EXTENDED REPORT

A CD4 T cell gene signature for early rheumatoid arthritis implicates interleukin 6-mediated STAT3 signalling, particularly in anti-citrullinated peptide antibody-negative disease

Arthur G Pratt,1,2 Daniel C Swan,3 Sarah Richardson,1 Gillian Wilson,2 Catharien M U Hilkens,1 David A Young,1 John D Isaacs1,2

ABSTRACT

Objective We sought clinically relevant predictive biomarkers present in CD4 T-cells, or in serum, that identified those patients with undifferentiated arthritis (UA) who subsequently develop rheumatoid arthritis (RA).

Methods Total RNA was isolated from highly purified peripheral blood CD4 T cells of 173 early arthritis clinic patients. Paired serum samples were also stored. Microarray analysis of RNA samples was performed and differential transcript expression among 111 ‘training cohort’ patients confirmed using real-time quantitative PCR. Machine learning approaches tested the utility of a classification model among an independent validation cohort presenting with UA (62 patients). Cytokine measurements were performed using a highly sensitive electrochemiluminescence detection system.

Results A 12-gene transcriptional ‘signature’ identified RA patients in the training cohort and predicted the subsequent development of RA among UA patients in the validation cohort (sensitivity 68%, specificity 70%). STAT3-inducible genes were over-represented in the signature, particularly in anti-citrullinated peptide antibody-negative disease, providing a risk metric of similar predictive value to the Leiden score in seronegative UA (sensitivity 85%, specificity 75%). Baseline levels of serum interleukin 6 (IL-6) (which signals via STAT3) were highest in anti-citrullinated peptide antibodies-negative RA and distinguished this subgroup from non-RA inflammatory synovitis (corrected p<0.05). Paired serum IL-6 measurements correlated strongly with STAT3-inducible gene expression.

Conclusion The authors have identified IL-6-mediated STAT3-signalling in CD4 T cells during the earliest clinical phase of RA, which is most prominent in seronegative disease. While highlighting potential biomarker(s) for early RA, the role of this pathway in disease pathogenesis awaits clarification.

The importance of prompt disease-modifying therapy in early rheumatoid arthritis (RA) is now established.1,2 Yet about 40% of patients with new-onset inflammatory arthritis present with disease that is unclassifiable at inception, having a so-called undifferentiated arthritis (UA).3 Timely intervention for the subset of these UA patients who subsequently develop RA therefore remains problematic. The issue is highlighted by the publication of updated RA classification criteria4 and a validated ‘prediction rule’ that foretells risk of UA progression to RA.5 Such approaches rely heavily on autoantibody status, emphasising the specificity of circualting anti-citrullinated peptide antibodies (ACPA) for RA.6 Consequently, the diagnosis of ACPA-negative RA remains challenging in the early arthritis clinic (EAC), being frequently delayed despite application of the prediction rule.7 The potential for the whole-genome transcription profiling to yield clinically relevant prognostic ‘gene signatures’ in autoimmune disease has been demonstrated.8,9 Applying a similar, prospective approach to the discovery of predictive biomarkers in UA should complement existing diagnostic algorithms, while providing new insights into disease pathogenesis.10 However, the use of peripheral blood (PB) mononuclear cells for transcriptional analysis may result in data that are biased by relative subset abundance.11 To address this, protocols for rapid ex vivo positive selection of cell subsets for the purpose of transcription profiling have been validated.12 Although no single cell-type is exclusively implicated in RA, many of its established and emerging genetic associations implicate the CD4 T cell.13 We therefore hypothesised that the PB CD4 T-cell transcriptome would provide a useful substrate for both biomarker discovery and a pathophysiological understanding of RA induction.

MATERIALS AND METHODS

A complete description of experimental and bioinformatics approaches are given in the online supplementary text.

Patients

Patients with recent onset arthritis, naïve to disease-modifying anti-rheumatic drugs and corticosteroids, were recruited between September 2006 and December 2008. An initial working clinical diagnosis was updated by the consulting rheumatologist at consecutive clinic visits for the duration of the study—median 28 months and >12 months in all cases. RA was diagnosed only where 1987 American College of Rheumatology classification criteria14 were fulfilled; UA was defined as a ‘suspected inflammatory arthritis where RA remained a possibility, but where established classification criteria for any rheumatological condition remained
unmet’ (see online supplementary text and supplementary table S1). Individuals whose arthritis remained undifferentiated at the end of the study were excluded. Patients gave written informed consent before inclusion into the study, which was approved by the local regional ethics committee.

