

Application of Microarrays to the Analysis of Gene Expression in Cancer

PASCALE F. MACGREGOR¹ and JEREMY A. SQUIRE^{2–4*}

Molecular diagnostics is a rapidly advancing field in which insights into disease mechanisms are being elucidated by use of new gene-based biomarkers. Until recently, diagnostic and prognostic assessment of diseased tissues and tumors relied heavily on indirect indicators that permitted only general classifications into broad histologic or morphologic subtypes and did not take into account the alterations in individual gene expression. Global expression analysis using microarrays now allows for simultaneous interrogation of the expression of thousands of genes in a high-throughput fashion and offers unprecedented opportunities to obtain molecular signatures of the state of activity of diseased cells and patient samples. Microarray analysis may provide invaluable information on disease pathology, progression, resistance to treatment, and response to cellular microenvironments and ultimately may lead to improved early diagnosis and innovative therapeutic approaches for cancer.

© 2002 American Association for Clinical Chemistry

Microarray Technology

Microarray methods were initially developed to study differential gene expression using complex populations of RNA (1). Refinements of these methods now permit the analysis of copy number imbalances and gene amplification of DNA (2) and have recently been applied to the systematic analysis of expression at the protein level (3). Many of the guiding principles of global analysis using microarrays are, in principle, applicable at the RNA, DNA, or protein level. In this review we focus our attention on microarray technologies applied to the anal-

ysis of RNA, with emphasis on the study of gene expression changes in tumors. Discussion covering details concerning the methods outlined below, and a glossary of terms commonly used in microarray analysis are provided in supplementary materials to this article (www.utoronto.ca/cancyto/CLINCHEM). Any microarray study typically involves six steps (see Fig. 1), and the next sections will summarize some of the critical parameters in the general design and optimization of microarray analysis of RNA for the study of gene expression.

MANUFACTURING OF MICROARRAYS

Spotted arrays are manufactured using *xyz* robots that use hollow pins to deposit cDNA (PCR products) or short oligonucleotides onto specially coated glass microscope slides (4). Spot sizes range between 80 and 150 μm in diameter, and arrays that contain up to 80 000 spots can be obtained. Gene sequences to be arrayed are selected from several public databases, which contain resources to access well-characterized genes and expressed sequence tags (ESTs)⁵ representative of genes of unknown function. The clones chosen are amplified from appropriate cDNA libraries by PCR and purified before spotting on the solid support.

In addition to their lower price and flexibility in design, spotted arrays offer the advantage of allowing the simultaneous expression analysis of two biological samples, such as test and control samples. This direct comparison of expression profiles of two biological samples, such as untreated cells compared with treated cells or healthy tissue compared with cancer, is an enormous advantage for any pairwise analysis. Furthermore, because these arrays can be spotted with thousands of sequenced expressed genes and ESTs of unknown function, they offer the potential for the discovery of new genes and defining their role in disease. One disadvantage of spotted arrays is that they provide information

¹ Microarray Centre, Clinical Genomics Center, University Health Network, Toronto, Ontario, M5G 1L7 Canada.

² Ontario Cancer Institute, Princess Margaret Hospital, University Health Network, Toronto, Ontario, M5G 2L9 Canada.

Departments of Medical ³Biophysics and ⁴Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, M5G 2L9 Canada.

*Author for correspondence. Fax 416-946-2065; e-mail jeremy.squire@utoronto.ca.

Received April 2, 2002; accepted May 21, 2002.

⁵ Nonstandard abbreviations: EST, expressed sequence tag; EOC, epithelial ovarian cancer; CGH, comparative genomic hybridization; ISH, in situ hybridization; and IHC, immunohistochemistry.

