



Immunol Allergy Clin N Am
22 (2002) 891–910

IMMUNOLOGY
AND ALLERGY
CLINICS OF
NORTH AMERICA

Functional genomics of immune responses

Kenneth Christopher, MD, Charlotte M. McKee, MD,
Thomas F. Mueller, MD, PhD, David L. Perkins, MD, PhD*

*Harvard Medical School, Brigham and Women's Hospital, 75 Francis Street, Boston,
MA 02115, USA*

Compilation of a draft of the sequence of the human genome [1,2] and concurrent technologic advances have made the simultaneous analysis of thousands of genes experimentally feasible. The study of the expression and function of many genes is called functional genomics. Technologies such as DNA microarrays, which quantitate the expression of all mRNAs (called the transcriptome), are becoming readily available. The expression of the complete transcriptome in the yeast *Saccharomyces cerevisiae* during biologic responses has been analyzed [3]. Although these experimental approaches are being applied to many different biologic responses and diseases, this article focuses on the analysis of the complex immune networks following transplantation. Similar approaches have been initiated to investigate the numerous genes that are presumed to comprise the alloimmune response following transplantation. The development of DNA microarray technology enables investigation of the complete transcriptome of expressed genes during allograft rejection. The goal of the global analysis of gene expression is to determine the molecular mechanisms of immunity and to discover new biologic targets for pharmaceutical agents and diagnostics. Paradigms of transplantation in lung and other solid organs indicate that the rejection response involves hundreds of genes that are controlled by complex regulatory networks. Models of transplantation in gene-deficient (knockout) mice support this concept. More than 100 strains of gene-deficient mice have been identified in which allograft survival is prolonged. With few exceptions, these strains eventually reject allografts, with delays observed in kinetics. These results indicate that the deleted genes are important, but not necessary, for allograft rejection. These observations indicate that the functions of many of these genes can be compensated for by alternative mechanisms that are, at least in part, functionally redundant. It is evident that allograft rejection in animal models involves the functions of numerous genes in concert. Studies of

* Corresponding author.

E-mail address: dperkins@rics.bwh.harvard.edu (D.L. Perkins).

graft survival and rejection in human studies typically have focused on a small number of genes or even a single gene. Given the animal data, however, it is likely that many genes, including some genes that have not been analyzed, function in complex networks in the human response to transplanted organs. Functional genomics provides a powerful tool for the study of transplantation rejection in the lung and other solid organs.

Immunonology of lung transplantation

Microarray analysis is beginning to be applied to allograft rejection in animal models and in clinical studies of solid organ transplants [4,5]. The authors' laboratory has initiated microarray analyses of rejection in clinical lung transplantation and in other animal models. Studies of clinical lung transplantation present several unique advantages compared with studies of other solid organs. The high incidence and rapid onset of rejection decrease the reliance on surrogate markers for rejection and reduces the size of the study population necessary to generate statistical power. There is a relatively low-risk access to intragraft immune cells with bronchoalveolar lavage (BAL). Studies that are limited to patients who receive transplants for cigarette smoking-related emphysema minimize the confounding factor of abnormal pulmonary immune responses in diseases with overt inflammatory components, such as cystic fibrosis, while taking advantage of the fact that this population makes up the single largest disease category in lung transplantation. Lung transplantation is an established therapy for patients with end-stage pulmonary disease caused by emphysema, cystic fibrosis, primary pulmonary hypertension, idiopathic pulmonary fibrosis, and a variety of other conditions. Almost 1000 lung transplants were performed in the United States in 2000 [6]; however, lung transplantation is complicated by a high incidence of acute and chronic rejection. To prevent rejection, patients with lung transplants are treated with high doses of immunosuppressive drugs, increasing the risk for infections, malignant tumors, other drug toxicities, and side effects. Because of these complications, the 3-year survival rate of patients with lung transplants is approximately 50% [6], a rate that is worse than that of patients with any other solid-organ transplant, including kidney, liver, and heart (Fig. 1). The poorer prognosis for recipients of lung transplants compared with that for recipients of other solid-organ transplants may be caused by the unique characteristics of immunity in the lung. The normal functions and immune responses of the lung ironically may make an alloimmune response (and rejection) more likely to occur in lung allografts than in other solid-organ allografts. As is true in all solid-organ transplants, allograft rejection in lung transplants is predicated on the recognition of foreign (transplantation) antigens of the allograft by the recipient's immune system. This response classically has been considered to be initiated by host CD4⁺ T cells, although CD8⁺ T cells also may be important in donor antigen recognition. Donor antigen may be presented to these cells in a variety of ways, including indirectly (as processed antigen presented on host

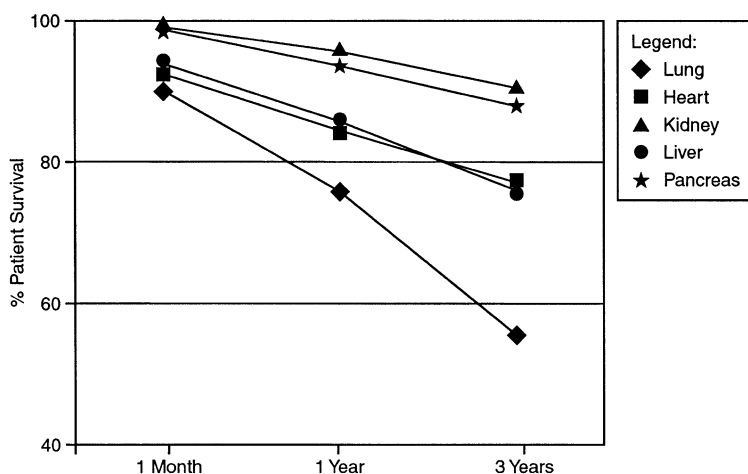


Fig. 1. Patient survival after lung transplantation is worse than survival after other solid-organ transplants. Survival rates for adult patients who receive lung, heart, kidney, liver, or pancreas transplants at 1 month, 1 year and 3 years. (Data from United Network for Organ Sharing, 2001 Annual report of the US Organ Procurement and Transplantation Network and the Scientific Registry for Transplant Recipients: transplant data 1991–2000. Richmond (VA): United Network for Organ Sharing; 2001. <http://www.unos.org>. Accessed December 2001.

