

Gene expression in juvenile arthritis and spondyloarthropathy: pro-angiogenic ELR⁺ chemokine genes relate to course of arthritis

M. G. Barnes, B. J. Aronow¹, L. K. Luyrink, M. B. Moroldo, P. Pavlidis², M. H. Passo, A. A. Grom, R. Hirsch³, E. H. Giannini, R. A. Colbert, D. N. Glass and S. D. Thompson

Objective. To evaluate the ability of microarray-based methods to identify genes with disease-specific expression patterns in peripheral blood mononuclear cells (PBMC) and synovial fluid mononuclear cells (SFMC) of juvenile arthritis patients and healthy controls.

Methods. Microarray data (Affymetrix U95Av2) from 26 PBMC and 20 SFMC samples collected from patients with active disease (classified by course according to ACR criteria) were analysed for expression patterns that correlated with disease characteristics. For comparison, PBMC gene expression profiles were obtained from 15 healthy controls. Real-time PCR was used for confirmation of gene expression differences.

Results. Statistical analysis of gene expression patterns in PBMC identified 378 probe sets corresponding to 342 unique genes with differing expression levels between polyarticular course patients and controls (*t* test, *P* < 0.0001). The genes represented by these probe sets were enriched for functions related to regulation of immune cell functions, receptor signalling as well as protein metabolism and degradation. Included in these probe sets were a group of CXCL chemokines with functions related to angiogenesis. Further analysis showed that, whereas angiogenic CXCL (ELR⁺) gene expression was elevated in polyarticular PBMC, expression of angiostatic CXCL (ELR⁻) chemokines was lower in polyarticular SFMC compared with corresponding pauciarticular samples (*t* test, *P* < 0.05).

Conclusions. This pilot study demonstrates that juvenile arthritis patients exhibit complex patterns of gene expression in PBMC and SFMC. The presence of disease-correlated biologically relevant gene expression patterns suggests that the power of this approach will allow better understanding of disease mechanisms, identify distinct clinical phenotypes in disease subtypes, and suggest new therapeutic approaches.

KEY WORDS: Human, Juvenile, Rheumatoid arthritis, Chemokines, Inflammation, Polyarticular course, JAK/STAT.

Chronic inflammatory arthritis is a source of morbidity for about 70 000 children in the USA alone. Currently three classification systems for the juvenile arthritides exist [1]. Juvenile rheumatoid arthritis (JRA) is defined by ACR criteria [2], while juvenile chronic arthritis (JCA) and juvenile idiopathic arthritis (JIA) correspond to criteria of the European League Against Rheumatism (EULAR) [1] and the International League of Associations for Rheumatology (ILAR) [3] respectively. Despite differences in classification systems, subtypes of juvenile arthritis are generally characterized by the number of affected joints within 6 months of onset. In the JRA classification, pauciarticular arthritis is defined by the involvement of four or fewer joints, while polyarticular arthritis affects five or more joints. Systemic-onset JRA (SOJRA) has distinct characteristics, including spiking fevers and rash, which may occur prior to the onset of clinical arthritis. Along with these classifications, disease subtypes can also be categorized based on age at onset, with early onset arthritis beginning before 6 yr of age and late onset at 6 yr or greater [4]. Furthermore, the course of disease can be described by the number of affected joints beyond

the first 6 months of disease, with four or fewer joints involved in the pauciarticular course and five or more joints involved in the polyarticular course. Pauciarticular patients sometimes develop a polyarticular course, while SOJRA, which may be pauciarticular at onset, typically progresses to a polyarticular course. Unfortunately, predicting the progression of disease in specific patients is problematic.

Although more common in adults, juvenile-onset spondyloarthritides (JSpA), which are similar to but distinct from the rheumatoid arthritides, can begin during childhood and may initially have a phenotype indistinguishable from late-onset pauciarticular JRA, especially in the absence of enthesitis and axial involvement [5]. Individuals who develop axial disease frequently progress to ankylosing spondylitis (AS), as defined by the modified New York criteria [6]. When this occurs during childhood, these patients are referred to as juvenile-onset AS (JAS). The presence of enthesitis is a useful predictor for JSpA, and HLA-B27 is frequently positive. However, many children with JSpA do not go on to develop axial disease, and identification of

William S. Rowe Division of Rheumatology and ¹Division of Pediatric Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, ²Genome Center and Department of Biomedical Informatics, Columbia University, New York, NY 10032 and ³Division of Rheumatology, Children's Hospital of Pittsburgh, Pittsburgh, PA 15213, USA.

Submitted 8 March 2004; revised version accepted 8 April 2004.

Correspondence to: S. D. Thompson, Division of Rheumatology, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA. E-mail: susan.thompson@cchmc.org

those whose disease will progress is not currently possible. The ILAR classification system for JIA includes a category for enthesitis-related arthritis (ERA), which attempts to identify patients with a spondyloarthropathy who do not fulfill criteria for JAS.

These arthropathies are thought to be autoimmune in origin and to represent complex genetic traits, albeit the genes involved and the precise aetiologies are poorly defined. Current subtype categorization is incomplete, as shown by the fact that the response to therapy varies within subtypes. Although JRA and RA have an arthritis phenotype in common, the diseases differ in many clinical features and with respect to age of onset, the presence of autoantibodies and genetic markers, which are largely different between RA and JRA [1, 7–9]. For example, only about 5% of children with JRA are positive for immunoglobulin M rheumatoid factor (RF) compared with over 70% of adult RA patients. While the expression of a variety of genes has previously been shown to be modulated in JRA subpopulations, particularly in synovial fluid, additional differentially expressed genes in peripheral blood mononuclear cells (PBMC) and synovial fluid mononuclear cells (SFMC) are likely to contribute to variability in phenotype, long-term outcome and response to treatment. It is possible that these yet to be identified factors can be used to distinguish previously unrecognized subtypes of patients, which will allow more tailored treatment regimens. The development of microarray technology may enable the identification of these additional factors by facilitating the analysis of global gene expression patterns in small amounts of sample acquired from PBMC and SFMC.

