Allergology International. 2005;54:345-349

REVIEW ARTICLE

Microarray as a Standard Laboratory Technique

Hirohisa Saito[1] and Kenji Matsumoto[1]

ABSTRACT

Rapid advances in microarray technologies and the related computational analyses have led to a paradigm shift in biological investigation. The number of studies using the microarray technology is logarithmically increasing. Thus, this technology is becoming a standard laboratory technique. Here, we introduce some successful microarray studies in relation to allergic diseases, and discuss why these authors succeeded, and discuss the strategy for microarray studies from data generation to data analysis.

KEY WORDS

DNA chip, microarray, system biology, transcriptome

INTRODUCTION

Recent advances in technology such as gene-targeting animal models or cultivation of single cell types enable us to understand the pathology of allergic diseases at molecular levels. However, it had not been possible to examine the levels of total molecules playing some roles in these diseases, especially regarding the whole protein molecules present in a cell or a body (the proteome). Microarrays, also called DNA chips, were first lauded in public by former USA President Clinton in his State-of-the-Union address in 1999,[1] and then have attracted tremendous interest among biologists and industries. Microarrays have been mainly used to examine the transcriptome (the whole transcripts present in a cell or a body), although some microarrays are now becoming available for detecting whole genetic variations present in the genome. The application of microarrays to the proteome is not yet practical when considering its meaning (the whole protein information present in a cell or a body) . In contrast, the number of studies using the microarray technology for the transcriptome is logarithmically increasing. Thus, this technology is becoming a standard laboratory technique. Here, we introduce some successful microarray studies in relation to allergic diseases and discuss why these authors succeeded.

GUIDELINES FOR MICROARRAY STUDIES

Microarray technologies can be classified into two types. One consists of oligonucleotide microarrays that use synthesized oligonucleotide as probes, while the other consists of cDNA microarrays that use whole cDNA molecules with irregular lengths as probes. Compared with oligonucleotide microarrays, the results of cDNA microarray assays are less reproducible because both the target RNA and probe sizes are irregular, resulting in variable binding capacity.[2] Thus, questions were raised in 2002 regarding the reproducibility of studies performed using microarrays, especially for cDNA microarrays.[3] Then, a document was generated outlining the minimal information that should be reported about a microarray experiment to enable its unambiguous interpretation and reproduction when submitting a microarray study to most journals (http://www.mged.org/Workgroups/MIAME/miame_checklist.html).[4] The guidelines demand the investigators to describe the following six factors, *i.e.*, experimental design, array design, sample types , hybridization methods , measurement methods, and normalization controls, in detail when submitting reports to journals (MGED Guide to authors, editors and reviewers of microarray gene expression papers). These microarray guidelines aim mainly at reproducing the results obtained using cDNA microarrays that had been hand-made in indi-

[1]Department of Allergy and Immunology, National Research Institute for Child Health & Development, Tokyo, Japan.
Correspondence: Hirohisa Saito, M.D., Ph.D., Department of Allergy and Immunology, National Research Institute for Child Health & Development, 2–10–1 Okura, Setagaya-ku, Tokyo 157–8535, Japan.
Email: hsaito@nch.go.jp
Received 30 March 2005.

vidual laboratories. For the microarray studies dealing with inflammatory tissues as in the field of allergy, the journal Arthritis and Rheumatism generated its own guideline.[5] These criteria were as follows: (1) Reproducibility must be demonstrated, including rigorous evaluation of the run-to-run variability for each gene. (2) Detailed statistical analysis is required, including appropriate corrections for repeated or multiple measurements. (3) Homogeneous cell populations should optimally be studied, to reduce the complexity of analysis. (4) A non-array method must confirm changes in the expression of key genes. These guidelines are applied to both types of microarrays.

EXPERIMENTS USING WELL-CONTROLLED ANIMAL MODELS

Most successful microarray-based results, especially in allergy-related fields, have been obtained in studies dealing with well-controlled animal model experiments. For example, Karp *et al.*[6] obtained eight new strains of mice by crossing A/J mice, which tend to have antigen-specific airway hypersensitivity, with C3H/HeJ mice. They used GeneChip (Affymetrix, Santa Clara, CA, USA), an oligonucleotide microarray platform, to examine >7000 genes expressed in the lungs after sensitization and challenge with allergens. They found that a genetic polymorphism of complement C5a, known to induce mast cell degranulation, was associated with the allergen-induced airway hypersensitivity. Zimmerman *et al.*[7] created an asthma model by sensitizing Balb/c mice with OVA and Aspergillus fumigatus and used GeneChip to analyze the lungs of these mice and controls after allergen challenge. As a result, they found 291 genes whose expression levels were commonly increased by challenge with both antigens. Among those genes, arginase I, arginase II and cationic acid transporter 2 were found to be dramatically increased. They suggested the importance of these molecules in the pathology of asthma by showing that they were also highly expressed in human asthmatic airway epithelium. It should be noted that these two groups have confirmed their results to be highly reproducible by comparing two samples after allergen challenge from each experimental condition.

