

Integrated Genomic and Transcriptomic Analysis of Ductal Carcinoma *In situ* of the Breast

Anne Vincent-Salomon,^{1,2} Carlo Lucchesi,² Nadège Gruel,^{2,3} Virginie Raynal,² Gaëlle Pierron,¹ Rémi Goudefroye,¹ Fabien Rey,⁴ François Radvanyi,⁴ Rémy Salmon,⁵ Jean-Paul Thiery,⁴ Xavier Sastre-Garau,¹ Brigitte Sigal-Zafrani,^{1,6} Alain Fourquet,⁷ and Olivier Delattre,^{1,2} for the breast cancer study group of the Institut Curie

Abstract Purpose: To gain insight into genomic and transcriptomic subtypes of ductal carcinomas *in situ* of the breast (DCIS).

Experimental Design: We did a combined phenotypic and genomic analysis of a series of 57 DCIS integrated with gene expression profile analysis for 26 of the 57 cases.

Results: Thirty-two DCIS exhibited a luminal phenotype; 21 were ERBB2 positive, and 4 were ERBB2/estrogen receptor (ER) negative with 1 harboring a *bona fide* basal-like phenotype. Based on a CGH analysis, genomic types were identified in this series of DCIS with the 1q gain/16q loss combination observed in 3 luminal DCIS, the mixed amplifier pattern including all ERBB2, 12 luminal and 2 ERBB2⁻/ER⁻ DCIS, and the complex copy number alteration profile encompassing 14 luminal and 1 ERBB2⁻/ER⁻ DCIS. Eight cases (8 of 57; 14%) presented a *TP53* mutation, all being amplifiers. Unsupervised analysis of gene expression profiles of 26 of the 57 DCIS showed that luminal and ERBB2-amplified, ER-negative cases clustered separately. We further investigated the effect of high and low copy number changes on gene expression. Strikingly, amplicons but also low copy number changes especially on 1q, 8q, and 16q in DCIS regulated the expression of a subset of genes in a very similar way to that recently described in invasive ductal carcinomas.

Conclusions: These combined approaches show that the molecular heterogeneity of breast ductal carcinomas exists already in *in situ* lesions and further indicate that DCIS and invasive ductal carcinomas share genomic alterations with a similar effect on gene expression profile.

Ductal carcinoma *in situ* (DCIS) represents 15% to 20% of all newly diagnosed cases of breast cancer, and its incidence is increasing as a result of mammography screening (1).

DCIS correspond to a proliferation of malignant epithelial cells that spare the basal membrane with no stromal invasion.

Authors' Affiliations: ¹Institut Curie, Department of Tumor Biology, ²Institut National de la Santé et de la Recherche Médicale Unit 830, Institut Curie, ³Institut Curie, Translational Research Department, ⁴Institut Curie, UMR144 Centre National de la Recherche Scientifique, ⁵Institut Curie, Department of Surgery, ⁶Head of the Breast Cancer Study Group of the Institut Curie, and ⁷Institut Curie, Department of Radiation Oncology, Paris, France

Received 6/14/07; revised 10/28/07; accepted 11/29/07.

Grant support: Institut National de la Santé et de la Recherche Médicale, the Institut Curie (Programme Incitatif et Coopératif), and the Ligue Nationale contre le Cancer. Construction of the BAC array was supported by grants from the Carte d'Identité des Tumeurs program of the Ligue Nationale Contre le Cancer. Dr Anne Vincent-Salomon was supported by an "Interface Institut National de la Santé et de la Recherche Médicale" grant.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

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

Requests for reprints: Olivier Delattre, Institut National de la Santé et de la Recherche Médicale Unit 830, Institut Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France. Phone: 33-1-42-34-66-81; Fax: 33-1-42-34-66-30; E-mail: olivier.delattre@curie.fr.

©2008 American Association for Cancer Research.
doi:10.1158/1078-0432.CCR-07-1465

They can be classified into three categories (low-, intermediate-, and high-grade) according to nuclear grade, differentiation, and presence of central necrosis (2). However, this classification of DCIS is poorly reproducible among pathologists (3). Molecular studies, still based on a limited number of cases, have revealed some characteristics of high-grade DCIS, including the high frequency of ERBB2 overexpression, *TP53* mutations, chromosomal 8p loss, and 1q gain (4–9). In contrast, positive estrogen receptors (ER), *BCL2* expression, and 16q loss have been reported to characterize low-grade DCIS (10–12).

Treatment, regardless of DCIS nuclear grade, is based on mastectomy or lumpectomy in combination with radiotherapy (1, 13). Factors known to be associated with a higher risk of recurrence are young age, positive margins, necrosis, and high nuclear grade (14, 15). Some emerging data suggest that ERBB2-amplified DCIS could present a higher risk of recurrence (16, 17).

Recent gene expression profiling studies have reported that consistent differences in expression of a subset of genes can be identified between low-grade and high-grade DCIS (18–20). Interestingly, *in situ* lesions preferentially cluster with invasive lesions of the same grade, suggesting that DCIS are precursors of their similar grade invasive counterpart (20). In support of this hypothesis, DCIS and invasive components of the same tumors exhibit similar patterns of genetic alterations (4, 7).

Gene expression profile analysis classified invasive breast carcinomas into five groups: luminal A or B, basal like, ERBB2, and normal like carcinomas (21, 22). These analyses have led

to the identification of specific immunophenotypic markers, which allow this classification to be used in clinical practice (23, 24). It is well-recognized that *ERBB2* is up-regulated more frequently in DCIS than in invasive carcinomas (6). The identification of a basal-like entity is emerging with four recent studies describing DCIS lesions with this phenotype (18, 25–27). It nevertheless remains to be determined whether basal-like DCIS show specific genomic/transcriptomic alterations.

Genomic profiles also provide criteria for a clinically relevant classification, as combined genomic and gene expression studies in invasive cancers have highlighted the prognostic effect of high level amplifications associated with gene over-expression at chromosome regions 8p12, 11q13, 17q12, and 20q13 independently of the previously described major gene expression classes. These cases are defined as amplifiers. Among nonamplifiers, one group with a good prognosis harbors few genetic alterations but recurrent 16q loss associated with 1q gain (1q/16q). Another group, which includes most of basal-like but also a subset of luminal cases, is characterized by numerous low-level copy number alterations (CNA; refs. 28, 29). In invasive breast cancer, this integrated genomic approach, based on genomic and transcriptomic data, is very promising to achieve better clinical management of patients.

In the present study, we analyzed the immunophenotype, the array CGH profiles, the *TP53* gene sequence, and the gene expression profile of DCIS frozen samples. We identified DCIS subtypes according to the molecular classification described for invasive ductal carcinomas (IDC), allowing comparison of the genomic, transcriptomic, and phenotypic profiles of *ERBB2*, luminal, and basal-like DCIS. Finally, correlations of genomic and transcriptomic data were used to identify candidate genes, the altered expression of which may play a critical role in early breast carcinogenesis via copy number changes.

Materials and Methods

Selection of tumor samples. Fifty-seven DCIS of the breast were retrospectively selected from our tumor bank based on the availability of a frozen sample. Samples were collected on fresh surgical specimens by a pathologist within 1 hour after surgery from patients who underwent lumpectomy ($n = 21$) or mastectomy ($n = 36$), flash frozen in liquid nitrogen, and stored at -80°C . Experiments were done in agreement with the French Bioethics Law 2004-800 and the Ethics Charter from the French National Institute of Cancer. All samples were reviewed for histologic classification according to nuclear grade and classified as low, intermediate, and high nuclear grade (Table 1; ref. 2).

