Human Molecular Genetics, 2000, Vol. 9, No. 20 **2981–2991**

Gene expression profiling of primary breast carcinomas using arrays of candidate genes

François Bertucci[1,3,4], Rémi Houlgatte[2], Athmane Benziane[1], Samuel Granjeaud[2], José Adélaïde[1], Rebecca Tagett[2], Béatrice Loriod[2], Jocelyne Jacquemier[5], Patrice Viens[3,4], Bertrand Jordan[2], Daniel Birnbaum[1,6,+] and Catherine Nguyen[2]

[1]Laboratoire de Biologie des Tumeurs, TAGC, Institut Paoli-Calmettes (IPC), IFR57, Marseille, France, [2]TAGC, CIML Luminy, IFR57, Marseille, France, [3]Département d'Oncologie Médicale, IPC, Marseille, France, [4]Université de la Méditerranée, Marseille, France, [5]Département d'Anatomie Pathologique, IPC, Marseille, France and [6]Laboratoire d'Oncologie Moléculaire, U119 INSERM, IFR57, Marseille, France

Received 19 July 2000; Revised and Accepted 19 October 2000

Breast cancer is characterized by an important histo-clinical heterogeneity that currently hampers the selection of the most appropriate treatment for each case. This problem could be solved by the identification of new parameters that better predict the natural history of the disease and its sensitivity to treatment. A large-scale molecular characterization of breast cancer could help in this context. Using cDNA arrays, we studied the quantitative mRNA expression levels of 176 candidate genes in 34 primary breast carcinomas along three directions: comparison of tumor samples, correlations of molecular data with conventional histoclinical prognostic features and gene correlations. The study evidenced extensive heterogeneity of breast tumors at the transcriptional level. A hierarchical clustering algorithm identified two molecularly distinct subgroups of tumors characterized by a different clinical outcome after chemotherapy. This outcome could not have been predicted by the commonly used histoclinical parameters. No correlation was found with the age of patients, tumor size, histological type and grade. However, expression of genes was differential in tumors with lymph node metastasis and according to the estrogen receptor status; *ERBB2* expression was strongly correlated with the lymph node status ($P < 0.0001$) and that of *GATA3* with the presence of estrogen receptors ($P < 0.001$). Thus, our results identified new ways to group tumors according to outcome and new potential targets of carcinogenesis. They show that the systematic use of cDNA array testing holds great promise to improve the classification of breast cancer in terms of prognosis and chemosensitivity and to provide new potential therapeutic targets.

INTRODUCTION

Pathologists and clinicians in charge of the management of breast cancer patients are facing two major problems, namely the extensive heterogeneity of the disease and the lack of factors—among conventional histological and clinical features—predicting with reliability the evolution of the disease and its sensitivity to cancer therapies. Breast tumors of the same apparent prognostic type vary widely in their responsiveness to therapy and consequent survival of the patient. New prognostic and predictive factors are needed to allow an individualization of therapy for each patient. Great hope is currently being placed on molecular studies which address the problem in a global fashion. Methods such as cytogenetics, comparative genomic hybridization and whole genome allelotyping have addressed the issue at the genome level. Currently, the modifications that take place in human tumors at the level of transcription can also be studied in a large, unprecedented scale, using new methods such as cDNA arrays that allow quantitative measurement of the mRNA expression levels of many genes simultaneously (1,2). Using arrays in which PCR products from cDNAs are arranged on a nylon membrane and hybridized with a radioactive probe made from relatively small amount of total RNA (3–5), we report the expression profiles of 176 candidate genes in a panel of 34 breast carcinomas and their analysis in correlation with conventional prognostic variables. The main objective of this study was to assess the capacity of cDNA array testing in clinical practice to better classify a heterogeneous cancer into tumor subtypes with more homogeneous clinical outcomes, and to identify new potential prognostic factors and therapeutics targets.

RESULTS

Data representation

Figure 1 shows examples of hybridizations of cDNA arrays with probes made from RNA extracted from normal breast tissue and breast tumors.

[+]To whom correspondence should be addressed at: Laboratoire d'Oncologie Moléculaire, U119 INSERM, IFR57, 27 Boulevard Leï Roure, 13009 Marseille, France. Tel: +33 4 91 75 84 07; Fax: +33 4 91 26 03 64; Email: birnbaum@marseille.inserm.fr

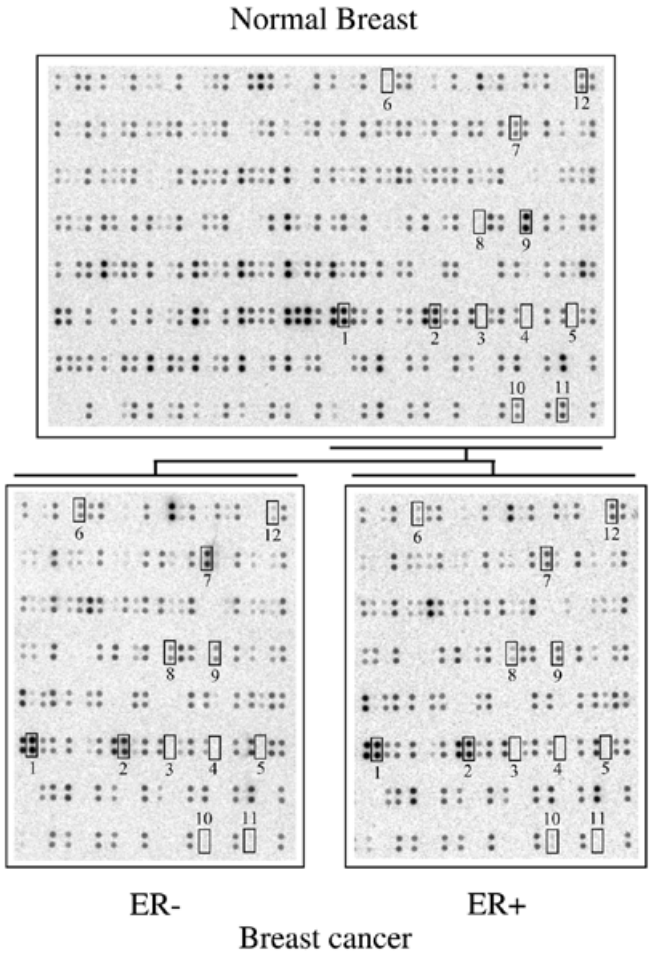


Figure 1. Example of differential gene expression between normal breast tissue (NB) and breast tumor samples. Each cDNA array on nylon filter was hybridized with a complex probe made from 5 µg of total RNA. The top image corresponds to the whole membrane. For the two bottom images, only the right portion of the membranes is shown. Numbers below the spots indicate housekeeping genes (1, *GAPDH*; 2, *actin*), negative control clones (3–5) and examples of genes differentially expressed between NB and breast tumor (6, *stromelysin3*; 7, *ERBB2*; 8, *MYBL2*; 9, *FOS*; 10, *TGFβR3*; 11, *desmin*) and between ER^- breast tumor and ER^+ breast tumor (12, *GATA3*).