antigen-presenting cells [APCs]) and directly (on donor APCs); nonclassical APCs, such as donor endothelial cells, also may present antigen and initiate a host immune response [7]. Recognition of donor alloantigen triggers a cascade (or, more accurately, an explosion) of cytokines and other inflammatory mediators, including the T helper cell 1 (T_H1) cytokines interleukin 2 (IL-2), interferon γ (IFN- γ), and tumor necrosis factor α (TNF- α). These mediators recruit and activate cytotoxic lymphocytes, macrophages, and other effector cells that cause allograft rejection. In addition to these mechanisms, interactions between normal pulmonary immune mechanisms and mechanisms of the alloimmune response to transplanted antigens may augment rejection in pulmonary transplants, because allograft rejection seems to occur by way of many (if not all) of the same immune mechanisms that are used by the host to fight infections. The lung interfaces directly with the outside environment and is a major portal for environmental agents, much like the gut [8]. Both organs are important sites of antigen clearance, antigen sensitization, and infection. Transplanted lungs contain large numbers of lymphocytes and APCs. Under normal circumstances, these cells provide constant antigen surveillance and maintain responses to aerosolized environmental antigens in the alveolae and at the epithelial surface. Dendritic cells are believed to be the most important APCs within the lung to perform this surveillance role [9]. Pulmonary dendritic cells and other leukocytes are transplanted with the lung allograft and efficiently may transport and present donor antigen to the host immune system in the setting of transplantation [10]. The normal function of pulmonary epithelium, as a physical and immunologic barrier to infection, may

become a liability in pulmonary transplantation. The normal activation and inflammation response of epithelial cells and their associated intraepithelial lymphocytes to invading infections may up-regulate donor major histocompatibility complex (MHC) antigens and other antigens on cell surfaces and can induce the same explosion of cytokines, chemokines, adhesion molecules, and other mediators that fosters allograft rejection.

The poor survival rate for lung transplantation largely is attributed to the vast majority of patients who develop obliterative bronchiolitis (OB). OB is believed to be a form of chronic allograft rejection that is characterized by the obliteration of small (< 1mm) airways with fibroblasts and collagen. The sequence of events that precede total obliteration include infiltration of T cells and macrophages, fibroblast proliferation with subepithelial collagen deposition, loss of epithelium and basement membrane followed by fibroblast, and collagen encroachment within the airway. Pathologic tissue demonstrates a full range of airway abnormalities, suggesting that this process is regulated locally with varying spatial and temporal progression. In OB, the alveolar interstitial compartment is spared from fibrosis. Although interstitial pulmonary fibrosis and OB might have similar features (eg, pathogenic mechanisms), there is strict spatial regulation of these processes.

By using microarray technology to obtain a better understanding of the immunology of rejection, more effective and less toxic therapies can be developed. It is hoped that this development will translate directly into improved survival and a better quality of life for patients undergoing lung transplantation. Similarly, microarray analysis of BAL (and other immune cells and tissues) may help unravel the immunologic responses in a wide variety of other pulmonary conditions.

Microarray methods

Multiple technical approaches for analyzing gene expression have been developed. In addition to microarrays, common approaches include filter arrays, serial analysis of gene expression (SAGE), and polymerase chain reaction (PCR), which have been reviewed previously [11,12]. This article focuses on the DNA microarray, an ordered layout of known DNA material that is bound to a glass or nylon surface available in commercial and custom formats. In most protocols, RNA is harvested from the experimental sample, converted to cDNA by reverse transcription, and transcribed to cRNA linked to a fluorochrome. The array surface is spotted with nucleic acid probes, which can consist of cDNA, PCR products, or oligonucleotides. Oligonucleotide arrays are produced commercially by photolithography techniques to synthesize oligomers directly on the glass slide. The fluorescently labeled target sample is hybridized with the array probes. After washing, fluorescence is detected by robotic confocal microscopy and quantitated by pixel-based algorithms. The results commonly are normalized to a reference standard that is linked to a second fluorochrome, or they may be normalized to a standard value, such as a mean of zero. With advances in automation

and analysis, the hospital diagnostic laboratory likely will include microarray technology in the future.

Data analysis

Bioinformatics

Studies involving microarrays easily can generate databases containing millions of data points. The challenge in microarray studies is to select data-mining tools that effectively normalize, filter, analyze, and present the results using appropriate algorithms [13]. Most commonly used data analyses demonstrate associations between a small subset of genes with known functions and expression profiles that are consistent with previous knowledge. Genes of unknown function may show expression profiles that are similar to profiles of known genes, enabling insight into possible gene function. Genes related by physical connections in a network have expression profiles that robustly correlate [14]. Mechanisms of unknown genes can be postulated by association to known genes and tested by hypothesis-driven studies. Alternatively, data can be analyzed in global terms, which has the advantage of retaining all of the available data, to generate complex models of biologic responses. This global approach is analogous to an ecologic study that investigates an entire forest, whereas the selection of a few genes is analogous to studying a single species. Microarray technology enables the practical study of complex biologic systems at a global level.

Differentiation of signal from noise

To generate quantitative data from microarray assays, it is crucial to produce replicates within acceptable limits of variability. Excess noise among the thousands of data points produces numerous false-positive differences of expression that prevent reliable analysis. Deviation among replicates can be caused by technical or biologic variation. Technical variation can be generated at any step of the microarray procedure. During the production of cDNA from sample mRNA, reverse-transcription bias occurs because of the differing efficiency of the reverse transcription of some mRNAs. This bias can affect the relative abundance of cDNA measured by an expression-array assay compared with alternative methods such as Northern blots or PCR. The reverse-transcription bias prevents quantitative comparison between different mRNAs on an array [15]. Preferential transcription of the 3' end of an mRNA, depending on the probe, can alter hybridization. Similarly, differential mRNA splicing can have marked effects on hybridization, particularly on short oligonucleotide probes.