DNA and RNA extractions. All tumor samples contained $>50\%$ of cancer cells, as assessed by H&E staining of histologic sections of the samples used for nucleic acid extraction. In this work, to avoid potential artifacts, which may be due to DNA or RNA amplification of microdissected samples, we privileged the investigation of samples large enough to extract good quality nucleic acids without need for preamplification steps. DNA and RNA extractions were done using standard procedures as previously described (30). RNA quality control was checked on Agilent 2100 bioanalyzer (Agilent Technologies). Good quality RNA was obtained in a subset of 26 cases among these 57 cases.

Immunohistochemistry. Immunostaining was done according to previously published protocols (31). The expression of ER (clone 6F11; 1/200; Novocastra), progesterone receptor (PR; clone 1A6; 1/200; Novocastra), *ERBB2* (clone CB11; 1/1,000; Novocastra), epidermal growth factor receptor (HER1; clone 31G7; 1/40; Zymed; Clinisciences), cytokeratin 5/6 (clone D5/16B4; 1/50; Dako), and cytokeratin 8/18 (clone DC10; 1/100; Zymed; Clinisciences) were evaluated. For each antibody, internal and external controls were included in the experiments.

Table 1. Pathologic characteristics and immunophenotype of the 57 DCIS

Variables	All DCIS $n = 57^*$ (%)	<i>ERBB2</i> $n = 21^*$ (%)	Luminal A $n = 32$ (%)	<i>ERBB2</i> ⁺ / <i>ER</i> ⁻ $n = 4$ (%)
Nuclear grade				
High	29 (51)	17 (81)	9 (28)	3 (75)
Nonhigh [†]	28 (49)	4 (19)	23 (72)	1 (25)
Necrosis				
≥10%	36 (63)	18 (86)	16 (50)	2 (50)
<10%	21 (37)	3 (14)	16 (50)	2 (50)
ER				
+	34 (60)	2 (10)	32 (100)	0
-	22 (40)	18 (90)	0	4 (100)
PR				
+	29 (52)	2 (10)	26 (81)	1 (25)
-	27 (48)	18 (90)	6 (19)	3 (75)
<i>ERBB2</i> (IHC) [‡]				
+	20 (36)	20 (100)	0 (0)	0
-	36 (64)	0	32 (100)	4 (100)
KRT8/18				
>50%	50 (89)	18 (90)	31 (97)	1 (25)
≤50%	6 (11)	2 (10)	1 (3)	3 (75)
KRT5/6				
≥5%	15 (27)	5 (25)	9 (28)	1 (25)
<5%	41 (73)	15 (75)	23 (72)	3 (75)
EGFR				
≥5%	6 (11)	3 (15)	2 (6)	1 (25)
<5%	50 (89)	17 (85)	30 (94)	3 (75)

NOTE: Cases were considered positive for ER and PR using $\geq 10\%$ of positive nuclei per carcinomatous duct and for *ERBB2* with a cutoff of $>60\%$ of positive cells.

Abbreviations: EGFR, epidermal growth factor receptor; KRT, cytokeratin.

*No fixed paraffin-embedded tissue sample was available for one case and immunophenotype was not assessed.

[†]Nonhigh-grade tumors corresponded to 4 low-grade and 24 intermediate-grade DCIS.

[‡]IHC, *ERBB2* status determined by immunohistochemistry.

***TP53* sequencing.** Exons 4 to 10 of *TP53* were PCR amplified and bidirectionally sequenced using Big Dye Terminator chemistry (Applied Biosystems) with an ABI PRISM 3700 DNA Analyzer (primer sequences previously published in ref. 32).

Array CGH. The 3.5K BAC array together with experimental procedures used for hybridization and washing were as previously published (32). All BAC and PAC were verified by end sequencing before spotting.

Analysis of array CGH data. Data analysis was based on the normalized ratios of Cy5/Cy3 signals observed for each BAC clone that previously passed the flag assessment procedure (32, 33). For autosomal chromosomes, the status of each BAC locus was defined according to the VAMP (Visualization and Analysis of Molecular Profiles) analysis procedure (34) using the GLAD (Gain and Loss Analysis of DNA) algorithm that, based on the Adaptive Weights Smoothing procedure, can be used to assign a status (gained, lost, or normal) to each chromosome region (33).

Regions of gain or loss and amplicons were defined as previously published (32). BACs known to harbor copy number variation in public databases (Database of Genomic Variants)⁸ or identified in the laboratory by array CGH experiments using normal/normal DNAs hybridizations were excluded from the analysis.

⁸ <http://projects.tcag.ca/variation/>

Gene expression profile analysis. The DNA microarrays used in this study were the Human Genome U133 Set (HG-U133; Affymetrix), consisting of one GeneChip array (A) and containing almost 20,000 probe sets. Microarray data were simultaneously normalized using the GCRMA package version 1.1.4 in the R environment (R Development Core Team). Unsupervised hierarchical clustering of the tumor sample was done using the dChip software (35) with standard Pearson correlation as similarity measure and centroid as linkage criteria. Clustering was done on a subset of probe sets with an interquartile range of >2 . Each probe set was preliminary standardized with respect to mean and SD. The comparison of mean gene expression between ERBB2-amplified and luminal DCIS was done using SAM (36) and Welch *t* tests. FDR for Welch *t* test was adjusted using the Benjamini-Hochberg corrections. Finally, for each probe set, the correlation with Cy5 to Cy3 ratios of the corresponding BAC clone was calculated using Pearson and Spearman correlation tests (for details see Supplementary Data 1).

Results

Patient and tumor characteristics. Fifty-seven DCIS cases with no associated invasive component were analyzed. The median age of patients was 52.5 years (range, 30-79 years). All patients were treated according to established protocols: mastectomy (36 patients) or conservative surgery followed by

radiation therapy (21 patients). The main tumor characteristics are indicated in Table 1. Twenty-nine cases (51%) were high nuclear grade, and 28 (49%) were nonhigh-grade, including 24 intermediate- and 4 low-grade tumors. For subsequent analyses, the intermediate and low nuclear grade DCIS cases were merged into a nonhigh-grade group. The immunophenotype of the 57 cases was determined to identify whether DCIS could be classified into subgroups previously identified in IDC: luminal ER-positive, ERBB2, basal-like triple-negative, and normal-like carcinomas (Table 1). Altogether, these markers were able to identify 32 luminal (ER positive) DCIS (most being nonhigh-grade group; 23 cases), 21 ERBB2 DCIS (most being ER negative; 18 ER-negative cases), and 17 of 21 being high nuclear grade, and finally 4 cases were ER and ERBB2 negative. Only one of these four ERBB2-negative/ER-negative DCIS was PR negative, expressed cytokeratin 5/6 basal cytokeratins and epidermal growth factor receptor, and could therefore be considered to be a *bona fide* basal-like DCIS (Fig. 1A).

Genomic analysis. We then investigated the *TP53* status and genomic profiles of these 57 DCIS lesions and their relationship with the three main phenotypic classes identified above. All samples were frozen DCIS lesions and comprised at least 50% of tumor cells. *TP53* mutations were detected by direct sequencing of exons 4 to 11. Mutations were found in eight