The crude results of all hybridizations were processed to be presented either as absolute or relative values in schematic figures. The normalization procedure allowed to display absolute values expressed in percentage abundance of mRNA in the probe (Fig. 2a). Each level of the blue color ladder represents a 3-fold interval of absolute abundance of mRNA. Each column corresponds to a tissue sample and each row to a gene. For graphic purpose, genes were ordered from top to bottom according to increasing median expression levels. Tumor samples were not ordered. The values in each sample displayed a wide range of intensities (3 decades in log scale) corresponding to expression levels ranging from ~0.002 to 5% of mRNA abundance. Many genes (for example *stromelysin 3*, *IGF2* and *GATA3*, arrows) displayed highly variable expression levels across all tumor samples, scattered over the whole dynamic range of values. A representation of relative values is shown in Figure 2b. Absolute values were log-transformed, omitting 18 clones whose median intensity was equal to zero across all tissues. Data for each of the 162 remaining clones were then median-centered, as well as data for each sample, so that the relative variation was shown, rather than the absolute intensity. A color scale was used to display data: red was used for expression level higher than the median and green for expression level lower than the median. The magnitude of the deviation from the median was represented by the color intensity. A hierarchical clustering program was then applied to group the 35 samples according to their overall gene expression profiles and to group the 162 clones on the basis of similarity of their expression levels in all tissues. This resulted in a picture highlighting groups of correlated tissues and groups of correlated genes as depicted by dendrograms.

Breast tumor classification

As shown in Figure 2b, the clustering algorithm identified two groups of samples, designated A [n = 15, including normal breast (NB)] and B (n = 20). These groups were similar with respect to patient age, menopausal status at diagnosis, Scarff–Bloom–Richardson (SBR) grading and tumor pathological size. However, 72% of tumors in group A were node positive and 75% in group B were node negative. Moreover, 80% of the tumors in group B were estrogen receptor (ER) positive and 50% in group A were ER negative. With a median follow-up of 44 months after diagnosis, overall survival was different between A and B groups: five women died in A (median follow-up 58 months) and one in B (median follow-up 40 months). But the frequency of metastatic relapse was relatively similar in the two groups, with five women who relapsed in A and six in B. Because the time between the diagnosis of metastasis and last follow-up is too short in B, a longer follow-up is needed to determine whether these two different groups, defined with expression profiles, really have a different outcome with respect to overall survival.

In group A of 15 samples, three samples (normal breast and two tumors) were different from each other and from the other 12 samples; the latter constituted two subgroups of tumors, A1 (n = 6) and A2 (n = 6), which could be further separated by clustering (Fig. 2c). The 12 tumors had a uniformly high risk of metastatic relapse according to conventional prognostic features (Table 1). Most of them had received comparable adjuvant anthracyclin-based chemotherapy after surgery, with more women treated in the A1 subgroup. Interestingly, these two subgroups, which could not be distinguished with the commonly used histoclinical features, had a very different clinical outcome. There were four metastatic relapses and four deaths in A1 (median follow-up 44 months). In contrast, and despite a longer median follow-up (90 months), no metastasis or death occurred in A2. This resulted in a significant better metastasis-free survival ($P < 0.01$) and overall survival ($P < 0.005$) for group A2 than for group A1 tumors. No such subgrouping could be done in B.

We searched for genes responsible for group A substructure; these are potentially relevant to the prognosis and the sensitivity to chemotherapy in these tumors. Thirty-two of 176 genes were identified by comparing their median expression level in A1 versus A2. We then reclustered the 12 tumors using the expression profiles of these genes (Fig. 3). The same subgroups A1 and A2 were evident and separated by two groups of genes: as

expected from the literature, high expression of *ERBB2*, *MYC* and *EGFR* was associated with bad prognosis subgroup A1 (6–8) and that of *E-cadherin* and the proto-oncogene *MYB* with good prognosis subgroup A2 (9,10). For most of the other genes, literature data are scarce and these results may stimulate new investigations. Differentiation state is a good prognostic factor in breast cancer and, accordingly, genes associated with cell differentiation, such as *GATA3* (11) and *CRABP2* (12), had a high level of expression in the better outcome group. The high expression of Ephrin-A1 mRNA in the bad prognosis subgroup suggests a role of this growth factor in breast cancer and can be paralleled with its upregulation during melanoma progression (13).

Differential gene expression between normal breast and breast tumors

To identify genes differentially expressed between breast tumors (T) and NB, the NB value for each gene was compared with its expression level in each tumor. When the expression level of a gene in NB was undetectable, only qualitative information could be deduced and the mRNA was considered as differentially expressed if the signal intensity in the tumor was superior to the reproducibility threshold (0.002% of mRNA abundance). In the other cases, differential expression was defined by an at least 2-fold expression difference. We also measured for each gene the number of tumors where it was over- or underexpressed. Table 2 shows a list of the top 20 over- and underexpressed genes. For these genes, we reported the T:NB ratio, where T represented their median expression value in the 34 tumors. This ratio ranged from 2.70 (*ABCC5*) to 17.76 (*GATA3*) for the overexpressed genes and from 0.00 (*desmin*) to 0.29 (*APC*) for the underexpressed genes.

High expression of *mucin 1*, *NM23, ERBB2*, *FGFR1* and *FGFR2*, *MYC*, *stromelysin3, cathepsin D* and downregulation of *FOS*, *APC*, *RBL2*, *FAS*, *BCL2* were found, reflecting what is known about their biology in cancer. *GATA3*, which codes for a member of the GATA family of zinc finger transcription factors, and *CRABP2*, encoding one of the two cellular retinoic acid-binding proteins, showed high expression of mRNA, extending previous results on cDNA arrays (4).

Differential gene expression among various breast tumors and correlation with histoclinical prognostic parameters

To search for potential prognostic markers in breast cancer, we looked for genes with expression levels correlated with conventional histoclinical prognostic parameters: age of patients, axillary node status, tumor size, histological grade and ER status. No significant correlation was found with age, tumor size and histological grade. However, the expression profiles of some genes correlated with ER status and axillary node involvement.

To identify genes potentially relevant to the hormone-responsive phenotype, we compared the gene expression profiles in ER-positive breast cancers ($n = 23$) versus ER-negative breast cancers ($n = 11$). Sixteen clones displayed a median intensity of zero in both groups. Twenty-five presented a fold change superior to 2. Table 3 displays the top 10 over- and underexpressed genes. Among them, the most differentially expressed was *GATA3* with a median intensity ratio ER^+:ER^- of 28.6 and a value for the first quartile of ER-positive tumors superior (5-fold) to the value of the third quartile of the ER-negative tumors (Fig. 4a). The high expression of *GATA3* in ER-positive tumors was statistically significant using a Mann–Witney test ($P < 0.001$). All ER-positive tumors and only 18% of ER-negative tumors displayed a *GATA3* expression level greatly superior (fold change > 3) to the normal breast value. We further analyzed *GATA3* expression by northern blot hybridization (Fig. 4b) in a panel of 79 breast cancers (21 ER-negative tumors and 58 ER-positive tumors), including 22 of the tumors analyzed with cDNA arrays; it confirmed the array results for those 22 tumors as well as the strong correlation between ER status and *GATA3* RNA expression (Mann–Witney test, $P < 0.0001$).

To search for genes whose expression profile was correlated with axillary lymph node status, a strong prognostic factor in breast cancer, we compared the group of node-negative tumors ($n = 19$) with the group of tumors with massive axillary extension (10 or more positive nodes). Furthermore, because survival decreases with the increase of the number of tumor-involved lymph nodes and because our expression measurements were quantitative, we looked for a correlation between the expression levels of these genes and the number of tumor-involved nodes (quantitative variables). Table 3 shows a list of the top 10 over- and underexpressed genes between these two groups. Most of these genes have not been previously reported as associated with node status, but some of these results are in agreement with literature data. The gene encoding the tyrosine kinase receptor ERBB2 was the most significantly overexpressed gene in node-positive tumors and displayed the highest correlation coefficient ($r = 0.68$; $P < 0.0001$).