We report observations from a pilot study demonstrating the utility of PBMC- and SFMC-derived gene expression patterns with respect to the childhood autoimmune arthritides, indicating the potential of this technology to differentiate those with the most severe polyarticular course from those with less severe disease or controls. The apparent power of this approach to aid in the clinical evaluation of these arthropathies is supported by both statistical correlation with disease entities and the occurrence of functional relationships within gene groups that are correlated with disease entities. In addition, gene expression analysis is shown to be effective in categorizing patients. Other studies have reported PBMC gene expression differences in RA [10], spondyloarthropathy [11] and systemic lupus erythematosus [12, 13], but this is among the first [14] reports using microarray gene expression analysis to specifically investigate the chronic childhood arthritides.

Materials and methods

Patients and samples

For the purpose of this study, 26 chronic arthritis patients seen at the Cincinnati Children's Hospital Medical Center Rheumatology Clinic were assessed for clinical phenotype by retrospective chart review. The ACR diagnosis and classification criteria are the common system in the USA and in all instances there was sufficient information in the charts to classify patients by course according to these criteria: pauciarticular ($n = 5$; age 12.9 ± 3.9 yr) and polyarticular ($n = 15$; age 15.9 ± 4.9 yr) course or JSpA ($n = 6$; age 21.2 ± 4.9 yr) (Supplementary Table 1, available online only). Of the 15 patients with polyarticular course, three had pauciarticular onset, nine had polyarticular onset and three had systemic onset. Given the retrospective retrieval of clinical data, along with the fact that ILAR criteria are not routinely used in this clinic, our ability to fully classify these patients by the ILAR JIA classification criteria is limited. Information about the presence or absence of RF in seven patients was not available. These patients have been identified in Supplementary Table 1 and were classified as if they were RF-negative with respect to JIA classification.

Following informed consent (approved by Cincinnati Children's Hospital Medical Center Institutional Review Board) and during a

TABLE 1. Classification of 342 genes regulated between polyarticular and controls by ontological groups (corrected for duplicate probes from a single gene)

Ontology class	No. in class	No. from 342 ^a	<i>P</i> ^b
Chemoattractant activity (GO:0042056)	23	6	0.00036
JAK-STAT cascade (GO:0007259)	23	6	0.00036
Regulation of cell adhesion (GO:0030155)	11	4	0.00043
Single-stranded RNA binding (GO:0003727)	13	4	0.00091
Amino acid metabolism (GO:0006520)	99	13	0.0016
Glutamine metabolism (GO:0006541)	9	3	0.0018
Amino biosynthesis (GO:0009309)	39	7	0.0018
Transmembrane receptor protein tyrosine phosphatase activity (GO:0005001)	10	3	0.0026
Ubiquitin conjugating enzyme activity (GO:0004840)	33	6	0.0028
Aromatic compound metabolism (GO:0006725)	55	8	0.0043
Hearing (GO:0007605)	28	5	0.0053
Amino acid biosynthesis (GO:0008652)	28	5	0.0053
Perception of sound (GO:0009592)	29	5	0.0062
Mitochondrial inner membrane (GO:0005743)	92	11	0.0063
Response to heat (GO:0009408)	39	6	0.0069
Protein tyrosine phosphatase activity (GO:0004725)	50	7	0.0078
Proton-transporting two-sector ATPase (GO:0016469)	14	3	0.0083
Aromatic amino acid family metabolism (GO:0009072)	22	4	0.0084
Embryonic development (GO:0009790)	22	4	0.0084
Proteasome endopeptidase activity (GO:0004299)	32	5	0.0098
Prenylated protein tyrosine phosphatase activity (GO:0004727)	32	5	0.0098

^aNumber of unique genes represented by the 378 probe sets.

^bHypergeometric distribution.

scheduled clinical visit, whole blood was collected in EDTA (ethylene diamine tetraacetate). Matched synovial fluid samples were available from 20 patients (10 polyarticular, five pauciarticular and five JSpA) at the same visit. In addition, peripheral blood was collected from 11 healthy controls (age 14.1 ± 4.6 yr). Cells were isolated by Ficoll gradient centrifugation and frozen in 90% fetal calf serum/10% dimethyl sulphoxide ($8-10 \times 10^6$ cells/ml) prior to total RNA isolation with Trizol (Invitrogen Life Technologies, Carlsbad, CA, USA).

Gene expression

For microarray analysis, biotinylated cRNA was synthesized from total RNA (Enzo, Farmingdale, NY, USA). After processing according to the Affymetrix GeneChip Expression Analysis Technical Manual (Affymetrix, Santa Clara, CA, USA), labelled cRNA was quantitated by hybridization to Affymetrix U95Av2 chips. Quality was assessed and expression values derived using Microarray Suite 5.0 (MAS 5.0; Affymetrix). For confirmation of expression patterns, real-time reverse transcription-PCR

(Assays-on-Demand; Applied Biosystems, Foster City, CA, USA), using total RNA prepared separately from that for the microarray analysis, was performed for two genes: CXCL3 (part number: Hs00171061_m1) and CXCL8 (part number Hs00174103_m1). Results were normalized to GAPDH (part number Hs99999905_m1).