CD4 T-cell RNA processing and array analysis
Whole blood drawn between 13:00 and 16:30 was stored at room temperature for ≤4 h before processing. After monocyte depletion by immuno-rosetting, an automated magnetic bead-based positive selection protocol was used to isolate CD4 cells (Stemcell Technologies, Vancouver, Canada). Using this approach, a median CD4 T-cell purity of 98.9% was achieved (range 95–99.7%), which was determined using flow cytometry (see online supplementary figure S1). Total CD4 T-cell RNA was immediately extracted, then quality controlled using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, California, USA). Raw data were normalised and expressed relative to the housekeeping gene PPIA (number; per cent) 29 (62) – 18 (29) <0.001 30 (18–60)

CRP (g/l; median, IQR) 5.75 (0.75–2) 1 (0–3) <0.001 1 (0–3)

Morning stiffness (hours; median, IQR) 1 (0.75–2) 0.007 1 (0.5–2)

ESR (c; median, IQR) 56 (30–78) 24 (14–52) <0.001 30 (18–60)

ACPA+ (number; per cent) 29 (62) – 18 (29) <0.001 30 (18–60)

RA 47 (100) 0 (0) – 25 (40)

Seronegative spondyloarthropathy – 22 (34) – 8 (13)

Self-limiting inflammatory – 12 (19) – 9 (15)

Other inflammatory – 3 (5) – 2 (3)

OA/non-inflammatory – 27 (42) – 18 (29)

Outcome diagnoses (number, per cent)

Training cohort

RA (n=47) Non-RA (n=64) p*

Age (years; mean, SD range) 60 (46–74) 48 (34–62) 0.01 52 (37–67)

% Female 65 61 NS 77

% White Caucasian 96 92 NS 90

Symptom duration (weeks; median, IQR) 12 (8–24) 21 (10.5–52) 0.026 14 (12–26)

Tender joint count (median, IQR) 10 (4–15) 7 (2–14) NS 8 (3–16.5)

Swollen joint count (median, IQR) 6 (2–10) 0 (0–2) <0.001 1 (0–3)

Morning stiffness (hours; median, IQR) 1 (0.75–2) 0.75 (0.1–2) 0.007 1 (0.5–2)

ESR (c; median, IQR) 56 (30–78) 24 (14–52) <0.001 30 (18–60)

CRP (g/l; median, IQR) 17 (9–62) 5 (2.5–19) <0.001 8.5 (0–17)

ACPA+ (number; per cent) 29 (62) 0 (0) <0.001 13 (21)

SF (number; per cent) 36 (77) 3 (6) <0.001 20 (32)

DAS28 (median) 5.37 NA –

Leiden prediction score (median, IQR) NA NA – 6.4 (5–7.6)

Outcome diagnoses (number, per cent)

RA 47 (100) 0 (0) – 25 (40)

Seronegative spondyloarthropathy – 22 (34) – 8 (13)

Self-limiting inflammatory – 12 (19) – 9 (15)

Other inflammatory – 3 (5) – 2 (3)

OA/non-inflammatory – 27 (42) – 18 (29)

Outcome diagnoses (number, per cent)

Values are mean (1 SD range), median (IQR) or % for normally distributed, skewed or dichotomous data, respectively.

* Statistical tests for significant difference between RA and non-RA groups: t-test, Mann–Whitney U or Fisher’s exact test for normally distributed, skewed or dichotomous data, respectively.

† Demographic, clinical and serological parameters are given for UA–RA and UA–non-RA subgroups in online supplementary table S2.

ACPA, anti-citrullinated peptide antibodies; CRP, C reactive protein; DAS28, disease activity score (incorporating 28-swollen/tender joint counts); ESR, erythrocyte sedimentation rate; NS, not significant; OA, osteoarthritis; RA, rheumatoid arthritis; RF, rheumatoid factor; UA, undifferentiated arthritis.

Serum cytokine measurement
Between 13:00 and 16:30, baseline serum was drawn and frozen at ~80°C until use. Serum interleukin 6 (IL-6), soluble IL-6 receptor (sIL6R), tumour necrosis factor α (TNFα), leptin and granulocyte colony stimulating factor concentrations were measured using a highly sensitive electrochemiluminescence immunosorbance detection system (Meso Scale Discovery, Gaithersberg, Maryland, USA), assays having been validated as outlined (online supplementary text and figure S3).