Other common sources of variability include quality-control issues such as probe selection and preparation, spotting, and fluorescence detection and quantitation. In a study designed to limit variability, data from microarrays that were custom made by the researchers were noted to have many misclassifications and

numerous false-positive results, and repetition produced different lists of expressed genes [16]. This variability likely is caused by background noise that is inherent in scanning, measuring, and comparing the fluorescence of the hybridized array. Repetition of custom-made microarray experiments was shown to decrease false-positive and false-negative results and the misclassification rate [16]. Whether this experiment using custom-made microarrays can be generalized to all custom-made arrays or to commercial microarrays is not clear.

The commercial GeneChip array (Affymetrix, Santa Clara, CA) includes 16 to 20 tethered oligonucleotide probes per target gene. To control for variability of hybridization to small oligonucleotides, paired oligomers that contain a single mismatched nucleotide are spotted on the array to serve as internal controls. Expression levels are quantitated by comparison of the fluorescence levels of matched and mismatched probes. In the authors' laboratory, hybridization of a single sample to two GeneChips resulted in a correlation coefficient of 0.98 between the expression data of the two arrays. Chip-to-chip variation is common, however. Differences may exist among the tethered probes used to measure the same gene [17]. A statistical model has been developed to account for individual probe-specific effects and automated detection of image artifacts and outliers [17]. To control for technical variation among microarrays, the values of each array often are normalized to a standard. The results are normalized to a mean of zero and a standard deviation of one by using a function known as the *z* value. Alternatively, in dual fluorescent approaches, experimental and reference samples are linked to different fluorescent probes, and the results are reported as a ratio.

Biologic variation is dependent on the type of samples, complexity of the biologic process, sample population, and environment. In animal models, environmental variation can include ambient temperature, time of day, and external stressors before and during the experiment. It is important that investigators determine the technical and biologic variation for each experimental protocol.

Filtering and normalization of data

Various strategies to differentiate signal from noise have been applied to array data. A common approach is to establish an arbitrary threshold value of absolute expression or change in expression to classify genes as expressed or absent. Classification is performed with algorithms that are generated by several laboratories and commercial sources [18]. Affymetrix Microarray Suite 4.0 classifies genes as present, absent, or indeterminate. After filtering out genes that are classified as absent, the positively expressed subset can be analyzed by statistical algorithms. A major advantage of this approach, in addition to its simplicity, is that much of the experimental noise is eliminated; however, a concomitant disadvantage is its difficulty in differentiating values of low abundance mRNAs from noise, causing many mRNAs that are expressed at low copy number to be eliminated from the dataset. Important and meaningful data can be lost with this approach. After filtering a dataset with exclusion criteria, expression values of the selected genes often are normalized. The set of values for each gene can be normalized to a

common central tendency (eg, to a mean of zero). During subsequent analyses, this approach emphasizes the comparison of relative, rather than absolute, changes in expression. A disadvantage of this approach is that some genes with large absolute changes combined with high constitutive expression in the control samples show small relative changes. The selection of a normalization protocol should be guided by the underlying biologic question.

Graphical representation of data with pseudocolored images

A simple, but often effective, approach to data analysis is the graphical representation of gene expression values by pseudocolor, with up- and down-regulated genes indicated with graduated intensities of opposing colors [19]. The Cluster program, after grouping expression profiles by a similarity metric, creates a graphical color-coded pattern of expression levels. Patterns of differential gene expression have been used to discriminate breast cancers with *BRCA1* mutations from those with *BRCA2* mutations [20], distinguish between acute myeloid leukemia and acute lymphoblastic leukemia [21], and provide a prognostic classification in clear-cell renal carcinoma [22]. When analyzed in this fashion, the expression array can be likened to the painting technique of the French neo-impressionist George Seurat. Seurat's technique of pointillism used small dots of contrasting color to build large complicated compositions. As with an expression array, Seurat's single dots are often too small to provide meaning by themselves. Each dot in the painting has a purpose in the building of the scene. In a model of transplantation analyzing the effects of CD40 ligand blockade, subsets of genes up-regulated at the time of rejection in the treated group compared with the untreated control group could be differentiated using color-coded images [23]. This study defined a distinct subset of genes expressed during acute rejection that differed from another subset that was up-regulated during delayed rejection after treatment.

Selection of genes based on expression ratios

The identification of genes that are up- or down-regulated significantly in experimental and reference samples is an important step in the analysis of expression array data. A common approach used in numerous reports is the determination of fold induction by calculating the ratio of expression in a experimental sample compared with expression in a control sample. This approach is straightforward and has been used in numerous published reports; however, it has several major limitations. Genes with low levels of absolute expression have more variation than do genes with high levels of expression; the ratios of low expressers are less accurate. If low expressers are excluded, large amounts of important data can be lost. Correlations between biologic significance and high ratios of change in gene expression have not been established for most genes. Unresolved questions exist regarding the biologic importance of a large relative change versus a large absolute change and whether a large relative change at the peak of a biologic process is more

important than a small change at the initiation. Expression ratios commonly are used to select a few genes with the largest relative changes in expression. These ratios reduce the power of the experimental design by concentrating on a few genes. An alternative to analysis that is based on simple ratios is the incorporation of maximum likelihood estimates [24]. The maximum likelihood estimate is a complicated statistical method that “estimates the parameters to make the probability of the observed data as large as possible” [25]. The major advantage of this approach is that genes with low absolute levels of expression can be shown to have regulated significantly. This approach was used to identify differentially expressed yeast genes by direct comparison of a series of repeated measurements of the two fluorescent dye intensities for each gene rather than relying on expression ratios [24].

Global profiles of biologic processes

A unique advantage of microarray experiments is the potential to analyze biologic responses at the level of the entire transcriptome. Although the identification of a few highly regulated genes may provide clues to a biologic response, the inclusion of all genes in an analysis may be crucial to a complete understanding. This theory has been demonstrated in several studies of *S. cerevisiae* in which all genes were identified and can be spotted on a single chip. In studies that examined nearly all of the yeast genome, 800 genes regulated by the yeast cell cycle were identified in cell cultures that were synchronized by three independent methods [26]. The rationale for global analysis is that this approach can produce important serendipitous observations. Previous studies of growth-arrested mammalian fibroblasts that were stimulated with serum have focused on cell-cycle progression. A microarray experiment of the fibroblast model confirmed many previous observations involving cell-cycle progression and detected the induction of many stress genes [27]. These results suggest that serum, which is produced by the clotting process, induces programmed biologic processes in response to the perceived stress. Ultimately, the scientific goal for each experimental condition is to understand the complete biologic response rather than just identify a few highly induced genes. A global perspective stresses global pattern recognition over single-gene recognition and function. Biologic research by way of hypothesis-driven research investigates the gene-protein-function model a few genes at a time. In addition to hypothesis-driven research, the global view can incorporate the information of the complex networks of a biologic response. Observing the shifts in patterns of gene expression between two experimental conditions facilitates a broad understanding of the biology and contributes to the appreciation of the complexity of a system.

