ANALYSIS OF HUMAN IMMUNE MEDIATED DISEASES AND THEIR MURINE MODELS BY GENE EXPRESSION PROFILING

By

Zheng Liu

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Microbiology and Immunology

August, 2006

Nashville, Tennessee

Approved:

Professor James W. Thomas

Professor Thomas. M. Aune

Professor Andrew J. Link

Professor Marylyn D. Ritchie

Professor Mark Boothby

For my mother

ACKNOWLEDGEMENTS

I wish to cordially thank my mentor Thomas M. Aune for his guidance, encouragement, support and friendship during my graduate career. I am very thankful for the guidance and advice from my other present and past thesis committee members: Mark Boothby, James W. Thomas, Andrew J. Link, Marylyn D. Ritchie, Nancy J. Olsen and Jason Moore. I have had wonderful time in Aune laboratory during this thesis work; I wish to express my thanks to all the present and past members of this laboratory. The thesis work reported herein is a collaborative effort and could not be accomplished without the participation of Kevin Maas, Robert W. Yelverton, Bryan Kraft, S. Bobo Tanner, and Tuulikki Sokka.

I must thank the health-care providers and patients who contributed to these studies for their help and cooperation. The patients' samples are extremely important for this thesis study, without their support, these works cannot be accomplished.

I also want to thank my friends: Alex Eshaghian, Frank Q. Zhan, and Kimberly G. Norman. They give me the encouragement and support for my graduate career.

Finally, I want to thank my loving wife, Min Liao (Tracy). Her support helps me be able to focus on my research and to accomplish the studies of this thesis.

Financial support for the author was provided from: National Institutes of Health (AI53984, AI44924, and DK58765), Vanderbilt University Medical Center Discovery Grant, and the Morgan Family Foundation, the McGee Foundation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BCR:	B cell receptor
DMARDs:	Disease-modifying anti-rheumatic drugs
ERA:	early rheumatoid arthritis
GDM:	gene distance matrix
IDDM:	insulin dependent diabetes mellitus
MHC:	major histocompatibility complex
MS:	multiple sclerosis
NOD:	non obese diabetic
NZM:	new zealand mixed
BB rats:	BioBreeding rats
PBMC:	peripheral blood mononuclear cell
PCA:	principal components analysis
PCR:	polymerase chain reaction
RA:	rheumatoid arthritis
SAM:	significance analysis of microarray
SIT:	specific immunotherapy
SLE:	system lupus erythematosus
SVM:	support vector machines
TCR:	T cell receptor
TNF:	tumor necrosis factor

CHAPTER I

INTRODUCTION

Autoimmune diseases and atopic diseases

The immunological dysfunctions may result from uncontrolled excessive immune responses to either self-antigens (autoimmune diseases) or environmental innocuous antigens (atopic diseases). Both autoimmune diseases and atopic diseases are manifestations of such hypersensitivity of immune responses:

Autoimmune diseases

Autoimmune diseases affect 3-5% of the human population (1). Clinical manifestations are the damage of specific organs, such as in insulin dependent diabetes mellitus (IDDM) and multiple sclerosis (MS) or damage to multiple organs and tissues, such as in rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE), resulting from inappropriate immune-mediated inflammation (2). In this thesis, we tried to using gene expression profiling method to study these four human autoimmune diseases: IDDM, insulin dependent diabetes also called type I diabetes (formerly known as "childhood" or "juvenile" diabetes) is most commonly diagnosed in children and adolescents. It is an autoimmune disorder, in which the immune system attacks the self-beta cells in the islets of the pancreas, thus reducing the insulin production (3, 4). MS, multiple sclerosis, is a chronic autoimmune disease that affects the brain and spinal cord in which the immune system attacks the myelin (sheath) surrounding nerve cells (5, 6).

RA, rheumatoid arthritis, is a chronic, inflammatory autoimmune disorder in which the immune system attacks the self joints. The disease is also systemic in that it often also affects the skin, blood vessels, heart, lungs and muscles (7, 8). ERA, early rheumatoid arthritis, is a subject of the study in this thesis, which refers to the rheumatoid arthritis whose disease duration is less than two years. SLE, system lupus erythematosus, is a chronic, potentially fatal autoimmune disease in which the immune system attacks the self organs and tissue, resulting in inflammation and tissue damage. SLE can systemically affect the body, and most often harms the kidneys (lupus nephritis), heart, joints, skin, lungs, blood vessels and nervous system (9, 10).

The failure of self-tolerance is believed to be the underlying cause of the formation of immune responses to self organ or tissues. Self-tolerance can be divided into two broad categories: central tolerance (11, 12) and peripheral tolerance (13, 14). In central tolerance, autoreactive immature lymphocytes in generative lymphoid organs (bone marrow for B cells and thymus for T cells) are typically eliminated by apoptosis resulting in clonal deletion or are inactivated. Peripheral tolerance controls the mature autoreactive lymphocytes that escape the central tolerance checkpoints and migrate to periphery. Basic peripheral self tolerance mechanisms include: clonal deletion of autoreactive lymphocytes via apoptosis, lymphocytes anergy due to intrinsic regulation to increase the activation threshold of autoreactive lymphocytes to self antigens (15), eliminating or reducing antibodies or BCRs or TCRs self reactivity by V(D)J recombination or somatic hypermutation, and extrinsic active suppression by T regulatory cells (16).

Recently, a new theory has been proposed to explain the origins of autoimmune disease by studying the murine model of human IDDM – non obese diabetic (NOD) mice (17). It has been recognized for some time that NOD mice are lymphopenic (reduced numbers of lymphocytes) compared to other strains. The model proposes that the host recognizes the state of lymphopenia and this drives naïve T cells to undergo homeostatic expansion in an attempt to fill the empty space. Proliferation of naïve T cells is dependent upon recognition of antigen and the only antigens in the host in the uninfected state are self-antigens. Therefore, this model proposes that the tendency to correct lymphopenia supersedes the need to avoid self-reactivity resulting in the generation of a population of self-reactive T cells, which migrate to extra-lymphoid sites and increase the susceptibility to autoimmune diseases at these sites. This may explain why lymphopenia resulting from immunotherapies may predispose people to autoimmune disease (18), such as induction of thyroid autoimmunity in 1/3 of patients with multiple sclerosis after lymphocyte depleting therapy (19). Studies of another type I diabetes rodent model, the BB (BioBreeding) rat, also suggested that lymphopenia may be an essential factor of inducing autoimmunity in this model (20-22).

Environmental factors may also predispose individuals to develop autoimmune disease. Infectious agents can mimic the structure of self-antigens, which may induce cross-reactivity of lymphocytes to self-antigens and finally result in autoimmune disease. This phenomenon has been observed in a variety of cases, such as relapses in MS are often triggered by common viral infections, herpesvirus, influenza, measles, papilloma virus and Epstein-Barr Virus which all have genes encoding sequences that mimic those found in the major structural proteins of myelin (6); the spirochete *Borellia burgdorferi*.

shares an antigenic determinant with LFA-1 which is targeted by autoimmune response in lyme arthritis (23); a single antigenic determinant on Epstein-Barr virus is also shared with one of the auto-antigens of SLE (24). The components of infectious microorganisms such as lipopolysaccharide (LPS), bacterial DNA may substantially improve immune responses to unrelated antigens (1, 25, 26), thus acting as adjuvants to augment immune responses and overriding checkpoints that prevent responses to self-antigens. Adjuvants components of infectious microorganisms such as LPS can stimulating the innate immune system to produces inflammatory cytokines and increases the activity of antigen-presenting cells by inducing class II MHC and costimulatory proteins (27, 28).

Autoimmune diseases also have an underlying genetic basis that is believed to increase disease susceptibility. The genetic predisposition to autoimmune diseases is supported by epidemiologic evidence. Studies showed when compared autoimmune disease incidence between monozygotic twins, between dizygotic twins, among family members, and within the general population, the greatest risk of developing autoimmune disease is if one's monozygotic twin has an autoimmune disease, the second greatest is if one's dizygotic twin has an autoimmune disease, and the person also has an increased risk of developing autoimmune disease if his/her first-degree relative is with an autoimmune disease (2, 29-31). The genetic predisposition to autoimmune disease can be divided into two categories: single-gene defect conferring susceptibility to an autoimmune disease and polygenic defects that confer susceptibility to develop an autoimmune disease (32). There are examples in both human autoimmune disease and animal models of autoimmune disease demonstrating that a single gene defect can induce

autoimmune disease (32). Examples include mutations of *AIRE* in human autoimmune polyendocrine syndrome (APS-1) (33); induced knockout or spontaneous mutation of mouse *Foxp3* gene, which exhibits high expression in $CD4^+CD25^+$ T regulatory cells, leads to systemic autoimmune diseases resulting from absence of $CD4^+CD25^+$ T regulatory cells (16). Mutation of *FOXP3* is associated with human IPEX syndrome (immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome). Mutation of *Fas* or *Fasl* in lpr/lpr and gld/gld mice results in severe lymphoproliferative syndromes, similar mutations are also found in humans with lymphoproliferative syndrome (ALPS) (34, 35).

Single gene defects that result in autoimmune syndromes may be straightforward and simple to explain. However, single gene defects leading to human autoimmune syndromes are very rare and more and more evidence supports the concept that the major autoimmune diseases result from a complex combination of polygenic effects (32, 36). Genetic linkage studies of human autoimmune diseases and their murine counterparts provide strong support for the existence of multigenic effects that contribute to autoimmune disease susceptibility (37-39). Comparative genetic mapping of human autoimmune diseases and their murine counterparts suggest that a common set of susceptibility loci contribute increased susceptibility to several different autoimmune diseases (2, 40, 41). The overlapping loci not only exist within each species but also exist between human and murine models when susceptibility loci are placed on the same genetic map (1, 2, 37, 38, 40-44). The fact that there are common loci shared by different autoimmune diseases is the basic rationale for our study to identify common gene expression profiles in autoimmune diseases.

Therefore, the combination of the genetic components, environmental factors and stochastic events predispose individuals to autoimmune diseases. The increased number of susceptibility alleles may increase the risk and incidence of autoimmune diseases and reflect a "threshold liability". This "threshold liability" model postulates the existence of a continuous trait, termed liability, and a threshold value of the liability is corresponding to different phenotype states. Individuals whose liability exceeds a given threshold will exhibit a certain phenotype, such as a certain disease; those with liability value below the threshold exhibit a opposite phenotype, such as lack of the diseases (45). The mechanism of "threshold liability" has been reported to contribute to establish susceptibility in a variety of diseases such as respiratory diseases, certain types of cancers, and Alzheimer disease (46-48). The "threshold liability" model in autoimmune diseases has been reviewed by Wandstrat et al (2) (Figure 1-1, Wandstrat et al 2001).

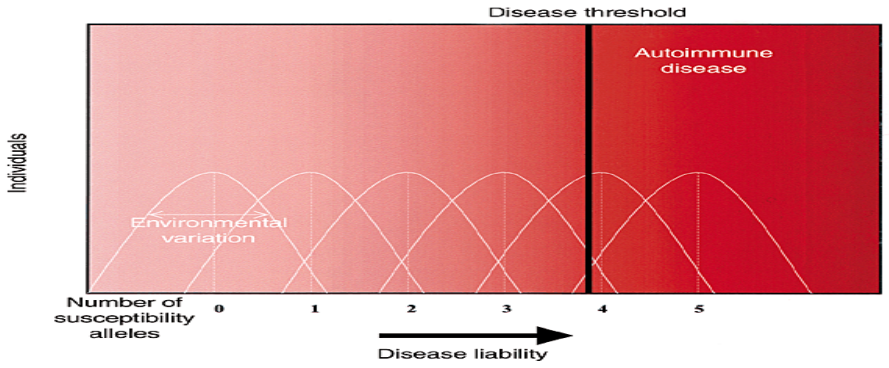


Figure 1-1. Threshold liabilities in autoimmune disease. In this model, only individuals located to the right of the disease threshold line will develop disease. The *x* axis represents increasing liability to disease, individuals located on the *x* axis based on the degree of their predisposition to disease. An incremental increase in the number of susceptibility alleles progressively increases liability of disease, resulting in movement toward the disease threshold at the right on the *x* axis. The disease liability introduced by environmental and stochastic effects is represented by the normal distribution curve around the location of individuals with specific degrees of genetic predisposition for disease.

Murine models of autoimmune diseases have been widely used to study disease pathogenesis. In the study of this thesis, we have used the NOD mouse that is a type I diabetes murine model and the NZM2410 mouse that is a model of human systemic lupus erythematosus. We have also used the NOD congenic strain NOD.H2^{h4} mice to perform genomic comparisons and the C57BL/6 strain was used as a control strain.

The NOD strain was originally developed in Japan during selection of a cataract-prone strain derived from outbreed Jc1: ICR mice. The NOD strain was established as a subline by inbreeding of brother-sister mating with increased incidence of spontaneous diabetes (49, 50). The incidence of diabetes in NOD mice is 60%-80% in females and 20%-30% in males, and interestingly, the highest incidence of diabetes occurs when mice are maintained in a pathogen free environment but incidence is decreased when mice are maintained in a relatively "dirty" environment (51, 52). The underlying mechanism of this effect is not clear. However, this phenomenon may reflect that the exposure to foreign antigens may help gradually maturing the immune system. Therefore, this matured immune system may further protect the host from allergic or autoimmune diseases (53). Onset of diabetes generally begins at 12 to 14 weeks of age in female mice and is a bit delayed in male mice. Mononuclear infiltration surrounding islets (peri-insulitis) begins at 3 to 4 weeks of age. Progression and invasion of islets (insulitis) proceeds during the following weeks and most mice demonstrate insulitis by 10 weeks of age. Marked decreases in pancreatic insulin content occur in females at about 12 weeks of age and several weeks later in males (50). The delayed incidence of diabetes in male mice may suggest late regulatory events (50). Genetic analysis of NOD mice has demonstrated that the phenotype is controlled by loci distributed over at least 11 different

chromosomes and 19 different disease susceptibility loci have been designated (*Idd* loci, refer to insulin-dependent diabetes (41)). Although all of these susceptibility loci contribute to development of diabetes of NOD mice, the *Idd1* locus confers the greatest risk (1, 50, 54-58). Genes within this locus encode major histocompatability complex (MHC) proteins of the class II type, I-A and I-E. The NOD haplotype I-A^{g7} has been intensively studied and confirmed to be the major genetic risk factor for susceptibility of diabetes (41, 54, 59, 60). The pathogenesis of diabetes in NOD mice has been studied for decades. Several mechanisms, other than genetic predisposition to diseases susceptibility, have been proposed such as break of central and peripheral tolerance, defects in regulatory systems and recently proposed mechanisms of generation of autoimmunity by homeostatic expansion under lymphopenic pressure (50). The evidence identified in another diabetes rodent model-BB rat (BioBreeding rat) supports that lymphopenia may drive to autoimmunity. The profound lymphodeficiency of the BB rat is demonstrated to contribute to diseases pathogenesis (20, 21). However, all these above mechanisms contributing to autoimmune diseases may also develop from the genetic background of NOD mouse.

NOD mice have been used as a surrogate for human diabetes for several decades and our understanding of pathogenesis of type I diabetes (IDDM, insulin dependent diabetes mellitus) is advanced by analyses of this model. However, because of the genus difference, limitations of this modeling exist, such as, deafness, the absence of C5 complement, resistance to ketoacidosis and the absence of the murine homolog of HLA-DR molecules on antigen-presenting cells in NOD mouse (55). We must realize that this highly inbred type I diabetes model can only be viewed as analogous to a single “case

study" in humans. It can not fully model the heterogeneous genetics of human type I diabetes.

Congenic strains of NOD mouse represent a powerful tool to dissect the polygenic effects of multiple susceptible loci of type I diabetes. The congenic strains are special types of inbred strains in which part of the genome is transferred from another inbred strain by backcrossing the donor mouse and recipient mouse with a phenotypic or genotypic selection. In this study (see chapter VI: part I), we used a NOD congenic strain, NOD.H2^{h4}, which contains I-A regions (*Idd1*) transferred from B10.H2^{h4} mouse onto the NOD background. Thus, this congenic strain changes the composition of the loci, *Idd1,* that confers the greatest disease liability of NOD mouse and this congenic strain does not develop type I diabetes.

NZM2410 is a model of human systemic lupus erythematosus. It is one of the NZM strains originated from crossing NZB and NZW strains (61). This inbred strain develops glomerulonephritis by one year of age in ~80% of males and females. Anti-dsDNA IgG antibodies are also present in the sera of ~80% of animals by six months of age. The strain is a valuable tool to investigate autoimmunity and the genetic components and pathology of murine systemic lupus erythematosus (SLE). Spontaneous SLE in NZM models is characterized by serum autoantibodies to chromatin, double-stranded DNA (dsDNA), histone proteins, and other antigens (61).

Atopic diseases

Atopic diseases are another group of immune-mediated human diseases studied in this thesis. They comprise a group of syndromes that include asthma, allergic rhinitis, and atopic dermitis. These diseases are caused by allergic response belonging to immediate hypersensitivity, which is the pathologic process that results from immunologically specific interactions between antigens (exogenous or endogenous) and humoral antibodies or sensitized lymphocytes. Immediate hypersensitivity is an immune reaction that is initiated by antigen binding to IgE on mast cells or basophils and leading to release of inflammatory mediators. These mediators cause allergic responses such as increased vascular permeability, vasodilation, bronchial and visceral smooth muscle contraction, and local inflammation. Figure 1-2 shows the cascade of allergic responses.

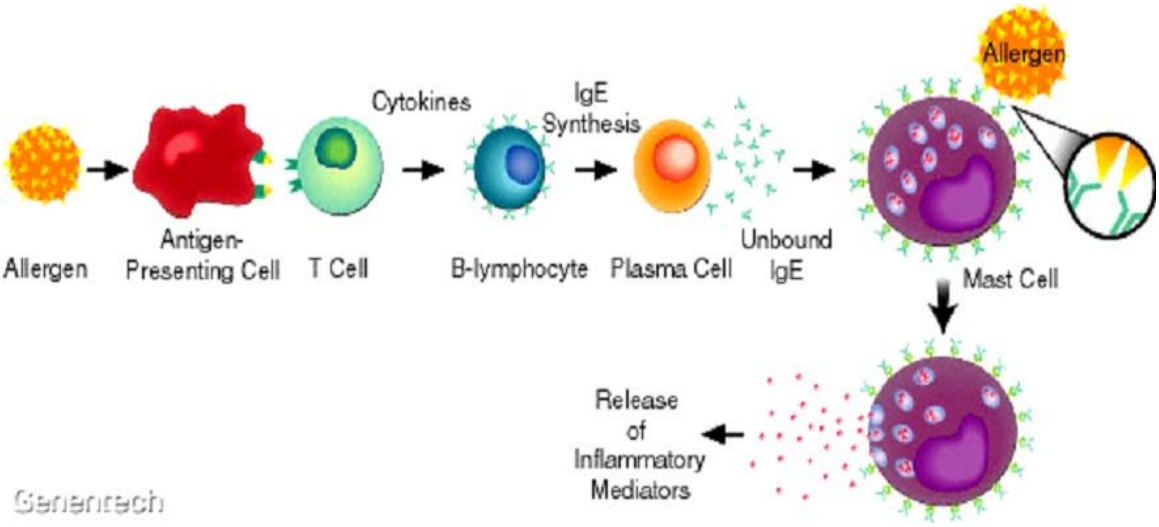


Figure 1-2. Cascade of allergic responses.

This reaction is called immediate hypersensitivity, because it begins rapidly, within minutes of antigen challenge and because it has major pathologic consequences

(hypersensitivity). In the most extreme systemic form of this reaction, called anaphylaxis, mast cell-derived or basophil-derived mediators can cause bronchial constriction to the point of asphyxiation, massive tissue edema and cardiovascular collapse leading to death (62).

As with the mechanisms of autoimmune diseases discussed above, the genetic analysis of atopic diseases also indicates that the susceptibility to atopy also arises from multi-genenic interactions and effects of environmental factors (63-65). Epidemiologic evidence from first-degree family members of individuals with atopic diseases and disease predisposition studies of large numbers of twins clearly indicate a strong genetic component to atopic disease (63, 64, 66, 67). The prevalence of allergic diseases including asthma, allergic rhinitis, atopic dermatitis, and allergic conjunctivitis in first-degree relatives of diseased individuals is significantly higher than in relatives of non-diseased individuals. Also, in these studies, the genetic impact can be viewed as independent from geographical components, since these studies were performed on different continents (66-68). The analysis of monozygotic and dizygotic twins demonstrates that the quantifiable traits of allergic diseases, such as serum IgE levels and skin test reactivity, show much stronger association with allergic disease in monozygotic twins than those in dizygotic twins (69). Similar to autoimmune genetic studies, whole genome screens and linkage analyses have revealed several susceptible loci, and related candidate genes have been identified in both human and murine models (63). Besides genetic effects, multiple environmental factors, such as temperature, molds, air pollutants, and smoke, also contribute to susceptibility to allergic diseases (65, 70-72).

The allergen specific immunotherapy (SIT) is a common and effective treatment for atopic diseases. However, its specific mechanism of action is not well understood. Generally, the SIT is thought to modulate the immune response and decrease the sensitivity to environmentally innocuous antigens (73). The operation of SIT is that small but increasing quantities of antigen are administered subcutaneously over a period of hours or more gradually over weeks or months. As a result of this treatment, specific IgE levels decrease and IgG titers often rise, perhaps further inhibiting IgE production by neutralizing the antigen and by antibody feedback (74). It is also possible that immunotherapy may work by inducing specific T cell tolerance or by changing the predominant phenotype of T helper cells from Th2 to Th1 (62, 75, 76). Recently, IL-10-secreting T regulatory cells have been proposed to modulate allergic responses and the SIT may, at least, in part restore the normal function of regulatory T cells to control the allergic responses (77, 78).

Gene expression profiling by microarray

As discussed above, because of the multigenic features of autoimmune diseases and atopic diseases, the microarray technology opened new avenues with great advantages to studying multigenic diseases. Instead of studying expression levels of a single gene or a handful of genes, the microarray can screen the expression pattern of tens of thousands genes simultaneously and provide an expression signature of a specific state in certain types of cells, the microarray data has been used to provide valuable information about disease phenotype, progression, prognosis, and responses to certain

therapies (79-86). Thus, the understanding of the molecular signatures of immune-mediated diseases may open a variety of potential diagnostic and therapeutic avenues.

Autoimmune and atopic diseases arise from complex interactions between multi-genetic products and environmental factors. Microarray technology has the advantage of screening levels of tens of thousands gene transcripts in a given experiment simultaneously. Therefore, this high throughput technology provides a powerful platform to study the complex network of gene expression in these immune-mediated diseases. It is thought that changes in gene expression profiles may reflect disease classification or identity, disease progression, disease activity, other disease manifestations, and may predict the direction of future developments. Therefore, gene expression profiling may provide valuable information about underlying mechanisms or pathways of diseases, diseases subtypes or classification, responses to treatment or environmental factors and information about future diseases activity. Thus, the gene expression profiling technology may shed light on discovering new biomarkers for diseases typing, new diagnostic or prognostic methods, and may help identify underlying causes that could lead to design of improved therapies or drugs to treat these diseases.

Application to autoimmune diseases

Gene expression profiling identified a common profile in the peripheral blood mononuclear cells (PBMC) among four different autoimmune diseases-IDDM, SLE, MS and RA (79), and the genes most differentially expressed are those that encode proteins involved in apoptosis, cell cycle progression, cell differentiation and cell migration. Four groups have independently discovered a specific “IFN signature” gene expression profile

in PBMC and whole blood in individuals with SLE (83, 87-90). Microarray analysis of RA has suggested that there are specific molecular signatures of gene expression that are present in the PBMC of individuals with RA (91) and a specific gene expression signature exists in recent onset RA (early RA, ERA), which shows overlap with a subset of patients with systemic lupus erythematosus (80). The specific signature for ERA has potential clinical application since it may permit early identification of patients that will develop RA and provide for earlier treatments that may prevent or delay morbidity and mortality. In addition to PBMC, synovial tissue samples have also been examined to identify unique gene expression profiles associated with RA (92, 93). Autopsy samples were also used to perform gene expression profiling in MS (81, 94). The comparison of gene expression in PBMC between SLE and MS showed a common pattern (95). Therefore, all the results of gene expression profiling of autoimmune diseases indicates that gene expression signatures are both common to several autoimmune diseases and unique to a given property of an individual autoimmune disease. Theses results are consistent with genetic linkage studies that demonstrate the existence of shared susceptibility loci among several different autoimmune diseases and loci that are unique for a given autoimmune disease.

Application to atopic diseases

There are a few gene expression profiling studies in atopic disease. The gene expression profiling of nasal mucosal biopsy samples from patients with allergic rhinitis demonstrates that this expression profile is dominated by immunoglobulin, especially IgA, that may play a role in eosinophil degranulation (96). Another study comparing gene

expression profiles in PBMC between atopic individuals and unaffected individuals suggests up-regulation of pro-survival signals in atopic individuals compared to healthy controls, which may be the underlying mechanism of long-lasting hypersensitivity responses in atopic individuals (97). Microarray based on the genes that are reported to be relevant in pathogenesis of atopic diseases were designed and were successfully used to diagnose atopy and asthma and assess disease activity (98). A differential gene expression profile has also been identified in bronchial tissues of asthmatic individuals, and strikingly, one third of the differential genes expression profile is corrected by inhaled corticosteroid treatment (99). This result suggested that the therapy did change the specific gene expression profile, but only partial of them, and the expression changes of these genes may be used to monitor the responses to this therapy.

Since the first application of microarray technology more ten years ago (100), the approach has rapidly become a standard technology used in almost every area of biomedical research. Most applications have been performed in cancer research. The gene expression profiling in cancer has been successfully used to diagnoses and classify the types/subtypes of cancers (85, 86, 101-103), evaluate prognosis of cancers (104-109) , predict the responses to certain treatments (110-113). Under controlled conditions, such as the same type tissue used for the microarray, the same hybridization protocols and the same microarray platforms-DNA array or oligo arrays, the results of gene expression profiling from different groups are fairly consistent regardless of the iteration of microarray experiments between or within the laboratories (114). Although clinical uses of microarrays are still experimental and complicated, they are beginning to show promise in clinical application. The first microarray-based tool for cancer diagnosis

(Oncotype DX, made by Genomic Health Inc.) has entered the market and used for clinical diagnosis of breast cancer. It is for helping to predict the likelihood of breast cancer recurrence in women with newly diagnosed, early-stage invasive breast cancer.

In contrast, only a few studies of gene expression profiling have been performed in autoimmune and atopic disease research. As describe above, its high throughput feature makes it an ideal tool to study these multifactoral diseases. With the development of bioinformatics tools and easy handling protocols, the microarray technology should gain more and more attention to investigators pursuing research into autoimmune diseases and atopic diseases. It cannot only help to make scientific discoveries but will also improve clinical interventions to these diseases.

Gene expression levels as quantitative genetic traits

Gene expression profiling methods have been widely used in a variety of areas of biomedical science. Recently, several groups have combined gene expression profiling with genetic linkage analysis (115-119). In these studies, they treated the gene expression level as a quantitative trait and identified the genetic loci that control this trait by linkage analysis. These studies clearly demonstrate in several model systems that the transcript level of a given gene in a given tissue can be an inherited trait (115-119). In our study to examine origins of gene expression signatures in autoimmune disease, we found that unaffected first-degree relatives of individuals with autoimmune disease share a portion of the gene expression profile observed in individuals with autoimmune disease (120). The genes that make up this overlapping profile are among the same genes that are differentially expressed in a number of different autoimmune diseases (79). This argues

that a portion of the gene expression profile observed in autoimmune disease is a familial or genetic trait rather than a product of the disease process. The evaluation of the correlation of parent-offspring pairs by levels of gene expression also confirms that expression levels of previously identified autoimmune signature genes correlate between parent and offspring indicating that they are governed by heredity.

Specific Aims

Autoimmune diseases and atopic diseases are immune-mediated diseases which exhibit the clinical manifestations of uncontrolled excessive immunological response to self-antigens or innocuous environmental allergens. The underlying molecular and cellular mechanisms of diseases pathogenesis have been pursued for decades, a large body of evidence has supported the idea that multigenetic interaction, environmental factors and stochastic events all contributed to disease susceptibility (1, 2, 18, 32, 121). Genetic linkages studies have been performed in humans and rodent models, the results indicated that multiple disease susceptibility genes exist, and some autoimmune disease susceptible loci are even co localized between human and rodent models.

In order to take the challenge of understanding this multigenetic regulation network in these immune-mediated diseases, the gene expression profiling method has been introduced into the autoimmune diseases and atopic diseases research. The gene expression profiling can determine the transcript level of thousands of genes simultaneously, thus providing the systemic image of these multigenetic interaction networks that may involve in diseases pathogenesis.

With application of microarray technology, we compared the gene expression profiles of four different autoimmune diseases-RA, MS, IDDM, and SLE. Strikingly, all these four different autoimmune diseases shared a common gene expression profile, and interestingly, the unaffected family members of autoimmune diseases individuals also shared this profile (79).

Since the gene expression signature has also been shared with unaffected family members of individuals with autoimmune diseases and further report indicates that a portion of these genes expression levels exhibit significant correlation between family pairs than non-family pairs (120). Therefore, the influence of familial resemblance may mask the specific expression pattern only due to autoimmune diseases phenotype. Thus, in chapter III of this thesis, we identified the gene expression signature of PBMC exclusively associated with the autoimmune disease phenotype and without the influence of familial resemblance.

Atopic diseases are also a group of immune-mediated diseases, and multigenetic interaction also contributes to susceptibility of the atopic diseases. We already identified a unique gene expression signature of PBMC existing in autoimmune diseases, in specific aim 2 of this thesis, we tried to determine if a unique gene expression signature of PBMC also exists in individuals with atopic diseases, and if this expression signature is similar as the signature of autoimmune diseases.

Gene expression profiling has been used not only to determine the specific disease phenotype, but also been used to monitor the responses to certain therapies and to evaluate the prognosis of patients. All these applications have been actively performed in cancer research, but not in autoimmune diseases or atopic diseases study. Therefore, in

chapter IV and V, we tried to advance our gene expression profiling method to the application of monitoring the transcripts level changes upon immunotherapy in individuals with allergic rhinitis, and we also tried to identify the predictor genes that can be used to evaluate the disease course of early RA patients.

The murine models of autoimmune disease have been greatly used in studying the pathogenesis of autoimmune disease and several susceptible loci overlap between human and mouse. Therefore, in chapter VI of this thesis, first, we tried to determine if there is a unique gene expression profile in T cells of autoimmune diseases murine models-NOD mice and NZM mice, and second, we compared them with each other and with their human counterpart-human type I diabetes and systemic lupus erythematosus to examine if there is any overlap of differentially expressed genes or pathways between human autoimmune diseases and murine models, which may be the important contributor to disease pathogenesis.

CHAPTER II

METHODS OF GENE EXPRESSION PROFILING

Microarrays can be grouped into two categories according to the probes on the solid support media- cDNA arrays and oligonucleotide arrays. As for cDNA microarray: the target mRNA is reverse-transcribed to cDNA, and during the reverse transcription process, the cDNA is labeled with a fluorescent dye (Cy3 or Cy5)-two channel microarray, or radioactive nucleotide (^{33}P)-single channel microarray. Then the labeled cDNA is hybridized to the probes (cDNA clones) on the microarray slides or membranes. After washing away the unbound cDNAs, the array is scanned and an image of the array is generated. In the image, the fluorescent intensity or the radioactive intensity of every spot on the array directly correlates the expression level of this cDNA clone in the sample (122). As for oligonucleotide microarray, there are two types of microarrays, two-channel microarray which is similar as the two-channel cDNA microarray described above, and single channel microarray such as commercial microarrays from Affymetrix Inc. As for the Affymetrix single channel microarray, the target RNA is reversely transcribed into cDNA and further was synthesized as biotin-labeled cRNA and hybridized onto this microarray, after staining with antibody conjugated phycoerythrin, intensity of each probe spot is identified (Expression analysis technical manual, Affymetrix, Inc). Also on this microarray, each probe-oligonucleotide is designed to be perfectly complementary to a target sequence and a partner probe is generated that is identical except for a single base mismatch in its center. These probe pairs, called the Perfect Match probe (PM) and the

Mismatch probe (MM), allows the quantization and subtraction of signals caused by non-specific cross-hybridization. The difference in hybridization signals between the partners, as well as their intensity ratios, serves as indicators of specific target gene transcript level (Data analysis fundamentals manual, Affymetrix Inc).

In this study, we used two types of cDNA microarray : fluorescent labeled cDNA microarrays (cDNA clone set printed on the glass) in murine model study, and radioactive labeled cDNA microarrays (cDNA clone set printed on the nylon membrane) in human diseases study. To generate data from the fluorescent-labeled cDNA microarray, two dye tags (Cy5, Cy3) are used to label the experimental sample (Cy5) and the reference sample (Cy3). After the image is generated, the ratio of intensities of Cy5 and Cy3 is calculated and this represents the relative expression level of a cDNA clone or gene in the test sample compared to the reference control. A schematic of this process is shown as in Figure 2-1:

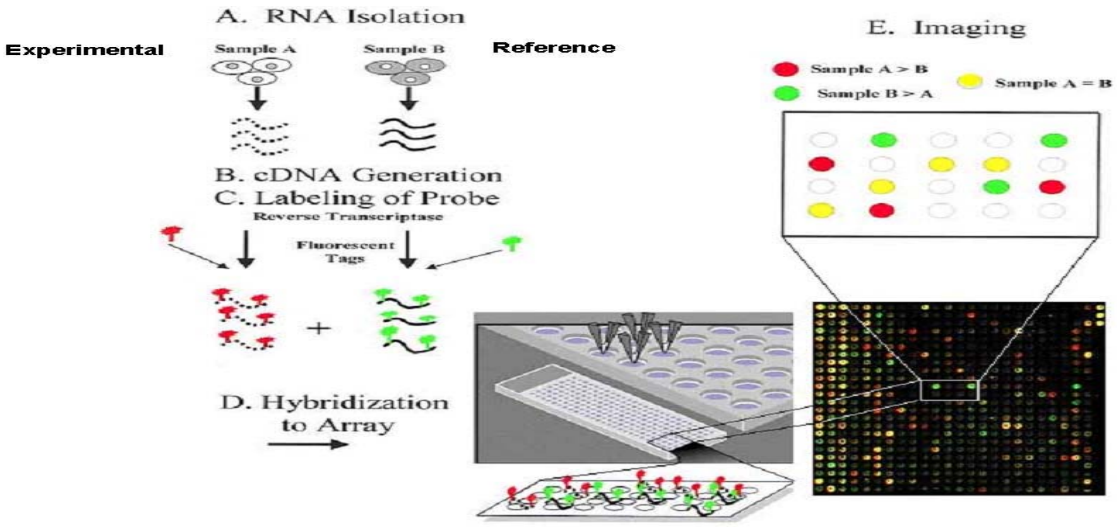


Figure 2-1. Flow chart of process of fluorescent microarray**.**

We also used radioactive labeled cDNA microarray membranes. As above, total RNA is reverse-transcribed and simultaneously radioactively labeled by incorporation of ^{33}P dCTP in the reverse transcription reactions. These labeled cDNAs are purified and hybridized to cDNA probes on nylon membranes. A phosphorimaging system detects the intensity of the each spot on the array, and the image is created and loaded into software (Pathways^{TM}) for basic preprocessing according to the software manual, such as importing image, aligning the image, identifying intensity of the spots on the filer and normalization. After processed by Pathways software, the expression value of each cDNA clone on the array is divided by the average intensity of the whole dataset (normalization); therefore, the average intensity of all the cDNA clones for each experiment is 1.0. The flow chart of this process is shown as Figure 2-2.