Gene clusters

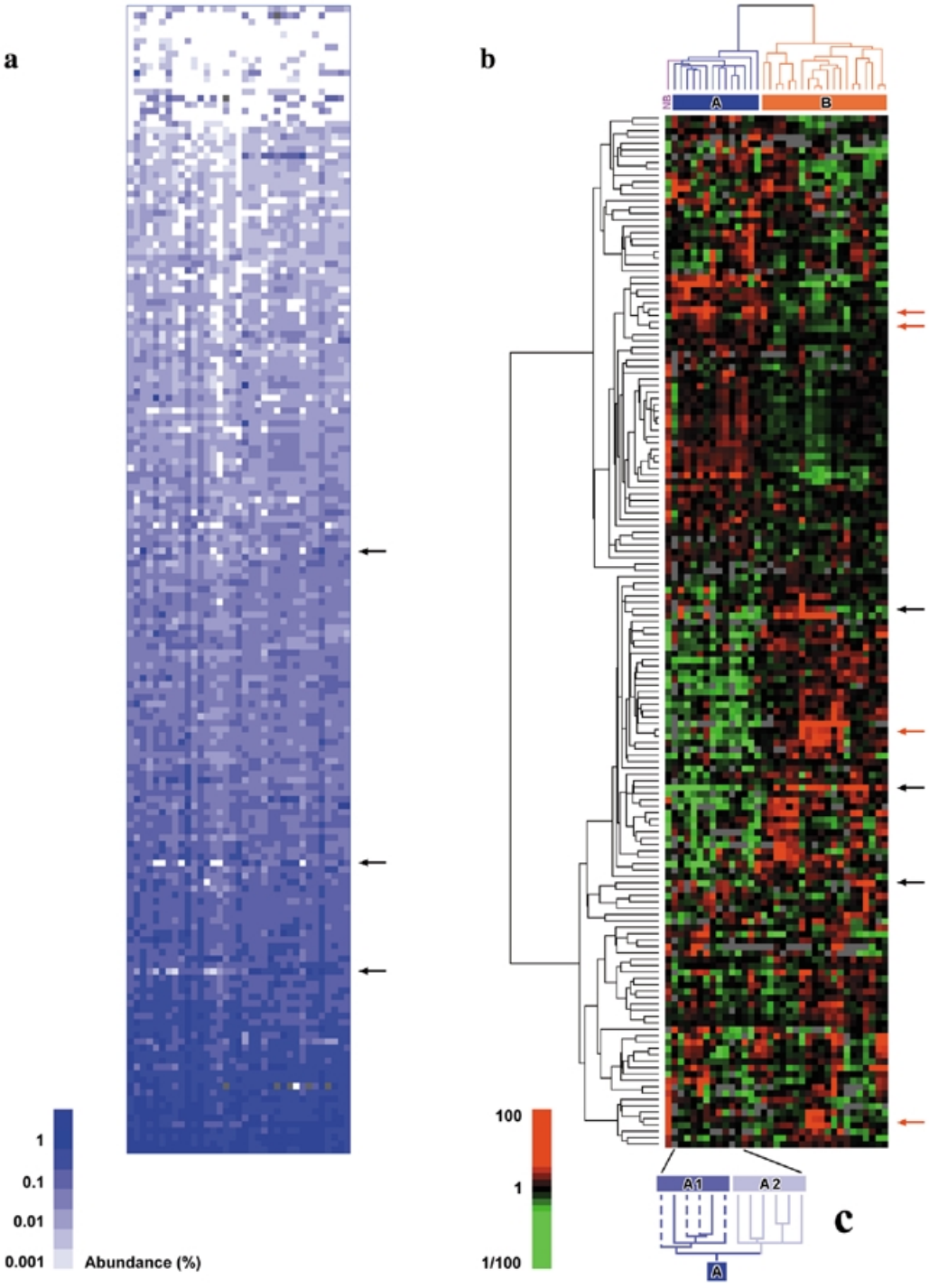
Gene clustering (Fig. 2b) showed groups of genes with correlated expression across samples. When different clones represented the same gene, they were clustered next to each other (red arrows). Correlation coefficients between gene pairs in the 34 tumors were often high (1% of the 13 041 gene pairs showed a correlation coefficient superior to 0.95; data not shown). An example of highly correlated gene expression is that of *BCL2* and *RBL2*. Such correlated expression, although it has not been described in the literature, probably reflects a common mechanism of regulation for these two genes. Furthermore, these genes also exhibited significant correlated expression with other genes such as *PPP2CA, AKT2, PRKCSH* and *TNFRSF6/FAS*. In particular, a striking correlated expression between *BCL2* and *FAS* could be observed ($r = 0.91$; data not shown). The exact meaning of this correlation is unknown, although it may reflect the necessary balance between apoptosis and anti-apoptosis for cell survival.

DISCUSSION

Although in human cancer the proportion of changes that is reflected at the RNA level is not known, monitoring gene expression patterns appears to be a very promising way of increasing our knowledge of the disease. Several different types of cancer have been investigated using cDNA arrays: cervical (14), hepatocellular (15), ovarian (16), colon (17) and renal carcinomas (18), glioblastomas (19), melanomas (20,21), rhabdomyosarcomas (22), acute leukemias (23) and lymphomas (24). In breast cancer, pioneering studies have yielded the first

expression patterns (4,25–31). They have in particular addressed the important issue of molecular differences in hormone responsive and non-responsive breast tumors. Thus, Yang *et al*. (28) and Hoch *et al.* (25) compared expression profiles of breast carcinoma cell lines known to represent these two categories and identifed a few genes with differential expression. One of these genes was *GATA3*. In these studies, cell lines were mostly used and tumor samples were rarely tested and generally in small numbers. The first study analyzing the expression profiles of a large series of breast cancers was published recently, but no correlation with clinical outcome was mentioned by the authors (32). To our knowledge, our study presents one of the largest reported series of breast cancer samples analyzed with cDNA array technology and the first one with an analysis of correlation with survival.

Several interesting points can be made at this stage. First, the differences in expression patterns among the tumors provided molecular transcriptional evidence of the histoclinical heterogeneity of breast cancer: this diversity was multifactorial, linked to many different genes, highlighting the interest of high



throughput analysis in this context. It was possible, with a hierarchical clustering program integrating the expression profiles, to separate normal breast tissue from most tumors and, moreover, to identify two different groups of tumors. Most importantly, we identified by clustering two different subgroups of tumors with a very distinct clinical outcome that could not be predicted with classical prognostic factors. Indeed, all these tumors had a theoretically bad prognosis as evaluated by current histoclinical tools. All these patients would be at the present time treated with adjuvant chemotherapy, but without the capacity for the physicians to identify patients who will benefit from this treatment and those who will not benefit. Gene expression profiles were apparently able to make this discrimination. Such predictive tools would have important therapeutic implications; patients with features of poor prognosis would be candidates for other treatment than standard chemotherapy, avoiding loss of time and toxicities related to first-line chemotherapy. Our results, although obtained with a small series of patients, suggest that the histoclinical category of poor prognosis breast cancer, currently treated with adjuvant anthracyclin-based chemotherapy, groups together at least two molecularly distinct subgroups of tumors with different outcomes which would require distinct chemotherapy regimens. Expression profiles could thus provide a new and more accurate way of classifying breast tumors of poor prognosis and of managing patients. Similar results were recently reported in one study concerning diffuse large B cell lymphomas (24).

