Award Number: W81XWH-14-1-0224

TITLE: Environmental Mycobiome Modifiers of Inflammation and Fibrosis in Systemic Sclerosis

PRINCIPAL INVESTIGATOR: Michael L. Whitfield, Ph.D

CONTRACTING ORGANIZATION: Dartmouth College
HANOVER, NH 03755

REPORT DATE: Nov 2018

TYPE OF REPORT:FINAL

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE		*Form Approved* *OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**		
1. REPORT DATE **NOVEMBER 2018**	**2. REPORT TYPE** Final	**3. DATES COVERED** 20AUG2014 - 19AUG2018

4. TITLE AND SUBTITLE
Environmental Mycobiome Modifiers of Inflammation and Fibrosis in Systemic Sclerosis

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-14-1-0224

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Michael Whitfield, Ph.D, Patricia A. Pioli, Ph.D
Robert Lafyatis, MD
Sarah Arron, MD/Ph.D
E-Mail: Michael.L.Whitfield@dartmouth.edu

5d. PROJECT NUMBER
PR130908

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Dartmouth COLLEGE, Hanover, NH, Boston University Medical Center, MA, University of California, San

8. PERFORMING ORGANIZATION REPORT

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project is focused on Systemic Sclerosis (SSc), a progressive fibrotic disease characterized by skin fibrosis and damage to internal organs. While a wide range of environmental and biological triggers have been proposed, no definitive etiologic agents have yet been identified. Metagenomic analysis of non-human sequences in SSc RNA-seq data was used to detect microbial sequences in human tissues in an unbiased, quantitative manner. Our studies suggest that disease pathogenesis includes a common environmental fungal trigger, *Rhodotorula glutinis*, which we hypothesize elicits immune activation in a permissive host genetic background.

Skin biopsies have been collected from SSc patients and analyzed by high-throughput sequencing, providing substantial gene expression data as well as detailed information regarding the host microbiome. Data have been compared against that of healthy control samples.

15. SUBJECT TERMS: IMSA, systemic sclerosis, scleroderma, mycobiome, fibrosis, gene, genetics, RNA-seq, *R. glutinis*, Metagenomics

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF**	**18. # OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	UU	80	

Table of Contents

COVER PAGE 1

REPORT DOCUMENTATION PAGE 2

1. INTRODUCTION 5

2. KEYWORDS: 5

3. ACCOMPLISHMENTS 5

Milestone 1 5
Milestone 2 6
Milestone 3 6
PRELIMINARY RESULTS BY MILESTONE 6
KEY RESEARCH ACCOMPLISHMENTS Summary (Aug 2014- Aug 2017) 15
The next reporting period: 15

4. IMPACT 15

5. CHANGES/PROBLEMS 16

6. PRODUCTS: 16

Oral Presentations: (Chronological Order) 16
Abstracts and Presentations: (Chronological Order) 18
Manuscripts: 18
Degrees obtained that are supported by this award 20
Development of cell lines, tissue or serum repositories 20

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS 20

8. SPECIAL REPORTING REQUIREMENTS 20

9. REFERENCES 20

10. APPENDIX 20

1. INTRODUCTION

Systemic sclerosis (SSc) is a heterogeneous disease of fibrosis and inflammation, concomitant with significant autoimmunity. SSc often presents with skin manifestations and Raynaud's phenomenon; the extent and location of fibrotic lesions in people with SSc contributes to the diagnoses of disease subtypes and prognosis. My laboratory has pioneered the use of gene expression subsets in SSc [1-5]. Most recently we have demonstrated enrichment of a mycobiome component (Rhodotorula glutinis) in SSc patient skin [6].

We describe our studies from all four years of the grant below. This work was accomplished by researchers at Geisel School of Medicine at Dartmouth, Boston University Medical Center and University of California, San Francisco under the partnering PI option.

2. KEYWORDS:

IMSA, systemic sclerosis, scleroderma, SSc, mycobiome, microbiome, fibrosis, gene, genetics, RNA-seq, Next Generation Sequencing, skin, *R. glutinis*, *Rhodotorula,* Metagenomics

3. ACCOMPLISHMENTS

Milestones were assigned to this proposal, with tasks to be accomplished by each investigator. The overall **summary** of our progress relative to these tasks is given below, followed by a complete discussion of our work during the four years of this grant.

Milestone 1 Determine the identity and distribution of microbiome components across SSc skin.

Task 1 (Months 1-36) Whitfield Laboratory to perform RNA-seq analysis of SSc skin biopsies.

Lesional forearm skin biopsies were collected from 23 SSc patients; seven patients also provided biopsies of non-lesional back skin. Forearm skin biopsies were also obtained from 6 age- and gender-matched healthy controls. Samples included both clinically limited (lSSc) and diffuse (dSSc) disease, with disease duration ranging from 0 – 35 years. The patient population consisted primarily of early stage patients (disease duration ≤ 2 years), though a handful of very late-stage patients (disease duration > 10 years) were also included to assess microbiome changes over time.

Task 2 (Months 6-36) Whitfield Laboratory to perform RNA-seq analysis for differentially expressed mRNAs and non-coding RNAs.

RNA-seq was performed on 36 skin biopsies, from 29 unique patients, resulting in an average of 83 million reads per sample (range: 51,278,817 – 112,643,430). Raw sequencing reads were aligned to the human genome (hg19) using STAR aligner [7], and the expression level of each gene was expressed as fragments per million mapped reads (FPKM). Intrinsic gene expression subset designations were determined based on support vector machine classification using normalized FPKM values [8]. Hierarchical clustering using the gene list from Johnson, et al.[9], resulting in a total of 1010 overlapping genes, revealed distinct molecular subsets of disease, characterized by strong immune activation, lipid signaling, and proliferation signals, consistent with previous publications [2, 3, 10]. Together, these data suggest our patient cohort is representative of the four major intrinsic gene expression subsets of SSc. Additionally, we find that forearm and back samples largely tend to cluster together, consistent with previous analyses [1]

Task 3 (Months 6-36) Arron group to perform IMSA and determine the identity of microbiome components.

Filtering of human sequence reads and microbiome annotation was performed using Integrated Metagenomic Sequence Analysis (IMSA) [11], yielding an average of 18,794 informative hits, defined as sequences mapping to five or fewer species, per skin biopsy (range: 3,098 – 74,429) across 1870 genera. To adjust for library-specific effects, all data were rarefied to the level of the lowest sample, followed by median centering of each genus by library preparation batch. This approach substantially reduced batch effects associated with library preparation, enabling direct comparisons of sample outputs across patients.

Task 4 (Months 1-24) Arron group to create scaffolds from aligned reads for each microbiome component and develop nested PCR followed by targeted multiplexed sequencing assays for cost-effective screening.

