

Expression profiling: Opportunities and pitfalls and impact on the study and management of allergic diseases

Santa Jeremy Ono, PhD, Takao Nakamura, MD, PhD, Masaharu Ohbayashi, PhD,
Maria Dawson, MPhil, Yoshifumi Ikeda, MD, Alex K. Nugent, BSc,
Masako Toda, PhD, and Gilbert Jay, PhD, DSc *London, United Kingdom*

Key word: *Gene profiling*

Eleven years have passed since Liang and Pardee¹ reported gene profiling through differential display-PCR technology and 8 years since Schena et al² reported the profiling of gene expression in yeast by using robotically printed DNA microarrays on the basis of Affymetrix VLSIPS technology (Santa Clara, Calif). The subsequent development of high-density arrays and critical bioinformatics programs has made this potentially powerful technology broadly available to the biomedical research community in a remarkably short time. Gene profiling has already delivered stunning insights into evolution, cancer, and the immune response, and there is guarded optimism that the approach will deliver additional insights throughout biomedical science. However, many of the emerging experiments have yielded disappointing results because of remaining imperfections in essential technologies, flawed experimental design, and/or poor data analysis. In this editorial we consider the potential and pitfalls of gene profiling, examine a few examples of data emerging from such studies in allergy research and other fields, and indicate safeguards that can enhance the chances that meaningful data are obtained.

Over the past couple of years, there has been overwhelming interest in the expression profiling of different human disorders by using DNA microarray technologies.^{3,4} Despite that interest, however, the findings that came out of many of those studies have not been particularly revealing, at least in terms of their increasing our understanding of the respective disease process. At best, they provided the prospect of an incomplete molecular fingerprint to facilitate diagnosis of the disease.⁵

Why is there such interest in the expression profiling of disease? The discovery of genes and their functions is the foundation of biotechnology. Although a disease might initiate with the faulty expression of a single gene, as demon-

strated in numerous transgenic animal studies, the molecular and clinical manifestation of that disease is the consequence of the involvement of successive biochemical events and numerous gene products. High-throughput approaches, such as the use of gene chips, are now required to facilitate the discovery of all the molecular players and to interrelate them in their role in the disease process.

The availability of the almost finished human genome sequence and, more recently, the draft mouse genome sequence has certainly enhanced the prospects of successful expression profiling.^{6,7} Although current computational algorithms have failed to definitively identify all the coding exons and properly associate them with individual genes, the genome sequences have provided ample numbers of potentially expressed sequences for high-throughput comparative screening of mRNA from cells under different conditions or physiologic states. After all, the expression characteristics of a segment of DNA might, at the end, serve to confirm the completeness of a gene that has been predicted by computational means.

Let us start this commentary with a brief description of DNA microarray technology. The technique determines the relative expression of multiple (up to tens of thousands) genes within biologic samples. The careful analysis of the data is accomplished through bioinformatics. Much of the utility of the experiment stems from the potential of comparing the expression of many genes between samples.

The typical DNA microarray experiment (see depiction in Fig 1) begins with the isolation of mRNA from 2 or more biologic samples (cells or tissues). This necessitates that meticulous care is taken at this stage because variability in mRNA preparation or RNA degradation can confound subsequent data. For this reason, it is important to monitor the quality of prepared RNAs before moving forward with the analyses. Once high-quality RNA has been prepared from the samples, reverse transcriptase is used to generate cDNA. During the synthesis of cDNA, radiolabeled, fluorescent, or biotinylated nucleotides are incorporated into the second strand of the DNA helix. These modifications permit the detection of immobilized cDNAs by means of phosphorimager analysis, fluorescence detectors, or labeled streptavidin. Comparative analyses can be performed on the same array (a grid of oligonucleotides or DNA fragments arrayed on a membrane or chip) by using cDNAs from different samples labeled with fluorescent dyes that are excitable at distinct wavelengths. Complementary RNA probes can also be used for DNA microarray experiments.

From the Department of Immunology, University College London, University of London, Institutes of Ophthalmology and Child Health and Moorfields Eye Hospital, NHS Trust.

Received for publication September 5, 2003; revised September 15, 2003; accepted for publication September 17, 2003.

Reprint requests: Santa Jeremy Ono, PhD, University College London, Institute of Ophthalmology, 11-43 Bath St, London, EC1V 9EL, Great Britain. *J Allergy Clin Immunol* 2003;112:1050-6.