Data analysis

Microarray expression values were analysed using algorithms available in GeneSpring 4.2.1 and 6.0 (Silicon Genetics, Redwood City, CA, USA). Chip-to-chip variation in expression intensities was controlled with per-chip normalization using the 50th percentile. Each probe set was then normalized such that the median expression for each gene was 1.0. Expression values were analysed by ANOVA and/or the *t* test with *P* values as indicated in the text. Real time PCR results were analysed with the Mann–Whitney test. Fold change was calculated as the ratio of geometric means in all cases. Ontological classification was performed using the program classScore (<http://microarray.cpmc.columbia.edu/classScore>). Gene designations from National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) were used.

Cross-validation

Sample identity was predicted by ‘hold-one-out’ cross-validation. This algorithm (Genespring version 6) used 50 genes, eight nearest neighbours, and significance was judged based on a *P* value of 0.2. This technique begins by removing one sample from the data set and identifying the 50 genes which best discriminate the two classes, polyarticular or control, considering only the remaining samples. Using these 50 genes, the eight closest samples (measured in Euclidian distance) to the ‘held-out’ sample are identified. Using these nearest neighbours, two *P* values are calculated to determine the likelihood of the closest eight samples being of a specific class, considering the proportion of samples in the data set (calculated for both control and polyarticular samples in our case). The *P* value ratio is the ratio (lowest *P* value)/(other *P* value). The sample is assigned to the class providing a *P* value ratio ≤ 0.2 . The algorithm then starts again by removing a different sample and continues iteratively until all samples have been held out of one analysis and their identity has been predicted. Because each iteration derives a new list of 50 genes, no single gene list is available.

Results

To investigate global gene-expression patterns specific to the course of juvenile arthritis, PBMC- and SFMC-derived RNA from a cohort of patients representing the spectrum of JRA/JSpA disease courses was analysed using U95Av2 Affymetrix GeneChips. The PBMC cohort included 20 JRA and six JSpA patients with long-standing disease (median duration 9.4 yr) and 11 healthy controls (Supplementary Table 1, available at *Rheumatology* online). The patient and control groups had similar age and gender distributions. An average of $42 \pm 5\%$ of the total probe sets were detected per sample with a total of 7670 probe sets (of 12 651 probe sets contained on the U95Av2 microarray) identified as ‘present’ in at least one of the 37 PBMC samples. Further analyses were limited to these 7670 probes. The SFMC cohort included samples from 15 JRA and five JSpA patients obtained at the same clinical visit as the paired PBMC sample.

Genes are expressed differentially between polyarticular JRA and control PBMC

The ability of microarrays to identify differentially expressed genes between juvenile arthritis subtypes and control samples in PBMC

and SFMC was evaluated by ANOVA or *t* test ($P < 0.0001$). Considering disease course subtypes defined by ACR criteria, the majority of differences in gene expression were found between polyarticular and control PBMC samples while few differences were observed for other comparisons using this conservative statistical cut-off (data not shown). The paucity of differences in the other PBMC comparisons may be due to the more extreme nature of polyarticular disease but nonetheless would also be harder to detect because of the smaller numbers of pauciarticular ($n = 5$) and JSpA ($n = 6$) samples. For SFMC, few genes were identified with significant differences in expression between disease groups. This is not surprising because comparisons with control samples are not available, and only small sample sizes were again available for pauciarticular ($n = 5$) and JSpA ($n = 5$) samples.

Further analysis was limited to polyarticular JRA vs control PBMC comparisons, the origin of the majority of the expression differences. A comparison of global gene expression levels (*t* test, $P \leq 0.0001$) between polyarticular ($n = 15$) and control ($n = 11$) PBMC identified 378 probe sets representing 342 unique genes with differential expression (Supplementary Table 2). Compared with the 0.77 probes expected by chance ($7670 \text{ probe sets} \times 0.0001$), the polyarticular vs control comparison represented a false-discovery rate of 0.2%. Inspection of the data for these 378 probe sets showed that 312 had decreased expression in polyarticular patients compared with controls, while 66 had increased expression. It should be noted that many of the genes determined as differentially expressed in this study had only a modest shift in level of expression. For example, the relative level of expression for MVK (mevalonate kinase) is on average 1.18-fold higher in polyarticular than in control, while YWHAQ (tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, theta polypeptide) has a slight decrease (polyarticular/control, 0.78).

In order to maximize the potential for discovery of genes relevant to disease in this exploratory study, we chose to include all 342 genes in our analysis and did not include an additional minimum fold-change cut-off (Fig. 1).

Ontological classification of genes that distinguish polyarticular JRA and control PBMC

One expected result of microarray experiments is a better understanding of the causes of phenotypic differences between disease groups. To begin to address this issue, we performed an ontological analysis of these 342 unique genes. This analysis allowed us to

TABLE 2. Expression of genes with chemokine and JAK/STAT cascade functions

Official name	Alternative name	Pauci ^a	JSpA ^a	Poly ^a	Ontology
AMSH	AMSH	0.27	0.62	0.20 ^c	JAK-STAT
CCL2	MCP-1	1.80	1.63	5.79 ^c	Chemo, JAK-STAT
CXCL1	Gro- α	2.15	5.29 ^b	11.97 ^c	Chemo
CXCL2	Gro- β	1.57	7.97 ^b	37.34 ^c	Chemo
CXCL3	Gro- γ	2.19	6.04 ^b	44.00 ^c	Chemo
CXCL8	IL-8	1.13	3.30 ^b	5.09 ^c	Chemo
CCL7	MCP-3	1.67	1.28	3.21 ^c	Chemo
NMI	NMI	0.72	0.81	0.57 ^c	JAK-STAT
STAT1	STAT1	0.79	0.64 ^b	0.45 ^c	JAK-STAT
STAT2	STAT2	0.75	0.56	0.42 ^c	JAK-STAT
STAT5B	STAT5B	0.71	0.88	0.49 ^c	JAK-STAT
STAT6	STAT6	0.58	0.70	0.40 ^c	JAK-STAT

^aMean (geometric) of expression values for disease course categories relative to control.