Similarly, despite molecular heterogeneity, significant correlations between the expression level of genes (*GATA3*, *ERBB2*) and histological tumor parameters were identified. The ER-positivity in breast cancer has been correlated with tumor differentiation, low proliferating rate, favorable prognosis and response to hormonal therapy. The relation between hormone sensitivity of breast cancer and ER status is not perfect and it is possible that some genes related to ER expression are more important than ER to characterize the hormone-sensitive phenotype. These genes could serve as predictive factors to guide the therapy and could stimulate research leading to biochemical intervention on new therapeutic targets. *GATA3* mRNA expression was highly correlated with ER status. GATA3, which is not estrogen regulated (25), is a transcription factor which could regulate the expression of genes involved in the ER-positive phenotype. Among the other genes that we found associated with ER status, some, such as *MYB* (10), *stromelysin 3* (33) and *CRABP2* (34), have been previously reported expressed at high levels in ER-positive breast tumors. The higher levels of *TP53* mRNA in our ER-positive tumors were surprising, although in agreement with a recent study (27). Most studies concerning TP53 expression analyzed the protein level rather than the mRNA level and TP53 protein levels are classically negatively correlated with the ER status (35). The high expression of *CRABP2* could be related to the better differentiated status of the ER-positive tumors. The low expression of the three immunity-related genes *IL2RB*, *IL2RG* and *CD3G* may be related to the low lymphoid infiltration in these well differentiated tumors. *ERBB2* high expression in breast cancer has been associated with a poor prognosis and some resistance to hormonal therapy and chemotherapy (36). It is involved in the regulation of cellular differentiation, adhesion and motility. The motility-enhancing activity of ERBB2 (37) could be responsible for the increased metastatic potential and the unfavorable prognosis of the breast tumors that overexpress *ERBB2*. The low expression of *E-cadherin* and *thrombospondin 1* in node-positive tumors are consistent with their putative role in different steps of metastatic spread: E-cadherin is an epithelial cell adhesion molecule whose disturbance is a prerequisite for the release of invasive cells in carcinomas (38) and thrombospondin 1 inhibits angionesis (39). Similarly, the high expression of the molecule surface antigen Mucin 1 in node-positive tumors (40) can reduce cell–cell interactions facilitating cell detachment and metastasis. *CD44*, encoding a transmembrane glycoprotein involved in cell adhesion and lymph node homing (41) was expressed at high levels in node-positive tumors as well as *GSTP1* (glutathione *S*-transferase Pi), recently reported associated with increased tumor size (27).

Second, there were a number of genes with highly correlated expression patterns. Gene correlations have already been reported with larger series of genes, essentially under dynamic experimental conditions (42) and recently in steady states (17). Here, correlations were based on expression profiles of a relatively small but selected series of genes and in steady states represented by different breast tumors. Gene correlations are potentially useful tools for cancer research in two ways: (i) they can provide information about the general regulation circuitry of a cancerous cell, allowing the identification of regulatory elements controlling expression networks; and (ii) they offer the possibility of reducing the complexity of the system analyzed by replacing, for example, the intensities of a large number of genes present in a gene cluster by their respective mean intensities.

Finally, our results, like other recent studies, highlight the great potential of cDNA array in cancer research. The gene

Figure 2. Representation of expression levels of 176 genes in normal breast tissue and 34 samples of breast carcinoma. Each column corresponds to a single tissue, and each row to a single gene. (**a**) The results are expressed as percentage abundance of individual mRNA within the sample, and are represented using a blue color scale. The color scale (log scale with a 3-fold interval) indicated at the bottom left ranges from light blue (expression level ≥ 0.001%) to dark blue (expression level ≥ 3%). White squares indicate clones with undetectable expression levels and gray squares indicate missing data. The tissue samples are arbitrarily ordered and the clones are ordered from top to bottom according to increasing median expression levels. Horizontal black arrows on the right of the figure mark three clones with highly variable expression levels between the tumors (*stromelysin3*, *IGF2* and *GATA3* from top to bottom). (**b**) The results are shown as relative expression levels (relative to the median value of each row and each column) and are represented with a color scale indicated at the bottom left ranging from 1/100 to 100-fold changes (gray squares indicate missing data). Eighteen clones with median expression level equal to zero in the 34 tumors are omitted. The clustering program arranges samples (n = 35) along the horizontal axis so that those with the most similar expression profiles are placed adjacent to each other. Similarly, clones (n = 162) are near each other along the vertical axis if they show a strong expression profile correlation across all tissues. The length of the branches of the dendrograms capturing respectively the samples (top) and the clones (left) reflects the similarity of the related elements. Two groups of tumors are separated and color coded: group A (blue) and group B (orange). Horizontal black and red arrows on the right of the figure, respectively, mark three genes with highly variable expression levels between the tumors (*IGF2*, *GATA3* and *stromelysin3* from top to bottom) and four pairs of different clones representing four genes. (**c**) Zoom representation of group A from (b), excluding the two outlier tumors at the right. The clustering separates two subgroups of tumors, A1 and A2. The dotted branches correspond to tumors associated with metastatic relapse and death. Follow-up was longer in A2 (median 90 months) than in A1 (median 44 months).

Table 1. Patient characteristics in subgroups A1 and A2

Subgroup	A1						A2					
Tumor position in the cluster	1	2	3	4	5	6	7	8	9	10	11	12
Age (years)	46	58	60	63	51	58	46	47	50	47	46	66
Nodal status	1	0	0	16	13	37	10	4	1	2	0	0
Histological size (mm)	60	20	26	35	20	30	27	25	30	25	20	22
SBR grade	II	III	II	III	II	III	II	II	II	II	II	III
ER status	–	–	–	–	–	–	+	-	+	+	+	+
Adjuvant chemotherapy	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No
Metastasis	Yes	No	Yes	Yes	No	Yes	No	No	No	No	No	No
Follow-up (months)	58	106	35	47	41	31	85	98	95	49	19	141
Patient status[a]	D	A	D	D	A	D	A	A	A	A	A	A

The 12 tumors are numbered from 1 to 12 according to their position from left to right in the clustering graphic displayed in Figure 3.
Adjuvant chemotherapy was anthracyclin based.
[a]A, alive; D, death from cancer progression.

expression profiles confirmed the heterogeneity of breast cancer and most importantly allowed us to identify, among a series of poor prognosis breast tumors, two subtypes of the disease not yet recognized with usual histoclinical parameters but with a different clinical outcome after adjuvant chemotherapy. Furthermore, we detected genes of which expression was correlated with classical prognostic factors. Obviously, given the high molecular heterogeneity of breast cancer, more samples and more genes will be necessary to evaluate the actual capacity of the method to classify all breast tumors into reliable prognostic subtypes and to identify the most discriminant genes. That will also require selection of more homogeneous series of patients in regard to their tumor characteristics, treatment and long follow-up, in order to answer precise questions. But as yet, these results are very encouraging.

