

MINIREVIEW

DNA Arrays in Clinical Oncology: Promises and Challenges

François Bertucci, Patrice Viens, Rebecca Tagett, Catherine Nguyen, Rémi Houlgatte, and Daniel Birnbaum

Department of Molecular Oncology (FB, DB), Institut Paoli-Calmettes, Department of Medical Oncology (FB, PV), Institut Paoli-Calmettes, University of Méditerranée-UFR Medicine (FB, PV), Ipsogen SA (RT), ERM206 INSERM (CN, RH), and U119 INSERM (DB), IFR57, Marseille, France.

SUMMARY: Cancer is a complex genetic disease characterized by the accumulation of multiple molecular alterations. Current diagnostic and prognostic classifications, based on clinical and pathologic factors, are insufficient to reflect the whole clinical heterogeneity of tumors. Most current anticancer agents do not differentiate between cancerous and normal cells, leading sometimes to disastrous adverse effects. Recent advances in human genome research and high-throughput molecular technologies make it possible finally to tackle the molecular complexity of malignant tumors. With DNA array technology, mRNA expression levels of thousands of genes can be measured simultaneously in a single assay. Oncology is benefiting on multiple fronts. Gene expression profiles are revealing new biologically and clinically relevant tumor subclasses previously indistinguishable and are identifying new diagnostic and prognostic biomarkers as well as new potential therapeutic targets. Here, we review the technology and present clinical applications for which promising results have been obtained. Finally, we discuss issues that must be resolved in the near future to allow DNA arrays to translate into benefits for cancer patients. (*Lab Invest* 2003, 83:305–316).

Cancer is a complex, polygenic disease that arises and progresses from the accumulation of multiple genetic and epigenetic anomalies. This combinatorial origin, the heterogeneity of malignant cells, and a variable host background produce multiple tumor subclasses. Today, cancer classifications are principally based on clinical and morphologic features that only partially reflect this heterogeneity, reducing the probability of the most appropriate diagnostic and therapeutic strategy for each patient. Most current anticancer agents do not differentiate between cancerous and normal cells, resulting in sometimes disastrous toxicity and an inconstant efficiency. The development of innovative drugs that selectively target cancer cells while sparing normal tissues is very promising as suggested by successful recent examples such as the use of mAb therapy against the ERBB2 receptor in breast cancer (O'Dwyer and

Druker, 2000), or the tyrosine kinase inhibitor STI571 in chronic myelogenous leukemia (Pegram et al, 1999).

Although a huge amount of biologic studies has been published so far, only a few molecular markers are used routinely in clinical practice. Besides the disease heterogeneity, there are two main reasons for such a paradox. First, traditional molecular analyses are reductionist, assessing only one or a few genes at a time, thus working with a biologic model too specific and limited to confront a process whose clinical outcome is likely governed by the combined influence of many genes. Second, most clinical study designs have resulted in conflicting and nonconclusive results. Current research aims to decipher, within the complex machinery of genes that dictate tumor evolution, new molecular alterations that would make diagnostic and/or therapeutic interventions more accurate and more specific.

Modern high-throughput RNA expression measurements combined with improved genomic information allow the genetic complexity of cancer to be investigated by simultaneous analysis of tens of thousands of genes in a single step. During the past few years, several gene expression profiling methods have emerged and have been applied successfully to cancer research. These include differential display (Cheng et al, 2002), serial analysis of gene expression (SAGE) (Velculescu et al, 1995), and DNA arrays (Granjeaud et al, 1999). DNA arrays have become prominent be-

DOI: 10.1097/01.LAB.0000059936.28369.19

Received November 25, 2002.

Work in our laboratories is supported by INSERM, CNRS, Institut Paoli-Calmettes, and grants from Association pour la Recherche contre le Cancer, Fédération Nationale des Centres de Lutte Contre le Cancer, and the Ligue Contre le Cancer (Comité National et Comité de Corse du Sud).

Address reprint requests to: Dr. Daniel Birnbaum, Department of Molecular Oncology, Institut Paoli-Calmettes and U119 INSERM, 27 Bd. Léi Roure, 13009 Marseille, France. E-mail: birnbaum@marseille.inserm.fr

cause they are easier to use, do not require large-scale DNA sequencing, and allow the parallel quantification of thousands of genes from multiple samples. Compared with conventional methods (eg, Northern blot, PCR analysis of reverse-transcribed RNA), these “molecular portraits” provide a holistic account of the molecular state of the cell type under scrutiny. Because the biologic diversity of tumors can often be accounted for by changes in gene expression, DNA arrays are expected to result in the identification of new diagnostic and prognostic markers and the development of new specific targeted therapies. This review describes DNA array–based gene expression profiling and its clinical applications in oncology, as well as the challenges that remain for the rapid transfer of technology and/or discoveries to clinical practice.

DNA Array Technology for Gene Expression Profiling

Principle. DNA array technology relies on nucleic acid hybridization between labeled free targets derived from a biologic sample and an array of many DNA fragments (the probes, representing genes of interest), tethered to a solid surface (Fig. 1). The

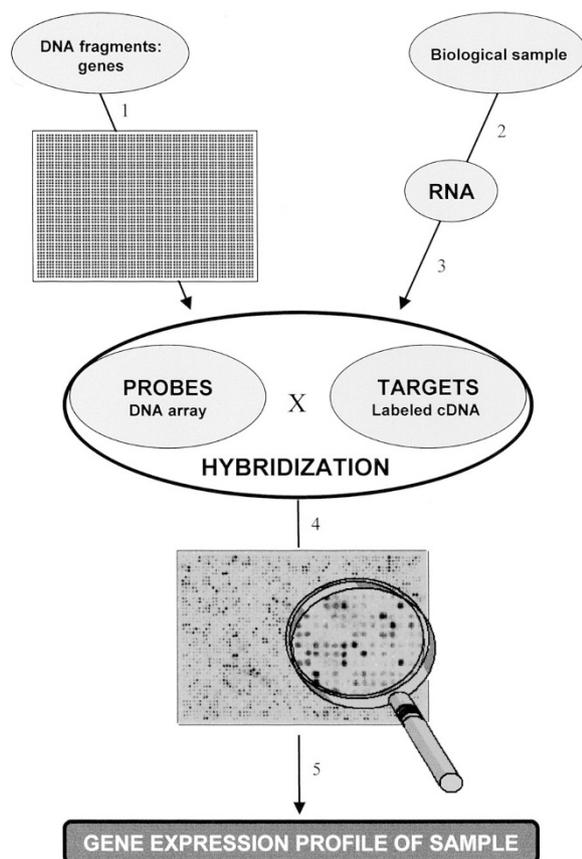


Figure 1.