© 2003 American Academy of Allergy, Asthma and Immunology
0091-6749/2003 \$30.00 + 0
doi:10.1016/j.jaci.2003.09.022

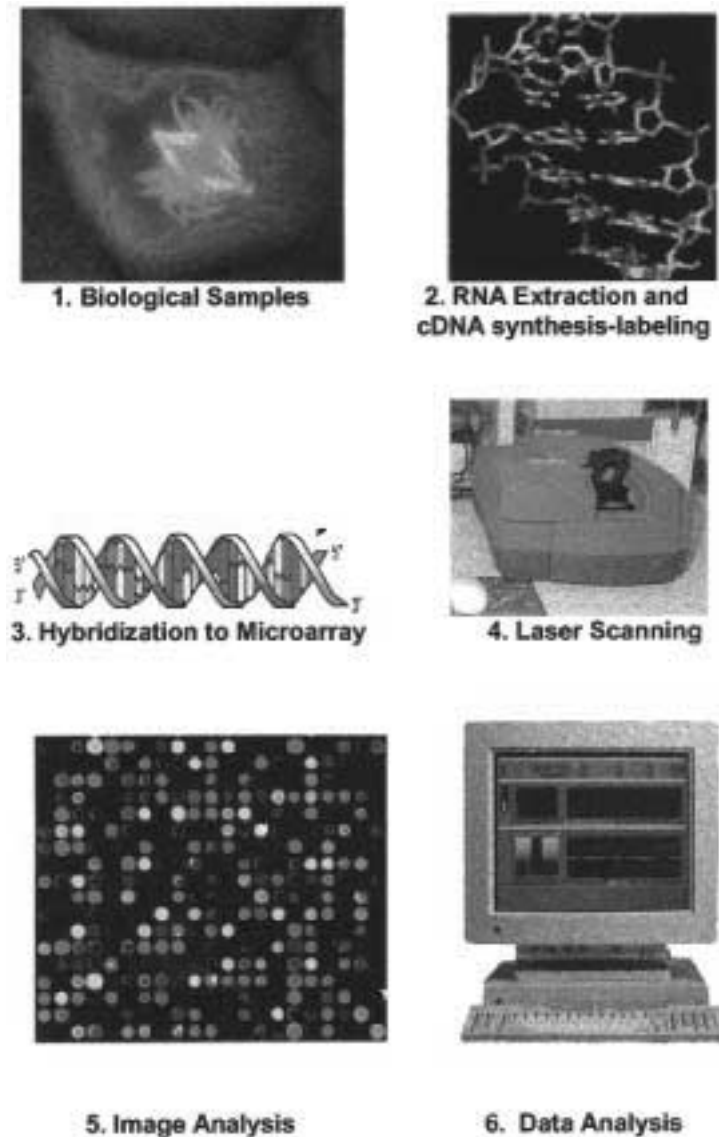


FIG 1. Typical sequence of events required for a DNA microarray experiment. 1, RNA is first isolated from the biologic samples. It is important to verify that the RNA is of good quality before proceeding further with the analysis. 2, RNAs are then reverse transcribed to generate labeled cDNA. This usually involves the incorporation of nucleotides with fluorescent dyes resonant at distinct frequencies. 3, The pools of labeled cDNA probes (labeled once again with different dyes) are then hybridized to either cDNA or oligonucleotides that have been spotted in an ordered, high-density array onto glass, metal, or nylon surfaces or chips. Once hybridization has occurred, the gene chips are washed to remove unbound probe. 4, The array is then imaged with a laser scanning fluorimeter. 5, A typical microarray image is shown. Each spot corresponds to a specific gene. The nucleic acid printed on the chip can be either cDNA generated by means of PCR or chemically synthesized oligonucleotides. Genes expressed in one sample are detected as green fluorescence, and genes expressed in the other sample are detected as red fluorescence. Genes expressed in both appear in different shades of yellow-orange. 6, The signals are quantified and downloaded to bioinformatics programs. These permit critical analyses of reproducibility and ranking of genes that are differentially expressed (by degree of differential expression).

