of each subtype for these 476 probe sets in the 122 samples from Sorlie et al. (4) by computing the median expression for each gene in the corresponding core samples. We then measured the correlation of each of our 202 samples with each centroid: 80 samples were close to the luminal A centroid, 2 to the luminal B centroid, 60 to the basal centroid, 27 to the ERBB2 centroid, and 19 to the normal-like centroid. Fourteen samples displayed a correlation inferior to 0.15 with any centroid and were not attributed to any subtype with this 476-gene set. Although basal and ERBB2 subtypes have been repeatedly recognized in independent data sets, identification of the other subtypes remains difficult (35); the intrinsic gene set is not optimal to classify all samples (2–4). In all sets, some samples remain unclassified because they have a low correlation with subtype centroids (4). Other studies have attempted to improve the stratification of breast cancer patients (see ref. 35 for details). In our study, we kept the standard definition given by the initial intrinsic gene of Sorlie et al. (4) to subtype tumors in the same

7 Data are available at http://yeti.marseille.inserm.fr/divers/Adelaide2007/Samples-LuminalBasalRData.
conditions as in recent integrated studies. Thus, a total of 142 samples were either basal or luminal in our series (data not shown).

**Genome analysis of basal and luminal breast tumors.** To identify genomic alterations in basal and luminal breast cancers, we used high-resolution aCGH. DNA was available for 42 basal and 32 luminal cases as defined above. We selected 19 additional cases (2 basal and 17 luminal) from immunohistochemical data. We thus analyzed gene copy number aberrations (CNA) in 93 primary breast tumors corresponding to 45 basal and 48 luminal tumors. CNAs are reported on frequency plots in Supplementary Fig. S1A to D. We distinguished low-level CNAs, which can derive from aneuploidy, from high-level CNAs, which result from focal amplification of chromosome regions, by using two different threshold values (log2 ratio >0.5 and [1]) in basal (Supplementary Fig. S1A and C) and luminal (Supplementary Fig. S1B and D) tumors.

In agreement with previous observations (9–11), we observed various and common alterations in both basal and luminal breast cancers. Basal tumors had higher frequencies (F) of copy number gains (with $F_{max} = 45\%$ for q22) compared with luminal tumors (with $F_{max} = 38\%$ for q13-q21). The highest CNA frequencies (>20%) were observed in 11 and 17 chromosomal regions in basal and luminal tumors, respectively (Supplementary Fig. S1A and B). In basal tumors, the highest CNA frequencies (>20%) were associated with copy number gains of 1q, 3q24-q27, 8p11-12, 8q21-ter, 10p, 12p13-ter, and 17q25-ter regions and with copy number losses of 4q13-14, 4q34-35, 5q, 8p21-ter, and 14q21 regions. In luminal tumors, frequent CNAs were associated with copy number gains of 1q, 8p12-ter, 11q13-14, 14q11, 16p13, and 17q23 regions and copy number losses of 1p36-ter, 2qter, 3p14, 4q34-35, 6q, 8p21-ter, 9qter, 16q, 17p13-ter, 19pter, and 22q.

Regional, high-level amplifications (with a threshold log2 ratio >1) were observed with differences in levels and in frequencies between the two molecular subtypes. High-level amplifications targeted chromosome regions 8p11, 8q11-ter, 10p13-ter, 15q26, 18q11, and 19q11-12 in basal tumors (Supplementary Fig. S1C) and 5q35, 8p11-12, 8q11-ter, 10q11, 11q13-14, and 17q12-13 in luminal tumors (Supplementary Fig. S1D). High-level amplification of 20q13 region was not observed. The exclusion of the ERBB2 subtype from this analysis explains the conspicuous absence of the 17q21 amplicon.

Thus, basal and luminal breast cancers show both common and specific alterations.

