

Review

Microarray analysis of gene expression in lupus

Mary K Crow¹ and Jay Wohlgemuth²

¹Mary Kirkland Center for Lupus Research, Hospital for Special Surgery, New York, NY, USA

²Expression Diagnostics, Inc, South San Francisco, CA, USA

Corresponding author: Mary K Crow (e-mail: crowm@hss.edu)

Received: 26 Aug 2003 Revisions requested: 5 Sep 2003 Revisions received: 22 Sep 2003 Accepted: 1 Oct 2003 Published: 13 Oct 2003

Arthritis Res Ther 2003, **5**:279-287 (DOI 10.1186/ar1015)

© 2003 BioMed Central Ltd (Print ISSN 1478-6354; Online ISSN 1478-6362)

Abstract

Recent advances in the study of global patterns of gene expression with the use of microarray technology, coupled with data analysis using sophisticated statistical algorithms, have provided new insights into pathogenic mechanisms of disease. Complementary and reproducible data from multiple laboratories have documented the feasibility of analysis of heterogeneous populations of peripheral blood mononuclear cells from patients with rheumatic diseases through use of this powerful technology. Although some patterns of gene expression, including increased expression of immune system cell surface activation molecules, confirm previous data obtained with other techniques, some novel genes that are differentially expressed have been identified. Most interesting is the dominant pattern of interferon-induced gene expression detected among blood mononuclear cells from patients with systemic lupus erythematosus and juvenile dermatomyositis. These data are consistent with long-standing observations indicating increased circulating interferon- α in the blood of patients with active lupus, but draw attention to the dominance of the interferon pathway in the hierarchy of gene expression pathways implicated in systemic autoimmunity.

Keywords: gene expression, interferon, microarray, statistical algorithms, systemic lupus erythematosus

Introduction: the dawning of the microarray era

The concept that the identification of genes that are differentially expressed in a disease state will elucidate disease mechanisms has driven the development of new technology. Earlier approaches, including Northern blotting, polymerase chain reaction (PCR), and RNase protection, have permitted analysis of small numbers of gene transcripts, but the value of characterizing a broad spectrum of gene products expressed in a cell population or in a disease state has stimulated the invention of more sophisticated tools. Subtractive hybridization and representational difference analysis, comparing gene expression in two cell populations, are time-intensive approaches used in the late 1990s to assist in gene discovery and to identify molecular pathways relevant to a disease. Microarray analysis, a system in which thousands of oligonucleotide sequences are spotted on a solid substrate, usually a glass slide, and

RNA-derived material from a cell population is hybridized to the gene array, is an innovative technology that has already changed our understanding of the mechanisms that underlie disease [1].

The utility of microarray analysis of gene expression was demonstrated impressively in 2000 when Alizadeh and colleagues used this technique to study the malignant cell population of patients with diffuse large B cell leukemia [2]. Although individual patient samples were not readily differentiated on the basis of traditional cell surface phenotypic markers, microarray analysis discerned two discrete tumor groups: those with a gene expression profile similar to that of germinal center B cells from healthy individuals and those with a profile similar to that of activated mature B cells. Significantly, these two groups were characterized by markedly different clinical courses: B cell lymphomas of the germinal center type had a 5-year survival

CNTRF = ciliary neurotrophic factor receptor; IFN = interferon; JCA = juvenile chronic arthritis; JDM = juvenile dermatomyositis; PBEF = pre-B-cell colony-enhancing factor; PBMC = peripheral blood mononuclear cells; PCR = polymerase chain reaction; RA = rheumatoid arthritis; SAM = significance analysis of microarrays; SLE = systemic lupus erythematosus.

of 76%, whereas lymphomas of the activated B cell type were associated with a 16% 5-year survival. As striking as this study was, at that time confidence was not high that microarray analysis of gene expression could be successfully applied to heterogeneous populations of cells, cells that were not monoclonal.

Numerous investigations over the past several years have demonstrated that significant and useful microarray data can be derived from more complex cell samples, including cell populations from peripheral blood. Although such studies face challenges in data interpretation, several laboratories have used microarray analysis to study mononuclear cells from patients with autoimmune diseases. When studying mixed cell populations, gene expression profiling can successfully detect differential gene expression in specific cell types present in the samples under comparison. However, it can also measure and reflect the cellular composition of the sample. The contribution of variable enrichment of a cell population in a sample can be sorted out by combining cell sorting or histology with expression profiling experiments. Cell sorting can also be used as an initial step in sample preparation to enrich for specific cell types in a cell mixture to overcome sensitivity and specificity limitations of arrays.

The recent advances in the analysis of broad gene expression patterns have rapidly led to new insights into pathogenic mechanisms of rheumatic diseases and are supporting new initiatives in therapeutic drug development. Most striking are microarray data that have refocused attention on the interferon (IFN) pathway in systemic lupus erythematosus (SLE) [3–6].

Analysis of microarray data

A statistical analysis of appropriately normalized microarray data is as important as the cell preparation, initial hybridization, and data extraction in deriving valuable information from this technology. Arrays are useful because they allow large-scale screening for differential gene expression. However, the thousands of variables in the analyses represent a significant statistical problem. Theoretically, when using an array with 1000 genes and a standard *t*-test with a significance at the $P < 0.05$ level, one would expect to identify 50 genes that are 'differentially expressed' by chance alone. Corrections for multiple comparisons, such as the Bonferroni method, can be used to address this problem but tend to lose efficacy with very large numbers of comparisons, and some differentially expressed genes might not appear as significant [7,8]. Significance analysis of microarrays (SAM) is an approach that uses an estimate of the false detection rate to make an estimate of significance under conditions in which thousands of data points are being compared [9]. We have learned that no one statistical algorithm is sufficient to derive an accurate view of the genes significantly over-

expressed or underexpressed in one population compared with another.