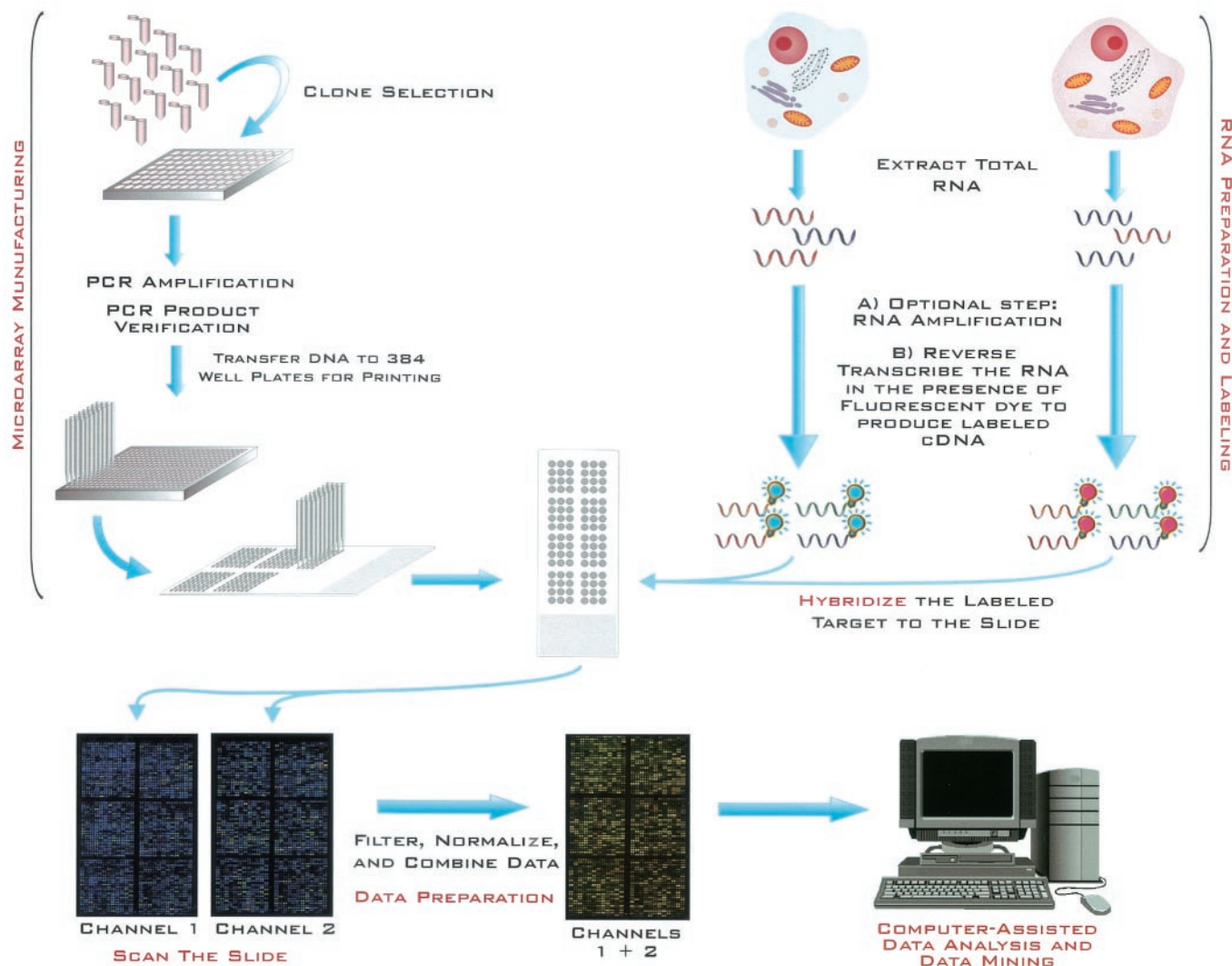


Fig. 1. The six steps in a microarray experiment.

Following standard nomenclature for microarray procedures (44), in this review we refer to the nucleic acids bound to the microarrays as the "probe" and the fluorescently labeled or radiolabeled cDNA hybridized to the array as the "target".

only on the relative gene expression between specific cells or tissue samples as opposed to direct quantification of RNA expression.

Affymetrix GeneChips™ are produced by synthesizing tens of thousands of short oligonucleotides in situ onto glass wafers, one nucleotide at a time, using a modification of semiconductor photolithography technology (1, 5). Generally, GeneChips are designed with 16–20 oligonucleotides representing each gene on the array. Each oligonucleotide on the chip is matched with an almost identical one, differing only by a central, single base mismatch. This allows determination of the degree of nonspecific binding by comparison of target binding intensity between the two partner oligonucleotides. The main advantage of Affymetrix GeneChips is their ability to measure the absolute expression of genes in cells or tissues. Their disadvantages, in addition to their higher

costs, include their current inability to simultaneously compare, on the same array, the degree of expression of two related biological samples. In addition, oligonucleotide-based microarrays require a priori knowledge of the gene sequences and require complex computational manipulation to convert the 40 feature signals into an actual expression value. More recently, oligonucleotide arrays have been developed that combine some the flexibilities and qualitative advantages associated with the use of synthetic probe arrays with the benefits of simultaneous analysis afforded by spotted glass array (6).

In our laboratories, we use the cDNA microarrays spotted with 1700 or 19 200 genes and ESTs manufactured at the University Health Network Microarray Centre (<http://www.microarrays.ca>) to study tumor progression and patient response to treatment in several human solid tumors.

EXPERIMENTAL DESIGN AND CHOICE OF REFERENCE

Careful design at the outset is crucial to the success of microarray experiments. In cancer research, case-control, blocked, and random profile designs predominate. In a case-control study, two samples from a single individual, e.g., tumor tissue and healthy tissue, are compared directly. Because patient variability and genetic heterogeneity are key issues in microarray data analysis, the case-control design is an excellent solution when feasible.

