

Integrated, multicohort analysis of systemic sclerosis identifies robust transcriptional signature of disease severity

Shane Lofgren,^{1,2} Monique Hinchcliff,³ Mary Carns,³ Tammara Wood,⁴ Kathleen Aren,³ Esperanza Arroyo,³ Peggie Cheung,⁵ Alex Kuo,⁵ Antonia Valenzuela,⁵ Anna Haemel,⁶ Paul J. Wolters,⁷ Jessica Gordon,⁸ Robert Spiera,⁸ Shervin Assassi,⁹ Francesco Boin,¹⁰ Lorinda Chung,^{5,11,12} David Fiorentino,¹² Paul J. Utz,^{1,5} Michael L. Whitfield,⁴ and Purvesh Khatri^{1,2}

¹Institute for Immunity, Transplantation, and Infection, ²Division of Biomedical Informatics Research, Department of Medicine, Stanford University, California, USA. ³Division of Rheumatology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA. ⁴Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA. ⁵Division of Rheumatology, Department of Medicine, Stanford University School of Medicine, Stanford, California, USA. ⁶Department of Dermatology, ⁷Pulmonary Division, Department of Medicine, University of California, San Francisco, California, USA. ⁸Department of Rheumatology, Hospital for Special Surgery, New York, New York, USA. ⁹Division of Rheumatology and Clinical Immunogenetics, The University of Texas Health Science Center Houston, Houston, Texas, USA. ¹⁰Division of Rheumatology, Department of Medicine, University of California, San Francisco, California, USA. ¹¹Department of Veterans Affairs Palo Alto Health Care System, Palo Alto, California, USA. ¹²Department of Dermatology, Stanford University School of Medicine, Stanford, California, USA.

Systemic sclerosis (SSc) is a rare autoimmune disease with the highest case-fatality rate of all connective tissue diseases. Current efforts to determine patient response to a given treatment using the modified Rodnan skin score (mRSS) are complicated by interclinician variability, confounding, and the time required between sequential mRSS measurements to observe meaningful change. There is an unmet critical need for an objective metric of SSc disease severity. Here, we performed an integrated, multicohort analysis of SSc transcriptome data across 7 datasets from 6 centers composed of 515 samples. Using 158 skin samples from SSc patients and healthy controls recruited at 2 centers as a discovery cohort, we identified a 415-gene expression signature specific for SSc, and validated its ability to distinguish SSc patients from healthy controls in an additional 357 skin samples from 5 independent cohorts. Next, we defined the SSc skin severity score (4S). In every SSc cohort of skin biopsy samples analyzed in our study, 4S correlated significantly with mRSS, allowing objective quantification of SSc disease severity. Using transcriptome data from the largest longitudinal trial of SSc patients to date, we showed that 4S allowed us to objectively monitor individual SSc patients over time, as (a) the change in 4S of a patient is significantly correlated with change in the mRSS, and (b) the change in 4S at 12 months of treatment could predict the change in mRSS at 24 months. Our results suggest that 4S could be used to distinguish treatment responders from nonresponders prior to mRSS change. Our results demonstrate the potential clinical utility of a novel robust molecular signature and a computational approach to SSc disease severity quantification.

Authorship note: S. Lofgren and M. Hinchcliff contributed equally to the manuscript.

Conflict of interest: The authors have declared that no conflict of interest exists.

Submitted: June 15, 2016

Accepted: November 15, 2016

Published: December 22, 2016

Reference information:

JCI Insight. 2016;1(21):e89073.
doi:10.1172/jci.insight.89073.

Introduction

Systemic sclerosis (SSc, scleroderma) is a rare autoimmune disease whose outward clinical manifestation is skin fibrosis. SSc is associated with high disease-related mortality with no known cures (1). Although SSc patients with progressive skin disease are currently prescribed many treatments, including immunosuppressants such as cyclophosphamide, methotrexate, and mycophenolate mofetil (MMF), only a subset of patients appears to improve during therapy (2). Changes in the modified Rodnan skin score (mRSS), generated by pinching and scoring skin fibrosis over 17 body areas (0 = no fibrosis to 3 = severe fibrosis; 0–51 scale), is

the most commonly used outcome measure. Although validated, the mRSS is problematic and subjective because it requires a long duration between measurements to assess interval change, and is confounded by obesity, edema, and interclinician variance. Although a number of previous studies have identified genomic biomarkers of skin disease (3–5), these studies have been performed on small numbers of patients and the biomarkers have not been robustly validated in independent cohorts from multiple clinical centers.

A growing body of evidence supports the use of skin, lung, and peripheral blood transcriptomes as SSc biomarkers to quantify interpatient heterogeneity, to identify dysregulated molecular pathways underlying disease, and to identify appropriate patients for specific SSc therapies (6–9). These studies have identified remarkable homogeneity in skin biopsies obtained from the same patient such that transcriptomic profiles of clinically unaffected back and clinically affected forearm biopsies are nearly identical (10–12). Based on the molecular heterogeneity observed in skin biopsies from SSc patients, 4 SSc subtypes, called intrinsic subsets, have been proposed: (a) normal-like, (b) limited, (c) fibroproliferative, and (d) inflammatory (10, 11).

One of the important features of these studies is that they are controlled experiments, which aim to reduce the effect of various biological and technical confounding factors as much as possible. For instance, strict inclusion and exclusion criteria restrict study participation to reduce the number and effects of confounding factors (8, 10). However, this controlled experimental approach typically does not capture the heterogeneity observed in the real-world SSc patient population. The analysis of individual carefully controlled datasets increases the risk of transcriptome profiles being dominated by tissue- or treatment-specific expression or confounded by other unknown biological and/or technical factors, all of which could obscure the pathways that may be common across phenotypically diverse SSc patients.

We have previously described an integrated, multicohort analysis framework that leverages the heterogeneity present in public data to increase reproducibility (13). We have repeatedly demonstrated the utility of this approach in identifying novel drug targets, diagnostic and prognostic biomarkers, and repurposing FDA-approved drugs in a broad spectrum of diseases including organ transplant, cancer, sepsis, and bacterial and viral infections (14–21). Here, we applied our multicohort analytical method to 2 SSc gene expression datasets, obtained from 158 skin biopsies from SSc patients, referred to as the UCSF1 cohort (GSE9285) (11) and the Boston cohort (GSE32413) (10) to identify a 415-gene signature. We validated this signature in 5 independent cohorts of skin biopsies from SSc patients in 5 clinical centers, and showed that the signature is not confounded by (a) intrinsic subset or (b) drug treatment (patients received MMF, nilotinib, or rituximab at different clinical centers). Further, we demonstrated across all skin biopsy cohorts that the 415-gene signature correlated with SSc severity as defined by mRSS. To test the diagnostic power of this signature, we collected a cohort, referred to as the Northwestern cohort, of 22 healthy controls and 67 SSc patients who underwent 183 serial skin biopsies at 5 time points over 3 years. We showed that longitudinal change in the 415-gene expression signature is strongly associated with longitudinal change in mRSS for each patient, and thus the 415-gene signature could enable real-time monitoring of SSc patients' responses to treatment. Finally, in a subset of patients in the Northwestern cohort, we showed that changes in the 415-gene signature predict subsequent changes in mRSS. The 415-gene signature is thus an objective and robustly validated signature of disease severity quantification.