**Gene alterations differentially associated with basal and luminal breast cancer.** Unsupervised hierarchical clustering (Fig. 1) of the 93 samples and the 225,388 oligonucleotide probes (excluding X and Y probes) revealed significant differences in the spectrum of alterations in basal and luminal breast cancers. A total of 30,600 probes were differently represented in basal and luminal cases ($P < 0.05$, Fisher's exact test, FDR corrected). Specific regional aberrations (gains and losses defined with a threshold value of log2 ratio >0.5) are shown for basal (red) and luminal (blue) tumors in Supplementary Fig. S2. Genes with significant differences of CNA frequencies are listed in Supplementary Tables S2A and B and summarized in Supplementary Table S2C.

We distinguished two types of gain alterations: copy number gains and high-level amplifications.

Copy number gains were found in chromosomal regions 1q22-23.1, 3q23-27.1, 6p21, 10p11.21-15.3, 12p12.1-13.33, and 17q25.3 and regions 8p12-21.3, 11q12.2-13.3, and 16p12.2-12.1 in basal and luminal tumors, respectively. They reflected the copy number gains of 432 genes differentially represented in the two subtypes (305 and 127 genes in basal and luminal, respectively; Supplementary Tables S2A and C). Gains at 11q13-14 and 12p13 regions (Supplementary Figs. S3 and S4; Fig. 2), respectively associated with luminal and basal subtypes, agreed with reported FISH analyses (34), serial analysis of gene expression results (13), and other studies (36). In addition to confirming recently reported findings (9–11, 13, 34, 36) we found gain of 10p as a new, basal-specific alteration present in ~20% of basal breast cancers (Fig. 3). Gain frequencies of a large number of 10p genes ($n > 100$) correlated with basal subtype.

High-level amplifications (with a threshold log2 ratio >1) targeted 43 genes (Supplementary Table S2D), mainly correlated with luminal tumors. Strikingly, 40 of these were located at 8p11-12 (Supplementary Table S2C) and 11q13-14. The telomeric part of the 8p11-12 amplification [A1 amplicon and part of A2 amplicon (37)] was found specifically in luminal tumors (Supplementary Fig. S5). These results agree with previous data (10). Rare high-level amplifications targeting PIK3CA, IGFR1, and CCNE1 loci at 3q26, 15q26, and 19q11-12, respectively, were found in basal samples (Supplementary Figs. S6-S8A). Other sharp high-level amplifications targeted NOTCH2 (1p11), EGFR (7p11), FGFR2 (10q26), and BCL2L2 (14q11; Supplementary Fig. S8B-D).

Losses of regions 4p15.32-34.3, 5q11.1-23.2, 7p14.3-22.3, 10q23.31, 14q32.33, and and 20q13.33 and regions 1p36.11-36.33, 2p31.3, 2p31.21, 6p21.32-24.1, 9q34.3, 14q22.3, 16p13.3-23.4, 17p12.3-13.3, and 22q11.22-13.33 were observed in basal and luminal tumors, respectively. They reflected the copy number losses of 574 genes differentially represented in the two subtypes (168 in basal and 405 in luminal; Supplementary Tables S2B and C).

To characterize the presence of potential gene breakages, we identified genome alterations associated with copy number transition profiles within a chromosome as described previously (31). Copy number transitions are likely to reflect DNA strand breakages that may lead to nonreciprocal translocations. We identified transition profiles with a threshold of aberration score (log2 ratio >0.5) in 1,975 genes in at least one of the 93 tumors. A total of 501 genes showed copy number transitions in at least two different tumors (Supplementary Table S3A). Results were coherent with our previous analysis of NRG1 (38), FHIT (39), and PARK2 (40), the two latter being associated with fragile sites. Eight other large genes associated with fragile sites showed transitions: CSMD1 (5%), CSM3 (3%), NBEA (2%), NRXN3 (2%), CDH13 (2%), ACCCN1 (2%), DCC (2%), and DMD (2%; ref. 41). Copy number transitions in PTEN were detected in 4% of tumors. We also observed transition in ETV6 (Supplementary Fig. S9A), in agreement with previous FISH results (42), and in FLT3, suggesting that these genes frequently targeted in leukemia could also be affected in breast cancer (Supplementary Fig. S9B). The density of probes allowed to map the breakpoints (Supplementary Fig. S9C). Statistical analysis showed that frequent copy number transitions targeted 6 and 15 genes in basal and luminal tumors, respectively (Supplementary Table S3B). Copy number variations (32) were identified in 11 of them (Supplementary Table S3B).

