0013-7227/02/$15.00/0
*Printed in U.S.A.*

Endocrinology 143(6):1983–1989
Copyright © 2002 by The Endocrine Society

# Perspective: Microarray Technology, Seeing More Than Spots

### *Introduction*

The publication in the mid-1990s of a new molecular tool, the DNA microarray, has led to a revolution in the way scientists approach the investigation of gene expression and regulation. This technology has made its impact upon many basic scientific disciplines including cancer biology, developmental biology, toxicology, investigation of growth-factor and hormonal signaling, and the applied areas of disease diagnostics and drug development. The ability to assay thousands of genes simultaneously in a high-throughput manner across RNA samples derived from various biological sources and treatments has increased the need for integration of higher-order statistical analyses and data management schema into molecular biology laboratories. Many investigators are struggling with interpreting large, complex datasets, communicating their findings, and developing strategies for following up numerous potential leads indicated in datasets. This perspective provides an overview of the technology and its impact, as well as strategies for study design, data analysis, and avoidance of potential pitfalls.

### *Overview of microarray technology*

In the mid-1990s, a multidisciplinary team at Stanford University published a landmark paper initiating a new era in the realm of gene expression studies (1). This paper described a new technology allowing the quantitative, simultaneous monitoring of the expression of thousands of genes using a new tool termed a DNA microarray. Since then, numerous papers have been published reviewing the technology and its potential applications for scientific discovery in a broad range of biological disciplines. While reviewing all of the literature and findings related to this exciting technology is not within the scope of this perspective, a broad overview of the technology will be provided with a focus on study design and data analysis. Thus, investigators who have not yet ventured into this new world of the DNA chip will have a guide to develop a satisfying microarray study.

Although the DNA microarray chip is a tool most commonly used to monitor the level of expression of a gene at the RNA level (1–3), it also documents DNA copy number (4–6) and DNA protein interactions (7, 8) and sequencing applications (polymorphism detection). This perspective concentrates on the application of DNA microarrays for monitoring the expression of a gene by measurement of its ability to hybridize to a target sequence localized to a specific region on a chip. To measure this hybridization, RNA, extracted from a biological sample of interest, is reverse-transcribed into cDNA that ideally represents a quantitative copy of genes expressed at the time of sample collection. This cDNA is labeled with a tracking molecule such as a radioactive or a fluorescent nucleotide, or an affinity molecule like biotin. The labeled cDNA is then hybridized to the DNA chip that contains thousands of gene targets. Ideally, each molecule in the labeled cDNA will only bind to its appropriate complementary target sequence on the array. Quantitative imaging coupled with clone database information allows measurement of the amount of labeled cDNA that hybridized to each target sequence, resulting in the identification and relative quantification of the genes expressed in the original biological sample (3, 9, 10).

Several varieties of the DNA microarray are used frequently throughout the scientific community. These include deposition or spotted cDNA, spotted oligomer, or synthesized oligomer chips. The cDNA microarray is comprised of a collection of partial gene sequences that are spotted individually into precise locations within the DNA chip (11). These sequences usually range in size from 500–2000 bp and may be chosen from different regions of the gene depending on the goal of the project. The selection of sequences from the 3′ untranslated region of the gene confers the advantage of specificity to the genes being measured. Multiple genes within a conserved functional family may have a high degree of sequence similarity, especially in domains responsible for catalytic functions. When using a microarray to monitor gene expression, ideally the investigator must detect the precise gene that is affected without cross-reactivity with other family members. By selecting a sequence that contains a majority of 3′ untranslated region, a researcher can take advantage of sequence diversity in this region that may be gene specific and not conserved among related family members. Furthermore, in recent applications, investigators have selected promoter sequences to generate custom chips to monitor binding to specific promoter regions (7, 8). While the long target sequences of cDNA microarrays provide an advantage for sensitivity of detection of gene expression, the potential problems with specificity must be considered (12). If there is significant sequence overlap between clones representative of multiple genes (>70%), such as in related enzyme families (*i.e.* kinases, cytochrome p450s, etc.), then the discrete transcript may not be specifically detected; rather, multiple, related sequences might be simultaneously detected (12). An additional advantage of these cDNA chips is their ease of use and attainability. Complete instructions on their manufacture within a laboratory are freely available (http://cmgm.stanford.edu/pbrown/mguide/nospecs2.html). The ability to manufacture chips within a research laboratory or an institute provides the advantages of flexibility and customization of design as necessitated by the scientific goal of the project.

Another form of DNA microarrays is based on deposition or on-chip synthesis of oligonucleotides for generation of targets. Several approaches to the manufacturing of these chips (13–15) lead to the same end result—a chip that con-

tains short oligomers ranging from 25–80 bases as the target sequences. While these shorter sequences can confer high specificity, they may have decreased sensitivity/binding compared with cDNA arrays. Users of these arrays compensate for sensitivity problems by employing multiple sequences for each gene. One disadvantage of these oligomer-based chips is their availability only from commercial manufacturers. The cost of custom development of an oligomer-based chip is beyond the budget of a majority of researchers; but more affordable, premade oligomer-based chips are available and may be sufficiently informative for investigators seeking to employ this type of array.

