American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis & Rheumatology
Vol. 73, No. 4, April 2021, pp 660–670
DOI 10.1002/art.41570

Large-Scale Characterization of Systemic Sclerosis Serum Protein Profile: Comparison to Peripheral Blood Cell Transcriptome and Correlations With Skin/Lung Fibrosis

Chiara Bellocchi,[1] Jun Ying,[2] Ellen A. Goldmuntz,[3] Lynette Keyes-Elstein,[4] John Varga,[5] Monique E. Hinchcliff,[6] Marka A. Lyons,[2] Peter McSweeney,[7] Daniel E. Furst,[8] Richard Nash,[7] Leslie J. Crofford,[9] Beverly Welch,[3] Jonathan G. Goldin,[10] Ashley Pinckney,[4] Maureen D. Mayes,[2] Keith M. Sullivan,[11] and Shervin Assassi[2]

Objective. To provide a large-scale assessment of serum protein dysregulation in diffuse cutaneous systemic sclerosis (dcSSc) and to investigate serum protein correlates of SSc fibrotic features.

Methods. We investigated serum protein profiles of 66 participants with dcSSc at baseline who were enrolled in the Scleroderma: Cyclophosphamide or Transplant Trial and 66 age- and sex-matched healthy control subjects. A panel of 230 proteins, including several cytokines and chemokines, was investigated. Whole blood gene expression profiling in concomitantly collected samples was performed.

Results. Among the participants with dcSSc, the mean disease duration was 2.3 years. All had interstitial lung disease (ILD), and none were being treated with immunosuppressive agents at baseline. Ninety proteins were differentially expressed in participants with dcSSc compared to healthy control subjects. Similar to previous global skin transcript results, hepatic fibrosis, granulocyte and agranulocyte adhesion, and diapedesis were the top overrepresented pathways. Eighteen proteins correlated with the modified Rodnan skin thickness score (MRSS). Soluble epidermal growth factor receptor was significantly down-regulated in dcSSc and showed the strongest negative correlation with the MRSS, being predictive of the score's course over time, whereas α_1-antichymotrypsin was significantly up-regulated in dcSSc and showed the strongest positive correlation with the MRSS. Furthermore, higher levels of cancer antigen 15-3 correlated with more severe ILD, based on findings of reduced forced vital capacity and higher scores of disease activity on high-resolution computed tomography. Only 14 genes showed significant differential expression in the same direction in serum protein and whole blood RNA gene expression analyses.

Conclusion. Diffuse cutaneous SSc has a distinct serum protein profile with prominent dysregulation of proteins related to fibrosis and immune cell adhesion/diapedesis. The differential expression for most serum proteins in SSc is likely to originate outside the peripheral blood cells.

INTRODUCTION

Systemic sclerosis (SSc; scleroderma) is a complex autoimmune disorder in which vascular involvement, immune dysregulation with autoantibody production, and fibrosis are the main pathologic processes (1). As is evident from a standardized mortality ratio of 3.5 (2), SSc is associated with a substantial mortality and disease burden. Development of effective

Supported by the Karen Brown Scleroderma Foundation, the NIH (grants R01-AR-073284 and UL1-TR-000371 and National Institute of Arthritis and Musculoskeletal and Skin Diseases grant P30-AR-061271), and the Department of Defense (grant W81XWH-16-1-0296). The Scleroderma: Cyclophosphamide or Transplantation (SCOT) study was supported by awards from the National Institute of Allergy and Infectious Diseases, NIH, to the study contract holder Duke University (grants N01-AI-05419 and HHSN 272201100025C).

[1]Chiara Bellocchi, MD: The University of Texas Health Science Center at Houston and McGovern Medical School, Houston, and Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy; [2]Jun Ying, MS, Marka A. Lyons, MA, Maureen D. Mayes, MD, MPH, Shervin Assassi, MD, MS: The University of Texas Health Science Center at Houston and McGovern Medical School, Houston; [3]Ellen A. Goldmuntz, MD, PhD, Beverly Welch, RN, MSN: National Institute of Allergy and Infectious Diseases, NIH, Rockville, Maryland; [4]Lynette Keyes-Elstein, DrPH, Ashley Pinckney, MS: Rho Federal Systems Division, Inc., Durham, North Carolina; [5]John Varga, MD: Northwestern University, Chicago, Illinois; [6]Monique E. Hinchcliff, MD: Yale University School of Medicine, New Haven, Connecticut; [7]Peter McSweeney, MD, Richard Nash, MD: Colorado Blood Cancer Institute, Denver; [8]Daniel E. Furst, MD: University of California Los Angeles, University of Washington, Seattle, and University of Florence, Florence, Italy; [9]Leslie J. Crofford, MD: Vanderbilt University Medical Center, Nashville, Tennessee; [10]Jonathan G. Goldin, MD, PhD: University of California Los Angeles; [11]Keith M. Sullivan, MD: Duke University, Durham, North Carolina.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Shervin Assassi, MD, MS, 6431 Fannin, MSB 5.266, Houston, TX 77030. Email: shervin.assassi@uth.tmc.edu.

Submitted for publication February 19, 2020; accepted in revised form October 27, 2020.

treatment strategies for SSc has been hampered by an incomplete understanding of disease pathogenesis and the underlying molecular heterogeneity. Genome-wide association studies (3,4) and whole genome microarray studies (5,6) have provided new insights into disease pathogenesis at the DNA and RNA levels. An interferon (IFN) signature is the most prominent gene expression signature in SSc peripheral blood cells (5), and IFN-inducible chemokines correlate with disease severity (7).

Although changes in serum protein profiles, as opposed to findings at the DNA or RNA level, may be more closely associated with disease pathogenesis, data on large-scale examination of serum proteins in SSc are still scarce. Investigations performed on small panels of proteins showed that chemokines, vascular growth factors, and adhesion molecules such as interleukin-8 (IL-8), vascular endothelial growth factor (VEGF), angiopoietin 2 (Ang-1), platelet endothelial cellular adhesion molecule 1 (PECAM-1) are markers of early disease and reflect endothelial dysregulation in SSc (8–10). IL-6 is related to lung fibrosis and might be predictive of disease progression in early SSc interstitial lung disease (ILD) (11). Moreover, levels of adipokines, such as leptin and adiponectin, have a negative correlation with changes in SSc skin fibrosis (12–14).

Recently, a study performed with the use of SOMAscan aptamer technology in 34 patients with diffuse cutaneous SSc (dcSSc) and 15 control subjects identified tumor necrosis factor (TNF), IFNγ, transforming growth factor β (TGFβ), and IL-13 as potential upstream regulators in SSc (15). Few other studies have recently explored larger panels of serum proteins in SSc (16,17).

In the present study, we investigated an extended panel of 230 serum proteins in serum samples obtained at baseline from individuals with dcSSc (18) who were enrolled in the Scleroderma: Cyclophosphamide or Transplantation (SCOT) Trial and compared these findings at a 1:1 ratio to serum proteins from matched healthy control subjects in order to provide a hypothesis-generating assessment of serum protein dysregulation and its clinical correlates in SSc (19). Moreover, availability of concomitantly collected whole blood RNA samples enabled direct comparison between SSc serum protein and whole blood gene expression profiles, showing that the differential expression for most serum proteins in SSc is likely to originate outside the peripheral blood cells.

