

# Gene Expression Profiling: From Microarrays to Medicine

ASHANI T. WEERARATNA,<sup>1</sup> JAMES E. NAGEL,<sup>1</sup> VALERIA DE MELLO-COELHO,<sup>1</sup> and DENNIS D. TAUB<sup>1,2</sup>

Accepted: October 31, 2003

With the mapping of the human genome comes the ability to identify genes of interest in specific diseases and the pathways involved therein. Laboratory technology has evolved in parallel, providing us with the ability to assay thousands of these genes at once, a technique known as microarray analysis. The main question that this type of technology raises is how we can apply this powerful technology to clinical medicine. Recently, advances in data analysis, as well as standardization of the technology, have allowed us to examine this question, and indeed a few clinical trials currently being performed include microarrays as part of their protocol. In this review we outline the microarray technique and describe these types of studies in further detail.

**KEY WORDS:** Microarray; clinic; gene expression; review.

## INTRODUCTION

One of the promises of the Human Genome Project is that through knowledge of genomic organization and chromosomal location, it will be possible to identify and link specific genes to susceptibility to various human diseases. In the past, gene expression information has been obtained on a one-by-one, single-gene basis typically through the use of Northern Blot Analysis; however, the introduction of hybridization to nucleotide arrays now permits the rapid, simultaneous screening of the expression of several thousand individual genes at a given time. The two most common forms of gene expression profiling used today are the serial analysis of gene expression (SAGE) and microarray analysis. The SAGE technique is based on the principle that a 10- to 14-bp sequence referred to as a "tag" can uniquely identify a transcript, provided that the tag is

obtained from a unique position within a transcript (1). This method of profiling allows researchers to examine the changes in the absolute levels of transcripts in a cell and, because it does not require an *a priori* knowledge of the transcriptome, can uncover novel genes expressed therein. However, this technique is quite labor-intensive and technically challenging, and the costs involved with the generation and sequencing of SAGE libraries are beyond the scope of many laboratories. Microarray technology, the older of the two techniques, is intrinsically more "user-friendly." The first recorded instance of this technology is often overlooked, but was published in a study by Augenlicht *et al.* where, in 1987, investigators used a nylon membrane, containing 4000 complementary DNA (cDNA) sequences to examine changes in gene expression in colon cancer (2). Since these early studies, microarray profiling has been significantly refined and modified to optimize the sensitivity of the assay as well as the number of genes examined in a given experiment.

## MICROARRAY ANALYSIS: A PRIMER

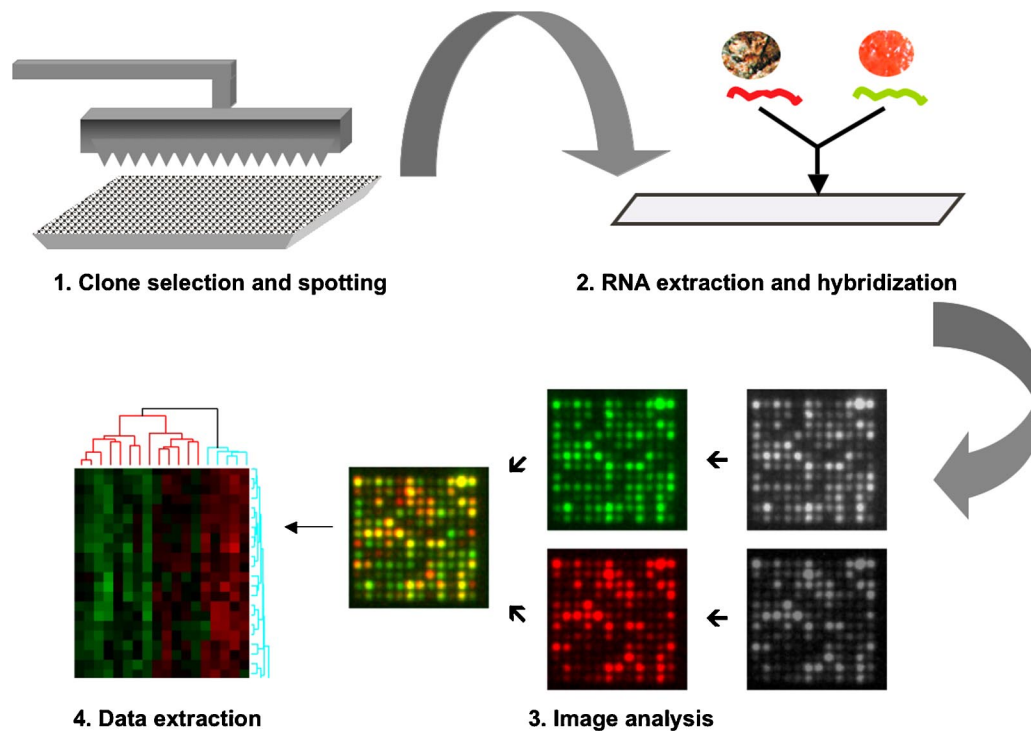
Gene expression profiling may provide valuable insights into the molecular mechanisms underlying disease. To perform a successful experiment, there is a need to identify clones of interest for arraying, isolate high-quality RNA from tissues of interest, and analyze the data in the most informative manner possible (Fig. 1). Each of these steps will be examined in detail below.

### *Experimental Design*

In a microarray experiment, gene expression is often compared in two samples of RNA. This typically means comparing "normal" to "diseased" tissues or "treated" and "untreated" cells or samples derived from various experimental conditions. What has become quite clear through its development and application is that microarray analysis is an exquisitely sensitive technique and prone to a

<sup>1</sup>Clinical Immunology Section, Laboratory of Immunology, Gerontology Research Center, National Institute on Aging, Nathan Shock Dr, Baltimore, Maryland.

<sup>2</sup>To whom correspondence should be addressed at Clinical Immunology Section, Laboratory of Immunology, Gerontology Research Center, National Institute on Aging, NIH, 5600 Nathan Shock Dr, Baltimore, Maryland 21224; e-mail: taubd@grc.nia.nih.gov.



**Fig. 1.** Microarray technology: From bedside to bench and back again. (1) Clones or oligonucleotides are selected and robotically spotted onto a glass slide or nylon filter. (2) RNA is extracted from cells or tissues of interest and labeled either with cy3 or cy5 (glass slides) or with  $P^{33}$  (nylon filters) and hybridized. (3) Images are analyzed using programs such as ArrayPro or IPLAB. (4) Data is clustered and information extracted using bioinformatics.

myriad of unavoidable variability, which leads to difficulty when designing an experiment. The first hurdle one must overcome is to select an appropriate disease state, or experimental condition as a reference against which all samples in a given test set can be compared. One major problem begins in simply defining “what is normal?” or obtaining a specimen to which a diseased tissue can be legitimately compared. Skin from different areas of the body can significantly vary, just as cells derived from different tissues or region of a given organ may significantly differ. Tumors are frequently a heterogenous mixed population displaying varying degrees of anaplasia, necrosis, and vascular proliferation. Thus, even comparing a single tumor cell type derived from different patients can yield quite varied gene expression profiles.