*Scientific impact of microarray studies*

The impact of the microarray technology has proven tremendous because it has enabled investigators to progress from studying the expression of one gene in several days to hundreds of thousands of gene expressions in a single day. For the first time, investigators can relatively quickly measure the expression of a complete genome (16–19) across a large number of environmental stimuli. This awe-inspiring technological breakthrough has the potential to impact some previously intractable scientific realms and aid in the elucidation of complex models and systems. Tracking the numbers of publications in PubMed reveals that, since 1995, over 2000 papers have included this technology. Scientists in the arena of cancer biology have been some of the most industrious at incorporating this technology into their studies (Fig. 1). While generous funding in this arena may have contributed to greater usage of microarrays, this technology is, nonetheless, ideally suited for the comparison of gene expression in various stages of tumor development (2, 20–24) and monitoring expressions of premalignant and tumorigenic cells following exposure to anticancer agents or tumor promoters (25–32). Several laboratories (25, 33) demonstrated that gene expression profiling of patient lymphoma samples improved distinction of tumor classification and provided insight on clinical outcome. These and other data indicate that one important benefit of this technology might be to inform clinicians of better, specific markers for cancer diagnosis, prognosis and treatment (23–25, 31, 34).

Other basic scientific disciplines have also benefited from global gene expression profiling. Investigations using microarrays have provided a more comprehensive view on the complex regulation of the cell cycle (16, 35), differentiation (36–38), and aging (39–41). Other studies involving organisms associated with infectious or other noncancerous diseases have elucidated alterations in gene expression that may prove beneficial in treatment (42–44). An understanding of the mechanism of action of pharmaceutical agents, their potential adverse outcomes, and the prediction of activity of unknown agents will undoubtedly significantly impact the development of safer and more efficacious drugs (45, 46).

*General considerations for study design and data analysis*

While microrarrays allow scientists to gaze across the genome at the response of an organism to a biological stimulus, they have an equal potential to mislead if false assumptions are made or flaws in study design are overlooked. At present, there is no one standard guide or consensus for the design and conduct of an expression-profiling experiment because the ultimate goal of the project dictates the study design. Some considerations, however, are worth noting (Table 1).

The data from a single microarray hybridization, while containing a plethora of valuable information, will also be filled with many potential false positives and negatives resulting from technical processing steps or part of the data analysis process. To achieve confidence in a dataset and the most sensitive limit of detection, replicate hybridizations—both biological and chip—should be analyzed (47–50). In all disciplines of the biological sciences, including gene profiling, the reproducibility of an observation has been and must

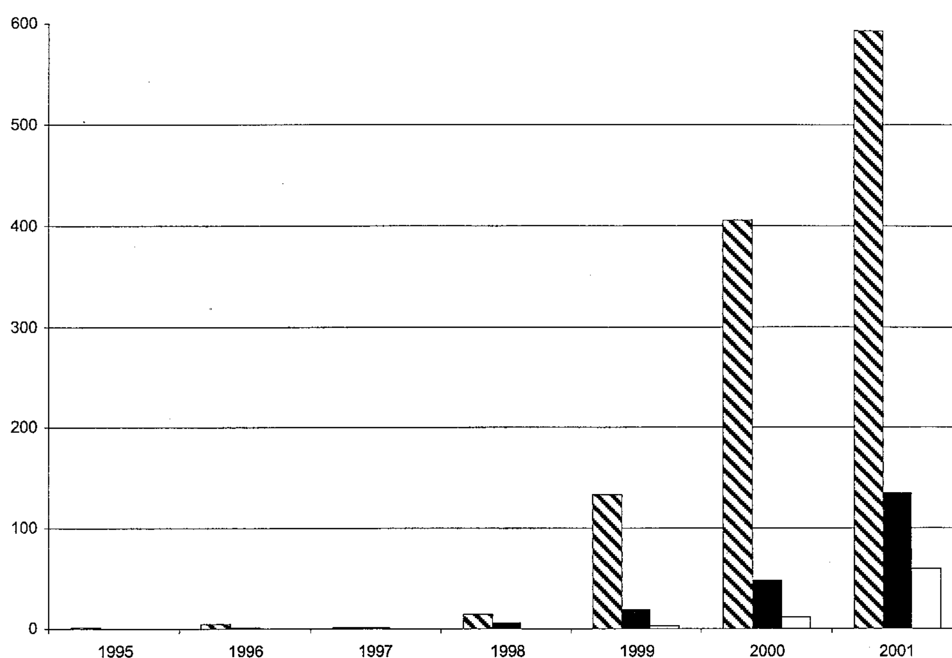


FIG. 1. Increased numbers of scientific studies employing microarray technology. A search of Medline by the following key words revealed a growing number of publications referencing microarray technology. Search terms were: DNA microarray (▧), DNA microarray and cancer (■), and DNA microarray and hormone (□).

