Research article

Open Access

Novel approaches to gene expression analysis of active polyarticular juvenile rheumatoid arthritis

James N Jarvis*[1], Igor Dozmorov*[2], Kaiyu Jiang[1], Mark Barton Frank[2], Peter Szodoray[3], Philip Alex[2] and Michael Centola[2]

[1]Department of Pediatrics, University of Oklahoma College of Medicine, Oklahoma City, OK, USA
[2]Department of Arthritis and Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA
[3]Broegelmann Research Laboratory, The Gade Institute, University of Bergen, Bergen, Norway
*Drs Jarvis and Dozmorov contributed equally to this work.

Correspondence: James N Jarvis (james-jarvis@ouhsc.edu)

Abstract

Juvenile rheumatoid arthritis (JRA) has a complex, poorly characterized pathophysiology. Modeling of transcriptosome behavior in pathologic specimens using microarrays allows molecular dissection of complex autoimmune diseases. However, conventional analyses rely on identifying statistically significant differences in gene expression distributions between patients and controls. Since the principal aspects of disease pathophysiology vary significantly among patients, these analyses are biased. Genes with highly variable expression, those most likely to regulate and affect pathologic processes, are excluded from selection, as their distribution among healthy and affected individuals may overlap significantly. Here we describe a novel method for analyzing microarray data that assesses statistically significant changes in gene behavior at the population level. This method was applied to expression profiles of peripheral blood leukocytes from a group of children with polyarticular JRA and healthy control subjects. Results from this method are compared with those from a conventional analysis of differential gene expression and shown to identify discrete subsets of functionally related genes relevant to disease pathophysiology. These results reveal the complex action of the innate and adaptive immune responses in patients and specifically underscore the role of IFN-γ in disease pathophysiology. Discriminant function analysis of data from a cohort of patients treated with conventional therapy identified additional subsets of functionally related genes; the results may predict treatment outcomes. While data from only 9 patients and 12 healthy controls was used, this preliminary investigation of the inflammatory genomics of JRA illustrates the significant potential of utilizing complementary sets of bioinformatics tools to maximize the clinical relevance of microarray data from patients with autoimmune disease, even in small cohorts.

Keywords: arthritis, autoimmunity, bioinformatics, juvenile rheumatoid arthritis, microarray

Introduction

'Juvenile rheumatoid arthritis' (JRA), a term for the most prevalent form of arthritis in children, is applied to a family of illnesses characterized by chronic inflammation and hypertrophy of the synovial membranes. The term overlaps, but is not completely synonymous, with the family of illnesses referred to as juvenile chronic arthritis and/or juvenile idiopathic arthritis in Europe. We [1] and others [2] have proposed that the pathogenesis of rheumatoid disease in adults and children involves complex interactions between innate and adaptive immunity. This complexity lies at the core of the difficulty of unraveling disease pathogenesis. Both innate and adaptive immune systems use multiple cell types, a vast array of cell-surface and secreted proteins, and interconnected networks of positive and negative feedback [3]. Furthermore, while separable in thought, the innate and adaptive wings of the immune system are functionally intersected [4], and pathologic events occurring at these intersecting points are likely to be highly relevant to our understanding of pathogenesis of adult and childhood forms of chronic arthritis [5].

DFA = discriminant function analysis; ELISA = enzyme-linked immunosorbent assay; GM-CSF = granulocyte/macrophage-colony-stimulating factor; HV = hypervariable; ICAM-1 = intercellular adhesion molecule-1; IFN = interferon; JRA = juvenile rheumatoid arthritis; SD = standard deviation; TGF = transforming growth factor; TNF = tumor necrosis factor.

Polyarticular JRA is a distinct clinical subtype character-ized by inflammation and synovial proliferation in multiple joints (four or more), including the small joints of the hands [6]. This subtype of JRA may be severe, because of both its multiple joint involvement and its capacity to progress rapidly over time. Although clinically distinct, polyarticular JRA is not homogeneous, and patients vary in disease manifestations, age of onset, prognosis, and therapeutic response. These differences very likely reflect a spectrum of variation in the nature of the immune and inflammatory attack that can occur in this disease [1].

Gene expression profiling using microarrays provides a highly parallel assay for assessing molecular pathophysiol-ogy in a comprehensive manner. It holds the potential to refine our understanding of complex disease states. However, microarray data analysis is commonly limited to a simple assessment of a single behavioral change in gene expression, genes that are up- or down-regulated on average among distinct populations. This approach has been used to identify groups of genes that are prognosti-cally or diagnostically relevant, but the predictive power of these gene sets for autoimmune disease has proved limited [7–9]. Changes in gene behavior among individu-als in diseased populations are complex and may reflect both the unique genetic makeup of individuals and distinct subclasses of disease.

In this preliminary investigation of the inflammatory genomics of JRA, we report the application of a novel bioinformatics approach to microarray data for the identifi-cation of genes whose expression behavior is modulated by disease in a complex manner at the population level. Accordingly, genes whose expression within a population changes from stable to variable are identified. This measure of gene behavior emulates at the molecular level the loss of homeostasis characteristic of disease patho-genesis. The method identified a significant number of genes relevant to the pathophysiology of polyarticular JRA distinct from those identified by standard differential gene expression analysis. In addition, we followed a subset of patients during therapy to characterize temporally depen-dent changes in gene expression. Using discriminant func-tion analysis (DFA) to analyze this cohort, we identified gene expression changes characteristic of therapeutic response approximately one month before the time at which full clinical response occurred. A clinical assay could be created from this data that may predict soon after initiation of therapy which patients will respond and which will not. The predictive potential of this data is pred-icated on the fact that within 2 to 4 weeks after the start of therapy, gene expression in responsive patients, as mea-sured by DFA, became more like that in healthy controls, while gene expression in nonresponsive patients became less like that in healthy controls. Moreover, the genes iden-tified by DFA to be predictive of therapeutic response were, for the most part, known regulators and effectors of the immune system. Taken together, these data suggest that successful therapy was able to reset immune response homeostasis to a significant extent in this cohort.

Materials and methods
Patients, patient selection, preparation of clinical specimens
We studied nine children newly diagnosed with polyarticu-lar JRA. Diagnosis was based on accepted and validated criteria endorsed by the American College of Rheumatol-ogy [10]. Children were excluded if they had been treated with corticosteroids or methotrexate, or if they had received therapeutic doses of nonsteroidal anti-inflamma-tory drugs for more than 3 weeks before the study. Patients with active disease ranged in age from 4 to 15 years and presented with proliferative synovitis of multi-ple joints and erythrocyte sedimentation rates ranging from 35 to 100 mm/hour. Control subjects ($n = 12$) were laboratory volunteers under 25 years of age. Leukocyte buffy coat preparations were made from peripheral blood and total RNA extracted with Trizol reagent (Invitrogen, Carlsbad, CA, USA). Fluorescent labeling of cDNA was undertaken using the Micromax TSA-labeling kit (PerkinElmer Life Sciences, Boston, MA, USA). Labeled cDNAs were hybridized with PerkinElmer Micromax human cDNA microarray containing 2,382 human genes, and arrays were scanned using an Affymetrix 428 Array Scanner (Affymetrix, Durham, NC, USA).

Five of these nine patients were followed up longitudinally (for 6–12 months) from the onset of therapy as they either responded or failed to respond to therapy. In this portion of the study, disease severity was scored for the degree of synovitis using a linear scoring system used previously in our laboratory [11]. This system is based on criteria used in clinical trials in JRA [12]. For purposes of comparison and analysis, untreated children were categorized as having active disease. Children treated for more than 6 weeks who had a ≥30% reduction in their disease sever-ity score were categorized as having had a partial response to therapy, while children with <30% reduction in their severity scores were categorized as having acute, persistent disease. Children were categorized as being fully responsive to therapy if they showed synovial thicken-ing in ≤3 joints, without warmth or tenderness in those joints and with no more than 30 minutes' morning stiffness per day. These criteria for full responsiveness have been validated in previous studies we have published examining markers of inflammation in JRA [13,14]. The patients' char-acteristics are summarized in Table 1.

Serum cytokine levels.
Serum IFN-γ levels were measured using the BioPlex system, a biometric sandwich ELISA assay from BioRad Inc (Hercules, CA, USA) in accordance with the manufac-

Table 1

Data for patients with polyarticular juvenile rheumatoid arthritis

Patient	Age (years)	Sex	Treatment	Final outcome
1	15	F	NSAIDs, corticosteroids, MTX	Full response
2	11	F	NSAIDs, hydroxychloroquine, MTX	Studied once during active disease
3	4	M	NSAIDs	Full response
4	15	F	NSAIDs, MTX, corticosteroids	Studied once during active disease
5	7	F	N/A	Studied once during active disease
6	10	M	N/A	Studied once during active disease
7	7	F	NSAIDs, methotrexate, corticosteroids	Full response
8	15	F	NSAIDs	Full response
9	12	M	NSAIDs, MTX, corticosteroids	Persistent disease (values taken 4 times in an 8-week interval)

F, female; M, male; MTX, methotrexate; N/A, not applicable; NSAIDs, nonsteroidal anti-inflammatory drugs.

turer's instructions. Serum from four patients during periods of attack and before treatment (denoted 'patients with active disease') and from 12 healthy control subjects was collected, stored at –80°C, and assayed in duplicate.

Normalization of array data

Normalization to correct for technical variation among individual microarray hybridizations was conducted using a two-step procedure described in detail elsewhere [15]. In brief, the procedure is based on the fact that spot intensities from genes not expressed by the samples of interest constitute noise and are therefore normally distributed. The method models the signals from nonexpressed genes to a normal distribution with a mean of 0 and standard deviation (SD) of 1, using an iterative nonlinear curve-fitting procedure.

