Review

Gene expression signatures for autoimmune disease in peripheral blood mononuclear cells
Nancy J Olsen, Jason H Moore and Thomas M Aune

Division of Rheumatology, T-3219 Medical Center North, Vanderbilt University, Nashville, Tennessee, USA

Corresponding author: Nancy J Olsen; nancy.olsen@vanderbilt.edu

© 2004 BioMed Central Ltd

Abstract

The relatively new technology of DNA microarrays offers the possibility to probe the human genome for clues to the pathogenesis and treatment of human disease. While early studies using this approach were largely in oncology, many new reports are emerging in other fields including infectious diseases and pharmacology, and applications in autoimmunity have been recently reported by our group and others. Some of these investigations have examined animal models of autoimmune disease, but a number of human studies have also been carried out. Of special interest are those that have used peripheral blood samples because, unlike tissue biopsies, these are readily available from all subjects. Using this approach, patterns of gene expression can be detected that distinguish patients with autoimmune conditions from normal subjects. Furthermore, the genes that are identified provide clues to possible pathogenetic mechanisms and are likely to be useful in developing tests to establish diagnostic categories and predict therapeutic responses.

Keywords: autoimmune disease, DNA microarrays, gene expression, systemic lupus erythematosus

Introduction
The relatively new technology of DNA microarrays has made it feasible to measure the expression levels of thousands of genes in small biological samples [1]. It has been suggested that this methodology might be especially useful in analyzing the complex and parallel changes that occur within cells and tissues of the immune system in normal and pathologic states [2]. Much of the early work using DNA microarrays was in the field of oncology; other studies have examined host responses to infectious agents or drugs [3]. The gene array approach is especially well-suited to the type of multifactorial analysis that is needed to unravel the causes of human autoimmune disorders that involve both complex genetics and environmental factors [4,5]. Studies in autoimmune disease have included the use of biopsy samples from affected patients, targeting tissues such as synovium, brain or skin [6–9]. While this approach can offer insights for some disease subsets, it does not permit study of all afflicted patients and cannot be applied to early phases of disease when therapeutic interventions are most likely to be useful. As an alternative, we and others have hypothesized that due to the systemic nature of autoimmune disease, clinically relevant changes in gene expression should be observed in peripheral blood mononuclear cells (PBMCs). Using peripheral blood as the source of gene expression material offers the possibility of sampling any individual at any time and also has the potential to detect early pathogenetic and prognostic factors. This review will examine studies in autoimmune disease, focusing on the utility of peripheral blood samples to identify genes of interest. The potential for this approach to provide insights into disease pathogenesis and to aid with diagnosis and management are also discussed.

Approach to the study of autoimmune disease
A relatively small number of microarray studies in autoimmunity have been reported [3]. Some of these have

IDDM = insulin-dependent diabetes mellitus; IFN = interferon; IL = interleukin; MS = multiple sclerosis; PCR = polymerase chain reaction; PBMC = peripheral blood mononuclear cell; RA = rheumatoid arthritis; SDA = Symbolic discriminant analysis; SLE = systemic lupus erythematosus; SLEDAI = SLE disease activity index.

used animal models, such as for alopecia areata [7] and experimental systemic lupus erythematosus (SLE) [10]. In human autoimmunity, biopsy samples from tissues such as rheumatoid synovium [6,9] and skin [7] have yielded disease insights. Other groups of investigators have concentrated on the possibility that peripheral blood might show gene expression correlations with disease states [11]. Six published reports have described results obtained using microarray analysis of PBMC populations from patients with various autoimmune disorders (Table 1). Two of these studies were in multiple sclerosis (MS) [12,13] and three were in SLE, including one that used only juvenile subjects [14–16]. In a study from our own laboratory, four different autoimmune diseases, rheumatoid arthritis (RA), SLE, MS and Type-I or insulin-dependent diabetes mellitus (IDDM), were studied [17]. The diseases represented in these reports span a broad spectrum within the rubric of autoimmunity, including both generalized (RA and SLE) and tissue-specific (MS and IDDM) pathologies. Three of these syndromes (RA, SLE and MS) show a female predominance, while IDDM in humans has no significant gender dimorphism. Treatments also differ, with RA and SLE usually requiring long-term continuous immune suppression, while MS often shows quiescent phases requiring no treatment and IDDM therapies are for glucose control rather than immune suppression. In most published studies, autoimmune samples have been compared to unaffected control individuals who are generally matched for the age and gender characteristics of the study population. Our group also investigated the relationship between a normal immune response and the autoimmune response by examining control subjects before and after routine influenza vaccination [17]. It is notable that in most of these reports, the numbers of samples are relatively small, with significant results reported using study groups of 10–20 subjects each, indicating the strength of the multiparameter approach to analysis of array data. Many investigators have used frozen samples, which permit added flexibility for approaches such as studies of longitudinal responses to a therapeutic intervention [13].