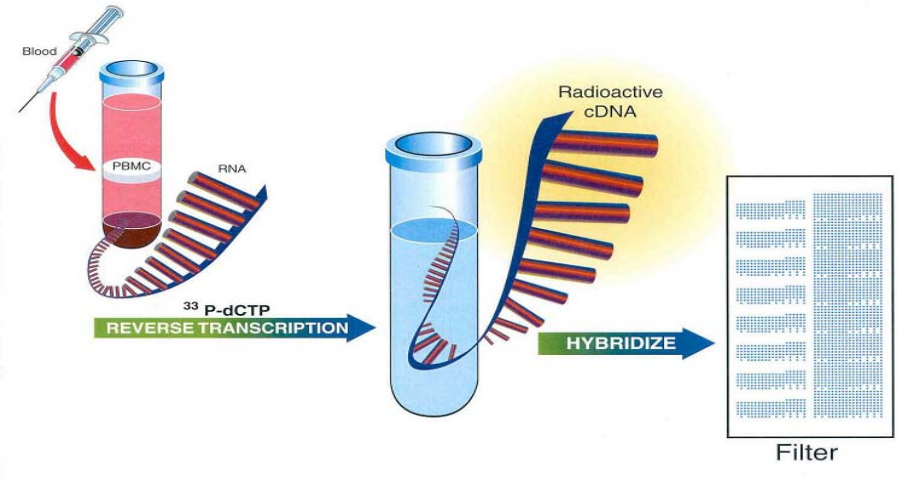


Figure 2-2. Flow chart of microarray with radioactive labeled probe.

Microarray data analysis- preprocessing

The raw data obtained from Pathways software needs further processing before performing data analysis. In this study, we used Cluster 3.0 software (123) to perform preprocessing because of the ease of data manipulation.

Data filtering: the first step is data filtering which removes certain genes that do not have the desired properties. In our case, genes with constant levels of expression across all samples are removed to decrease the dataset to a smaller size. In general, filtering strategies utilize the threshold of variance of gene expression across all the samples or the threshold of a statistic (124). The filtering conditions are arbitrary and there is a general danger of "losing data". We used the first strategy, filtered the data by selecting those genes whose standard deviation of expression levels across all samples exceeded 1.0. These genes were used for further processing.

Log-transformation: The filtered data next undergoes $\log_2$ transformation. Log-transformation is the general step to preprocess normalized data (125, 126) because log-transformation makes the quantitative distribution of over- and under-expressed genes symmetrical (126). Log-transformation also provides a good approximation of the normal distribution (127) and can be applied to original intensity values or ratio values. Different log bases can be used and will generate identical results. In our studies, we used $\log_2$ transformation of the normalized data.

Centralization: centralization is sometimes categorized into the normalization step. It is used to correct the systemic bias arising from differences in mRNA concentration or quality and for differences in labeling efficiency. These biases have the effect of

multiplying values for all genes by a fixed scalar. Mean or median centering of the data in log-space has the effect of correcting this bias (122, 128).

Normalization: All the raw microarray data are normalized to permit comparison among the different groups of microarray experiments, such as a treated group and a control group. There are several normalization methods for microarray data analysis, such as intensity averaging, ratio averaging, Lowess normalization, exogenous control genes, and house keeping genes (128). In this study, we used the built-in normalization method in Cluster 3 software. This method multiplies all values in each row of data by a scale factor S to normalize the log transformed mean centralized data so that the sum of the squares of the values in each row is 1.0 (a separate *S* is computed for each row) (129).

After all these preprocessing steps, we obtain a log-transformed, mean polished, and normalized dataset. It is suitable for further advanced microarray analysis.

Microarray data analysis-advanced microarray analysis

Useful terms: gene expression matrix and expression vectors. Gene expression matrix refers to a dataset of a combination of a number of microarray experiments, as shown in Figure 2-3. Rows represent hybridization intensities to different cDNA clones across all the experiments and columns represent several microarray experiments. The color of each cell of the matrix represents the relative expression of a gene in an experiment.

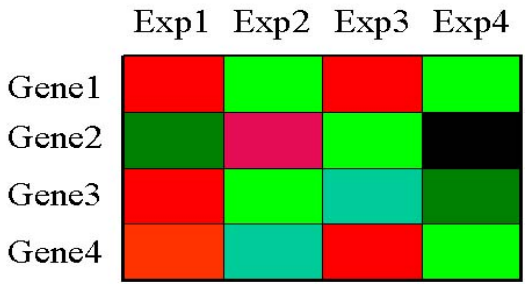


Figure 2-3. Gene expression matrix. The color codes denote the relative expression level of a gene in a microarray experiment. The red color represents the relative expression level is greater than 1, and green color represents the expression level is less than 1. The black color represents the relative expression level is 1 ($\log_2 1=0$).

Gene expression vectors are represented by the values of expression of an individual gene over the range of experiments. Similarly, experiment vectors are represented by the values of all the genes in this experiment (Figure 2-4).

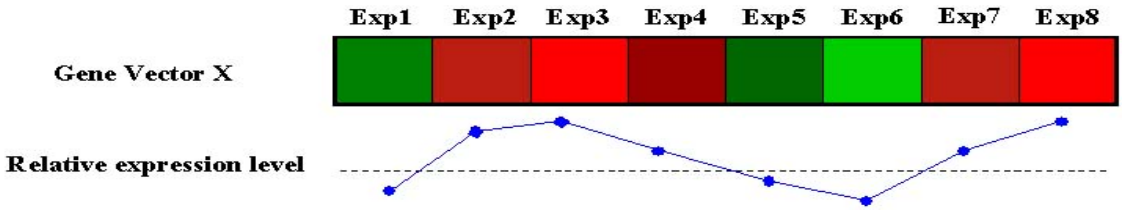


Figure 2-4. Gene Expression Vectors encapsulate the expression of a gene over a number of microarray experiments.

As for advanced analysis of microarray data, we generally have three types of tools. First, supervised or unsupervised methods are used to categorize the data into different groups that may correspond to certain biological classifications. In this study, we used unsupervised methods: hierarchical clustering, principal component analysis, and gene distance matrix demonstration; supervised methods: significance analysis of microarray, support vector machine. Support tress analysis was also used to evaluate the reliability of the clustering results (130). Based upon the information we developed, we

also identified the genes whose expression profile can be used to predict phenotypes of other samples. This mission is carried out by support vector machine analysis. The leave-one-out cross-validation (LOOCV) by support vector machine analysis can be used to internally validate original data and evaluate predictability. Third, for certain sets of gene signatures obtained by microarray, we used EASE analysis to reveal the intrinsic biological themes, which may lead to further functional assays to identify the potential biologically relevant pathways. Basic mechanisms about these methods are as below:

HCL (Hierarchical clustering): The algorithm of HCL begins with each data point as a separate cluster, and during the clustering process of HCL, similar genes are joined into the same groups based upon differences or similarities in expression levels. The iterative process continues with the joining of resulting groups based upon their similarity until all groups are connected in a hierarchical tree. The similarity of expression of genes or clusters during HCL process is determined by the program setting of distance metrics to measure the distance of two gene vectors and linkage methods to determine the distance between two clusters (128). In this study, euclidean distance is calculated to measure the distance metric and average linkage is used as the linkage method to determine the distance between two clusters. There are three linkage methods in the software: single linkage, complete average, and average linkage that can be viewed as the comprised method of the former two linkage methods. The average linkage uses the average of all pair-wise distances between gene vectors in one cluster and vectors in the other cluster (131).

ST (support tree clustering): ST is used to identify hierarchical trees and to show the statistical support for the cluster of the trees, based upon Jackknife re-sampling of the

data (130). Jackknife re-sampling takes each expression vector and randomly omits a sample. This method produces expression vectors that have one fewer sample and this is often done to minimize the effect of single outlier values (130, 132). For each re-sampling process, a HCL is performed and the result is compared to the original clustering result. The percentage of the original clustering results that occur during the number of re-samplings indicates the level of reliability or support for the clustering result.

GDM (gene distance matrix): GDM is used to generate a distance metric matrix to demonstrate the similarity between samples (gene vectors or experimental vectors) and thus the groups that contain the samples (gene group or experiment group). The distance metric (Euclidean Distance) is calculated between two samples (vectors) to provide an intuitive and comprehensive view of the distance (or similarity) between any two samples. These distances are depicted as a colored matrix representing all sample-to-sample distances. The intensity of the color represents the distance or similarity between the two samples.

PCA (Principal Component Analysis): PCA is an exploratory multivariate statistical technique that allows the identification of key variables (or combinations of variables), principal components, which represent a multi-dimensional data set (for example, a number of gene expression variables in a number of samples). Therefore, the PCA reduces the dimensions of the original dataset and magnifies the tendency of the data. In this way, the underlying overall relationships of the data can be uncovered. Generally, the two or three most representative principal components are used to project all the data points to a two- or three- dimensional space (133, 134).

SAM (Significant Analysis of Microarray): SAM analysis is used to identify the significant genes based on differential expression between different groups according to initial classification. SAM was specifically developed for analysis of genome-wide expression data (135). Briefly, SAM uses the standard deviation of repeated gene expression measurements to assign a score to each gene. It estimates a false discovery rate (FDR, the proportion of genes likely to be identified as significant genes by chance) by permutation of the data for a particular score. SAM analysis ascertains that genes identified as differentially expressed do not arise from a random fluctuation of the large quantity of data generated (135). FDR is used to evaluate the statistical significance of the gene set identified by SAM. It is defined as the expected proportion of falsely identified genes (V) among the total number of identified significant genes (R) (FDR=V/R, V and R are calculated by the program). In this study, the FDR=0.00000, which means FDR< 10^{-5}.

SVM (Support Vectors Machine): SVM is a supervised learning method. The first step for SVM is to establish the training set. For example, samples can be divided into two separate classes, such as under two different treatment methods, two different prognosis statuses, these samples are used to form a training set in which any sample can be labeled positively if it's a member of a certain class or negatively if it's a non-member of that class but a member of the other class. Using the training set, the SVM is capable of learning to determine if a sample is a member or non-member of a given class. By complicated mathematic computations, SVM records the expression features of the class, which it learned from training set. These features can be used to determine if new

samples (test sets) are members or non-members of a given class according to the expression pattern of the samples in test set (136, 137).

LOOCV (leave one out cross validation): in this study, we used the LOOCV feature of SVM to validate our results (128, 130). LOOCV analysis employs SVM to perform multiple permutations. At every permutation, one of the samples is removed and the remaining samples are used as the training set to determine the classification of the removed sample. After the LOOCV process, the final classification of each sample is determined by the rest of the samples. LOOCV is a very useful statistial tool for validating the classification results when there is no independent test set available for validation (102, 109, 138).

EASE (Expression Analysis Systemic Explorer): The Expression Analysis Systemic Explorer (EASE version 2.0) is used to search for significant biological categories within specified gene lists. EASE analysis will report a group of over-represented biological categories that are represented in the gene list (test list, such as differentially expressed genes). A statistic reports the probability that the prevalence of the particular biological categories within the test gene list is due to chance alone given the prevalence of the biological categories in the population of genes under study (all "genes" loaded into software for analysis) (139-141). For detailed description, please refer to the following chapters. EASE analysis provides a powerful tool to identify the over-represented biological categories in a given gene list, with the information of significant biological categories within the genes, we can advance our investigation from transcripts level measurement to biological functional studies. This may shed light on

identifying new biological pathways or genes that may be important for pathogenesis or status of a given disease.

CHAPTER III

IDENTIFICATION OF GENE EXPRESSION SIGNATURES IN AUTOIMMUNE DISEASE WITHOUT THE INFLUENCE OF FAMILIAL RESEMBLANCE

Abstract

Even though autoimmune diseases are heterogeneous, believed to result from the interaction between genetic and environmental components, patients with these disorders exhibit reproducible patterns of gene expression in their peripheral blood mononuclear cells. A portion of this gene expression profile is a property of familial resemblance rather than autoimmune disease. Here we wanted to identify the portion of this gene expression profile that is independent of familial resemblance and determine if it is a product of disease duration, disease onset, or other factors. By employing supervised clustering algorithms, we identified 100 genes whose expression profiles are shared in individuals with various autoimmune diseases but are not shared by unaffected family members of individuals with autoimmune disease or by controls. Individuals with early disease (1 yr after onset) and established disease (10 yr after onset) exhibit a near identical expression pattern suggesting that this unique profile is a product of disease onset rather than disease duration.

Introduction

Autoimmune diseases are heterogeneous diseases believed to arise from immune-mediated attack against self-antigens. Both genetic and environmental factors play important roles in their onset and pathogenesis (1, 142, 143). Epidemiologic data along

with genetic linkage studies clearly support the presence of a genetic contribution to susceptibility to autoimmune disease (2, 37, 40, 41, 144-148). Linkage studies have demonstrated the presence of susceptibility loci that are shared among multiple autoimmune diseases (2, 41, 144) and those that are unique for a given autoimmune disease (40, 145, 146).

Specific gene expression profiles have also been found in the peripheral blood mononuclear cells (PBMC) of individuals with autoimmune diseases (79-81, 83, 87-91, 149, 150). Some of these, for example, an "IFN-signature" has been found in systemic lupus erythematosus (SLE) that is a function of disease severity (87) and an early disease signature has been found in rheumatoid arthritis (RA) (80), seem to be unique to a given type of autoimmune disease. In addition, we have described a gene expression signature that is shared among several autoimmune diseases, including RA, SLE, type 1 diabetes (IDDM), and multiple sclerosis (MS) (79). Thus, gene expression signatures that are both common to several autoimmune diseases and unique to a given property of an individual autoimmune disease exist.

Studies on the genetics of gene expression have been performed in model systems and demonstrate that a portion of the gene expression profile or transcript level in a given tissue is an inherited trait (115-117). Thus, transcript levels are heritable phenotypes and this has been demonstrated under a variety of conditions across a range of species (118). A power of this approach is that linkage analysis can identify loci that control variation in transcript levels. Knowledge of the genes that encode these transcripts will undoubtedly stimulate identification of "candidate genes" that reside in loci identified by linkage analysis. Along these lines, we have asked if unaffected first-degree relatives of

individuals with autoimmune disease share gene expression profiles observed in individuals with autoimmune disease. We find that these two groups share an overlapping gene expression profile (120). The genes that make up this overlapping profile are among the same genes that are differentially expressed in a number of different autoimmune diseases. This argues that a portion of the gene expression profile observed in autoimmune disease is a family trait that may arise through genetic mechanisms.

Therefore, we wanted to determine if we could identify exclusive gene expression signatures shared by individuals with autoimmune disease but not by unaffected family members (refer to unaffected family members of individuals with autoimmune diseases, hereafter) or by control individuals. We also wanted to determine if this gene expression signature was a function of onset of autoimmune disease or duration of autoimmune disease.

Materials and methods

Patient populations

This study consists of the following groups as described in the text:

(1) Eight healthy control individuals without active infection or family history of autoimmunity.

(2) Fifty-four individuals with autoimmune diseases: SLE (systemic lupus erythematosus, n=19), RA (established rheumatoid arthritis, average disease duration of 10.5 ± 2.6 yr, n=9), ERA (early rheumatoid arthritis, average disease

duration of 1.1± 0.3 yr, n=17), IDDM (insulin dependent diabetes mellitus, n=5), and MS (multiple sclerosis, n=4).

(3) Eight unaffected family members of individuals with autoimmune diseases (SLE and RA).

All autoimmune patients satisfied established criteria for diagnosis of their respective diseases. Human subject studies were approved by the committee for the protection of human subjects of the Vanderbilt University Institutional Review Board.

Sample preparation and microarray procedures

Analysis procedures presented here comply with MIAME (minimal information about a microarray experiment) guidelines established by the Microarray Gene Expression Data Society (www.mged.org). PBMC were isolated from 20 ml heparinized blood on a Ficoll-Hypaque gradient. All samples were processed within one hour of blood collection. Total RNA was isolated with Tri-Reagent (Molecular Research Center. Inc., Cincinnati OH), 5 μg RNA was reverse-transcribed by reverse transcriptase (Superscript II, Invitrogen Corporation, Carlsbad, CA) in the presence of ^{33}P-dCTP. Labeled probes were purified using a Bio-Spin 6 Chromatography Column (Bio-Rad Laboratories, Inc., Hercules, CA). Before hybridization, GeneFilters membranes (GF-211,4133 human cDNA clones were printed on this membrane, Research Genetics/Invitrogen Corporation, Carlsbad, CA) were washed in boiled 0.5% SDS, saturated with 5.0 ml Microhyb solution (HYB125.GF, Research Genetics/Invitrogen Corporation, Carlsbad, CA). Filters were treated with pre-hybridization reagents (5.0 μg Human Cot-1 DNA and 5.0 μg Poly dA, Invitrogen Corporation, Carlsbad, CA) in a

hybridization roller tube (Midwest Scientific, St. Louis MO) for 2 hours at 42 ºC. Purified, labeled probes were denatured and added to roller bottles containing filters and pre-hybridization solution. GeneFilters membranes were hybridized overnight at 42 ºC. After hybridization, membranes were washed (1st and 2nd wash: 2X SSC, 1% SDS at 50 ºC for 20 minutes; 3rd wash: 0.5X SSC, 1% SDS at 55 ºC for 15 minutes). After washing, membranes were exposed to imaging screens for 24 hr and the screens were scanned by a phosphorimager (Molecular Dynamics/Amersham Biosciences, Piscataway NJ). Acquired images by the phosphorimager were loaded into Pathways 4.0 software (Research Genetics/Invitrogen Corporation, Carlsbad, CA). The relative intensity of each spot on the membrane was determined and the microarray dataset was subjected to further analysis using the different analytical platforms. Microarray data have been deposited into GEO database, accession number: GSE3447.

Data Analysis

Raw microarray data were imported into Cluster (version 3.0) (123) for pre-processing. For the analysis, we filtered the data by selecting those genes whose standard deviation of expression levels across all samples exceeded 1.0. These genes were used for normalization and analysis. Filtered data were normalized and loaded into the TIGR MultiExperiment Viewer (MEV version 3.1) (130) from the Institute of Genomic Research (Rockville, MD, USA). The following data analysis modules of MEV were used for the analysis: HCL (hierarchical clustering), SAM (significant analysis of microarray), ST (support tree clustering), PCA (principal components analysis) and GDM (gene distance matrix).

SAM (significant analysis of microarray): Microarray data were statistically analyzed using the SAM algorithm, which was specifically developed for analysis of genome-wide expression data (135). Briefly, SAM uses the standard deviation of repeated gene expression measurements to assign a score to each gene. It estimates a FDR by permutation of the data for a particular score. SAM analysis ascertains that genes identified as differentially expressed do not arise from a random fluctuation of the large quantity of data generated (135).

FDR was used to evaluate the statistical significance of the gene set identified by SAM. It is defined as the expected proportion of falsely identified genes (V) among the total number of identified significant genes (R) (FDR=V/R, V and R are calculated by the program). In this study, the FDR=0.00000.

ST (support tree clustering): ST was used to identify hierarchical trees and to show the statistical support for the nodes of the trees, based upon Jackknife re-sampling of the data.

PCA (principal components analysis): PCA is used to identify the key variables (or combination of variables), principal components, which represent a multi-dimensional data set (for example, a number of gene expression variables in a number of samples). The three most representative principal components are used to map each element to a three dimensional viewer (134).

GDM (gene distance matrix): We used the GDM to calculate the distance metric (Euclidean Distance) between two samples to provide an intuitive and comprehensive view of the distance (or similarity) between any two samples. These distances are

depicted as a colored matrix representing all sample-to-sample distances. The intensity of the color represents the distance or similarity between the two samples.

Detailed descriptions of the applications of these modules are provided in the results section and are also available in the TIGR MEV manual (130).

Biological process categorization by gene ontology

EASE: the Expression Analysis Systemic Explorer (EASE version 2.0) was used to search for common biological themes within gene lists generated by our microarray analysis. EASE assigns identified genes to "GO: Biological Process" categories of the Gene Ontology Consortium (www.geneontology.org) (139-141) and categories are tested statistically (EASE: Fisher's Exact Test) to identify over-represented categories of identified genes within the biological process system. Significant functional categories are those with the number of list hits (LH) of at least 2 with a P value < 0.05 (EASE, Fisher's Exact Test with Bonferroni correction).

Statistical analysis

Euclidean Distance: In our study, Euclidean distance was used to build the hierarchical clustering (unsupervised, SAM supervised, and support tree). It was also used to calculate the distance of groups in the gene distance matrix (GDM) (130). Euclidean distance was used to calculate the distance metric that reflects the distance between two objects in space. Euclidean distance extends to as many dimensions as present in the expression vectors to be compared. In our case, the expression vectors are represented by the values of expression of an individual gene over the range of samples

(or in GDM analysis of this study, the expression vectors are represented by the values of expression over the range of identified genes for an individual sample). The calculation of Euclidean distance is:

$$d(u,v) = \sqrt{\sum_{i=1}^{m}(u_i - v_i)^2}$$

Where distance, d (u, v) can range from 0 to positive infinity and u_i and v_i represent the two vector values, m represents the number of samples (or in GDM analysis of this study, m represents the number of identified genes) used in this calculation.

Jackknife re-sampling: Jackknifing takes each expression vector and randomly omits a sample. This method produces expression vectors that have one fewer sample and this is often done to minimize the effect of single outlier values (130, 132).

Results

Supervised gene expression profiling permits separation of unaffected family members from individuals with autoimmune diseases

Previously we reported that individuals with different autoimmune diseases share a common gene expression profile and that unaffected first-degree relatives of individuals with autoimmune disease share a portion of this profile (79, 120). Here, we wanted to determine if we could identify a second set of genes whose expression profiles permit separation of individuals with autoimmune disease from unaffected family members. To do so, we increased the number of autoimmune samples from 8 to 54, and used numbers of controls (N=8) and unaffected family members (N=8) similar to our previous analysis. Of the 4133 genes for which we had expression data, 752 genes passed filtering condition

at a standard deviation of 1 using the Cluster software. After filtering, expression data for the 752 genes were normalized and loaded into the TIGR MEV software. We performed a significant analysis of microarray (SAM) study using these 752 genes. For the first step, we set initial classifications: control individuals group, unaffected family members group, and autoimmune individuals group. Since the false discovery rate (FDR, the proportion of genes likely to be identified by chance) was set to 0.00000 (FDR<10^{-5}), statistically, the number of genes identified by chance is less than 1. The expression levels of the SAM identified 100 genes separated the samples into different groups according to their initial classification. We applied hierarchical clustering algorithm to the gene expression profile of these 100 genes (Fig. 3-1).

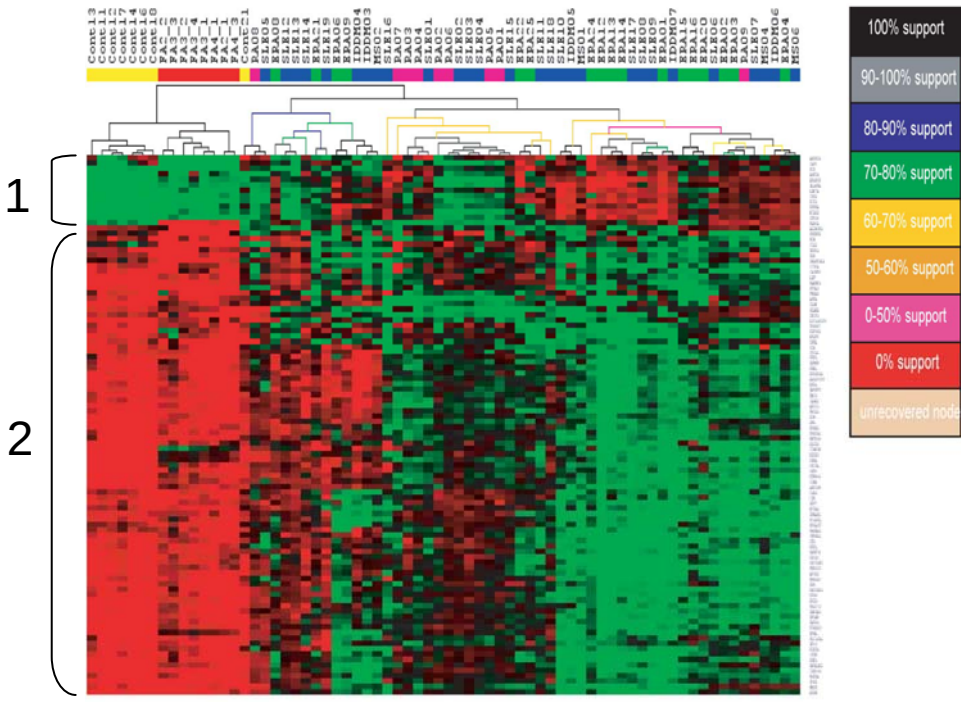


Figure 3-1. Hierarchical clustering analysis of expression data using the SAM identified 100 genes with support tree validation. Jackknife re-sampling was used with 1000 times to establish the support levels for the separate nodes or branches. Color codes correspond to a given level of support for a given node. 1, 2 refer to different variation patterns of genes across the samples.

The unaffected family members were grouped into a separate branch from the individuals with autoimmune disease, but within the same branch as the controls. The only exception was that control 21 clustered with two autoimmune patients (RA08 and SLE05), and this small cluster was grouped together with the autoimmune disease group. Two major patterns of gene expression were seen. One pattern contained genes that were highly expressed in controls and unaffected family members and weakly expressed in individuals with autoimmune disease (Fig. 3-1, pattern 2). The second pattern contained genes that exhibited the opposite expression pattern, weak expression in controls and unaffected family members and high expression in individuals autoimmune disease (Fig. 2-1, pattern 1).

Support tree analysis of the reliability of hierarchical clustering

By increasing the sample pool size and with the assistance of supervised SAM analysis, we successfully separated the unaffected family members from individuals with autoimmune disease. In order to test the reliability of the hierarchical clustering results we obtained after SAM analysis, we performed Jackknife re-sampling 1000 times. For each re-sampling process, a hierarchical clustering was performed and the result was compared to the original clustering result. The percentage of the original clustering results that occurred during the 1000 re-samplings indicates the level of reliability or support of the clustering result (Fig. 3-1). The color of the nodes denotes the support level of the clustering, i.e. the frequency this clustering node appeared among the 1000 re-sampling processes. By performing support tree analysis, we found that the branch containing the unaffected family members and control individuals was separated from the

branch of individuals with autoimmune disease and achieved greater than 90% support by the Jackknife re-sampling test. Therefore, we concluded that the separation between unaffected family members, control individuals and individuals with autoimmune diseases was very reliable.

Principal components analysis (PCA)

From the above analysis, the 100 genes identified by SAM analysis permitted reliable separation among control individuals, unaffected family members, and individuals with autoimmune diseases. In order to gain an intuitive overview of how these three groups were distributed in three-dimensional space according to expression profiles of these 100 genes, we performed principal components analysis (PCA). PCA projected these three groups of individual samples (individuals with autoimmune diseases, healthy controls and unaffected family members) into a three-dimensional space and their positions in this space were determined by their gene expression profile.

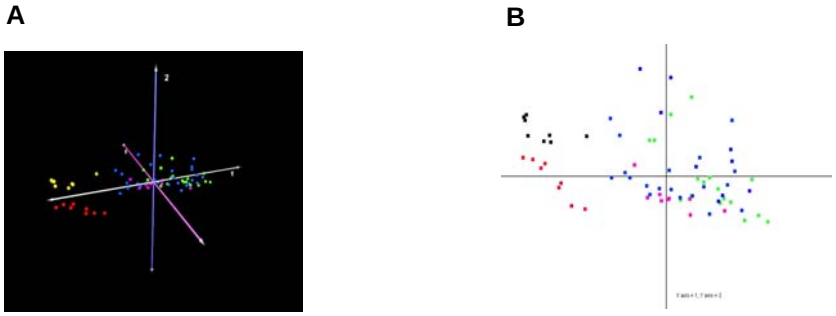


Figure 3-2. PCA analysis of distribution of samples according to expression levels of SAM identified 100 genes. (A) Projection of samples in three-dimensional space: yellow spheres represent control individuals, red spheres represent unaffected family members and other color spheres represent individuals with autoimmune disease. (B) Two-dimensional projection of samples onto an X-Y plane according to expression levels of SAM identified 100 genes. Black squares represent control individuals, red squares represent unaffected family members and other color squares represent individuals with different autoimmune diseases.

PCA clearly separated the samples into three distinct groups in space: unaffected family members (red spheres), healthy controls (yellow spheres) and individuals with autoimmune disease (other color spheres) (Fig. 3-2 A, 3-D space, B, 2-D space). This analysis demonstrated two points. First, as individual samples, the control samples were more like other control samples than the non-control samples, the unaffected family member samples were more like other unaffected family samples than the other samples, and the autoimmune samples were more like other autoimmune samples than the non-autoimmune samples. Second, as a group, the unaffected family member samples were more like the control samples than they were like the autoimmune samples. This result essentially confirmed our hierarchical clustering results and demonstrated that the separation pattern may be determined by the similarity of the different groups.

Gene Distance Matrix

In order to further examine the similarity of these three groups-controls, unaffected family members and individuals with autoimmune disease, we examined the similarity among the three groups by calculating the distance metric (Euclidean Distance, see statistical analysis) between the samples from each of two different groups (unaffected family members versus controls, unaffected family members versus individuals with autoimmune diseases, and controls versus individuals with autoimmune diseases). The result is shown in the Gene Distance Matrix (GDM) (Fig.3-3). Each square element within the matrix is rendered a color that represents the distance or similarity between the two samples associated with the element. Color similarity indicates that two samples have a high degree of actual similarity. Thus, samples with

dark blue colors indicate that the corresponding two samples have a higher degree of similarity than the samples denoted by pale blue colors. These data indicate that the expression profiles of the 100 genes identified by SAM analysis have a higher overall similarity between the unaffected family members and the controls than that between the unaffected family members and the individuals with autoimmune diseases or the controls and the individuals with autoimmune diseases.

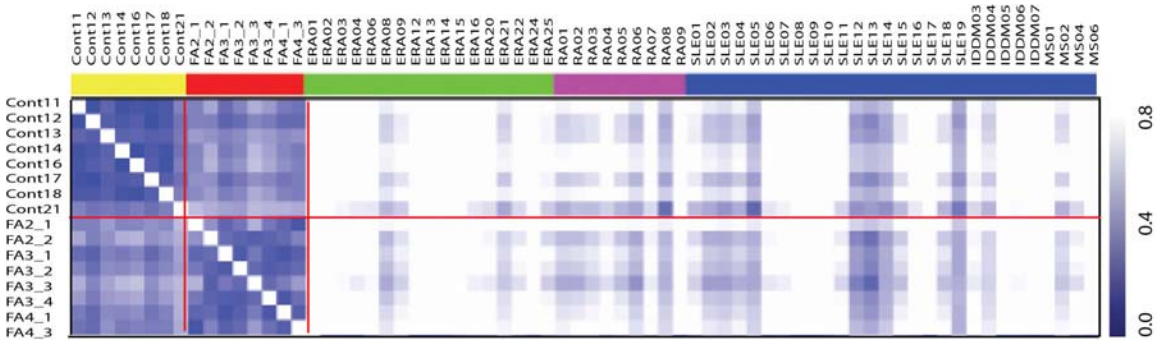


Figure 3-3. Gene Distance Matrix (GDM). The similarity among the three groups was measured by calculating the distance metric between the samples from each of two different groups (unaffected family members versus controls, unaffected family members versus individuals with autoimmune disease, and controls versus individuals with autoimmune disease). Each square element within the matrix is rendered a color that represents the distance or similarity between the two samples associated with the element.

This result is consistent with the PCA analysis (Fig. 3-2) and can be considered as the underlying rationale for the distribution pattern of these samples in the PCA space.

Composition of SAM identified 100 genes

In our previous study, we identified a shared gene expression signature in individuals with different autoimmune diseases (79). We examined the 100 genes identified by SAM. We found that 64% (64/100) of these genes represent previously

identified autoimmune signature genes (79). We also examined the distribution of the autoimmune signature genes in the total 752 genes dataset (110/752) identified by the initial filtering. The chi-squared test revealed that the autoimmune signature genes were significantly over-represented in the set of genes identified by SAM (χ^2 =132.39, P<0.001, Chi-Squared Test).

In order to gain an overview of the potential biological relevance of alterations in expression of the 100 genes identified by SAM, we performed EASE analysis to categorize the 100 genes identified by SAM into their biological ontology (see methods). We considered categories as significant if they contained at least two genes with $P < 0.05$ (Fisher's exact test with Bonferroni correction).

Table 3-1. Over-represented categories in 100 genes. Biological process categories significantly over-represented ($P < 0.05$; Fisher's Exact Test, Bonferroni correction applied) in 100 genes identified by SAM.

Gene category	LH	EASE	Symbols
RNA splicing	3	0.042	*SF3A3; SIP1; SNRP70*
DNA damage response	2	0.042	*BRCA1; TP53*
amino acid catabolism	3	0.042	*ASL; ASPA; BDH*
regulation of CDK activity	2	0.042	*CDKN1B; CKS2*
response to toxin	2	0.042	*BPHL; EPHX2*

LH: list hits, the number of identified genes within the specific category. EASE: *P*-value, Fisher's Exact Test, Bonferroni correction applied.

The over-represented categories of the 100 genes identified by SAM included "RNA splicing", "DNA damage response genes", "amino acid catabolism" "regulation of CDK

activity", and "response to toxin" (Table 3-1). These results are consistent with the idea that lymphocytes from individuals with autoimmune disease exhibit defects in responses to DNA damage and cell cycle control, as we have shown previously, as well as defects in other processes (79, 151)

Quantitative differences in gene transcript levels among the three groups

From the above analysis we found that the 100 genes identified by SAM were over-represented by autoimmune signature genes and that "DNA damage response genes" and "regulation of CDK activity" were over-represented biological categories. Next we wanted to examine the quantitative variation in expression levels of the 100 genes identified by SAM in the different groups. We divided the 100 genes into two groups: "autoimmune signature" genes and non-"autoimmune signature" genes according to our previous studies (79).