Blocked designs are typically used to study the effect of a treatment or growth condition on a sample such as a cell line. They have been successfully used to examine cell lines grown under different conditions (e.g., cultured in the presence or absence of an anticancer drug) or different related cell lines (e.g., wild type vs mutant, nontransfected cells vs transfected cells). Random profile designs are widely used in microarray experiments when cell lines or patient samples are selected and profiled. Most of the "profiling papers" have used this design, which offers the ability to use data from many different individuals but offers no intrinsic control for bias in the patient populations or cell populations used.

In both the blocked and randomized profile designs, the sample is typically compared with a common or "universal" reference, which should have adequate representation of the majority of genes on the array being profiled and be easily available. Commercially available reference RNA is often a good choice because of wide gene representation (e.g., Stratagene and Clontech). The use of a common reference also offers the advantage of allowing longitudinal comparative analysis among several microarray projects between different research groups interested in a common aspect of cancer research, such as tumor progression or resistance to anticancer drugs. We have recently used a pool of 9 cell lines to establish the expression profiles of a series of 15 ovarian cancer samples (7).

The importance of replicates cannot be overemphasized because variability can be very high in microarray experiments. Many groups, including ours, also choose to carry out so-called "dye reversals", in which one replicate array is hybridized with the experimental sample labeled with one fluorophore and the reference sample with the other dye. The corresponding duplicate array is then hybridized with experimental samples and reference samples labeled with the opposite fluorophores. This strategy generates replicate data while balancing the possible differential efficiency of dye incorporation among RNA samples.

TARGET PREPARATION AND HYBRIDIZATION

Both total RNA and mRNA can be used for microarray experiments and allow the attainment of high-quality data with a high degree of confidence. High-quality RNA is crucial for successful microarray experiments. Different standard RNA extraction methodologies have been used successfully, and the choice of protocol is largely a ques-

tion of personal experience. Quantitative and qualitative evaluation of the RNA obtained can be carried out by standard techniques, such as agarose gel electrophoresis, but is limited by the relatively large amounts of sample required. More recently, assessment of RNA quality and quantity has been greatly facilitated by the use of microcapillary-based devices such as the Agilent Bioanalyzer (Agilent Technologies), which can be used with as little as 5 ng of total RNA.

One of the current limitations in the routine application of microarray technology to patient samples is sufficient RNA availability. Thus, there has been considerable interest in the development of RNA amplification strategies that facilitate RNA extraction from laser capture microdissected (LCM) samples, such as fine-needle biopsies. For standard microarray experiments, the isolated RNA is reverse-transcribed into target cDNA in the presence of fluorescent (generally Cy3-dNTP or Cy5-dNTP) or radio-labeled deoxynucleotides ($[^{33}\text{P}]$ - or $[^{32}\text{P}]\text{-}\alpha\text{-dCTP}$). After purification and denaturation, the labeled targets are hybridized to the microarrays at a temperature determined by the hybridization buffer used. After hybridization, the arrays are washed under stringent conditions to remove nonspecific target binding and are air-dried.

IMAGE ACQUISITION AND QUANTIFICATION

Microarray image processing uses differential excitation and emission wavelengths of the two fluorors to obtain a scan of the array for each emission wavelength, typically as two 16-bit grayscale TIFF images. These images are then analyzed to identify the spots, calculate their associated signal intensities, and assess local background noise. Most image acquisition software packages also contain basic filtering tools to flag spots such as extremely low-intensity spots, ghosts spots (where background is higher than spot intensity), or damaged spots (e.g., dust artifacts). These results allow an initial ratio of the evaluated channel/reference channel intensity to be calculated for every spot on the chip. The products of the image acquisition are the TIFF image pairing and a quantified data file that has not yet been normalized. An excellent assessment of different image analysis methods can be found at <http://oz.berkeley.edu/tech-reports/>.

DATABASES AND NORMALIZATION

The quantity of data generated in a microarray experiment typically requires a dedicated database system to store and organize the microarray data and images. The first role of a local microarray database is the storage and annotation (description of experimental parameters) of microarray experiments by the investigator who designed and carried out the microarray experiments. In addition, there is currently an increasing global interest in making microarray data sets publicly available in a standardized format. This would allow other investigators to reproduce published microarray experiments, to thereby independently verify them, to compare data sets across different

microarray platforms, and importantly, to interrogate published microarray data sets by use of various bioinformatics tools to explore different biological problems. To answer this need, the Minimal Information about a Microarray Experiment, or MIAME standard, has been proposed by the MGED (<http://www.mged.org>) organization as a series of criteria that should be used when defining microarray experiment parameters. In our group, we enter all microarray data into a local microarray database (GeneTraffic; Iobion Informatics), which holds all of the microarray data files and TIFF images, as well as a MIAME supportive annotation of our experiments.