Quantitative real-time PCR
CD4 T cell total RNA samples were reverse transcribed using superscript II reverse transcriptase and random hexamers, according to the manufacturer’s instructions (Invitrogen, Carsbad, California, USA). Real-time PCR reactions were performed as part of a custom-made TaqMan Low Density Array (7900HT real-time PCR system, Applied Biosystems, Foster City, California, USA). Raw data were normalised and expressed relative to the housekeeping gene β-actin as 2−ΔCt values.

General statistics
Parametric and non-parametric analyses of variance, Mann–Whitney U tests, Pearson’s correlation coefficients, intra-class correlations, multivariate analyses and the construction of receiver operator characteristic (ROC) curves were performed, as described, using SPSS version.15.0 (SPSS, Chicago, Illinois, USA). The derivation of Leiden prediction rules and

Table 1  Clinical characteristics of the rheumatoid arthritis (RA) and non-RA comparator groups used to generate list of differentially expressed genes, which together comprise a training cohort for machine-learning (total n=111) and the independent undifferentiated arthritis (UA) validation cohort (n=62)
transcriptional ‘risk metrics’ for ACPA-negative RA is outlined in the online supplementary text.

**RESULTS**

**Patient groups**

A total of 173 patient samples were retrospectively selected for microarray analysis. One hundred and eleven of these originated from patients assigned definitive diagnoses at inception, confirmed at a median 28 months follow-up (minimum 1 year); an RA versus non-RA discriminatory ‘signature’ was derived from this ‘training cohort’ alone. The remaining 62 samples, all representing UA patients, formed an independent ‘validation cohort’ for testing the utility of the ‘signature’ according to diagnostic outcome at the end of the study (see online supplementary text and table S5), and applying it to the independent ACPA-negative UA cohort, we could improve the utility of the prediction rule for seronegative UA patients (AU ROC=0.84; SEM=0.06; figure 1D).

**Gene signature adds value to existing tools in diagnosing ACPA-negative UA**

Next, we tested the value of our 12-gene signature in comparison with the existing ‘Leiden prediction rule’ as a predictor of RA among UA patients.5 While the discriminatory utility achieved by the prediction rule in our UA cohort was comparable with that previously reported (n=62; AU ROC curve=0.86; SEM=0.05, data not shown), its performance diminished among the ACPA-negative sub-cohort (n=49; AU ROC curve=0.74; SEM=0.08; figure 1D). Employing a 12-gene risk metric, as described above, equivalent discriminatory utility was found in this sub-cohort (AU ROC curve=0.78; SEM=0.08, data not shown). However, by deriving a modified risk metric, which combined all features of the Leiden prediction rule with our 12-gene risk metric (see online supplementary text and table S5), and applying it to the independent ACPA-negative UA cohort, we could improve the utility of the prediction rule for seronegative UA patients (AU ROC=0.84; SEM=0.06; figure 1D).

**STAT3 transcription profile is most prominent in ACPA-negative RA**

All 173 patients studied were now grouped into four categories based on outcome diagnosis alone: ACPA-positive RA, ACPA-negative RA, inflammatory non-RA controls and osteoarthritis(OA); their demographic and clinical characteristics are presented for comparison (online supplementary table S4).
with independently predicted STAT3-inducible gene sets (see online supplementary text and supplementary gene list 9) confirmed a preponderance of STAT3-inducible genes (hypergeometric p-values <0.005 in both cases; see online supplementary text) – which was not seen for genes deregulated only in ACPA-positive RA (p=0.19).