PATIENTS AND METHODS

Selection of study population. Of the 75 participants with dsSSc who were included in the SCOT trial, 66 had a serum sample obtained at baseline available for analysis. Samples obtained at baseline were examined in the present study. None of the participants were receiving immunosuppressive agents except for ≤10 mg per day of prednisone or its equivalent during blood sample collection at baseline. However, 27 participants had received immunosuppressive agents in the 2 months prior to baseline sample collection.

Briefly, the inclusion criteria included diffuse cutaneous involvement, lung or kidney involvement, and a disease duration of <5 years (calculated from the onset of the first non–Raynaud's symptom). Exclusion criteria included significant prior treatment with cyclophosphamide (CYC; either oral or intravenous), presence of clinically significant rheumatic diseases other than SSc, any active uncontrolled infection, or HIV, hepatitis C virus, and hepatitis B virus infections. All SCOT participants provided informed consent, and the SCOT protocol was approved by the Institutional Review Board of all participating institutions. Detailed information on the selection of participants and study design have been published previously (19). Additionally, serum from healthy controls that were matched at a 1:1 ratio based on age (±10 years) and sex was also investigated in the present study.

Severity of lung involvement was evaluated by forced vital capacity percent predicted (FVC%) (19). As another surrogate for ILD severity, standardized volumetric high-resolution computed tomography (HRCT) scanning was performed, and quantitative interstitial lung disease (QILD) score was measured using an established algorithm. The QILD score (expressed as a percentage) represents the sum of quantitative lung fibrosis, quantitative ground glass, and quantitative honeycombing for all lung lobes (20,21). The MRSS was used to assess severity of skin involvement (22). Antibody profiles were determined using commercial laboratories at each site.

Serum protein determination. Serum protein assays were performed by Myriad Rules-Based Medicine using Human Discovery Multi-Analyte Profiling (MAP) (https://myriadrbm.com/scientific-media/multiplex-assay-development-white-paper/) multiplexed immune assay version 2, which was the most comprehensive panel of proteins available with this technology at the time of study. This panel includes an extensive list of cytokines, chemokines, metabolic markers, hormones, growth factors, tissue remodeling proteins, angiogenesis markers, acute-phase reactants, and cancer markers that can reliably and reproducibly be measured with this technology. Levels of 228 serum proteins were determined with this assay.

All samples were stored at a temperature lower than –70°C and had not been previously thawed. An aliquot of each sample was added to individual microsphere multiplexes of the selected MAP and blocker. After incubation, multiplexed cocktails of biotinylated reporter antibodies were added. Multiplexes were labeled using an excess of streptavidin–phycoerythrin solution. The resulting data were interpreted using proprietary software developed by Myriad Rules-Based Medicine. In addition, 2 cytokines considered to be pertinent to SSc pathogenesis, IL-10 and IL-6, were determined by ultrasensitive Simoa Assays (Quanterix) (23).

For each assay, the lower limit of quantification (LLOQ) and upper limit of quantification (ULOQ) were determined by Myriad Rules-Based Medicine, representing the concentration at the lower and upper limit of the linear range of the standard curve,

respectively (i.e., the lowest and highest amount of protein that can be detected accurately).

Analysis of gene expression and its relationship to serum proteins. Global gene expression studies on whole blood RNA samples obtained at baseline (stored in Tempus tubes) from SCOT participants and from matched controls were examined using Illumina Human HT-12 bead arrays. All microarray experiments were performed in a single batch (24). Data were normalized according to the quantile method. A corresponding transcript was present on the microarray platform for the majority of investigated proteins. We focused on the transcripts that corresponded to the investigated proteins. Specifically, 282 transcripts corresponding to 172 examined proteins were identified. Only 10 (5.5%) of 182 serum proteins did not have a matching probe on the microarray platform. For the transcript analysis, the data set was filtered by the list of corresponding 282 transcripts. Differentially expressed transcripts identified at a false discovery rate (FDR) of <0.1 in SSc participants compared to controls were analyzed by 2-sample *t*-test. We used a less stringent FDR cutoff for the analysis of differentially expressed transcripts than was used for the protein analysis (FDR of <0.1 versus FDR of <0.05) as the effect size (i.e., fold change) tends to be lower at the RNA level than at the protein level. For example, fold changes in transcript levels in SSc participants compared to controls among the 282 transcripts investigated ranged from 0.61 to 1.41 whereas the fold change in the corresponding proteins ranged from 0.4 to 3.68. Subsequently, the list of differentially expressed transcripts was intersected with the list of differentially expressed proteins. Microarray analysis was performed with BRB ArrayTools (National Cancer Institute, National Institutes of Health) (25).

Interferome database search. We examined whether the differentially expressed serum proteins in SSc participants versus control subjects were IFN-inducible using the Interferome version 2.01 database (http://interferome.its.monash.edu.au/interferome/) (26). For the Interferome database, *Homo sapiens* was chosen as species and lung, skin, and blood were selected as organs. A list of type I IFN–inducible proteins was generated.

Statistical analysis. Proteins with levels below the LLOQ in >50% of SSc samples collected at baseline were excluded from the analysis. A total of 182 proteins (79.1%) had a detectable level in >50% of SSc samples. Of these 182 proteins, 128 proteins (70.2%) had measurements above the LLOQ in all samples. For the remainder of proteins, levels below the LLOQ were replaced by the LLOQ. Moreover, protein measurements above the ULOQ were replaced by the ULOQ. As shown in Supplementary Figure 1 (http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract), raw values of the majority of the serum proteins showed a right-skewed distribution, and protein expression data were natural log–transformed to approximately conform to normality. Principal components analysis (PCA) was performed to identify outliers. *T*-tests were used to estimate differential expression for each protein between SSc participants and control subjects. *P* values were adjusted for multiple testing using the Benjamini-Hochberg method (27).

Proteins with an FDR of <0.05 were considered to be differentially expressed in the comparison of SSc participants to control subjects. Subsequently, differentially expressed proteins were modeled using Ingenuity Pathway Analysis software (Qiagen [https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/]) to identify the overrepresented canonical pathways and to predict activated upstream cytokines/growth factors. The goal of Upstream Regulator Analysis is to identify upstream regulators of a molecular profile and predict whether they are activated or inhibited. This analysis is based on expected causal effects between upstream cytokines/growth factors and targets; the expected causal effects are derived from the literature compiled in Ingenuity Knowledge Base (28). A Z score algorithm is used to make predictions. The

Table 1. Demographic and clinical characteristics of the SCOT participants and control subjects*