Once data have been loaded into the database, they are normalized, and aggregate statistics are calculated. Normalization is a process that scales spot intensities such that the normalized ratios provide an approximation of the ratio of gene expression between the two samples. Discussion of the different strategies for normalization of microarray data is beyond the scope of this review article, but the choice of a robust and adequate normalization method is as crucial for the quality of the data obtained as the experimental design of the microarray experiment itself. A discussion of normalization methods is provided in supplementary materials to this article (www.utoronto.ca/cancyo/CLINCHEM).

STATISTICAL ANALYSIS AND DATA MINING

Analysis of large gene expression data sets is a new area of data analysis with its own unique challenges. Data mining methods typically fall into one of two classes: supervised and unsupervised. In unsupervised analysis, the data are organized without the benefit of external classification information. Hierarchical clustering (8), K-means clustering (9, 10), or self-organizing maps (11) are examples of unsupervised clustering approaches that have been widely used in microarray analysis (8, 12–15).

Supervised analysis uses some external information, such as the disease status of the samples studied. Supervised analysis involves choosing from the entire data set a training set and a testing set and also involves construction of classifiers, which assign predefined classes to expression profiles. Once the classifier has been trained on the training set and tested on the testing set, it can then be applied to data with unknown classification. Supervised methods include k-nearest neighbor classification, support vector machines, and neural nets. Golub et al. (16) used a k-nearest neighbor strategy to classify the expression profiles of leukemia samples into two classes: acute myeloid leukemia and acute lymphocytic leukemia. Recently Su et al. (17) used large-scale RNA profiling and supervised machine learning algorithms to construct a molecular classification for 10 carcinomas (prostate, lung, ovary, colorectum, kidney, liver, pancreas, bladder/ureter, and gastroesophagus). Similarly, neural network analysis has been used by Khan et al. (18) to delineate consistent patterns of gene expression in cancer.

Tusher et al. (19) recently proposed a strategy called SAM (significance analysis of microarrays), which allows the determination of significantly differentially expressed genes between groups of samples analyzed by expression arrays. We have used this approach to narrow down the analysis to a subset of genes that were also shown to be differentially expressed when analyzed by conventional two-dimensional hierarchical clustering. As discussed below, we have recently identified genes that show differential expression between early-stage epithelial ovarian cancer (EOC), late-stage EOC, and healthy ovary (Fig. 2).

Expression Profiling Applied to Cancer Biology

Cancer is caused by the accumulation of genetic and epigenetic changes resulting from the altered sequence or expression of cancer-related genes, such as oncogenes or tumor suppressor genes, as well as genes involved in cell cycle control, apoptosis, adhesion, DNA repair, and angiogenesis. Because gene expression profiles provide a snapshot of cell functions and processes at the time of sample preparation, comprehensive combinatorial analysis of the gene expression patterns of thousands of genes in tumor cells and comparison to the expression profile obtained with healthy cells should provide insights concerning consistent changes in gene expression that are associated with tumor cellular dysfunction and any concomitant regulatory pathways. Microarray technology has been widely used in the past 3 years to investigate tumor classification, cancer progression, and chemotherapy resistance and sensitivity. In this section we provide three examples to demonstrate that expression arrays can be used to gain a better understanding of the basic biology, diagnosis, and treatment of cancer.

MOLECULAR TUMOR CLASSIFICATION

Improvements in tumor classification are central to the development of novel and individualized therapeutic approaches. Histologically indistinguishable tumors often show significant differences in clinical behavior, and subclassification of these tumors based on their molecular profiles may help explain why these tumors respond so differently to treatment. In a landmark study, Golub et al.

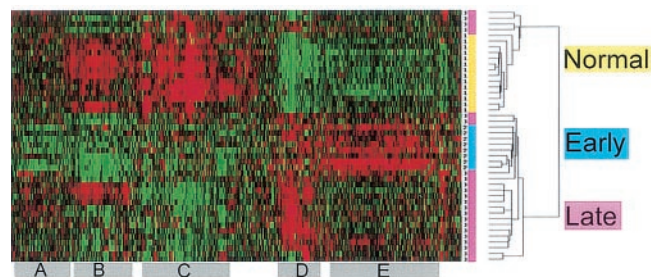


Fig. 2. Two-dimensional hierarchical clustering of microarray data obtained with 22 ovarian tissue samples.

Samples include 15 early- or late-stage serous EOC and 7 healthy ovaries. A–E represent distinct clusters of genes whose expression permits distinction among early EOC, late EOC, and healthy ovary (Normal).