Principle of gene expression measurement with DNA arrays. Each number defines an experimental step. 1, DNA fragments are represented by cDNA clones robotically spotted or by oligonucleotides in situ synthesized or robotically spotted onto solid surface; 2, RNA is extracted from biologic sample; 3, RNA is simultaneously reverse transcribed and labeled; 4, hybridization image is acquired by scanning; 5, hybridization signals are automatically quantified and data are normalized.

targets, produced by reverse transcription and simultaneous labeling of RNA molecules, are part of a complex mixture of distinct cDNA fragments that hybridize with their cognate probes during the assay. The signal generated on each probe reflects the mRNA expression level of the corresponding gene in the sample. After detection, quantification, and integration of signals with specialized software, intensities are normalized for technical deviations, providing a “gene expression profile” for each sample, comparable to profiles from other samples.

Technologic Platforms. Two main implementations of DNA arrays have been applied with success. The first uses arrays of cDNA clones robotically spotted on a solid surface in the form of PCR products. Several versions exist, depending on the type of support (nylon, glass) and the type of target labeling (radioactivity, colorimetry, fluorescence) (Bertucci et al, 1999b; Chen et al, 1998; Schena et al, 1995). This approach is flexible, allowing researchers to make arrays with their own gene sets, but it requires accurate collection and storage of cDNA clones and PCR products, which may be avoided by using commercially available arrays. Versions such as nylon arrays with radioactive detection are relatively cheap.

The second implementation uses arrays of oligonucleotides either directly synthesized in situ on a support (Hughes et al, 2001; Lockhart et al, 1996) or robotically spotted (Kane et al, 2000). Probes design requires knowledge of the gene sequences. Their length (oligonucleotides of 20 to 80 bp) allows differential detection of members of gene families or alternative transcripts not distinguishable by cDNA spotted arrays. The main drawback remains the elevated cost.

Data Analysis. DNA arrays deliver a new type of data with several thousands of measurements per experiment. The analysis, interpretation, and meaningful display and storage of such a huge volume of data are particularly challenging. Although genes that display extreme expression changes between samples may require specific analysis, the true strength of high-throughput experiments in unraveling the complexity of cancer comes through the mathematical identification of expression patterns or signatures within profiling data. Sophisticated dedicated software developed for this task includes the “unsupervised” and “supervised” varieties (Brazma and Vilo, 2000). Unsupervised methods define classes without any a priori, organizing data by clustering genes and/or samples simply by similarities in their expression profiles. Once clusters are delineated, correlations are searched for between sample clusters and histoclinical parameters, and between gene clusters and functional annotation, chromosomal location, or cell type. One of the most popular clustering tools is hierarchic clustering (Eisen et al, 1998) coupled with a graphic visualization tool (Fig. 2A). Other methods include K-means clustering and self-organizing maps. The resulting sample classification often correlates with a histologic characteristic defined by large sets of genes and not necessarily with the clinical distinction of interest (eg, drug resistance) generally governed by a smaller set. By defining

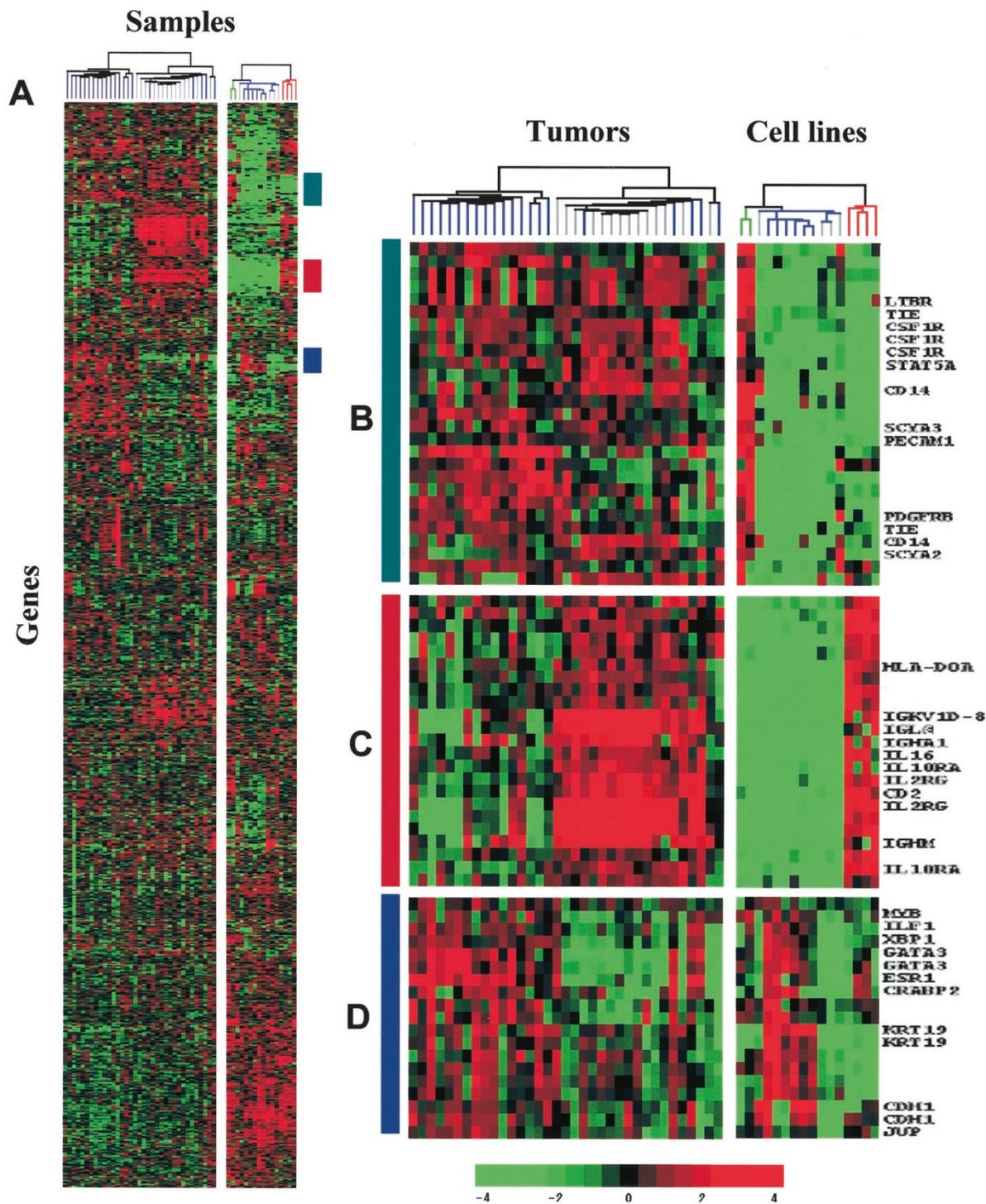


Figure 2.