MATERIALS AND METHODS

Tumor samples and RNA extraction

To avoid any bias of selection as to the type and size of the tumors, the RNAs to be tested were prepared from unselected samples. Samples of primary invasive breast carcinomas were collected from 34 patients undergoing surgery at the Institut Paoli-Calmettes. After surgical resection, the tumors were macrodissected: a section was taken for the pathologist's diagnosis and an adjacent piece was quickly frozen in liquid nitrogen for molecular analyses. The median age of patients at the time of diagnosis was 55 years (range 39–83) and most of them were post-menopausal. Tumors were classified according to the WHO histological typing of breast tumors: 29 ductal carcinomas, 2 lobular carcinomas, 1 mixed ductal and lobular carcinoma and 2 medullary carcinomas. They had various sizes, inferior or equal to 20 mm (n = 13), between 20 and 50 mm (n = 18) or superior to 50 mm (n = 3), axillary lymph node status (negative: 19 tumors; positive: 15 tumors), SBR grading (I: 3 tumors; II: 20 tumors; III: 10 tumors; not evaluable: 1 tumor) and ER status evaluated by immunohistochemical assay (23 ER positive, 11 ER negative). The ER-positivity cutoff value was 10%. Adjuvant treatment with radiotherapy and when necessary multi-agent anthracyclin-based chemotherapy (n = 16) was given to patients according to local practice.

Total RNA was extracted from tumor samples by standard methods, as previously described (43). Total RNA from NB tissue was obtained from Clontech (Palo Alto, CA). RNA was isolated from eight tissue specimens from Caucasian females, age range 23–47 years. RNA integrity was controlled by denaturing formaldehyde agarose gel electrophoresis and northern blots using a 28S-specific oligonucleotide.

cDNA arrays preparation

Gene expression was analyzed by hybridization of arrays with radioactive probes. The arrays contained PCR products of five control clones and 180 IMAGE human cDNA clones selected with practical criteria (3′ sequence of mRNA, same cloning vector, host bacteria and insert size). This represented 176 genes (four genes were represented by two different clones): 121 with proven or putative implication in cancer and 55 implicated in immune reactions (the list is available at http:/ tagc.univmrs.fr/pub/Cancer/). Their identity was verified by 5′ tag-sequencing of plasmid DNA and comparison with sequences in the expressed sequence tag (dbEST) and nucleotide (GenBank) databases at the NCBI. Identity was confirmed for all but 14 clones without significant gene similarity, which were referenced by their GenBank accession number. The control clones were: *Arabidopsis thaliana* cytochrome *c554* gene (used for hybridization signal normalization), three poly(A) sequences of different sizes and the vector pT7T3D (negative controls).

PCR amplification, purification and robotical spotting of PCR products onto Hybond-N$^+$ membranes (Amersham Pharmacia Biotech, Little Chalfont, UK) were done according to described protocols (4). All PCR products were spotted in

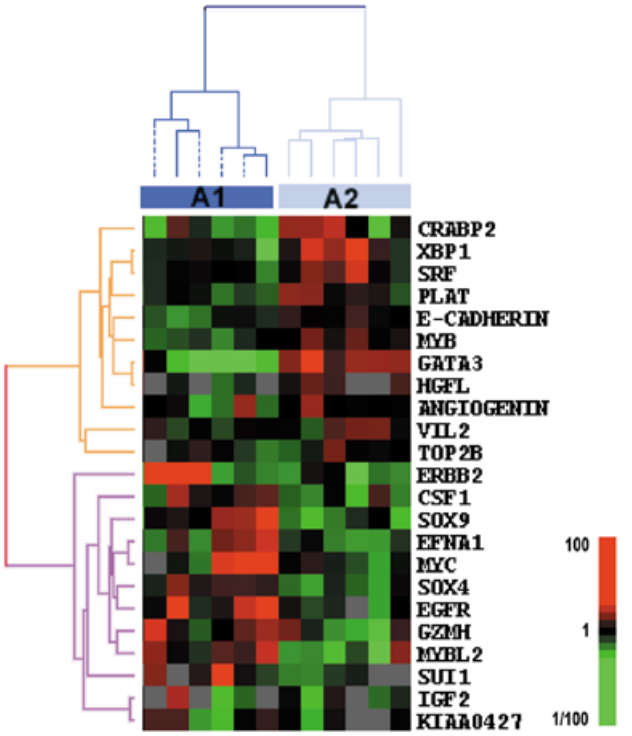


Figure 3. Prognostic classification of breast cancer by gene expression profiling. The 12 samples of group A (see Fig. 2b and c) were reclustered using the top 32 differentially expressed genes between A1 and A2 subgroups. Data were displayed as in Figure 2b and shown with the same color key. The hierarchical clustering was applied to expression data from the 23 clones (of 32) of which expression levels presented at least a 2-fold change in two or more samples (out of 12). Two subgroups of tumors, A1 and A2, are shown as well as two groups of differentially expressed clones. The dotted branches of tumor cluster A1 correspond to samples associated with metastatic relapse and death.

duplicate. For normalization purpose, the *c554* gene was spotted 96-fold scattered over the whole membrane.

cDNA array hybridizations

Hybridizations were done successively with a vector oligonucleotide (to precisely determine the amount of target DNA accessible to hybridization in each spot), then after stripping of vector probe, with complex probes made from the RNAs (4). Each complex probe was hybridized to a distinct filter. Probes were prepared from total RNA with an excess of oligo(dT25) to saturate the poly(A) tails of the messengers and to insure that the reverse transcribed product did not contain long poly(T) sequences. A precise amount of *c554* mRNA was added to the total RNA before labelling to allow normalization of the data.

Five micrograms of total RNA (~100 ng of mRNA) from tissue samples were used for each labeling. Probe preparation and hybridization of the membranes were done according to previously described procedures (http:/tagc.univ-mrs.fr/pub/Cancer/). Hybridization was done in excess of target (~15 ng of DNA in each spot) and binding of cDNAs to the targets was linear and proportional to the quantity of cDNA in the probe.

Detection and quantification of cDNA array hybridization signals

Quantitative data were obtained using an imaging plate device. Hybridization signal detection with a Fuji Bas 1500 machine (Raytest, Paris, France) and quantification with the HDG Analyser software (Genomic Solutions, Ann Arbor, MI) were done as previously described (http:/tagc.univ-mrs.fr/pub/Cancer/). Quantification was calculated by integrating all spot pixel intensities and substracting a spot background value determined in the neighboring area. Spots were located with a LaPlacian transformation. Spot background level was the median intensity of all the pixels present in a small window centered on the spot and which were not part of any spot (44). Quantified data were normalized in three steps and expressed as absolute gene expression levels (i.e. percentage abundance of individual mRNA with respect to mRNA within the sample) as described (4).

Array data analysis

Before analysis of the results, the reproducibility of the experiments was verified by comparing duplicate spots, or one hybridization with the same probe on two independent arrays, or two independent hybridizations with probes prepared from the same RNA. In every case, the results showed good reproducibility with respective correlation coefficients of 0.95, 0.98 and 0.98 (data not shown). Moreover, genes represented by two different clones on the array, such as *CDK4* or *ETV5*, displayed similar expression profiles for the two clones in all samples. This reproducibility was sufficient enough to consider a 2-fold expression difference as significantly differential.

For graphical representation, data were displayed as absolute expression levels (Fig. 2a). For better visualization of clustering, results were log-transformed and displayed as relative values median-centered in each row and in each column (Fig. 2b). Hierarchical clustering was applied to the tissue samples and the genes using the Cluster program developed by Eisen *et al.* (45) (average linkage clustering using Pearson correlation as similarity metric). Results in Figures 2 and 3 were displayed with the TreeView program (45).

