



# Multiplex immune serum biomarker profiling in sarcoidosis and systemic sclerosis

P. Beirne<sup>\*,#,#,#</sup>, P. Pantelidis<sup>\*,#,#,#</sup>, P. Charles<sup>†,+,#,#</sup>, A.U. Wells<sup>\*</sup>, D.J. Abraham<sup>§</sup>, C.P. Denton<sup>§</sup>, K.I. Welsh<sup>\*</sup>, P.L. Shah<sup>\*</sup>, R.M. du Bois<sup>f</sup> and P. Kelleher<sup>\*,†,\*\*,\*</sup>

**ABSTRACT:** Multiplex protein technology has the potential to identify biomarkers for the differentiation, classification and improved understanding of the pathogenesis of interstitial lung disease. The aim of this study was to determine whether a 30-inflammatory biomarker panel could discriminate between healthy controls, sarcoidosis and systemic sclerosis (SSc) patients independently of other clinical indicators. We also evaluated whether a panel of biomarkers could differentiate between the presence or absence of lung fibrosis in SSc patients.

We measured 30 circulating biomarkers in 20 SSc patients, 21 sarcoidosis patients and 20 healthy controls using Luminex bead technology and used Fisher's discriminant function analysis to establish the groups of classification mediators.

There were significant differences in median concentration measurements between study groups for 20 of the mediators but with considerable range overlap between the groups, limiting group differentiation by single analyte measurements. However, a 17-analyte biomarker model correctly classified 90% of study individuals to their respective group and another 14-biomarker panel correctly identified the presence of lung fibrosis in SSc patients.

These findings, if they are corroborated by independent studies in other centres, have potential for clinical application and may generate novel insights into the modulation of immune profiles during disease evolution.

**KEYWORDS:** Chemokines, cytokines, fibrosis, growth factors, sarcoidosis, scleroderma

Sarcoidosis, a multisystem granulomatous disease of unknown origin, which characteristically affects the lungs, is associated with upregulation of the monocyte/macrophage-T cell helper (Th) type 1 cell axis and polyclonal B cell activation [1]. Systemic sclerosis (SSc) is a systemic autoimmune disease characterised by tissue fibrosis (skin and lung in particular), endothelial cell injury and persistent cellular immune activation (Th1, Th2 and B cells) [2–4]. Currently, SSc is commonly classified according to the degree of skin involvement and the presence of SSc-related autoantibodies that correlate well with the clinical phenotypes. Several studies have shown that serum levels of a number of cytokine and chemokine are either up- or downregulated in patients with sarcoidosis and systemic sclerosis. Some of these have the potential to serve as markers of disease development since they also correlate with disease status and organ involvement [5, 6]. Pulmonary fibrosis, an important complication of both conditions, responds poorly to current therapies. Therefore, there is considerable interest in identifying biomarkers that may support early therapeutic intervention and delineate a favourable response [7].

The use of analytical platforms capable of simultaneous analysis of large numbers of proteins (protein microarrays) can improve our understanding of the mechanistic basis of immune diseases and can also identify useful biomarkers for clinical diagnosis, subclassification of disease or response to therapy [8]. The use of protein microarrays in immunology has largely concentrated on profiling secreted proteins (cytokines, chemokines or growth factors) or autoantibodies [9]. Multiplex analysis of either secreted proteins or autoantibodies has been reported for a number of inflammatory diseases; however, there have been very few reports of the use of this technology in patients with respiratory disease in general and interstitial lung disease in particular [8, 10]. Serum protein microarrays have enormous potential for risk stratification and as surrogate markers of response to treatment, given the inaccessibility of the lung to regular sampling. However, changes in the levels of such inflammatory biomarkers are not unique to individual clinical conditions and tend to overlap considerably between conditions. In addition, usually few mediators are concomitantly studied in such investigations and subsequently only mediators with demonstrable significant

## AFFILIATIONS

\*Dept of Respiratory Medicine, Royal Brompton & Harefield NHS Trust, #Interstitial Lung Disease Unit, Dept of Population Genetics and Gene Therapy, National Heart and Lung Institute, Imperial College School of Medicine,

†Division of Immunology, Imperial College NHS Trust,

‡Kennedy Institute of Rheumatology, Imperial College,

§Division of Medicine, Research Dept of Inflammation, Centre for Rheumatology and Connective Tissue Diseases, Royal Free and University College Medical School, University College London, and,

\*\*Dept of Immunology, Imperial College, Chelsea & Westminster Hospital Campus, London, UK.

<sup>f</sup>National Jewish Health, Denver, CO, USA.

#,# These authors contributed equally to this work.

## CORRESPONDENCE

P. Kelleher  
Dept of Respiratory Medicine  
Royal Brompton & Harefield NHS Trust  
London  
UK  
E-mail: p.kelleher@imperial.ac.uk

Received:

Feb 17 2009

Accepted after revision:

May 28 2009

First published online:

June 18 2009

European Respiratory Journal  
Print ISSN 0903-1936  
Online ISSN 1399-3003

median changes are used to establish clinical differentiation by applying a threshold cut-off method outside the limits of normality. This tends to produce strict cut-off criteria that can only be fulfilled by a small subset of subjects. However, compared with individual measurements, it is possible that greater specificity and inclusivity may be achieved by a combination of mediators.