**Correlation between gene expression and genome CNA.** To identify genes whose expression levels were significantly modified in relation to DNA copy number changes, we studied 74 samples (42 basal and 32 luminal) by both aCGH and RNA expression analyses. The genomic and gene expression profiles were then compared across subtype. To define subtype-specific candidates, aCGH data from 7,694 probes with CNA frequencies different between the two subtypes were compared with the Affymetrix data of their 1,834 corresponding independent genes (Fig. 4). The expression of 197 genes was deregulated in relation to CNA with
differences between basal and luminal breast cancers (Supplementary Tables S4A and S5A). Combined genomic and gene expression status of these 197 candidate genes is shown for each individual tumor in Supplementary Tables S4B and S5B. The most significant candidates are shown in Fig. 5A and B.

A total of 106 genes showed gains correlated with up-regulated mRNA expression (Supplementary Table S4A): 73 and 33 genes (corresponding to 77 and 34 probes, respectively) were associated with basal and luminal subtypes, respectively. A total of 91 genes showed losses correlated with down-regulated mRNA expression (Supplementary Table S5A): 28 and 61 genes (corresponding to 30 and 61 probes, respectively) were associated with basal and luminal subtype, respectively. These results suggest the existence of potential oncogenes and tumor suppressor gene (TSG) differentially associated with each subtype.

Among 43 highly amplified genes associated with either basal or luminal subtype (Supplementary Table S2D), 8 genes exhibited up-regulated expression in relation to amplification: ZNF703, PROSC, ASH2L at 8p12, CCND1, PPFA1, CTNN at 11q13.3, HBXAP (11q13.5), and NARS2 (11q14.1; Supplementary Table S4C). They were all associated with luminal subtype. Six of these eight genes have already been defined as potential oncogenes.

Conversely, 237 candidates (152 potential oncogenes and 85 potential TSG; Supplementary Table S6) exhibited correlation between CNA and mRNA expression without specificity for basal or luminal subtype. The presence of common candidate genes in the two subtypes suggests the existence of common pathways of oncogenesis.

For each individual tumor exhibiting the most frequent copy number transitions, gene CNA and expression profiles for 12 of the 21 targeted genes are shown in Supplementary Table S7 (Affymetrix data were not available for the other 9 targeted genes). No basal tumor displayed multiple transition profiles, whereas luminal tumors showed 4 to 13 potential gene breakages. The effect of these alterations in the luminal tumors was not significant on the 5-year metastasis-free survival (MFS) and other clinical features (data not shown). Among the 12 genes exhibiting copy number transitions, none presented a down-regulated expression. Five genes, CHRD, TNXB, SPTBN5, RAB11FIP3, and BC012355, were up-regulated in

Figure 1. Recurrent abnormalities in 93 primary breast tumors. Left, unsupervised hierarchical clustering of genome copy number profiles measured for 93 primary breast tumors by aCGH on 225,388 probes (without X and Y). Red, increased genome copy number; green, decreased genome copy number. Bar to the left, chromosome locations with chromosome 1pter to the top and 22pter to the bottom. Vertical orange line, limits of the two major clusters. The location of the odd-numbered chromosomes is indicated. Below the dendrogram, the first seven rows indicate biological and clinical features of the tumors, whereas the last row indicates subtypes. Color codes and corresponding legends are indicated in the box located to the right at the top of the figure. MED, medullary; LOB, lobular; MIX, mixed lobular and ductal; DUC, ductal; NA, nonassigned. CNA frequencies discriminated basal (red) and luminal (blue) tumors. Right, frequencies of genome copy number gains and losses are plotted as a function of chromosome location with the same previous sequential order for, from the left to the right, both (global) basal and luminal subtypes. Horizontal lines, chromosome boundaries. Positive and negative values indicate frequencies of tumors showing copy number increase and decrease, respectively, with gains and losses as described in Materials and Methods.
38%, 38%, 50%, 40%, and 72% of the corresponding tumors, respectively.