Overall, sample selection is a conundrum. Peripheral blood, the prototypical clinical specimen, is seldom a useful source of informative specimens simply because localized gene expression changes in tissues are not represented in RNA made from peripheral blood leukocytes. Moreover, the percentage of white blood cells within a given patient can vary from sample to sample affecting the RNA recovered. Because of the exquisite sensitivity

inherent in microarray analysis, the use of mixed cell populations is a tenuous proposition. This is especially true for more complex organ structures such as the brain, where, for example, dopaminergic cells that display pathogenesis in several neurodegenerative and addiction disorders, are very sparsely represented in the total organ cell population. Most commonly, inclusion of a pure population of the suspected infiltrating cell types in the experiment can assist in the identification of genes associated with infiltration or contamination, and then statistical analysis can be used to exclude these genes from the analysis. In addition, research techniques such as FACS or laser capture microdissection can further enrich these heterogenous cell populations, resulting in more isolated and defined cell subpopulations for profiling. However, such enrichments must also consider the fact that inflammatory infiltrates or cells present in an adjacent tissue may themselves be part of the disease process and therefore are an appropriate component of the specimen.

The development of many diseases may occur over an extended period of time and some may even include an orderly progression of stages. The originating event(s) leading to clinical symptoms or findings may have been

initiated many years prior to diagnosis so that specimens obtained when symptoms develop may have limited informational value in predicting pathoformic disease. Despite these caveats, over the past few years, a steady stream of reports has appeared describing the use of microarray technology in a variety of research areas including cancer, autoimmune or infectious disease, and a variety of inherited disorders, all with the intent of identifying and understanding their molecular origins and mechanisms. Some have indeed identified potentially valuable markers for diagnosis. More recently, as the genomes of various bacteria, viruses, parasites, and other pathogens are sequenced, studies have been directed toward elucidating specific genes involved in microbial pathogenicity and virulence with the obvious expectation that such genes may serve as potential therapeutic targets in disease treatment.

#### *Microarray Construction and Hybridization*

The first step in the construction of a microarray is to identify and collect clones (cDNAs) or short oligonucleotides that encode genes important for research purposes. cDNA arrays can be designed and constructed with a number of different goals in mind. Such arrays may be focused on a particular tissue, chromosome, developmental stage, gene family, disease, or functional characteristic (e.g., signaling molecules, cytokines, apoptotic-mediators), or may be unfocused. Oligonucleotide microarrays are manufactured by *in situ* synthesis on glass using a combination of photolithography and oligonucleotide chemistry. The result is a panel of short oligonucleotides that, depending on the particular array, identify up to about 33,000 discrete human genes. Recently, other manufacturers have begun to produce what are being called "spotted" oligonucleotide arrays. Rather than the oligonucleotide being directly synthesized on the array substrate, these arrays are constructed using a robotic pin-based microarrayer to spot conventionally synthesized 40- to 80-bp oligonucleotides onto glass slides or nylon filters (3).

Genes of interest can be identified using the public UniGene database (<http://www.ncbi.nlm.nih.gov/UniGene/>). UniGene is an experimental system for automatically partitioning nucleotide sequence data deposited in GenBank into a nonredundant set of gene-oriented clusters (4). Each UniGene cluster contains sequences that represent a unique gene as well as related information such as the tissue types in which the gene has been expressed and its chromosomal location. Unigene numbers may also correlate to hypothetical proteins (i.e., proteins identified by *in silico* analysis of genetic sequences) or as yet uncharacterized transcripts obtained from random-primed

cDNA libraries, referred to as expressed sequence tags (ESTs). Each UniGene cluster typically includes a number of clones that may be potentially used for cDNA array construction. A useful public source of cDNA clones are those made available by the Integrated Molecular Analysis of Genomes and their Expression (IMAGE) Consortium that also places sequence, map, and expression data about these clones into the public domain (4). At the present time, there are approximately 4.5 million ESTs in the NCBI databases (<http://www.ncbi.nlm.nih.gov/>) and most are available from commercial distributors. Additional cDNA clones are available from commercial enterprises and from research laboratories that have constructed and sequenced unique cDNA libraries.

After choosing the appropriate genes and ESTs for a given array, these genes can be cloned into plasmid vectors suitable for transforming bacteria. Bacterial clones containing cDNAs of interest are propagated and the DNA extracted and purified. The gene-specific inserts are amplified in microtiter plates by PCR. The best arrays contain only clones that are sequence-verified, to ensure accuracy and quality. Such verification is crucial to the reliability of the data obtained from such arrays. Using a multiwell format, these cDNA inserts or oligonucleotides can be spotted by a robotic microarrayer onto glass slides, or nylon or nitrocellulose membranes that have been pretreated to augment their surface charge and increase the adherence of the DNA (5). Currently, depending on whether the format is nylon-based or glass-chip-based, an array may contain anywhere from 500 (nylon) to over 30,000 (glass chip) genes. Glass- and nylon-based arrays are often regarded as alternative technologies. However, they have both strong and weak points that are often complementary. Filter-based arrays used with radioactivity generally require less total RNA, although with current protocols such as dendrimer or amino-allyl labeling, small amounts of RNA can be used for fluorescence arrays as well (6). However, filter-based arrays have lower per filter cost, making them an attractive choice for smaller laboratories (7). On the other hand, fluorescent labeling allows control and experimental RNA to be hybridized together, allowing for the significant advantage of a direct comparison.

The process of DNA hybridization involves the reassociation of single-stranded DNA to form double-stranded DNA with one strand originating from a cell or tissue under study and the other strand with the target sequence that has been printed or synthesized on the microarray. A crucial factor for successful hybridization is the purity and quality of the RNA extracted from the cells or tissue of interest. Contamination of this RNA with genomic DNA, proteins or detergent residues, or its degradation by ubiquitous ribonucleases may cause serious problems

during the RT-PCR steps of the procedure. The method of labeling probe RNA depends on the particular type of microarray being used for the study. With microarrays printed on glass slides, it is customary to label during reverse transcription one sample with the dye cyanine-3 (Cy3) that, when excited by light, yields green fluorescence and, the other sample with cyanine-5 (Cy5) that yields red fluorescence (8). Synthesized oligonucleotide arrays typically use biotinylated probes and are stained posthybridization with streptavidin conjugated to phycoerythrin. For microarrays using nylon membranes, the target RNA is typically radioactively labeled by incorporation of [<sup>33</sup>P] dCTP or [<sup>32</sup>P] dCTP nucleotides during reverse transcription (9). While not commonly performed, arrays on glass slides may also be queried with radiolabeled probes. Irrespective of labeling method, the probes are purified and incubated in a suitable buffer for 16–24 h with the microarray. Posthybridization, the arrays are washed and quantity of signal incorporated in each spot is measured using either a specialized slide reader or an imaging system.

