Am J Physiol Gastrointest Liver Physiol 282: G397–G402, 2002;
10.1152/ajpgi.00519.2001.

themes

Cutting-edge technology
I. Global gene expression profiling using DNA microarrays

MAYI ARCELLANA-PANLILIO AND STEPHEN M. ROBBINS
Southern Alberta Microarray Facility, Departments of Oncology and Biochemistry and Molecular Biology, Faculty of Medicine, University of Calgary, Calgary, Alberta T2N 4N1, Canada

Panlilio, Mayi Arcellana, and Stephen M. Robbins. Cutting-edge technology. I. Global gene expression profiling using DNA microarrays. *Am J Physiol Gastrointest Liver Physiol* 282: G397–G402, 2002; 10.1152/ajpgi.00519.2001.—Having the complete human genomic sequence poses a new challenge: to use genomic structural information to display and analyze biological processes on a genome-wide scale to assign gene function. DNA microarrays are a miniaturized, ordered arrangement of nucleic acid fragments from individual genes located at defined positions on a solid support, enabling the analysis of thousands of genes in parallel by specific hybridization. This review describes technical aspects, discusses relevant applications, and suggests factors affecting the use of this technology and how it fits in the grand scheme of meeting the needs of the postgenomic era.

genome; human; gene expression profiling; oligonucleotide array sequence analysis; genetic techniques

HAVING THE FULL HUMAN GENOME sequence in hand has been likened to being given a dictionary of words with very few definitions. The challenge of the postgenomic era is to find those missing definitions, to use genomic structural information to display and analyze biological processes on a genome-wide scale, and to assign gene function. Although molecular biology has traditionally taken a reductionist approach to biological questions, it has long been recognized that genes act in concert with other genes, often in separate dimensions of time and space. To fully understand the underlying biology, these molecular interactions have to be studied from the conceptual framework of the entire genome, the logistical challenges of which approach are being overcome with the advent of high throughput technologies. Although a definitive count of human genes must await further experimental and computational analysis, the human genome may be comprised of <40,000 genes, making the goal of elucidating the functional roles of all genes surprisingly attainable.

There are several technologies that can be used to define the gene expression profile of a particular cell type or tissue. These include differential screening of cDNA libraries (15), subtractive cDNA hybridization (24), differential display of RNA (17), serial analysis of gene expression (23), and cDNA/oligonucleotide microarrays (http://genetics.nature.com/web_specials). All of these techniques have intrinsic advantages and limitations, but for the purposes of this review, we will focus on DNA microarrays, because they currently provide the most effective way of comparing numerous samples for the expression of thousands of genes.

BACKGROUND OF DNA MICROARRAYS

DNA microarrays may be defined as a miniaturized, ordered arrangement of nucleic acid fragments derived from individual genes located at defined positions on a solid support, enabling the analysis of thousands of genes in parallel by specific hybridization. The choice of which genes should be represented in a given microarray tends to run from the global (entire genome on a slide) to the specific (such as a certain pathway or cell type). In general, DNA microarrays are of two kinds, depending on the material arrayed: cDNA or oligonucleotide. For cDNA arrays, the nucleic acid fragments are spotted robotically, using protocols pioneered at Stanford Univ. (http://cmgm.stanford.edu/pbrown/mguide/). The cDNAs used for spotting are usually derived by PCR amplification of cDNA libraries. For oligonucleotide arrays, in situ synthesis produces short 20–25 mers by photolithography (Affymetrix; www.affymetrix.com) or lengths of up to 60 nt by inkjet technology (Agilent Technologies; DNA_microarray@agilent.com). Both of these technologies were developed originally for the computer industry and have

Address for reprint requests and other correspondence: S. M. Robbins, Faculty of Medicine, Univ. of Calgary, 3330 Hospital Drive, N.W., Calgary, Alberta T2N 4N1, Canada (E-mail: srobbins@ucalgary.ca).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

been adapted for the manufacture of microarrays. A third type of oligonucleotide array is made by spotting presynthesized oligonucleotides on glass slides. The oligonucleotides in this case are usually longer (~70 nt) than their in situ synthesized counterparts.