Hierarchical clustering analysis of cell lines and breast cancer samples. A, Scaled-down matrix representation of expression levels of ~1000 genes in 35 breast cancer samples (*left*) and 16 cell lines (*right*). Each row represents a gene, each column represents a sample, and each cell represents the expression level of a gene in a sample (relative to its median level across all samples) using a color-code scale that ranges from bright green for underexpressed genes to bright red for overexpressed genes. Samples are clustered together along the horizontal axis according to similarity in their expression profiles across all genes, and, conversely, genes are clustered along the vertical axis on the basis of similarity in their expression across all samples. A dendrogram displayed above the samples reflects the degree of similarity between the connected samples (the same dendrogram displayed along the genes has been deleted here for graphic purpose). This classification delineates classes of samples defined by the correlated expression patterns of gene subsets. The colorization allows groups of correlated genes and groups of samples to be highlighted, providing a faster and more intuitive view of the information. The branches of dendrograms are colored, showing the different types of tumors (blue, estrogen receptor (ER)-positive tumors; gray, ER-negative tumors) and cell lines (green, macrophage cell lines; red, B-lymphocyte cell lines; blue or gray, epithelial mammary cell lines). Colored bars on the right indicate the gene clusters shown in B, C, and D. B to D, Identification of gene clusters characteristic of a cell type ("virtual microdissection"). Clusters B (macrophage cluster), C (B-lymphocyte cluster), and D (epithelial cluster) are overexpressed in the corresponding cell lines and include genes classically expressed in these lines (CD14 and other genes involved in macrophage functions, Ig, cytokines, and cytokine receptors for the B-lymphocyte lines, *ESR1* (coding for ER) or keratin 19 for the epithelial mammary cancer cells).

relevant classes before analysis, supervised techniques bypass this issue. These algorithms incorporate external information related to samples studied (eg, survival) to identify the optimal set of genes that best discriminate between the tumor classes. They result in "molecular classifiers" capable of assigning membership of an unknown sample to a predefined class of interest. Generally, a classifier is derived through training on a random subset of chosen samples with known class memberships (learning set) and then validated on the remaining subset, or testing set (of known classification). Once validated, it is applied to classify samples of unknown classification. Supervised methods include support vector machines, weighted votes (Golub et al, 1999), and neural networks (Khan et al, 2001).

Clinical Applications of Gene Expression Profiling in Oncology

DNA arrays are tools that are capable of confronting tumor heterogeneity and of enhancing molecular diagnoses. Current classifications are insufficient to reflect the diversity of cancer. Ideally, they should produce subclasses of tumors defined by a common mechanism of malignant transformation. In this case, patients within the same category would probably display a more uniform clinical outcome. Because of the complexity of disease, a combination of markers identified by comprehensive molecular analysis is likely to be more accurate than a single marker. Although there are multiple studies of cancer profiling using DNA arrays, we present here those that particularly stress a clinical interest for oncology. More fundamental applications are beyond the scope of this review or have been reviewed elsewhere (Clarke et al, 2001).

Diagnosis and Screening. The diagnostic interest was initially suggested by Khan et al (1998), whose expression profiling clusters of cancer cell lines agreed with the organ type of origin. This interest was reinforced by Golub et al (1999), whose tissue sample profiles distinguished between acute myeloid leukemia and acute lymphoblastic leukemia (ALL). It was encouraging that diagnostic power was soon being demonstrated on a variety of other cancers known for diagnostic difficulty in clinical practice. Small round blue-cell tumors of childhood include at least four types: neuroblastoma, rhabdomyosarcoma, non-Hodgkin's lymphoma, and Ewing sarcoma. Therapeutic options, response to therapy, and prognosis depend on the type, but unfortunately their appearance is similar on routine histology. Their distinction is thus primordial and requires sophisticated techniques that sometimes fail. Using gene expression signatures of 83 samples and neural networks, Khan et al (2001) correctly classified the four types of small round blue-cell tumors of childhood samples and identified the genes most relevant to the classification. Similarly, Gordon et al (2002) described a classifier of eight genes whose expression was capable of accurately

distinguishing between malignant pleural mesothelioma and adenocarcinoma of the lung.

New subclasses of tumors with biologic relevance have been revealed and characterized by gene expression profiling. Alizadeh et al (2000) identified two subclasses of diffuse large B-cell lymphoma derived from different stages of B-cell maturation. Armstrong et al (2002) confirmed that ALL with translocation involving the mixed-lineage leukemia gene *MLL* is a unique entity with an expression profile distinct from both ALL and acute myeloid leukemia. Sorlie et al (2001) defined five subclasses of breast cancer, including new ones such as a myoepithelial and a luminal epithelial subclass. Bittner et al (2000) identified two subclasses of cutaneous malignant melanoma that further showed distinct aggressive potential.

Another application is to identify, among the thousands of tested genes, screening or diagnosis markers by comparing gene expression profiles from normal, premalignant, and malignant tissues from the same organ. Kim and colleagues (Wong et al, 2001) profiled ovarian cancer cell lines and healthy human ovarian surface epithelial cell cultures and found that osteopontin had a high cancer/human ovarian surface epithelial ratio. In a subsequent study, they found association between osteopontin plasma levels and ovarian cancer, suggesting its possible role as a screening marker (Kim et al, 2002). Similarly, comparison of gene expression profiles of prostate samples from different types (normal tissue, benign hyperplasia, localized cancer and metastatic hormone-refractory cancer) identified α -methylacyl-CoA race-mase as a potential cancer marker (Rubin et al, 2002). Another original use of DNA arrays for identifying screening markers was the comparison of gene expression profiles of blood samples from women with breast cancer and healthy volunteers (Martin et al, 2001). A panel of genes that were overexpressed in the patients' blood and accurately separated the two populations were identified.

DNA arrays may also contribute to the diagnosis of metastases by identifying the tumor tissue origin. Carcinoma of unknown primary site is relatively rare in the clinic but problematic. Determination of the nature of the original tumor is important for the delivery of an appropriate treatment, but histologically these carcinomas are often identical. Better tumor markers are required to assign metastases to likely primary sites. Studies using cancer cell lines or tumor tissues have shown the capacity of gene expression profiles to classify cancers according to their tissue of origin (Giordano et al, 2001; Khan et al, 1998; Ramaswamy et al, 2001; Scherf et al, 2000; Su et al, 2001). For example, Giordano et al (2001) measured mRNA expression of ~7200 genes in 154 adenocarcinomas of the lung, colon, and ovary. Analysis of data correctly classified all but two samples in agreement with the pathologic assessment of the primary site. Interestingly, further immunohistochemical analysis of the two outliers revealed a diagnosis consistent with the molecular classification in one case (ovarian metastasis

of colon cancer) and a sarcoma in the second case. These studies also showed that metastases generally conserve the expression profile from the tissue of origin, thus suggesting the potential of technology for identifying the tissue origin of carcinoma of unknown primary site. Similarly, Bhattacharjee et al (2001) demonstrated that DNA arrays could discriminate between primary lung adenocarcinomas and metastases of extrapulmonary origin.