Results

Integrated, multicohort analysis of SSc skin samples identifies 415 differentially expressed genes. We identified 4 clinical microarray gene expression datasets from the NCBI's Gene Expression Omnibus (GEO) database comprising 282 samples obtained from skin (Figure 1 and Supplemental Table 1; supplemental material available online with this article; doi:10.1172/jci.insight.89073DS1) (7, 10–12), which were derived from healthy control subjects, patients with diffuse cutaneous SSc or limited cutaneous SSc, and with various disease duration. Two of these datasets profiled longitudinally collected samples from study participants during treatment with rituximab (10) and nilotinib (7). Furthermore, we generated whole-genome transcriptional profiles for 3 additional cohorts, referred to as the Northwestern cohort, the UCSF2 cohort, and the Stanford cohort. Supplemental methods provide further description of each dataset.

We hypothesized that despite the clinical and molecular heterogeneity observed in skin biopsies between SSc patients, the set of genes that are significantly differentially expressed across multiple cohorts compared with healthy individuals would constitute a robust signature of SSc skin disease. We chose 2 independent cohorts, referred to as the UCSF1 cohort (GSE9285) and the Boston cohort (GSE32413), as our discovery cohorts (10, 11) because they were obtained at different centers, and

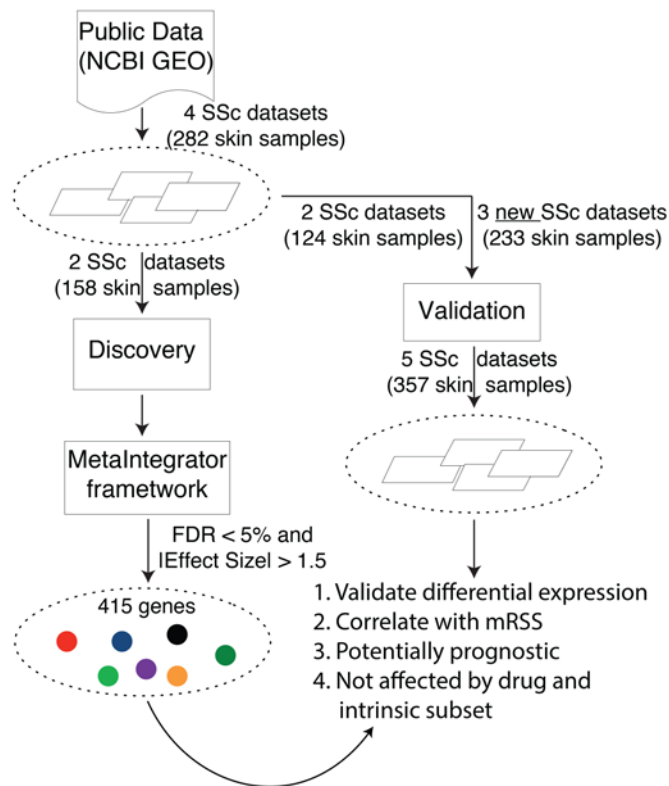


Figure 1. Workflow for integrated multiple-cohort analysis of systemic sclerosis. mRSS, modified Rodnan skin score; SSc, systemic sclerosis.

contained SSc samples across all intrinsic subsets. Hence, they represented known molecular heterogeneity of SSc patients. Our goal was to identify differentially expressed genes in SSc, irrespective of whether they are differentially expressed in other diseases or not. Therefore, we excluded 6 biopsies from patients with morphea or eosinophilic fasciitis in the discovery cohorts because these are skin diseases related to, but distinct from, SSc. Further, because forearm and back biopsies from SSc patients have been repeatedly shown to be molecularly indistinguishable from each other (8, 10, 11), we used both forearm and back skin biopsies from SSc patients as cases. Overall, we utilized expression data from 158 out of 164 skin biopsies from the 2 discovery cohorts (Supplemental Table 1).

We identified 415 differentially expressed genes (211 overexpressed, 204 underexpressed) in SSc patients compared with healthy controls using the following criteria: |Effect size| greater than or equal to 1.5, FDR less than 5% (Figure 2A and Supplemental Table 2). We applied the same multicohort analysis framework to 4 independent validation cohorts (the Houston cohort, the Northwestern cohort, the UCSF2 cohort, and the Stanford cohort). The Hospital for Special Surgery (HSS) cohort lacked healthy controls, and was excluded from validation and

other analyses that required healthy controls. Out of 415 differentially expressed genes identified in the discovery cohorts, 186 genes (120 overexpressed, 66 underexpressed) were still differentially expressed in the validation cohorts (FDR < 5%; Figure 2B).

Functional characterization of the 415-gene SSc signature. We applied 3 approaches to functionally characterize the 415-gene signature. First, we utilized data from the NIH LINCS Program, which quantifies the effects of stimulation with each of the 314 extracellular signaling proteins on the transcriptome across at least 18 cell lines. Our goal was to find perturbations that most resembled our SSc signature to elucidate the signaling pathways that might be driving it in patients. We correlated the effect size of the 415-gene signature with the expression profiles of the 314 extracellular signaling proteins. We identified 6 positively correlated (FDR \leq 5%) signaling proteins. There was significant enrichment for *EGFR* ligands (4 out of 11 known *EGFR* ligands — *HBEGF*, *HRG*, *EPR*, *BTC*; hypergeometric $P = 1.2 \times 10^{-5}$; Supplemental Table 3). Furthermore, *EGF* was also correlated with the 415-gene set, although the correlation was marginally significant after multiple hypothesis correction ($P = 3.9 \times 10^{-3}$; FDR = 7%).

Second, we compared the 415-gene set with expression profiles of 2,473 genes that are individually overexpressed across different cell lines in LINCS. Overexpression of 10 genes correlated (FDR < 5%; Supplemental Table 4) with the 415-gene set. One gene, *SMAD4*, is an intracellular effector of TGF- β signaling. Elevated nuclear localization of *SMAD4* in primary dermal fibroblasts from SSc patients independent of TGF- β receptor signaling has been observed previously (22).

Finally, to learn more about the processes that were being upregulated in SSc, we performed analysis of the 211 overexpressed genes using Enrichr (23) for the KEGG pathways. This analysis identified 21 significant pathways, including focal adhesion, regulation of actin cytoskeleton, chemokine signaling pathway, and NF- κ B signaling as the top 4 most significantly enriched (FDR \leq 5%; Supplemental Table 5). Importantly, the PI3K/Akt signaling pathway, which is downstream of EGFR signaling, was strongly enriched (FDR = 17%). These results, combined with the results from LINCS, suggest that EGFR signaling through PI3K/Akt may be a significantly dysregulated pathway in SSc patients.