Whether spotting cDNAs or presynthesized oligonucleotides, the nucleic acid fragments are deposited onto substrate-coated glass slides using highly precise x-y-z robotic systems to produce arrays with thousands of elements, each ~100 μm in diameter, spotted within an area of a few square centimeters. The main advantage of these dense arrays is their requirement for small volumes for hybridization and proportionately smaller amounts of sample for analysis. The fundamental aspects of working in a small format at high density and even the idea of specific binding to immobilized material as a quantitative measure are by no means unique nor original to DNA microarrays, given examples in the chemistry field of the microspot ligand assays developed by Ekins and Chu (8).

In 1995, Brown and his colleagues at Stanford Univ. published the first paper on DNA microarrays (19), describing them as a high-capacity system developed to monitor RNA levels of numerous genes simultaneously, using two-color immunofluorescence. These investigators were able to rapidly expand the technology from the original analysis of 45 Arabidopsis genes (19) to study gene expression in human samples by hybridization to DNA microarrays with over 1,000 elements (7, 20). It has since become evident that given current estimates of the total number of human genes, the entire genome might be analyzed in a single experiment.

The study of gene expression using microarrays is based on the competitive hybridization of differently labeled populations of cDNAs. Fluorescent dyes, usually Cy3 and Cy5, are used to distinguish cDNA pools reverse transcribed from different mRNA samples that have been isolated from cells or tissues. The labeled cDNAs are applied to the microarray and allowed to hybridize under conditions analogous to those established for Southern blotting. After the slide is washed to remove nonspecific hybridization, it is read in a confocal laser scanner that can differentiate between Cy3- and Cy5-signals, collecting fluorescence intensities to produce a separate 16-bit TIFF image for each channel. Image-processing software is used to find the spots and quantify them, which are not trivial operations, involving considerations such as how background corrections should be made. The relative intensities obtained for each channel are then normalized to adjust for differences in labeling and detection efficiencies so that the two data sets become comparable and ratios of intensity for each spot can be calculated.

Early cDNA-labeling protocols called for including fluorescently tagged nucleotides in the reverse transcription reaction. It was observed, however, that the fluorescent moieties conferred different capacities for incorporation in the growing DNA strand, resulting in a dye bias that was partially resolved by running reciprocal labeling experiments, facetiously referred to as "fluor flips." More recently, an indirect labeling procedure addressed the dye-bias problem by separating the incorporation of label from the DNA synthesis. In this protocol, amino allyl nucleotide is added to the reverse transcription reactions of individual sample RNAs to produce cDNAs that uniformly incorporate the modified nucleotide, then monoreactive dye is coupled with the modified nucleotide in a chemical reaction that presumably goes to completion, thereby producing cDNAs with similar levels of label. A representative example of a microarray experiment comparing the patterns of gene expression between two RNA samples is depicted in Fig. 1.

DNA microarrays are being used extensively to generate comprehensive data on gene expression patterns in a wide variety of tissues, under different experimental conditions and contexts, to uncover new players, and reveal new molecular mechanisms underlying these expression patterns. Although often criticized as a "fishing expedition," these experiments lead to the formulation of new hypotheses that can be tested using microarrays in an iterative strategy more reminiscent of the classic scientific method than of the global screening strategy invariably applied at the outset. True to the scientific method, this systematic acquisition of information results in an ongoing refinement of existing models of gene interaction and a closer and closer approximation of natural truth.

To fully realize the potential of microarrays, the information obtained must be of consistently high quality, because poor data will lead to poor biological conclusions. The minimum information about a microarray experiment standards, established by the Microarray Gene Expression Databases group (http://www.mged.org/), designed "to specify the minimum information that must be reported about a microarray (or any DNA array) based gene expression monitoring experiment to ensure the interpretability, as well as potential verification of the results by third parties," are the result of ongoing collaborative efforts to guard the quality of reported data. To compare gene expression profiles across a large number of samples, it is most efficient to use a standard reference RNA for all experiments. This reference would be chosen based on the experimental design and may be specific to the project. On the other hand, comparing expression profiles from different laboratories would require a universal reference RNA. There is not a real consensus among researchers in the field as to what that universal reference should be, although there are universal reference RNAs commercially available even now (http://www.stratagene.com/).