Although a comparative analysis of the current statistical approaches to microarray data is beyond the scope of this review, we will direct the reader to several of the most useful algorithms that we and others have used to derive a comprehensive data set that includes genes most significantly differentially expressed between cell populations (Table 1). When analyzing microarray data from two sets of patient-derived samples, we first use the SAM algorithm with parameters set to determine the genes significantly overexpressed or underexpressed in the patient group of interest compared with the control group [9]. The false detection rate is set at approximately 5% (indicating that in only 5 cases out of 100 will a gene be erroneously identified as significantly differentially expressed). Additional algorithms, including supervised harvesting classification and a method termed 'shrunken centroids', are then applied to the data set [10–16]. Genes that either are identified in several of the algorithms or are most highly ranked in one of the algorithms are then used in a hierarchical clustering approach to identify additional genes that are coexpressed with the most significantly differentially expressed genes.

Array data can be internally validated by, for example, carrying multiple probes for the same gene. When data from all such probes in a particular experiment are identified as correlating in an array experiment, the probability of the finding being real is increased. Further, when data from multiple genes that are known to be coexpressed in a cluster or cellular pathway are correlated, the result has greater significance than data from a single gene. Microarray analysis cannot be relied on to develop a definitive rank order of the most significant gene products associated with a particular disease or cell state. Rather, the technology is most useful in drawing attention to pathways, and sometimes individual genes, that are most relevant to the recent *in vivo* experience of the cell population being studied.

As revealing as microarray data can be, confirmation of the data using more accurate, quantitative, and less variable approaches, such as real-time PCR, and validation in a second patient population are essential for drawing meaningful conclusions [17].

Early microarray data from the use of SLE peripheral blood

In early 2003, Rus and colleagues reported data from a study of gene expression in peripheral blood mononuclear cells (PBMC) from 21 lupus patients and 12 controls [18]. The microarray assay used (Panorama Cytokine Gene Array membranes; Sigma Genosys, Inc) included 375 genes enriched in cytokines, chemokines, cell surface

Table 1**Statistical algorithms used in analysis of microarray data**

Statistical algorithms	Characteristics	References	Sources
Significance analysis of microarrays (SAM)	Identifies differentially expressed genes between sample sets; estimates significance for genes; considers large numbers of genes in array experiments	[9]	Stanford University, http://www-stat.stanford.edu/~tibs/x-mine , Brisbane, CA, http://www.x-mine.com/
Hierarchical clustering	Unsupervised clustering; clusters genes with similar expression patterns; clusters samples with similar expression patterns	[15]	University of California, Berkeley, http://rana.lbl.gov/EisenSoftware.htm
Supervised harvesting classification	Class prediction; identifies subset of genes that best classify samples as gene sets; estimates accuracy of gene set on prospective population	[10]	x-mine, Brisbane, CA, http://www.x-mine.com/
Classification and regression trees (CART), multiple additive regression trees (MART)	Class prediction; develops decision trees to classify samples using the expression of a subset of genes; estimates accuracy of the gene panel on a prospective set	[12,14]	CART: Salford Systems, http://www.salford-systems.com/ MART: Stanford University, http://www-stat.stanford.edu/~jhf/R-MART.html or Salford Systems
Shrunken centroids (prediction analysis for microarrays, PAM)	Class prediction; identifies subset of genes that best classify samples as gene sets; estimates accuracy of gene set on prospective population	[11]	Stanford University, http://www-stat.stanford.edu/~tibs/PAM/index.html
Affymetrix MAS 5.0			Affymetrix
GeneSpring			Silicon Genetics
Pathways 3.0			Research Genetics

receptors, and other immune-system cell surface molecules, including adhesion molecules. Data analysis comprised several approaches: selection of genes with mean expression in the SLE group that was more than 2.5-fold that observed in the healthy control group; genes that were significantly different between groups with the use of the Mann-Whitney *U*-test at a significance level of $P < 0.05$; and SAM analysis with a false detection rate of 8%. Fifty genes were identified as differing by more than 2.5-fold, and 20 genes differed between groups on the basis of the Mann-Whitney *U*-test, of which 15 differed by more than 2.5-fold between groups. The importance of confirming microarray data with an additional technique was clearly illustrated in this report. The gene that showed the greatest fold difference between SLE and control groups by microarray, that encoding ciliary neurotrophic factor receptor α (*CNTFR*), could not be confirmed by reverse transcriptase PCR (RT-PCR). Increased expression of two other genes that were also studied by RT-PCR, *CXCR2* and that encoding pre-B-cell colony-enhancing factor (*PBEF*), was confirmed, although the fold increase as assessed by RT-PCR was less than estimated by microarray. Of the genes significantly overexpressed, the products of some, including *MMP3*, *TNFRSF1B* (TNF receptor 2), *IL1B*, and *FCGR1A*, have been documented to be increased in SLE. Among other genes with increased expression were *TNFSF10* (encoding TRAIL), *TNFRSF10C* and *TNFRSF10D* (TRAIL recep-

tors 3 and 4), *IL1RAP* (interleukin-1 receptor accessory protein), *TGFBR3* (transforming growth factor- β receptor type III), and *CCR7* (a T cell chemokine receptor important for the recruitment of CD4 T cells to the periarterial lymphoid sheath). Although the Rus study did not provide an opportunity to detect the expression of genes that were not obviously directly related to immune system function, it drew attention to several genes that had not been studied previously in detail in SLE. Overall, the data presented in this study supported the value of the microarray approach for detecting genes differentially expressed among PBMC from lupus patients.

