Published in final edited form as:

Lupus. 2011 March ; 20(3): . doi:10.1177/0961203310383072.

A systemic lupus erythematosus gene expression array in disease diagnosis and classification: a preliminary report

Y-T Juang^{1,*}, C Peoples^{1,*}, R Kafri^{2,*}, VC Kyttaris¹, K Sunahori¹, K Kis-Toth¹, L Fitzgerald¹, S Ergin¹, M Finnell¹, and GC Tsokos¹

¹Division of Rheumatology, Beth Israel Deaconess Medical Center, Harvard Medical School, USA

²Department of Systems Biology, Harvard Medical School, USA

Abstract

Systemic lupus erythematosus (SLE) is a clinically heterogeneous disease diagnosed on the presence of a constellation of clinical and laboratory findings. At the pathogenetic level, multiple factors using diverse biochemical and molecular pathways have been recognized. Succinct recognition and classification of clinical disease subsets, as well as the availability of disease biomarkers, remains largely unsolved. Based on information produced by the present authors' and other laboratories, a lupus gene expression array consisting of 30 genes, previously claimed to contribute to aberrant function of T cells, was developed. An additional eight genes were included as controls. Peripheral blood was obtained from 10 patients (19 samples) with SLE and six patients with rheumatoid arthritis (RA) as well as 19 healthy controls. T cell mRNA was subjected to reverse transcription and PCR, and the gene expression levels were measured. Conventional statistical analysis was performed along with principal component analysis (PCA) to capture the contribution of all genes to disease diagnosis and clinical parameters. The lupus gene expression array faithfully informed on the expression levels of genes. The recorded changes in expression reflect those reported in the literature by using a relatively small (5ml) amount of peripheral blood. PCA of gene expression levels placed SLE samples apart from normal and RA samples regardless of disease activity. Individual principal components tended to define specific disease manifestations such as arthritis and proteinuria. Thus, a lupus gene expression array based on genes previously claimed to contribute to immune pathogenesis of SLE may define the disease, and principal components of the expression of 30 genes may define patients with specific disease manifestations.

Keywords

gene array; principal component analysis; systemic lupus erythematosus

Introduction

Systemic lupus erythematosus (SLE) presents with fascinating clinical heterogeneity underlined by equally diverse pathogenic factors and immune system abnormalities. Immune cell abnormalities converge to the production of autoantibodies, mostly against nuclear antigens, immune complexes and T cells, which contribute to disease pathology.¹⁻⁶ Disease management still relies on the use of indiscriminate immunosuppression and

©The Author(s), 2010

Correspondence to: George C. Tsokos, Division of Rheumatology, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA, gtsokos@bidmc.harvard.edu.

*These authors contributed equally to this work

treatment of arising complications.³ Progress has been undermined by the absence of tools to classify the disease and measure its activity and proper disease-specific treatment targets.

Several laboratories, including ours, have reported distinctly aberrant expression of genes of the adaptive and innate immune system.^{2,4–7} Aberrant expression of several genes has been implicated in vitro to contribute to the abnormal function of immune cells. For example, correction of the decreased levels of CD3 ζ in SLE T cells results in increased production of interleukin 2 (IL-2),⁸ inhibition of the increased spleen tyrosine kinase (Syk) levels in SLE T cells results in normal CD3-mediated cell signaling,⁹ and inhibition or silencing of increased protein phosphatase 2A (PP2A) results in corrected IL-2 production.¹⁰

Wishing to capture simultaneously the aberrant expression of all reported genes at a given time point of disease progression using a sensible amount of peripheral blood, we constructed a gene expression array in which we included 30 genes. In this preliminary communication, we report that we can capture gene expression variations similar to those reported using classical bio-chemical approaches. More importantly, we present first evidence that principal component analysis (PCA) of the expression levels of the included 30 genes placed SLE patients apart from normal subjects and patients with rheumatoid arthritis (RA).¹¹ Furthermore, distinct clinical manifestations were defined by individual principal components. Our gene expression array should facilitate the diagnosis of SLE with improved sensitivity and specificity, and, when larger cohorts of patients have been studied, it may enable a molecular classification of patients that better dictates treatment.

