Exploiting DNA Microarrays in Renal Transplantation
Mei-Sze Chua and Minnie Sarwal
*Graft* 2002; 5; 223

The online version of this article can be found at:
http://gft.sagepub.com/cgi/content/abstract/5/4/223

Published by:
SAGE Publications
http://www.sagepublications.com

Additional services and information for *Graft* can be found at:

Email Alerts: http://gft.sagepub.com/cgi/alerts

Subscriptions: http://gft.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav
Exploiting DNA Microarrays in Renal Transplantation

Mei-Sze Chua and Minnie Sarwal

The advent of DNA microarray technology is revolutionizing the way biomedical research is conducted. Many disciplines have started to apply this high-throughput screening of gene expression to improve their understanding of the molecular basis and to aid in more accurate classification, diagnosis, and/or prognosis of human diseases. We are doing pioneering work in using cDNA microarrays to generate highly informative gene expression profiles from pediatric patients undergoing renal transplantation. Specifically, we will be able to address important issues related to acute and chronic allograft rejection and/or dysfunction, as well as improve on current steroid-based immunosuppressive therapy. This article will provide an overview of DNA microarray technology and discuss how this powerful technique can be exploited in the field of renal transplantation.

Introduction
The analysis of gene expression using DNA microarrays is rapidly transforming many interdisciplinary fields of science and medicine. The human genome project has identified thousands of genes that may potentially be useful for the diagnosis, prognosis, and/or therapy of human diseases; this exciting and challenging task is greatly facilitated by the advent of DNA microarray technology. In general, the term DNA microarray refers to a high-density array of oligonucleotide or complementary DNA (cDNA) immobilized onto a structural support; it is based on the principle that complementary sequences of DNA can be used to probe and hybridize to the immobilized DNA molecules. Thus, unlike traditional methods of quantifying mRNAs by northern blotting or quantitative PCR, which can only measure a few genes at a time, DNA microarrays allow the rapid and accurate analysis of global gene expression in an overnight hybridization. Since current technology can produce a microarray containing up to 50,000 genes on a single microscopic glass slide, it is possible to analyze almost all sequenced genes in the human genome at one time (with some redundancy of the gridded clones). Because of the extreme versatility of this technology, it can be readily adapted for use in many investigations, ranging from the basic (e.g., comparative sequence analysis) to clinical sciences (e.g., tumor classification) (Fig. 1).

One of the most widespread biomedical applications of DNA microarrays is in cancer research, where it is used to provide valuable insights into the molecular mechanisms of cancer development and the clinical classification of molecular subclasses of malignancies, and for the identification of prognostic markers and candidate gene targets for the design of novel therapeutic agents. For example, the National Cancer Institute has used microarrays to characterize their panel of 60 human tumor cell lines according to their tumor origin, and have also studied the effects of more than 70,000 chemical compounds on these cell lines with the goal of identifying genes responsible for drug sensitivity or resistance. Other biomedical disciplines have been slower to adopt this approach; the areas of immunology, nephrology, and circulatory research are only beginning to see the benefits of using such high-throughput analysis of gene expression. There is also extensive application of

ABBREVIATIONS
aRNA amplified antisense RNA
CAN Chronic allograft nephropathy
cDNA Complementary DNA
RT-PCR Reverse transcription-PCR
SAM Significance Analysis of Microarrays
SOMs Self-Organizing Maps
DNA microarrays to pharmacology, specifically in terms of toxicology, drug discovery, and drug design. In particular, gene expression “fingerprints” for standard toxicological agents may be generated to aid the identification of potential toxic properties of novel agents. Last but not least, it is also an attractive and powerful technique in the emerging field of pharmacogenomics, where the main concern is the identification and/or development of safer and more effective therapeutics based on an individual’s specific genetic profile.

In this review, we will add on to the expanding literature on DNA microarray application by discussing its potential benefits in the area of renal transplantation.