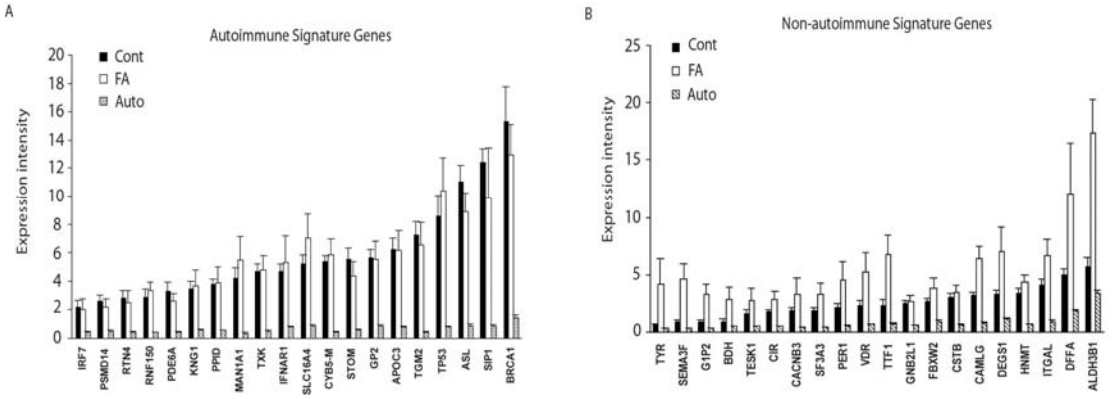


Figure 3-4. Quantitative differences in gene transcript levels among the three groups. **(A)** Average expression levels (mean and standard error) of representative "autoimmune signature" genes within in SAM identified 100 genes among three groups are shown. (Cont vs Auto and FA vs Auto: $P<0.001$, Student's T test) **(B)** Average expression levels (mean and standard error) of representative "non-autoimmune signature" genes in SAM identified 100 genes among three groups are shown (Cont vs Auto and FA vs Auto: $P<0.001$, Student's T test).

Comparison of expression levels of representative "autoimmune signature" genes demonstrated that these genes were uniformly expressed at similar levels in controls and unaffected family members and were expressed at considerably lower levels in autoimmune individuals (Fig. 3-4 A). The non-"autoimmune signature" genes displayed a more heterogeneous expression pattern (Fig. 3-4 B). However, in most cases, expression levels of individual genes were higher in unaffected family members than in the controls or the autoimmune samples.

Expression levels of "autoimmune signature" genes in individuals with early RA (ERA)

We have previously analyzed two populations on patients with RA, one population with an average disease duration of 10.5 ± 2.6 yr and a second population with an average disease duration of 1.1± 0.3 yr (80). Therefore, we compared the gene expression profiles between early RA and established RA to determine if each group exhibited the same expression patterns of the 100 genes identified by SAM. We found that these genes were highly expressed in controls and unaffected family members but weakly expressed in both ERA and RA samples with similar expression patterns seen in the two RA groups (Fig. 3-5).

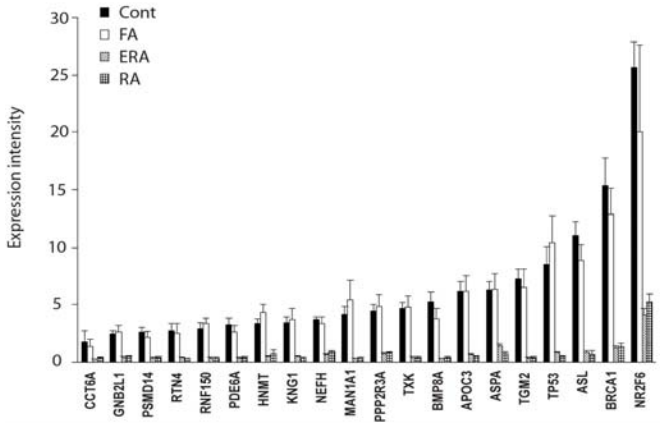


Figure 3-5. Comparison of gene expression profiles between early RA and established RA samples. Expression levels (mean and standard error) of representative SAM identified genes were compared in individuals with early RA, individuals with established RA, unaffected family members and controls.

We conclude from this comparison that expression levels of these genes change early after onset of rheumatoid arthritis and remain relatively constant thereafter. We presume that gene expression patterns in the other autoimmune diseases follow a similar pattern.

Validation analysis

In order to validate our results, we arbitrarily divided our samples into two sets, a training set and a test set; each set contained 4 controls, 4 unaffected family members and 27 individuals with autoimmune disease. We applied the same analysis described above to the training set: of the 4133 genes for which we had expression data for the training set, 722 passed filtering conditions at a standard deviation of 1. We performed SAM using these 722 genes, as FDR equals to 0.00000, SAM identified 99 genes whose expression data accurately separated the three groups (control, unaffected family and autoimmune disease) with > 90% support by support tree hierarchical clustering analysis (Fig. 3-6 A). Next, we determined if the gene expression profile of these 99 genes could be used to

discriminate the three groups in the test set. To do so, we applied support tree hierarchical clustering to the test set using the expression levels of the 99 genes identified by the training set analysis (Fig. 3-6 B). The autoimmune disease group was separated from the control and unaffected family member groups with > 90% support, except control 21, which clustered with RA08 in the autoimmune groups. Under the node shared by the control and unaffected family groups, the control group was clustered together and separated from the unaffected family member group. Next we used the gene expression profile of the 99 genes to perform support tree hierarchical clustering with the total sample pool (training set and test set). Under these conditions, individuals with autoimmune disease clustered together with more than 90% support and samples from the

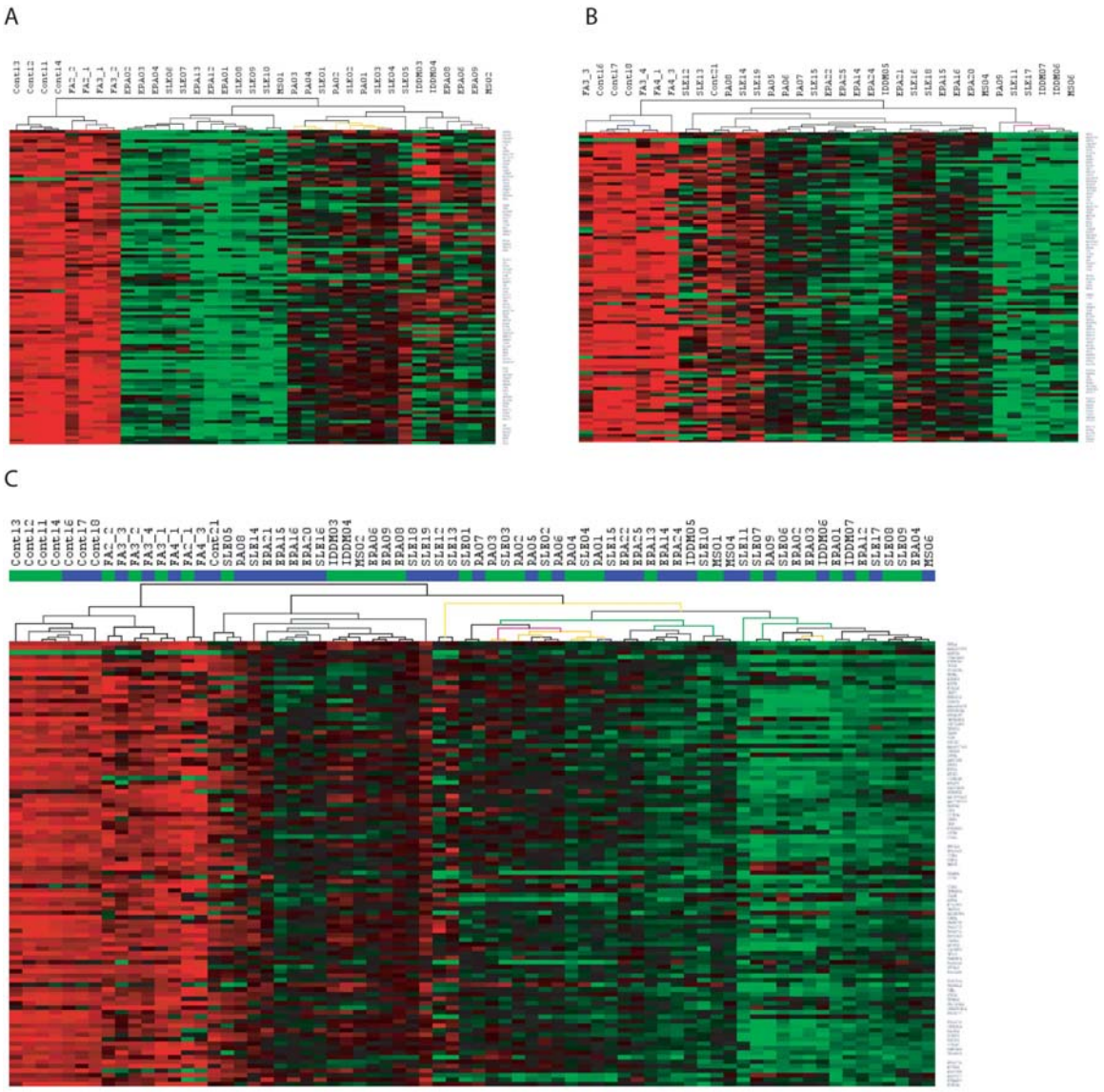


Figure 3-6. Validation analysis. **(A).** Hierarchical clustering with support tree validation of SAM identified 99 genes from training set. **(B)** Hierarchical clustering with support tree validation of SAM identified 99 genes' in the test set. **(C).** Hierarchical clustering with support tree validation of SAM identified 99 genes' in both training and test sets. The green bar indicates samples from the training set and the blue bar indicates samples from the test set. Jackknife re-sampling (1000 times) was used to establish support levels for the separate nodes or branches. Color codes correspond to a given level of support for a given node as in Fig. 2-1.

training and test sets were distributed with each other under the autoimmune disease group node (Fig. 3-6 C). Similarly, the control group was clustered together and separated from the unaffected family member group, except for control 21, which was clustered with SLE05 and RA08 in the autoimmune group. Within each group, samples from training and test sets were intermingled with each other. Strikingly, 80 of the 100 genes identified by SAM using the entire dataset were also identified from the training set only. EASE analysis of the 99 genes identified by SAM in the training set revealed the same over-represented functional categories ($P<0.05$, Fisher's Exact Test, Bonferroni correction applied) as identified from the previous 100 genes (Table 3-1). These results indicate that the gene expression signature identified by these methods will discriminate individuals with autoimmune disease from unaffected individuals, both family members and non-family members of affected individuals.

Discussion

Previously, we characterized a single common gene expression profile present in the PBMC of individuals with four different autoimmune diseases, RA, SLE, MS and IDDM (79). A portion of this gene expression profile is also present in unaffected first-degree relatives of individuals with autoimmune disease (120). Our interpretation of these results is that a portion of these variations in gene transcript levels is associated with familial resemblance rather than clinical manifestation of autoimmune disease suggesting that it may be of genetic origin. In this report, we tried to identify a set of genes whose expression profile exclusively reflected the presence of autoimmune disease without the influence of familial resemblance. To do this, we pre-classified control individuals (N=8),

unaffected family members (N=8) and individuals with autoimmune disease (N=54) as three independent groups. We reasoned that increasing the sample size of the autoimmune class might help discriminate between the autoimmune class and the control and unaffected family member classes. With the assistance of supervised SAM, we successfully identified 100 genes whose expression profiles can discriminate individuals with autoimmune disease from unaffected family members and controls using unsupervised hierarchical clustering algorithms. By validation analysis, we confirmed that the gene expression signatures identified by this method could accurately discriminate individuals with autoimmune disease from unaffected individuals who are either family members or non-family members of affected individuals.

PCA is another measure that provides an overview of gene expression data. PCA essentially confirms the hierarchical clustering results. This analysis segregates samples into three groups: control individuals, unaffected family members and individuals with autoimmune disease, based upon their distribution in three- or two-dimensional space. Samples from control individuals and unaffected family members are also grouped in closer proximity to one another than are individuals with autoimmune disease and unaffected family members or controls. The similarity among different groups measured in GDM also demonstrates the similarity of gene expression profiles among different groups determined their distribution in the PCA three-dimensional space.

The 100 genes identified by SAM consist largely of previously identified autoimmune signature genes (79) and their expression levels are much lower in individuals with autoimmune disease than in control individuals or unaffected family members. EASE analysis reveals that one of the most over-represented biological

process categories in the 100 genes identified by SAM is the "DNA damage response" category that includes both *TP53 and BRCA1* genes (*TP53* is the gene name of p53). The p53 protein, which is also under-expressed in RA lymphocytes, is a central mediator of cellular responses to stress able to induce either cell cycle arrest or apoptosis depending upon the degree of stress or damage (152, 153). Defects in lymphocyte apoptosis may contribute to development of autoimmunity. For example, the MRL murine strain has a mutation in *Fas* and exhibits defects in apoptosis and develops an autoimmune-like syndrome (34). Studies in murine models of collagen-induced arthritis demonstrate that loss of p53 function contributes to more severe lymphocytic infiltration into tissues and more severe joints destruction (154). In addition, both viral and cellular factors can interfere with p53 expression and function (155), and the consequences may affect the normal expression of p53 as well as p53 effector genes- *Gadd45a* or *p21*, and this may also contribute to the development of autoimmunity (156, 157).

Interestingly, there was a report about T lymphocytes *BRCA1* knockout mice (158). In this conditional *BRCA1* knockout mouse, the peripheral T cell number greatly decreased due to T cell apoptosis, the author reasoned that Brca1 was responsible for maintaining the stability of genome, without Brca1, the cell is easy to lose its genome stability due to environmental stress and is susceptible to apoptosis. However, when this mouse was crossed to *TP53* knockout mouse, the double knockout mouse (*BRCA1* knockout and *TP53* knockout) almost fully restored its peripheral T cell number (158). From this results, also considering *BRCA1* and *TP53* are both substantially underexpressed in control group than that of autoimmune group (more than 15 fold), we have the hypothesis: T cells have increased population undergoing apoptosis in

autoimmune individuals compared to those of control individuals, the increased apoptosis is due to substantially underexpressed Brca1 in autoimmune individuals; thus, the peripheral T cell number would be decreased due to increased apoptosis, however, p53 was also substantially underexpressed, in parallel as double knockout mouse model(158), the peripheral T cell number could be partially restored; therefore, the autoimmune individual may experience chronic lymphopenia; finally the lymphopenia driven T cells homeostatic proliferation may finally result autoimmunity under autoimmune individual genetic setting or effects of other factors (17, 159, 160). This may suggest that lymphopenia can be relevant contributor to establish autoimmunity. It's not uncommon that some autoimmune diseases accompany with lymphopenia (161), our results may give one possible explanation for lymphopenia existing in autoimmune individuals, however, it needs further investigation.

In this analysis, we have RA patients with early disease and established disease. Expression levels of the 100 genes identified by SAM are highly under-expressed in both groups compared to that of the control and unaffected family member groups and exhibit similar expression patterns in the two RA groups. We conclude from this comparison that changes in gene transcript levels must occur at the time of disease onset or very shortly after disease onset. Although we do not know the precise mechanism that causes these changes in gene expression after disease onset, it may arise from cell intrinsic or extrinsic mechanisms. If one considers how this may occur in T lymphocytes, intrinsic changes may arise from alterations in lymphoid progenitor cells, selection processes during development in the thymus, or perhaps establishment of a chronic infection in progenitor or peripheral T cells. Extrinsic mechanisms may result from alterations in the

host environment such as changes in the cytokine or chemokine milieu that lymphocytes face in the periphery or changes in other cell types that interact with lymphocytes. We believe that this will be a fruitful avenue of future investigation.

Gene expression signatures in autoimmune disease have been widely described (79, 81, 83, 87-91, 149, 150). Differential expression of these genes has the potential to affect both onset and pathogenesis of autoimmune disease. However, because of the complex genetic characteristics of autoimmune disease, the high variability of the genetic character of the human population and the contribution of genetics to gene expression (115-118), the differential expression pattern of these genes cannot be only attributed to the presence of autoimmune disease. It may also reflect familial resemblance. Both components should be considered in any study of gene expression profiling in human disease. This may help us more accurately identify changes in gene expression that are strictly associated with disease without masking of the genetic components. However, it must be considered that the gene expression pattern that is strictly associated with disease onset may not arise without the contribution of the gene expression pattern that is governed by genetics or familial resemblance.

CHAPTER IV

HIGHLY CONSERVED GENE EXPRESSION PROFILES IN HUMANS WITH ALLERGIC RHINITIS ALTERED BY IMMUNOTHERAPY

Abstract

Atopic diseases, resulting from hypersensitivity to a wide variety of allergens, affect 10-20% of the population. Immunotherapy is an effective treatment for atopic diseases, but its mechanisms are not fully understood. We studied gene expression profiles in the peripheral blood mononuclear cells and examined whether the individuals with allergic rhinitis have a unique gene expression profile and how the immunotherapy affect the gene expression profiles. A highly conserved gene expression profile exists in atopic subjects, which permitted their accurate segregation from control or autoimmune subjects. A major feature of this profile was the under-expression of a variety of genes that encode proteins required for apoptosis and over-expression of genes that encode proteins critical for stress responses and signal transduction. We also identified 563 genes that can segregate individuals with allergic rhinitis based upon receipt of immunotherapy. This profile can be used to identify individuals with allergic rhinitis and to evaluate responses to immunotherapy. Quantitative endpoints, such as gene expression, may assist clinicians faced with clinical decisions in the diagnosis of patients and the evaluation of response to therapy. The knowledge of the possible genetic basis for immunotherapy efficacy may also lead to novel therapeutic approaches for atopic diseases.

Introduction

Atopic disease represents one of the major categories of immune-mediated disease in the human population (162, 163). Both environmental and genetic factors contribute to disease onset and severity (164, 165). One common type of atopic disease, allergic rhinitis, is a consequence of immediate hypersensitivity to common airborne allergens localized to the upper respiratory tract (73, 166). De-sensitization by immunotherapy ("allergy shot") is a common clinical treatment for atopic diseases(77) and can provide a possible permanent improvement in the clinical disease. However, its precise mechanism is not fully understood.

Gene expression profiling offers a powerful tool to classify human diseases, identify common features of clinically distinct diseases, and contribute to disease diagnoses. Alterations in gene expression profiles in a common cellular or tissue source may also permit more accurate monitoring of changes in disease activity after therapy, may predict patient responses to a given type of treatment, or may predict risks of unwanted side-effects in response to treatments (105, 112, 138, 167). Identification of genes whose expression levels fulfill these goals could have significant impact on the decisions made by clinicians in patient treatment.

Several groups including our own have recently reported initial results of these gene expression studies in human autoimmune diseases (79, 89, 149). These results clearly demonstrate that it is possible to use gene expression profiling to classify human diseases and to identify common features present in clinically distinct autoimmune diseases. Therefore, in this study, using similar approaches, we aimed to determine if we also found conserved gene expression profiles in peripheral blood mononuclear cells

(PBMC) in individuals with allergic rhinitis and if these profiles shared common features with those of human autoimmune disease. We also wanted to determine if and how gene expression profiles in these individuals might change after immunotherapy. Here, we demonstrate that individuals with allergic rhinitis have unique gene expression profiles in their PBMC compared to those who are not allergic or have autoimmune diseases without allergic rhinitis. We also demonstrate that these gene expression profiles change in a characteristic way after immunotherapy. Thus, we provide direct support for the concept that identification of "biomarkers" in human disease is possible and that they can be used to monitor responses to therapy. The identification of which gene expression pathway may underlie atopic diseases and is altered with immunotherapy, an effective clinical treatment, may provide insight into development of novel approaches to disease management.

Materials and methods

Patient populations

Patients with allergic rhinitis (N=8; 22-58 yrs of age) had blood drawn at their first clinic visit before receiving allergy immunotherapy in spring and summer of 2003, none of them had received immunotherapy previously. Four patients (female; 3 Caucasians 1 African American; mean age 30, range 23-34 yrs) had blood collected on the day of RUSH immunotherapy(168, 169) and mean 129 days after RUSH (range 41-188 days) when the maintenance dose was achieved. Concentration of RUSH (w/v) at last sample is 1:100. None of these 4 patients have asthma, all skin tested positive to a broad

range of perennial and seasonal allergens. Allergy extracts used in skin tests and immunotherapy were as follows:

Greer Laboratories™. Mold Mixes: alternaria, aspergillus, helminthosporium, ormodendrum, penicillium, curvularia, fusarium, mucor, pullularia, rhizopus. Weed mixes: cocklebur, lambs quarter, rough pigweed, giant ragweed, short ragweed, sage mix (prairie sage, common sagebrush), dock sorrel mix (yellow dock, sheep sorrel), kochia, English plantain, nettle. Tree Mixes: eastern sycamore, sweet gum, white mulberry, elm mix (American, Chinese), eastern cottonwood, oak mix (black, red, white), sugar maple-box elder mix, birch mix (black red, white), red cedar, pecan pollen-shagbark hickory mix, ash mix (green, white), acacia, mountain cedar, olive. Hollister-Stier (Bayer)™. Dust mites (D. Farinae and D. Pteronyssinus mites), cat hair, dog hair & dander. Cockroach mix (American & German). Grass mixes: Kentucky Bluegrass, Orachrd Grass, Redtop, Timothy, Sweet Vernal, Meadow Fescue, Perennial Rye, Johnson Grass, Bermuda Grass.

All of these 4 patients had nasal (sneezing, running, blocked), ocular (itching, tear flow, redness), ear, headache symptoms prior to immunotherapy and reported symptom improvement 1-3 months after immunotherapy. Symptoms intensity was documented as 0 = no symptoms, 1 = mild, 2 = moderate, and 3 = severe. All these patient's symptoms were improved and medication usages were reduced after immunotherapy. The detailed information of the patients is summarized in table 3-1.

Table 4-1. Patients information

age	gender	disease duration*	skin test+[†]	specific IT[†]	pre-IT ID #	pre-IT score[‡]	post-IT ID #	post-IT score[‡]	interval between microarray (days)
22	M	>10	b	b	001	3	-		
58	F	>10	a	a	002	3	-		
32	M	>10	a	a	004	3	-		
34	F	16	b	b	007	3	-		
23	F	13	a	a	006	3	015	0	167
30	F	>10	a	a	010	3	016	1	41
31	F	>10	b	b	013	3	017	0	119
34	F	15	c	c	009	3	019	0	188

[*]disease duration in years; [†] skin test activity: a= cat, molds, grasses, trees, weeds, b= dust mite, cat, dog, molds, grasses, trees, weeds; [‡] symptoms score. c (RAST, radioallergosorbent testing)= cat, moulds, grass, tree and weed pollen
IT, immunotherapy

As a control population, we employed eight individuals without a clinical diagnosis and no known family history of atopic or autoimmune disease. Our analysis also included samples from our established cohort of autoimmune individuals which has been described before (79). These studies were approved by the Vanderbilt University Institutional Review Board. All individuals in diseases groups and control group gave written informed consent.

Sample preparation and Microarray procedures.

Analysis procedures presented here comply with MIAME (minimal information about a microarray experiment) guidelines established by the Microarray Gene Expression Data Society (www.mged.org). PBMC were isolated from 20 ml heparinized blood on a Ficoll-Hypaque gradient. All samples were processed within one hour of blood collection. Total RNA was isolated with Tri-Reagent (Molecular Research Center.

Inc., Cincinnati OH), 5 µg RNA was reverse-transcribed by reverse transcriptase (Superscript II, Invitrogen Corporation, Carlsbad, CA) in the presence of ^{33}P-dCTP. Labeled probes were purified using a Bio-Spin 6 Chromatography Column (Bio-Rad Laboratories, Inc., Hercules, CA). Before hybridization, GeneFilters membranes (GF-211, Research Genetics/Invitrogen Corporation, Carlsbad, CA) were washed in boiled 0.5% SDS, saturated with 5.0 ml Microhyb solution (HYB125.GF, Research Genetics/Invitrogen Corporation, Carlsbad, CA). Filters were treated with pre-hybridization reagents (5.0 µg Human Cot-1 DNA and 5.0 µg Poly dA, Invitrogen Corporation, Carlsbad, CA) in a hybridization roller tube (Midwest Scientific, St. Louis MO) for 2 hours at 42 ºC. Purified, labeled probes were denatured and added to roller bottles containing filters and pre-hybridization solution. GeneFilters membranes were hybridized overnight at 42 ºC. After hybridization, membranes were washed (1st and 2nd wash: 2X SSC, 1% SDS at 50 ºC for 20 minutes; 3rd wash: 0.5X SSC, 1% SDS at 55 ºC for 15 minutes). After washing, membranes were exposed to imaging screens for 24 hr and the screens were scanned by a phosphorimager (Molecular Dynamics/Amersham Biosciences, Piscataway NJ). Acquired images by the phosphorimager were loaded into Pathways 4.0 software (Research Genetics/Invitrogen Corporation, Carlsbad, CA). The relative intensity of each spot on the membrane was determined and the microarray dataset was subjected to further analysis using the different analytical platforms. Microarray data have been deposited into GEO database, accession number GSE1964 and series 1-5 (GSM35074-GSM35142).

Data analysis

Cluster (version 3.0) (123)and JAVA Treeview(170) software programs were used to identify similarities among individual samples. Datasets were analyzed using hierarchical clustering algorithms. As a separate approach, we treated expression levels of individual genes from the microarray datasets as separate variables. We calculated averages and standard deviations. Because of its familiarity, we used the Student's two-tailed, unpaired t-test to determine statistical significance, P value < 0.05.

Biological process categorization by gene ontology

EASE: the Expression Analysis Systemic Explorer (http://david.niaid.nih.gov/david/ease.htm, EASE version 2.0) was used to search for common biological themes within gene lists generated by our microarray analysis. EASE assigns identified genes to "GO: Biological Process" categories of the Gene Ontology Consortium (www.geneontology.org) (139-141) and categories are tested statistically (EASE: Fisher's Exact Test) to identify over-represented categories of identified genes within the biological process system. Significant functional categories are those with the number of list hits (LH) of at least 2 with a P value (EASE) < 0.05 (Fisher's Exact Test).

Statistical analysis

Data were normalized to yield an average intensity of 1.0 for each clone (4132) represented on the microarray. Reproducibility of the method was established by performing replicate hybridizations to separate microarrays. Linear regression analysis demonstrated that separate hybridizations yielded R^2 values ranging from 0.87 to 0.96.

Different exposure lengths of identical filters also produced high R^2 values (0.99). The Student's t-test was used to determine statistical significance.

Results

Hierarchical clustering of gene expression profiles

We selected individuals with allergic rhinitis and performed gene expression profiling. Application of hierarchical clustering to gene expression profiles permitted accurate segregation of the allergic rhinitis group from control subjects and autoimmune subjects by using the entire microarray dataset that included 4132 genes. We also identified 68 over-expressed genes, and 165 under-expressed genes that exhibited a greater than 3-fold difference in expression between control individuals and individuals with allergic rhinitis (before immunotherapy) ($P < 0.05$, Student's t-test). Using this smaller dataset, we were also able to perfectly segregate the allergic rhinitis group, independent of receipt of immunotherapy, from the other groups (Fig. 4-1 A, B, and C). These results demonstrate that each individual with allergic rhinitis has a common gene expression profile that is distinct from control individuals and individuals with autoimmune diseases. The ratio of expression levels (allergic/control) of genes that were differentially expressed in individuals with allergic rhinitis (before immunotherapy) compared to controls varied from a minimum of -4.64 (log2) to a maximum of 5.36 (log2).

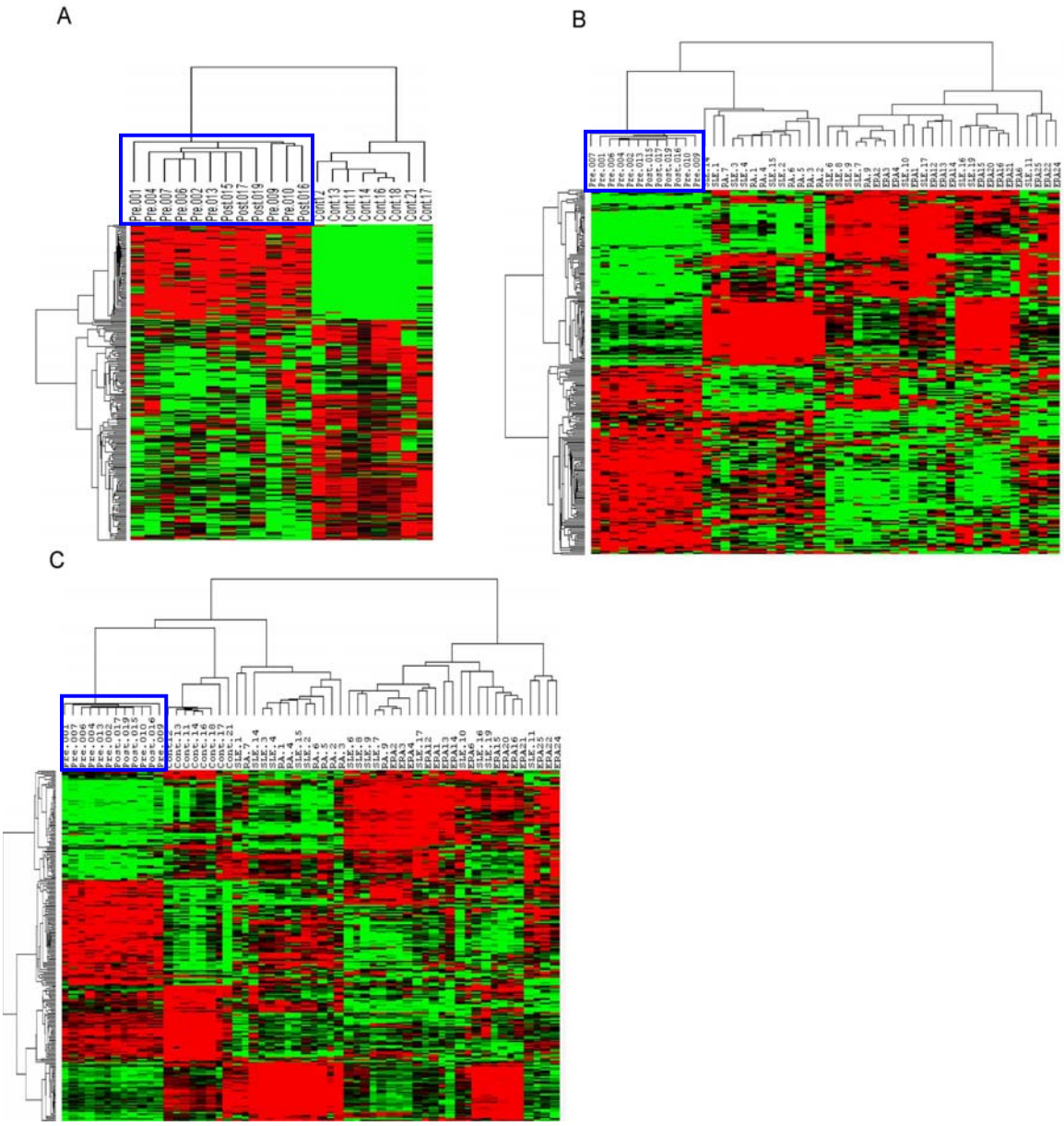


Figure 4-1. Hierarchical clustering of gene expression profiles separates the allergic rhinitis group from other groups-control and autoimmune. Hierarchical clustering of microarray data using subjects with allergic rhinitis before (pre-) and after (post-) receipt of immunotherapy and age-matched controls (Cont) **(A)**, and subjects with allergic rhinitis and autoimmune disease (SLE, RA and ERA (early RA)) **(B),** and subjects with allergic rhinitis, control subjects and subjects with autoimmune disease **(C)**. the allergic rhinitis group is in the blue square.

Given the small sample size of our base of subjects, we also wanted to use standard methods to evaluate if these differences between allergic and control subjects would achieve statistical significance. To do this, we treated the expression level of each gene as an individual variable and used, because of its familiarity, the student's T test to determine statistical significance. We found a number of under-expressed and over-expressed genes that exhibited a high degree of difference between the two subject groups with a high level of statistical significance (Table 4-2). These same genes were also identified by the hierarchical clustering methods in the heat maps in figure 4-1.

Table 4-2. Quantitative gene expression differences between the allergic and control groups. Results are expressed as the average expression level (control, n=8, allergic, n=8) ± standard deviation. P values were determined by the student's T test.

Gene	Control	Allergy	*p*-value
	Under-expressed genes:		
CPNE1	11.5±3.9	0.5±0.2	<0.001
MARCH-VI	9.2±6.0	0.5±0.2	0.005
CASP6	9.4±2.1	0.6±0.2	<0.001
DIPA	7.1±1.8	0.5±0.2	<0.001
CDKN1B	8.9±2.4	0.6±0.2	<0.001
BPHL	6.2±2.0	0.5±0.2	<0.001
SST	8.0±5.8	0.6±0.2	0.009
EFNB1	7.8±3.9	0.6±0.2	0.001
RIOK3	5.2±1.0	0.4±0.2	<0.001
SLC1A5	8.2±4.3	0.7±0.2	0.002
PON3	6.7±4.0	0.6±0.3	0.003
GGTLA1	4.5±2.5	0.4±0.2	0.002
POR	4.1±2.2	0.4±0.2	0.002
PAR5	5.0±2.3	0.5±0.2	0.001
LMAN1	4.2±2.8	0.5±0.2	0.006
RAB7	4.4±2.3	0.5±0.2	0.002
UBE2I	15.8±12.4	1.8±0.8	0.015
GNAO1	4.3±2.0	0.5±0.2	0.001
TAF1C	6.1±2.7	0.7±0.4	0.001
	Over-expressed genes:		
RAB6	0.3±0.1	13.3±11.7	0.016
RPL10	0.5±0.2	4.5±3.0	0.007
APEH	0.3±0.1	2.4±0.7	<0.001
CETN1	0.3±0.1	2.2±1.7	0.014
CD22	0.3±0.1	2.1±0.9	0.001
ESR1	0.3±0.1	1.7±1.4	0.024
CD53	0.3±0.1	1.4±0.6	0.001
MCSP	1.4±0.4	7.8±4.2	0.003
IL8	0.3±0.1	1.8±0.6	<0.001
RPML3	0.3±0.1	1.2±0.6	0.003
PLGL	0.4±0.2	2.0±0.7	0.001
PDGFB	0.2±0.1	0.9±0.4	0.002
RGS10	0.3±0.1	1.4±0.6	0.001
HTR3A	0.3±0.1	1.2±0.5	0.001
PTP4A1	0.3±0.1	1.2±0.6	0.005
PROS1	0.2±0.1	0.8±0.1	<0.001
RPL7A	0.9±0.3	3.5±1.5	0.002
UTX	0.3±0.1	1.2±0.5	0.002
SEMA5A	0.2±0.0	0.6±0.2	<0.001

According to recent discussions(171-177), we searched our microarray dataset for genes of possible relevance to allergic rhinitis, such as cytokines: *IL3, IL4, IL5, IL9, IL10, IL12, IL13, GM-CSF*; cytokine receptors: *IL4R, IL10RA, IL10RB, IL13RA1, IL13RA2*; chemokines and chemokine receptors; *T-bet, GATA3*; all the chemokines and chemokine receptors; immunoglobulin genes: *IgE, IgG, IgA*; cell surface markers: *CD4, CD8* and etc. However, these gene expression levels only exhibit modest differences between allergic rhinitis patients and healthy controls, except *IL8* and *IL15*. (GEO database: GSE1964: GSM35074-GSM35142)

Our criteria for a differentially expressed gene is that this gene exhibited a greater than 3-fold difference in expression between control individuals and individuals with allergic rhinitis (before immunotherapy) with $P < 0.05$ (Student's t-test). Therefore, applying these criteria, the genes I discussed above can not be considered as differentially expressed genes. In fact, in a previous report, Benson et al used microarrays to study gene expression in nasal biopsies from allergic rhinitis patients, they analyzed genes of these cytokines, growth factors and their receptors, but the differences in gene expression between patients and controls were generally modest (96). However, even the gene expression difference between allergic patients and healthy controls is modest; it can still have a significant effect on pathogenesis of allergic rhinitis.

