Lobular and ductal carcinomas of the breast have distinct genomic and expression profiles

François Bertucci [1], Béatrice Orsetti [2], Vincent Nègre [2], Pascal Finetti [1], Carole Rougé [2], Jean-Charles Ahomadegbe [3], Frédéric Bibeau [2 4], Marie-Christine Mathieu [5], Isabelle Treilleux [6], Jocelyne Jacquemier [1], Lisa Ursule [2], Agnès Martinec [7], Qing Wang [8], Jean Bénard [5 9], Alain Puisieux [8 10], Daniel Birnbaum [1], Charles Theillet [2 *]

[1] Cancérologie INSERM : U599 , Institut Paoli Calmettes , Université de la Méditerranée - Aix-Marseille II , Institut Paoli-Calmettes 27, Boulevard Lei Roure 13009 MARSEILLE,FR

[2] IRCM, Institut de recherche en cancérologie de Montpellier INSERM : U896 , Université Montpellier I , CRLC Val d'Aurelle - Paul Lamarque F-34298 Montpellier,FR

[3] UPRES 3535 Institut Gustave Roussy , Université Paris Sud - Paris XI , Villejuif,FR

[4] Laboratoire d'anatomopathologie CRLC Val d'Aurelle-Paul Larmarque , FR

[5] Département de Biopathologie Institut Gustave Roussy, , Villejuif,FR

[6] Laboratoire d'Anatomopathologie Centre Léon Bérard , Lyon,FR

[7] Ipsogen S.A. Ipsogen S.A. , Marseille-Luminy,FR

[8] Laboratoire d'Oncologie Moléculaire Centre Léon Bérard , Lyon,FR

[9] IMC, Interactions moléculaires et cancer CNRS : UMR8126 , Institut Gustave Roussy , Université Paris Sud - Paris XI , pavillon de recherche 1 39 Rue Camille Desmoulins 94805 VILLEJUIF CEDEX,FR

[10] Oncogénèse et progression tumorale INSERM : U590 , Université Claude Bernard - Lyon I , CRLCC Léon Bérard , 28, Rue Laennec 69373 Lyon Cedex 08,FR

* Correspondence should be adressed to: Charles Theillet <charles.theillet@valdorel.fnclcc.fr >

Abstract

Invasive ductal carcinomas (IDCs) and invasive lobular carcinomas (ILCs) are the two major pathological types of breast cancer. Epidemiological and histoclinical data suggest biological differences, but little is known about the molecular alterations involved in ILCs. We undertook a comparative large-scale study by both array-CGH and cDNA microarray of a set of 50 breast tumors (21 classic ILCs and 29 IDCs) selected on homogeneous histoclinical criteria. Results were validated on independent tumor sets, as well as by quantitative RT-PCR. ILCs and IDCs presented differences at both the genomic and expression levels with ILCs being less rearranged and heterogeneous than IDCs. Supervised analysis defined a 75-BACs signature discriminating accurately ILCs from IDCs. Expression profiles identified two subgroups of ILCs: typical ILCs (~50%), which were homogeneous and displayed a normal-like molecular pattern, and atypical ILCs, more heterogeneous with features intermediate between ILCs and IDCs. Supervised analysis identified a 75-gene expression signature that discriminated ILCs from IDCs, with many genes involved in cell adhesion, motility, apoptosis, protein folding, extracellular matrix, and protein phosphorylation. Although ILCs and IDCs share common alterations, our data show that ILCs and IDCs could be distinguished on the basis of their genomic and expression profiles suggesting that they evolve along distinct genetic pathways.

MESH Keywords Breast Neoplasms ; genetics ; metabolism ; pathology ; Cadherins ; genetics ; metabolism ; Carcinoma, Ductal, Breast ; genetics ; metabolism ; pathology ; Carcinoma, Lobular ; genetics ; metabolism ; pathology ; Chromosomes, Artificial, Bacterial ; Female ; Gene Expression Profiling ; Gene Expression Regulation, Neoplastic ; Humans ; Mutation ; genetics ; Nucleic Acid Hybridization ; Oligonucleotide Array Sequence Analysis ; RNA, Messenger ; genetics ; metabolism ; RNA, Neoplasm ; genetics ; metabolism ; Reverse Transcriptase Polymerase Chain Reaction ; Tumor Suppressor Protein p53 ; genetics

Author Keywords breast cancer ; DNA microarray ; genetic profiles ; array-CGH

Introduction

Breast cancer is a complex and heterogeneous disease, which, despite important efforts, remains difficult to describe comprehensively and, therefore, to treat appropriately. Up to 20 pathological types have been defined, but two of them, invasive ductal (IDCs) and invasive lobular carcinomas (ILCs), account for about 90% of all breast tumors. Median incidence of ILCs is about 12% and increases disproportionately compared to IDCs in western countries (Li et al., 2003). ILCs and IDCs differ from each other with respect to various histological, biological and clinical features. Remarkably ILCs are less cohesive than IDCs and tend to form single files of invading cells. This feature has been associated with the frequent inactivation of the E-cadherin gene (CDH1) (Berx et al., 1995). ILCs are predominantly estrogen receptor (ER), and progesterone receptor (PR) positive, and thus presumably more homogeneous than IDCs. Their pathological grade is generally lower than that of IDCs and they show a lower proliferation index (Sastre-Garau et al., 1996). ILCs are less sensitive to chemotherapy (Katz et al., 2007) and are more prone to form bone, gastrointestinal, peritoneal and ovarian metastases than IDCs (Lamovec & Bracko, 1991). Despite these differences, ILCs show similar prognoses as IDCs (Toikkanen et al., 1997), and the treatment

of ILCs and IDCs is similar. Patients would benefit from a better tailored treatment. Therefore, it appears crucial to gain insight in the molecular differences that distinguish the two pathological types.

There are a number of reasons to suspect that ILCs and IDCs represent distinct molecular entities. Cytogenetic-based studies have suggested that they differ at the karyotype level, with ILCs being specified by a combination of gains at 1q and losses at 16q (Flagiello et al., 1998). However, chromosomal CGH-based studies have shown contradictory results (Gunther et al., 2001 ; Loveday et al., 2000), and only two studies based on array-CGH have compared ILCs and IDCs (Loo et al., 2004 ; Stange et al., 2006). Expression profiling studies have revealed the transcriptional heterogeneity and new molecular subtypes of breast cancer, but these studies were mainly performed on IDCs (Bertucci et al., 2006). Three studies (Korkola et al., 2003 ; Turashvili et al., 2007 ; Zhao et al., 2004) reported expression signatures that distinguish IDCs from ILCs with reasonable accuracy. However, save for CDH1 , the gene sets generated in either study show little overlap.

No comprehensive genomic and transcriptomic study comparing ILCs and IDCs has been reported yet. Because breast cancer is heterogeneous and different phenotypes may possibly intermingle making the comparisons delicate, we reasoned that working with stringently-defined tumor sets could prove crucial to establish clear cut genetic differences between IDCs and ILCs. We thus constituted a tumor training set selected on homogeneous and focused phenotypic criteria, comprising 21 classic ILCs and 29 IDCs. Molecular profiles were determined at the DNA and RNA levels using microarrays. Tumors were also analyzed for the presence of TP53 and CDH1 mutations. Our data support the idea that the two major histological types of breast cancer arise along distinct genetic routes.

Results

Phenotypic characteristics of the tumor training set

In order to limit the heterogeneity of the analyzed tumor set and avoid its dispersion in smaller entities we worked on a selected tumor collection. Our aim was to compare matched sets of tumors and because ILCs are predominantly grade 2 and hormone receptor-positive, we preferentially selected grade 2, pT2, ER+, invasive tumors with less than 3 involved axillary lymph nodes. A total of 21 ILCs and 29 IDCs were selected after cross-checking by four pathologists. All ILCs were of the classic subtype and voluntarily excluded other ILC subtypes, thus restricting our study to a subset of lobular cancers.

The 50 tumor samples were analyzed at both the genomic (array-CGH) and expression (cDNA microarrays) levels and for the presence of mutations in CDH1 and exons 4 through 10 of TP53 . Although some mutations may have been lost in our analysis, we detected 6 tumors with TP53 mutations and 13 with CDH1 mutations. TP53 mutations were restricted to the IDCs and CDH1 mutations to the ILCs. It must be mentioned that in addition to mutations and loss, CDH1 may be inactivated by methylation. Immunohistochemical study of E-cadherin expression in a subset of 33 tumors (14 ILCs and 19 IDCs) showed negative staining in 16 cases (12 ILCs and 4 IDCs), whereas 17 tumors (15 IDCs and 2 ILCs) were positive (Supplementary Table S1 ; $p=3\ 10^{-4}$).