Cluster algorithms

To perform global analyses, software has been developed that compares large datasets using similarity metrics and pattern recognition algorithms [13,28]. Two

common types of algorithms are discussed: (1) dendrograms that are generated by clustering algorithms and (2) self-organizing maps (SOMs), which are a form of neural network.

Dendrograms

Dendrograms that are generated by agglomerative clustering algorithms are analogous to evolutionary trees. In this approach, all experimental samples and genes are compared with each other using a similarity metric, such as the Pearson correlation coefficient. The two samples with the greatest similarity are linked, and the remaining sample with the greatest similarity is attached to the previously linked samples. The resulting dendrogram provides a graphical overview of the similarity among the samples (Fig. 2). An analysis of diffuse, large B-cell lymphoma tissue identified two distinct gene expression patterns associated with different survival patterns [29]. The major advantage of this approach is the generation of a graphical global overview of multiple experimental samples or genes.

There are several important limitations of the dendrogram approach. At each step, the two samples with the least variation are linked; however, the algorithm does not determine whether the variation is significant. The variation between groups is calculated as the sum of the differences between each component (ie, gene) in each sample. The variation is a single value that can produce misleading comparisons if multiple diverse biologic processes are occurring simultaneously. Dendrograms are generated by an agglomerative process and are susceptible to initial conditions, with repeat analyses producing different patterns of linkage. A variation of the clustering algorithm is the k-means clustering. In this algorithm,

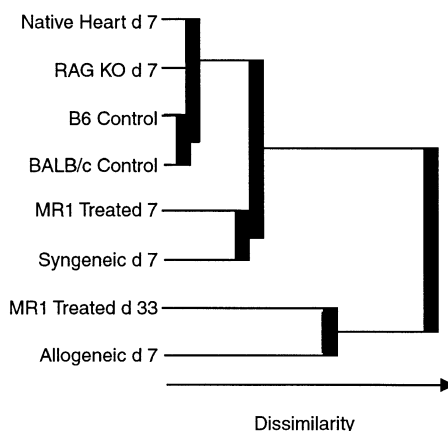


Fig. 2. Expression levels of candidate genes. Following murine cardiac transplantation, the level of expression of a panel of 60 candidate genes, including inflammatory molecules, cytokines, chemokines, chemokine receptors, and cell surface markers, was analyzed in graft hearts by RNase protection assay. Experimental models included allogeneic, syngeneic, alymphoid, and anti-CD40 ligand murine antibody (MR1) treated. A dendrogram was generated by Cluster and TreeView software [19]. Controls are untransplanted BALB/c and B6 hearts. The degree of dissimilarity is proportional to the total length of the horizontal axis between groups.

the data is divided into an arbitrary number (k) of clusters that is defined by the operator. If k is selected appropriately, the clusters can correlate with important biologic distinctions; however, because k must be selected empirically, correlations may be difficult to establish or erroneous.

SOMs

Another approach, which also associates a variable number of samples (or genes) in each cluster, is to use SOMs [30]. SOMs are artificial neural-network algorithms that have been applied successfully to high-dimensional and nonlinear problems in many areas of science, including microarray data [28,31]. To generate maps, the user selects an arbitrary number of nodes, and genes are assigned randomly to those nodes. In an iterative learning process, the genes are rearranged by the algorithm based on a similarity metric, such as geometric distance, and are reassigned to nodes with increasing similarity (Fig. 3). The number of genes assigned to each node is not predetermined but is variable based on an optimal arrangement. The number of iterations is defined by the user and should be increased until the gene assignments to nodes remain constant. The number of nodes and the geometry also is defined by the user and should be adjusted empirically by minimizing the standard deviations of the genes assigned to each node and eliminating nodes with few or no genes. The major advantage of SOMs compared with dendrograms is that the number of genes per node is

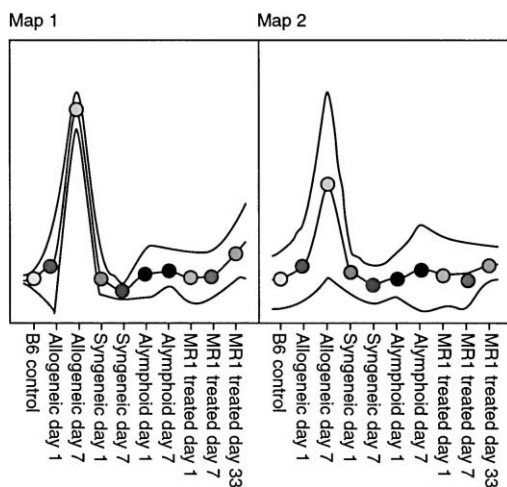


Fig. 3. SOMs of expression levels of candidate genes. The RNase protection assay data, which were generated in 60 graft genes after murine cardiac transplantation, were analyzed with GeneCluster [31] using SOMs and were separated into nine distinct maps (two of which are shown). The number of maps was determined empirically to minimize the standard deviations and maximize the number of genes per map. The open circles show the mean level of gene expression, and the lines indicate the standard deviations of the expression for each map. The y-axis represents relative expression calculated for each SOM. Genes of map 1 (up-regulated in allogeneic day 7 and MR1-treated mice after day 32) differed from the genes of map 2 (moderately up-regulated only at allogeneic day 7).

determined by the similarity of the cluster rather than by a predetermined hierarchical structure. If sufficient iterations are used, the algorithm is not prejudiced by initial conditions.

