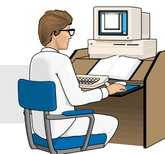

The Use of Microarrays in Medicine

Gady S. Cojocaru[1,2], Gideon Rechavi MD PhD[2] and Naftali Kaminski MD[1,3]

[1]Functional Genomics Unit, [2]Pediatric Hemato-Oncology Section, and [3]Institute of Respiratory Medicine, Sheba Medical Center, Tel-Hashomer, Israel

Key words: microarrays, gene expression, cluster analysis, genomics, disease classification

IMAJ 2001;3:292–296

The completion of the human genome project marks a major leap in our understanding of biological processes that will eventually transform all medical sciences. In a recent review in this journal, Ben Asher et al. [1] described the Israeli perspective of the human genome project, including microarray technology. In the present review, we will provide an overview of the application of microarray technology in research relevant to human disease.

Microarrays allow for the simultaneous monitoring of the behavior/expression of thousands of genes, thus providing a functional aspect to sequence information. Initially, the applications of the two main microarray technologies (to be described later) were limited to very simple or basic science applications such as describing the response of fibroblasts to serum, or providing a list of the genes induced during the yeast cell cycle [2,3]. However, from early on it was clear that these technologies are potentially applicable to research relevant to human disease and to clinical medicine [4,5]. Several approaches have recently emerged, such as the use of patient samples for disease classification, and of genetically modified animals in animal models of human disease [6–8]. In this review, we provide an overview of microarray technology. We describe the current applications of microarray technology in research relevant to human disease and present some of the future perspectives. In addition, we provide some practical advice to scientists who wish to use microarray technology in their field of research.

DNA microarray technologies

The detection of hybridization between a DNA sample and a known labeled probe was developed a quarter of a century ago by Ed Southern [9]. This principle has been used to detect single species of DNA (Southern blotting) or RNA (Northern blotting). The completion of the various (human and other organisms) genome projects required the development of large-scale high throughput technologies that can detect multiple copies and species of DNA or RNA. Microarrays, based on Southern's method of nucleotide hybridization, contain multiple DNA sequences (probes) spotted or synthesized on a relatively small surface. This feature of microarrays allows the simultaneous monitoring of multiple matching sequences in a given sample. Naturally, placing as many DNA sequences as possible on a small as possible surface is highly dependent upon technological advances such as photolithographic oligonucleotide synthesis and ink-jet based technology. Currently, there are two widely used microarray technologies:

- *in situ* synthesized oligonucleotide microarrays – mainly oligonucleotides synthesized by photolithography on a glass surface.
- spotted microarrays – mostly created by robotic printing of pre-prepared cDNAs or oligonucleotides.

In situ synthesized oligonucleotide microarrays

In situ synthesized oligonucleotide microarrays are mainly produced by photolithographic printing on glass substrate. In this method, the glass is treated with a photochemically removable protective group. Light is directed through a mask to specific sites on the glass in order to produce photo deprotection. The glass is then incubated with hydroxyl-protected deoxynucleosides, and the chemical coupling occurs at the sites illuminated in the previous step. This process is repeated until an oligonucleotide at the required length has been produced [10]. In expression microarrays (the most common application), each gene is represented by 16–20 probes perfectly matching the gene sequence and 16–20 similar probes containing point mutations. The perfect matching probes provide the expression level of each transcript, while the mismatching probes provide a measure of hybridization specificity. Total RNA or mRNA is extracted from experimental samples and is labeled with a fluorescent oligonucleotide. The labeled sample is hybridized on the array, the array is scanned, and absolute expression levels are obtained for each probe by using a dedicated laser scanner [10]. The measurements are highly reproducible and the signal is linear over a wide range of transcript copy numbers and concentrations.

Photolitographic oligonucleotide synthesis can be used also for other applications such as single nucleotide polymorphisms detection, linkage analysis (using genomic DNA rather then mRNA), and mutation detection. For example, the p53 mutation chip can detect known mutations, single base mutations and deletions in the p53 gene.

Spotted arrays

cDNA arrays are produced by spotting polymerase chain reaction products (usually 0.6–2.4 kb in length) representing specific genes onto a pretreated glass slide or other surface suitable for hybridization [11]. The spotting is performed using dedicated robotic printing systems (arrayer) that spot pre-prepared probes (cDNA clones, PCR products or oligonucleotides) on a pretreated nylon or glass matrix. This matrix is then probed with fluorescently labeled cDNA. In the most commonly used application, mRNA or total RNA is extracted from a sample and a control – one is labeled with cy-3 (red fluorescent dye) and the other with cy-5 (green fluorescent dye). "Competitive" hybridization is performed on the slide, the slide is scanned by a dual wavelength laser, and ratio signals are obtained [12]. Relative message abundance is estimated by direct comparison between the test sample and the reference [12,13]. This technology is robust and produces highly reproducible results, though standardization of the diverse platforms is a problem.