Features	Participants with dcSSc (n = 66)	Control subjects (n = 66)
Age, mean ± SD years	45.3 ± 10.3	46.3 ± 9.5
Female sex	40 (60.6)	40 (60.6)
Race		
African American	6 (9.1)	9 (13.6)
Asian	3 (4.5)	0 (0)
Other	6 (9.1)	0 (0)
White	51 (77.3)	57 (86.4)
Disease duration, mean ± SD years	2.31 ± 1.25	NA
MRSS score, mean ± SD	29.22 ± 9.35	NA
FVC%, mean ± SD	74.62 ± 15.61	NA
QILD score, mean ± SD	22.92 ± 11.63	NA
QILD score of >0	66 (100)	NA
Autoantibodies		
Positive		
ANAs	57 (86.4)	NA
ACAs	4 (6.1)	NA
ATAs	26 (39.4)	NA
RNPs	10 (15.2)	NA
Negative		
ANAs	7 (10.6)	NA
ACAs	56 (84.8)	NA
ATAs	39 (59.1)	NA
RNPs	55 (83.3)	NA
Testing not performed		
ANAs	2 (3)	NA
ACAs	6 (9.1)	NA
ATAs	1 (1.5)	NA
RNPs	1 (1.5)	NA

* Except where indicated, values are the number (%). SCOT = Scleroderma: Cyclophosphamide or Transplantation; dcSSc = diffuse cutaneous systemic sclerosis; ANAs = antinuclear antibodies; ACAs = anticentromere antibodies; ATAs = anti–topoisomerase I antibodies; MRSS = modified Rodnan skin thickness score; FVC% = forced vital capacity percent predicted; QILD = quantitative interstitial lung disease score; NA = not applicable.

primary purpose of the activation Z score is to infer the activation states of predicted expression regulators. Given the observed differential regulation of a molecule ("up" or "down") in the data set, the activation state of an upstream regulator is determined by the regulation direction associated with the relationship from the regulator to the molecule. In practice, Z scores >2 or <–2 can be considered significant.

For correlation with clinical variables (i.e., the MRSS, FVC%, HRCT–QILD score), Pearson's correlation was calculated, and proteins that reached the nominal significance level ($P < 0.05$) in the univariable analysis and a Pearson's correlation coefficient of ≥0.3 or ≤–0.3 were considered as significantly correlated. For this, we did not account for multiple comparisons as this was a hypothesis-generating analysis. Multivariable analyses with adjustment for age and sex were also performed. Partial correlation coefficients after adjustment for these demographic factors were also provided.

In an exploratory analysis, the predictive significance of serum proteins found to correlate with MRSS scores and FVC% was examined for the serial measurement of MRSS scores and FVC% obtained 3–14 months after randomization in the CYC arm (representing the active treatment period) and transplantation arm, separately. For this analysis, mixed-effects linear regression modeling was used after controlling for disease severity at baseline (i.e., the MRSS or FVC% at baseline) and time variable. Fixed effects were serum protein levels and the MRSS or FVC% at baseline as well as time point (all as continuous), and random effects were the intercept and time point. An unstructured correlation matrix was used. Analyses were performed using R Studio version 0.99.489 (RStudio Consortium) and SAS version 9.4 (SAS Institute Inc.).

RESULTS

Demographic and clinical characteristics. Demographic and clinical characteristics of study patients at baseline are presented in Table 1. As expected in diffuse disease, anti–topoisomerase I (Scl-70) was the most common disease-specific autoantibody observed (39.4%), followed by anti-RNP antibodies (15.2%). All SCOT participants had signs of alveolitis on HRCT as evidenced by visual confirmation of ground glass opacity, and mean disease duration was 2.3 years.

Serum protein levels. Ninety of 182 proteins were differentially expressed in SSc participants compared to control subjects, with an FDR of <0.05. A heatmap shows 90 differentially expressed proteins in samples from SCOT participants at baseline compared to healthy controls (Supplementary Figure 2, http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract). The 10 most up-regulated and down-regulated proteins based

Table 2. Top up-regulated and down-regulated serum proteins in the SCOT participants compared to control subjects*

Protein name	Gene name	Fold change	P_{raw}	P_{FDR}	Direction of difference
Growth hormone	GH1†	3.69	<0.001	<0.001	Up-regulated
Ferritin	FTH1	3.04	<0.001	<0.001	Up-regulated
C-reactive protein	CRP	2.98	<0.001	<0.001	Up-regulated
Chromogranin A	CHGA	2.77	<0.001	<0.001	Up-regulated
MIP-3β	CCL19†	2.48	<0.001	<0.001	Up-regulated
MCP-1	CCL2†	2.48	<0.001	<0.001	Up-regulated
Myoglobin	MB	2.38	<0.001	<0.001	Up-regulated
MIG	CXCL9†	2.30	<0.001	<0.001	Up-regulated
BLC	CXCL13†	2.19	<0.001	<0.001	Up-regulated
Prolactin	PRL	2.08	<0.001	<0.001	Up-regulated
Lactoylglutathione lyase	GLO1†	0.49	<0.001	<0.001	Down-regulated
Neuron-specific enolase	ENO2†	0.56	0.002	0.007	Down-regulated
Vitamin K–dependent protein S	PROS1†	0.56	0.005	0.013	Down-regulated
SOD1	SOD1	0.65	<0.001	<0.001	Down-regulated
Protein S100A6	S100A6†	0.69	0.002	0.006	Down-regulated
MIF	MIF	0.71	0.023	0.046	Down-regulated
Adiponectin	ADIPOQ	0.72	<0.001	<0.001	Down-regulated
Kallikrein 7	KLK7†	0.73	<0.001	<0.001	Down-regulated
IGFBP6	IGFBP6	0.73	<0.001	<0.001	Down-regulated
Tetranectin	CLEC3B†	0.75	<0.001	<0.001	Down-regulated

* Values of >1 refer to up-regulated expression of proteins, and values of <1 refer to down-regulated expression of proteins in Scleroderma: Cyclophosphamide or Transplantation (SCOT) study participants compared to values measured in control subjects. For example, a fold change of 2.30 is equivalent to an increase of 130% from the reference value, and a fold change of 0.65 is equivalent to a decrease of 35% from the reference value. FDR = false discovery rate; MIP-3β = macrophage inflammatory protein 3β; MCP-1 = monocyte chemotactic protein 1; MIG = monokine induced by interferon-γ; BLC = B lymphocyte chemoattractant; SOD1 = superoxide dismutase 1; MIF = macrophage migration inhibitory factor; IGFBP6 = insulin-like growth factor binding protein 6.
† Interferon-inducible proteins.