Materials and methods

Patients

Patients ($n=10$) fulfilling the four ACR-established criteria for the diagnosis of SLE were included, while six patients with an established diagnosis of RA served as disease controls (Table 1). In brief, the age range was 23–56 years old, 90% were women, and 30% were of Caucasian, 20% African, 20% Hispanic and 30% other origin. The age of the RA individuals ranged from 28 to 67 years of age. Nineteen samples from healthy subjects matched for age, sex and ethnicity served as controls. Six patients were studied on two or three occasions during the course of the study. The study was approved by the Institutional Review Board of Beth Israel Deaconess Medical Center.

Basic design of the SLE gene array

The array was manufactured on a 96-well plate. Each well was embedded with a pair of primers to PCR-amplify either eight housekeeping/control genes (including CD3 ζ , GAPDH, RTC, and HGDC) or a specific gene ($n=30$) chosen because of claimed importance in the expression of aberrant T cell function in SLE.^{1,2} Primers for an additional nine genes claimed to be aberrantly expressed in SLE were embedded but not included in the current analysis. A list of the included genes is shown in Table 2. SLE or RA samples were run in parallel to a normal sample on the 96-well plate.

Determination of gene expression levels

T cell-derived mRNA^{9,10} was reversely transcribed to cDNA using the RT2 First Strand Kit (SABiosciences, Frederick, MD) and placed in the wells of the 96-well plate. Quantitative real time PCR was subsequently performed using the RT² Real-Time SYBR Green PCR Master Mix (SABiosciences) and the product was evaluated utilizing a Roche LightCycler 480 PCR system (Roche, Indianapolis, IN), which allows gene expression detection within a 10-log interval. Gene expression levels were normalized against the housekeeping gene CD3 ϵ .

Statistical analysis

Student's t-test was applied to compare the expression of single genes between patients and normal individuals. Principal component analysis (PCA) was applied to identify directions (principal components) along which the variation of the data is maximal^{11,12} using the Matlab (7.0R14, MathWorks) software. In the initial dataset, two individuals displayed exceedingly higher expression values for all genes. To avoid bias, principal components were calculated after excluding these individuals. Representing these individuals on the principal component axes that were calculated in their absence preserved all recorded trends.

Results

Expression levels of genes detected by the gene array

The gene expression array was first designed as a tool to enable the simultaneous determination of the levels of expression of genes reported by us and others^{2,4,6,7} to be abnormally expressed and to contribute to the immunopathogenesis of the disease. In Figure 1 we show the expression levels of two representative genes, CD3 ζ and CREM, as determined by the SLE gene expression array. As expected, CD3 ζ ¹⁰ mRNA levels are decreased and CREM¹³ mRNA levels are increased in T cells from patients with SLE compared with T cells from sex and age-matched normal individuals. Accordingly, the SLE gene expression array can be used to detect simultaneously the levels of expression of 30 genes using a small amount (5 ml) of peripheral blood. The expression levels of all genes in T cells from patients with RA were comparable to those in normal T cells as predicted from previous studies.^{10,13}

PCA of expression levels of genes included in the SLE gene expression array

We considered that meaningful phenotypes of the disease would more likely be represented as a function of all genes rather than the separate expression values. This consideration was further supported by the fact that the included genes had been previously identified as contributors to the SLE immunopathology. Thus, it seemed natural, as a first attempt, to apply PCA, a mathematical algorithm that organizes data, e.g. gene expression values, into functions (principal components) that better represent the variation between individuals.¹¹ Each calculated principal component is a function, specifically a linear combination of all expression values. For example, if an individual was initially characterized by an expression level e_1 for gene 1 and e_2 for gene 2, a calculated PC would have the form $pc_1 = c_1e_1 + c_2e_2$, where c_1 and c_2 are values calculated such that pc_1 describes more of the variation in the sample than either e_1 or e_2 does independently.