A second normalization step is then performed using the genes significantly expressed above background (>3 SD above background). Gene expression values are log-transformed, with negative values replaced by the lowest positive logarithmic value obtained. Expression profiles of genes statistically significantly expressed above background are then adjusted to each other using a robust regression analysis. This analysis is based on the observation that the expression levels of the majority of genes do not change in compared samples, and that expression values are normally distributed around a regression line with a small proportion of differentially expressed 'outliers'. The outliers' contribution in the regression analysis is down-weighted in an iterative manner until the residuals are normally distributed as measured by deviations from the regression line calculated against the averaged profile. Expression profiles of both control and experimental groups are then scaled to the averaged profile of the control group.

The two main sources of heterogeneity in gene expression variations are the 'additive component', prominent at low expression levels, and the 'multiplicative component', prominent at high expression levels [16]. The intensity measurement $y_{i,j}$ for gene $i \in I = \{i_1,...,i_n\}$ in sample $j \in J = \{j_1,...,j_m\}$ is modeled by the equation $y_{i,j} = a_{i,j} + m_{i,j} \times e^h + e_{i,j}$ where a is the normal background (not dependent on gene expression), m is the expression level in arbitrary units, e is first error term (additive) – which represents the standard deviation of background – and h is the second error term, which represents the proportional error (the multiplicative component) [17,18]. The first error term is excluded from analysis by eliminating expression values at or below background levels. The second error term is transformed from multiplicative (and therefore expression-dependent, rising with expression level [18]), into additive (expression-independent) by log-transformation of data [16] using the equation $log(y) = log(m) + h$, where h is the residual for log-transformed data. The independence of h from individual gene expressions is confirmed with the Kolmogorov–Smirnov normality test in our experiments. We determine h for each sample as a deviation of the gene expression ordinates from a regression line calculated against of the averaged profile for gene expressions in all samples of the control group. The majority of these deviations follow a normal distribution. Genes of the control groups whose deviations belong to this distribution are expressed at similar levels among groups; this group is therefore denoted the 'reference group'. Variations in expression among samples of the genes within this group are due principally to technical variability and normal biologic variation. The parameters of variation defined by the reference group are used to identify differentially expressed genes and hypervariable genes whose expression levels vary in a statistically significant manner from the reference group (Fig. 1). A standard

Figure 1
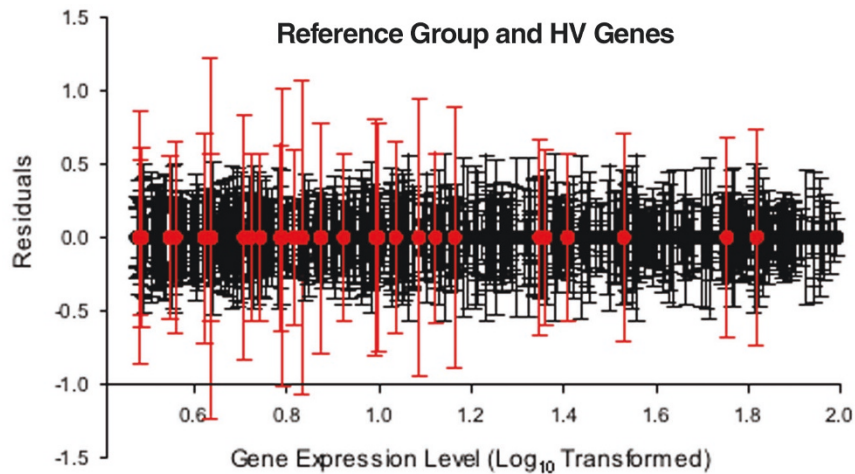
Graphical representation of hypervariable (HV) gene analysis in patients with juvenile rheumatoid arthritis (JRA) (*n*=9) and a reference group (*n*=12). A reference group of genes from the control group whose expression levels do not vary significantly on a population basis was identified as described in Materials and methods. Expression levels in this reference group, denoted the averaged profile, have a normal distribution. This group is represented by black lines on a plot of residuals (values representing expression level variance in the control population) vs average gene expression levels (log$_{10}$-transformed). Red lines represent genes whose variation in expression in healthy controls or untreated patients with acute disease was significantly greater than that of the reference group. These genes are defined as hypervariable (HV) genes.

F-test is used to determine if a given gene's expression is variable with respect to the reference group using Matlab software (Mathworks, Natick, MA, USA).

Identification of genes differentially expressed in patients vs control group

These analyses are performed using standard statistical analysis methods in Matlab software and include:

1. Selection of statistically different levels of expression using the Student's *t*-test with the commonly accepted significance threshold of *P*<0.05. Because of the large number of genes present on microarrays, a significant proportion of genes identified as differentially expressed in this manner will be false positive determinations at this threshold level.

2. An associative *t*-test, in which the replicated residuals for each gene in the experimental group are compared with the entire set of residuals from the reference group (defined above). The hypothesis that gene expression in the experimental group, presented as replicated residuals (deviations from averaged control-group profile), is distributed similarly to the several thousand members of the normally distributed set of residuals for gene expressions in the reference group is tested. The significance threshold is corrected to 1/(number of genes) to make it improbable that false positives arise. Only genes with *P* values below the threshold of both the Student's *t*-test and the associative *t*-test are then presented in tables as differentially expressed genes. Relative ratios of expression for

genes that are differentially expressed above background in both groups are calculated.

3. Genes expressed distinctively above background in one group and not in another are defined as uniquely expressed genes.

Selection of hypervariable (HV) genes

To have an opportunity to evaluate inhomogeneity in gene expression variability, it is necessary to normalize this variability to make it independent of the level of gene expression. The two main sources of heterogeneity in gene expression variations – additive and multiplicative components – are excluded in our analysis by eliminating expression values at or below background levels and by log-transformation of the data. Expression deviations η are determined for each sample as a deviation of the gene expression ordinates from regression line calculated against the averaged profile for gene expressions in all samples of the control group. The majority of these deviations follow a normal distribution. The SD of this distribution is used for identification of hypervariable genes whose expression levels vary in a statistically significant manner from the reference group of stable genes as determined using an F-test (Fig. 1).

Discriminant function analysis (DFA)

DFA was used for selection of the set of genes that maximally discriminate among the groups studied. A forward stepwise DFA was performed in accordance with the manufacturer's instructions, using the statistical software

package Statistica (StatSoft, Tulsa, OK, USA). In this analysis, the model for discrimination is built in a stepwise manner. Specifically, at each step all variables are reviewed to determine which will maximally discriminate among groups. This variable is then included in a discriminative function, denoted a root, which is an equation consisting of a linear combination of gene expression changes used for the prediction of group membership. An F test is used to determine the statistical significance of the discriminatory power of the selected genes. The stepwise procedure is 'guided' by a standard threshold for the F test (established by the analytical package). In general, variables will continue to be included in the model, as long as the respective F values for those variables are larger than this standard threshold. The 170 genes expressed statistically significant from background in all five groups of samples (expression levels >3 SD over background as defined above) were used for this analysis.

The discriminant potential of the final equations can be observed in a simple multidimensional plot of the values of the roots obtained for each group. This provides a graphical representation of the similarity among the various groups. The discriminative power of each gene can also be characterized by the partial Wilks λ coefficient. This value is equal to the ratio of within-group differences in expression to within- and between-group differences in expression. Its value ranges from 1.0 (no discriminatory power) to 0.0 (perfect discriminatory power).

Biochemical function and pathway analysis
The genes in the data tables presented herein are functionally annotated. Gene functions were obtained from the Swiss-Prot Protein knowledge base (when available). This database was created and is maintained by the Swiss Institute of Bioinformatics (Biozentrum - Basel University, Basel, Switzerland). Additionally, the software package Pathway Assist (Strategene, La Jolla, CA, USA) was used to identify functional interrelationships among the genes defined as JRA-related in the analyses described above. This software uses the KEGG, DIP, and BIND databases and natural language scans of Medline to define functionally related genes. These functional relationships were then graphically represented by the software as a network.

All original programs were written using MathLab and Statistica statistical software and are available on request from igor-dozmorov@omrf.ouhsc.edu.

Results
Differential gene expression analysis of active disease
Statistical analysis of the difference of gene expression in samples from 9 patients and 12 healthy controls gave the following results. 1716 genes of the total number of 2382 genes in the microarray were expressed distinctively from background ($P < 0.05$) in both groups. Of these, 78 were

statistically differentially expressed in either patients or controls. These genes passed the Student's t-test at the threshold of 0.05 and the associative t-test at the threshold of 0.0005, a stringency that results in the selection of less than one expected false positive and less than one expected false negative determination. This analysis classifies differentially expressed genes into four groups:
1. genes expressed at higher levels on average in untreated patients with active disease, relative to healthy controls (34 identified, Table 2A);
2. genes expressed at lower levels in treated patients with active disease, relative to healthy controls (15 identified, Table 2B).
3. genes whose expression was detected above background only in untreated patients with active disease (18 identified, Table 2C); and
4. genes whose expression was detected above background only in healthy controls (2 identified, Table 2D).

Differential gene expression analysis is a common means of identifying the genes involved in a given pathophysiology. Our analysis identified key regulators of innate immunity and inflammation including the proinflammatory mediators formyl peptide receptor 1, ICAM-1 (intercellular adhesion molecule-1), thymosin β4, and PLA-2 (phospholipase A$_2$), which were up-regulated in patients, and the anti-inflammatory mediator TNF receptor 1 (TNF-R1), which was down-regulated in patients. Genes regulating the adaptive immune response were also identified, including those for β2 microglobulin, MHC class I, GTP-binding protein-HSR1 (a polymorphic microsatellite marker in the human MHC class I region), and Semaphorin/CD100 (a B-cell and dendritic-cell surface receptor that modulates cellular activation), which were all up-regulated in patients, and the gene for transcription factor 8 (a repressor of IL-2 expression), which was down-regulated in patients. These data highlight the importance of these genes in regulating the immune and inflammatory response in JRA.