#### *Image Analysis and Data Extraction*

Analysis of microarray data continues in an evolutionary state with a number of different research groups analyzing their data in a variety of ways using combinations of various microarray-specific, spreadsheet, data display, and statistical software programs (10–13). To date, there is no universally accepted method to analyze microarray data and thus the analytic method selected is frequently directed toward the specific research question being asked. Often, microarray data is examined using several techniques with the method providing the most robust interpretation being utilized for publication and further pursuit.

One of the challenges in array data analysis is to distinguish specific physiologic changes in gene expression from the noise and variability inherent within the microarray technique. Although there is a paucity of data specifically addressing such variability in human tissue, current available information suggests that the normal variance of expression of tightly regulated genes in a given tissue may range up to 20–30%. The miniaturization of the assay and the ability to conduct thousands of experiments at a given time (for analysis purposes, hybridization to each array spot can be considered a small experiment) in parallel inherently produces considerable variability in a microarray experiment (14). The sources of fluctuation accumulate at each step of the microarray procedure from the initial processing of the tissue sample, through target and array preparation, hybridization, and image processing (15, 16).

Whereas fluorescent-labeling of spotted cDNAs allows both the experimental and control RNA to be hybridized on the same microarray, radioactively labeled samples require that each specimen be hybridized on a separate array. The arrays are queried using specific software that recognizes and assigns a numeric density value to each spot on the microarray. Irrespective of whether fluorescent- or radioactive-labeled microarrays are used, most research groups then apply a normalization procedure to the family of arrays included in an experiment to bring their signal range into an acceptable confidence interval and adjust the signals on each filter to approximate a normal distribution with a mean of 0 and a standard deviation of 1. Many different normalization techniques have been described but there is yet no agreement as to a “best” way to normalize microarray data (17). Normalization is important to eliminate artifacts and allow comparison between filters. However, normalization has limits that, when ignored, can result in the creation of false signals and misinterpretation of the data (18).

Following normalization, microarray data is examined to identify differences in gene expression. The simplest technique is the ratio of experimental to control or fold change. Many published studies have used the “twofold change” criterion as a measure of significance and it has been shown that this method can be reproducible even taking into account interlaboratory variability. For example, one study compared gene expression changes in yeast, in three different laboratories, and showed a greater than 95% concordance in genes increased over twofold (19). While this method is straightforward, it rapidly becomes apparent that this calculation may not be useful in all cases, most importantly because it eliminates all information about absolute gene expression levels. More significantly, fold change does not embrace any knowledge of biology. Succinctly, genes that are members of a defined pathway or that respond to a common challenge are likely to be coregulated and therefore could be expected to display similar patterns of expression.

A statistical technique generically termed “exploratory multivariate data” or “cluster analysis” has come into the forefront to identify groups of genes that display similar changes in expression. In general, classical clustering techniques start by creating a set of bidirectional distance vectors that represent the similarity between genes and between clusters of genes. An iterative process is then undertaken where each gene profile is compared to all the other gene profiles and clusters until eventually all genes are in one cluster (10). There are numerous hierarchical clustering algorithms that differ in their starting point and the manner in which they calculate and compare the distances between the existing clusters and the remainder of the data

set (17). Bottom-up (agglomerative) hierarchical clustering was first applied to microarray analysis by Eisen *et al.* (20). Because this technique produces readily visualized patterns of coordinately regulated genes and is supported by software programs such as Clusteru<sup>®</sup> and TreeView<sup>®</sup> created by Eisen (<http://rana.lbl.gov/>), it has become extremely popular for microarray analysis. Other types of cluster analysis include multidimensional cluster analysis, which uses the similarities between two samples to generate a Pearson's pairwise correlation coefficient. This gives an idea of the magnitude of difference between two samples and, when applied to three or more samples, also provides a direction of the difference between them. Once these samples have been mapped into a three-dimensional plot, the similarity between two samples can be assessed by the distance between them. The more tightly two samples cluster together, the more similar they are. Once these classes of genes have been identified, statistical analyses can be used to best determine which genes cause the samples to segregate as they do (21).

However, irrespective of the particular clustering method chosen, it quickly becomes apparent that microarrays can differentiate tens of thousands of genes, only a small subset, in the range of 5–10%, undergo significant change in expression, and are therefore worthy of additional study. This point led to the testing of another group of statistical techniques that included self-organizing maps (22) and *K*-means clustering (23) that organize the expression data before actual clustering (24, 25). More recently mathematical procedures such as probabilistic principal component analysis (26) and support vector machines (SVMs) (27–29), as well as models based on neural network designs (12, 30) or Bayesian inference (31, 32), have begun to be explored. In these techniques, an analysis algorithm is “trained” with a portion of the data set and these results used to heuristically select among various data-fitting models, one of which is then used to examine the entire data set. If an analysis technique can be developed and validated that can identify the genes that undergo a significant change in expression and remove those that do not, it could alter microarray design and construction in favor of smaller focused arrays that query only biologically relevant genes. Clearly, gene expression analysis remains a work in progress. The goal is to develop tools that can identify meaningful expression changes, evaluate the significance of these changes to determine whether they are different than what might occur by chance alone, and ultimately group genes to reveal and examine the combinatorial nature of transcriptional control.

Two important points to take into consideration when running a microarray experiment are the necessity to run

replicate experiments (33–35) and to validate the gene expression changes using other techniques such as real-time PCR. It may also be useful to analyze the expression levels of proteins encoded by the altered genes. This can be done by techniques such as immunocytochemistry or Western blot, using specific antibodies for the proteins of interest. Recently, members of the Microarray Gene Expression Data ([www.mged.org](http://www.mged.org)) society have advocated the adoption of the Minimal Information About a Microarray Experiment (MIAME) guidelines (36, 37) ([www.mged.org/Workgroups/MIAME/miame\\_checklist.html](http://www.mged.org/Workgroups/MIAME/miame_checklist.html)). One effect of these standards seems certain—there will be a move to the use of a single microarray product for all future clinical studies.

### *Bioinformatics and Data Mining*

Bioinformatics applies principles of information science and technology to make life science data more understandable and useful. In practice, when dedicated computer software is used to search for hidden patterns in groups of data and to link this information to other data, this is referred to as “data mining.” The usual first endpoint of a microarray experiment is a list of the genes or their GenBank accession numbers that have undergone a meaningful change in expression. More often than not, a few of the gene names are recognizable, but the majority are not. However, what we really want to know is the function(s) of the gene, how it is related to other biologic pathways and processes or defined clinical syndromes or diseases, how this gene was affected by how the microarray experiment was conducted, and to link this information with clinical data, treatment outcomes, and drug responses. Ultimately, we want to develop gene expression data into a prognostic or diagnostic tool.