^bSignificant with respect to controls (*t* test, $P \leq 0.05$).

^cSignificant with respect to controls (*t* test, $P \leq 0.0001$).

Chemo, chemotactic activity ontology.

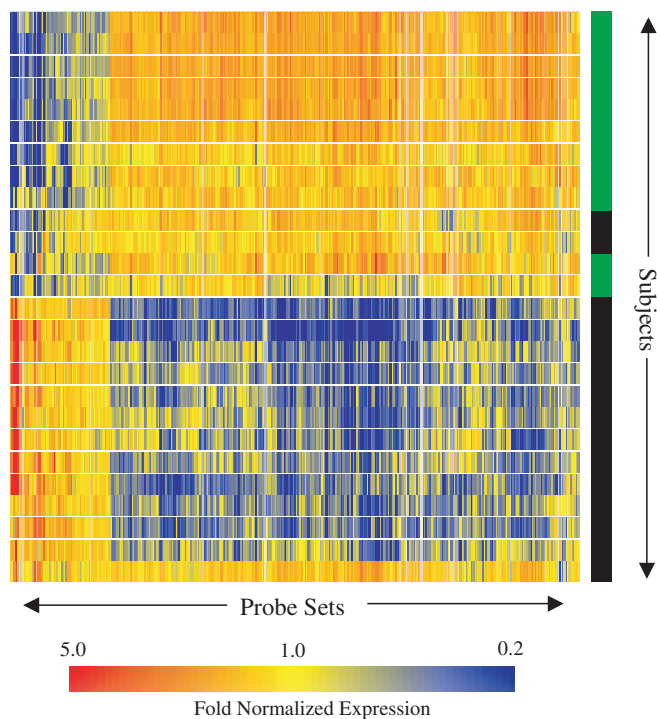


FIG. 1. Three hundred and seventy-eight probe sets identified as differentially expressed between 15 polyarticular and 11 control PBMC samples (Student's *t* test; $P < 0.0001$). Each column represents one probe set and each row represents one person. The colour bar on the right indicates the course of disease for the patients: green, controls; black, polyarticular. Colouring of squares (blue, yellow and red) indicates the relative expression of the probe sets from 0.20- to 5.0-fold change relative to the median of all samples, as indicated by the colour bar at the bottom. For display, samples and genes were ordered by hierarchical clustering.

divide the differentially expressed genes into several groups based on their functional properties as recorded in the gene annotations provided from the gene ontology consortium [15]. In order to ensure statistical relevance and to exclude categories that were overly broad, we considered only those ontologies with greater

than 8 and less than 100 members (Table 1). Statistical analysis (hypergeometric distribution; $P \leq 0.01$) confirmed that 21 of these functional groups were indeed over-represented among the 342 genes activated or repressed in polyarticular samples compared with their relative proportion in the 7670 probe sets on the microarray. These 21 functional categories were relevant to such biological processes as the regulation of immune cell function, receptor signalling and protein metabolism and degradation. Particularly interesting with respect to the pathogenic mechanisms of arthritis were functional categories defined by genes encoding components of the JAK-STAT signalling cascade, and chemokine genes (Table 1).

The genes identified as part of the JAK-STAT cascade in the gene ontology consortium database that were also present in the list of 342 genes include: STAT1, STAT2, STAT5B, CCL2, AMSH and NMI (Table 2). Following the statistical identification of this category, STAT6 was individually identified in the list of 342 genes and was added to this category [16]. These genes are involved in immune cell regulation, including functions related to the Th1/Th2 dichotomy, STAT6 being critical to the Th2 cytokine signalling cascades. In contrast to the chemokines discussed below, the expression of most of these genes was lower in polyarticular samples than controls. Values in the pauciarticular and JSpA samples were intermediate between polyarticular and controls (Fig. 2A, Table 2).

The gene ontology with the best *P* value included genes with chemoattractant activity (Table 1). The prominence of these genes among the most overexpressed genes was striking. In fact, of the 13 genes with levels of expression at least three-fold higher in polyarticular JRA than in controls, six encode chemokines (Supplementary Table 2) and three of the six (CXCL3, CXCL2 and CXCL1) were the most overexpressed (with 44, 37 and 12-fold over-expression respectively) of the total 342 genes. Expression of these genes in the pauciarticular and JSpA samples was intermediate between polyarticular and controls (Fig. 2B, Table 2). This high level of expression and the angiogenic properties of these three chemokines prompted a further investigation of this group.