Correlation of clusters with biologic processes

Dendrograms and SOMs produce clusters of genes that are associated by a similarity metric; however, they do not define a biologic function. Inspection of clustered genes may indicate functions based on previous reports. Based on similar patterns of expression, genes of unknown function and expressed sequence tags (ESTs) can be assigned hypothetical functions that can be tested. In a study of the response of *S. cerevisiae* to the change of a single nutrient, microarray analysis detected the genes that previously were identified by biochemical approaches [14]. The microarray results discovered several genes that previously were not associated with the metabolic pathway using previous experimental approaches. This study demonstrates the power of microarray analysis to extend the understanding of function to a more global view of a biologic response. Similar types of correlations have been detected in studies of sporulation and proliferation [32,33].

Although this approach has been successful in analyzing experimental designs focused on a single isolated perturbation, it becomes more difficult in analyzing complex processes. The authors' laboratory has analyzed the in vivo response to allograft transplantation in a murine model of heart transplantation. In this model, multiple perturbations, including stress, ischemia, innate immune stimuli, and alloantigen, trigger biologic responses. To produce a global view of the major biologic responses following transplantation, genes with significant changes in expression were assigned a biologic process based on the Gene Ontology database. A major advantage of this approach is that the interpretation is not biased by the user but is based on objective criteria. The Gene Ontology annotation system uses a structured language for specific biologic processes that provides a consistent classification of gene function in many organisms, including *Caenorhabditis elegans*, *S. cerevisiae*, mice, and humans [34]. By assigning unbiased biologic processes to individual genes, Gene Ontology provides a fast, updated database to quickly define the functions of genes of interest. Application of the Gene Ontology system to microarray data allows for the patterns of change among groups of genes to be given biologic relevance. The application of Gene Ontology classification to all of the data focuses the interpretation on a global or functional genomic view. This global view is analogous to an ecologic analysis of an ecosystem. In murine cardiac transplantation, genes that constitutively were up-regulated in the heart were determined using microarrays. After transplantation, these constitutively expressed genes were evaluated with microarrays for their up-regulation pattern. Assignment of biologic processes by way of Gene Ontology to the sets of genes before and after transplantation suggests an increase in the percentage of genes dedicated to cell communication and defense and a decrease in the percentage of genes involved in cell maintenance. This analysis underscores the complexity of the transcription response to transplantation.

Confirmation of microarray data

An optimum method to validate the results obtained by microarray data has not been found. As discussed earlier, oligo arrays and cDNA arrays measure fluorescence intensity levels rather than mRNA or transcripts of an individual gene. These intensity levels depend on several factors. Oligo arrays may produce hybridization of different genes to the same spot (cross-hybridization) or of one gene to different spots. This method may lead to a rather low specificity for a particular gene. Sensitivity of the assay is influenced by several factors that enter the process from the actual expression levels to the observed intensity level of phosphorescent dye. Differences in mRNA concentrations, exposure time, camera sensitivity, background, and saturation have a direct impact on the quantitation of actual expression levels [35,36]. These factors have to be taken into account when analyzing expression signatures of all genes and when quantitating the expression profile of a single candidate gene. It is mandatory to use well-designed experiments to reduce biologic variability and to test for reproducibility with replicates in the light of technical variability.

Alternative methods to confirm the expression profiles are needed. Although there is no gold-standard reference method available, different groups have used Northern blot analysis, quantitative competitive PCR, RNase protection assays, and in situ hybridization [5,36]. These methods permit a quantitation of gene expression; however, they are far from ideal, because they are rather time consuming, are technically demanding, and need a high amount of RNA.

Real-time reverse transcription PCR (RT-PCR) seems to be a promising method to validate expression profiles of individual genes [35,37]. As an amplification method, real-time RT-PCR needs only small amounts of RNA. Individually designed primers that are used with fluorogenic probes permit a high specificity for the particular gene. The kinetic analysis, using comparable threshold cycles for the different amplification plots, allows for relative quantitation of gene expression. Technically, this procedure is quick and robust; however, differences in the efficiency of the RT step and variability in the PCR have to be taken into account [11,38]. Real-time RT-PCR should not be seen as a standardized quantitative reference method but rather as a way to semiquantitatively validate expression profiles of single genes. A genome-wide analysis in yeast of the regulation of mRNA and protein expression in response to different nutrient sugars showed that approximately 95% of the induced genes were up-regulated at the level of mRNA and protein [14]. Although the correlation may be lower in mammals [2], these observations suggest that there is a strong correlation between increases in mRNA and increases in protein expression. Depending on the experimental question, however, it may be important to determine the correlation between mRNA and protein expression. Although proteomics is a rapidly developing field, the technology to screen all proteins, called the proteome, is technically more difficult than screening mRNA. For most laboratories, genome-wide analysis of protein expression is not feasible. For genome-wide protein study to become a reality, improvements are needed in the recombinant-protein-expression system, the

binding surface, and the detection of the binding of protein to the array. Progress in protein and antibody arrays has been applied to the monitoring of serum antibodies and to profiling protein expression levels [39]. Investigators have demonstrated the ability to screen for protein–protein interactions [40], identify the protein targets of small molecules [40], and isolate antibodies against impure proteins and complex antigens [41].

Transplantation

Microarray studies

The application of DNA microarrays to the analysis of transplantation is in its infancy. The technique will allow accurate stratification of patients according to their rejection stage and allow determination of whether graft dysfunction is caused by an immunologic mechanism. This section focuses on several studies that have investigated the alloimmune response by way of microarrays in human and animal models.

To enhance the understanding of the process of rejection in lung transplantation, differential gene expression is being studied in the small airways of patients undergoing a second lung transplantation for the indication of OB.(S.D. Shapiro, MD, personal communication, 2002). Portions of these patients' airways are used to stage OB and to isolate mRNA for microarray analysis. Preliminary results have demonstrated that within the airways of any individual patient, pathogenic changes can be found, including normal, inflamed airway, loss of basement membrane and epithelium, and airway obliteration with collagen and or fibroblasts. SOMs of gene profiles confirm the expected increase in collagen genes and decrease in epithelial genes as OB progresses. Specific to the stages of inflammation and basement-membrane destruction, there is up-regulation of several matrix metalloproteinases leading to the following hypothesis regarding the pathogenesis of OB. In a mismatched allograft, injury to epithelial cells initiates immune and inflammatory cell infiltration. Immune and inflammatory cells (but perhaps epithelial cells directly) cause fibroblast proliferation with subepithelial fibrosis. Epithelial cells and macrophages are activated to produce matrix metalloproteinases that ultimately degrade the basement membrane, leading to apoptosis and preventing repopulation with migrating epithelial cells. Once the barrier between the basement membrane and epithelial cells is lost, fibroblasts and collagen encroach the airway, resulting in total obliteration. This hypothesis is an example of a new focus of research resulting from a small but promising subset of genes identified from microarray studies.