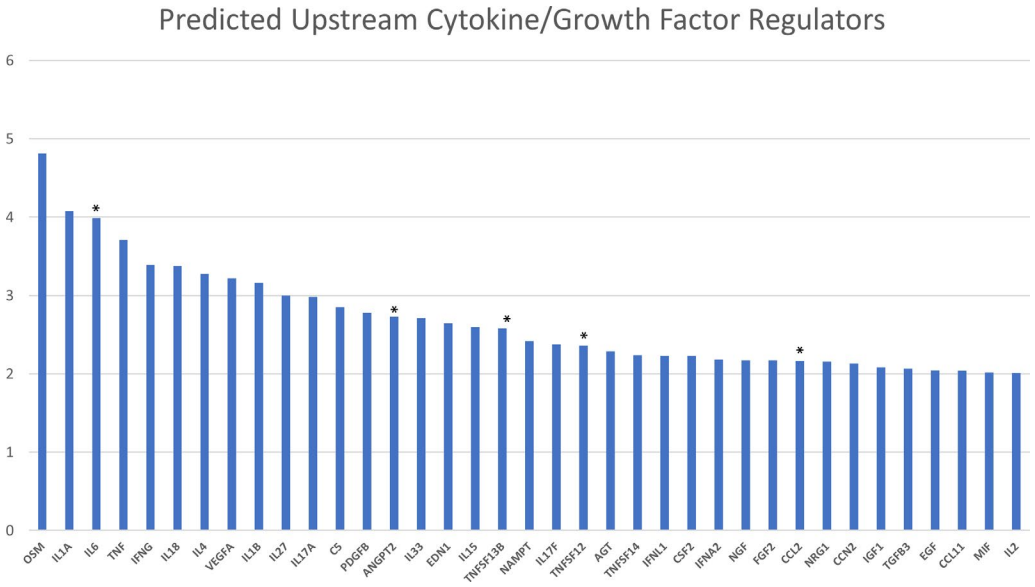

Figure 1. Top predicted upstream cytokine and growth factor regulators based on the Ingenuity Knowledge Base. Y axis shows the activation Z score calculated based on the Ingenuity Pathway Analysis for identifying upstream regulators (see Patients and Methods for further details). Proteins that were differentially expressed in participants with diffuse cutaneous systemic sclerosis compared to control subjects are shown (asterisks). Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract.

on the fold change are presented in Table 2. The complete list of 90 differentially expressed proteins in SSc is available in Supplementary Table 1 (http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract). Complete analysis results for all examined 182 serum proteins are available in an additional data file on the Scleroderma Program at McGovern Medical School website

Table 3. Significant correlations between serum protein expression and the modified Rodnan skin thickness score at baseline*

		Correlation			
		Univariable†		Multivariable‡	
Protein name	Gene name	r	*P*	Partial correlation	*P*
α_1-antichymotrypsin§	SERPINA3	0.42	0.001	0.42	0.001
NT-proBNP§	NPPB	0.38	0.002	0.38	0.002
Endostatin§	COL18A1	0.37	0.002	0.38	0.002
Osteopontin§	SPP1	0.34	0.006	0.38	0.002
Ang-2§	ANGPT2	0.33	0.005	0.33	0.007
SAP component	APCS	0.32	0.008	0.32	0.010
Tenascin-C§	TNC	0.31	0.011	0.32	0.010
α_1-microglobulin	AMBP	0.31	0.011	0.32	0.012
IGFBP4§	IGFBP4	0.30	0.018	0.34	0.009
IL-22	IL22	−0.30	0.016	−0.29	0.020
HGF receptor	MET	−0.30	0.016	−0.31	0.012
Tetranectin¶	CLEC3B	−0.31	0.011	−0.34	0.007
Kallikrein 5¶	KLK5	−0.33	0.007	−0.33	0.007
uPA	PLAU	−0.34	0.005	−0.35	0.005
TARC	CCL17	−0.35	0.003	−0.36	0.003
Tamm-Horsfall urinary glycoprotein	UMOD	−0.40	0.001	−0.42	0.001
MDC¶	CCL22	−0.43	<0.001	−0.44	<0.001
EGFR¶	EGFR	−0.43	<0.001	−0.43	0.001

* Correlations were determined using Pearson's correlation coefficients. NT-proBNP = N-terminal pro–brain natriuretic peptide; Ang-2 = angiopoietin 2; SAP = serum amyloid P; IGFBP4 = insulin-like growth factor binding protein 4; IL-22 = interleukin-22; HGF = hepatocyte growth factor; uPA = urokinase plasminogen activator; TARC = thymus and activation–regulated chemokine; MDC = macrophage-derived chemokine; EGFR = epidermal growth factor receptor.
† Calculated using a univariable model.
‡ Calculated using a multivariable model after adjustment for age and sex.
§ Proteins differentially expressed and up-regulated in participants with diffuse cutaneous systemic sclerosis (dcSSc) compared to control subjects.
¶ Proteins differentially expressed and down-regulated in participants with dcSSc compared to control subjects.

(https://www.uth.tmc.edu/scleroderma/), and individual level protein data are available at ImmPort (https://www.immport.org).

As shown in the Supplementary Figure 3 (http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract), PCA identified only 1 outlier. PCA showed that the majority of SCOT participants had a different serum protein profile compared to control subjects. Furthermore, the 27 individuals who received immunosuppressive therapy 2 months prior to sample collection did not group separately from other SCOT participants. Supplementary Tables 2 and 3 (http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract) show the demographic/clinical characteristics, as well as the list and duration of prior immunosuppressive treatment in SCOT participants dichotomized based on whether they were treated with immunosuppressive agents 2 months prior to sample collection.

Prominent role of profibrotic and granulocyte/agranulocyte extravasation pathways in SSc serum profile. An Ingenuity Canonical Pathway Analysis of differentially expressed serum proteins in SSc participants compared to control subjects revealed hepatic fibrosis, granulocyte adhesion and diapedesis, and agranulocyte adhesion and diapedesis as the top 3 overrepresented pathways. Of note, the same 3 pathways were found to be the top dysregulated pathways in our previously published global gene expression study on SSc skin (29). Interestingly, the top overrepresented canonical pathways in the concomitantly collected whole blood RNA samples were antigen presentation, IFN, and natural killer cell pathways. The complete list of overrepresented canonical pathways in both data sets is shown in the additional data file on the Scleroderma Program at McGovern Medical School website (https://www.uth.tmc.edu/scleroderma/).

As shown in Figure 1, the top predicted activated upstream cytokines/growth factors (based on an activation Z score of >2) for the observed SSc serum protein profile included prominent profibrotic proteins such as oncostatin M (OSM), IL-6, IL-18, IL-4, IL-33, B–cell activating factor (TNFSF13B), and monocyte chemotactic protein 1 (MCP-1; CCL2).

Correlation of serum protein expression with MRSS scores. In SSc participants compared to controls, levels of α_1-antichymotrypsin, N-terminal pro–brain natriuretic peptide (NT-proBNP), endostatin, osteopontin, tenascin, and insulin-like growth factor binding protein 4 (IGFBP-4) were up-regulated ($P < 0.05$) (Supplementary Table 1, http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract) and showed a positive correlation with the MRSS (Table 3). Epidermal growth factor receptor (EGFR), macrophage-derived chemokine, kallikrein 5, and tetranectin were down-regulated in SSc participants compared to control subjects ($P < 0.05$) (Supplementary Table 1) and were negatively correlated with the MRSS. All serum proteins that correlated with MRSS scores are shown in Table 3. Furthermore, complete analysis results for all proteins can be found at https://www.uth.tmc.edu/scleroderma/.

In an exploratory analysis, the predictive significance of serum proteins listed in Table 3 for the course of MRSS scores from 3–14 months after randomization in the CYC arm (representing active treatment period; n = 32) and transplantation arm (n = 30) was investigated. As shown in Supplementary Table 4 (http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract), NT-proBNP and Ang-2 serum levels at baseline predicted higher subsequent MRSS scores ($P = 0.013$ and $P = 0.038$, respectively), whereas EGFR and kallikrein 5 serum levels at baseline predicted lower subsequent MRSS scores ($P = 0.034$ and $P = 0.003$, respectively) in the CYC arm after adjustment for MRSS scores at baseline. Similarly, EGFR and kallikrein 5 predicted lower subsequent MRSS scores in the transplantation arm ($P = 0.019$ and $P = 0.004$, respectively).

