

## The Use of DNA Microarrays to Assess Clinical Samples: The Transition from Bedside to Bench to Bedside

JOHN A. COPLAND,\* PETER J. DAVIES,<sup>†</sup> GREGORY L. SHIPLEY,<sup>†</sup>  
CHRISTOPHER G. WOOD,<sup>‡</sup> BRUCE A. LUXON,<sup>||</sup> AND RANDALL J. URBAN\*

*\*Endocrinology Division, Internal Medicine, University of Texas Medical Branch, Galveston, Texas 77555; †Department of Pharmacology, University of Texas-Houston Medical School, Houston, Texas 77030; ‡Urology Department, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030; ||Human Biological Chemistry & Genetics Departments, University of Texas Medical Branch, Galveston, Texas 77555*

### ABSTRACT

The advent of gene array technology brings the ability to classify disease states to the molecular level by examining changes in all mRNAs expressed in cells or tissues. Comparing changes in gene expression patterns between normal and diseased cells and/or tissues has elucidated unique subsets of genes identifiable to a specific disease. Already, new subclassifications of specific cancers have been discovered, belying that genomic profiling can uniquely distinguish a specific disease state and tissue of origin. This technology bestows the ability to examine global changes occurring in a cell or tissue(s), thereby allowing the elucidation of alterations in dysregulated biological, biochemical, and molecular events leading to disease states such as diabetes, hypertension, infertility, obesity, osteoporosis, and atherosclerosis. Furthermore, genomic profiling will lead to new molecular targets for the development of drug therapeutics. Futuristically, one could envision personalized patient therapies based upon identification of specific aberrant signaling pathways that can be targeted for drug therapy.

### I. Introduction

The introduction of gene array technology in the mid 1990s (Skena *et al.*, 1995; DeRisi *et al.*, 1996) has led to a phenomenal ability to develop new subclasses of common diseases, predict disease outcome, and identify novel molecular targets for potential drug therapy. DNA arrays allow for simultaneous quantitative measurement of mRNA expression by thousands of genes in a biological sample. It is expected that this “gene profiling” will play a key role in understanding drug side effects at the molecular level and, one day, provide the rationale for individualized drug therapies for each patient, providing effective treatment and decreased side effects. Thus, clinicians with valuable patient tissue and blood samples are needed to use gene array technology to help create these novel molecular profiles, especially for endocrine-linked diseases such as diabe-

tes, hypertension, multiple endocrine neoplasia (MEN) syndromes, other cancers, infertility, obesity, osteoporosis, atherosclerosis, and leiomyomas (Korach, 2002). Large biological sample numbers will be required to develop precise gene profiles.

As shown in Figure 1, clinical applications demonstrate the uses of gene array data in which multivariate analyses via gene array detection supercede the conventional gene-by-gene analysis approach that is limited by biological insight. This systemic approach allows for elucidation of molecular complexities in alterations in signal transduction pathways that alter disease processes. Thus, the overall goal should be to complete construction of the roadmap identifying each molecule in all signaling pathways in each and every cell type known to regulate cellular functions as well as to characterize between signaling pathways (for 14 elegant reviews demonstrating signal transduction pathways important to cell function, see *Science* 296:1632–1657). Unique alterations in the roadmap should be predictive of specific diseases and the phenomena of different endocrine-driven stages of life, such as onset of puberty, pregnancy, and aging.

The premise of gene microarray technology is that steady-state mRNA levels are altered in diseased cells/tissues. These alterations determine changes in function and phenotype. Thus, it would be expected that tens to hundreds of genes out of the total expressed genome would be altered, depending upon the disease state. Therefore, it should be expected that multiple alterations within a single signaling pathway would occur as well as among signaling pathways due to changes in the dynamics of crosstalk between paths. The premise further dictates that, for each specific disease state, a unique subset of mRNAs will be

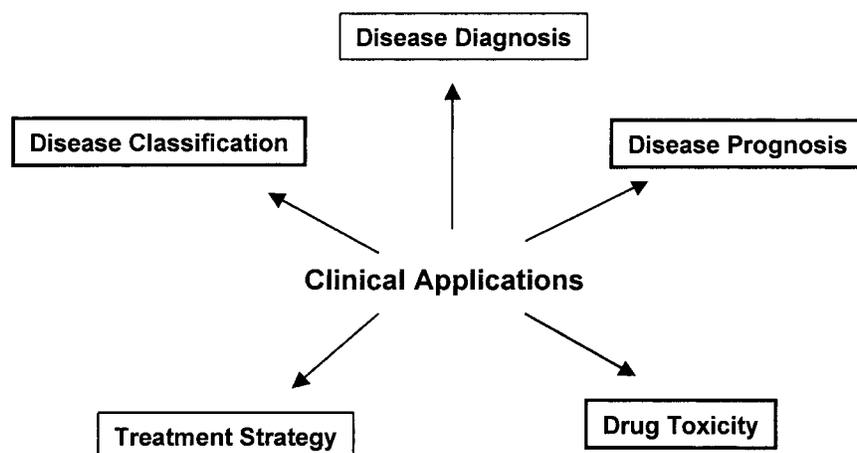


FIG. 1. Clinical applications of microarray technology.

altered, creating a signature or fingerprint pattern that identifies that disease as well as the cell type of origin. In the cancer arena, this has proven true: gene array profiles have been performed upon different types of cancers, each replicating a signature pattern correlating to tissue of origin (Ross *et al.*, 2000; Ramswamy *et al.*, 2001; Nielsen *et al.*, 2002). Moreover, new cancer subclassifications have been discovered using gene profiling for lung adenocarcinoma, breast carcinoma, leukemia, lymphoma, melanoma, and colon carcinoma (Khan *et al.*, 1998; Golub *et al.*, 1999; Alizadeh *et al.*, 2000; Ross *et al.*, 2000; Garber *et al.*, 2001; Notterman *et al.*, 2001; Sorlie *et al.*, 2001; Armstrong *et al.*, 2002). This strategy — combined with clinical parameters — is leading to well-defined disease classifications and the ability to predict clinical outcome (Bhattacharjee *et al.*, 2001; Lakhani and Ashworth, 2001; van't Veer *et al.*, 2002).

While gene array technology represents a marvelous opportunity to meld the clinical researcher with the basic scientist, three significant issues must be considered when attempting to derive meaningful array data from a clinical research protocol. These problems are 1) the quality and amount of tissue sample from which the cDNA for hybridization is derived, 2) the type of array to be used for the tissue derived-cDNA, and 3) the extent and nature of the data analysis.

## II. Tissue Samples

The source of human mRNA remains a major challenge, slowing widespread use of gene array technology in clinical research. The two major problems with obtaining tissue are the amount that can be obtained (which limits the quantity of mRNA that can be isolated) and the complexity of the tissue sample itself. Therefore, clinical studies that are published using DNA gene arrays are performed in settings where tissue/blood are abundant and readily obtainable. Typically, 1–2 micrograms of mRNA or 10 micrograms of total RNA are required for an oligo or cDNA array. Nylon microarrays, which use radioactive detection, need only nanogram quantities of mRNA (Bertucci *et al.*, 1999). Another solution is to subject the sample mRNA to linear amplification methods, which require  $\leq 50$  ng of mRNA for gene array analysis (Van Gelder *et al.*, 1990; Phillips and Eberwine, 1996; Wang *et al.*, 2000). Ambion (Austin, TX) sells a kit based upon Eberwine's linear amplification of mRNA into antisense RNA. This is an important development, since many clinical samples are small.

Although nearly all clinical gene array analyses have been performed on cancer specimens, to profile differences between tumor and normal tissues as well as between different tumors, we have included other examples of clinical samples that have utilized gene arrays. The reader is referred to the June 2002 issue of *Endocrinology*, which highlights the impact of the human genome upon endocrinology. Gene expression patterns were measured in subjects with scleroderma from inflammatory cells obtained by bronchoalveolar lavage (Luzina

*et al.*, 2002). Circulating leukocytes and peripheral blood mononuclear cells have been used to assess kidney diseases (Alcorta *et al.*, 2002), expression of cytokine- and chemokine-related genes in lupus patients (Rus *et al.*, 2002), and gene expression profiles of mononuclear cells in humans after infection with human immunodeficiency virus (HIV) type 1 RF (Vahey *et al.*, 2002). In some clinical settings, tissue may be readily available – for example, gene expression markers that have been measured in subjects with endometriosis (Eyster *et al.*, 2002), from osteoarthritic cartilage (Aihara *et al.*, 2002), and from brain tissue with gene expression patterns in schizophrenia (Mimmack *et al.*, 2002).

The tissue itself is a second problem in determining the value of gene array data. Processing tissue rapidly to maintain RNA integrity is crucial. Artificial gene array data are generated from degraded mRNA. Therefore, having access to a competent tissue bank linked to searchable databases that contain the clinical, biological, and biochemical characteristics of the sample is key to obtaining meaningful diagnostic interpretations from integration of these clinical correlates and reliable gene array data.

Most tissue samples obtained from humans are a mixture of different cell types. For example, a muscle biopsy sample taken from the vastus lateralis muscle will contain not only skeletal muscle but also blood vessels, connective tissue, nerve tissue, and stromal cells. Therefore, changes in gene expression patterns, when comparing two different muscle biopsy samples, are a reflection of all the cell types present in that sample. Many claim that this can confound the analysis and limit applicability of results. Methods such as laser capture microdissection that allow for isolation of individual cells still are limited technologically (Simone *et al.*, 1998; Brail *et al.*, 1999; Luo *et al.*, 1999; Best and Emmert-Buck, 2001). However, many others argue that all the cell types influence the function of the tissue in question and the gene expression pattern as a composite of the whole is more meaningful than any one isolated cell type. Clearly, this is an area of active debate. Yet, virtually everyone agrees that altered signaling and interactions between cell types are informative and diagnostic of a specific disease. In fact, this has proven true in profiling (Eisen and Brown, 1999; Young, 2000; Ramaswamy and Golub, 2002). The significance of this discussion can be determined only through comparing array data between individual cell types and the tissue as a whole (Alizadeh *et al.*, 2001).