EXPERIMENTS USING A CULTURED PURE CELL TYPE

There have been many successful microarray studies using cultured cells with or without stimulation. Yuyama *et al.* applied microarray analyses to find genes regulated by IL-4 and IL-13 in human bronchial epithelial cells. They identified a subset of disease-relevant genes, squamous cell carcinoma antigen-1 (SCCA1) and SCCA2, the cysteine and serine protease inhibitors by comparison with cDNA libraries derived from normal and asthmatic bronchial biopsies.[8] Qiao *et al.* used microarray to identify 8 upregulated genes in differentiated (type I) alveolar epithelial cells compared with type II primary type cells. They confirmed three of these genes, PAI-1, P2X4, and P15INK4B by using Northern blotting analysis and Western blotting analysis for corresponding proteins.[9] Zhang *et al.* also used oligonucleotide microarray to examine the effect of anti-viral agent ribavirin on gene expression in respiratory syncytial virus-infected human A549 pulmonary type II epithelial cells. They found that ribavirin up-regulated the transcriptional activity of a reporter gene selectively driven by the IFN-stimulated response element (ISRE), indicating the enhancement of IFN-related anti-viral proteins by the drug.[10] Cho *et al.* used oligonucleotide microarray to examine the up-regulated genes in activated human mast cells, and identified activin A to be one of the most up-regulated genes. Activin A was found to be expressed equally by mouse mast cells. They found that mast cell-derived activin A play an important role in airway smooth muscle cell hypertrophy by measuring activin A levels in asthma model mice and mast cell-deficient W/W^v mice.[11]

Wang *et al.*[12] and Okumura *et al.*[13] used oligonucleotide microarray to examine the up-regulated genes expressed by activated mast cells. These two groups separately identified amphiregulin as a transcript that is markedly increased following aggregation of FcεRI. Amphiregulin induced tissue remodeling, *i.e.*, proliferation of lung fibroblasts and marked induction of MUC5AC transcripts in a human respiratory epithelial cell line. Both groups showed that amphiregulin-positive mast cells are increased in the airways of asthmatics. These investigators employed culture-derived human mast cells for microarray as the first screening and confirmed the *in vivo* expression and function of this up-regulated gene using human tissue samples.

Thus, microarray experiments seem to be highly reproducible when well-controlled animal models or pure cells are employed.

EXPERIMENTS USING MIXED CELL POPULATIONS

The gene expression levels obtained using GeneChip are usually between 1 and approximately 25,000. However, the expression levels of certain abundantly or scantly expressed genes do not show linear relationships, so that the actual dynamic range is considered to be only 10^2. In our early GeneChip study using human clinical samples, we identified several increased transcripts in peripheral blood mononuclear cells (PBMC) obtained from severe atopic patients by calculating the ratio of their expression levels.[14] However, the ratio should have been more carefully calculated when the values were small even when accompanied by the presence call, which means an analysis

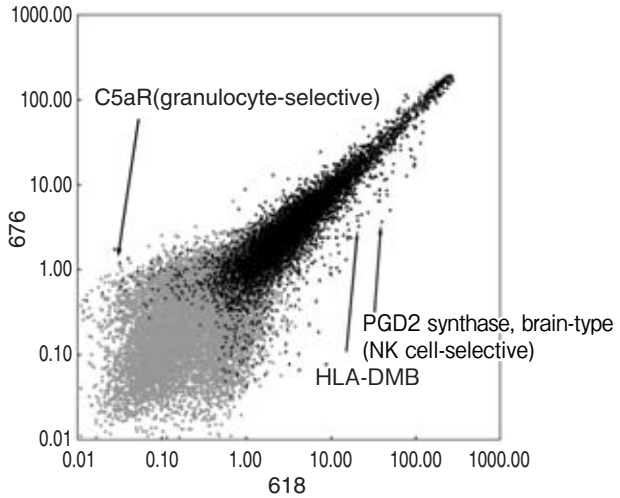

Fig. 1 A scatter plot of gene expression levels of CD4+ T cells from two normal donors (sample ID #676 and #686) on GeneChip U133A Arrays (Affymetrix). Black dots represent genes expressed with presence calls in at least either one sample and gray dots represent genes expressed with absence or marginal calls in both samples. Brain-type PGD_2 synthase (PGDS) was highly expressed by only one donor CD4 sample, while C5aR was only expressed with a presence call by another sample. Since C5aR is judged to be specific in granulocytes[20] and PGDS is selectively expressed by NK cells compared with other cell types (unpublished observation), these transcripts may be artifactually present among the whole CD4 population. The experimental design was previously described by us.[20]

that is statistically significant. Increased transcripts in PBMC included defensin-1 and ribosomal protein L 37. However, these genes may have been derived from contamination of the PBMC population by basophils, because our recent results regarding cell type-specific transcriptome data[15] found these genes to be highly expressed by basophils. Thus, contamination by a very small population of a different cell type having a certain highly expressed transcript may cause an artifactual presence of the transcript in the whole population even where the major cell type lacks it (Fig. 1).