**TABLE 1.** Important considerations in the design of a microarray experiment

| |
|---|
| Biological parameters |
| Sampling of appropriate number of replicates |
| Inclusion of other relevant samples, *i.e.* nonactive analogs, inhibitors, knockouts |
| Measurement of accompanying biological characteristics; physiology |
| Development of stringent treatment and harvest protocols |
| Technical parameters |
| Choice of chip, validation of identity of gene set |
| Hybridization of chip replicates, inclusion of dye swap |
| Choice of sample used as control in ratio analysis (universal, matched, other) |
| Scanning protocol that ensures proper laser power settings for optimal signals |
| Approach for validation of results |
| Data analysis parameters |
| Image analysis software, method of normalization and background subtraction |
| Data processing and long-term storage and backup |
| Data format, standardization, flexibility for import into multiple programs |
| Determination of appropriate visualization tools |
| Application of statistical approaches to determine gene significance |
| Gene annotation and ontology linkages |

continue to be a requisite for publication. A good study design will strategize to incorporate biological replicates into the chip analysis process; such inclusion is not apparent in all microarray-related publications. The use of biological replicates in the analysis schema enables detection of gene-expression changes that may be more related to environmental alterations than to the stimulus or biological parameter being monitored. In addition, examination of as many biological parameters as possible will enable the investigator to account for variations in experiments and obtain accurate results. For example, simultaneous measurements of DNA synthesis, cell cycle progress, viability, and apoptosis might be employed as standard measures to gauge the health and growth of cultured cells and ensure that these parameters are consistent within replicates. For clinical samples, notations on patient history and health, tissue handling upon necropsy or surgery, clinical chemistry, and histopathology will collectively aid in the ultimate interpretation of the gene-expression profiles by providing a phenotypic anchor or description of sample physiology that provides a biological context for the sample from which the RNA was derived (51). For example, these descriptions will provide additional information to aid in integration and interpretation of the gene expression profile as related to proliferation, apoptosis, necrosis, growth arrest, differentiation, or other physiological states.

### *Statistical and visualization considerations*

In the optimal study design, multiple chips for each biological sample should be assayed (50). This safeguard will provide the investigator the best dataset for balancing sensitivity with confidence. In the case of two-color fluorescent cDNA microarray hybridization, a dye swap should be conducted to filter out artifacts that could be attributed to unequal incorporation or quenching of the dye molecules.

Historically, many investigators have used an arbitrary fold cutoff as a measure of significance of a gene expression change. This approach poses many potential problems, including the gross overestimation of the number of significant gene changes in a dataset that contains a wide distribution of alterations (*e.g.* two diverse cell types like normal and cancerous tissue; Fig. 2A); or the severe underestimation of gene expression changes when a dataset is derived from a tightly linked comparison (Fig. 2B), as in the case of an early time point following hormonal stimulation (51a). Numerous, statistics-based, analytical approaches have been designed to gain the most sensitivity in detecting gene changes while simultaneously providing a measure of the potential error associated with the statistical method. Statistical measurements of significance can be employed at both the single- and multiarray levels (52–57).

The output from a microarray hybridization is usually a large data spreadsheet filled with numbers related to the distribution of the signal and background intensities for each gene on the chip. Most investigators convert these data to a more visual format for interpretation. Many visualization tools are available, both academically and commercially. One of the first tools applied to microarray data was clustering (58), which provides a two-dimensional hierarchical grouping of data points. In one dimension, assayed samples are grouped according to their similarity; in the second dimension, genes are grouped according to the overall similarity of their expression patterns across the samples. This type of analysis is useful for visualizing groups of genes that are similarly regulated across the biological samples under study (57, 59). Higher order computational approaches are needed to support the interpretation of data for discerning gene signatures or informative gene subsets. Methods such as genetic algorithms combined with supervised clustering (23, 31, 52, 57), principle component analysis, self-organizing maps, and linear discriminant analysis are all examples of analytical methods that might be used for discovering trends of expression across samples.

Contemplation of the ultimate research objective for the study prior in the context of visualization of microarray data will ensure that appropriate treatment groups are incorporated in the study design. For example, if an investigator wants to follow the signaling of a receptor-ligand interaction, inclusion of the profiling of a receptor-negative line, receptor pathway inhibitor, or a nonactive ligand to filter out potential effects not directly related to the activated pathway should allow clear determination of direct receptor mediated effects (Fig. 3). The use of a targeted treatment to help show biological specificity was demonstrated in a study (41, 60) in which the investigators wanted to examine effects of aging. They hypothesized that effects of metabolism and energy consumption contributed to the aging process in mice and used microarrays to determine age-associated gene-expression changes in these pathways. They benefited by adding a group of caloric-restricted mice in their study design and thus could filter their observed, age-associated gene-expression changes to determine precisely those genes that changed with normal aging and reversed in the extended-lifespan, calorie-restricted mice.