Expression levels for all 30 genes in all studied individuals are shown in Figure 2A. After applying PCA, principal components were identified and ordered according to their contribution to the overall variance (Figure 2B). Figure 2C demonstrates that 42% of the sample variation can be attributed to principal component 1, and as much as 71% of the overall variation can be accounted for by the first five principal components and 88% by the first 10 principal components. Figure 3 shows a scatter plot representation of individual samples with the first three principal components axes. This plot revealed a striking result whereby the control individuals are spatially separated from the SLE patients. In fact, the variation of control individuals was more constrained and was enclosed by a smaller volume, i.e. a smaller enclosing convex hull. In contrast, SLE patients were far more scattered in these representation axes. Illustrating the clinical and pathogenic complexity of the disease, SLE patient samples were not confined to any specific location and could be roughly classified as having high values in at least one of the principal component axes. Samples from patients with rheumatoid arthritis seemed to localize separately.

We next asked whether separate individual principal components may represent distinct disease manifestations. We should point out that the calculation of each principal component took place without inputting prior knowledge about the specific diagnosis (controls vs. patients) or clinical manifestation. It was therefore interesting to ask whether any principal component would define a clinically identified disease feature. It has been frequently demonstrated that principal components may better correlate with clinical features than separate gene expression values.^{14–16} Interestingly, despite our rather small sample size, different principal components appeared to uniquely report different clinical features (Figure 4). Specifically, Figure 4A shows that principal components 2 and 9 correlate significantly with SLEDAI scores. In addition, and more interestingly, arthritis is best defined by principal component 7 and proteinuria by principal component 3.

Discussion

We present here first evidence on a gene expression array consisting of 30 genes that various laboratories, including ours, had previously claimed to contribute to the immunopathogenesis of SLE, as follows: 1) the array faithfully reports on the gene expression abnormality in a fashion similar to that reported previously using traditional biochemical approaches; 2) the array separates in space (using the three first principal components derived from PCA) the location of SLE samples from those defined by samples from patients with RA and normal individuals; and 3) distinct principal components defined groups of patients with specific clinical manifestations.

While we and others have been studying immune cell biochemistry and molecular biology in patients with SLE in order to identify novel molecular treatment targets and biomarkers, we were challenged physically to record simultaneously the expression of all identified genes at a given time point of the disease. To overcome this difficulty we constructed a gene array, which, even in its first phase, can detect the expression of all genes. For brevity we report here that the mRNA levels of two genes, CD3 ζ and CREM (Figure 1), were found to be expressed as previously reported.^{10,13}

Because the genes that we chose had been studied in a logical manner in efforts to understand, among other things, why CD3-mediated T-cell signalling is increased^{17,18} and why IL-2 production is decreased in SLE T cells,^{19,20} we considered that the application of PCA would reduce the noise of the heat-map (Figure 2A) recorded expression levels and identify linear patterns, principal components, which would reduce the number of dimensions of the data to a number that is manageable. Reassuringly, we found that the first principal component contributed 42% to all variation and the first five principal components 81%. The most surprising finding was that, when the first three principal components were plotted in a three-dimensional scattergram, the position of the samples from normal individuals defined a restricted convex hull and only two of the 19 SLE samples were located within that space. The samples from RA patients defined a separate space. The 17 lupus samples were positioned outside the space defined by the normal samples regardless of the assigned SLEDAI score, suggesting that the 30-gene expression array may very well identify SLE patients who do not have any other clinical manifestations. It remains to be established, among other things, whether the expression array changes position in space as clinical manifestations are added and the ACR-established requirements for the diagnosis of SLE are met.