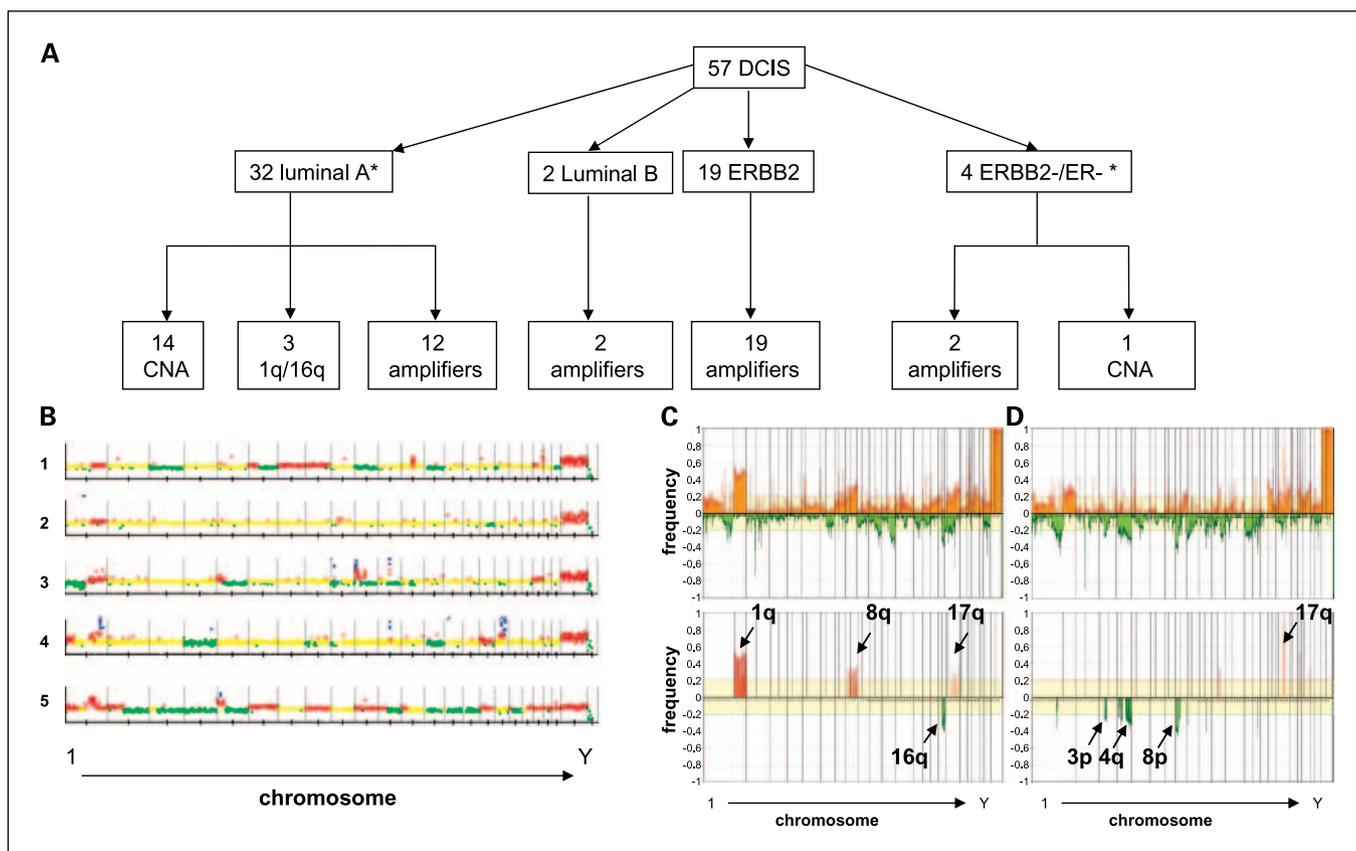
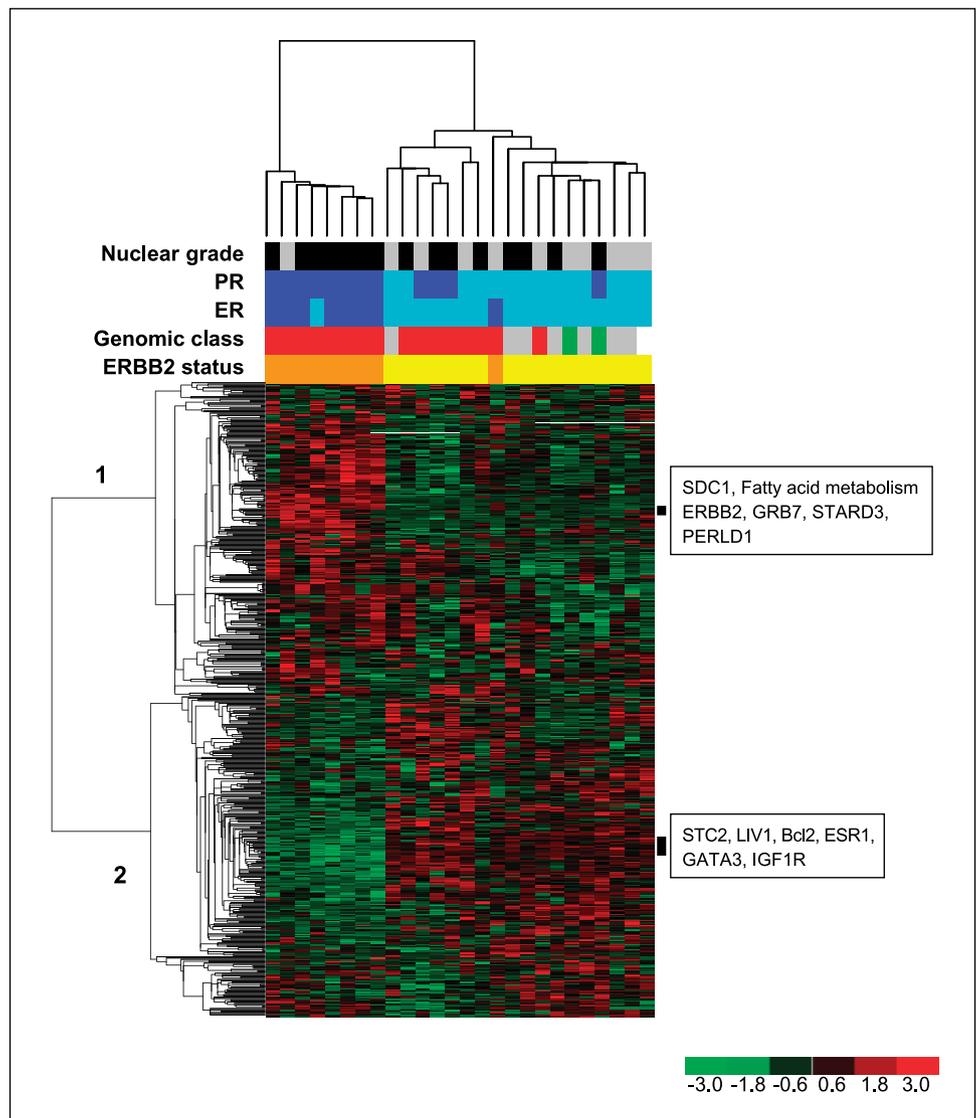


Fig. 1. Phenotypic and genomic analysis of DCIS. **A**, classification of the 57 DCIS according to the molecular subtypes previously identified for invasive breast carcinomas and their genomic status. *, three luminal DCIS and one ERBB2-negative, ER-negative DCIS presented a flat profile. **B**, examples of array CGH profiles from defined genomic status; **B1**, luminal with CNA; **B2**, luminal with 1q gain/16q loss profile; **B3**, luminal amplifier; **B4**, ERBB2 amplified; **B5**, triple-negative, basal-like amplifier with numerous CNA. **C** and **D**, frequencies of genome copy number gains and losses plotted as a function of genome location (**C**, luminal DCIS; **D**, ERBB2-amplified DCIS). Orange, gains; green, losses. From chromosome 1pter on the left to chromosome Xq on the right. Vertical lines, chromosome boundaries; vertical hemi-dashed lines, centromere locations. The BAC clones exhibiting significantly more frequent gains (red) or losses (green) between luminal (**C**) and ERBB2-amplified DCIS (**D**) are displayed underneath each frequency plot.