The 415-gene set distinguishes SSc from healthy skin. We performed a principal component analysis (PCA) on both discovery cohorts together using the 415-gene set. The first principal component (PC1) explained 20.7% of variance in the discovery cohorts, and represented differences between the 2 discovery datasets (Supplemental Figure 1A). Our goal was to assess the robustness of the PCs that distinguish SSc patients

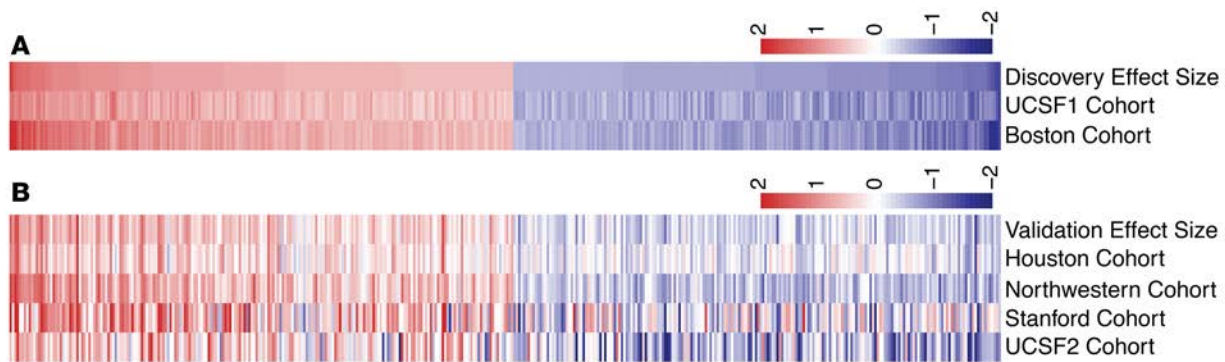


Figure 2. Discovery and validation of robust 415-gene systemic sclerosis signature. Effect-size heatmaps of 415-gene systemic sclerosis signature in (A) 2 discovery and (B) 4 validation cohorts. Each column is a gene and each row is a cohort. The first row in each heatmap shows the summary effect size for each gene in the discovery or validation cohorts. Genes in both heatmaps are sorted in decreasing order of their summary effect size in discovery cohorts. See Supplemental Table 1 for the number of case and control samples in each group.

from healthy subjects. Therefore, we did not analyze PC1 further to understand the differences between the 2 cohorts. The next 3 PCs (PC2, PC3, and PC4) explained 15.2%, 6.4%, and 4.3% of variance, respectively, and correctly separated SSc patients from healthy control subjects in the discovery cohorts (Supplemental Figure 1B). Next, we transformed the Northwestern cohort using the PCA loadings from the discovery cohorts. PC2, PC3, and PC4 continued to distinguish healthy controls and SSc samples in the Northwestern cohort (Supplemental Figure 1C). Collectively, these results demonstrate that the 415-gene set robustly distinguishes SSc patients from healthy control subjects across independent heterogeneous cohorts of skin biopsies.

The 415-gene signature permits molecular quantification of skin score (mRSS). Qualitative exploratory analysis of PCs demonstrated that patients with low mRSS are closer to the cluster of healthy control subjects compared with patients with higher mRSS (Supplemental Figure 1, A–C). Therefore, we defined a centroid of healthy subjects in the discovery cohorts as the median of their PC2, PC3, and PC4 coordinates. We defined a potentially novel metric, distance to health (DTH), as the Euclidean distance between an SSc biopsy and this centroid. There was a significant correlation between the mRSS and DTH in both discovery cohorts ($r = 0.340$ and $P = 0.012$ in the UCSF1 cohort; $r = 0.402$ and $P = 4.7 \times 10^{-4}$ in the Boston cohort; Supplemental Figure 2, A and B, additional summary statistics in Supplemental Table 6). The DTH also had high correlation with mRSS in 5 independent validation cohorts ($r = 0.502$ and $P = 2.4 \times 10^{-5}$ in the Houston cohort; $r = 0.672$ and $P < 2 \times 10^{-16}$ in the Northwestern cohort; $r = 0.712$ and $P = 1.4 \times 10^{-2}$ in the Stanford cohort; $r = 0.440$ and $P = 1.70 \times 10^{-2}$ in the UCSF2 cohort; $r = 0.489$ and $P = 2.1 \times 10^{-2}$ in the HSS cohort; Supplemental Figure 2, C–G). Collectively, our results showed a strong correlation between the DTH and SSc skin disease severity, as measured by mRSS, across 7 independent cohorts composed of 417 SSc skin samples from 224 SSc patients from 6 clinical centers.

The 415-gene signature permits precise objective monitoring of skin disease during treatment. The consistently high positive correlation between DTH and mRSS in SSc skin samples strongly suggested that in a given patient, the DTH would decrease as mRSS declines. We hypothesized that DTH would permit more precise and objective real-time monitoring of SSc skin disease in an individual patient.

We compared DTH and longitudinal mRSS for individual SSc patients in the Northwestern cohort, in which 43 out of 67 enrolled patients prospectively underwent skin biopsies from the clinically involved arm between 2008 and 2015. For these patients, we collected a biopsy prior to treatment initiation, and 6, 12, 24, and 36 or 48 months during or after treatment ($n = 135$ biopsies). Furthermore, we included 22 healthy patients (22 biopsies) as controls and 24 SSc registry patients (26 biopsies) who agreed to undergo skin biopsies but were not enrolled in the clinical study. Overall, we profiled 183 biopsies in the Northwestern cohort.

There was a significant correlation between the change in DTH (Δ DTH) and change in mRSS (Δ mRSS) from baseline to the last study visit ($r = 0.474$, $P = 0.011$; Supplemental Figure 3), demonstrating that DTH changed in the same direction as mRSS across all patients. Furthermore, in nearly every patient, directional change in the DTH (increase or decrease toward healthy centroid) between time points mirrored mRSS directional change. For instance, patient 4 had an mRSS of 9 at baseline, and 3 at 6, 12, and 24 months. The DTH for patient 4 decreased at 6 months when mRSS decreased from 9 to 3, and remained virtually