Applications of microarrays in medicine

The most straightforward approach in research relevant to human disease would be analysis and comparison of gene expression patterns in blood samples, surgically removed tissues, or biopsies of normal and diseased specimens. This simple approach can provide lists of genes correlated with a disease, a feature that may improve the diagnostic options of the disease and provide new targets for intervention [14]. A major contribution of such experiments will be class discovery and sub-classification of diseases. Sub-classification of a disease can help physicians decide upon treatment strategies and better predict outcome and survival [6]. Though powerful, there are some limitations to using this approach. Challenging technical aspects are sample quality and quantity. Since the starting material is total RNA, it is essential that surgical specimens be handled properly to assure minimal RNA degradation. RNA quantity may also be a limitation; most microarray applications require a minimum amount of 1–10 μg of RNA corresponding to 1–5 g of tissue that may be too much for biopsy material. A different challenging aspect is that the data obtained are descriptive while the mechanisms underlying the pathological condition remain elusive.

Several approaches have evolved to overcome these problems, mainly focusing on careful study design and combining data obtained from patients with animal models of human disease [6.8]. In the next paragraphs, we provide a few representative examples that demonstrate the use of microarrays in studies of human samples and animal models of human disease.

Human disease

Alizadeh et al. [6] studied diffuse large B cell lymphoma, the most common subtype of non-Hodgkin's lymphoma. Two distinct expression patterns were identified in a population of pathologically identical DLBCL samples. One of the patterns had characteristics of germinal center B cells, while the other had characteristics of activated B cells. Subsequent analysis demonstrated that the patterns correlated with distinct clinical outcome. Patients with the germinal center phenotype had significantly better survival rates than patients with the activated B cell phenotype [6]. Other studies using microarrays to analyze human disease are shown in Table 1.

Another use of expression microarrays is to identify transcriptional programs in normal human development. In a recent paper aimed at identifying genes whose expression is associated with aging, mRNA expression levels of skin fibroblasts obtained from young, middle-aged and old men, and a progeria patient were compared. The results of this study suggest that the underlying mechanism of the aging process involves increasing errors in mitotic machinery in the post-reproductive stage [21].

Animal models of human disease

Though informative, the use of human samples is limited by their availability as well as the insufficient mechanistic information. Using animal models of human diseases, one can monitor expression levels at different stages of the disease in the presence of disease-modifying agents and in highly controlled environments. Another advantage of animal models is the use of transgenic animals that can provide insight into the expression profile of susceptibility genes involved in disease. Kaminski et al. [8] analyzed the gene expression programs that underlie pulmonary fibrosis in response to bleomycin, a drug that causes lung inflammation and fibrosis, in two strains of susceptible mice (129 and C57bl/6). They then compared the gene expression patterns in these mice with 129 mice carrying a null mutation in the epithelial-restricted integrin β6 subunit (*β6-/-*), which develop inflammation but are protected from pulmonary fibrosis. Using hierarchical clustering [22] they identified two distinct groups of genes involved in the inflammatory and fibrotic responses, suggesting that the fibrotic process is activated independently after bleomycin-induced lung injury. Their results support the need to develop interventions targeted specifically at the fibrotic process instead of the commonly used but ineffective anti-inflammatory interventions [23].

A combination of an animal model and a human cell line is demonstrated in the work by Clark et al. [24]. The goal of their study was to determine the cellular events that distinguish the metastatic phenotypes of different melanoma cell lines. They injected mice with melanotic cells derived from a poorly