**Serum IL-6 is highest in ACPA-negative RA and independently predicts CD4 STST3-inducible gene expression**

Since one classical mechanism of STAT3 phosphorylation is via gp130 co-receptor ligation,29 we hypothesised that increased systemic levels of a key gp130 ligand and pro-inflammatory cytokine, IL-6, may be responsible for the STAT3-mediated transcriptional programme in early RA patients. Baseline serum IL-6 was measured in 131 of the 173 EAC patients which were subsequently grouped according to their ultimate diagnosis (ACPA-negative RA, ACPA-positive RA, non-RA inflammatory arthropathy or OA). IL-6 levels were low overall (generally <100 pg/ml), but were highest in the ACPA-negative RA group (figure 3C). Indeed, unlike the generic marker of systemic inflammation C reactive protein (CRP), baseline IL-6 discriminated ACPA-negative RA from non-RA inflammatory arthritides (figures 3C,D). Furthermore, among individuals for whom paired and contemporaneous serum IL-6 and PB CD4 T-cell RNA samples were available, clear correlations between IL-6 and the normalised expression of STAT3-inducible genes were seen (figures 4A–D; also online supplementary figures S5A–D); for example, serum IL-6 measurements correlated with normalised SOCS3 expression: Pearson’s R=0.57, p<0.001 (figure 4A). Multivariate analysis confirmed that IL-6, but not CRP or TNFα (which does not signal via STAT3), independently predicted PB CD4 T cell SOCS3 expression (β=0.53; p<0.001; see online supplementary table S6) excluding a more general influence of inflammation.
Functional analysis of array data. Genes differentially expressed (>1.2-fold change; p<0.05) between osteoarthritis (OA) and three separate inflammatory comparator groups were overlapped in a Venn diagram (see text, and online supplementary gene lists 2–4, for detailed list compositions). Genes uniquely deregulated in rheumatoid arthritis (RA) (anti-citrullinated peptide antibodies (ACPA)-negative, ACPA-positive or both) could thereby be identified and subjected to pathway analysis (see text). The top two over-represented biological functions identified for the three indicated sets are shown, along with the proportion of the set associated with the function in question, and a p value relating to the likelihood of given proportions occurring by chance (Fisher’s exact test). Online supplementary gene lists 5–7 summarise functionally related genes thereby identified. The three indicated sets were combined to identify canonical pathways over-represented among genes differentially expressed between RA and OA in general. Pathways of particular interest in the biological context are listed (genes in question are listed in online supplementary gene list 8), *hypergeometric p-values (Fisher's exact) in each case <0.01.

Given that only 30–50% of PB CD4 T cells are thought to express membrane-bound IL6R, we also measured sIL6R (as a surrogate of IL-6R trans-signalling) and two other gp130 ligands, granulocyte colony stimulating factor and leptin, both of which have been implicated in RA pathogenesis. However, levels in sera from a subset of 80 study patients correlated with neither the diagnostic outcome nor the STAT3 gene expression. Finally, IL-10 and IL-17, which are both STAT3 activators, were undetectable in the vast majority of sera (data not shown).

**STAT3-inducible, RA-associated expression signature is activated by IL-6 in primary CD4 T cells of healthy donors in vitro**

To confirm that the observed deregulated expression of STAT3 target genes among early RA patients was downstream of IL-6 signalling, primary human CD4 T cells were incubated in vitro with recombinant human IL-6 and the expression of relevant target genes measured at 1 and 6 h (see online supplementary text and figures S6–S7). Robust upregulation of SOCS3, PIM1, BCL3 and MYC was observed consistently 1 h after the addition of IL-6. A similar trend was seen for SBNO2, which became significant in the presence of recombinant soluble human IL-6 receptor. Conversely and consistent with prior observations, a distinct trend towards repression of ID3 was seen in response to IL-6 plus sIL6R.

The signature’s sensitivity and specificity (0.85 and 0.75) for predicting subsequent RA in seronegative UA patients equate to a positive likelihood ratio of 3.4, indicating that a prior probability of 25% for RA progression among this cohort (13 of the 49 patients progressed to RA) doubles to 53% for an individual who has been assigned a positive SVM classification. Moreover, of the 13 ACPA-negative UA patients who progressed to RA in our cohort, 8 fell into an ‘intermediate’ risk category for RA progression according to the validated Leiden prediction score. Encouragingly, all but one of these patients were correctly classified based on their 12-gene expression profile. Our proposal that this approach might add value to existing algorithms for the diagnosis of ACPA-negative UA is further supported by the construction of ROC curves comparing the Leiden prediction rule with a modified risk metric that incorporates features of our gene signature (figure 1D).