Prognosis. Several studies with cancer cell lines have suggested the capacity of DNA arrays for identifying gene sets associated with metastasis or response to treatment. Zajchowski et al (2001) compared gene expression profiles of breast cancer cell lines with varying potential for invasion. They identified 24 genes differentially expressed between weakly and highly invasive lines and showed that their RNA expression profiles were sufficient to predict the aggressiveness, as measured in vitro, of previously uncharacterized cell lines. Other studies have demonstrated correlations between gene expression profiles and profiles of sensitivity of cancer cell lines (NCI60 panel) to certain cytotoxic drugs, among the hundreds or thousands of tested chemical compounds (Scherf et al, 2000; Staunton et al, 2001).

Before any clinical application, these studies must be extended to patient tissue samples, but in vivo, additional complicating factors also govern clinical outcome, such as tumor environment or anatomic or pharmacologic parameters. However, retrospective studies on pretreatment tissue samples have suggested the power of gene expression profiles in prognostic classification of hematologic malignancies (Alizadeh et al, 2000; Devillard et al, 2002; Hofmann et al, 2002; Rosenwald et al, 2002; Shipp et al, 2002; Yeoh et al, 2002) and solid tumors (Ahr et al, 2002; Beer et al, 2002; Bertucci et al, 2002b; Fuller et al, 2002; Garber et al, 2001; Kihara et al, 2001; Pomeroy et al, 2002; Singh et al, 2002; Sorlie et al, 2001; Takahashi et al, 2001; van 't Veer et al, 2002). Characteristics of these studies are summarized in Table 1. The most frequently analyzed cancers are lymphomas and breast cancer. The first encouraging study was done on diffuse large B-cell lymphoma from 40 patients treated with anthracyclin-based chemotherapy (Alizadeh et al, 2000). The tumors with a pattern close to that of germinal center B cells had a significantly better survival than the tumors with expression patterns corresponding to activated B cells. Rosenwald et al (2002) recently confirmed and refined this prognostic stratification in an expanded study of 240 diffuse large B-cell lymphoma samples. They identified 17 genes for which expression predicted survival after chemotherapy, independent of the International Prognostic Index. Yeoh et al (2002) used oligonucleotide arrays to profile leukemia blasts from 360 pediatric ALL samples. Expression signatures identified the six major prognostically important subtypes of leukemia: T-ALL, and B-ALL with a hyperdiploid karyotype or a *BCR-ABL*, *E2A-PBX1*, *TEL-AML1*, or *MLL* gene rearrangement. The authors did not identify a single transcriptional signature that predicted re-

lapse irrespective of the genetic subtype. However, distinct expression profiles associated to relapse were defined within individual subtypes such as T-ALL and hyperdiploid B-ALL.

Similar results have been reported for different clinical forms of breast cancer (Ahr et al, 2002; Bertucci et al, 2002b; Sorlie et al, 2001; van 't Veer et al, 2002). No new histoclinical factor—except the protein overexpression of *ERBB2* and recently of *uPA/PAI-1*—has been validated as a prognostic and/or predictive factor during the past two decades. Although adjuvant chemotherapy improves survival in localized breast cancer, a number of issues remain. In particular, patients with good prognosis need to be more accurately identified to avoid potentially toxic treatment, and patients of poor prognosis who will or will not benefit from the standard adjuvant chemotherapy currently used need to be determined. Analyzing mRNA expression of ~1000 candidate genes in tumor samples from 55 women who were treated with adjuvant anthracyclin-based chemotherapy, we identified a 40-gene set whose expression distinguished three subclasses of tumors that, although balanced with respect to clinicopathologic features, showed significantly different 5-year survival (Bertucci et al, 2002b). van 't Veer et al (2002) measured the expression of ~25,000 unselected genes in tumor samples from women with lymph node-negative good prognosis breast cancer. They identified a predictor set of 70 genes that could discriminate between tumors that would be likely to metastasize and need adjuvant treatment and those that probably would not. Similarly, Sorlie et al (2001) defined five subclasses of locally advanced breast tumors with different survival after neoadjuvant doxorubicin. To explore further the validity of results, we compared the lists of discriminator genes identified in these breast cancer prognostic studies (Ahr et al, 2002; Bertucci et al, 2002b; Sorlie et al, 2001; van 't Veer et al, 2002). Despite several different methodologic aspects, 26 genes were found in at least two lists (Bertucci et al, 2002a). Reassuring is that some have a known prognostic value (eg, *ESR1*, *ERBB2*), but most are not yet associated with prognosis but have functions that make them prime candidates for novel therapeutic targets.

Another use of DNA arrays relies on the search for correlations between gene expression profiles and histoclinical prognostic factors. Each of the latter probably reflects the expression of hundreds of genes. Subtle molecular differences important for clinical outcome thus may be hidden by the rough estimate that these factors provide but be picked up by large-scale expression analyses. Two major prognostic factors of breast cancer have been investigated by comparing the molecular profiles of estrogen receptor (ER)-positive and ER-negative tumors (Bertucci et al, 2000; Gruvberger et al, 2001; Martin et al, 2000; van 't Veer et al, 2002) and profiles of tumors with and without axillary lymph node metastasis (Bertucci et al, 2000; West et al, 2001). The determination of lymph node status currently relies on surgical axillary lymph dissection, which is associated with significant mor-

Table 1. Cancer-DNA Arrays: Prognostic Studies

Reference	Cancer Type	Pathologic Type	Tested Samples ^a	DNA Array	Tested Genes ^b	Clinical End Point	Discriminator Genes ^c
Alizadeh et al, 2000	Lymphoma	Non-Hodgkin's lymphoma: DLBCL	40	cDNA glass fluo	~18,000	5-year survival	~100
Shipp et al, 2002	Lymphoma	Non-Hodgkin's lymphoma: DLBCL	58	oligo glass fluo	~7,000	5-year survival	13
Rosenwald et al, 2002	Lymphoma	Non-Hodgkin's lymphoma: DLBCL	240	cDNA glass fluo	~12,000	5-year survival	17
Devillard et al, 2002	Lymphoma	"Classical" Hodgkin's lymphoma	21	cDNA Nylon RA	~1,000	Survival	72
Yeoh et al, 2002	Leukemia	Pediatric T-ALL and HD B-ALL	43 and 64	oligo glass fluo	~13,000	Continuous CR vs relapse	7 and 20
Sorlie et al, 2001	Breast	Locally advanced adenocarcinoma	51	cDNA glass fluo	~8,000	Survival	496
Bertucci et al, 2002a	Breast	Localized poor prognosis adenocarcinoma	55	cDNA Nylon RA	~1,000	5-year survival	40
van't Veer et al, 2002	Breast	Localized good prognosis adenocarcinoma	97	oligo glass fluo	~25,000	Metastasis-free survival	70
Ahr et al, 2002	Breast	Adenocarcinoma	55	cDNA Nylon RA	~600	Metastatic relapse	41
Kihara et al, 2001	Esophageal	Squamous cell carcinoma	26	cDNA glass fluo	~9,000	Survival	52
Takahashi et al, 2001	Renal	Clear cell adenocarcinoma	29	cDNA glass fluo	~22,000	5-year survival	51
Garber et al, 2001	Lung	Adenocarcinoma	31	cDNA glass fluo	~24,000	Survival	149
Beer et al, 2002	Lung	Stage I adenocarcinoma	67	oligo glass fluo	~7,000	Survival	50
Pomeroy et al, 2002	CNS	Medulloblastoma	60	oligo glass fluo	~7,000	5-year survival	8
Singh et al, 2002	Prostate	Adenocarcinoma	21	oligo glass fluo	~13,000	Disease-free survival	5

