Gene expression profiling for molecular characterization of inflammatory breast cancer and prediction of response to chemotherapy

F Bertucci, P Finetti, J Rougemont, E Charafe-Jauffret, V Nasser, Béatrice B. Loriod, J Camerlo, R Tagett, C Tarpin, G Houvenaeghel, et al.

]

Gene Expression Profiling for Molecular Characterization of Inflammatory Breast Cancer and Prediction of Response to Chemotherapy

François Bertucci,[1,2,5] Pascal Finetti,[1] Jacques Rougemont,[6] Emmanuelle Charafe-Jauffret,[1,3,5] Valéry Nasser,[1] Béatrice Loriod,[6] Jacques Camerlo,[2] Rebecca Tagett,[7] Carole Tarpin,[2] Gilles Houvenaeghel,[4,5] Catherine Nguyen,[6] Dominique Maraninchi,[2,5] Jocelyne Jacquemier,[3] Rémi Houlgatte,[6] Daniel Birnbaum,[1] and Patrice Viens[2,5,8]

[1]Département d'Oncologie Moléculaire, [2]Département d'Oncologie Médicale, [3]Département de Biopathologie, and [4]Département de Chirurgie, Institut Paoli-Calmettes and UMR599 Institut National de la Santé et de la Recherche Médicale (INSERM), IFR137, Marseille, France; [5]Faculté de Médecine, Université de la Méditerranée, Marseille, France; [6]Laboratoire TAGC, ERM206 INSERM, Marseille, France; [7]Ipsogen SA, Marseille, France; and [8]Centre d'Investigation Clinique de Marseille Sainte-Marguerite, Marseille, France

ABSTRACT

Inflammatory breast cancer (IBC) is a rare but aggressive form of breast cancer with a 5-year survival limited to ~40%. Diagnosis, based on clinical and/or pathological criteria, may be difficult. Optimal systemic neoadjuvant therapy and accurate predictors of pathological response have yet to be defined for increasing response rate and survival. Using DNA microarrrays containing ~8,000 genes, we profiled breast cancer samples from 81 patients, including 37 with IBC and 44 with noninflammatory breast cancer (NIBC). Global unsupervised hierarchical clustering was able to some extent to distinguish IBC and NIBC cases and revealed subclasses of IBC. Supervised analysis identified a 109-gene set the expression of which discriminated IBC from NIBC samples. This molecular signature was validated in an independent series of 26 samples, with an overall performance accuracy of 85%. Discriminator genes were associated with various cellular processes possibly related to the aggressiveness of IBC, including signal transduction, cell motility, adhesion, and angiogenesis. A similar approach, with leave-one-out cross-validation, identified an 85-gene set that divided IBC patients with significantly different pathological complete response rate (70% in one group and 0% in the other group). These results show the potential of gene expression profiling to contribute to a better understanding of IBC, and to provide new diagnostic and predictive factors for IBC, as well as for potential therapeutic targets.

INTRODUCTION

Inflammatory breast cancer (IBC) is a rare (~5% of cases) but aggressive form of breast cancer. At diagnosis, a majority of patients show axillary lymph node involvement and ~35% have distant metastases. Although survival has been improved by the introduction of primary chemotherapy in the multimodality treatment, prognosis remains poor with 5-year survival ranging from 30 to 50% (1). Diagnosis is based on clinical and/or pathologic criteria. Clinical inflammatory symptoms arise quickly and involve more than one third of the breast. The disease is classified as T_{4d} according to the tumor-node-metastasis (TNM)-Union International Contre Cancer (UICC) classification (5th edition). The tumor is often of ductal type, with high histologic grade, negative for hormone receptors, and highly angiogenic and invasive (1). The presence of tumor emboli in dermal lymphatic vessels constitutes the pathologic hallmark of the disease. In some cases non-IBC (NIBC) may be difficult to distinguish, and yet this distinction is crucial for treatment. Prognostic features (2) remain contested. Response to primary chemotherapy is a strong, although imperfect, indicator of survival (3). Accurate treatment response predictors as well as optimal systemic therapy have to be defined to increase response rate and survival.

Because of its relative infrequency and the small size of diagnostic samples, IBC has rarely been investigated at the biological level, and little is known about the underlying molecular alterations such as those that could explain its poor prognosis (4). Most studies have focused on a single marker or a few markers such as hormone receptors [estrogen receptors (ERs), progesterone receptors (PRs)], growth factors (ERBB2, epidermal growth factor receptor) and tumor suppressors (P53; ref. 5). Experimental models have recently led to the identification of genes involved in IBC, such as *ARHC*, coding for the RhoC GTPase, and *WISP3*, coding for a *S*-(2-chloro-1,1,2-trifluoroethyl)glutathione (CTGF)-related protein (5).

New genomic approaches such as DNA microarrays (6) and serial analysis of gene expression (SAGE; ref. 7) provide unprecedented tools to tackle the complexity of cancers. Comprehensive gene expression profiles of NIBC defined with the use of DNA microarrays have revealed tumor subtypes (8, 9) and expression signatures that could improve prognostic classification (9–16). DNA microarrays have recently been used in IBC to identify genes deregulated by RhoC overexpression in cell lines (17) but have thus far not been applied to clinical specimens.

We used DNA microarrays for monitoring the RNA expression of ~8.000 genes in breast cancer samples from 81 patients, including 37 with IBC. Our objective was to investigate the transcriptional profiles of IBC and to search for molecular signatures that correlate with the IBC type and the pathological complete response to primary chemotherapy.

MATERIALS AND METHODS

Patients and Samples. Eighty-one cancer samples were profiled by using DNA microarrays. They were obtained from 81 patients with breast adenocarcinoma who had undergone initial surgery in Institut Paoli-Calmettes. Each patient gave written informed consent. Samples were macrodissected and were frozen in liquid nitrogen within 30 minutes of removal. All of the medical records and tumor sections were *de novo* reviewed before analysis. Profiled specimens contained more than 60% tumor cells as assessed before RNA extraction. The 81 samples included 37 pretreatment samples from 37 patients with IBC, selected by using the following criteria: T_{4d} tumor (TNM-UICC classification) and/or presence of superficial dermal lymphatic invasion. The 44 other samples represented histoclinical forms of NIBC: locally advanced (T_3, T_{4a}, T_{4b}, T_{4c} of the TNM-UICC classification, 14 cases), and localized with (18 cases) or without (12 cases) pathological axillary lymph node involvement. For each of these four clinical forms, samples were consecutive and selected on the above-cited criteria and on the availability of good quality RNA. Main histoclinical characteristics of patients are listed in Table 1. Immunohistochemical data collected included ER, PR, and P53 status (positivity cutoff values of 1%), and ERBB2 status (0 to 3+ score as illustrated by

Received 7/29/04; revised 9/22/04; accepted 10/1/04.