Gene expression changes are usually measured relative to

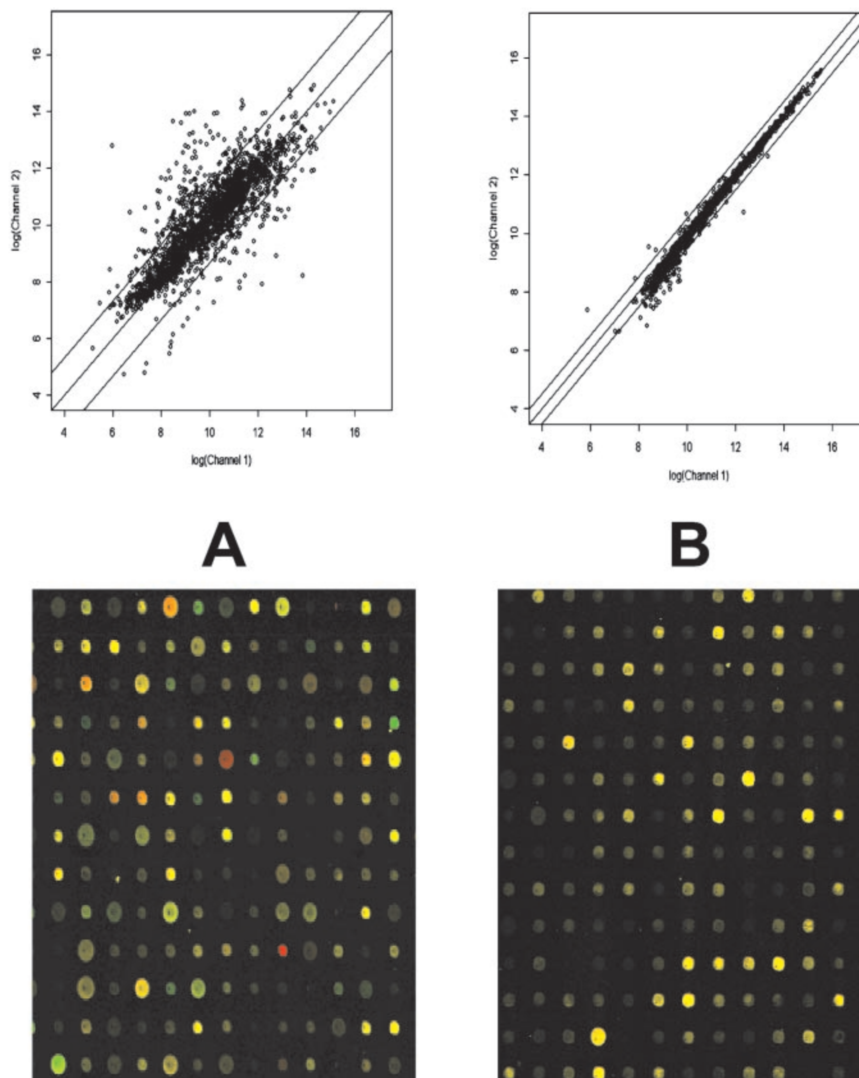


FIG. 2. Properties of data distribution in various sample comparisons. A, Distribution of nonrelated biological samples. The distribution of gene expression changes between nonrelated samples is wide. Use of a 99% confidence interval indicates a statistical significant cutoff at 2.51-fold. B, Distribution of related biological samples. The distribution of gene expression changes between closely related samples is very tight. Use of a 99% confidence interval indicates a statistical significant cutoff at 1.43-fold. Image analysis was conducted using Microarray suite of IP Labs (Scanalytics, Arlington, VA).

another sample. Therefore, another early consideration in the study design process is the strategy for pairing sample hybridizations, in the case of two-color cDNA arrays. (Traditionally in oligonucleotide arrays or arrays hybridized with radioactive probes, only one sample is hybridized per chip and sample pairing is made during the data analysis process.) Investigators have adopted several approaches in sample pairing. As in traditional experiments, many investigators use closely matched sample controls (*e.g.* vehicle or time matched). Other researchers choose to use universal control RNA derived from multiple samples that is designed to provide a baseline measurement on every gene being assayed (61). A third approach, employed when closely matched sample controls cannot be obtained, *e.g.* some clinical studies, uses a pooled-normal, or tissue-representative control (23). Each sample-pairing strategy has advantages and disadvantages, although matching as closely as possible the concurrent control to the sample under study may be most desirable (50). One advantage of the universal control is the ease of comparing all the samples assayed against it; however, comparison to datasets that were not analyzed against it may be difficult. In addition, the maintenance of a reproducible, universal RNA stock for an extended period of time is of paramount importance when employing this approach. Finally, a completely different hybridization design minimizes control measurements and implements a loop design that optimizes direct treatment comparisons (where sample A is compared with B, B is compared with C, C is compared with D, and D is compared with A).