We also searched our dataset to identify the changes of these allergy related genes before and after immunotherapy, only a few of their expressions are significantly changed ($P<0.05$, student's T test) after immunotherapy, such as *CCL16, CXCL13, CCL4, CCL25, CCL7*, but the fold changes are less than 2, and *TGFβ* exhibit a two fold

difference between before and after immunotherapy, however, the P value of the statistical calculation for *TGFβ* between before therapy and after therapy is 0.09.

In our dataset, several genes that have not been previously associated with atopic disease showed markedly different levels of expression between the two groups. These included *MARCH-VI, CASP6, RAB6,* and *MAN2A2*.

Individual variability in differential gene expression

We selected the ten most under- and over-expressed genes in the atopic population compared to non-atopic control individuals for further analysis and grouped control individuals (white bars), atopic individuals before therapy (black bars), and atopic individuals after therapy (hatched bars). We found that each under-expressed gene exhibited a similar expression level in each atopic individual independent of therapy (Fig 4-2 A). *P*-values for each of these differences in gene expression between the allergic and control groups treated as an individual variable (Student's t-test), ranged between 0.01 and 0.0001. Over-expressed genes, as a group, exhibited somewhat greater variability in the atopic population (Fig. 4-2 B).

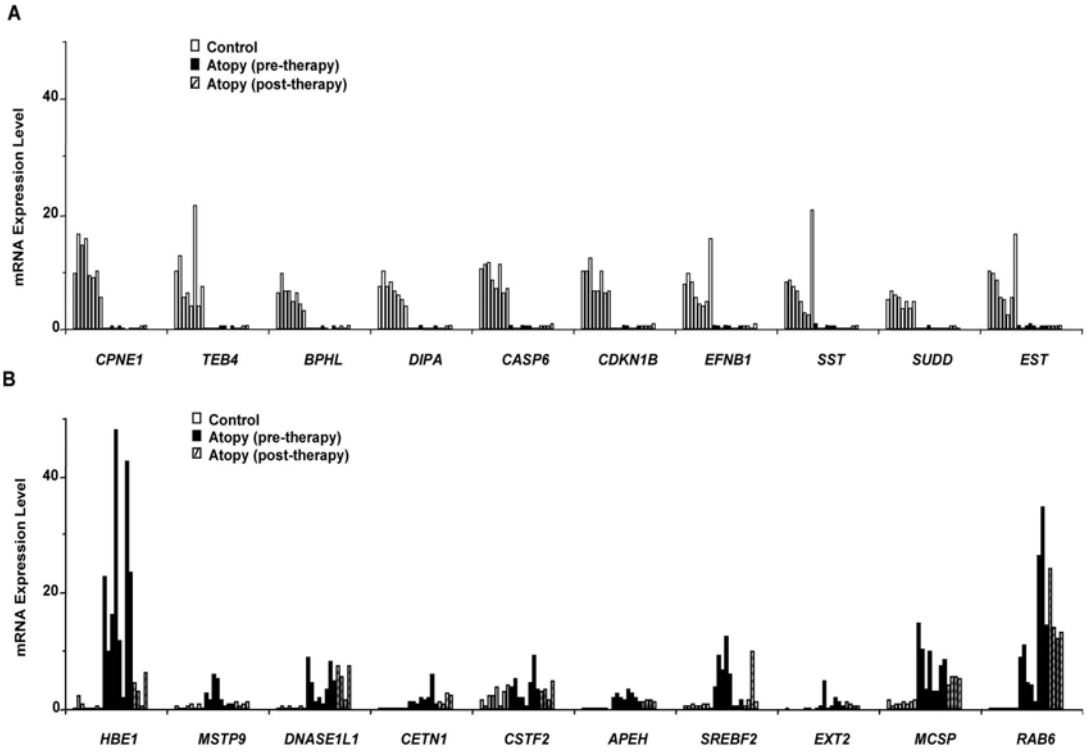


Figure 4-2. Individual variation in gene expression in allergy. Expression levels of individual genes were compared in 8 control individuals (white bars) and 12 allergic rhinitis individuals (8 pre-therapy, black bars, 4 post-therapy, hatched bars). Under-expressed **(A)** and over-expressed **(B)** genes were analyzed.

Some of the variability in the group of over-expressed genes appeared to reflect responses to therapy. For example, expression levels of *HBE1* and *MSTP9* were relatively high in individuals before immunotherapy but were reduced to control levels after immunotherapy.

EASE analysis of differentially expressed genes in individuals with allergic rhinitis

We performed EASE analysis to categorize the over- and under-expressed genes in the allergic rhinitis group. The Gene Ontology Biological Process categories in which over- or under- expressed genes were over-represented by EASE scores (Fisher's Exact Test, $P < 0.05$) are shown in Table 3-3.

Table 4-3. Biological process categories significantly over-represented by differentially expressed genes of individuals with allergic rhinitis

Over-expressed Genes			Under-expressed Genes		
Gene Category	**LH**	**EASE**	**Gene Category**	**LH**	**EASE**
cellular process	31	0.03	induction of apoptosis	5	0.03
cell communication	19	0.02	JNK cascade	3	0.01
signal transduction	15	0.04	activation of JUNK	2	0.01
response to stress	8	0.04	lipoprotein biosynthesis	2	0.01
cell motility	5	0.03	cellular morphogenesis	2	0.02
angiogenesis	3	0.00	lipoprotein metabolism	2	0.02
cell growth	3	0.02	glycerophospholipids	2	0.03
chemotaxis	3	0.03	DNA replication	2	0.04

LH: list hits, the number of identified differentially expressed genes within the specific category. EASE: Fisher's Exact Test.

The most over-represented categories in the over-expressed gene group included inflammatory responses, signal transduction and stress responses. In the under-expressed gene group, we surprisingly found that the most over-represented category was the biological process of apoptosis and activation of the JNK cascade that is also critical for many pathways of apoptosis.

Immunotherapy alters gene expression profiles

Immunotherapy is a common and effective treatment for atopic disease but its precise mode of action is not fully understood. We wanted to determine if immunotherapy affects the gene expression profile observed in the PBMC of subjects with allergic rhinitis. To do so, we performed hierarchical clustering analysis of subjects with allergic rhinitis pre- and post-immunotherapy using different standard deviations to evaluate different numbers of genes in the dataset. We were unable to perfectly discriminate between the pre- and post-immunotherapy groups after filtering the dataset

by using 1, 1.5 or 2 standard deviations (447, 228, and 123 genes, respectively) (data not shown). However, by decreasing the standard deviation of the filtering condition we were able to include more data for analysis and found that expression levels of 563 genes could significantly segregate allergic individuals based upon receipt of immunotherapy (P<0.05, Fisher's Exact Test) (Fig. 4-3 A).

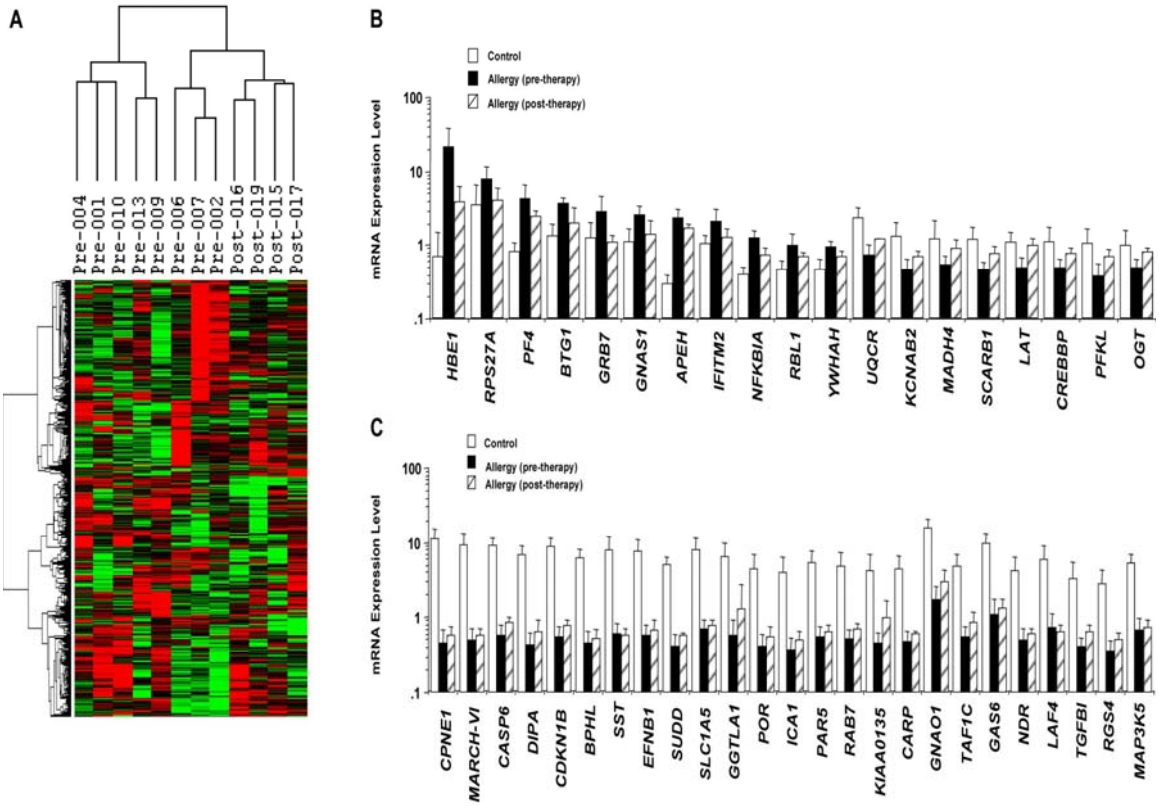


Figure 4-3. Immunotherapy alters gene expression profiles of individuals with allergic rhinitis. **A.** Hierarchical clustering analysis was performed with a group of atopic individuals before (Pre-) or after (Post-) immunotherapy. Expression levels of 563 genes accurately segregated the pre- and post-immunotherapy groups into separate branches. **B.** Genes that change expression levels after immunotherapy are shown. Levels of gene expression for control (white bars) and pre- (black bars) and post-immunotherapy (hatched bars) (normalized average expression ± S.D) are shown, the changes between control and pre-immunotherapy, and changes between pre-immunotherapy and post-immunotherapy are significant (P<0.05, Student's t-test). **C.** As in B, except that genes that do not change expression levels after immunotherapy are shown (changes between control and pre-immunotherapy are significant, P<0.05, and changes between post-immunotherapy and pre-immunotherapy are not statistically significant).

Next, we wanted to determine if immunotherapy resulted in detectable changes in gene expression in PBMC of individuals with allergic rhinitis. We searched our microarray data set to identify changes in expression levels of specific genes in allergic individuals before and after receipt of immunotherapy. We identified a group of genes whose expression levels were high or low in allergic individuals (before immunotherapy, $P < 0.05$) compared to controls that partially or completely returned to control levels after immunotherapy (Fig. 4-3 B, $P < 0.05$). In contrast, genes that were under-expressed in the allergic group before immunotherapy ($P < 0.05$) remained under-expressed after immunotherapy (Fig. 4-3 C, representative examples). We subjected these 563 discriminatory genes to EASE analysis as shown in Table 4-4.

Table 4-4. Biological process categories significantly over-represented in the 563 genes

Gene Category	LH	EASE
Response to external stimulus	77	0.005
Response to biotic stimulus	61	<0.001
Biosynthesis	58	0.004
Defense response	56	0.001
Macromolecule biosynthesis	50	0.002
Immune response	50	0.004
Protein biosynthesis	41	0.000
Cell motibility	26	0.015
Muscle contraction	14	0.009
Antigen presentation	8	<0.001
Ribosome biogenesis and assembly	7	0.002
Cell cycle checkpoint	7	0.004
DNA recombination	7	0.034
Smooth muscle contraction	4	0.029
Centrosome cycle	3	0.001
DNA damage response	3	0.012

LH: list hits, the number of identified genes within the specific category. EASE: Fisher's Exact Test.

The most over-represented category was “response to external stimulus”. In fact, many of the gene ontogeny groups selected by the EASE analysis were groups that are typically considered to participate in the innate or adaptive immune response. These results suggest that immunotherapy may alter the activities of these pathways to alleviate the symptoms of allergy.

Scoring system to classify atopic disease

As we have done with autoimmune disease (79, 178) , we determined if we could devise a scoring system based upon expression levels of small numbers of genes that would classify individuals as allergic and distinguish these individuals from control individuals and those with autoimmune disease. We selected 15 under-expressed genes and 5 over-expressed genes that, based upon our analysis of the data sets, appeared to have the greatest discriminatory power. To give each gene equal weight in the scoring system, we summed the average expression value in the control (N = 8) and allergic group (N=6) for each gene and divided by 2 ((control + allergy)/2). We inspected expression levels of each gene in each sample and assigned it a value of 0 if it was less than the average and 1 if it was greater than the average. The maximum possible score was 15 or 5 for the under- or over-expressed gene categories, respectively, and the minimum possible score was 0 (Fig. 4-4 A).

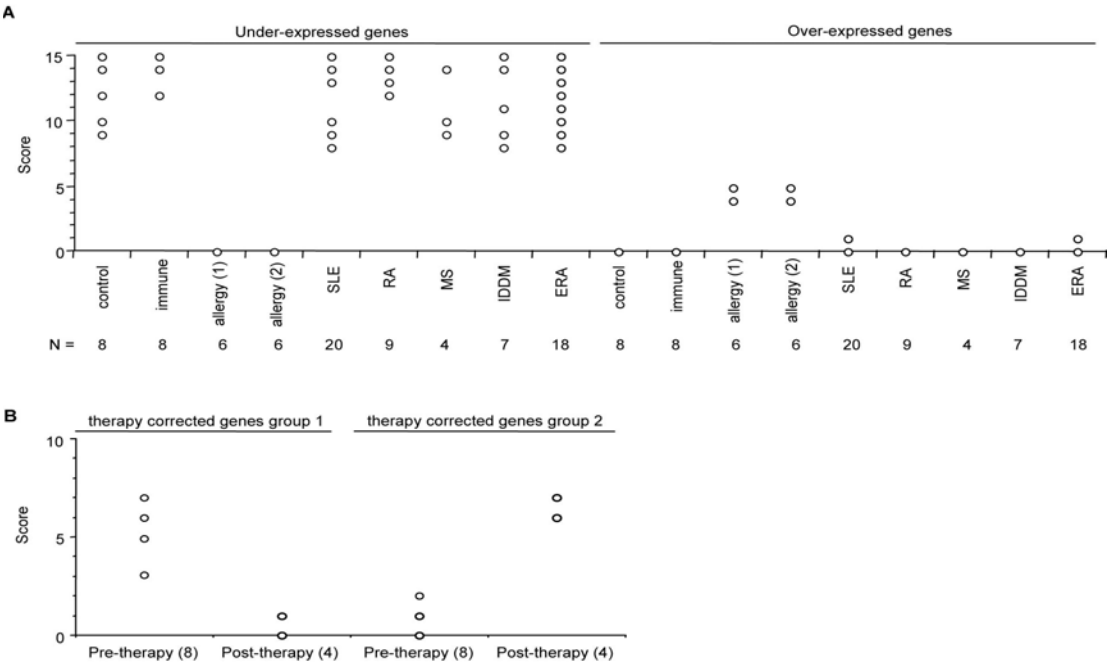


Figure 4-4. Gene expression profiles accurately discriminate allergy from autoimmunity and pre-therapy from post-therapy. **A.** The score (y-axis) is shown for each individual subject analyzed from the different populations (x-axis). We used 15 under-expressed genes (left panel) and 5 over-expressed genes (right panel) to develop the scoring system. Numbers of samples (N) for each category are shown below the graphs. The allergy (1) group was used to derive the scoring system and the allergy (2) group was used to test the scoring system. Under-expressed genes: *MARCH-VI, SST, EFNB1, SLC1A5, GGTLA1, POR, TAF1C, PAR5, RAB7, CARP, MAP3K5, GNAO1, LAF4, LILRB2,* and *TGFBI*, over-expressed genes: *RPML3, CD53, CD22, APEH,* and *CETN1*. **B.** As in A, we used genes whose expression levels have been corrected after immunotherapy to design scoring system. We have "therapy corrected genes group 1": *HBE1, PF4, BTG1, GRB7, APEH, NFKBIA, RBL1,* and *YWHAH*. And "therapy corrected genes group 2": *UQCR, KCNAB2, MADH4, LAT, CREBBP, PFKL* and *OGT*.

For under-expressed genes, the range of scores for non-allergic individuals (control and autoimmune) ranged from 8 to 15. All allergic individuals used to derive the initial scoring system achieved a score of 0 in the test. We also analyzed expression data from 6 additional allergic individuals who were not included to derive the initial scoring system. They also received a score of 0 in the test. Examination of over-expressed genes yielded similar results. Both groups of allergic individuals received scores of 4-5 while all other individuals received scores of 0-1. All these comparisons yielded high levels of

statistical significance ($P \leq 0.0001$). In addition, no individual was incorrectly assigned to or excluded from the allergy population based upon score results. Similarly, we designed scoring systems to discriminate the status based upon receipt of immunotherapy. We employed two groups of genes whose expression levels were high (group 1) or low (group 2) in allergic individuals (before immunotherapy, $P < 0.05$) compared to controls and which partially or completely returned to control levels after immunotherapy. We can clearly discriminate pre- and post-therapy groups: by using group 1 genes, the pre-therapy group received scores of 3-7, the post-therapy group received scores 0-1; by using group 2 genes, the pre-therapy group received scores of 0-2 and the post-therapy group received scores of 6-7 (Fig. 4-4 B)

Discussion

Our results indicate that there is a conserved gene expression profile in individuals with allergic rhinitis. In this study, the small sample size may be expected to limit our conclusions, especially since statistical methods for microarray analysis are oftentimes absent or unfamiliar. Therefore, we also wanted to evaluate statistical significance of the microarray results using more familiar methods. To do so, we also treated each gene as an individual variable and used the student's T test to determine statistical significance. We found that this approach was sufficient to identify a number of genes that were differentially expressed between the control and allergic groups with a high degree of statistical significance, even for such a small sample size. This supports the idea that atopic individuals have a conserved gene expression profile in their PBMC. Using similar approaches, we have also found a conserved gene expression profile in

individuals with autoimmune disease. As with autoimmune disease, we find that expression profiles in all individuals with allergic rhinitis are extremely consistent. There is < 5% overlap between these two gene expression profiles. Taken together, these results demonstrate that the two major immune-mediated diseases in the human population are associated with two unique gene expression profiles in PBMC. It is our hypothesis that these alterations in gene expression result in alterations in cellular functions that confer a liability that an individual will develop atopic or autoimmune disease, respectively.

The source we used for this study is PBMC instead of affected tissues. One of the reasons to use PBMC is easy access. Also, surgically removed affected tissues are generally only in small size that can not generate enough amount of RNA used for microarray study, and the positions to obtain the tissues may generate more variation compared to PBMC, because the changes of positions may result in variation of cell populations which can cause greater variations in microarray study, since there are increases in recruitment of certain inflammatory cells around affected tissues, which can contaminate the source for microarray study (179). In contrast, current protocol can generate enough RNA from PBMC and also the current protocol for isolating the PBMC from the whole blood can greatly decrease the contamination of other source.

In this study, we also designed scoring systems, based upon levels of expression of individual genes, one scoring system can correctly classify individuals as having allergic rhinitis and exclude controls and individuals with autoimmune disease from this category, similarly, and another system can correctly classify individuals based upon receipt of immunotherapy. We have taken a similar approach to the analysis of autoimmune

disease and find that we can also design a scoring system, based upon expression levels of small numbers of genes, which can correctly classify individuals as autoimmune. These scoring systems may aid clinicians by offering analyses that can correctly classify or exclude an individual from the atopic or autoimmune category and can monitor the responses to immunotherapy. Therefore, predictive tests that absolutely classify individuals into these disease categories or monitor their responses to therapy may aid the decision making process by the clinicians.

We have 4 patients who are monitored after immunotherapy, the limited sample size is always a issue needed to be handled in microarray study, generally complicated statistical methods are applied to validate the results(138), in our study, the hierarchical clustering algorithm can significantly separate the post-immunotherapy group from pre-immunotherapy group ($P< 0.05$, Fisher's Exact Test), also the gene expression profile of our scoring systems can discriminate the pre- and post-immunotherapy groups, this also cross validates our results.

We used a new computer program, EASE: the Expression Analysis Systemic Explorer, to search for common biological themes represented by the genes that are differentially expressed in atopy. EASE assigns the identified genes to biological processes based upon the Gene Ontology Consortium and employs statistical methods (Fisher's Exact Test) to identify over-represented biological process categories in a given gene list. We separated our gene lists into two categories, over- and under-expressed genes. The over-expressed gene groups in atopy could be broadly categorized as contributing to specific aspects of the innate or adaptive immune response. Genes in the under-expressed group could be broadly categorized as genes that encode proteins

required for apoptosis. Taken together, the observed alterations in these two groups suggest that both chronic activation (over-expressed gene group) and reduced apoptosis (under-expressed group) may contribute to the chronicity of allergic rhinitis. This is consistent with the suggestion that reduced apoptosis in inflammatory cells may contribute to the chronicity of inflammatory processes (180). Many different pathways are involved in apoptosis, including signaling through death receptors (*FAS*, *TNFR1*(181)), the p53 pathway (182), the Bcl-2 family of genes (183) and the caspase cascade of effector proteases (183). In our dataset, *CASP6* is under-expressed by about 20-fold and *CASP4* is under-expressed by about 3-fold. Components of other apoptosis pathways, such as, *FAS*, *TP53*, and *BCL2,* are not differentially expressed in allergic rhinitis PBMC.

A second goal of microarray-based research in human disease is the identification of genes whose expression levels change in response to a therapy. In this regard, these "biomarkers" may provide quantitative measurements that demonstrate whether or not a given treatment is having its desired effect. Measurements of biomarker levels could augment clinician and patient assessments of therapeutic effectiveness of a given treatment. Desensitization by immunotherapy is a common effective treatment for allergic rhinitis although its precise mechanisms are not fully understood (77).
The immunological changes associated with immunotherapy have been discussed that specific IgE levels decrease and IgG titers often rise, perhaps further inhibiting IgE production by neutralizing the antigen and by antibody feedback (74). It is also possible that immunotherapy may work by inducing specific T cell tolerance or by changing the predominant phenotype of antigen-specific T cells from Th2 to Th1 (75, 76). However,

most of these studies test the production of cytokines or immunoglobulins at proteins levels serologically, and it has also been described elsewhere that the Th2 to Th1 shift represented by mRNA expression changes of cytokines happened in local nasal but not in peripheral blood (184, 185). In our study, we used freshly isolated PBMC to examine the transcription levels of the genes that are suspected to have relevance in allergic rhinitis such as cytokines, chemokines and etc, these genes only exhibit modest differences at transcription levels between pre- and post- immunotherapy. However, our analysis of gene expression profiles of PBMC from atopic individuals before and after immunotherapy permitted accurate segregation of these two groups ($P<0.05$, Fisher's Exact Test). Surprisingly, measurement of the contribution of expression levels of a large number of genes (> 500) was required to achieve complete segregation of the atopic group into pre- and post-therapy sub-groups. This result implicates that immunotherapy can alter the gene expression profile, but the genes that are conventionally considered to be related to allergic rhinitis did not exhibit substantial changes upon immunotherapy in PBMC. And the allergy signature was a relatively stronger signature imprinted in the patients before or after immunotherapy, we have to have more gene expression signals to discriminate the gene expression profiles pre-immunotherapy from post-immunotherapy. When we looked at changes in expression levels of individual genes, we found that we could identify a number of genes whose expression levels were significantly high or low in the atopic group compared to the control group and returned to control levels after immunotherapy. These genes are potential biomarkers that may augment the ability of a clinician to monitor responses to immunotherapy. Whether expression levels of these same genes would also change in response to other forms of therapy that reduce the

symptoms of allergic rhinitis remains to be determined. In contrast to these results, expression levels of those genes that are low in atopy compared to control do not change after immunotherapy. A future goal of drug discovery may be to try to change expression levels of this group of genes, which may provide for novel approaches to treatment of these diseases. Nevertheless, these initial experiments demonstrate that it is feasible to use gene expression changes as biomarkers to monitor therapeutic responses to a complex disease such as allergic rhinitis.

We also applied EASE analysis to the group of genes chosen by computer that permitted accurate segregation of patients before and after immunotherapy. The most over-represented categories in this gene list may be broadly classified as those that contribute to an immune response. These categories include responses to external stimuli, defense responses, immune responses, biosynthesis, cell motility, and antigen presentation, all key elements of the innate and acquired immune response. Taken together, these results provide molecular support for the observed clinical role of immunotherapy, de-sensitization of the host.

In conclusion, there is a highly conserved gene expression profile among individuals with allergic rhinitis, and immunotherapy alters this gene expression profile. Knowledge of these gene lists and the changes following immunotherapy may identify potential therapeutic targets and potential biomarkers that may be used to monitor therapeutic responses. Both of these may contribute to better care of patients with atopic diseases.

CHAPTER V

PREDICTION OF DISEASE SEVERITY IN PATIENTS WITH EARLY RHEUMATOID ARTHRITIS

Abstract

A group of 17 patients with early rheumatoid arthritis (1 ± 0.2 years disease duration) was evaluated at baseline for gene expression profiles. Disease status in these patients was re-evaluated after a mean follow-up of 5 years using an index that combined pain, global and MHAQ scores. Baseline gene expression values were analyzed using unsupervised and supervised computer algorithms to identify "predictor genes" whose combined expression levels correlated with the follow-up disease severity score. Support vector machines (SVM) were used to validate the reliability of these predictor genes. The unsupervised clustering algorithms separated patients with early rheumatoid arthritis into two branches. The only significant difference between these two groups was the disease severity score; demographic variables and medication use was not different between the two groups. Supervised significance of microarray (SAM) analysis identified 6 predictor genes of future disease severity. These results were confirmed using SVM to perform leave-one out cross validation. Our results indicate that peripheral blood gene expression profiles may be a useful tool to predict future disease severity in patients with early rheumatoid arthritis.

Introduction

Rheumatoid arthritis (RA) is a chronic, inflammatory joint disease with autoimmune features. Substantial evidence suggests that early intervention in individuals with RA results in improved control of disease activity, decreased joint damage and fewer extraarticular manifestations (186-188). Early RA patients may benefit from early aggressive therapies, such as new biologic agents that block the activity of TNF-α (Tumor Necrosis Factor alpha), which control disease activity and joint destruction (189, 190). However, these biologic agents are generally expensive and up to 30% of RA patients have incomplete responses (191, 192). In addition, these drugs can have significant side effects including increases in severe infections and other autoimmune manifestations. In the approximately 30% of early RA patients who do not develop any erosion(192), treatment with TNF blockers may not be necessary, and other drugs such as methotrexate may be sufficient. These clinical issues highlight the need for new approaches that would permit individualization of therapy for patients with early RA including development of additional prognostic markers.

Current prognostic methods for RA are generally based on the integrated use of information derived from patient self-assessment questionnaires, the physical examination, routine laboratory studies such as ESR (erythrocyte sedimentation rate) and CRP (C-Reactive Protein) and radiographic findings. Titer of rheumatoid factor and anti-CCP (antibodies to Cyclic Citrullinated Peptides), imaging methods, including conventional X-rays, ultrasound and magnetic resonance, genetic markers, such as *HLA-DRB1* gene alleles, have also been employed (193-197). Most diagnostic methods are dependent on the diagnostic tests that are evaluated. This may result in circularity and

overestimation of the diagnostic properties of the tests (198-200). Considering the prevalence of rheumatoid arthritis, which is estimated as 1% of the US population (NIAMS, National Institute of Arthritis and Musculoskeletal and Skin Diseases 2004), even a small increase in the accuracy of disease severity prediction has the potential to benefit a substantial number of RA patients.

Microarrays provide a powerful tool to screen expression levels of thousands of genes in a single sample. We and others have used this approach to identify gene expression signatures in peripheral blood mononuclear cells (PBMCs) of individuals with autoimmune diseases, including rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), multiple sclerosis, and type I diabetes mellitus (79, 89-91, 149). We also have described a unique gene expression signature that distinguishes patients with early RA (ERA) from those with more established disease (80). The objective of the present study was to determine if gene expression signatures collected early in the course of RA could predict future disease severity.

Materials and methods

Patients

The 17 patients used for this study were included in a previous report detailing results of the initial microarray analyses (201). Follow-up clinical information was obtained by one of the investigators (T.S.) as part of ongoing longitudinal investigations.

The mean (S.E.M.) disease duration for these patients at the time of the initial microarray screening was 1 ± 0.2 year. Clinical course was evaluated after a mean (range) of 5.0 (1.5-7.4) years.

Self-assessment evaluations collected in the follow-up analysis included 100 mm visual analog scales for pain and global assessment and a modified health assessment questionnaire (re-coded MHAQ). Each of these measures was converted to an indexed score. "Pain" and "Global": 0= 0-9, 1= 10-29, 2= 30-59, 3= ≥60; "Re-coded MHAQ": 0= 0, 1= 0.13-0.25, 2= 0.38-0.50, 3= 0.63-1.0, 4= >1.0. An overall composite index score was then calculated as a sum of the three score components. The overall disease score had a possible range of 0-10, with scores of 0-3 considered mild and scores of 4-10 classified as severe(202).

This study was approved by the Vanderbilt University Institutional Review Board. All individuals participate in this study gave written informed consent.

Sample preparation and microarray procedures

PBMC were isolated from 20 ml heparinized blood on a Ficoll-Hypaque gradient. All samples were processed within 2-4 hours of blood collection. Total RNA was isolated with Tri-Reagent (Molecular Research Center. Inc., Cincinnati OH) and 5 μg RNA was used to prepare cDNA with reverse transcriptase (Superscript II, Invitrogen Corporation, Carlsbad, CA) in the presence of ^{33}P-dCTP. Labeled probes were purified using a Bio-Spin 6 Chromatography Column (Bio-Rad Laboratories, Inc., Hercules, CA). Before hybridization, GeneFilters membranes (GF-211, Research Genetics/Invitrogen Corporation, Carlsbad, CA) were washed in boiled 0.5% SDS, saturated with 5.0 ml

Microhyb solution (HYB125.GF, Research Genetics/Invitrogen Corporation, Carlsbad, CA). Filters were treated with pre-hybridization reagents (5.0 µg Human Cot-1 DNA and 5.0 µg Poly dA, Invitrogen Corporation, Carlsbad, CA) in a hybridization roller tube (Midwest Scientific, St. Louis MO) for 2 hours at 42 ºC. Purified, labeled probes were denatured and added to roller bottles containing filters and pre-hybridization solution. GeneFilters membranes were hybridized overnight at 42 ºC. After hybridization, membranes were washed three times, exposed to imaging screens for 24 hr and screens were scanned by a phosphorimager (Molecular Dynamics/Amersham Biosciences, Piscataway NJ). Acquired images were loaded into Pathways 4.0 software (Research Genetics/Invitrogen Corporation, Carlsbad, CA). The relative intensity of each spot on the membrane was determined and the microarray dataset was subjected to further analysis using the different analytical platforms. Data were normalized to yield an average intensity of 1.0 for each clone (4133) represented on the microarray. Reproducibility of the method was established by performing replicate hybridizations to separate microarrays (79). Original microarray data have been deposited into GEO database; accession number GSE1964 (GSM35124-GSM35142).

Data analysis

Cluster (version 3.0) (123) and TIGR microarray software MultiExperiment Viewer (MEV) (130) were used to identify significant genes that can discriminate patients according to their future disease severity. The following data analysis modules of MEV were used to perform further analyses: HCL (hierarchical clustering), ST (support tree clustering), SVM (support vector machines) and PCA (principal components

analysis). Detailed descriptions of the applications of these programs to the analysis are provided in the results section. Analysis procedures presented here comply with MIAME (minimal information about a microarray experiment) guidelines established by the Microarray Gene Expression Data Society (www.mged.org). Clinical variables are shown as mean ± S.E.M. Statistical analyses of the clinical data were carried out using Fisher's exact test or Student's T Test with a P value of <0.05 considered significant.

Results

Clustering analysis of ERA patient gene expression profiles

We first performed unsupervised hierarchical clustering of gene expression profiles of the ERA patients. The algorithm segregated the patient profiles into two major groups (Fig. 5-1).

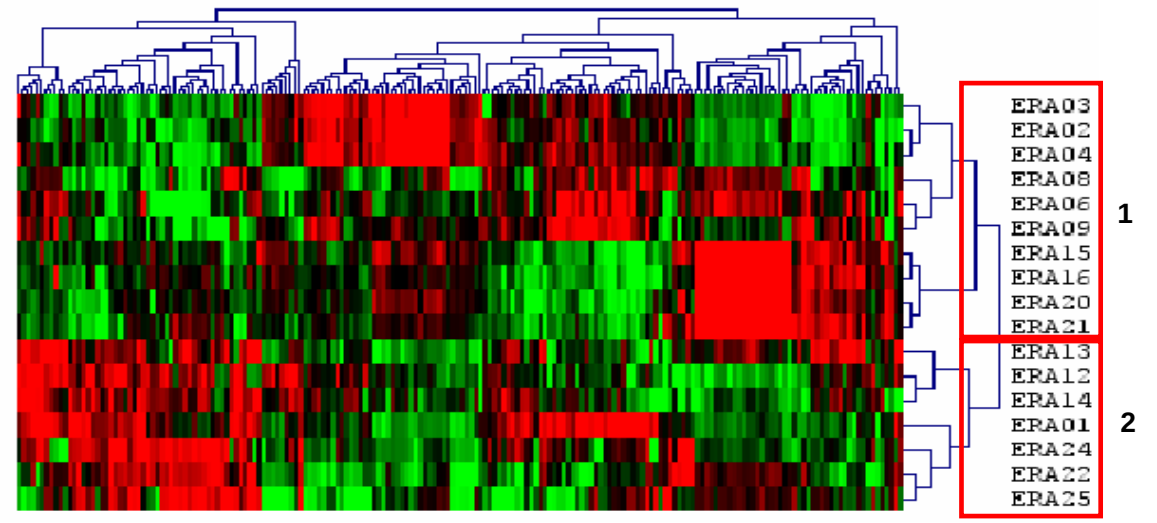


Figure 5-1. Unsupervised hierarchical clustering was applied to the gene expression profile of 192 genes that passed the filtering condition (SD=2). The patients were separated into two Clusters: 1 and 2 as indicated by the red square.