We developed a fast and efficient nested PCR reaction targeting microbiome components specific for fungal species identification. We then evaluated the identity and number of fungal reads by next generation sequencing (see Task 5). We have found the data from this targeted mycobiome method to be lower quality than our metagenomic RNA-seq analyses and we are exploring targeted microbiome methods. Although the method was developed we did not feel it had high enough accuracy to justify this method on a large set of independent samples, and instead will focus on metagenomic methods for the validation.

Our preliminary findings suggest significant bacterial dysbiosis in affected skin, with only modest changes in fungal abundance. Validation of these findings will therefore be run using targeted 16S sequencing to identify changes in bacterial composition. DNA has been collected for 116 skin samples (SSc and healthy controls) to date, with an additional 178 samples awaiting DNA purification. Targeted sequencing of the V2-4-8 and V3-6,7-9 hypervariable regions of 16S rRNA from these samples will be performed using the Ion 16S Metagenomics Kit (ThermoFisher) and sequenced on the Ion Torrent.

This approach enables direct assessment of all microbiome components associated with affected skin, including archaea, bacteria, viruses, fungi, and parasites, as well as comparisons between microbiome composition and gene expression of associated tissues using the same computational methods.

Task 5 (Months 1-12) Whitfield Laboratory to examine a larger population of archived skin biopsy RNA to determine the prevalence of microbiome components across the SSc population.

The goal of this task was validation. The Whitfield is routinely sequencing skin and esophageal biopsies of patients from other funded studies and now have more than 300 biopsies with RNA-seq data that can be analyzed. We are currently performing metagenomic analyses on these other datasets to validate the findings in this initial cohort.

Task 6 (Months 1-24) Culture microbiome components from the skin of SSc patients. Use of skin biopsies as a method for fungal culture was not successful.

Disappointingly, swabbing of affected skin of SSc patients has yet to result in recovery of clinically relevant fungi or other organisms. Swabbing and other culture-based methods of microbial detection will be revisited following completion of our RNA-seq analyses, enabling targeted isolation of organisms associated with SSc lesional skin.

Milestone 2 Identify the inflammatory infiltrates in SSc skin and their response to microbiome components

Task 1 (Months 1-6) Whitfield Laboratory to perform computational analysis/prediction of inflammatory cell infiltrates from whole genome expression data.

We have used single sample Gene Set Enrichment Analysis (ssGSEA) to identify the cellular subsets in SSc skin at different stages of disease.

Task 2 (Months 6-24) Perform immunohistochemistry to validate the computational predictions of task 1.

We have optimized markers for different cell types in SSc skin. We have used CD163 for macrophages, CD1c for myeloid dendritic cells (mDCs), and CD3 for T cells in a separate study. Current work focuses on the identification of T cell subsets (CD4 and CD8), and B cells (CD19 and CD20). We can now use these markers to look at innate and adaptive immune cells in the patients of this study.

Task 3 (Months 1-18) Whitfield Laboratory to develop protocols for the isolation and characterization of immune cells from skin using the sclerodermatous Graft-Versus Host Disease (sclGVHD) mouse including detailed characterization of cell types.

We established the sclGVHD model in the laboratory. We have performed initial cell isolations and phenotyping of these samples.

Task 4 (Months 6-18) Identify the secreted mediators of fibrosis / inflammation being produced (Whitfield / Pioli). Once cells are isolated, we will screen for secreted pro-fibrotic mediators.

We have demonstrated that co-culture of human SSc macrophages with fibroblasts results in fibrotic activation. Please see discussion below for details.

Task 5 (Months 12-36) Apply protocols to characterize the inflammatory infiltrate in the skin of SSc patients (Whitfield / Pioli). After cell isolation procedures have been optimized in the sclGVHD mouse, we will examine the infiltrate and profibrotic mediators in SSc skin biopsies.

Protocols for isolating these cells from mouse and human skin have been optimized.

Milestone 3 Determine if SSc patients have a specific immune response against *R. glutinis* that is different from healthy controls and if this response can drive fibrosis.

Task 1 (Months 1-24) Test patient sera for cross-reactivity against *R. glutinis* antigens (Whitfield/Lafyatis).

We have performed western blots using whole cell lysates and probed with sera collected from both healthy controls and SSc patients.

Task 2 (Months 1-24) Identify the cross-reacting proteins by mass spectrometry (Whitfield).

Serum-immunoprecipitation of R. glutinis and human HeLa cell whole cell lysates followed by mass spectrometry was performed to identify immunoreactive proteins associated with R. glutinis. We have written a manuscript on the human cross-reactivity. We are having difficulty with the annotation state of the R. glutinis genome for annotating those spectra.

Task 3 (Months 12-36) Use isolated PBMCs and isolated monocytes to examine the cytokines secreted and changes in gene expression when cells are exposed to *R. glutinis* or other putative micro / mycobiome triggers (Whitfield/Pioli).

Despite attempts to carry out this task, we encountered technical problems and were unable to complete this as planned. We were able to analyze the response of these cells to LPS (See Figure 11) as surrogate for microbiome exposure.

Task 4 (Months 12-24) Determine if chronic exposure to *R. glutinis* or other micro / mycobiome components stimulate a fibrotic response in a mouse model of SSc. (Whitfield).

Despite attempts to carry out this task, we encountered technical problems and were unable to complete this as planned.

PRELIMINARY RESULTS BY MILESTONE

Milestone 1: Determine the identity and distribution of microbiome components across SSc skin

Task 1: RNA-Seq analysis of SSc skin. *Lesional forearm skin biopsies were collected from 23 SSc patients; seven patients also provided biopsies of non-lesional back skin. Forearm skin biopsies were also obtained from 6 age- and gender-matched healthy controls. Samples included both clinically limited (lSSc) and diffuse (dSSc) disease, with disease duration ranging from 0 – 35 years. The patient population consisted primarily of early stage patients (disease duration ≤ 2 years), though a handful of very late-stage patients (disease duration > 10 years) were also included to assess microbiome changes over time.* (Table 1)

Table 1. Summary clinical information

	Control Subjects (*N* = 6)	**SSc Patients (*N* = 23)**
Age, median (range) years	53 (25 - 67)	53 (27 - 77)
Sex, *N* (%) female	4 (67%)	19 (83%)
Race, *N* (%) Caucasian	5 (83%)	20 (87%)
SSc subtype, *N* (%) diffuse	NA	15 (65%)
MRSS, median (range)	NA	16 (0 - 44)
Disease duration from first non-Raynaud's, median (range) years	NA	1.0 (0 - 35)
ILD/PAH, *N* (%)	NA	8 (35%)