### III. Type of Array

Another important decision that a clinical investigator must make is choosing the type of array to use with the clinical sample (Figure 2). The technique is based upon the same principle as Northern and southern blotting, in which a labeled (radioactive or fluorescent) complementary DNA (cDNA) probe has been created from reverse transcription of mRNA and hybridized to cDNA or short

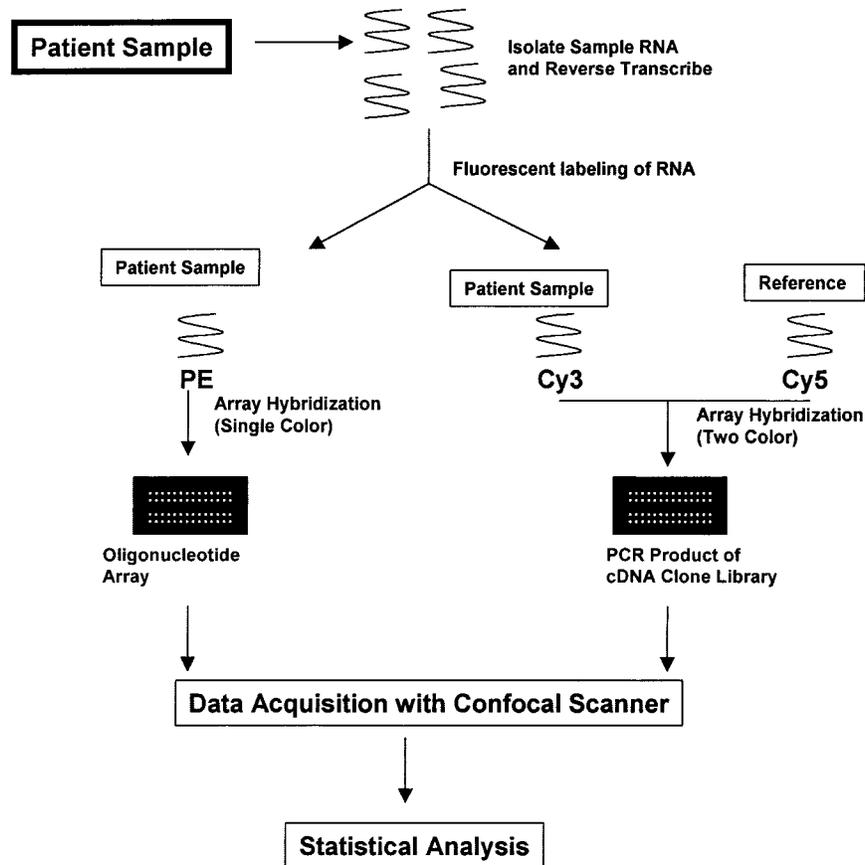


FIG. 2. Oligonucleotide and cDNA microarrays. Oligonucleotide microarrays utilize a direct synthesis of oligonucleotides (25–60 mer) onto solid surface with a single-color readout of gene expression from a patient sample. A cDNA microarray utilizes polymerase chain reaction (PCR)-generated products (1–2 Kb) from cDNA libraries onto a solid surface with simultaneous, two-color readout (Cy3 and Cy5) of gene expression from patient sample and a reference standard sample.

oligonucleotides representative of unique gene sequences immobilized on either a nylon membrane or glass slide (in excess of mRNA). Early on, only a few arrays were available. However, as the field has advanced, more-specific arrays that contain focused genes are being developed. A recent study identified Apo L proteins as a group that shows increased expression in the brain tissue of humans with schizophrenia (Mimmack *et al.*, 2002). This finding was made using a custom candidate gene cDNA array comprising 300 genes that were implicated in schizophrenia (Mimmack *et al.*, 2002). On the other end of the scale from

custom cDNA arrays are commercial oligonucleotide arrays such as the Human Genome U133 GeneChip<sup>®</sup> produced by Affymetrix (Santa Clara, CA) that contains 39,000 of the best-characterized human genes (Gershon, 2002). Clinical investigators must decide between a specialized array and a more-generalized one that would be more inclusive but also generate many more data for interpretation. Cost is another consideration. A number of excellent reviews describe design and benefits of various arrays (Alizadeh *et al.*, 2001; Bertucci *et al.*, 2001; Arcellana-Panlilio and Robbins, 2002; Grant *et al.*, 2002).

#### IV. Gene Array Data Analysis

As detailed above, many methods are available to allow investigators to begin to interpret the significance of the array data. Moreover, investigators must validate key gene expression patterns using methods such as Northern blotting, quantitative reverse transcription-polymerase chain reaction (QRT-PCR), or RNase protection assays. However, most clinical investigators must collaborate with 1) a basic scientist, to generate the array data, and 2) bioinformatics experts, to analyze the array results and indicate significant gene expression patterns. As the field continues to develop, this process will become more automated and, with this progression, clinical investigators without access to the multiple groups of investigators found at major medical institutions will begin to use array technology. An example of this continuing automation of a DNA analysis device that will both synthesize oligonucleotide probes and perform hybridization is being tested by Geniom, a German biotechnology company (Gershon, 2002).

A detailed analysis of differential expression, including clustering and profiling, requires a rich data field. Therefore, experiments should be set up to construct a data matrix or “dataframe” consisting of a row for each gene and a column for each chip. The first column contains the identification of the gene in each row. To achieve the best results when clustering, it is useful to have at least four columns of array data. Because very large volumes of data are generated by microarray experiments, it is important to select for analysis only those genes that appear to show differential expression due to the experimental conditions. If this is not done, the large number of genes whose expression was either not changing or was due to random fluctuations are likely to wash out many important experimental effects. To filter out these genes, a strategy should select only those genes fit for further analysis. Good experimental design makes this easier to do in a systematic manner (Tian *et al.*, 2002). Typically, we set up our experiments to be suitable for either a one-way or two-way analysis of variance (ANOVA) filter (Figure 3). Factorial designs are very useful for microarray experiments and often fit the experimental situation very well. Note that a  $2 \times 2$  factorial design can be satisfied with as few as four GeneChips and is ideal for a two-way ANOVA (though at least three replicates are preferred for statistical reasons) and

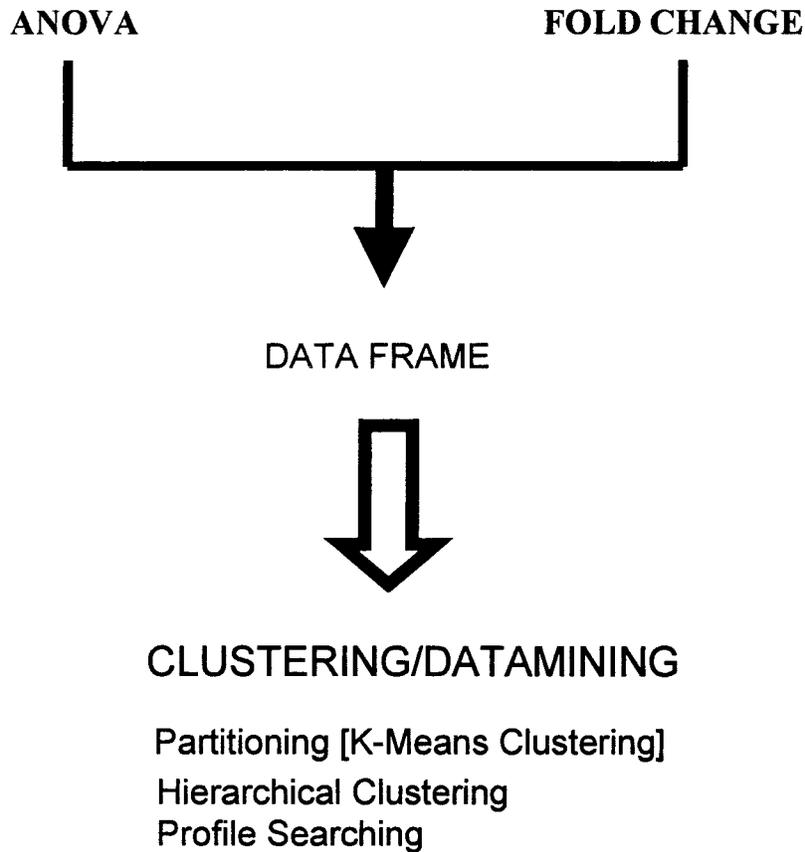


FIG. 3. Diagram demonstrating the flow and options in gene array data analysis. The initial filters to identify genes that are statistically different between comparison groups include analysis of variance (ANOVA) and fold change (by cutoff), thus creating a dataframe or base for further cluster analysis techniques.

construction of a good dataframe. For Affymetrix arrays, we first filter out genes that the initial analysis rates as “Absent” in each of the GeneChips. These are discarded as uninteresting (i.e., unresponsive) to this set of experiments. Next, we perform the ANOVA separately for each gene, keeping only those that show a probability that the F-ratio ( $\text{Pr}(F)$ ) is significant at some level of confidence such as 95% (i.e., having a  $\text{Pr}(F)$  value  $\leq 0.05$ ). This means that the differential expression for that gene is likely due to the experimental conditions rather than to random fluctuations at that level of confidence. Genes that don’t meet this criterion are discarded from further analysis. The remaining genes are most likely

to demonstrate responses that correlate in some manner with the experimental conditions and are most useful for further discovery. This method avoids the problems of multiple t-tests and is statistically more satisfying than simply requiring  $\geq 3$ -fold change as a cutoff. If no other method is available or suitable due to the experimental design, however, fold-change or log ratio cutoffs can be used to effectively filter out genes that appear to be unresponsive to the experimental conditions.

A primary goal of microarray analysis is discovery of hidden patterns of differential expression within the data field. Clustering methods especially are the tools of choice (Figure 3). This is usually best done iteratively using a partitioning method (e.g., k-means clustering, partitioning around medoids (PAM)). This will break the initial large group of genes into smaller subclusters, based on the similarity of their patterns of differential expression across all the experiments (GeneChips) simultaneously. Thus, a type of pattern-recognition algorithm groups genes because their expression is similar. Each gene within a subcluster will have enhanced or depressed expression in unison, when plotted across the different experimental conditions. It will be useful to give some of the details about how this is done because the underlying methodology is common to virtually all clustering techniques and it helps to understand why good experimental design resulting in a proper dataframe is usually the best approach.