Microarray experiments generate enormous amounts of data that require the development of software for data acquisition, quality scoring, and tracking of data points. Ultimately, archived data are useful only as far as they can be analyzed and interpreted, and, therefore, data analysis software is absolutely essential. There are numerous commercial packages available including ArrayPro (Media Cybernetics), ArrayStat (Imaging Research), and QuantArray (PerkinElmer)

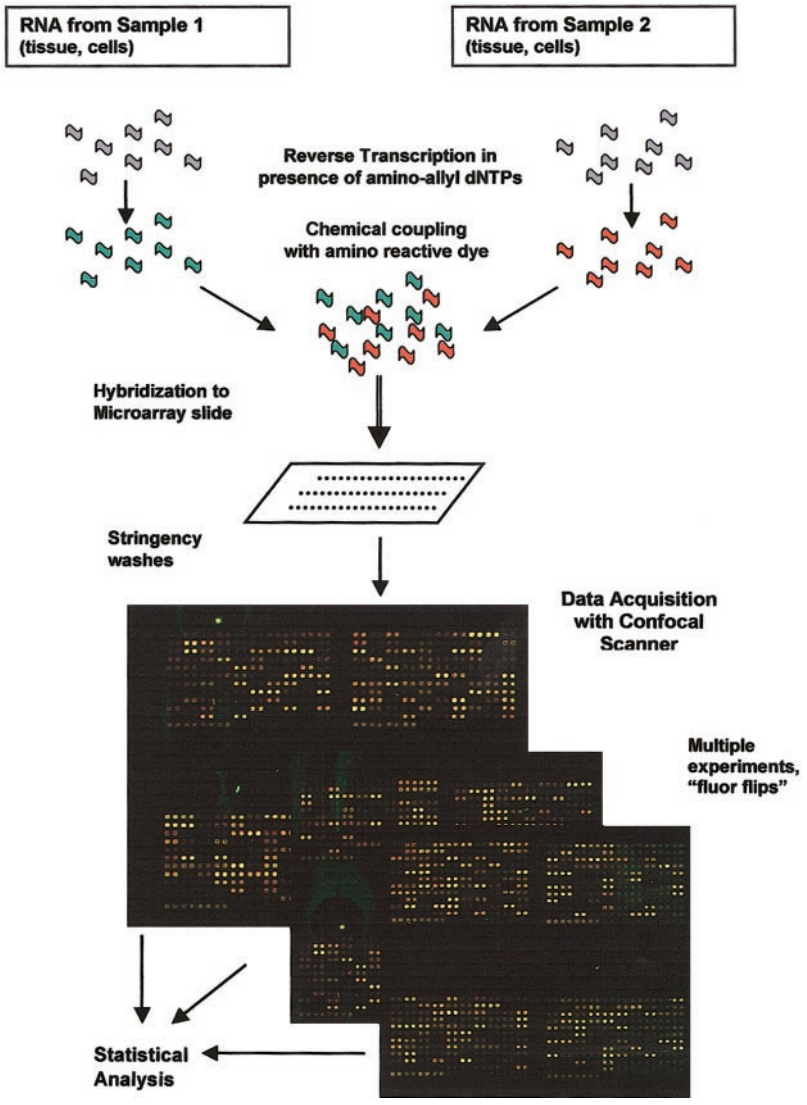

Fig. 1. Schema of a DNA microarray experiment showing how RNA from 2 samples are reverse transcribed and labeled with different fluors, hybridized to the microarrays slide, washed, and scanned to acquire data. The tiled scan images indicate the necessity for replicates and multiple experiments, producing data amenable to analysis.

for single experiment analysis, and others such as Gene Traffic (Iobion), GeneSpring (Silicon Genetics), and Resolver (Rosetta Inpharmatics) for analysis across experiments. There are also public sources of software, such as ScanAlyze, Cluster, and TreeView, which are all available for downloading from the Lawrence Berkeley National Laboratory (http://rana.lbl.gov/). The vast amount of data generated from these DNA microarray experiments may be a biologist's nightmare; however, it is a bioinformatician's dream.

Although originally designed to monitor patterns of gene expression, as clues to gene function and cell state (19, 20), DNA microarrays have since been used to study other indicators of gene action, including the genome-wide quantification of DNA copy number (21), the characterization of genetic mutation (11), and the detection of single-nucleotide polymorphisms (12).