While peripheral blood offers many advantages as a source of analysis material, one potential drawback is the small quantities of RNA that can be reasonably obtained. Surprisingly, information about the amount of blood needed to produce an analyzable sample has not been uniformly reported; one group used lymphocytopheresis, suggesting a need for large numbers of cells [12]. Early chip protocols often required more than 25 µg of total RNA, which could only be obtained by using large blood volumes. This could be problematic, especially in studies of children or seriously ill subjects. In our initial studies, the gene filter from Research Genetics (now Invitrogen, Carlsbad CA), which contained clones for approximately 4300 identified human genes, was chosen because only 5 µg of total RNA was required and we were interested in testing the feasibility of analyzing small blood samples. These gene filters are, however, no longer available. Current recommendations for other platforms, such as the Affymetrix Gene Chip Arrays®, require no more than 5 µg total RNA, probably due to improved efficiency of the labeling techniques, and this can be readily attained from blood samples without amplification. Sample size, therefore, is probably no longer a limiting factor in experimental design.

Methods for verifying data from microarrays have become familiar to most users. Reproducibility has been achieved

Table 1

Gene expression studies of peripheral blood mononuclear cells from patients with autoimmune diseases

Subjects studied	Gene arrays used	Summary	Reference
MS ($n = 27$) and normal controls ($n = 19$)	14,000 cDNA clones	MS patients distinguished from normal controls using 53 discriminatory genes	Bomprezzi et al., 2003 [12]
MS ($n = 10$)	Mini-lymphochip with >12,000 elements	Identification of a set of genes regulated by IFN-β	Sturzebecher et al., 2003 [13]
SLE ($n = 21$) and normal controls ($n = 12$)	Cytokine gene array with 375 genes in duplicate	Clustering distinguished most patients from controls	Rus et al., 2002 [14]
SLE ($n = 48$) and normal controls ($n = 42$)	Affymetrix U95A array with >10,000 genes	Dysregulation of genes in the IFN pathway present in SLE patients with active disease	Baechler et al., 2003 [15]
Pediatric SLE ($n = 30$); JCA ($n = 12$); normal children ($n = 9$)	Affymetrix U95AV2 array with >12,000 genes	SLE patients had overexpression of granulopoiesis-related and IFN-induced genes related to the presence of active disease	Bennett et al., 2003 [16]
SLE, RA, MS, IDDM ($n = 53$) and normal controls ($n = 9$)	Research Genetics/Invitrogen with >4300 genes	Autoimmune patients clustered separately from normal and immunized controls	Maas et al., 2002 [17]

IDDM, insulin-dependent diabetes mellitus; IFN, interferon; JCA, juvenile chronic arthritis; MS, multiple sclerosis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus.

by performing replicate hybridizations of the same sample on different arrays [14,17]. However, in general, replicate analyses are not required [18]. In some studies, confirmation of the microarray findings has been accomplished using independent methods such as real-time PCR [14,19] or detection of the encoded proteins [20]. Of interest in human studies are clinical correlations made with gene expression levels that fit with predicted changes. For example, in a study of childhood SLE, the only patient in complete remission was clustered with the healthy controls, suggesting that the signature expressed in the ill patients was disease-related [16], and in an MS trial of interferon-β (IFN-β) clinically-defined responders and non-responders showed differences in gene expression profiles [13].

The large amount of data generated in microarray experiments necessitates the use of filtering to permit focus on the genes of interest. Approaches to this issue have included requiring that each gene have a minimal intensity across all conditions [12,15], and that genes without significant changes be eliminated from further analysis [17]. For studies in PBMC populations, analyses are generally limited to the approximately 5000 genes that are expressed in these cells [16]. Other investigators have applied additional requirements, such as eliminating genes that show changes in expression levels with collection or shipping of the samples [15], although the advent of RNA stabilization tubes for blood collection may make this less of a concern in future studies.

Studies of the normal immune response

Most sudies of autoimmune disease have used normal controls that reflect the demographics of the patient population of interest. Disease controls have also been used, as in the case of juvenile polyarthritis patients who were compared to juvenile SLE patients [16]. We considered that prior to embarking on an analysis of immune responses in disease states, it would be of interest to establish parameters of the normal host response to an exogenous antigen challenge. This approach permitted verification of the feasibility of the design as well as establishment of a comparator for auto-immune diseases.

We chose to use the response of healthy control subjects to routine vaccination with the inactivated influenza vaccine to test whether this approach would produce measurable alterations in gene expression [17]. For these studies, the subjects ($n = 9$) were sampled before and at various times after immunization. The postvaccine samples were collected at three different time points: early (3 days), middle (6 to 9 days), and late (19 to 21 days). A self-organizing map algorithm was employed to compare the preimmune to the postimmune group. Not surprisingly, when all of the genes were included, this approach largely

showed each vaccinated individual to most closely resemble his or her own pretreatment pattern due to the large number of genes that did not change. To focus on the distinctions most specific for the vaccine intervention, the gene set was subjected to a filter using the Pathways program (Research Genetics) to remove genes that did not show significant variation, and this revealed groups of genes that were readily distinguished on the basis of pre- and post-immune status (Fig. 1). The results of these studies suggested that changes in expression levels of a subset of the measured genes could distinguish the pre- and post-immunization status. It also verified that PBMC populations could be readily used for gene expression analysis and that the changes observed were likely to reflect immune response parameters.