A second early report came from Maas and colleagues, who showed PBMC microarray data from a small number of patients with SLE ($n=9$), rheumatoid arthritis (RA; $n=9$), type I diabetes mellitus ($n=5$) or multiple sclerosis ($n=4$), along with nine control subjects before and after immunization with influenza vaccine [19]. The array used (Research Genetics GF-211 membrane) included more than 4000 genes, and the statistical analysis was based on Eisen's Cluster and Treeview software, as well as the Research Genetics Pathways 3.0 program used to locate differentially expressed genes in pathways related to the immune system [15]. As in the Rus study, microarray analysis of unfractionated PBMC provided data that seemed to be reproducible and statistically significant, and characterization of the cell composition of the PBMC

indicated that a variable presence of mononuclear cell populations could not account for differential gene expression. However, the analysis was not able to distinguish a gene expression profile that was distinct for SLE as compared with RA or other autoimmune diseases. Ninety-five genes were identified that distinguished the samples from all four autoimmune diseases from healthy controls, including those encoding the cell surface receptors TGFBR2, CSF3R, and BMPR2, which were overexpressed in the autoimmune patients, and several genes implicated in apoptosis (*TRADD*, *TRAF2*, *CASP6*, *CASP8*), which were underexpressed. It is of interest that *ADAR*, an IFN-induced gene encoding the RNA-specific adenosine deaminase, was highly overexpressed in most patients with autoimmune disease. Increased RNA editing, dependent on the ADAR protein, has been identified in T cells in patients with SLE [20]. The patterns seen in the patients with autoimmune disease were distinct from those observed in samples from healthy subjects who had received influenza vaccination, suggesting that the pathways involved in the immunopathogenesis of autoimmune disease do not simply reflect an active immune response to an antigen. This study relied mainly on clustering algorithms to identify genes that were differentially expressed among the study groups and also eliminated from analysis genes whose expression levels did not vary by more than 3SD from their means, supporting our view that multiple approaches, including algorithms such as SAM, are useful in discerning gene expression relevant to disease-specific molecular pathways.

Interferon-induced gene expression in rheumatic diseases

A gene expression profile identified by microarray analysis and consistent with IFN-mediated gene transcription in a rheumatic disease was first reported by Tezak and colleagues in a study of muscle biopsy tissue from four patients with juvenile dermatomyositis (JDM) [21]. Data from biopsies were compared with gene expression data from two healthy control peripheral blood samples as well as previously obtained microarray data from muscle biopsy samples from patients with Duchenne muscular dystrophy by using Affymetrix HuFL GeneChips. Genes showing at least a twofold difference in comparisons of JDM samples with each of two control samples were used to develop a list of differentially expressed genes. Ninety-one genes showed more than twofold increased expression, and 87 genes showed more than twofold lower expression. It should be noted that in studies using small numbers of patient samples, although genes might be identified that are truly differentially expressed between the samples, the variable expression patterns might be unrelated to the disease state. In this study, fold differences between patient samples and controls were highly variable (ranging from 3.8-fold to 96.4-fold higher in patients than in controls), but the list of differentially expressed genes was

striking for its enrichment in those that have been linked to IFN. *MX1*, *MX2*, *G1P3*, *IRF7*, and *C1ORF29* were among the IFN-induced genes overexpressed in the JDM muscle samples. Increased expression of several genes, including *G1P3* (encoding IFN-induced 6-16) and *CDKN1A* (p21 cyclin kinase inhibitor), were confirmed by real-time PCR, and *G1P3* expression was also identified in JDM peripheral blood. It is of interest that there was no significant increase in mRNA for either IFN- α or IFN- γ , although IFN- γ protein was seen in JDM muscle by using immunohistochemistry. The authors interpreted their data as consistent with effects of both IFN- α/β and IFN- γ on gene expression. In addition to this set of IFN-induced genes, gene expression profiles indicating ischemia and myofiber degeneration and regeneration were detected.

Data supporting a pathogenic role for type I IFN (predominantly IFN- α) in SLE have been available for 25 years, with type II IFN (IFN- γ) also being implicated in murine models of lupus [22–31]. It has only been recently that microarray data from several laboratories have refocused attention on this important cytokine and its downstream targets. Studies from our laboratory and from others have used more extensive microarrays than those used in the two SLE studies described above to characterize the broad gene expression profile operative in the peripheral blood of patients with SLE [3–6]. Most striking is an 'IFN signature', a prominent overexpression of mRNAs encoded by genes regulated by IFNs and similar to that detected in JDM muscle.

Detection of an interferon signature in SLE by using microarray technology

Ambitious projects aimed at characterizing broad gene expression profiles in large numbers of SLE PBMC samples have come to fruition in 2003. Baechler and colleagues have reported microarray data from 48 SLE patients and 42 healthy controls with Affymetrix U95A GeneChips [4]. After eliminating genes that were highly sensitive to induction *ex vivo*, 4566 genes remained for analysis. An initial set of genes was selected on the basis of an unpaired Student's *t*-test (apparently without correction for multiple comparisons) and further selection based on a more than 1.5-fold difference in expression between lupus samples and controls, a difference of at least 100 units in the expression value, and $P < 0.001$ by *t*-test. SAM analysis was not used in this study. In the full data set, 161 genes fulfilled all of these criteria. Hierarchical clustering of these genes was then performed to identify gene expression patterns among the study samples. As in the JDM study, a striking increase in the expression of a group of genes previously reported to be induced by IFN was observed in about half of the SLE subjects. The authors identified 23 of the 161 differentially expressed genes as targets of IFN by determining gene expression of PBMC cultured for 6 hours either with IFN- α plus IFN- β or with

IFN- γ . It should be noted that the patterns of gene expression induced in PBMC by type I and type II IFNs can vary with time after initial stimulation. The data from Baechler, then, provide only a partial assessment of IFN-regulated genes at a single point in time.