We compared patient clinical features to determine if the patients that segregated into the two clusters exhibited any common clinical feature. We found that age of onset, race, gender and RF positivity were not statistically different between individuals in the two groups. Medication usages for steroids and DMARDs or their combinations at the time when samples were collected were also not statistically different (Table 5-1).

Table 5-1. Clinical Phenotypes Corresponding to Cluster Designations

Phenotype	**Cluster**	**1**	**2**	***P* value**[a]
Age of onset		61	65	ns
Severity	Mild	8	1	***0.015***
	Severe	2	6	
Race	Black	0	2	ns
	White	10	5	
Gender	Female	9	6	ns
	Male	1	1	
Rf	Positive	7	4	ns
	Negative	3	3	
GLUCOCORTICOIDs and DMARDs				
PRED		5/10	5/7	ns
MTX		9/10	6/7	ns
HCQ		1/10	2/7	ns
LEF		1/10	0/7	ns
ENB		1/10	1/7	ns
REMIC		0/10	1/7	ns
PRED+MTX +HCQ+LEF		9/10	6/7	ns
ENB+REMIC		1/10	2/7	ns
All DMARDs		9/10	6/7	ns

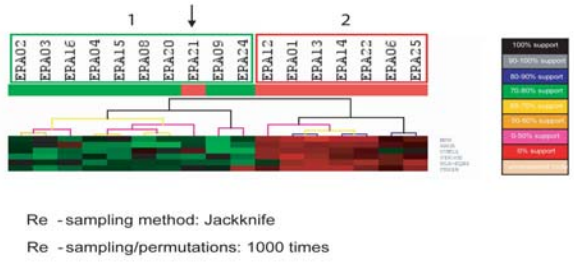
a: the difference of "age of onset" is calculated by students' T test, the differences of other phenotypes are calculated by Fisher's exact test. ns: not significant

The only significant difference was the disease index distribution. In Cluster 1, 8 of the 10 patients (80%) were classified as having mild disease while in Cluster 2, only 1 of 6 (17%) was in the mild class (P=0.015).

Significance analysis of microarray (SAM)

The supervised method- Significance Analysis of Microarray (SAM) was used to identify genes that were differentially expressed according to their disease severity (Fig. 5-2 A). To do this, we divided patients into a severe disease group and a mild disease group according to their clinical evaluation (see Patients section) and performed SAM analysis with the false discovery rate set to zero. Under these stringent conditions, 6 genes were identified. We then performed unsupervised support tree clustering using expression levels of these 6 genes (Jackknife algorithm re-sampling by 1000 permutations).

A



Re -sampling method: Jackknife

Re -sampling/permutations: 1000 times

B

Severity	Branch		p value
	1	2	
Mild	9	0	0.0004
Severe	1	7	

Figure 5-2. Support tree hierarchical clustering was applied to the six predictor genes expression profile among 17 patients. A. Jackknife re-sampling was used with permutation 1000 times. The color codes correspond to a given level of support. B. Table shows that branch 1 and 2 are significantly separated according to their disease severity (P=0.0004, Fisher's Exact Test).

As shown in Figure 5-2A, two major branches were derived. Patients separated into two groups by their future disease severity. Branch 1 contained 10 patients, and only 1 of these was in the severe group while Branch 2 had 7 patients, all of whom were classified as severe (P=0.0004, Fisher's exact test) (Fig 5-2 B). This analysis classified each patient according to the clinical evaluation with the exception of patient 21. According to the clinical evaluation, patient 21 should be in the severe group. However, the clustering results grouped this patient into the mild group.

Leave one out cross validation of support vector machines (SVM)

We next wanted to validate that the 6 identified genes were predictors of future disease severity. Unfortunately, we did not have a new early RA patient group to use for validation. As an alternative, we used SVM to cross validate the predictors. SVM uses kernel function to search a hyperplane in multi-dimensional space that maximizes the distance of training samples to the plane from either side and then classifies the test samples based on their orientation to the hyperplane (28). We used leave one out cross validation (LOOCV) to validate predictor genes. LOOCV analysis employs SVM to perform multiple permutations. At every permutation, one of the samples is removed and the remaining samples are used as predictors to re-classify the removed sample.

Figure 5-3 shows the results of LOOCV of SVM analysis. We first assigned numbers to each patient according to their disease severity: "1" = "severe" and "-1" = "mild" based upon their clinical evaluation. The LOOCV analysis reclassified patient 21 from the "severe" group to the "mild" group but kept all the other classifications the same. In panel II, we assigned the scores according to our clustering results. Here patient 21

was given a score of "-1". The LOOCV analysis maintained the original classification of "-1" for patient 21 as well as the classification of remaining patients. In order to further confirm our classification by the 6 genes we changed patient 1's score from "1" to "-1", and patient 21's score from "1" to "-1". After LOOCV analysis, patient 1's score was corrected to "1" but patient 21's score remained "-1" (Fig.5-3, panel III). Next, we changed patient 1's score to "-1" and left the patient 21's score "1" or unchanged. After LOOCV analysis, patient 1's score was corrected to "1" and patient 21's score was corrected to "-1" (Fig.5-3, panel IV).

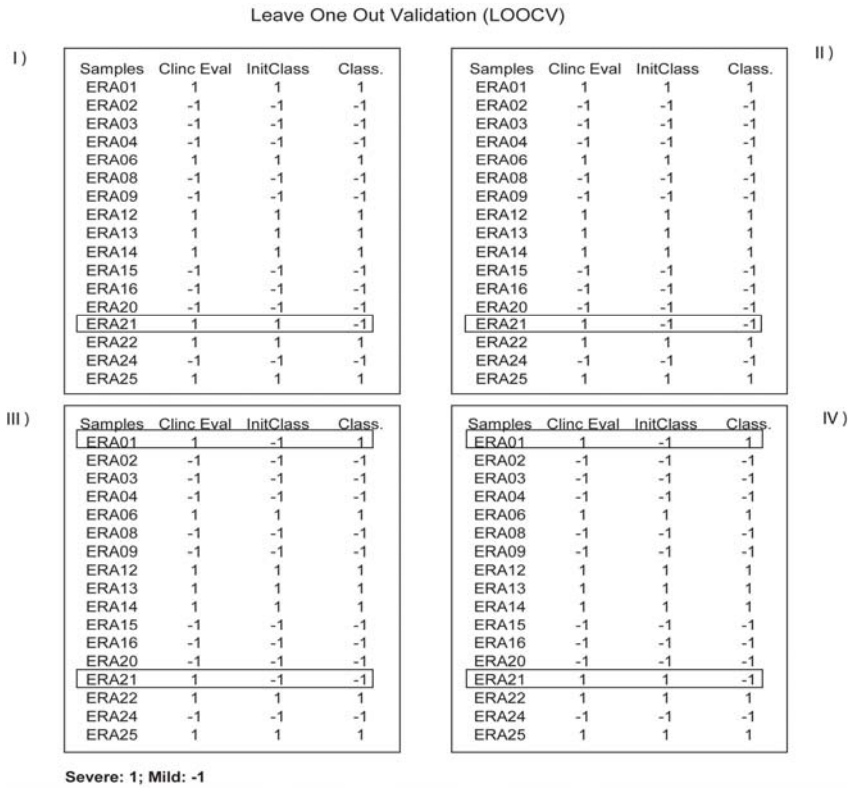


Figure 5-3. LOOCV of Support Vector Machines. Leave one out cross validation of support vector machines was used to validate our original testing of patients. The scores of disease severity are assigned to two groups: severe=1, mild=-1. "Clinc Eval" column shows scores assigned to each patient according to clinical evaluation index (see patients and methods); "Init Class" column shows scores used as initial classification for SVM; "Class" column shows scores assigned to each patient after cross validation.

Taken together, the above LOOCV analysis supports the results from the clustering analysis, indicating that the 6 genes may be appropriate candidates as predictors of future disease severity in early rheumatoid arthritis patients.

Principal components analysis of 6-predictor gene expression profiles

By using SAM analysis, we identified 6 genes as predictors of future disease severity in ERA patients. After unsupervised clustering (support tree), we classified the patients into two groups, a severe group and a mild group, except for patient 21. Therefore, we used Principal Component Analysis (PCA) to examine the clusters by using the 6 gene expression profiles. PCA projected all the patients into a 3 dimensional space according to their 6 genes expression profiles (Fig. 5-4 A). The severe group (red spheres) and mild group (green spheres) were separated into two polar positions. Next, we projected all the patients onto a two dimensional plane (X, Y) (Fig. 5-4 B).

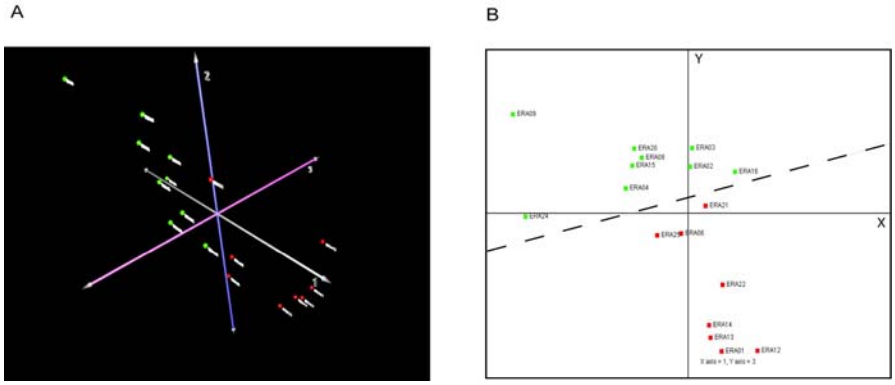


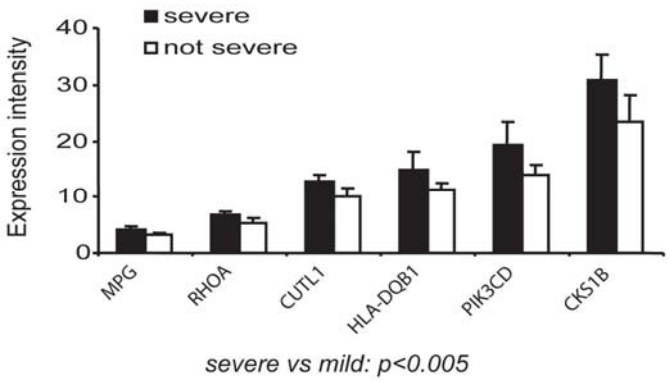
Figure 5-4. PCA validation of classification by 6 predicator genes. A, three-dimensional projection of 17 patients by expression levels of 6 predictor genes. Green spheres represent the mild group. Red spheres represent the severe group. B. 2-dimensional projection of 17 patients by gene expression levels of 6 predictor genes onto (X, Y) plane. Dashed line was drawn between severe group and mild group of ERA patients. Green squares represent the mild group. Red squares represent the severe group.

Patients were separated into two areas by the dashed line. Patient 21 aligned very close to the mild group, which may be the reason why we cannot classify this subject in accordance with the clinical evaluation.

The 6-predictor genes

Each of the 6-predictor genes: *MPG*, *RHOA*, *CUTL1*, *HLA-DQB1*, *PIK3CD*, and *CKS1B*, was over-expressed in the severe patient group compared to the mild patient group (Fig. 5-5 A).

A



B

Acc	Symbol	severe	not severe	p value	Name
N26769	MPG	4.20	3.30	0.0002	N-methylpurine-DNA glycosylase
AA676955	RHOA	6.90	5.47	0.0004	Ras homolog gene family, member A
AA292536	CUTL1	12.82	10.08	0.0003	Cut-like 1, CCAAT displacement protein (Drosophila)
AA458472	HLA-DQB1	14.89	11.30	0.0027	Major histocompatibility complex, class II, DQ beta 1
AA281784	PIK3CD	19.23	13.78	0.0015	Phosphoinositide-3-kinase, catalytic, delta polypeptide
AA459292	CKS1B	30.69	23.49	0.0031	CDC28 protein kinase regulatory subunit 1B

Figure 5-5. Average gene expression intensity in severe group and mild group is shown by bar graph in A for the 6 predictor genes. The value represented is the mean ± standard deviation in each group. Each gene's gene bank accession number, symbol, name and average expression intensity in severe and mild group are shown in B, students' T test was used to measure the significance of difference.

Although the average difference in transcript levels of each gene was < 2-fold between the two groups, the difference in transcript levels of each gene achieved a high level of statistical significance ($P < 0.005$) (Fig. 5-5 B).

Discussion

Disease-modifying therapy early in the course of RA can lead to improved disease control and decreased joint damage (203, 204). However, the uncertain course of RA in some patients coupled with the adverse effects and high cost of newer therapies make decisions regarding treatment strategies complex. About 30% of early RA patients will not develop severe disease and a small number may even undergo remission without treatment (191, 192). The current prognostic factors are relatively powerful tools, including measurement of HAQ scores, autoantibody levels and genetic markers. However, the prognostic sensitivity of the combined tests is in range of 80%-90% which is not sufficient to predict outcome in an individual patient (192, 205-208).

In the present study, we identified a combination of 6 expressed genes that correlated with clinical course in RA patients. We can achieve > 94% (16/17) accuracy of prediction of future disease severity if our clinical follow up assessment is 100% accurate. We propose that addition of this analysis to other measures, such as HAQ score, HLA genotyping, and rheumatoid factor may be useful to predict future disease severity in early RA patients. This information may then be useful for making decisions about therapies.

In our study, we did not have an independent population to validate our predictor genes and considering the small sample size of the data set, as an alternative, we applied

LOOCV. LOOCV has been widely used to validate classification and prediction models of disease and is considered to be statistically sufficient validation (102, 109, 138). We found that our results from LOOCV were consistent with our clustering analysis. However, independent validation is still the gold standard to verify our results.

The difference in the expression pattern of the 6-predictor genes is remarkably consistent in both the severe and mild patient group. Interestingly, MPG protein levels are know to be increased in patients with RA compared to controls (209). Most analyses of *HLA-DR* and *HLA-DQ* have focused on genetic polymorphisms rather than differences in expression levels (197, 210). In addition to these sequence polymorphisms, transcript levels of *HLA-DRB1* are also elevated in RA (211). Our results confirm that transcript levels of *HLA-DQB1* are elevated in RA and demonstrate that they also help discriminate mild from severe disease. We did not find direct evidence to link *RHOA*, *CUTL1*, *PIK3CD*, and *CKS1B* with the disease activity of rheumatoid arthritis. However, proteins encoded by these genes are involved in critical cellular functions such as signal transduction, regulation of gene expression, and cell cycle progression and may have unknown effects on disease course of RA.

Our 6-predictor gene system accurately discriminates the severe group and mild group achieving more than 94% accuracy. These results suggest that it is possible to predict future disease severity using this type of approach. At the time of sampling, between sampling and clinical follow-up, and at the time of clinical follow-up, all patients were on some type of anti-rheumatic therapy. Clinically, RA is a very heterogeneous disease and our system may discriminate between individuals who will develop aggressive or mild disease. Alternatively, our system may discriminate between

individuals who exhibit good responses to anti-rheumatic therapies and therefore develop mild disease and those patients who exhibit poor responses to anti-rheumatic therapies and therefore develop severe disease. Other unrecognized co-variables may also contribute to the precise mechanism that allows expression levels of the 6-predictor gene set to predict if an individual will develop severe or mild disease.

In summary, we identified 6 genes whose expression levels in ERA predict future disease severity. By using LOOCV and PCA, we validated this classification. Taken together, these results suggest it may be possible to use gene expression profiling in ERA to predict future disease severity.

CHAPTER VI

STUDY OF MURINE MODELS OF AUTOIMMUNE DISEASE BY GENE EXPRESSION PROFILING

Abstract

A general view is that critical genes involved in biological pathways are highly conserved among species. In order to understand human autoimmune diseases, a great deal of effort has been devoted to the study of murine models that mirror many pathologic properties observed in a given human disease. In our study, we first identified a specific gene expression profile of T lymphocytes in NOD mice, in which a deregulated system was identified. This feature of lymphocytes in NOD mice may explain their susceptibility to apoptosis compared to non-diabetic C57BL6 or NOD.H2^{h4} mice and it may represent the underlying mechanism causing lymphopenia in NOD mice believed to contribute to onset of autoimmunity.

Second, since we found that lymphocytes from humans with different autoimmune diseases all carry a common conserved gene expression profile, we wanted to determine if lymphocytes from common murine models of autoimmune disease carried a gene expression profile similar to the human profile and if both IDDM and SLE mouse models carried a shared gene expression profile. We identified numerous differentially expressed genes in the autoimmune strains compared to non-autoimmune strains. However, we found very little overlap in the gene expression profile between human autoimmune disease and murine models of autoimmune disease and between different

murine autoimmune models. Our research further demonstrates that murine models of autoimmunity do not perfectly match human autoimmune diseases.

Part I: Deregulated stress system of lymphocytes in NOD mice measured by gene expression profiling

Introduction

Autoimmune diseases are believed to arise from complex interactions between genetic and environmental factors(1, 142). In order to understand the pathogenesis of autoimmune diseases, investigators have devoted considerable effort to the study of animal models of autoimmune disease. For diabetes, these include diabetic prone BB rats and non-obese diabetic (NOD) mice that spontaneously develop an insulin-dependent diabetes mellitus (IDDM)-like syndrome (54, 55, 212, 213). A profound lymphopenia has been reported to lead to diabetes in BB rats (20, 21), and intrinsic defects in lymphopenia and homeostatic expansion in NOD mice may contribute to establish autoimmunity (17). In this study, we try to determine if apoptosis defects exist in lymphocytes of NOD mice, which could represent an underlying mechanism leading to lymphopenia. We also wanted to determine if we could find alterations in transcript levels of specific genes that may contribute to alterations in sensitivity to apoptosis. We revealed that NOD mice had a higher apoptotic population of lymphocytes than that of B6 mice or the non-diabetic congenic strain, NOD.H2^{h4}, *in vivo*. NOD T cells are also more susceptible to apoptosis-inducing stimuli than B6 or NOD.H2^{h4} T cells. Gene expression profiling analysis identified a set of genes with a unique expression profile in NOD T cells. The apoptosis

prone characteristic of this gene expression profile was identified by the over-represented biological categories, especially the de-regulated heat shock proteins. Therefore, this specific expression pattern of these genes may contribute to establishing susceptibility to apoptosis that may increase the likelihood of developing autoimmune diabetes.

Materials and methods

Mice: NOD, C57BL/6 and B6.*Idd1,5* female mice were obtained from Jackson Labs (Bar Harbor, ME). Procedures and care of animals were in accordance with Vanderbilt University guidelines provided by the Institutional Animal Care and Use Committee. B6.*Idd1,5* is a double congenic strain that contains the NOD MHC *Idd1* and NOD *Idd5* loci on the C57BL/6 background. This strain develops periinsulitis, but insulitis is extremely rare (Jackson Labs Mice Database). The experiments were performed by using the female, age matched mice at age of 12 weeks old.

Reagents: complete RPMI 1640 medium supplemented with 10% FBS (HyClone Laboratories, Logan, UT), 100 U/ml of penicillin, 100 μg/ml of streptomycin, 2mM L-glutamine, and 0.05 mM 2-ME (J. T. Baker, Phillipsburg, NJ) was used for cell culture. The mAbs used in this study were purified from tissue culture fluids of hybridoma cells obtained from American Type Culture Collection (Manassas, VA): anti-I-A (10-3.6.2, specific for I-A g7, k, r, f, or s haplotypes; Y3P, specific for I-A b, f, p, q, r, s, u and v haplotypes). BioMag® goat anti-mouse IgG coated magnetic beads were obtained from QIAGEN (Valencia, CA). Tri Reagent was obtained from Molecular Research Center, Inc (Cincinnati, OH). Superscript II transcriptase was obtained from Invitrogen Corporation (Carlsbad, CA). Annexin-V PE Apoptosis Detection Kit I, FITC-conjugated

anti-CD3 antibody and APC-conjugated anti-CD19 antibody were purchased from BD Pharmingen (San Diego, CA 92121). PCR purification kits were obtained from QIAGEN. CMT^{TM} hybridization chambers were obtained from CORNING (Corning, NY). Cy5, Cy3 and other reagents used for cDNA labeling and hybridization were obtained from VMSR (Vanderbilt Microarray Shared Resource). NIA mouse15K microarray chips were made by VMSR.

NIA Mouse 15K cDNA clone set: the NIA mouse 15K cDNA clone set is a 15,247-element clone set. The 15K clone set was developed at the National Institutes of Aging (NIA) and has been sequence verified from both the 5' and 3' termini. Approximately 15,000 unique cDNA clones were derived from 52,374 ESTs from pre- and periimplantation embryos, E12.5 female gonad/mesonephros, and new born ovary(214). Every cDNA clone is assigned a unique clone ID in the mouse 15K NIA clone set.

$CD3^+$ T cells purification: $CD3^+$ T lymphocytes were purified from spleen by negative selection. Red blood cells were removed by hypotonic lysis. I-A expressing cells were removed by incubation with anti-I-A antibody (10-3.6.2 for NOD and B6.*Idd1,5*; Y3P for C57BL/6) for 30 min at 0°-4°C and with goat anti-mouse IgG coated magnetic beads at room temperature for 30 min. A magnetic column removed cells bound to beads. Average purity of $CD3^+$ T cells was approximately 90-95% as determined by flow cytometry.

Total RNA isolation: Tri Reagent was used to isolate total RNA according to the manufacture's protocol. We pooled total RNA from purified T cells of 2-3 mice for each microarray experiment. This provided us with sufficient total RNA and decreased bias

due to biological variation. We used 5×10^7 cells to isolate total RNA for each group (experimental and control) for each microarray experiment.

cDNA labeling, microarray hybridization, and data analysis: Labeled cDNA was prepared from 10 μg of total RNA. RNA was transcribed into fluorescently labeled cDNA using an anchored oligo-dT primer ($dT_{20}MN$), unlabeled dNTPs, superscript II reverse transcriptase, and dCTP labeled with one of the two fluorophores, Cy5 or Cy3, according to the standard protocol of VMSR. The Cy5 and Cy3 labeled cDNAs were combined, purified together, and dissolved in 2x hybridization buffer (50% formamide, 10 × SSC and 0.2% SDS). Microarray slides were prehybridized in 1% BSA, 5 × SSC, and 0.1% SDS at 55°C for 45 min. After prehybridization, slides were rinsed 5 times in MilliQ water, 1 time in isopropanol, and allowed to air dry. Combined samples were placed on the prehybridized slides and covered with lifterslips. Microarray slides were hybridized at 42°C for 14-16 hours in a humidified hybridization chamber. After hybridization, the lifterslips were removed and microarray slides were washed at room temperature with serial 5-min washes with 2 × SSC, 0.1%SDS, 1 × SSC, and 0.1 × SSC and dried by centrifugation at 50 × g for 5 min. Microarray slides were scanned using a GenePix® 4000B scanner (Axon Instrument Inc., Union City, CA). Image files were analyzed with GenePix® Pro 3.0 software (Axon Instrument Inc., Union City, CA). The software package was used to identify individual clones from the microarray image and to calculate signal intensity values. Ratios (fluorescent intensity of experimental sample/ fluorescent intensity of control sample) from the two channels (red: Cy5, green: Cy3) represent the fold change in gene expression levels between the two samples. Each

microarray was performed at least twice and results were averaged. Microarray original data were deposited into GEO database: GSE2524

Definition of differentially expressed genes-DEGs: DEGs are defined as a clone that exhibited a greater than 3-fold difference in expression between subject strains and reference strains with a $P < 0.05$ (the Students' T test).

Biological process categorization by gene ontology: EASE: the Expression Analysis Systemic Explorer (http://david.niaid.nih.gov/david/ease.htm, EASE version 2.0) was used to search for common biological themes within gene lists generated by our microarray analysis. EASE assigns identified genes to "GO: Biological Process" categories from the Gene Ontology Consortium (www.geneontology.org) (139-141). Significant over-represented categories are those with the number of list hits (LH) of at least 2 with a $P < 0.05$ (Fisher's Exact Test).

Reciprocal bone marrow transfer experiment: Two MHC class II matched strains: NOD and B6.*Idd1,5* were used. Recipient mice were irradiated by gamma irradiation at a lethal dose of 10 Gy and received donor bone marrow (1×10^7 cells) by I.V. injection at the base of the tail. T cells were harvested 6 weeks after bone marrow transfers and used for microarray analysis.

Endogenous and stimuli-induced apoptosis: To induce apoptosis dependent upon DNA strand breaks, spleen cells were exposed to a total dose of 2 Gy of gamma irradiation using a Marker 1 Irradiator (J.L.Shepherd & Associates, Glendale, Calif.) at a dose rate of 1.897 Gy/min at room temperature. To induce a heat shock response, cells were transferred into 5 ml medium (2×10^6/ml) and incubated at 42 ºC for 30 min. After 16 hours of culture at 37 ºC, cells were harvested. Freshly isolated and cultured cells

were stained with PE-conjugated Annexin-V, FITC-conjugated anti-CD3 antibody and APC-conjugated anti-CD19 antibody according to the manufacture's protocol. Flow cytometry data were collected on a FACScan using CellQuest Software (BD Bioscience, San Jose, CA) and analyzed by WinMDI software packages (provided by J. Trotter, The Scripps Research Institute, La Jolla, CA).

Reverse transcription and Real-time PCR analysis: Total RNA was prepared using Tri-Reagent and used for first strand synthesis with Superscript III First-Strand Synthesis System for RT-PCR (Invitrogen Corporation, Carlsbad, CA). 2 μg of the total RNA of NOD and B6 mice was reverse-transcribed. In order to evaluate the transcribed *Hsp105* gene expression, we performed TaqMan RT-PCR using ICycler IQ Multicolor Real-time PCR Detection System (Bio-Rad laboratories, Inc., Hercules, CA). Reactions were performed in triplicate. TaqMan reactions were setup in 96-well PCR plates (Bio-Rad laboratories, Inc., Hercules, CA 94547) according to manufacturer's protocols. For each reaction, added 22.5 μl cDNA template (50ng RNA used), 25 μl 2X TaqMan Universal Master Mix (Applied Biosystems, Branchburg, NJ) and 2.5 μl 20X TaqMan Gene Expression Assay. Running the plate by default PCR thermal conditions according to manufacturer's protocol (Applied Biosystems, TaqMan Gene Expression Assay Protocol 4333458B). TaqMan Gene Expression Assays: Mm00607939_s1, *Actb*; Mm00442864_m1, *Hsp105/Hsp110* (Applied Biosystems, Branchburg, NJ). Changes in cycle threshold (ΔC_t) values of the samples were determined by subtracting the average of the triplicate C_t values of the target gene (*Hsp105*) from the average of the triplicate C_t values of the reference gene (*ACTB*). The relative gene expression levels were determined by subtracting the average ΔC_t value of the target from the average ΔC_t

value of the calibrator. The amount of target (expressed as fold change), normalized to an endogenous reference and relative to a calibrator, was expressed as $2^{-\Delta\Delta C_t}$ (User Bulletin #2, Applied Biosystems, Branchburg, NJ).

Western blot assay: Cell lysates were prepared from purified $CD3^+$ T cells of NOD, and B6 mice by incubation in 250 µl RIPA buffer (Tris-HCl: 50 mM, pH 7.4 NP-40: 1% Na-deoxycholate: 0.25% NaCl: 150 mM EDTA: 1 mM). Protein (80 µg/lane) was separated on 10% precast SDS-PAGE gels (Bio-Rad Laboratories, Hercules, CA), under reducing conditions and blotted onto Sequi-Blot PVDF membrane (Bio-Rad Laboratories, Hercules, CA). Hsp90 α/β and Hsp105 were identified using the rabbit polyclonal anti-Hsp90α/β, anti-Hsp105 antibodies (Santa Cruz Biotechnology Inc, Santa Cruz, CA), followed by HRP-conjugated bovine anti-rabbit-IgG (Santa Cruz Biotechnology Inc, Santa Cruz, CA) as the secondary antibody. Membranes were exposed to ECL plus reagent (Amersham Biosciences UK) according to the manufacture's protocol, developed with a Kodak Image Station 440CF, and analyzed using one-dimensional Image Analysis software (Kodak Molecular Imaging System, USA).

Results

Apoptosis in NOD T cells

Lymphopenia in NOD mice may result from increased apoptosis or other mechanisms. We used two approaches to explore this hypothesis. First, we determined if strain dependent differences in the portion of apoptotic lymphocytes exist, *in vivo*. We

used Annexin-V staining and flow cytometry to determine the percentage of apoptotic T cells and B cells in age-matched female NOD, B6 and NOD.H2^{h4} mice. We found a larger fraction of apoptotic T cells in NOD compared to B6 and NOD.H2^{h4} (Figure. 6-1a). Interestingly, we also identified a similar pattern in B cells. The percentage of apoptotic B cells was higher in NOD mice than in B6 and NOD.H2^{h4} mice (Fig. 6-1b).

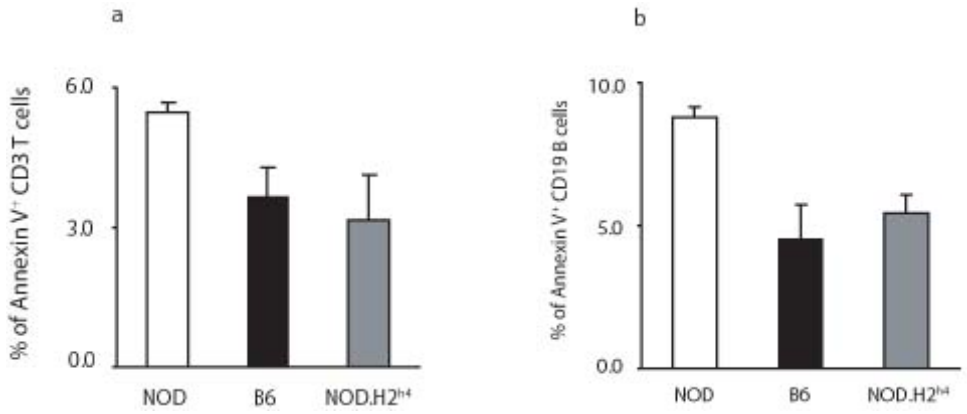


Figure 6-1. Lymphocyte apoptosis, *in vivo*. Spleen cells were harvested from female age-matched mice of the indicated strains. Lymphocytes were stained with FITC-labeled anti-CD3, APC-labeled anti-CD19, and PE-labeled annexin V and analyzed by flow cytometry. Results are expressed as the mean ± standard deviation of 6 mice in each group. (a) % apoptotic T cells, NOD vs B6, NOD vs NOD.H2^{h4}: $P<0.05$ (Students' T test). (b) % apoptotic B cells, NOD vs B6, NOD vs NOD.H2^{h4}: $P<0.05$ (Students' T test).

Second, we performed cell culture experiments to determine sensitivity to cell death after exposure to stimuli known to induce apoptosis, gamma irradiation and heat shock. After 16 hours, cultures were harvested and cells labeled with PE-conjugated Annexin-V and FITC-conjugated anti-CD3. The percentage of apoptotic cells was determined by flow cytometry. We found that NOD T cells were more sensitive to gamma radiation induced apoptosis than B6 or NOD.H2^{h4} T cells (Fig. 6-2a). The rate of apoptosis in NOD T cells was 30% higher than in B6 T cells.

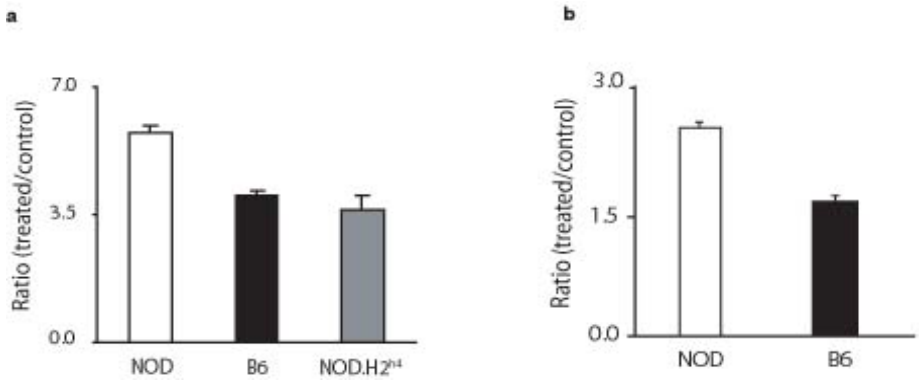


Figure 6-2. Rate of T cell apoptosis after (a) gamma irradiation induced apoptosis or (b) heat shock induced apoptosis, in vitro. T cells were harvested from female age-matched mice of the indicated strains and subjected to gamma irradiation or heat shock as outlined in Methods. After 16 hr, cultures were harvested, labeled with FITC-labeled anti-CD3 and PE-labeled Annexin V and analyzed by flow cytometry. The ratio represents the % of Annexin-V positive cells in the treated group compared to the untreated group. The values represent mean ± standard deviation of 3 animals in each group. Student's t-test was used to calculate statistical significance, in (a): NOD vs B6, NOD vs NOD.H2^{h4}: $P<0.05$ (Students' T test) in (b) NOD vs B6: $P<0.05$ (Students' T test).

We also compared apoptosis between T cells of NOD and B6 mice after heat shock. Apoptosis of NOD T cells was almost 40% higher than that of B6 T cells (Fig. 6-2b). This suggests that the increased apoptotic fraction observed, *in vivo*, may reflect an intrinsic property of NOD T and B lymphocytes rather than an external property of the host environment.

Unique gene expression profiles in NOD T cells

From the apoptosis assays, we learned that NOD lymphocytes have an apoptosis-prone characteristic and it may lead to lymphopenia in NOD mice. Next, we tried to search for an underlying mechanism that may explain susceptibility to apoptosis in NOD lymphocytes. We used microarrays to attempt to identify differentially expressed genes between NOD mice and control strains-B6 mice to determine if these differentially

expressed genes can be candidates to confer a functional liability upon NOD T cells to make them more susceptible to apoptosis than B6. We identified 109 differentially expressed genes (DEGs) between the two strains. Most DEGs were over-expressed (81 DEGs) rather than under-expressed (28 DEGs) in NOD T cells. The level of over-expression ranged from approximately 3- to 15-fold, the level of under-expression ranged from approximately 3- to 6- fold (Fig. 6-3).