Array-CGH profiling

Gains and losses in ILC and IDC

Genome-wide array-CGH analysis identified copy number changes (CNC) in all but one tumor of the training set. Genomic imbalances were more frequent in IDCs than ILCs (17.4% vs . 11% of the BACs showing CNC, p=0.004) (Figure 1A–B). The two pathological types shared common aberrations, with frequent (occurrence > 20%) gains and some peaks (>40%) at 1q41-q43, 8q13 and 8q24, 16p13, 17q23 and 20q13. Frequent losses exceeded 20% occurrence and were found at 6q, 8p, distal 11q, 13, 16q. However, differences between IDCs and ILCs were apparent and could be visualized on frequency difference plots (Figure 1C–D). In IDCs, most prevalent CNCs were gains at 8q, 16p, 17q and 20q, and losses at 3p, 4q, 7p, 8p, 15q, 18q and X. In ILCs, the most prevalent changes were gains at 1q, 7p12, 11q13, 16p13, Xp11, and losses at 11q21-qter, 13, 17q and 22. DNA amplification at 11q13 was evenly distributed throughout ILCs and IDCs (40% and 20–33% at CCND1 and PAK1 , respectively), whereas that at 17q24.1, (THRAP1 and SMURF2) was restricted to IDCs (37%). A twosample Wilcoxon test identified 114 BACs differently involved in ILCs and IDCs (Figure 1E).

Copy number profiles may be used to stratify breast cancers in three groups referred to as simplex, complex and amplifier (Fridlyand et al., 2006 ; Hicks et al., 2006). Simplex profiles are characterized by infrequent gains or losses involving whole chromosomal arms, complex by highly rearranged patterns involving multiple regions of gains and losses and infrequent amplification, and amplifier by high-level amplification associated to moderately rearranged patterns. We found simplex, complex or amplifiers in both IDCs and ILCs (see Supplementary Table 1). However, simplex tumors were more frequent in ILCs (47.6%) than IDCs (31%; difference not significant).

Genomic imbalances discriminating ILC from IDC and definition of a genomic classifier

To identify regions of CNC that discriminate ILCs from IDCs, we applied a supervised analysis based on a combination of signal-to-noise (S2N) and support vector machine (SVM). S2N was used to select differential features, SVM to classify tumors, and LOOCV (leave-one-out cross-validation) to estimate the performance of the classifier. By LOOCV, 43 tumors of the training set were

correctly classified (86% overall accuracy; Figure 2), with 25/29 (86.3%) for IDCs and 18/21 (85.7%) for ILCs. Most IDCs bearing a TP53 mutation (5/6) were classified as IDC. Only 2/13 ILCs with a CDH1 mutation were misclassified as IDC (Figure 2A). The retained genomic signature corresponded to 75 BACs identified in 50/50 iterations of the LOOCV procedure (Table 1). These BACs were located on 16 chromosomes with largest clusters at 1q32.1-q42.3, 15q11.2-q22.2, 17q23.2-q24.3 and 20q11.21-q13.33.

We used this 75-BACs signature to classify the tumors by hierarchical clustering, producing two major clusters strongly correlated with the pathological type (Figure 2B), with IDCs predominantly found in cluster I and ILCs in cluster II. We next tested the relevance of our 75-BACs signature on an independent validation group of 23 grade 2 tumors. Eighteen of 23 tumors were correctly classified resulting in an overall accuracy of 78% ranging from 75% for IDCs to 85.7% for ILCs (Table 2).

Gene expression profiling

Tumors were profiled using cDNA microarrays comprising 5407 genes and 2898 ESTs.

Global transcriptional profiles

Unsupervised hierarchical clustering was applied to the 7782 genes/ESTs showing significant variation in expression levels across the 50 samples of the training set (present in at least 80% of the samples with standard deviation >0.1). As reflected by the dendrogram, the tumors displayed heterogeneous expression profiles (Figure 3A–B), and were sorted into two major groups showing differential pathological type distribution. Whereas ILCs were predominantly found in group II (18/21 ILCs clustered in this group), IDCs distributed more evenly with 16/29 IDCs in group I and 13/29 in group II. Interestingly, group II subdivided in two subgroups (IIa and IIb) comprising 17 and 14 tumors respectively. While group IIa was almost evenly composed of ILCs (8/17) and IDCs (9/17), group IIb comprised 10 out of 14 ILCs. These results suggested a split in the ILC population, with a fraction (subgroup IIb) being more homogeneous than those in subgroup IIa. By reference to Zhao et al (Korkola et al., 2003 ; Turashvili et al., 2007 ; Zhao et al., 2004), we defined ILCs from subgroup IIb as typical ILCs, whereas those clustering in subgroups I and IIa corresponded to atypical (or IDC-like) ILCs. Noticeably, there was no difference in the incidence of CDH1 mutation in typical and atypical ILCs (Supplementary Table 1).

Several clusters of genes were evidenced corresponding to specific cell types or pathways (Figure 3A). These gene clusters were differentially expressed in the three subgroups. Striking features of ILCs, notably in subgroup IIb, were low levels of expression of the proliferation and luminal clusters and relatively high expression of the adipose cluster. Moreover, all ILCs, from subgroup IIa or IIb, displayed low expression of the ERBB2 and the CDH1 clusters. CDH1 mRNA expression levels correlated well with CDH1 IHC status (Figure 3B). We did not identify any correlation between the typical vs atypical character of ILCs and the following histoclinical features: age of patients, morphology, pathological tumor size, CDH1 IHC and mutation status. However, it was interesting to see that 3/11 (27%) patients with atypical ILC displayed a relapse vs 1/10 (10%) patients with typical ILC. It is of note that follow up time was equivalent in both typical and atypical ILCs (>72 months). We then analyzed the distribution of our tumor set according to the molecular subtypes (luminal A, luminal B, basal, ERBB2 **+**, and normal-like) identified by Sorlie and coworkers (Sorlie et al., 2001) in IDCs. These subtypes were defined on the basis of ~500 “intrinsic genes” of which 169 were common to our gene set. Based on these genes and the Sorlie and coworkers' samples (Sorlie et al., 2003), we defined five sets of centroids representing the average expression of each subtype. By measuring the correlation of each of our 50 samples with each centroid (Supplementary section), we assigned each tumor to a molecular subtype (Figure 3B ; Supplementary Table 1). IDCs and ILCs were differently distributed in the 5 molecular subtypes ($p=0.04$, Fisher exact test). ILCs presented no luminal B, a smaller proportion of luminal A (5 cases), basal (1 case) and ERBB2 (1 case), and an increase in normallike subtype (8 cases). Interestingly, 7/10 ILCs from subgroup IIb were of the normal-like subtype, while ILCs from subgroup IIa and I distributed in the 5 subtypes. This confirms that ILCs are less heterogeneous than IDCs and can be split into two subsets, one homogeneous, predominantly of the normal-like subtype, and the other, more diverse in terms of molecular subtypes, presenting IDC-like features.

Comparison of ILCs and IDCs

The same supervised approach as for array-CGH (combining signal-to-noise and support vector machine) identified a set of genes discriminating ILCs and IDCs. Carried out on the tumor training set, it resulted in an accurate segregation of 29/29 (100%) IDCs and 17/21 ILCs (81%) (Figure 4A). It is of note that the 4 ILCs predicted as IDCs were atypical ILCs, whereas all typical ILCs were accurately classified. The expression signature contained the 75 genes/ESTs (71 characterized genes and 4 ESTs) identified in 50/50 LOOCV iterations, with 48 genes overexpressed and 27 genes underexpressed in ILCs. Genes are distributed on 30 chromosomal arms, of which 1q, 11q, 17q concentrate a larger number of genes than others (Table 3). As expected, CDH1 was among the genes underexpressed in ILCs, whereas the 17q12 ERBB2 -GRB7 -C17orf37 cluster was overexpressed in IDCs. Association of the genes with biological processes according to Gene Ontology (GO) is shown in Table 4 . Six processes were significantly overrepresented: cell adhesion, cell motility, apoptosis, protein folding, extracellular matrix, and protein phosphorylation. Genes involved in fatty acid or basic metabolism, transcription, molecule transport were also included in the signature.

The classification power of our signature is also illustrated by hierarchical clustering (Figure 4B). Two distinct tumor clusters were defined with only 3 misclassified samples (2 IDCs and 1 ILC). It is of note that 8/10 typical ILCs clustered together in a close branch of the dendrogram, confirming their homogeneity as well as their difference with the atypical ILCs.

These results were validated in two sequential steps. The technical validation of cDNA microarrays data was done by quantitative RT-PCR on 45 samples (26 IDCS, 19 ILCs) from the original training set. As shown in Figure 5 , quantitative RT-PCR results confirmed significant differential expression ($p<10^{-4}$, t-test) between ILCs and IDCs for all 5 genes, substantiating the reliability of our microarray results. We next verified the performance of our signature on an independent set of 199 tumors previously profiled on the same microarray platform (Bertucci et al., 2004). SVM classification resulted in the accurate assignment of 88% (151/171) IDCs and 75% (21/28) ILCs, resulting in an 86% overall accuracy (Table 5).