DLBCL, diffuse large B-cell lymphoma; fluo, fluorescence; RA, radioactivity; ALL, acute lymphoblastic leukemia; HD, hyperdiploid; CR, complete remission; CNS, central nervous system.

^a Number of patients included in prognosis analysis (the table is limited to studies including 20 or more patients).^b Number of genes tested.^c Number of discriminator genes identified.

bility. Sentinel lymph node biopsy is being evaluated to replace classical invasive dissection. However, in both cases, false-negative results are possible. Accurate prediction of the axillary status from analysis of tumors would obviate the recourse to lymph node surgery. Among differentially expressed genes that we identified between tumors with and without node metastasis, some had a function in agreement with a potential role in invasion (eg, *ERBB2*, *CDH1*), whereas for others (eg, *SOX4*, *GSTP1*), the connection was not clear, calling for further investigations (Bertucci et al, 2000).

Potential Pitfalls in Gene Expression Profiling of Cancer Samples

Altogether, these studies have revealed the great transcriptional heterogeneity of tumors. They have shown the potential of DNA arrays to discriminate, from RNA expression level of dozens of genes and among classically indistinguishable tumors, new biologically and/or clinically relevant subclasses that probably represent different diseases that require different management. Results, obtained on a relatively small number of samples, must now be validated and refined. Future studies with more samples and more genes will tell us whether it is possible to improve predictive power to 100% accuracy, but there is no guarantee (Ince and Weinberg, 2002). Such stratification, together with the increasing availability of new alternative diagnostic and therapeutic options, is expected to guide patients toward the strategy most likely to succeed for them. The characterization of discriminator genes will provide new markers that are useful for screening, diagnosis, prognosis, and monitoring and will help in deciphering the pathways involved in malignant transformation and in developing new molecularly targeted anticancer drugs. However, the remaining obstacles must not go unnoticed amid the enthusiasm generated. Beyond the yet limited access and complexity of DNA array technology, both of which are now rapidly improving, several experimental issues and pitfalls still may complicate investigations with clinical specimens in cancer research and blur the results.

Tumor Specimens. DNA array experiments require high-quality RNA. Unfortunately, although current RNA extraction methods work well with frozen tumor specimens, they perform poorly with formalin-fixed, paraffin-embedded tissues, which constitute the bulk of pathology archives. Although large numbers of archival frozen samples are available in many clinical institutions, they are often suboptimal with respect to RNA quality, preservation, or clinical information. Today, the development of high-throughput molecular analyses makes researchers even more aware of the crucial need to collect, identify, and store high-quality specimens in tumor banks. Careful and rapid processing of specimens from the clinic to the laboratory to the freezer should obey strict standardized protocols. Banks should be linked to a searchable database that contains all appropriate histoclinical information, in-

cluding treatment and outcome. Organized institutional ethics, informed patient consents, and patient confidentiality are important logistical challenges. The collection of adequate specimens will be improved if physicians and patients are better educated on the one hand and if these procedures and requirements become a component of all ongoing clinical trials on the other hand.

The small size of many clinical specimens from early diagnoses and new minimally invasive diagnostic procedures is another critical issue. Efforts are under way to reduce the amount of sample required for analysis. Most platforms of DNA array work with a few micrograms of mRNA, except for nylon microarrays with radioactive detection, which use only a few nanograms (Bertucci et al, 1999a). One solution is to amplify the sample mRNA using linear amplification methods before labeling (Luo et al, 1999).

Variability of Data. There are potential sources of experimental and biologic variability in DNA array experiments that may affect and complicate analyses. Because of the numerous error-prone steps in experiments, acquisition of valid data requires many quality controls aimed at ensuring excellent hybridization conditions, correct signal-to-noise ratio, dynamic and linear range, sensitivity, and reproducibility. These issues require experiments to be replicated with non-precious biologic material (eg, cell lines) to understand better and eliminate the sources of errors before analysis of precious tumor specimens.

Factors of biologic variability related to tumor tissue samples include their handling and the heterogeneity of tumor cells. Another factor is that solid tumors contain several cell types ("normal" and malignant cells) in different proportions and functional status. An expression profile from such tissue in fact represents a snapshot of the genes expressed by many cell types at a given moment. Taking this into consideration, the pathologists should carefully macrodissect zones enriched in tumor cells before RNA extraction. This solution may further be reinforced during data analysis by confronting, with clustering techniques, expression profiles of heterogeneous specimens with those of cell lines that represent the cell types present in the sample. Such an approach allows the isolation of independent gene clusters characteristic of a cell type (Fig. 2, B to D). Another strategy, more drastic but also more difficult and labor-intensive, lies in the use of laser microdissection, which allows the procurement of pure cell subpopulations from frozen or fixed tissue (Emmert-Buck et al, 1996; Lechner et al, 2001). However, the low amount of resulting RNA and its poor quality currently make difficult its application to DNA array analyses. Moreover, depending on the question, isolation of pure cancer cells may not be desirable because tumor development is highly influenced by interactions of malignant cells with surrounding non-malignant cells.

Handling of Data. All types of DNA array experiments depend on the statistical significance of observed correlations (Nadon and Shoemaker, 2002), but the great dissymmetry of variables poses a statistical

problem: the number of hybridized samples is greatly inferior to the number of genes being tested (multiple hypothesis testing). This problem can be encountered by analyzing several hundreds of samples, by confirming correlations on an independent set of cases, or by randomly permutating the labels of classes and comparing the correlations obtained with random data and with actual data. The biologic interpretation of statistically validated expression profiles may be enhanced by parallel analysis of cell models that represent different stages of tumor maturation or different exposures to stimuli as well as by the functional annotation of the discriminator genes (Alizadeh et al, 2000).