The pattern of expression of the IFN-regulated genes was complex (Table 2). Eleven of the IFN-regulated genes that were differentially expressed between SLE and control PBMC clustered together and were preferentially induced by a combination of IFN- α and IFN- β . However, two genes preferentially induced by IFN- γ (*SERPING1* and *FCGR1A*) were also significantly increased in the SLE patients and clustered with the IFN- α/β -induced genes. Additional genes, preferentially induced by IFN- α/β but not clustering with the major IFN-induced group (*CD69*, *RGS1*, *IL1RN*, and *AGRN*), were also increased in the SLE group, and two genes repressed by IFN- α/β were significantly underexpressed in the SLE patients. Three genes significantly overexpressed in SLE were decreased in expression by both IFN- γ and IFN- α/β (*EREG*, *THBS1*, and *ETS1*). These are decreased somewhat more by IFN- γ than by IFN- α/β . Taken together, these SLE data, along with the very useful information about genes induced by type I and type II IFNs, support the type I IFNs as being particularly important in the gene expression pattern that distinguishes SLE PBMC from those of healthy controls. Confirmation of these data by a more quantitative technique will be essential to determine more precisely the relative expression of this gene set in patients with SLE, as well as those with other autoimmune diseases.

In addition to the IFN-regulated genes, the Baechler study documented an increased expression of genes encoding immune system activation antigens, including *TNFR6* (encoding Fas), *CD54* (ICAM-1), and *CD69* in the SLE samples, along with *FCGR1A* (as observed in the Rus study) and *FCGR1A* (Table 3) [4]. Other genes detected by Rus and colleagues were also identified (*IL1B*, *IL1RB*). Genes with decreased expression in the SLE samples included *TCF3* (encoding transcription factor E2 α), important in B cell development; *TCF7* (transcription factor 7 [T-cell specific, HMG box]), a polymorphism of which has been associated with type I diabetes; and *LCK* (lymphocyte-specific tyrosine kinase), which mediates T cell activation.

A second extensive gene expression study by Bennett and colleagues also used Affymetrix U95AV2 microarrays to analyze PBMC from 30 pediatric lupus patients ranging in age from 6 to 18 years (mean age 13), 12 patients with juvenile chronic arthritis (JCA), and 9 healthy control children [5]. Of the 30 SLE patients, 18 were studied no more than 1 year after diagnosis, reflecting the gene expression profile at a point in time closer to the onset of symptoms than is usually possible to document in adult

SLE patients. Data were analyzed with a correction for multiple comparisons. With this approach, 15 genes were found differentially expressed between SLE patients and healthy controls, and 14 of those 15 were identified as targets of IFN. Several of the most significantly differentially expressed genes were among those identified in the Baechler study (*C1ORF29*, *MX1*, *LY6E*, *PLSCR1*, and *APOBEC3B*) and others identified with less stringent statistical criteria were also found in the Baechler study (Table 2) or are closely related to genes identified by Baechler (*OAS1*, *OAS2*, *MX2*). In addition, Bennett identified *IFI44* (encoding hepatitis C-associated microtubular aggregate protein), *IFIT4* (termed *CIG49* by those authors), *CIG5* (viperin), and *C1ORF29* as nearly universally expressed in their SLE subjects. The JCA samples did not demonstrate overexpression of IFN-induced genes.

Although neither the Baechler study nor the Bennett study confirmed the IFN-regulated gene expression signature by using more quantitative techniques, the similarity of results derived from the two studies is remarkable, strongly supporting the significance of the IFN pathway in SLE and also demonstrating the power of microarray technology, even when applied to heterogeneous populations of peripheral blood cells.

Additional observations by Bennett and colleagues included the significant overexpression of *DEFA3*, encoding the neutrophil-specific $\alpha 3$ defensin that is predominantly expressed in precursors of mature polymorphonuclear cells [5]. Sorting of granular cells identified by flow cytometry confirmed a population of early granulocyte cells not present in mononuclear cell populations from controls. *DEFA3* and another neutrophil gene *FPRL1* (encoding formyl peptide receptor-like-1), along with several of the IFN-induced genes, were highly correlated with disease activity as measured by the Systemic Lupus Erythematosus Disease Activity Index.

Our laboratories have performed microarray analysis on PBMC samples from 22 SLE, 15 RA, 8 osteoarthritis, 2 JCA, and 9 control PBMC samples and have detected a gene expression signature virtually identical to that described by the other groups (Table 2) [3]. Importantly, our data were derived from a microarray (proprietary to Expression Diagnostics, Inc) distinct from that used by Baechler and Bennett and included more than 8000 gene sequences, most of which were identified from subtracted and normalized cDNA libraries isolated from resting and activated leukocytes. We have found that, in contrast to the RA, osteoarthritis, and JCA patients, and healthy controls, most adult SLE patients express the IFN gene signature (example data shown in Fig. 1). Moreover, we have confirmed the increased expression of several of those overexpressed genes by using real-time PCR analysis and in a second cohort of SLE patients [3]. Of great interest is the