### *Potential pitfalls*

Heralding the advantages of expression profiling must be fairly balanced by stating the limitations of the technology. Obviously, this is an expression-based technology, capable only of monitoring cellular responses at the RNA level. Some critical signaling changes may occur only at the protein or posttranslational level and therefore would not be detected with gene arrays. However, at this point microarray measurements of RNA provide the advantage of being able to use one biological sample to measure thousands of defined sequences simultaneously. Protein analyses sometimes require breaking one sample into multiple cellular fractions and cannot always provide definition of the moiety being measured.

FIG. 3. Illustration of how experimental design and visualization work together in data interpretation. Example of a clustering diagram generated from an experiment where tissue was isolated from an animal treated with a compound, a compound inhibitor, and dual treatment of compound and inhibitor together. *Red* signifies genes that are induced, and *green* indicates genes that are repressed by treatment relative to vehicle control treated animals. The clustering visualization indicates a group of genes that are highly induced by the treatment (shown in *bright red*) and then are attenuated by cotreatment with the inhibitor (shown by *less intense red color*), indicating that affects on these genes are directly related to the compound/inhibitor pathway.

However, as signaling maps and networks are more finely mapped, predicting potential protein changes or activation of signaling pathways based on upstream or downstream RNA changes should become possible.

One challenge in interpreting gene expression profiling data will be distinguishing between initial and secondary effects. An initial perturbation of a biological system will induce gene expression changes that will be followed by more alterations related to secondary, cellular changes. The dissection of the cause-and-effect relationship will be a challenge that will be possible by careful experimental design incorporating a broad range of time points or disease progression.

At the technical level, several pitfalls are easily circumvented by careful follow-up of microarray data. First, errors in sequence databases sometime lead to errors in gene annotation; therefore, confirmation of clone identification is essential. For cDNA microarray technology, routine resequencing may be advantageous in confirming the assumed identity of a gene of interest (62). Second, microarrays indicate a quantitative assessment of the level of gene expression, a value that for some genes may be imprecise (63). Qualitative interpretation may suffice for some investigations; however, if precise quantitation is important, as for certain clinical diagnostic applications, the level of gene expression should be validated by another technology. One may wish to employ real-time quantitative PCR for a high-throughput, quantitative measure (64). Alternately, Northern blots or ribonuclease protection assays provide the benefit of determining not only a quantitative measure, but also the number of potential transcripts detected with the chip sequence (65). Finally, gene-expression arrays can provide a sensitive measure of gene changes within a subpopulation of cells. We have found that microarrays can detect gene expression in merely 5% of the total population (66). This finding signifies that, if a cell culture or tissue is heterogeneous, significant gene changes may be related to only a small fraction of the cells; and investigators may wish to confirm localization with an *in situ* technique (67).

### *Future promise*

The future for microarrays is bright, as they will undoubtedly continue to be used in well-funded industry for high-throughput screening of compound targets, diagnostic development, and drug development (45, 46). As the cost of conducting experiments decreases, more academic investigators will include this technology in their arsenal of tools (68). Microarray analysis should not be considered as the conclusion to an experiment, but as a discovery mechanism to help determine which avenues to pursue. More variations of the classic microarray paradigm, like the search for promoter elements (7, 8) or screening of DNA/chromosomes for expressed genes (4–6, 69, 70), may arise. Scientific discoveries documented within the ever-expanding databanks will be facilitated by the adoption of standard data formats (71), common databases (72–79), more detailed global network mapping (80–86), and development of stronger interactions of biologists with computer scientists and statisticians. We can expect that the format of microarrays will change to become denser, more miniaturized, and technically standardized, and the desire to assay genome-wide cellular changes using this innovative technology will only be enhanced.

Cynthia A. Afshari
Laboratory of Molecular Carcinogenesis
National Institute of Environmental Health Sciences
Research Triangle Park, North Carolina 27709

## Acknowledgments

The author expresses her appreciation to Drs. Ed Lobenhofer, Hisham Hamadeh, Alex Merrick, Ben Van Houten, and Robert Maronpot for their critical reading and advice in the preparation of this article; to Jennifer Collins, Lee Bennett, and Sherry Grissom for preparation of the figures; and to JoAnne Johnson for detailed editing.

Received February 28, 2002. Accepted March 4, 2002.

Address all correspondence and requests for reprints to: Cynthia Afshari, Ph.D., National Institute of Environmental Health Sciences, P.O. Box 12233, MD2-04, 111 T. W. Alexander Drive, Research Triangle Park, North Carolina 27709. E-mail: afshari@niehs.nih.gov.