APPLICATIONS OF DNA MICROARRAYS

Of the over 1,500 references that have cited the use microarray technology in the new millennium, a major proportion has some relation to cancer. DNA microarrays have helped to define molecular features of cancer progression as well as to distinguish between the nonmetastatic and the metastatic phenotype, such as in melanoma (6) and medulloblastoma (18). Gene expression profiling using cDNA or oligonucleotide microarrays has permitted the molecular characterization of a number of different cancers into more specialized subgroups (1a). Cancers of similar histopathology often respond differently to a given therapeutic protocol and have divergent clinical outcomes suggesting that these cancers are, in fact, different. Comparing their patterns of global gene expression is a way to identify those differences. It may be overly simplistic to believe that we can obtain a genetic blueprint that defines a given tumor cell type because of the complex and inherently variable nature of tumors. Nevertheless, these types of experiments are providing a framework around which to build a hierarchy of genetic and clinical information for any given cancer. There is already some evidence of the feasibility of this approach. Com-

puter-based algorithms, such as artificial neural networks, have been applied successfully to microarray data to decipher gene-expression signatures to aid diagnostic classification of small, round, blue-cell tumors and identify targets for therapy (16).

We will now focus on a few examples of microarray applications more germane to the readership of this journal. Although the idea of using this technology is still in its infancy, cDNA/oligonucleotide arrays have been used in physiology research. It should be emphasized, however, that most of these studies profile the expression of a few thousand genes by hybridization to arrays spotted onto Nylon filters. Because the equipment requirements for the use of filter arrays are found in a typical molecular biology laboratory, first forays into the field of microarrays have often been to study gene expression in these somewhat less dense arrays of a few thousand genes spotted within an area that is an order of magnitude larger than in glass slide arrays.

Drug-induced hepatotoxicity and liver damage cause significant morbidity and mortality and represent a major challenge in the area of drug development and design. Despite an extensive body of literature describing various hepatotoxicants, the fundamental mechanisms of hepatotoxicity remain largely unknown. There are now a number of studies that address this issue, showing changes in gene expression associated with toxic exposure such as that to carbon tetrachloride (14) and dioxins (9). It is worth noting that these studies rely quite heavily on the use of cultured hepatoma cells, HepG2, which may not be the most appropriate setting. These studies nevertheless illustrate the potential that microarrays may help to unravel underlying molecular mechanisms. This technology may also be appropriate for monitoring liver toxicity to various drug candidates in animal model systems (10).

DNA microarrays are being analyzed to understand the innate immune system, including the cellular response to various bacterial endotoxins. Despite the availability of antibiotics, bacteremia remains a serious health problem and is a leading cause of death in intensive care units. For gram-negative bacteria, these effects are mediated largely by the bacterial endotoxin lipopolysaccharide (LPS). It is clear from numerous studies that LPS can induce the expression of a number of genes, including adhesion molecules, chemokines, and other proinflammatory cytokines (1). However, clinical trials to abrogate the biological effects of many of these mediators have met with little success. DNA microarrays are now being used to identify other genes that may be regulated in response to endotoxins, such as LPS, with the ultimate hope of providing new drug targets for sepsis (25). With the use of human umbilical vein endothelial cells as a model, investigators found various genes that appear to be regulated in response to LPS. Many of the genes identified confirm already well-established LPS-regulated genes, whereas a few may represent new gene targets. These studies evaluated the expression of ~4,000 genes, suggesting that many other targets may remain to be identified. It should also be mentioned that gene expression profiles investigating the effects of different bacterial endotoxins are showing potential differences in the cellular response to various bacteria.

Microarray technology has also been used to monitor changes in metabolic processes including aging. Although it is clear that normal cells have a finite replicative lifespan, the mechanisms that control lifespan and the age-related phenotypes associated with it are not well understood. In a recent study by Tollet-Egnell and colleagues (22), DNA microarrays were used to analyze changes in gene expression in rat liver during aging, in the presence or absence of hormone replacement [growth hormone (GH)]. In their study, they observed that of the 3,000 genes examined, 1,000 were expressed in the rat liver and 47 of these showed altered expression during the aging process. Most of the affected genes encoded proteins involved in intermediary metabolism, mitochondrial respiration, and drug metabolism. Interestingly, a large number (~40%) of these transcripts, most of which had never before been linked to GH control, behaved normally in the presence of the hormone replacement. It is worth noting that several of the same genes seemed to be altered during the cellular senescence of fibroblasts. Many of those found to be upregulated during cellular senescence mimic inflammatory wound repair processes. Thus these microarray studies may influence how clinicians advise patients about age-related diseases and lifestyle decisions that have to be made with respect to the consequences of various physiological processes, including aging.