Once labeled cDNAs are generated, they are hybridized simultaneously to the microarray. As indicated above, the microarray can be printed on various media (eg, glass, metal, or nylon) and consist of spots of genetic information (oligonucleotides or cDNA fragments) of

approximately 10 to 200 μm in size. PCR-generated cDNA arrays are usually printed onto glass or nylon surfaces. Oligonucleotide arrays are synthesized directly on a chip by using a technology pioneered by Affymetrix, Inc. After washing (to remove nonspecifically bound

labeled probe), the extent of hybridization to each gene spot is determined by means of image scanning. The data are then downloaded into a database for analyses. This analysis is quite complex and requires the assistance of experts familiar with the analysis of gene-profiling experiments. For additional information on DNA microarray technology, the reader is referred to <http://www.gene-chips.com>.

Now that we have discussed the basics of gene profiling, it is pertinent to ask why the current DNA microarray technology has not generated more important findings. In many of the reported studies, more than 40% to 50% of the analyzed DNA sequences have shown altered levels of mRNA expression under the experimental conditions used. To arbitrarily select those genes with the highest levels of detectable change is not necessarily appropriate. In retrospect, it is understandable why such a high percentage of genes have altered expression when the physiologic state of the cell is changed or manipulated. The phenotype that we can score is frequently the manifestation of many interrelated biochemical pathways within the cell, which cumulate to give a detectable change that might be many steps removed from the initial event. In our attempts to understand the development of a disease, many of the genes that are discovered to have altered expression are likely to be the consequence of the disease, such as a change in the growth or differentiation state of the cells involved, rather than being responsible for the initiating events that lead to the disease.

Perhaps the choice of experimental systems has to be blamed in at least some cases. For example, gross clinical specimens from patients are generally considered not ideal for molecular studies. Postmortem changes and normal tissue contamination, not to mention variance in the stages of disease, are factors that have to be considered. The argument that a high-throughput technology should be able to deal with individual sample differences is likely to be incorrect, particularly when it involves biologic samples.

What would constitute good experimental samples for expression analysis? If one's interest in using DNA microarrays is for more than a cataloging purpose, then "purity of cells" is an important consideration. Apart from changes in gene expression effected by intracellular signaling, one has to be reminded of molecular changes initiated by intercellular interactions. For example, the cross-talk between the epithelial nurse cells and the trafficking lymphocytes in the thymus or between the stoma and the epithelial compartment in the prostate induces significant changes in gene expression between cells of different lineages.⁸ These changes are particularly relevant late in the disease process and culminate in significant alterations in the cellular morphology of the tissue in question. Given this concern and our inability to assign certain molecular changes to specific cell types within tissues, efforts to ensure the purity of cells has to be taken seriously.

Apart from the use of tissue culture cells under differential treatment regimens, attempts to microdissect a clinical sample to select for a specific subpopulation of

cells and to remove extraneous contamination have generated mRNA of improved utility. For example, such a technology should prove rewarding for the analysis of ductal carcinoma in situ in the study of breast cancer.⁹ In addition, the use of tissues from animals carrying a transgene compared with those from control animals that do not might also be attractive in generating important findings.¹⁰ Mice transgenic for a rearranged T-cell receptor gene have been used successfully in vivo to generate memory T cells induced by the specific antigen.¹¹ In each of these scenarios, purity of cells is achieved by a different means and should facilitate efforts directed at expression profiling.

Although much of the early work on expression profiling has been carried out in developmental biology and oncology research, it is now coming to the fore in many fields. Expression profiling is emerging front and center in allergy research. Four examples of such studies are detailed in this issue of *The Journal of Allergy and Clinical Immunology* (JACI), and several others have or will appear shortly in other allergy-immunology journals. An article from Dean Metcalfe's group at the National Institutes of Health reports the gene profiling of RNA samples obtained from bone marrow mononuclear cells obtained from patients with mastocytosis and normal subjects.¹² This work is notable, not only in light of its findings but also its high standard of experimentation and data analysis. The analyses are performed on RNA from isolated bone marrow mononuclear cells, and this likely reduces noise from less focused analyses. Tremendous care is taken to control for variability between groups and between analyses. Finally, both hierarchical and K-means clustering analyses are used to identify 10 genes expressed in patients with mastocytosis. Attention to each of these criteria (see the comments in the remainder of this editorial) is what makes this such an exemplary study. From a scientific standpoint, the correlation of the expression levels of these genes with serum tryptase levels and the functions of several of the identified genes (eg, the genes encoding ATF-3, GATA-2, TSSC3, and maf-F) lend strong support to the reliability of the work and the likelihood that some of these genes participate in the pathogenesis of mastocytosis.