Correspondence between genomic and expression data

We first determined the overlap between copy number changes and genes discriminating the two pathological types. Ten of the 75 genes (13%) of the expression signature (CD34 , 1q32.2; MARCH7 , 2q24.2; TGFBR2 , 3p24.1; ALDH1L1 , 3q21.2; EFCBP1 , 8q21.3; STUB1 , 16p13.3; PECAM , 17q23.3; ABCA6 , 17q24.2; MMP24 , 20q11.2; YWHAB , 20q13.1) mapped either within or at close proximity of a BAC included in the genomic signature. We were also interested in verifying whether typical and atypical ILCs presented differential genomic patterns (normal, simplex, complex and amplifier). It was remarkable that atypical ILCs presented a larger proportion of complex or amplifier patterns whereas most typical ILCs were simplex or normal (p=0.08, Fisher exact test). We also found a significant correlation (p=0.02, Fisher exact test) between genomic patterns and molecular subtype (luminal, basal, ERBB2 and normal-like) with more normal or simplex patterns in luminal A or normal-like tumors, and more complex or amplifier patterns within luminal B, basal or ERBB2 samples.

Discussion

We aimed at identifying molecular differences between ILCs and IDCs. For the first time to our knowledge, this was done at both the genomic and expression levels by means of array-CGH and cDNA microarray profiling and in a homogeneous series of samples with respect to several pathological features (Scarff Bloom Richardson grade, pT, hormone receptor and axillary lymph node status). Although these stringent criteria may have put the focus on a specific subset of breast cancer we noted that they allowed the identification of molecular differences independent from these features. We identified two molecular signatures, one at the genomic level (75 BAC clones), the second at the transcriptional level (75 genes/ESTs). Both signatures were accurate (86 and 92%, respectively) in classifying tumors from the original training set and, noticeably, performed well on independent validation sets (78 and 86% respectively). Quantitative RT-PCR further confirmed our results.

Genomic differences between ILCs and IDCs

Of the two studies (Loo et al., 2004 ; Stange et al., 2006) that looked for copy number differences between ILCs and IDCs by means of array-CGH, only Stange and coworkers (Stange et al., 2006) identified a significantly discriminating set of BAC clones. Five anomalies are common to our work and that of Stange: they involve 16p13.3, 16q12-q21, 17q23.2-q24.3 and 20q13.1-q13.3 regions. All these locations correspond to gains, which occur more frequently in IDCs than ILCs or are restricted to IDCs (17q23-q24). The somewhat restricted overlap between the discriminator BAC clones may reflect the differences in tumor samples respectively analyzed in both studies. Anomalies selected in our genomic signature correspond predominantly to events occurring more frequently in IDCs. This predominance reflects the higher level of rearrangements in IDCs. We found that events occurring at a high frequency are rare in ILCs. Some chromosomal locations showed inverse patterns. For instance, chromosomes 16, 17, 20 showed a predominance of gains in IDCs and of losses in ILCs; conversely, 7 and X were preferentially gained in ILCs and lost in IDCs.

Our data agree with classical CGH-based studies that showed the differential involvement of 17q and 20q in IDCs and ILCs (Gunther et al., 2001). However, they are in contrast with results indicating that ILCs are specified by increased frequency of losses at 16q (Stange et al., 2006). The 16q22 region harboring the CDH1 gene was not differentially involved in ILCs and IDCs in our dataset. Concomitant gain at 1q and loss at 16q were frequently found in a subset of ER-positive IDCs. Similarly, it was proposed that 11q13 amplification was more frequent in ILCs than IDCs (Stange et al., 2006). This contrasts with our data showing that 11q13 amplification, involving principally the CCND1 locus, was evenly distributed in ILCs and IDCs, likely because of the selection of ER-positive IDCs in our analysis. Our data show that, while it was possible to determine genomic anomalies discriminating lobular and ductal carcinomas, some ILCs shared a number of anomalies with ER-positive IDCs.

Differential expression between ILCs and IDCs

Expression analysis revealed two populations of ILCs, which differ with respect to their global expression profile, their molecular subtype as well as the expression profile for the 75-gene signature. This result was in agreement with Zhao and coworkers (Zhao et al., 2004) who identified typical ILCs and atypical “ductal-like” ILCs. Typical ILCs likely correspond to our homogeneous subgroup IIb ILCs,

while atypical correspond to more heterogeneous ILCs from group I and subgroup IIa. Korkola and coworkers (Korkola et al., 2003) also evidenced two groups of ILCs based on their ILCs vs IDCs expression signature.

Three previous expression profiling studies (Korkola et al., 2003 ; Turashvili et al., 2007 ; Zhao et al., 2004) have reported lists of genes with differential expression between ILCs and IDCs. The overlap between these lists and ours is low (Supplementary Table 2) with CDH1 being the only gene in common. Of the 75 genes selected in our expression signature, 11 genes (ALDH1A1, CAV1, CDH1, ERG, FABP4, IGF1, PDK4, TF, TGFBR2, VWF, YWHAB) were present in at least one of the three published lists, the best overlap being found with the list by Zhao and coworkers (Zhao et al., 2004). The three studies differ from ours by several aspects: no matching based on tumor characteristics was done to select samples, the number of which ranged from 5 to 21 for ILCs and 5 to 109 for IDCs, different technological microarray platforms and different analytic methods were used to generate the lists of discriminator genes and, finally, no validation tumor set was provided. This small overlap between the gene signatures in our and previous studies may also be explained by the lack of whole genome coverage. It is of note that biological processes or functions show greater concordance across these studies.

In our study, discriminator genes are involved in several cellular processes. Functional annotation of genes helps generate hypotheses about the biological mechanisms that sustain the differences in histoclinical properties of ILCs and IDCs. In particular genes overexpressed in IDCs correspond preferentially to promoters of cell proliferation (e.g. tyrosine kinase receptor ERBB2 , JAK2 , transcription factor ANKRD32 and calmodulin-binding NRGN), whereas those overexpressed in ILCs code for proteins involved in cell adhesion (VWF , ELN , DPT , EMCN) or lipid (FABP4 , CAV1 , ADIPOG) and retinoic acid metabolism (ALDH1A1). SFRP1 , TGFBR2 and IGF1 , whose functions are associated with cell differentiation rather than proliferation, were also upregulated in ILCs. This was further comforted by a search for functional pathways by means of the Ingenuity Pathway analysis (Ingenuity Systems, www.ingenuity.com). Two networks were identified, showing highest scores with cancer, tissue morphology and organismal injury. Network 1 was centered around CDH1, with direct interactions with MMP3, TGFBR2 and transcriptional activator TFAP2A and indirect links with p38-MAPK and NFKB (both of which are reported to be downregulated in this link). This network is thus clearly related to ILCs. Network 2 is centered around ERBB2-JAK2 with strong links to the heat shock protein system and apparent cross-regulations at the post-tanslational level. Its relation to IDC appears unequivocal. Overall, these data suggest that ILCs are less proliferative and characterized by a higher degree of differentiation than IDCs.

Correspondence between genomic and expression data

The degree of concordance between the genomic and expression signatures was 13%. It is in agreement with the 10–15% rate of the variation in gene expression estimated to be linked to genomic gains and losses (Pollack et al., 2002). Although this concordance may appear relatively low and might have been improved using whole genome expression and high resolution CGH arrays, it suggests a link between copy number and expression changes in the two tumor types.

In conclusion, our data show that ILCs and IDCs, while showing distinct genetic pathways share common rearrangements or expression patterns. These common genetic features define a subgroup of tumors intermediate between ILCs and IDCs. The existence of two subsets of ILCs was further substantiated by the genomic patterns defined as simplex, complex and amplifier (Fridlyand et al., 2006 ; Hicks et al., 2006). ILCs were predominantly of the simplex type, however, when we split ILCs into two subgroups “typical ILCs” and “ atypical ILCs”, it was clear that most simplex ILCs were of the typical subgroup, while atypical ILCs comprised a larger number of complex and amplifier cases as did IDCs. These data suggest that atypical ILCs may correspond to a more aggressive subset of ILCs that have acquired genomic characteristics in common with IDCs. This idea is reinforced by our data showing that 3 of 4 ILCs associated to a relapse and in some cases fatal outcome corresponded to atypical ILCs.

Material and methods

Tumor material

Primary breast cancers were collected in four French cancer hospitals: Centre Léon Bérard (Lyon), Institut Paoli-Calmettes (Marseille), Centre Val d'Aurelle (Montpellier) and Institut Gustave Roussy (Villejuif). Tumor biopsies were snap-frozen in liquid nitrogen upon surgical removal and stored at −80°C until nucleic acids extraction. All tumor sections were de novo reviewed prior to analysis by four pathologists (F.B., M.C.M., I.T., J.J.), and all profiled specimens contained more than 60% of tumor cells. DNA and RNA were isolated using respectively QIAamp DNA Midi Kit and Rneasy Mini Kit (Qiagen). Three series of tumors were assembled and analyzed in parallel. A “training set” of 50 samples, including 29 IDCs and 21 classic ILCs, exclusively composed of Scarff Bloom Richardson (SBR) grade 2 tumors, pT2 (pathological tumor size between 2 and 5 cm), ER+, with less than 3 involved axillary lymph nodes. These criteria limited the dispersion and increased the chances to determine genetic differences discriminating ILCs and IDCs. Forty-five if the 50 tumors (26 IDCS, 19 ILCs) were also analyzed by quantitative RT-PCR to validate cDNA microarrays results. A second set composed of 23 SBR grade 2 tumors (16 IDCs, 7 ILCs) was used to validate the genomic signature. A third set, previously published (Bertucci et al., 2004), consisting of 199 unselected invasive tumors (171 IDCs, 28 ILCs) was used to validate the expression signature. Description of these tumor sets is presented in the Supplementary data (Supplementary Table 1) .