Table 2

Interferon-induced genes identified in large-scale microarray analyses of SLE PBMC

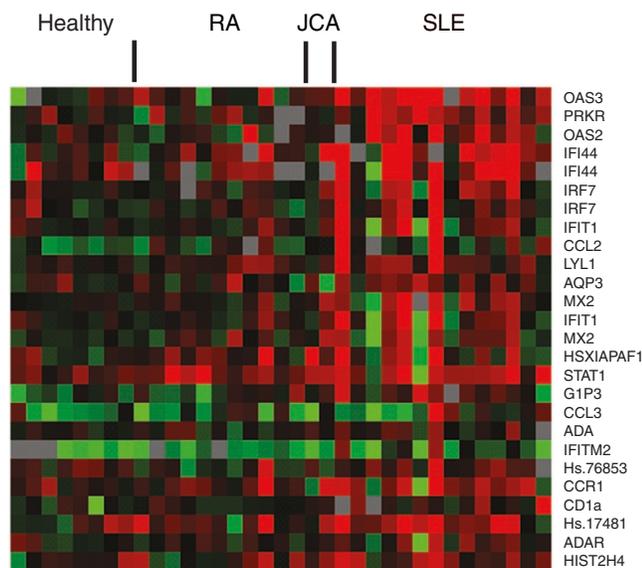
Gene	Protein	Response of PBMC to IFN- α/β (type I)/IFN- γ (type II)		Expression in SLE compared with control PBMC		
		Reference: [4] [*]	[4] [†]	[5] [‡]	[3] [§]	[6]
<i>IFIT1</i>	Interferon-induced protein with tetratricopeptide repeats-1	18.3		Up	Up	Up
<i>OASL</i>	2'-5'-oligoadenylate synthetase-like	16.3	Up	Up		Up
<i>LY6E</i>	Lymphocyte antigen 6 complex, locus E	14.9	Up	Up	Up	Up
<i>OAS2</i>	2'-5'-oligoadenylate synthetase	12.7		Up		Up
<i>OAS3</i>	2'-5'-oligoadenylate synthetase	NA	NA		Up	
<i>IFI44</i>	Hepatitis C microtubular aggregate protein	10.7		Up	Up	
<i>MX1</i>	Myxovirus resistance 1	9.8	Up	Up	Up	
<i>G1P3</i>	Interferon, alpha-inducible protein (IFI-6-16)	7.1	Up		Up	
<i>PRKR</i>	Protein kinase, interferon- α -inducible double-stranded RNA-dependent	6.9	Up		Up	
<i>IFIT4</i>	Interferon-induced protein with tetratricopeptide repeats 4	6.8		Up	Up	Up
<i>PLSCR1</i>	Phospholipid scramblase 1	6.6	Up	Up	Up	
<i>C1ORF29</i>	Hypothetical protein expressed in osteoblasts; similar to IFI44	6.1	Up	Up	Up	
<i>HSXIAPAF1</i>	XIAP-associated factor-1	6.1	Up	Up	Up	
<i>G1P2</i>	Interferon, alpha-inducible protein (IFI-15K)	5.9	Up	Up		
Hs. 17518 (<i>Cig5</i>)	Viperin	5.5		Up		
<i>IRF7</i>	Interferon regulatory factor 7	4.6		Up	Up	
<i>CD69</i>	Early T-cell activation antigen	4.1	Up			
<i>LGALS3BP</i>	Lectin, galactoside-binding, soluble, 3 binding protein	3.8	Up	Up		
<i>IL1RN</i>	Interleukin-1 receptor antagonist	3.5	Up			Up
<i>APOBEC3B</i>	Phorbolin 1-like	2.0	Up	Up	Up	
<i>RGS1</i>	Regulator of G-protein signaling 1	1.8	Up			
<i>AGRN</i>	Agurin	> 71.2 (α/β and $\gamma \rightarrow < 0$)	Up	Up		
<i>EREG</i>	Epiregulin	1.3 (α/β and $\gamma \rightarrow < 0$)	Up			
<i>THBS1</i>	Thrombospondin 1	1.3 (α/β and $\gamma \rightarrow < 0$)	Up			
<i>ETS1</i>	v-ets erythroblastosis virus E26 oncogene homolog 1	1.2 (α/β and $\gamma \rightarrow < 0$)	Up			
<i>ADAM9</i>	A disintegrin and metalloproteinase domain 9	1.1 (α/β and $\gamma \rightarrow < 0$)	Up			
<i>SERPING1</i>	Serine (or cysteine) proteinase inhibitor (C1 inhibitor)	0.85	Up		Up	
<i>USP20</i>	Ubiquitin specific protease 20	0.25 (α/β and $\gamma \rightarrow < 0$)	Down			
<i>MATK</i>	Megakaryocyte-specific tyrosine kinase	< 0.13 ($\alpha/\beta \rightarrow < 0$)	Down			
<i>FCGR1A</i>	Fc fragment of IgG, high-affinity Ia receptor	0.08	Up		Up	