In this study, we describe the use of a protein microarray panel to discriminate between healthy controls and patients with two disparate immunologically driven conditions: pulmonary sarcoidosis or systemic sclerosis (conditions where immune activation may precede the onset of pulmonary fibrosis). The main aim of this study was to investigate inflammatory profiles using multibiomarkers consisting of cytokines, chemokines and growth factors and assess whether individual mediators or panels of mediators could classify study participants into their three distinct groups (healthy controls, sarcoidosis and systemic sclerosis) on the basis of biological measurements alone. Secondly, we tried to explore whether a multibiomarker panel could differentiate between the presence and absence of lung fibrosis in SSc. Thirdly, we assessed the complexity of correlations between mediators in health and disease. We found a multibiomarker panel of 17 analytes that could classify the subjects to their respective clinical group. We suggest that independent validation of these findings and further development of protein microarrays may provide new insights into the pathogenesis of fibrotic lung diseases and also the development of clinically informative biomarker panels.

## MATERIALS AND METHODS

### Patient recruitment

Serum was obtained from sarcoidosis and SSc patients. All patients satisfied the accepted international diagnostic criteria for both conditions [11, 12]. Demographic information is given in table 1. Given the epidemiology of both diseases it was not possible to match patient groups for age or sex. Patients were excluded if they were taking corticosteroid therapy >10 mg of prednisolone per day or equivalent, or if they were on any of the powerful second-line immunosuppressants, such as cyclophosphamide, azathioprine or methotrexate. Biomarker results were later analysed for any confounding effects of sex, age or treatment. Serum from healthy controls was obtained from an anonymised research serum bank held within the Kennedy Institute, Charing Cross Hospital, Imperial College Healthcare NHS Trust (London, England) and a proportion of these were collected during this study. Sample handling and storage were identical for historical and recent controls and patient groups. Ethical permission for this study was obtained from the Brompton, Harefield and NHLI Ethics Committee (REC No 02-002 and 01-257).

### Experimental method

30 proteins were measured in serum using a fluorescent bead-based technology (Luminex® Corporation, TX, USA). The assays were carried out in compliance with the kit manufacturer. In brief, antibody-conjugated beads were prepared and aliquoted into a 96-well pre-wetted filter plate, before addition of either 100 µL of standards solution in the designated wells or 100 µL of serum, diluted 1:2 with assay diluent. Subsequent steps involved washes interspersed by the addition of

biotinylated detector antibody and later Streptavidin-RPE solution. Finally, the plate was placed on the XY platform of the Luminex 100 instrument for analysis. STarStation software was used for data acquisition and analysis. Standard curves were generated for each analyte, and the mean fluorescence intensity value of each analyte in each well was converted into a concentration using the linear portion of the standard for all detected values. This value was then multiplied by the dilution coefficient 2 to give the concentration of the analyte in the original serum.

The fluorescent beads (FIDIS™) for this study were obtained from Biomedical Diagnostics (BMD, Croissy Beaubourg, France). Biomarkers assayed were classified according to their

**TABLE 1** Clinical characteristics of the sarcoidosis and systemic sclerosis (SSc) patients in this study

	Sarcoidosis	SSc
<b>Patients n</b>	21	20
<b>Sex females/males n</b>	9/12	14/6
<b>Age at sampling yrs</b>	37 (27–63)	58 (17–74)
<b>Sarcoid CXR stage<sup>#</sup></b>		
0	1	
I	7	
II	5	
III	2	
IV	6	
<b>SSc additional information</b>		
lcSSc/dcSSc n		11/9
Polymyositis overlap n		3
FA/no FA n		14/6
<b>SSc autoantibody</b>		
ACA		1
ANA+ENA		7
Jo-1		2
Scl70		9
Scl/PM		1
<b>Immunosuppression treatment</b>		
0	14	12
1	7	8

20 healthy controls (12 male and 8 female; 18–60 yrs of age) and 41 subjects were studied: 21 sarcoidosis and 20 SSc. Sarcoidosis chest radiographs (CXR) were staged (stage 0: normal; stage I: bilateral hilar lymphadenopathy (BHL); stage II: BHL with pulmonary infiltrates; stage III: infiltrates without BHL; stage IV: pulmonary fibrosis). Scleroderma patients were divided clinically into those with limited cutaneous systemic sclerosis (lcSSc) or diffuse cutaneous systemic sclerosis (dcSSc), those with scleroderma/polymyositis overlap and those with or without fibrosing alveolitis (FA/no FA). Antibody status is given for the scleroderma patients (ANA: antinuclear antibody; ENA: extractable nuclear antigen; ACA: anticentromere antibody; Scl 70: anti-topoisomerase antibody; Scl/PM: scleroderma polymyositis overlap antibody; Jo-1: anti-histadyl tRNA synthetase antibody). The use of immunosuppressive treatment at the time of venesection is recorded as 0 (no immunosuppressive treatment) or 1 (immunosuppressive treatment including corticosteroids up to a maximum dose of 10 mg of prednisolone or equivalent per day, but excluding powerful second line immunosuppressant agents such as azathioprine, methotrexate or cyclophosphamide). #: includes one patient on penicillamine.