A study of rejection in human recipients of kidney allografts analyzed gene expression in biopsies of renal allografts from patients on triple immunotherapy (cyclosporine, mycophenolate mofetil, prednisone) with a confirmed diagnosis of acute rejection [4]. Expression of gene transcripts from seven renal biopsies with histopathologic evidence of acute cellular rejection were compared with expres-

sion in tissue from three normal renal biopsies. Using a minimum threshold of four-fold induction, four genes were identified in each acute rejection sample: human monokine induced by IFN- γ (a CXC chemokine that is crucial for murine acute allograft rejection) [42], T cell receptor (TCR) active β -chain protein, IL-2–stimulated phosphoprotein (expressed predominantly in T cells), and Abcb2 (ATP-binding cassette, sub-family B, member 2; important in intracellular antigen processing) [43]. Other potentially important genes that were up-regulated in six of the seven rejection biopsies include interferon-stimulated growth factor-3, which is involved in signal transduction; complement factor 3; the CC chemokine macrophage inflammatory protein-3 β ; myeloid differentiation protein; and the adhesion molecule CD18. Kidney epidermal growth factor and kininogen were found to be down-regulated in five of the rejection biopsies.

The observation that multiple IFN- γ –responsive genes were highly up-regulated in this study illustrates the difficulty of interpreting microarray data by focusing on a small subset of genes. Although the sample size was small, this evaluation provides a window into the complex acute rejection process in the context of immunosuppression. In murine models of transplantation, IFN- γ –deficient mice reject allografts with kinetics similar to those in wild-type recipients; however, IFN- γ –deficient mice have resistance to tolerance induction [44]. Although IFN- γ may be highly induced, its major function may be to provide negative regulatory signals rather than promote rejection.

To understand the temporal pattern of differential gene expression in acute rejection without immunosuppression, murine cardiac allografts, isografts, and native hearts were studied comparatively with microarrays [5]. One day after transplantation, the most profoundly induced genes in cardiac allografts versus isografts included Ig light-chain-variable region, Ig-rearranged κ -chain, cathepsin L precursor, apolipoprotein E precursor (a modulator of immune response) [45], and complement factor 3. Five days after transplantation, the most profoundly induced genes in cardiac allografts versus isografts were monokine induced by IFN- γ , IFN- γ by induced GTPase, large multifunctional protease 2 (processor of MHC class I antigens) [46], RANTES (regulated on activation non–T-cell expressed and secreted) (a CC chemokine), IFN- γ , and MHC class II genes. Most of the 20 most highly up-regulated genes at day 5 were associated with IFN- γ induction. The studies of human renal and murine cardiac transplant rejection identified highly induced genes based on increased ratios of expression; however, the reports did not attempt a global view of the rejection process.

The authors' laboratory is studying the differential gene expression of rejection in allogeneic, syngeneic, and alymphoid (mice that are deficient in the recombinaase-activating gene and lack T and B lymphocytes) grafts in a murine model of cardiac transplantation. This study was designed to investigate antigen-independent and antigen-dependent responses. Using murine arrays (Affymetrix), the authors analyzed gene expression in response to ischemia, systemic stress, and innate and adaptive immune stimuli. Of the 6500 genes and ESTs that were studied following transplantation, the authors identified 181 genes with significantly modulated expression in at least one of the experimental groups

(T.F. Mueller, unpublished data 2001). To characterize the global response following transplantation in terms of biologic responses, the Gene Ontology database was used to categorize the biologic processes of all modulated genes (Fig. 4). The relative contribution of each biologic process changed in different experimental groups, altering the patterns of the percentages of genes involved in defense, stress, metabolism, and communication. Transplantation biologists traditionally have focused on the defense response, or more specifically, the

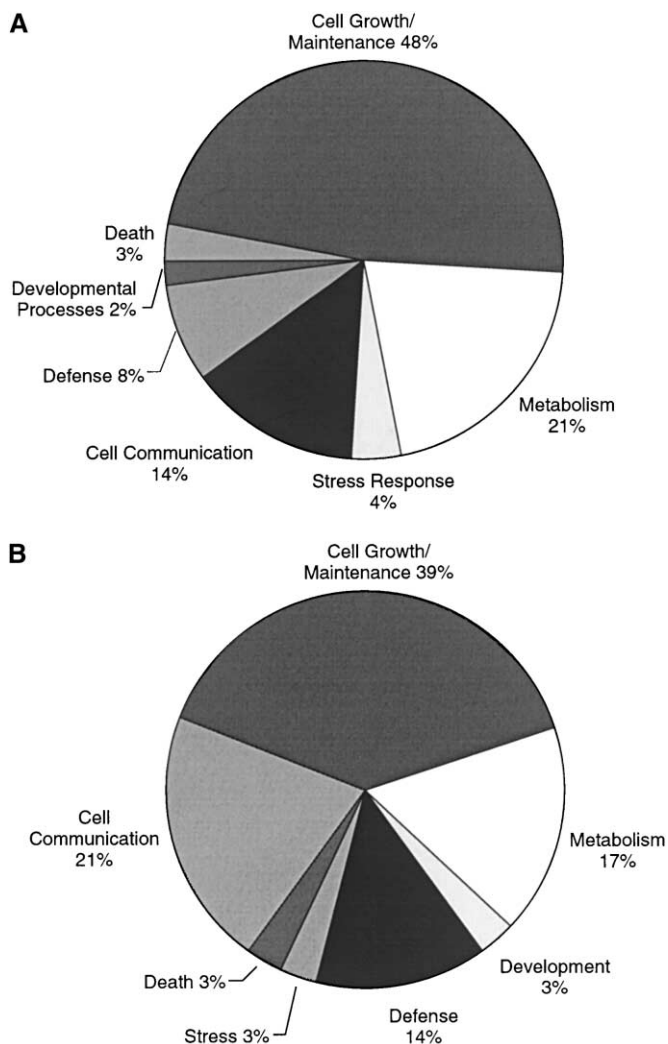


Fig. 4. Biologic processes of modulated genes after cardiac transplantation. (A) Gene Ontology-defined biologic processes constitutively expressed in control untransplanted heart as measured by DNA microarrays. (B) Gene Ontology-defined biologic processes up-regulated subsequent to murine cardiac transplantation as measured by DNA microarrays.

adaptive immune response; however, the authors' results show that additional biologic processes, including metabolism and communication, may be components of the concerted response during rejection. These biologic processes provide fertile areas for future research and eventual targets for possible novel therapeutic options.