PCR = polymerase chain reaction

Table 1. Studies using microarrays to analyze human disease

Disease	Microarray type	Summary	Ref.
Melanoma	cDNA	Molecular classification of cutaneous malignant melanoma	[15]
Colon cancer	Oligonucleotide	Two-way clustering analysis of colon cancer, grouped together class of similar-expression gene families and segregated the normal samples from the pathological ones	[16]
Diabetes type I	cDNA	Ineffective differentiation of diabetic Vα24JαQ T cells and a role for invariant T cells in the recruitment and activation of cells from the myeloid lineage	[17]
Ovarian cancer	cDNA	Comparison of expression patterns between carcinomas and the corresponding normal ovarian tissues enabled identification of 55 genes that were commonly up-regulated and 48 genes that were down-regulated in the cancer specimens	[18]
Development	oligonucleotide	Comparison of adult vs. fetal tissue identified 11 genes with similar-expression patterns. These genes are candidates as cell maintenance genes	[19]
AML and ALL	cDNA	A class discovery procedure automatically discovered the distinction between AML and ALL without previous knowledge of these classes	[14]
Alzheimer's disease	cDNA	Relative to normal CA1 neurons, NFT neurons in AD brains showed significant reductions in several classes of genes that are known to encode proteins implicated in AD neuropathology, including phosphatases/kinases, cytoskeletal proteins, synaptic proteins, glutamate receptors, and dopamine receptors	[20]

AML = acute myeloid leukemia, ALL = acute lymphoblastic leukemia, AD = Alzheimer's disease, NFT = neurofibrillary tangles.

metastatic cell line. They then analyzed gene expression patterns in cells harvested from the rare lung metastases and compared them to the original cell line. About 30 genes were differentially expressed. The most highly expressed genes in the metastatic phenotype were fibronectin, thymosin β4 and RhoC. They confirmed the central role of RhoC in the metastatic phenotype by demonstrating that poorly metastatic cells became highly metastatic when transfected with RhoC. Furthermore, expression of a dominant-negative form of RhoC in the highly metastatic cells transformed them into the poorly metastatic phenotype [24].

Infectious disease

Viruses and bacteria affect their host cell by altering mRNA expression levels in specific ways. RNA extracted from infected tissues can be used as a "signature" for identifying the pathogen [25].

Challenges of microarray technology

Data management and mining

The crucial stage in both oligonucleotide and cDNA arrays experiments is data management and analysis. Microarray experiments produce vast amounts of information that require the use of specialized data management tools. The diversity and scope of the data require the creation of multidisciplinary teams consisting of physicians, biologists and bioinformaticians (mathematicians, computational biologists and database managers).

Specifically, data handling is composed of several layers. Once experimental data have been obtained, the data have to be normalized in order to standardize all experiments to variations in microarray quality, sample handling and biological diversity. The data must be stored in relational databases linked to publicly available genomic and functional databases as well as to results of previous experiments, allowing easy data retrieval and manipulation. Advanced analysis tools such as clustering applications and data mining tools [22,26] are needed in order to retrieve all the functional information stored in the data. Standardization of the data (raw and processed) is a major problem due to the plethora of available technologies. Standard expression database management systems are being developed to permit data sharing by scientists worldwide [27]. This will facilitate better communication among the different disciplines using microarray technology and will lead to better experimental designs and data interpretation.

Cost

Microarray technology is highly cost-intensive. Setting up a system ranges in price from US$ 60,000 to 500,000. The cost of each chip ranges from $50 for self-made chips to $2,500 for ready-made ones. Though self-made chips are significantly cheaper one has to consider the cost of the system setup, including acquisition of clones for printing, quality control of printing and other technical concerns that are obstacles to immediate implementation. The computational hardware and software can range in cost from $10,000 to more than $100,000.

DLBCL = diffuse large B cell lymphoma

Therefore, joint ventures of medical and academic institutes together with biotechnological and pharmaceutical industries are crucial.

Closing remarks

In this paper, we have provided an overview of the two most commonly used microarray technologies – oligonucleotide and spotted cDNA arrays. Microarray technology has been widely used in basic molecular biology and in research relevant to human disease. The scope of this review was to demonstrate its growing application in medicine – particularly in molecular classification of disease and in analyzing animal models of human disease. We believe that many more applications will emerge in the near future, such as early detection of disease, drug toxicity prediction, mutation and SNP analysis, and diagnosis of infections.

Several practical suggestions may make the transition to the use of microarrays in research relevant to human disease less painful. They are:

- If you are thinking of microarray-based experiments – stop! Consider well if this approach is the most suitable. For many biological questions conventional molecular biology is often more satisfactory, and definitely cheaper.
- Have a well-defined scientific question – although microarrays are used as hypothesis-generating tools, the results are highly dependent on the careful design of the experiment.
- Make sure that you can deal with the data, dedicate a data analysis and storage workstation, and check your Internet access.
- Collaborate with bioinformatics groups.
- Handle the samples as obsessively as possible. Variable RNA degradation is a critical factor in surgically obtained samples. Samples must be snap-frozen in liquid nitrogen immediately after removal.
- Repeat your experiments. If repeating an experiment using microarrays is too expensive, repeat and verify the results using alternative cost-efficient methods. Do not overlook validating target genes at the protein expression level.
- Good Luck.