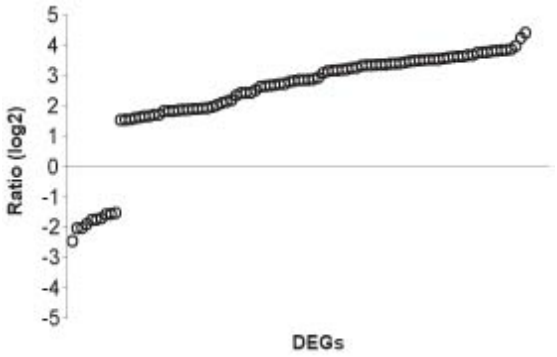


Figure 6-3. Gene expression profiles of differentially expressed genes of NOD T cells. Total RNA was isolated from 10-12 week old NOD and C57BL/6 T cells; labeled with Cy5 or Cy3 dyes and hybridized to the NIA mouse 15K microarray slide. The ratio (log2) (NOD/B6) of expression of individual DEGs (○) is shown.

This set of genes (NOD DEGs, hereafter) can be viewed as a specific profile of NOD T cells.

EASE analysis of NOD DEGs

NOD DEG expression profiles may result in strain dependent differences in functional properties. Therefore, we performed EASE analysis to categorize over- and under-expressed genes in NOD DEGs into functional categories that are over-represented compared to the entire NOD DEG list (Fisher's Exact Test, $P < 0.05$). The most over-

represented categories in the over-expressed gene group were apoptosis-inducing genes including *BNIP3L*, *PDCD7*, and *TIA1* (Table 6-1).

Table 6-1. Biologic process categories significantly over-represented (P < 0.05; Fisher's Exact Test) by differentially expressed genes in NOD mice.

Over expressed genes			
Gene Category	**LH**	**EASE**	**Gene Symbol**
apoptosis	3	0.046	*Pdcd7; Bnip3l; Tia1*
proton transport	2	0.012	*Atp1a1;Atp6v0d1*
chromatin assembly	2	0.019	*Nap1l4; Cbx5*
coenzyme, prosthetic group synthesis	2	0.021	*Atp6v0d1; Alas2*
Under expressed genes			
Gene Category	**LH**	**EASE**	**Gene Symbol**
protein folding	5	0.000	*Hsp90α/β;Cct3; Hspa8; Dnajb1*
response to external stimulus	5	0.000	*Hsp90α/β; Hsp105;Hspa8; Ly6e*
response to stress	4	0.002	*Hsp90α/β; Hsp105;Hspa8*
protein metabolism	6	0.030	*Hsp90α/β;Cct3;Hspa8;Dnajb1; Ly6e*

LH: list hits, the number of identified differentially expressed genes within the specific category. EASE: P value of Fisher's Exact Test.

In the under-expressed gene group, we found that all of the most over-represented categories were stress response related heat shock proteins or chaperones, including *Hsp90α/β* and *Hsp105*. Both *Hsp90 α/β* and *Hsp105* are known regulators of cellular apoptosis and, under certain experimental conditions, will inhibit cell apoptosis (215-218). The EASE analysis results suggested that the NOD DEGs are associated with apoptosis pathways.

Validation of microarray data

Next we performed western blotting to determine expression levels of Hsp90α/β and Hsp105 proteins in NOD and B6 T cells. Hsp90 protein expression was lower in NOD T cells compared to B6 T cells (Fig. 6-4a). Similarly, Hsp105 protein expression was lower in NOD compared to B6 T cells (Fig. 6-4b). We also used real-time PCR to measure the expression level of *Hsp105*. We found it to be higher in B6 than in NOD T cells (mean fold changes of 4.4, range 3.5-5.3). These results confirmed our microarray results.

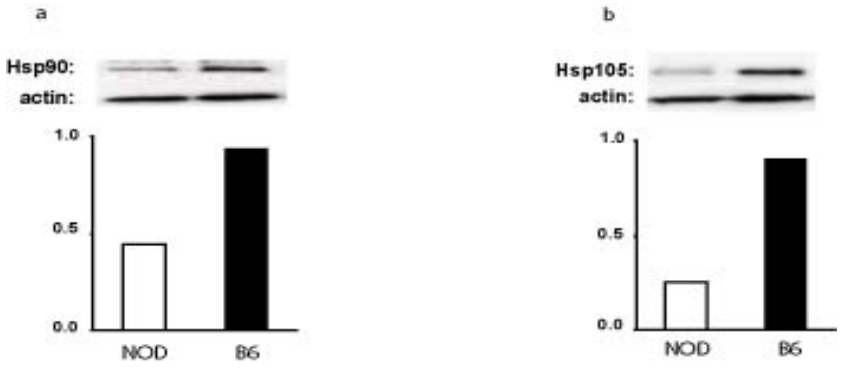


Figure 6-4. Western blot analysis of Hsp90 and Hsp105 protein levels in NOD, B6 T cells. (a) Hsp90α/β protein levels in NOD and B6 detected by western blot assay**.** β-actin was used as loading control. The bar graph represents Hsp90 expression normalized to β-actin. (b) Hsp105 protein expression in NOD and B6 detected by western blot assay**.** β-actin was used as loading control. The bar graph represents Hsp105 expression normalized to β-actin.

NOD T cells DEGs are donor- rather than host-dependent in bone marrow transplants

We performed bone marrow transfer experiments to determine if the unique NOD DEG expression profile was dependent upon donor cells of origin or the host environment. Splenic T cells were harvested from bone marrow chimeras ("B6toNOD" refers to chimeric mice whose donor is B6.*Idd1,5* and recipient is NOD; similarly, "NODtoB6" refers to bone marrow chimeras whose donor is NOD and recipient is B6.*Idd1,5*). We compared NOD DEG profiles in T cells between NOD mice and

"NODtoB6" mice. We found that most of the DEG expression profile in "NODtoB6" mice was similar to that in NOD mice (Fig. 6-5a).

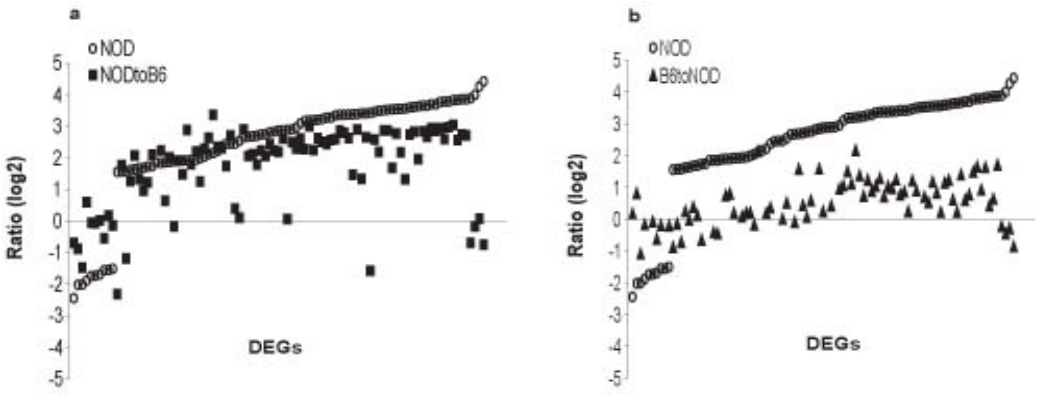


Figure 6-5. Gene expression profiles of NOD DEGs developing in different host environments. Total RNA was isolated from parental T cells and T cells from recipient mice after adoptive transfer: (a) NODtoB6 (donor: NOD mice, recipient: B6.*Idd1, 5* mice). (b) B6toNOD (donor: B6.*Idd1, 5* mice, recipient: NOD mice).

However, when we compared the gene expression profiles of DEGs in T cells between NOD mice and "B6toNOD" mice, we found that the expression levels of most of the DEGs in "B6toNOD" mice were similar to expression levels of B6 T cells (Fig. 6-5b). Therefore, these bone marrow transfer experiments clearly support the notion that, at least for splenic T cells, the gene expression profile is a strain dependent trait and does not vary significantly when T cells develop in a different strain with an identical MHC. However, we cannot rule out the possibility that MHC also contributes to the unique gene expression profile of T cells.

Association of Gene expression profiles with MHC

From the bone marrow transplantation experiments, we demonstrated that the gene expression profile of NOD T cells seems to be an intrinsic trait rather than a

reflection of the host environment. Next we performed microarray analysis to compare gene expression profiles in T cells from diabetic NOD and the non-diabetic congenic strain- NOD.H2^{h4}. We used the same RNA pool from C57BL/6 T cells for comparison and found that the gene expression level of NOD DEGs was greatly altered by the exchange of the MHC (*Idd1* loci). Expression levels of NOD DEGS in NOD.H2^{h4} T cells were more like B6 T cells than of NOD T cells (Fig. 6-6).

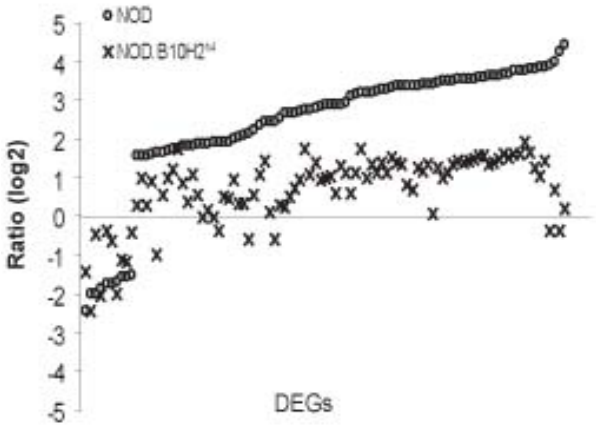


Figure 6-6. Comparison of gene expression profiles of splenic T cells between NOD and congenic strain-NOD.H2^{h4}. Expression ratios of NOD DEGs in NOD (O) or NOD.H2^{h4} (×) T cells were compared to C57BL/6 T cells. T cells were hybridized with RNA from C57BL/6 T cells to microarray slides containing the 15 K NIA clone set.

These results show that, in NOD congenic strains, at least for T cells, replacement of the MHC region in NOD can lead to marked changes in gene expression profiles. This altered gene expression profile may result in altered functional properties of NOD.H2^{h4} T cells.

Discussion

In this study, we demonstrated that lymphocytes of NOD mice exhibit a higher percentage of apoptotic cells *in vivo*. Although defects in phagocytosis of apoptotic cells by macrophages in NOD mice may contribute to the increase in numbers of apoptotic cells (219), in this study, we showed that NOD T cells exhibit greater sensitivity to apoptosis-inducing stimuli in cell cultures than those of B6 and NOD.H2^{h4} strains. Thus we propose that increased sensitivity to apoptosis, at least in part, contributes to the increased numbers of apoptotic cells observed *in vivo*. Our gene expression profiling analysis identified a specific gene expression profile for NOD T cells, which implicated increased susceptibility to apoptosis as a potential defect in NOD T cells. Therefore, enhanced apoptosis in NOD T cells maybe a potential cause for lymphopenia in NOD mice. It has been discussed in autoimmune animal models that lymphopenia can be an underlying mechanism to generate autoimmune diseases (17, 20, 21). The elevated apoptosis of T cells in diabetic BB rats has been described, and in peripheral lymphoid organs of diabetic prone BB rats have severe T cell lymphopenia (21). The lymphopenia in NOD T cells induced compensatory homeostatic proliferation may generate autoimmunity (17).

Given our results, it is difficult to conclude that a difference in expression of a single gene or protein in NOD lymphocytes accounts for the observed differences in susceptibility to apoptosis. Rather, we find that a collection of genes and proteins that have known roles in inducing or preventing apoptosis are differentially expressed in NOD T cells, this specific gene expression profile may represent the specific trait of NOD T cells. Among these genes, one category is apoptosis-inducing genes and these

genes are over-expressed and include *Bnip3l* and *Pdcd7* (220, 221). In under expressed genes, the most over-represented category is heat shock proteins, how the heat shock proteins regulate the apoptosis has been reviewed(222). The heat shock protein Hsp90, one of the most over-represented genes in under-expressed NOD DEGs, is also an important regulator of apoptosis (215-217, 223, 224). More interesting, an inhibitor of apoptosis (IAP) family member-survivin(225), which was recently reported to be required for both proliferation and inhibition of apoptosis of expanding T cells(226), is also regulated by Hsp90 and global suppression of Hsp90 results in proteasomal degradation of survivin and apoptosis(227). It has also been suggested that Hsp105 protects microtubules and has anti-apoptotic function(228). Therefore, these proteins probably contribute to a generalized increase in sensitivity to apoptosis as we observed in apoptotic cells measurement in vivo and in vitro cell culture assay, which in turn, may contribute to the observed lymphopenia in NOD (17).

Bone marrow transplantation experiments revealed that the gene expression profile of T cells in NOD is probably an independent property of NOD T cells, not varying according to different host environment. Non-diabetic congenic strain NOD.H2^{h4} exhibit similar apoptosis pattern in vivo or in vitro as that of B6 mice, the expression of NOD DEGs in NOD.H2^{h4} also show the similar pattern as that of B6 mice. All these evidences suggested that the specific gene expression profile of T cells in NOD mice is an independent property that may contribute to establish the elevated apoptosis in NOD lymphocytes. In addition, the process of lymphopenia leading to autoimmunity may map to *Idd3* locus(17), however, the congenic mice-NOD.H2^{h4} with altered *Idd1* loci did not exhibit apoptosis prone characteristic, and NOD DEGs gene expression in NOD.H2^{h4}

was also modulated to control level. These results suggested that *Idd1* loci may also contribute to establish the susceptibility to apoptosis in NOD mice.

In summary, our study demonstrates that the NOD T cell unique gene expression profile confers a deregulated stress response system, especially in the heat shock protein family, that may increase susceptibility of NOD T cells to apoptosis under various environmental stresses, which may result in lymphopenia. Our results implicate that clinical intervention to the components of the stress response system, especially heat shock family, such as Hsp90 or Hsp105, may pave a new way to diagnosis, prevents and treatment to type I diabetes.

Part II: Comparison of gene expression profiles between human autoimmune diseases and murine models

Introduction

We have recently employed cDNA microarray techniques to screen gene expression profiles in human autoimmune disease and have found that each individual with one of four different autoimmune diseases shares a common profile (79). A dominant feature of these profiles is the under-expression of key genes that encode proteins required for apoptosis, inhibition of cell cycle progression, and inhibition of cell differentiation and migration (79). In general, critical genes that control biological processes are highly conserved among species. It has been argued that it is possible to identify these genes by comparing conserved biological processes among species. Therefore, we hypothesized that critical differentially expressed genes (DEGs) that may regulate autoimmune disease may be shared among human autoimmune diseases and murine models of spontaneous autoimmune disease.

To address this question we used cDNA microarray analysis to compare gene expression profiles in resting T cells among two autoimmune strains, NOD and NZM, and a non-autoimmune strain, C57BL/6. We asked two questions. First, we compared gene expression profiles between the two murine autoimmune strains to determine if they, like human autoimmune disease, shared overlapping profiles of gene expression. Second, we compared gene expression profiles in the murine models to the unique gene expression profiles observed in human autoimmune disease to identify genes that displayed common differential expression. Our results show that there is very little

overlap between gene expression profiles in T cells in human and murine autoimmune disease and between two different autoimmune murine models.

Materials and methods

Mice: NOD, NZM (2410), and C57BL/6 mice were obtained from Taconic (Germantown, NY 12526) and housed in the Vanderbilt University animal facilities. Female mice were used between 8-24 weeks of the age.

Reagents: Complete RPMI 1640 medium supplemented with 10% FBS (Lot ALK 14837; HyClone Laboratories, Logan, UT), 100 U/ml of penicillin, 100 μg/ml of streptomycin, 2mM L-glutamine, and 0.05 mM 2-ME (J. T. Baker, Phillipsburg, NJ) was used for cell culture. The mAbs used in this study were purified from tissue culture fluids of hybridoma cells obtained from American Type Culture Collection (Manassas, VA/Rockville, MD): anti-I-A (10-3.6.2, specific for I-A g7, k, r, f, or s haplotypes; Y3P, specific for I-A b, f, p, q, r, s, u and v haplotypes,). BioMag® Goat anti-Mouse IgG coated magnetic beads were obtained from QIAGEN (Valencia, CA). Tri Reagent was obtained from Molecular Research Center, INC. Superscript II transcriptase was obtained from Invitrogen (Carlsbad, CA). PCR purification kits were obtained from QIAGEN. CMT^{TM} hybridization chambers were obtained from CORNING (Corning, NY). Cy5, Cy3 and other reagents used for cDNA labeling and hybridization were obtained from the Vanderbilt Microarray Shared Resource. The Vanderbilt Microarray Shared Resource made NIA mouse 15K microarray chips.

NIA Mouse 15K cDNA clone set: The NIA mouse 15K cDNA clone set is a 15,247-element clone set. The 15K clone set was developed at the National Institutes of

Aging (NIA) and has been sequence verified by sequencing from both the 5' and 3' termini. Approximately 15,000 unique cDNA clones were derived from 52,374 ESTs from pre- and periimplantation embryos, E12.5 female gonad/mesonephros, and new born ovary (214). Every cDNA clone is assigned a unique clone ID in the mouse 15K NIA clone set.

$CD3^+$ T cells purification: $CD3^+$ T lymphocytes were purified from spleen by negative selection. Red blood cells were removed by hypotonic lysis. I-A expressing cells were incubated with anti-I-A mAb (10-3.6.2 for NOD, Y3P for NZM 2410 and C57BL/6) for 30 min at 0°-4°C. Labeled cells were depleted by incubation with goat-anti-mouse IgG coated magnetic beads at room temperature for 30 min. Cells bound to beads were removed by a magnet column. Average purity of $CD3^+$ T cells was approximately 90-95% as determined by flow cytometry.

Total RNA isolation: Tri Reagent was used to isolate total RNA from splenic T cells according to the manufacture's protocol. We pooled total RNA from the T cells of 2-3 mice for each microarray experiment. This provided us with sufficient total RNA and decreased bias due to biological variation. We used 5×10^7 cells to isolate total RNA for each group (experimental and control) for each microarray experiment.

Labeling of cDNA: Labeled cDNA was prepared from 30 μg aliquots of total RNA. RNA was transcribed into fluorescently labeled cDNA using an anchored oligo-dT primer ($dT_{20}MN$), unlabeled dNTPs, superscript II reverse transcriptase, and dCTP labeled with one of the two fluorophores, Cy5 or Cy3, according to the standard protocol of Vanderbilt Microarray Shared Resource. Cy5 was used to label the experimental samples (NOD or NZM); Cy3 was used to label the control samples (C57BL/6). The

Cy5 and Cy3 labeled cDNAs were combined in Buffer PB and purified together by using the PCR purification kit. Finally, the combined samples were dissolved in 2x hybridization buffer (50% formamide, 10x SSC and 0.2% SDS).

Microarray hybridization and data analysis: Microarrays were prehybridized in prehybridization solution (1% BSA, 5x SSC, and 0.1% SDS) at 55°C for 45 min. After prehybridization, arrays were rinsed 5 X in MilliQ water, 1X in isopropanol, and allowed to air dry. Combined samples were placed on the prehybridized slide and covered with a lifterslip. Microarrays were hybridized at 42°C for 14-16 hours in a humidified hybridization chamber. After hybridization, the lifterslip was removed and the microarray slide was washed at room temperature with serial 5-min washes with 2x SSC, 0.1%SDS, 1x SSC, and 0.1x SSC. Microarray slides were dried by centrifugation at 50 x g for 5 min. Microarray slides were scanned using a GenePix® 4000B scanner (Axon Instrument Inc., Union City, CA). Image files were analyzed with GenePix® Pro 3.0 software (Axon Instrument Inc., Union City, CA). The software package was used to identify individual clones from the microarray image and to calculate signal intensity values after normalization. Ratios (fluorescent intensity of experimental sample/ fluorescent intensity of control sample) from the two channels (red: Cy5, green: Cy3) represent the fold difference in gene expression levels between the two samples. Each microarray was performed at least in triplicate and results were averaged.

Definition of DEGs: We arbitrarily defined a DEG as a gene or clone that exhibited a greater than 3-fold difference in expression between two samples with a $P < 0.05$ (using the student's T test).

Comparison of human and murine mice microarray data: To identify overlapping DEGs in mouse and human datasets, we used human clone accession numbers to search the GeneCard^{TM} database (229) to identify corresponding human DEGs. We used the same database to determine if there was a known homologous murine gene. If so, we searched our mouse microarray dataset to find the corresponding genes. Expression levels of the corresponding murine genes were obtained from microarray datasets. If the gene was also a DEG in murine datasets and had the same expression pattern (under-expressed or over-expressed) as that in the human microarray dataset, we called this gene an overlapping DEG between human autoimmune disease and the corresponding murine autoimmune disease model.

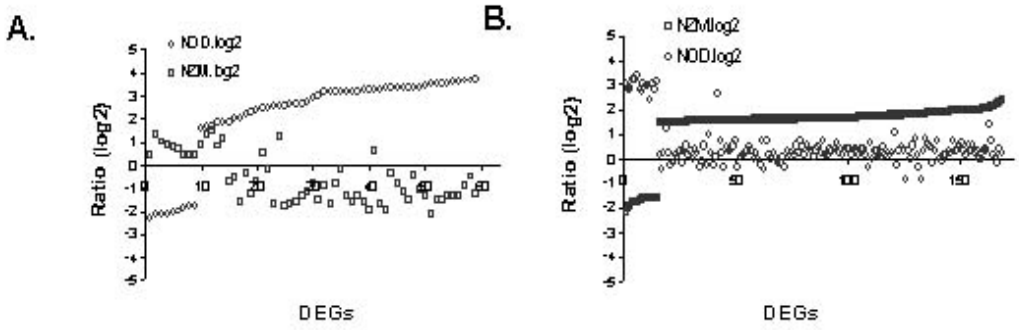
Gamma irradiation induced apoptosis: T cells were purified from the murine strains: NOD, NZM and C57BL/6, and were cultured in complete RPMI 1640 media with L-glutamine (2mM), FCS (fetal calf serum 10%) and 2-ME (0.2mM) in 6-well plates at a concentration 1×10^6 cells/ml. Cells were exposed to a total of 4 Gy of gamma irradiation using a Marker 1 Irradiator (J.L.Shepherd & Associates, Glendale, Calif.) at a dose rate of 1.897 Gy/min at room temperature. Cell viability was calculated by determining live cell numbers after gamma irradiation to non-irradiated controls.

Results

Gene expression profiles in T cells from autoimmune NOD and NZM strains

A striking feature of our microarray analyses of human autoimmune disease was that most DEGs were identical in each of four distinct autoimmune diseases, IDDM, SLE,

RA, and MS(79). Therefore, we wanted to determine if T cells from NOD and NZM strains shared common DEGs when compared to C57BL/6. We used microarrays to compare gene expression profiles between autoimmune-prone NOD, NZM T cells and non-autoimmune C57BL/6 T cells. First, we identified DEGs in the NOD versus C57BL/6 data set (Fig. 6-7 A).



C.

Common NOD & NZM DEGs

Acc#	Name	NOD	NZM
		Ratio (log2)	
BG071984	unknown	1.8	1.2
BG069482	c13orf1 homologue	1.7	1.6
BG072801	*S100A9*	1.6	1.3

Figure 6-7. Comparison of gene expression profiles in NOD and NZM T cells. **A.** NOD T cell DEGs were identified from our microarray data set. The corresponding expression levels of these genes in NZM T cells are shown. **B.** The reciprocal experiment shows NZM T cell DEGs and their corresponding expression levels in NOD T cells of NZM in NOD dataset. **C.** Overlapping DEGs between NOD and NZM data sets.

Next we checked their expression levels in the NZM data set to determine if NOD DEGs were also NZM DEGs. We performed the reciprocal analysis by identifying NZM DEGs and checking their expression level in the NOD data set. We found 59 DEGs in the NOD data set for which we could find corresponding expression data in NZM data set, and 172 DEGs in NZM data set with corresponding expression data in the NOD data set. But, in

contrast to the situation with different human autoimmune diseases, we only found three overlapping DEGs between the two autoimmune mouse models (Fig. 6-7 A, B and C).

Comparison of DEGs between murine autoimmune models and human autoimmune disease

From human public databases, we identified DEGs for each human autoimmune disease. There was considerable overlap in expression levels among this set of genes in each human autoimmune disease (Fig. 6-8 A).

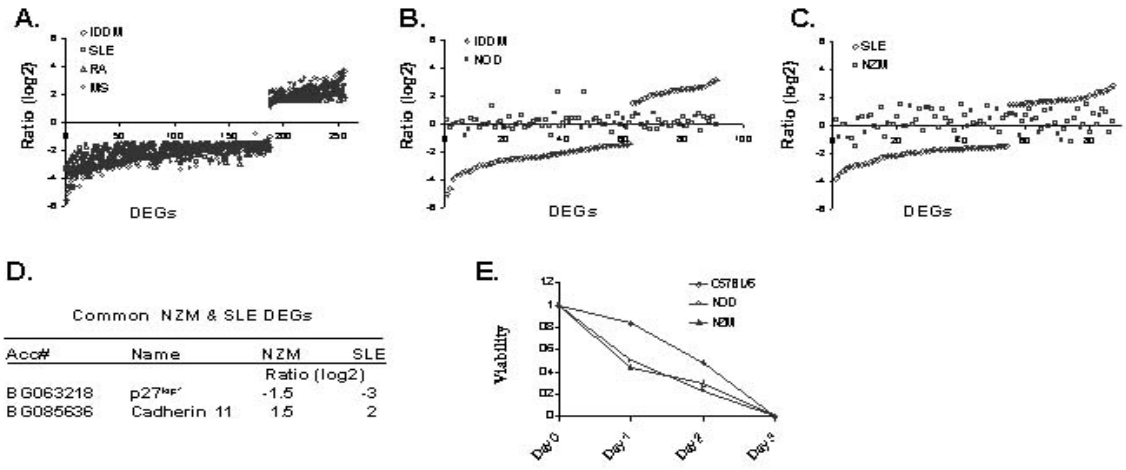


Figure 6-8. Comparison of gene expression profiles between human autoimmune diseases and murine models of autoimmune disease. **A.** Overlapping DEGs in four different human autoimmune diseases (IDDM, SLE, RA, MS). **B.** Expression profiles of human IDDM DEGs in NOD T cells. **C.** Expression profiles of human SLE DEGs in NZM T cells. **D.** Overlapping DEGs between the NZM mouse model and human SLE. **E.** Comparison of gamma irradiation induced cell death in NZM, NOD and B6 T cells.

Next, we searched public databases (229) for the murine homologues and examined our microarray database to determine if we had expression data of a corresponding gene or clone. We were able to identify 129 murine clones represented in our microarray dataset that were homologous to human autoimmune disease DEGs. Next, we determined if these DEGs were also differentially expressed in T cells from NOD and NZM (Fig. 6-8 B

& C). When we compared human IDDM with NOD mouse data, we did not find any overlapping DEGs between human and mouse data sets. We compared the DEGs in SLE to our NZM microarray data set and found two overlapping DEGs (Fig. 6-8 C & D).

Gamma-irradiation induced apoptosis

Analysis of DEGs in human autoimmune disease suggests that lymphocytes that have this gene expression signature should have defects in cell cycle control and apoptosis, especially p53-mediated damage response pathways (79). Gamma-irradiation induced cell death is known to be dependent upon p53. Lymphocytes from individuals with human autoimmune disease express very low levels of basal *TP53* mRNA and p53 protein (79, 230). They also fail to undergo gamma-irradiation induced cell death (230). Therefore, we wanted to determine if lymphocytes from autoimmune NOD and NZM (2410) strains also exhibited defects in gamma-irradiation induced cell death when compared to non-autoimmune strains. T cells from NOD and NZM murine strains did not show noticeable resistance to gamma-irradiation induced apoptosis compared with non-autoimmune strains (Figure 6-8 E). This result is consistent with our microarray data that demonstrated that expression levels of *Trp53* were equivalent in T cells from autoimmune and non-autoimmune strains.

Discussion

We have found a conserved gene expression profile in the lymphocytes of humans with autoimmune disease. This profile is also seen in unaffected first-degree relatives arguing that it arises, at least in part, from genetic factors rather than the disease process

(79, 178). We had two goals in this study. First, we wanted to determine the extent to which this conserved gene expression pattern was conserved in murine models of spontaneous autoimmune disease. Second, we wanted to determine the extent to which strains with distinct autoimmune diseases shared common gene expression profiles. We did not find overlapping DEGs in T cells between human autoimmune disease and NOD mice. We only found 2 overlapping DEGs in T cells between human autoimmune disease and NZM 2410 mice, and we found 3 overlapping DEGs in T cells between NOD and NZM mice. After gamma irradiation, a p53 dependent response, T cells from human autoimmune patients have strong resistance to gamma irradiation induced apoptosis (230) and T cells from murine autoimmune and non-autoimmune strains show similar sensitivity to gamma-irradiation induced cell death. Therefore, these autoimmune models may not perfectly match their corresponding human autoimmune diseases.

The NOD mice we used are at the age of 8-12 weeks old. We know that the NOD mouse exhibits autoimmune pancreatic insulitis at the age of four weeks (231) and insulin-dependent diabetes is seen in females as early as 12 weeks of age (142). Most female NOD mice at 20-24 weeks of age have elevated blood glucose levels. In fact, we compared gene expression profiles of different age NOD mice to determine if they change as the autoimmunity progresses from recognition of self-antigen to insulitis to diabetes. To do so, we performed microarray analysis of T cells from age-matched mice of 4-6 week, 8-12 week and 20-24 week of age. In this experiment, we did not observe differences in gene expression profiles between older and younger mice. Most of the DEGs among three different age mice showed the same gene expression pattern. These

results suggest that differences in gene expression profiles in the total T cell population do not change dramatically during autoimmune disease progression.

NZM (2410) mice develop glomerulonephritis by one year of age in ~80% of males and females. Anti-dsDNA IgG antibodies are also present in the sera of ~80% of animals by six months of age (61). The mice we used for our analysis were 8-15 weeks old. We do not exactly know if the gene expression profiles in NZM (2410) T cells will change with disease progression. However, based upon our studies in NOD T cells, it is reasonable to suspect that the gene expression profile in NZM (2410) T cells will not change with disease progression. In addition, the gene expression profile observed in humans with autoimmune disease is also observed in unaffected first-degree relatives of affected individuals suggesting that this gene expression profile may not reflect disease activity.

Since we did not find overlapping DEGs between human autoimmune diseases and their murine counterparts, we searched for murine DEGs that may confer a similar cellular phenotype. In fact, our microarray data support this notion. It has been described elsewhere that there are multiple apoptosis defects in animal models and human autoimmune diseases (232-235). Many gene expression differences in lymphocytes from human autoimmune disease patients predict defects in cell cycle control and apoptosis. Although we did not see similar DEGs in cell cycle control and apoptosis in NOD T cells, we found that one gene, Ptpn13, which encodes the FAP-1 protein, is over-expressed. This protein inhibits Fas-mediated apoptosis (236-239). Therefore, these results may indicate that physiological defects in apoptosis are a

common property of both human autoimmune disease and murine models of autoimmune disease.

When we compared NZM data with SLE data, we found two overlapping DEGs. One was the down-regulated gene *Cdkn1b*, which encodes $p27^{kip1}$ that is a cyclin-dependent kinase inhibitor that is a regulator of cell cycle progression. $p27^{kip1}$ is also an essential anergy factor for the blockade of clonal expansion of alloreactive human and mouse helper T lymphocytes through inhibiting interleukin 2 transcription (240). And it is described elsewhere that $p27^{kip1}$ is down regulated in an autoimmune disease-Hashimoto's thyroiditis (HT) and its down regulation is independent of the proliferative status and of changes in the proteins it regulates, Skp2 and cyclin D3 (241). $p27^{kip1}$ also has a redundant function with p21 (242-244). p21 is indispensable for maintaining tolerance and preventing a lupus-like syndrome in female mice, as p21 knockout mice develop severe lupus-like disease (157). We do not have gene expression data from our NIA mouse 15K cDNA clone set or human microarray dataset for the gene that encodes p21 (*Cdkn1a*) but we do know that basal and inducible levels of p21 are depressed in lymphocytes from patients with autoimmune disease (79). Taken together, it is likely that the defects in expression of both cyclin kinase inhibitors, p21 and p27, contribute to the pathogenesis of human autoimmune disease, including SLE, and lupus-related pathogenesis observed in the NZM model. One major difference between the two species is that defects in the human can be linked to defects in the p53 damage response pathway while defects in the mouse occur in the presence of an intact p53 damage response pathway.

In summary, our results show that there is not a significant overlap between DEGs in T cells from human autoimmune disease and murine models of autoimmune disease. In contrast to human autoimmune disease, there is also not significant overlap between DEGs from two distinct murine autoimmune diseases. However, the overlap that does exist may suggest that defects in conserved pathways produce a functional liability that may contribute to development of autoimmune disease. For example, defects in apoptosis are predicted from our analysis of human autoimmune disease and these have been confirmed experimentally (230). Our findings suggest that NOD T cells may also have a defect in apoptosis but this may arise from increased expression of an inhibitor of apoptosis, *Ptpn13*, rather than a defect in expression of a gene that is required for apoptosis. Our analysis of NZM T cells shows a more direct parallel. Defects in expression of cyclin kinase inhibitors are seen in T cells from both human autoimmune disease and murine models including SLE and NZM. Other data indicate that these inhibitors of cell cycle progression are required to establish a state of tolerance or anergy and their absence leads to development of sex-linked lupus in mice (157).

Our results indicate that murine models do not perfectly model their corresponding human autoimmune diseases when gene expression profiles are considered. Other investigators have previously recognized limitations of using rodent models to study human diseases. Failure of DNA vaccine technology in humans demonstrates that success in murine models does not necessarily translate to success in human trials. Study of murine models has benefited our understanding of basic principles of immunology research. However, using one inbred, engineered species to study the diseases of outbred humans is a big limitation. It has been suggested that large animals may better model

human diseases (245). There is relative agreement that type 1 diabetes in humans and spontaneous type 1 diabetes-like syndromes in animal models are autoimmune disorders. But, interventions to prevent diabetes based on mice research has not been successfully transferred to humans (246). Our results further confirmed that autoimmune murine models do not perfectly model their corresponding human autoimmune disease at least in the perspective of gene expression profile. But they do have some similar phenotypes that may result from conserved pathological pathways.

CHAPTER VII

GENERAL DISCUSSION AND CONCLUSION REMARKS

Gene expression profiling with microarray can screen expression levels of thousands of genes simultaneously and help identify comprehensive molecular signatures of certain states of disease. Therefore, it is very powerful tool to study multigenic diseases, such autoimmune and atopic diseases. The work of this thesis tests the possibility of performing gene expression profiling using peripheral blood mononuclear cells (PBMC) as source to characterize diseases. The application of gene expression profiling can be basically divided into three objectives-classifying disease phenotypes, monitoring or predicting responses to treatment, and predicting disease outcome. In this chapter, I will use the following flowchart (Fig. 7-1) as a simplified outline to discuss the relevant points of this thesis and important implications of our work.