Global gene expression profiling by microarrays may be useful in evaluating population-based studies such as those investigating the impact of nutritional status. The nutrient zinc, for example, affects a wide variety of physiological processes including those necessary for growth and proper immune function. Cousins and colleagues (3) previously identified some zinc-regulated genes from the small intestine, using a combination of subtractive hybridization and differential display, and more recently, they demonstrated the potential usefulness of DNA filter arrays to profile global gene expression in rodents in the early stages of zinc deficiency. It is clear that nutritional status has pleiotropic effects, regulating gene function in diverse pathways that include cellular signaling, growth, transcription, redox, and energy metabolism. The effects of caloric restriction on gene expression in the liver of mice have also been studied using DNA microarray expression analysis (5).

The task of determining whether the regulation of many of these genes is physiologically relevant is made more daunting because evaluating gene expression on a global scale may reveal unlikely candidate genes and/or pathways that had not been considered before. Thus far, array data have provided investigators with a cataloguing of various differences in gene expression within their biological system. Initially, these differences in gene expression provide little mechanistic insight to explain the associated phenotypes, and tremendous effort will be required to assign specific bio-

logical consequences of the gene profiles revealed for a given cell type or in response to particular stimuli. However, there is more to this than simply slugging through the validation of hypothesized interactions, which might be achieved by buying bigger and better machines or by hiring more graduate students. A large measure of effort will be required to modify how seemingly illogical results are construed. This is not to say that every red herring ought to be investigated but that an open mind be maintained, because only then will it be receptive to whatever wondrous revelations might be in the offing.

PRESENT AND FUTURE CHALLENGES OF DNA MICROARRAYS

Notwithstanding the enormous power of this technology, the appreciation of its shortcomings is critical so as not to ask questions of it that cannot be answered in a valid manner. One limitation of microarrays is the tendency to pick up on genes that are more abundantly expressed, which may be a technical hurdle that simply needs to be overcome by improved methods of labeling and signal detection. Related to this issue is the observation that despite the miniaturization of microarrays, a considerable amount of sample is still required such that protocols will recommend labeling cDNA equivalent to 10 μg of total RNA, which may be trivial in some systems but difficult to achieve in others, such as for microdissected or flow cytometer sorted cells. A whole area of research focusing on obtaining valid data from exceedingly small amounts of material has arisen, leading to the development of such methods as for in vitro amplification of RNA (2), reverse transcription from single cells (4), and the use of tyramide-based techniques that can amplify fluorescent signals up to 1,000-fold (13).

Although probably true in a global sense, the assumption of the close correlation between RNA expression and gene function must nevertheless be tested for genes that become candidates for further study. The apparent levels of RNA expression should be confirmed independently either by Northern hybridization, quantitative PCR methods, or RNAse protection assays. The complementary methods of analysis, such as in situ hybridization and immunohistochemistry, will continue to be valuable in providing spatial information likely to surpass even the most patient practice of microdissection of cells from tissue.

From the point of view of the technology, microarrays seem to be headed toward becoming more comprehensive, not only increasing the number of elements that can be arrayed, but expanding the kinds of materials that are arrayed, such as the spotting of peptides and other small molecules at one end and the arraying of microsections of cells and tissue at the other. At first blush, these trends appear to be driven solely by technology, and perhaps the cynic would continue to see these as indicative of a technology testing its limits purely for the fun (and consequential financial gain) of it. We propose that these developments are driven by a realization that gene expression, however global, is only part of what defines a gene and that it will be necessary to integrate information obtained from different realms, such as from genomic structure, RNA expression profiles, protein interaction mapping, and compartmentalization of gene function in time and space, to formulate increasingly meaningful hypotheses that can be tested at several fronts and, finally, to arrive at truly defining those words in the dictionary of the human genome.

We apologize that many references were not included due to space limitations.

S. M. Robbins holds a Canada Research Chair in Cancer Biology and is a Senior Scholar of the Alberta Heritage Foundation for Medical Research. Both authors direct the Southern Alberta Microarray Facility that receives funding from the Alberta Cancer Board.