The genes and corresponding proteins listed were identified as significantly differentially expressed in SLE and healthy control PBMC in the study by Baechler and colleagues [4], or were among the most commonly overexpressed transcripts among SLE PBMC in the study by Bennett and colleagues [5]. The significant overexpression of these genes in SLE compared with control PBMC in microarray data sets from Crow and colleagues [3] or Han and colleagues [6] is also noted. In addition, genes identified by Baechler and colleagues as both regulated by type I (IFN- α/β) or type II (IFN- γ) and differentially expressed by SLE PBMC are noted. A ratio of gene expression induced in healthy PBMC by IFN- α/β (1000 U/ml for 6 hours) compared with IFN- γ (1000 U/ml for 6 hours) for each gene was calculated by determining the net microarray score for each of four control samples studied by Baechler and colleagues (stimulated microarray score minus unstimulated microarray score) for both IFN- α/β and IFN- γ stimulation, determining the average of the four scores, and dividing the IFN- α/β score by the IFN- γ score for each gene. In some cases, scores for both IFN- α/β - and IFN- γ -induced gene expression were less than background [indicated as (α/β and $\gamma \rightarrow < 0$)]. When the score for either IFN- α/β -induced or IFN- γ -induced gene expression was less than the background, the score for the lower value was replaced with a score of 50, to permit the calculation of an approximate ratio. Abbreviations: JCA, juvenile chronic arthritis; NA, not available; OA, osteoarthritis; RA, rheumatoid arthritis. ^{*}The microarray system used was an Affymetrix U95A GeneChip (Affymetrix, Santa Clara, CA); 4566 genes were analyzed. Four healthy donors of PBMC were studied. [†]The microarray system used was an Affymetrix U95A GeneChip; 4566 genes were analyzed. The donors of PBMC studied were 48 with SLE and 42 healthy. [‡]The microarray system used was an Affymetrix U95AV2 GeneChip; about 4600 genes were analyzed. The donors of PBMC studied were 30 pediatric with SLE, 12 with JCA, and 9 healthy children. [§]The microarray system used was a proprietary microarray from Expression Diagnostics, Inc; 8143 oligonucleotides are represented. The donors of PBMC studied were 22 with SLE, 15 with RA, 8 with OA, 2 with JCA, and 9 healthy. ^{||}The microarray system used was a Mergen ExpressChip DNA Microarray (Mergen, Ltd, San Leandro, CA); 3002 genes are represented. The donors of PBMC studied were 10 with SLE and 18 healthy. ^{||}Identified as capicua homolog in some studies.

Table 3**Selected gene families overexpressed or underexpressed in SLE PBMC**

Gene families overexpressed	Examples	Gene families underexpressed	Examples
IFN target genes	See Table 2	Transcription factors	<i>TCF3, TCF7</i>
TNF and TNF receptor families	<i>TNFSF10</i> (TRAIL), <i>TNFRSF10C</i> (TRAIL receptor 3), <i>TNFR6</i> (Fas)	Kinases	<i>LCK</i>
Chemokines and chemokine receptors	<i>CCR7, CXCR2</i>	T cell receptors	<i>TCRB, TCRD</i>
Cell surface activation antigens	<i>CD69</i>	C-type lectins	<i>KLRB1</i>
Fc receptors	<i>FCGR1A, FCGR2A</i>		
Metalloproteinases	<i>MMP3, MMP9</i>		
Defensins	<i>DEFA3, F2RPA</i>		

Genes listed have been identified in microarray studies described in [3–6,8,19].

Figure 1

Exemplary gene sequences that cluster with PRKR and OAS3. Hierarchical clustering was performed on the total study population to determine genes that cluster with PRKR and OAS3. A visual demonstration of the expression of a selection from those genes, comprising a partial IFN signature, is shown. Data are shown from a subset of SLE samples tested ($n=14$) and from rheumatoid arthritis (RA) ($n=11$), juvenile chronic arthritis (JCA) ($n=2$), and control samples ($n=8$). Relative expression compared with an internal control ranged from approximately -0.5 (bright green) to 0.5 (bright red).

identity of the stimulus for the IFN-induced gene expression signature. It is noteworthy that none of the SLE or JDM studies identified significantly increased expression of type I or type II IFN mRNA, with the exception of a recent report by Han and colleagues [6] that detected IFN- ω , another type I IFN species, by using a distinct microarray system (Mergen ExpressChip DNA Microarray System

Human HO4) in a study of 10 SLE patients and 8 healthy controls. The Han study also detected an increased expression of several of the genes identified in the other three studies (Table 2). In contrast to the Baechler and Bennett reports, which note that type I and II IFNs are undetectable in serum from most SLE patients, we have measured plasma IFN- α protein by ELISA and are currently analyzing those data in relation to the IFN target gene expression data, as well as measures of clinical disease activity. Importantly, our data provide direct support for IFN- α in the gene expression profile observed in SLE PBMC (KA Kirou, C Lee, S George, K Louca, M Peterson, MK Crow, unpublished observations). However, it should be noted that the IFN signature is complex, as described in detail for the Baechler study, and additional work will be required to determine the relative roles of type I and type II IFN in the gene expression profile observed in SLE peripheral blood. Analysis of the age of study patients, time from diagnosis, organ system involvement, and current therapy, along with gene expression data, will be essential for an understanding of the place of IFN family members in disease pathogenesis.

Beyond the impressive IFN gene signature, each of the studies identified additional genes that were differentially expressed in SLE and control samples. In view of the biological, patient, and assay variation encountered in gene profiling of SLE, a diagnostic gene signature might prove most useful clinically. An algorithm such as shrunken centroids can be used to develop a robust multigene classifier that can be tested in clinical studies as a measure of diagnosis or clinical outcome [11].

Microarray analysis of gene expression at sites of tissue damage

The next generation of reports describing global gene expression patterns based on microarray assays will prob-

ably derive from studies of tissue samples from sites of disease activity. Unpublished data presented at scientific meetings demonstrate the feasibility of microarray analysis of glomeruli from lupus nephritis kidneys and from skin lesions of lupus patients. As valuable as those data will be, we are encouraged that peripheral blood seems to provide a reasonable sampling of those gene pathways that are activated and relatively disease specific.