A second, well-defined study comes from the laboratory of Joshua Boyce at the Brigham & Women's Hospital.¹³ In this work the investigators test the hypothesis that T_H2 cytokines alter the expression of functionally relevant genes in activated mast cells. In a manner analogous to that previously tested by Steve Galli (Stanford University) and coworkers,¹⁴ Boyce compared activation-dependent gene expression of cord blood-derived human mast cells maintained in stem cell factor alone or when treated with IL-4, IL-5, or IL-9. The data clearly show that although the 3 ILs promote stem cell factor-dependent mast cell growth, they have very different effects on FcεRI-stimulated mast cell growth and transcriptome composition. IL-4 priming is particularly favorable for cytokine gene expression by FcεRI-stimulated mast cells. Thus the data fit well with the view that

tissue-resident mast cells might serve primarily in innate immunity, whereas IL-4-primed mast cells might undergo a program of differentiation that arms them for helminth infection or allergic inflammation. Thus this study begins with a clear hypothesis, is well designed, and leads to data that support the initial hypothesis.

Gene-profiling experiments of much more complex tissues can also lead to interesting findings but, by their very nature, are more difficult to interpret. In these cases it is particularly important for differences in gene expression to be verified, followed up, or both by alternative methods, such as real time-PCR or immunohistochemistry, or by studying mice deficient in a specific gene. Two examples of such challenging experiments can be found in this issue. Donald Leung and coworkers from the National Jewish Medical and Research Center profile the expression of 6000 genes in skin biopsy specimens from patients with atopic dermatitis and psoriasis.¹⁵ Interestingly, the data suggest that different panels of chemokines are expressed in skin biopsy specimens obtained from patients with these 2 conditions. Importantly, the data are followed up by real-time PCR analyses. The distinct panels of chemokines expressed are hypothesized to help explain the very different cell infiltrates observed in these disorders. The data presented will certainly form the foundation of future experiments to probe the role each of these genes plays in the pathogenesis of these diseases.

The second example of a more complex analysis comes from the group of Nicholas Beeson at the University of Michigan.¹⁶ In this case the investigators sought to identify genes that might contribute to the pathogenesis of nasal polyposis associated with allergic rhinitis. The expression of 12,000 genes was analyzed in nasal mucosa biopsy specimens from 10 patients with and 10 without nasal polyps. Once again, the study is well performed, with careful statistical analysis. The differentially expressed genes are verified by means of quantitative PCR analysis and, in one important case, by means of immunohistochemistry. The finding that the greatest differential expression is restricted to genes associated with neoplasia supports the view that nasal polyposis stems from deregulated cell growth. Finally, the identification of enhanced mammaglobulin expression in nasal polyps (also observed in breast neoplasia) underscores the view that this unknown gene product might participate in cell growth regulation.

Because gene-profiling experiments, such as those described above, generate considerable amounts of data and supporting analyses, it is anticipated that some of this information cannot be presented in print. However, such data are likely to be useful to the scientific community and will be available for readers from a repository within the JACI Web site. The editors have established the repository as a service to authors who have supplemental material relevant to a published article. We believe that it is important that JACI articles be self-contained and present all relevant information at a single source, the JACI Web site.

Having described strong studies in the field of allergy research dealing with both well-defined and complex systems, we would like to end this editorial with some general comments about study design and our view on the future evolution of gene profiling.

First, how does one select which expression-profiling methodology to use? Many of the more laborious approaches, such as differential display analysis, serial analysis of gene expression, and representational difference analysis, have by and large been replaced by the DNA microarray technology, which uses either oligonucleotides or cDNAs. The ease of use of this latter technology is offset, arguably, by its lack of sensitivity for lowly expressed transcripts. For example, many of the computationally predicted G protein-coupled receptor genes have undetectable expression, unless their expression profiles are determined by using PCR-based technologies in otherwise lower throughput approaches. In support of the DNA microarray approach is the fact that many, although not all, human genes have already been predicted by means of analysis of the human genome sequence, albeit not everyone is complete or correct. Hence the microarrays should be or can be fairly well represented in terms of gene coverage and will continue to improve as different spliced forms become discovered.