Grant support: Grants from the Ministries of Health and Research (Cancéropôle) and from the Association pour le Recherche contre le Cancer (Printemps 2002, N°4719) and Ligue Nationale contre le Cancer (label).

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: This paper is dedicated to the memory of Dr. Anne-Chantal Braud. Supplementary data for this article can be found at Cancer Research Online (http://cancerres.aacrjournals.org).

Requests for reprints: Daniel Birnbaum, Département d'Oncologie Moléculaire, UMR599 INSERM, IFR137, 27 Boulevard Leï Roure, 13009 Marseille, France. Phone: 33-4-91-75-84-07; Fax: 33-4-91-26-03-64; E-mail: birnbaum@marseille.inserm.fr.

Table 1 *Tumor characteristics*

Characteristics *	Total ($N = 81$)	NIBC ($n = 44$)	IBC ($n = 37$)	P value †
Age, median (range), years	58 (24–86)	62 (39–86)	55 (24–81)	0.12
Pev status (81)				<0.01
Pev 0–1	53 (65%)	44 (100%)	9 (24%)	
Pev 2	14 (17.5%)	0 (0%)	14 (38%)	
Pev 3	14 (17.5%)	0 (0%)	14 (38%)	
Clinical axillary node status (78)				0.07
N_0	31 (40%)	21 (48%)	10 (29%)	
N_1, N_2, N_3	47 (60%)	23 (52%)	24 (71%)	
Histological type (81)				0.19
ductal	67 (83%)	34 (77%)	33 (89%)	
lobular	11 (14%)	7 (16%)	4 (11%)	
other	3 (3%)	3 (7%)	0 (0%)	
SBR grade (81)				0.07
1–2	38 (47%)	25 (57%)	13 (35%)	
3	43 (53%)	19 (43%)	24 (65%)	
Dermal lymphatic emboli (81)				<0.01
Present	26 (32%)	0 (0%)	26 (70%)	
Absent	55 (68%)	44 (100%)	11 (30%)	
Angioinvasion (79)				<0.01
Present	42 (53%)	16 (36%)	26 (74%)	
Absent	37 (47%)	28 (64%)	9 (26%)	
ER status (81)				0.36
Negative	32 (39%)	15 (34%)	17 (46%)	
Positive	49 (61%)	29 (66%)	20 (54%)	
PR status (81)				0.06
Negative	30 (37%)	12 (27%)	18 (49%)	
Positive	51 (63%)	32 (73%)	19 (51%)	
ERBB2 status (81)				0.79
Negative 0–1+	61 (75%)	34 (77%)	27 (73%)	
Positive 2–3+	20 (25%)	10 (23%)	10 (27%)	
P53 status (81)				0.35
Negative	51 (63%)	29 (66%)	22 (59%)	
Positive	30 (37%)	15 (34%)	15 (41%)	

Note. For all characteristics except age, the values are number of patients with percentage of evaluated cases in parentheses.

Abbreviations: Pev 0–1, no inflammatory sign; Pev 2 and Pev 3, localized and extensive inflammatory signs, respectively; SBR, Scarff-Bloom-Richardson.

* In this column in parentheses, the number of evaluated cases among 81 patients.

† Not significant when $P > 0.05$.

the HercepTest kit (DakoCytomation, Glostrup, Denmark) scoring guidelines: 0 to 1+, negative; 2 to 3+, positive). Our series of IBCs displayed characteristics similar to those in other series in the literature (1). Seventy-one percent of our patients had clinically palpable axillary lymph nodes. Most tumors were ductal carcinoma, with relatively frequent high Scarff-Bloom-Richardson grade, negative hormone receptor status, and positive ERBB2 and P53 status. After diagnostic biopsy, IBC patients were treated by using a multimodality approach, including anthracycline-based chemotherapy followed by mastectomy (for clinically nonprogressive and consenting patients) then radiotherapy. Mastectomy specimens were examined to determine the pathological response to chemotherapy. Analysis concerned several tissue sections (a minimum of 20 per specimen) taken from each quadrant, from the nipple areolar complex, and from suspected areas of tumor involvement. Response was scored in four grades as described previously (18). Grades 1 and 2 were considered as pathological complete response-positive, and grades 3 and 4 as failures (pathological complete response-negative). Axillary lymph nodes, when available postchemotherapy (for some patients, lymph node dissection was made at diagnosis before any chemotherapy), were examined to determine the presence or absence of tumor residue. Response was classified as failure in case of grade 1 or 2 response with positive lymph node.

In addition, two RNA samples from normal breast (NB) tissue (Clontech, Palo Alto, CA), representing pools of six (NB1) or four (NB2) whole normal breasts, were profiled, as well as RNA from 14 cell lines that provided models for cell types of mammary tissues—IBC epithelial (SUM-149), NIBC luminal epithelial (BT-474, MCF-7, MDA-MB-134, SK-BR-3, T47D), NIBC stromal-like epithelial (MDA-MB-231, BT-549), and basal-like epithelial (HME-1) cell lines—and for cell types of nonmammary tissues—endothelial (human umbilical vein endothelial-cell), fibroblastic (HFFB), lymphocytic B and T (Daudi and T-act, respectively) and macrophage (THP1) cell lines. All of the cell lines were obtained from American Type Culture Collection (Manassas, VA),[9] except for SUM-149 (a kind gift of S. P. Ethier, Department of Radiation Oncology, University of Michigan Comprehensive Cancer Center, Ann Arbor, MI), and were grown as recommended by the supplier.

RNA Isolation. Total RNA was extracted from frozen samples by using guanidium isothiocyanate and cesium chloride gradient, as described previously (19). RNA integrity was controlled by denaturing formaldehyde agarose gel electrophoresis and by microanalysis (Agilent Bioanalyzer, Palo Alto, CA).

DNA Microarray Production and Hybridization. Gene expression analyses were done with home-made cDNA-spotted nylon microarrays and radioactive detection. Microarrays were produced as described previously (20). Briefly, cDNA clones were polymerase-chain-reaction–amplified in microtiter plates, then spotted onto Hybond-$n + 2 \times 7$ cm^2 membranes (Amersham) with a 64-pin print head on a MicroGrid II microarrayer (Apogent Discoveries, Cambridge, England). They contained 8,016 spotted cDNA clones, representing 7,874 IMAGE clones and 142 control clones. The IMAGE clones included 6,664 named genes (155 Unigene release), ~2,500 selected specifically to be related to oncogenesis, and 1,210 expressed sequence tags. All of the membranes belonged to the same printing batch. Before RNA hybridization, the quality of spotting, including the determination of target DNA amount accessible for each spot, was controlled by hybridization with a ^{33}P-oligonucleotide sequence common to all polymerase chain reaction products. After stripping, microarrays were hybridized with ^{33}P-probes made from total RNA (21, 22). To verify the reproducibility of the experiments, six samples (one cell line, two IBCs, three NIBCs) were hybridized twice on different microarrays, resulting in a total of 103 hybridizations. Probe preparations, hybridizations, and washes were done as described previously (20). Briefly, 2 μg of total RNA were retrostranscribed (oligo-d_T priming) in the presence of [α-^{33}P]dCTP (Amersham Pharmacia Biotech, Little Chalfont, United Kingdom). Hybridizations were carried out during 48 hours at 68°C in a volume of 500 μL of buffer. After washes, arrays were exposed for 24 to 72 hours to phosphorimaging plates. Signal detection was done with a FUJI BAS 5000 machine at 25-μm resolution (Raytest, Paris, France) and quantification with the ArrayGauge software (Fuji Ltd, Tokyo, Japan). The full list of clones and more details on the preparation of the microarrays and probes and hybridizations are available on the web.[10]