A microarray dataframe can be described as an  $n$ -by- $p$  matrix containing  $n$  rows corresponding to the “objects” (i.e., genes, probe sets) on the microarray and  $p$  columns, each corresponding to a separate microarray (or the average of a set of replicates) corresponding to a different experimental time or condition (e.g., tissue sample, temperature, dosage). Thus, each gene/probe set is identified with a row,  $\mathbf{n}_i$ , which is a vector of order  $p$  where each point,  $x_{ij}$ , describes the fluorescent intensity of that probe set or gene in microchip experiment  $p_j$ . Associated with the  $n$ -by- $p$  matrix of experimental measurements is an  $n$ -by- $n$  table that is a collection of proximities describing the comparisons of all possible pairs of objects (i.e., genes). For the purposes of clustering microarray data, these proximities most often describe the dissimilarities between the differential expression patterns of two genes (or its conjugate, similarity) or covariance. Most simply, dissimilarity  $\mathbf{d}_{ij}$  can be explained as the Euclidean distance (i.e., crossproduct) between two vectors in a data set, each representing the differential expression of a particular gene (e.g., any two of the  $\mathbf{n}_i$ ). Using the  $\mathbf{d}_{ij}$  table, usually called a dissimilarity matrix, as input, a variety of clustering methods can be used to identify those objects (genes or probe sets) that behave most alike across a given set of experiments. Two of the more-common methods for doing this are k-means clustering and hierarchical clustering.

The k-means clustering continues iteratively, partitioning the genes into a growing number of smaller and smaller subclusters, until the pattern of expression for all members of the subcluster are not significantly dissimilar. Once the

subclusters are identified, they can be clustered hierarchically to show the juxtaposition of each gene within the subcluster, based upon the similarity of their differential expression patterns. This is shown graphically as a dendrogram (similar to a family tree). Hierarchical clustering techniques can be either agglomerative or divisive, depending upon whether they start with each of the member genes as an individual, then group them together into families, or whether they start with one large family and divide it up progressively into smaller and smaller subfamilies, until each gene is a separate branch. Usually, there are small qualitative differences between the results of the two methods but occasionally larger differences show up that need to be reconciled. We typically use an agglomerative nesting technique called AGNES. An often-insightful use of hierarchical clustering is to cluster the transpose of the dataframe associated with a specific subcluster. This shows the interrelationships between the columns (GeneChips) of the dataframe rather than the rows (genes). From this, we can see how genes within the subcluster differentiate the experiments. It is becoming increasingly common to perform hierarchical clustering for both the rows and columns and to show their dendrograms on the same graphic aligned along the top and side of a “heat map” (Figures 4–7). A heat map is a graphical matrix where each cell corresponds to the signal intensity of a specific gene in a specific experiment. The rows and columns of a heat map are arranged to show simultaneously the interrelationship between the different experiments and the genes within the subcluster. The color of each cell is significant and is selected from a gradient of colors (typically, red to green), where the shade of the color is proportional to the signal intensity (or log ratio) of that gene in that experiment. These values often are normalized or scaled to z-scores for best effect. Normally, we represent high values as shades of red, intermediate values as shades of gray to black, and low values as shades of green. (A few authors do it the opposite way, so be sure to check the legend when reading articles containing heat maps.) Due to color limitations in this review, we have used white to represent high values, black to represent low values, and shades of gray to represent gradations of gene expression levels between high and low.

Another, oftentimes very useful way to generate the information for a heat map is called profiling. This method is used in Example 1 (see Section VA below) to identify genes commonly upregulated and downregulated by two thiazolidinedione compounds and genes uniquely regulated by these compounds (Figures 4–6). In this method, one constructs a model profile of how one would expect some genes to behave across the columns of a dataframe to demonstrate a pattern of meaningful biological significance. Each of the rows (genes) of the dataframe is ranked according to how similar its expression profile is to the model profile, then assigned a similarity coefficient between +1 and -1. A coefficient of +1 represents a perfect correlation, while -1 represents perfect anticorrelation. Zero means that there is no correlation. In practice, one chooses

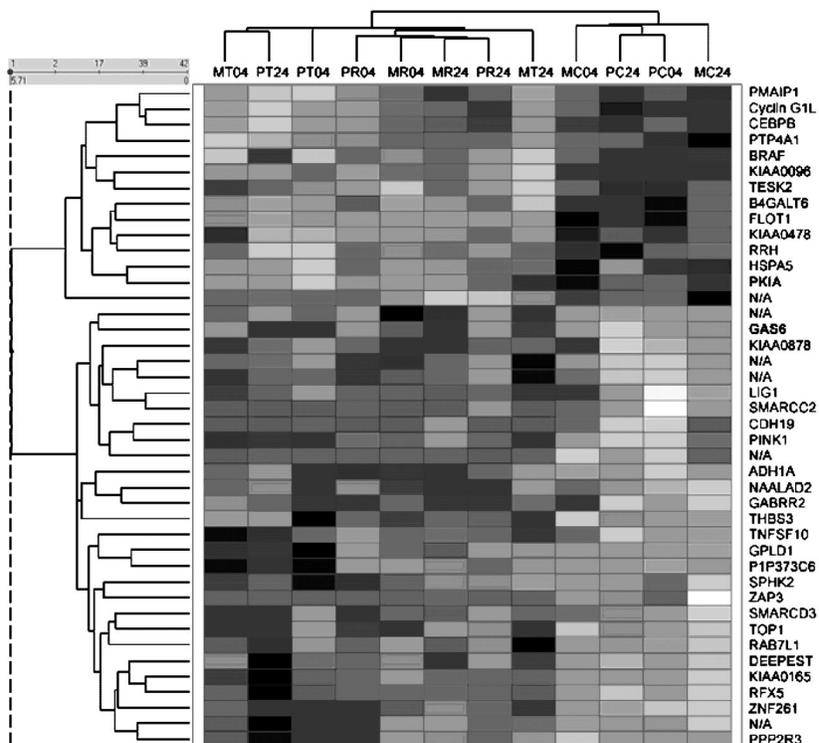


FIG. 4. Heat map of genes identified as commonly regulated by troglitazone (TRO) and rosiglitazone (ROSI) in Mia Paca-1 and Panc-1 cells. Cells were treated for either 4 or 24 hours with 20  $\mu$ M TRO or ROSI. White bars indicate high gene expression, while black bars indicate low gene expression. Shades of gray represent variations between high and low gene expression levels. Abbreviations: Panc-1 (P); Mia Paca-2 (M); TRO (T); ROSI (R); 4-hour treatment (04); and 24-hour treatment (24). Therefore, 24-hour TRO treated Panc-1 cells is represented as "PT24." N/A represents unknown genes.

genes that are highly correlated (and usually, highly anticorrelated) to the model profile, then constructs a heat map showing their hierarchical distribution among the genes and across the experiments. The choice of a suitable cutoff for the similarity coefficient is somewhat arbitrary but, similar to an  $R^2$  correlation coefficient in regression analyses, absolute values below 0.70 often will not show good correlation with the model profile. This is something to trial and error on data using a spreadsheet and charting program.

Once a subset of genes that service a working hypothesis has been selected, it is very instructive to search out their functions (using PubMed, gene ontology entries (GO), etc.) and to characterize the subcluster for the similarity or

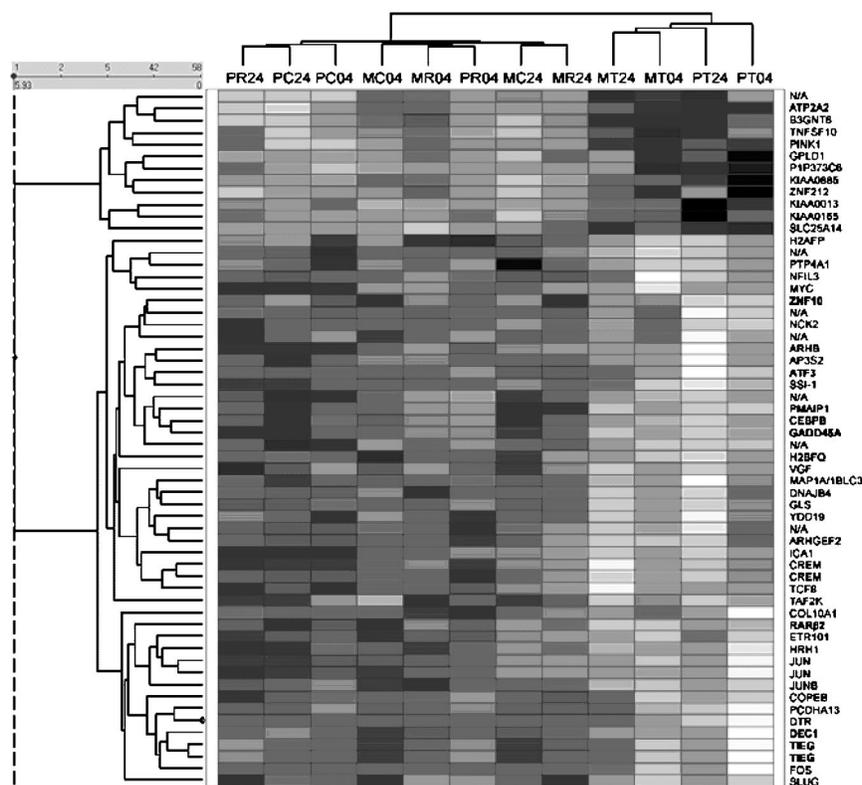


FIG. 5. Heat map of genes uniquely regulated by TRO in Mia Paca-1 and Panc-1 pancreatic cells. [See Figure 4 for detailed descriptions.]

juxtaposition by their function or pathways. This can be an especially enlightening part of the discovery process.