It is well accepted that an unmet need in the field of SLE is the requirement to classify patients in a more accurate manner, better reflecting underlying biochemical abnormalities, which may enable properly targeted treatment. When we asked whether any of the calculated principal components define distinct clinical manifestations, we observed that,

although the SLEDAI score was better represented by principal components 2 and 9, arthritis was defined by principal component 7 and proteinuria by principal component 3. We acknowledge the small number of entries, and verification of our findings with larger numbers of patients is in order, yet the principal component-defined presence of distinct clinical manifestations is significant (Figure 4).

Our approach to the identification of gene expression signature is conceptually different from that reported by the Pascual group²¹ and that reported by McKinney *et al.*²² using CD8⁺ T cell gene expression arrays in that we included only genes claimed in previous in vitro studies to be part of the aberrant SLE T cell function. Yet the approaches at this point are complementary, and the ideal tool to properly diagnose and classify patients with SLE is still to be conceived and developed.

Expansion of the SLE samples to larger numbers is needed to identify possible effects of treatment and to determine whether principal components can accurately define patients with distinct clinical or laboratory abnormalities. Inclusion of larger numbers representing various ethnic groups is necessary, and prospective studies are also needed to determine whether clinical variation in any given patient affects the patient's position in the three-dimensional space defined by the first three or any other combination of principal components.

In conclusion, in this preliminary communication we present evidence that a gene expression array consisting of 30 genes, selected because of their reported importance in the pathogenesis of the disease, can identify SLE patients and define those with distinct clinical manifestations.

Acknowledgments

Funding

This study was supported by NIH grants R01AI42269, R01AI68787, R01AI49954, R01AI85567 and K23AR55672 and the Kirkland Scholar Program.

References

1. Crispin JC, Liossis SN, Kis-Toth K, et al. Pathogenesis of human systemic lupus erythematosus: recent advances. *Trends Mol Med.* 2010; 16:47–57. [PubMed: 20138006]
2. Kammer GM, Perl A, Richardson BC, Tsokos GC. Abnormal T cell signal transduction in systemic lupus erythematosus. *Arthritis Rheum.* 2002; 46:1139–1154. [PubMed: 12115215]
3. Davidson A, Diamond B. Autoimmune diseases. *N Engl J Med.* 2001; 345:340–350. [PubMed: 11484692]
4. Perl A, Fernandez DR, Telarico T, Doherty E, Francis L, Phillips PE. T-cell and B-cell signaling biomarkers and treatment targets in lupus. *Curr Opin Rheumatol.* 2009; 21:454–464. [PubMed: 19550330]
5. Perl A. Systems biology of lupus: mapping the impact of genomic and environmental factors on gene expression signatures, cellular signaling, metabolic pathways, hormonal and cytokine imbalance, and selecting targets for treatment. *Autoimmunity.* 2010; 43:32–47. [PubMed: 20001421]
6. Pernis AB. Th17 cells in rheumatoid arthritis and systemic lupus erythematosus. *J Intern Med.* 2009; 265:644–652. [PubMed: 19493058]
7. Ballestar E, Esteller M, Richardson BC. The epigenetic face of systemic lupus erythematosus. *J Immunol.* 2006; 176:7143–7147. [PubMed: 16751355]
8. Nambiar MP, Fisher CU, Warke VG, et al. Reconstitution of deficient T cell receptor zeta chain restores T cell signaling and augments T cell receptor/CD3-induced interleukin-2 production in