Data Analysis. Before analysis, a filter procedure eliminated noninformative genes; 2,300 genes were retained on the basis of being measured (expression level $>2 \times$ background signal) in at least 50% of the samples in at least one of the three classes (IBC, NIBC, NB). Signal intensities were normalized for the amount of spotted DNA (21), then the variability of experimental conditions were normalized by using a local weighted scatter plot smoother analysis (LOWESS) for each print-tip group (23). Data were then $\log_2$-transformed and were analyzed by unsupervised and supervised methods. Expression data and histoclinical variables of tumor samples, in a format conforming to the MIAME guidelines[11] are available on the Worldwide Web.[10] Unsupervised hierarchical clustering [Cluster program (24)] investigated the relationships between the genes and between the samples by using data that were median-centered on genes, Pearson correlation, and centroid linkage clustering. Results were displayed by using the TreeView program (R.D.M. Page, University of Glasgow, Glasgow, United Kingdom) (24). To identify the gene clusters that were most responsible for the resulting subdivision of samples, we used the method of quality-threshold clustering (25). We first selected the gene clusters with minimal size and minimal correlation of 10 and 0.6, respectively. Their average expression profile was then computed and submitted to supervised analysis (see below), to identify the most discriminating profile between the two predominant sample clusters. Supervised analysis was applied to identify and rank genes that discriminate between two relevant subgroups of samples. A discriminating score (*DS*) was calculated for each gene (26) as $DS = (M1 - M2)/(S1 + S2)$, where *M1* and *S1*, respectively, represent mean and SD of expression levels of the gene in subgroup 1, and *M2* and *S2* represent those levels in subgroup 2. Because of multiple hypotheses testing, confidence levels were estimated by 200 iterative random permutations of samples as described previously (27) and by computing the proportion of permutations in which the number of genes selected exceeds the observed number of genes. Once identified, the classification power of the discriminator

[9] Internet address: http://www.atcc.org/.

[10] Internet address: http://tagc.univ-mrs.fr/pub/.

[11] Internet address: http://www.mged.org/miame.

signature was illustrated by classifying samples according to the correlation coefficient of their expression profile with the median profile of the IBC samples ("IBC signature") or of the pathological complete response–positive IBC samples ("PCR signature"). The "IBC signature" was validated on an independent sample set. A "leave-one-out" procedure (26) estimated the accuracy of prediction of the "PCR signature." Statistical analyses were done by using the SPSS software (version 10.0.5, SPSS Inc., Chicago, IL). Correlations between sample groups and histoclinical parameters were calculated with the Fisher exact test or χ^2 test when appropriate. A P value < 0.05 was considered significant.

RESULTS

A total of 103 samples representing 97 different cases (81 cancer tissue samples including 37 IBC and 44 NIBC, 2 NB samples, and 14 cell lines) were profiled by using DNA microarrays.

Unsupervised Hierarchical Clustering Based on Global Gene Expression Profiles. Before clustering, a filter procedure eliminated genes with uniformly low expression or with low expression variation across the experiments, retaining 2,300 genes/expressed sequence tags. Results of hierarchical clustering are shown in Fig. 1. The cancer tissue samples displayed heterogeneous expression profiles (Fig. 1*A* and *B*). Overall, they fell in two groups that significantly differed with respect to the IBC or NIBC type. In the left group, which included the two NB samples, 25 of the 42 cancer samples (60%) were IBC, whereas in the right group, 14 of the 44 samples (31%) were IBC ($P = 0.008$, Fisher exact test). Correlations existed between the two groups and the IHC status of tumors for ER ($P < 0.001$), ERBB2 ($P = 0.003$), P53 ($P = 0.001$), and angioinvasion ($P = 0.051$). As expected, all of the cell lines represented separate branches of the dendrogram.

Gene clustering revealed groups of coordinately expressed genes. Some of these represented signatures of biological processes or cell types (see *colored bars* on the right of Fig. 1*A* and *zooms* in Fig. 1*C*). A cluster with a prominent role in the classification of samples included *ESR1,* which codes for ER-α, several transcription factor genes (*GATA3, XBP1, ILF1, GLI3, PBX1*), and genes associated with ER-positive status (*KRT19*, *CCND1*, *EMS1*, *MUC1*). This cluster, overexpressed in luminal ER-positive NIBC cell lines as compared with the basal-like cell line and the IBC cell line, was designated "luminal/ER+ cluster." Variation in expression of *ESR1* mRNA correlated well with IHC ER status. The "ERBB2-related cluster," overexpressed in cell lines with amplification of *ERBB2,* included *ERBB2, GRB7,* and *PPARBP,* identified as part of an ERBB2 gene expression signature (28). As reported elsewhere (8, 10), the "early response cluster" included immediate-early genes (*JUNB*, *FOS*, *ATF3*, *EGR1*, *NR4A1*, *DUSP1*) and was overexpressed in normal samples overall as compared with cancer samples. A "proliferation cluster" was globally overexpressed in cell lines as compared with tissues. It included *PCNA,* which codes for a proliferation marker used in clinical practice, and genes involved in glycolysis (*GAPD*, *LDHA*, *ENO1*), metabolism (*ALDH3A1*, cytochrome *c* oxidase and ATP synthase subunits), cell cycle and mitosis (tubulin genes, *CDK4*, *BUB3*, *CCNB2*), and protein synthesis (ribosomal proteins; not shown). The "immune cluster" was rich in genes expressed in B- or T-cells and macrophages (immunoglobulin genes, HLA class I and II, *CD69*, *IL16, CD14*, *CSF1, CSF1R*, and genes regulated by interferon such as *STAT1*, *B2M*, *IFI27*). It was globally negatively correlated with the "luminal/ER+ cluster," reflecting the strong lymphoid infiltrate of ER-negative tumors. The "stromal cluster" was rich in genes related to extracellular matrix remodeling (collagen genes, *MMP2*, *PRSS3*, *SPARC*, *EDN1*) and strongly expressed in the fibroblastic cell line. The "vascular cluster," strongly expressed in the human umbilical vein endothelial-cell cell line, contained genes related to endothelial cells (*ENG*, *VWF*, *CD31*, *CDH5*, *HSPG2*, *FN1*). Another cluster, designated "basal cluster," included cytokeratins (*KRT5*, *6, 7, 13, 14, 15, 16*), integrins (*ITGA2, ITGA6, ITGB4*), and other genes (*COL17A1*, *EGFR*, *laminin*, *TRIM29, CRYAB, SLPI*) and was overexpressed in the basal-like cell line. This cluster was also overexpressed in the IBC cell line as compared with the NIBC cell lines.