## V. Practical Examples

### A. EXAMPLE 1. DISTINGUISHING MOLECULAR ACTIONS OF DRUGS OF THE SAME CLASS, THEN LINKING THESE ACTIONS TO BIOLOGICAL ACTIVITY

Thiazolidinediones (Tzds) belong to a class of compounds that are ligand agonists for peroxisome proliferator-activated receptor gamma (PPAR $\gamma$ ), a nuclear transcriptional factor. Tzds are clinically used in the treatment of type 2 diabetes as insulin sensitizers (Horikoshi and Yoshioka, 1998; Willson *et al.*,

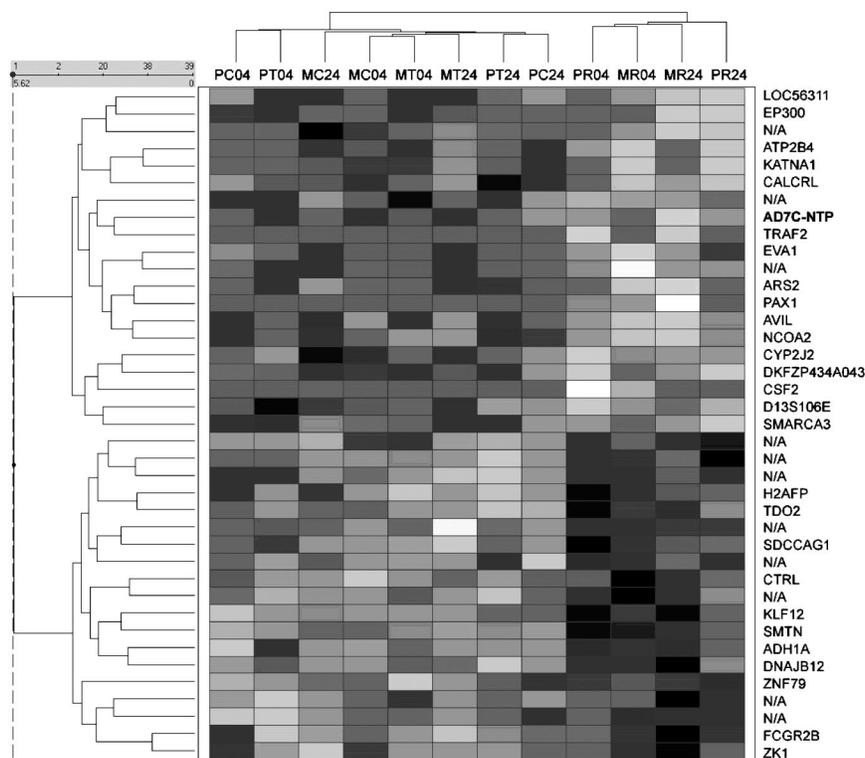


FIG. 6. Heat map of genes uniquely regulated by ROSI in Mia Paca-1 and Panc-1 pancreatic cells. [See Figure 4 for detailed descriptions.]

2001). However, these compounds have different profiles with regard to side effects (Fujiwara and Horikoshi, 2002). PPAR $\gamma$  agonists are known to control adipocyte differentiation pathways via activation and suppression of key regulatory molecules determining adipocyte phenotype (Kliewer *et al.*, 2002; Walczak and Tontonoz, 2002). Tzds also recently have been described to have antiproliferative as well as chemopreventive activity against human tumors (Debril *et al.*, 2001; Sporn *et al.*, 2001). We have found that troglitazone (TRO) and rosiglitazone (ROSI) have dissimilar profiles with regard to growth inhibitory profiles and induction of apoptosis in two human pancreatic tumor cell lines, Mia Paca-1 and Panc-1 (Cowey *et al.*, 2001). Thus, we were interested in identifying and determining whether genes that control cell proliferation and cell death were regulated by these compounds. Cells were treated in culture with 20  $\mu$ M of either TRO or ROSI for 4 and 24 hours, followed by RNA isolation and analysis of gene expression by Affymetrix analysis. The heat maps shown in

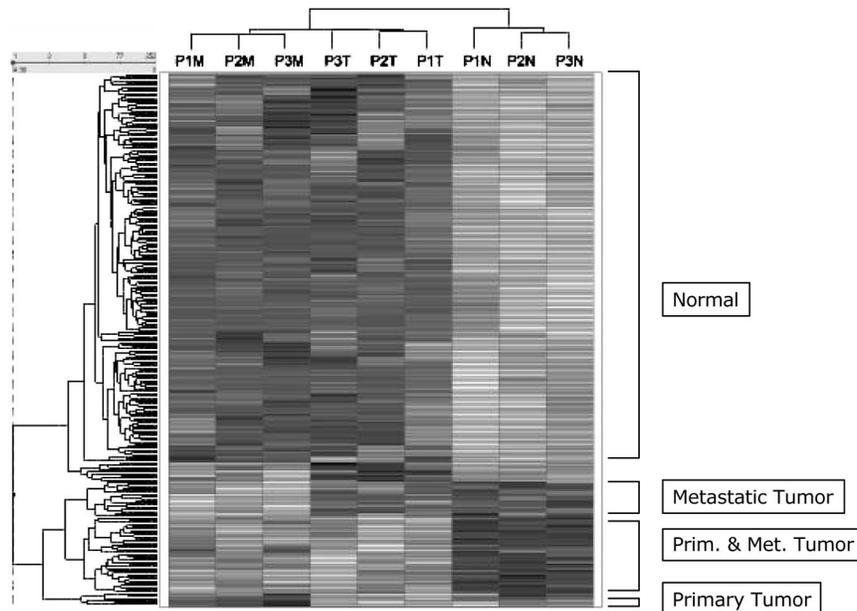


FIG. 7. Identifying genes related to tumorigenesis in patients with renal cell carcinoma (RCC). RNA was isolated from tissue of normal kidney, primary tumor, and metastasis of three patients undergoing radical nephrectomy for the treatment of RCC. Abbreviations: P1, P2, P3 = patient 1, 2, or 3; N = normal ; T = primary tumor; M = metastasis.

Figures 4–6 highlight the ability of TRO and ROSI to identify commonly and divergently regulated genes. Of the 12,558 probe sets on the human Affymetrix gene chip, 8249 genes were marked as “Present” (e.g., expressed). A pairwise comparison of control to treatment group using a 3-fold change cutoff retained 4158 probe sets. Using idealized profiles, we performed profile comparisons utilizing Spotfire DecisionSite 7.0 to identify correlates (upregulated genes) and anticorrelates (downregulated genes) with a similarity to the idealized profile of 0.7. The three profile searches performed are reflected in Figure 4 (genes commonly regulated by TRO and ROSI and different from their appropriate control), Figure 5 (genes altered uniquely by TRO), and Figure 6 (genes uniquely altered by ROSI). Thus, we were able to identify 42 genes commonly regulated by TRO and ROSI, 58 genes altered uniquely by TRO, and 39 genes altered by ROSI treatment at 4 and 24 hours.

The next challenge is to determine which of these genes are known to be regulated by Tzds. Bioinformatics programs written to query PubMed, Medline, and other databases can rapidly make the connection by probing the gene set identified against key words such as thiazolidinedione and PPAR $\gamma$ . Another

challenge is to identify functions of genes in a data set and link them to biochemical pathways that can be interpreted in a meaningful manner related to the demonstrated action of the compound: in our case, inhibition of cell proliferation (TRO and ROSI) and apoptosis (TRO). We have written a program that links gene name, GenBank number, chromosome location, and functions (biochemical, biological, organismal, and molecular; [www.bioinfo.utmb.edu](http://www.bioinfo.utmb.edu)). Table I shows genes that play a role in inhibiting cell proliferation and stimu-

TABLE I  
*Thiazolidinedione-regulated Genes That Control Cell Proliferation and Apoptosis  
Identified Using Gene Arrays*

Function	Gene	Troglitazone				Rosiglitazone			
		M04	P04	M24	P24	M04	P04	M24	P24
Apoptosis	Retinoic acid receptor alpha	++	++	++	++	+	+	0	-
Apoptosis	Gadd45	++	++	++	++	+	+	0	0
Cell cycle arrest	Protein kinase A inhibitor	++	+	-	+	++	0	++	0
Antiapoptosis	Gas6	0	-	-	--	-	-	--	-
Apoptosis	Transforming growth factor beta-inducible early protein	++	++	0	+	0	+	0	+
Tumor suppressor	DEC1	++	++	+	+	0	0	0	-
Apoptosis	AD7c-neuronal thread protein	+	-	+	-	++	++	++	0
Transcriptional repressor	Activating transcription factor-3	++	++	++	++	+	+	0	+
Transcriptional repressor	Zinc nuclear finger 10	++	+	++	+	+	+	-	-

[Genes were identified by cluster analysis or cutoff techniques described in the text for Mia Paca-1 (M) and Panc-1 (P) cells treated with TRO, ROSI, or vehicle control for 4 (04) or 24 hours (24). A program linked to PubMed/Gene Ontology was used to identify genes in this subset that regulate apoptosis and cell proliferation. Visualization of shades of gray from the heatmaps of Figures 4–6 correlates with expression levels of these identified genes. ++ = highly expressed (+ = increased expression), compared to matched untreated control, and correlates with white on heat map. 0 = no difference from control. - = a decrease in gene expression of treated group, compared to its matched control value. M = Mia Paca-1 cells; P = Panc-1 cells; 04 = 4 hours of treatment; 24 = 24 hours of treatment.]

lating apoptosis. For the first time, we demonstrated that retinoic acid receptor alpha ( $RAR\alpha$ ) is regulated by Tzds. Also not previously identified to be regulated by Tzds are genes that are transcriptionally regulated by  $TGF\beta$  and p53 (Conner *et al.*, 1999; Kimura *et al.*, 2001; McDonald and El-Deiry, 2001; Kondo *et al.*, 2002; Yun *et al.*, 2002; Zawel *et al.*, 2002). Incidentally, Panc-1 and Mia Paca-1 cells express nonfunctional mutant forms of p53. Thus, we can identify multiple genes in a specific signaling pathway. Clearly, there is overlapping yet distinct regulation of these genes by TRO and ROSI. Thus, one can test the role of each of the newly identified genes by either overexpressing the respective gene in cells or by selectively blocking expression of each gene (antisense RNA, small interfering RNA (siRNA), or dominant-negative expression constructs). Yet another challenge is to identify unknown genes or expressed sequence tags (ESTs). Multiple strategies can be taken in this respect using data search bases and programs. Using the Genbank number for the unknown gene, a DNA sequence can be copied and pasted into programs (e.g., [www.ncbi.nlm.nih.gov/genome/seq/HsBlast.html](http://www.ncbi.nlm.nih.gov/genome/seq/HsBlast.html)) that will perform searches for homologies to the unknown gene.