Subsequent analysis was done using Excel software (Microsoft) and statistical analyses with SPSS software. Metastasis-free survival and overall survival were measured from diagnosis until the first metastatic relapse or death, respectively. They were estimated using the Kaplan–Meier method and compared between groups with the log-rank test. Correlations of gene pairs based on expression profiles were measured with the correlation coefficient *r*. The search for genes with expression levels correlated with tumor parameters was done in several successive steps. First, genes were detected by comparing their median expression level in the two subgroups of tumors discordant according to the parameter of interest. We used the median values rather than the mean values because of the high variability of the expression levels for many genes, resulting in a standard deviation of expression level similar or superior to the mean value and making comparisons with means impossible. Second, these detected genes were inspected visually on graphics and, finally, an appropriate statistical analysis was applied to those that were convincing to validate the correlation. Comparison of *GATA3* expression between ER-positive tumors and ER-negative tumors was validated using a Mann–

Table 2. List of the genes that show the most frequent differential expression between normal breast tissue and 34 breast carcinomas as measured by cDNA array analysis

Clone ID	Gene/protein identity	Gene symbol	Chromosome location	*n*	T:NB
Overexpressed genes					
154343	*Granzyme H*	GZMH	14q11.2	32	9.51
235947	*Stromelysin 3*	STMY3	22q11.2	31	15.92
207378	*MYB related protein B*	MYBL2	20q13.1	31	[a]
153275	*Cellular retinoic acid binding protein 2*	CRABP2	1q21.3	29	7.16
129757	*GATA-binding protein 3*	GATA3	10p15	28	17.76
120649	*T-lymphocyte surface CD2 antigen*	CD2	1p13.1	28	7.54
109677	*CREB binding protein*	CREBBP	16p13.3	28	5.08
172152	*EGFR-binding protein GRB2*	GRB2	17q24–q25	28	5.00
66969	*Transcription factor RELB*	RELB	19	28	3.61
182007	*ETS-related transcription factor ELF1*	ELF1	13q13	27	3.58
153446	*LIM domain protein RIL*	RIL	5q31.1	26	4.03
203394	*ETS variant gene 5* (*ETS-related molecule*)	ETV5	3q28	25	3.67
160963	*Thrombospondin 1*	THBS1	15q15	25	3.39
188393	*POU domain, class 2, transcription factor 2*	POU2F2	19	24	4.02
187822	*Integrin, beta 2*	ITGB2	21q22.3	24	3.01
243907	*Nuclear factor of activating T cell subunit p45*	NF45	1	24	2.84
158347	EST H27202	EST		23	2.91
230933	EST AW184517	EST		22	2.85
212366	*ATP-binding cassette, sub-family C* (*CFTR/MRP*), *5*	ABCC5	3q27	22	2.70
149401	*Cathepsin D*	CTSD	11p15.5	21	2.97
Underexpressed genes					
153854	*Desmin*	DES	2q35	34	0.00
208717	*P55-C-FOS proto-oncogene protein*	FOS	14q24.3	33	0.05
159093	*Transcription factor AP4*	TFAP4	16p13	33	0.11
124340	*Tenascin XA*	TNXA	6p21.3	33	0.14
133738	*Prolactin*	PRL	6p22.2–p21.3	32	0.00
133891	*Chorionic somatomammotropin hormone 1*	CSH1	17q22–q24	32	0.00
151501	*Tyrosine kinase receptor TEK*	TEK	9p21	32	0.00
183030	*Activating transcription factor 3*	ATF3	1	32	0.07
120916	*Phosphodiesterase I*	PDNP2	8q24.1	32	0.14
155716	EST R72075	EST		31	0.00
208118	*Transforming growth factor beta receptor type III*	TGFBR3	1p33–p32	31	0.14
187547	*Diphtheria toxin receptor*	DTR	5q23	31	0.17
108490	*HIV-1 Rev binding protein*	HRB	2q36	31	0.20
147002	*B-cell CLL/lymphoma 2*	BCL2	18q21.3	31	0.26
182610	*Microsomal glutathione S transferase 1*	MGST1	12p12.3–p12.1	31	0.28
152802	*Phospholipase A2 membrane associated, group IIA*	PLA2G2A	1p35	30	0.03
183087	*Interleukin 3 receptor alpha chain*	IL3RA	Xp22.3;Yp13.3	30	0.24
108571	*Retinoblastoma-like 2 (p130)*	RBL2	16q12.2	29	0.28
125294	*Adenomatous polyposis coli protein*	APC	5q21–q22	29	0.29
151767	*FASL receptor*	TNFRSF6	10q24.1	28	0.27

n, number of tumor samples in which the gene is dysregulated (fold change ≥ 2) compared with normal breast tissue; T:NB, ratio of median expression level in 34 breast tumors to expression level in normal breast.
[a]*MYBL2* transcript displayed a median expression level of 0.025% in breast tumors and was undetectable in NB.

Table 3. List of genes differentially expressed between ER-positive and ER-negative breast tumors and between axillary lymph node-negative tumors and tumors with 10 or more involved lymph nodes

Clone ID	Gene/protein identity	Gene symbol	ER^+:ER^-
129757	*GATA-binding protein 3*	GATA3	28.6
356763	*Granzyme A*	GZMA	5.7
248613	*MYB proto-oncogene*	MYB	3.4
211999	*KIAA1075 protein*	KIAA1075	3.3
235947	*Stromelysin 3*	STMY3	3.1
229839	*Macrophage stimulating 1*	MST1	2.8
153275	*Cellular retinoic acid binding protein 2*	CRABP2	2.7
301950	*X-box binding protein 1*	XBP1	2.7
205314	*Tumor protein p53*	TP53	2.5
126233	*Insulin-like growth factor 2*	IGF2	2.4
66322	*CD3G antigen, gamma*	CD3G	0.0
195022	*Interleukin 2 receptor gamma chain*	IL2RG	0.0
111461	*SOX4 protein*	SOX4	0.4
151475	*Epidermal growth factor receptor*	EGFR	0.5
195022	*Interleukin 2 receptor beta chain*	IL2RB	0.5
130788	*Topoisomerase (DNA) II beta* (180 kDa)	TOP2B	0.6
323948	*SOX9 protein*	SOX9	0.6
183641	*S100 calcium-binding protein beta*	S100B	0.6
246620	EST N53133	EST	0.6
231424	*Glutathione S transferase Pi*	GSTP1	0.6
Clone ID	Gene/protein identity	Gene symbol	N^-:$10N^+$
129757	*GATA-binding protein 3*	GATA3	11.0
160963	*Thrombospondin 1*	THBS1	6.6
151475	*Epidermal growth factor receptor*	EGFR	5.4
120916	*Phosphodiesterase I*	PDNP2	4.9
183030	*Activating transcription factor 3*	ATF3	4.6
211999	*KIAA1075 protein*	KIAA1075	4.5
110480	*Nuclear factor 1 A-type*	NF1A	4.5
182264	*P-selectin*	SELP	4.4
356763	*Granzyme A*	GZMA	4.3
214008	*E-cadherin*	CDH1	4.0
147016	*ERBB2 receptor protein-tyrosine kinase*	ERBB2	0.2
179197	*Protein phosphatase PP2A*, 55 kDa subunit	PP2A BR gamma	0.2
231424	*Glutathione S transferase Pi*	GSTP1	0.4
111461	*SOX4 protein*	SOX4	0.4
195022	*Interleukin 2 receptor beta chain*	IL2RB	0.4
220451	*Zinc finger protein 144*	ZNF144	0.5
125413	*Mucin 1*	MUC1	0.6
290007	*CD44 antigen, epithelial form*	CD44	0.6
108571	*Retinoblastoma-like 2 (p130)*	RBL2	0.7
130788	*Topoisomerase (DNA) II beta* (180 kDa)	TOP2B	0.7