**TABLE 2** Median values of the 20 differentially regulated biomarkers in each population

Cytokine functional group	Cytokine	Normal		Sarcoid			Scleroderma			
		Median	Range	Median	Range	p-value	Median	Range	p-value	
<b>Innate immunity</b>	IL-1 $\beta$	0	0–42.3	2.5	0–499.1	0.003	9.4	0–255.8	0.004	
	IL-6	0	0–20.6	19.9	0–633.4	<0.001	31.7	0–183.7	<0.001	
	IL-8	0	0–19.5	42.0	10.0–1505.3	<0.001	21.1	0–1962.7	<0.001	
	TNF- $\alpha$	7.1	3.8–44.9	7.1	0.3–179.3	NS	7.1	3.8–217.2	NS	
	TNF-RI	814.2	421.6–1208.8	1284.1	437.2–4006.9	<0.001	1,430.6	586.5–4715.7	<0.001	
	TNF-RII	519.0	298.1–927.9	1628.0	155.0–2262.6	<0.001	962.4	127.0–3867.3	0.002	
	IL-12 p40	77.3	23.7–234.3	138.8	49.3–387.4	0.001	169.8	43.8–2406.9	<0.001	
	IL-15	0.9	0–17.6	3.2	0–252.0	0.049	5.7	0–674.7	0.001	
<b>Adaptive immunity</b>	T regulatory	IL-10	1.0	0.3–52.5	1.0	0–5.9	NS	2.4	0.3–14.8	<0.001
	Th17	IL-17	0	0–127.0	0	0–40.3	NS	11.2	0–634.0	0.003
<b>Chemokines</b>	Th1 CXCR3 ligands	IP-10	9.4	4.5–27.1	49.3	10.7–297.7	<0.001	32.8	7.6–2468.0	<0.001
		MIG	27.6	22.1–52.4	108.5	29.9–1193.0	<0.001	41.1	28.8–3068.0	<0.001
	Th1 CCR5 ligands	MIP-1 $\alpha$	16.0	0–63.2	42.9	0–273.7	0.012	16.0	0–290.0	0.008
		MIP-1 $\beta$	121.5	53.5–276.6	340.4	33.5–1604.2	0.001	311.2	42.2–1107.1	<0.001
		RANTES <sup>#</sup>	8.2	1.1–43.6	27.8	5.3–43.6	0.005	7.3	2.1–43.6	NS
	Th2 CCR2	MCP-1	420.7	164.0–794.8	425.6	226.7–1417.6	NS	859.8	235.6–5334.6	<0.001
	Th2 CCR3	Eotaxin	156.3	97.3–271.1	171.3	67.4–347.3	NS	208.6	83.4–344.1	NS
	<b>Growth factors</b>	EGF	166.6	33.3–374.1	421.0	83.5–593.9	0.001	325.6	179.4–711.3	<0.001
HGF		145.1	44.1–223.8	327.4	114.8–734.6	<0.001	416.4	188.2–932.1	<0.001	
<b>Miscellaneous</b>	DR5	0	0–198.7	148.1	0–4976.6	<0.001	104.5	0–1820.2	<0.001	

The 20 biomarkers have been arranged into categories based on their principal function in the immune response. In each of the three groups, the median and range (pg·mL<sup>-1</sup> except <sup>#</sup>: ng·mL<sup>-1</sup>) are given for each protein. In each of the two disease groups, p represents the probability that the median level of that cytokine is different from controls (Mann–Whitney test). Th: T helper cell; IL: interleukin; IP: interferon-inducible protein; TNF: tumour necrosis factor; MIG: monokine induced by interferon- $\gamma$ ; MIP: macrophage inflammatory protein; RANTES: regulated on activation, normal T cell expressed and secreted; MCP: monocyte chemoattractant protein; EGF: epidermal growth factor; HGF: hepatocyte growth factor; DR: death receptor; ns: nonsignificant.

predominant immune function as follows: innate immunity (interleukin (IL)-1, -6, -8, -12p40, -15, interferon (IFN)- $\alpha$ , tumour necrosis factor (TNF)- $\alpha$ , TNF-RI and TNF-RII), adaptive immunity (IL-2, -4, -5, -10, -13, -17 and IFN- $\gamma$ ), chemokines interferon-inducible protein (IP-10), monokine induced by IFN- $\gamma$  (MIG), macrophage inflammatory protein (MIP)-1 $\alpha$ , MIP-1 $\beta$ , regulated on activation, normal T cell expressed and secreted (RANTES), monocyte chemoattractant protein (MCP)-1 and eotaxin, growth factors (granulocyte colony-stimulating factor (G-CSF), granulocyte–macrophage colony-stimulating factor (GM-CSF), epidermal growth factor (EGF), hepatocyte growth factor (HGF), fibroblast growth factor basic (FGFb) and vascular endothelium growth factor (VEGF)), and to death receptor (DR)5.

### Data analysis

For individual serum mediators, differences between study groups were detected using the Kruskal–Wallis test. When the p-value was <0.05 a Mann–Whitney test was used to compare pairs from all three. Spearman's rank correlation was used to establish the interrelationship between analyte levels in each study group. Here we report all correlations and we highlight “strong correlations” that survive Bonferroni correction for multiple comparisons (p<0.001).

Fisher's discriminant function analysis was used for multigroup classification. A variable was entered into the “model” if the significance level of its F-value was <0.05 and was removed if the significance level was >0.1. The model was then used to classify each case into a diagnostic group. Cross-validation was also included in which the phenotype of each case was classified by the functions derived from all cases other than the case. We measured the percentage accuracy of the model, *i.e.* the percentage of correct predictions made by the model, and the receiver operator curve (ROC) to gain insight into the decision-making ability of the model, *i.e.* how likely is the model to accurately assign an individual to a correct group in the context of the other two groups. The probability from the discriminant analysis for a case belonging to a particular diagnostic category was used to build the ROC for the model. Discriminant function analysis was also used to determine which variables can differentiate between scleroderma patients with and without lung fibrosis. All analysis was conducted using the SPSS v14 software package (SPSS, Chicago, IL, USA).