Studies in autoimmune disease
Multiple sclerosis

Multiple sclerosis is an organ-specific autoimmune disorder targeting myelinated fibers in the central nervous system. The disease may have many modes of presentation and has clinically distinct subsets. One report has described microarray findings in MS brain lesions using autopsy samples from human subjects [8]. This study highlighted several genes, including some involved in T-cell activation and neurotransmitters, which have potential relevance in designing targeted therapies. More relevant to the current discussion are studies that have been done using PBMCs from MS patients. Since brain lesions are not readily available for biopsy, the possibility of obtaining useful information from the peripheral blood in this disorder has great potential for clinical applications. In one study comparing PBMCs from MS patients to normal control subjects more than a thousand differentially expressed genes were found; 53 of these were used to discriminate normal subjects from MS patients [12]. Genes in the upregulated category included several encoding components of the tumor necrosis factor signaling pathway. Downregulated genes included heat shock protein-70, and others encoding proteins involved in cell cycling. A second report generated using MS patients is the only one that has examined longitudinal specimens from patients enrolled in a clinical trial [13]. Patients treated with IFN-β in this trial could be separated on the basis of MRI scan results into responders and nonresponders. Most responders showed changes in IFN-β-regulated genes, while few of the nonresponders showed these changes. Genes of interest encoded cytokines and chemokines (IL-8, granulocyte-macrophage colony-stimulating factor, IL-3 receptor) and signaling molecules (JNK1, Jun B, PKC-β). An implication of this study is that patients who have a greater chance of responding to IFN-β might be identified by gene expression patterns in PBMCs. This hypothesis remains to be verified in a larger group of patients.

Figure 1
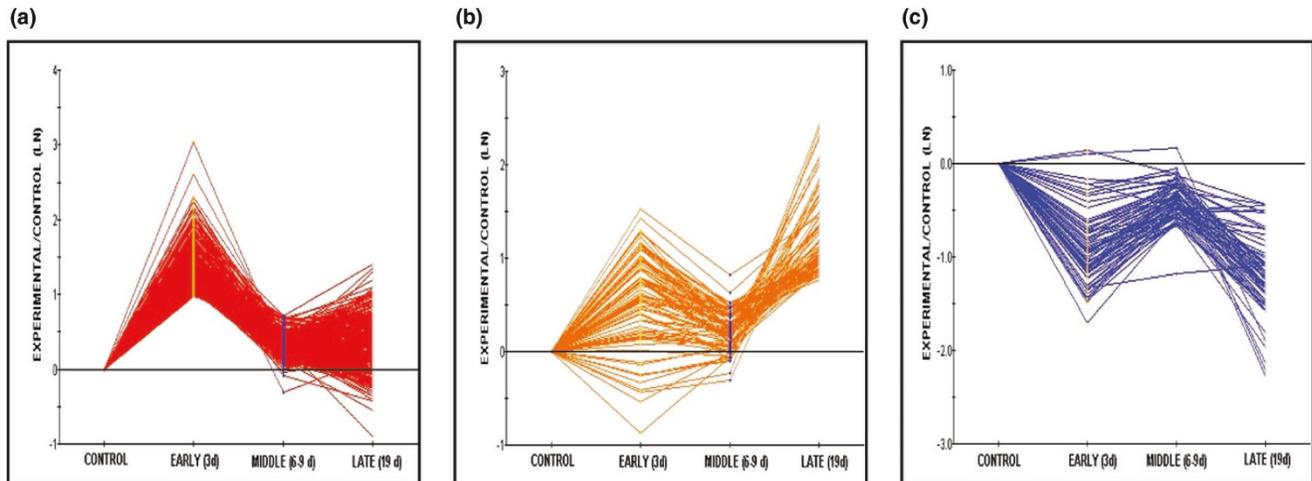
Kmeans analysis of the normal immune response in studies of Maas *et al.* [17]. Data are presented as the natural logarithm of the ratio of the experimental group to the control group indicated on the x-axis. Individual lines in the plot represent expression ratios of individual genes over the time course. Clustering patterns correspond to early upregulated genes **(a)**, late upregulated genes **(b)** and genes that were downregulated throughout the followup period **(c)**.

Systemic lupus erythematosus

Three studies of PBMC gene expression in lupus patients have been published [14–16]. The earliest of these, published in 2002, used a cytokine gene array. Most of the changes observed were in genes that had not previously been identified as contributing to the pathogenesis of SLE, consistent with the view that microarray experiments are a method of data discovery [3]. A major finding of this study in 21 SLE patients was that clustering analyses permitted clear separation of the patients from the controls, even with the relatively small number of genes (375) available on the array. No correlation with clinical disease status as measured by the SLE disease activity index (SLEDAI) score was seen, suggesting that the differences were related to the disease state itself and not to activity variables or medications.