To identify objectively the gene clusters most responsible for the subdivision of samples into two main groups (Fig. 1*B*), we applied supervised analysis to the 24 gene clusters identified by quality-threshold clustering analysis. Two subclusters from the "luminal cluster" were significantly overexpressed in the right group (rich in NIBC). Two subclusters from the "basal cluster," one subcluster from the "immune cluster" and another from the "vascular cluster" were significantly overexpressed in the left group (rich in IBC). The two most discriminating subclusters came from the "luminal" and the "basal cluster."

Gene Expression Signature for Inflammatory Breast Cancer Identified by Supervised Analyses. To identify a gene expression signature that discriminated IBC from NIBC samples, we applied supervised analysis by using two independent (learning and validation) tumor sets. The assignment of samples to each set was random, but preserved the IBC/NIBC ratio. The learning set (55 samples: 25 IBCs, 30 NIBCs) was used to define the gene expression signature. Using a discriminating score combined with permutation tests, we identified 109 cDNA clones differentially expressed between IBCs and NIBCs. The significance threshold used produced fewer than five false positives and ensured that the number of genes selected by chance, given 200 iterative random permutations, never exceeded 109. Sixty-four clones were overexpressed and 45 were underexpressed in IBC samples. They represented 90 characterized genes and 19 other sequences or expressed sequence tags (Supplementary Table 1).[12] The classification of 55 samples based on these 109 genes is shown in Fig. 2*A*. A threshold of 0 (*solid line* in Fig. 2*A*) sorted the samples into two classes ("predicted IBC class," positive scores; "predicted NIBC class," negative scores) that strongly correlated with the observed histoclinical type: 79% of the 28 "predicted IBC class" samples were IBC, whereas 89% of the 27 "predicted NIBC class" samples were NIBC [odds ratio (OR) = 26.78; 95% confidence interval (CI), 5.59–187.9; $P = 4.10^{-7}$, Fisher exact test]. A more stringent threshold improved the accuracy of discrimination: for example, with a cutoff of 0.2 and -0.2 (*dashed lines* in Fig. 2*A*), 85% of the "predicted IBC class" samples were IBC and 90% of the "predicted NIBC class" samples were NIBC, leaving some samples unclassifiable.

To estimate its robustness, we tested this gene expression signature on a set of 26 independent samples (12 IBC, 14 NIBC). None of these samples had been included in the learning set, which allowed for the estimation of the true predictive accuracy. Samples were classified by using the same procedure (Fig. 2*B*). The two predicted classes strongly correlated with the distinction between IBC and NIBC ratio. There were 10 IBCs (83%) of the 12 samples in the "predicted IBC class" and 12 NIBCs (86%) of the 14 samples in the "predicted NIBC class" (OR = 24.43; 95% CI , 2.71–414.2; $P = 0.001$, Fisher exact test), with a prediction accuracy of 85%. The IBC cell line was within the "predicted IBC class." These results suggest the robustness of our model for discriminating IBCs and NIBCs.

Gene Expression Signature for Pathological Complete Response in Inflammatory Breast Cancer. Among the 37 IBC samples, 26 mastectomy specimens were available for assessment of pathological complete response: 9 were defined as pathological complete response-positive and 17 as pathological complete response-

[12] Supplementary data for this article can be found at Cancer Research Online (http://cancerres.aacrjournals.org).