Fig. 2. Unsupervised hierarchical clustering of 26 DCIS tumors using intrinsically variable gene expression of 502 genes (interquartile range, >2). Each row represents a different tumor, and each column represents one of 502 genes used for this unsupervised clustering (orange, *ERBB2* amplified; yellow, *ERBB2* not amplified), indicating the genomic type (red, amplifiers; gray, CNA; green, 1q/16q; white flat profile). Cyan, ER or PR positive; dark blue, ER and PR. NG, nuclear grade; black, high-grade; gray, nonhigh-grade. The genes are divided into two main clusters. The first cluster (1) contains genes involved in fatty acid metabolism, extracellular matrix, and the *ERBB2* amplicon. The second cluster (2) includes genes involved in apoptosis and ER signaling.



cases (14%) and corresponded to two frameshift, two splice site, two missense, and two stop mutations. These mutations were located in exon 5 (three cases); exon 6 (two cases); and exons 7, 8, and 9 (one case each). They occurred more frequently in high-grade DCIS (7 of 29; 24%) than in nonhigh-grade DCIS (1 of 28; 4%; $P = 0.051$) and significantly more frequently in *ERBB2* (6 of 17; 35%) than in luminal and *ERBB2/ER* DCIS (2 of 36; 6%; $P = 0.0047$).

CNA (amplicons, gains, and losses) were identified using a one-megabase resolution BAC array. All samples but four exhibited altered profiles, indicating that the amount of stromal or other normal cells in the samples was not sufficient to mask copy number changes. Fifty-nine different amplicons were observed in the 57 DCIS (Supplementary Data 2). Fourteen amplicons were recurrent (observed in at least two tumors), including *ERBB2* on chromosome 17q12, *Cyclin D1* at 11q13, *FGFR1* at 8p12, *BMP7/BCAS1* at 20q13, but also five 17q amplicons distinct from *ERBB2*. Interestingly, the chromosomal localization of amplicons (distinct from *ERBB2*) differed between the *ERBB2*-amplified and the luminal DCIS groups, as

half of the amplicons (13 of 26) of the *ERBB2* group associated with *ERBB2* were also localized on chromosome 17q, but only 5 of 29 amplicons of the luminal group were localized on this chromosome arm ($P = 2.10^{-2}$).

The *ERBB2* amplicon size differed from one tumor to another, ranging from 32.9×10^6 bp to 35.6×10^6 bp with a minimal common region of amplification located between 34.9×10^6 bp and 35.2×10^6 bp centered on the *ERBB2* gene sequence but also containing other genes including *STARD3*, *GRB7*, and *PERLD1*.

Regions of gains and losses observed with a frequency of $>20\%$ in the luminal and *ERBB2*-amplified groups were identified. Common regions of gain were observed in both groups (Fig. 1C and D). They were located on chromosomes 1p, 1q, 5p, 7p, 7q, 8q, 14q, 16p, 19p, 19q, 17q, and 20q. Only one common and recurrent region of loss on 11q was identified. The *ERBB2*-amplified group presented specific regions of 3p, 4p, 4q, and 8p losses and 17q gains compared with the luminal group that presented specific regions of gains on 1q, 8p, and 17q and losses on 16q (Supplementary Data 3).

According to the criteria recently proposed by Fridlyand and Chin in 2006 (28, 29), DCIS cases could be classified into three categories: tumors characterized by the presence of amplicons (mixed amplifiers), tumors characterized by 1q gain and 16q loss (1q/16q group), and tumors with complex CNA. *ERBB2*-amplified DCIS never showed 16q loss. The luminal group exhibited genetic complexity with 12 cases classified as amplifiers, 3 cases harboring a 1q/16q profile, 14 cases exhibiting a complex pattern of CNA, and 3 cases characterized by a flat profile. Within the *ERBB2*/ER-negative DCIS group, two belonged to the amplifier category with amplicons located on chromosomes 5p, 8p, and 22q, one harbored CNA and one case was characterized by a flat profile (Fig. 1A and B).

Gene expression analysis. We also investigated transcriptomic differences between *ERBB2*-amplified and luminal DCIS. A subset of 26 cases for which good quality RNAs were available, including 9 *ERBB2*-amplified and 17 luminal DCIS, were analyzed on Affymetrix U133A. A total of 502 genes exhibited an interquartile range of >2 (Supplementary Data 4). These genes discriminated two major clusters of tumors mainly organized according to *ERBB2* and hormonal receptor status rather than nuclear grade. One branch included most (8 of 9) *ERBB2*-amplified, ER-negative cases, and the other branch comprised all (17 of 17) luminal cases and one *ERBB2*-amplified, ER-negative/PR-positive case. As shown in Fig. 2, genes coamplified with *ERBB2* (*GRB7*, *PERLD1*, and *STARD3*) and genes of the ER pathway (*ESR1*, *GATA3*, *STC2*, *LIV1*, *IGF1R*...) were major contributors of this clustering. The most discriminant genes between *ERBB2*-amplified and luminal DCIS (Welch and Sam tests results, $P < 0.01$) are indicated in Supplementary Data 5. Unsupervised and supervised analyses showed that genes involved in estrogens receptor pathways, extracellular matrix remodeling (*syndecan*, *ezrin*, ...), immune response (*TRBV19*, ...), antiapoptotic pathways, cell adhesion (*CXCR4*, *syndecan*, and *ITGB6*), or fatty acid metabolism (*LDRL*) were differentially expressed between the *ERBB2*-amplified and luminal groups (Fig. 2). Finally, genes encoding important signal transduction and transcription factors molecules were differentially expressed. *FGFR2*, *RET*, *KIF5*, and *IGF1R* were overexpressed in luminal cases, whereas *ERBB2*, *GRB7*, *PTPRC* were overexpressed in *ERBB2*-amplified cases.

Interestingly, displaying genomic characteristics of tumors on the unsupervised cluster revealed that among the two branches of the luminal cluster, one was enriched in amplifiers distinct

from the *ERBB2* amplicon; the second was enriched in CNA (Fig. 2).

Genomic and transcriptomic correlations. Given the strong relationship between the genomic classification and the unsupervised clustering of gene expression profiles and to determine the effect of genomic changes on gene expression levels, we combined the analyses of the Affymetrix signals and array CGH ratios for the 26 DCIS analyzed with both approaches. Using the Pearson and Spearman correlation tests, we observed that the expression level of 416 genes was significantly correlated with the ratio of their corresponding BAC locus (372 BAC clones; Fig. 3). Correlations corresponding to amplicon were mostly located on 17q and corresponding to either the variable extent of the *ERBB2* amplicon from one tumor to another or to other 17q amplicons, distinct from but co-occurring with *ERBB2* (Supplementary Data 2). Other correlations due to amplicons mainly involved 8p (from 37 to 40 Mb), 11p12-p13 (from 34 to 36 Mb), 11q13 (from 69 to 78 Mb), and 20q13.32 (from 46 to 55 Mb; Supplementary Data 6). The main correlations due to low copy gains or losses were observed on chromosomes 1q, 8q, and 16q (Table 2; Supplementary Data 6). The analysis of genome expression correlations indicates that the amount of stromal or other normal cells in the samples was not sufficient to hamper the detection of a strong copy number effect upon gene expression.

Similar correlations have recently been described in IDC (28). Interestingly, when the lists of genes correlated with copy number changes in DCIS and IDC (28) were compared, we observed a strong overlap, as 204 of 416 (49%) genes that correlated with genomic status in DCIS were also shown to be correlated with genomic status in IDC. Strikingly, this overlap represented 103 of 170 genes (61%) within the genomic regions differentially altered between the *ERBB2* and luminal DCIS groups (Table 2; Supplementary Data 3).

Discussion

We report the first combined phenotypic, genomic, and transcriptomic analysis of a series of DCIS frozen samples. The most striking conclusions of this work are the high degree of phenotypic, genomic, and transcriptomic similarities between DCIS and IDC and high diversity of breast tumors existing already *in situ* early phase.