Clinical implications

In addition to deciphering the complex biologic responses to transplantation, studies of gene expression that use microarrays and alternative technologies, such as real-time PCR, may provide accurate diagnostic and prognostic information for clinical management. In current clinical practice, acute rejection typically is diagnosed by histologic analysis of graft biopsies. The application of these new technologies could produce a molecular profile of each biopsy specimen, which could provide more precise diagnostic criteria. Based on the analysis of graft rejection in knockout mice, the authors believe that acute rejection includes a diverse spectrum of molecular profiles that reflect multiple biologic mechanisms of rejection. Using algorithms such as SOMs, the authors have identified overlapping and distinct components of these profiles. These observations have important clinical implications. Some profiles may be associated with smoldering rejection or progression to chronic rejection (a major problem in lung transplantation), and different profiles may suggest different therapies. Therapeutic decisions could be based on rational analysis of the molecular profiles of rejection.

These exciting possibilities suggest the practical questions of how many genes must be analyzed, whether analyzation of the complete transcriptome is necessary, or whether a selected subset is sufficient. In a murine cardiac transplant model, the authors analyzed the expression of 60 genes with known immune functions at the time of acute rejection in a group treated with CD40 ligand blockade [23], a treatment shown to prolong graft survival in murine and primate models [47]. Dendrograms clearly differentiated the CD40 ligand blockade (MR1) group from the untreated control group (Fig. 2). SOMs identified nine clusters of genes with distinct patterns of expression in the treated and control groups (see Fig. 3). A total of nine genes, one selected from each cluster, were sufficient to differentiate the treated and control groups in repeat dendrograms. These results suggest that the SOM algorithm can be used to select a small subset of genes that is sufficient to differentiate between different mechanisms of rejection based on different molecular profiles of gene expression. Future studies of clinical specimens should provide associations between gene expression and diagnosis, prognosis, and optimal therapeutic regimens.

Models of rejection

Current textbook models of rejection typically emphasize the role of T cells in the generation of regulatory and effector mechanisms that produce rejection [48]; however, the authors' studies, plus the work of many laboratories, indicate that

there are additional crucial components of the biologic response following transplantation. One important focus of the authors' research is the role of antigen-independent mechanisms in modulating rejection. Studies have shown correlations between graft survival and factors such as brain death, ischemia, reperfusion, and stress. The authors' have identified subsets of genes that were up-regulated significantly because of antigen-independent mechanisms. Specifically, the authors have identified genes that were up-regulated in response to ischemia, stress, and antigen-independent mechanisms presumably caused by innate immune responses. Consistent with this interpretation, the analysis of biologic processes of regulated genes using the Gene Ontology database demonstrated that processes in addition to defense, including communication, were up-regulated during rejection (Fig. 4).

Because of technical limitations, most previous studies of transplantation have not produced quantitative analyses of gene expression; however, the authors' recent studies suggest that quantitative changes in expression may be a common response to therapeutic and genetic perturbations. In murine transplant studies, an important observation based on the research of many laboratories is that knockout mice that are deficient in immune genes often demonstrate a phenotype of delayed rejection. More than 100 gene-deficient mice have been reported to eventually reject grafts with delayed kinetics. In contrast, only a few knockout strains have been shown to have indefinitely prolonged allograft survival. The authors' analysis of gene expression in several of these strains with delayed rejection, including the strain deficient in TNF receptor 1, revealed quantitative changes in gene expression levels; however, few genes were qualitatively absent. This observation also has important clinical implications. The molecular profiles of tissue specimens likely will require quantitative determinations of the expression levels to generate effective diagnostic and prognostic interpretations.

Future considerations

Expression array technology is becoming an increasingly powerful and accepted tool for experimentation; in the future, it will likely become an important diagnostic tool in clinical laboratories. The unique advantage of array technology is its power to measure simultaneous expression of thousands of genes. Current biologic knowledge has been established using reduction techniques that focus on the analysis of a few, or even a single, gene per experiment. With the explosion of data produced by array technology that analyze the complete transcriptome, the future challenge will be to integrate the data into global views of biologic responses, including understanding the complex regulatory networks modulating the responses. Proof of principle studies in yeast has demonstrated the feasibility of producing global interpretations. These studies have confirmed previous biologic observations and have revealed new information, such as linking genes of unknown function with specific biochemical pathways.

To facilitate interpretation of the large amount of data by the scientific community, standards for data entry in databases (MIAME [minimum information about a microarray experiment] format) [49] and databases that are linked to the Internet for easy access must be established, similar to databases for gene identification and genome sequences. The combination of microarray technology, bioinformatics, and central databases should produce an explosion in discovery science that generates new paradigms of biologic responses.