- patients with systemic lupus erythematosus. *Arthritis Rheum.* 2003; 48:1948–1955. [PubMed: 12847689]
9. Krishnan S, Juang YT, Chowdhury B, et al. Differential expression and molecular associations of Syk in systemic lupus erythematosus T cells. *J Immunol.* 2008; 181:8145–8152. [PubMed: 19018007]
10. Katsiari CG, Kyttaris VC, Juang YT, Tsokos GC. Protein phosphatase 2A is a negative regulator of IL-2 production in patients with systemic lupus erythematosus. *J Clin Invest.* 2005; 115:3193–3204. [PubMed: 16224536]
11. Ringner M. What is principal component analysis? *Nat Biotechnol.* 2008; 26:303–304. [PubMed: 18327243]
12. Rencher, AC. *Methods of multivariate analysis.* Edited by Wiley-Interscience; 2002. p. 380-407.
13. Solomou EE, Juang YT, Gourley MF, Kammer GM, Tsokos GC. Molecular basis of deficient IL-2 production in T cells from patients with systemic lupus erythematosus. *J Immunol.* 2001; 166:4216–4222. [PubMed: 11238674]
14. Khan J, Wei JS, Ringner M, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med.* 2001; 7:673–679. [PubMed: 11385503]
15. Bullinger L, Dohner K, Bair E, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med.* 2004; 350:1605–1616. [PubMed: 15084693]
16. Saal LH, Johansson P, Holm K, et al. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proc Natl Acad Sci U S A.* 2007; 104:7564–7569. [PubMed: 17452630]
17. Liossis SN, Ding XZ, Dennis GJ, Tsokos GC. Altered pattern of TCR/CD3-mediated protein-tyrosyl phosphorylation in T cells from patients with systemic lupus erythematosus. Deficient expression of the T cell receptor zeta chain. *J Clin Invest.* 1998; 101:1448–1457. [PubMed: 9525988]
18. Tsokos GC, Nambiar MP, Tenbrock K, Juang YT. Rewiring the T-cell: signaling defects and novel prospects for the treatment of SLE. *Trends Immunol.* 2003; 24:259–263. [PubMed: 12738420]
19. Lieberman LA, Tsokos GC. The IL-2 defect in systemic lupus erythematosus disease has an expansive effect on host immunity. *J Biomed Biotechnol.* 2010; 2010:740619. [PubMed: 20625413]
20. Tenbrock K, Tsokos GC. Transcriptional regulation of interleukin 2 in SLE T cells. *Int Rev Immunol.* 2004; 23:333–345. [PubMed: 15204092]
21. Allantaz F, Chaussabel D, Banchereau J, Pascual V. Microarray-based identification of novel biomarkers in IL-1-mediated diseases. *Curr Opin Immunol.* 2007; 19:623–632. [PubMed: 18036805]
22. McKinney EF, Lyons PA, Carr EJ, et al. A CD8+ T cell transcription signature predicts prognosis in autoimmune disease. *Nat Med.* 2010; 16:586–591. [PubMed: 20400961]

Figure 1. Systemic lupus erythematosus gene expression array determines faithfully the levels of studied genes. (A) CD3 ζ mRNA levels in T cells. (B) CREM mRNA levels in T cells.

Figure 2. Gene expression in systemic lupus erythematosus T cells. (A) Gene expression values in patients with SLE. (B) First 10 principal components for all patients. (C) The percentage of variation accounted for by each of the principal components.

Figure 3. Variation between individuals represented on the axes of the first three principal components. The grey shaded conclave (convex hull) is defined by the position of the entries for the normal individuals. The green shaded conclave is defined by the position of the entries of samples from patients with rheumatoid arthritis.

Figure 4. Correlation between individual principal components and clinical manifestations. (A) SLEDAI, (B) arthritis and (C) proteinuria. Perpendicular lines represent standard errors.

Table 1

Demographic, clinical and laboratory features of research subjects

Patient	Age	Sex	Race/Ethnicity	SLEDAI	C3	C4	Anti-dsDNA	Neuro-psychiatric	Nephritis	MSK	Skin	Serositis	Hematologic	Other
#1	28	F	C	12			N		Y				Y	
#2	56	F	C	0	90	28				Y	Y			Y
#3	36	F	C/A	0	106	42					Y			Y
#4	30	F	A/H	4	133	28	Y		Y	Y				Y
#5*	37	F	A/A	10	161	38	N	Y		Y		Y		Y
				0				Y		Y				
#6*	24	F	I	0	111	18	N		Y	Y	Y		Y	
				0	99	15	N		Y	Y	Y		Y	
				4	91	13	N		Y	Y	Y		Y	
#7*	23	F	A	35	0	6	N	Y	Y	Y	Y	Y		
				0	118	35	N	Y	Y	Y	Y	Y		
#8*	54	F	C	14	161	37	N	Y		Y	Y		Y	
				0				Y	Y	Y	Y	Y	Y	
#9*	26	M	A	2	75	4	N	Y			Y		Y	
				4	66	4	Y	Y			Y		Y	
				4	80	4	Y	Y			Y		Y	
#10*	39	F	C/H	0	104	20	N							
				10	86	11	Y		Y	Y				Y
				2	102	18	Y		Y	Y				Y