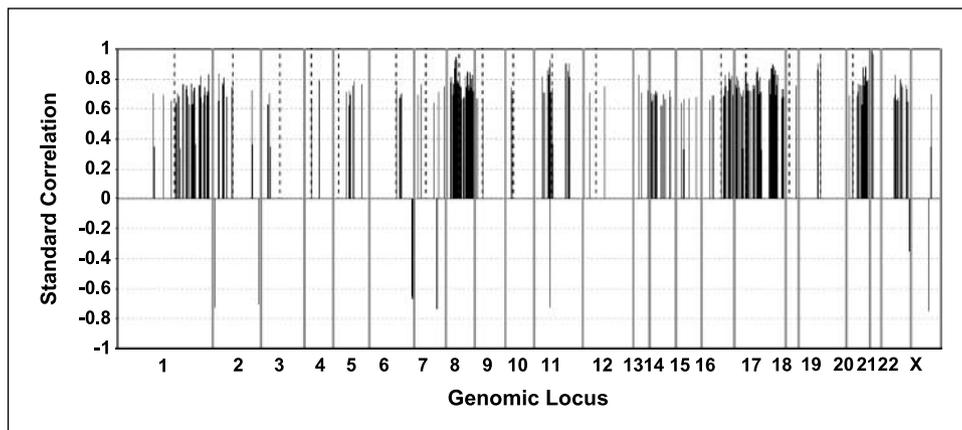


Fig. 3. Chromosomal localization of the most significant genome transcriptome correlations. The results of Spearman and Pearson correlations according to probe sets and BAC loci are plotted as a function of genome location from chromosome 1pter to chromosome Xq.

Table 2. Genes showing significant expression copy number correlations and localized within regions discriminating ERBB2 and luminal DCIS

Chr	Cytoband	Startlocus	Endlocus	GeneID	Correlation coefficient	Correlation in IDC
1	q21.1	144150231	144356105	POLR3C	0.64	+
1	q21.1	144150231	144356105	PEX11B	0.63	+
1	q21.3	150870525	152158474	INTS3	0.68	
1	q21.3	149888549	150060705	MRPL9	0.7	+
1	q23.1	154888337	155047911	MRPL24	0.76	+
1	q23.1	154888337	155047911	ISG20L2	0.63	+
1	q23.2	158452438	158743347	COPA	0.73	+
1	q23.2	158452438	158743347	NCSTN	0.66	+
1	q23.3	159044725	159381234	DEDD	0.75	+
1	q23.3	159514689	160713118	DUSP12	0.68	+
1	q23.3	159044725	159381234	PFDN2	0.66	+
1	q23.3	159381234	159514689	PPOX	0.76	+
1	q23.3	159381234	159514689	USP21	0.71	+
1	q23.3	159381234	159514689	B4GALT3	0.7	+
1	q23.3	159381234	159514689	NDUFS2	0.7	+
1	q24.1	165426500	165600956	POU2F1	0.62	
1	q24.3	169943186	170838266	PIGC	0.77	+
1	q24.3	169943186	170838266	KIAA0859	0.71	+
1	q25.1	173402605	173547236	KIAA0040	0.63	+
1	q25.1	173189878	173350442	CACYBP	0.62	+
1	q25.3	180754428	181784986	LAMC2	0.74	
1	q25.3	179011386	180565371	STX6	0.72	+
1	q32.1	204138581	204930767	IKBKE	0.82	
1	q32.1	203342491	203495694	RIPK5	0.75	
1	q32.1	202027575	202349129	ZC3H11A	0.76	
1	q32.1	205087762	205277242	C1orf116	0.67	
1	q32.1	203164295	203342491	RBBP5	0.75	
1	q32.1	203963475	204138581	NUCKS1	0.71	
1	q32.1	205277242	205431667	PFKFB2	0.67	
1	q32.3	210152024	210772193	NENF	0.64	+
1	q41	219896304	221353763	MIA3	0.69	
1	q42.11	221568617	223326349	NVL	0.74	+
1	q42.11	221568617	223326349	WDR26	0.74	+
1	q42.13	227197070	228112708	NUP133	0.72	+
1	q42.2	228271115	229195892	COG2	0.68	
1	q42.3	232853266	233928437	RBM34	0.71	
1	q42.3	232853266	233928437	TOMM20	0.72	
1	q42.3	232853266	233928437	ARID4B	0.65	
1	q43	234794031	243579104	HEATR1	0.83	+
8	p23.1	8827798	9852904	TNKS	0.82	+
8	p22	15734104	17364572	CNOT7	0.84	
8	p21.3	22230663	22882560	BIN3	0.78	+
8	p21.2	24851358	26612249	PPP2R2A	0.81	+
8	p21.2	24851358	26612249	BNIP3L	0.68	+
8	p21.1	28844096	29353654	KIF13B	0.69	
8	p12	36836067	38260917	GPR124	0.87	
8	p12	36836067	38260917	ASH2L	0.81	+
8	p12	36836067	38260917	BRF2	0.8	+
8	p12	36836067	38260917	LSM1	0.8	+
8	p12	36836067	38260917	BAG4	0.79	+
8	p12	36836067	38260917	DDHD2	0.75	+
8	p12	29507083	31056165	DCTN6	0.77	
8	p12	36836067	38260917	PROSC	0.69	+
8	p12	29507083	31056165	PPP2CB	0.64	
8	p12	38335306	38425958	WHSC1L1	0.95	+
8	p12	38335306	38425958	FGFR1	0.92	+
8	p11.23	38497241	40010617	ADAM3A	0.79	
8	p11.21	40185770	41825747	GINS4	0.79	
8	p11.21	40185770	41825747	GOLGA7	0.73	
8	p11.21	42044320	42283104	AP3M2	0.67	+
8	p11.21	42850482	43166289	FNTA	0.82	+
8	p11.21	42287374	42449080	VDAC3	0.78	+
8	p11.21	41825747	42044313	MYST3	0.77	+
8	q11.23	55029042	55683872	TCEA1	0.75	+

(Continued on the following page)

Table 2. Genes showing significant expression copy number correlations and localized within regions discriminating ERBB2 and luminal DCIS (Cont'd)