Data comparison with recent integrated genomic studies. The 197 basal- and luminal-specific candidate genes and the 237 common candidate genes were compared with those obtained in two recent integrated whole-genome analyses of breast cancers (Supplementary Table S8; refs. 10, 11). A total of 1,183 candidate genes (11) and 66 highly amplified and overexpressed genes (10) were compared with the here-defined candidates.

Among the 72 candidate genes in common with the data of Neve et al. (11), 19 genes exhibited correlation between CNA and gene expression as well as association with basal or luminal subtype. Gains and overexpression of \textit{PRCC} (1q23.1), \textit{UPF2} (10p14), \textit{C10orf7} (10p14), \textit{HSPA14} (10p14), \textit{RPP38} (10p14), \textit{FOXM1} (12p13), \textit{TULP3} (12p13), and \textit{TEAD4} (12p13) were associated with basal breast cancer (Supplementary Table S4A). Gains and overexpression of \textit{PROSC}, \textit{BRF2}, \textit{ASH2L}, and \textit{DDHD2} at 8p12 and \textit{CCND1}, \textit{FADD}, \textit{PPFIA1}, and \textit{CTTN} at 11q13 (Supplementary Table S4A) as well as loss and down-regulated expression of \textit{SNX3} (6q21), \textit{COX4II} (10p14), and \textit{RPL13} (12p13) genes were associated with luminal subtype (Supplementary Table S5A).

Sixteen highly amplified and overexpressed candidate genes were common to the three studies. None was correlated with basal subtype, whereas eight were associated with luminal subtype: \textit{PROSC}, \textit{BRF2}, \textit{ASH2L}, and \textit{DDHD2} and \textit{CCND1}, \textit{FADD}, \textit{PPFIA1}, and \textit{CTTN} located at 8p12 and 11q13, respectively. Eight other 8p11-12 genes, \textit{LSM1}, \textit{ADAM9}, \textit{MYS3T}, \textit{AP3M}, \textit{POLB}, \textit{VDAC3}, \textit{SCL20A2}, and \textit{THAPI}, were frequently found in the two subtypes (Supplementary Table S6).

In conclusion, our study, using a higher-resolution approach, generated a list of 73 and 33 candidate oncogenes exclusively associated with basal or luminal tumors, respectively (see Fig. 5 and Supplementary Tables S4A and C). Among them, 65 and 25 genes were newly defined as candidate oncogenes associated with basal and luminal subtypes, respectively. \textit{HORMAD1} \((P = 6.5 \times 10^{-10})\) and \textit{ZNF703} \((P = 7 \times 10^{-4})\) were the most significant basal and luminal potential oncogenes, respectively.

Genome profiling of MBC. Gene expression profiling has defined MBCs as a subset of basal breast cancers (24). To determine whether genomic alterations may also distinguish MBCs \((n = 14)\) from ductal grade 3 carcinomas of basal subtype (ductal carcinomas-BSGIII; \(n = 19\)), we compared CNA frequency plots obtained from genomic profiles in these two groups of breast cancers (Supplementary Fig. S10). Unsupervised hierarchical clustering separated the two groups and revealed differences in CNA frequencies \((P < 0.05, \text{Fisher’s exact test, FDR corrected; data not shown})\). Gains of chromosomal regions 4q34 and 12p13 and loss of 4q32-33, 11q13, and 15q11 were associated with MBCs.
whereas gains of 1q23-24, 1q42, 5p13, 5q35, 8p23, 15q13, and 16p12 correlated with ductal carcinomas-BSGIII. Genes targeted are listed in Supplementary Tables S9A and B. Among genes exhibiting copy number gains (log2 ratio >0.5), 86 presented gain frequencies different between MBCs and ductal carcinomas-BSGIII (9 for MBCs and 79 for basal non-MBC; Supplementary Table S9A). Only four genes presented copy number losses correlated with MBCs (Supplementary Table S9B).