DNA Microarray Methodology

Two main types of DNA microarrays, oligonucleotide and cDNA arrays, exist depending on the way they are made. In the 1st method, oligonucleotides corresponding to a short fragment (20-25 nucleotides) of single-stranded DNA are directly synthesized on a solid surface by photolithography; alternatively, presynthesized oligonucleotides can also be printed onto glass slides by mechanical spotting or ink-jet technology. This method allows very high density arrays (>250 000 oligonucleotide spots/cm²) and a large number of identical arrays to be made, but it is extremely expensive and lacks flexibility in design. The 2nd method involves spotting solutions of individual PCR-amplified double-stranded cDNA fragments (500-600 base pairs) in small spots on microscopic glass slides using capillary action-based printing tips. Current technology allows over 50,000 genes to be printed on a single glass slide, and each array may be custom designed for specific investigations. Such cDNA arrays are by far the more frequently used in academic laboratories.

Summary/main points:

- DNA microarrays allow high-throughput, parallel screening of gene expressions across multiple experimental sets in an overnight hybridization.
- Microarray technology is extremely powerful and versatile and can be applied to many disciplines ranging from the basic to clinical sciences.
- The tremendous amount of data points generated from DNA microarrays requires a combination of computational analytical programs to generate meaningful comparisons.
- In renal transplantation, we can generate highly informative gene expression profiles from specific patient groups, thereby allowing us to address important issues related to acute and chronic allograft rejection and/or dysfunction, as well as improve on current steroid-based immunosuppressive therapy.
Most spotted cDNA microarrays use a 2-color hybridization scheme to visualize and measure the gene expression levels reproducibly when comparing multiple samples (Fig. 2). The mRNA extracted from test and reference samples are 1st fluorescently labeled during reverse transcription prior to hybridization. Commonly used dyes include Cy3 and Cy5 as well as fluorescein and rhodamine; the fluorophore used should have a narrow excitation and emission peak, a high level of photon-emission and resistance to photobleaching. Typically, test sample cDNA is labeled red (with Cy5), whereas reference sample cDNA is labeled green (with Cy3); they are then competitively hybridized to each cDNA microarray and the ratio of red to green fluorescence (R/G ratio) measured by scanning the microarray slide using two different wavelengths corresponding to the dyes used. The relative abundance of test versus reference mRNA is therefore indicated by the color of the individual spots: red if test mRNA is more abundant than reference mRNA, green if reference mRNA is more abundant than test mRNA, and yellow if they are equally abundant.

In our investigations, we use a common reference sample of constant composition throughout all experiments; large numbers of individual test samples can then be compared with one another in ratio units relative to the common reference. The level of red versus green fluorescence at a particular cDNA spot reflects the level of mRNA for this gene in the test sample compared to the common reference sample. By repeating this process for multiple samples while keeping the common reference constant, different samples can be compared with each other using the reference sample as a consistent comparison standard. Our common reference consists of a mixture of diverse cell types contained in 11 human cell lines derived from a wide variety of human malignancies. The common reference will show some minimum level of fluorescence intensity for the majority of the genes spotted on the microarray, thereby allowing a meaningful ratio of red to green signal to be obtained.

**Data Analysis**

Since an overnight hybridization of a single microarray gives rise to thousands of data points, a typical experiment can produce a tremendous amount of complex data sets that has to be interpreted; the use of data analysis software is therefore essential to achieve meaningful analyses. Data analysis tools are currently available from either public sources (e.g., http://genome-www4.stanford.edu/MicroArray/SMD/restech.html) or commercial suppliers (such as GeneSpring from SiliconGenetics).