A second, relatively large study compared 48 adult SLE patients to 42 healthy control subjects [15]. Clustering analysis grouped 37 of the 48 SLE patients together while the remaining patients were clustered with the control subjects. Most of the discriminatory genes were those that had higher expression levels in the SLE patients than in the controls. Especially notable was the finding that genes in the IFN-regulated pathway were upregulated in about half of the patients while control subjects expressed low levels. Furthermore, high levels of IFN-regulated gene expression could be used to identify patients who had more severe disease manifestations. A similar IFN signature has been described in pediatric SLE patients [16]. Children with SLE were generally clustered separately from controls, with the only exception being a

patient who did not have active disease. These two studies suggest that the IFN signature is related to disease activity and that blocking IFN pathways might have therapeutic efficacy in SLE.

Studies in four autoimmune diseases

Our group has studied gene expression in adult subjects with four autoimmune disorders: RA, MS, IDDM, and SLE [17]. In each instance, the patients were not restricted in any way other than satisfying appropriate diagnostic criteria (RA, SLE; [21,22]) or being identified by a specialist physician (MS, IDDM). Individuals were not excluded on the basis of any clinical variables such as what medications they were taking or how long they had had disease. PBMC samples from the autoimmune patients were compared to control subjects by flow cytometry and no significant differences were seen. In addition, expression levels of genes encoding activation markers (CD54, CD38, CD71) were not significantly different in the autoimmune patients than in control subjects [17]. These findings suggested that it was valid to compare gene expression levels in PBMC preparations between the two groups of subjects.

When unsupervised clustering was applied to the dataset, most of the autoimmune patients were clustered separately from normal or immunized controls (Fig. 2). The study group included a set of identical twins with lupus (SLE 8 and 9) and it was of interest that they were very closely related to each other, as would be predicted. We were not able to separate the autoimmune diseases from each other, although two of the IDDM patients did not

Figure 2
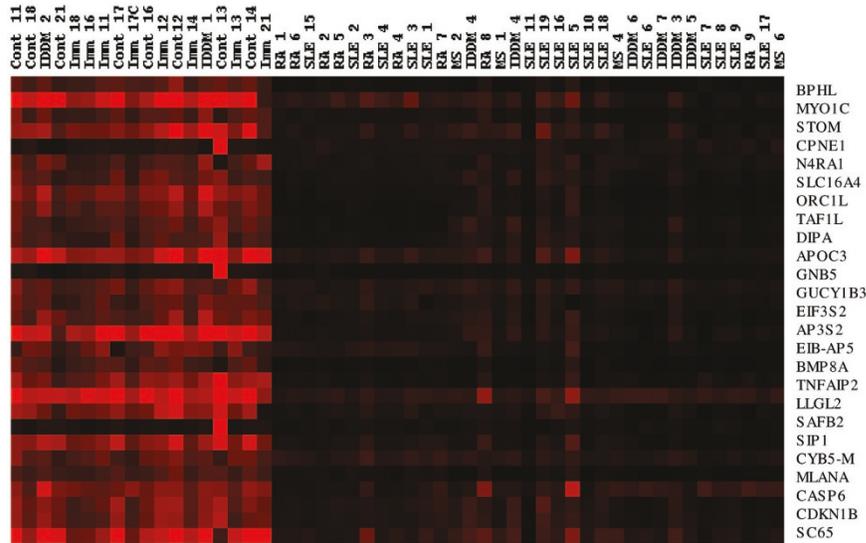
Cluster analysis using the self-organizing map algorithm of subjects studied by Maas *et al.* [17] including normal controls before (cont) and after vaccination (imm) compared to four types of autoimmune patients. Data were filtered to exclude genes that did not show significant change (2 S.D.) under any of the conditions. A portion of the total number of genes analyzed is shown.

cluster with the other autoimmune subjects. Whether in fact these patients may have had a form of diabetes intermediate between type I and II is a possibility that is currently under study in a larger group of subjects. Our inability to separate RA and SLE patients on the basis of the gene expression data differs from the results reported in a study of juvenile subjects, in which patients with chronic polyarthritis clustered separately from those with SLE [16]. Reasons for this difference are not apparent, but might include the larger number of genes analyzed in the juvenile subjects or basic differences in pathogenesis between childhood and adult forms of these disorders.

We further analyzed the data to identify genes that were most differentially expressed in autoimmune diseases relative to the normal immune response. For this analysis, the control data were from normal individuals prior to influenza vaccination and at the midpoint of the post-vaccine response (6–9 days). Two clusters of differentially expressed genes were identified that distinguished between all patients with autoimmune disease and normal individuals. One downregulated cluster included 117 genes that were consistently underexpressed in all four autoimmune groups compared to unimmunized or immunized controls. Several of these were related to apoptosis pathways and others to ubiquitin/proteasome function (Table 2). Inhibitors of various cellular functions were also present in this cluster. One of the most consistently underexpressed genes was *TP53*, which is an important component of the cellular response to damage and regulates normal cell responses including apoptosis [23].