As previously noted, the trend in microarray analysis is toward various unsupervised clustering techniques. More recently, supervised techniques such as SVM or neural networks that allow nonexpression data to be incorporated into the clustering model have shown added promise. However, it is accepted that there is really no single technique that is appropriate for all data sets, leaving the interpretation of microarray data an inexact science (17). Depending on how the data is processed, different relationships may be revealed which in and of themselves may be informative. Data-mining software continues to evolve with several dozen commercial and academic products available. So far, no “one program does it all,” and one is typically left with moving the expression data between various database, statistical, graphics, and annotation packages during analysis.

Numerous new high-quality databases have been constructed and other previously existing databases such as GenBank significantly augmented by data produced by the Human Genome Project (<http://www.ncbi.nlm.nih.gov/Sitemap/index.html>). Together, these databases contain a truly amazing amount of information that is being updated and expanded on a regular basis. The effect of this continuous updating is that there is a degree of impermanence associated with the data. This has produced the secondary effect of turning the retrieval of information about individual genes from the various online genomic databases such as LocusLink ([www.ncbi.nlm.nih.gov/LocusLink/](http://www.ncbi.nlm.nih.gov/LocusLink/)), OMIM ([www.ncbi.nlm.nih.gov/omim/](http://www.ncbi.nlm.nih.gov/omim/)), Aceview ([www.ncbi.nlm.nih.gov/IEB/Research/Acembly/](http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/)), UniGene ([www.ncbi.nlm.nih.gov/unigene/](http://www.ncbi.nlm.nih.gov/unigene/)), KEGG ([www.genome.ad.jp/kegg/kegg2.html](http://www.genome.ad.jp/kegg/kegg2.html)), and GeneCards<sup>TM</sup> (<http://nciarrray.nci.nih.gov/cards/>) into complex hot-linked spreadsheets and databases, that are nonetheless user-friendly. This situation will persist into the foreseeable future as new gene function and pathway data becomes available over the Internet.

#### *Single-Nucleotide Polymorphisms (SNPs)*

Another concept receiving a great deal of recent attention, particularly by the pharmaceutical industry, is the observation that there are genetically based differences in drug and immune responses between individuals that may be utilized to optimize a person's therapy and that are responsible for adverse or suboptimal drug responses. Further, a logical corollary to this idea is that through sequence analysis, it will be possible to identify disease-susceptibility genes that might represent potential targets for future drug development or other interventional therapy (38). This has created a new field, pharmacogenomics, that examines inherited gene variations that dictate drug response and studies their effect on clinical drug responses (39). Presently, the identification and cataloguing of SNPs is the most popular method to investigate these complex genetic associations. SNPs are the most common form of genetic variation, occurring approximately once every 1000 bp throughout the 3 billion base pair human genome. Although their incidence varies substantially across the genome, the total number of human SNPs is estimated to be over 10 million (40). As of September 2002, 4.3 million SNPs have been deposited in the public dbSNP database (<http://www.ncbi.nlm.nih.gov/SNP/>) and approximately 1.25 million SNPs have been mapped *in silico* to the human genome by The SNP Consortium (TSC; a private, not-for-profit alliance of 13 major multinational companies and the Wellcome Trust; <http://snp.cshl.org>) (41, 42).

Since DNA possesses only four nucleotides, the number of potential SNP variations is quite restricted, making SNPs well-suited for high-throughput automated or parallel analysis (43). However, the bottleneck in SNP genotyping is the sample preparation; i.e., purifying hundreds of thousands of different loci so that SNP genotyping can be done at each site. Many different SNP detection technologies are under commercial development (44). Presently, most large-scale projects examining genome-wide SNP expression are based on differential hybridization affinity using either spotted or *in situ* synthesized oligonucleotide arrays (45, 46) or utilize mass spectroscopy for genotype analysis (7, 47). However, a new spotted thiol-modified oligonucleotide gold thin film array appears capable of substantially improving detection speed and sensitivity (48). SNP data analysis is much simpler than microarray data analysis because the readout can be designed to be binary (i.e., fluorescent or nonfluorescent) rather than scalable. At present, SNPs have provided the most clinically relevant data linked to the HGP but continue to be a work in progress. A number of clinical studies are already under way using SNP-based genetic testing to identify patients who are at increased risk for diabetes, cardiovascular disease, adverse drug reactions, cancer, or deep vein thrombosis. SNP genotyping holds great promise as a breakthrough technology that will introduce the era of personalized medicine. However, the clinical SNP genotyping studies that are now in progress are largely, if not exclusively, based on gene-disease associations and gene polymorphisms that were discovered 5–10 years previously demonstrating that considerable important work is still needed to identify meaningful disease-causing and modifying genes. Linking these discoveries with public data produced by large genome-wide SNP discovery and validation initiatives can be expected to promote a gradual introduction of SNP genotyping into diagnostic medicine and gene-based pharmacotherapeutics.

#### FROM BEDSIDE TO BENCH AND BACK AGAIN: MICROARRAYS IN THE CLINIC

Although there are several potential pitfalls associated with microarray technology, it is a powerful technique, and as evidenced by the surge in the medical literature over the past few years, has become increasingly popular (Table I). It is nevertheless accepted that the widespread inclusion of clinical data into microarray analysis algorithms will not be simple. Several decades of research and development of clinical record systems has shown that a data model needed to capture the broad array of clinical parameters is extremely complex and difficult to standardize.