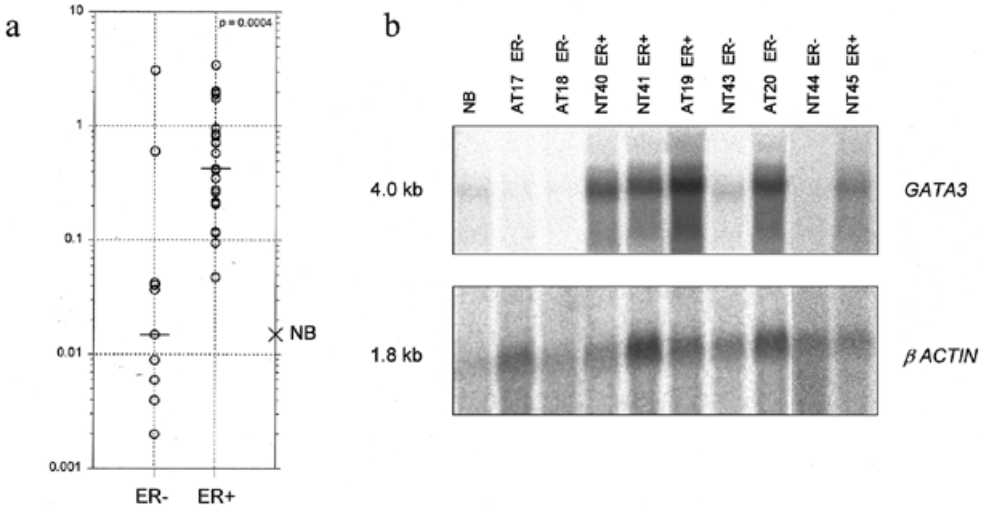


Figure 4. Correlation of *GATA3* expression with ER phenotype. (**a**) The expression levels of *GATA3* in 34 breast cancer samples (*y*-axis) monitored by cDNA array analysis are reported in percentage of abundance of individual mRNA with respect to mRNA within the sample (log scale). *GATA3* is significantly overexpressed in the ER-positive tumors ($n = 23$) versus the ER-negative tumors ($n = 11$) using the Mann–Witney test ($P = 0.0004$). The expression level of *GATA3* in normal breast tissue is reported on the right (NB). (**b**) Northern blot analysis of *GATA3* in normal breast sample (NB) and nine breast cancer samples (AT, tumor analyzed with cDNA array and northern blot; NT, tumor analyzed with northern blot). Blots were probed successively with cDNA from *GATA3* (top) and β–*actin* (bottom). ER status is indicated for each tumor sample.

Witney test. Correlation coefficients were used to compare the gene expression levels with the number of axillary nodes involved.

Northern blot analysis

Seventy-nine breast tumors, including 22 of the 34 tested on the arrays, were analyzed for *GATA3* expression by northern blot hybridization. *RNA* extraction from tumor samples and northern blots was performed as previously described (43). The *GATA3* probe was prepared from the IMAGE cDNA clone 129757, which corresponds to the 3′ region (from +843 to +1689) of the *GATA3* cDNA sequence (GenBank accession no. X55122). The insert (846 bp) was obtained by digestion of the clone with *Eco*RI and *Pac*I enzymes. Northern blots were stripped and re-hybridized using a β-*actin* probe (46).

ACKNOWLEDGEMENTS

We are grateful to Drs D. Maraninchi and C. Mawas for encouragement. This work was supported by INSERM, CNRS, Institut Paoli-Calmettes, Association Française contre la Myopathie, Association pour la Recherche sur le Cancer, Comité de la Corse du Sud de la Ligue Nationale Contre le Cancer and Fédération Nationale des Centres de Lutte contre le Cancer. F.B. was supported by a fellowship from Association pour la Recherche contre le Cancer.

REFERENCES

1. DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A. and Trent, J.M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.*, **14**, 457–460.
2. Jordan, B.R. (1998) Large-scale expression measurement by hybridization methods: from high-density membranes to 'DNA chips'. *J. Biochem. (Tokyo)*, **124**, 251–258.
3. Nguyen, C., Rocha, D., Granjeaud, S., Baldit, M., Bernard, K., Naquet, P. and Jordan, B.R. (1995) Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics*, **29**, 207–216.
4. Bertucci, F., Van Hulst, S., Bernard, K., Loriod, B., Granjeaud, S., Tagett, R., Starkey, M., Nguyen, C., Jordan, B. and Birnbaum, D. (1999) Expression scanning of an array of growth control genes in human tumor cell lines. *Oncogene*, **18**, 3905–3912.
5. Bertucci, F., Bernard, K., Loriod, B., Chang, Y.C., Granjeaud, S., Birnbaum, D., Nguyen, C., Peck, K. and Jordan, B.R. (1999) Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. *Hum. Mol. Genet.*, **8**, 1715–1722.
6. Ross, J.S. and Fletcher, J.A. (1999) The HER-2/neu oncogene: prognostic factor, predictive factor and target for therapy. *Semin. Cancer Biol.*, **9**, 125–138.
7. Scorilas, A., Trangas, T., Yotis, J., Pateras, C. and Talieri, M. (1999) Determination of c-myc amplification and overexpression in breast cancer patients: evaluation of its prognostic value against c-erbB-2, cathepsin-D and clinicopathological characteristics using univariate and multivariate analysis. *Br. J. Cancer*, **81**, 1385–1391.
8. Fox, S.B., Smith, K., Hollyer, J., Greenall, M., Hastrich, D. and Harris, A.L. (1994) The epidermal growth factor receptor as a prognostic marker: results of 370 patients and review of 3009 patients. *Breast Cancer Res. Treat.*, **29**, 41–49.
9. Heimann, R., Lan, F., McBride, R. and Hellman, S. (2000) Separating favorable from unfavorable prognostic markers in breast cancer: the role of E-cadherin. *Cancer Res.*, **60**, 298–304.
10. Guerin, M., Sheng, Z.M., Andrieu, N. and Riou, G. (1990) Strong association between c-myb and oestrogen-receptor expression in human breast cancer. *Oncogene*, **5**, 131–135.
11. Lim, K.C., Lakshmanan, G., Crawford, S.E., Gu, Y., Grosveld, F. and Douglas Engel, J. (2000) Gata3 loss leads to embryonic lethality due to noradrenaline deficiency of the sympathetic nervous system. *Nature Genet.*, **25**, 209–212.
12. Mills, K.J., Vollberg, T.M., Nervi, C., Grippo, J.F., Dawson, M.I. and Jetten, A.M. (1996) Regulation of retinoid-induced differentiation in embryonal carcinoma PCC4.aza1R cells: effects of retinoid-receptor selective ligands. *Cell Growth Differ.*, **7**, 327–337.