Interestingly, several of the immunoregulatory genes that were up-regulated in patients are known to be induced by interferon γ (IFN-γ), including those for thymosin β4, MHC class I, and ICAM-1, suggesting that this cytokine is increased in patients. To test this hypothesis, serum IFN-γ levels were assessed by ELISA in 4 patients with active disease and in a group of 12 healthy controls. Patient serum IFN-γ levels were significantly higher than in healthy controls ($P < 0.00067$). Values ranged from 60 to 1,626 pg/ml in patients and from <1.4 (the level of sensitivity of the assay) to 9.6 pg/ml in healthy controls (Fig. 2), implicating IFN-γ in the pathophysiology of polyarticlular JRA.

To more fully disclose the pathways relevant to JRA pathogenesis, the genes identified as differentially expressed in patients were grouped according to function using

Table 2

Differentially expressed genes in patients with polyarticular rheumatoid polyarthritis (*n* = 9) and healthy controls (*n* = 12)

A. Genes overexpressed in acute untreated patients

Gene bank	Name	Description	AverAD	AverHC	AD/HC	Summary of function
S54761	B2M	β_2-mu, β_2-microglobulin	266.2	57.5	4.6	β_2-microglobulin; major component of the hemodialysis-associated amyloid fibrils
L20941	FTHL6	Ferritin heavy chain	264.8	90.3	2.9	Ferritin heavy polypeptide 1; iron-storage protein
M17733	TMSB4X	Thymosin β_4	251.7	56.1	4.5	Thymosin β_4; sequesters actin monomers and inhibits actin polymerization
M11147	FTL	Ferritin L chain	230.3	82.9	2.8	Ferritin light polypeptide; iron storage protein
		Homologous to elongation factor 1α 1 (PTI-1)	224.7	59.1	3.8	Homologous with a truncated and mutated form of human elongation factor 1α subunit
X04098	ACTG	Cytoskeletal γ-actin	219.4	49.4	4.4	γ-actin; member of the non-muscle family of actins
X52008	GLR	$\alpha2$ subunit of inhibitory glycine receptor	215.9	74.5	2.9	$\alpha2$ subunit of the glycine receptor chloride channel; binds strychnine and is important for inhibitory neurotransmission
M11354		H3.3 histone, class B	213.1	62.5	3.4	Member of the H3 histone family; involved in compaction of DNA into nucleosomes
Y14040	CASH	CASH β protein	206.9	71.2	2.9	Caspase-like apoptosis regulatory protein; lacks caspase catalytic activity
Y13829	EXP40	MBNL protein	184.1	79.4	2.3	Strongly similar to uncharacterized KIAA0428
		CD74	158.4	67.0	2.4	HLA-DR antigens associated invariant chain
		Coactiosin-like protein	131.6	62.0	2.1	Interacts with 5-lipoxygenase
AF010187	FIBP	FGF-1 intracellular binding protein (FIBP)	127.6	43.4	2.9	Acidic fibroblast growth factor intracellular binding protein; may mediate the mitogenic properties associated with acidic FGF1
M60627	FMLP	*N*-formylpeptide receptor (fMLP-R26)	122.7	64.7	1.9	Formyl peptide receptor 1, a G protein-coupled receptor; binds bacterial *N*-formyl-methionyl peptides
X91257	SERRS	Seryl-tRNA synthetase	111.8	59.2	1.9	Cytosolic seryl-tRNA synthetase; class II aminoacyl tRNA synthetase, aminoacylates its cognate tRNAs with serine during protein biosynthesis
L13463	G0S8	Helix-loop-helix basic phosphoprotein (G0S8)	108.9	65.2	1.7	Regulator of G-protein signalling 2; negatively regulates G protein-coupled receptor signalling; has a basic helix-loop-helix motif
M77693	SSAT	Spermidine/spermine *N*1-acetyltransferase	108.0	34.2	3.2	Spermidine/spermine *N*1-acetyltransferase; catalyzes rate-limiting step in polyamine catabolism
J03077	SAP1	Co-β-glucosidase (proactivator)	107.3	37.0	2.9	Prosaposin; precursor of saposins A-D, may bind and transport gangliosides, cleavage products activate lysosomal hydrolysis of sphingolipids
X16478		5′ fragment for vimentin N-terminal fragment	101.0	42.7	2.4	Intermediate filament subunit
J00068	NEM2	Adult skeletal muscle α-actin mRNA	91.7	39.8	2.3	$\alpha1$ actin; skeletal muscle-specific actin
M63603	PLB	Phospholamban	89.4	46.3	1.9	Phospholamban; regulates the sarcoplasmic reticulum calcium pump
K00558	K-ALPHA-1	α-tubulin	77.9	36.9	2.1	α-tubulin (k-α-1); may be part of a heterodimer that polymerizes to form microtubules; member of a family of microtubule structural proteins

Table 2 *(Continued)*

A. Genes overexpressed in acute untreated patients *(Continued)*

Gene bank	Name	Description	AverAD	AverHC	AD/HC	Summary of function
D76444	KF1	hkf-1	68.0	36.0	1.9	May be associated with membranous protein sorting; contains a zinc finger domain
M27110	PLP	Proteolipid protein mRNA (PLP)	51.2	28.1	1.8	Proteolipid protein; predominant protein in myelin
AF001434	HPAST	Hpast (HPAST)	16.1	3.2	5.1	Very strongly similar to murine Ehd; may be involved in ligand-initiated endocytosis
M33882	IFI-78K	p78 protein	11.3	1.4	8.0	Similar to murine Mx; may be a guanine nucleotide-binding protein
AB006190	AQPap	mRNA for aquaporin adipose	8.4	1.8	4.8	Aquaporin 7; water and glycerol channel expressed predominantly in adipose tissue
D49489	P5	Protein disulfide isomerase-related protein P5	7.4	3.2	2.3	Member of the protein disulfide isomerase superfamily; contains two thioredoxin-like domains
X06990	BB2	Intercellular adhesion molecule-1 ICAM-1	5.8	1.5	3.8	Surface glycoprotein; binds the integrin LFA-1 (ITGB2) and promotes adhesion; member of the immunoglobulin superfamily
U39317	UBE2D2	E2 ubiquitin conjugating enzyme UbcH5B	5.7	2.4	2.4	Member of the ubiquitin-conjugating enzyme E2 subfamily; may catalyze ubiquitination of cellular proteins prior to degradation
L16842	UQCRC1	Ubiquinol cytochrome-*c* reductase core I protein	5.5	2.7	2.1	Core I protein; subunit of the ubiquinol-cytochrome-*c* oxidoreductase in the mitochondrial respiratory chain
U45448	P2X1	P2x1 receptor	4.7	1.7	2.8	Purinergic receptor 1; ligand-gated ion channel that may be gated by extracellular adenosine 5′-triphosphate (ATP)
AF083255	RHELP	RNA helicase-related protein	4.3	2.2	2.0	Moderately similar to human P72; may be an ATP-dependent helicase; member of DEAD/H box family, has conserved C-terminal helicase domain
U68536	ZNF24	Zinc finger protein	4.0	1.4	2.8	Zinc finger protein 24; contains zinc fingers

B. Genes overexpressed in healthy controls

Gene bank	Name	Description	AverAD	AverHC	HC/AD	Summary of function
U00968	SREBP1	SREBP-1	36.0	85.2	2.4	Transcription factor; activates genes involved in lipid metabolism, translocates to the nucleus and activates transcription of the LDL receptor and H MG CoA synthase genes in sterol-depleted cells
M36072	SURF-3	Ribosomal protein L7a (surf 3) large subunit	43.5	78.2	1.8	Ribosomal protein L7a; component of the 60-S ribosomal subunit
X80909	NACA	α NAC mRNA	32.9	68.0	2.1	Nascent-polypeptide-associated complex α subunit; binds nascent polypeptides and promotes the interaction between signal recognition particle and signal peptide
M15661	RPL36A	Ribosomal protein L36a	35.2	59.6	1.7	Ribosomal protein L36a; component of the large 60-S ribosomal subunit
U10248	HUMRPL29	Ribosomal protein L29 (humrpl29)	27.9	48.4	1.7	Ribosomal protein L29; component of the large 60-S ribosomal subunit, also functions as a cell surface heparin/heparan sulfate (HP/HS)-binding protein
M33294	TNF-R	Tumor necrosis factor receptor	10.6	32.4	3.1	Type I tumor necrosis factor receptor; mediates proinflammatory cellular responses; contains a juxtamembrane domain
D15050	AREB6	Transcription factor AREB6	14.6	32.4	2.2	Transcriptional modulator; inhibits interleukin-2 expression in T lymphocytes; contains a zinc finger domain