Chr	Cytoband	Startlocus	Endlocus	GeneID	Correlation coefficient	Correlation in IDC
8	q11.23	54844920	55029042	ATP6V1H	0.73	+
8	q12.1	58525613	59889959	LOC137886	0.74	
8	q12.1	61401378	61814156	RAB2A	0.66	
8	q21.11	74891351	76034750	TCEB1	0.67	
8	q21.13	81648505	82728066	PMP2	0.67	
8	q22.1	97062467	98198501	MTERFD1	0.81	+
8	q22.1	97062467	98198501	UQCRB	0.82	+
8	q22.1	97062467	98198501	PTDSS1	0.66	+
8	q22.1	95496349	95679222	RAD54B	0.77	+
8	q22.1	98198501	98375057	PGCP	0.74	
8	q22.2	98375057	99735538	HRSP12	0.74	+
8	q22.2	100848701	102004873	POLR2K	0.74	+
8	q22.2	100738763	100848701	VPS13B	0.7	
8	q22.3	103891151	104422933	ATP6V1C1	0.85	+
8	q22.3	103891151	104422933	AZIN1	0.7	+
8	q22.3	100848701	102004873	ANKRD46	0.7	
8	q22.3	102195771	103646239	ZNF706	0.84	
8	q23.2	110623681	110825964	EBAG9	0.76	+
8	q24.11	117853649	118046959	RAD21	0.67	+
8	q24.12	120401806	121248223	TAF2	0.73	+
8	q24.13	124908744	126133308	TRMT12	0.84	+
8	q24.13	124908744	126133308	RNF139	0.76	+
8	q24.13	123977707	124708761	DERL1	0.77	+
8	q24.13	124908744	126133308	SQLE	0.64	+
8	q24.13	123977707	124708761	C8orf32	0.65	+
8	q24.13	126133308	126334773	NSMCE2	0.74	
8	q24.22	131964334	133156759	KIAA0143	0.8	+
8	q24.3	143909082	145536611	SIAHBP1	0.85	+
8	q24.3	140031123	141603135	NIBP	0.83	
8	q24.3	143909082	145536611	EXOSC4	0.82	+
8	q24.3	143909082	145536611	CYC1	0.8	+
8	q24.3	143909082	145536611	GPAA1	0.73	+
8	q24.3	143909082	145536611	SHARPIN	0.72	
8	q24.3	146133391		C8orf33	0.72	+
8	q24.3	143909082	145536611	PYCRL	0.72	+
8	q24.3	143909082	145536611	GRINA	0.71	
8	q24.3	143909082	145536611	SCRIB	0.68	+
8	q24.3	143909082	145536611	PLEC1	0.68	
8	q24.3	143909082	145536611	ZNF696	0.68	+
8	q24.3	143909082	145536611	ZC3H3	0.66	+
8	q24.3	141603135	141808565	PTK2	0.79	+
8	q24.3	146029133	146133391	ZNF7	0.71	+
8	q24.3	143739886	143909082	C8orf55	0.69	
8	q24.3	146029133	146133391	ZNF16	0.66	+
16	q12.1	46027393	46605539	PHKB	0.69	+
16	q12.1	48832865	49022383	BRD7	0.74	
16	q13	56003087	56156112	COQ9	0.75	
16	q13	56003087	56156112	CIAPIN1	0.71	
16	q13	54836075	56003087	ARL2BP	0.68	
16	q21	56619483	57476831	CNOT1	0.83	+
16	q21	56619483	57476831	GOT2	0.71	+
16	q21	65020904	65209402	CKLF	0.64	
16	q22.1	65209402	65623027	APPBP1	0.75	
16	q22.1	65209402	65623027	FAM96B	0.65	+
16	q22.1	67478745	69245806	AARS	0.72	+
16	q22.1	67478745	69245806	DDX19A	0.72	+
16	q22.1	67478745	69245806	SF3B3	0.72	+
16	q22.1	67478745	69245806	WWP2	0.67	+
16	q22.1	69245806	69419474	VAC14	0.7	+
16	q22.3	69419474	71533759	KIAA0174	0.85	+
16	q22.3	69419474	71533759	DHX38	0.71	+
16	q22.3	71533759	74776431	PSMD7	0.8	+
16	q22.3	71533759	74776431	RFWD3	0.72	+
16	q23.1	71533759	74776431	KARS	0.79	+

(Continued on the following page)

Table 2. Genes showing significant expression copy number correlations and localized within regions discriminating ERBB2 and luminal DCIS (Cont'd)

Chr	Cytoband	Startlocus	Endlocus	GeneID	Correlation coefficient	Correlation in IDC
16	q23.1	71533759	74776431	ADAT1	0.73	
16	q23.1	71533759	74776431	WDR59	0.68	
16	q23.1	71533759	74776431	TERF2IP	0.64	+
16	q23.2	79052806	80639519	ASCIZ	0.72	+
16	q24.1	83335109	84444005	COX4NB	0.82	+
16	q24.1	83335109	84444005	KIAA0182	0.71	
16	q24.1	83335109	84444005	COX4I1	0.69	+
16	q24.1	83335109	84444005	ZDHHC7	0.69	+
17	p13.3	2077107	3789432	ITGAE	0.8	
17	q11.2	24125209	25739938	ERAL1	0.66	
17	q11.2	24125209	25739938	FLOT2	0.75	
17	q11.2	22465093	23921490	IFT20	0.71	
17	q11.2	23991111	24125209	RPL23A	0.77	
17	q11.2	23926892	24057159	SUPT6H	0.76	
17	q11.2	27772588	27941694	PSMD11	0.76	+
17	q11.2	23991111	24125209	C17orf63	0.76	
17	q11.2	23944011	24039852	SPAG5	0.75	
17	q11.2	23926892	24057159	SDF2	0.68	
17	q11.2	26430455	26574472	NF1	0.72	
17	q11.2	23991111	24125209	TRAF4	0.63	
17	q12	34434812	34987922	CRKRS	0.82	
17	q12	34434812	34987922	RPL19	0.85	
17	q12	34434812	34987922	PPARBP	0.71	
17	q12	34987922	35196413	GRB7	0.87	+
17	q12	35246223	35393977	GSDML	0.87	+
17	q12	34987922	35196413	PERLD1	0.86	+
17	q12	34987922	35196413	STARD3	0.82	+
17	q12	34987922	35196413	ERBB2	0.82	+
17	q12	35246223	35393977	PSMD3	0.81	+
17	q12	30384409	30599689	NLE1	0.76	+
17	q12	34201212	34324738	CCDC49	0.75	
17	q12	30270215	30599689	LIG3	0.73	+
17	q12	34201212	34324738	RPL23	0.71	
17	q12	34201212	34324738	PIP5K2B	0.68	
17	q12	34987922	35196413	PNMT	0.65	+
17	q21.1	35465293	35659097	CASC3	0.66	+
17	q21.1	35465293	35659097	THRA	0.66	
17	q21.2	35920386	37443438	SMARCE1	0.66	+
17	q21.2	35682074	35839525	CDC6	0.81	
17	q21.2	35682074	35839525	WIPF2	0.75	+
17	q21.2	35737736	35915957	RARA	0.71	
17	q21.2	35682074	35839525	TOP2A	0.7	

NOTE: The genes are ordered by chromosomes and are located in regions differentially gained or lost between ERBB2 and luminal DCIS (listed in Supplementary Data 3). The correlation coefficients (Pearson and Spearman) were considered significant when higher than 0.6 with a Benjamini-Hochberg corrected *P* value lower than 0.05. In case of multiple probe sets corresponding to one gene, the highest correlation coefficient is indicated. Similarly, for probe sets that matched with two flanking (backward and forward) BACs, the highest correlation coefficient is shown. The last column indicates the genes exhibiting expression copy number correlations in infiltrating ductal carcinomas (28). The complete list of genes exhibiting significant expression copy number correlations is provided in Supplementary Data 6.

DCIS can be classified into the luminal, ERBB2, and basal-like groups previously described in IDC (21, 22). In our series of 57 DCIS, most tumors (56%) belonged to the luminal group based on ER expression and the absence of ERBB2 overexpression. The second group of tumors comprised ERBB2 overexpressing DCIS (37%) that were mostly high grade (81%) and ER negative/PR negative (90%) in agreement with previous reports showing that ERBB2 overexpression is rare in ER-positive DCIS (37). A third small group was composed of three triple-negative DCIS (5%) with only one case demonstrating a clear basal-like immunophenotype (2%) with positive staining for keratin 5/6 and epidermal growth factor receptor, confirm-

ing that the recently recognized basal-like DCIS group is a rare entity among pure DCIS (6-10% of cases; refs. 25, 26, 38, 39). This was also shown by a recent gene expression profile analysis of a series of DCIS (18).

When compared with the luminal group, the ERBB2-amplified cases presented a particular array CGH profile with more frequent 17q gains, 3p, 4p, 4q, and 8p losses. In contrast, the luminal DCIS group harbored significantly more frequent 16q losses. These observations refine previously published classic CGH data showing a significantly higher number of 17q gains in high nuclear grade DCIS (which correspond to most cases of our ERBB2-amplified, ER-negative group) and 16q

losses in low-grade DCIS (which are represented in the present series by the luminal cases; refs. 4, 7, 11).

We also show that DCIS can be classified according to the genomic criteria recently proposed for IDC (28), as a first group of 34 tumors presented high-level amplification of genomic regions (amplifiers). All *TP53* mutations were observed in this group. Remarkably, the recurrent amplicons of DCIS were located on 8p12 (*FGFR1*), 11q12 (*Cyclin D1*), 17q12 (*ERBB2*), and 20q13 (*BMP7/BCAS1*), corresponding also to the most frequent amplicons of IDC (28). Compared with other amplifier cases and as recently observed in IDC (28), half amplicons associated with *ERBB2* were also located on chromosome 17q. In contrast, in the luminal group, only 17% of amplicons involved 17q. A second group comprised tumors with a 16q deletion associated with 1q gain but no additional CNA. All these cases presented a luminal phenotype. The last group comprised cases with more complex CNA. In this series, the three triple-negative DCIS did not present any recurrent genomic profile, but two were amplifiers including the *bona fide* basal-like DCIS, which also harbored a *TP53* gene mutation.