Our data indicate that PB CD4 T cells in early RA are characterised by a predominant upregulation of biological pathways involved in cell cycle progression (ACPA-positive) and survival, death and apoptosis (ACPA-negative) (figure 2; also online supplementary gene lists 5–6). Pathway analysis also suggested that T-cell development and differentiation were deregulated in both RA serotypes (online supplementary gene list 7). These findings concur with previous observations of impaired T-cell homeostasis in RA, characterised by increased turnover, telomere shortening and immunosenescence. Given the well-characterised importance of the STAT3 signalling pathway in both oncogenesis and T-cell survival, it was notable that five genes from our statistically robust 12-gene RA signature are downstream of STAT3 signalling. The degree to which these genes sub-cluster according to the expression pattern among individuals in both the training and validation cohorts (figure 1A,C) presumably reflects their co-regulation by STAT3. Their upregulation was generally most pronounced in ACPA-negative RA (figure 3A,B; also online supplementary figure S4A–C), explaining why the predictive utility of the 12-gene signature was optimal in this disease subset.
Our observation that increased serum IL-6 levels among EAC attendees may predict a diagnosis of RA versus alternative arthritides is consistent with findings of previous biomarker studies, but ours is the first demonstration of a particular association with ACPA-negative disease (figure 3C). Striking correlations were seen between PB CD4 T-cell expression of several STAT3-inducible genes and paired, contemporaneous serum IL-6 concentrations, which were independent of alternative acute phase markers (figures 4A–D; also online supplementary figures S5A–D and table S6). STAT3 phosphorylation and downstream transcription is initiated by ligation of the cell-surface gp130 co-receptor by a range of ligands, including IL-6. We measured IL-6 in particular because of its recognised role as a pro-inflammatory cytokine in RA, and we excluded similar relationships with sIL6R (a surrogate of IL-6R trans-signalling) and other relevant substrates of STAT3 signalling. Therefore, the STAT3-inducible gene expression signature that we have identified does appear to be downstream of IL-6 signalling. The capacity of IL-6 alone to induce the STAT3-regulated elements of our early RA gene expression signature in primary CD4 T cells was confirmed in vitro (online supplementary figures S6 and S7).

In conclusion, our data provide strong evidence for the induction of an IL-6-mediated STAT3 transcription programme in PB CD4 T cells of early RA patients, which is most prominent in ACPA-negative individuals and which contributes to a gene expression ‘signature’ that may have diagnostic utility. Furthermore, our findings could pave the way for a novel treatment paradigm, whereby emerging drugs targeting the IL-6-gp130-STAT3 ‘axis’ find a rational niche as first choice agents in the management of ACPA-negative RA. Studies, such as ours, should ultimately contribute to the realisation of true ‘personalised medicine’ in early inflammatory arthritis, in which complex heterogeneity is stratified into pathophysiologically and therapeutically relevant subsets, with clear benefits in terms of clinical outcome and cost.

Contributors AGP was involved in the study’s conception and design, sample and data collection; he conducted the majority of the laboratory work, performed a substantial part of the analysis and drafted the manuscript. DCS helped design the study and made a major contribution to data analysis, also contributing intellectually to revising the manuscript. SR made substantial contributions to the laboratory work, data analysis and manuscript revision. GW was involved in the study design and made a major contribution to the recruitment of patients, sample handling and the recording of clinical data, also critically appraising the manuscript draft. CH helped conceive and design the study, also contributing to the laboratory work and manuscript drafting. DAY helped conceive and design the study and performed some of the laboratory work; he also made a substantial contribution to data analysis, and contributed to manuscript drafting. JDI conceived and designed the study and contributed to data analysis; he provided intellectual input and supervision throughout the study and made a substantial contribution to manuscript drafting.

Acknowledgements AGP’s work was supported by a clinical research fellowship from the Arthritis Research Campaign, UK. This work was supported by the UK NIHR Biomedical Research Centre for Ageing and Age-Related Disease Award to the Newcastle upon Tyne Hospitals NHS Foundation Trust. Clinical and translational research in the Musculoskeletal Research Group is supported by the Northumberland, Tyne and Wear Comprehensive Local Research Network. The authors would like to thank the clinical staff at The Freeman Hospital who co-operated with the recruitment phase of the study, and, of course, the many patient volunteers who contributed so willingly.

Funding This study was supported by Arthritis Research UK (grant number 17983).

Competing interests None.

Ethics approval Ethics approval was provided by the Newcastle and North Tyneside Local Research Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.
Figure 4  (A–D) Serum interleukin 6 (IL-6) concentrations correlate with STAT3-inducible gene expression in peripheral blood CD4 T cells (see online supplementary figure S5, for additional examples). Data are shown for 131 individuals in whom paired, contemporaneous samples were available at baseline; Pearson’s r and associated p values are shown.

REFERENCES