Three broad steps are involved in data analysis: data normalization, data filtering, and pattern identification. Data must first be normalized to effectively compare expression levels; it is then reduced by eliminating genes expressed below a defined
threshold value (in our case, we use a 4-fold threshold value, i.e., genes that are expressed at less than 4-fold from the common reference were not considered in further analyses). Finally, clustering and visualization programs such as hierarchical clustering and \( K \)-means clustering are used to generate fundamental gene expression patterns inherent in the massive data sets, thereby allowing possible biological or clinical relevance to be inferred. In hierarchical clustering, similarity or distance measures are used to distinguish between groups; this is a form of unsupervised method of data analysis where there is no prior knowledge of the true functional classes. This is opposed to the supervised method of data analysis, which specifies which data should cluster together through some sort of training (e.g., Support Vector Machines) (Fig. 3).

One of the most reliable methods for studying gene expression patterns is hierarchical clustering. Genes with similar expression profiles across a set of experimental samples are clustered together on the vertical axis, and in the same manner, experimental samples are clustered together on the horizontal axis based on their similarity in expression behavior across a specified set of genes. Thus, the data are displayed in a tabular form with each row representing the data for a single gene and each column representing the data for a single experimental sample (Fig. 4). A hierarchical tree or dendogram is displayed next to the clustered genes and above the clustered experimental samples to graphically denote the degrees of relatedness between adjacent samples and genes: the closer the 2 samples are together, the greater the similarity between them. In this colored tabular scheme, the fold deviation from average expression of each gene across the set of samples studied is represented by a range of colors from red (above-average level of mRNA present for that gene) through black (average expression of that gene) to green (below-average level of mRNA present for that gene). Through such clustering of data points, groups of genes that are either highly expressed or relatively underexpressed in different subsets of samples are highlighted. The biological or clinical significance of differential gene expression across experimental samples may then be inferred and further tested.

Hierarchical clustering has a number of shortcomings that can be overcome by the implementation of Self-Organizing Maps (SOMs). SOMs have a number of features that make them well suited for the clustering and analysis of gene expression patterns.
patterns: in contrast to the rigid structure of hierarchical clustering, SOMs allow one to impose partial structure on the clusters and facilitate easy visualization and interpretation; they also have good computational properties and are easy to implement, reasonably fast, and scalable to large data sets.

Cluster analysis of microarray data produces coherent patterns of gene expression but provides little information about statistical significance. Conventional statistical $t$ tests are limited in that they provide the probability that a difference in gene expression occurred by chance in only a small numbers of genes. A statistical method, Significance Analysis of Microarrays (SAM), was therefore developed to be specifically adapted for use in microarrays. SAM allows identification of genes with statistically significant changes in expression by assimilating a set of gene-specific $t$ tests. Each gene is assigned a score on the basis of its change in gene expression relative to the standard deviation of repeated measurements for that gene; genes with scores greater than a threshold are considered potentially significant. The cutoff for significance is determined by adjusting a delta parameter, chosen by the user based on the false positive rate. This analytical method is robust, straightforward, and can be adapted to a broad range of experimental situations; it is available for use at http://www-stat-class.stanford.edu/SAM/SAMServlet.

The use of Singular Value Decomposition additionally provides a useful mathematical framework for processing and modeling genome-wide expression data. In this analysis, the expression data in the form of genes × arrays space is linearly transformed to “eigengenes” × “eigenarrays” space, where the eigengenes or eigenarrays are unique orthonormal superpositions of the genes or arrays. Normalizing the data by filtering out the eigengenes (and the corresponding eigenarrays) that are inferred to represent noise or experimental artifacts enables meaningful comparison of the expression of different genes across different arrays in different experiments; such normalization will improve any further analysis of the expression data.