Unsupervised analysis of gene expression profiles supports the observation that *ERBB2*-amplified and luminal DCIS constitute two distinct entities. Gene ontology analysis of the *ERBB2* DCIS signature showed that the genes up-regulated in this group belonged to *ERBB2* signal transduction pathways but also to fatty acid metabolism and other protein kinase and transmembrane signaling pathways. Genes down-regulated mainly belonged to apoptotic and ER regulation pathways. The list of genes differentially expressed between *ERBB2* and luminal DCIS in this series was compared with the recently published list of genes that discriminate DCIS according to their differentiation status (18). Although the two comparisons were based on slightly different criteria, they both highlighted major genes involved in breast carcinogenesis such as *ESR1*, *GRB7*, and *ERBB2*. A link between the *ERBB2* pathway and fatty acid metabolism has been previously reported in cell lines (40) and invasive breast carcinomas (41). The present study also showed that these pathways are activated right from the earliest stages of *ERBB2*-driven breast carcinogenesis. Interestingly, as reported in IDC (42, 43), overexpression of 17q12 genes *STARD3*, *GRB7*, and *PERLD1* is also associated with *ERBB2* overexpression in DCIS, which raises the question of their putative synergistic role during early breast carcinogenesis. Finally, *Integrin $\beta 6$* , a receptor for fibronectin and cytotactin, is overexpressed in the *ERBB2* DCIS group. Recent data have highlighted molecular interactions between *ERBB2* receptors and integrin networks (41, 44). Moreover, experiments conducted with *ITGB6* knockout mice indicate that *ITGB6* may regulate transforming growth factor β and metalloprotease, in particular matrix metalloproteinase 12 expression (45). Further investigations are needed to understand the role of *ITGB6* in *ERBB2*-amplified DCIS.

It is noteworthy that genes recently involved in angiogenesis or invasion such as *SDC1* (*syndecan*; refs. 46–48) or *CXCR4* are up-regulated in *ERBB2*-positive *in situ* lesions (49). Reciprocally, *FGFR2*, a gene recently linked to a susceptibility locus in breast cancer (50, 51), is up-regulated in luminal DCIS.

The correlation between array CGH and expression data shows that amplifications but also more subtle copy number changes lead to modifications of the gene expression profile.

Strikingly, a large part of the genes, for which a strong correlation between expression level and genomic status was observed, also exhibited a similar correlation in IDC (28), indicating that genome alterations have similar consequences on gene expression in both IDC and DCIS. In line with the data reported by Chin et al. (28), these results identify genes located in frequently altered regions in breast cancer, strongly influenced by copy number variations, and whose gene dosage effects may play a role in early phases of breast carcinogenesis. Two genes on chromosome 16 seem to be of particular interest: *ASCIZ* (*ATM/ATR-substrate ChK2-interacting Zinc finger*; ref. 52), which encodes a protein involved in lesion-specific Rad51 focus formation, a process that also involves *BRCA1* and *BRCA2*, two master genes in breast carcinogenesis; and *CNOT1*, which encodes a member of the *CCR4-NOT* complex, a global regulator of transcription that may act as a repressor of endogenous estrogens target genes (53). Interestingly, *CNOT7*, another member of the *CCR4-NOT* complex, is regulated by gene copy number variation on chromosome 8p, a chromosome frequently altered in *ERBB2* DCIS tumors. Based on these observation, the hypothesis of haploinsufficiency of the *CCR4-NOT* complex playing a role in early breast tumorigenesis may now be tested.

Altogether, these integrated genomic and transcriptomic data confirm previous results showing strong similarities between IDC and DCIS (20, 54). Although we cannot exclude the possibility that subtle genetic alterations that may have escaped detection in this study could also play a role in the evolution from DCIS to IDC, this study raises the hypothesis that the difference between DCIS and IDC is not based on specific genetic alterations or on their consequences on gene expression profile. Because a recent report suggested that a 35-gene expression classifier might distinguish IDC and DCIS (18), the role of epigenetic modifications in the progression from DCIS to IDC will need to be evaluated.

In conclusion, our data show that DCIS already displays the molecular diversity observed in IDC and, therefore, can be classified according to molecular criteria distinguishing *ERBB2*-amplified DCIS, usually high-grade, ER negative with frequent *TP53* mutations, from luminal DCIS corresponding to low/intermediate-grade, ER positive with a very low rate of *TP53* mutations. In our series, only three cases were classified as triple negative and only one was classified as basal-like DCIS, which confirms that this last entity is rare among DCIS. Further studies are needed to define whether a classification of DCIS based on molecular markers may help to more accurately define cases associated with a higher risk of recurrence. And finally, genomic/transcriptomic correlations represent a promising tool to identify new genes and pathways important in early breast carcinogenesis.

Acknowledgments

We thank Dr Paul Fréneaux for his constant support during our study, Céline Grassart for her excellent technical assistance, and Ingrid Lebigot and her team from the Institut Curie's tumor bank.

Members of the Breast Cancer Study group in addition to coauthors: Bernard Asselain, Alain Aurias, Emmanuel Barillot, Marc Bollet, François Campana, Paul Cottu, Patricia de Cremoux, Véronique Diéras, Laurent Mignot, Jean-Yves Pierga, Marie-France Poupon, Dominique Stoppa-Lyonnet, Anne Tardivon, Fabienne Thibault, and Pascale This.