Changes in cell populations are especially crucial in microarray analysis of tissues characterized by recruitment of inflammatory cells. Indeed, arginase I is highly expressed by neutrophils among the human leukocyte types. Up-regulation of transcriptional levels in a crude tissue is often simply caused by an increase in recruitment of a certain inflammatory cell type. Therefore, in order to avoid difficulty in interpreting the results, it is important, for comparison, to purify the target cell population as much as possible. However, mRNA is unstable, meaning that complicated procedures for purification should be avoided.

Guajardo *et al*. have recently succeeded in identifying nasal epithelial-specific up-regulated genes by differentiating the cell types obtained by scraping the nasal membranes of allergic rhinitics.[16]

However, computational identification of the cell-type specificity may be preferable when a certain important transcript is found in crude tissue samples when cell types can not be identified. For such purpose, microarray data may be used by downloading from a public on-line database.

USE OF PUBLIC DATABASES AND BIOINFORMATICS

It has been useful and is becoming increasingly necessary to obtain microarray data from public databases, especially for comparison of a certain transcript with other cell types. Regarding asthmatic CD4 T cells, all the microarray data can be downloaded from Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/). The experiments were done using oligonucleotide microarrays (Affymetrix GeneChip U133A) and CD4 T cells obtained from 5 patients with severe asthma, 15 patients with mild asthma and 6 controls. The Laboratory for System Biology and Medicine at the University of Tokyo have made available the microarray data of 55 different types of human tissue cell types which were obtained using Affymetrix GeneChip U133A (http://www.lsbm.org/site_e/index.html). It is possible to combine the data obtained from public databases and our own data. For analyzing large-scale gene expression data, hierarchical clustering analysis is becoming a common laboratory technique.[17,18] Such a clustering tool can be freely obtained at http://rana.lbl.gov/EisenSoftware.htm. It should be noted that clustering analysis is just bringing one of the possibilities to profile the data but not a statistical significance to the results. Kaminski and Friedman have introduced in their review article[19] some methods for identifying genes with statistically significant changes in gene expression. Statistical calculation will be required after suitable normalization of data obtained from different laboratories.

APPLICATIONS TO SYSTEM BIOLOGY

cDNA microarrays which use largely unknown sequences or differential display PCR which uses random primers had been useful for searching unidentified genes compared with oligonucleotide microarrays which use synthesized oligonucleotides with known sequences as probes. However, after complete sequencing of the human genome or mouse genome, we understand that the number of genes is limited to only <30,000. One of the oligonucleotide microarrays, GeneChip, covers the measurement of >45,000 transcripts at once. That means that we have obtained a tool which can measure nearly the total "transcriptome".

Microarray studies have been appreciated as a tool which can measure large-scale gene expression at once. However, this is changing now that the tool for understanding all transcriptional information has been obtained. Now, we may say that there is nothing left in the genome. It will soon be applied to pharmaceutical development. For example, theoretically, we will be able to predict all transcripts related to an adverse effect of a certain drug, and change the drug structure in order to avoid it. For example, we have identified mast cell-, eosinophil- and basophil-specific transcripts, which are expected to be potential therapeutic targets for allergic diseases because these cells play crucial roles in allergic inflammation.[20] A certain gene that is highly expressed in mast cells, for example, can be easily found by searching the mast cell transcriptome. Comparison with the transcriptomes expressed by other cell types, however, may reveal that said gene is not truly mast cell-specific. Thus, it is especially important to elucidate information regarding the cell type-specificity when developing a new drug. Such information may help to predict drug-related adverse events caused by interaction of a drug with responsible molecules present in important organs. In the future, the safety of candidate drugs could be evaluated by comparing their efficacy on these granulocytes with their toxicity to vital organs.

The increasing prevalence of allergic diseases in developed countries is considered by many to be caused, at least in part, by rapid improvement of hygiene. In human beings, the immune system developed as an ingenious device for defending against frequent attacks by microbes. Therefore, our immune system seems to have become deranged in our recent, unprecedented hygienic environment. It is now necessary to understand all functional elements comprising the immune system, not just a single molecule present in an immunocyte working in our immune system. Until now, microarrays have been used as a high-throughput assay method. However, it should be stressed that microarrays can detect whole transcripts present in a cell and can be used for understanding system biology.[21] It is anticipated that as this and related technology continues to advance, it will help our understanding of deranged human immunity as a system.