REFERENCES

1. **Aderem A and Ulevitch RJ.** Toll-like receptors in the induction of the innate immune response. *Nature* 406: 782–787, 2000.
1a. **Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, and Staudt LM.** Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403: 503–511, 2000.
2. **Baugh L, Hill AA, Brown EL, and Hunter CP.** Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res* 29: E29, 2001.
3. **Blanchard RK, Moore JB, Green CL, and Cousins RJ.** Inaugural article: modulation of intestinal gene expression by dietary zinc status: effectiveness of cDNA arrays for expression profiling of a single nutrient deficiency. *Proc Natl Acad Sci USA* 98: 13507–13513, 2001.
4. **Brail LH, Jang A, Billia F, Iscove NN, Klamut HJ, and Hill RP.** Gene expression in individual cells: analysis using global single cell reverse transcription polymerase chain reaction (GSC RT-PCR). *Mutat Res* 406: 45–54, 1999.
5. **Cao SX, Dhahbi JM, Mote PL, and Spindler SR.** Genomic profiling of short- and long-term caloric restriction effects in the liver of aging mice. *Proc Natl Acad Sci USA* 98: 10630–10635, 2001.
6. **Clark EA, Golub TR, Lander ES, and Hynes RO.** Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* 406: 532–535, 2000.
7. **DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, and Trent JM.** Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 14: 457–460, 1996.
8. **Ekins RP and Chu FW.** Multianalyte microspot immunoassay–microanalytical "compact disk" of the future. *Clin Chem* 37: 1955–1967, 1991.
9. **Frueh FW, Hayashibara KC, Brown PO, and Whitlock JP Jr.** Use of cDNA microarrays to analyze dioxin-induced changes in human liver gene expression. *Toxicol Lett* 122: 189–203, 2001.
10. **Gerhold D, Lu M, Xu J, Austin C, Caskey CT, and Rushmore T.** Monitoring expression of genes involved in drug metabolism and toxicology using DNA microarrays. *Physiol Genomics* 5: 161–170, 2001.
11. **Hacia JG, Brody LC, Chee MS, Fodor SP, and Collins FS.** Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat Genet* 14: 441–447, 1996.
12. **Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, Brody LC, Wang D, Lander ES, Lipshutz R, Fodor SP, and Collins FS.** Determination of ancestral alleles for human single-nucleotide

polymorphisms using high-density oligonucleotide arrays. *Nat Genet* 22: 164–167, 1999.
13. **Heiskanen MA, Bittner ML, Chen Y, Khan J, Adler KE, Trent JM, and Meltzer PS.** Detection of gene amplification by genomic hybridization to cDNA microarrays. *Cancer Res* 60: 799–802, 2000.
14. **Holden PR, James NH, Brooks AN, Roberts RA, Kimber I, and Pennie WD.** Identification of a possible association between carbon tetrachloride-induced hepatotoxicity and interleukin-8 expression. *J Biochem Mol Toxicol* 14: 283–290, 2000.
15. **Hoog C.** Isolation of a large number of novel mammalian genes by a differential cDNA library screening strategy. *Nucleic Acids Res* 19: 6123–6127, 1991.
16. **Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, and Meltzer PS.** Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 7: 673–679, 2001.
17. **Liang P and Pardee AB.** Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257: 967–971, 1992.
18. **MacDonald TJ, Brown KM, LaFleur B, Peterson K, Lawlor C, Chen Y, Packer RJ, Cogen P, and Stephan DA.** Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease. *Nat Genet* 29: 143–152, 2001.
19. **Schena M, Shalon D, Davis RW, and Brown PO.** Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467–470, 1995.
20. **Schena M, Shalon D, Heller R, Chai A, Brown PO, and Davis RW.** Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* 93: 10614–10619, 1996.
21. **Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D, and Albertson DG.** Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet* 29: 263–264, 2001.
22. **Tollet-Egnell P, Flores-Morales A, Stahlberg N, Malek RL, Lee N, and Norstedt G.** Gene expression profile of the aging process in rat liver: normalizing effects of growth hormone replacement. *Mol Endocrinol* 15: 308–318, 2001.
23. **Velculescu VE, Zhang L, Vogelstein B, and Kinzler KW.** Serial analysis of gene expression. *Science* 270: 484–487, 1995.
24. **Welcher AA, Torres AR, and Ward DC.** Selective enrichment of specific DNA, cDNA and RNA sequences using biotinylated probes, avidin and copper-chelate agarose. *Nucleic Acids Res* 14: 10027–10044, 1986.
25. **Zhao B, Bowden RA, Stavchansky SA, and Bowman PD.** Human endothelial cell response to gram-negative lipopolysaccharide assessed with cDNA microarrays. *Am J Physiol Cell Physiol* 281: C1587–C1595, 2001.