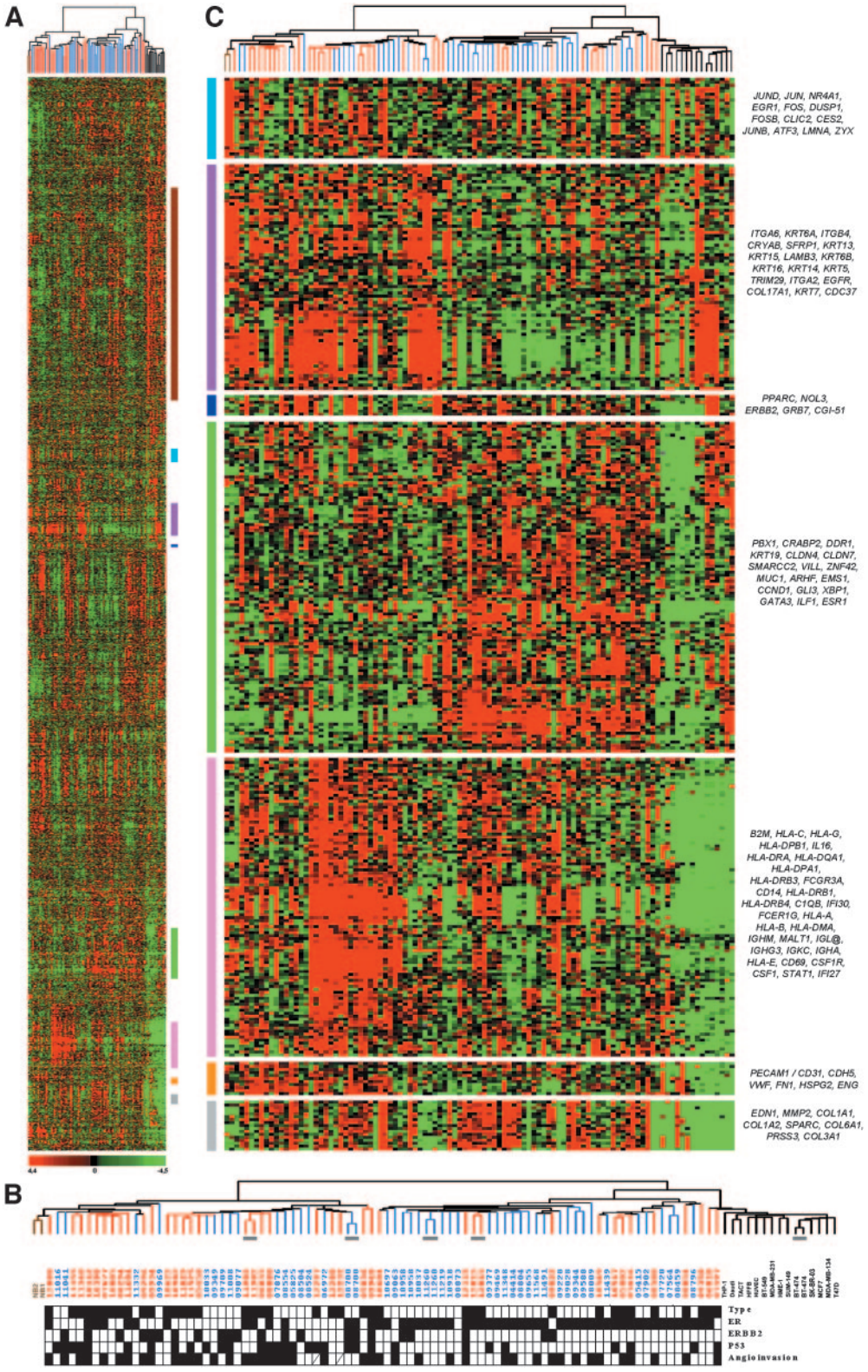
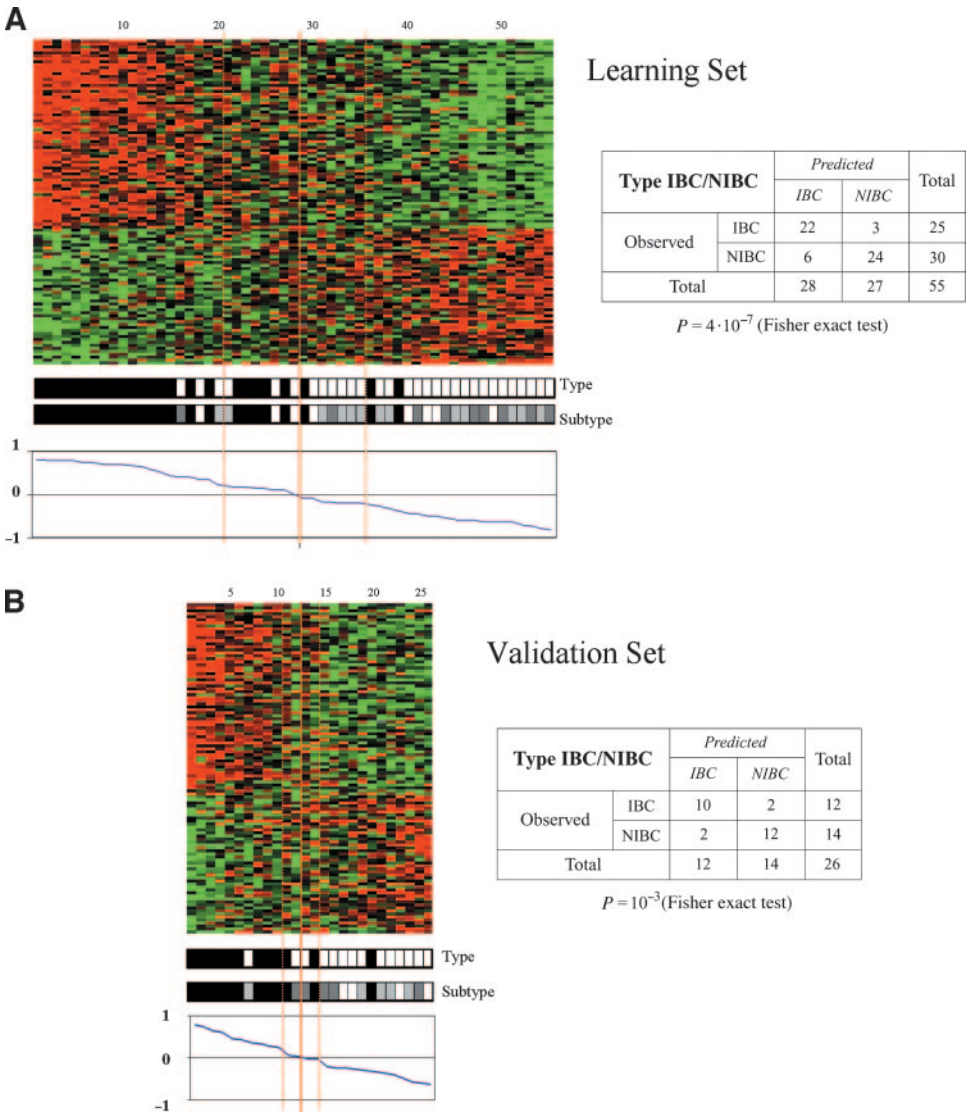


Fig. 1. Global gene expression profiles in inflammatory and noninflammatory breast cancer. *A*, hierarchical clustering of 103 samples and 2,300 cDNA clones based on mRNA expression levels. *Each row*, a clone; *each column*, a sample. *Color scale across the bottom,* the expression level of each gene in a single sample relative to its median abundance across all samples; *red*, expression level above the median; *green*, expression level below the median; *color saturation*, the magnitude of deviation from the median; *gray*, missing data. *Above matrix, dendrogram of samples*, overall similarities in gene expression profiles (zoomed in *B*). *Colored bars to the right*, the locations of eight gene clusters of interest. [These clusters, except the "proliferation cluster" (*brown bar*), are zoomed in *C*.] *B*, *top*, dendrogram of samples (*from dendrogram in A*): *red branches*, IBC samples ($n = 39$); *blue branches*, NIBC samples ($n = 47$); *brown branches*, NB samples ($n = 2$); *black branches*, cell lines ($n = 15$); *numbers below the dendrogram*, tissue samples; *five small gray horizontal bar*s, five pairs of duplicate samples clustered together. *Black and white chart below the numbers*, some relevant features of numbered samples (▨, unavailable): *Type* (*white,* NIBC; *black,* IBC); *ER*, ER immunohistochemical (IHC) status (*white,* negative; *black,* positive); *ERBB2*, ERBB2 IHC status (*white*, negative; *black*, positive); *P53*, P53 IHC status (*white*, negative; *black*, positive); *Angioinvasion* (*white*, negative; *black*, positive). *C*, expanded view of selected gene clusters. *On right side*, names of genes; some genes included in these clusters are referenced by their Human Genome Organization (HUGO) abbreviation as used in "Entrez Gene" (http://www.ncbi.nih.gov/entrez). *On left side*, *from top to bottom*: *light blue bar*, early response; *dark pink bar*, basal; *dark blue bar*, ERBB2-related; *green bar*, luminal/ER+; *light pink bar*, immune; *orange bar*, vascular; *gray bar*, stromal.