Table 2 *(Continued)*

B. Genes overexpressed in healthy controls *(Continued)*

Gene bank	Name	Description	AverAD	AverHC	HC/AD	Summary of function
U54559	EIF3S3	Translation initiation factor eIF3 p40 subunit	12.6	30.5	2.4	Translation initiation factor 3, subunit 3 (γ, 40kDa); subunit of the complex that stabilizes initiator Met-tRNA binding to 40-S subunits
U46751	P60	Phosphotyrosine independent ligand p62	9.5	29.1	3.1	Ubiquitin-binding protein; binds SH2 domain of p56lck and ubiquitin; contains G-protein-binding region, PEST and cys-rich zinc-finger-like motifs
AF017305	Unph	Deubiquitinating enzyme UnpEL (UNP)	11.4	21.7	1.9	Strongly similar to murine Unp; removes ubiquitin from ubiquitin-conjugated proteins; member of the ubiquitin-specific cysteine (thiol) protease family
M57567	ARF5	ADP-ribosylation factor (hARF5)	6.5	15.4	2.4	ADP-ribosylation factor 5, a GTP-binding protein; stimulates cholera toxin activity, may be involved in vesicular intracellular transport
U02609	TBL3	Transducin-like protein	5.3	9.3	1.8	Contains WD40 repeats
		NEDD5	3.0	9.0	3.0	Role of Nedd5 in neurite outgrowth
		CAPON	4.1	8.1	2.0	C-terminal PDZ domain ligand of neuronal nitric oxide synthase.
		Adenylate cyclase, type VII	3.4	6.4	1.9	Adenylate cyclase (type 7), an ATP-pyrophosphate lyase; converts ATP to cAMP

C. Genes expressed in active untreated patients only

| Gene bank	Name	Description	AverAD	AverHC	Summary of function
D14874	PROAM-N20	Adrenomedullin	6.8	ND	Precursor of adrenomedullin (AM) and the putative 20-amino-acid peptide proAM-N20; regulates blood pressure and heart rate
X86556	ACADVL	HVLCAD gene	2.8	ND	Very-long-chain-acyl-coenzyme-A dehydrogenase; oxidizes straight-chain acyl-CoAs
X78873	PPP1R2	Inhibitor 2 gene	2.6	ND	Inhibitory subunit 2 of protein phosphatase 1; associates with the γ isoform of protein phosphatase 1
M28099	FBP	Folate-binding protein (FBP)	1.4	ND	Adult folate-binding protein 1 (folate receptor α); binds and initiates transport of folate and methotrexate
U60800	CD100	Semaphorin (CD100)	1.4	ND	Member of the semaphorin family of chemorepellant proteins; induces B lymphocytes to aggregate and promotes their differentiation
M83233	HTF4A	Transcription factor (HTF4A)	1.2	ND	Transcriptional activator; binds to the immunoglobulin enhancer E-box consensus sequence; contains a basic helix-loop-helix domain
M22430	PLA2L	RASF-A PLA2	0.9	ND	Group IIA secretory phospholipase A_2; hydrolyzes the phospholipid sn-2 ester bond, releasing a lysophospholipid and a free fatty acid; similar to murine Pla2g2a
AF005080	XP5	Skin-specific protein (xp5)	0.9	ND	Skin-specific protein
U96759	VBP-1	von-Hippel–Lindau-binding protein (VBP-1)	0.8	ND	von-Hippel–Lindau-binding protein; binds tumor suppressor VHL and forms a complex with VHL protein; has a consensus site for tyrosine phosphorylation
X59498	TBPA	Ttr mRNA for transthyretin	0.7	ND	Transthyretin (prealbumin); carrier protein, transports thyroid hormones and retinol in the plasma
L25665	HSR1	GTP-binding protein (HSR1)	0.6	ND	Putative GTP-binding protein
U24163	FZRB	Frizzled related protein Frzb precursor (fzrb)	0.6	ND	Frizzled-related protein; similar to frizzled family of receptors

Table 2 *(Continued)*

C. Genes expressed in active untreated patients only *(Continued)*

Gene bank	Name	Description	AverAD	AverHC	Summary of function
X78031	FUCT-VII	α-1,3-fucosyl-transferase	0.4	ND	Leukocyte α-1,3-fucosyltransferase; functions in selectin ligand synthesis
L11924	MST1	Macrophage-stimulating protein (MST1)	0.4	ND	Proapoptotic when overexpressed; binds p53
M26393		Short-chain-acyl-CoA dehydrogenase	0.4	ND	Short-chain-acyl-coenzyme-A dehydrogenase; may act in the first step in beta-oxidation of C4–C6 fatty acids; strongly similar to
murine Acads					
U25033	NNAT	Neuronatin α	0.4	ND	Neuronatin; possibly functions to regulate ion channels during brain development
X95073	TRAX	Translin-associated protein X	0.4	ND	Interacts with translin (TSN)
U63336	CAT56	MHC class I region proline-rich protein	0.4	ND	Undefined

D. Genes expressed in healthy controls only

Gene bank	Name	Description	AverAD	AverHC	Summary of function
X57637	GGTA	mRNA involved in tapetochoroidal dystrophy	ND	1.3	Component A of geranylgeranyl transferase; modifies Rab proteins; has similarity to guanine nucleotide dissociation inhibitors
Z11566	PR22	Pr22 protein	ND	0.5	Stathmin (oncoprotein 18), a cytosolic phosphoprotein

AverAD, AverHC, average expression level (defined as the number of standard deviations from mean of background) in untreated patients with active disease and in healthy controls, respectively. ND, none detected

Figure 2
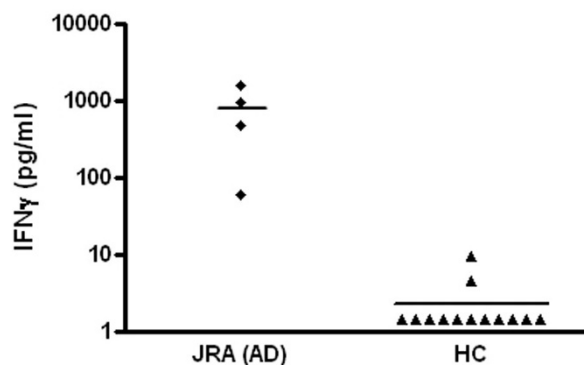
Serum IFN-γ levels in untreated patients with active juvenile rheumatoid arthritis (JRA) and healthy controls (HC). A scatter plot of serum IFN-γ concentrations in 4 patients with active disease (AD) and 13 HC is shown. The values for 11 HC that were <1.4 pg/ml (the limit of detection of the assay) are represented by triangular symbols that appear as the lowest value in the distribution. Average values in a given population are represented as a horizontal line. Concentrations are shown in pg/ml on a log scale.

recently developed commercial software (Pathway Assist, Ariadne Genomics, Rockville, MD, USA). A subset of functionally interrelated genes was identified and this network of genes graphically represented (Fig. 3). This analysis highlighted the importance of inflammatory and immune modulation, as well as such basic cellular processes relevant to leukocyte function as apoptosis, motility, and proliferation. The network of functionally related genes generated by this software allows the connections among these basic physiologic processes to be identified, demonstrating that the pathophysiologic response of these patients is highly coordinated.

Higher variability of genes in active disease

A novel analytical method was applied to the microarray data: identification of genes whose expression is relatively unchanging in the control population and becomes HV in JRA patients with active disease. The logical basis of this approach was based on the hypothesis that the loss of homeostasis characteristic of active autoimmune disease can be used to identify genes whose expression regulates the processes involved. For example, temperature is tightly regulated in healthy controls and is relatively stable on a population level. In patients with active polyarticular JRA, low-grade fever is relatively common and temperature levels vary on a population basis to a greater degree than in healthy controls. Therefore, the genes that code for regulators of pathophysiologic processes such as temperature control, or, by analogy, inflammatory response, may likewise be expected to vary on a population level in patients more than in healthy controls.

Figure 3
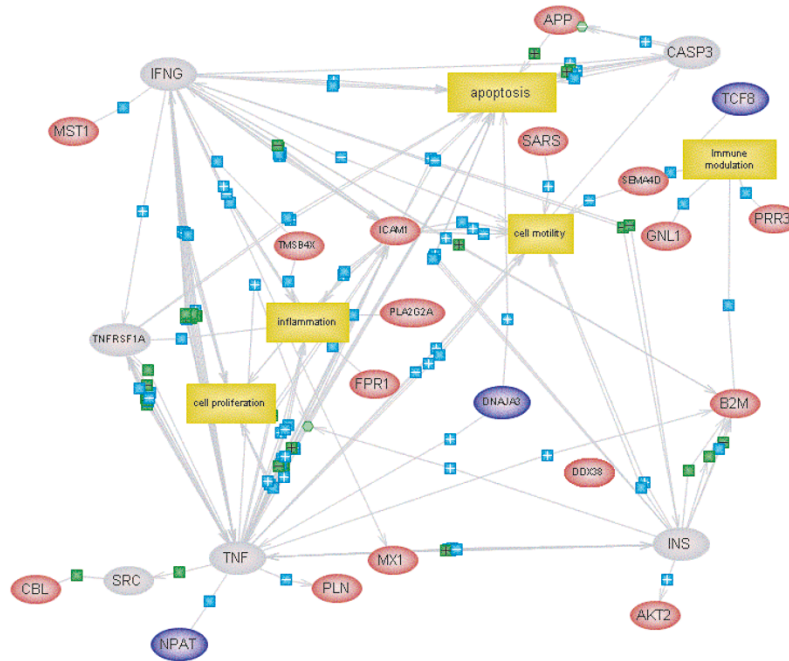
Functional associations of genes selected as differentially expressed in patients with juvenile rheumatoid arthritis (JRA) and normal controls. Tabular data from differential expression analysis were analyzed using Pathway Assist software. The graphical output delineating a functionally related network of genes is shown. Genes that were expressed at higher levels in JRA patients are represented as red ovals. Genes expressed at higher levels in controls are represented as blue ovals. Major biologic processes related to these genes are represented as yellow rectangles. White ovals represent genes that are functionally related to the genes used for analysis. Upon addition of these genes, several functional connections among the genes being analyzed can be observed. Green squares signify that a defined regulatory relationship exits between genes. Blue squares signify that a putative regulatory relationship between genes has been identified but not biochemically defined. +, positive regulation; −, negative regulation.