## References

1. **Schena M, Shalon D, Davis RW, Brown PO** 1995 Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470
2. **DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM** 1996 Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 14:457–460
3. **Shalon D, Smith SJ, Brown PO** 1996 A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res 6:639–645
4. **Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO** 1999 Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet 23:41–46
5. **Stephan DA, Chen Y, Jiang Y, Malechek L, Gu JZ, Robbins CM, Bittner ML, Morris JA, Carstea E, Meltzer PS, Adler K, Garlick R, Trent JM, Ashlock MA** 2000 Positional cloning utilizing genomic DNA microarrays: the Niemann-Pick type C gene as a model system. Mol Genet Metab 70:10–18
6. **Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, Burchard J, Dow S, Ward TR, Kidd MJ, Friend SH, Marton MJ** 2000 Widespread aneuploidy revealed by DNA microarray expression profiling. Nat Genet 25: 333–337
7. **Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA** 2000 Genome-wide location and function of DNA binding proteins. Science 290:2306–2309
8. **Ren B, Cam H, Takahashi Y, Volkert T, Terragni J, Young RA, Dynlacht BD** 2002 E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev 16:245–256
9. **Khan J, Saal LH, Bittner ML, Chen Y, Trent JM, Meltzer PS** 1999 Expression profiling in cancer using cDNA microarrays. Electrophoresis 20:223–229
10. **Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J** 2000 A concise guide to cDNA microarray analysis. Biotechniques 29:548–550; 552–556
11. **Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW** 1996 Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA 93:10614–10619
12. **Evertsz EM, Au-Young J, Ruvolo MV, Lim AC, Reynolds MA** 2001 Hybridization cross-reactivity within homologous gene families on glass cDNA microarrays. Biotechniques 31:1182–1186
13. **Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ** 1999 High density synthetic oligonucleotide arrays. Nat Genet 21:20–24
14. **Singh-Gasson S, Green RD, Yue Y, Nelson C, Blattner F, Sussman MR, Cerrina F** 1999 Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. Nat Biotechnol 17:974–978
15. **Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephaniants SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, Linsley PS** 2001 Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat Biotechnol 19:342–347
16. **Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B** 1998 Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol Biol Cell 9:3273–3297
17. **Brown PO, Botstein D** 1999 Exploring the new world of the genome with DNA microarrays. Nat Genet 21:33–37
18. **Roberts CJ, Nelson B, Marton MJ, Stoughton R, Meyer MR, Bennett HA, He YD, Dai H, Walker WL, Hughes TR, Tyers M, Boone C, Friend SH** 2000 Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science 287:873–880
19. **Jelinsky SA, Estep P, Church GM, Samson LD** 2000 Regulatory networks revealed by transcriptional profiling of damaged *Saccharomyces cerevisiae* cells: Rpn4 links base excision repair with proteasomes. Mol Cell Biol 20:8157–8167
20. **Khan J, Simon R, Bittner M, Chen Y, Leighton SB, Pohida T, Smith PD, Jiang Y, Gooden GC, Trent JM, Meltzer PS** 1998 Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. Cancer Res 58:5009–50013
21. **Alizadeh A, Eisen M, Davis RE, Ma C, Sabet H, Tran T, Powell JI, 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** 1999 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
22. **Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR** 2001 Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci USA 98:15149–15154
23. **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** 2001 Gene-expression profiles in hereditary breast cancer. N Engl J Med 344: 539–548
24. **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, Eystein Lonning P, Borresen-Dale AL** 2001 Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 98:10869–10874
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** 1999 Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286:531–537
26. **Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, Kohn KW, Reinhold WC, Myers TG, Andrews DT, Scudiero DA, Eisen MB, Sausville EA, Pommier Y, Botstein D, Brown PO, Weinstein JN** 2000 A gene expression database for the molecular pharmacology of cancer. Nat Genet 24:236–244
27. **Kudoh K, Ramanna M, Ravatn R, Elkahloun AG, Bittner ML, Meltzer PS, Trent JM, Dalton WS, Chin KV** 2000 Monitoring the expression profiles of doxorubicin-induced and doxorubicin-resistant cancer cells by cDNA microarray. Cancer Res 60:4161–4166
28. **Kihara C, Tsunoda T, Tanaka T, Yamana H, Furukawa Y, Ono K, Kitahara O, Zembutsu H, Yanagawa R, Hirata K, Takagi T, Nakamura Y** 2001 Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles. Cancer Res 61:6474–6479
29. **Staunton JE, Slonim DK, Coller HA, Tamayo P, Angelo MJ, Park J, Scherf U, Lee JK, Reinhold WO, Weinstein JN, Mesirov JP, Lander ES, Golub TR** 2001 Chemosensitivity prediction by transcriptional profiling. Proc Natl Acad Sci USA 98:10787–10792
30. **Zembutsu H, Ohnishi Y, Tsunoda T, Furukawa Y, Katagiri T, Ueyama Y, Tamaoki N, Nomura T, Kitahara O, Yanagawa R, Hirata K, Nakamura Y** 2002 Genome-wide cDNA microarray screening to correlate gene expression profiles with sensitivity of 85 human cancer xenografts to anticancer drugs. Cancer Res 62:518–527
31. **van 't Veer LJ, Dai H, van De Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van Der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH** 2002 Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530–536
32. **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** 2002 Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med 8:68–74
33. **Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson Jr J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM** 2000 Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503–511
34. **Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta KJ, Rubin MA, Chinnaiyan AM** 2001 Delineation of prognostic biomarkers in prostate cancer. Nature 23:412:822–826
35. **Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JC, Trent JM, Staudt LM, Hudson Jr J, Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO** 1999 The transcriptional program in the response of human fibroblasts to serum. Science 283:83–87
36. **Dietz AB, Bulur PA, Knutson GJ, Matasic R, Vuk-Pavlovic S** 2000 Maturation of human monocyte-derived dendritic cells studied by microarray hybridization. Biochem Biophys Res Commun 275:731–738
37. **Guo X, Liao K** 2000 Analysis of gene expression profile during 3T3-L1 preadipocyte differentiation. Gene 251:45–53
38. **Dupont J, Khan J, Qu BH, Metzler P, Helman L, LeRoith D** 2001 Insulin and IGF-1 induce different patterns of gene expression in mouse fibroblast NIH-3T3 cells: identification by cDNA microarray analysis. Endocrinology 142: 4969–4675
39. **Funk WD, Wang CK, Shelton DN, Harley CB, Pagon GD, Hoeffler WK** 2000 Telomerase expression restores dermal integrity to *in vitro*-aged fibroblasts in a reconstituted skin model. Exp Cell Res 258:270–278
40. **Lee HM, Greeley GH, Jr, Englander EW** 2001 Age-associated changes in gene