Figure 4. Representative example of data clustering output. Dendograms indicate the degree of relatedness between adjacent genes and samples.
Limitations of DNA Microarrays

The variability of microarray results can be significant, especially for genes with low expression levels. Replication is recommended to establish a high degree of confidence, and also to reduce the number of potential false positive results. However, this may be difficult due to high cost or limitation of the sample amount. Confounding factors specific to microarray experiments that add to data variability include insufficient total RNA from samples, therefore requiring amplification steps that may introduce bias (see later); unequal efficiency of fluorescent dye labeling during reverse transcription; and reduced ability or failure of certain DNA elements on the array to detect the right transcripts as a result of cross-hybridization or adverse secondary structure. Alternative and more conventional techniques such as northern blotting, RNase protection, or PCR preceded by reverse transcription (RT-PCR), in particular quantitative RT-PCR, may be used to verify a subset of results, thereby helping to establish an estimate of the variability of a given experimental system.

DNA microarrays provide results on mRNA expression levels that do not necessarily correlate with protein expression levels or function, thus these results may provide only an incomplete view of the functional significance of differentially expressed genes in the experiments. Techniques for protein analysis such as western blotting, RNase protection, or PCR preceded by reverse transcription (RT-PCR), in particular quantitative RT-PCR, may be used to verify a subset of results, thereby helping to establish an estimate of the variability of a given experimental system.

DNA microarrays provide results on mRNA expression levels that do not necessarily correlate with protein expression levels or function, thus these results may provide only an incomplete view of the functional significance of differentially expressed genes in the experiments. Techniques for protein analysis such as western blotting, two-dimensional polyacrylamide gel electrophoresis, radioligand receptor binding, chromatographic separation and detection, and mass spectrometry remain indispensable for elucidating protein levels or function.

With the rapid advance of various technologies, the development and use of protein microarrays to address these questions may soon be possible.

DNA Microarrays in Renal Transplantation

DNA microarray is an extremely versatile and powerful technique that can be and has been applied to many areas of science and medicine. However, its potential in the field of transplantation, particularly in renal transplantation, has yet to be fully exploited. A recent study of murine cardiac allografts using DNA microarray technology identified the up-regulation of interferon-γ gene as an important factor in the late phase of acute rejection in vivo.22 Similarly, we expect there to be changes in the expression of multiple immune response genes in acute renal allograft rejection; indeed, we have preliminary results to support this hypothesis (Sarwal et al., manuscript in preparation). We are currently investigating 3 broad areas pertinent to renal transplantation by using cDNA microarrays to generate gene expression profiles from clinical samples: acute allograft rejection; chronic allograft nephropathy; and differential immune response in patients on a steroid-free versus steroid-based immunosuppression protocol. As we are primarily interested in genes involved in the immune system, our cDNA microarrays ("lymphochips" of about 35,000 DNA clones representing about 12,000 genes) are custom designed to contain an enriched population of immune response genes. Our research is greatly supported by the Stanford Microarray Database.23

Gene Expression Profiling in Acute Rejection

With the help of gene expression profiling of acute rejection allografts from patients, we are able to identify biomarkers that can more accurately diagnose acute rejection episodes and distinguish them from other causes of graft dysfunction such as acute tubular necrosis, chronic allograft nephropathy (CAN), and drug nephrotoxicity. Indeed, we have observed up-regulation of genes previously known to be important in acute allografts, such as T-cell receptor alpha (13.9-fold), granulysin (13.4-fold), MHC class II DP beta (5.3-fold), CD8 alpha (3.4-fold), and IL-2R (3.0-fold); additionally, we observed novel gene expression patterns and up-regulation (above 2-fold) of previously unidentified genes.24 Altogether, SAM analysis (using a delta value of zero, which gives us no false positives) predicted more than 200 genes that are significantly over- or underexpressed in acute rejection samples compared to normal kidneys (Sarwal et al., manuscript in preparation). We have confirmed with further biochemical analysis that granulysin is a useful predictor of acute rejection and steroid resistance in renal transplantation25; this illustrates that information extracted from DNA microarrays can be accurate and insightful.