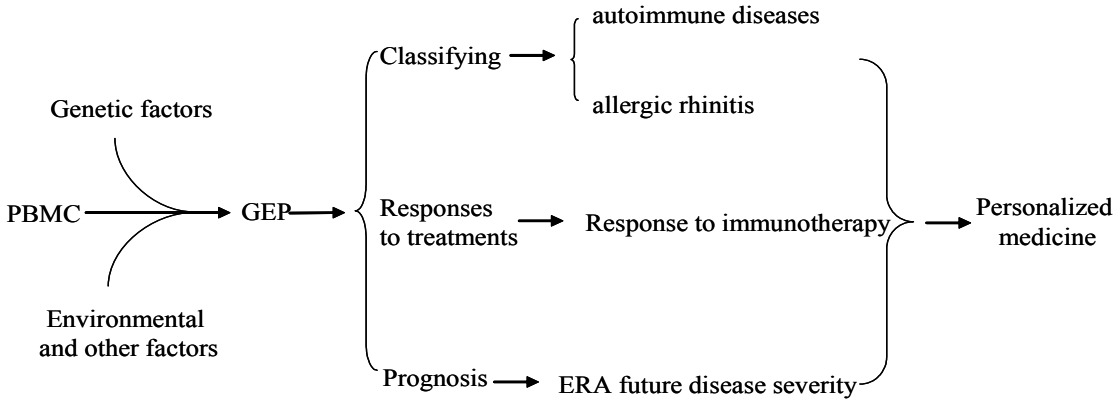


Figure 7-1. Flow chart of from gene expression profiling to personalized medicine, GEP, gene expression profiling.

Classifying disease types or subtypes is an objective of gene expression profiling. Previously, we identified a common gene expression profile that can distinguish individuals with autoimmune disease from the healthy control group (79). Unaffected first-degree relatives of individuals with autoimmune disease share a portion of this profile suggesting it is of genetic origin independent of disease onset (120). Therefore, genetic factors and other factors such as environmental factors and regulatory events may combine to establish this disease specific gene expression profile. In chapter III, we identified a common gene expression profile which distinguished a cohort of autoimmune individuals from both the healthy control group and the unaffected family member group, thus, this gene expression profile is common feature of autoimmune disease and is without influence of genetic components. Gene expression profiles specific for a particular disease or a specific state of disease have also been demonstrated, such as the unique gene expression profile in MS (149) which can distinguish individuals with MS from healthy controls, and a unique gene expression profile in early RA (about one year disease duration) which can distinguish early RA from established RA (about 10 years disease duration). Recently, an interferon signature in SLE was identified by several groups (83, 89, 90, 247). The expression profile of these interferon signature genes seems specific for SLE and may be correlated with SLE disease activity (83), but is not a common gene expression profile for autoimmune disease as we identified in chapter III, and they can not be found in other autoimmune disease-RA (247).

The common autoimmune disease expression profile identified in chapter III is shared by individuals with both early RA (about one year disease duration) and established RA (about 10 years disease duration), this result implies that this common

autoimmune diseases expression profile is acquired shortly after disease onset and is sustained independent of disease duration. How this specific gene expression profile is maintained is not fully understood. One possible explanation is that lymphocytes in blood are continuously exposed to cytokines as a result of the chronic inflammation and the unique gene expression signature is a product of cytokine stimulation. In fact, expression of a unique gene expression signature in SLE is attributed to exposure to interferon (87, 88). Our results do not rule out this possibility. Presence of a common infectious agent could also explain this altered gene expression profile. Exhaustive searches for infectious agents in autoimmune disease have failed to confirm this possibility. However, our results do not rule it out either. Somatic mutation of genomic DNA may cause a defect in some master transcriptional regulator resulting in variation of downstream gene expression levels. This seems less likely since, unlike tumors, lymphocytes continuously die and are replenished in the periphery making it unlikely that a single or small number of somatic mutations could affect the majority of circulating lymphocytes. Another mechanism of maintaining a specific gene expression profile may not involve DNA mutation, but rather alteration of gene transcription scaffold-chromatin structure under certain environmental conditions. Examples include histone modifications or methylation of CpG DNA (248-250). These epigenetic modifications have been tied to alterations in gene expression inherited by daughter cells. Although it is generally thought that specific mechanisms sustain this memory or imprinting, these mechanisms are not clearly understood. Nevertheless, our results provide a framework to permit more careful examination of the precise mechanisms involved in sustaining altered gene expression profiles in individuals with either autoimmune or atopic diseases.

In parallel to gene expression profiling of PBMC in human autoimmune diseases, we also performed gene expression profiling in T cells of a human type I diabetes murine model- NOD (chapter VI). We identified NOD signature genes-NOD DEGs, which are differentially expressed, compared to B6 T cells. The most over-represented over-expressed NOD DEGs are apoptosis inducing genes, and strikingly, almost all of the over-represented under-expressed genes are heat shock family genes including *HSP90* which is one of the most under-expressed genes in NOD T cells and is thought to be a master regulator of apoptosis (215-217, 223, 224) and may be involved in regulating T cell apoptosis (226, 227). Therefore, we postulate that NOD T cells may have a de-regulated stress response system making NOD T cells more susceptible to apoptosis. The difference in expression of NOD DEGs may result from the NOD specific host environment since the NOD mouse is progressing toward a diabetic status resulting from decreased insulin levels. Therefore, this complex environment may cause unknown effects on expression levels of certain genes such as heat shock family genes. However, the bone marrow transplantation experiments demonstrate that the differential expression pattern is independent of host environment, therefore, the differentially expressed NOD DEGs is rather an intrinsic trait than the result from NOD host environment. In fact, the in vitro cell culture experiments also support this idea, since in vitro cultured NOD T cells exhibit less resistance to gamma irradiation or heat shock induced apoptosis than that of B6 T cells cultured under the same conditions. Therefore, the NOD T cells are susceptible to apoptosis under environmental stress and it may cause lymphopenia in NOD mouse, and lymphopenia driven NOD T cells homeostatic proliferation may result in diabetes (17, 159, 160). When we compared the NOD signature genes and human

autoimmune disease signature genes, we found very little overlap between differential gene expression profiles in T cells from NOD mouse disease and people with autoimmune disease. However, among the human autoimmune signature genes, *BRCA1* and *TP53* are substantially underexpressed, this may also result in chronic lymphopenia in autoimmune individuals (chapter III, discussion), and similarly, lymphopenia may finally contribute to onset of autoimmunity in these patients. The NOD mouse and human autoimmune diseases do not have a common gene expression signature, but may have a common outcome resulting from these differentially expressed signature genes-lymphopenia, which may be a potential contributor to establishing autoimmunity.

We have described a common gene expression profile of autoimmune disease in chapter III. We wanted to further test the possibility if we can use gene expression profiling to identify specific gene expression profiles in PBMC in another immune-mediated disease-allergic rhinitis. In contrast to chapter III in which we identified a common signature for four autoimmune diseases, in chapter IV, we only examined a single atopic disease, allergic rhinitis. However, the expression profile of these signature genes in allergic rhinitis was completely distinguished from the healthy control group and the autoimmune disease group, indicating the presence of a unique signature in allergic rhinitis. Further work will have to be performed to identify the genetic and disease-dependent portions of the atopic gene expression profile as we did in chapter III.

Identifying the gene expression profile that can be used to monitor or predict individuals' response to certain treatments is one of the objectives of gene expression profiling analysis. In allergic rhinitis, we demonstrated that gene expression profiling can be used to monitor responses to immunotherapy. Key findings are that immunotherapy

does not affect the entire signature gene expression profile for allergic rhinitis. Rather, only a subset of genes alters expression levels after therapy. Nevertheless, these results demonstrate that the goal of using gene expression profiling in PBMC to monitor responses to therapies in human disease is feasible.

In addition to classifying diseases and monitoring response to treatments, another objective of gene expression profiling is to predict disease course. With the assistance of supervised methods, we identified a gene expression signature in early RA capable of predicting future disease severity. Although it was not possible to replicate these studies with a second population to validate our results, they suggest it may be possible to predict disease course with gene expression profiling in PBMC thus aiding the physician in the clinical management of their patients. To our knowledge, this is the first gene expression profiling study to identify prognostic factors in rheumatoid arthritis research (chapter V).

One of the most important implications of our study is that using PBMC as the source to perform gene expression profiling to characterize diseases is feasible, even though the blood tissue is not the major site affected by the disease. Gene expression profiling has been performed with different tissue source; most of them are tissues of the major site where the diseases originate or directly affected by the diseases, such as microarray study in cancers using tissue samples at the site of the tumor (86, 102-104, 108, 109); microarray studies in autoimmune diseases using kidney tissues of SLE (251), brain tissues of MS (81, 252), synovial tissues of RA (92, 93); microarray studies in atopic diseases using nasal tissues (96, 253, 254). Our results indicate that PBMC is a reliable source for gene expression profiling; even when PBMC are not the site that is directly affected by the diseases, such as the autoimmune diseases and atopic diseases. In

fact, affected tissues used for microarray are generally only in small size that can not generate sufficient RNA for microarray study, and the positions to obtain the tissues may generate more variation compared to PBMC, because the changes of positions may result in variation of cell populations which can cause greater variations in microarray study. In contrast, current protocols can generate enough RNA from PBMC and also the current protocol for isolating the PBMC from the whole blood can greatly decrease the contamination from other sources. Also, PBMC is a very easily accessible human tissue sample, as for patients, surgically removing a piece of affected tissues for microarray study may not be as easily accepted as collecting a tube of blood. However, PBMC as common source for diseases profiling needs further evaluation in other diseases such as certain cancers whose major affected sites are not blood.

With the advance of technology, we can imagine that a tube of blood or even a drop of blood from patient can be used to extract comprehensive gene expression profiles, with referencing to the recorded phenotype signature profiles database for the diseases, we can determine individual's disease phenotype, subtype, reveal the diseases progression, monitor the responses to different therapies, design novel personalized therapies or drugs, and evaluate the disease outcomes. Further, we can evaluate the liability of diseases in healthy individuals and help prevent the onset of disease. All of these will radically changed modern medicine: the prescriptions based upon individuals' symptoms will be changed to be based upon personal genomic information; the clinical treatments will be changed to be tailored to pre-clinical prevention. This personalized medicine will revolutionize modern medicine and advance human health care to a bright new age.

LIST OF PUBLICATIONS

1. ***Liu, Z.****, K. Maas, and T. M. Aune.* 2004. "Comparison of differentially expressed genes in T lymphocytes between human autoimmune disease and murine models of autoimmune disease". Clinical Immunology 112:225-230.

2. ***Zheng Liu****, Robert W. Yelverton, Brian Kraft, S. Bobo Tanner, Nancy J. Olsen and Thomas M. Aune.* 2005. "Highly Conserved Gene Expression Profiles in Humans with Allergic Rhinitis Altered by Immunotherapy". Clinical and Experimental Allergy. 35:1581-90.

3. ***Zheng Liu,*** *Kevin Maas and Thomas M. Aune*. 2006 "Identification of Gene Expression Signatures in Autoimmune Disease without the Influence of Familial Resemblance". Human Molecular Genetics. 15:501-509.

4. ***Zheng Liu*** *and Thomas M. Aune*. 2006. "De-regulated stress system in lymphocyte of NOD mice may associate with NOD lymphopenia". Genes and Immunity *In press*.

5. ***Zheng Liu,*** *Tuulikki Sokka, Kevin Maas, Nancy J. Olsen and Thomas M. Aune.* 2006. "Prediction of Disease Severity in Patients with Early Rheumatoid Arthritis by Gene Expression Profiling". Annals of the Rheumatic Diseases. *Submitted.*

REFERENCES

1. Marrack, P., Kappler, J. and Kotzin, B.L. (2001) Autoimmune disease: why and where it occurs. *Nat Med,* **7,** 899-905.

2. Wandstrat, A. and Wakeland, E. (2001) The genetics of complex autoimmune diseases: non-MHC susceptibility genes. *Nat Immunol,* **2,** 802-9.

3. (2002) *Immunology of diabetes: autoimmune mechanisms and the prevention and cure of type 1 diabetes*. GYAT/PCP, New York.

4. Wucherpfennig, K.W. and Eisenbarth, G.S. (2001) Type 1 diabetes. *Nat Immunol,* **2,** 767-8.

5. Steinman, L. (1996) Multiple sclerosis: a coordinated immunological attack against myelin in the central nervous system. *Cell,* **85,** 299-302.

6. Steinman, L. (2001) Multiple sclerosis: a two-stage disease. *Nat Immunol,* **2,** 762-4.

7. Kavanaugh, A.F.L., P. E. (1996) Clinical Immunology: Principles and Practice. In, Mosby, St. Louis, pp. 1093-1115.

8. Feldmann, M. (2001) Pathogenesis of arthritis: recent research progress. *Nat Immunol,* **2,** 771-3.

9. Kotzin, B.L. (1996) Systemic lupus erythematosus. *Cell,* **85,** 303-6.

10. Lipsky, P.E. (2001) Systemic lupus erythematosus: an autoimmune disease of B cell hyperactivity. *Nat Immunol,* **2,** 764-6.

11. Ohashi, P.S. (2003) Negative selection and autoimmunity. *Curr Opin Immunol,* **15,** 668-76.

12. Venanzi, E.S., Benoist, C. and Mathis, D. (2004) Good riddance: Thymocyte clonal deletion prevents autoimmunity. *Curr Opin Immunol,* **16,** 197-202.

13. Walker, L.S. and Abbas, A.K. (2002) The enemy within: keeping self-reactive T cells at bay in the periphery. *Nat Rev Immunol,* **2,** 11-9.

14. Mathis, D. and Benoist, C. (2004) Back to central tolerance. *Immunity,* **20,** 509-16.

15. Schwartz, R.H. (2003) T cell anergy. *Annu Rev Immunol,* **21,** 305-34.

16. Kronenberg, M. and Rudensky, A. (2005) Regulation of immunity by self-reactive T cells. *Nature,* **435,** 598-604.

17. King, C., Ilic, A., Koelsch, K. and Sarvetnick, N. (2004) Homeostatic expansion of T cells during immune insufficiency generates autoimmunity. *Cell,* **117,** 265-77.

18. Goodnow, C.C., Sprent, J., Fazekas de St Groth, B. and Vinuesa, C.G. (2005) Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature,* **435,** 590-7.

19. Coles, A.J., Wing, M., Smith, S., Coraddu, F., Greer, S., Taylor, C., Weetman, A., Hale, G., Chatterjee, V.K., Waldmann, H. *et al.* (1999) Pulsed monoclonal antibody treatment and autoimmune thyroid disease in multiple sclerosis. *Lancet,* **354,** 1691-5.

20. Hernandez-Hoyos, G., Joseph, S., Miller, N.G. and Butcher, G.W. (1999) The lymphopenia mutation of the BB rat causes inappropriate apoptosis of mature thymocytes. *Eur J Immunol,* **29,** 1832-41.

21. Jung, C.G., Kamiyama, T. and Agui, T. (1999) Elevated apoptosis of peripheral T lymphocytes in diabetic BB rats. *Immunology,* **98,** 590-4.

22. Hornum, L., Romer, J. and Markholst, H. (2002) The diabetes-prone BB rat carries a frameshift mutation in Ian4, a positional candidate of Iddm1. *Diabetes,* **51,** 1972-9.

23. Gross, D.M., Forsthuber, T., Tary-Lehmann, M., Etling, C., Ito, K., Nagy, Z.A., Field, J.A., Steere, A.C. and Huber, B.T. (1998) Identification of LFA-1 as a candidate autoantigen in treatment-resistant Lyme arthritis. *Science,* **281,** 703-6.

24. James, J.A. and Harley, J.B. (1998) B-cell epitope spreading in autoimmunity. *Immunol Rev,* **164,** 185-200.

25. Chiller, J.M., Skidmore, B.J., Morrison, D.C. and Weigle, W.O. (1973) Relationship of the structure of bacterial lipopolysaccharides to its function in mitogenesis and adjuvanticity. *Proc Natl Acad Sci U S A,* **70,** 2129-33.

26. Kearney, E.R., Pape, K.A., Loh, D.Y. and Jenkins, M.K. (1994) Visualization of peptide-specific T cell immunity and peripheral tolerance induction in vivo. *Immunity,* **1,** 327-39.

27. Vella, A.T., Mitchell, T., Groth, B., Linsley, P.S., Green, J.M., Thompson, C.B., Kappler, J.W. and Marrack, P. (1997) CD28 engagement and proinflammatory cytokines contribute to T cell expansion and long-term survival in vivo. *J Immunol,* **158,** 4714-20.

28. Tanaka, H., Demeure, C.E., Rubio, M., Delespesse, G. and Sarfati, M. (2000) Human monocyte-derived dendritic cells induce naive T cell differentiation into T helper cell type 2 (Th2) or Th1/Th2 effectors. Role of stimulator/responder ratio. *J Exp Med,* **192,** 405-12.

29. Silman, A.J., MacGregor, A.J., Thomson, W., Holligan, S., Carthy, D., Farhan, A. and Ollier, W.E. (1993) Twin concordance rates for rheumatoid arthritis: results from a nationwide study. *Br J Rheumatol,* **32,** 903-7.

30. Redondo, M.J., Rewers, M., Yu, L., Garg, S., Pilcher, C.C., Elliott, R.B. and Eisenbarth, G.S. (1999) Genetic determination of islet cell autoimmunity in monozygotic twin, dizygotic twin, and non-twin siblings of patients with type 1 diabetes: prospective twin study. *Bmj,* **318,** 698-702.

31. Kalman, B. and Lublin, F.D. (1999) The genetics of multiple sclerosis. A review. *Biomed Pharmacother,* **53,** 358-70.

32. Rioux, J.D. and Abbas, A.K. (2005) Paths to understanding the genetic basis of autoimmune disease. *Nature,* **435,** 584-9.

33. Bjorses, P., Aaltonen, J., Horelli-Kuitunen, N., Yaspo, M.L. and Peltonen, L. (1998) Gene defect behind APECED: a new clue to autoimmunity. *Hum Mol Genet,* **7,** 1547-53.

34. Nagata, S. and Suda, T. (1995) Fas and Fas ligand: lpr and gld mutations. *Immunol Today,* **16,** 39-43.

35. Fischer, A., Rieux-Laucat, F. and Le Deist, F. (2000) Autoimmune lymphoproliferative syndromes (ALPS): models for the study of peripheral tolerance. *Rev Immunogenet,* **2,** 52-60.

36. Ghosh, S., Palmer, S.M., Rodrigues, N.R., Cordell, H.J., Hearne, C.M., Cornall, R.J., Prins, J.B., McShane, P., Lathrop, G.M., Peterson, L.B. *et al.* (1993) Polygenic control of autoimmune diabetes in nonobese diabetic mice. *Nat Genet,* **4,** 404-9.

37. Becker, K.G., Simon, R.M., Bailey-Wilson, J.E., Freidlin, B., Biddison, W.E., McFarland, H.F. and Trent, J.M. (1998) Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. *Proc Natl Acad Sci U S A,* **95,** 9979-84.

38. Bergsteinsdottir, K., Yang, H.T., Pettersson, U. and Holmdahl, R. (2000) Evidence for common autoimmune disease genes controlling onset, severity, and chronicity based on experimental models for multiple sclerosis and rheumatoid arthritis. *J Immunol,* **164,** 1564-8.

39. Todd, J.A., Aitman, T.J., Cornall, R.J., Ghosh, S., Hall, J.R., Hearne, C.M., Knight, A.M., Love, J.M., McAleer, M.A., Prins, J.B. *et al.* (1991) Genetic analysis of autoimmune type 1 diabetes mellitus in mice. *Nature,* **351,** 542-7.

40. Jawaheer, D., Seldin, M.F., Amos, C.I., Chen, W.V., Shigeta, R., Monteiro, J., Kern, M., Criswell, L.A., Albani, S., Nelson, J.L. *et al.* (2001) A genomewide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. *Am J Hum Genet,* **68,** 927-36.

41. Vyse, T.J. and Todd, J.A. (1996) Genetic analysis of autoimmune disease. *Cell,* **85,** 311-8.

42. Theofilopoulos, A.N. and Kono, D.H. (1999) The genes of systemic autoimmunity. *Proc Assoc Am Physicians,* **111,** 228-40.

43. Griffiths, M.M., Encinas, J.A., Remmers, E.F., Kuchroo, V.K. and Wilder, R.L. (1999) Mapping autoimmunity genes. *Curr Opin Immunol,* **11,** 689-700.

44. Encinas, J.A. and Kuchroo, V.K. (2000) Mapping and identification of autoimmunity genes. *Curr Opin Immunol,* **12,** 691-7.

45. Falconer, D.S. (1989) *Introduction to Quantitative Genetics.* Third edn ed. John Wiley & Sons, New York, NY.

46. (1997) Genes for asthma? An analysis of the European Community Respiratory Health Survey. *Am J Respir Crit Care Med,* **156,** 1773-80.

47. Czene, K., Lichtenstein, P. and Hemminki, K. (2002) Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer,* **99,** 260-6.

48. Meyer, J.M. and Breitner, J.C. (1998) Multiple threshold model for the onset of Alzheimer's disease in the NAS-NRC twin panel. *Am J Med Genet,* **81,** 92-7.

49. Makino, S., Kunimoto, K., Muraoka, Y., Mizushima, Y., Katagiri, K. and Tochino, Y. (1980) Breeding of a non-obese, diabetic strain of mice. *Jikken Dobutsu,* **29,** 1-13.

50. Anderson, M.S. and Bluestone, J.A. (2005) The NOD mouse: a model of immune dysregulation. *Annu Rev Immunol,* **23,** 447-85.

51. Bach, J.F. (1994) Insulin-dependent diabetes mellitus as an autoimmune disease. *Endocr Rev,* **15,** 516-42.

52. Singh, B. and Rabinovitch, A. (1993) Influence of microbial agents on the development and prevention of autoimmune diabetes. *Autoimmunity,* **15,** 209-13.

53. Bach, J.F. (2002) The effect of infections on susceptibility to autoimmune and allergic diseases. *N Engl J Med,* **347,** 911-20.

54. Wicker, L.S., Todd, J.A. and Peterson, L.B. (1995) Genetic control of autoimmune diabetes in the NOD mouse. *Annu Rev Immunol,* **13,** 179-200.

55. Atkinson, M.A. and Leiter, E.H. (1999) The NOD mouse model of type 1 diabetes: as good as it gets? *Nat Med,* **5,** 601-4.

56. Miyazaki, T., Uno, M., Uehira, M., Kikutani, H., Kishimoto, T., Kimoto, M., Nishimoto, H., Miyazaki, J. and Yamamura, K. (1990) Direct evidence for the contribution of the unique I-ANOD to the development of insulitis in non-obese diabetic mice. *Nature,* **345,** 722-4.

57. Ridgway, W.M., Ito, H., Fasso, M., Yu, C. and Fathman, C.G. (1998) Analysis of the role of variation of major histocompatibility complex class II expression on nonobese diabetic (NOD) peripheral T cell response. *J Exp Med,* **188,** 2267-75.

58. Wicker, L.S., Appel, M.C., Dotta, F., Pressey, A., Miller, B.J., DeLarato, N.H., Fischer, P.A., Boltz, R.C., Jr. and Peterson, L.B. (1992) Autoimmune syndromes in major histocompatibility complex (MHC) congenic strains of nonobese diabetic (NOD) mice. The NOD MHC is dominant for insulitis and cyclophosphamide-induced diabetes. *J Exp Med,* **176,** 67-77.

59. Wicker, L.S. (1997) Major histocompatibility complex-linked control of autoimmunity. *J Exp Med,* **186,** 973-5.

60. Tisch, R. and McDevitt, H. (1996) Insulin-dependent diabetes mellitus. *Cell,* **85,** 291-7.

61. Rudofsky, U.H., Evans, B.D., Balaban, S.L., Mottironi, V.D. and Gabrielsen, A.E. (1993) Differences in expression of lupus nephritis in New Zealand mixed H-2z homozygous inbred strains of mice derived from New Zealand black and New Zealand white mice. Origins and initial characterization. *Lab Invest,* **68,** 419-26.

62. Abbas, A.K., Lichtman, A.H. and Pober, J.S. (2000) *Cellular and Molecular Immunology.* Fourth Edition ed. W.B. SAUNDERS COMPANY, Philadelphia London New York St. Louis Sydney Toronto.

63. Ono, S.J. (2000) Molecular genetics of allergic diseases. *Annu Rev Immunol,* **18,** 347-66.

64. Anderson, G.G. and Morrison, J.F. (1998) Molecular biology and genetics of allergy and asthma. *Arch Dis Child,* **78,** 488-96.

65. Kleeberger, S.R. and Peden, D. (2005) Gene-environment interactions in asthma and other respiratory diseases. *Annu Rev Med,* **56,** 383-400.

66. Sibbald, B., Horn, M.E., Brain, E.A. and Gregg, I. (1980) Genetic factors in childhood asthma. *Thorax,* **35,** 671-4.

67. Luoma, R., Koivikko, A. and Viander, M. (1983) Development of asthma, allergic rhinitis and atopic dermatitis by the age of five years. A prospective study of 543 newborns. *Allergy,* **38,** 339-46.

68. Taylor, B. and Broom, B.C. (1981) Atopy in medical students. *Ann Allergy,* **47,** 197-9.

69. Hopp, R.J., Bewtra, A.K., Watt, G.D., Nair, N.M. and Townley, R.G. (1984) Genetic analysis of allergic disease in twins. *J Allergy Clin Immunol,* **73,** 265-70.

70. Nafstad, P., Oie, L., Mehl, R., Gaarder, P.I., Lodrup-Carlsen, K.C., Botten, G., Magnus, P. and Jaakkola, J.J. (1998) Residential dampness problems and symptoms and signs of bronchial obstruction in young Norwegian children. *Am J Respir Crit Care Med,* **157,** 410-4.

71. Platts-Mills, T.A. and Carter, M.C. (1997) Asthma and indoor exposure to allergens. *N Engl J Med,* **336,** 1382-4.

72. McConnell, R., Berhane, K., Gilliland, F., London, S.J., Islam, T., Gauderman, W.J., Avol, E., Margolis, H.G. and Peters, J.M. (2002) Asthma in exercising children exposed to ozone: a cohort study. *Lancet,* **359,** 386-91.

73. Holgate, S.T. and Broide, D. (2003) New targets for allergic rhinitis--a disease of civilization. *Nat Rev Drug Discov,* **2,** 902-14.

74. Wachholz, P.A. and Durham, S.R. (2004) Mechanisms of immunotherapy: IgG revisited. *Curr Opin Allergy Clin Immunol,* **4,** 313-8.

75. Ebner, C., Siemann, U., Bohle, B., Willheim, M., Wiedermann, U., Schenk, S., Klotz, F., Ebner, H., Kraft, D. and Scheiner, O. (1997) Immunological changes during specific immunotherapy of grass pollen allergy: reduced lymphoproliferative responses to allergen and shift from TH2 to TH1 in T-cell clones specific for Phl p 1, a major grass pollen allergen. *Clin Exp Allergy,* **27,** 1007-15.

76. Secrist, H., Chelen, C.J., Wen, Y., Marshall, J.D. and Umetsu, D.T. (1993) Allergen immunotherapy decreases interleukin 4 production in CD4+ T cells from allergic individuals. *J Exp Med,* **178,** 2123-30.

77. Till, S.J., Francis, J.N., Nouri-Aria, K. and Durham, S.R. (2004) Mechanisms of immunotherapy. *J Allergy Clin Immunol,* **113,** 1025-34; quiz 1035.

78. Hawrylowicz, C.M. and O'Garra, A. (2005) Potential role of interleukin-10-secreting regulatory T cells in allergy and asthma. *Nat Rev Immunol,* **5,** 271-83.

79. Maas, K., Chan, S., Parker, J., Slater, A., Moore, J., Olsen, N. and Aune, T.M. (2002) Cutting edge: molecular portrait of human autoimmune disease. *J Immunol,* **169,** 5-9.

80. Olsen, N., Sokka, T., Seehorn, C.L., Kraft, B., Maas, K., Moore, J. and Aune, T.M. (2004) A gene expression signature for recent onset rheumatoid arthritis in peripheral blood mononuclear cells. *Ann Rheum Dis,* **63,** 1387-92.

81. Lock, C., Hermans, G., Pedotti, R., Brendolan, A., Schadt, E., Garren, H., Langer-Gould, A., Strober, S., Cannella, B., Allard, J. *et al.* (2002) Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. *Nat Med,* **8,** 500-8.

82. Chabas, D., Baranzini, S.E., Mitchell, D., Bernard, C.C., Rittling, S.R., Denhardt, D.T., Sobel, R.A., Lock, C., Karpuj, M., Pedotti, R. *et al.* (2001) The influence of the proinflammatory cytokine, osteopontin, on autoimmune demyelinating disease. *Science,* **294,** 1731-5.

83. Baechler, E.C., Batliwalla, F.M., Karypis, G., Gaffney, P.M., Ortmann, W.A., Espe, K.J., Shark, K.B., Grande, W.J., Hughes, K.M., Kapur, V. *et al.* (2003) Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci U S A,* **100,** 2610-5.

84. Rus, V., Atamas, S.P., Shustova, V., Luzina, I.G., Selaru, F., Magder, L.S. and Via, C.S. (2002) Expression of cytokine- and chemokine-related genes in peripheral blood mononuclear cells from lupus patients by cDNA array. *Clin Immunol,* **102,** 283-90.

85. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A. *et al.* (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science,* **286,** 531-7.

86. Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X. *et al.* (2000) Distinct types of diffuse

large B-cell lymphoma identified by gene expression profiling. *Nature,* **403,** 503-11.

87. Crow, M.K., Kirou, K.A. and Wohlgemuth, J. (2003) Microarray analysis of interferon-regulated genes in SLE. *Autoimmunity,* **36,** 481-90.

88. Crow, M.K. and Wohlgemuth, J. (2003) Microarray analysis of gene expression in lupus. *Arthritis Res Ther,* **5,** 279-87.

89. Bennett, L., Palucka, A.K., Arce, E., Cantrell, V., Borvak, J., Banchereau, J. and Pascual, V. (2003) Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med,* **197,** 711-23.

90. Han, G.M., Chen, S.L., Shen, N., Ye, S., Bao, C.D. and Gu, Y.Y. (2003) Analysis of gene expression profiles in human systemic lupus erythematosus using oligonucleotide microarray. *Genes Immun,* **4,** 177-86.

91. Batliwalla, F.M., Baechler, E.C., Xiao, X., Li, W., Balasubramanian, S., Khalili, H., Damle, A., Ortmann, W.A., Perrone, A., Kantor, A.B. *et al.* (2005) Peripheral blood gene expression profiling in rheumatoid arthritis. *Genes Immun,* **6,** 388-97.

92. Devauchelle, V., Marion, S., Cagnard, N., Mistou, S., Falgarone, G., Breban, M., Letourneur, F., Pitaval, A., Alibert, O., Lucchesi, C. *et al.* (2004) DNA microarray allows molecular profiling of rheumatoid arthritis and identification of pathophysiological targets. *Genes Immun,* **5,** 597-608.

93. van der Pouw Kraan, T.C., van Gaalen, F.A., Huizinga, T.W., Pieterman, E., Breedveld, F.C. and Verweij, C.L. (2003) Discovery of distinctive gene expression profiles in rheumatoid synovium using cDNA microarray technology: evidence for the existence of multiple pathways of tissue destruction and repair. *Genes Immun,* **4,** 187-96.

94. Tajouri, L., Mellick, A.S., Ashton, K.J., Tannenberg, A.E., Nagra, R.M., Tourtellotte, W.W. and Griffiths, L.R. (2003) Quantitative and qualitative changes in gene expression patterns characterize the activity of plaques in multiple sclerosis. *Brain Res Mol Brain Res,* **119,** 170-83.

95. Mandel, M., Gurevich, M., Pauzner, R., Kaminski, N. and Achiron, A. (2004) Autoimmunity gene expression portrait: specific signature that intersects or differentiates between multiple sclerosis and systemic lupus erythematosus. *Clin Exp Immunol,* **138,** 164-70.

96. Benson, M., Svensson, P.A., Carlsson, B., Jernas, M., Reinholdt, J., Cardell, L.O. and Carlsson, L. (2002) DNA microarrays to study gene expression in allergic airways. *Clin Exp Allergy,* **32,** 301-8.

97. Brutsche, M.H., Brutsche, I.C., Wood, P., Brass, A., Morrison, N., Rattay, M., Mogulkoc, N., Simler, N., Craven, M., Custovic, A. *et al.* (2001) Apoptosis signals in atopy and asthma measured with cDNA arrays. *Clin Exp Immunol,* **123,** 181-7.

98. Brutsche, M.H., Joos, L., Carlen Brutsche, I.E., Bissinger, R., Tamm, M., Custovic, A. and Woodcock, A. (2002) Array-based diagnostic gene-expression score for atopy and asthma. *J Allergy Clin Immunol,* **109,** 271-3.

99. Laprise, C., Sladek, R., Ponton, A., Bernier, M.C., Hudson, T.J. and Laviolette, M. (2004) Functional classes of bronchial mucosa genes that are differentially expressed in asthma. *BMC Genomics,* **5,** 21.

100. Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science,* **270,** 467-70.

101. Bullinger, L., Dohner, K., Bair, E., Frohling, S., Schlenk, R.F., Tibshirani, R., Dohner, H. and Pollack, J.R. (2004) Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med,* **350,** 1605-16.

102. Borczuk, A.C., Shah, L., Pearson, G.D., Walter, K.L., Wang, L., Austin, J.H., Friedman, R.A. and Powell, C.A. (2004) Molecular signatures in biopsy specimens of lung cancer. *Am J Respir Crit Care Med,* **170,** 167-74.

103. Li, Y., St John, M.A., Zhou, X., Kim, Y., Sinha, U., Jordan, R.C., Eisele, D., Abemayor, E., Elashoff, D., Park, N.H. *et al.* (2004) Salivary transcriptome diagnostics for oral cancer detection. *Clin Cancer Res,* **10,** 8442-50.

104. Chang, H.Y., Nuyten, D.S., Sneddon, J.B., Hastie, T., Tibshirani, R., Sorlie, T., Dai, H., He, Y.D., van't Veer, L.J., Bartelink, H. *et al.* (2005) Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A,* **102,** 3738-43.

105. Huang, E., Cheng, S.H., Dressman, H., Pittman, J., Tsou, M.H., Horng, C.F., Bild, A., Iversen, E.S., Liao, M., Chen, C.M. *et al.* (2003) Gene expression predictors of breast cancer outcomes. *Lancet,* **361,** 1590-6.

106. van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T. *et al.* (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature,* **415,** 530-6.

107. van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J. *et al.* (2002) A gene-

expression signature as a predictor of survival in breast cancer. *N Engl J Med,* **347,** 1999-2009.