Correlation of serum protein levels with ILD severity. Serum protein correlates of FVC and HRCT–QILD score are shown in Supplementary Tables 5 and 6 (http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract). Notably, cancer antigen 15-3 (CA 15-3) and growth-regulated α protein were associated with more severe involvement (i.e., lower FVC and higher HRCT–QILD scores) in both analyses (Figure 2). As shown in Supplementary Table 7 (http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract), no proteins that correlated with FVC at the

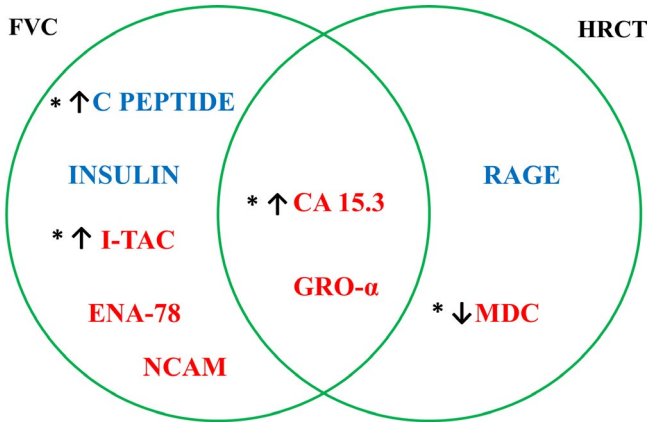

Figure 2. Venn diagram showing serum proteins that correlate with forced vital capacity (FVC) and high-resolution computed tomography quantitative interstitial lung disease (HRCT–QILD) score in participants with diffuse cutaneous systemic sclerosis (dcSSc). Associations with a better FVC and HRCT–QILD score (blue font) and associations with a worse FVC and HRCT–QILD score (red font) are shown. **Asterisks** indicate proteins that were differentially expressed and up-regulated (**up arrows**) or down-regulated (**down arrows**) in participants with dcSSc compared to control subjects. I-TAC = interferon-inducible T cell α chemoattractant; CA 15.3 = cancer antigen 15-3; RAGE = receptor for advanced glycosylation end products; ENA-78 = epithelial-derived neutrophil activating protein 78; GRO-α = growth-regulated α protein; MDC = macrophage-derived chemokine; NCAM = neuronal cell adhesion molecule.

baseline visit predicted the course of FVC 3–14 months after randomization in both treatment arms (n = 32 each for CYC and transplantation arms).

Whole blood gene expression versus serum proteins. After filtering the whole blood gene expression data set by 282 transcripts corresponding to 172 examined proteins, we compared transcript profiles between SSc participants and control subjects. A total of 52 transcripts were differentially expressed (FDR of <0.1) in dcSSc participants versus controls. Among this list, 24 transcripts had a corresponding differentially expressed protein, of which only 17 transcripts (belonging to 14 genes) showed a differential expression in the concordant direction in its corresponding protein (Table 4). Of note, 7 transcripts were differentially expressed in the opposite direction, indicating that the observed differential expression in the serum proteins does not stem from peripheral blood cells. Among concordantly expressed transcript–serum protein pairs, IL-1β was the only one with a decreased number at both whole blood gene expression and serum protein levels whereas the remainder were up-regulated in SSc at both levels. Moreover, endostatin was up-regulated both at RNA transcript and protein levels and positively correlated with the MRSS (Table 3). CA 15-3 was up-regulated both at transcript and serum protein levels and was positively correlated with the presence of lung fibrosis (Supplementary Tables 5 and 6, http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract).

Type I IFN–inducible proteins. Among the 90 differentially expressed proteins in SSc participants, 40 up-regulated molecules are known to be type I IFN–inducible (44.4%) whereas 10 down-regulated proteins were type I IFN–inducible (11.1%), according to the Interferome database search (Supplementary Table 1, http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract). MCP-1, monokine induced by IFNγ (MIG), IFNγ-induced protein 10 (IP-10), IFN-inducible T cell α chemoattractant (I-TAC), as well as CA 15-3 are among these up-regulated type I IFN–inducible proteins (Supplementary Table 8, http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract).

DISCUSSION

The present study represents a large-scale analysis of serum proteins in participants with dcSSc compared at a 1:1 ratio to matched healthy control subjects. Ninety differentially expressed proteins were identified among the 230 assayed by the utilized platform (Supplementary Table 1, http://onlinelibrary.wiley.com/doi/10.1002/art.41570/abstract). Candidate proteins emerged from correlation analysis with the MRSS, FVC, and QILD score. Ingenuity Pathway Analysis revealed fibrosis and extravasation-related pathways as the top overrepresented biologic processes. Lastly, transcripts and proteins showing differential expression at whole blood RNA and serum protein levels were identified, showing that only a small portion of differentially expressed serum proteins were also differentially expressed in a concordant direction in the whole blood RNA samples.

Table 4. Whole blood gene expression versus serum protein expression at baseline*

			Whole blood gene expression			Serum proteins		
Analyte name	Gene name	Direction†	Fold change	P_{raw}	P_{FDR}	Fold change	P_{raw}	P_{FDR}
IL-1β	IL1B	Down-regulated	0.71	<0.001	<0.001	0.89	0.015	0.032
α_1-antitrypsin	SERPINA1	Up-regulated	1.04	0.018	0.092	1.22	<0.001	0.001
MIG	CXCL9	Up-regulated	1.05	0.001	0.010	2.30	<0.001	<0.001
IL-2Rα	IL2RA	Up-regulated	1.05	0.017	0.091	1.97	<0.001	<0.001
Cancer antigen 15-3	MUC1‡	Up-regulated	1.06	<0.001	0.006	1.67	0.001	0.004
Cancer antigen 15-3	MUC1‡	Up-regulated	1.11	<0.001	<0.001	1.67	0.001	0.004
TNF ligand superfamily member 13	TNFSF13‡	Up-regulated	1.08	0.001	0.008	1.23	0.017	0.037
TNF ligand superfamily member 13	TNFSF13‡	Up-regulated	1.1	<0.001	0.005	1.23	0.017	0.037
MCP-1	CCL2	Up-regulated	1.12	<0.001	0.001	2.48	<0.001	<0.001
IL-16	IL16	Up-regulated	1.15	<0.001	0.005	1.08	0.019	0.040
LOX-1	OLR1	Up-regulated	1.15	0.006	0.040	1.38	<0.001	<0.001
Endostatin	COL18A1	Up-regulated	1.16	<0.001	0.001	1.28	<0.001	<0.001
BAFF	TNFSF13B‡	Up-regulated	1.17	0.015	0.081	1.83	<0.001	<0.001
BAFF	TNFSF13B‡	Up-regulated	1.23	0.001	0.009	1.83	<0.001	<0.001
IP-10	CXCL10	Up-regulated	1.26	<0.001	0.001	2.03	<0.001	<0.001
Haptoglobin	HP	Up-regulated	1.29	0.001	0.008	1.71	<0.001	<0.001
MPO	MPO	Up-regulated	1.41	<0.001	0.006	1.84	<0.001	<0.001

* FDR = false discovery rate; IL-1β = interleukin-1β; MIG = monokine induced by interferon-γ; IL-2Rα = IL-2 receptor α; TNF = tumor necrosis factor; MCP-1 = monocyte chemotactic protein 1; LOX-1 = lectin-like oxidized low-density lipoprotein receptor 1; IP-10 = interferon-γ–inducible 10-kd protein; MPO = myeloperoxidase.
† Concordant direction of whole blood RNA expression versus serum protein expression.
‡ Two transcript variants of this gene were differentially expressed.