## RESULTS

### Analysis of individual serum proteins

20 serum proteins were elevated in one or both patient groups compared with controls (table 2). Innate immune system

**TABLE 3** Pattern of differentially regulated biomarkers

Cytokine functional group	Cytokine	Sarcoidosis	Scleroderma	
<b>Innate immunity</b>	IL-1 $\beta$	+	+	
	IL-6	+	+	
	IL-8	+	+	
	TNF-RI	+	+	
	TNF-RII	++	+	
	IFN- $\alpha$		+	
	IL-12 p40	+	+	
	IL-15	+	+	
<b>Adaptive immunity</b>				
	T regulatory		++	
	Th17		++	
<b>Chemokines</b>				
	Th1 CXCR3 ligands	IP-10	+	+
		MIG	+	+
	Th1 CCR5 ligands	MIP-1 $\alpha$	+	+
		MIP-1 $\beta$	+	+
		RANTES	++	
	Th2 CCR2	MCP-1		++
	Th2 CCR3	Eotaxin		+ <sup>#</sup>
	<b>Growth factors</b>	HGF	+	+
		EGF	+	+
<b>Apoptosis</b>	DR5	+	+	

The table profiles serum concentrations of all identified differentially expressed biomarkers in sarcoidosis and scleroderma, classified according to their predominant immune function. Th: T helper cell; IL: interleukin; TNF: tumour necrosis factor; IP: interferon-inducible protein; MIG: monokine induced by interferon- $\gamma$ ; MIP: macrophage inflammatory protein; RANTES: regulated on activation, normal T cell expressed and secreted; MCP: monocyte chemotactic protein; HGF: hepatocyte growth factor; EGF: epidermal growth factor; DR: death receptor. +: median serum concentration was significantly higher than the healthy control population. ++: median serum concentration was significantly higher than the healthy control population and the other disease group; #: in untreated patients only.

cytokines IL-1, -6, -8, TNF-RI, TNF-RII and growth factors EGF and HGF were elevated in both diseases. In scleroderma there was elevation of Th1 chemokines (IP-10, MIG, MIP-1 $\alpha$ , MIP-1 $\beta$ ), Th2 chemokines (MCP-1 and eotaxin), and the Th17 cytokine IL-17. In sarcoidosis there was elevation of Th1 chemokines (IP-10, MIG, MIP-1 $\alpha$ , MIP-1 $\beta$  and RANTES) (table 3). There was considerable overlap in serum cytokine levels between the two disease groups so that it was not possible to classify patients using individual protein levels alone (see fig. 1).

We did find some differences in cytokine levels between groups according to complications. When sarcoidosis patients were classified according to the presence or absence of pulmonary infiltrates/fibrosis, only IL-6 was significantly increased ( $p=0.026$ ) in patients with pulmonary infiltrates/fibrosis (chest radiograph stage II–IV) compared with patients without pulmonary infiltrates/fibrosis (stage 0–I). In scleroderma, we observed a significant increase of EGF ( $p=0.027$ ) in those positive for the anti-Scl-70 autoantibody compared with the

anti-Scl-70 negative group. We looked for any confounding effects of treatment. In sarcoidosis, IL-6 was significantly raised in treated patients compared with those on no therapy (median 44.2 pg·mL<sup>-1</sup> versus 9.0 pg·mL<sup>-1</sup>;  $p=0.036$ ) suggesting it may be a marker for more severe disease. In scleroderma, FGFb was significantly higher in untreated patients than in treated (median 31.0 versus 0.1 pg·mL<sup>-1</sup>;  $p=0.032$ ) and eotaxin was elevated compared with healthy controls in untreated patients only ( $p=0.002$ ), suggesting that serum levels of both may be reduced by glucocorticoid therapy. However, the initial use of the Kruskal–Wallis for these analyses denotes a *post hoc* approach and when corrected for multiple comparisons these associations do not survive correction. Nevertheless, the tendency towards significance is reported here as it can form the basis for independent confirmatory studies.

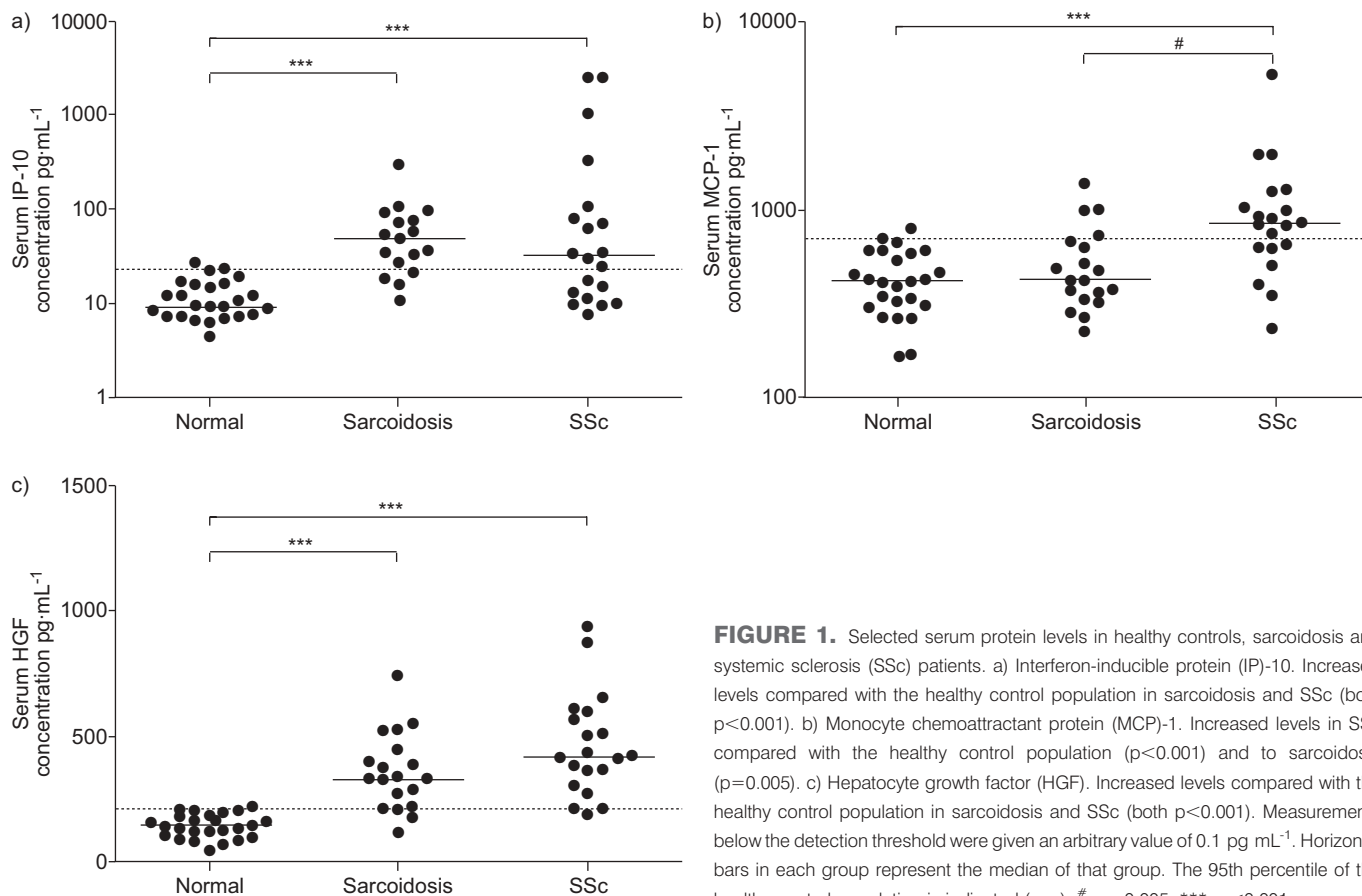
### Correlation between serum biomarkers levels