Conclusion

Since the dramatic illustration of the clinical utility of microarray technology in patients with malignant disease 3 years ago, efforts to study mixed mononuclear cell populations in patients with autoimmune diseases have been remarkably successful. Not only are microarray-derived data interpretable and significant, but they have drawn our attention back to a key cytokine pathway. Increased expression of IFN in patients with active SLE was first reported in 1979, but relatively few investigators have pursued the role of IFN in SLE [22]. An important exception is the group led by Ronnblom and Alm, who noted the induction of lupus autoantibodies and clinical lupus in patients receiving therapeutic IFN- α [26]. That group has gone on to characterize the IFN-producing cells as well as some of the properties of immune complexes that induce the production of IFN by plasmacytoid dendritic cells [32–37]. In addition, Bennett's collaborators, led by Banchereau, have presented functional data showing the induction of efficient stimulators of allogeneic T cell responses by IFN- α in SLE sera [28]. Now, the repeated appearance of IFN-induced genes among the most significantly overexpressed genes in data derived from multiple laboratories using distinct microarrays raises the profile of IFN as a pathogenic mediator of the myriad alterations to the immune system seen in SLE.

In view of the important effects of IFN on immune system function, including activities that could contribute to the development of systemic autoimmunity, this cytokine system might represent an excellent target for therapeutic modulation [38–44]. At the same time, both type I and type II IFNs are important, if not essential, for effective host defense against pathogenic microbes. The weight of data are consistent with the action of type I IFNs as primary mediators of the observed gene expression signature in SLE, with significant consequences for the development of clinical disease. However, additional potential mediators of the IFN gene signature, including CpG DNA or double-stranded RNA, are candidates that should be explored.

Competing interests

MKC is a scientific advisor to Expression Diagnostics, Inc, and JW is Vice-President, Research and Development, Expression Diagnostics, Inc. Research conducted by MKC was supported through a contract between Expression Diagnostics, Inc. and Hospital for Special Surgery.

Acknowledgements

We thank our collaborators at the Hospital for Special Surgery and Expression Diagnostics, Inc, particularly Dr Kyriakos Kirou and Ms Christina Lee and Ms Sandhya George, for their contributions to the work described in this review. MKC is supported by a Target Identification in Lupus Grant from the Alliance for Lupus Research, by a Novel Research Grant from the Lupus Research Institute, and by the Mary Kirkland Center for Lupus Research.

References

1. Brown PO, Botstein D: **Exploring the new world of the genome with DNA microarrays.** *Nat Genet* 1999, **21**:33-37.
2. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JJ, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Staudt LM: **Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.** *Nature* 2000, **403**:503-511.
3. Crow MK, George S, Paget SA, Ly N, Woodward R, Fry K, Wohlgenuth J: **Expression of an interferon-alpha gene program in SLE.** *Arthritis Rheum* 2002, **46**:S281.
4. 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 USA* 2003, **100**:2610-2615.
5. 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.
6. Han GM, Chen SL, Shen N, Ye S, Bao CD, Gu YY: **Analysis of gene expression profiles in human systemic lupus erythematosus using oligonucleotide microarray.** *Genes Immun* 2003, **4**:177-186.
7. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Statist Soc* 1995, **57**:289-300.
8. Dudoit S, Gentleman RC, Quackenbush J: **Open source software for the analysis of microarray data.** *Biotechniques* 2003, **Mar Suppl**:45-51.
9. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98**:5116-5121.
10. Hastie T, Tibshirani R, Botstein D, Brown P: **Supervised harvesting of expression trees.** *Genome Biol* 2001, **2**:research0003.1-0003.12.
11. Tibshirani R, Hastie T, Narasimhan B, Chu G: **Diagnosis of multiple cancer types by shrunken centroids of gene expression.** *Proc Natl Acad Sci USA* 2002, **99**:6567-6572.
12. Friedman JH: *Stochastic Gradient Boosting. Technical report.* Stanford, CA: Department of Statistics, Stanford University; 1999.
13. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: **Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.** *Science* 1999, **286**:531-537.
14. Breiman L, Friedman JH, Olshen RA, Stone CJ: *Classification and Regression Trees.* Belmont, CA: Wadsworth; 1984.
15. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.
16. Sherlock G: **Analysis of large-scale gene expression data.** *Curr Opin Immunol* 2000, **12**:201-205.
17. Kirou K, Lee C, Crow MK: **Measurement of cytokines in autoimmune disease.** *Methods Mol Med*, in press.
18. Rus V, Atamas SP, Shustova V, Luzina IG, Selaru F, Magder LS, Via CS: **Expression of cytokine- and chemokine-related genes in peripheral blood mononuclear cells from lupus patients by cDNA array.** *Clin Immunol* 2002, **102**:283-290.
19. 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.
20. Laxminarayana D, Khan IU, Kammer G: **Transcript mutations of the alpha regulatory subunit of protein kinase A and up-regulation of the RNA-editing gene transcript in lupus T lymphocytes.** *Lancet* 2002, **360**:842-849.