For the same samples (14 MBCs and 19 ductal carcinomas-BSGIII), gene expression was also analyzed and compared with genomic profiles to identify candidate genes. Only gains associated with overexpression of LOC440731 (1q42.2) were associated with MBCs.

Validation of candidate oncogenes. First, we focused our analysis on gain of 10p, a new alteration in basal breast cancer. Twenty 10p genes were deregulated in relation to copy number gain with a significant difference in relation with basal subtype (Supplementary Table S4A). To begin to validate 10p candidates in our samples (40 basal and 31 luminal), we measured by immunohistochemistry the protein expression of the CDC123/C10orf7 growth promoter, for which a reliable antibody is available. CDC123 was expressed in 16 basal breast cancers and in none of the luminal samples (Fig. 6A). CDC123 protein expression was associated with basal breast cancer (P = 0.000203, χ² test) and its level correlated with CDC123 mRNA expression (Spearman test rho = 0.296; P = 0.012), which is itself influenced by the genomic status of the gene (Supplementary Tables S4A and B). The effect of CDC123 protein expression in the 16 basal tumors was not significant on the 5-year MFS and other clinical features (data not shown). No correlation between the CDC123 protein expression and the MBC type was established (P = 0.2464, Fisher’s exact test).

Second, we considered important to analyze protein expression from genes targeted by the high-level amplifications because, although rare, they may provide interesting therapeutic targets. As an example, we studied the EGFR amplification in greater details (Supplementary Fig. S8B). EGFR protein expression was analyzed in the same samples (42 basal and 32 luminal) by immunohistochemistry (Fig. 6B). Two basal tumors showed EGFR amplification and EGFR loss, respectively (Supplementary Fig. S8B). EGFR mRNA was expressed at high level in 3 basal cases (including the amplified case), at normal level in 14 basal (35.8%) and 1 luminal cases, and at low level (or inferior to twice the normal breast tissue expression level) in 25 basal (60%) and 31 luminal tumors (96.6%). Membrane positivity was observed in 31 basal cases (79.4%) with a quick score
higher than 10, including 17 cases with >50% of positive cells. The basal case with EGFR gene loss did not show EGFR expression. The EGFR-amplified case had the strongest immunohistochemical signal, with the maximum quick score of 300, suggesting a high production of protein as a consequence of gene amplification (Fig. 6B). EGFR protein expression was associated with basal cases ($P = 1.496 \times 10^{-9}$, $\chi^2$ test) as reported previously (42).

**Discussion**

We have described a combined, high-resolution analysis of gene expression and genome copy number in primary human breast cancers of basal and luminal subtypes. It was performed to characterize genomic events and identify candidate cancer genes and potential therapeutic targets that may help identify and treat two major molecular breast cancer subtypes. We used the 244K aCGH matrix from Agilent Technologies, which allows a better resolution than BAC-based aCGH (43).

**Genomic alterations and molecular subtypes.** We observed the existence of multiple genomic alterations supporting the idea that breast cancers are highly heterogeneous. Several regions and genes targeted by these alterations were different in basal and luminal subtypes. We confirmed some strong associations between subtype and genome profiles reported in previous studies (9–11); good examples are gains at 12p13 and 11q13 regions associated with basal and luminal subtypes, respectively. We identified gain of 10p as a new alteration of basal breast cancer and found that gains of 8p11-12 targeting the previously defined A1 and A2 amplicons (37) are associated with luminal subtype.