In this analysis, 444 genes were identified as HV genes in untreated patients with active disease and stable or expressed below background in healthy controls (see Additional file 1).

Among the 122 genes identified as HV genes in both groups, 27 had a statistically significant higher level of variation in untreated patients with active disease (Table 3). Many of the genes identified as increasing in variability in these patients have a direct role in inflammation and immune regulation and are known to be involved in inflammatory arthritis. These genes provide a more concise picture of the molecular pathophysiology of JRA than is obtained in a traditional analysis of differentially expressed genes and include: IL-8, MHC class I, regulators of TNF-α (e.g. TGFβ1-induced anti-apoptotic factor 1) and granulocyte/macrophage-colony-stimulating factor (GM-CSF) (e.g. cold shock protein A), and human cartilage protein gp-39 (a major secretory product of articular chondrocytes and synovial cells). It is of note that none of these 27 genes were identified by differential expression analysis.

Pathway analysis software was used to reveal the principal biologic processes revealed by these data. Interestingly,

while the genes identified by HV analysis were distinct from those identified by differential expression analysis, the physiologic processes identified, such as inflammation and immune modulation, apoptosis, and cell motility, were similar (Fig. 4).

Discriminant function analysis (DFA)
In the above analyses, genes with behavior that varies between patients with active disease and control individuals were identified. DFA is distinct from the above analyses in that it identifies a set of genes whose expression levels, as a group, vary among populations. In this analysis, genes with the most significant power to discriminate among groups when used as variables in a linear equation, denoted a root, were identified. The groups of genes identified by DFA are statistically interrelated and may therefore be functionally interrelated. For this analysis, the following groups were used: nine untreated patients with acute disease; five of these nine patients were followed up prospectively during treatment, with partially responsive, fully responsive, and nonresponsive patients defined as independent groups; and six healthy controls.
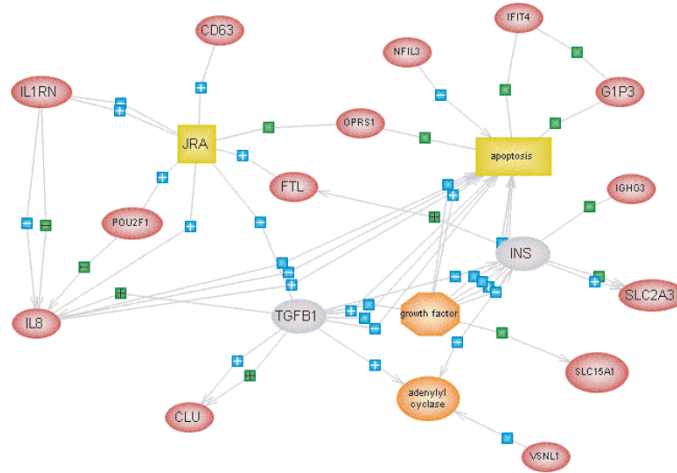
Table 3

Genes that distinguish patients with active juvenile rheumatoid arthritis by hypervariable gene analysis

Gene bank	Name	Description	F-test[a]	Summary of function
U26173	IL3BP1	bZIP protein NF-IL3A (IL3BP1)	4.5E-06	Basic leucine zipper (bZIP) transcription factor; activates IL3 gene expression and can also repress transcription; binds to regulatory sequences in the promoters of the adenovirus E4, γ interferon (IFNG), and interleukin 3 (IL3) genes
X87689	G6	Putative p64 CLCP protein	0.00245	Nuclear chloride channel-27; intracellular chloride channel
Y00787	IL-8	IL-8	0.00307	Interleukin 8; cytokine that plays a role in chemoattraction and activation of neutrophils; has similarity to several platelet-derived factors
M62831	ETR101	Transcription factor ETR101	0.00336	Immediate early gene product induced by TPA stimulation in promyelocytic leukemia cell line HL-60 and in other leukemia cell lines
X83394	HLA-C	HLA-Cw*0704	0.00553	Heavy chain that associates with β_2-microglobulin; forms MHC class I complex that binds peptides to present them to CTLs
D88378	PSMF1	Proteasome inhibitor hPI31 subunit	0.00556	A protein inhibitor of the 20-S proteasome
U08815	SPA61	Spliceosomal protein (SAP 61)	0.0059	Spliceosome-associated protein 3a, subunit 3; component of the essential heterotrimeric splicing factor SF3a; contains a zinc finger
M55067	NCF-1	47-kDa autosomal granulomatous disease protein	0.0068	Neutrophil cytosol factor 1; cytosolic component of NADPH oxidase, required for neutrophil superoxide production
D42063	RANBP2	RanBP2	0.0073	Ran binding protein 2; component of the nuclear pore complex; has leucine-rich and cyclophilin-like domains, and zinc finger motifs
M80927	YKL40	Glycoprotein (GP-39 synovial protein)	0.00793	Cartilage glycoprotein-39; has similarity to chitinases
X06272	SRPR	Docking protein	0.00826	Signal recognition particle receptor (docking protein); binds the SRP complexed with the translating ribosome to prepare for translocation of the polypeptide across the rough endoplasmic reticulum membrane
D44497	Clabp	Actin-binding protein p57	0.00981	Coronin 1A; binds actin, involved in mitosis, cell motility, formation of phagocytic vacuoles and phagocytosis; has five WD domains
X86693	HEVIN	Hevin-like protein	0.01158	Expressed in high endothelial venules, may have antiadhesive properties and a role in leukocyte extravasation.
D86970	MAJN	TGFβ1-induced antiapoptotic factor	0.01174	Protects against tumor necrosis factor-α (TNF)-induced apoptosis
X69391		Ribosomal protein L6	0.01251	mRNA for ribosomal protein L6
X76105	DAP	DAP-1 mRNA	0.01319	Death-associated protein; may mediate apoptosis induced by interferon-γ; has proline-rich regions
Y08110	SORLA	Mosaic protein LR11	0.02155	Sortilin-related receptor; may be involved in the uptake of lipoproteins and proteases
D88153	HYA22	HYA22	0.02189	Protein with similarity to *Saccharomyces cerevisiae* Psr1p and Psr2p
D21267	SNAP-25	Synaptosomal-associated protein, 25kDa	0.02199	Synaptosomal-associated protein, 25-kDa; involved in membrane fusion during exocytosis; member of the SNAP family of proteins
M11233	CPSD	Cathepsin D	0.02328	Cathepsin D; lysosomal aspartyl protease (acid protease)
D88827	ZNF#	Zinc finger protein FPM315	0.0244	C2H2-type zinc finger protein 263; may act as a transcriptional repressor; contains C2H2-type zinc fingers, and KRAB-A and LeR domains
U06452	MART-1	Melanoma antigen recognized by T cells	0.02575	Melan-A (melanoma antigen recognized by T cells 1)
X13403	OTF1	Octamer-binding protein Oct-1	0.02722	Ubiquitously expressed POU homeodomain transcription factor 1; binds to the octamer motif ATGCAAAT
X56976	UBE1X	Ubiquitin-activating enzyme E1	0.02928	Ubiquitin-activating enzyme E1; activates ubiquitin to mark cellular proteins for degradation; very strongly similar to murine Ube1x, which may function in DNA repair
X95325	CSDA	DNA-binding protein A variant	0.03421	Member of a family of transcriptional regulators; binds and represses the promoter of the GM-CSF gene; contains a cold-shock domain
X58536	HLA-C	HLA class I locus C heavy chain	0.035	Heavy chain that associates with β_2-microglobulin; forms MHC class I complex that binds peptides to present them to CTLs
U16031	IL-4-STAT	Transcription factor IL-4 Stat	0.03632	Signal transducer and activator of transcription 6, interleukin-4 induced; activates expression of the interleukin-4 receptor gene in response to interleukin-4 and mediates JAK kinase signal transduction; member of the STAT family

[a]Statistical parameter assessing the relative variation in patients with active disease vs healthy controls.

Figure 4

Functional associations of hypervariable (HV) genes in patients with juvenile rheumatoid arthritis (JRA). Tabular data from the HV gene analysis was analyzed using Pathway Assist software. The graphical output delineating a functionally related network of genes is shown. Genes that displayed increased variation in expression in JRA patients are represented as red ovals. Major biologic processes related to these genes are represented as yellow rectangles. JRA is represented as a yellow square to delineate genes directly associated with pathology. Intracellular objects upon which a given set of genes acts are represented as yellow ovals. White ovals represent genes that are functionally related to the genes used for analysis. White hexagons represent organelles functionally associated with the genes analyzed. Orange hexagons represent classes of small molecules associated with the genes analyzed. Green squares signify that a defined regulatory relationship exits between genes. Blue squares signify that a putative regulatory relationship between genes has been identified but not biochemically defined. +, positive regulation; −, negative regulation.

Interestingly, literature searches and pathway analysis revealed that nearly all of the 19 genes identified by this analysis are either directly or indirectly associated with immune modulation and autoimmune disease (Table 4; Fig. 5). These genes include: IL-1 receptor antagonist, an anti-inflammatory cytokine that has been recently approved by the FDA for use as a biologic therapy in inflammatory arthritis [19] and which also plays a significant role in the pathogenesis of polyarticular JRA [20]; TGFβ, another potent anti-inflammatory cytokine, which is involved in many biologic processes including immune homeostasis, regulation of apoptosis, and self-tolerance [21]; IL-8, which has been shown to be elevated in JRA patients and plays a key role in joint inflammation by recruiting neutrophils to synovial tissue and fluid [22,23]; ferritin, highlighting the significance of anemia of chronic disease associated with JRA [24]; transcription factor octamer-binding protein (Oct-1), which is expressed in synovial cells in the majority of RA patients and may modulate synovial outgrowth [25]; CD63, a leukocyte adhesion molecule and marker of dendritic cell maturation and neutrophil activation [26–28]; and GLUT/SLC2A, a glucose transporter protein expressed by chondrocytes that is regulated by TNF-α, IL-1β, IGF-1, and a key modulator of

skeletal development, chondrogenesis, and cartilage degradation in osteoarthritis [29].