Expression patterns of angiogenic (ELR^+) and angiostatic (ELR^-) CXCL genes in PBMC and SFMC suggest mechanistic differences between disease subtypes

In addition to overall subtype differences, the ability of this data set to identify mechanistic differences among the disease subtypes in both PBMC and SFMC was tested. An interesting group of

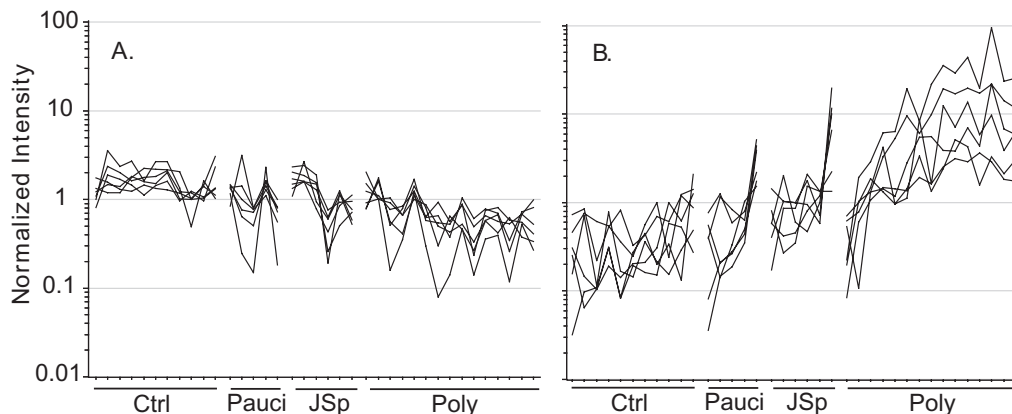


FIG. 2. Expression pattern of specific genes. Ontological analysis of the 342 unique genes identified several functional categories with altered expression, including the JAK-STAT cascade and chemokine activity. Expression levels, as determined by microarrays, normalized to the median of all samples, are presented in the graph. Each line represents one gene. Each vertex on the line represents an individual patient. Samples are first grouped by course then ordered roughly by expression. (A) JAK-STAT ontology modified as follows: removed CCL2, which is present in chemokine activity category; added STAT6, which is not present in the gene ontology consortium category. (B) Chemokine activity ontology.

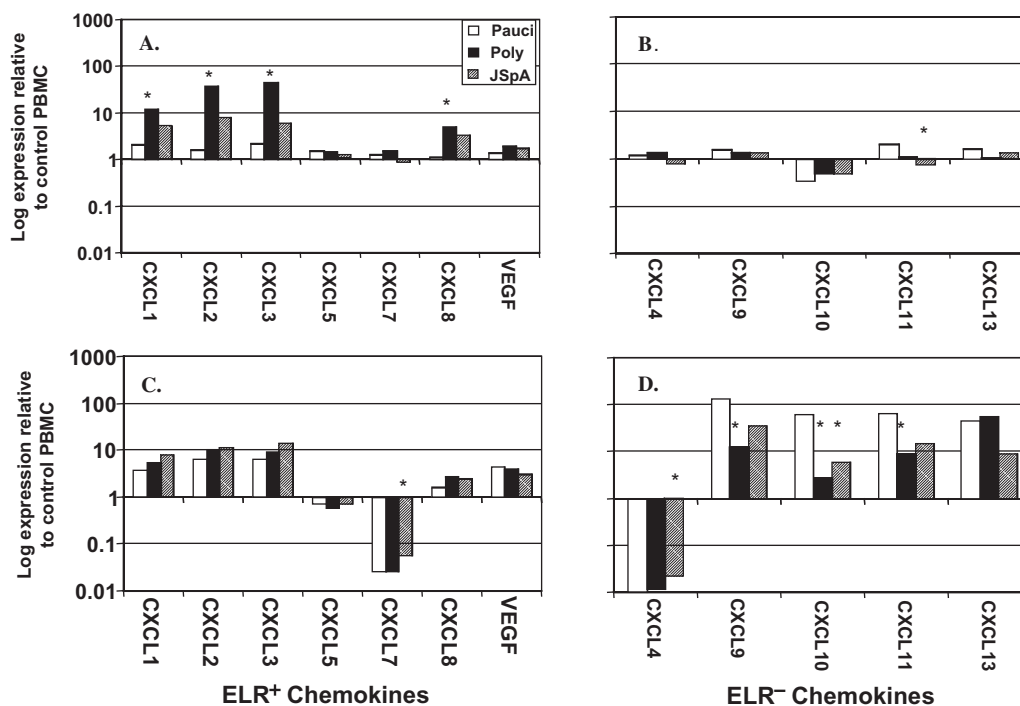


FIG. 3. Expression of angiogenic (ELR⁺) and angiostatic (ELR⁻) genes in PBMC and SFMC. (A) Expression of ELR⁺ genes and VEGF in PBMC. (B) Expression of ELR⁻ genes in PBMC. (C) Expression of ELR⁺ genes and VEGF in SFMC. (D) Expression of ELR⁻ genes in SFMC. * $P < 0.05$ compared with respective pauciarticular samples (Student's t test).

CXCL chemokines was identified as part of the group of genes with chemotactic activity that also had high levels of expression in polyarticular and JSpA patients compared with controls (Fig. 2). In addition to their chemotactic activity, the CXCL chemokines can generally be categorized as angiogenic or angiostatic, based on the presence or absence (respectively) of an ELR (glutamic acid, lysine and arginine) motif. Based on our interest in angiogenesis, we chose to examine this family of genes for expression differences in PBMC and SFMC [17]. Vascular endothelial growth factor (VEGF), a prototypical angiogenic factor, was also included in this analysis. Of the 14 genes (CXCL1–13 and VEGF) related to angiogenesis available on the U95Av2 microarray, 12 were detected in at least one sample (six angiogenic chemokines, five angiostatic chemokines and VEGF). As there were only 12 genes in this analysis, we used the more liberal cut-off of $P < 0.05$ to distinguish genes that were differentially expressed between disease subtypes. Therefore, we were able to use all the disease subtypes, even those with limited sample size.