References

- [1] Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
- [2] Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science* 2001; 291:1304–51.
- [3] Gasch AP, Spellman PT, Kao CM, et al. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* 2000;11:4241–57.
- [4] Akalin E, Hendrix RC, Polavarapu RG, et al. Gene expression analysis in human renal allograft biopsy samples using high-density oligoarray technology. *Transplantation* 2001;72: 948–53.
- [5] Saiura A, Mataki C, Murakami T, et al. A comparison of gene expression in murine cardiac allografts and isografts by means DNA microarray analysis. *Transplantation* 2001;72:320–9.
- [6] United Network for Organ Sharing. 2001 Annual report of the US Organ Procurement and Transplantation Network and the Scientific Registry for Transplant Recipients: transplant data 1991–2000. Richmond (VA): United Network for Organ Sharing; 2001. <http://www.unos.org>. Accessed December 2001.
- [7] Nguyen D, Shennib H. Early immunological events in the lung allograft. In: Shennib H, editor. *Immunology of the lung allograft*. Austin, TX: R.G. Landes Company; 1995. p. 1–28.
- [8] Hayday A, Viney JL. The ins and outs of body surface immunology. *Science* 2000;290:97–100.
- [9] Holt PG. Antigen presentation in the lung. *Am J Respir Crit Care Med* 2000;162(Suppl): S151–6.
- [10] Sekine Y, Bowen LK, Heidler KM, et al. Role of passenger leukocytes in allograft rejection: effect of depletion of donor alveolar macrophages on the local production of TNF- α , T helper 1/T helper 2 cytokines, IgG subclasses, and pathology in a rat model of lung transplantation. *J Immunol* 1997;159:4084–93.
- [11] Freeman WM, Walker SJ, Vrana KE. Quantitative RT-PCR: pitfalls and potential. *Biotechniques* 1999;26:112–22, 124–5.
- [12] Yamamoto M, Wakatsuki T, Hada A, et al. Use of serial analysis of gene expression (SAGE) technology. *J Immunol Methods* 2001;250:45–66.
- [13] Sherlock G. Analysis of large-scale gene expression data. *Curr Opin Immunol* 2000;12:201–5.
- [14] Ideker T, Thorsson V, Ranish JA, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 2001;292:929–34.
- [15] Buhler J. Anatomy of a comparative gene expression study. Available at: <http://www.cs.wustl.edu/~jbuhler/research/array/>. Accessed January 15, 2002.
- [16] Lee ML, Kuo FC, Whitmore GA, et al. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci U S A* 2000;97:9834–9.
- [17] Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* 2001;98:31–6.
- [18] Shi L. DNA microarray (genome chip): monitoring the genome on a chip. Available at: <http://www.gene-chips.com/>. Accessed January 15, 2002.

- [19] 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;95:14863–8.
- [20] Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 2001;344:539–48.
- [21] Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531–7.
- [22] Takahashi M, Rhodes DR, Furge KA, et al. Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. *Proc Natl Acad Sci U S A* 2001;98: 9754–9.
- [23] Damrauer S, DeFina RA, He H, et al. Molecular profiles of allograft rejection following inhibition of CD40 ligand costimulation differentiated by cluster analysis. *J Leukoc Biol* 2002;71: 348–58.
- [24] Ideker T, Thorsson V, Siegel AF, et al. Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data. *J Comput Biol* 2000;7:805–17.
- [25] Fisher LD, VanBelle G. Biostatistics a methodology for the health sciences. New York: John Wiley & Sons; 1993.
- [26] Spellman PT, Sherlock G, Zhang MQ, et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 1998; 9: 3273–97.
- [27] Iyer VR, Eisen MB, Ross DT, et al. The transcriptional program in the response of human fibroblasts to serum. *Science* 1999;283:83–7.
- [28] Dopazo J, Zanders E, Dragoni I, et al. Methods and approaches in the analysis of gene expression data. *J Immunol Methods* 2001;250:93–112.
- [29] Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;403:503–11.
- [30] Kohonen T. Self organizing maps. 3rd edition. New York: Springer; 2001. p. 86–113, 217.
- [31] Tamayo P, Slonim D, Mesirov J, et al. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci U S A* 1999;96:2907–12.
- [32] Pilpel Y, Sudarsanam P, Church GM. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat Genet* 2001;29:153–9.
- [33] Raghuraman MK, Winzeler EA, Collingwood D, et al. Replication dynamics of the yeast genome. *Science* 2001;294:115–21.
- [34] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25–9.
- [35] Mills JC, Roth KA, Cagan RL, et al. DNA microarrays and beyond: completing the journey from tissue to cell. *Nat Cell Biol* 2001;3:E175–8.
- [36] Wu TD. Analysing gene expression data from DNA microarrays to identify candidate genes. *J Pathol* 2001;195:53–65.
- [37] Steuervald N, Cohen J, Herrera RJ, et al. Analysis of gene expression in single oocytes and embryos by real-time rapid cycle fluorescence monitored RT-PCR. *Mol Hum Reprod* 1999;5: 1034–9.
- [38] Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* 2000;25:169–93.
- [39] Cahill DJ. Protein and antibody arrays and their medical applications. *J Immunol Methods* 2001;250:81–91.
- [40] MacBeath G, Schreiber SL. Printing proteins as microarrays for high-throughput function determination. *Science* 2000;289:1760–3.
- [41] de Wildt RM, Mundy CR, Gorick BD, et al. Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nat Biotechnol* 2000;18:989–94.
- [42] Koga S, Auerbach MB, Engeman TM, et al. T cell infiltration into class II MHC-disparate allografts and acute rejection is dependent on the IFN- γ -induced chemokine Mig. *J Immunol* 1999;163:4878–85.

- [43] Bahram S, Arnold D, Bresnahan M, et al. Two putative subunits of a peptide pump encoded in the human major histocompatibility complex class II region. *Proc Natl Acad Sci U S A* 1991; 88:10094–8.
- [44] Lakkis FG, Arakelov A, Konieczny BT, et al. Immunologic “ignorance” of vascularized organ transplants in the absence of secondary lymphoid tissue. *Nat Med* 2000;6:686–8.
- [45] Mistry MJ, Clay MA, Kelly ME, et al. Apolipoprotein E restricts interleukin-dependent T lymphocyte proliferation at the G1A/G1B boundary. *Cell Immunol* 1995;160:14–23.
- [46] Hisamatsu H, Shimbara N, Saito Y, et al. Newly identified pair of proteasomal subunits regulated reciprocally by interferon γ . *J Exp Med* 1996;183:1807–16.
- [47] Pierson 3rd RN, Chang AC, Blum MG, et al. Prolongation of primate cardiac allograft survival by treatment with ANTI-CD40 ligand (CD154) antibody. *Transplantation* 1999;68:1800–5.
- [48] Coffman T. Inflammatory response to allografts. In: Norman D, Suki WN, editors. *Primer on transplantation*. Thorofare, NJ: American Society of Transplant Physicians; 1998. p. 33–41.
- [49] Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;29:365–71.