Fourteen of the 19 genes identified by DFA were uniquely identified by this analysis; 2 of the 19, Ribosomal protein L37 and Ferritin light chain, were also identified by differential expression analysis, and 2 of the 19, IL-8 and Oct-1, were also identified by HV gene analysis, demonstrating that these three analyses are complementary, yielding predominantly nonoverlapping results (Fig. 6).

The values of the roots obtained by DFA analysis can be used to graphically represent the differences of the gene expression values obtained for the groups analyzed. Values obtained for individuals in a given group tended to cluster (Fig. 7), suggesting that the phenotypic changes occurring in a given group during treatment are common to all individuals of that group. Moreover, this graphical representation demonstrates that as patients respond to therapy, they tend to have expression profiles that are less like those of untreated patients with active disease and more like those of healthy controls (Fig. 7). In fact, the distance a given group of treated patients moved towards the normal controls was proportional to their level of response,

Figure 5

Functional associations of genes identified by discriminant function analysis (DFA) in patients with juvenile rheumatoid arthritis (JRA) who were undergoing treatment. Tabular data from the this analysis was analyzed using Pathway Assist software. The graphical output delineating a functionally related network of genes is shown. Genes that were associated with disease or treatment response are represented as red ovals. Major biologic processes related to these genes are represented as yellow rectangles. White ovals represent genes that are functionally related to the genes used for analysis. Orange hexagons represent classes of small molecules associated with the genes analyzed, and orange circles represent specific small molecules associated with the genes analyzed. Green squares signify that a defined regulatory relationship exits between genes. Blue squares signify that a putative regulatory relationship between genes has been identified but not biochemically defined. +, positive regulation; −, negative regulation.

with fully responsive patients moving closer to healthy controls than partially responsive patients (Fig. 7). In contrast, the position on the graph of the values obtained from a patient who was nonresponsive to standard therapy, whose disease severity actually increased after initiation of therapy (values for four separate samples taken between 2 and 8 weeks after therapy are represented in Fig. 7), changed in a manner that is distinct from those responding to therapy (responders were positioned away from the active disease group towards the healthy controls in a clockwise manner, and the four values obtained from the nonresponsive patient were positioned away from the active disease group in a counterclockwise manner) (Fig. 7). This method of pharmacogenomic analysis therefore provides a means of developing a clinical assay that may predict patients' response to therapy early in the course of polyarticular JRA treatment.

Discussion

We have used three distinct statistical methods for analyzing microarray data from children with polyarticular JRA. These methods include a standard analysis of differential gene expression, a novel means of assessing genes whose behavior dynamics are modulated in populations (denoted hypervariable gene analysis), and DFA to identify the genes and molecular pathways involved in the pathogenesis of polyarticular JRA. Each method identified both well established and novel mechanisms of disease pathogenesis, and because the biologic basis of each of the

methods is unique, the methods are complementary, with each highlighting a different aspect of the disease.

Analysis of differentially expressed genes demonstrated that in patients with active disease, principal modulators of both adaptive and innate immunity were up-regulated, consistent with the hyperactivation of the immune and inflammatory responses characteristic of this disease. Disease-specific up-regulation of a group of interferon-induced genes, including thymosin β4, MHC class I, and ICAM-1, was observed. Moreover, serum IFN-γ levels were dramatically increased in patients with active disease relative to healthy controls. These data provide the first broad-based molecular evidence for the existing hypothesis that JRA is a Th-1-biased disorder and demonstrate the pathologic immunomodulatory role of IFN-γ in these patients [30]. These data also suggest that therapy directed specifically at reducing IFN-γ levels may prove efficacious in treating JRA.

The fact that several genes involved in oxidative stress are among the genes most significantly up-regulated during active disease suggests that this process also plays a significant role in the pathophysiology of polyarticular JRA, as suggested by previous studies of oxidation products and free radical scavengers in patients [31]. These findings are consistent with our previous work demonstrating the importance of immune complexes (potent triggers to neutrophil superoxide release) [32,33] in the pathophysiology

Table 4

Genes that distinguish children with active juvenile rheumatoid arthritis in discriminant function analysis

Gene bank	Name	Description	AverAD	AverHC	Partial Wilks λ	Summary of function
M11147	FTL	Ferritin L chain	23.25	82.87	0.008	Ferritin light polypeptide; iron storage protein
D23661	RPL37	Ribosomal protein L37	65.72	73.62	0.027	Ribosomal protein L37; component of the large 60-S ribosomal subunit
M59907	GRANULOPHYSIN CD63		27.03	21.75	0.006	Melanoma-associated antigen; may function as a blood platelet activation marker
M20681	SLC2A3P	SLC2A3 glucose transporter	26.23	29.56	0.004	Facilitated glucose transporter
Y14737	HDC	Immunoglobulin heavy constant γ 3	22.24	56.99	0.002	Constant region of heavy chain of IgG3
U14747	VSNL1	VSNL1	22.05	12.47	0.174	Visinin-like protein 1; may bind calcium
X02812	TGFB1	TGFB1	20.11	30.1	0.005	Transforming growth factor β1; regulates cell proliferation, differentiation, and apoptosis
Y00787	IL-8	IL-8	19.36	54.2	0.018	Interleukin 8; cytokine that plays a role in chemoattraction and activation of neutrophils
AF026939	RIGG	IFNIT4 interferon-induced protein	16.68	1.97	0.026	Interferon-induced protein
X13403	OTF1	Octamer-binding protein Oct-1	15.46	14.96	0.005	Ubiquitously expressed POU homeodomain transcription factor 1
Y08110	SORLA	Sortilin-related receptor	14.48	28.66	0.002	Sortilin-related receptor; may be involved in the uptake of lipoproteins and proteases
U75283	SR-BP1	Sigma receptor	11.46	5.34	0.004	Type I sigma receptor; transmembrane protein that interacts with psychotomimetic drugs, including cocaine and amphetamines
X02492	15-Jun	Interferon-inducible fragment	10.3	1.88	0.024	Induced by α and β interferon; hydrophobic
M64722	APOJ	Clusterin	9.74	4.84	0.003	Clusterin, glycoprotein found in high-density lipoproteins and endocrine and neuronal granules; has a role in the terminal complement reaction
M55646	IL1RN	IL-1 receptor antagonist	7.03	3.69	0.056	Interleukin 1 receptor antagonist; binds to and inhibits the IL-1 receptor
X57198	TFIIS	Transcription elongation factor A	6.38	7.9	0.005	Transcription elongation factor A (SII); stimulates the activity of the RNA polymerase II elongation complex
AF043233	HPECT1	Caco-2 oligopeptide transporter	5.49	2.87	0.004	H(+)-coupled peptide transporter; absorbs small peptides produced by digestion of dietary proteins and may transport β-lactam antibiotics
U26173	IL3BP1	Interleukin 3 regulated nuclear factor	5.4	6.58	0.033	Basic leucine zipper (bZIP) transcription factor; activates IL3 gene expression and can also repress transcription; binds to regulatory sequences in the promoters of the adenovirus E4, γ interferon (IFNG), and interleukin 3 (IL3) genes
X05997		Gastric lipase	5.27	28.58	0.008	Digestion of dietary triglycerides in the gastrointestinal tract

AverAD, AverHC, average expression level (defined as the number of standard deviations from mean of background) in untreated patients with active disease and in healthy controls, respectively. The discriminative power of each gene can also be characterized by the partial Wilks λ coefficient. This value is equal to the ratio of within-group differences in expression to within- and between-group differences in expression. It ranges from 1.0 (no discriminatory power) to 0.0 (perfect discriminatory power). ND, none detected.
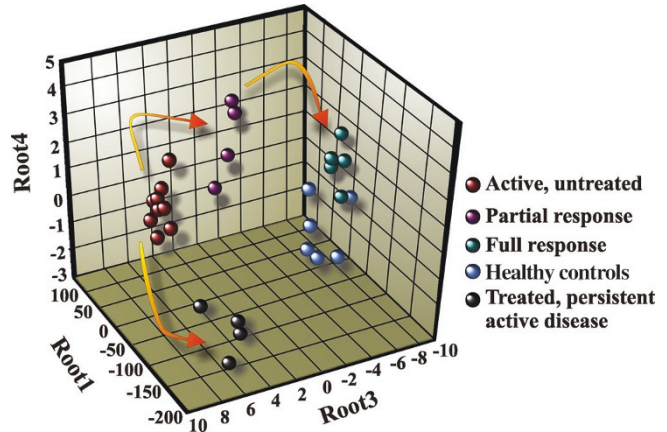
Figure 6
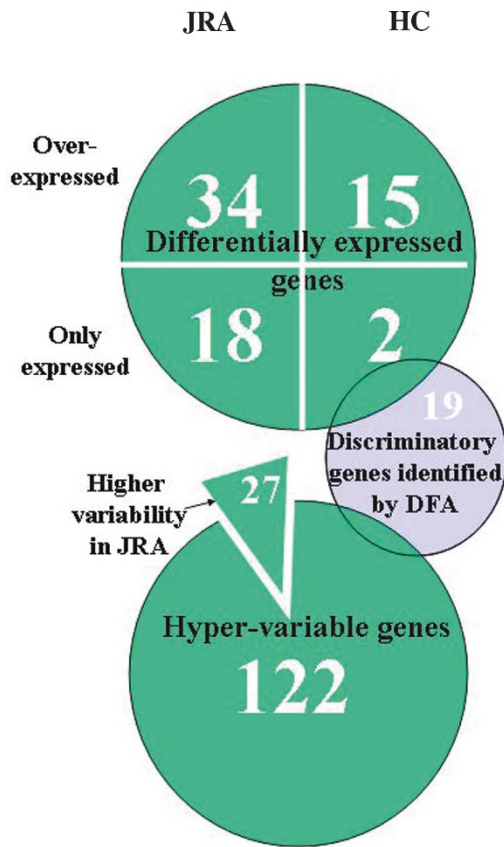
An overview of the results from the three analytical methods. The numbers of genes that were identified by differential expression analysis, hypervariable (HV) gene analysis, and discriminant function analysis (DFA) are represented in a Venn diagram. The numbers of genes identified uniquely in a given analysis as relevant to patients with juvenile rheumatoid arthritis (JRA) are shown in nonoverlapping regions. The numbers of genes identified in more than one analysis are shown in overlapping regions. HC, genes expressed at higher levels in healthy controls; JRA, genes expressed at higher levels in patients.