We conducted a correlation analysis to establish the underlying protein networks in each study group. To visualise these correlations we used the nodes/vertices format in which serum proteins are considered as nodes and the significant correlations as the lines/vertices in the graphs displayed (see fig. 2). In healthy controls there were strong positive correlations ( $p<0.001$  after Bonferroni correction) between IL-17 and TNF- $\alpha$  and IFN- $\alpha$  and negative correlations between granulocyte colony stimulating factor (G-CSF) and MIP1 $\beta$  and EGF (fig. 2a). In patients with sarcoidosis the topology of the inflammatory protein network was different and there appeared to be significant positive correlations between IL-8 and EGF, TNF- $\alpha$  and IL-1 (fig. 2b). IL-1 levels also strongly correlated with IL-6 levels. The most complex inflammatory network was observed in scleroderma patients (fig. 2c), with three significant relationships; positive correlations between IL-1 and GM-CSF, MIP1- $\alpha$  and IL-15, positive correlations between TNF- $\alpha$  with IL-15 and IL12p40 and negative correlations between RANTES and IL-15, IL-2, IL-4 and IFN- $\alpha$ .

### Differentiation of the study groups by combination of mediators

We used discriminant analysis to build a “model” that can best predict to which group (healthy controls, sarcoidosis, SSc) a case belongs. Using all cases in this study we identified a group of 17 mediators (IL-1, TNF- $\alpha$ , IL-15, TNF-RI, TNF-RII, IFN- $\gamma$ , IL-2, VEGF, IP-10, FGFb, MCP1, MIP1 $\alpha$ , RANTES, G-CSF, eotaxin, HGF, MIG) that can be used to group cases into healthy controls, sarcoidosis or SSc. The 17-cytokine profile correctly classified 89.5% of cases. A limitation of this type of analysis is that the cytokine profile may be biased by results obtained from a small number of patients. We then cross-validated our findings by constructing a new analytical model whereby each individual patient was in turn removed from the original data set and the 17-cytokine profile obtained from the remaining cases was then used to determine whether the omitted case could be correctly classified as belonging to either the healthy control, sarcoidosis or SSc patient group. This analysis correctly classified 84.2% of omitted cases. We also randomly assigned original cases into a training cohort (60% of original cases) to obtain new classification function coefficients for the 17 analytes and then applied the new parameter obtained to a test cohort that consisted of the remaining cases. The training cohort correctly classified 94.2% of cases correctly





**FIGURE 1.** Selected serum protein levels in healthy controls, sarcoidosis and systemic sclerosis (SSc) patients. a) Interferon-inducible protein (IP)-10. Increased levels compared with the healthy control population in sarcoidosis and SSc (both  $p < 0.001$ ). b) Monocyte chemoattractant protein (MCP)-1. Increased levels in SSc compared with the healthy control population ( $p < 0.001$ ) and to sarcoidosis ( $p = 0.005$ ). c) Hepatocyte growth factor (HGF). Increased levels compared with the healthy control population in sarcoidosis and SSc (both  $p < 0.001$ ). Measurements below the detection threshold were given an arbitrary value of  $0.1 \text{ pg mL}^{-1}$ . Horizontal bars in each group represent the median of that group. The 95th percentile of the healthy control population is indicated (.....). #:  $p = 0.005$ . \*\*\*:  $p < 0.001$ .

and 75.6% cases in the test cohort were correctly classified. Overall, the highest percentage of classification accuracy was observed for healthy controls and SSc patients with healthy controls showing the best classification accuracy. The percentage of classification accuracy by the algorithms in the three groups is shown in table 4. The area under the ROC based on the discriminant function analysis probability of a case belonging to a particular diagnostic category was healthy controls=0.97; sarcoidosis=0.96; SSc=0.98.