Type IBC/NIBC		Predicted		Total
		IBC	NIBC	
Observed	IBC	22	3	25
	NIBC	6	24	30
Total		28	27	55

$P = 4 \cdot 10^{-7}$ (Fisher exact test)

Type IBC/NIBC		Predicted		Total
		IBC	NIBC	
Observed	IBC	10	2	12
	NIBC	2	12	14
Total		12	14	26

$P = 10^{-3}$ (Fisher exact test)

Fig. 2. Supervised classification of 81 samples based on the molecular signature IBC/NIBC. *A*, *left*, classification of 55 samples from the learning set with the 109-gene expression signature. *Top panel*, *each row of the data matrix*, a gene; *each column*, a sample. Expression levels are depicted according to the color scale used in Fig. 1. Genes *from top to bottom*, ordered by their decreasing discriminating score. Tumor samples are numbered from 1 to 55 and are ordered *from left to right* according to the correlation coefficient of their expression profile with the median profile of the IBC samples (*bottom panel*). *Solid orange line*, the threshold 0 that separates the two classes of samples: IBC class (*left of the line*) and NIBC class (*right of the line*). *Dashed orange line*, optimized threshold. *Middle panels*, the histoclinical type of tumors (■, IBC; □, NIBC) and their histoclinical subtype (■, IBC; *dark gray square*, locally advanced NIBC; *light gray square*, node-positive localized NIBC; □, node-negative localized NIBC). *Right*, correlation between the molecular grouping based on the combined expression of the 109 genes and the histoclinical type of samples. *B*, *left* and *right*, see legend for *A*, but applied to the 26 independent samples from the validation set.

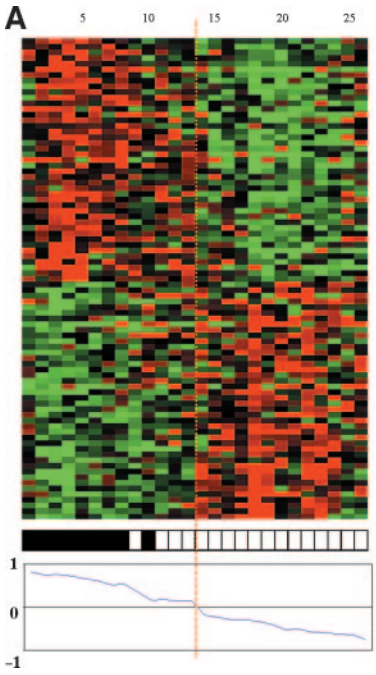
negative. No significant correlation existed between the pathological complete response rate and Scarff-Bloom-Richardson grade, dermal lymphatic emboli, angioinvasion, immunohistochemical data (Supplementary Table 2). Using a supervised analysis, we identified 85 genes/expressed sequence tags (Supplementary Table 3) that discriminated between pathological complete response-positive and pathological complete response-negative tumors (probability that this number of genes would be selected by chance, 0.09). Forty-three of them were overexpressed in pathological complete response-positive samples and 42 were underexpressed. The resulting classification of samples is shown in Fig. 3. The cutoff value of 0 defined two classes strongly correlated with the rate of pathological complete response. Nine of 13 "predicted pathological complete response-positive class" samples (70%; positive scores) experienced pathological complete response, as compared with 0 of 13 of the "predicted pathological complete response-negative class" (negative scores; $P = 0.0004$, Fisher exact test), leading to a classification accuracy of 85%.

We estimated the validity of our procedure by the "leave-one-out" cross-validation method (26). Iteratively, one of the 26 samples was removed, and a multigene predictor was generated from the remaining samples. The "leave-one-out" sample was then classified by using this predictor and the procedure described above. The process was repeated for each of the samples, and the rate of correct classification was calculated. Sixty-two percent of samples were correctly assigned, with a rate of 67% for pathological complete response-positive and 59% for pathological complete response-negative samples, and positive and negative predictive values for PCR of 46 and 77%, respectively. Although the predictive gene set generated at each cross-validation loop was slightly different, on average, 85% of the genes of the gene expression signature were conserved.

DISCUSSION

Unsupervised Hierarchical Clustering Identifies Subclasses of Breast Cancer. Both NIBC and IBC samples showed great transcriptional heterogeneity, indicating the existence of molecular subclasses among each of them. Yet despite this diversity, the unsupervised global approach produced two large groups, one with 2-fold more IBC samples than the other, suggesting a global expression difference between IBC and NIBC in most cases. Despite the differences in microarray platforms, all of the identified gene clusters were similar to those previously reported (8–10, 13), suggesting the validity of the data and reliability of the technology.

As compared with the group rich in NIBC, the group rich in IBC exhibited overexpression of the basal, the immune, and the vascular gene clusters and underexpression of the luminal cluster. These ex-



B

Pathological Response to primary chemotherapy		Predicted		Total
		PCR+	PCR–	
Observed	PCR+	9	0	9
	PCR–	4	13	17
Total		13	13	26

$P = 4 \cdot 10^{-4}$ (Fisher exact test)

Fig. 3. Supervised classification of 26 IBC samples based on the molecular signature for the pathological complete response (*PCR*). *A*, classification of 26 samples by using the 85-gene expression signature. *Top panel*, expression levels according to the color scale used in Fig. 1. Genes are ordered from top to bottom by their decreasing discriminating score. Tumor samples are numbered from 1 to 26 and are ordered from left to right according to the correlation coefficient of their expression profile with the median profile of the *PCR+* samples (*bottom panel*). *Solid orange line* (threshold = 0), separates the two classes of samples, "*PCR+* class" (at the *left of the line*) and "*PCR−* class" (*right of the line*). The *middle panels* indicate the pathological response to primary chemotherapy (■, *PCR+*; □, *PCR−*). *B*, correlation between the molecular grouping based on the combined expression of the 85 genes and the pathologic response.

pression changes were in agreement with the phenotypical characteristics of IBC and NIBC, and suggest that IBC is related to the basal lineage more frequently than is NIBC. In fact, none of the differential expressions revealed by this approach appeared completely specific to IBC or NIBC but, rather, reflected the luminal or basal-like phenotypes.

Identification of a Gene Expression Signature for Inflammatory Phenotype in Breast Cancer. By supervised analysis, we identified a 109-gene signature that discriminated IBC and NIBC. This classifier had comparable prediction accuracy (85%) in an independent set of samples, providing evidence of its robustness. The molecular distinction was not strict for all samples, with a large range of intermediate profiles between the respective "typical profiles" for IBC and NIBC. Among all these profiles, there was no particular organization of samples with respect to clinically or pathologically defined IBC or according to the histoclinical form of NIBC (Fig. 2). The correction for testing multiple hypotheses may have caused the elimination of potentially interesting genes. For example, *ARHC* (29) did not pass our stringent threshold, although it was up-regulated in our series of IBC ($P = 0.002$ and 0.03 for the two corresponding clones, Mann–Whitney test without correction for multiple comparisons). Similarly, *CTGF,* which codes for a protein with 57% similarity to *WISP3* (29), was also underexpressed ($P = 0.004$, Mann–Whitney test without correction), but not included in our signature.