Another criticism of the high-density DNA microarrays, at least on an operational level, is the uncertainty of what to do next after those transcripts that are the most upregulated or downregulated have been identified. Because many of these selected genes are unrelated in function and are derived from a diversity of biochemical pathways, it is not always clear how to move the project forward. Fortunately, this is not always the case, as observed in some of the contributions found in this issue of JACI. It is with this concern that some investigators have opted to use low-density but focused arrays, such as those derived from specific pathways. For example, an assembly of tyrosine kinases, src-homology domain 2 (SH2) and src-homology domain 3 (SH3) proteins, might help define the molecular partners involved in specific pathways of signal transduction.

The following are additional issues that should be considered in the design of gene-profiling experiments:

1. Controls. (A) Have they been included and are they appropriate? (B) Have universal calibrators of gene expression been carefully analyzed? How do they compare with those proposed by the National Institute of Standards and Technology? For consistency and assurance that reports will be well developed, the editors believe that all authors submitting gene chip microarray studies, beginning in January 2004, should conform to the MIAME (Minimum Information About a Microarray Experiment) guidelines put forward by the Microarray Gene Expression Data Society (<http://www.mged.org/>).
2. Experimental design. (A) In patient sample-oriented studies, how extensive is the controlling for variability in gene expression between subjects in each group-

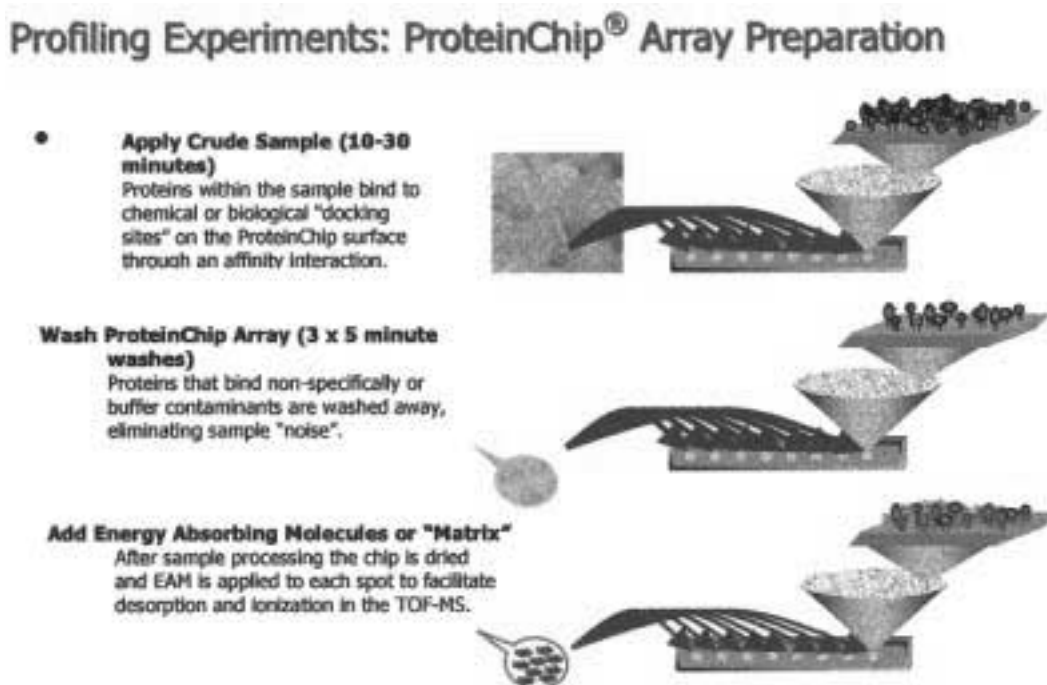


FIG 2. Isolation of subsets of the proteome through Ciphergen protein chip arrays. The technology makes use of chip-based chromatography coupled to linear time-of-flight spectroscopy. The protein chip is aluminum based with a derivatized surface. Different chemical surfaces (eg, hydrophobic and hydrophilic) are used to purify subsets of the proteome on the basis of affinity chromatography. After sample application, nonspecific interactions are broken by means of washing, the chip is dried, and energy-absorbing molecules are added to permit desorption and ionization of specifically bound proteins.

ing? (B) Has the effect of sex differences-cellularity been considered in their data interpretation? (C) Are there a sufficient number of subjects studied to be certain that expression differences are of biologic relevance to the condition or disease state? (D) Have the investigators tested for technical reproducibility on the same sample run at different times? (E) Has the work received appropriate bioinformatics support?