TP53 and CDH1 mutation identification

Tumor DNA was subjected to PCR amplification of individual exons: exons 1–16 of CDH1 and exons 4–10 of TP53 which correspond to the DNA binding domain of the p53 protein and concentrate over 90% of TP53 mutations affecting breast cancer. PCRamplified products were purified and subsequently analyzed by direct sequencing using PRISM Dye Terminator (Applied Biosystems, Foster City, CA) with an automated sequencer ABI 373 (Applied Biosystems, Foster City, CA). Specific primers used for PCR reactions and sequencing are available upon request.

CDH1 immunostaining

Tissue microarray (TMA) preparation, immunohistochemical staining and scoring were done as described (Jacquemier et al., 2005). The E Cadherin monoclonal antibody at 1/2000° (Transduction laboratories, Lexington, KY.) was used according the supplier's recommendations. Slides were evaluated under a light microscope by two independent observers on the Spot Browser device (Alphelys). A cut-off of 1% for the quick-score classified samples into two classes: negative (Q <1%) and positive (Q ≥1%).

Array-CGH profiling

We used human Integrachip V2 to establish genomic profiles (IntegraGen SA, Evry. France, http://www.integragen.com). IntegraChip V2 is composed of 3172 bacterial artificial chromosome (BAC) clones including 2862 sequenced clones with a median gap of 1 clone/0.8 Mb. DNA labeling, hybridization, were done as previously described (Orsetti et al., 2006). Image processing and analysis are detailed in supplementary information . Clones with missing values in over 50% of the tumors were discarded. Gains and losses were defined respectively at 0.25 and −0.25 as $\log^2$ ratio thresholds.

Gene expression profiling with cDNA microarrays

Expression profiles were defined using Ipsogen DiscoveryChip cDNA microarrays (Ipsogen, Marseille, France; http://www.ipsogen.fr/). Nylon microarrays contained PCR products from a total of 8305 Image clones. Clones represented 2898 expressed sequence tags (ESTs) and 5407 known genes, ~3000 of which were related with oncogenesis. Microarrays, probe labelling, hybridization, signal capture and data normalization were as described (Bertucci et al., 2004).

Supervised and unsupervised data analyses

Identical analytic methods were applied for array-CGH and expression profiles. Supervised analysis methodology is described in supplementary data (Supplementary Figure 1) . Unsupervised analysis was based on hierarchical clustering performed using Cluster and TreeView software (Eisen et al., 1998) with median-centered values and Pearson correlation as similarity metrics.

Quantitative RT-PCR

Quantitative RT-PCR was as described by Applied Biosystems (Foster City, CA USA). The primers, fluorescent probes and reagents used for quantifications were from Applied Biosystems. All reactions were performed in duplicate. Each sample was normalised on the content of ribosomal RNA.

Statistical analysis

Correlations between sample groups and histoclinical parameters were calculated with the Fisher's exact test. All statistical tests were two-sided at the 5% level of significance. Statistical analysis was done using the SPSS software (version 10.0.5).

Ackowledgements:

This study was developped as part of a joint program « Développement d'outils de diagnostic moléculaire en Cancérologie: Applications aux cancers du sein » Ministère de l'Enseignement Supérieur, de la Recherche et de la Technologie and Fédération Nationale des Centres de Lutte Contre le Cancer and was supported by funds from INSERM, the Association de Recherche sur le Cancer (ARC), grant 5102, Institut National du Cancer, Cancéropoles PACA and Grand Sud Ouest. The help of the Génopole Montpellier Languedoc-Roussillon is gratefully acknowledged. The authors thank Pr Dominique Maraninchi and Dr Claude Mawas for setting up this work and Mrs Sophie Tourpin for technical help.

References:

- Bertucci F , Borie N , Ginestier C , Groulet A , Charafe-Jauffret E , Adelaide J . 2004 ; Identification and validation of an ERBB2 gene expression signature in breast cancers . Oncogene . 23 : 2564 - 2575
- Bertucci F , Finetti P , Cervera N , Maraninchi D , Viens P , Birnbaum D . 2006 ; Gene expression profiling and clinical outcome in breast cancer . Omics . 10 : 429 - 443
- Berx G , Cleton-Jansen AM , Nollet F , de Leeuw WJ , van de Vijver M , Cornelisse C , van Roy F . 1995 ; E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers . Embo J . 14 : 6107 - 6115
- Eisen MB , Spellman PT , Brown PO , Botstein D . 1998 ; Cluster analysis and display of genome-wide expression patterns . Proc Natl Acad Sci U S A . 95 : 14863 - 14868

- Flagiello D , Gerbault-Seureau M , Sastre-Garau X , Padoy E , Vielh P , Dutrillaux B . 1998 ; Highly recurrent der(1;16)(q10;p10) and other 16q arm alterations in lobular breast cancer . Genes Chromosomes Cancer . 23 : 300 - 306
- Fridlyand J , Snijders AM , Ylstra B , Li H , Olshen A , Segraves R . 2006 ; Breast tumor copy number aberration phenotypes and genomic instability . BMC Cancer . 6 : 96 -
- Gunther K , Merkelbach-Bruse S , Amo-Takyi BK , Handt S , Schroder W , Tietze L . 2001 ; Differences in genetic alterations between primary lobular and ductal breast cancers detected by comparative genomic hybridization . J Pathol . 193 : 40 - 47
- Hicks J , Krasnitz A , Lakshmi B , Navin NE , Riggs M , Leibu E . 2006 ; Novel patterns of genome rearrangement and their association with survival in breast cancer . Genome Res . 16 : 1465 - 1479
- Jacquemier J , Ginestier C , Rougemont J , Bardou VJ , Charafe-Jauffret E , Geneix J . 2005 ; Protein expression profiling identifies subclasses of breast cancer and predicts prognosis . Cancer Res . 65 : 767 - 779
- Katz A , Saad ED , Porter P , Pusztai L . 2007 ; Primary systemic chemotherapy of invasive lobular carcinoma of the breast . Lancet Oncol . 8 : 55 - 62
- Korkola JE , DeVries S , Fridlyand J , Hwang ES , Estep AL , Chen YY . 2003 ; Differentiation of lobular versus ductal breast carcinomas by expression microarray analysis . Cancer Res . 63 : 7167 - 7175
- Lamovec J , Bracko M . 1991 ; Metastatic pattern of infiltrating lobular carcinoma of the breast: an autopsy study . J Surg Oncol . 48 : 28 - 33
- Li CI , Anderson BO , Daling JR , Moe RE . 2003 ; Trends in incidence rates of invasive lobular and ductal breast carcinoma . Jama . 289 : 1421 - 1424
- Loo LW , Grove DI , Williams EM , Neal CL , Cousens LA , Schubert EL . 2004 ; Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes . Cancer Res . 64 : 8541 - 8549
- Loveday RL , Greenman J , Simcox DL , Speirs V , Drew PJ , Monson JR , Kerin MJ . 2000 ; Genetic changes in breast cancer detected by comparative genomic hybridisation . Int J Cancer . 86 : 494 - 500
- Orsetti B , Nugoli M , Cervera N , Lasorsa L , Chuchana P , Rouge C . 2006 ; Genetic profiling of chromosome 1 in breast cancer: mapping of regions of gains and losses and identification of candidate genes on 1q . Br J Cancer . 95 : 1439 - 1447
- Pollack JR , Sorlie T , Perou CM , Rees CA , Jeffrey SS , Lonning PE . 2002 ; Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors . Proc Natl Acad Sci U S A . 99 : 12963 - 12968
- Sastre-Garau X , Jouve M , Asselain B , Vincent-Salomon A , Beuzeboc P , Dorval T . 1996 ; Infiltrating lobular carcinoma of the breast. Clinicopathologic analysis of 975 cases with reference to data on conservative therapy and metastatic patterns . Cancer . 77 : 113 - 120
- Sorlie T , Perou CM , Tibshirani R , Aas T , Geisler S , Johnsen H . 2001 ; Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications . Proc Natl Acad Sci U S A . 98 : 10869 - 10874
- Sorlie T , Tibshirani R , Parker J , Hastie T , Marron JS , Nobel A . 2003 ; Repeated observation of breast tumor subtypes in independent gene expression data sets . Proc Natl Acad Sci U S A . 100 : 8418 - 8423
- Stange DE , Radlwimmer B , Schubert F , Traub F , Pich A , Toedt G . 2006 ; Highresolution genomic profiling reveals association of chromosomal aberrations on 1q and 16p with histologic and genetic subgroups of invasive breast cancer . Clin Cancer Res . 12 : 345 - 352
- Toikkanen S , Pylkkanen L , Joensuu H . 1997 ; Invasive lobular carcinoma of the breast has better short- and long-term survival than invasive ductal carcinoma . Br J Cancer . 76 : 1234 - 1240
- Turashvili G , Bouchal J , Baumforth K , Wei W , Dziechciarkova M , Ehrmann J . 2007 ; Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis . BMC Cancer . 7 : 55 -
- Zhao H , Langerod A , Ji Y , Nowels KW , Nesland JM , Tibshirani R . 2004 ; Different gene expression patterns in invasive lobular and ductal carcinomas of the breast . Mol Biol Cell . 15 : 2523 - 2536