Considering the angiogenic (six ELR⁺ and VEGF) genes [18], significantly higher expression levels were found for four (CXCL1, CXCL2, CXCL3 and CXCL8) in PBMC from polyarticular compared with pauciarticular patients and healthy controls (Fig. 3A). In contrast, while expression in SFMC was increased compared with control PBMC, expression was generally equivalent between all subtypes (Fig. 3C). Conversely, expression of three angiostatic (ELR⁻) genes [18] (CXCL9, CXCL10 and CXCL11), while increased compared with control PBMC, was lower in polyarticular SFMC compared with pauciarticular SFMC (Fig. 3D). CXCL10 was also expressed at lower levels in JSpA SFMC compared with pauciarticular SFMC. Expression of the ELR⁻ genes was essentially the same among subtypes in PBMC (Fig. 3B).

Real-time PCR was used to confirm the microarray measurements for two angiogenic genes. Expression of CXCL3 and CXCL8 was 55-fold and 26-fold higher respectively in polyarticular than control PBMC (polyarticular $n = 5$; control $n = 4$) (Fig. 4). These results are consistent with the microarray findings,

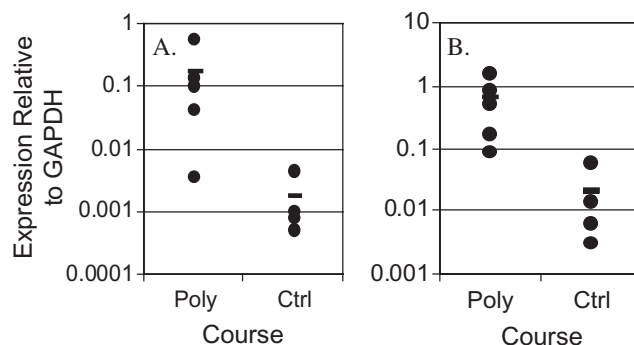


FIG. 4. Real-time PCR confirmation of gene expression. Scatter plots of expression levels in PBMC from five polyarticular patients and four controls as determined by real-time PCR are shown. All values are normalized to expression of GAPDH and are presented on a log scale. Each point represents one patient. Bars represent median values. (A) Expression levels of CXCL3. (B) Expression levels of CXCL8.

which showed increases in expression of 44- and 5-fold for CXCL3 and CXCL8 respectively (Supplementary Table 2).

Classification of samples by expression patterns

In addition to identifying genes with subtype-specific expression patterns, the data set was used to assess the possibility of using microarray gene expression differences to classify samples. Using the cross-validation technique, we found that most samples were predicted accurately (Table 3). All 11 control samples were identified correctly by this method and 13/15 polyarticular samples were identified correctly. Not surprisingly, the two misclassified polyarticular samples, 801 and 824, have P value ratios strongly

TABLE 3. Prediction of sample identities

ID	Actual course	Predicted course	<i>P</i> value ratio ^a	Control votes ^b	Poly votes ^b
1082	Ctrl	Ctrl	4.16E-05	8	0
1083	Ctrl	Ctrl	0.0221	6	2
1085	Ctrl	Ctrl	0.00171	7	1
1087ctrl	Ctrl	Ctrl	0.00171	7	1
1089	Ctrl	Ctrl	0.00171	7	1
1095	Ctrl	Ctrl	0.00171	7	1
7021.31	Ctrl	Ctrl	0.0221	6	2
7113.3	Ctrl	Ctrl	0.00171	7	1
7118.3	Ctrl	Ctrl	0.00171	7	1
7149.3	Ctrl	Ctrl	4.16E-05	8	0
813.31	Ctrl	Ctrl	0.131	5	3
1073	Poly	Poly	0.00278	0	8
1081	Poly	Poly	0.00278	0	8
1087	Poly	Poly	0.00278	0	8
19	Poly	Poly	0.00278	0	8
801	Poly	Ctrl	0.000153	8	0
824	Poly	Ctrl	0.00442	7	1
831	Poly	Poly	0.00278	0	8
850	Poly	Poly	0.00278	0	8
872	Poly	Poly	0.00278	0	8
878	Poly	Poly	0.00278	0	8
894	Poly	Poly	0.00278	0	8
912	Poly	Poly	0.00278	0	8
9137	Poly	Poly	0.00278	0	8
9272	Poly	Poly	0.00278	0	8
993	Poly	Poly	0.00278	0	8

^aWith respect to predicted value.

^bNumber out of eight nearest (Euclidean distance) samples of each category.

indicating membership in the control class of samples. In addition, upon visual examination they appear to have patterns distinct from the other polyarticular samples and similar to the controls (Fig. 1; black boxes in the group indicating controls). Future studies will be performed to evaluate the ability of gene expression patterns to refine the list of probes that differentiate disease subtypes as well as to identify independent samples (i.e. independent validation).

Discussion

Assignment to clinical subtypes based on the pattern of joint involvement (e.g. pauciarticular or polyarticular JRA, or JSpa) has some prognostic value, but does not allow accurate identification of children who will develop severe destructive arthritis and would benefit most from aggressive treatment started early in disease. By focusing on markers in the peripheral blood and synovial fluid late in disease, we have begun to address this issue and to provide a better understanding of the chronic nature of disease pathology. Using a variety of methods, this pilot study has demonstrated the utility of microarray technology to identify biologically relevant, subtype-specific differences in gene expression in PBMC and SFMC from patients with juvenile arthritis. In addition, this study showed that expression of genes in PBMC and SFMC is not always consistent; however, data from the two types of sample are complementary and may be used to better understand subtype-specific mechanisms of disease.