ANA primary pattern, *N* (%) patients		
Homogenous	NA	1 (4%)
Nucleolar	NA	5 (22%)
Speckled	NA	6 (26%)
Centromere	NA	2 (9%)
SSc-specific antibodies, *N* (%)		
Anti-centromere	NA	3 (13%)
Scl-70	NA	3 (13%)
RNA Polymerase III	NA	5 (22%)
Current Therapies, *N* (%)	NA	17 (74%)
Prior Therapies, *N* (%)		
Amlodipine	NA	4 (17%)
Methotrexate	NA	4 (17%)
Prednisone	NA	3 (13%)

Abbreviations: SSc, systemic sclerosis; ANA, anti-nuclear antibodies; MRSS, modified Rodnan skin score; ILD, interstitial lung disease; PAH, pulmonary arterial hypertension; NA, not applicable

Current and prior therapies include all treatments observed in 3 or more patients

Table 2. Statistics of alignment

Sample ID	Disease Type	Total Reads	Uniquely Mapped	Mapped length	% multi-mapped	% Unmapped
MK01-FA	SSc	79642964	0.6	148	0.37	0.03
MK01-B	SSc	87465540	0.45	148	0.53	0.02
AM02-FA	SSc	82459754	0.7	148.1	0.28	0.02
AM02-B	SSc	89078313	0.64	148.1	0.34	0.02
KB03-FA	SSc	110881994	0.8	148.1	0.17	0.03
KB03-B	SSc	80684936	0.76	148	0.14	0.1
JP04-FA	SSc	86911206	0.73	148	0.15	0.12
JP04-B	SSc	87609067	0.54	148	0.36	0.1
KB05-FA	SSc	78252845	0.72	148	0.16	0.12
KB05-B	SSc	105842068	0.67	147.7	0.27	0.06
KL06-FA	SSc	103150299	0.81	148	0.15	0.04
KL06-B	SSc	75717290	0.64	148.2	0.34	0.02
SH07-FA	SSc	85776797	0.67	145.7	0.13	0.2
SH07-B	SSc	113066228	0.75	148	0.2	0.04
N01_Base	SSc	82992452	0.67	148	0.31	0.03
N05_Base	SSc	77151907	0.74	148.1	0.23	0.02
N07-Base	SSc	87025838	0.8	148.1	0.18	0.02
N09_Base	SSc	86468461	0.73	148.1	0.25	0.02
N10_Base	SSc	100362240	0.62	147.7	0.31	0.07
N11_Base	SSc	105056505	0.47	147.9	0.47	0.05
N15_Base	SSc	81209030	0.81	147.8	0.12	0.07
N18_Base	SSc	72275968	0.8	147.6	0.13	0.07
N13-2	Control	80255529	0.64	148.1	0.34	0.02
N15-3	Control	92668648	0.67	147.4	0.31	0.02
N15-5	Control	84968648	0.59	147.9	0.38	0.04
N15-15	Control	92222389	0.32	148	0.61	0.06
N15-21	Control	103634414	0.32	147.6	0.58	0.06
N15-24	Control	75379182	0.78	148	0.14	0.08

Task 2 RNA-seq analysis for differentially expressed mRNAs.

RNA-seq was performed on 36 skin biopsies, from 29 unique patients, resulting in an average of 83 million reads per sample (range: 51,278,817 – 112,643,430). Raw sequencing reads were aligned to the human genome (hg19) using STAR aligner [7], and the expression level of each gene was expressed as fragments per million mapped reads (FPKM). Batch biases generated by the inclusion of previously sequenced samples from a separate study (N_Base samples) was performed with ComBat (**see prior progress reports)**.

As an initial analysis of these data, we examined the consensus genes from Mahoney et al. [4]. These are genes that were consistently and reproducibly associated with individual SSc intrinsic gene expression subsets across three independent patient cohorts. Expression of these genes in our RNA-seq data reveals increased expression in the inflammatory and fibroproliferative subsets of patients (**Figure 1**). Expression of these genes is shown both before and after batch correction. Intermixing of samples is clearly evident after ComBat correction, indicating that batch correction was successful.

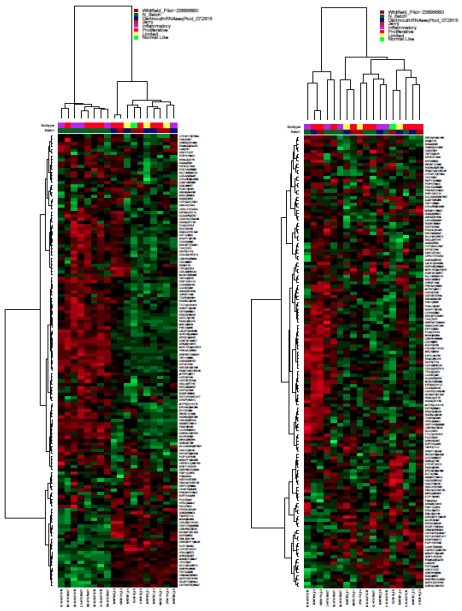


Figure 1. Heatmap of Mahoney_267modules before and after correction. (Left) no ComBat correction; (Right) with ComBat correction.

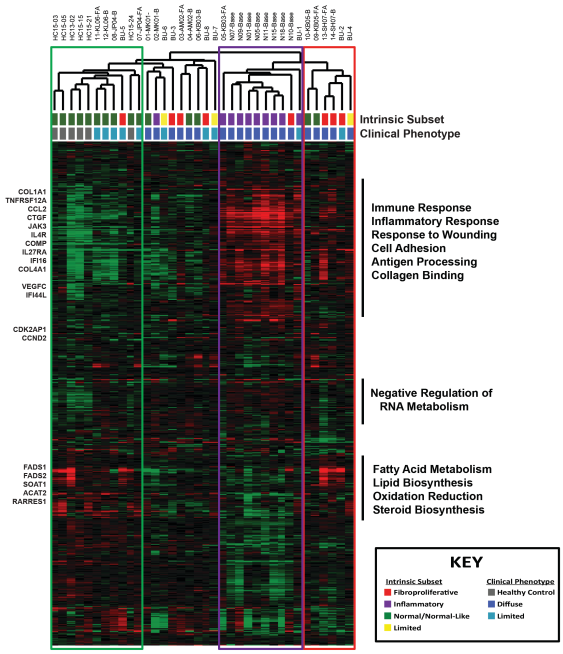


Figure 2. Intrinsic Subset Analysis of RNA-seq Reads from SSc Skin. Assignment of intrinsic molecular subsets for SSc patients was performed using a Support Vector Machine (SVM) developed for the purpose. Displayed are the 1010 genes from Johnson, et al. collapsed on gene ID and extracted from the normalized FPKM values for all 36 RNA-seq samples. Hierarchical clustering revealed distinct molecular subsets of disease, consistent with previous publications. The sample dendrogram is colored to indicate intrinsic subset designations: normal-like (green), limited (yellow), inflammatory (purple), proliferative (red). Hash marks indicate SSc clinical diagnosis associated with each sample. Black bars indicate genes that clustered together hierarchically; the most significantly overrepresented GO terms are listed.