3. Data Analysis. (A) What confirming technology for gene expression changes has been included? There should be at least some RT-PCR, Northern blot, or Western blot confirmations of representative transcripts forming a focus for the publication. (B) How much thought has gone into evaluating the gene expression changes? Do the changes in expression have precedence in the literature? Do they make sense with respect to known biochemical pathways and interactions?

Apart from mRNA profiling, we should also note that technologies are now becoming available for protein profiling.^{17,18} Such proteomics approaches are necessary because mRNA accumulation is not always analogous to protein accumulation, and more importantly, posttranslational modifications, such as phosphorylation and dephosphorylation of target proteins, frequently play cru-

cial roles in regulating specific functions in the cell. Protein profiling is likely to be more difficult to perform, demand higher technical skills, and be more expensive to carry out. However, there are remarkable advances in proteomics technology that will undoubtedly facilitate such studies in the future. For example, although proteome analysis has traditionally relied on the laborious and demanding technique of 2-dimensional gel electrophoresis of proteins, new technologies pioneered by companies such as Ciphergen Biosystems, Inc (Fremont, Calif; www.ciphergen.com) permit the systematic purification of subsets of the proteome through protein chips. These make use of chips coated with specific chemistries, which permit the isolation of protein subsets (Fig 2). These are then resolved by using laser desorption and ionization. Ionized proteins are detected, and their mass is accurately determined by using time-of-flight mass spectrometry (Fig 3). Such proteomics approaches have already had a significant effect in the study of complex diseases, such as AIDS and cancer.^{19,20}

Certainly we are on the right track. As technologies continue to be developed, the experimental findings will become more relevant and, no doubt, have a more significant effect. There is also a future potential of using gene and protein profiling in clinical diagnosis and management. Either broad fingerprints of transcriptomes or spe-

Surface Enhanced Laser Desorption/Ionization (SELDI)- Derivatized 'Affinity' surfaces

- Retained proteins are "eluted" from the ProteinChip Array by Laser Desorption and Ionization
- Ionized proteins are detected and their mass accurately determined by Time-of-Flight Mass Spectrometry

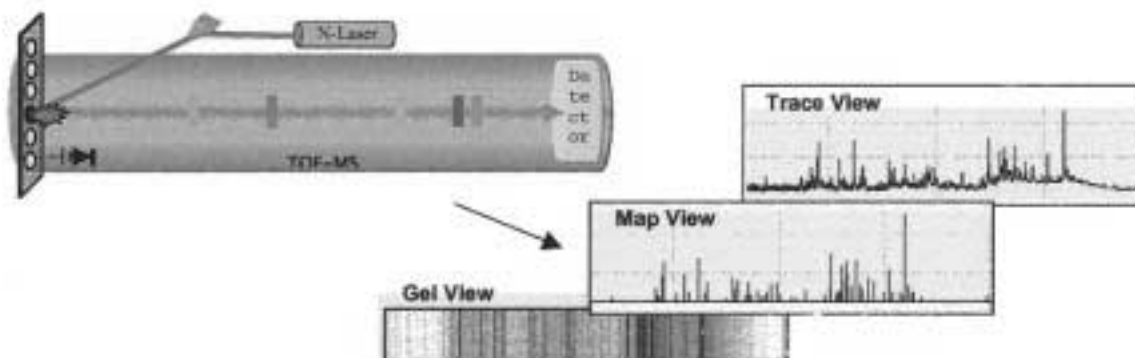


FIG 3. Identification of specifically bound proteins and analysis of expression levels. Proteins from different biologic samples are effectively eluted from the protein chips by means of surface-enhanced laser desorption/ionization (SELDI). Ionized proteins are detected, and their masses are determined by means of time-of-flight mass spectrometry. The data are then presented either as map or trace views or in the form of a gel. Cluster analysis can identify proteins that are upregulated or downregulated in particular samples. The identities of differentially expressed proteins can be determined by means of peptide sequencing. This can either involve chip tryptic digestion or tryptic digestion of SDS gel pieces, followed by on-chip peptide mapping.

cific diagnostic tests for the expression of a restricted set of genes might eventually be of value in diagnosis or clinical management (eg, in the case of mastocytosis).¹⁹ However, the developing stage of the field (as outlined above) indicates that further developments are essential before gene profiling can be used confidently in the clinical management of allergic disease.