13. Easty, D.J., Hill, S.P., Hsu, M.Y., Fallowfield, M.E., Florenes, V.A., Herlyn, M. and Bennett, D.C. (1999) Upregulation of ephrin-A1 during melanoma progression. *Int. J. Cancer*, **84**, 494–501.
14. Shim, C., Zhang, W., Rhee, C.H. and Lee, J.H. (1998) Profiling of differentially expressed genes in human primary cervical cancer by complementary DNA expression array. *Clin. Cancer Res.*, **4**, 3045–3050.
15. Tsou, A.P., Wu, K.M., Tsen, T.Y., Chi, C.W., Chiu, J.H., Lui, W.Y., Hu, C.P., Chang, C., Chou, C.K. and Tsai, S.F. (1998) Parallel hybridization analysis of multiple protein kinase genes: identification of gene expression patterns characteristic of human hepatocellular carcinoma. *Genomics*, **50**, 331–340.
16. Schummer, M., Ng, W.V., Bumgarner, R.E., Nelson, P.S., Schummer, B., Bednarski, D.W., Hassell, L., Baldwin, R.L., Karlan, B.Y. and Hood, L. (1999) Comparative hybridization of an array of 21 500 ovarian cDNAs for the discovery of genes overexpressed in ovarian carcinomas. *Gene*, **238**, 375–385.
17. Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D. and Levine, A.J. (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl Acad. Sci. USA*, **96**, 6745–6750.
18. Moch, H., Schraml, P., Bubendorf, L., Mirlacher, M., Kononen, J., Gasser, T., Mihatsch, M.J., Kallioniemi, O.P. and Sauter, G. (1999) High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma. *Am. J. Pathol.*, **154**, 981–986.
19. Rhee, C.H., Hess, K., Jabbur, J., Ruiz, M., Yang, Y., Chen, S., Chenchik, A., Fuller, G.N. and Zhang, W. (1999) cDNA expression array reveals heterogeneous gene expression profiles in three glioblastoma cell lines. *Oncogene*, **18**, 2711–2717.
20. Huang, F., Adelman, J., Jiang, H., Goldstein, N.I. and Fisher, P.B. (1999) Identification and temporal expression pattern of genes modulated during irreversible growth arrest and terminal differentiation in human melanoma cells. *Oncogene*, **18**, 3546–3552.
21. Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A. *et al.* (2000) Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature*, **406**, 536–540.
22. Khan, J., Simon, R., Bittner, M., Chen, Y., Leighton, S.B., Pohida, T., Smith, P.D., Jiang, Y., Gooden, G.C., Trent, J.M. and Meltzer, P.S. (1998) Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res.*, **58**, 5009–5013.
23. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A. *et al.* (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531–537.
24. Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X. *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, **403**, 503–511.
25. Hoch, R.V., Thompson, D.A., Baker, R.J. and Weigel, R.J. (1999) GATA-3 is expressed in association with estrogen receptor in breast cancer. *Int. J. Cancer*, **84**, 122–128.
26. Hilsenbeck, S.G., Friedrichs, W.E., Schiff, R., O'Connell, P., Hansen, R.K., Osborne, C.K. and Fuqua, S.A. (1999) Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. *J. Natl Cancer Inst.*, **91**, 453–459.
27. Martin, K.J., Kritzman, B.M., Price, L.M., Koh, B., Kwan, C.P., Zhang, X., Mackay, A., O'Hare, M.J., Kaelin, C.M., Mutter, G.L. *et al.* (2000) Linking gene expression patterns to therapeutic groups in breast cancer. *Cancer Res.*, **60**, 2232–2238.
28. Yang, G.P., Ross, D.T., Kuang, W.W., Brown, P.O. and Weigel, R.J. (1999) Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. *Nucleic Acids Res.*, **27**, 1517–1523.
29. Perou, C.M., Jeffrey, S.S., van de Rijn, M., Rees, C.A., Eisen, M.B., Ross, D.T., Pergamenschikov, A., Williams, C.F., Zhu, S.X., Lee, J.C. *et al.* (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl Acad. Sci. USA*, **96**, 9212–9217.
30. Nacht, M., Ferguson, A.T., Zhang, W., Petroziello, J.M., Cook, B.P., Gao, Y.H., Maguire, S., Riley, D., Coppola, G., Landes, G.M. *et al.* (1999) Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. *Cancer Res.*, **59**, 5464–5470.
31. Sgroi, D.C., Teng, S., Robinson, G., LeVangie, R., Hudson Jr, J.R. and Elkahloun, A.G. (1999) In vivo gene expression profile analysis of human breast cancer progression. *Cancer Res.*, **59**, 5656–5661.
32. Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A. *et al.* (2000) Molecular portraits of human breast tumours. *Nature*, **406**, 747–752.
33. Hahnel, E., Harvey, J.M., Joyce, R., Robbins, P.D., Sterrett, G.F. and Hahnel, R. (1993) Stromelysin-3 expression in breast cancer biopsies: clinicopathological correlations. *Int. J. Cancer*, **55**, 771–774.
34. Skoog, L., Humla, S., Klintenberg, C., Pasqual, M. and Wallgren, A. (1985) Receptors for retinoic acid and retinol in human mammary carcinomas. *Eur. J. Cancer Clin. Oncol.*, **21**, 901–906.
35. Thor, A.D., Moore II, D.H., Edgerton, S.M., Kawasaki, E.S., Reihsaus, E., Lynch, H.T., Marcus, J.N., Schwartz, L., Chen, L.C., Mayall, B.H. *et al.* (1992) Accumulation of p53 tumor suppressor gene protein: an independent marker of prognosis in breast cancers. *J. Natl Cancer Inst.*, **84**, 845–855.
36. Allred, D.C., Harvey, J.M., Berardo, M. and Clark, G.M. (1998) Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod. Pathol.*, **11**, 155–168.
37. Spencer, K.S., Graus-Porta, D., Leng, J., Hynes, N.E. and Klemke, R.L. (2000) ErbB2 is necessary for induction of carcinoma cell invasion by ErbB family receptor tyrosine kinases. *J. Cell Biol.*, **148**, 385–397.
38. Behrens, J. (1993) The role of cell adhesion molecules in cancer invasion and metastasis. *Breast Cancer Res. Treat.*, **24**, 175–184.
39. Roberts, D.D. (1996) Regulation of tumor growth and metastasis by thrombospondin-1. *FASEB J.*, **10**, 1183–1191.
40. Taylor-Papadimitriou, J., Burchell, J., Miles, D.W. and Dalziel, M. (1999) MUC1 and cancer. *Biochim. Biophys. Acta*, **1455**, 301–313.
41. Sneath, R.J. and Mangham, D.C. (1998) The normal structure and function of CD44 and its role in neoplasia. *Mol. Pathol.*, **51**, 191–200.
42. Iyer, V.R., Eisen, M.B., Ross, D.T., Schuler, G., Moore, T., Lee, J.C.F., Trent, J.M., Staudt, L.M., Hudson Jr, J., Boguski, M.S. *et al.* (1999) The transcriptional program in the response of human fibroblasts to serum. *Science*, **283**, 83–87.
43. Theillet, C., Adelaide, J., Louason, G., Bonnet-Dorion, F., Jacquemier, J., Adnane, J., Longy, M., Katsaros, D., Sismondi, P., Gaudray, P. *et al.* (1993) FGFRI and PLAT genes and DNA amplification at 8p12 in breast and ovarian cancers. *Genes Chromosomes Cancer*, **7**, 219–226.
44. Granjeaud, S., Nguyen, C., Rocha, D., Luton, R. and Jordan, B.R. (1996) From hybridization image to numerical values: a practical, high throughput quantification system for high density filter hybridizations. *Genet. Anal.*, **12**, 151–162.
45. Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **95**, 14863–14868.
46. Ferrari, S., Battini, R. and Cossu, G. (1990) Differentiation-dependent expression of apolipoprotein A-I in chicken myogenic cells in culture. *Dev. Biol.*, **140**, 430–436.