Intrinsic gene expression subset designations were determined based on support vector machine classification using normalized FPKM values [8]. Hierarchical clustering using the gene list from Johnson, et al. [9], resulting in a total of 1010 overlapping genes, revealed distinct molecular subsets of disease, characterized by strong immune activation, lipid signaling, and proliferation signals, consistent with previous publications [2-4] **(Figure 2)**. Together, these data suggest our patient cohort is representative of the four major intrinsic gene

expression subsets of SSc. Additionally, we find that forearm and back samples largely tend to cluster together, consistent with previous analyses [1].

Task 3: IMSA analysis to identity microbiome components. Filtering of human sequence reads and microbiome annotation was performed using Integrated Metagenomic Sequence Analysis (IMSA) [11], yielding an average of 18,794 informative hits, defined as sequences mapping to five or fewer species, per skin biopsy (range: 3,098 – 74,429) across 1870 genera. To adjust for library-specific effects, all data were rarefied to the level of the lowest sample, followed by median centering of each genus by library preparation batch. This approach substantially reduced batch effects associated with library preparation, enabling direct comparisons of sample outputs across patients.

Antimicrobial gene expression is suppressed in SSc lesional skin

Antimicrobial peptides (AMPs), including cathelicidin (CAMP/LL-37), α-defensins, and β-defensins, are an essential component of epithelial barrier defenses. To assess the role of AMPs in SSc, we compared gene expression levels between SSc and controls, as well as between lesional forearm and non-lesional back skin. Among the major AMPs, dermcidin (DCD) is highly expressed across samples, regardless of disease type, while other major AMPs, including cathelicidin (CAMP) and the α-defensins, were virtually undetected, with no difference in expression between SSc and controls. In contrast, β-defensin 1 (DEFB1), an AMP produced by epithelial cells, is expressed across all samples; however, these levels are significantly lower is SSc lesional skin compared to healthy controls ($p < 0.001$ by unpaired *t*-test), as well as in lesional forearm compared to non-lesional back skin ($p = 0.007$ by paired *t*-test). Similar results were also seen between SSC lesional skin and healthy controls in in a previous SSc skin RNA-seq dataset, suggesting a potential mechanism underlying microbiome differences in SSc patients.

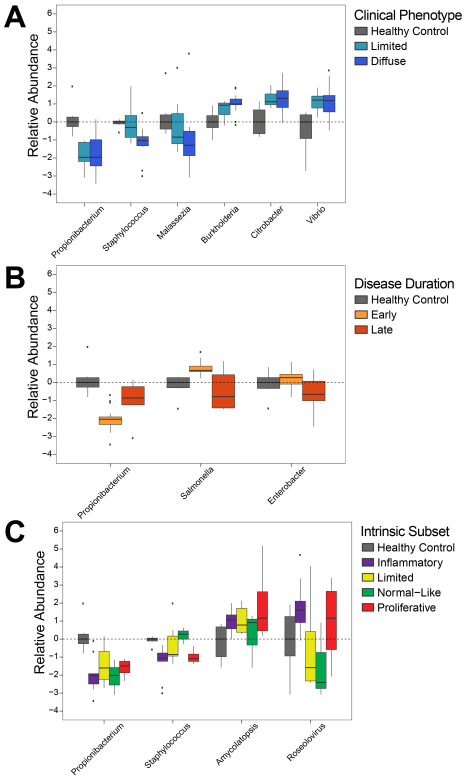


Figure 3. Differential Abundance of Major Skin Taxa. SSc lesional skin exhibits significant changes in microbiome composition, relative to controls. Differential abundance of select genera, relative to controls, based on **A.** clinical subtype, **B.** disease duration (early, < 5 years; late, > 5 years), and **C.** intrinsic molecular subset

Microbiome genus level differences are correlated with SSc clinical phenotypes

We used a process called rarefaction to account for the variable read depth of different samples. We used Quantitative Insights Into Microbial Ecology (QIIME) to perform rarefaction of outputs, a process by which taxa are randomly sampled without replacement; this process is necessary to ensure even sampling depth across patients. Alpha and beta diversity measures are calculated from these data.

SSc patients exhibited large changes in microbiome composition relative to controls, characterized by decreases in lipophilic taxa, such as *Propionibacterium* and *Staphylococcus*, combined with increases in a wide range of Gram-negative bacteria, including *Burkholderia*, *Citrobacter*, and *Vibrio* ($p < 0.05$ for all; **Figure 3A**). These differences were not associated with clinical subtype, with limited and diffuse disease exhibiting broadly similar abundances of major taxa **(Figure 3A)**. Decreases were also observed in the fungus *Malassezia* relative to controls, with the greatest decrease occurring in dSSc patients.

Associations between disease duration and genus-level abundance were also evident, with significant ($p < 0.05$) or near-significant ($p < 0.10$) differences in 6 of the top 21 genera, including *Propionibacterium*, *Salmonella*, and *Enterobacter* **(Figure 3B)**. Relative decreases in *Propionibacterium* were evident for both early and late stage patients, relative to controls. We observed differential directions for the relative abundance of *Salmonella* including significant increases for early stage patients and reduced abundance in late stage patients. Comparisons between the four intrinsic molecular subsets of disease revealed modest differences associated with the normal-like and inflammatory subsets, with normal-like patients broadly mimicking differences seen between SSc and

controls, while the inflammatory group was characterized by decreased *Staphylococcus* and increased *Roseolovirus*, relative to other subsets (**Figure 3C**). The absence of more acute genus-level distinctions between subsets is likely the result of high levels of some genera in both inflammatory and proliferative patients (**Figure 3C)**, thereby limiting the diagnostic value of any single genus. Other clinical cofactors, including sex and autoantibody status, were not statistically different between groups.