Identifying Noninvasive Markers of Allograft Dysfunction

Predicting allograft rejection has traditionally been a difficult task despite advances in immuno-
suppressant therapy. Noninvasive markers of acute rejection are currently being studied in peripheral blood and urine by RT-PCR analysis, with the aim of using these genes as rapid and reliable diagnostic tools for acute rejection in the clinic.\(^{25,26,27}\) We have confirmed the expression of these and many other novel genes during acute rejection, with distinct molecular heterogeneity among acute rejection episodes. Additionally, we are studying the gene expression profiles of peripheral blood lymphocytes isolated from patients with acute rejection, and correlation studies with results from tissue biopsy samples may aid in the identification of noninvasive, peripheral markers that are diagnostic of rejection, therefore allowing us to obviate the need for tissue biopsy in future diagnosis. Correlation between these markers and the clinical responses of rejection episodes is currently under way (Sarwal et al., manuscript in preparation). Taken together, these results may allow us to better understand the molecular basis and pathogenesis of acute rejection, thus leading to improvements in therapeutic intervention and long-term graft survival, and the possible identification of novel targets for immunomodulation.

Gene Expression Profiling in Chronic Allograft Nephropathy

The pathogenesis and molecular basis of chronic renal transplant rejection or CAN is poorly understood; improved diagnosis and therapy are urgently sought. The etiology of CAN may either be immune or nonimmune; clarification of the cause is an important step to its prevention and/or treatment. We believe data from DNA microarray experiments can help us with more accurate classification, thereby allowing us to design more effective treatment strategies. We are using a baboon chronic vascular injury model (in collaboration with Pekka Hayry, University of Finland, Helsinki) to help us gain insight into the molecular basis of nonimmune chronic vascular injury.\(^{28}\) The choice of the baboon as a model species for this study arises from several considerations: human samples are difficult to obtain and sample volume is often inadequate for study; the baboon carotid (unlike the rodent vessel) has a defined intima area similar to that in humans, and the baboon is genetically homologous to humans and thus cross-hybridization of baboon samples to human arrays would be likely. Total RNA was extracted from carotid artery samples obtained after balloon denudation injury to the left carotid artery, amplified one round, fluorescently labeled, and successfully hybridized to human cDNA microarrays. So far, we are able to identify multiple factors involved in the early, intermediate, and late phases of chronic vascular injury, which are being correlated with human chronic vascular injury in CAN biopsy samples. We hope to eventually extend this study to more completely understand the molecular basis of chronic rejection in transplant patients by generating an immune model of primate vascular injury.

Gene Expression Profiling of Immunosuppressive Therapy

The use of corticosteroids has been a mainstay in the immunosuppressive therapy for renal transplantation; however, their use is limited by multiple undesirable side effects including Cushingoid habitus, impeded growth, infection, hypertension, hyperlipidemia, glucose intolerance, diabetes mellitus, bone loss, cataracts, acne, and changes in mood and behavior. Some of these side effects are especially dilapidating for adolescents, and avoidance of steroid-based therapy would be greatly beneficial. However, it has proven to be difficult to withdraw steroids in patients without increasing the risks of acute and chronic rejection. In a pilot study, we successfully implemented a steroid-free immunosuppressive regime involving daclizumab (a humanized anti-IL-2 receptor antibody), mycophenolate mofetil (an inhibitor of T- and B-lymphocyte proliferation in response to allospecific stimulation), and tacrolimus (an inhibitor of calcineurin and therefore cytokine production).\(^{29}\) Patients on these alternative immunosuppressive drugs have improved renal function and decreased incidence of clinical and subclinical rejection. Specific genes show differential regulation in response to various agents used in the immunosuppressive regime of transplant patients and are being studied at the time of graft dysfunction; a comparison of the gene expression profiles of patients on a steroid-free versus a steroid-based protocol will help to delineate the molecular and cellular responses involved in the
steroid response, thus providing new insights into transplant immunopharmacology.