(16) applied microarray technology to develop innovative classifications of leukemias, using microarray analysis based on "neighborhood analysis" and the utilization of tumor class predictors. This strategy was able to distinguish between acute myeloid leukemia and acute lymphocytic leukemia without supervisory analysis. Other groups have also used gene expression pattern analysis to classify, at the molecular level, breast tumors (20, 21), B-cell lymphoma (14), cutaneous melanoma (22), and lung adenocarcinoma (23, 24). Likewise, in a recent study analyzing molecular profiles of 50 nonneoplastic and neoplastic prostate samples, Dhanasekaran et al. (25) established signature expression profiles of healthy prostate, benign prostatic neoplasia, localized prostate cancer, and metastatic prostate cancer. These studies established the feasibility of combining large-scale molecular analysis of expression profiles with classic morphologic and clinical methods of staging and grading cancer for better diagnosis and outcome prediction.

DRUG SENSITIVITY

Despite considerable advances in cancer treatment, acquired resistance to chemotherapeutic drugs continues to be a major obstacle in patient treatment and overall outcome. Anticancer drug resistance is thought to occur through numerous mechanisms, and microarrays offer a new approach to studying the cellular pathways implicated in these mechanisms and in predicting drug sensitivity and unexpected side effects. Most array studies have been carried out using cancer cell lines that are rendered resistant to commonly used anticancer drugs. For example, Kudoh et al. (26) monitored the expression profiles of doxorubicin-induced and -resistant cancer cells in an attempt to obtain molecular fingerprinting of anticancer drugs in cancer cells. Scherf et al. (27) analyzed a subset of 1400 genes from a study reported by Ross et al. (28) and studied the correlation between expression profiles and drug mechanism of action of a panel of 118 anticancer drugs. Obtaining further insights into the mechanism of action of anticancer drugs and the diverse pathways involved in drug resistance may eventually be invaluable for design of more strategic treatments that are most appropriate for an individual tumor.

IDENTIFICATION OF TUMOR-SPECIFIC MOLECULAR MARKERS

Several research groups have focused on identifying subsets of genes that show differential expression between healthy tissues or cell lines and their tumor counterparts to identify biomarkers for several solid tumors, including ovarian carcinomas (7, 29–32), oral cancer (33), melanoma (34), colorectal cancer (35), and prostate cancer (36). In our recent study (7) carried out on a cohort of 13 patients with EOC, we identified a subset of genes that show differential expression between healthy ovaries and ovarian tumors (Fig. 2). Some of these genes, such as metallothionein 1G, which was found to be up-regulated

in tumor samples, are implicated in resistance to the anticancer drug cisplatin and might be an indicator of pretreatment resistance of these tumors to cisplatin. Other genes identified in our study, such as the osteopontin gene, which was strongly up-regulated in some tumor samples and which has been shown (37) to be secreted in the serum of patients with metastatic cancer, might be an excellent candidate for biomarkers of tumor progression in EOC. One of the most important challenges facing investigators using microarray analysis is determining which of the plethora of new differentially expressed genes is biologically relevant to the tumor system being studied. Even when rigorous efforts are made to minimize the number of variables in a microarray study, there may be an unmanageable number of differentially expressed genes that will contribute excessive background values. Therefore, combining expression microarray analysis with other approaches, particularly cytogenetics techniques, such as spectral karyotyping and chromosome and array comparative genomic hybridization (CGH) (2), offers the possibility to focus on significantly smaller subsets of genes of direct relevance to tumor biology (7). Monni et al. (38) and Barlund et al. (39) recently used a combination of expression arrays and CGH array techniques on breast cancer cell lines and have identified a limited number of genes that are both amplified and overexpressed. [For a review, see Monni et al. (40), as illustrated in Fig. 3].

Finally, validation of the relative expression obtained from genome-wide microarray analysis is critical. Several approaches can be chosen, from basic Northern analysis or semiquantitative reverse transcription-PCR to *in situ* hybridization (ISH) using tissue microarrays. Mousset et al. (41) recently analyzed the expression of several candidate genes associated with prostate cancer that they had previously identified by cDNA microarray analysis. Tissue microarrays constructed from 544 histologic biopsies were analyzed by ISH using RNA probes and/or by immunohistochemistry (IHC) using antibodies. There was excellent correlation between the cDNA microarray results and the results obtained with ISH and Northern blot analysis. In addition, protein expression assessed by IHC was also consistent with RNA expression. Similarly, Dhanasekaran et al. (25) used comparable technologies to confirm overexpression of hepsin and PIM-1 in prostate cancer (Fig. 4).

PRACTICAL AND FUTURE APPLICATIONS OF MICROARRAY TECHNOLOGY

The numbers of microarray-based studies identifying new genes or molecular pathways involved in tumor classification, cancer progression, or patient outcome are growing exponentially. We are now approaching what is being referred to as the "postgenomic era", during which the diagnostic, prognostic, and treatment response biomarker genes identified by microarray screening will be interrogated to provide personalized management of patients.