We observed rare high-level amplifications in basal tumors, suggesting again that breast cancers are highly heterogeneous even within a given subtype. They affected small regions, including

![Figure 4](https://www.aacrjournals.org/11571/CancerRes2007;67:24.11571)

**Figure 4.** Correlation between gene expression and genome alterations. Genomic and gene expression profiles were established for 74 breast tumors (42 basal and 32 luminal). Left, unsupervised hierarchical clustering of genome copy number profiles measured for the 74 breast tumors by aCGH with 7,694 discriminant probes identified to significantly differentiate basal and luminal tumors and to cover genes commonly analyzed in genomic and gene expression studies. Vertical orange line, limits of the two major clusters. Data are displayed as described in Fig. 1. To the immediate right are plotted the corresponding significant CNA frequency differences observed for luminal (blue) and basal (red) subtypes. Horizontal lines on frequency plots, chromosome boundaries; horizontal dashed lines, centromere locations. Gene expression of the 1,834 corresponding independent genes was analyzed on the same 74 tumors and allowed to draw an unsupervised hierarchical clustering of genes (shown to the right of the figure) with regard to their respective chromosomal location. The aCGH data from 7,694 probes that revealed significant differences of CNA frequencies between the two subtypes (genomic profiles) were compared with the expression of their 1,834 corresponding independent genes (expression profiles). To the immediate left of the expression profiles are plotted (to the right or to the left part, respectively) the candidate genes that showed significant differences in copy number gain correlated with up-regulated expression or in copy number loss correlated with down-regulated expression between basal (red) and luminal (blue) subtypes.
Figure 5. Most significant candidate cancer genes in basal and luminal tumors. One hundred sixty candidate oncogenes and 89 putative TSGs are listed in Supplementary Tables S4A and S5A, respectively. The most significant candidate oncogenes (A) and TSG (B) are shown associated with their corresponding chromosome CNA frequency plot defined in basal and luminal tumors. For a better visualization of the targeted region, a threshold value of log2 ratio >0.3 was used to draw chromosome CNA frequency plots using CGH Analytics software. For each subtype, all the candidate genes were significantly defined with a FDR-adjusted t test $P < 0.05$, but some of them were associated with a significance of $P < 0.001$ (noted with asterisk). Underlined genes have been associated with oncogenesis (see Supplementary Tables S4A and S5A for references).
PIK3CA (3q26), IGF1R (15q26), and CCNE1 (19q11-12), but also single genes, such as EGFR (7p11), FGFR2 (10q26), and BCL2L2 (14q11). EGFR, FGFR2, EGFR, and IGF1R are tyrosine kinase receptors and thus potential therapeutic targets. Knowledge of CNAs can have immediate clinical use in diagnosis and can, in some cases, provide useful prognostic information (see ref. 44 for review). The existence of these amplifications and such high degree of heterogeneity confirms that molecular profiling will be paramount to select the appropriate treatment.

Specific genomic losses were also detected in basal and luminal subtypes. Our results agree with loss of heterozygosity (LOH) studies: LOH at 4p and 5q defines a subclass of basal breast cancers, whereas LOH at 1p and 16q occurs preferentially in ER-positive breast cancers (45). Losses of 4p and 5q associated with basal breast cancers targeted several genes (Supplementary Tables S2B and C), including candidate or known TSG: SLIT2 (4p15.31), GPR125 (4p15.31), RASAI (5q14.3), and APC (5q22.2). Genes coding for P53, its repressor PRDM1 (6q21), and its effector PERP (6q23.3) were among genes with losses associated with luminal subtype (Supplementary Tables S2B and C). However, expression of these three genes was not significantly deregulated in relation to copy number loss.