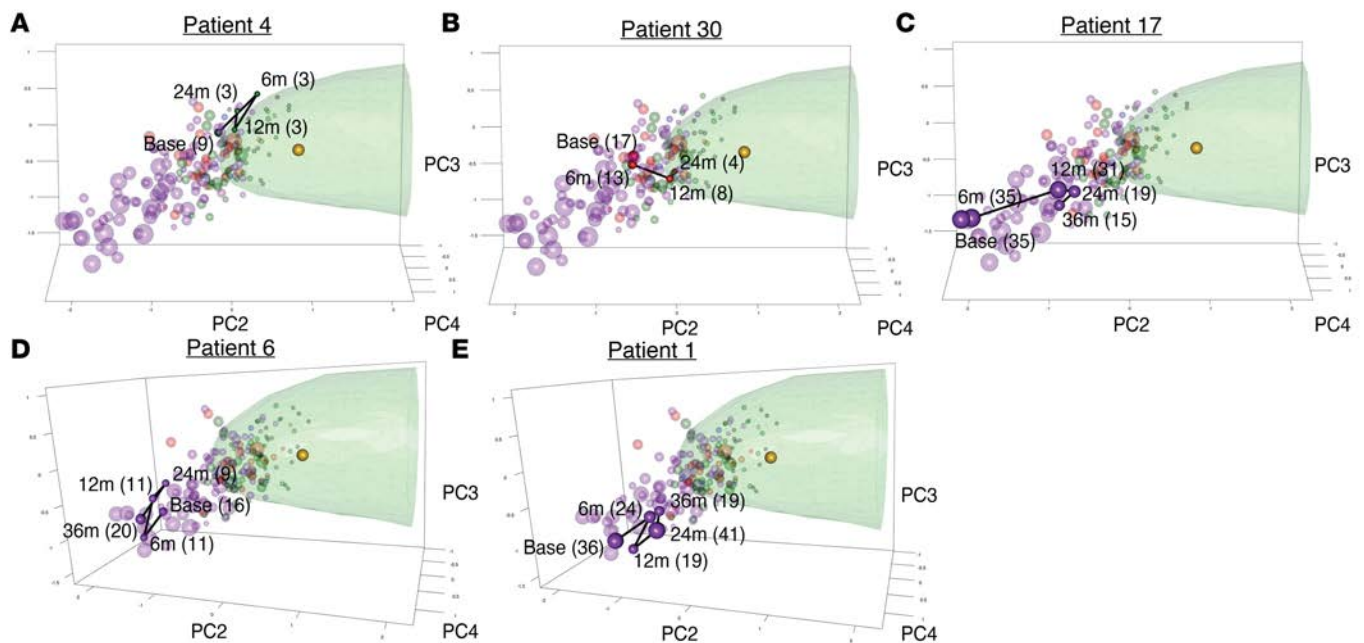


Figure 3. Principal component analysis of 415-gene signature allows monitoring individual patients during mycophenolate mofetil treatment. Each sphere represents a skin biopsy from the Northwestern cohort, where the size of a sphere size is proportional to the modified Rodnan skin score (mRSS) and color indicates its assigned intrinsic subset (black, healthy control; green, normal-like; blue, limited; red, fibroproliferative; purple, inflammatory). The green region represents the health bubble. The golden sphere indicates the centroid of healthy controls. Numbers in brackets represent the mRSS score at a given time point for a given patient. Each panel in **A** through **E** ($n = 183$ for each) displays the same set of samples with different patient trajectory highlighted. m, months; PC, principal component.

the same, as mRSS persisted at 3 (Figure 3A). Similarly, patient 30 had an mRSS of 17, 13, 8, and 4 at baseline, 6, 12, and 24 months, respectively. The DTH for patient 30 was similar at baseline and 6 months, but decreased significantly as mRSS decreased from 13 to 8 at 12 months, and continued to decrease as mRSS decreased from 8 to 4 at 24 months (Figure 3B). Importantly, in patients whose disease improved and then worsened again, we were able to track changes in their disease severity. For example, patient 6 initially improved from an mRSS of 16 at baseline, to 11, 11, and 9 at 6, 12, and 24 months, respectively, during which time their DTH also decreased. At 36 months, the mRSS worsened again to 20, and the patient had increased DTH that was almost equal to DTH at baseline (Figure 3C). Patient 17 had an mRSS of 35, 35, 31, 19, and 15 at baseline, 6, 12, 24, and 48 months, respectively. The DTH for patient 17 continued to decrease concomitantly with mRSS (Figure 3D). It is interesting to note that, for patient 17, although mRSS only decreased by 4 from baseline to 12 months, the DTH decreased by 42%. This large decrease in DTH was reflected in a similarly large reduction in mRSS at the next time point, when mRSS decreased by 12 from 12 months to 24 months. A similar trend was also observed in patient 1, who was enrolled in the study following stem cell transplant, and had an mRSS of 36, 24, 19, 41, and 19 at baseline, 6, 12, 24, and 36 months, respectively (Figure 3E; Supplemental Figure 4). At first glance for patient 1, the DTH increased when mRSS decreased from 24 to 19, and decreased when mRSS increased from 19 to 41. However, further analysis showed that in patient 1 the change in DTH for a given time point was prognostic of change in mRSS for the next time point. This prognostic trend in the DTH compared with mRSS is further analyzed below.

Transforming DTH into a clinically useable score. PCs are difficult to implement into clinical practice owing to variance introduced by different treatment and technology protocols in individual PCs. Therefore, similar to our previous results (14, 17, 19, 20), we defined SSc skin severity score (4S) for a skin biopsy as the difference between the mean of overexpressed genes and mean of underexpressed genes in the 415-gene set. 4S distinguished SSc skin samples from healthy skin biopsies with very high accuracy in both discovery and validation cohorts (range = 0.88–1 in validation cohorts; Supplemental Figure 5, A and B). Further, similar to DTH, 4S was positively correlated with mRSS in discovery cohorts ($r = 0.69$ in UCSF1 and $r = 0.42$ in