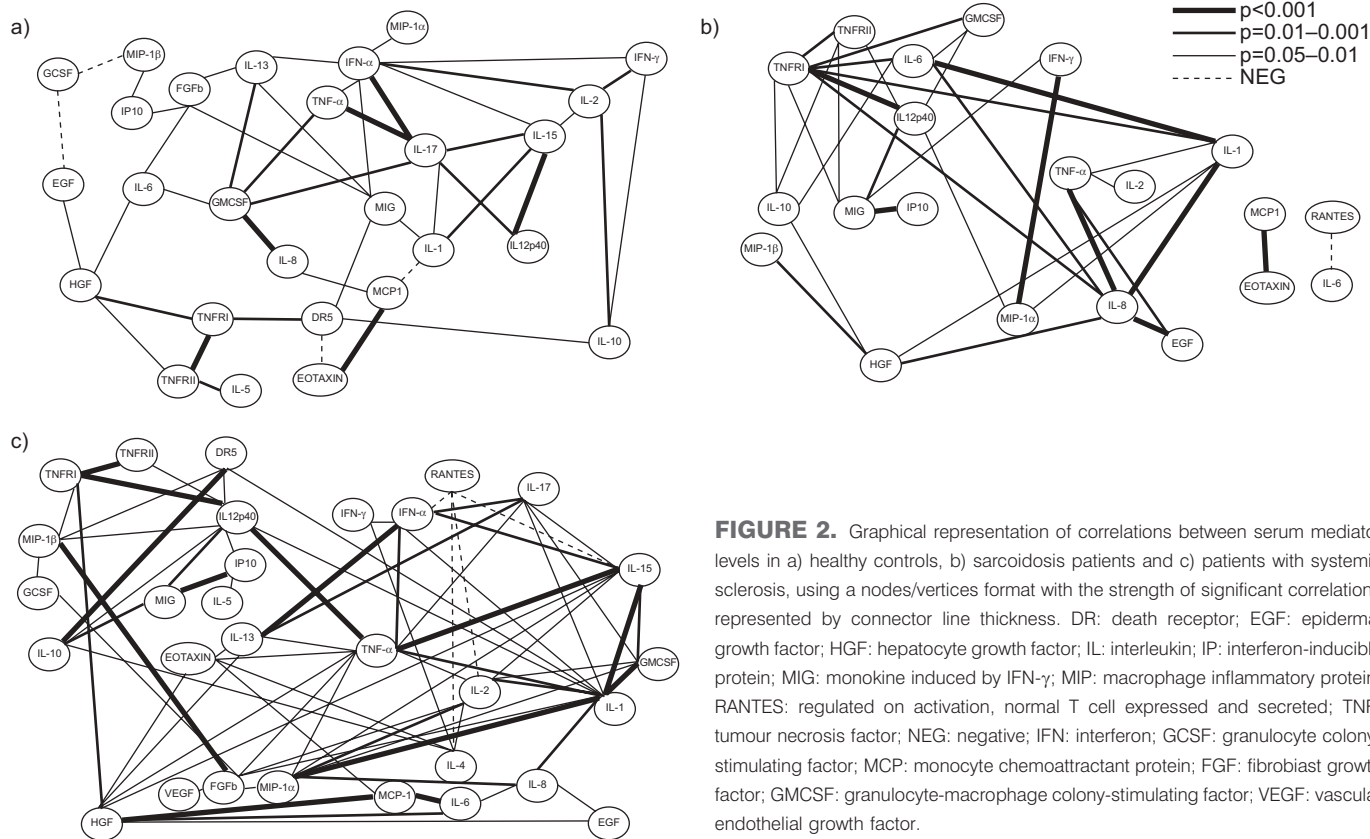
We also evaluated which of the analytes contributed most to the accuracy in the generated predictive algorithms. We identified that six analytes, TNF-RII, IFN- $\gamma$ , IP-10, RANTES, HGF and MIG, were the major contributors to the classification models with 78.9% of original grouped cases being correctly classified to their corresponding groups on these analytes alone. The ROC area under the curve for this model was healthy controls=0.97, sarcoidosis=0.90, SSc=0.93.

We also wanted to explore whether, as a proof of principle due to the small number of cases, this approach could also be applied in differentiating between the presence or absence of lung fibrosis in SSc patients. 14 variables, IL-6, TNF- $\alpha$ , IFN- $\alpha$ , IL-17, IP-10, MIG, HGF, VEGF, TNF-RII, G-CSF, IFN- $\gamma$ , MCP-1, RANTES and IL-5, classified with 100% accuracy those patients with or without lung fibrosis, with 90% of the cases correctly classified on cross-validation, although the small number of patients did not permit any further validation.

## DISCUSSION

This study shows that a serum protein microarray has the capacity to distinguish healthy individuals from patients with two conditions: sarcoidosis and systemic sclerosis. In both diseases, respiratory complications are a major cause for morbidity and mortality. However, the extent of involvement is variable and markers that classify cases likely to progress or that may be candidates for immune directed therapy are needed.

Analysis of median levels of individual cytokines confirmed previously published studies showing upregulation of IL-1 $\beta$ , IL-6, IL-8, IL-10, IL-12p40, IL-15, IL-17, TNF-RI, TNF-RII, MIP-1 $\alpha$ , MIP-1 $\beta$ , MCP-1, HGF and IP-10 in SSc patients [13–23] and upregulation of IL-8, TNF-RI, TNF-RII, IL-12p40, MIP-1 $\alpha$ , MIP-1 $\beta$ , MIG and IP-10 in sarcoidosis [24–27]. Novel findings, as far as we are aware, for other individual biomarkers included increased serum IFN- $\alpha$ , MIG, IP-10, DR5 and eotaxin levels in SSc patients and elevated serum levels of RANTES, DR5, EGF and HGF in sarcoidosis compared with healthy controls. For some proteins (notably VEGF in both diseases) and MCP-1 in sarcoidosis, we were unable to replicate previous reports [26, 28] and the reasons for these discrepant findings could include assay sensitivity and/or patient population choice. Another interesting observation was the variable complexity of correlations between mediators in the three groups. From the nodes/vertices representation, it is obvious that in SSc there is a much more complex correlation profile between mediators than in the other two groups.