expression patterns in the duodenum and colon of rats. Mech Ageing Dev 122:355–371
41. **Lee C-K, Klopp RG, Weindruch R, Prolla TA** 1999 Gene expression profile of aging and its retardation by caloric restriction. Science 285:390–393
42. **Shaheduzzaman S, Krishnan V, Petrovic A, Bittner M, Meltzer P, Trent J, Venkatesan S, Zeichner S** 2002 Effects of HIV-1 Nef on cellular gene expression profiles. J Biomed Sci 9:82–96
43. **Kato-Maeda M, Gao Q, Small PM** 2001 Microarray analysis of pathogens and their interaction with hosts. Cell Microbiol 3:713–719
44. **Wilson M, DeRisi J, Kristensen HH, Imboden P, Rane S, Brown PO, Schoolnik GK** 1999 Exploring drug-induced alterations in gene expression in Mycobacterium tuberculosis by microarray hybridization. Proc Natl Acad Sci USA 96:12833–12838
45. **Afshari CA, Nuwaysir EF, Barrett JC** 1999 Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. Cancer Res 59:4759–4760
46. **Ulrich R, Friend SH** 2002 Toxicogenomics and drug discovery:will new technologies help us produce better drugs? Nature Rev Drug Disc 1:84–88
47. **Lee ML, Kuo FC, 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
48. **Pritchard CC, Hsu L, Delrow J, Nelson PS** 2001 Project normal: defining normal variance in mouse gene expression. Proc Natl Acad Sci USA 98:13266–13271
49. **Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS** 2001 Assessing gene significance from cDNA microarray expression data via mixed models. J Comput Biol 8:625–637
50. **Churchill GA, Oliver B** 2001 Sex, flies and microarrays. Nat Genet 29:355–356
51. **Tennant RW** 2002 EHP 110–1, 2002, Editorial: the national center for toxicogenomics: using new technologies to inform mechanistic toxicology. Environ Health Perspect 110:A8–A10