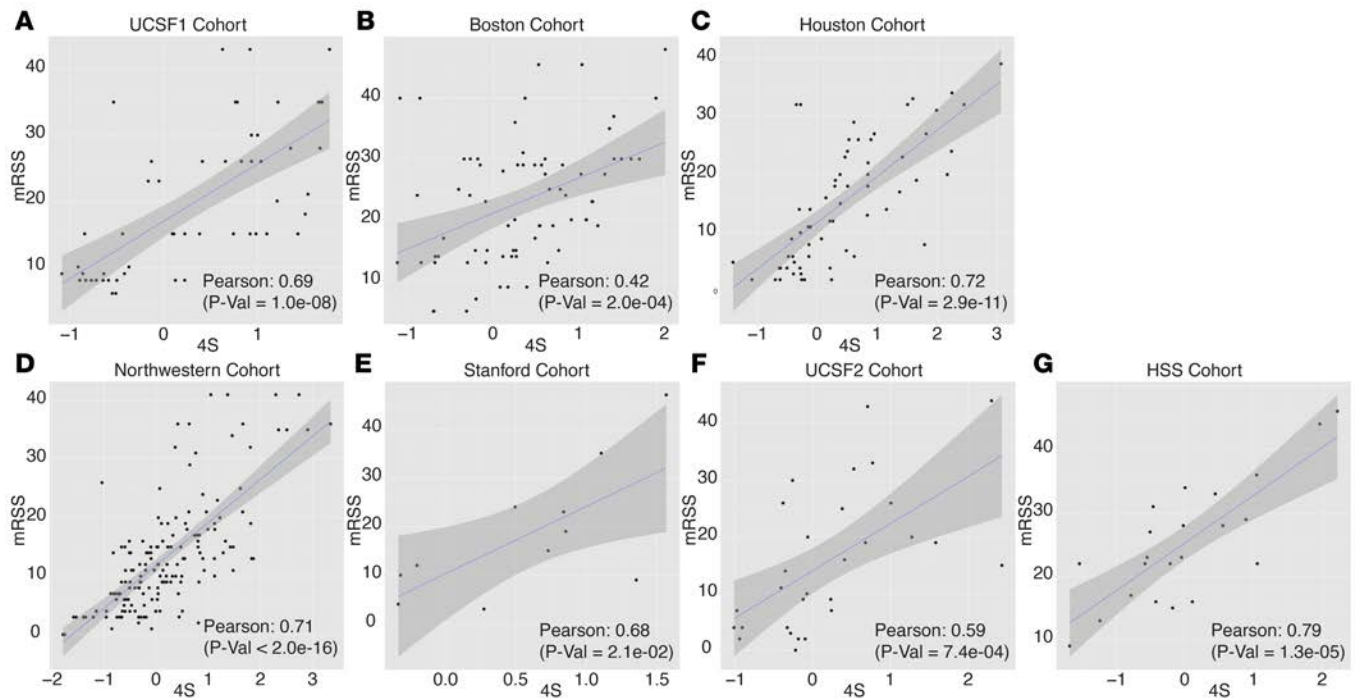


Figure 4. 4S is highly correlated with modified Rodnan skin score (mRSS) across all datasets. Correlation plots comparing mRSS with the systemic sclerosis skin severity score (4S) for all of the datasets. Panels **A** and **B** ($n = 54$ and 72 , respectively) are discovery cohorts and panels **C–G** ($n = 66, 161, 13, 29$, and 22 , respectively) are validation cohorts. Each data point represents a patient skin biopsy. The blue line is the line of best fit and the gray region represents its 95% CI. HSS, Hospital for Special Surgery.

the Boston cohort; $P < 0.0002$; Figure 4, A and B), and 5 validation cohorts ($r = 0.72, 0.71, 0.68, 0.59$, and 0.79 in the Houston, Northwestern, Stanford, UCSF2, and HSS cohorts, respectively, $P < 0.02$; Figure 4, C–G), demonstrating that 4S also maintained the strong correlation observed between the DTH and mRSS. We further explored whether the association between 4S and mRSS is confounded by any demographic or clinical variable in each of the 6 skin biopsy cohorts including age, sex, disease duration, ethnicity, antibody positivity (anti-centromere [ACA], anti-nuclear [ANA], anti-RNA pol III, and anti-Scl-70 antibodies), and percentage predicted lung forced vital capacity (FVC) and diffusion capacity (DLCO) values. There were no consistent differences or correlations with any of these variables for 4S (Supplemental Table 7).

Collectively, these results demonstrate the potential clinical utility of our validated 415-gene signature as a quantitative and unbiased clinical biomarker for SSc skin disease severity, which unlike mRSS, is not confounded by factors such as obesity, edema, drug treatment, or inter- and intrarelator variability as well as other demographic and clinical variables, and has the potential to enable precise quantification of molecular SSc skin disease severity for individual patients.

4S is a prognostic marker of change in mRSS. The significant reduction in the DTH prior to reduction in mRSS in individual SSc patients (Figure 3, C and D) and high correlation between the DTH and 4S (Supplemental Figure 6) suggested that a change in 4S may be prognostic of future change in mRSS. Therefore, in the 19 patients that had biopsies performed at baseline, 12 months, and 24 months, we correlated the change in 4S from baseline to 12 months with the subsequent change in mRSS from 12 months to 24 months. We found a significant correlation ($r = 0.56, P = 0.013$; Figure 5). These results suggest that 4S could be used as a prognostic marker to allow monitoring individual SSc patients, although further validation is needed.

Comparison of 4S and DTH with intrinsic subsets. Several studies using whole-genome expression profiles of SSc skin biopsies have demonstrated the molecular heterogeneity between patients with SSc (10–12). Based on this heterogeneity, SSc patients can be classified by their intrinsic subset (10, 11). Therefore, we explored whether there is any relationship between 4S and the intrinsic subsets.

The intrinsic subset had been determined for patients in 3 independent cohorts (UCSF1, Boston, and Northwestern cohorts). Across these 3 cohorts, normal-like SSc patients had the lowest 4S ($P < 0.05$), whereas fibroproliferative and inflammatory subsets had the highest 4S (Figure 6). There was no difference in 4S

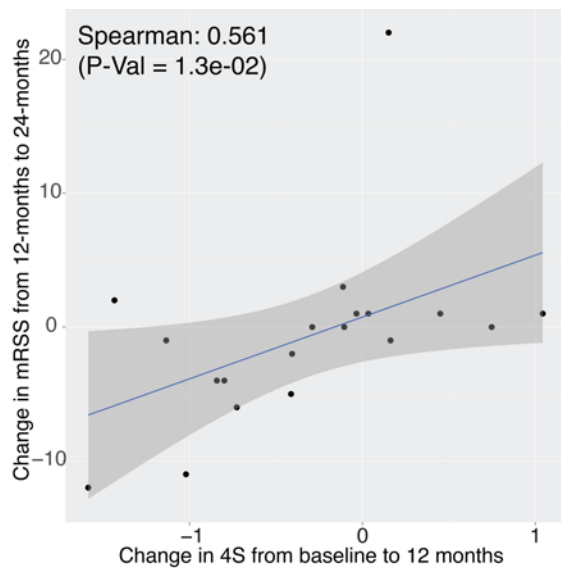


Figure 5. Change in 4S predicts subsequent change in modified Rodnan skin score (mRSS). The x axis represents the change in the systemic sclerosis skin severity score (4S) from baseline to 12 months, whereas the y axis represents the change in mRSS from 12 months to 24 months in the Northwestern cohort ($n = 19$). Each data point represents a patient. The blue line is the line of best fit and the gray region represents its 95% CI.

between fibroproliferative and inflammatory intrinsic subsets except in the Boston cohort. Inflammatory SSc patients showed a wide range of 4S that completely overlapped with fibroproliferative SSc patients. Similar results were also observed when comparing the DTH with intrinsic subsets (Supplemental Figure 7).

Collectively, our results show that despite the repeatedly observed clinical and molecular heterogeneity in SSc patient populations, 4S robustly and reproducibly distinguishes healthy and diseased biopsies and correlates with SSc disease severity as measured by mRSS. It may also be prognostic of future changes in disease severity. However, unlike mRSS, 4S is not susceptible to a number of clinical confounding factors. Furthermore, the distribution of 4S across SSc intrinsic subsets (lower scores in normal-like SSc patients, and no difference in 4S between fibroproliferative and inflammatory SSc patients) suggests that SSc patients may fall along a continuum defined by 4S.