#### B. EXAMPLE 2. IDENTIFYING GENE SIGNATURES RELATED TO TUMORIGENESIS IN PATIENTS AND DEMONSTRATING UNIQUE GENE EXPRESSION PATTERNS IN EACH PATIENT

In Figure 7, a heat map demonstrates the ability to identify differences in gene expression between normal tissue and that of matched primary tumor and metastasis in three patients diagnosed with metastatic clear cell renal cell carcinoma (RCC). Hierarchical clustering was used to determine whether genes were differentially expressed in tumors versus normal kidney tissue. As shown in Figure 7, 253 genes were identified, clearly separating out gene expression of normal tissue of patients 1–3 (see right top of heatmap in Figure 7). Although genes were different in primary tumor and metastasis samples, compared to their respective controls, few genes delineated a difference between the two groups. This had been observed previously by others; it was suggested that the primary tumor already had been programmed to metastasize and that gene expression levels will be similar in the primary tumor and metastasis tissues (Bhattacharjee *et al.*, 2001). However, expression signatures are being used to diagnose cancer subtypes (van't Veer *et al.*, 2002). Golub and colleagues have designed supervised learning algorithms to classify tumor samples into specific diagnostic categories based on their gene expression signatures. They and others have demonstrated successfully that an unknown tumor sample can be classified correctly, based upon its signature (Khan *et al.*, 2001; Ramaswamy *et al.*, 2001; Yeang *et al.*, 2001). Using these strategies, the number of genes needed to diagnose a disease can be narrowed down to only those that act as clear indicators

of that disease state. Thus, the three-patient population sample shown in Figure 7 must be expanded by hundreds of matched normal and tumor samples in order to define the gene signature to appropriately define diagnosis. These data linked to patient outcome will lead to predicting prognoses based upon gene expression patterns. Even this small sample allowed us to identify dysregulation of multiple signaling pathways and multiple genes within the same signal pathway (data not shown). This has been verified by RT-PCR and immunohistochemical analysis in a larger set of matched tissue samples.

Another strategy that can be used to identify gene expression patterns that change in similar fashions is to profile a particular gene, then search the database for genes that have the same or opposite profile (anticorrelate). Again, genes that change in similar or diametrically opposite patterns may interact in a signaling pathway or coregulate a phenotypic change in the cell or tissue (Figure 8). Using the gene expression of pattern for E-cadherin (downregulated 50- to 100-fold in all primary and metastatic tumors, compared to matched normal controls) demonstrates the ability to identify genes in the data set from Figure 7 that are downregulated similarly to E-cadherin or are upregulated (anticorrelate). These strategies should allow one to identify key signaling pathways and crosstalk between pathways in regulating proliferation, differentiation, and cell death (Figure 9). Filling in the pieces to this puzzle and understanding all the functions of each gene product will provide the roadmap for developing effective treatment regimens.

### C. EXAMPLE 3. IDENTIFYING MUSCLE GENES REGULATED BY ANDROGENS IN OLDER ADULTS

Figure 10 displays an analysis of muscle biopsy samples taken from older men given testosterone injections for 6 months (Ferrando *et al.*, 2002). This study found that administration to older men to increase their testosterone concentrations to those of younger men increased muscle strength, lean body mass, and the net balance of muscle metabolism (Ferrando *et al.*, 2002). To further investigate the cause of the increase in strength, Atlas™ cDNA Expression Arrays made by CLONTECH Laboratories, Inc. (Palo Alto, CA) were hybridized with the tissue samples. The CLONTECH DNA array chip can determine the expression of gene targets from five major subgroupings. However, for this analysis, insufficient muscle biopsy tissue was available to analyze each sample. Therefore, for the baseline, 1-month, and 6-month time points, total RNA from the seven subjects that received testosterone for 6 months was pooled for the hybridization. Of the 2300 genes screened, significant changes were detected in 230. One group of genes that responded to testosterone was mitochondrial proteins involved in the oxidative phosphorylation pathway. Two of the protein complexes (IV and V) showed a response of the nuclear-encoded mitochondrial proteins. From these

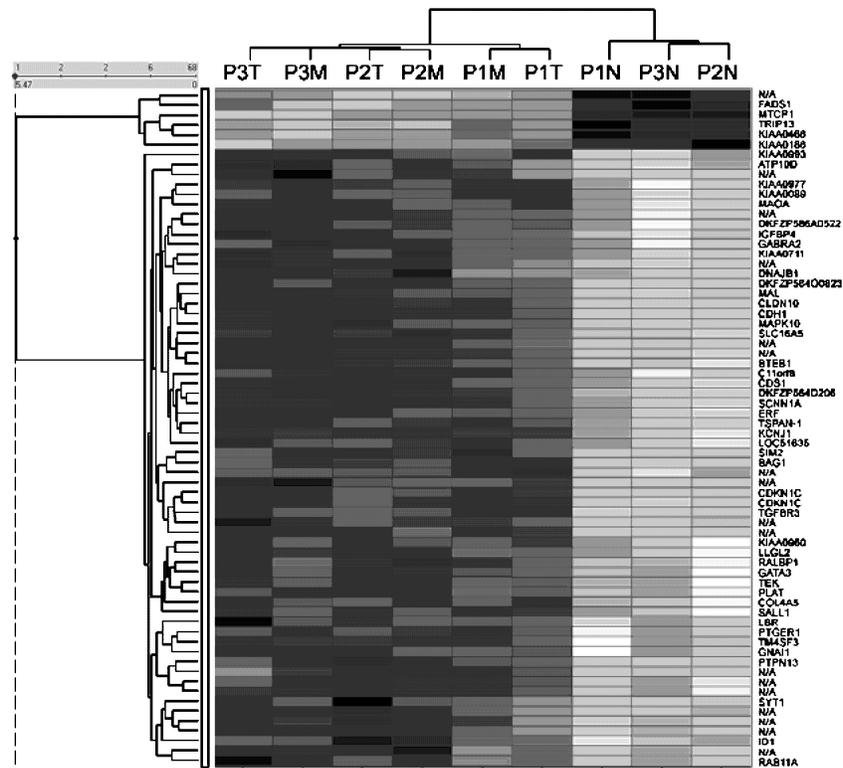


FIG. 8. Profiling genes with similar or opposite expression patterns. E-cadherin is downregulated 50- to 100-fold in all primary and metastatic tumors, compared to matched control normal renal tissue. Using software from Spotfire 7.0, genes with the same expression patterns and genes with the opposite (anticorrelate) patterns to that of E-cadherin were identified using the genes identified in Figure 7. Abbreviations are the same as shown in Figure 7. N/A represents unknown genes.

data, three representative genes were selected for verification. Western blot analysis was performed on each of the seven subjects individually, with antibodies to the genes being verified. As shown in Figure 10, an increase in protein expression occurred at 1 month that returned to pretreatment levels by 6 months. Cytochrome c oxidase Vb is one of the smaller subunits of the cytochrome c oxidase complex that is nuclear encoded and is one of the key regulatory subunits of the complex (Basu *et al.*, 1997; Lenka *et al.*, 1998). Adenosine triphosphate (ATP) synthase  $\gamma$  forms the central core unit of the ATP synthase complex around which other subunits orient themselves (Konno *et al.*, 2000; Tsunoda *et al.*, 2001), including ATP synthase  $\alpha$ . These mitochondrial oxidative phos-

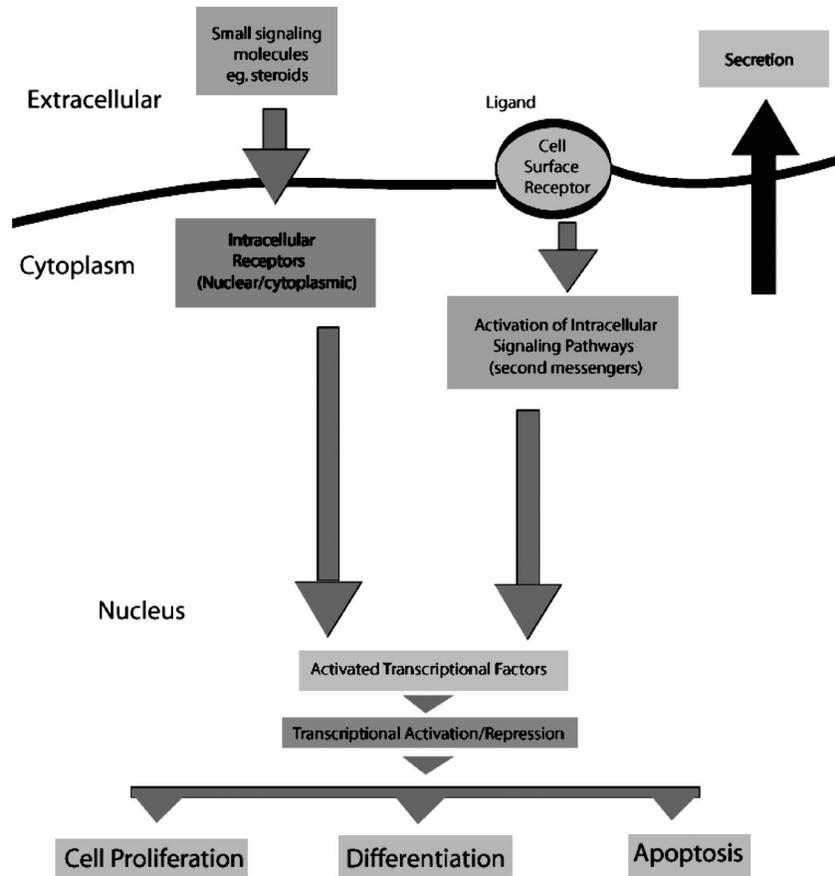


FIG. 9. Diagram illustrating signaling pathways in the regulation of cell proliferation, differentiation, apoptosis, adhesion, and secretion. Multiple signaling pathways regulate these phenotypic and ever-changing processes directly and indirectly via crosstalk with other signaling cascades.

phorylation proteins are nuclear encoded, indicating that they are responding to the administration of testosterone in these older men. This pattern of expression, which follows that of the androgen receptor (Ferrando *et al.*, 2002), indicates that further studies are needed to assess the effects of cycling testosterone on muscle mass and strength.