We also detected genomic alterations through analysis of copy number transitions, which could be associated with chromosome breaks. They targeted 501 genes in at least two different tumors. Results were consistent with previous studies (38–40), underlining the quality and the resolutive power of the aCGH analysis. Ten large genes associated with fragile sites were targeted, with the higher frequency for CSMD1 (41). Copy number transitions with frequencies different between subtypes targeted 6 and 15 genes in basal and luminal tumors, respectively. The difference in the number of involved genes may suggest a different mode of genomic instability in luminal comparatively to basal tumors. Among genes with frequent transitions in basal subtype (Supplementary Table S3B) was PUM2, a human homologue of Pumilio, which codes for a protein required to maintain germ-line stem cells in protostomians (46). Among genes with frequent copy number transitions in luminal samples (Supplementary Table S3B), few are characterized except GLI2 (GLI-Kruppel family member GLI2), CHRD (chordin), DOM3Z, and TNXB.

**Figure 6.** CDC123 and EGFR, two basal candidates. A, CDC123 expression was analyzed by immunohistochemistry and quantified. As illustrated with luminal L5342 sample, no luminal case expressed CDC123, whereas medullary B11614 was the most positive basal case. CDC123 was also found expressed in the basal B12478 sample. B, EGFR expression was analyzed by immunohistochemistry and quantified in 12 basal tumors. EGFR expression was negative in the case with EGFR gene loss (B11442), whereas the EGFR-amplified case (B8595) displayed the strongest immunohistochemical signal intensity, suggesting a high production of EGFR protein as a consequence of gene amplification (Supplementary Fig. S8B). Increased EGFR expression was also observed in cases without EGFR amplification (B9752 and B10033). EGFR and CDC123 showed membrane and cytoplasmic signal, respectively. CDC123 and EGFR scoring associated the percentage of positive cells (P, from 0% to 100%) and the intensity of the staining (I, from 0 to 3) as described in Materials and Methods.
between basal (73 genes) and luminal (33 genes) subtypes. The extent of concordance between high-level amplification and increased gene expression has been analyzed previously (47). In basal and luminal subtypes, several genes have been associated with oncogenesis or with the control of various cellular processes (see Supplementary Table S4A and Fig. 5 for details and references). Interestingly, among basal candidate genes, several code for proteins potentially involved in transcription or cell cycle regulation, whereas luminal candidates encode proteins involved in cellular differentiation. Among the eight genes with high expression in relation to high-level amplification in luminal tumors, six have been associated with cancer (Supplementary Table S4C). The genes targeted by low-level CNAs could be selected during oncogenesis to increase basal metabolism of the tumor cells, whereas genes deregulated by high-level amplifications are predominantly involved in signaling and cell cycle regulation (10). IKBKE (1q32.1) has been recently identified as an oncogene in breast cancers (48). In our work, IKBKE was not found among specific and common candidate oncogenes. This indicates that our list is not comprehensive and further supports the great heterogeneity of breast cancer.

Taken together, our and previous studies (10, 11) support the existence of a combined activation of several potential oncogenes in 1q21-23, 8p12, 10p14, 11q13, and 12p13 regions involved in specific basal and luminal oncogenic pathways. Our high-resolution study pointed to additional candidate oncogenes such as 10p genes and HORMAD1 and ZNF703.

A total of 89 genes with copy number losses correlated with down-regulated mRNA expression were associated with basal (28 genes) and luminal (61 genes) subtypes, respectively. In basal and luminal subtypes, only 7 and 11 genes have been thus far associated with cancer as potential TSG or as coding for proteins involved in growth inhibition pathways (see Supplementary Table S5A and Fig. 5 for details and references), respectively.

A total of 237 common candidate cancer genes were associated with both molecular subtypes. These may participate in general oncogenic mechanisms. One possibility is that these genes are targeted in a stem cell background, whereas subtype-specific alterations occur in committed progenitors (49).