51a. **Lobenhofer EK, Bennett L, Cable PL, Li L, Bushel PR, Afshari CA,** Regulation of DNA replication fork genes by 17$\beta$-estradiol. Mol Endocrinol, in press
52. **Chen D, Chen Y, Hu S** 1997 A pattern classification procedure integrating the multivariate statistical analysis with neural networks. Comput Chem 21:109–113
53. **Kerr MK, Churchill GA** 2001 Statistical design and the analysis of gene expression microarray data. Genet Res 77:123–128
54. **Chen Y, Yakhini Z, Ben-Dor A, Dougherty E, Trent JM, Bittner M** 2001 Analysis of expression patterns: the scope of the problem, the problem of scope. Dis Markers 17:59–65
55. **Brazma A, Vilo J** 2001 Gene expression data analysis. Microbes Infect 3:823–829
56. **Bushel PR, Hamadeh H, Bennett L, Sieber S, Martin K, Nuwaysir EF, Johnson K, Reynolds K, Paules RS, Afshari CA** 2001 MAPS: a microarray project system for gene expression experiment information and data validation. Bioinformatics 17:564–565
57. **Quackenbush J** 2001 Computational analysis of microarray data. Nat Rev Genet 2:418–427
58. **Eisen MB, Spellman PT, Brown PO, Botstein D** 1998 Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95: 14863–14868
59. **Sherlock G** 2001 Analysis of large-scale gene expression data. Brief Bioinform 2:350–362
60. **Weindruch R, Kayo T, Lee CK, Prolla TA** 2001 Microarray profiling of gene expression in aging and its alteration by caloric restriction in mice. J Nutr 131:918S–923S
61. **Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D** 2000 Molecular portraits of human breast tumours. Nature 406:747–752
62. **Taylor E, Cogdell D, Coombes K, Hu L, Ramdas L, Tabor A, Hamilton S, Zhang W** 2001 Sequence verification as quality-control step for production of cDNA microarrays. Biotechniques 31:62–65
63. **Amundson SA, Bittner M, Chen Y, Trent J, Meltzer P, Fornace Jr AJ** 1999 Fluorescent cDNA microarray hybridization reveals complexity and heterogeneity of cellular genotoxic stress responses. Oncogene 18:3666–13672
64. **Walker NJ** 2001 Real-time and quantitative PCR: applications to mechanism-based toxicology. J Biochem Mol Toxicol 15:121–127
65. **Taniguchi M, Miura K, Iwao H, Yamanaka S** 2001 Quantitative assessment of DNA microarrays—comparison with Northern blot analyses. Genomics 71:34–39
66. **Hamadeh HK, Bushel P, Tucker CJ, Martin K, Paules R, Afshari CA** 2002 Detection of diluted gene expression alterations using cDNA microarrays. Biotechniques 32:322–329
67. **Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP** 1998 Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat Med 4:844–847
68. **Khan J, Bittner ML, Chen Y, Meltzer PS, Trent JM** 1999 DNA microarray technology: the anticipated impact on the study of human disease. Biochim Biophys Acta 1423:M17–M28
69. **Phillips JL, Hayward SW, Wang Y, Vasselli J, Pavlovich C, Padilla-Nash H, Pezullo JR, Ghadimi BM, Grossfeld GD, Rivera A, Linehan WM, Cunha GR, Ried T** 2001 The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. Cancer Res 61:8143–8149
70. **Sudbrak R, Wieczorek G, Nuber UA, Mann W, Kirchner R, Erdogan F, Brown CJ, Wohrle D, Sterk P, Kalscheuer VM, Berger W, Lehrach H, Ropers HH** 2001 X chromosome-specific cDNA arrays: identification of genes that escape from X-inactivation and other applications. Hum Mol Genet 10:77–83
71. **Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge 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** 2001 Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29:365–371
72. **Ermolaeva O, Rastogi M, Pruitt KD, Schuler GD, Bittner ML, Chen Y, Simon R, Meltzer P, Trent JM, Boguski MS** 1998 Data management and analysis for gene expression arrays. Nat Genet 20:19–23
73. **Kellam P** 2001 Microarray gene expression database: progress towards an international repository of gene expression data. Genome Biol 2:4011
74. **Sherlock G, Hernandez-Boussard T, Kasarskis A, Binkley G, Matese JC, Dwight SS, Kaloper M, Weng S, Jin H, Ball CA, Eisen MB, Spellman PT, Brown PO, Botstein D, Cherry JM** 2001 The Stanford microarray database. Nucleic Acids Res 29:152–155
75. **Geschwind DH** 2001 Sharing gene expression data: an array of options. Nat Rev Neurosci 2:435–438
76. **Gardiner-Garden M, Littlejohn TG** 2001 A comparison of microarray databases. Brief Bioinform 2:143–158
77. **Edgar R, Domrachev M, Lash AE** 2002 Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30: 207–210
78. **Dwight SS, Harris MA, Dolinski K, Ball CA, Binkley G, Christie KR, Fisk DG, Issel-Tarver L, Schroeder M, Sherlock G, Sethuraman A, Weng S, Botstein D, Cherry JM** 2002 Saccharomyces genome database (SGD) provides secondary gene annotation using the gene ontology (GO). Nucleic Acids Res 30:69–72
79. **Bono H, Kasukawa T, Hayashizaki Y, Okazaki Y** 2002 READ: RIKEN expression array database. Nucleic Acids Res 30:211–213
80. **Schulze-Kremer S** 1998 Ontologies for molecular biology. Pac Symp Biocomput 695–706
81. **Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G** 2000 Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29
82. **Karp PD** 2000 An ontology for biological function based on molecular interactions. Bioinformatics 16:269–285
83. **Karp PD** 2001 Pathway databases: a case study in computational symbolic theories. Science 293:2040–2044
84. **Masys DR, Welsh JB, Lynn Fink J, Gribskov M, Klacansky I, Corbeil J** 2001 Use of keyword hierarchies to interpret gene expression patterns. Bioinformatics 17:319–326
85. **Bono H, Kasukawa T, Furuno M, Hayashizaki Y, Okazaki Y** 2002 FANTOM DB: database of functional annotation of RIKEN mouse cDNA clones. Nucleic Acids Res 30:116–118
86. **Kanehisa M, Goto S, Kawashima S, Nakaya A** 2002 The KEGG databases at GenomeNet. Nucleic Acids Res 30:42–46