A=African American, C=Caucasian, F=female, H=Hispanic, I=Indian, M=male, MSK=Musculoskeletal, N=no, Y=yes.

* patients studied on a second or third occasion.

Table 2

Layout of the SLE gene expression array

	1	2	3	4	5	6
A	IFNA1	CD247	CREM	HDAC1	NFATC2	PTGS2
B	IFNA5	CD3E	CTLA4	ICAM1	PDCD1	ROCK1
C	IL10	CD40LG	FASLG	IFNG	PPP2CA	SYK
D	IL23A	CD44	FCER1G	IL17A	PPP2CB	EZR
E	FASLG	CD44V3	FOS	IL17F	PRKAR1B	GAPDH
F	GAPDH	CD44V6	FOXP3	IL2	PRKAR2B	HGDC
G	HGDC	CD70	GATA3	IL21	PRKCD	RTC
H	CALM3	CREB1	RELA	IL6	PRKCQ	NTC

IFNA1=Interferon alpha 1, CD247=CD247 molecule, CREM=cAMP responsive element modulator, HDAC=histone deacetylase 1, NFATC2= nuclear factor of activated T cells, cytoplasmic, calcineurin-dependent 2, PTGS2 =prostaglandin-endoperoxide synthase 2 (prosta-glandin G/H synthase and cyclooxygenase), IFNA5=interferon alpha 5, CD3E=CD3e molecule, epsilon (CD3-TCR complex), CTLA4 =cytotoxic T-lymphocyte-associated protein 4, ICAM1 = intercellular adhesion molecule 1 (CD54), human rhinovirus receptor, PDCD1=programmed cell death 1, ROCK1=rho-associated, coiled-coil containing protein kinase 1, IL10=interleukin 10, CD40LG=CD40 ligand (TNF superfamily member 5, hyper-IgM syndrome), FASLG=Fas ligand (TNF superfamily member 6), IFNG=interferon gamma, PPP2CA = protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform, SYK = spleen tyrosine kinase, IL23A = interleukin 23, alpha subunit p19, CD44= CD44 molecule (Indian blood group), FCER1G=Fc fragment of IgE, high affinity 1, receptor for gamma polypeptide, IL17A= interleukin 17A, PPP2CB =protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform, EZR = ezrin, CD44V3 =v3 variant of CD44, FOS=V-fos FBI murine osteosarcoma viral oncogene homolog, IL17F=interleukin 17F, PRKAR1B=protein kinase, cAMP-dependent, regulatory, type I, beta, GAPDH=glyceraldehyde-3-phosphate dehydrogenase, CD44V6=v6 variant of CD44, FOXP3=forkhead box P3, IL2=interleukin 2, PRKAR2B=protein kinase, cAMP-dependent, regulatory, type II, beta, HGDC=human genomic DNA contamination, CD70=CD70 molecule, GATA3=GATA binding protein 3, IL21=interleukin 21, PRKCD=protein kinase C, delta, RTC=reverse transcription control, CALM3 = calmodulin 3 (phosphorylase kinase, delta), CREB1=cAMP response element binding protein 1, RELA = V-rel reticuloendotheliosis viral oncogene homolog A, nuclear factor of kappa light polypeptide gene enhancer in B-cells, p65 (avian), IL6= interleukin 6, PRKCQ=protein kinase C, theta, NTC=no template control.