This gene expression signature represents a bar code signature of the IBC phenotype. Whether the discriminator genes are causative or even predictive of the phenotype in a biological sense or whether they reflect another associated phenomenon remain to be explored. Several genes are related to signal transduction, cell motility, invasion, and angiogenesis. Genes overexpressed in IBC included *ARHQ*, a member of the Rho GTPase family involved in cytoskeletal organization and cell motility (30); *RAB1A,* a small GTPase; tyrosine kinase *SYK*; and *FNTA*. *FNTA* encodes the farnesyltransferase α subunit; van Golen *et al.* (31) reported that treatment of the SUM149 cell line and HME-RhoC transfectants with a farnesyl transferase inhibitor reversed the RhoC-induced phenotype, with a significant decrease in motility and invasion. The same authors suggested the involvement of the mitogen-activated protein kinase (MAPK) pathway in RhoC-induced motility, invasion, and angiogenesis in IBC (32). Here, we identified genes encoding MAPK1 and STK24, a serine/threonine kinase that functions upstream of the MAPK cascade, as overexpressed in IBC. Overexpressed genes also included genes from the "basal cluster," *CDC37*, involved in cell cycle regulation, and *ITGB4*, which promotes carcinoma invasion (33). Some of the proteins encoded by other overexpressed genes also stimulate cell motility: VASP plays a role in integrin-mediated cell adhesion (34), and CACNB1 (calcium channel, voltage-dependent, β 1 subunit), AKAP1, and AKAP7 (A kinase (PRKA) anchor protein 1 and 7) are involved in calcium signaling (35). Other genes found overexpressed are involved in local inflammatory processes (*CXCL2*, *BMP4*, *SCGB1A1*, *FPRL1*, *VCAM1*), cell cycle (*CCNG2*, *CDC37*, *CCT2*), apoptosis (*DAD1*, *ALS2CR2*), transport (*CRABP1*, *SLC18A2*, *SLC22A4*, *SLC2A12*), and transcription (*ARNT*, *DTR*, *NPAS2, SIX3*). *ARNT* encodes the β subunit of hypoxia-inducible factor 1 (HIF1), involved in angiogenesis and tumor progression (36). Increased expression of several genes involved in carbohydrate metabolism (*PDPK1*, *FUCA1*, *GAPD*, *RPN2*) and protein synthesis (*RPL13A*, *RPS2*, *RPS6KA4*, *MBNL1*) was associated with the IBC phenotype, possibly related to increased metabolism and cell proliferation in IBC. Genes found underexpressed in IBC, such as *BRE* (37), *GPC4* (38), *THBS4* (39), and *PTPRA* also encoded proteins involved in negative regulation of cell motility, invasion, or angiogenesis. Finally, the analysis pointed to some interesting chromosomal regions. Five genes down-regulated in IBC (*NDUFS4*, *THBS4*, *BTF3*, *COX7C*, *RIOK2*) were located at 5q11–14. This result, combined with the higher frequency of basal-like tumors among IBCs, may be related to a significantly higher rate of loss of heterozygosity at 5q in basal-like breast cancers (40). Two genes down-regulated in IBC (*PSMB8*, *CSNK2B*) are found on 6p21, which harbors loss of heterozygosity more frequently in IBC than in NIBC (41). Conversely, three genes up-regulated in IBC (*CXCL2*, *CCNG2*, *MASA/E-1*) were located at 4q21.1, in a 10-Mb-long region that contains many genes encoding pro-inflammatory cytokines and growth factors. A 2-Mb-long region at 22q11.21 contains three genes (*RTN4R, PIK4CA, MAPK1*) up-regulated in IBC. These regions of co-up-regulated or -down-regulated genes may correspond to genome alterations specific to IBC.

Identification of a Gene Expression Signature for Pathological Complete Response in Inflammatory Breast Cancer. IBC is mostly treated with anthracycline-based primary chemotherapy. Results are disappointing, with 5-year survival ranging from 30 to 50% (1). No clinical or molecular marker has been found that reliably predicts pathological complete response to such chemotherapy (42, 43). As a consequence, chemotherapy is delivered empirically to all patients. Current efforts are, therefore, aimed at discovering molecular markers that would help clinicians to select an alternative chemotherapy or another systemic treatment that would improve response rate.

In our series, pathological complete response after anthracycline-based chemotherapy was observed in 35% of mastectomy specimens. No correlation was found between response and any tested histoclinical condition. Global unsupervised hierarchical clustering showed no separation between pathological complete response-positive and pathological complete response-negative samples, which suggests that the response to chemotherapy is governed by a smaller set of genes. Supervised analysis identified a set of 85 genes the expression of which divided patients in two groups with, respectively, 70% and 0% of pathological complete response. Because of the small number of cases, estimation of the classifier performance was not done in an independent series, but by the use of leave-one-out cross-validation, a classical alternative method. Because this method may overestimate accuracy, it will be necessary to validate this signature with a larger independent sample series. Regardless of the small sample size, the respective positive and negative predictive values for pathological complete response of ~50 and ~80% are highly encouraging in the current clinical context, in which the expectation of any unselected IBC patient to achieve pathological complete response is less than 15 to 25% (1). This classifier is a first step in achieving a criterion on which to negatively or positively select for the most efficient therapy for patients.

Additional experiments are required to investigate the role of some of the discriminator genes in response to therapy. It was interesting to find genes already reported as associated with drug sensitivity. For example, a high expression of *CDKN1B* (p27) was associated with pathological complete response, as previously reported in acute myeloid leukemia (44). Two recent reports identified gene expression signature associated with response to primary chemotherapy in locally advanced breast cancer (15, 16). Comparison of the discriminator genes with ours revealed several genes (*AK3*, *ATP6V1F*, *EIF3S9*, *MRPL4*, *APOD*, *PPP5C*) that belong to the same families. *CCL3* encoding the MIP1A chemokine is overexpressed in sensitive B-cell chronic lymphocytic leukemia cell samples (45). The up-regulation in pathological complete response-positive samples of genes encoding other chemokines, cytokines, and cytokine receptors (*CSF1R*, *CCL2*, *CCL3*, *MMP9*) is consistent with a role of the host immune system in tumor eradication after chemotherapy.

Our study is the first example of high-throughput gene profiling applied to clinical specimens of IBC. Although obtained on a small series of samples (IBC is a rare disease), our results are encouraging. They show the potential for a better understanding of this particular and aggressive form of breast cancer and for the identification of new diagnostic and predictive factors and potential therapeutic targets. Further validation on a larger and multicentric series of samples is warranted, as well as additional investigations of our discriminator genes to determine their relevance in the aggressiveness of IBC and possible therapeutic utility.

ACKNOWLEDGMENTS

We thank L. Bachelart and S. Jeangirard for assistance at the beginning of this work.