Acknowledgements. The authors thank the Arison Dorsman family donation for the Center of DNA microarrays in Pediatric Oncology and to Dr. Ninnette Amarilio for her helpful remarks.
Dr. Kaminski is a recipient of a research grant from the Israel Lung Association. Dr. Rechavi holds the Gregorio and Dora Shapiro Chair for Hematologic Malignancies, Sackler Faculty of Medicine, Tel Aviv University.
The work was performed as part of the academic requierments for the PhD degree of G. Cojocaru, Sackler Faculty of Medicine, Tel Aviv University.

SNP = single nucleotide ploymorphism

References

1. Ben-Asher E, Califa-Caspi V, Horn-Saban S, Avidan N, Olender Z, Adato A, Glusman G, Safran M, Rubinstein M, Lancet D. Harvesting the human genome: the Israeli perspective. *IMAJ* 2000;2:657–64.
2. Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JCF, Trent JM, Staudt LM, Hudson J, Jr., Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO. The transcriptional program in the response of human fibroblasts to serum. *Science* 1999;283:83–7.
3. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell* 1998;9:3273–97.
4. Madden SL, Wang CJ, Landes G. Serial analysis of gene expression: from gene discovery to target identification. *Drug Discov Today* 2000;5:415–25.
5. Masters JR, Lakhani SR. How diagnosis with microarrays can help cancer patients. *Nature* 2000;404:921.
6. 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, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;403:503–11.
7. Alizadeh AA, Staudt LM. Genomic-scale gene expression profiling of normal and malignant immune cells. *Curr Opin Immunol* 2000;12:219–25.
8. Kaminski N, Allard JD, Pittet JF, Zuo F, Griffiths MJ, Morris D, Huang X, Sheppard D, Heller RA. Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis. *Proc Natl Acad Sci USA* 2000;97:1778–83.
9. Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 1975;98:503–17.
10. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nat Genet* 1999;21:20–4.
11. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nat Genet* 1999;21:10–14.
12. Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G. Making and reading microarrays. *Nat Genet* 1999;21:15–19.
13. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nat Genet* 1999;21:33–7.
14. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531–7.
15. Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D, Sondak V. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 2000;406:536–40.
16. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, Levine AJ. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci USA* 1999;96:6745–50.
17. Wilson SB, Kent SC, Horton HF, Hill AA, Bollyky PL, Hafler DA, Strominger JL, Byrne MC.Multiple differences in gene expression in regulatory Valpha 24Jalpha Q T cells from identical twins discordant for type I diabetes. *Proc Natl Acad Sci USA* 2000;97:7411–16.
18. Ono K, Tanaka T, Tsunoda T, Kitahara O, Kihara C, Okamoto A, Ochiai K, Takagi T, Nakamura Y. Identification by cDNA microarray of genes involved in ovarian carcinogenesis. *Cancer Res* 2000;60:5007–11.
19. Warrington JA, Nair A, Mahadevappa M, Tsyganskaya M. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol Genomics* 2000;2:143–7.
20. Ginsberg SD, Hemby SE, Lee VM, Eberwine JH, Trojanowski JQ. Expression profile of transcripts in Alzheimer's disease tangle-bearing CA1 neurons. *Ann Neurol* 2000;48:77–87.

21. Ly DH, Lockhart DJ, Lerner RA, Schultz PG. Mitotic misregulation and human aging. *Science* 2000;287:2486–92.
22. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998;95:14863–8.
23. King TE, Jr. Update in pulmonary medicine. *Ann Intern Med* 1998;129:806–12. [published erratum appears in *Ann Intern Med* 1999;131:74].
24. Clark EA, Golub TR, Lander ES, Hynes RO. Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* 2000;406:532–5.
25. Manger ID, Relman DA. How the host 'sees' pathogens: global gene expression responses to infection. *Curr Opin Immunol* 2000;12:215–18.
26. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci* USA 1999;96:2907–12.
27. Brazma A, Vilo J. Gene expression data analysis. *FEBS Lett* 2000;480:17–24.

Correspondence: Dr. N. Kaminski, Director, Functional Genomics Unit, Institute of Respiratory Medicine, Sheba Medical Center, Tel-Hashomer 52621, Israel. Phone: (972-3) 530-5351, Fax: (972-3) 530-2377, email: kamins@itsa.ucsf.edu