In identifying genes of potential biological interest, the choice of statistical tests and acceptable significance levels must consider the large number of comparisons inherent in microarray data [18]. A *P* value cut-off of 0.0001 was used initially in this study and was expected to return 0.77 probe sets by chance. With our result of 378 regulated probe sets, this corresponds to a 0.2% false-discovery rate. On the other hand, a *P* value of 0.05 is generally considered a cut-off for statistical significance. Using $P < 0.05$, a *t* test of

our microarray data returned 3395 probe sets with differential expression between polyarticular and control samples with 384 expected by chance, providing an unacceptably high 11.3% false-discovery rate. In this exploratory study, all of the genes identified by our statistical cut-off were retained as they may represent genes with large expression changes in minor cell populations or small expression changes in major cell populations. Either possibility may be important for disease and merit further investigation.

The enrichment of genes differentially expressed between JRA patients and controls for those defined as a part of the JAK-STAT signalling cascade is consistent with the biologic role of these molecules. For instance, relatively lower levels of expression of STAT6, a molecule that is critical for the development of Th2 T-cell differentiation, was one of the distinguishing features of polyarticular JRA samples. This is consistent with previous reports suggesting that, while the Th1 immune response is disease-promoting in JRA in general, the presence of Th2 cytokines in some patients may be associated with a better outcome [8].

The prominence of genes encoding chemokines is also probably relevant pathogenically. One of the most important properties of chemokines is the ability to direct trafficking of inflammatory cells and retain the cells at a site of inflammation. Therefore, the differences in the levels of chemokine gene expression may explain differences in the degree of severity of inflammation between patients with poly- and pauciarticular disease. In addition, many chemokines, including those of the CXCL family, influence angiogenesis.

The CXCL chemokines can be subdivided based on the presence or absence of an ELR (glu-leu-arg) motif. This structural difference is functionally important; in addition to influencing the type of cell that a chemokine attracts, the ELR motif also affects the ability to influence vascular remodelling seen in inflammation [18–21]. Thus, ELR⁺ chemokines generally promote neovascularization, which occurs in inflamed tissues, while ELR⁻ chemokines block it (with the exception of CXCL12 [22], which we did not observe in our study). Extensive neovascularization of the inflamed synovium is an important pathogenic mechanism in arthritis. Newly formed blood vessels provide not only the source of nutrients for the growing pannus, but also increased access for inflammatory cells to infiltrate the synovium and thus promote joint damage (reviewed in [23, 24]). The overall outcome of vascular remodelling depends on the balance between the pro- and anti-angiogenic factors present in the tissues, while levels of angiogenic factors have been correlated with arthritic inflammation [17, 25, 26]. Therefore, we hypothesize that the increased expression of angiogenic factors in the PBMC and SFMC of all patients, especially polyarticular, may increase the angiogenic potential in the system. At the same time, expression of angiostatic factors is increased to a greater extent in SFMC samples from pauciarticular patients than either polyarticular or JSpa, thereby potentially limiting angiogenesis in the pauciarticular joint. Taken together, these expression patterns may help shift the overall balance towards relatively higher angiogenic activity in polyarticular patients, consistent with the more aggressive synovial expansion and joint destruction seen in polyarticular patients compared with those with pauciarticular JRA and JSpa.

While a number of microarray studies have focused on sorted cell populations, others have used mixed populations, as was done in this study [12, 13, 27]. Each sample type is useful for answering different questions. For instance, sorted cells can be used to ask how gene expression in activated macrophages is affected by disease state. On the other hand, using the same sorted population would be unlikely to provide information about T cells. Because we cannot define the exact population of cells which is important in arthritis (multiple types are probably important) and because we are primarily interested in identifying genes to use as clinical markers of disease, we chose to use the mixed PBMC and SFMC populations. Our present study has indicated the likelihood of identifying genes useful for the classification of clinical samples in

these sample types, a result that was not entirely expected, given the joint localization of the major pathology.

Two of the polyarticular patients have expression patterns that are consistently different from those of the other polyarticular patients. These expression differences could not be explained by clinical findings such as the number of active joints, duration of disease, medication, age or gender. These expression findings are consistent with distinct mechanistic processes for these patients. Using traditional methods of RNA analysis, which only look at one or a few genes at a time, it would not be possible to appreciate the coordinated expression of this many genes and it is likely that these samples would simply be classified as outliers for the specific gene in question. Limitations in sample quantity would also probably prohibit the analysis of sufficient numbers of genes from the same sample to identify this coordinated regulation. A larger patient population will be necessary to distinguish any previously unrecognized subpopulations and to allow identification of genes that are regulated specifically in these subpopulations.

It is hoped that, separately or in conjunction with genomic screening, we will eventually be able to use gene expression profiling to identify patients at greatest risk of destructive disease and to predict their response to therapy at the earliest time-point possible. Future studies with more samples and purified cell populations will allow us to investigate mechanisms of the disease process, to more rigorously define genes with the power to categorize patients, and to validate any classification schemes with independent cohorts.


<i>Rheumatology</i>	Key messages
	<ul style="list-style-type: none"> • Subtype-specific gene expression profiles are present in PBMC from JRA patients. • The balance of angiogenic and angiostatic CXCL chemokines in polyarticular JRA favours angiogenesis.