108. O'Donnell, R.K., Kupferman, M., Wei, S.J., Singhal, S., Weber, R., O'Malley, B., Cheng, Y., Putt, M., Feldman, M., Ziober, B. *et al.* (2005) Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. *Oncogene,* **24,** 1244-51.

109. Pomeroy, S.L., Tamayo, P., Gaasenbeek, M., Sturla, L.M., Angelo, M., McLaughlin, M.E., Kim, J.Y., Goumnerova, L.C., Black, P.M., Lau, C. *et al.* (2002) Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature,* **415,** 436-42.

110. Panelli, M.C., Wang, E., Phan, G., Puhlmann, M., Miller, L., Ohnmacht, G.A., Klein, H.G. and Marincola, F.M. (2002) Gene-expression profiling of the response of peripheral blood mononuclear cells and melanoma metastases to systemic IL-2 administration. *Genome Biol,* **3,** RESEARCH0035.

111. Mariadason, J.M., Arango, D., Shi, Q., Wilson, A.J., Corner, G.A., Nicholas, C., Aranes, M.J., Lesser, M., Schwartz, E.L. and Augenlicht, L.H. (2003) Gene expression profiling-based prediction of response of colon carcinoma cells to 5-fluorouracil and camptothecin. *Cancer Res,* **63,** 8791-812.

112. Chang, J.C., Wooten, E.C., Tsimelzon, A., Hilsenbeck, S.G., Gutierrez, M.C., Elledge, R., Mohsin, S., Osborne, C.K., Chamness, G.C., Allred, D.C. *et al.* (2003) Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet,* **362,** 362-9.

113. Spentzos, D., Levine, D.A., Kolia, S., Otu, H., Boyd, J., Libermann, T.A. and Cannistra, S.A. (2005) Unique gene expression profile based on pathologic response in epithelial ovarian cancer. *J Clin Oncol,* **23,** 7911-8.

114. Dobbin, K.K., Beer, D.G., Meyerson, M., Yeatman, T.J., Gerald, W.L., Jacobson, J.W., Conley, B., Buetow, K.H., Heiskanen, M., Simon, R.M. *et al.* (2005) Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. *Clin Cancer Res,* **11,** 565-72.

115. Cheung, V.G., Conlin, L.K., Weber, T.M., Arcaro, M., Jen, K.Y., Morley, M. and Spielman, R.S. (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet,* **33,** 422-5.

116. Monks, S.A., Leonardson, A., Zhu, H., Cundiff, P., Pietrusiak, P., Edwards, S., Phillips, J.W., Sachs, A. and Schadt, E.E. (2004) Genetic inheritance of gene expression in human cell lines. *Am J Hum Genet,* **75,** 1094-105.

117. Morley, M., Molony, C.M., Weber, T.M., Devlin, J.L., Ewens, K.G., Spielman, R.S. and Cheung, V.G. (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature,* **430,** 743-7.

118. Schadt, E.E., Monks, S.A., Drake, T.A., Lusis, A.J., Che, N., Colinayo, V., Ruff, T.G., Milligan, S.B., Lamb, J.R., Cavet, G. *et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature,* **422,** 297-302.

119. Brem, R.B., Yvert, G., Clinton, R. and Kruglyak, L. (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science,* **296,** 752-5.

120. Maas, K., Chen, H., Shyr, Y., Olsen, N.J. and Aune, T. (2005) Shared gene expression profiles in individuals with autoimmune disease and unaffected first-degree relatives of individuals with autoimmune disease. *Hum Mol Genet,* **14,** 1305-14.

121. Ermann, J. and Fathman, C.G. (2001) Autoimmune diseases: genes, bugs and failed regulation. *Nat Immunol,* **2,** 759-61.

122. Bowtell, D.a.S., Joseph (2003) *DNA microarrays : a molecular cloning manual*. Cold Spring Harbor Laboratory Press, cold Spring Harbor, N.Y.

123. de Hoon, M.J., Imoto, S., Nolan, J. and Miyano, S. (2004) Open source clustering software. *Bioinformatics,* **20,** 1453-4.

124. Shannon, W., Culverhouse, R. and Duncan, J. (2003) Analyzing microarray data using cluster analysis. *Pharmacogenomics,* **4,** 41-52.

125. Kerr, M.K., Martin, M. and Churchill, G.A. (2000) Analysis of variance for gene expression microarray data. *J Comput Biol,* **7,** 819-37.

126. Quackenbush, J. (2001) Computational analysis of microarray data. *Nat Rev Genet,* **2,** 418-27.

127. Hoyle, D.C., Rattray, M., Jupp, R. and Brass, A. (2002) Making sense of microarray data distributions. *Bioinformatics,* **18,** 576-84.

128. Berrar, D.P., Dubitzky, Werner, Granzow, Martin (2003) *A practical approach to microarray data analysis*. Kluwer Academic Publishers, Boston, MA.

129. Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A,* **95,** 14863-8.

130. Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M. *et al.* (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques,* **34,** 374-8.

131. Gordon, A.D. (1999) *Classification.* 2nd Ed ed. Chapman & Hall, New York.

132. Low, L.Y. (1983) The Jackknife, the Bootstrap and Other Resampling Plans - Efron,B. *Journal of the American Statistical Association,* **78,** 987-987.

133. Holter, N.S., Mitra, M., Maritan, A., Cieplak, M., Banavar, J.R. and Fedoroff, N.V. (2000) Fundamental patterns underlying gene expression profiles: simplicity from complexity. *Proc Natl Acad Sci U S A,* **97,** 8409-14.

134. Raychaudhuri, S., Stuart, J.M. and Altman, R.B. (2000) Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac Symp Biocomput***,** 455-66.

135. Tusher, V.G., Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A,* **98,** 5116-21.

136. Gaasterland, T. and Bekiranov, S. (2000) Making the most of microarray data. *Nat Genet,* **24,** 204-6.

137. Brown, M.P., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares, M., Jr. and Haussler, D. (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci U S A,* **97,** 262-7.

138. Ntzani, E.E. and Ioannidis, J.P. (2003) Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet,* **362,** 1439-44.

139. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* **25,** 25-9.

140. Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. and Lempicki, R.A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol,* **4,** P3.

141. Hosack, D.A., Dennis, G., Jr., Sherman, B.T., Lane, H.C. and Lempicki, R.A. (2003) Identifying biological themes within lists of genes with EASE. *Genome Biol,* **4,** R70.

142. Todd, J.A. (1991) A protective role of the environment in the development of type 1 diabetes? *Diabet Med,* **8,** 906-10.

143. Theofilopoulos, A.N. (1995) The basis of autoimmunity: Part II. Genetic predisposition. *Immunol Today,* **16,** 150-9.

144. Bias, W.B., Reveille, J.D., Beaty, T.H., Meyers, D.A. and Arnett, F.C. (1986) Evidence that autoimmunity in man is a Mendelian dominant trait. *Am J Hum Genet,* **39,** 584-602.

145. Prokunina, L. and Alarcon-Riquelme, M. (2004) The genetic basis of systemic lupus erythematosus--knowledge of today and thoughts for tomorrow. *Hum Mol Genet,* **13 Spec No 1,** R143-8.

146. MacGregor, A.J., Snieder, H., Rigby, A.S., Koskenvuo, M., Kaprio, J., Aho, K. and Silman, A.J. (2000) Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum,* **43,** 30-7.

147. Lawrence, J.S., Martins, C.L. and Drake, G.L. (1987) A family survey of lupus erythematosus. 1. Heritability. *J Rheumatol,* **14,** 913-21.

148. Deapen, D., Escalante, A., Weinrib, L., Horwitz, D., Bachman, B., Roy-Burman, P., Walker, A. and Mack, T.M. (1992) A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum,* **35,** 311-8.

149. Bomprezzi, R., Ringner, M., Kim, S., Bittner, M.L., Khan, J., Chen, Y., Elkahloun, A., Yu, A., Bielekova, B., Meltzer, P.S. *et al.* (2003) Gene expression profile in multiple sclerosis patients and healthy controls: identifying pathways relevant to disease. *Hum Mol Genet,* **12,** 2191-9.

150. Jarvis, J.N., Dozmorov, I., Jiang, K., Frank, M.B., Szodoray, P., Alex, P. and Centola, M. (2004) Novel approaches to gene expression analysis of active polyarticular juvenile rheumatoid arthritis. *Arthritis Res Ther,* **6,** R15-R32.

151. Maas, K., Westfall, M., Pietenpol, J., Olsen, N.J. and Aune, T. (2005) Reduced p53 in peripheral blood mononuclear cells from patients with rheumatoid arthritis is associated with loss of radiation-induced apoptosis. *Arthritis Rheum,* **52,** 1047-57.

152. Kuerbitz, S.J., Plunkett, B.S., Walsh, W.V. and Kastan, M.B. (1992) Wild-type p53 is a cell cycle checkpoint determinant following irradiation. *Proc Natl Acad Sci U S A,* **89,** 7491-5.

153. Stewart, Z.A. and Pietenpol, J.A. (2001) p53 Signaling and cell cycle checkpoints. *Chem Res Toxicol,* **14,** 243-63.

154. Yamanishi, Y., Boyle, D.L., Pinkoski, M.J., Mahboubi, A., Lin, T., Han, Z., Zvaifler, N.J., Green, D.R. and Firestein, G.S. (2002) Regulation of joint destruction and inflammation by p53 in collagen-induced arthritis. *Am J Pathol,* **160,** 123-30.

155. Prives, C. and Hall, P.A. (1999) The p53 pathway. *J Pathol,* **187,** 112-26.

156. Salvador, J.M., Hollander, M.C., Nguyen, A.T., Kopp, J.B., Barisoni, L., Moore, J.K., Ashwell, J.D. and Fornace, A.J., Jr. (2002) Mice lacking the p53-effector gene Gadd45a develop a lupus-like syndrome. *Immunity,* **16,** 499-508.

157. Balomenos, D., Martin-Caballero, J., Garcia, M.I., Prieto, I., Flores, J.M., Serrano, M. and Martinez, A.C. (2000) The cell cycle inhibitor p21 controls T-cell proliferation and sex-linked lupus development. *Nat Med,* **6,** 171-6.

158. Mak, T.W., Hakem, A., McPherson, J.P., Shehabeldin, A., Zablocki, E., Migon, E., Duncan, G.S., Bouchard, D., Wakeham, A., Cheung, A. *et al.* (2000) Brcal required for T cell lineage development but not TCR loci rearrangement. *Nat Immunol,* **1,** 77-82.

159. Theofilopoulos, A.N., Dummer, W. and Kono, D.H. (2001) T cell homeostasis and systemic autoimmunity. *J Clin Invest,* **108,** 335-40.

160. Baccala, R. and Theofilopoulos, A.N. (2005) The new paradigm of T-cell homeostatic proliferation-induced autoimmunity. *Trends Immunol,* **26,** 5-8.

161. Schulze-Koops, H. (2004) Lymphopenia and autoimmune diseases. *Arthritis Res Ther,* **6,** 178-80.

162. McFadden, E.R., Jr. and Gilbert, I.A. (1992) Asthma. *N Engl J Med,* **327,** 1928-37.

163. Umetsu, D.T., McIntire, J.J., Akbari, O., Macaubas, C. and DeKruyff, R.H. (2002) Asthma: an epidemic of dysregulated immunity. *Nat Immunol,* **3,** 715-20.

164. Cloutier, M.M., Wakefield, D.B., Hall, C.B. and Bailit, H.L. (2002) Childhood asthma in an urban community: prevalence, care system, and treatment. *Chest,* **122,** 1571-9.

165. Blumenthal, J.B. and Blumenthal, M.N. (2002) Genetics of asthma. *Med Clin North Am,* **86,** 937-50.

166. Skoner, D.P. (2001) Allergic rhinitis: definition, epidemiology, pathophysiology, detection, and diagnosis. *J Allergy Clin Immunol,* **108,** S2-8.

167. Brugarolas, J., Haynes, B.F. and Nevins, J.R. (2001) Towards a genomic-based diagnosis. *Lancet,* **357,** 249-50.

168. PA., G. (1994) Rush injection therapy for asthma and allergic rhinitis. *Ann Allergy,* **73,** 378-80.

169. DK., G. (1996) Risk management in allergen immunotherapy. *J Allergy Clin Immunol.,* **98,** 330-4.

170. Saldanha, A.J. (2004) Java treeview--extensible visualization of microarray data. *Bioinformatics*.

171. Cameron, L., Hamid, Q., Wright, E., Nakamura, Y., Christodoulopoulos, P., Muro, S., Frenkiel, S., Lavigne, F., Durham, S. and Gould, H. (2000) Local synthesis of epsilon germline gene transcripts, IL-4, and IL-13 in allergic nasal mucosa after ex vivo allergen exposure. *J Allergy Clin Immunol,* **106,** 46-52.

172. Benson, M., Strannegard, I.L., Strannegard, O. and Wennergren, G. (2000) Topical steroid treatment of allergic rhinitis decreases nasal fluid TH2 cytokines, eosinophils, eosinophil cationic protein, and IgE but has no significant effect on IFN-gamma, IL-1beta, TNF-alpha, or neutrophils. *J Allergy Clin Immunol,* **106,** 307-12.

173. Durham, S.R., Ying, S., Varney, V.A., Jacobson, M.R., Sudderick, R.M., Mackay, I.S., Kay, A.B. and Hamid, Q.A. (1992) Cytokine messenger RNA expression for IL-3, IL-4, IL-5, and granulocyte/macrophage-colony-stimulating factor in the nasal mucosa after local allergen provocation: relationship to tissue eosinophilia. *J Immunol,* **148,** 2390-4.

174. Bradding, P., Mediwake, R., Feather, I.H., Madden, J., Church, M.K., Holgate, S.T. and Howarth, P.H. (1995) TNF alpha is localized to nasal mucosal mast cells and is released in acute allergic rhinitis. *Clin Exp Allergy,* **25,** 406-15.

175. Nouri-Aria, K.T., O'Brien, F., Noble, W., Jabcobson, M.R., Rajakulasingam, K. and Durham, S.R. (2000) Cytokine expression during allergen-induced late nasal responses: IL-4 and IL-5 mRNA is expressed early (at 6 h) predominantly by eosinophils. *Clin Exp Allergy,* **30,** 1709-16.

176. Varney, V.A., Jacobson, M.R., Sudderick, R.M., Robinson, D.S., Irani, A.M., Schwartz, L.B., Mackay, I.S., Kay, A.B. and Durham, S.R. (1992) Immunohistology of the nasal mucosa following allergen-induced rhinitis. Identification of activated T lymphocytes, eosinophils, and neutrophils. *Am Rev Respir Dis,* **146,** 170-6.

177. KleinJan, A., Vinke, J.G., Severijnen, L.W. and Fokkens, W.J. (2000) Local production and detection of (specific) IgE in nasal B-cells and plasma cells of allergic rhinitis patients. *Eur Respir J,* **15,** 491-7.

178. Aune, T.M., Maas, K., Moore, J.H. and Olsen, N.J. (2003) Gene expression profiles in human autoimmune disease. *Curr Pharm Des,* **9,** 1905-17.

179. Saito, H., Abe, J. and Matsumoto, K. (2005) Allergy-related genes in microarray: an update review. *J Allergy Clin Immunol,* **116,** 56-9.

180. Vignola, A.M., Chanez, P., Chiappara, G., Siena, L., Merendino, A., Reina, C., Gagliardo, R., Profita, M., Bousquet, J. and Bonsignore, G. (1999) Evaluation of apoptosis of eosinophils, macrophages, and T lymphocytes in mucosal biopsy specimens of patients with asthma and chronic bronchitis. *J Allergy Clin Immunol,* **103,** 563-73.

181. Ashkenazi, A. and Dixit, V.M. (1998) Death receptors: signaling and modulation. *Science,* **281,** 1305-8.

182. Evan, G. and Littlewood, T. (1998) A matter of life and cell death. *Science,* **281,** 1317-22.

183. Adams, J.M. and Cory, S. (1998) The Bcl-2 protein family: arbiters of cell survival. *Science,* **281,** 1322-6.

184. Wachholz, P.A., Nouri-Aria, K.T., Wilson, D.R., Walker, S.M., Verhoef, A., Till, S.J. and Durham, S.R. (2002) Grass pollen immunotherapy for hayfever is associated with increases in local nasal but not peripheral Th1:Th2 cytokine ratios. *Immunology,* **105,** 56-62.

185. Klimek, L., Dormann, D., Jarman, E.R., Cromwell, O., Riechelmann, H. and Reske-Kunz, A.B. (1999) Short-term preseasonal birch pollen allergoid immunotherapy influences symptoms, specific nasal provocation and cytokine levels in nasal secretions, but not peripheral T-cell responses, in patients with allergic rhinitis. *Clin Exp Allergy,* **29,** 1326-35.

186. Lard, L.R., Visser, H., Speyer, I., vander Horst-Bruinsma, I.E., Zwinderman, A.H., Breedveld, F.C. and Hazes, J.M. (2001) Early versus delayed treatment in patients with recent-onset rheumatoid arthritis: comparison of two cohorts who received different treatment strategies. *Am J Med,* **111,** 446-51.

187. Egsmose, C., Lund, B., Borg, G., Pettersson, H., Berg, E., Brodin, U. and Trang, L. (1995) Patients with rheumatoid arthritis benefit from early 2nd line therapy: 5 year followup of a prospective double blind placebo controlled study. *J Rheumatol,* **22,** 2208-13.

188. van der Heide, A., Jacobs, J.W., Bijlsma, J.W., Heurkens, A.H., van Booma-Frankfort, C., van der Veen, M.J., Haanen, H.C., Hofman, D.M., van Albada-Kuipers, G.A., ter Borg, E.J. *et al.* (1996) The effectiveness of early treatment

with "second-line" antirheumatic drugs. A randomized, controlled trial. *Ann Intern Med,* **124,** 699-707.

189. Bathon, J.M., Martin, R.W., Fleischmann, R.M., Tesser, J.R., Schiff, M.H., Keystone, E.C., Genovese, M.C., Wasko, M.C., Moreland, L.W., Weaver, A.L. *et al.* (2000) A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis. *N Engl J Med,* **343,** 1586-93.

190. Genovese, M.C., Bathon, J.M., Martin, R.W., Fleischmann, R.M., Tesser, J.R., Schiff, M.H., Keystone, E.C., Wasko, M.C., Moreland, L.W., Weaver, A.L. *et al.* (2002) Etanercept versus methotrexate in patients with early rheumatoid arthritis: two-year radiographic and clinical outcomes. *Arthritis Rheum,* **46,** 1443-50.

191. Klareskog, L. and McDevitt, H. (1999) Rheumatoid arthritis and its animal models: the role of TNF-alpha and the possible absence of specific immune reactions. *Curr Opin Immunol,* **11,** 657-62.

192. Morel, J. and Combe, B. (2005) How to predict prognosis in early rheumatoid arthritis. *Best Pract Res Clin Rheumatol,* **19,** 137-46.

193. Weyand, C.M., Hicok, K.C., Conn, D.L. and Goronzy, J.J. (1992) The influence of HLA-DRB1 genes on disease severity in rheumatoid arthritis. *Ann Intern Med,* **117,** 801-6.

194. Weyand, C.M., McCarthy, T.G. and Goronzy, J.J. (1995) Correlation between disease phenotype and genetic heterogeneity in rheumatoid arthritis. *J Clin Invest,* **95,** 2120-6.

195. Gorman, J.D., Lum, R.F., Chen, J.J., Suarez-Almazor, M.E., Thomson, G. and Criswell, L.A. (2004) Impact of shared epitope genotype and ethnicity on erosive disease: a meta-analysis of 3,240 rheumatoid arthritis patients. *Arthritis Rheum,* **50,** 400-12.

196. Morel, J., Roch-Bras, F., Molinari, N., Sany, J., Eliaou, J.F. and Combe, B. (2004) HLA-DMA*0103 and HLA-DMB*0104 alleles as novel prognostic factors in rheumatoid arthritis. *Ann Rheum Dis,* **63,** 1581-6.

197. van der Horst-Bruinsma, I.E., Visser, H., Hazes, J.M., Breedveld, F.C., Verduyn, W., Schreuder, G.M., de Vries, R.R. and Zanelli, E. (1999) HLA-DQ-associated predisposition to and dominant HLA-DR-associated protection against rheumatoid arthritis. *Hum Immunol,* **60,** 152-8.

198. Visser, H., le Cessie, S., Vos, K., Breedveld, F.C. and Hazes, J.M. (2002) How to diagnose rheumatoid arthritis early: a prediction model for persistent (erosive) arthritis. *Arthritis Rheum,* **46,** 357-65.

199. Ransohoff, D.F. and Feinstein, A.R. (1978) Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med,* **299,** 926-30.

200. Visser, H. (2005) Early diagnosis of rheumatoid arthritis. *Best Pract Res Clin Rheumatol,* **19,** 55-72.

201. Sokka, T. and Pincus, T. (2003) Eligibility of patients in routine care for major clinical trials of anti-tumor necrosis factor alpha agents in rheumatoid arthritis. *Arthritis Rheum,* **48,** 313-8.

202. Pincus, T., Amara I, Sokka T, Koch G (2004) A simple Apgar-like patient score (ALPS) based on 3 Core Data Set patient questionnaire self-report measures, physical function, patin and globla status, yields results similar to ACR 20 in a rheumatoid arthritis clinical trial. *Arthritis and Rheumatism,* **50,** S315.

203. Korpela, M., Laasonen, L., Hannonen, P., Kautiainen, H., Leirisalo-Repo, M., Hakala, M., Paimela, L., Blafield, H., Puolakka, K. and Mottonen, T. (2004) Retardation of joint damage in patients with early rheumatoid arthritis by initial aggressive treatment with disease-modifying antirheumatic drugs: five-year experience from the FIN-RACo study. *Arthritis Rheum,* **50,** 2072-81.

204. Sokka, T., Kautiainen, H., Hakkinen, A. and Hannonen, P. (2004) Radiographic progression is getting milder in patients with early rheumatoid arthritis. Results of 3 cohorts over 5 years. *J Rheumatol,* **31,** 1073-82.

205. Combe, B., Dougados, M., Goupille, P., Cantagrel, A., Eliaou, J.F., Sibilia, J., Meyer, O., Sany, J., Daures, J.P. and Dubois, A. (2001) Prognostic factors for radiographic damage in early rheumatoid arthritis: a multiparameter prospective study. *Arthritis Rheum,* **44,** 1736-43.

206. Dixey, J., Solymossy, C. and Young, A. (2004) Is it possible to predict radiological damage in early rheumatoid arthritis (RA)? A report on the occurrence, progression, and prognostic factors of radiological erosions over the first 3 years in 866 patients from the Early RA Study (ERAS). *J Rheumatol Suppl,* **69,** 48-54.

207. van der Heijde, D.M., van Riel, P.L., van Leeuwen, M.A., van 't Hof, M.A., van Rijswijk, M.H. and van de Putte, L.B. (1992) Prognostic factors for radiographic damage and physical disability in early rheumatoid arthritis. A prospective follow-up study of 147 patients. *Br J Rheumatol,* **31,** 519-25.

208. Young, A., Corbett, M., Winfield, J., Jaqueremada, D., Williams, P., Papasavvas, G., Hay, F. and Roitt, I. (1988) A prognostic index for erosive changes in the hands, feet, and cervical spines in early rheumatoid arthritis. *Br J Rheumatol,* **27,** 94-101.

209. Belpaire, F.M. and Bogaert, M.G. (1990) Binding of diltiazem to albumin, alpha 1-acid glycoprotein and to serum in man. *J Clin Pharmacol,* **30,** 311-7.

210. Bradley, D.S., Nabozny, G.H., Cheng, S., Zhou, P., Griffiths, M.M., Luthra, H.S. and David, C.S. (1997) HLA-DQB1 polymorphism determines incidence, onset, and severity of collagen-induced arthritis in transgenic mice. Implications in human rheumatoid arthritis. *J Clin Invest,* **100,** 2227-34.

211. Kerlan-Candon, S., Combe, B., Vincent, R., Clot, J., Pinet, V. and Eliaou, J.F. (2001) HLA-DRB1 gene transcripts in rheumatoid arthritis. *Clin Exp Immunol,* **124,** 142-9.

212. Chappel, C.I. and Chappel, W.R. (1983) The discovery and development of the BB rat colony: an animal model of spontaneous diabetes mellitus. *Metabolism,* **32,** 8-10.

213. Marliss, E.B., Nakhooda, A.F., Poussier, P. and Sima, A.A. (1982) The diabetic syndrome of the 'BB' Wistar rat: possible relevance to type 1 (insulin-dependent) diabetes in man. *Diabetologia,* **22,** 225-32.

214. Tanaka, T.S., Jaradat, S.A., Lim, M.K., Kargul, G.J., Wang, X., Grahovac, M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R. *et al.* (2000) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. *Proc Natl Acad Sci U S A,* **97,** 9127-32.

215. Holt, S.E., Aisner, D.L., Baur, J., Tesmer, V.M., Dy, M., Ouellette, M., Trager, J.B., Morin, G.B., Toft, D.O., Shay, J.W. *et al.* (1999) Functional requirement of p23 and Hsp90 in telomerase complexes. *Genes Dev,* **13,** 817-26.

216. Akalin, A., Elmore, L.W., Forsythe, H.L., Amaker, B.A., McCollum, E.D., Nelson, P.S., Ware, J.L. and Holt, S.E. (2001) A novel mechanism for chaperone-mediated telomerase regulation during prostate cancer progression. *Cancer Res,* **61,** 4791-6.

217. Pandey, P., Saleh, A., Nakazawa, A., Kumar, S., Srinivasula, S.M., Kumar, V., Weichselbaum, R., Nalin, C., Alnemri, E.S., Kufe, D. *et al.* (2000) Negative regulation of cytochrome c-mediated oligomerization of Apaf-1 and activation of procaspase-9 by heat shock protein 90. *Embo J,* **19,** 4310-22.

218. Sreedhar, A.S., Mihaly, K., Pato, B., Schnaider, T., Stetak, A., Kis-Petik, K., Fidy, J., Simonics, T., Maraz, A. and Csermely, P. (2003) Hsp90 inhibition accelerates cell lysis. Anti-Hsp90 ribozyme reveals a complex mechanism of Hsp90 inhibitors involving both superoxide- and Hsp90-dependent events. *J Biol Chem,* **278,** 35231-40.

219. O'Brien, B.A., Huang, Y., Geng, X., Dutz, J.P. and Finegood, D.T. (2002) Phagocytosis of apoptotic cells by macrophages from NOD mice is reduced. *Diabetes,* **51,** 2481-8.

220. Chen, G., Cizeau, J., Vande Velde, C., Park, J.H., Bozek, G., Bolton, J., Shi, L., Dubik, D. and Greenberg, A. (1999) Nix and Nip3 form a subfamily of pro-apoptotic mitochondrial proteins. *J Biol Chem,* **274,** 7-10.

221. Park, E.J., Kim, J.H., Seong, R.H., Kim, C.G., Park, S.D. and Hong, S.H. (1999) Characterization of a novel mouse cDNA, ES18, involved in apoptotic cell death of T-cells. *Nucleic Acids Res,* **27,** 1524-30.

222. Takayama, S., Reed, J.C. and Homma, S. (2003) Heat-shock proteins as regulators of apoptosis. *Oncogene,* **22,** 9041-7.

223. Sato, S., Fujita, N. and Tsuruo, T. (2000) Modulation of Akt kinase activity by binding to Hsp90. *Proc Natl Acad Sci U S A,* **97,** 10832-7.

224. Basso, A.D., Solit, D.B., Chiosis, G., Giri, B., Tsichlis, P. and Rosen, N. (2002) Akt forms an intracellular complex with heat shock protein 90 (Hsp90) and Cdc37 and is destabilized by inhibitors of Hsp90 function. *J Biol Chem,* **277,** 39858-66.

225. Ambrosini, G., Adida, C. and Altieri, D.C. (1997) A novel anti-apoptosis gene, survivin, expressed in cancer and lymphoma. *Nat Med,* **3,** 917-21.

226. Song, J., So, T., Cheng, M., Tang, X. and Croft, M. (2005) Sustained survivin expression from OX40 costimulatory signals drives T cell clonal expansion. *Immunity,* **22,** 621-31.

227. Fortugno, P., Beltrami, E., Plescia, J., Fontana, J., Pradhan, D., Marchisio, P.C., Sessa, W.C. and Altieri, D.C. (2003) Regulation of survivin function by Hsp90. *Proc Natl Acad Sci U S A,* **100,** 13791-6.

228. Saito, Y., Yamagishi, N., Ishihara, K. and Hatayama, T. (2003) Identification of alpha-tubulin as an hsp105alpha-binding protein by the yeast two-hybrid system. *Exp Cell Res,* **286,** 233-40.

229. GeneCard™ database. Weizmann Institute of Science http://bioinfo.weizmann.ac.il/cards/index.html.

230. Maas, K.H., Westfall, M., Pietenpol, J.A., Olsen, N. and Aune, T.M. (2003) Cellular Damage Response Defects in Human Rheumatoid Arthritis. *Arthritis Rheum/Abstracts,* **48,** S147.

231. Dahlen, E., Dawe, K., Ohlsson, L. and Hedlund, G. (1998) Dendritic cells and macrophages are the first and major producers of TNF-alpha in pancreatic islets in the nonobese diabetic mouse. *J Immunol,* **160,** 3585-93.

232. Watanabe-Fukunaga, R., Brannan, C.I., Copeland, N.G., Jenkins, N.A. and Nagata, S. (1992) Lymphoproliferation disorder in mice explained by defects in Fas antigen that mediates apoptosis. *Nature,* **356,** 314-7.

233. Colucci, F., Bergman, M.L., Penha-Goncalves, C., Cilio, C.M. and Holmberg, D. (1997) Apoptosis resistance of nonobese diabetic peripheral lymphocytes linked to the Idd5 diabetes susceptibility region. *Proc Natl Acad Sci U S A,* **94,** 8670-4.

234. Emlen, W., Niebur, J. and Kadera, R. (1994) Accelerated in vitro apoptosis of lymphocytes from patients with systemic lupus erythematosus. *J Immunol,* **152,** 3685-92.

235. Lorenz, H.M., Grunke, M., Hieronymus, T., Herrmann, M., Kuhnel, A., Manger, B. and Kalden, J.R. (1997) In vitro apoptosis and expression of apoptosis-related molecules in lymphocytes from patients with systemic lupus erythematosus and other autoimmune diseases. *Arthritis Rheum,* **40,** 306-17.

236. Itoh, N., Yonehara, S., Ishii, A., Yonehara, M., Mizushima, S., Sameshima, M., Hase, A., Seto, Y. and Nagata, S. (1991) The polypeptide encoded by the cDNA for human cell surface antigen Fas can mediate apoptosis. *Cell,* **66,** 233-43.

237. Itoh, N. and Nagata, S. (1993) A novel protein domain required for apoptosis. Mutational analysis of human Fas antigen. *J Biol Chem,* **268,** 10932-7.

238. Sato, T., Irie, S., Kitada, S. and Reed, J.C. (1995) FAP-1: a protein tyrosine phosphatase that associates with Fas. *Science,* **268,** 411-5.

239. Zhou, Y.W., Komada, Y., Inaba, H., Azuma, E. and Sakurai, M. (1998) Down-regulation of Fas-associated phosphatase-1 (FAP-1) in interleukin-2-activated T cells. *Cell Immunol,* **186,** 103-10.

240. Boussiotis, V.A., Freeman, G.J., Taylor, P.A., Berezovskaya, A., Grass, I., Blazar, B.R. and Nadler, L.M. (2000) p27kip1 functions as an anergy factor inhibiting interleukin 2 transcription and clonal expansion of alloreactive human and mouse helper T lymphocytes. *Nat Med,* **6,** 290-7.

241. Troncone, G., Iaccarino, A., Caleo, A., Bifano, D., Pettinato, G. and Palombini, L. (2003) p27 Kip1 protein expression in Hashimoto's thyroiditis. *J Clin Pathol,* **56,** 587-91.

242. Sherr, C.J. and Roberts, J.M. (1999) CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev,* **13,** 1501-12.

243. Cheng, M., Olivier, P., Diehl, J.A., Fero, M., Roussel, M.F., Roberts, J.M. and Sherr, C.J. (1999) The p21(Cip1) and p27(Kip1) CDK 'inhibitors' are essential activators of cyclin D-dependent kinases in murine fibroblasts. *Embo J,* **18,** 1571-83.

244. Coats, S., Whyte, P., Fero, M.L., Lacy, S., Chung, G., Randel, E., Firpo, E. and Roberts, J.M. (1999) A new pathway for mitogen-dependent cdk2 regulation uncovered in p27(Kip1)-deficient cells. *Curr Biol,* **9,** 163-73.

245. Hein, W.R. and Griebel, P.J. (2003) A road less travelled: large animal models in immunological research. *Nat Rev Immunol,* **3,** 79-84.

246. Greiner, D.L., Rossini, A.A. and Mordes, J.P. (2001) Translating data from animal models into methods for preventing human autoimmune diabetes mellitus: caveat emptor and primum non nocere. *Clin Immunol,* **100,** 134-43.

247. Kirou, K.A., Lee, C., George, S., Louca, K., Papagiannis, I.G., Peterson, M.G., Ly, N., Woodward, R.N., Fry, K.E., Lau, A.Y. *et al.* (2004) Coordinate overexpression of interferon-alpha-induced genes in systemic lupus erythematosus. *Arthritis Rheum,* **50,** 3958-67.

248. Berger, S.L. (2002) Histone modifications in transcriptional regulation. *Curr Opin Genet Dev,* **12,** 142-8.

249. Jaenisch, R. and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet,* **33 Suppl,** 245-54.

250. Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature,* **403,** 41-5.

251. Peterson, K.S., Huang, J.F., Zhu, J., D'Agati, V., Liu, X., Miller, N., Erlander, M.G., Jackson, M.R. and Winchester, R.J. (2004) Characterization of heterogeneity in the molecular pathogenesis of lupus nephritis from transcriptional profiles of laser-captured glomeruli. *J Clin Invest,* **113,** 1722-33.

252. Mycko, M.P., Papoian, R., Boschert, U., Raine, C.S. and Selmaj, K.W. (2004) Microarray gene expression profiling of chronic active and inactive lesions in multiple sclerosis. *Clin Neurol Neurosurg,* **106,** 223-9.

253. Guajardo, J.R., Schleifer, K.W., Daines, M.O., Ruddy, R.M., Aronow, B.J., Wills-Karp, M. and Hershey, G.K. (2005) Altered gene expression profiles in nasal respiratory epithelium reflect stable versus acute childhood asthma. *J Allergy Clin Immunol,* **115,** 243-51.

254. Liu, Z., Kim, J., Sypek, J.P., Wang, I.M., Horton, H., Oppenheim, F.G. and Bochner, B.S. (2004) Gene expression profiles in human nasal polyp tissues studied by means of DNA microarray. *J Allergy Clin Immunol,* **114,** 783-90.