CONCLUSION

Here, we have introduced some successful microarray studies. Rapid advances in microarray technologies and the related computational analyses have led to a paradigm shift in biological investigation, such that the bottleneck in research is shifting from data generation to data analysis. As Kaminski and Friedman described,[19] one of the most painful stages in the analysis of the results of microarray experiments is the postcomputational stage. There is no final resolution for this stage in the analysis. Moreover, microarray results often demand more knowledge about medicine and biology from us than before. Therefore, collaboration is necessary by sharing large-scale data which may bring different answers to each investigator working in different research fields.

REFERENCES

1. Editorial. Getting hip to the chip. *Nat. Genet.* 1998;**18**: 195-197.
2. Southern E, Mir K, Shchepinov M. Molecular interactions on microarrays. *Nat. Genet.* 1999;**21**:5-9.
3. Editorial. Microarray standards at last. *Nature* 2002;**419**: 323.
4. Brazma A, Hingamp P, Quackenbush J *et al.* Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* 2001; **29**:365-371.
5. Firestein GS, Pisetsky DS. DNA microarrays: boundless technology or bound by technology? Guidelines for studies using microarray technology. *Arthritis Rheum.* 2002; **46**:859-861.
6. Karp CL, Grupe A, Schadt E *et al.* Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. *Nat. Immunol.* 2000;**1**:221-226.
7. Zimmermann N, King NE, Laporte J *et al.* Dissection of experimental asthma with DNA microarray analysis identifies arginase in asthma pathogenesis. *J. Clin. Invest.* 2003;**111**:1863-1874.
8. Yuyama N, Davies DE, Akaiwa M *et al.* Analysis of novel disease-related genes in bronchial asthma. *Cytokine* 2002; **19**:287-296.
9. Qiao R, Zhou B, Liebler JM, Li X, Crandall ED, Borok Z. Identification of three genes of known function expressed by alveolar epithelial type I cells. *Am. J. Respir. Cell Mol. Biol.* 2003;**29**:98-105.
10. Zhang Y, Jamaluddin M, Wang S *et al.* Ribavirin treatment up-regulates antiviral gene expression via the interferon-stimulated response element in respiratory syncytial virus-infected epithelial cells. *J. Virol.* 2003;**77**:5933-5947.
11. Cho SH, Yao Z, Wang SW *et al.* Regulation of activin a expression in mast cells and asthma: its effect on the proliferation of human airway smooth muscle cells. *J. Immunol.* 2003;**170**:4045-4052.
12. Wang SW, Oh CK, Cho SH *et al.* Amphiregulin expression in human mast cells and its effect on the primary human lung fibroblasts. *J. Allergy Clin. Immunol.* 2005;**115**: 287-294.
13. Okumura S, Sagara H, Fukuda T, Saito H, Okayama Y. FcεRI-mediated amphiregulin production by human mast cells increases mucin gene expression in epithelial cells. *J. Allergy Clin. Immunol.* 2005;**115**:272-279.
14. Heishi M, Kagaya S, Katsunuma T *et al.* High-density oligonucleotide array analysis of mRNA transcripts in peripheral blood cells of severe atopic dermatitis patients. *Int. Arch. Allergy Immunol.* 2002;**129**:57-66.
15. Nakajima T, Iikura M, Okayama Y *et al.* Identification of granulocyte subtype-selective receptors and ion channels by high-density oligonucleotide probe array. *J. Allergy Clin. Immunol.* 2004;**113**:528-535.
16. Guajardo JR, Schleifer KW, Daines MO *et al.* Altered gene expression profiles in nasal respiratory epithelium reflect stable versus acute childhood asthma. *J. Allergy Clin. Immunol.* 2005;**115**:243-251.

17. Troyanskaya OG, Garber ME, Brown PO *et al*. Nonparametric methods for identifying differentially expressed genes in microarray data. *Bioinformatics* 2002;**18**:1454-1461.
18. Eisen MB, Spellman PT, Brown PO *et al*. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 1998;**95**:14863-14868.
19. Kaminski N, Friedman N. Practical approaches to analyzing results of microarray experiments. *Am. J. Respir. Cell Mol. Biol.* 2002;**27**:125-132.
20. Nakajima T, Iikura M, Okayama Y *et al*. Identification of granulocyte subtype-selective receptors and ion channels by using a high-density oligonucleotide probe array. *J. Allergy Clin. Immunol.* 2004;**113**:528-535.
21. Saito H. Translation of the human genome into clinical allergy. Part 2. *Allergol. Int.* 2004;**53**:87-92.