REFERENCES

1. Jaiyesimi IA, Buzdar AU, Hortobagyi G. Inflammatory breast cancer: a review. J Clin Oncol 1992;10:1014–24.
2. Buzdar AU, Singletary SE, Booser DJ, et al. Combined modality treatment of stage III and inflammatory breast cancer. M. D. Anderson Cancer Center experience. Surg Oncol Clin N Am 1995;4:715–34.
3. Feldman LD, Hortobagyi GN, Buzdar AU, Ames FC, Blumenschein GR. Pathological assessment of response to induction chemotherapy in breast cancer. Cancer Res 1986;46:2578–81.
4. Cristofanilli M, Buzdar AU, Hortobagyi GN. Update on the management of inflammatory breast cancer. Oncologist 2003;8:141–8.
5. Kleer CG, van Golen KL, Merajver SD. Molecular biology of breast cancer metastasis. Inflammatory breast cancer: clinical syndrome and molecular determinants. Breast Cancer Res 2000;2:423–9.
6. Bertucci F, Houlgatte R, Nguyen C, et al. Gene expression profiling of cancer by use of DNA arrays: how far from the clinic? Lancet Oncol 2001;2:674–82.
7. Polyak K, Riggins GJ. Gene discovery using the serial analysis of gene expression technique: implications for cancer research. J Clin Oncol 2001;19:2948–58.
8. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature (Lond) 2000;406:747–52.
9. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 2003;100: 8418–23.
10. Bertucci F, Nasser V, Granjeaud S, et al. Gene expression profiles of poor-prognosis primary breast cancer correlate with survival. Hum Mol Genet 2002;11:863–72.
11. van 't Veer LJ, Dai H, van De Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature (Lond) 2002;415:530–6.
12. van de Vijver MJ, He YD, van't Veer LJ et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.
13. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci USA 2003;100:10393–8.
14. Huang E, Cheng SH, Dressman H, et al. Gene expression predictors of breast cancer outcomes. Lancet 2003;361:1590–6.
15. Chang JC, Wooten EC, Tsimelzon A, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. Lancet 2003;362:362–9.
16. Ayers M, Symmans WF, Stec J, et al. Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. J Clin Oncol 2004;22:2284–93.
17. Wu M, Wu ZF, Kumar-Sinha C, Chinnaiyan A, Merajver SD. RhoC induces differential expression of genes involved in invasion and metastasis in MCF10A breast cells. Breast Cancer Res Treat 2004;84:3–12.
18. Chevallier B, Chollet P, Merrouche Y, et al. Lenograstim prevents morbidity from intensive induction chemotherapy in the treatment of inflammatory breast cancer. J Clin Oncol 1995;13:1564–71.
19. Theillet C, Adelaide J, Louason G, et al. FGFRI and PLAT genes and DNA amplification at 8p12 in breast and ovarian cancers. Genes Chromosomes Cancer 1993;7:219–26.
20. Bertucci F, Salas S, Eysteries S, et al. Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. Oncogene 2004;23: 1377–91.
21. Bertucci F, Van Hulst S, Bernard K, et al. Expression scanning of an array of growth control genes in human tumor cell lines. Oncogene 1999;18:3905–12.
22. Bertucci F, Bernard K, Loriod B, et al. Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. Hum Mol Genet 1999;8:1715–22.
23. Yang YH, Dudoit S, Luu P, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 2002;30:e15.
24. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998;95:14863–8.
25. Wang Y, Jatkoe T, Zhang Y, et al. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. J Clin Oncol 2004;22:1564–71.
26. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science (Wash DC) 1999;286:531–7.
27. Magrangeas F, Nasser V, Avet-Loiseau H, et al. Gene expression profiling of multiple myeloma reveals molecular portraits in relation to the pathogenesis of the disease. Blood 2003;101:4988–5006.
28. 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.
29. van Golen KL, Davies S, Wu ZF, et al. A novel putative low-affinity insulin-like growth factor-binding protein, LIBC (lost in inflammatory breast cancer), and RhoC GTPase correlate with the inflammatory breast cancer phenotype. Clin Cancer Res 1999;5:2511–9.
30. Hall A. Rho GTPases and the actin cytoskeleton. Science (Wash DC) 1998;279: 509–14.
31. van Golen KL, Bao L, DiVito MM, et al. Reversion of RhoC GTPase-induced inflammatory breast cancer phenotype by treatment with a farnesyl transferase inhibitor. Mol Cancer Ther 2002;1:575–83.
32. van Golen KL, Bao LW, Pan Q, et al. Mitogen activated protein kinase pathway is involved in RhoC GTPase induced motility, invasion and angiogenesis in inflammatory breast cancer. Clin Exp Metastasis 2002;19:301–11.

33. Shaw LM, Rabinovitz I, Wang HH, Toker A, Mercurio AM. Activation of phosphoinositide 3-OH kinase by the alpha6beta4 integrin promotes carcinoma invasion. Cell 1997;91:949–60.
34. Samarin S, Romero S, Kocks C, et al. How VASP enhances actin-based motility. J Cell Biol 2003;163:131–42.
35. Huang JB, Kindzelskii AL, Clark AJ, Petty HR. Identification of channels promoting calcium spikes and waves in HT1080 tumor cells: their apparent roles in cell motility and invasion. Cancer Res 2004;64:2482–9.
36. Tacchini L, Matteucci E, De Ponti C, Desiderio MA. Hepatocyte growth factor signaling regulates transactivation of genes belonging to the plasminogen activation system via hypoxia inducible factor-1. Exp Cell Res 2003;290:391–401.
37. Bharti AC, Aggarwal BB. Nuclear factor-kappa B and cancer: its role in prevention and therapy. Biochem Pharmacol 2002;64:883–8.
38. Sanderson RD. Heparan sulfate proteoglycans in invasion and metastasis. Semin Cell Dev Biol 2001;12:89–98.
39. de Fraipont F, Nicholson AC, Feige JJ, Van Meir EG. Thrombospondins and tumor angiogenesis. Trends Mol Med 2001;7:401–7.
40. Wang ZC, Lin M, Wei LJ, et al. Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers. Cancer Res 2004;64:64–71.
41. Lerebours F, Bertheau P, Bieche I, et al. Evidence of chromosome regions and gene involvement in inflammatory breast cancer. Int J Cancer 2002;102:618–22.
42. Vincent-Salomon A, Carton M, Freneaux P, et al. ERBB2 overexpression in breast carcinomas: no positive correlation with complete pathological response to preoperative high-dose anthracycline-based chemotherapy. Eur J Cancer 2000; 36:586–91.
43. Viens P, Penault-Llorca F, Jacquemier J, et al. High-dose chemotherapy and haematopoietic stem cell transplantation for inflammatory breast cancer: pathologic response and outcome. Bone Marrow Transplant 1998;21:249–54.
44. Radosevic N, Delmer A, Tang R, Marie JP, Ajchenbaum-Cymbalista F. Cell cycle regulatory protein expression in fresh acute myeloid leukemia cells and after drug exposure. Leukemia (Baltimore) 2001;15:559–66.
45. Vallat L, Magdelenat H, Merle-Beral H, et al. The resistance of B-CLL cells to DNA damage-induced apoptosis defined by DNA microarrays. Blood 2003;101: 4598–606.