**FIGURE 2.** Graphical representation of correlations between serum mediator levels in a) healthy controls, b) sarcoidosis patients and c) patients with systemic sclerosis, using a nodes/vertices format with the strength of significant correlations represented by connector line thickness. DR: death receptor; EGF: epidermal growth factor; HGF: hepatocyte growth factor; IL: interleukin; IP: interferon-inducible protein; MIG: monokine induced by IFN-γ; MIP: macrophage inflammatory protein; RANTES: regulated on activation, normal T cell expressed and secreted; TNF: tumour necrosis factor; NEG: negative; IFN: interferon; GCSF: granulocyte colony-stimulating factor; MCP: monocyte chemoattractant protein; FGF: fibroblast growth factor; GMCSF: granulocyte-macrophage colony-stimulating factor; VEGF: vascular endothelial growth factor.

Sarcoidosis and SSc share the features of immune activation and the development of pulmonary fibrosis. Despite the identification of increased inflammatory mediator levels in sarcoidosis and SSc, no single protein proved capable of discriminating between controls, sarcoidosis and SSc, which is consistent with similar observations in other interstitial lung disease studies highlighted in a recent review [29]. This was mainly due to the fact that, even for those mediators in which the majority of the patients were above the 95th percentile of the normal range,

there was considerable overlap between disease groups. Therefore, we investigated beyond the individual analyte level by exploring whether a combination of mediators was more informative in terms of clinical group differentiation. For this purpose we built a 17-analyte differentiation model using the ROC to gain insight into the decision-making ability of the model and examined the accuracy of correct predictions made by the model since AUC and accuracy provide different information [30]. We observed that the combination of the 17-multimarker panel was able to distinguish, with 89.5% accuracy, between the three clinically distinct groups on the basis of sera inflammatory mediator levels alone. The robustness of this panel of analytes was verified by cross-validation and, separately, by the random selection of subsets of patients. Even with a reduction to six “core” analytes, a model with 78.9% clinical group classification accuracy was constructed. Interestingly, many of the analytes contributing to the models were not upregulated in a disease-specific manner. This raises the issue whether the common practice of restricting analyte selection only to those with demonstrable median levels differences is sound, especially since little additional discriminatory information is gained by selecting mediators that are up- or downregulated together. We also wanted to explore, as a proof of principle, whether the same approach could be used to discriminate in the SSc group between patient subsets with and without lung involvement. The observation, however tentative, that biological measurements distinguish between the presence and absence of lung fibrosis in SSc suggests that inflammatory biomarkers might be useful in assessing the development of lung fibrosis in this disease. These data, though tantalising, need

**TABLE 4** Percentage of original and cross-validated cases correctly classified into different diagnostic categories by the 17-analyte model

Original diagnosis category	Predicted diagnostic category		
	Healthy controls	Sarcoidosis	SSc
<b>Original group cases</b>			
Healthy controls	100.0	0.0	0.0
Sarcoidosis	5.9	82.4	11.8
SSc	5.0	5.0	90.0
<b>Cross-validated group cases</b>			
Healthy controls	90.0	10.0	0.0
Sarcoidosis	11.8	70.6	17.6
SSc	5.0	5.0	90.0

SSc: systemic sclerosis.

confirmation requiring serial measurements in a large cohort and, it can be surmised, with the incorporation of other proteins (for example, transforming growth factor- $\beta$ 1, connective tissue growth factor and platelet-derived growth factor) that have been implicated in the pathology of this disease.

In conclusion, we show that serum protein profiles can distinguish between sarcoidosis, SSc and healthy controls. In addition, multiplex protein technology may discriminate between the presence and absence of pulmonary fibrosis in patients with SSc. Although this tool is more complex than the usual indices used to make a clear diagnosis, the identification of disease-specific mediator profiles carries the distinct possibility that changes in these profiles can be explored in the future as a means of tracking disease evolution, response to treatment or differentiation of phenotypically unusual or overlapping patterns. These findings, if they are corroborated by separate, prospective validation cohorts from other centres, have the potential for clinical application and may generate novel insights into the immune basis of respiratory disease. In future work, the range of proteins that can be detected is likely to be expanded and studies are warranted to evaluate the prognostic value of variations in disease-specific profiles.

#### SUPPORT STATEMENT

This work was supported by generous funding from the UK Raynaud's and Scleroderma Association, the Arthritis Research Campaign and the Asmarley Charitable Trust.

#### STATEMENT OF INTEREST

None declared.