Acknowledgements

This work was supported in part by the Children's Hospital Research Foundation of Cincinnati, the Schmidlapp Foundation, and National Institutes of Health Grants AR39979, AR47363, AR47784, AR48372, AR46177 and AR48929.

The authors have declared no conflicts of interest.

Supplementary data

 Supplementary data are available at *Rheumatology* Online.

References

- Petty RE, Cassidy JT. The juvenile idiopathic arthritides. In: Cassidy JT, Petty RE, eds. *Textbook of pediatric rheumatology*, 4th edn. St Louis: W.B. Saunders, 2001:214–7.
- Cassidy JT, Levinson JE, Bass JC *et al.* A study of classification criteria for a diagnosis of juvenile rheumatoid arthritis. *Arthritis Rheum* 1986;29:274–81.
- Petty RE, Southwood TR, Baum J *et al.* Revision of the proposed classification criteria for juvenile idiopathic arthritis: Durban, 1997. *J Rheumatol* 1998;25:1991–4.
- Murray KJ, Moroldo MB, Donnelly P *et al.* Age-specific effects of juvenile rheumatoid arthritis-associated HLA alleles. *Arthritis Rheum* 1999;42:1843–53.
- Petty RE, Cassidy JT. Juvenile ankylosing spondylitis. In: Cassidy JT, Petty RE, eds. *Textbook of pediatric rheumatology*, 4th edn. St Louis: W.B. Saunders, 2001:323.
- van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum* 1984;27:361–8.
- de Kleer IM, Kamphuis SM, Rijkers GT *et al.* The spontaneous remission of juvenile idiopathic arthritis is characterized by CD30+ T cells directed to human heat-shock protein 60 capable of producing the regulatory cytokine interleukin-10. *Arthritis Rheum* 2003;48:2001–10.
- Murray KJ, Grom AA, Thompson SD, Lieuwen D, Passo MH, Glass DN. Contrasting cytokine profiles in the synovium of different forms of juvenile rheumatoid arthritis and juvenile spondyloarthropathy: prominence of interleukin 4 in restricted disease. *J Rheumatol* 1998;25:1388–98.
- Glass DN, Giannini EH. Juvenile rheumatoid arthritis as a complex genetic trait. *Arthritis Rheum* 1999;42:2261–8.
- Maas K, Chan S, Parker J *et al.* Cutting edge: molecular portrait of human autoimmune disease. *J Immunol* 2002;169:5–9.
- Gu J, Marker-Hermann E, Baeten D *et al.* A 588-gene microarray analysis of the peripheral blood mononuclear cells of spondyloarthropathy patients. *Rheumatology* 2002;41:759–66.
- Bennett L, Palucka AK, Arce E *et al.* Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med* 2003;197:711–23.
- Baechler EC, Batliwalla FM, Karypis G *et al.* Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci USA* 2003;100:2610–5.
- Jarvis J, Dozmorov I, Jiang K *et al.* Novel approaches to gene expression analysis of active polyarticular juvenile rheumatoid arthritis. *Arthritis Res Ther* 2003;6:R15–R32.
- Ashburner M, Ball CA, Blake JA *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25–9.
- Leonard WJ, O'Shea JJ. Jaks and STATs: biological implications. *Annu Rev Immunol* 1998;16:293–322.
- Scola MP, Imagawa T, Boivin GP *et al.* Expression of angiogenic factors in juvenile rheumatoid arthritis: correlation with revascularization of human synovium engrafted into SCID mice. *Arthritis Rheum* 2001;44:794–801.
- Arenberg DA, Polverini PJ, Kunkel SL, Shanafelt A, Strieter RM. *In vitro* and *in vivo* systems to assess role of C-X-C chemokines in regulation of angiogenesis. *Methods Enzymol* 1997;288:190–220.
- Strieter RM, Polverini PJ, Kunkel SL *et al.* The functional role of the ELR motif in CXC chemokine-mediated angiogenesis. *J Biol Chem* 1995;270:27348–57.
- Keane MP, Arenberg DA, Moore BB, Addison CL, Strieter RM. CXC chemokines and angiogenesis/angiostasis. *Proc Assoc Am Physicians* 1998;110:288–96.
- Strieter RM, Polverini PJ, Arenberg DA, Kunkel SL. The role of CXC chemokines as regulators of angiogenesis. *Shock* 1995;4:155–60.
- Pablos JL, Santiago B, Galindo M *et al.* Synovial cell-derived CXCL12 is displayed on endothelium and induces angiogenesis in rheumatoid arthritis. *J Immunol* 2003;170:2147–52.
- Koch AE. Review: angiogenesis: implications for rheumatoid arthritis. *Arthritis Rheum* 1998;41:951–62.
- Firestein GS. Starving the synovium: angiogenesis and inflammation in rheumatoid arthritis. *J Clin Invest* 1999;103:3–4.
- Maeno N, Takei S, Imanaka H *et al.* Increased circulating vascular endothelial growth factor is correlated with disease activity in polyarticular juvenile rheumatoid arthritis. *J Rheumatol* 1999;26:2244–8.
- Ballara S, Taylor PC, Reusch P *et al.* Raised serum vascular endothelial growth factor levels are associated with destructive change in inflammatory arthritis. *Arthritis Rheum* 2001;44:2055–64.
- Ye S, Pang H, Gu YY *et al.* Protein interaction for an interferon-inducible systemic lupus associated gene, IFIT1. *Rheumatology* 2003;42:1155–63.