**Table I.** Microarray Use in Clinical Research

Disease	Refs.
Alcoholism	
profiles alcoholic frontal cortex on UniGEM V array	66
profiles alcoholic motor and frontal cortex on UniGEM V array	67
Cardiovascular	
profiles cardiomyopathy on CardioChip	68
profiles heart failure on 10,272 gene cardiac cDNA array	69
Diabetes	
profiles human islet cells on Affymetrix U95A array	70
profiles human muscle cells on Affymetrix Hu6800 array	71
Endocrinology	
profiles skeletal muscle from T3-Rx pts on 22,640 spot cDNA array	72
profiles differentiation of preadipocytes on Mu Gene Discovery array	73
Gerontology	
review of microarray applied to aging	74
profiles progeria fibroblasts on Affymetrix HuGeneFL array	75
Immunology	
profiles innate immune response to bacteria on NCI LymphoChip	76
profiles PBMC from lupus patients on Panorama gene array	77
Neuroscience	
profiles 24 gliomas on ATLAS gene expression array	78
profiles brain from patients with Alzheimer's disease on UniGene Lifearray	79
profiles schizophrenic prefrontal cortex on Hu UniGEM V array	80
Oncology	
Lymphoma and Leukemia	
profiles 96 B cell lymphomas and CLL on NCI LymphoChip	81
profiles CML hematopoietic stem cells on ExpressChip <sup>TM</sup> oligo array	82
Melanoma	
profiles neural crest cells on 22,889 melanocyte EST cDNA array	83
uses CAST algorithm to subclassify histopathologically similar tumors	84
gene expression profiling of melanoma	85
Colorectal	
profiles colon carcinoma cell line on Affymetrix HuGeneFL array	86
profiles 10 primary and metastatic colon carcinomas on 9191 cDNA array	87
Breast	
profiles hereditary breast cancer on cDNA and tissue arrays	88
profiles 97 primary breast carcinomas on Hu 25K array	89
Gastric, Liver	
profiles 26 gastric carcinomas on Affymetrix HuGeneFL array	90
profiles 20 hepatocellular carcinomas on 23,040 spot cDNA array	91
Prostate, Renal	
profiles prostate cancer xenografts on ATLAS and ResGen cDNA arrays	92
profiles hormonal-Rx prostate cancer xenografts on 2 cDNA arrays	93
profiles kidney disease and renal cell carcinomas on Affymetrix U95A array	94
Other	
profiles alveolar rhabdomyosarcoma cell lines on 1238 spot cDNA array	95
profiles ovarian tissue and cell lines on MICROMAX human cDNA array	96
Toxicogenomics	
describes use of microarrays in toxicology and NIEHS Hu ToxChip array	97
describes the use of gene expression profiling in toxicology studies	98

Consequently, unlike most genomic data, there are few shared online national clinical databases (49). Serious concerns about privacy issues are also certain to have significant impact on universal access to clinical data. Despite these concerns, several studies have emerged that indicate that microarray technology may have value in the clinic and a few clinical trials even include microarrays as part of their protocol. Many studies have used this type of experiment to classify disease states, in the hopes of being able to predict clinical outcome, response to treatment,

and perhaps new molecular targets for improved therapy. For example, molecular profiling of cutaneous melanoma allowed for the identification of a more motile group of tumors histopathologically indistinguishable from their less aggressive counterparts (21). In addition, potential usefulness of microarray-derived gene expression data has been shown in several recent studies of lymphoma, leukemia (25, 50), and multiple myeloma (51) where modeling techniques that incorporate outcome and drug response during treatment were used to define tumor types or patient

groups and to suggest rational targets for drug therapy or development. It is becoming increasingly evident that these techniques may have a significant impact on current diagnostic methods (52).

Although many studies have demonstrated that current clinical parameters are reliable predictors of outcome, a few studies are beginning to reveal that certain prognostic indicators can more efficiently be derived from profiling studies. Analysis of breast cancer samples by microarray revealed that where standard clinical and histological criteria can be useful in predicting disease outcome in patients with early onset breast cancer, these patients had a very distinct gene expression signature that acted as an even more powerful predictor (53). This sort of robust prediction can also be made on the basis of microarray analysis of children presenting with medulloblastomas, in whom, again, outcome can be determined on the basis of their molecular signature (54).

In addition to classifying disease states and predicting outcomes, microarray analysis can also be used to analyze the effects of treatments and patient response to therapy. A recent study demonstrated that the effects of diverse regulators of breast cancer growth on breast cancer cells in culture linked the behavior of these cells to important clinical properties seen in *in vivo* specimens (55). Array analysis has been used to examine the probability of the rejection of renal allografts, by studying the gene expression profiles of patients who rejected their grafts as compared to those who did not. Again, accurate predictions as to whether or not a patient would reject a graft could be made on the basis of their molecular profiles. Other examples include using an oligonucleotide array, encoding several variations for the gene which encodes debrisoquine hydroxylase (CYP2D6) that metabolizes various psychotropic medications, and thus similar to an SNP chip, where researchers were able to determine which patients might need adjustments in dosage due to their ability to metabolize these drugs, based on the alleles expressed (56). In another study examining the changes in skeletal muscle tissues of patients with and without insulin treatment, several genes were differentially expressed. These genes were associated with muscle insulin resistance and complications associated with insulin metabolism, allowing again for the reassessment of treatment of patients with differing profiles (57). When genes are identified that are useful as prognostic indicators, or markers of response, as indicated by the aforementioned studies, smaller arrays can be custom made to reflect these discoveries, and to aid with patient assessment, diagnostically or therapeutically. In at least two cases, a chip has been made to aid in diagnosis. One of these, the OvaChip, contains genes involved in ovarian cancer as identified by SAGE analy-

sis, and the utility of this chip is under investigation by several different groups (58). Another, the lymphochip, uses a custom array to help diagnose tumors of lymphoid origin (59, 60). In cancers where tumor type and origin can be difficult to diagnose, this type of chip could have great utility. In addition, using a microarray platform to analyze viral DNA has proven its effectiveness. An HPV DNA chip uses a microarray platform to screen patients for possible human papilloma virus (HPV) infection, by spotting several different HP viruses on a microarray, allowing clinicians and researchers to determine the possibility of probable complications (such as cervical cancer) depending on the type of HPV present in the patient (61, 62). Furthermore, the most recent array application of clinical significance was the use of microarray technology to identify the virus responsible for SARS. To do this, researchers created a chip containing over 12,000 different viral gene signatures (the ViroChip) and only a few spots on this chip, all of which correlated to corona virus, showed positive expression (63–65). The time saved using this method of analysis may significantly advance the discovery of a treatment for this epidemic and others of its kind.

#### THE FUTURE OF MICROARRAYS

To date, microarray analysis has existed almost exclusively as research tool that requires considerable effort and time by skilled individuals to prepare high-quality RNA, label and hybridize the arrays, and read and analyze the data. Although microarray technology has begun to enter clinical medicine, several significant hurdles need to be overcome. For routine clinical lab use, significant improvements are needed in microarray fabrication, hybridization methodology, and analysis that will allow much or all of the processes to be fully automated and thus increase reproducibility within and across experiments. Microfluidics and nanofabrication technologies that range from the use of DNA as a construction material for mechanical devices to the use of carbon nanotubes to produce microarray-like device may have greater potential for full automation as well as increasing throughput speed and accuracy in the study of gene expression. In addition, the field of proteomics is a rapidly burgeoning one, and the identification of proteins and antigens for therapeutic use will be of high significance in the future. Several commercial software vendors have already announced they plan to modify their data-mining software to link nucleotide and protein databases and tools that in the future may allow both individual gene transcription and translation to be readily evaluated. Upon the accumulation of these



technologies and data-mining tools, it is likely that the promise of microarrays as a tool for the clinician may one day be realized.