Figure 7

A graphical representation of the discriminatory potential of discriminant function analysis (DFA). DFA was used to identify, in a cohort of treated patients with juvenile rheumatoid arthritis (JRA) and healthy controls, a subset of genes whose expression values can be linearly combined in an equation, denoted a root, whose overall value is distinct for a given characterized group. The expression values of the individuals in the cohort were plotted in three dimensions to visually represent the relative differences in gene expression among the distinct populations. Gene expression values were plotted on this graph for nine untreated patients with active disease. Five of these nine patients were followed up prospectively during treatment. Four patients responded to treatment and one patient was nonresponsive. The values obtained for the four responsive patients at the time of partial response (after 2–4 weeks of treatment) and full response (6–8 weeks) are shown, as are four independent values obtained from a nonresponsive patient (taken during an 8-week interval). Values for healthy controls are also represented. Values obtained for individuals from each of these groups tended to cluster. Response to therapy was reflected in the spatial relationships among groups.

of polyarticular JRA [14] and provide a logical basis to begin investigations of novel treatment modalities targeting these pathways.

This work represents the first application of hypervariable gene analysis to the study of human disease of which we are aware. This statistical measure of gene variability was designed to identify genes that have lost their normal regulation in a manner that mimics the loss of homeostasis characteristic of autoimmune disease. Regulation of genes that are involved in such processes as inflammatory cell and immune cell activity, which will vary in groups of affected individuals more than in healthy controls, would therefore be expected to become increasingly chaotic at the population level in groups of affected individuals. This hypothesis is validated by the fact that a significant number of genes identified by this analysis are involved in

immune and inflammatory regulation and have been implicated as mediators of autoimmune disease. Importantly, the overall picture of disease is more accurately depicted when the genes identified by HV analysis are added to those identified as differentially expressed. For example, neutrophils play a principal role in disease pathogenesis [34–36,1]. However, genes that regulate neutrophil function, such as those for IL-8, 47-kDa autosomal granulomatous disease protein, DNA-binding protein A variant (which regulates GM-CSF production), and cathepsin D, a potent neutrophil collagenase found to be highly expressed in the synovium of JRA, RA, and osteoarthritis patients [37–39], were identified by HV gene analysis and not by identification of genes that are up- or down-regulated in patients or controls. Moreover, the genes for YKL-40 (or cartilage protein-39), a cartilage-derived autoantigen that is up-regulated in RA and thought to be a direct target of autoimmune attack in that disease [40–43], EDG-1, a pro-angiogenic protein essential for vascular maturation [44], and Oct-1, a transcription factor expressed in RA synovial cells and which regulates gene transcription [21], were more highly variable in JRA than in

controls, indicating that these proteins may play a major role in joint-specific pathology, including erosions, in polyarticular JRA, although they were not detected as significantly differentially regulated. On the basis of these findings, it is interesting to speculate that polyarticular JRA may be kept in check by the synchrony of a distinct subset of disease-specific genes whose dysregulation can predispose a patient to a flare in disease activity. These results suggest that hypervariable gene analysis will be a useful adjunct to traditional analyses of differential gene expression.

Among the most important aspects of the results reported here was the fact that DFA of microarray data may predict therapeutic response to specific therapies. The predictive potential of this analysis was based on the finding that the dynamics of gene expression in patients with active disease were modulated towards levels characteristic of healthy controls in proportion to a given individual's response to therapy. These data suggest that successful immunosuppressive therapy re-establishes some level of normality in these patients. This hypothesis is consistent with the fact that, in effectively treated patients, the disease is suppressed for some time, as if patients have re-established normal, or near-normal, homeostasis. It is also consistent with the clinical observation that disease activity can flare after periods of relative quiescence, as if this re-established homeostasis requires some trigger to be disrupted.

It is useful to note that there was uniform regression towards a normal profile, whether the response was achieved using nonsteroidal anti-inflammatory drugs alone or required more potent agents such as methotrexate, suggesting that the actions of these agents on the immune and inflammatory response in a successfully treated patient are similar. Given that treatment response can be predicted by DFA relatively early in the treatment course, these results demonstrate the potential for using pharmacogenomic analyses to identify the most efficacious treatment modality for a given patient.

We do not assert that this analysis is the 'final say' for predicting therapeutic response in polyarticular JRA. These preliminary investigations will need to be followed up longitudinally with a larger group of patients, and individual response profiles will need to be developed for specific agents (e.g. nonsteroidal anti-inflammatory drugs, methotrexate, etc). Despite these limits, these data conclusively demonstrate the power of analyzing gene expression behavior using DFA in analysis of complex diseases such as polyarticular JRA.

Conclusion
We have demonstrated that the relevance of microarray data can be maximized by applying bioinformatics tools that specifically address the nature of the data obtained. Standard differential gene expression analysis was used to illuminate specific pathways relevant to disease pathophysiology. Moreover, a novel statistical method was created specifically to exploit the fact that autoimmune disease is characterized by loss of homeostasis, and therefore that expression of immune and inflammatory genes that regulate and affect these pathophysiologic processes will vary more in patients than in healthy controls. Lastly, DFA was used to analyze prospective data from patients treated with conventional therapy, as this is the most powerful analytical tool for obtaining data from a complex cohort. These later methods proved complementary to the conventional analysis of microarray data and highlighted previously characterized aspects and identified novel aspects of disease pathophysiology. Importantly, the results of the DFA could be used to develop a clinical assay that may predict therapeutic response in patients early in the treatment course, demonstrating that molecular outcomes, when measured on a comprehensive scale, can be used as prognostically relevant biologic markers of disease activity.

Additional file

The following Additional file is available online:

Additional file 1
An Excel file containing a table that lists all the genes identified as hypervariable (HV) in patients vs controls. This includes 45 genes HV in healthy controls and stable or not expressed in untreated patients with active disease, 444 genes HV in untreated patients with active disease and stable or not expressed in healthy controls, and 122 genes HV in both groups. The 122 genes HV in both groups were used for subsequent analysis as described in the text. Gene names, accession number, and variability status in groups are shown.
See http://arthritis-research.com/content/supplementary/ar1018-s1.xls

Competing interests
None declared.

References
1. Jarvis JN: **Pathogenesis and mechanisms of inflammation in the childhood rheumatic diseases.** *Curr Opin Rheumatol* 1998, **10:**459-467.

2. Arend WP: **The innate immune system in rheumatoid arthritis.** *Arthritis Rheum* 2001, **44:**2224-2234.

3. Lo D, Feng L, Li L, Carson MJ, Crowley M, Pauza M, Nguyen A, Reilly CR: **Integrating innate and adaptive immunity in the whole animal.** *Immunol Rev* 1999, **169:**225-239.

4. Fearon DT, Locksley RM: **The instructive role of innate immunity in the acquired immune response.** *Science* 1996, **72:**50-53.

5. Warrington KJ, Seisuke T, Goronzy JJ, Weyand CM: **CD4+, CD28– T cells in rheumatoid arthritis patients combine features of innate and adaptive immune systems.** *Arthritis Rheum* 2001, **44:**13-20.

6. Jarvis, JN: **Juvenile rheumatoid arthritis: a guide for practitioners.** *Pediatr Ann* 2002, **31:**437-446.

7. Bowcock AM, Shannon W, Du F, Duncan J, Cao K, Aftergut K, Catier J, Fernandez-Vina MA, Menter A: **Insights into psoriasis and other inflammatory diseases from large-scale gene expression studies.** *Hum Mol Genet* 2001, **10:**1793-1805.

8. Lawrance IC, Fiocchi C, Chakravarti S: **Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes.** *Hum Mol Genet* 2001, **10:**445-456.

9. Heller RA, Schena M, Chai A, Shalon D, Bedilion T, Gilmore J, Woolley DE, Davis RW: **Discovery and analysis of inflammatory disease-related genes using cDNA microarrays.** *Proc Natl Acad Sci USA* 1997 **94:**2150-2155.

10. Cassidy JT, Levinson JE, Brewer EJ: **The development of classification criteria for children with juvenile rheumatoid arthritis.** *Bull Rheum Dis* 1989, **38:**1-7.