Core microbiome by patient is predictive of clinical involvement

To identify changes in microbiome composition associated with clinical covariates, we calculated the number of taxa that accounted for 90% of the annotated reads across our entire dataset, which we collectively refer to as the SSc skin core microbiome. The SSc skin core microbiome was composed of 103 genera, and included representatives from bacteria, fungi, and viruses. Organisms not included in the core microbiome were exclusively low abundance taxa found in only a small number of samples. Hierarchical clustering of the SSc skin core microbiome revealed patterns of microbial abundance closely mimicking that seen within an individual, characterized by clear differences between SSc and controls **(Figure 4A)**. Organisms of the SSc skin microbiome formed distinct branches within the dendrogram. Lipophilic commensals (*Malassezia*, *Propionibacterium*, and *Cutibacterium*) were the predominant genera in normal-like patients, Gram-negative bacteria (*Veillonella*, *Prevotella*, *Neisseria*, and *Actinomycetes*) were abundant in the limited and proliferative subsets, and viruses (*Roseolovirus* and *Cyprinivirus*) were highest in inflammatory patients **(Figure 4A)**. These patterns are consistent with the various environmental niches associated with each class of organisms and are suggestive of changes in skin morphology and immune activation associated with each subset.

SSc skin microbiome profiles were analyzed using principal component analysis (PCA) to identify the broad, population-based changes associated with clinical covariates. Lesional forearm and non-lesional back skin were not significantly different among SSc patients (p = 0.097; **Figure 4B**). Similarly, no significant differences were evident based on SSc clinical subtype (p = 0.156; **Figure 4C**). In contrast, microbiome profiles were strongly correlated with intrinsic subset, with the strongest differences seen in normal-like and inflammatory patients, indicative of a link between disease activity of microbial abundance (p = 0.014; **Figure 4A, D**).

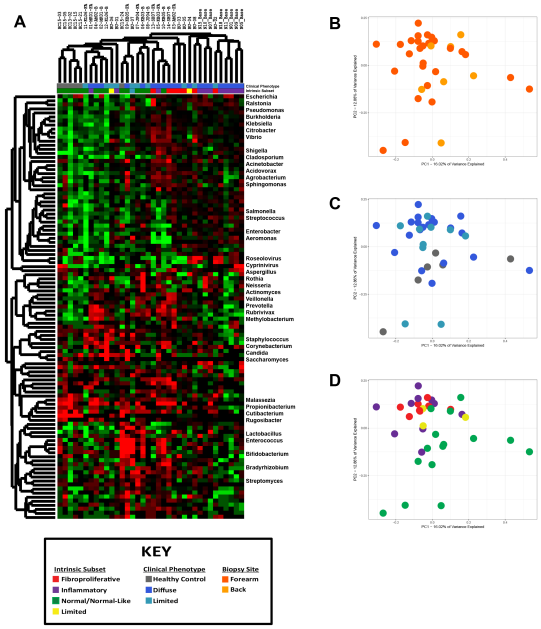


Figure 4. Distribution of the SSc Skin Core Microbiome. The distribution and relative abundance of the SSc skin core microbiome was calculated by rarefaction to the depth of the lowest sample, and filtering to retain the fewest taxa necessary to account for 90% of all reads, resulting in a total of 103 unique genera. Data were then log_2-transformed and median centered by library preparation. **A.** Hierarchical clustering of the core microbiome. Hash marks below the dendrogram indicate intrinsic subset designations and SSc clinical diagnosis for each sample. Principal component analysis of the core microbiome was performed to identify associations between microbiome composition and **B.** biopsy location, **C.** clinical diagnosis, and **D.** intrinsic subset.

Microbiome composition is correlated with inflammatory pathway activation in SSc skin biopsies

Given the close association seen between clinical subtype and microbiome composition, we next sought to identify relationships between relevant molecular pathways and taxonomic abundance using single sample gene set enrichment analysis (ssGSEA). ssGSEA analysis generates a single value quantifying the extent to which a given gene set is coordinately up- or down-regulated in a sample. This analysis was repeated for all available KEGG pathways, generating a table of pathway activation scores for each patient sample. Using this data, we then used Pearson's correlations to compare each of these individual pathways against all genera in the SSc skin core microbiome, the resulting correlation matrix allows for a direct comparison of gene expression and microbiome composition (**Figure 5A**).

Hierarchical clustering of this dataset revealed strong associations between human gene expression and microbial abundance. Processes such as T-cell, B-cell, chemokine, and TGFβ-signaling in the absence of fatty acid signaling are strongly indicative of the inflammatory subset **(Figure 5A, cluster 1)**. Lower immune activation signals in combination with major fatty acid metabolism processes are commonly seen in the proliferative subset **(Figure 5A, cluster 2)**, while fatty acid signaling in the absence of immune activation is most commonly seen in normal-like patients **(Figure 5A, cluster 3)**.

Taxonomic abundance was strongly associated with the molecular processes of immune activation, lipid metabolism, cell proliferation, and Notch/Wnt signaling (**Figure 5A**). Clustering of these processes was strongly correlated with differences in microbial abundance between SSc and controls, with statistically significant differences evident in 5 of 7 clusters (paired t-test, $p < 0.05$ for all; **Figure 5A and B**). Among the most significant clusters were cluster 1, dominated by major lipophilic taxa, such as *Malassezia* and *Propionibacterium*, along with numerous Gram-positive Actinobacteria species. These organisms were significantly more abundant in healthy controls ($p < 0.001$ for Actinobacteria and *Propionibacterium* by paired *t*-test) and exhibited strong positive correlations to lipid metabolism and cell proliferation KEGG pathways **(Figure 5A and B)**. In contrast, cluster 5 exhibits substantial increases in a wide range of Proteobacteria and other Gram-negative taxa in SSc patients ($p < 0.001$ by paired *t*-test), and is strongly correlated with KEGG immune activation pathways, including Toll-like receptor (TLR) and transforming growth factor beta (TGF-β) signaling **(Figure 5A and D)**. Cluster 3 shows strong, positive correlations with immune activation, lipid metabolism, and Notch/Wnt signaling, and is associated with statistically significant decreases in Bacteroidetes levels in SSc patients ($p = 0.028$ by paired *t*-test), combined with modest increases in Proteobacteria, relative to controls ($p = 0.085$ by paired *t*-test; **Figure 5A and E**). These data demonstrate a strong association between underlying gene expression and the composition of the skin microbiome in SSc.