In the present proteomics analysis, the majority of samples obtained from participants with early-stage dcSSc had a distinct serum protein profile compared to samples obtained from control subjects, confirming the presence of a prominent IFN signature in SSc (5,7,30). Among the 65 serum proteins that were up-regulated, 40 were type I IFN–inducible proteins, with 10 of those being chemokines (MCP-1, macrophage inflammatory protein 1β [MIP-1β], MCP-2, MCP-4, MIP-3β, myeloid progenitor inhibitory factor 1, MIG, IP-10, I-TAC, and B lymphocyte chemoattractant) as well as 30 other proteins, including osteopontin (SPP-1) and β_2-microglobulin. Additionally, several type I IFN–inducible molecules were commonly dysregulated at both the RNA and protein level (*CXCL9*/MIG, *CCL2*/MCP-1, *MUC1*/CA 15-3, *IL16*/IL-16, *CXCL10*/IP-10, *SERPINA1*/α_1-antitrypsin, *TNFSF13*/TNF ligand superfamily, *OLR1*/lectin-like oxidized low-density lipoprotein receptor 1, and *TNFSF13B*/B cell–activating factor). Of note, plasma IP-10 and I-TAC were previously found to be up-regulated in 266 individuals with early-stage SSc enrolled in the GENISOS cohort, and these levels correlated with a peripheral blood cell IFN gene expression score (7). Moreover, in a phase I open-label clinical trial of anifrolumab (an anti–IFNα receptor 1 monoclonal antibody) conducted in 34 patients with SSc, levels of SPP-1 correlated with IFN activity (whole blood type I IFN gene signature score) whereas β_2-microglobulin, IP-10, and MCP-4 were suppressed after treatment with anifrolumab, supporting the notion that these proteins are regulated by type I IFN (16).

In our correlation analysis, serum soluble EGFR showed the strongest negative correlation with the MRSS and was significantly down-regulated in participants with SSc compared to controls. Soluble EGFR can inhibit the activation of its transmembrane receptor by binding EGF or by directly binding the transmembrane receptor itself, which can disrupt EGF/EGFR cell signaling (31). Decreased soluble EGFR in SSc might lead to an activation of EGF pathways, as already described for some subtypes of lung cancers (32). Indeed, a recent multicohort analysis of SSc skin transcriptome data across 7 data sets from 6 centers comprising 515 samples identified 6 signaling proteins which positively correlated with the SSc signature, 4 of which were EGFR ligands (33). Our data provide further support for EGFR signaling as a potential driver of fibrosis in SSc skin. Of note, a correlation between serum soluble EGFR and lung FVC or QILD score was not observed.

In our study, α_1-antichymotrypsin showed the strongest positive correlation with the MRSS. This protein is an acute-phase reactant produced by the liver (34). Its biologic function is to inhibit several serine proteases, mainly cathepsin G, which is contained in the neutrophil granules and released at the site of inflammation. Notably, an excess of cathepsin G function is linked to tissue damage (35). Moreover, 2 proteins mainly associated with SSc vascular manifestation, endostatin and NT-proBNP, also showed a moderately positive correlation with the MRSS. Endostatin is a peptide derived from the C-terminus of type XVIII collagen produced by fibroblasts with antiangiogenic properties. Previous studies have shown that endostatin is up-regulated in SSc serum (36,37). Its antiangiogenic role suggests a feedback loop between endostatin and features of vascular impairment such as digital ulcers, pulmonary arterial hypertension (PAH), and scleroderma renal crisis (38–40). Vascular involvement and extensive skin involvement are not mutually exclusive. In fact, 2 previous studies have shown an association between serum levels of endostatin and more extensive skin involvement (39,41). Of note, endostatin-derived peptides have exhibited antifibrotic properties and were able to prevent and reverse dermal TGFβ–induced fibrosis in both ex vivo human skin and in vivo mouse models (42). Similarly, NT-proBNP has a more established link with vascular abnormalities in SSc, particularly with PAH, cardiac damage, and mortality (43), but previous studies have also shown a positive correlation of NT-proBNP with MRSS scores (44–46).

In our correlative analyses with SSc–ILD features, the availability of both FVC and QILD scores at the baseline visit in all SCOT participants enabled us to identify serum proteins that correlate with functional lung volume, as well as scleroderma-related radiographic findings. Significant correlations were observed in CA 15-3 and growth-regulated α protein (GROα) with both FVC and QILD scores in a clinically concordant direction, but only CA 15-3 also showed significant up-regulation in SSc participants compared to controls. CA 15-3 significantly correlated with lower FVC and higher QILD scores on HRCT. CA 15-3 is a mucin encoded by the gene *MUC1*, which also encodes Krebs von den Lungen 6 protein. CA 15-3 is produced by epithelial cells, including type 2 pneumocytes, and is commonly used as a tumor marker in clinical practice for breast and ovarian cancer (47,48). In a previous retrospective study of 221 individuals with SSc, CA 15-3 was a useful marker in identifying individuals with significant ILD and also correlated with decreased FVC and higher lung fibrosis scores (49). Of note, CA 15-3 did not show a significant positive correlation with the MRSS, underscoring its value as a lung-specific marker. GROα is a neutrophil chemoattractant, and consistent with our results, a previous study indicated that this protein was up-regulated in SSc sera and was associated with lung impairment in SSc, correlating with decreased diffusing capacity for carbon monoxide and FVC (50).

Building on the availability of concomitantly collected serum and whole blood RNA samples, we performed a direct comparison between these 2 sample types, showing that the differential expression for most proteins in SSc serum is most likely to originate outside peripheral blood cells. Our studies focused on serum and peripheral blood RNA, which can be readily obtained and are practical sources of biomarker development during routine clinical care. The 3 overrepresented pathways in SSc serum were exactly the same 3 overrepresented pathways previously identified in our global SSc skin gene expression study (hepatic fibrosis, granulocyte adhesion and diapedesis, and agranulocyte adhesion and diapedesis) (29). To further investigate this

finding, we compared the list of 90 differentially expressed serum proteins with the differentially expressed transcripts in concomitantly collected peripheral blood cell RNA samples. This analysis yielded only 14 molecules that were differentially expressed in both sample types in the concordant direction. There were even 7 molecules that were differentially expressed in the opposite direction. These results support the notion that the source for the majority of differentially expressed serum proteins is likely to be outside of peripheral blood cells.

In line with our findings, a recently published SOMAscan proteome analysis in 2 cohorts of 14 and 20 patients with SSc showed that most of the differentially expressed serum proteins overlapped with serum proteins from 2 previously published SSc skin messenger RNA expression data sets (15). Prominently affected end organs in SSc such as the skin and lungs are potential sources for the SSc serum protein signature, although it is possible that other organs such as the liver are also contributing to the SSc protein profile.