References

1. Burstein HJ, Polyak K, Wong JS, et al. Ductal carcinoma *in situ* of the breast. *N Engl J Med* 2004;350:1430–41.
2. Recht A, Rutgers EJ, Fentiman IS, et al. The fourth eortc DCIS Consensus Meeting (Chateau Marquette, Heemskerk, the Netherlands, 23-24 January 1998)–Conference Report. *Eur J Cancer* 1998;34:1664–9.
3. Sneige N, Lagios MD, Schwarting R, et al. Interobserver reproducibility of the Lagios nuclear grading system for ductal carcinoma *in situ*. *Hum Pathol* 1999;30:257–62.
4. O'Connell P, Pekkel W, Fuqua S, et al. Analysis of loss of heterozygosity in 399 premalignant breast lesions at 15 genetic loci. *J Natl Cancer Inst* 1998;90:697–703.
5. Simpson JF, Page DL. The p53 tumor suppressor gene in ductal carcinoma *in situ* of the breast. *Am J Pathol* 2000;156:5–6.
6. Van de Vijver MJ, Peterse JL, Mooi WJ, et al. Neu-protein overexpression in breast cancer. Association with comedo-type ductal carcinoma *in situ* and limited prognostic value in stage II breast cancer. *N Engl J Med* 1988;319:1239–45.
7. Hwang ES, DeVries S, Chew KL, et al. Patterns of chromosomal alterations in breast ductal carcinoma *in situ*. *Clin Cancer Res* 2004;10:5160–7.
8. Done S, Eskandarian S, Bull S, et al. p53 missense mutations in microdissected high-grade ductal carcinoma *in situ* of the breast. *J Natl Cancer Inst* 2001;93:700–4.
9. Buerger H, Mommers E, Littmann R, et al. Correlation of morphologic and cytogenetic parameters of genetic instability with chromosomal alterations in *in situ* carcinomas of the breast. *Am J Clin Pathol* 2000;114:854–9.
10. Zafrani B, Leroyer A, Fourquet A, et al. Mammographically-detected ductal *in situ* carcinoma of the breast analyzed with a new classification. A study of 127 cases: correlation with estrogen and progesterone receptors, p53 and c-erbB-2 proteins, and proliferative activity. *Semin Diagn Pathol* 1994;11:208–14.
11. Buerger H, Otterbach F, Simon R, et al. Comparative genomic hybridization of ductal carcinoma *in situ* of the breast—evidence of multiple genetic pathways. *J Pathol* 1999;187:396–402.
12. Quinn CM, Ostrowski JL, Harkins L, et al. Loss of bcl-2 expression in ductal carcinoma *in situ* of the breast relates to poor histological differentiation and to expression of p53 and c-erbB-2 proteins [in Process Citation]. *Histopathology* 1998;33:531–6.
13. Leonard GD, Swain SM. Ductal carcinoma *in situ*, complexities and challenges. *J Natl Cancer Inst* 2004;96:906–20.
14. Sanders ME, Schuyler PA, Dupont WD, et al. The natural history of low-grade ductal carcinoma *in situ* of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. *Cancer* 2005;103:2481–4.
15. Wong JS, Kaelin CM, Troyan SL, et al. Prospective study of wide excision alone for ductal carcinoma *in situ* of the breast. *J Clin Oncol* 2006;24:1031–6.
16. Kepple J, Henry-Tillman RS, Klimberg VS, et al. The receptor expression pattern in ductal carcinoma *in situ* predicts recurrence. *Am J Surg* 2006;192:68–71.
17. Provenzano E, Hopper JL, Giles GG, et al. Biological markers that predict clinical recurrence in ductal carcinoma *in situ* of the breast. *Eur J Cancer* 2003;39:622–30.
18. Hannemann J, Velds A, Halfwerk JB, et al. Classification of ductal carcinoma *in situ* by gene expression profiling. *Breast Cancer Res* 2006;8:R61.
19. Adeyinka A, Emberley E, Niu Y, et al. Analysis of gene expression in ductal carcinoma *in situ* of the breast. *Clin Cancer Res* 2002;8:3788–95.
20. Ma X-J, Salunga R, Tuggle JT, et al. Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci U S A* 2003;100:5974–9.
21. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747–52.
22. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869–74.
23. Nielsen TO, Hsu FD, Jensen K, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* 2004;10:5367–74.
24. van de Rijn M, Perou CM, Tibshirani R, et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. *Am J Pathol* 2002;161:1991–6.
25. Bryan BB, Schnitt SJ, Collins LC. Ductal carcinoma *in situ* with basal-like phenotype: a possible precursor to invasive basal-like breast cancer. *Mod Pathol* 2006;19:617–21.
26. Dabbs DJ, Chivukula M, Carter G, et al. Basal phenotype of ductal carcinoma *in situ*: recognition and immunohistologic profile. *Mod Pathol* 2006;19:1506–11.
27. Livasy CA, Karaca G, Nanda R, et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod Pathol* 2006;19:264–71.
28. Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* 2006;10:529–41.
29. Fridlyand J, Snijders AM, Ylstra B, et al. Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* 2006;6:96.
30. Reyat F, Stransky N, Bernard-Pierrot I, et al. Visualizing chromosomes as transcriptome correlation maps: evidence of chromosomal domains containing co-expressed genes—a study of 130 invasive ductal breast carcinomas. *Cancer Res* 2005;65:1376–83.
31. Azoulay S, Lae M, Freneaux P, et al. KIT is highly expressed in adenoid cystic carcinoma of the breast, a basal-like carcinoma associated with a favorable outcome. *Mod Pathol* 2005;18:1623–31.
32. Vincent-Salomon A, Gruel N, Lucchesi C, et al. Identification of typical medullary breast carcinoma as a genomic sub-group of basal-like carcinomas, a heterogeneous new molecular entity. *Breast Cancer Res* 2007;9:R24.
33. Hupe P, Stransky N, Thiery JP, et al. Analysis of array CGH data: from signal ratio to gain and loss of dna regions. *Bioinformatics* 2004;20:3413–22.
34. La Rosa P, Viara E, Hupe P, et al. VAMP: visualization and analysis of array-CGH, transcriptome and other molecular profiles. *Bioinformatics* 2006;22:2066–73.
35. Li C, Wang WH. DNA-Chip Analyzer (dChip). In: "The Analysis of Gene Expression Data: Methods and Software". New York: Springer.
36. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98:5116–21.
37. Collins LC, Schnitt SJ. HER2 protein overexpression in estrogen receptor-positive ductal carcinoma *in situ* of the breast: frequency and implications for tamoxifen therapy. *Mod Pathol* 2005;18:615–20.
38. Livasy CA, Perou CM, Karaca G, et al. Identification of a basal-like subtype of breast ductal carcinoma *in situ*. *Hum Pathol* 2007;38:197–204.
39. Paredes J, Lopes N, Milanezi F, et al. P-cadherin and cytokeratin 5: useful adjunct markers to distinguish basal-like ductal carcinomas *in situ*. *Virchows Arch* 2007;450:73–80.
40. Kumar-Sinha C, Ignatowski KW, Lippman ME, et al. Transcriptome analysis of HER2 reveals a molecular connection to fatty acid synthesis. *Cancer Res* 2003;63:132–9.
41. Bertucci F, Borie N, Ginestier C, et al. Identification and validation of an ERBB2 gene expression signature in breast cancers. *Oncogene* 2004;23:2564–75.
42. Kauraniemi P, Kallioniemi A. Activation of multiple cancer-associated genes at the ERBB2 amplification in breast cancer. *Endocr Relat Cancer* 2006;13:39–49.
43. Kauraniemi P, Kuukasjärvi T, Sauter G, et al. Amplification of a 280-kilobase core region at the ERBB2 locus leads to activation of two hypothetical proteins in breast cancer. *Am J Pathol* 2003;163:1979–84.
44. Guo W, Pylayeva Y, Pepe A, et al. $\beta 4$ integrin amplifies ERBB2 signaling to promote mammary tumorigenesis. *Cell* 2006;126:489–502.
45. Morris DG, Huang X, Kaminski N, et al. Loss of integrin $\alpha(v)\beta 6$ -mediated TGF- β activation causes Mmp12-dependent emphysema. *Nature* 2003;422:169–73.
46. Maeda T, Desouky J, Friedl A. Syndecan-1 expression by stromal fibroblasts promotes breast carcinoma growth *in vivo* and stimulates tumor angiogenesis. *Oncogene* 2006;25:1408–12.
47. Baba F, Swartz K, van Buren R, et al. Syndecan-1 and syndecan-4 are overexpressed in an estrogen receptor-negative, highly proliferative breast carcinoma subtype. *Breast Cancer Res Treat* 2006;98:91–8.
48. Gotte M, Kersting C, Radke I, et al. An expression signature of syndecan-1 (cd138), E-cadherin and c-met is associated with factors of angiogenesis and lymphangiogenesis in ductal breast carcinoma *in situ*. *Breast Cancer Res* 2007;9:R8.
49. Li YM, Pan Y, Wei Y, et al. Upregulation of CXCR4 is essential for HER2-mediated tumor metastasis. *Cancer Cell* 2004;6:459–69.
50. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;447:1087–93.
51. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 2007;39:870–4.
52. McNeese CJ, Conlan LA, Tennis N, et al. ASCIZ regulates lesion-specific rad51 focus formation and apoptosis after methylating DNA damage. *EMBO J* 2005;24:2447–57.
53. Winkler GS, Mulder KW, Bardwell VJ, et al. Human Ccr4-Not complex is a ligand-dependent repressor of nuclear receptor-mediated transcription. *EMBO J* 2006;25:3089–99.
54. Porter D, Lahti-Domenici J, Keshaviah A, et al. Molecular markers in ductal carcinoma *in situ* of the breast. *Mol Cancer Res* 2003;1:362–75.