## VI. Verification of Gene Array Analysis

DNA microarray technology represents a powerful tool for expression profiling and the scanning of very large numbers of genes for potential differ-

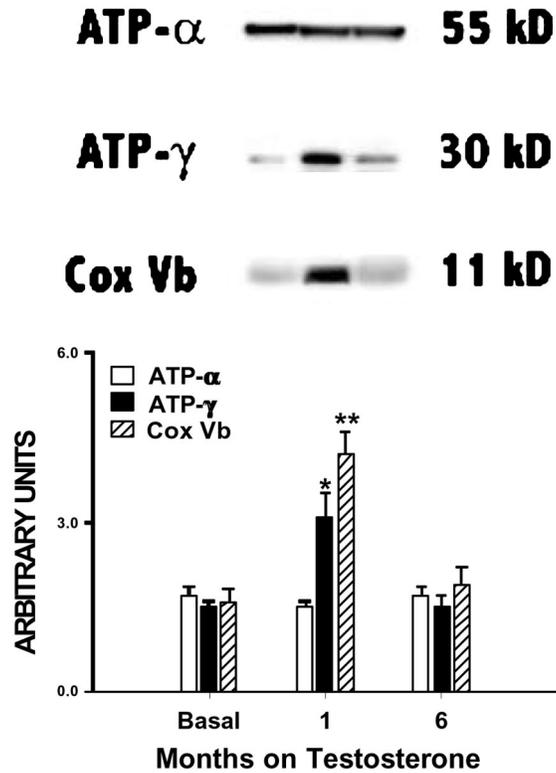


FIG. 10. Western analysis of mitochondrial proteins during testosterone administration. Three nuclear-encoded mitochondrial proteins in the oxidative phosphorylation pathway were analyzed by western blots. The top panel is a representative subject and the graph below represents the mean  $\pm$  standard error from the seven subjects. Actin (not shown) was used to standardize the blots. ATP, adenosine triphosphate; Cox Vb, cytochrome c oxidase Vb. \*,  $p \leq 0.05$ , as determined by ANOVA.

ences in gene expression. However, data generated by microarray-based studies have inherent limitations that complicate its use as a stand-alone technology. The problems with microarray-based studies stem largely from their dependence on differential hybridizations to detect differences in transcript abundance. Many factors may affect nucleic acid hybridization reactions. The standardized hybridization conditions of array experiments have been developed to minimize, on average, the impact of these factors. However, for any individual transcript, the hybridization conditions may not be optimal and apparent differences in gene expression may, on occasion, be biased by these effects. In some instances, it may be possible to use sufficient replications, coupled with statistical tools, to

demonstrate confidence in the results obtained (Lee *et al.*, 2000; Tusher *et al.*, 2001). More commonly, however, investigators resort to an independent measurement of transcript abundance (e.g., Northern blot, quantitative RT-PCR) to confirm their results. “Real-time” quantitative RT-PCR (q-PCR) has proven to be an invaluable tool not only for *confirming* microarray-based results but also for *extending* them into a more-general context.

#### A. REAL-TIME QUANTITATIVE RT-PCR

q-PCR is an ingenious technique that allows for the continuous measurement of products generated during the course of a multicycle PCR reaction (Gibson *et al.*, 1996; Heid *et al.*, 1996). A variety of experimental strategies are available that use changes in fluorescence emission intensities to monitor the progress of PCR reactions. The Taqman procedure, one of the earliest real-time q-PCR techniques developed, is based upon measuring the increase in fluorescence that results from the template-dependent hydrolysis of a fluorescent hybridization probe. This fluorescent hybridization probe is a sequence-specific oligonucleotide that is complementary to the amplicon being assayed. The probe contains two fluorescent dyes covalently attached to specific nucleotides such that fluorescence emission of the dye being excited is quenched by intramolecular fluorescence resonance energy transfer (FRET). During each PCR cycle, Taq polymerase hydrolyzes the hybridization probe molecules that are bound to the template, releasing the fluorescent dyes and dequenching the fluorescence. The increase in fluorescence is directly proportional to the number of probe molecules hydrolyzed that, in turn, under appropriate conditions, are directly proportional to the number of template molecules in the PCR reaction tube in that particular PCR cycle. During the geometric amplification phase of the PCR reaction, the amount of template generated is geometrically related to the number of template molecules present at the start of the reaction and the number of PCR cycles completed. Thus, by continuously monitoring the level of fluorescence in the PCR reaction in real-time PCR instrumentation, it is possible to accurately quantitate the number of templates present in the unknown sample at the beginning of the PCR reaction.

#### B. THE VALUE OF COMBINING q-PCR WITH MICROARRAY-BASED STUDIES

The value of real-time q-PCR as a technique for confirming microarray results stems from its ability to provide an accurate measure of the abundance of specific transcripts in an RNA preparation. Under most circumstances, microarray studies provide a semiquantitative estimate of transcript abundance and differential gene expression. Once a transcript of interest has been identified, it is relatively straightforward to design a specific real-time q-PCR assay for the

transcript, then use the assay to accurately measure the level of the transcript in the two RNA samples being compared. Not only will the technique allow for confirmation of an apparent differential level of expression but it also provides an accurate measure of the *degree* of differential expression. This is particularly important and useful in the analysis of transcripts that show modest levels of induction in the microarray studies. Transcripts that show large (i.e., > 3-fold) and consistent changes in expression in microarray studies generally are readily confirmed by an independent technique such as q-PCR, Northern blots, or RNase protection assays. However, many of the most important changes in gene expression, particularly in *in vivo* and clinical studies, are represented by more-moderate changes (i.e.,  $\geq 1.5$ -fold) in groups of related genes such as those comprising key metabolic pathways. These levels of change often are difficult to distinguish from noise in the microarray studies but can be readily evaluated using q-PCR. The precision and reproducibility of q-PCR assays allow investigators to apply relevant statistical tests to the data to confirm the differential expression of genes with modest inductions. The combination of microarray and q-PCR measurements allows investigators to work with confidence in the “gray” zones of 1.5- to 3-fold changes, where most of the important biology is occurring (Singh and Liu, 2001).

A particular value of real-time q-PCR assays is that they are readily adapted to the analysis of multiple samples and can be run in a high-throughput mode. Since the progress of the PCR reaction is monitored optically, a variety of instruments have been developed that allow for simultaneous quantitation of transcripts in either 96-well or 384-well format. Thus, in one experiment, even allowing for replicates and controls, it is possible to quantitate transcripts for large numbers of samples (i.e., 20–200). Given that the usual run time for a real-time q-PCR reaction is less than 2 hours, this capability means that it is possible to quantitate many transcripts in many samples in a relatively brief period of time. This high-throughput capability is particularly useful in clinical studies. Very often, it is not logistically feasible to run microarrays on large numbers of individual patient samples. The clinical investigator may be restricted to running only a few samples, due to either limited availability of RNA requiring pooling of specimens or the expense of running large numbers of individual chips. To “validate” the results of the microarray study, it often is desirable to extend the analysis to a much larger pool of samples or subjects. We have found that real-time q-PCR analysis of a much larger series of subjects than could be included in the microarray study leads to a much more accurate and useful estimation of the extent to which the changes detected on the microarray can be applied to the patient population as a whole.

In addition to its suitability for high-throughput analyses, real-time q-PCR assays are particularly useful adjuncts to microarray studies. Their extreme sensitivity permits analysis of transcripts in very small amounts of input total

RNA. Under normal conditions, it is easy to develop assays with a lower limit of detection of  $10^2$ – $10^3$  transcript molecules. This level of sensitivity permits detection of even low-abundance transcripts in nanogram quantities of total RNA. Since most microarray techniques perform best utilizing micrograms of RNA, it is often convenient to carry out the initial microarray experiments on pools of patient-derived samples, then switch to the much more-sensitive PCR-based technologies to confirm the array results in the panel of individual samples that contributed to the pool. In this way, it is possible not only to “confirm” the array result but also to acquire quantitative information on the distribution of differential gene expression in the patient population of interest.

A further useful aspect of the real-time q-PCR technique is that it does not require intact RNA to provide meaningful data on transcript abundance. The quality of data recovered from most microarray-based procedures depends heavily on the quality of the RNA used to generate cDNAs. On the other hand, it is possible to develop real-time q-PCR assays based on very short amplicons (60–80 nt). These amplicons will remain intact in RNA preparations that have been subjected to extensive degradation. This is particularly useful in a clinical context, where it is often difficult to control for the handling of biological specimens at the time of collection. This feature of real-time q-PCR assays can be applied to RNA recovered from formalin-fixed tissue blocks (Uray and Connelly, 2001). RNA fragments (usually several hundred nucleotides in length) can be extracted from formalin-fixed, paraffin-embedded tissue blocks such as those routinely maintained in pathology archives. This capability can provide a powerful complement to microarray-based studies. For instance, we have used RNAs derived from a limited number of surgical specimens with sufficient material to permit microarray-based analysis. We have then used real-time q-PCR assays to measure the expression of the transcripts identified as of potential interest by the microarray studies, in a much larger series of cases for which archival formalin-fixed specimens exist. An additional benefit of this approach is that it can be combined with laser capture microdissection techniques that allow for recovery of RNA from specific cellular subsets of diseased tissues. In this way, it is possible to extend the results of the microarray studies to a detailed analysis of the pattern of gene expression in large numbers of well-documented clinical cases.

In summary, we have found real-time q-PCR to be one of the most-useful approaches to first confirm and then extend the results obtained from microarray-based analyses.

### C. APPLICATION PROCEDURES

1. Real-time q-PCR assays can be designed using target gene sequences accessible from genomic databases such as GenBank. Although a number of

primer design algorithms can be used for this purpose, we routinely employ Primer Express (Applied Biosystems), since it allows for the simultaneous design of both the PCR primers and fluorescent Taqman probe. We generally design our amplicons to be  $< 100$  nt in length (usually, 60–80 nt). The PCR primers, the fluorescent probe (usually with a fluoroscein amidite (FAM) reporter dye and either tetramethyl rhodamine (TAMRA) or a “black hole” quencher dye) and a single-stranded sDNA of the amplicon (a long oligonucleotide for use as a standard) are ordered from one of several commercial vendors.

2. Assay conditions are optimized with the sDNA amplicon standards by adjusting primer and  $Mg^{+2}$  concentrations to generate assays with a slope of the standard curve (Ct versus log template molecules) of  $-3.2$  to  $-3.5$  and a lower limit of detection of  $10^2$  amplicon molecules.