The present study has some weaknesses. It is mainly hypothesis-generating and does not include mechanistic experiments. Moreover, although we used the most comprehensive proteomics MAP panel provided by the Myriad Rules-Based Medicine at the time of study, we cannot provide a comprehensive view of protein dysregulation, in contrast to the findings in genome-wide association and genome-wide gene expression studies, due to the technical limitations of the available proteomics assays. It is likely that a more comprehensive proteomics platform will lead to identification of additional candidate biomarkers. Furthermore, though we had access to concomitantly collected whole blood RNA samples, samples from affected end organs (skin or lung) were not available in the SCOT trial. Moreover, an independent validation cohort was not included in the present study. Future studies are needed to confirm the association of identified serum proteins with SSc fibrotic features.

The present study also has several strengths. To our knowledge, this investigation represents the largest serum protein study in SSc with validated and robust multiplex protein assays. We analyzed a well-characterized subset of dcSSc participants with early progressive fibrotic disease. SCOT participants were matched at a 1:1 ratio for age and sex with control subjects in order to avoid the potential confounding effect of differences in demographic characteristics and to generate sufficient power for identification of differentially expressed proteins. Moreover, the availability of FVC as well as QILD scores on HRCT enabled us to identify serum proteins that correlate with lung function as well as the extent of radiographic involvement. Lastly, to our knowledge, the present study is the first to directly compare a large-scale SSc serum protein profile to the concurrently obtained whole blood transcriptome.

In conclusion, 4 important observations emerged from the present study. Namely, SSc serum samples from SSc participants showed a distinct proteomics profile compared to samples from control subjects, which includes an activation of prominent profibrotic cytokines. Moreover, up-regulation of several type I IFN–inducible proteins was also observed, confirming previous genetic and gene expression studies and demonstrating a prominent IFN signature in SSc. Furthermore, a direct comparison between the serum protein expression profile and peripheral blood gene expression profile indicated that the primary source for the SSc serum proteomics profile lies outside peripheral blood cells. Lastly, we were able to identify serum protein correlates of the MRSS and ILD severity, suggesting that EGFR, α_1-antichymotrypsin, and CA 15-3 are candidate proteins for future mechanistic studies in SSc.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Bellocchi had full access to all the study data and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Bellocchi, Goldmuntz, Keyes-Elstein, McSweeney, Furst, Crofford, Mayes, Sullivan, Assassi.

Acquisition of data. Bellocchi, Goldmuntz, Keyes-Elstein, Varga, Hinchcliff, McSweeney, Furst, Nash, Crofford, Welch, Goldin, Pinckney, Mayes, Sullivan, Assassi.

ADDITIONAL DISCLOSURES

Author Pinckney is an employee of Rho Federal Systems Division.

REFERENCES

1. Varga J, Trojanowska M, Kuwana M. Pathogenesis of systemic sclerosis: recent insights of molecular and cellular mechanisms and therapeutic opportunities. J Scleroderma Relat Disord 2017;2:137–52.
2. Elhai M, Meune C. Trends in mortality in patients with systemic sclerosis over 40 years: a systematic review and meta-analysis of cohort studies. Rheumatology (Oxford) 2012;51:1017–26.
3. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet 2010;42:426–9.
4. Allanore Y, Saad M, Dieudè P, Avouac J, Distler JH, Amouyel P, et al. Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. PLoS Genet 2011;7:e1002091.
5. Assassi S, Mayes MD, Arnett FC, Gourh P, Agarwal SK, Mcnearney TA, et al. Systemic sclerosis and lupus: points in an interferon-mediated continuum. Arthritis Rheum 2010;62:589–98.
6. Whitfield ML, Finlay DR, Murray JI, Troyanskaya OG, Chi JT, Pergamenschikov A, et al. Systemic and cell type-specific gene expression patterns in scleroderma skin. Proc Natl Acad Sci U S A 2003;100:12319–24.
7. Liu X, Mayes MD, Tan FK, Wu M, Reveille JD, Harper BE, et al. Correlation of interferon-inducible chemokine plasma levels with disease severity in systemic sclerosis. Arthritis Rheum 2013;65:226–35.
8. Bandinelli F, del Rosso A, Gabrielli A, Giacomelli R, Bartoli F, Guiducci S, et al. CCL2, CCL3 and CCL5 chemokines in systemic sclerosis: the correlation with SSc clinical features and the effect of prostaglandin E1 treatment. Clin Exp Rheumatol 2012;30:S44–9.