A major and critical complication is to produce data that can be combined and compared within and between laboratories. An ideal marker (either single gene or group of genes) should be consistent across all assays, obtained in a reproducible way by any laboratory, using any platform, but the intrinsic variability of DNA array data, the use of different technological platforms with different experimental conditions, different gene sets, different normalization procedures without any all-encompassing standard, and the analysis of samples from different patients make it difficult to combine and compare independent data. Ideally, different platforms should be compared and the protocols should be standardized using the same clinical specimens. This is difficult because of the value and rarity of human cancer samples. All data should be freely accessible in the public domain, in an ordered, comprehensive, and standardized form. International efforts are under way notably by the Microarray Gene Expression Database society to create a universal public expression database in a fully annotated format, with enough data and experimental information to allow everyone to reproduce the experiments or analyses (Brazma et al, 2001). These efforts will facilitate comparison of the different experimental procedures and development of new analytic tools and will help scientists to validate observed correlations and to perform meta-analyses.

Another issue is how to use expression profiles of discriminator genes for classifying new samples into diagnostic or prognostic categories. Current software classifies samples with reference to other samples, the data of which come from previous measurements on the same platform. A recently published alternative uses pairwise expression ratios of the most discriminator genes and may avoid platform-related problems (Gordon et al, 2002).

Altogether, these issues are leading to new collaborations between researchers across disciplines (physicians, biologists, mathematicians, statisticians, and computer scientists) for the development of adequate laboratory information management systems that are capable of confronting the clinical parameters of tumors with expression data. Large-scale multicenter cancer genomics projects are attempting to take on all of the potential pitfalls. The International Genomics Consortium (www.intgen.org/), for example, aims to perform gene expression profiles of ~10,000 tumor

samples during the next 3 years using standardized procedures and storing data in public databases.

Challenges before Clinical Transfer

In addition to the above cited issues, other important steps must be addressed before any routine clinical application.

Clinical Trials. Clinical trials will be the first to use DNA arrays for molecular diagnosis. Technology should be systematically incorporated both to identify new markers and to investigate the probability of observing an activity of the drug under investigation. Inclusion of patients in therapeutic trials is currently based on both clinical and pathologic criteria. The interpretation of results—and the future of the tested drug—will greatly benefit from gene expression data that may define subclasses of patients for which the drug seems particularly efficient. Once identified, such molecular markers will accelerate and refine the process of subsequent clinical trials by allowing the definition of smaller but truly homogeneous groups of patients.

In the past, many barriers have prevented the transfer of molecular tests from research to patient management. Statistically valid conclusions could rarely be reached for several reasons, including the global incomparability of studies. Divergences could not be interpreted, and small data sets could not be merged. Today, researchers are poised to learn from past mistakes and to face the challenges of high-throughput technologies such as DNA arrays. Ongoing retrospective large-scale studies will have to confirm published data and demonstrate the value of the technology if patient treatment is to be improved. If confirmed, that clinical utility will have to be assessed in prospective randomized clinical trials. Sensitivity, specificity and reproducibility, technical feasibility outside large academic centers, and cost will have to be addressed, and experimental conditions will have to be standardized. Adequate design will also require a sufficient number of carefully selected samples, sufficiently long follow-up, and meaningful end points (response rate or survival). A successful example of a well-designed approach is the assessment of uPA/PAI-1 in breast cancer (Janicke et al, 2001).

Combining DNA Arrays with Other High-Throughput Molecular Analyses. In the future, gene expression profiling of clinical specimens will have to be associated with other emergent high-throughput genome and proteome analyses. Among promising technologies are comparative genomic hybridization arrays (Pinkel et al, 1998), two-dimensional gel electrophoresis and mass spectrometry, and protein arrays (Lawrie et al, 2001). Coordinated strategies thus are required to manage and store samples such that DNA, RNA, and proteins are preserved.

Data Validation. Once a potential marker or target has been identified, and before clinical application, it needs to be validated on a wider scale. The first step is to select, among the numerous identified molecules, the candidate(s) that must be prioritized for future

investigations. Genes coding for secreted proteins are good candidates for screening markers. Genes coding for membrane-associated proteins or enzymes may offer therapeutic targets. Once candidates are selected, their clinical value must be evaluated on a larger series of specimens with long follow-up. Tissue microarrays offer a potent tool to assess rapidly and efficiently the correlations discovered with DNA arrays (Hoos and Cordon-Cardo, 2001; Nocito et al, 2001; Rimm et al, 2001). These consist of small tissue cores (0.6 mm in diameter) of up to 1000 different formalin-fixed specimens that are arrayed on a glass slide and may be queried at the DNA (fluorescence in situ hybridization), RNA (in situ hybridization), and protein levels (immunohistochemistry). Several studies have shown the potential of combining such technology with DNA array-based gene expression profiling (Barlund et al, 2000; Ginestier et al, 2002). When multiple candidate genes must be validated or the antibody is not available, another recently developed tool that can quantify RNA expression level of many separate genes in many samples simultaneously is quantitative PCR of reverse-transcribed RNA (RQ-PCR) (Gibson et al, 1996).

The Future of DNA Arrays

The type of gene expression profiling-based diagnostic platform that will be used in routine clinical practice is not yet defined. Recent studies have shown that combined RNA expression of only a few dozen genes may provide sufficient information for diagnostic and prognostic purposes. Specialized disease-specific DNA arrays that contain only discriminator genes are presently being developed. Quantitative PCR of reverse-transcribed RNA is an alternative solution if the number of genes is manageable. The strenuous need for high-quality RNA extracted from frozen specimens might make immunohistochemistry more practical in routine, but RNA and protein levels do not always correlate and immunohistochemistry has some disadvantages, too: it is not quantitative and may lack diagnostic information provided by RNA expression profiles, it requires production and validation of antibodies, and analysis of combinations of proteins on the same specimen is difficult.

The recent development of high-throughput technologies has opened a new era of biomedical research and hopefully will boost the exciting field of molecular medicine. DNA arrays provide unprecedented tools that confront the complexity of tumors. Preliminary results promise a better approach to cancer management and cure. The current challenge is to demonstrate the clinical benefits for patients, but the potential is enormous. It is anticipated that measuring the genetic activity of tumors will lead to more correct disease diagnoses and treatments. Indeed, future therapies may be based on the vulnerable points of tumor progression identified through expression analyses. Successful implementation of technology in clinical practice will depend on progress in collection of specimens and in technology, on collaborations be-

tween scientists from different disciplines, and on well-designed clinical studies. It is likely that such molecular approaches will affect the current generation of physicians in charge of cancer patients, hopefully transforming cancer management into a more structured and logical science and a successful medicine.

Acknowledgments

We thank Dominique Maraninchi and Claude Mawas for encouragement.