11. Jarvis JN, Pousak T, Krenz M: **Detection of IgM rheumatoid factors by ELISA in children with juvenile rheumatoid arthritis: correlation with articular disease and laboratory abnormalities.** *Pediatrics* 1992, **90:**945-949.

12. Brewer EJ, Giannini EH, Kuzmina N, Alekseev L: **Penicillamine and hydroxychloroquine in the treatment of severe juvenile rheumatoid arthritis. Results of the USA–USSR double-blind controlled trial.** *N Engl J Med* 1986, **314:**1269-1276.

13. Jarvis JN, Pousak T, Krenz M, Iobidze MV, Taylor H: **Complement activation and immune complexes in juvenile rheumatoid arthritis.** *J Rheumatol* 1993, **20:**114-117.

14. Jarvis JN, Iobidze M, Taylor H, Krenz M: **Complement activation and immune complexes in children with polyarticular JRA: a longitudinal study.** *J Rheumatol* 1994, **21:**1124-1127.

15. Dozmorov I, Centola M: **An associative of gene expression array data.** *Bioinformatics* 2003, **19:**204-211.

16. Rocke DM, Durbin B: **A model for measurement error for gene expression arrays.** *Comput Biol* 2001, **8:**557-569.

17. Zien A, Aigner T, Zimmer R, Lengauer T: **Centralization: a new method for the normalization of gene expression data.** *Bioinformatics* 2001, **Suppl 1:**S323-331.

18. Lee ML, Kuo FC, Whitmore GA, Sklar J: **Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations.** *Proc Natl Acad Sci USA* 2000, **97:**9834-9839.

19. Chikanza IC: **Juvenile rheumatoid arthritis: therapeutic perspectives.** *Paediatr Drugs* 2002, **4:**335-348.

20. Muzaffer MA, Dayer JM, Feldman BM, Pruzanski W, Roux-Lombard P, Schneider R, Laxer RM, Silverman ED: **Differences in the profiles of circulating levels of soluble tumor necrosis factor receptors and interleukin 1 receptor antagonist reflect the heterogeneity of the subgroups of juvenile rheumatoid arthritis.** *J Rheumatol* 2002, **29:**1071-1078.

21. Bommireddy R, Ormsby I, Yin M, Boivin GP, Babcock GF, Doetschman T: **TGFbeta1 inhibits Ca(2+)-calcineurin-mediated activation in thymocytes.** *J Immunol* 2003, **170:**3645-3652.

22. Khalkhali-Ellis Z, Bulla GA, Schlesinger LS, Kirschmann DA, Moore TL, Hendrix MJ: **C1q-containing immune complexes purified from sera of juvenile rheumatoid arthritis patients mediate IL-8 production by human synoviocytes: role of C1q receptors.** *J Immunol* 1999, **163:**4612-4620.

23. De Benedetti F, Pignatti P, Bernasconi S, Gerloni V, Matsushima K, Caporali R, Montecucco CM, Sozzani S, Fantini F, Martini A: **Interleukin 8 and monocyte chemoattractant protein-1 in patients with juvenile rheumatoid arthritis. Relation to onset types, disease activity, and synovial fluid leukocytes.** *J Rheumatol* 1999, **26:**425-431.

24. Kirel B, Yetgin S, Saatci U, Ozen S, Bakkaloglu A, Besbas N: **Anaemia in juvenile chronic arthritis.** *Clin Rheumatol* 1996, **15:**236-241.

25. Wakisaka S, Suzuki N, Takeno M, Takeba Y, Nagafuchi H, Saito N, Hashimoto H, Tomita T, Ochi T, Sakane T: **Involvement of simultaneous multiple transcription factor expression, including cAMP responsive element binding protein and OCT-1, for synovial cell outgrowth in patients with rheumatoid arthritis.** *Ann Rheum Dis* 1998, **57:**487-494.

26. Matsunaga T, Ishida T, Takekawa M, Nishimura S, Adachi M, Imai K: **Analysis of gene expression during maturation of immature dendritic cells derived from peripheral blood monocytes.** *Scand J Immunol* 2002, **56:**593-601.

27. Lopez S, Halbwachs-Mecarelli L, Ravaud P, Bessou G, Dougados M, Porteu F: **Neutrophil expression of tumour necrosis factor receptors (TNF-R) and of activation markers (CD11b, CD43, CD63) in rheumatoid arthritis.** *Clin Exp Immunol* 1995, **101:**25-32.

28. Koyama Y, Suzuki M, Yoshida T: **CD63, a member of tetraspan transmembrane protein family, induces cellular spreading by reaction with monoclonal antibody on substrata.** *Biochem Biophys Res Commun* 1998, **246:**841-846.

29. Mobasheri A, Vannucci SJ, Bondy CA, Carter SD, Innes JF, Arteaga MF, Trujillo E, Ferraz I, Shakibaei M, Martin-Vasallo P: **Glucose transport and metabolism in chondrocytes: a key to understanding chondrogenesis, skeletal development and cartilage degradation in osteoarthritis.** *Histol Histopathol* 2002, **17:**1239-1267.

30. Scola MP, Thompson SD, Brunner HI, Tsoras MK, Witte D, Van D, Grom AA, Passo MH, Glass DN: **Interferon-gamma:interleukin 4 ratios and associated type 1 cytokine expression in juvenile rheumatoid arthritis synovial tissue.** *J Rheumatol* 2002, **29:**369-378.

32. Araujo V, Arnal C, Boronat M, Ruiz E, Dominguez C: **Oxidant-antioxidant imbalance in blood of children with juvenile rheumatoid arthritis.** *Biofactors* 1998, **8:**155-159.

32. Robinson JJ, Watson F, Bucknall RC, Edwards SW: **Stimulation of reactive oxidant production in neutrophils by soluble and insoluble immune complexes occurs via different receptors/signal transduction systems.** *FEMS Immunol Med Microbiol* 1994 **8:**249-257.

33. Amit A, Kindzelskii AL, Zanoni J, Jarvis JN, Petty HR: **Complement deposition on immune complexes reduces the frequencies of metabolic, proteolytic, and superoxide oscillations of migrating neutrophils.** *Cell Immunol* 1999, **194:**47-53.

34. Lin SJ, Huang JL, Chao HC, Lee WY, Yang MH: **A follow-up study of systemic-onset juvenile rheumatoid arthritis in children.** *Acta Paediatr Taiwan* 1999, **40:**176-181.

35. Sikora A, Brozik H, Sikora JP, Golebiowska M: **Chemiluminescence of peripheral blood leukocytes and activity of an inflammatory process in juvenile chronic arthritis (JCA).** *Acta Univ Carol [Med] (Praha)* 1994, **40:**75-79.

36. Egger G, Klemt C, Spendel S, Kaulfersch W, Kenzian H: **Migratory activity of blood polymorphonuclear leukocytes during juvenile rheumatoid arthritis, demonstrated with a new whole-blood membrane filter assay.** *Inflammation* 1994, **18:**427-441.

37. Taubert H, Riemann D, Kehlen A, Meye A, Bartel F, John V, Brandt J, Bache M, Wurl P, Schmidt H, Weber E: **Expression of cathepsin B, D and L protein in juvenile idiopathic arthritis.** *Autoimmunity* 2002, **35:**221-224.

38. Keyszer GM, Heer AH, Kriegsmann J, Geiler T, Trabandt A, Keysser M, Gay RE, Gay S: **Comparative analysis of cathepsin L, cathepsin D, and collagenase messenger RNA expression in synovial tissues of patients with rheumatoid arthritis and osteoarthritis, by in situ hybridization.** *Arthritis Rheum* 1995, **38:**976-984.

39. Poole AR, Hembry RM, Dingle JT, Pinder I, Ring EF, Cosh J: **Secretion and localization of cathepsin D in synovial tissues removed from rheumatoid and traumatized joints. An immunohistochemical study.** *Arthritis Rheum* 1976, **19:**1295-1307.

40. Matsumoto T, Tsurumoto T: **Serum YKL-40 levels in rheumatoid arthritis: correlations between clinical and laborarory parameters.** *Clin Exp Rheumatol* 2001, **19:**655-660.

41. Vos K, Miltenburg AM, van Meijgaarden KE, van den Heuvel M, Elferink DG, van Galen PJ, van Hogezand RA, van Vliet-Daskalopoulou E, Ottenhoff TH, Breedveld FC, Boots AM, de

Vries RR: **Cellular immune response to human cartilage glyco-protein-39 (HC gp-39)-derived peptides in rheumatoid arthritis and other inflammatory conditions.** *Rheumatology (Oxford)* 2000, **39:**1326-1331.

42. Patil NS, Sonderstrup G, Mellins ED: **Autoantigenic HCgp39 epitopes are presented by the HLA-DM-dependent presentation pathway in human B cells.** *J Immunol* 2001, **166:**33-41.

43. Hakala BE, White C, Recklies AD: **Human cartilage gp-39, a major secretory product of articular chondrocytes and synovial cells, is a mammalian member of a chitinase protein family.** *J Biol Chem* 1993, **268:**25803-25810.

44. Liu Y, Wada R, Yamashita T, Mi Y, Deng CX, Hobson JP, Rosenfeldt HM, Nava VE, Chae SS, Lee MJ, Liu CH, Hla T, Spiegel S, Proia RL: **Edg-1, the G protein-coupled receptor for sphingosine-1-phosphate, is essential for vascular maturation.** *J Clin Invest* 2000, **106:**951-961.

Correspondence

Dr James Jarvis, Pediatric Rheumatology Research, Basic Sciences Education Building #235A, University of Oklahoma College of Medicine, Oklahoma City, OK 73104. Tel: +1 405 271 4755; fax: +1 405 271 2281; e-mail: James-Jarvis@ouhsc.edu