#### REFERENCES

- 1 Minshall EM, Tscopoulos A, Yasrael Z, *et al.* Cytokine mRNA gene expression in active and nonactive pulmonary sarcoidosis. *Eur Respir J* 1997; 10: 2034–2039.
- 2 Lamblin C, Bergoin C, Saelens T, *et al.* Interstitial lung diseases in collagen vascular diseases. *Eur Respir J* 2001; 18: Suppl. 32, 69s–80s.
- 3 Hasegawa M, Fujimoto M, Takehara K, *et al.* Pathogenesis of systemic sclerosis: altered B cell function is the key linking systemic autoimmunity and tissue fibrosis. *J Dermatol Sci* 2005; 39: 1–7.
- 4 Valentini G, Baroni A, Esposito K, *et al.* Peripheral blood T lymphocytes from systemic sclerosis patients show both Th1 and Th2 activation. *J Clin Immunol* 2001; 21: 210–217.
- 5 Suga M, Iyonaga K, Ichiyasu H, *et al.* Clinical significance of MCP-1 levels in BALF and serum in patients with interstitial lung diseases. *Eur Respir J* 1999; 14: 376–382.
- 6 Antonelli A, Ferri C, Fallahi P, *et al.* CXCL10 (alpha) and CCL2 (beta) chemokines in systemic sclerosis—a longitudinal study. *Rheumatology (Oxford)* 2008; 47: 45–49.
- 7 Diot E, Diot P, Valat C, *et al.* Predictive value of serum III procollagen for diagnosis of pulmonary involvement in patients with scleroderma. *Eur Respir J* 1995; 8: 1559–1565.
- 8 Balboni I, Chan SM, Kattah M, *et al.* Multiplexed protein array platforms for analysis of autoimmune diseases. *Annu Rev Immunol* 2006; 24: 391–418.
- 9 Fathman CG, Soares L, Chan SM, *et al.* An array of possibilities for the study of autoimmunity. *Nature* 2005; 435: 605–611.
- 10 Bowler RP, Ellison MC, Reisdorph N. Proteomics in pulmonary medicine. *Chest* 2006; 130: 567–574.
- 11 Statement on Sarcoidosis. Joint Statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG) Adopted by the ATS Board of Directors and by the ERS Executive Committee, February 1999. *Am J Respir Crit Care Med* 1999; 160: 736–755.
- 12 Preliminary Criteria for the Classification of Systemic Sclerosis (Scleroderma). Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. *Arthritis Rheum* 1980; 23: 581–590.
- 13 Kurasawa K, Hirose K, Sano H, *et al.* Increased interleukin-17 production in patients with systemic sclerosis. *Arthritis Rheum* 2000; 43: 2455–2463.
- 14 Needleman BW, Wigley FM, Stair RW. Interleukin-1, interleukin-2, interleukin-4, interleukin-6, tumor necrosis factor alpha, and interferon-gamma levels in sera from patients with scleroderma. *Arthritis Rheum* 1992; 35: 67–72.
- 15 Furuse S, Fujii H, Kaburagi Y, *et al.* Serum concentrations of the CXC chemokines interleukin 8 and growth-regulated oncogene-alpha are elevated in patients with systemic sclerosis. *J Rheumatol* 2003; 30: 1524–1528.
- 16 Majewski S, Wojas-Pelc A, Malejczyk M, *et al.* Serum levels of soluble TNF alpha receptor type I and the severity of systemic sclerosis. *Acta Derm Venereol* 1999; 79: 207–210.
- 17 Sato S, Hasegawa M, Takehara K. Serum levels of interleukin-6 and interleukin-10 correlate with total skin thickness score in patients with systemic sclerosis. *J Dermatol Sci* 2001; 27: 140–146.
- 18 Sato S, Hanakawa H, Hasegawa M, *et al.* Levels of interleukin 12, a cytokine of type 1 helper T cells, are elevated in sera from patients with systemic sclerosis. *J Rheumatol* 2000; 27: 2838–2842.
- 19 Zielonka TM, Demkow U, Bialas B, *et al.* Modulatory effect of sera from sarcoidosis patients on mononuclear cell-induced angiogenesis. *J Physiol Pharmacol* 2007; 58: Suppl. 5, 753–766.
- 20 Heilig B, Fiehn C, Brockhaus M, *et al.* Evaluation of soluble tumor necrosis factor (TNF) receptors and TNF receptor antibodies in patients with systemic lupus erythematoses, progressive systemic sclerosis, and mixed connective tissue disease. *J Clin Immunol* 1993; 13: 321–328.
- 21 Suzuki J, Morimoto S, Amano H, *et al.* Serum levels of interleukin 15 in patients with rheumatic diseases. *J Rheumatol* 2001; 28: 2389–2391.
- 22 Scala E, Pallotta S, Frezzolini A, *et al.* Cytokine and chemokine levels in systemic sclerosis: relationship with cutaneous and internal organ involvement. *Clin Exp Immunol* 2004; 138: 540–546.
- 23 Matsushita T, Hasegawa M, Hamaguchi Y, *et al.* Longitudinal analysis of serum cytokine concentrations in systemic sclerosis: association of interleukin 12 elevation with spontaneous regression of skin sclerosis. *J Rheumatol* 2006; 33: 275–284.
- 24 Ziegenhagen MW, Fitschen J, Martinet N, *et al.* Serum level of soluble tumour necrosis factor receptor II (75 kDa) indicates inflammatory activity of sarcoidosis. *J Intern Med* 2000; 248: 33–41.
- 25 Shigehara K, Shijubo N, Ohmichi M, *et al.* Increased circulating interleukin-12 (IL-12) P40 in pulmonary sarcoidosis. *Clin Exp Immunol* 2003; 132: 152–157.
- 26 Hashimoto S, Nakayama T, Gon Y, *et al.* Correlation of plasma monocyte chemoattractant protein-1 (MCP-1) and monocyte inflammatory protein-1alpha (MIP-1 $\alpha$ ) levels with disease activity and clinical course of sarcoidosis. *Clin Exp Immunol* 1998; 111: 604–610.
- 27 Takeuchi M, Oh I, Suzuki J, *et al.* Elevated serum levels of CXCL9/monokine induced by interferon-gamma and CXCL10/interferon- $\gamma$ -inducible protein-10 in ocular sarcoidosis. *Invest Ophthalmol Vis Sci* 2006; 47: 1063–1068.
- 28 Choi JJ, Min DJ, Cho ML, *et al.* Elevated vascular endothelial growth factor in systemic sclerosis. *J Rheumatol* 2003; 30: 1529–1533.
- 29 Tzouveleki A, Kouliatsis G, Anevlaivis S, *et al.* Serum biomarkers in interstitial lung diseases. *Respir Res* 2005; 6: 78.
- 30 Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms lower classification accuracy. *IEEE Trans Knowl Data Eng* 2005; 17: 299–310.