References

- Ahr A, Karn T, Solbach C, Seiter T, Strebhardt K, Holtrich U, and Kaufmann M (2002). Identification of high risk breast-cancer patients by gene expression profiling. *Lancet* 359: 131–132.
- Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, and Staudt LM (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403:503–511.
- Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, Sallan SE, Lander ES, Golub TR, and Korsmeyer SJ (2002). MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* 30:41–47.
- Barlund M, Forozaan F, Kononen J, Bubendorf L, Chen Y, Bittner ML, Torhorst J, Haas P, Bucher C, Sauter G, Kallioniemi OP, and Kallioniemi A (2000). Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. *J Natl Cancer Inst* 92:1252–1259.
- Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, Lizyness ML, Kuick R, Hayasaka S, Taylor JM, Iannettoni MD, Orringer MB, and Hanash S (2002). Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 8:816–824.
- Bertucci F, Bernard K, Loriod B, Chang YC, Granjeaud S, Birnbaum D, Nguyen C, Peck K, and Jordan BR (1999a). Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. *Hum Mol Genet* 8:1715–1722.
- Bertucci F, Eisinger F, Houlgatte R, Viens P, and Birnbaum D (2002a). Gene-expression profiling and identification of patients at high risk of breast cancer. *Lancet* 360:173–174.
- Bertucci F, Houlgatte R, Benziane A, Granjeaud S, Adelaide J, Tagett R, Loriod B, Jacquemier J, Viens P, Jordan B, Birnbaum D, and Nguyen C (2000). Gene expression profiling of primary breast carcinomas using arrays of candidate genes. *Hum Mol Genet* 9:2981–2991.
- Bertucci F, Nasser V, Granjeaud S, Eisinger F, Adelaide J, Tagett R, Loriod B, Giaconia A, Benziane A, Devillard E, Jacquemier J, Viens P, Nguyen C, Birnbaum D, and Houlgatte R (2002b). Gene expression profiles of poor-prognosis primary breast cancer correlate with survival. *Hum Mol Genet* 11:863–872.
- Bertucci F, Van Hulst S, Bernard K, Loriod B, Granjeaud S, Tagett R, Starkey M, Nguyen C, Jordan B, and Birnbaum D

- (1999b). Expression scanning of an array of growth control genes in human tumor cell lines. *Oncogene* 18:3905–3912.
- Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, Lander ES, Wong W, Johnson BE, Golub TR, Sugarbaker DJ, and Meyerson M (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 98:13790–13795.
- Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D, and Sondak V (2000). Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 406:536–540.
- Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansong W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FCP, Kim IF, Markowitz V, Matese JC, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Ronald Taylor R, Vilo J, and Vingron M (2001). Minimum Information About a Microarray Experiment (MIAME): Toward standards for microarray data. *Nat Genet* 29:365–371.
- Brazma A and Vilo J (2000). Gene expression data analysis. *FEBS Lett* 480:17–24.
- Chen JJ, Wu R, Yang PC, Huang JY, Sher YP, Han MH, Kao WC, Lee PJ, Chiu TF, Chang F, Chu YW, Wu CW, and Peck K (1998). Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. *Genomics* 51:313–324.
- Cheng Q, Lau WM, Tay SK, Chew SH, Ho TH, and Hui KM (2002). Identification and characterization of genes involved in the carcinogenesis of human squamous cell cervical carcinoma. *Int J Cancer* 98:419–426.
- Clarke PA, te Poele R, Wooster R, and Workman P (2001). Gene expression microarray analysis in cancer biology, pharmacology, and drug development: Progress and potential. *Biochem Pharmacol* 62:1311–1336.
- Devilard E, Bertucci F, Trempat P, Bouabdallah R, Loriod B, Giaconia A, Brousset P, Granjeaud S, Nguyen C, Birnbaum D, Birg F, Houlgatte R, and Xerri L (2002). Gene expression profiling defines molecular subtypes of classical Hodgkin's disease. *Oncogene* 21:3095–3102.
- Eisen MB, Spellman PT, Brown PO, and Botstein D (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863–14868.
- Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA, and Liotta LA (1996). Laser capture microdissection. *Science* 274:998–1001.
- Fuller GN, Hess KR, Rhee CH, Yung WK, Sawaya RA, Bruner JM, and Zhang W (2002). Molecular classification of human diffuse gliomas by multidimensional scaling analysis of gene expression profiles parallels morphology-based classification, correlates with survival, and reveals clinically-relevant novel glioma subsets. *Brain Pathol* 12:108–116.
- Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van de Rijn M, Rosen GD, Perou CM, Whyte RI, Altman RB, Brown PO, Botstein D, and Petersen I (2001). Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci USA* 98:13784–13789.
- Gibson UE, Heid CA, and Williams PM (1996). A novel method for real time quantitative RT-PCR. *Genome Res* 6:995–1001.
- Ginestier C, Charaffe-Jauffret E, Bertucci F, Eisinger F, Geneix J, Bechlian D, Conte N, Adélaïde J, Toiron Y, Nguyen C, Viens P, Moziconacci MJ, Houlgatte R, Birnbaum D, and Jacquemier J (2002). Distinct and complementary information provided by use of tissue and DNA microarrays in the study of breast tumor markers. *Am J Pathol* 161:1223–1233.
- Giordano TJ, Shedden KA, Schwartz DR, Kuick R, Taylor JM, Lee N, Misek DE, Greenson JK, Kardias SL, Beer DG, Rennett G, Cho KR, Gruber SB, Fearon ER, and Hanash S (2001). Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. *Am J Pathol* 159:1231–1238.
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, and Lander ES (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286:531–537.
- Gordon GJ, Jensen RV, Hsiao LL, Gullans SR, Blumenstock JE, Ramaswamy S, Richards WG, Sugarbaker DJ, and Bueno R (2002). Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. *Cancer Res* 62:4963–4967.
- Granjeaud S, Bertucci F, and Jordan BR (1999). Expression profiling: DNA arrays in many guises. *Bioessays* 21:781–790.
- Gruvberger S, Ringner M, Chen Y, Panavally S, Saal LH, Borg A, Ferno M, Peterson C, and Meltzer PS (2001). Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res* 61:5979–5984.
- Hofmann WK, de Vos S, Elashoff D, Gschaidmeier H, Hoelzer D, Koeffler HP, and Ottmann OG (2002). Relation between resistance of Philadelphia-chromosome-positive acute lymphoblastic leukaemia to the tyrosine kinase inhibitor ST1571 and gene-expression profiles: A gene-expression study. *Lancet* 359:481–486.
- Hoos A and Cordon-Cardo C (2001). Tissue microarray profiling of cancer specimens and cell lines: Opportunities and limitations. *Lab Invest* 81:1331–1338.
- Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephanians SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, and Linsley PS (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 19:342–347.
- Ince TA and Weinberg RA (2002). Functional genomics and the breast cancer problem. *Cancer Cell* 1:15–17.
- Janicke F, Prechtel A, Thomssen C, Harbeck N, Meisner C, Untch M, Sweep CG, Selbmann HK, Graeff H, and Schmitt M (2001). Randomized adjuvant chemotherapy trial in high-risk, lymph node-negative breast cancer patients identified by urokinase-type plasminogen activator and plasminogen activator inhibitor type 1. *J Natl Cancer Inst* 93:913–920.
- Kane MD, Jatko TA, Stumpf CR, Lu J, Thomas JD, and Madore SJ (2000). Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* 28:4552–4557.