Table 2

Differentially expressed autoimmune genes

Category	Gene symbol
Underexpressed in autoimmune	
Apoptosis	*TRADD, TRAP1, TRIP, TRAF2, CASP6, CASP8, TP53, SIVA*
Ubiquitin/proteasome	*UBE2M, UBE2G2, POH1*
Cell cycle inhibitors	*CDKN1B, CDKN2A, BRCA1*
Differentiation inducers	*LIF, CD24*
Enzyme inhibitors	*APOC3, KAL1*
Overexpressed in autoimmune	
Receptors	*CSF3R, HLA-DMB, HLALS, TGFBR2, BMPR2*
Inflammatory mediators	*MSTP9, BDNF, CES1, CYR61*
Signaling/2nd messenger	*FASTK, DGKA, DGKD*
Autoantigens	*GARS, GAD2*

Downregulation of p53 explains a significant portion of the differentially expressed genes in the autoimmune signature, suggesting that this single gene may be central to the autoimmune state. In other ongoing studies we have confirmed, by independent methods, that cellular damage response pathways that are dependent on p53 are defective in patients with RA [20] and MS (S Sriram and

T Aune, unpublished data). A cluster of 95 overexpressed genes was more heterogeneous, representing several distinct functional categories, including receptors, inflammatory mediators, signaling molecules and autoantigens (Table 2).

Our findings extended those of other groups (working in MS and SLE) by including patients with RA and IDDM, and by offering direct comparisons between them. The similar gene expression findings in these clinically diverse conditions are consistent with the hypothesis that autoimmune disorders share an underlying pathogenesis. Furthermore, the differences between autoimmune and vaccinated subjects suggest that autoantigens elicit responses that are distinct from the normal host defenses to exogenous antigens.

Classification and prediction based on gene expression profiles

Prediction of disease class is a major goal of microarray studies [18]. Since gene expression patterns permit clustering of patients with autoimmune disease from normal control subjects, there has been significant interest in using the gene expression data to classify disease subsets and predict responses to treatments. In one study of MS patients, two genes (*HIF2* and *CKS2*) were used to discriminate between MS patients and controls [12]. Although the correct prediction rate was 80%, the separation was not complete and some samples were misclassified. In SLE patients, Baechler et al. showed a significant correlation between gene expression data and number of SLE criteria ($r = 0.51$; $P = 0.002$), and they were also able to use the IFN score to distinguish SLE patients from controls with a high degree of accuracy ($P = 2.7 \times 10^{-7}$) [15]. More than half of the SLE patients, however, had IFN scores that were not distinguishable from controls. The IFN score could, therefore, only be used to distinguish patients with more active SLE. Studies in children with SLE also showed a correlation between the IFN signature and disease activity as measured by the SLEDAI, reinforcing the hypothesis that the IFN signature is a measure of disease activity or severity [16].

Inspection of gene expression data generated in our study of autoimmune subjects suggested that it was the downregulated set of genes that was most consistently altered, while the upregulated genes showed a greater degree of heterogeneity [17]. We were interested in the possibility that an equation derived from this set of downregulated genes might predict whether or not an unknown subject belonged to the autoimmune class. A set of 35 genes, that were significantly underexpressed in the SLE population compared to the control population, was selected [17]. To combine these 35 genes into an equation, we used an approach similar to linear discriminant analysis, which permits combination of many

Figure 3
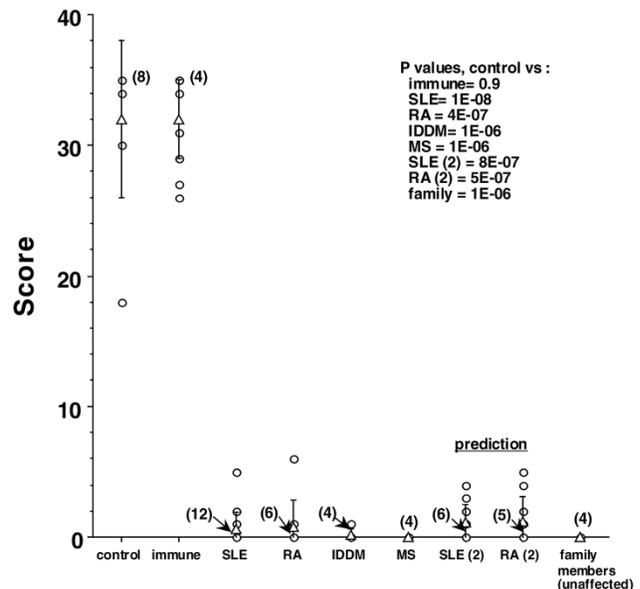
Classification and prediction of autoimmune disease. The score (y-axis) is shown for each individual sample analyzed form the different patient groups (x-axis). The 35 genes used to derive this score are found in the original reference (Maas *et al*; [17]).