Figure 1

Frequency of genomic imbalances in IDCs and ILCs and differences according to histological types

Gains and losses were calculated using 0,25 and − 0,25 as $\log^2$ ratio thresholds. The overall frequency of gains (black) and losses (grey) in the whole training set of 50 tumors was calculated for the 2872 filtered BACs and plotted against their genomic position (Hg18).. A /IDC, B /ILC. The absolute difference corresponding to the subtype specific frequencies was calculated by substracting the frequency of one subtype by the other; C /IDC-ILC, D /ILC-IDC. E /p-values associated to the differences were computed using Wilcoxon two-sample test (CGHtest, http://www.few.vu.nl/~mavdwiel/CGHtest.html). Only significant p-values were represented ($p<0.05$): we plotted 1-p-value for gains, and − (1-p-value) for losses.

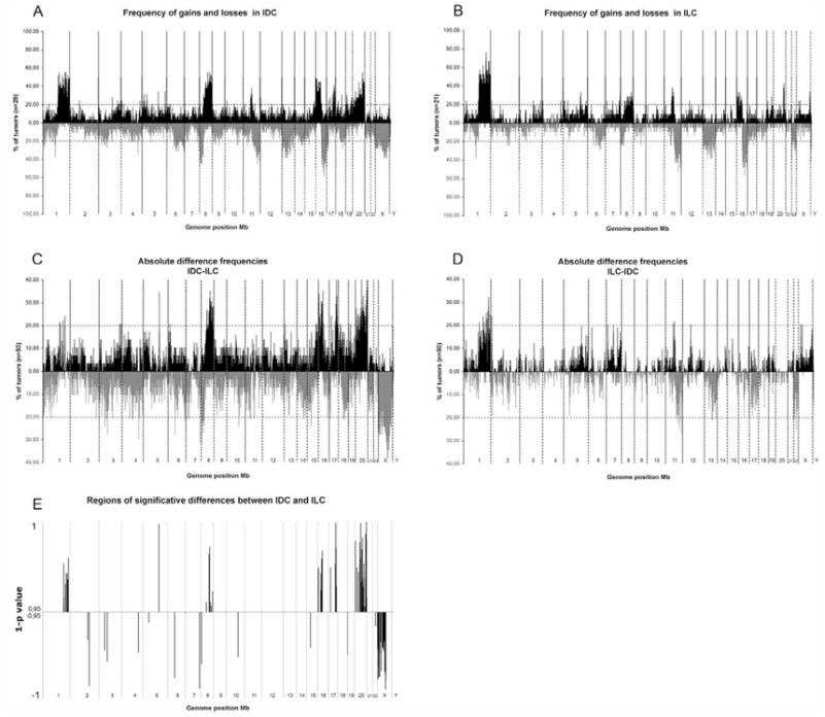

Figure 2

Classification of the training tumor set on the basis of the 75 BACs genomic signature

A/ The 50 tumors of the training set were classified by SVM and plotted according to their probability to belong to the IDC subclass. A probability > 0.5 signs for IDC, < 0.5 signs for ILC classification. IDCs are indicated by circles and ILCs by triangles. Black circles correspond to IDCs bearing a TP53 mutation, black triangles to ILCs with a CDH1 mutation. B/ The same 50 tumors were classified using hierarchical clustering based on the the 75 BAC genomic signature. Each column represents a tumor, each row represents a BAC clone. Each cell in the matrix represents the DNA copy number of a BAC clone in a single sample relative to its median abundance across all samples. Red and green indicate levels respectively above and below the median. The magnitude of deviation from the median is represented by the colour saturation. Tumors are separated into two major clusters (I and II). Histological types are shown under the dendrogram: blue boxes indicate ILCs and yellow boxes indicate IDCs.

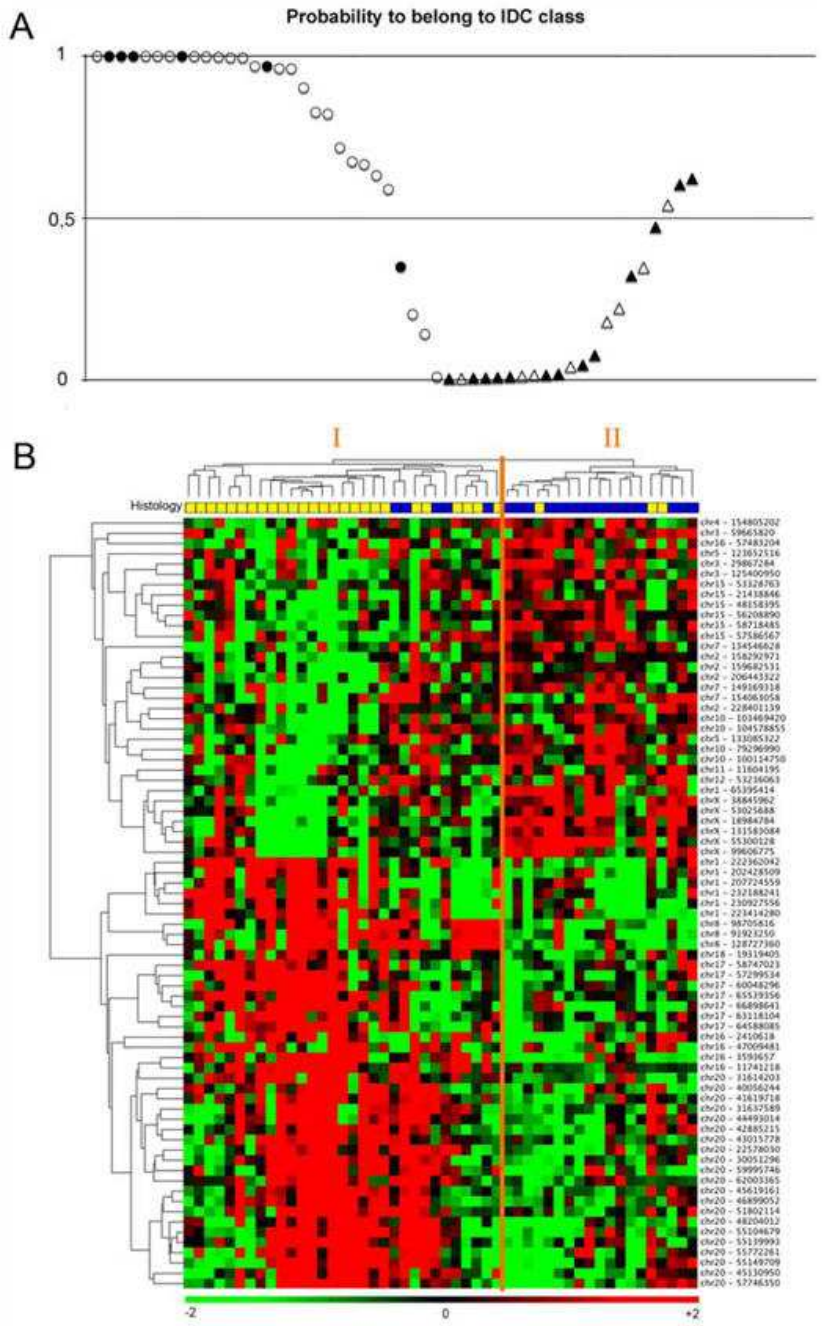

Figure 3

Global gene expression profiling in lobular and ductal breast cancer

A/ Hierarchical clustering of 50 samples and 7782 genes/ESTs with significant variation in mRNA expression level across the samples. Representation is as in Figure 3, except that color code represents gene expression level relative to its median abundance across the samples. The dendrogram of samples (above matrixes) represents overall similarities in gene expression profiles and is zoomed in B. Colored bars to the right indicate the locations of 9 gene clusters of interest that are zoomed in B. B/ Dendrograms of samples and gene clusters. Top , Two large groups of tissue samples (designated I and II), and three subgroups (I, IIa and IIb) are evidenced by clustering and delimited by dashed orange vertical lines. Middle , some relevant features of samples are represented according to a color ladder (unavailable, oblique feature): pathological type (IDC, yellow; ILC, blue), CDH1 IHC status (negative, white; positive, black), and molecular subtype of samples based on the intrinsic gene set (dark blue, luminal A; light blue, luminal B; pink, ERBB2-overexpressing; red, basal; green, normallike; white, not assigned with a correlation inferior to 0.15 with each centroid). Down , expanded view of selected gene clusters named from top to bottom: CDH1 (black bar), luminal/ER (dark blue bar), proliferation (grey bar), ERBB2-related (pink bar), immune (green bar), basal (red bar), adipose (orange bar), early response (light blue bar), stromal (brown bar).