Discussion

The primary goal of our multicohort analysis was to integrate gene expression data from multiple heterogeneous datasets to define a conserved, quantitative, and robust signature of SSc despite the significant molecular heterogeneity observed in the SSc patient population. Such a signature will lay an important foundation for future studies designed to identify novel treatments. We have previously demonstrated the utility of leveraging biological and technological heterogeneity in publicly available data for identifying robust and reproducible signatures of various diseases including organ transplant, sepsis, viral infection, and tuberculosis (14–17, 19, 20). However, these diseases are more prevalent, and corresponding analyses used hundreds of samples in discovery cohorts. In contrast, SSc is a rare disease with publicly available data from a small number of skin biopsies and blood samples. Despite the significantly smaller number of discovery cohorts and biopsies (158 biopsies in 2 cohorts) compared with our previous work, we identified a 415-gene signature composed of differentially expressed genes in skin biopsies from SSc patients compared with healthy control subjects, which we validated in 357 skin biopsies from 5 independent SSc cohorts. We further defined a disease severity measure called 4S using the signature. We showed that 4S significantly correlated with SSc severity as measured by mRSS across all skin biopsy datasets, irrespective of treatment, and is not confounded by clinical factors (obesity, edema, experience performing mRSS). Changes in 4S between baseline and 12 months showed significant correlation with subsequent changes in skin disease severity out to 24 months. Longitudinal studies are needed to further validate this finding. However, if the 4S prognostic ability is validated, it will enable clinicians to identify treatment response earlier and change therapy in nonresponders to avoid unnecessary exposure to ineffective treatment. Furthermore, it may also be useful in clinical trials, particularly in proof-of-concept (biological or clinical) investigations, where novel drugs are administered only for short intervals (e.g., 6–8 weeks) and the changes in mRSS are not able to show a clear treatment effect.

Given the repeatedly observed clinical and molecular heterogeneity in SSc patients across multiple independent cohorts (8, 10–12), the robustness of the SSc signatures identified in our analysis is unexpected and significant. Until now, the lack of such a robust signature has hampered development of novel diagnostics and prognostics for monitoring SSc patients in clinics. It should be possible in future studies to reduce the 415-gene signature to a smaller gene set that is clinically possible to measure. Further, owing to this molecular heterogeneity, SSc patients have been grouped into intrinsic subsets that are stable in serial skin biopsies (10). In contrast, across all cohorts, our results show that 4S is significantly positively correlated with mRSS. The DTH, 4S, and mRSS have significantly increasing trends with the intrinsic subsets, such that the normal-like group has the lowest mean for each measure and the inflammatory group has the highest mean for each measure. PCA of these samples showed that the lower 4S for normal-like samples reflected the fact that they were significantly closer to healthy controls than those in fibroproliferative or

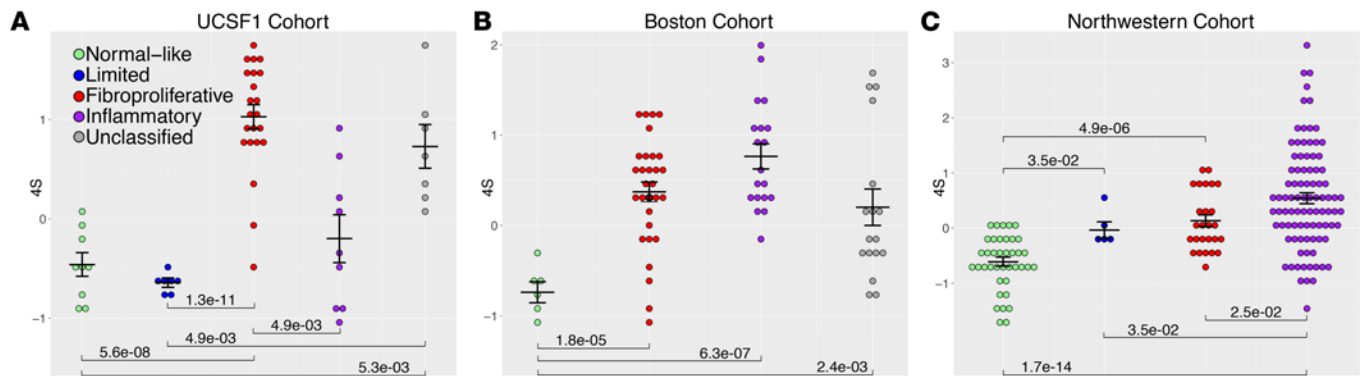


Figure 6. Comparison of 4S across intrinsic subsets. Beeswarm dot plots (mean \pm SEM) comparing the systemic sclerosis skin severity score (4S) across intrinsic subsets in (A) the UCSF1 cohort, (B) the Boston cohort, and (C) the Northwestern cohort. Bars between groups represent FDR-corrected P values (q values) from Student's unpaired, 2-sided t test of 4S between the intrinsic subsets. Bars between a pair of intrinsic subsets is shown only when $q < 5\%$. Unclassified samples were not assigned to any intrinsic subsets in the original publications. (A) $n = 9, 7, 22, 9,$ and 7 ; (B) $n = 6, 0, 30, 19,$ and 17 ; (C) $n = 39, 5, 26, 91,$ and 0 for normal-like, limited, fibroproliferative, inflammatory, and unclassified, respectively.

inflammatory subsets. These results across all SSc skin cohorts strongly suggest that the molecular heterogeneity of SSc patients can be represented as a continuum defined by 4S.

Arguably, this may not be surprising given that mRSS is a continuum and there is a significant correlation between mRSS and 4S across 7 independent SSc skin biopsy cohorts. However, recent analyses suggest there may be stages in the progressive disease (24). Our results in longitudinal serial skin biopsy samples in the Northwestern cohort also suggest that patients may go through different intrinsic subsets as part of their progressive diseases. Collectively, these results further suggest that the intrinsic subset assignment for a patient is not stable over time, and may change (10). This observation suggests that while the intrinsic subsets are indicative of current pathway activation, they may be along a continuum in this progressive disease.

Our comparison of the 415-gene signature with transcriptional profiles of 314 ligand stimulations across different cell lines in LINCS showed that it is positively correlated with multiple EGFR ligands. In fact, out of the 11 known EGFR ligands, 4 are among the 6 positively correlated ligands with our signature. A number of in vitro and in vivo studies have suggested that EGF, among other cytokines, regulates dermal fibroblast proliferation and extracellular matrix (ECM) deposition, which is essential to fibrosis (25). Furthermore, pathway analysis of our 415-gene signature identified PI3K/Akt as significant, which is downstream of the EGFR signaling pathway. Indeed, autoantibodies against the extracellular domain of EGFR have been described in SSc patients (26). Collectively, these multiple lines of evidence strongly suggest an important role of EGFR and its downstream pathways, including PI3K/Akt, in SSc. Because differential expression of genes in this signature is found in all SSc patients independent of serum autoantibodies, disease subtype (limited cutaneous and diffuse cutaneous), disease duration, and skin score, drugs that broadly target this gene signature and its associated pathways may be especially useful as initial or background therapies that could be effective for all SSc patients.

Finally, our results described here provide yet another example of, and underscore the importance of, leveraging the heterogeneity present in publicly available data to identify reproducible disease signatures that are diagnostic and provide novel mechanistic insights into disease pathogenesis. In addition to inherent confounding factors associated with any experimental design, the analysis of individual datasets to identify a gene signature can also be affected by subjective factors such as a choice of a threshold for statistical significance, or bias towards a specific known biological pathway (27). On the other hand, the multicohort analysis of independently generated public data that we use here allows the identification of consistent and robust gene signatures while better accounting for the biological and technical heterogeneity observed in the real world. This approach has been found to be very effective at uncovering genes with consistent expression profiles that are mechanistic, diagnostic, and therapeutic (14–17, 19, 20).