- Khan J, Simon R, Bittner M, Chen Y, Leighton SB, Pohida T, Smith PD, Jiang Y, Gooden GC, Trent JM, and Meltzer PS (1998). Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res* 58:5009–5013.
- Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, and Meltzer PS (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 7:673–679.
- Kihara C, Tsunoda T, Tanaka T, Yamana H, Furukawa Y, Ono K, Kitahara O, Zembutsu H, Yanagawa R, Hirata K, Takagi T, and Nakamura Y (2001). Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles. *Cancer Res* 61:6474–6479.
- Kim JH, Skates SJ, Uede T, Wong KK, Schorge JO, Feltmate CM, Berkowitz RS, Cramer DW, and Mok SC (2002). Osteopontin as a potential diagnostic biomarker for ovarian cancer. *JAMA* 287:1671–1679.
- Lawrie LC, Fothergill JE, and Murray GI (2001). Spot the differences: Proteomics in cancer research. *Lancet Oncol* 2:270–277.
- Lechner S, Muller-Ladner U, Neumann E, Dietmaier W, Welsh J, Scholmerich J, Ruschoff J, and Kullmann F (2001). Use of simplified transcripts for the analysis of gene expression profiles in laser-microdissected cell populations. *Lab Invest* 81:1233–1242.
- Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, and Brown EL (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14:1675–1680.
- Luo L, Salunga RC, Guo H, Bittner A, Joy KC, Galindo JE, Xiao H, Rogers KE, Wan JS, Jackson MR, and Erlander MG (1999). Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat Med* 5:117–122.
- Martin KJ, Graner E, Li Y, Price LM, Kritzman BM, Fournier MV, Rhei E, and Pardee AB (2001). High-sensitivity array analysis of gene expression for the early detection of disseminated breast tumor cells in peripheral blood. *Proc Natl Acad Sci USA* 98:2646–2651.
- Martin KJ, Kritzman BM, Price LM, Koh B, Kwan CP, Zhang X, Mackay A, O'Hare MJ, Kaelin CM, Mutter GL, Pardee AB, and Sager R (2000). Linking gene expression patterns to therapeutic groups in breast cancer. *Cancer Res* 60:2232–2238.
- Nadon R and Shoemaker J (2002). Statistical issues with microarrays: Processing and analysis. *Trends Genet* 18:265–271.
- Nocito A, Kononen J, Kallioniemi OP, and Sauter G (2001). Tissue microarrays (TMAs) for high-throughput molecular pathology research. *Int J Cancer* 94:1–5.
- O'Dwyer ME and Druker BJ (2000). Status of bcr-abl tyrosine kinase inhibitors in chronic myelogenous leukemia. *Curr Opin Oncol* 12:594–597.
- Pegram M, Hsu S, Lewis G, Pietras R, Beryt M, Sliwkowski M, Coombs D, Baly D, Kabbinnar F, and Slamon D (1999). Inhibitory effects of combinations of HER-2/neu antibody and chemotherapeutic agents used for treatment of human breast cancers. *Oncogene* 18:2241–2251.
- Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, and Albertson DG (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 20:207–211.
- Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C, Allen JC, Zagzag D, Olson JM, Curran T, Wetmore C, Biegel JA, Poggio T, Mukherjee S, Rifkin R, Califano A, Stolovitzky G, Louis DN, Mesirov JP, Lander ES, and Golub TR (2002). Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 415:436–442.
- Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, and Golub TR (2001). Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* 98:15149–15154.
- Rimm DL, Camp RL, Charette LA, Costa J, Olsen DA, and Reiss M (2001). Tissue microarray: A new technology for amplification of tissue resources. *Cancer J* 7:24–31.
- Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, Lopez-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, and Staudt LM (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 346:1937–1947.
- Rubin MA, Zhou M, Dhanasekaran SM, Varambally S, Barrette TR, Sanda MG, Pienta KJ, Ghosh D, and Chinnaiyan AM (2002). Alpha-methylacyl coenzyme A racemase as a tissue biomarker for prostate cancer. *JAMA* 287:1662–1670.
- Schena M, Shalon D, Davis RW, and Brown PO (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–470.
- Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, Kohn KW, Reinhold WC, Myers TG, Andrews DT, Scudiero DA, Eisen MB, Sausville EA, Pommier Y, Botstein D, Brown PO, and Weinstein JN (2000). A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 24:236–244.
- Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, Ray TS, Koval MA, Last KW, Norton A, Lister TA, Mesirov J, Neuberg DS, Lander ES, Aster JC, and Golub TR (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 8:68–74.
- Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D'Amico AV, Richie JP, Lander ES, Loda M, Kantoff PW, Golub TR, and Sellers WR (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 1:203–209.
- Sortie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P, and Borresen-Dale AL (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98:10869–10874.

Staunton JE, Slonim DK, Collier HA, Tamayo P, Angelo MJ, Park J, Scherf U, Lee JK, Reinhold WO, Weinstein JN, Mesirov JP, Lander ES, and Golub TR (2001). Chemosensitivity prediction by transcriptional profiling. *Proc Natl Acad Sci USA* 98:10787–10792.

Su AI, Welsh JB, Sapinoso LM, Kern SG, Dimitrov P, Lapp H, Schultz PG, Powell SM, Moskaluk CA, Frierson HF Jr, and Hampton GM (2001). Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Res* 61:7388–7393.

Takahashi M, Rhodes DR, Furge KA, Kanayama H, Kagawa S, Haab BB, and Teh BT (2001). Gene expression profiling of clear cell renal cell carcinoma: Gene identification and prognostic classification. *Proc Natl Acad Sci USA* 98:9754–9759.

van 't Veer LJ, Dai H, van De Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van Der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, and Friend SH (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415:530–536.

Velculescu VE, Zhang L, Vogelstein B, and Kinzler KW (1995). Serial analysis of gene expression. *Science* 270:484–487.

West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zuzan H, Olson JA Jr, Marks JR, and Nevins JR (2001). Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci USA* 18:18.

Wong KK, Cheng RS, and Mok SC (2001). Identification of differentially expressed genes from ovarian cancer cells by MICROMAX cDNA microarray system. *Biotechniques* 30:670–675.

Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, Behm FG, Raimondi SC, Relling MV, Patel A, Cheng C, Campana D, Wilkins D, Zhou X, Li J, Liu H, Pui CH, Evans WE, Naeve C, Wong L, and Downing JR (2002). Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1:133–143.

Zajchowski DA, Bartholdi MF, Gong Y, Webster L, Liu HL, Munishkin A, Beauheim C, Harvey S, Ethier SP, and Johnson PH (2001). Identification of gene expression profiles that predict the aggressive behavior of breast cancer cells. *Cancer Res* 61:5168–5178.