Note: As indicated earlier in this commentary, readers and potential authors of gene-profiling experiments should refer to the JACI Online Repository for their future publication of gene array data and for the viewing of supplemental information for published articles.

We thank Dr Ben Reed from CIPHER for Figs 2 and 3.

REFERENCES

- Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 1992;257:967-71.
- Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci U S A* 1996;93:10614-9.
- Morrison DA, Ellis JT. The design and analysis of microarray experiments: applications in parasitology. *DNA Cell Biol* 2003;22:357-94.
- Lou XJ, Schena M, Horrigan FT, Lawn RM, Davis RW. Expression monitoring using cDNA microarrays. A general protocol. *Methods Mol Biol* 2001;175:323-40.
- Covell DG, Wallqvist A, Rabow AA, Thanki N. Molecular classification of cancer: unsupervised self-organizing map analysis of gene expression microarray data. *Mol Cancer Ther* 2003;2:317-32.
- Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, et al. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002;420:520-62.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860-921.
- Aguilar LK, Aguilar-Cordova E, Cartwright J Jr, Belmont JW. Thymic nurse cells are sites of thymocyte apoptosis. *J Immunol* 1994;152:2645-51.
- Middleton LP, Amin M, Gwyn K, Theriault R, Sahin A. Breast carcinoma in pregnant women: assessment of clinicopathologic and immunohistochemical features. *Cancer* 2003;98:1055-60.
- Arlotta P, Tai AK, Manfioletti G, Clifford C, Jay G, Ono SJ. Transgenic mice expressing a truncated form of the high mobility group I-C protein develop adiposity and an abnormally high prevalence of lipomas. *J Biol Chem* 2000;275:14394-400.
- Jacob J, Baltimore D. Modelling T-cell memory by genetic marking of memory T cells in vivo. *Nature* 1999;399:593-7.
- D'Ambrosio C, Akin C, Wu Y, Magnusson MK, Metcalfe DD. Gene expression analysis in mastocytosis reveals a highly consistent profile with candidate molecular markers. *J Allergy Clin Immunol* 2003;112:1162-70.
- Lora JM, Al-Garawi A, Pickard MD, Price KS, Bagga S, Sicoli J, et al. FcεRI-dependent gene expression in human mast cells is differentially controlled by T helper type 2 cytokines. *J Allergy Clin Immunol* 2003;112:1119-26.
- Tsai M, Tam SY, Galli SJ. Distinct patterns of early response gene expression and proliferation in mouse mast cells stimulated by stem cell factor, interleukin-3, or IgE and antigen. *Eur J Immunol* 1993;23:867-72.

15. Nomura I, Gao B, Boguniewicz M, Darst MA, Travers JB, Leung DYM. Distinct patterns of gene expression in the skin lesions of atopic dermatitis and psoriasis: a gene microarray analysis. *J Allergy Clin Immunol* 2003;112:1195-202.
16. Fritz SB, Terrell JE, Conner ER, Kukowska-Latallo JF, Baker JR. Nasal mucosal gene expression in patients with allergic rhinitis with and without nasal polyps. *J Allergy Clin Immunol* 2003;112:1057-63.
17. Toda M, Ono SJ. Genomics and proteomics of allergic disease. *Immunology* 2002;106:1-10.
18. Ono SJ. Molecular genetics of allergic diseases. *Annu Rev Immunol* 2000;18:347-66.
19. Zhang L, Yu W, He T, Yu J, Caffrey RE, Dalmasso EA, et al. Contribution of human alpha-defensin 1, 2, and 3 to the anti-HIV-1 activity of CD8 antiviral factor. *Science* 2002;298:995-1000.
20. Wulfkuhle JD, Paweletz CP, Steeg PS, Petricoin EF 3rd, Liotta L. Proteomics approaches to the diagnosis, treatment, and monitoring of cancer. *Adv Exp Med Biol* 2003;532:59-68.