In summary, the generalizability of the 415-gene set across multiple independent cohorts suggests that it could be useful for SSc diagnostics and prognostics, and to identify novel drug targets. Because this gene set is found in all SSc patients independent of serum autoantibodies, disease subtype (limited cutaneous and diffuse cutaneous), disease duration and skin score, drugs that broadly target these gene signatures and

their associated pathways may be a useful initial or background therapy for all SSc patients. Most of the existing SSc treatments suppress immune responses broadly (e.g., cyclophosphamide, autologous stem cell transplant, MMF, etc.); however, our results suggest that the 415-gene set could serve as a starting point to identify novel drug targets for SSc treatment.

Methods

Data collection and preprocessing. We downloaded 4 gene expression datasets of SSc skin biopsies from GEO. For each study we used the sample phenotypes as defined by the corresponding original published study. We mapped microarray probes in each dataset to Entrez Gene identifiers (IDs). If a probe matched more than 1 gene, the expression data for the probe were expanded to add 1 record for each mapped gene (28).

Meta-analysis by combining effect size. We analyzed the 2 skin biopsy datasets used for discovery using 2 different meta-analysis methods as described before (13, 14): (a) combining effect size and (b) combining P values. We estimated the effect size for each gene in each dataset as Hedges' adjusted g , which accounts for small sample bias. If multiple probes mapped to a gene, we summarized the effect size for each gene using the fixed-effect inverse-variance model.

Then, we combined the study-specific effect sizes for each gene into a single meta-effect size using a linear combination of study-specific effect sizes, f_p , where each study-specific effect size was weighted by the inverse of the variance in the corresponding study (Equation 1). After computing the meta-effect size, we identified significant genes using the Z statistic, and corrected P values for multiple hypothesis testing using Benjamini-Hochberg FDR correction (29).

$$f_{meta} = (f_1 w_1 + f_2 w_2 + \dots + f_k w_k) / (w_1 + w_2 + \dots + w_k); w_i = 1 / var(f_i) \quad (\text{Equation 1})$$

Meta-analysis by combining P values. We used Fisher's sum of logs method (30) for meta-analysis by combining P values. For each gene, we summed the logarithm of the 1-sided hypothesis testing P values across k studies, and compared the result with a χ^2 distribution with $2k$ degrees of freedom. This process allowed us to identify significant genes:

$$\chi_{2k}^2 = -2 \sum_{i=1}^k \log(p_i)$$

(Equation 2)

Selection of differentially expressed genes. We selected 415 genes that satisfied the following criteria: (a) absolute summary effect-size greater than 1.5, (b) FDR less than or equal to 5% across all datasets when combining effect size, (c) measured in all discovery datasets, (d) when combining P values using Fisher's test, FDR less than or equal to 5%.

LINCS perturbagen signature comparisons. We downloaded the level-4 (median absolute deviation normalized) data from the LINCS c3 server using the LINCS sig_query tool. We processed the L1000 data, which measured expression of 1,000 select genes, as follows: we removed the signatures that did not pass LINCS internal quality control standards and were not labeled as gold. We removed the probes that were not in the bing probe set. We computed the expression of a gene by averaging the expression of all probes that mapped to a given gene. Each perturbagen in LINCS has a number of signatures representing gene expression changes it induces in different cell lines, at different doses, and measured at different times. We used the median of gene expression changes for each signature to create a single consensus signature. Next, we computed the Pearson correlation of the SSc signature with the perturbagen signatures. We corrected P values for multiple hypothesis testing using Benjamini-Hochberg FDR correction and selected perturbagens with FDR less than or equal to 5%.

PCA. As preprocessing, we imputed any missing values using the impute.knn function from the impute package (version 1.42.0; <http://www-stat.stanford.edu/~hastie/Papers/missing.pdf>) in R (<http://www.R-project.org/>). We then retained only the gene expression measurements for our 415-gene signature and discarded the rest. We then rank normalized each sample to reduce batch effects. Then, we concatenated both discovery cohorts into a single matrix of expression ranks and performed PCA using the prcomp function from the stats package in R.

Sample-level principal component values were determined in R using the predict function, supplying the output of prcomp and the samples to be transformed. After determining the sample principal component values, we use the rgl package in R (version 0.95.1337; <http://CRAN.R-project.org/package=rgl>) to create 3-dimensional plots of the cohorts.

Defining the health bubble, centroid of health, and DTH. The health bubble represents the space defined by PC2, PC3, and PC4 where a sample is more likely to be from a healthy control than from an SSc patient. To determine the boundaries of the health bubble, we used support vector machine (SVM) to build a classifier to distinguish between healthy controls and SSc patients in the discovery cohorts. Specifically, in the discovery cohorts, we used PC2, PC3, and PC4 as feature variables to predict the class labels. We used the ksvm function in the e1071 package in R with C-svc as the type parameter (version 1.6-7; <http://CRAN.R-project.org/package=e1071>). We determined the boundary between healthy and SSc patients as the convex hull connecting the set of points where the probability of being an SSc patient, as predicted by the SVM model, was 50%. We determined these points by computing probabilities for an equally spaced grid of points in our PC space and selecting only the points with log-odds between -0.1 and 0.1 . We used the ashape3d function in the alphashape3d package to determine and plot the convex hull (version 1.1; <http://CRAN.R-project.org/package=alphashape3d>). Next, we defined the centroid of health as the median of PC2, PC3, and PC4 for the healthy samples in the discovery cohorts. Finally, we defined DTH for an SSc sample as the Euclidean distance between the centroid of health and the corresponding skin sample.

Statistics. All P values reported in this analysis are 2-tailed. All comparisons between means of discrete groups were performed using the 2-tailed unpaired Student's t test. Student's t test P values were determined in R using the t.test function, which follows standard procedures for determining significance.

All correlation coefficients were Pearson correlation coefficients, except where noted. Pearson and Spearman correlation coefficients and P values were determined using the cor.test function in R, which follows the standard procedures. More concretely, the standard error of a Pearson correlation coefficient is given by Equation 3.

$$SE_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

(Equation 3)

With this value, we determine the t statistic and compute the P value from the Student's t distribution with degrees of freedom $n - 2$. Spearman correlation coefficients are determined as the Pearson correlation values of the ranks of the data. To determine Spearman correlation P values, we determine the Pearson correlation of the ranks and then follow the same procedures noted above.

Study approval. Northwestern cohort: The Northwestern University IRB approved the study (STU00004428). Informed consent was obtained from all research subjects prior to study participation. Stanford cohort: The Stanford IRB reviewed and approved the study protocol (IRB protocol number 12407) and all subjects provided written informed consent prior to participating in the study. UCSF2 cohort: Skin samples were collected as part of study number 11-08170, which was approved by the UCSF IRB. All subjects provided informed consent prior to their participation in the study. HSS cohort: The protocol was approved by the IRB (IRB protocol number 2014-268) at the Hospital for Special Surgery. Patients provided written informed consent before enrollment.