9. Cossu M, Andracco R, Santaniello A, Marchini M, Severino A, Caronni M, et al. Serum levels of vascular dysfunction markers reflect disease severity and stage in systemic sclerosis patients. Rheumatology (Oxford) 2016;55:1112–6.
10. Riccieri V, Stefanantoni K, Vasile M, Macrì V, Sciarra I, Iannace N, et al. Abnormal plasma levels of different angiogenic molecules are associated with different clinical manifestations in patients with systemic sclerosis. Clin Exp Rheumatol 2011;29:S46–52.
11. De Lauretis A, Sestini P, Pantelidis P, Hoyles R, David M, Goh NS, et al. Serum interleukin 6 is predictive of early functional decline and mortality in interstitial lung disease associated with systemic sclerosis. J Rheumatol 2013;40:435–46.
12. Masui Y, Asano Y, Shibata S, Noda S, Aozasa N, Akamata K, et al. Serum adiponectin levels inversely correlate with the activity of progressive skin sclerosis in patients with diffuse cutaneous systemic sclerosis. J Eur Acad Dermatol Venereol 2012;26:354–60.
13. Fantuzzi G. Adiponectin and inflammation: consensus and controversy. J Allergy Clin Immunol 2008;121:326–30.
14. Lakota K, Wei J, Carns M, Hinchcliff M, Lee J, Whitfield ML, et al. Levels of adiponectin, a marker for PPAR-γ activity, correlate with skin fibrosis in systemic sclerosis: potential utility as biomarker? Arthritis Res Ther 2012;14:R102.
15. Rice LM, Mantero JC, Stifano G, Ziemek J, Simms RW, Gordon J, et al. A proteome-derived longitudinal pharmacodynamic biomarker for diffuse systemic sclerosis skin. J Invest Dermatol 2017;137:62–70.
16. Guo X, Higgs BW, Bay-Jensen AC, Karsdal MA, Yao Y, Roskos LK, et al. Suppression of T cell activation and collagen accumulation by an anti-IFNAR1 mAb, anifrolumab, in adult patients with systemic sclerosis. J Invest Dermatol 2015;135:2402–9.
17. Beirne P, Pantelidis P, Charles P, Wells AU, Abraham DJ, Denton CP, et al. Multiplex immune serum biomarker profiling in sarcoidosis and systemic sclerosis. Eur Respir J 2009;34:1376–82.
18. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988;15:202–5.
19. Sullivan KM, Goldmuntz EA, Keyes-Elstein L, McSweeney PA, Pinckney A, Welch B, et al. Myeloablative autologous stem-cell transplantation for severe scleroderma. N Engl J Med 2018; 378:35–47.
20. Kim HJ, Li G, Gjertson D, Elashoff R, Shah SK, Ochs R, et al. Classification of parenchymal abnormality in scleroderma lung using a novel approach to denoise images collected via a multicenter study. Acad Radiol 2008;15:1004–16.
21. Khanna D, Nagaraja V, Tseng C, Abtin F, Suh R, Kim G, et al. Predictors of lung function decline in scleroderma-related interstitial lung disease based on high-resolution computed tomography: implications for cohort enrichment in systemic sclerosis–associated interstitial lung disease trials. Arthritis Res Ther 2015;17:372.
22. Clements P, Lachenbruch P, Siebold J, White B, Weiner S, Martin R, et al. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. J Rheumatol 1995;22:1281–5.
23. Rivnak AJ, Rissin DM, Kan CW, Song L, Fishburn MW, Piech T, et al. A fully-automated, six-plex single molecule immunoassay for measuring cytokines in blood. J Immunol Methods 2015;424:20–7.
24. Assassi S, Wang X, Chen G, Goldmuntz E, Keyes-Elstein L, Ying J, et al. Myeloablation followed by autologous stem cell transplantation normalises systemic sclerosis molecular signatures. Ann Rheum Dis 2019;78:1371–8.
25. Simon R, Lam A, Li MC, Ngan M, Menenzes S, Zhao Y. Analysis of gene expression data using BRB-ArrayTools. Cancer Inform 2007;3:11–7.
26. Rusinova I, Forster S, Yu S, Kannan A, Masse M, Cumming H, et al. Interferome v2.0: an updated database of annotated interferon-regulated genes. Nucleic Acids Res 2013;41:1040–6.
27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995;57:289–300.
28. Krämer A, Green J, Pollard J, Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics 2014;30: 523–30.
29. Assassi S, Swindell WR, Wu M, Tan FD, Khanna D, Furst DE, et al. Dissecting the heterogeneity of skin gene expression patterns in systemic sclerosis. Arthritis Rheumatol 2015;67:3016–26.
30. López-Isac E, Martín JE, Assassi S, Simeón CP, Carreira P, Ortego-Centeno N, et al. IRF4 newly identified as a common susceptibility locus for systemic sclerosis and rheumatoid arthritis in a cross-disease meta-analysis of genome-wide association studies. Arthritis Rheumatol 2016;68:2338–44.
31. Basu A, Raghunath M, Bishayee S, Das M. Inhibition of tyrosine kinase activity of the epidermal growth factor (EGF) receptor by a truncated receptor form that binds to EGF: role for interreceptor interaction in kinase regulation. Mol Cell Biol 1989;9:671–7.
32. Lococo F, Paci M, Rapicetta C, Rossi T, Sancisi V, Braglia L, et al. Preliminary evidence on the diagnostic and molecular role of circulating soluble EGFR in non-small cell lung cancer. Int J Mol Sci 2015;16:19612–30.
33. Lofgren S, Hinchcliff M, Carns M, Wood T, Aren K, Arroyo E, et al. Integrated, multicohort analysis of systemic sclerosis identifies robust transcriptional signature of disease severity. JCI Insight 2016;1:e89073.
34. Bode JG, Albrecht U, Häussinger D, Heinrich PC, Schaper F. Hepatic acute phase proteins–regulation by IL-6- and IL-1-type cytokines involving STAT3 and its crosstalk with NF-κB-dependent signaling [review]. Eur J Cell Biol 2012;91:496–505.
35. Baker C, Belbin O, Kalsheker N, Morgan K. SERPINA3 (aka α-1-antichymotrypsin). Front Biosci 2007;12:2821–35.
36. Sedie AD, Riente L, Maggiorini L, Pratesi F, Tavoni A, Migliorini P, et al. Potential biomarkers in patients with systemic sclerosis. Int J Rheum Dis 2018;21:261–5.
37. Distler O, del Rosso A, Giacomelli R, Cipriani P, Conforti ML, Guiducci S, et al. Angiogenic and angiostatic factors in systemic sclerosis: increased levels of vascular endothelial growth factor are a feature of the earliest disease stages and are associated with the absence of fingertip ulcers. Arthritis Res 2002;4:R11.
38. Reiseter S, Molberg Ø, Gunnarsson R, Lund MB, Aalokken TM, Aukrust P, et al. Associations between circulating endostatin levels and vascular organ damage in systemic sclerosis and mixed connective tissue disease: an observational study. Arthritis Res Ther 2015;17:231.
39. Hebbar M, Peyrat JP, Hornez L, Hatron PY, Hachulla E, Devulder B. Increased concentrations of the circulating angiogenesis inhibitor endostatin in patients with systemic sclerosis. Arthritis Rheum 2000;43:889–93.
40. Mecoli CA, Shah AA, Boin F, Wigley FM, Hummers LK. The utility of plasma vascular biomarkers in systemic sclerosis: a prospective longitudinal analysis. Arthritis Rheumatol 2020;72:1341–9.
41. Farouk HM, Hamza SH, El Bakry SA, Youssef SS, Aly IM, Moustafa AA, et al. Dysregulation of angiogenic homeostasis in systemic sclerosis. Int J Rheum Dis 2013;16:448–54.
42. Yamaguchi Y, Takihara T, Chambers RA, Veraldi KL, Larregina AT, Feghali-Bostwick CA. A peptide derived from endostatin ameliorates organ fibrosis. Sci Transl Med 2012;4:136ra71.
43. Allanore Y, Komocsi A, Vettori S, Hachulla E, Hunzelmann N, Distler J, et al. N-terminal pro-brain natriuretic peptide is a strong predictor of mortality in systemic sclerosis. Int J Cardiol 2016;223:385–9.
44. Carlo-Stella N, Belloli L, Biondi ML, Marasini B. Serum N-terminal pro-brain natriuretic peptide, a marker of skin thickness in systemic sclerosis? [letter]. Clin Rheumatol 2009;28:241–2.
45. Choi HJ, Shin YK, Lee HJ, Kee JY, Shin DW, Lee EY, et al. The clinical significance of serum N-terminal pro-brain natriuretic peptide in systemic sclerosis patients. Clin Rheumatol 2008;27:437–42.

46. Elshamy HA, Ibrahim SE, Farouk HM, Moustafa AA, Aly IM, Osman WM. N-terminal pro-brain natriuretic peptide in systemic sclerosis: new insights. Eur J Dermatol 2011;21:686–90.
47. Apostolopoulos V, Stojanovska L, Gargosky SE. MUC1 (CD227): a multi-tasked molecule. Cell Mol Life Sci 2015;72:4475–500.
48. Lakshmanan I, Ponnusamy MP, Macha MA, Haridas D, Majhi PD, Kaur S, et al. Mucins in lung cancer: diagnostic, prognostic, and therapeutic implications. J Thorac Oncol 2015;10:19–27.
49. Celeste S, Santaniello A, Caronni M, Franchi J, Severino A, Scorza R, et al. Carbohydrate antigen 15.3 as a serum biomarker of interstitial lung disease in systemic sclerosis patients. Eur J Intern Med 2013;24:671–6.
50. Furuse S, Fujii H, Kaburagi Y, Fujimoto M, Hasegawa M, Takehara K, et al. Serum concentrations of the CXC chemokines interleukin 8 and growth-regulated oncogene-α are elevated in patients with systemic sclerosis. J Rheumatol 2003;30:1524–8.