## REFERENCES

1. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: Serial analysis of gene expression. *Science* 270:484–487, 1995
2. Augenlicht LH, Wahrman MZ, Halsey H, Anderson L, Taylor J, Lipkin M: Expression of cloned sequences in biopsies of human colonic tissue and in colonic carcinoma cells induced to differentiate in vitro. *Cancer Res* 47:6017–6021, 1987
3. Nakanishi T, Oka T, Akagi T: Recent advances in DNA microarrays. *Acta Med Okayama* 55:319–328, 2001
4. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, Wagner L: Database resources of the National Center for Biotechnology. *Nucleic Acids Res* 31:28–33, 2003
5. Schulze A, Downward J: Navigating gene expression using microarrays—A technology review. *Nat Cell Biol* 3:E190–E195, 2001
6. Manduchi E, Searce LM, Brestelli JE, Grant GR, Kaestner KH, Stoeckert CJ, Jr: Comparison of different labeling methods for two-channel high-density microarray experiments. *Physiol Genomics* 10:169–179, 2002
7. Bertucci F, Bernard K, Liorid B, Chang YC, Granjeaud S, Birnbaum D, Nguyen C, Peck K, Jordan BR: Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. *Hum Mol Genet* 8:1715–1722, 1999
8. Schena M, Shalon D, Davis RW, Brown PO: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–470, 1995
9. Lennon GG, Lehrach H: Hybridization analyses of arrayed cDNA libraries. *Trends Genet* 7:314–317, 1991
10. Carr KM, Bittner M, Trent JM: Gene-expression profiling in human cutaneous melanoma. *Oncogene* 22:3076–3080, 2003
11. Peterson LE: CLUSFAVOR 5.0: Hierarchical cluster and principal-component analysis of microarray-based transcriptional profiles. *Genome Biol* 3:SOFTWARE0002, 2002
12. Ringner M, Peterson C: Microarray-based cancer diagnosis with artificial neural networks. *Biotechniques (Suppl)* 34:30–35, 2003
13. Ringner M, Peterson C, Khan J: Analyzing array data using supervised methods. *Pharmacogenomics* 3:403–415, 2002
14. Pietu G, Alibert O, Guichard V, Lamy B, Bois F, Leroy E, Mariage-Sampson R, Houlgatte R, Soularue P, Auffray C: Novel gene transcripts preferentially expressed in human muscles: revealed by quantitative hybridization of a high density cDNA array. *Genome Res* 6:492–503, 1996
15. Schuchhardt J, Beule D, Malik A, Wolski E, Eickhoff H, Lehrach H, Herzl H: Normalization strategies for cDNA microarrays. *Nucleic Acids Res* 28:E47, 2000
16. Herzl H, Beule D, Kielbasa S, Korb J, Sers C, Malik A, Eickhoff H, Lehrach H, Schuchhardt J: Extracting information from cDNA arrays. *Chaos* 11:98–107, 2001
17. Quackenbush J: Computational analysis of microarray data. *Nat Rev Genet* 2:418–427, 2001
18. Beissbarth T, Fellenberg K, Brors B, Arribas-Prat R, Boer J, Hauser NC, Scheideler M, Hoheisel JD, Schutz G, Poustka A, Vingron M: Processing and quality control of DNA array hybridization data. *Bioinformatics* 16:1014–1022, 2000
19. Piper MD, Daran-Lapujade P, Bro C, Regenberg B, Knudsen S, Nielsen J, Pronk JT: Reproducibility of oligonucleotide microarray transcriptome analyses. An interlaboratory comparison using chemostat cultures of *Saccharomyces cerevisiae*. *J Biol Chem* 277:37001–37008, 2002
20. Eisen MB, Spellman PT, Brown PO, Botstein D: Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863–14868, 1998
21. Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, Samps 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, Sondak V: Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 406:536–540, 2000
22. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitarawan S, Dmitrovsky E, Lander ES, Golub TR: Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 96:2907–2912, 1999
23. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: Systematic determination of genetic network architecture. *Nat Genet* 22:281–285, 1999
24. Sherlock G: Analysis of large-scale gene expression data. *Curr Opin Immunol* 12:201–205, 2000
25. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286:531–537, 1999
26. Crescenzi M, Giuliani A: The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data. *FEBS Lett* 507:114–118, 2001
27. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M, Jr, Haussler D: Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci USA* 97:262–267, 2000
28. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D: Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 16:906–914, 2000
29. 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, Golub TR: Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 8:68–74, 2002
30. Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS: Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 7:673–679, 2001
31. Friedman N, Linial M, Nachman I, Pe'er D: Using Bayesian networks to analyze expression data. *J Comput Biol* 7:601–620, 2000
32. Broet P, Richardson S, Radvanyi F: Bayesian hierarchical model for identifying changes in gene expression from microarray experiments. *J Comput Biol* 9:671–683, 2002
33. Carlisle AJ, Prabhu VV, Elkahlon A, Hudson J, Trent JM, Linehan WM, Williams ED, Emmert-Buck MR, Liotta LA, Munson PJ, Krizman DB: Development of a prostate cDNA microarray and statistical gene expression analysis package. *Mol Carcinog* 28:12–22, 2000
34. Coombes KR, Highsmith WE, Krogmann TA, Baggerly KA, Stivers DN, Abruzzo LV: Identifying and quantifying sources of variation