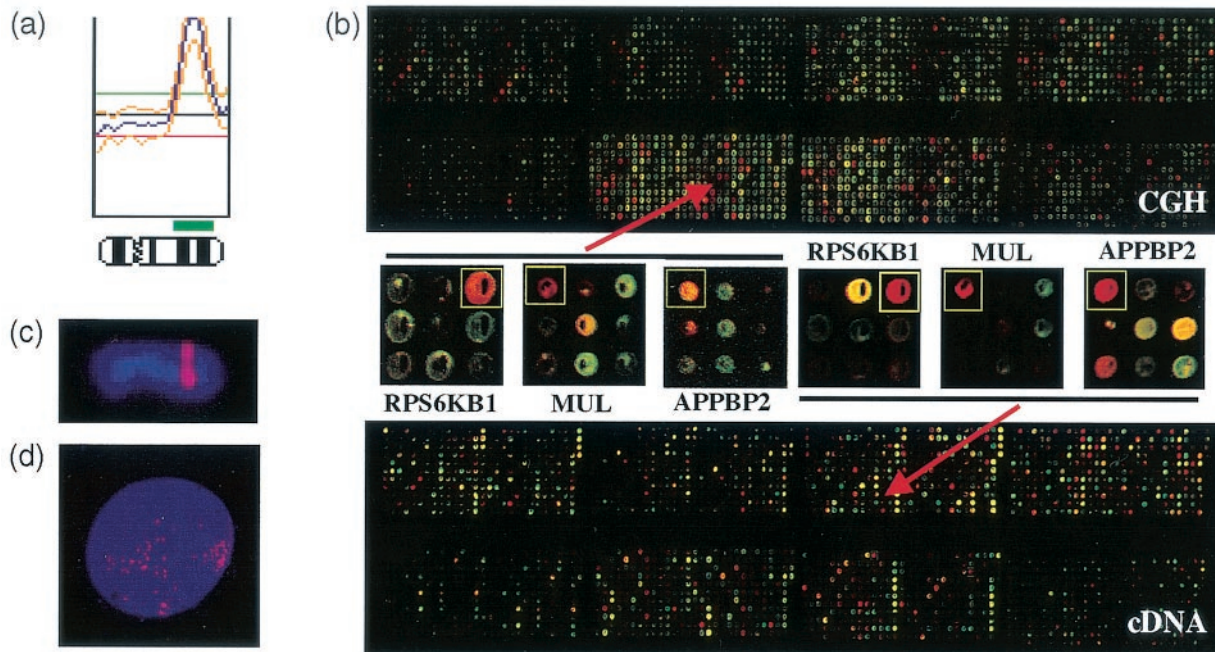


Fig. 3. Detection of amplified and overexpressed genes by cDNA and CGH microarray techniques in the MCF7 breast cancer cell line. (a), copy number ratio profile for chromosome 17 obtained from conventional CGH analysis indicates a large region of high-level amplification at 17q23. (b), CGH microarray (top) and cDNA microarray (bottom) analyses of the MCF7 breast cancer cell line. The same cDNA microarray format containing chromosome 17-specific genes and ESTs was used for both analyses. After hybridization with red-labeled tumor DNA (CGH microarray; top) or cDNA (cDNA microarray; bottom) against a green-labeled reference specimen, genes that are amplified and overexpressed are visualized as red dots. The insets show three regions on the cDNA microarray at a higher magnification, visualizing the amplification (left panels) and overexpression (right panels) of three genes, MUL, RPS6KB1, and APPBP2, that are located at 17q23 by fluorescence ISH. This corresponds to the same region in which amplification was seen by CGH. (d), high-level RPS6KB1 amplification in MCF7 cells as visualized by interphase fluorescence ISH. From Monni et al. (40).