21. Tezak Z, Hoffman EP, Lutz JL, Fedczyna TO, Stephan D, Bremer EG, Krasnoselska-Riz I, Kumar A, Pachman LM: **Gene expression profiling in DQA1*0501+ children with untreated dermatomyositis: a novel model of pathogenesis.** *J Immunol* 2002, **168**:4154-4163.
22. Hooks JJ, Moutsopoulos HM, Geis SA, Stahl NI, Decker JL, Notkins AL: **Immune interferon in the circulation of patients with autoimmune disease.** *N Engl J Med* 1979, **301**:5-8.
23. Hooks JJ, Jordan GW, Cupps T, Moutsopoulos HM, Fauci AS, Notkins AL: **Multiple interferons in the circulation of patients with systemic lupus erythematosus and vasculitis.** *Arthritis Rheum* 1982, **25**:396-400.
24. Preble OT, Black RJ, Friedman RM, Klippel JH, Vilcek J: **Systemic lupus erythematosus: presence in human serum of an unusual acid-labile leukocyte interferon.** *Science* 1982, **216**:429-431.
25. Rich SA: **Human lupus inclusions and interferon.** *Science* 1981, **213**:772-775.
26. Ronnblom LE, Alm GV, Oberg KE: **Possible induction of systemic lupus erythematosus by interferon-alpha treatment in a patient with a malignant carcinoid tumour.** *J Intern Med* 1990, **227**:207-210.
27. Wandl UB, Nagel-Hiemke M, May D, Kreuzfelder E, Kloke O, Kranzhoff M, Seeber S, Niederle N: **Lupus-like autoimmune disease induced by interferon therapy for myeloproliferative disorders.** *Clin Immunol Immunopathol* 1992, **65**:70-74.
28. Blanco P, Palucha AK, Gill M, Pascual V, Banchereau J: **Induction of dendritic cell differentiation by IFN-alpha in systemic lupus erythematosus.** *Science* 2001, **294**:1540-1543.
29. Peng SL, Moslehi J, Craft J: **Roles of interferon-gamma and interleukin-4 in murine lupus.** *J Clin Invest.* 1997, **99**:1936-1946.
30. Lawson BR, Prud'homme GJ, Chang Y, Gardner HA, Kuan J, Kono DH, Theofilopoulos AN: **Treatment of murine lupus with cDNA encoding IFN- γ R/Fc.** *J Clin Invest* 2000, **106**:207-215.
31. Santiago-Raber ML, Baccala R, Haraldsson KM, Choubey D, Stewart TA, Kono DH, Theofilopoulos AN: **Type-I interferon receptor deficiency reduces lupus-like disease in NZB mice.** *J Exp Med* 2003, **197**:777-788.
32. Svensson H, Johannisson A, Nikkila T, Alm GV, Cederblad B: **The cell surface phenotype of human natural interferon- α producing cells as determined by flow cytometry.** *Scand J Immunol* 1996, **44**:164-172.
33. Vallin H, Blomberg S, Alm GV, Cederblad B, Ronnblom L: **Patients with systemic lupus erythematosus (SLE) have a circulating inducer of interferon-alpha (IFN- α) production acting on leukocytes resembling immature dendritic cells.** *Clin Exp Immunol* 1999, **115**:196-202.
34. Vallin H, Peters A, Alm GV, Ronnblom L: **Anti-double-stranded DNA antibodies and immunostimulatory plasmid DNA in combination mimic the endogenous IFN-alpha inducer in systemic lupus erythematosus.** *J Immunol* 1999, **163**:6306-6313.
35. Bave U, Alm GV, Ronnblom L: **The combination of apoptotic U937 cells and lupus IgG is a potent IFN-alpha inducer.** *J Immunol* 2000, **165**:3519-3526.
36. Bave U, Vallin H, Alm GV, Ronnblom L: **Activation of natural interferon-alpha producing cells by apoptotic U937 cells combined with lupus IgG and its regulation by cytokines.** *J Autoimmun* 2001, **17**:71-80.
37. Magnusson M, Magnusson S, Vallin H, Ronnblom L, Alm GV: **Importance of CpG dinucleotides in activation of natural IFN-alpha-producing cells by a lupus-related oligodeoxynucleotide.** *Scand J Immunol* 2001, **54**:543-550.
38. Ronnblom L, Alm GV: **A pivotal role for the natural interferon α -producing cells (plasmacytoid dendritic cells) in the pathogenesis of lupus.** *J Exp Med* 2001, **194**:59-63.
39. Aman MJ, Tretter T, Eisenbeis I, Bug G, Decker T, Aulitzky WE, Tilg H, Huber C, Peschel C: **Interferon-alpha stimulates production of interleukin-10 in activated CD4+ T cells and monocytes.** *Blood* 1996, **87**:4731-4736.
40. Brinkmann V, Geiger T, Alkan S, Heusser CH: **Interferon alpha increases the frequency of interferon gamma-producing human CD4+ T cells.** *J Exp Med* 1993, **178**:1655-1663.
41. Kirou KA, Vakkalanka RK, Butler MJ, Crow MK: **Induction of Fas ligand-mediated apoptosis by interferon-alpha.** *Clin Immunol* 2000, **95**:218-226.
42. Radvanyi LG, Banerjee A, Weir M, Messner H: **Low levels of interferon-alpha induce CD86 (B7.2) expression and accelerate dendritic cell maturation from human peripheral blood mononuclear cells.** *Scand J Immunol* 1999, **50**:499-509.
43. Luft T, Pang KC, Thomas E, Herzog P, Hart DN, Trapani J, Cebon J: **Type I IFNs enhance the terminal differentiation of dendritic cells.** *J Immunol* 1998, **161**:1947-1953.
44. Crow MK: **Interferon-alpha: a new target for therapy in SLE?** *Arthritis Rheum* 2003, **48**:2396-2401.

Correspondence

Mary K Crow MD, Hospital for Special Surgery, 535 East 70th Street, New York, NY 10021, USA. Tel: +1 212 606 1397; fax: +1 212 774 2337; e-mail: crowm@hss.edu