measurements into a single value or score [24]. First, average gene expression values for the SLE and control groups were determined. The overall average was calculated from these two means [(SLE + control)/2]. Expression levels for each of the 35 genes in individual subjects from each group were then compared to this calculated mean. If the value was greater than the overall mean for that gene, then it was assigned a score of 1; if less than the mean it was assigned a score of 0. Thus, the maximum possible score was 35 and the minimum possible score was 0. For the unimmunized controls, scores were 18 or greater; most scores were close to the upper limit of 35; there was no significant difference for scores in these subjects after immunization (Fig. 3). In contrast, all four autoimmune groups showed scores close to the minimum value of 0. Differences between the autoimmune and control groups were highly significant ($P < 10^{-6}$).

To further test the predictive value of this equation, new sets of SLE and RA patients that were not included in the initial data analysis were subjected to scoring; none of these individuals had a score greater than 6, confirming that they belonged to the autoimmune set. Although normal subjects had generally high scores, this was not the case for four individuals tested who were first-degree relatives of autoimmune patients. All four of these normal individuals had a score of 0 in the 35-gene equation, indicating that they also carried the autoimmune signature (Fig. 3).

The autoimmune signature defined by the 35-gene equation most likely represents an inherited liability for development of disease rather than a consequence of the disease or its treatment. The are two reasons for this, the first of which is derived from the results in first-degree relatives, as described above. These persons did not have a clinical disease and were not being treated with immunosuppressive medications yet carried the entire set of downregulated genes. The second is the observation that MS and diabetes patients were not receiving the same drugs as those in the SLE and RA groups. Many of the MS patients were being treated with IFN without glucocorticoids; none of the IDDM patients were on immunosuppressives. In addition, since patients represented a broad range of disease activity and severity, it appears that these clinical variables did not impact on expression of this signature. Thus, the autoimmune signature may confer a liability for development of disease, while the specific disease syndrome that develops is likely dependent on additional factors that might include other genes or environmental stimuli like microbes or hormones [25]. The concept that different autoimmune diseases share basic features is, in fact, not new [26]. Furthermore, in clinical practice it is not uncommon for patients with features of more than one autoimmune syndrome to present diagnostic dilemmas [27]. Occurrence of multiple autoimmune diseases within a family is also relatively common, suggesting that similar genetic components can underlie very different clinical syndromes [28]. Studies in progress in autoimmune families suggest that many of the genes in the autoimmune signature display high levels of heritability (K Maas and T Aune, unpublished data).

Symbolic discriminant analysis of microarray data

Relationships between gene expression variables are very complex and patterns that have clinical significance may take many forms. It is likely that combinations of variables that utilize operators other than addition and subtraction will be revealing of significant relationships. Symbolic discriminant analysis (SDA) is an alternative approach that has been developed to identify complex gene relationships which may be nonlinear [29]. We used SDA to compare the gene expression data derived from normal individuals to patients with either RA or SLE [30]. Personal computers are not sufficiently powerful to cope with SDA requirements so analyses were carried out on four processors of the Vanderbilt Multi-Processor Integrated Research Engine (VAMPIRE), a 110-processor computer system running the Linux operating system. Cross validation was used to verify that associations detected were reproducible and not due to chance alone.

Results of these analyses yielded sets of genes that were repeatedly detected in the SDA-derived models. For RA compared to control 8 genes were identified and for SLE

Table 3

**Differentially expressed autoimmune genes detected by symbolic discriminant analysis**

Gene symbol	Gene description
RA versus control	
SR-BP1	Sigma receptor
PPP2R5A	Protein phosphatase 2, regulatory subunit B,
UFDIL	Ubiquitin fusion degradation 1-like
INADL	PDZ domain protein
CAPZAZ	Capping protein
CSPG2	Chondroitin sulfate proteoglycan 2
–	EST similar to 26s protease regulatory subunit 6
DEFA-1	Defensin-like
SLE versus control	
BMPR2	Bone morphogenic protein receptor type II
S100A11	Calguizzarin, calgranulin family member
NOL1	Nucleolar protein I
PPOX	Protophyrinogen oxidase
–	EST, similar to platelet factor 4
PPP3CA	Calcineurin A subunit

EST, expressed sequence tag; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus.

6 were identified (Table 3). These genes were used in model equations that were able to classify patients versus controls without overlap. Of note, the genes appearing in the models included a defensin and two calcium-binding-related proteins, both of which have been observed by other investigators to contribute to the PBMC signature for lupus [16]. It was not surprising that the genes identified by the SDA approach differed from those identified using clustering analyses. This is most likely due to the fact that the SDA method primarily considers gene–gene interactions rather than individual genes. Thus, the results suggest that some of the genes identified by the clustering method did not contribute to the discriminatory power beyond that which was provided by the pattern detected by SDA.