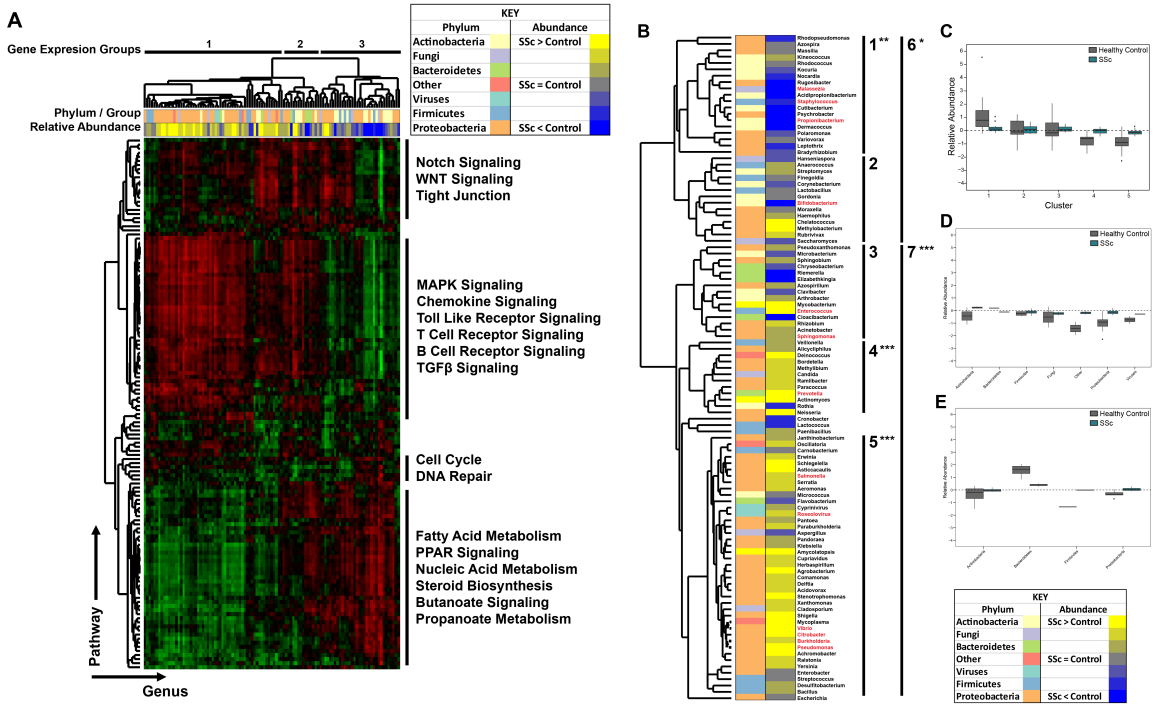


Figure 5. Microbiome Composition is Associated with Pathway Activation in SSc Skin. Single sample gene set enrichment analysis (ssGSEA) was run against normalized FPKM values for all 36 patient samples, using curated KEGG pathways as the probe gene sets. A correlation matrix was then generated by calculating Pearson's correlations for all combinations of ssGSEA values and genus-level abundance across all patients. **A.** Hierarchical clustering of the correlation matrix revealed strong associations between SSc-associated gene expression pathways and microbial composition. **B.** Taxonomic clustering based on gene expression. Hash marks indicate phylum/group associated with each sample. Relative abundance indicates the degree to which each genus is differentially present in SSc patients, relative to controls with yellow indicating abundance is higher in SSc, while blue indicates abundance is higher in controls. Black bars indicate KEGG pathways that clustered together hierarchically, with representative pathways listed alongside each cluster (*$p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ by paired *t*-test). Clinically relevant genera are highlighted in red. **C.** Relative abundance of all genera by taxonomic cluster. **D and E.** Distribution of taxa for cluster 5 (**D**) and cluster 3 (**E**).

Tasks 4 and 5. Develop a nested PCR-based assay followed by targeted multiplexed sequencing as a cost-effective method for screening archived skin biopsy RNA to determine the prevalence of microbiome components across the SSc population. Improvements in our sample-processing pipeline now allow for simultaneous extraction of DNA, RNA, and miRNA from all patient biopsies. DNA is being used as a template for targeted sequencing of the intergenic transcribed spacer regions (ITS), a region widely regarded as the gold standard for fungal species identification. To date, targeted ITS sequencing libraries have been analyzed from 48 archived samples (39 SSc and 9 controls), which includes both paired lesional and non-lesional skin as well as multiple time points from a single patient (Figure 5). Sequencing outputs are being analyzed by IMSA to identify differences in microbial diversity and species abundance between patients and controls, between lesional and non-lesion skin, as well as how these populations change over time.

Our preliminary findings from our RNA-seq data suggest significant bacterial dysbiosis in affected skin, with only modest changes in fungal abundance. Validation of these findings will therefore be run using targeted 16S sequencing to identify changes in bacterial composition. DNA has been collected for 116 skin samples (SSc and healthy controls) to date, with an additional 178 samples awaiting DNA purification. Targeted sequencing of the V2-4-8 and V3-6,7-9 hypervariable regions of 16S rRNA from these samples will be performed using the Ion 16S Metagenomics Kit (ThermoFisher) and sequenced on the Ion Torrent.