**Genome profiling discriminates between MBCs and other basal breast cancers.** Among the 79 genes with significant gains in ductal carcinomas-BSGIII, some have potential oncogenic roles in cancer progression, such as CDC41 (Supplementary Table S9A). We showed that gains in MBCs target significantly GALNT17 and CR5900663 at 4q34 and LRRC23, ENO2, ATN1, C12orf57, PTPN6, PHB2, and BCL2L14 at 12p13. Interestingly, 12p13 genes targeting BCL2L14 are consistent with an amplicon associated with MBCs (14) and with gene expression analyses (24). Losses of 4q32-33, 11q13, and 15q11 are associated with MBCs and target KLHL2 (4q32-33), FLJ26534 (4q32-33), ORAOV1 (11q13), and A1941978 (15q11), respectively (Supplementary Table S9B). We identified only LOC440731 (1q42.2) as a candidate oncogene associated with MBCs.

**Validation of candidate oncogenes.** We found that the 10p candidate oncogene CDC123/C10orf7, which is potentially involved in cell cycle regulation, is expressed in basal breast cancer at the protein level (see Supplementary Table S4A). Similarly, we confirmed the strong relation between EGFR protein level and basal subtype. Frequent EGFR positivity has been observed in metaplastic carcinomas, a subset of basal carcinomas (50). Pursuing these analyses with the list of potential candidate breast cancer genes may provide therapeutic targets.

In conclusion, we have studied genome copy number and gene expression in primary human breast cancers. We have identified genomic events that distinguish basal and luminal tumors and have identified various rare high-level amplifications that may provide therapeutic targets. We have characterized gain of 10p as a new alteration in basal breast cancer and have shown that a limited region of 8p11-12 amplicon is found in luminal rather than in basal tumors. Our study pointed to basal and luminal candidate oncogenes, such as HORMAD1 and ZNF703, defined as the best significant basal and luminal potential cancer genes, respectively. We have described 73 candidate oncogenes and 28 candidate TSG as well as 33 candidate oncogenes and 61 candidate TSG that may contribute to the pathophysiology of basal and luminal breast tumors, respectively. At last, among 10p candidate oncogenes associated with basal subtypes, we validated CDC123/C10orf7 protein as a potential basal marker ($P = 2 \times 10^{-5}$).

Integrated genomics help understand mammary oncogenesis and develop future treatments associating therapeutic targets with various molecular categories of tumors. Such treatments could target candidates associated with a specific subtype, candidates common to all subtypes, and the rare highly amplified candidates with high protein expression level. In this study, six potential oncogenes associated with basal or luminal subtype (Supplementary Table S5A), nine commonly found (Supplementary Table S8), as well as several rare highly amplified candidates are potentially druggable.

**Acknowledgments**

Received 7/6/2007; revised 9/6/2007; accepted 10/15/2007.

**Grant support:** Institut National de la Santé et de la Recherche Médicale, Institut Paoli-Calmettes, L'Association pour la Recherche sur le Cancer (N° 7931), Ligue Nationale Contre le Cancer (Label 2007-2009), Caisse Primaire d'Assurance Maladie des Professions Libérales de Province, and Institut National du Cancer (PI 2006). I. Bekhouche is supported by a fellowship from the Algerian Government. F. Sircoulomb is supported by a fellowship from the French Government.

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.

We thank F. Birg and J.P. Borg for encouragement, B. Esterni for statistical analysis, and J.M. Durey for iconographic assistance.

**References**

Integrated Profiling of Basal and Luminal Breast Cancers

José Adélaïde, Pascal Finetti, Ismahane Bekhouche, et al.


Updated version Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/67/24/11565

Supplementary Material Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2007/12/13/67.24.11565.DC1

Cited articles This article cites 50 articles, 18 of which you can access for free at: http://cancerres.aacrjournals.org/content/67/24/11565.full.html#ref-list-1

Citing articles This article has been cited by 36 HighWire-hosted articles. Access the articles at: /content/67/24/11565.full.html#related-urls

E-mail alerts Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.