Other gene expression signatures

The autoimmune signature is only one of many gene expression patterns that will be of importance in patients with diseases that involve the immune system (Fig. 4). It is likely that some expression patterns will be specific for clinically distinct disorders within the autoimmune category. Consistent with this hypothesis, we have found that clustering analyses can be used to separate patients with early RA of 2 years' duration or less from patients with longstanding RA [19]. All of these RA patients fit the 35-

Figure 4
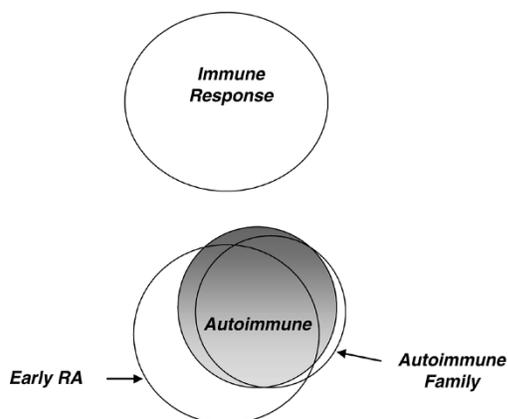
Schematic presentation of how the autoimmune gene expression signature might be compared to other gene expression patterns. Overlaps of the autoimmune patients with autoimmune families and with early RA patients have been described, while normal immune responses appear to be distinct.

gene autoimmune equation demonstrated here, and were subsequently clustered into two groups that correlated with disease duration but not with medications or disease activity (data not shown). It has been proposed that other signatures may be predictive of responses to drugs, which would provide direction for therapies [13]. It is also likely that other immune system disorders that are not associated with autoimmunity will have distinct patterns that fall outside of the autoimmune signature. As an example, preliminary studies of patients with allergic diseases or asthma indicate that there is a distinctly different pattern of PBMC gene expression consistent with altered immunopathogenesis for these disorders [31]. The rapid and robust changes in gene expression following influenza vaccination suggest that PBMC profiles also might be useful in early detection of infections.

Conclusion

The application of gene expression analysis to the study of human disease is a new and rapidly evolving area of investigation. These are powerful techniques that permit capture of many simultaneous processes and convert the findings into quantitative, reproducible data. The large numbers of data points that can be generated from a single individual make it likely that significant findings can emerge from smaller groups of subjects compared with previous approaches. Furthermore, the patterns and associations that are developed between individual genes afford a close look at molecular processes. One might consider that in diseases with complex etiologies, the study of one gene in 5000 people may be less informative about the disease process than the study of 2500 genes a few individuals. The former approach is critically dependent on choosing the correct gene for study. The

latter approach may reveal new genes of interest, and in that way has the potential to generate novel hypotheses. The fact that these powerful studies can be carried out in human subjects and not just in animal models is likely to advance discovery of new ways to diagnose, classify and treat human autoimmune disease.

Competing interests

N Olsen and T Aune hold equity interest in ArthroChip LLC, which is exploring the use of gene expression in diagnostics.

References

1. Brown PO, Botstein D: **Exploring the new world of the genome with DNA microarrays**. *Nat Genet* 1999, **21:**33-37.
2. Staudt LM, Brown PO: **Genomic views of the immune system**. *Annu Rev Immunol* 2000, **18:**829-859.
3. Moser KL, Gaffney PM, Grandits ME, Emamian ES, Machado DB, Baechler EC, Rhodus NL, Behrens TW: **The use of microarrays to study autoimmunity**. *J Investig Dermatol Symp Proc* 2004, **9:**18-22.
4. Marrack P, Kappler J, Kotzin BL: **Autoimmune disease: why and where it occurs**. *Nat Med* 2001, **7:**899-905.
5. Davidson A, Diamond B: **Autoimmune diseases**. *N Engl J Med* 2001, **345:**340-350.
6. Huh SJ, Paik DJ, Chung HS, Youn J: **Regulation of GRB2 and FLICE2 expression by TNF-alpha in rheumatoid synovium**. *Immunol Lett* 2003, **90:**93-96.
7. Carroll JM, McElwee KJ, King E, Byrne MC, Sundberg JP: **Gene array profiling and immunomodulation studies define a cell-mediated immune response underlying the pathogenesis of alopecia areata in a mouse model and humans**. *J Invest Dermatol* 2002, **119:**392-402.
8. Lock C, Hermans G, Pedotti R, Brendolan A, Schadt E, Garren H, Langer-Gould A, Strober S, Cannella B, Allard J, Klonowski P, Austin A, Lad N, Kaminski N, Galli SJ, Oksenberg JR, Raine CS, Heller R, Steinman L: **Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis**. *Nat Med* 2002, **8:**500-508.
9. van der Pouw Kraan TC, van Gaalen FA, Huizinga TW, Pieterman E, Breedveld FC, Verweij CL: **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* 2003, **4:**187-196.
10. Alexander JJ, Saxena AK, Bao L, Jacob A, Haas M, Quigg RJ: **Prominent renal expression of a murine leukemia retrovirus in experimental systemic lupus erythematosus**. *J Am Soc Nephrol* 2002, **13:**2869-2877.
11. Alcorta D, Preston G, Munger W, Sullivan P, Yang JJ, Waga I, Jennette JC, Falk R: **Microarray studies of gene expression in circulating leukocytes in kidney diseases**. *Exp Nephrol* 2002, **10:**139-149.
12. Bomprezzi R, Ringner M, Kim S, Bittner ML, Khan J, Chen Y, Elkahloun A, Yu A, Bielekova B, Meltzer PS, Martin R, McFarland HF, Trent JM: **Gene expression profile in multiple sclerosis patients and healthy controls: identifying pathways relevant to disease**. *Hum Mol Genet* 2003, **12:**2191-2199.
13. Sturzebecher S, Wandinger KP, Rosenwald A, Sathyamoorthy M, Tzou A, Mattar P, Frank JA, Staudt L, Martin R, McFarland HF: **Expression profiling identifies responder and non-responder phenotypes to interferon-beta in multiple sclerosis**. *Brain* 2003, **126:**1419-1429.
14. Rus V, Atamas SP, Shustova V, Luzina IG, Selaru F, Magder LS, Via CS: **Expression of cytokine- and chemokine-related genes**