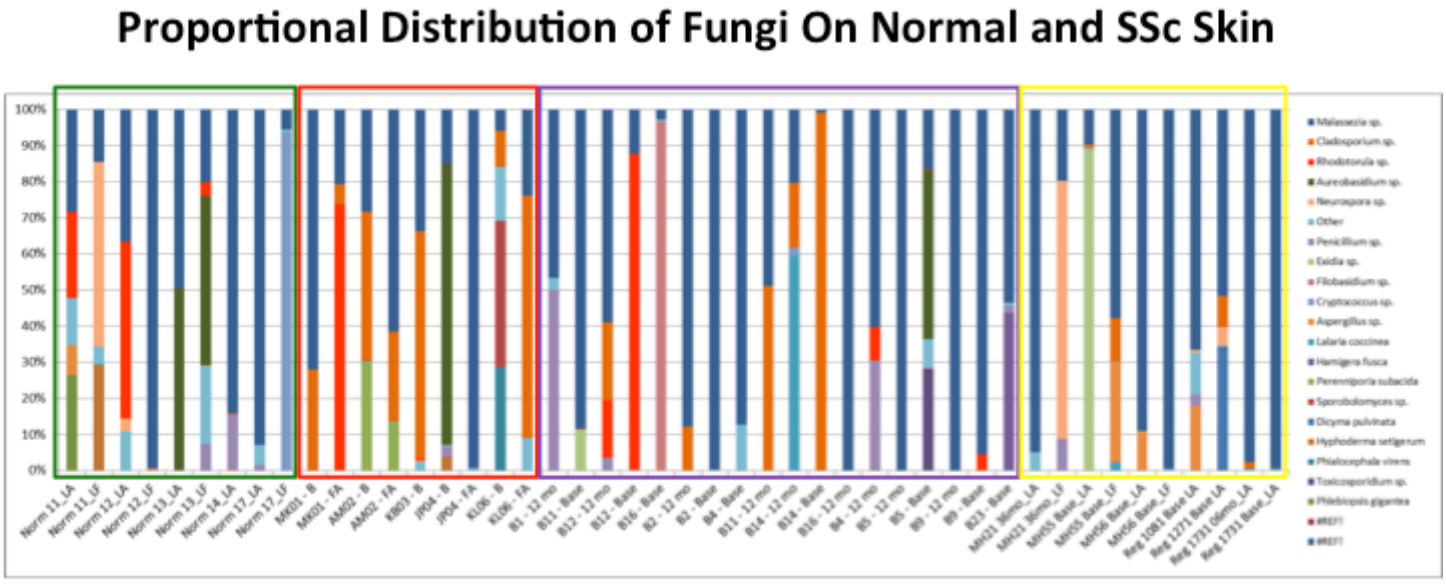


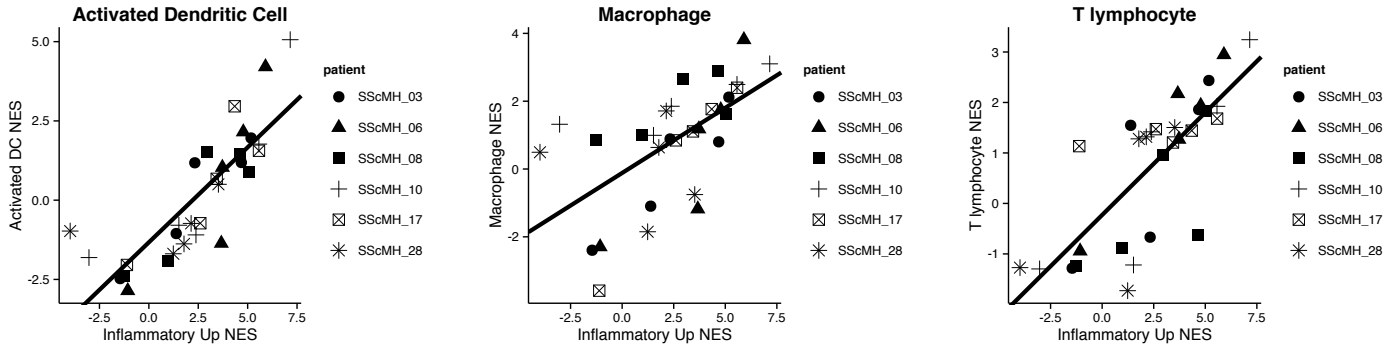
Figure 6. Targeted ITS sequencing of normal and SSc skin biopsies. Below is a preliminary analysis of targeted fungal ITS sequencing which shows a subset of patients have increased *R. glutinis* sequences (red). The most prominent fungal species detected on skin were *Malassezia* spp. (blue), the most common genus of skin commensal fungi.

Task 6: Culture microbiome components from the skin of SSc patients. As mentioned above, *swabbing of affected skin of SSc patients has unfortunately yet to result in recovery of clinically relevant fungi or other organisms. Swabbing and other culture-based methods of microbial detection will be revisited following completion of our RNA-seq analyses, enabling targeted isolation of organisms associated with SSc lesional skin.* This may be due to the use of antiseptics prior to biopsy collection as a means of preventing infection of the biopsy site.

Milestone 2: Identify the inflammatory infiltrates in SSc skin and their response to microbiome components

Task 1: Computational prediction of inflammatory cell infiltrates from genomic expression data. We have used single sample Gene Set Enrichment Analysis (ssGSEA) to identify the cellular subsets in SSc skin at different stages of disease. We first benchmarked the ssGSEA method in my laboratory using publicly available gene expression data from pools of cell lines that had a known composition (data not shown). These data demonstrated that ssGSEA accurately predicted cell type enrichment. We then analyzed a set of patients for whom we had whole genome expression data and that had strong expression of the inflammatory signature. We find the inflammatory signature is most strongly correlated with gene expression signatures from activated Dendritic Cells (DCs) and macrophages (MØs) (Figure 6). These methods are being applied in conjunction with the samples being analyzed in milestone 1.

Figure 7. Correlation of cell type signatures with a patient's inflammatory signature normalized enrichment score (NES). The inflammatory signature in SSc skin is most highly correlated with activated DCs and MØs.



Task 2: Perform immunohistochemistry to validate the computational predictions of task 1 above. We have optimized markers for different cell types in SSc skin.

We have optimized markers for different cell types in SSc skin. We have used CD163 for macrophages, CD1c for myeloid dendritic cells (mDCs) and CD3 for T cells in a separate study. We can now use these markers to look at innate and adaptive immune cells in the patients of this study. We have optimized these stains in a set of SSc samples and will now be performing these stains in samples for this study. T cell subsets will be identified using antibodies against CD4 and CD8, and CD19 and CD20 will be used to determine B cell localization.

Task 3: Develop protocols for the isolation and characterization of immune cells from skin using the sclerodermatous Graft-Versus Host Disease (sclGVHD) mouse including detailed characterization of cell types.

We established the sclGVHD model in the laboratory and can recapitulate both skin thickening (Figure 8) and the aberrant gene expression profiles observed in our prior studies (Figure 9). We have performed initial cell isolations and phenotyping of these samples (Figure 10).

A. Allogeneic Transfer (sclGVHD) **B. Syngeneic Transfer (control)**

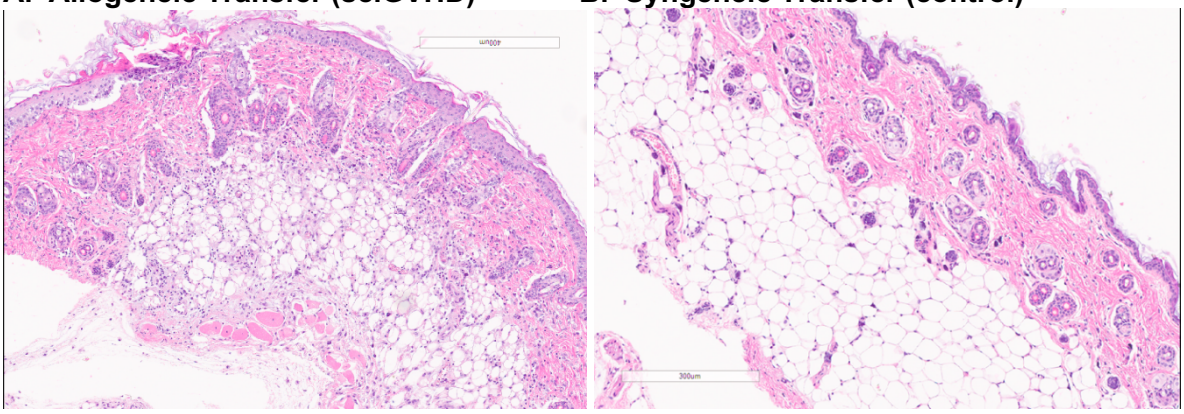
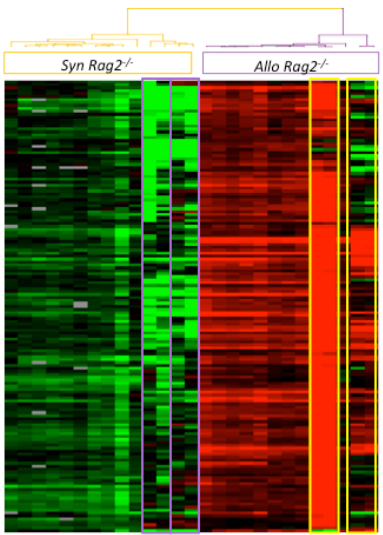


Figure 8. Immunohistochemistry was performed to show increased fibrosis at 2 weeks after disease initiation. As expected we observe skin thickening in the sclGVHD mouse that is not observed in controls.