- in microarray data using high-density cDNA membrane arrays. *J Comput Biol* 9:655–669, 2002
35. Lee ML, Kuo FC, Whitmore GA, Sklar J: Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci USA* 97:9834–9839, 2000
  36. Ball CA, Sherlock G, Parkinson H, Rocca-Sera P, Brooksbank C, Causton HC, Cavalieri D, Gaasterland T, Hingamp P, Holstege F, Ringwald M, Spellman P, Stoeckert CJ, Jr, Stewart JE, Taylor R, Brazma A, Quackenbush J: Standards for microarray data. *Science* 298:539, 2002
  37. 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 FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M: Minimum information about a microarray experiment (MIAME) toward standards for microarray data. *Nat Genet* 29:365–371, 2001
  38. Roses AD: Pharmacogenetics and the practice of medicine. *Nature* 405:857–865, 2000
  39. McCarthy JJ, Hilfiker R: The use of single-nucleotide polymorphism maps in pharmacogenomics. *Nat Biotechnol* 18:505–508, 2000
  40. Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han JH, Duan J, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windemuth A, Xu C, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G, Vovis GF: Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293:489–493, 2001
  41. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D: A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928–933, 2001
  42. Holden AL: The SNP consortium: Summary of a private consortium effort to develop an applied map of the human genome. *Biotechniques (Suppl)*32:22–24, 2002
  43. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES: Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22:231–238, 1999
  44. Wypgant M: SNP market view: Opportunities, technologies, and products. *Biotechniques (Suppl)*32:78–86, 2002
  45. Hirschhorn JN, Sklar P, Lindblad-Toh K, Lim YM, Ruiz-Gutierrez M, Bolk S, Langhorst B, Schaffner S, Winchester E, Lander ES. SBE-TAGS: An array-based method for efficient single-nucleotide polymorphism genotyping. *Proc Natl Acad Sci USA* 97:12164–12169, 2000
  46. Fan JB, Chen X, Halushka MK, Berno A, Huang X, Ryder T, Lipshutz RJ, Lockhart DJ, Chakravarti A: Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res* 10:853–860, 2000
  47. Buetow KH, Edmonson M, MacDonald R, Clifford R, Yip P, Kelley J, Little DP, Strausberg R, Koester H, Cantor CR, Braun A: High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc Natl Acad Sci USA* 98:581–584, 2001
  48. Lu M, Shortreed MR, Hall JG, Wang L, Berggren T, Stevens PW, Kelso DM, Lyamichev V, Neri B, Smith LM: A surface invasive cleavage assay for highly parallel SNP analysis. *Hum Mutat* 19:416–422, 2002
  49. Kohane IS: Bioinformatics and clinical informatics: The imperative to collaborate. *J Am Med Inform Assoc* 7:512–516, 2000
  50. Ohmine K, Ota J, Ueda M, Ueno S, Yoshida K, Yamashita Y, Kirito K, Imagawa S, Nakamura Y, Saito K, Akutsu M, Mitani K, Kano Y, Komatsu N, Ozawa K, Mano H: Characterization of stage progression in chronic myeloid leukemia by DNA microarray with purified hematopoietic stem cells. *Oncogene* 20:8249–8257, 2001
  51. Zhan F, Hardin J, Kordsmeier B, Bumm K, Zheng M, Tian E, Sanderson R, Yang Y, Wilson C, Zangari M, Anaissie E, Morris C, Muwalla F, van Rhee F, Fassas A, Crowley J, Tricot G, Barlogie B, Shaughnessy J, Jr: Global gene expression profiling of multiple myeloma, monoclonal gammopathy of undetermined significance, and normal bone marrow plasma cells. *Blood* 99:1745–1757, 2002
  52. Korenberg MJ: Prediction of treatment response using gene expression profiles. *J Proteome Res* 1:55–61, 2002
  53. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Bernards R, Friend SH: Expression profiling predicts outcome in breast cancer. *Breast Cancer Res* 5:57–58, 2002
  54. 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, Golub TR: Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 415:436–442, 2002
  55. Cunliffe H, Ringnér M, Bilke S, Walker RL, Cheung JM, Chen Y, Meltzer PS: The gene expression response of breast cancer to growth regulators: Patterns and correlation with tumor expression profiles. *Cancer Res* 63:7158–7166, 2003
  56. Murphy GM, Jr, Pollock BG, Kirshner MA, Pascoe N, Cheuk W, Mulsant BH, Reynolds CF, III: CYP2D6 genotyping with oligonucleotide microarrays and nortriptyline concentrations in geriatric depression. *Neuropsychopharmacology* 25:737–743, 2001
  57. Sreekumar R, Halvatsiotis P, Schimke JC, Nair KS: Gene expression profile in skeletal muscle of type 2 diabetes and the effect of insulin treatment. *Diabetes* 51:1913–1920, 2002
  58. Sawiris GP, Sherman-Baust CA, Becker KG, Cheadle C, Teichberg D, Morin PJ: Development of a highly specialized cDNA array for the study and diagnosis of epithelial ovarian cancer. *Cancer Res* 62:2923–2928, 2002
  59. Alizadeh A, Eisen M, Davis RE, Ma C, Sabet H, Tran T, Powell JJ, Yang L, Marti GE, Moore DT, Hudson JR, Jr, Chan WC, Greiner T, Weisenburger D, Armitage JO, Lossos I, Levy R, Botstein D, Brown PO, Staudt LM: The lymphochip: A specialized cDNA microarray for the genomic-scale analysis of gene expression in normal and malignant lymphocytes. *Cold Spring Harb Symp Quant Biol* 64:71–78, 1999
  60. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JJ, 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, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown P, Staudt LM: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403:503–511, 2000
  61. An HJ, Cho NH, Lee SY, Kim IH, Lee C, Kim SJ, Mun MS, Kim SH, Jeong JK: Correlation of cervical carcinoma and precancerous