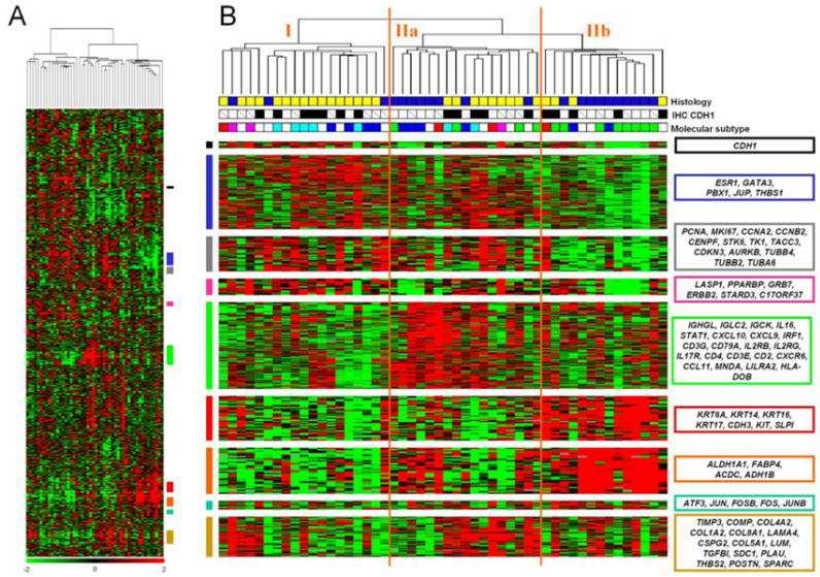

Figure 4

Classification of the training tumor set on the basis of the 75 ESTs/gene expression signature

A/ and B/Classification of the 50 tumors of the training set. Representation is as in Figures 2A–B .

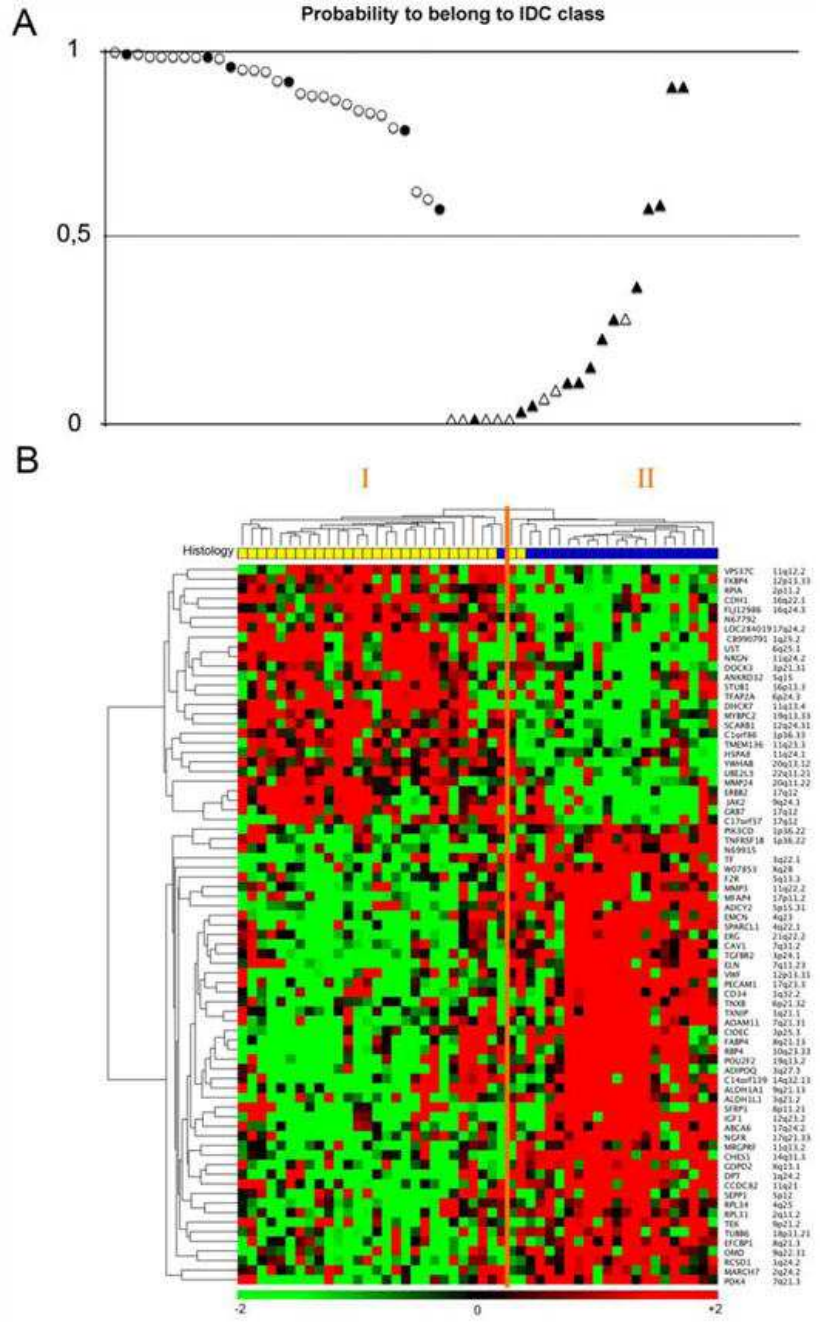

Figure 5

Validation of cDNA microarray data with quantitative RT-PCR

Boxplots of the expression of 5 genes in IDCs and ILCs (45 tumors from the training set) measured by quantitative RT-PCR. Expression is given in arbitrary units. P-values are strongly significant (t-test). The horizontal black line represents the median expression level.

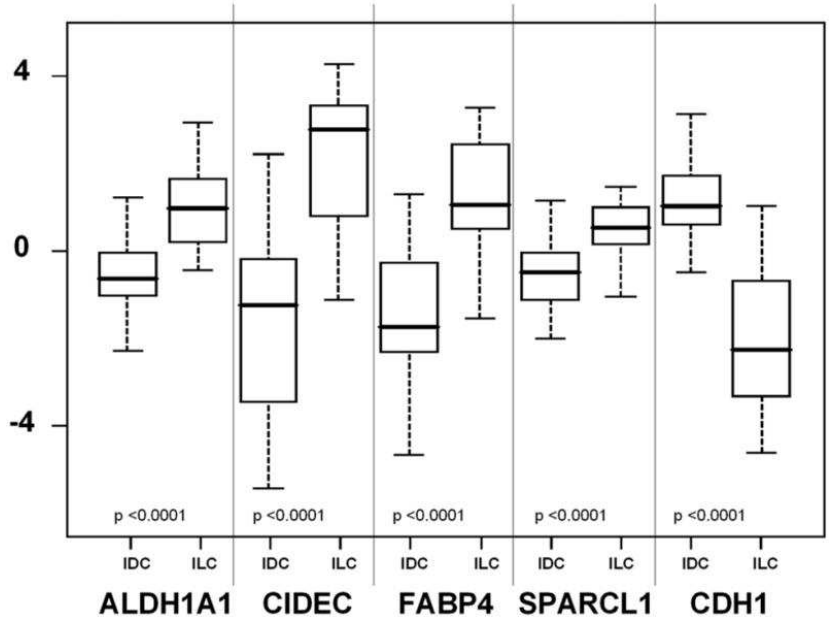

Table 1

75 BACs of the genomic signature.

Clone name	Mb Start	Mb End	CytoBand
RP11-86K19	65395414	65593152	1p31.3
RP11-219P13	202428509	202603375	1q32.1
RP11-123O6	207724559	207869795	1q32.2
FE0DBACA14ZD04	222362042	222512173	1q42.12
RP11-79O22	223414280	223593296	1q42.13
FE0DBACA14ZG05	230927556	231077745	1q42.2
CTB-17M17	232188241	232278477	1q42.3
FE0DBACA27ZF07	158292971	158432169	2q24.1
RP11-343J7	159682531	159882013	2q24.1
CTA-388C7	206443322	206542163	2q33.3
RP11-90L9	228401139	228557350	2q36.3
FE0DBACA17ZF06	29867284	30018475	3p24.1
CTD-2175D15	59665820	59809076	3p14.2
RP11-25L9	125400950	125564390	3q21.2
CTB-21J19	154805202	154966122	4q31.3
RP11-1103G8	123652516	123817681	5q23.2
RP11-4E3	133085322	133272664	5q31.1
RP11-140I14	134546628	134697478	7q33
RP11-45C4	149169318	149330056	7q36.1
RP11-80J22	154063058	154225480	7q36.2
FE0DBACA12ZH10	91923250	92073564	8q21.3
FE0BPADA8ZF10	98705816	98887966	8q22.1
RP11-237F24	128727360	128877662	8q24.21
RP11-469G7	79296990	79458206	10q22.3
CTD-2022D20	100114750	100279334	10q24.2
CTD-2062E5	103469420	103604807	10q24.32
CTA-224D5	104578855	104583204	10q24.32
RP11-82L24	11604195	11758360	11p15.3
CTB-18O12	53236063	53367863	12q13.2
FE0DBACA8ZD01	21438846	21638239	15q11.2
RP11-49K3	48158395	48330577	15q21.2
RP11-527E24	53328763	53478859	15q21.3
CTB-4N21	56208890	56341570	15q21.3
FE0DBACA3ZH04	57586567	57757359	15q22.2
FE0DBACA8ZE02	58718485	58903081	15q22.2
FE0DBACA28ZE01	2410618	2614456	16p13.3
RP11-499F21	3593657	3759566	16p13.3
RP11-394B14	11741218	11933423	16p13.13
RP11-428C9	47009481	47172659	16q12.1