Author contributions

SL, MH, and PK conceived the project and wrote the manuscript. SL and PK analyzed the data. MH, MC, KA, EA created the registry, gathered patient biopsies, and kept clinical records for the Northwestern cohort. AV, DF, and LC gathered patient biopsies and kept clinical records for the Stanford cohort. AH, PW, and FB gathered patient biopsies and kept clinical records for the UCSF2 cohort. JG and RS gathered patient biopsies and kept clinical records for the HSS cohort. SA gathered patient biopsies and kept clinical records for the Houston cohort. PK oversaw the project.

Acknowledgments

PK is supported by grants from the Bill and Melinda Gates Foundation, and National Institute of Allergy and Infectious Diseases (NIAID) grants 1U19AI109662, U19AI057229, and U54AI117925 outside the submitted work. MH is supported by the Scleroderma Foundation, NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases grants K12 HD055884, K23 AR059763, R21 AR068035, and the Scleroderma Research Foundation. FB is supported by the Nina Ireland Program for Lung Health (NIPLH) Award. JG is supported by the Scleroderma Foundation New Investigator Grant and the Hospital for Special Surgery/Kellen Foundation Clinician Scientist Development Award. The HSS Cohort trial was funded by Novartis. SA is supported by the Scleroderma Foundation SCORE grant and NIH grant K23AR0314636. LC is supported by the Scleroderma Foundation, the Scleroderma Research Foundation, and the Scleroderma Clinical Trials Consortium. PJU is the recipient of a gift from the Henry Gustav Floren Foundation, and a gift from The Ben May Charitable Trust of Mobile, Alabama, USA. PJU is supported by NIH grants NIAID U19 AI1110491, Stanford Autoimmunity Center of Excellence (ACE); NIAID 1 UM1AI110498 01, ACE Collaborative Project; NIAID 1 R01 AI125197-01; and Alliance for Lupus Research, grant number 21858. MLW and TAW were supported by grants from the Scleroderma Research Foundation (SRF), by a grant from the Dr. Ralph and Marian Falk Medical Research Trust and by NIH grants P50 AR060780 and P30 AR061271.

Address correspondence to: Purvesh Khatri, 279 Campus Dr., Beckman Center B235A, Stanford University, Stanford, California 94305, USA. Phone: 650.497.5281; E-mail: pkhatri@stanford.edu.

- Desbois AC, Cacoub P. Systemic sclerosis: An update in 2016. *Autoimmun Rev*. 2016;15(5):417–426.
- Volkman ER, Furst DE. Management of systemic sclerosis-related skin disease: a review of existing and experimental therapeutic approaches. *Rheum Dis Clin North Am*. 2015;41(3):399–417.
- Farina G, Lafyatis D, Lemaire R, Lafyatis R. A four-gene biomarker predicts skin disease in patients with diffuse cutaneous systemic sclerosis. *Arthritis Rheum*. 2010;62(2):580–588.
- Rice LM, Stifano G, Ziemek J, Lafyatis R. Local skin gene expression reflects both local and systemic skin disease in patients with systemic sclerosis. *Rheumatology (Oxford)*. 2016;55(2):377–379.
- Rice LM, et al. A longitudinal biomarker for the extent of skin disease in patients with diffuse cutaneous systemic sclerosis. *Arthritis Rheumatol*. 2015;67(11):3004–3015.
- Chung L, et al. Molecular framework for response to imatinib mesylate in systemic sclerosis. *Arthritis Rheum*. 2009;60(2):584–591.
- Gordon JK, et al. Nilotinib (Tasigna™) in the treatment of early diffuse systemic sclerosis: an open-label, pilot clinical trial. *Arthritis Res Ther*. 2015;17:213.
- Hinchcliff M, et al. Molecular signatures in skin associated with clinical improvement during mycophenolate treatment in systemic sclerosis. *J Invest Dermatol*. 2013;133(8):1979–1989.
- Chakravarty EF, et al. Gene expression changes reflect clinical response in a placebo-controlled randomized trial of abatacept in patients with diffuse cutaneous systemic sclerosis. *Arthritis Res Ther*. 2015;17:159.
- Pendergrass SA, Lemaire R, Francis IP, Mahoney JM, Lafyatis R, Whitfield ML. Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. *J Invest Dermatol*. 2012;132(5):1363–1373.
- Milano A, et al. Molecular subsets in the gene expression signatures of scleroderma skin. *PLoS One*. 2008;3(7):e2696.
- Assassi S et al. Dissecting the heterogeneity of skin gene expression patterns in systemic sclerosis. *Arthritis Rheumatol*. 2015;67(11):3016–3026.
- Sweeney TE, Haynes WA, Vallania F, Ioannidis JP, Khatri P. Methods to increase reproducibility in differential gene expression via meta-analysis [published online ahead of print September 14, 2016]. *Nucleic Acids Res*. doi: 10.1093/nar/gkw797.
- Khatri P, et al. A common rejection module (CRM) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation. *J Exp Med*. 2013;210(11):2205–2221.
- Mazur PK, et al. SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer. *Nature*. 2014;510(7504):283–287.
- Chen R, et al. A meta-analysis of lung cancer gene expression identifies PTK7 as a survival gene in lung adenocarcinoma. *Cancer Res*. 2014;74(10):2892–2902.
- Sweeney TE, Shidham A, Wong HR, Khatri P. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. *Sci Transl Med*. 2015;7(287):287ra71.
- Sweeney TE, Khatri P. Comprehensive validation of the FAIM3:PLAC8 ratio in time-matched public gene expression data. *Am J Respir Crit Care Med*. 2015;192(10):1260–1261.
- Andres-Terre M, et al. Integrated, multi-cohort analysis identifies conserved transcriptional signatures across multiple respiratory viruses. *Immunity*. 2015;43(6):1199–1211.
- Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. *Lancet Respir Med*. 2016;4(3):213–224.
- Li MD, Burns TC, Morgan AA, Khatri P. Integrated multi-cohort transcriptional meta-analysis of neurodegenerative diseases. *Acta Neuropathol Commun*. 2014;2:93.
- Mori Y, Chen SJ, Varga J. Expression and regulation of intracellular SMAD signaling in scleroderma skin fibroblasts. *Arthritis Rheum*. 2003;48(7):1964–1978.

23. Draghici S, et al. A systems biology approach for pathway level analysis. *Genome Res.* 2007;17(10):1537–1545.
24. Mahoney JM, et al. Systems level analysis of systemic sclerosis shows a network of immune and profibrotic pathways connected with genetic polymorphisms. *PLoS Comput Biol.* 2015;11(1):e1004005.
25. Ihn H. Scleroderma, fibroblasts, signaling, and excessive extracellular matrix. *Curr Rheumatol Rep.* 2005;7(2):156–162.
26. Planque S, et al. Autoantibodies to the epidermal growth factor receptor in systemic sclerosis, lupus, and autoimmune mice. *FASEB J.* 2003;17(2):136–143.
27. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol.* 2012;8(2):e1002375.
28. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med.* 2008;5(9):e184.
29. 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(1):289–300.
30. Fisher RA. *Statistical Methods for Research Workers.* Edinburgh, Scotland: Oliver & Boyd; 1934.