**Study Limitations and Modifications**

In the above studies, only a limited amount of total RNA (well below the 50 µg required for a single microarray experiment) can be extracted from a typical biopsy sample. We have overcome this limitation by optimizing a modified RNA amplification protocol. A small amount of total RNA (as low as 10 ng, with a mean extraction of 3.4 µg of total RNA) is first reverse-transcribed into cDNA, which can then be double amplified to produce up to 100-fold of the estimated amount of original starting mRNA; the resulting amplified antisense RNA (aRNA) is then used for microarray analysis (with the optimal amount for hybridization being 5 µg of aRNA). The systematic bias that may be introduced by RNA amplification has been assessed by comparing the expression profiles generated by aRNA and those generated by total RNA, and a very strong correlation between the 2 was obtained. We have also proven the robustness of this amplification strategy ($R^2 = 0.87$ between 1st- versus 2nd-round amplification aRNA; Sarwal et al., unpublished data) and have been routinely using double-amplified aRNA, for both biopsy samples and the common reference, in our microarray analyses.

Core needle biopsies may generate samples from either the cortex or the medulla; we have controlled the variation of gene expression profiles due to tissue sampling by obtaining the gene expression signatures of cortex versus medulla and comparing them across all study samples (manuscript in preparation).

Renal biopsy samples used for analysis contain a mixture of different cell types; thus, with the exception of cell-specific genes (e.g., E-selectin), the source of mRNA is unknown and limits our ability to interpret the cellular signatures relating to the gene expression patterns of our data. There are various options to address this issue: laser-capture microdissection of cellular subtypes of interest and microarray analysis after message amplification have been attempted. Alternatively, the gene expression profiles from specific cell types can be compared with those of the whole tissue; in our study, we have generated data from resting and activated T and B cells, the major group of cells infiltrating the graft during the alloimmune response. We are also characterizing cellular signatures by analyzing allograft biopsy tissues using immunohistochemistry (with antibodies generated to specific genes of interest identified by cDNA microarray analysis); this will further allow us to study cellular subtype localization in the tissue and confirm gene expression at the protein level.

Taken together, transplantation studies using cDNA microarrays will provide valuable insights into the molecular mechanisms of allograft dysfunction, chronic and acute rejection, and graft acceptance, thereby helping to direct future studies in identifying better surrogate markers of graft dysfunction (more specific for the etiology of graft dysfunction than the serum creatinine level) and helping to design better preventive and therapeutic strategies in the field of transplantation.

**Concluding Remarks**

We are beginning to generate encouraging and highly informative data from our patient samples; however, we are aware that microarray technology is only a tool to help us identify novel genes and interesting patterns of gene expression. Subsequently, we will need further molecular and biochemical assays before we can fully decipher the biological and clinical relevance of our data. In this postgenomic era, increased cooperation and sharing of information among investigators is vital; to this end, public gene expression databases have been made available, for example, Stanford Microarray Database (SMD), ExpressDB, The Gene Expression Database (GXD), and Gene Expression Omnibus (GEO).

In conclusion, DNA microarrays can be used in an analogous manner to combinatorial chemistry and high-throughput screening assays to highlight potential candidates for further studies; sieving through the huge volume of data to obtain meaningful results may be daunting, but it is often rewarding. The successful development of protein microarrays in the future will be the next important step to help scientists unravel the functions of many genes, thus giving us a more complete understanding of the molecular basis of many dis-
cases. It is probable that exciting new therapeutic targets will be discovered through the application of these technologies.

Acknowledgments

This project is done in collaboration with Dr. Patrick Brown at the Howard Hughes Medical Institute, Stanford University. We gratefully acknowledge the assistance of the Brown Laboratory in the data processing and bioinformatics analysis, especially Xin Chen, Ash Alizadeh, Jennifer Boldrick, and Orly Alter. This project is funded by grants from the NIH (NIH5P3-05 and NIH3P3-0551), Clinical Center for Immunological Studies at Stanford University (CCIS), Packard Foundation, and Roche Pharmaceuticals.

REFERENCES