FE0DBACA22ZF01	57483204	57483382	16q21
RP11-535J5	57299534	57496091	17q23.2
RP11-300L16	58747023	58922997	17q23.3
RP11-81D7	60048296	60066600	17q24.1
FE0DBACA16ZC02	63118104	63295003	17q24.2
RP11-293K20	64588085	64772817	17q24.2
RP11-300G13	65539356	65725424	17q24.3
RP11-108I3	66898641	67067383	17q24.3
FE0DBACA23ZE11	19319405	19415886	18q11.2
RP11-566G8	22578030	22779817	20p11.21
FE0DBACA26ZA01	30051296	30201447	20q11.21
FE0DBACA25ZB10	31614203	31771032	20q11.22
CTD-2061E8	31637589	31771792	20q11.22
FE0DBACA27ZG02	40056244	40190025	20q12
FE0DBACA23ZA08	41619718	41780058	20q13.12
CTD-2200K24	42885215	42976796	20q13.12
FE0DBACA25ZH08	43015778	43166028	20q13.12
FE0DBACA4ZC12	44493014	44673426	20q13.12
FE0DBACA25ZA08	45130950	45339408	20q13.12
FE0DBACA24ZC11	45619161	45769373	20q13.12
RP11-523P17	46899052	47096954	20q13.13
CTD-2112O20	48204012	48335040	20q13.13
RP11-55E1	51802114	51966525	20q13.2
FE0DBACA25ZC11	55104679	55224851	20q13.31
CTD-2045C24	55139993	55256243	20q13.31
FE0DBACA25ZE11	55149709	55241492	20q13.31
RP11-402F1	55772261	55922598	20q13.32
RP11-335C6	57746350	57934405	20q13.33
RP11-460D14	59995746	60208710	20q13.33
FE0DBACA5ZH02	62003365	62160261	20q13.33
RP11-482C1	18984784	19152238	Xp22.13
FE0DBACA1ZH07	38845962	39018790	Xp11.4
RP11-258C19	53025688	53204103	Xp11.22
RP11-126A13	55300128	55467787	Xp11.21
RP11-508G21	99606775	99832145	Xq22.1
RP11-665G20	131583084	131750233	Xq26.2

Table 2

SVM classification of an independent validation set of 23 breast carcinomas

CGH-array profiles of these tumors were determined and classsified by means of the 75 BACs genomic signature. Rows correspond to the different subclasses determined on the basis of the pathology report. Accuracy corresponds to ratio of tumors correctly classified on the total number of cases in the subclass.

		SVM classification		
Histological typing	**N**	**IDC**	**ILC**	**Accuracy (%)**
IDC	16	12	4	75.0
ILC	7	1	6	85.7
Total	23	13	10	

Table 3

75 ESTs/genes of the expression signature.

HUGO Gene symbol	Gene Definition	Clone image	Accession number	Unigene	RefSeq	Start	End	CytoBand	Status
JAK2	Janus kinase 2 (a protein tyrosine kinase)	image:789379	BX106933	Hs.591081	NM_004972.2	4975245	5117994	9p24.1	UP DUC
ANKRD32	Ankyrin repeat domain 32	image:182580	H42196	Hs.556673	NM_032290.2	94040330	94057327	5q15	UP DUC
C17orf37	Chromosome 17 open reading frame 37	image:4283597	BC006006	Hs.333526	NM_032339.3	35138937	35140314	17q12	UP DUC
C1orf86	Chromosome 1 open reading frame 86	image:150515	H01457	Hs.107101	NM_182533.1	2110848	2116074	1p36.33	UP DUC
CDH1	Cadherin 1, type 1, E-cadherin (epithelial)	image:214008	H72404	Hs.461086	NM_004360.2	67328696	67426945	16q22.1	UP DUC
DHCR7	7-dehydrocholesterol reductase	image:153009	R50345	Hs.503134	NM_001360.2	70823105	70837125	11q13.4	UP DUC
DOCK3	Dedicator of cytokinesis 3	image:33799	R44552	Hs.476284	NM_004947.3	50687676	51396668	3p21.31	UP DUC
ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro	image:756253	AA480116	Hs.446352	NM_001005862.1	35097919	35138440	17q12	UP DUC
FKBP4	FK506 binding protein 4, 59kDa	image:341237	W58661	Hs.524183	NM_002014.2	2774414	2783383	12p13.33	UP DUC
FLJ12986	Hypothetical protein FLJ12986	image:306077	N91481	Hs.54713	AK023048.1	87827474	87829903	16q24.3	UP DUC
GRB7	Growth factor receptor-bound protein 7	ipso:0000177	BE246692	Hs.86859	NM_005310.2	35147713	35157063	17q12	UP DUC
HSPA8	Heat shock 70kDa protein 8	image:884719	AA629567	Hs.180414	NM_006597.3	122433411	122438054	11q24.1	UP DUC
LOC284019	Hypothetical protein LOC284019	image:250619	H90075	Hs.370140	AL832149.1	62497017	62503952	17q24.2	UP DUC
MMP24	Matrix metallopeptidase 24 (membrane-inserted)	image:325088	W46985	Hs.567417	NM_006690.3	33278117	33328218	20q11.22	UP DUC
MYBPC2	Myosin binding protein C, fast type	image:306239	N78998	Hs.85937	NM_004533.2	55628004	55661389	19q13.33	UP DUC
N_A	Full length insert cDNA YH97B03	image:30336399	CB990791	Hs.496139		175181399	175181998	1q25.2	UP DUC

N_A		image:291523	N67792			194902888	194903369	3q29	UP DUC
NRGN	Neurogranin (protein kinase C substrate, RC3)	image:177718	H46419	Hs.524116	NM_006176.1	124114952	124122307	11q24.2	UP DUC
RPIA	Ribose 5-phosphate isomerase A (ribose 5-phosphate epimerase)	image:263097	N20072	Hs.469264	NM_144563.2	88772291	88831566	2p11.2	UP DUC
SCARB1	Scavenger receptor class B, member 1	image:51976	H23199	Hs.298813	NM_005505.3	123828129	123914287	12q24.31	UP DUC
STUB1	STIP1 homology and U-box containing protein 1	image:154442	R54831	Hs.592081	NM_005861.2	670116	672768	16p13.3	UP DUC
TFAP2A	Transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)	image:149884	H00651	Hs.519880	NM_001042425.1	10504903	10527783	6p24.3	UP DUC
TMEM136	Transmembrane protein 136	image:298746	N74690	Hs.643516	NM_174926.1	119701226	119706556	11q23.3	UP DUC
UBE2L3	Ubiquitin-conjugating enzyme E2L 3	image:327216	W02771	Hs.108104	NM_003347.2	20251957	20308323	22q11.21	UP DUC
UST	Uronyl-2-sulfotransferase	image:220199	H82656	Hs.557541	NM_005715.1	149110157	149439818	6q25.1	UP DUC
VPS37C	Vacuolar protein sorting 37 homolog C (S. cerevisiae)	image:50238	H16997	Hs.523715	NM_017966.4	60654304	60685492	11q12.2	UP DUC
YWHAB	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide	image:262546	H99319	Hs.651212	NM_003404.3	42947758	42970574	20q13.12	UP DUC
ABCA6	ATP-binding cassette, sub-family A (ABC1), member 6	image:112127	T84930	Hs.647403	NM_080284.2	64586442	64649610	17q24.2-q24.3	UP LOB
ADAM11	ADAM metallopeptidase domain 11	image:184240	H43855	Hs.6088	NM_002390.4	40192094	40214738	7q21.31	UP LOB
ADCY2	Adenylate cyclase 2 (brain)	image:282977	N45141	Hs.481545	NM_020546.2	7449343	7883194	5p15.31	UP LOB
ADIPOQ	Adiponectin, C1Q and collagen domain containing	image:183476	H45617	Hs.80485	NM_004797.2	188043157	188058944	3q27.3	UP LOB
ALDH1A1	Aldehyde dehydrogenase 1 family, member A1	image:309697	N94546	Hs.76392	NM_000689.3	74705408	74757789	9q21.13	UP LOB
ALDH1L1	Aldehyde dehydrogenase 1 family, member L1	image:153982	R67615	Hs.434435	NM_012190.2	127305098	127382175	3q21.2	UP LOB
C14orf139	Chromosome 14 open reading frame 139	image:265829	N20974	Hs.41502	CR457337	94945159	94945729	14q32.13	UP LOB
CAV1	Caveolin 1, caveolae protein, 22kDa	image:377461	AA055368	Hs.74034	NM_001753	115952075	115988466	7q31.2	UP LOB
CCDC82	Coiled-coil domain containing 82	image:277621	N49389	Hs.525088	NM_024725.2	95725589	95762710	11q21	UP LOB
CD34	CD34 molecule	image:770858	AA434387	Hs.374990	NM_001773.2	206126507	206151306	1q32.2	UP LOB
CHES1	Checkpoint suppressor 1	image:221846	H84982	Hs.434286	NM_005197.2	88692278	88953127	14q31.3-q32.11	UP LOB
CIDEC	Cell death-inducing DFFA-like effector c	image:155655	R71842	Hs.567562	NM_022094.2	9883399	9895740	3p25.3	UP LOB
DPT	Dermatopontin	image:153505	R48303	Hs.80552	NM_001937.3	166931331	166965052	1q24.2	UP