- lesions with human papillomavirus (HPV) genotypes detected with the HPV DNA chip microarray method. *Cancer* 97:1672–1680, 2003
62. Kim CJ, Jeong JK, Park M, Park TS, Park TC, Namkoong SE, Park JS: HPV oligonucleotide microarray-based detection of HPV genotypes in cervical neoplastic lesions. *Gynecol Oncol* 89:210–217, 2003
  63. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, Rollin PE, Dowell SF, Ling AE, Humphrey CD, Shieh WJ, Guarner J, Paddock CD, Rota P, Fields B, DeRisi J, Yang JY, Cox N, Hughes JM, LeDuc JW, Bellini WJ, Anderson LJ: A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 348:1953–1966, 2003
  64. Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, DeRisi JL: Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci USA* 99:15687–15692, 2002
  65. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Penaranda S, Bankamp B, Maher K, Chen MH, Tong S, Tamin A, Lowe L, Frace M, DeRisi JL, Chen Q, Wang D, Erdman DD, Peret TC, Burns C, Ksiazek TG, Rollin PE, Sanchez A, Liffick S, Holloway B, Limor J, McCaustland K, Olsen-Rasmussen M, Fouchier R, Gunther S, Osterhaus AD, Drosten C, Pallansch MA, Anderson LJ, Bellini WJ: Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 300:1394–1399, 2003
  66. Lewohl JM, Wang L, Miles MF, Zhang L, Dodd PR, Harris RA: Gene expression in human alcoholism: microarray analysis of frontal cortex. *Alcohol Clin Exp Res* 24:1873–1882, 2000
  67. Mayfield RD, Lewohl JM, Dodd PR, Herlihy A, Liu J, Harris RA: Patterns of gene expression are altered in the frontal and motor cortices of human alcoholics. *J Neurochem* 81:802–813, 2002
  68. Barrans JD, Allen PD, Stamatiou D, Dzau VJ, Liew CC: Global gene expression profiling of end-stage dilated cardiomyopathy using a human cardiovascular-based cDNA microarray. *Am J Pathol* 160:2035–2043, 2002
  69. Hwang JJ, Allen PD, Tseng GC, Lam CW, Fananapazir L, Dzau VJ, Liew CC: Microarray gene expression profiles in dilated and hypertrophic cardiomyopathic endstage heart failure. *Physiol Genomics* 10:31–44, 2002
  70. Shalev A, Pise-Masison CA, Radonovich M, Hoffmann SC, Hirshberg B, Brady JN, Harlan DM: Oligonucleotide microarray analysis of intact human pancreatic islets: identification of glucose-responsive genes and a highly regulated TGFbeta signaling pathway. *Endocrinology* 43:3695–3698, 2002
  71. Sreekumar R, Halvatsiotis P, Schimke JC, Nair KS: Gene expression profile in skeletal muscle of type 2 diabetes and the effect of insulin treatment. *Diabetes* 51:1913–1920, 2002
  72. Clement K, Viguier N, Diehn M, Alizadeh A, Barbe P, Thalamas C, Storey JD, Brown PO, Barsh GS, Langin D: In vivo regulation of human skeletal muscle gene expression by thyroid hormone. *Genome Res* 12:281–291, 2002
  73. Guo X, Liao K: Analysis of gene expression profile during 3T3-L1 preadipocyte differentiation. *Gene* 251:45–53, 2000
  74. Han E, Hilsenbeck SG: Array-based gene expression profiling to study aging. *Mech Ageing Dev* 122:999–1018, 2001
  75. Ly DH, Lockhart DJ, Lerner RA, Schultz PG: Mitotic misregulation and human aging. *Science* 287:2486–2492, 2000
  76. Boldrick JC, Alizadeh AA, Diehn M, Dudoit S, Liu CL, Belcher CE, Botstein D, Staudt LM, Brown PO, Relman DA: Stereotyped and specific gene expression programs in human innate immune responses to bacteria. *Proc Natl Acad Sci U S A* 99:972–977, 2002
  77. Rus V, Atamas SP, Shustova V, Luzina IG, Selaru F, Magder LS, Via CS: Expression of cytokine- and chemokine-related genes in peripheral blood mononuclear cells from lupus patients by cDNA array. *Clin Immunol* 102:283–290, 2002
  78. Beitner-Johnson D, Seta K, Yuan Y, Kim H, Rust RT, Conrad PW, Kobayashi S, Millhorn DE: Identification of hypoxia-responsive genes in a dopaminergic cell line by subtractive cDNA libraries and microarray analysis. *Parkinsonism Relat Disord* 7:273–281, 2001
  79. Lock C, Hermans G, Pedotti R, Brendolan A, Schadt E, Garren H, Langer-Gould A, Strober S, Cannella B, Allard J, Klonowski P, Austin A, Lad N, Kaminski N, Galli SJ, Oksenberg JR, Raine CS, Heller R, Steinman L: Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. *Nat Med* 8:500–508, 2002
  80. MacDonald TJ, Brown KM, LaFleur B, Peterson K, Lawlor C, Chen Y, Packer RJ, Cogen P, Stephan DA: Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease. *Nat Genet* 29:143–152, 2001
  81. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JJ, 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, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403:503–511, 2000
  82. Ohmine K, Ota J, Ueda M, Ueno S, Yoshida K, Yamashita Y, Kirito K, Imagawa S, Nakamura Y, Saito K, Akutsu M, Mitani K, Kano Y, Komatsu N, Ozawa K, Mano H: Characterization of stage progression in chronic myeloid leukemia by DNA microarray with purified hematopoietic stem cells. *Oncogene* 20:8249–8257, 2001
  83. Loftus SK, Chen Y, Gooden G, Ryan JF, Birznieks G, Hilliard M, Baxevasis AD, Bittner M, Meltzer P, Trent J, Pavan W: Informatic selection of a neural crest-melanocyte cDNA set for microarray analysis. *Proc Natl Acad Sci U S A* 96:9277–9280, 1999
  84. 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, Sondak V: Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 406:536–540, 2000
  85. Clark EA, Golub TR, Lander ES, Hynes RO: Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* 406:532–535, 2000
  86. Gupta RA, Polk DB, Krishna U, Israel DA, Yan F, DuBois RN, Peek RM Jr: Activation of peroxisome proliferator-activated receptor gamma suppresses nuclear factor kappa B-mediated apoptosis induced by *Helicobacter pylori* in gastric epithelial cells. *J Biol Chem* 276:31059–31066, 2001
  87. Yanagawa R, Furukawa Y, Tsunoda T, Kitahara O, Kameyama M, Murata K, Ishikawa O, Nakamura Y: Genome-wide screening of genes showing altered expression in liver metastases of human colorectal cancers by cDNA microarray. *Neoplasia* 3:395–401, 2001
  88. Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Kallioniemi OP, Wilfond B, Borg A, Trent J: Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 344:539–548, 2001
  89. Van T Veer LJ, Dai H, Van De Vijver MJ, He YD, Hart A, Mao M, Peterse HL, Van Der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415:530–535, 2002

90. Hippo Y, Taniguchi H, Tsutsumi S, Machida N, Chong JM, Fukayama M, Kodama T, Aburatani H: Global gene expression analysis of gastric cancer by oligonucleotide microarrays. *Cancer Res* 62:233–240, 2002
91. Okabe H, Satoh S, Kato T, Kitahara O, Yanagawa R, Yamaoka Y, Tsunoda T, Furukawa Y, Nakamura Y: Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: Identification of genes involved in viral carcinogenesis and tumor progression. *Cancer Res* 61:2129–2137, 2001
92. Bubendorf L, Kolmer M, Kononen J, Koivisto P, Mousset S, Chen Y, Mahlamaki E, Schraml P, Moch H, Willi N, Elkhoulou AG, Pretlow TG, Gasser TC, Mihatsch MJ, Sauter G, Kallioniemi OP: Hormone therapy failure in human prostate cancer: analysis by complementary DNA and tissue microarrays. *J Natl Cancer Inst* 91:1758–1764, 1999
93. Mousset S, Wagner U, Chen Y, Kim JW, Bubendorf L, Bittner M, Pretlow T, Elkhoulou AG, Trepel JB, Kallioniemi OP: Failure of hormone therapy in prostate cancer involves systematic restoration of androgen responsive genes and activation of rapamycin sensitive signaling. *Oncogene* 20:6718–6723, 2001
94. Skubitz KM, Skubitz AP: Differential gene expression in renal-cell cancer. *J Lab Clin Med* 140:52–64, 2002
95. Khan J, Simon R, Bittner M, Chen Y, Leighton SB, Pohida T, Smith PD, Jiang Y, Gooden GC, Trent JM, Meltzer PS: Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res* 58:5009–5013, 1998
96. Mok SC, Chao J, Skates S, Wong K, Yiu GK, Muto MG, Berkowitz RS, Cramer DW: Prostatein, a potential serum marker for ovarian cancer: Identification through microarray technology. *J Natl Cancer Inst* 93:1458–1464, 2001
97. Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA: Microarrays and toxicology: The advent of toxicogenomics. *Mol Carcinog* 24:153–159, 1999
98. Waring JF, Halbert DN: The promise of toxicogenomics. *Curr Opin Mol Ther* 4:229–235, 2002