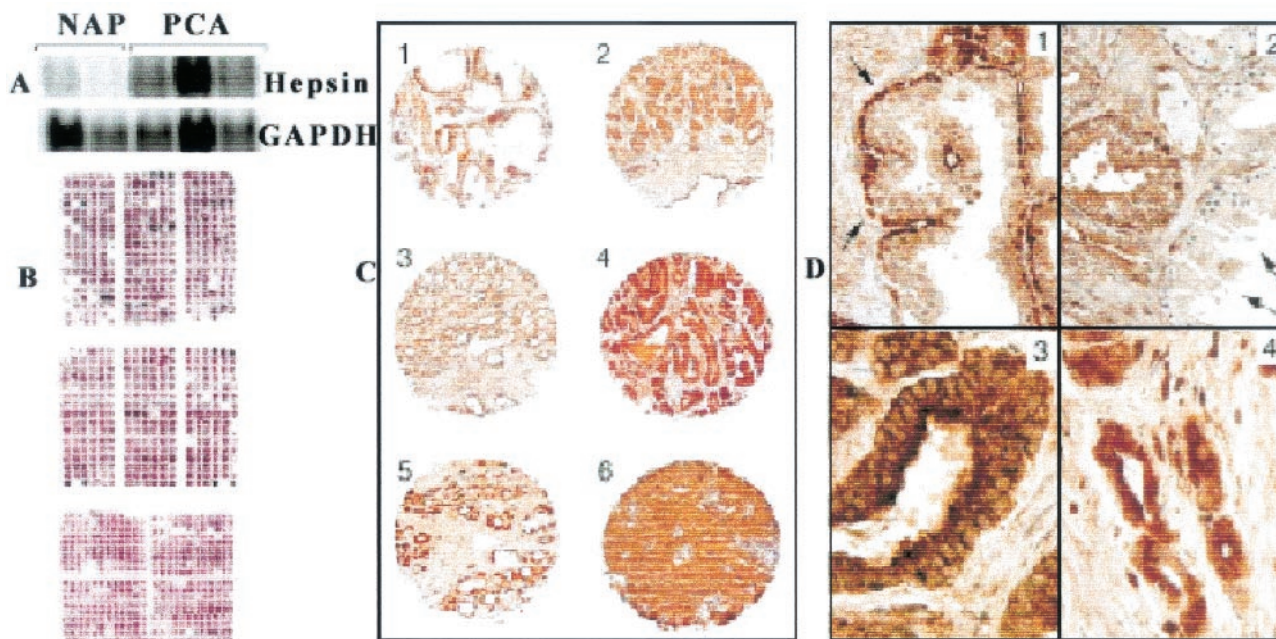


Fig. 4. Hepsin is overexpressed in prostate cancer. (A), Northern blot analysis of human hepsin and normalization with GAPDH. NAP, normal adjacent prostate; PCA, localized prostate cancer. (B), tissue microarrays used for hepsin analysis (stained with hematoxylin and eosin). (C), representative elements of a tissue microarray stained with anti-hepsin antibody. IHC demonstrates absent or weak staining of benign prostate and strong staining in localized prostate cancer. (D), benign prostate glands demonstrate strong basal cell staining (panel 1) but weak expression in the secretory luminal cells (panel 2). From Dhanasekaran et al. (25).

Clinicians will be able to use microarrays during early clinical trials to confirm the mechanisms of action of drugs and to assess drug sensitivity and toxicity. Coupled with more conventional biochemical analysis such as IHC and ELISA, microarrays will be used for diagnostic and prognostic purposes. A recent example of such a potential "bench to bedside" translation was published by Kim et al. (42). The osteopontin gene, which encodes a calcium-binding glycoprotein, had been identified by cDNA microarray analysis as being up-regulated in ovarian cancer (43). In their study, Kim et al. (42) showed that screening of plasma samples from ovarian cancer patients revealed that osteopontin protein concentrations in plasma were significantly higher in a majority of patients with ovarian cancer compared with healthy controls. This study demonstrated the potential value of cDNA microarray analysis in identifying biomarker genes in cancer and the feasibility of subsequently testing these genes at the protein level by conventional biochemical assays. Although the major limiting factors for routine use in a clinical setting at present are cost and access to the microarray technology, it is likely that costs will decrease in the near future and that the technology will become increasingly user friendly and automated.

Conclusion

The range of applications of microarray technology is enormous. Recent studies in human cancer have demonstrated that microarrays can be used to develop a new molecular taxonomy of cancer, including clustering of cancers according to prognostic groups on the basis of gene expression profiles. The list of potential uses of this technique is not limited to cancer research. For example, the temporal impact on gene expression by drugs, environmental toxins, or oncogenes may be elucidated, and regulatory networks and coexpression patterns can then be deciphered. In the 6 years since its inception, microarray technology has become a major tool for the investigation of global gene expression of all aspects of human disease and in biomedical research.

We thank Dr. Jim Woodgett and Jason Gonçalves for critically reviewing this manuscript, and Monique Albert for help in figure design and manuscript preparation.