									LOB
EFCBP1	EF-hand calcium binding protein 1	image:282100	N51496	Hs.560892	NM_022351.2	91872954	92040805	8q21.3	UP LOB
ELN	Elastin (supravalvular aortic stenosis, Williams-Beuren syndrome)	image:810934	AA459308	Hs.647061	NM_000501.1	73080454	73120965	7q11.23	UP LOB
EMCN	Endomucin	image:272630	N36136	Hs.152913	NM_016242.2	101538009	101658202	4q23	UP LOB
ERG	V-ets erythroblastosis virus E26 oncogene homolog (avian)	image:302929	N90107	Hs.473819	NM_182918.2	38675671	38792267	21q22.2	UP LOB
F2R	Coagulation factor II (thrombin) receptor	image:813254	AA455910	Hs.482562	NM_001992.2	76047547	76067054	5q13.3	UP LOB
FABP4	Fatty acid binding protein 4, adipocyte	image:162654	CR744520	Hs.391561	NM_001442.1	82553490	82558004	8q21.13	UP LOB
GDPD2	Glycerophosphodiester phosphodiesterase domain containing 2	image:192521	H41285	Hs.438712	NM_017711.2	69559716	69569955	Xq13.1	UP LOB
IGF1	Insulin-like growth factor 1 (somatomedin C)	image:813179	AA456321	Hs.160562	NM_000618.2	101313807	101398454	12q23.2	UP LOB
MARCH7	Membrane-associated ring finger (C3HC4) 7	image:327461	W20438	Hs.529272	NM_022826.2	160277256	160333329	2q24.2	UP LOB
MFAP4	Microfibrillar-associated protein 4	image:759163	AA496022	Hs.296049	NM_002404.1	19227350	19231086	17p11.2	UP LOB
MMP3	Matrix metallopeptidase 3 (stromelysin 1, progelatinase)	image:324700	BX117609	Hs.375129	NM_002422.3	102211738	102219552	11q22.2	UP LOB
MRGPRF	MAS-related GPR, member F	image:324543	W52061	Hs.118513	NM_145015.2	68528443	68537311	11q13.2	UP LOB
N_A	Transcribed locus, weakly similar to XP_848633.1 similar to LINE-1 reverse transcriptase homolog	image:301068	W07853	Hs.433075		151073054	151383976	Xq28	UP LOB
N_A		image:297752	N69915						UP LOB
NGFR	Nerve growth factor receptor (TNFR superfamily, member 16)	image:154790	R55303	Hs.415768	NM_002507.1	44927666	44947360	17q21.33	UP LOB
OMD	Osteomodulin	image:258606	N32201	Hs.94070	NM_005014.1	94216359	94226381	9q22.31	UP LOB
PDK4	Pyruvate dehydrogenase kinase, isozyme 4	image:594120	AA169469	Hs.8364	NM_002612.3	95050749	95063861	7q21.3	UP LOB
PECAM1	Platelet/endothelial cell adhesion molecule (CD31 antigen)	image:127201	R08228	Hs.652132	NM_000442.3	59753596	59817743	17q23.3	UP LOB
PIK3CD	Phosphoinositide-3-kinase, catalytic, delta polypeptide	image:712401	AA281784	Hs.518451	NM_005026.2	9634390	9711556	1p36.22	UP LOB
POU2F2	POU domain, class 2, transcription factor 2	image:186924	H43361	Hs.646363	NM_002698.2	47284513	47328470	19q13.2	UP LOB
RBP4	Retinol binding protein 4, plasma	image:78644	T61830	Hs.50223	NM_006744.3	95341584	95350983	10q23.33	UP LOB
RCSD1	RCSD domain containing 1	image:309244	N93864	Hs.493867	NM_052862.2	165865954	165942109	1q24.2	UP LOB
RPL31	Ribosomal protein L31	image:309643	N98434	Hs.469473	NM_000993.2	100985183	100989312	2q11.2	UP LOB

RPL34	Ribosomal protein L34	image:178137	H47015	Hs.438227	NM_000995.3	109761171	109771086	4q25	UP LOB
SEPP1	Selenoprotein P, plasma, 1	ipso:0000013		Hs.652198	NM_005410.2	42835740	42847781	5p12	UP LOB
SFRP1	Secreted frizzled-related protein 1	image:783700	AA446824	Hs.213424	NM_003012.3	41238636	41286137	8p11.21	UP LOB
SPARCL1	SPARC-like 1 (mast9, hevin)	image:289272	N73696	Hs.62886	NM_004684.2	88613514	88669530	4q22.1	UP LOB
TEK	TEK tyrosine kinase, endothelial (venous malformations, multiple cutaneous and mucosal)	image:151501	H02848	Hs.89640	NM_000459.2	27099286	27220171	9p21.2	UP LOB
TF	Transferrin	image:246018	N52306	Hs.518267	NM_001063.2	134947925	134980325	3q22.1	UP LOB
TGFBR2	Transforming growth factor, beta receptor II (70/80kDa)	image:309647	N98436	Hs.82028	NM_001024847.1	30622998	30710635	3p24.1	UP LOB
TNFRSF1B	Tumor necrosis factor receptor superfamily, member 1B	image:307628	N92967	Hs.256278	NM_001066.2	12149647	12191863	1p36.22	UP LOB
TNXB	Tenascin XB	image:124340	R00971	Hs.485104	NM_019105.5	32116911	32185131	6p21.32	UP LOB
TUBB6	Tubulin, beta 6	image:264801	N21031	Hs.193491	NM_032525.1	12298257	12316567	18p11.21	UP LOB
TXNIP	Thioredoxin interacting protein	image:5893349	BQ001703	Hs.533977	NM_006472.1	144149939	144153880	1q21.1	UP LOB
VWF	Von Willebrand factor	image:154475	R54854	Hs.440848	NM_000552.3	5928301	6104097	12p13.31	UP LOB

UP DUC more expressed in ductal than in lobular carcinomas
UP LOB more expressed in lobular than in ductal carcinomas

Table 4

Biological processes (Gene Ontology) associated with genes differentially expressed between ILCs and IDCs.

Extracellular region and matrix p= 1.13E-07	ADIPOQ	↑	OMD	↑
	DPT	↑	SEPP1	↑
	ELN	↑	SPARCL1	↑
	IGF1	↑	TF	↑
	MFAP4	↑	TNXB	↑
	MMP3	↑	VWF	↑
Protein kinase p= 0.02	ERG	↑	PIK3CD	↑
	JAK2	↓	TEK	↑
	PDK4	↑	TGFBR2	↑
	ERBB2	↓		
Apoptosis p= 0.01	CIDEC	↑	NGFR	↑
	F2R	↑	TNFRSF1B	↑
Cell motility p= 4.73E-04	IGF1	↑	JAK2	↑
	F2R	↑	PECAM1	↑
Cell adhesion p= 0.002	CD34	↑	OMD	↑
	CDH1	↓	PECAM1	↑
	DPT	↑	SCARB1	↓
	MFAP4	↑	TNXB	↑
	MYBPC2	↓	VWF	↑
Protein folding p= 0.015	FKBP4	↓	STUB1	↓
	HSPA8	↓		

↑ overexpression in ILCs;
↓ underexpression in ILCs.

Table 5

SVM classification of an independent validation set of 199 breast carcinomas

Rows correspond to both subclasses determined on histopathological criteria, columns to the prediction of belonging to one class or another by SVM and the expression signature.

		SVM classification		
Histological typing	**N**	**IDC**	**ILC**	**Accuracy (%)**
IDC	171	151	20	88.3
ILC	28	7	21	75.0
Total	199	158	41	