127

in peripheral blood mononuclear cells from lupus patients by cDNA array. *Clin Immunol* 2002, **102:**283-290.

15. Baechler EC, Batliwalla FM, Karypis G, Gaffney PM, Ortmann WA, Espe KJ, Shark KB, Grande WJ, Hughes KM, Kapur V, Gregersen PK, Behrens TW: **Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus**. *Proc Natl Acad Sci U S A* 2003, **100:**2610-2615.

16. Bennett L, Palucka AK, Arce E, Cantrell V, Borvak J, Banchereau J, Pascual V: **Interferon and granulopoiesis signatures in systemic lupus erythematosus blood**. *J Exp Med* 2003, **197:**711-723.

17. Maas K, Chan S, Parker J, Slater A, Moore J, Olsen N, Aune TM: **Cutting edge: molecular portrait of human autoimmune disease**. *J Immunol* 2002, **169:**5-9.

18. Simon R, Radmacher MD, Dobbin K: **Design of studies using DNA microarrays**. *Genet Epidemiol* 2002, **23:**21-36.

19. Olsen NJ, Sokka T, Seehorn CL, Kraft B, Maas K, Moore J, Aune TM: **A Gene Expression Signature for Recent Onset Rheumatoid Arthritis in Peripheral Blood Mononuclear Cells**. *Ann Rheum Dis* 2004, in press.

20. Maas KH, Westfall M, Pietenpol J, Olsen N, Aune T: **Cellular damage response defects in human rheumatoid arthritis**. *Arthritis Rheum* 2003, **48**:277.

21. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, et al: **The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis**. *Arthritis Rheum* 1988, **31:**315-324.

22. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, Schaller JG, Talal N, Winchester RJ: **The 1982 revised criteria for the classification of systemic lupus erythematosus.** *Arthritis Rheum* 1982, **25:**1271-1277.

23. Schuler M, Green DR: **Mechanisms of p53-dependent apoptosis**. *Biochem Soc Trans* 2001, **29:**684-688.

24. Fisher RA: **The use of multiple measurements in taxonomic problems**. *Ann Eugen* 1936, **7:**179-188.

25. Aune TM, Maas K, Moore JH, Olsen NJ: **Gene expression profiles in human autoimmune disease**. *Curr Pharm Des* 2003, **9:** 1905-1917.

26. Moutsopoulus HM, Talal N: **Connective tissue diseases: one disease or many?** *Lupus* 1994, **3:**5-10.

27. Panush RS, Edwards NL, Longley S, Webster E: **'Rhupus' syndrome**. *Arch Intern Med* 1988, **148:**1633-1636.

28. Lin JP, Cash JM, Doyle SZ, Peden S, Kanik K, Amos CI, Bale SJ, Wilder RL: **Familial clustering of rheumatoid arthritis with other autoimmune diseases**. *Hum Genet* 1998, **103:**475-482.

29. Moore JH, Parker JS, Hahn LW: **Symbolic discriminant analysis for mining gene expression patterns**. *Lecture Notes in Artificial Intelligence* 2001, **2167:**372-381.

30. Moore JH, Parker JS, Olsen NJ, Aune TM: **Symbolic discriminant analysis of microarray data in autoimmune disease**. *Genet Epidemiol* 2002, **23:**57-69.

31. Tanner SB, Yelverton R, Kraft B, Aune TM, Olsen NJ: **A gene expression signature for allergic rhinitis in peripheral blood**. *Clinical Immunology* 2003, **Suppl 1:**S133.