3. Total RNA samples are assayed after DNase I pretreatment. The range of RNA concentrations will vary based on the amount of material available and the anticipated transcript abundance. We routinely use 10–100 ng of total RNA per determination. Each RNA sample is assayed in triplicate, with a fourth aliquot that is run without reverse transcriptase (–RT control) to control for signal generated by genomic DNA rather than RNA.

4. Samples are subjected to RT prior to PCR amplification. Although we use the reverse PCR primer for the reverse transcriptase reaction, other investigators report equivalent success with random primed RT reactions.

5. In parallel with the unknown RNA samples, we run a standard curve with known amounts of amplicon ranging from  $10^3$ – $10^7$  molecules. The values of template molecules in the unknown samples are determined by interpolation of the  $C_t$  (PCR cycles to reach an arbitrarily set threshold) of the unknown samples on the amplicon-specific standard curve. The data analysis protocols are embedded in the software included with the commercially available real-time PCR instrumentation.

6. We use robotics to assemble both the RT and PCR reactions, since robotics enhances both the throughput of the assays and the precision of the data that are generated.

## VI. Conclusions

The era is dawning in which integration of large data sets (molecular, biochemical, and histological) create a multidisciplinary approach to accurately determine diagnosis, prognosis, and effective treatment regimens for each patient. Clearly, organizing the large quantity of data generated from gene array profiling of mRNA from biological samples linked to analyses tools will be an effective approach toward these goals. It is expected that an understanding of molecular signaling pathways will result in identifying each and every disease by its unique signature. It should also be expected that, in identifying key molecules

in each signaling pathway, drugs able to regulate biological activity will be identified. Thus, one could envision diagnosis by gene array and individualizing drug treatment for each patient (Figure 11).

The clinician has a critical role in identifying patient samples and creating the link to have these unique biological samples analyzed to create gene signatures. Currently, the application of array technology to the clinical investi-

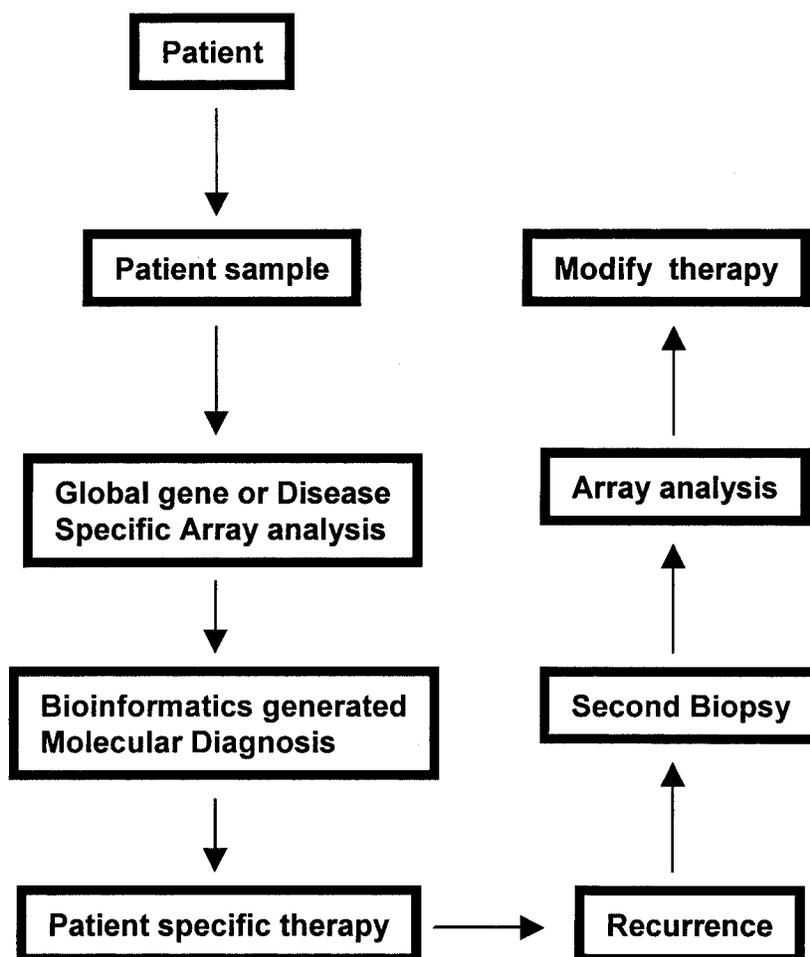


FIG. 11. Idealized patient diagnosis and treatment protocol. This model assumes that each patient has a similar gene profile that identifies a specific disease subtype but also contains a unique genomic profile, compared to others with similar diagnosis. This unique profile would aid in the design of patient-specific therapeutic intervention.

gator is limited by the tissue sample, available arrays, and analysis of data generated from the array. A clinical investigator must have access to a collaborative group of scientists that assist in utilizing the gene array technology. This is a unique time for the clinician to help create diagnoses and prognoses tools for her/his patients and to be involved in bedside to bench to bedside experience. In the near future, advancing automation in the array field may reduce this dependence and broadly expand the use of array technology in clinical investigation. One should expect clinical diagnosis to follow a pattern described in Figure 11.

## VI. Website Resources

### Commercial Gene Arrays

Agilent Technologies: [DNA\\_microarray@agilent.com](mailto:DNA_microarray@agilent.com)

Affymetrix: [www.affymetrix.com](http://www.affymetrix.com)

Clontech: [www.clontech.com](http://www.clontech.com)

Incyte Pharmaceuticals: [www.incyte.com/reagents/catalog/support/](http://www.incyte.com/reagents/catalog/support/)

### Facilities Performing Gene Arrays/RT-PCR/Tissue Arrays/Databases

National Institutes of Health (NIH) genomics:

<http://www.nhgri.nih.gov/DIR/Microarray/main.html>

<http://www.ncbi.nlm.nih.gov/genome/seq/>

Stanford University: <http://cmgm.Stanford.edu/pbrown/mguide>

<http://genome-www4.stanford.edu/MicroArray/SMD/>

Massachusetts Institute of Technology (MIT) cancer genomics website:

[www.genome.wi.mit.edu/MPR/](http://www.genome.wi.mit.edu/MPR/)

University of Texas Houston-Medical School RTPCR/arrays:

<http://girch2.med.uth.tmc.edu/>

University of Texas Medical Branch (UTMB):

Affymetrix gene array facility: [www.scms.utmb.edu/genomics](http://www.scms.utmb.edu/genomics)

Bioinformatics Group: [www.bioinfo.utmb.edu/](http://www.bioinfo.utmb.edu/)

Nature: <http://genetics.nature.com/>

### Analysis Software

ArrayPro (Media Cybernetics): <http://www.mediacy.com/arraypro.htm>

ArrayStat (Imaging Research): <http://www.imagingresearch.com/>

Spotfire: [www.spotfire.com](http://www.spotfire.com)

### Public Sources of Software

University of Texas Houston-Medical School RTPCR/arrays:

<http://girch2.med.uth.tmc.edu/>

University of Texas Medical Branch (UTMB):

Affymetrix gene array facility: <http://www.scms.utmb.edu/genomics>

Bioinformatics Group: [www.bioinfo.utmb.edu/](http://www.bioinfo.utmb.edu/)

Lawrence Berkeley National Laboratory: <http://rana.lbl.gov/>  
 Stanford University: <http://genome-www4.stanford.edu/MicroArray/SMD/restech.html>

*Single Experiment Analysis*

Gene Traffic (Iobion): <http://www.iobion.com/>  
 GeneSpring (Silicon Genetics): <http://www.silicongenetics.com/>  
 Resolver (Rosetta Inpharmatics): <http://www.rii.com/>

ACKNOWLEDGMENTS

A special thank you is extended to Shauna LeGrand for the design of Figure 9 and to Joanna Burris for editorial comments. This work was supported in part by grants from the National Institutes of Health AG/AR-11000 (RJU) and M01-RR-0073 (General Clinical Research Center, University of Texas Medical Branch) as well as John Sealy Memorial Endowment Fund for Biomedical Research at UTMB (JAC) and American Cancer Society Institutional Research Grant IRG-96-152-04 (UTMB).

REFERENCES

- Aihara T, Zien A, Gehrsitz A, Gebhard PM, McKenna L** 2002 Anabolic and catabolic gene expression pattern analysis in normal versus osteoarthritic cartilage using complementary DNA-array technology. *Arthrit Rheum* 44:2777-2789
- Alcorta D, Preston G, Munger W, Sullivan P, Yang JJ, Waga I, Jeanette JC, Falk R** 2002 Microarray studies of gene expression in circulating leukocytes in kidney diseases. *Exp Nephrol* 10:139-149
- Alizadeh AA, Eisen MB, Davis RE** 2000 Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403:503-511
- Alizadeh AA, Ross DT, Perou CM, Van de Rijn M** 2001 Towards a novel classification of human malignancies based on gene expression patterns. *J Pathol* 195:41-52
- Arcellana-Panlilio M, Robbins SM** 2002 Cutting-edge technology I. Global gene expression profiling using DNA microarrays. *Am J Physiol Gastrointest Liver Physiol* 282:G397-G402
- Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, Sallan SE, Lander ES, Golub TR, Korsmeyer SJ** 2002 MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* 30:41-47
- Basu A, Lenka N, Mullick J, Avadhani NG** 1997 Regulation of murine cytochrome oxidase Vb gene expression in different tissues and during myogenesis. Role of a YY-1 factor-binding negative enhancer. *J Biol Chem* 272:5899-5908
- Bertucci F, Bernard K, Loriod B** 1999 Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. *Hum Mol Genet* 8:1715-1722
- Bertucci F, Houlgatte R, Nguyen C, Viens P, Jordan BR, Birnbaum D** 2001 Gene expression profiling of cancer by use of DNA arrays: how far from the clinic? *Lancet Oncol* 2:674-682
- Best CL, Emmert-Buck MR** 2001 Molecular profiling of tissue samples using laser capture microdissection. *Exp Rev Mol Diagn* 1:53-60
- 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, Meyerson M** 2001 Classification of human lung carcinomas by mRNA

- expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 98:13790–13795
- Brail LH, Jang A, Billia G, Iscove NN, Klamut HJ, Hill RP** 1999 Gene expression in individual cells: analysis using global single cell reverse transcription polymerase chain reaction (GSC RT-PCR). *Mutat Res* 406:45–54
- Conner EA, Teramoto T, Wirth PJ, Kiss A, Garfield S, Thorgeirsson SS** 1999 HGF-mediated apoptosis via p53/bax-independent pathway activating JNK1. *Carcinogenesis* 4:583–590
- Cowey CL, Fleming DR, Copland JA** 2001 Vitamin E analogs, troglitazone, and thiazolidinediones are effective inhibitors of cell proliferation in human pancreatic tumor cells. *Am Assoc Cancer Res* 42:173
- Debril M-B, Renaud J-P, Fajas L, Auwerx J** 2001 The pleiotropic functions of peroxisome proliferator-activated receptor gamma. *J Mol Med* 79:30–47
- DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA** 1996 Use of a cDNA microarray to analyze gene expression patterns in human cancer. *Nat Genet* 14:457–460
- Eisen MB, Brown PO** 1999 DNA arrays for analysis of gene expression. *Meth Enzymol* 303:179–205
- Eyster KM, Boles AL, Brannian JD, Hansen KA** 2002 DNA microarray analysis of gene expression markers on endometriosis. *Fertil Steril* 77:42
- Ferrando A, Sheffield-Moore M, Yeckel CW, Gilkison C, Jiang J, Achacosa, A, Lieberman SA, Tipton K, Wolfe RR, Urban RJ** 2002 Testosterone administration in older men improves muscle function: Molecular and physiological mechanisms. *Am J Physiol: Endocrinol Metab* 282:E601–E607
- Fujiwara T, Horikoshi H** 2002 Troglitazone and related compounds: therapeutic potential beyond diabetes. *Life Sci* 67:2405–2416
- 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, Petersen I** 2001 Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci USA* 98:13784–13789
- Gershon D** 2002 Microarray technology: an array of opportunities. *Nature* 416:885–891
- Gibson UE, Heid CA, et al.** 1996 A novel method for real time quantitative RT-PCR. *Genome Res* 6:995–1001
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES** 1999 Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531–537
- Grant EP, Pickard MD, Briskin MJ, Gutierrez-Ramos J-C** 2002 Gene expression profiles: creating new perspectives in arthritis research. *Arthrit Rheumat* 46:874–884
- Heid CA, Stevens J, Livak KJ, Williams PM** 1996 Real time quantitative RT-PCR. *Genome Res* 6:986–994
- Horikoshi H, Yoshioka T** 1998 Troglitazone — a novel antidiabetic drug for treating insulin resistance. *Drug Discov Today* 3:79–88
- Khan J, Simon R, Bittner M, Chen Y, Leighton SB, Pohida T, Smith PD, Jiang Y, Gooden GC, Trent J, Meltzer P** 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, Meltzer PS** 2001 Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 7:673–679

- Kimura SH, Ikawa M, Ito A, Okabe M, Nojima H** 2001 Cyclin G1 is involved in G2/M arrest in response to DNA damage and in growth control after damage recovery. *Oncogene* 20:3290–3300
- Kliewer SA, Xu HE, Lambert MH, Willson TM** 2002 Peroxisome proliferator-activated receptors: from genes to physiology. *Recent Prog Horm Res* 56:239–263
- Kondo M, Shibata T, Kumagai T, Osawa T, Shibata N, Kobayashi M, Sasaki S, Iwata M, Noguchi N, Uchida K** 2002 15-Deoxy-delta(12,14)-prostaglandin J(2): the endogenous electrophile that induces neuronal apoptosis. *Proc Natl Acad Sci USA* 99:7367–7372
- Konno H, Yodogawa M, Stumpp MT, Kroth P, Strotmann H, Motohashi K, Amano T, Hisabori T** 2000 Inverse regulation of F1-ATPase activity by a mutation at the regulatory region on the gamma subunit of chloroplast ATP synthase. *Biochem J* 352:783–788
- Korach KS** 2002 Bringing genomics research to endocrinology. *Endocrinology* 143:1982
- Lakhani SR, Ashworth A** 2001 Microarray and histopathological analysis of tumours: the future and the past? *Nat Rev* 1:151–157
- Lee ML, Kuo GC, Whitmore GA, Sklar J** 2000 Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci USA* 97:9834–9839
- Lenka N, Vijayarathy C, Mullick J, Avadhani NG** 1998 Structural organization and transcription regulation of nuclear genes encoding the mammalian cytochrome c oxidase complex. *Prog Nucleic Acid Res Mol Biol* 61:309–344
- Luo L, Salunga RC, Guo H** 1999 Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat Med* 5:117–122
- Luzina IG, Atamas SP, Wise R, Wigley FM, Xiao HQ, White B** 2002 Gene expression in bronchoalveolar lavage cells from scleroderma patients. *Am J Respir Cell Mol Biol* 26:549–557
- McDonald ER, El-Deiry WS** 2001 Checkpoint genes in cancer. *Ann Med* 33:113–122
- Mimmack ML, Ryan M, Baba H** 2002 Gene expression analysis in schizophrenia: reproducible up-regulation of several members of the apolipoprotein L family located in a high-susceptibility locus for schizophrenia on chromosome 22. *Proc Natl Acad Sci USA* 99:4680–4685
- Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, O'Connell JX, Zhu S, Fero M, Sherlock G, Pollack JF, Brown PO, Botstein D, Van de Rijn M** 2002 Molecular characterisation of soft tissue tumours: a gene expression study. *Lancet* 359:1301–1307
- Notterman DA, Alon U, Sierk AJ, Levine AJ** 2001 Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res* 61:3124–3130
- Phillips J, Eberwine JH** 1996 Antisense RNA amplification: a linear amplification method for analyzing the mRNA population from single living cells. *Methods: A Companion to Meth Enzymol* 10:283–288
- Ramaswamy S, Golub TR** 2002 DNA microarrays in clinical oncology. *J Clin Oncol* 7:1932–1941
- Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang C-H, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR** 2001 Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* 98:15149–15154
- Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, Iyer V, Jeffrey SS, Van de Rijn M, Waltham M, Pergamenschikov A, Lee JCF, Lashkari D, Shalon D, Myers TG, Weinstein JN, Botstein D, Brown PO** 2000 Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 24:227–235
- Rus V, Atamas SP, Shustova V, Luzina IG, Selaru F, Magder LS, Via CS** 2002 Expression of cytokine and chemokine-related genes in peripheral blood mononuclear cells from lupus patients by cDNA array. *Clin Immunol* 102:290

- Schena M, Shalon D, Davis RW, Brown PO** 1995 Quantitative monitoring of gene expression patterns with complementary DNA microarray. *Science* 270:467–470
- Simone NL, Bonner RG, Gillespie JW, Emmert-Buck MR, Liotta LA** 1998 Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends Genet* 14:272–276
- Singh A, Liu S, Crombie DL, Boehm M, Liebowitz MD, Heyman RA, Depre C, Nagy L, Tontonoz P, Davies PJ** 2001 Differential effects of rexinoids and thiazolidinediones on metabolic gene expression in diabetic rodents. *Mol Pharmacol* 59:765–773
- Sorlie 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, Lonning PE, Borresen-Dale A-L** 2001 Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98:10869–10874
- Sporn MB, Suh N, Mangelsdorf DJ** 2001 Prospects for prevention and treatment of cancer with selective PPAR $\gamma$  modulators (SPARMs). *Trends Mol Med* 7:395–400
- Tian B, Zhang Y, Luxon BA, Garofalo RP, Casola A, Sinha M, Brasier AR** 2002 Identification of NF- $\kappa$ B-dependent gene networks in respiratory syncytial virus-infected cells. *J Virol* 76:6800–6814
- Tsunoda SP, Aggeler R, Yoshida M, Capaldi RA** 2001 Rotation of the c subunit oligomer in fully functional F1F0ATP synthase. *Proc Natl Acad Sci USA* 98:898–902
- Tusher BG, Tibshirani R, Chu G** 2001 Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98:5116–5121
- Uray IP, Connelly JH, Frazier O, Taegtmeier H, Davies PJ** 2001 Altered expression to tyrosine kinase receptors Her2/neu and GP130 following left ventricular assist device (LVAD) placement in patients with heart failure. *J Heart Lung Transplant* 20:210
- Vahey MT, Nau ME, Jagodzinski LL, Yalley-Ogunro J, Taubman M, Michael NL, Lewis NG** 2002 Impact of viral infection on the gene expression profiles of proliferating normal human peripheral blood mononuclear cells infected with HIV type 1RF. *AIDS Res Hum Retroviruses* 18:179–192
- Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JK, Eberwine JH** 1990 Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* 87:1663–1667
- van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH** 2002 Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415:530–536
- Walczak R, Tontonoz P** 2002 PPARadigms and PPARadoxes: expanding roles for PPAR  $\gamma$  in the control of lipid metabolism. *J Lipid Res* 43:177–186
- Wang E, Miller LL, Ohnmacht GA** 2000 High fidelity mRNA amplification for gene profiling. *Nat Biotechnol* 18:457
- Willson TM, Lambert MH, Kliewer SA** 2001 Peroxisome proliferator-activated receptor  $\gamma$  and metabolic disease. *Annu Rev Biochem* 70:341–367
- Yeang C-H, Ramaswamy S, Tamayo P, Mukherjee S, Rifkin R, Angelo M, Reich M, Lander ES, Mesirov JP, Golub TR** 2001 Molecular classification of multiple tumor types. *Bioinformatics* 17(suppl 1):S316–S322
- Young RA** 2000 Biomedical discovery with DNA arrays. *Cell* 102:9–15
- Yun Z, Maecker HL, Johnson RS, Giaccia AJ** 2002 Inhibition of PPAR  $\gamma$  2 gene expression by the HIF-1 regulated gene DEC1/Stra13: a mechanism for regulation of adipogenesis by hypoxia. *Dev Cell* 2:331–341
- Zawel L, Yu J, Torrance CJ, Markowitz S, Kinzler KW, Vogelstein B, Zhou S** 2002 DEC1 is a downstream target of TGF- $\beta$  with sequence-specific transcriptional repressor activities. *Proc Natl Acad Sci USA* 99:2848–2853