References

1. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nat Genet* 1999;21:20–4.
2. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 1999;23:41–6.
3. Haab BB. Advances in protein microarray technology for protein expression and interaction profiling. *Curr Opin Drug Discov Dev* 2001;4:116–23.
4. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467–70.
5. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. *Science* 1991;251:767–73.
6. Okamoto T, Suzuki T, Yamamoto N. Microarray fabrication with covalent attachment of DNA using bubble jet technology. *Nat Biotechnol* 2000;18:438–41.
7. Bayani J, Brenton JD, Macgregor PF, Beheshti B, Albert M, Nallainathan D, et al. Parallel analysis of sporadic primary ovarian carcinomas by spectral karyotyping, comparative genomic hybridization and expression microarrays. *Cancer Res* 2002;62:3466–76.
8. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998;95:14863–8.
9. Hartigan JA. Clustering algorithms. New York: John Wiley & Sons, 1975:351pp.
10. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. Systematic determination of genetic network architecture. *Nat Genet* 1999;22:281–5.
11. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, et al. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci U S A* 1999;96:2907–12.
12. DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, et al. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 1996;14:457–60.
13. Alizadeh AA, Staudt LM. Genomic-scale gene expression profiling of normal and malignant immune cells. *Curr Opin Immunol* 2000;12:219–25.
14. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;403:503–11.
15. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci U S A* 1999;96:6745–50.
16. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531–7.
17. Su AI, Welsh JB, Sapinoso LM, Kern SG, Dimitrov P, Lapp H, et al. Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Res* 2001;61:7388–93.
18. Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001;7:673–9.
19. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98:5116–21.
20. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747–52.
21. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869–74.
22. Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 2000;406:536–40.
23. Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, et al. Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A* 2001;98:13784–9.
24. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P,

- et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* 2001;98:13790–5.
25. Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, et al. Delineation of prognostic biomarkers in prostate cancer. *Nature* 2001;412:822–6.
 26. Kudoh K, Ramanna M, Ravatn R, Elkahloun AG, Bittner ML, Meltzer PS, et al. Monitoring the expression profiles of doxorubicin-induced and doxorubicin-resistant cancer cells by cDNA microarray. *Cancer Res* 2000;60:4161–6.
 27. Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, et al. A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000;24:236–44.
 28. Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000;24:227–35.
 29. Ismail RS, Baldwin RL, Fang J, Browning D, Karlan BY, Gasson JC, et al. Differential gene expression between normal and tumor-derived ovarian epithelial cells. *Cancer Res* 2000;60:6744–9.
 30. Mok SC, Chao J, Skates S, Wong K, Yiu GK, Muto MG, et al. Prostatein, a potential serum marker for ovarian cancer: identification through microarray technology. *J Natl Cancer Inst* 2001;93:1458–64.
 31. Shridhar V, Lee J, Pandita A, Iturria S, Avula R, Staub J, et al. Genetic analysis of early- versus late-stage ovarian tumors. *Cancer Res* 2001;61:5895–904.
 32. Shridhar V, Sen A, Chien J, Staub J, Avula R, Kovats S, et al. Identification of underexpressed genes in early- and late-stage primary ovarian tumors by suppression subtraction hybridization. *Cancer Res* 2002;62:262–70.
 33. Alevizos I, Mahadevappa M, Zhang X, Ohyama H, Kohno Y, Posner M, et al. Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. *Oncogene* 2001;20:6196–204.
 34. Clark EA, Golub TR, Lander ES, Hynes RO. Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* 2000;406:532–5.
 35. Hegde P, Qi R, Gaspard R, Abernathy K, Dharap S, Earle-Hughes J, et al. Identification of tumor markers in models of human colorectal cancer using a 19,200-element complementary DNA microarray. *Cancer Res* 2001;61:7792–7.
 36. Luo J, Duggan DJ, Chen Y, Sauvageot J, Ewing CM, Bittner ML, et al. Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. *Cancer Res* 2001;61:4683–8.
 37. Singhal H, Bautista DS, Tonkin KS, O'Malley FP, Tuck AB, Chambers AF, et al. Elevated plasma osteopontin in metastatic breast cancer associated with increased tumor burden and decreased survival. *Clin Cancer Res* 1997;3:605–11.
 38. Monni O, Barlund M, Mousset S, Kononen J, Sauter G, Heiskanen M, et al. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc Natl Acad Sci U S A* 2001;98:5711–6.
 39. Barlund M, Forozaan F, Kononen J, Bubendorf L, Chen Y, Bittner ML, et al. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. *J Natl Cancer Inst* 2000;92:1252–9.
 40. Monni O, Hyman E, Mousset S, Barlund M, Kallioniemi A, Kallioniemi OP. From chromosomal alterations to target genes for therapy: integrating cytogenetic and functional genomic views of the breast cancer genome. *Semin Cancer Biol* 2001;11:395–401.
 41. Mousset S, Bubendorf L, Wagner U, Hostetter G, Kononen J, Conelison R, et al. Clinical validation of candidate genes associated with prostate cancer progression in the CWR22 model system using tissue microarrays. *Cancer Res* 2002;62:1256–60.
 42. Kim JH, Skates SJ, Uede T, Wong Kk KK, Schorge JO, Feltmate CM, et al. Osteopontin as a potential diagnostic biomarker for ovarian cancer. *JAMA* 2002;287:1671–9.
 43. Wong KK, Cheng RS, Mok SC. Identification of differentially expressed genes from ovarian cancer cells by MICROMAX cDNA microarray system. *Biotechniques* 2001;30:670–5.
 44. The chipping forecast. *Nat Genet* 1999;21(Suppl).