Gene expression analyses were performed on skin biopsies from the sclGVHD mouse and compared to our prior study of this model [12]. We find gene expression changes were produced and consistent with those observed when the model was generated at Harvard (Figure 9). Therefore, we can clearly reproduce this model faithfully, including the molecular SSc phenotype.

Figure 9: Gene expression analysis of the sclGVHD mouse. We performed gene expression microarray analyses of the skin of the sclGVHD mouse generated in Dr. Whitfield's lab for this study with data from the sclGVHD taken from Greenblatt et al. 2012 [12]. We find that

samples from mice generated in this project (highlighted boxes) faithfully recapitulate the aberrant gene expression observed in our original study. These gene expression data are consistent with what we observe in the inflammatory subset of SSc.

Skin tissue was minced and digested with Collagenase D and DNase I and filtered through 70- and 40-micron mesh to facilitate cell dispersion. Single cell suspensions were stained with antibodies directed against the pan-leukocyte marker CD45, myeloid cell markers CD11b and CD11c, and CD115, CD206, and murine MHC-Class II (IA/IE) for flow cytometric analysis (Figure 10). Gating of positively stained cells was determined using fluorescence-minus-one (FMO) controls. CD45 positive live cells were gated and surface expression of CD11c, which is a murine dendritic cell marker, and CD11b, which is highly expressed on mouse macrophages, was analyzed on the $CD45^+$ cell population (Figures 10A and 10B). Consistent with previous reports [12, 13], there is an increase in skin macrophages ($CD11b^+CD11c^-$) in sclGVHD mice compared with syngeneic transplant controls. As demonstrated in Figure 10C, the macrophage cell population is characterized by significantly increased expression of the CSF-1R CD115 and murine MHC Class II (IA/IE). Surface levels of the mannose receptor CD206 were also elevated, although they did not reach statistical significance. These findings are consistent with results obtained by Greenblatt et al., indicating that we have established this model for our future analysis. Furthermore, these results suggest that the activation profile of SSc macrophages is unlikely to conform to a uniformly pro- or anti-inflammatory polarization state, as CD115 and CD206 are typically expressed by alternatively activated macrophages and enhanced IA/IE surface levels are characteristic of pro-inflammatory macrophages.

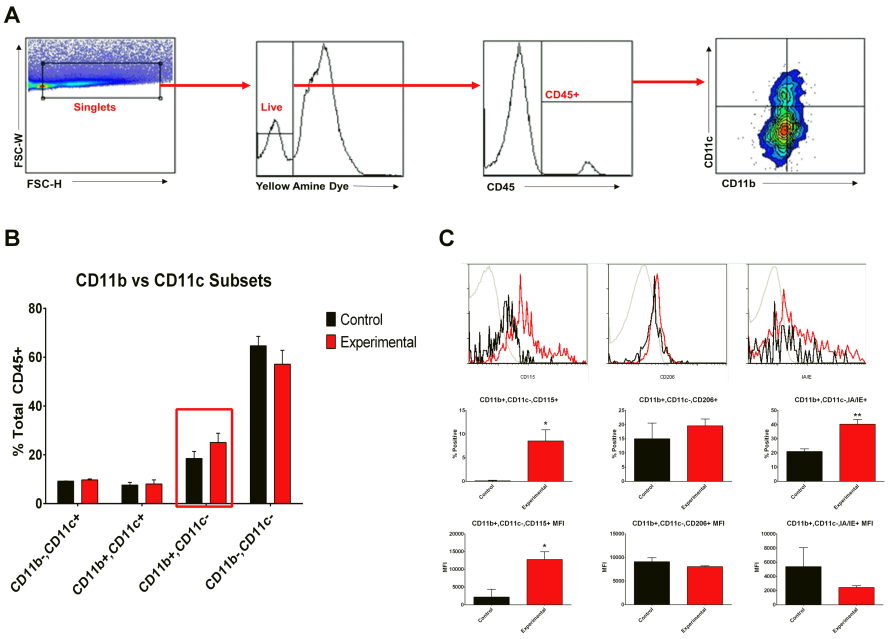


Figure 10. Characterization of the Myeloid Cell Population in skin of the 2-week-old sclGVHD mouse. Flow cytometric analysis of $CD45^+$ cells derived from back skin of BALB/c *Rag2*$^{-/-}$ hosts that received either syngeneic BALB/c or allogeneic B10.D2 splenocytes two weeks prior to cell harvest (n=3 per group): A. Gating strategy for selection of CD11b vs. CD11c positive cells. B. Percentage of CD11b/CD11c cell populations in syngeneic (control) vs. allogeneic (experimental) recipients measured in panel A. and C. Percentages and mean fluorescence intensity (MFI) of three characteristic macrophage markers (CD115, CD206, and IA/IE) on gated macrophage population ($CD11b^+/CD11c^-$) in syngeneic (black bars) vs. allogeneic (red bars) recipients.

Task 4: Identify the secreted mediators of fibrosis / inflammation being produced (Whitfield / Pioli). Once cells are isolated, we will screen for secreted pro-fibrotic mediators. In related studies, we have demonstrated that co-culture of human SSc macrophages with fibroblasts results in fibrotic activation. To identify candidate factors released by macrophages that potentially mediate this activation, RNA and supernatants were collected from differentiated SSc macrophages and analyzed by qRT-PCR and ELISA. As demonstrated in Figure 11, both mRNA and secreted protein levels of TGF-beta, IL-6, and CCL2 were elevated under basal conditions in SSc macrophages compared with healthy control macrophages. Notably, each of these factors has been implicated in the regulation of fibrosis. Multiplex analysis of SSc macrophage supernatants suggests additional factors may underlie fibrotic activation, including PDGF. Current studies are focused on the use of blocking antibodies to determine the relative contribution of these (and potentially other regulators) to the induction and maintenance of fibrotic activation. We find that LPS exposure does not increase the activation of SSc derived macrophages.

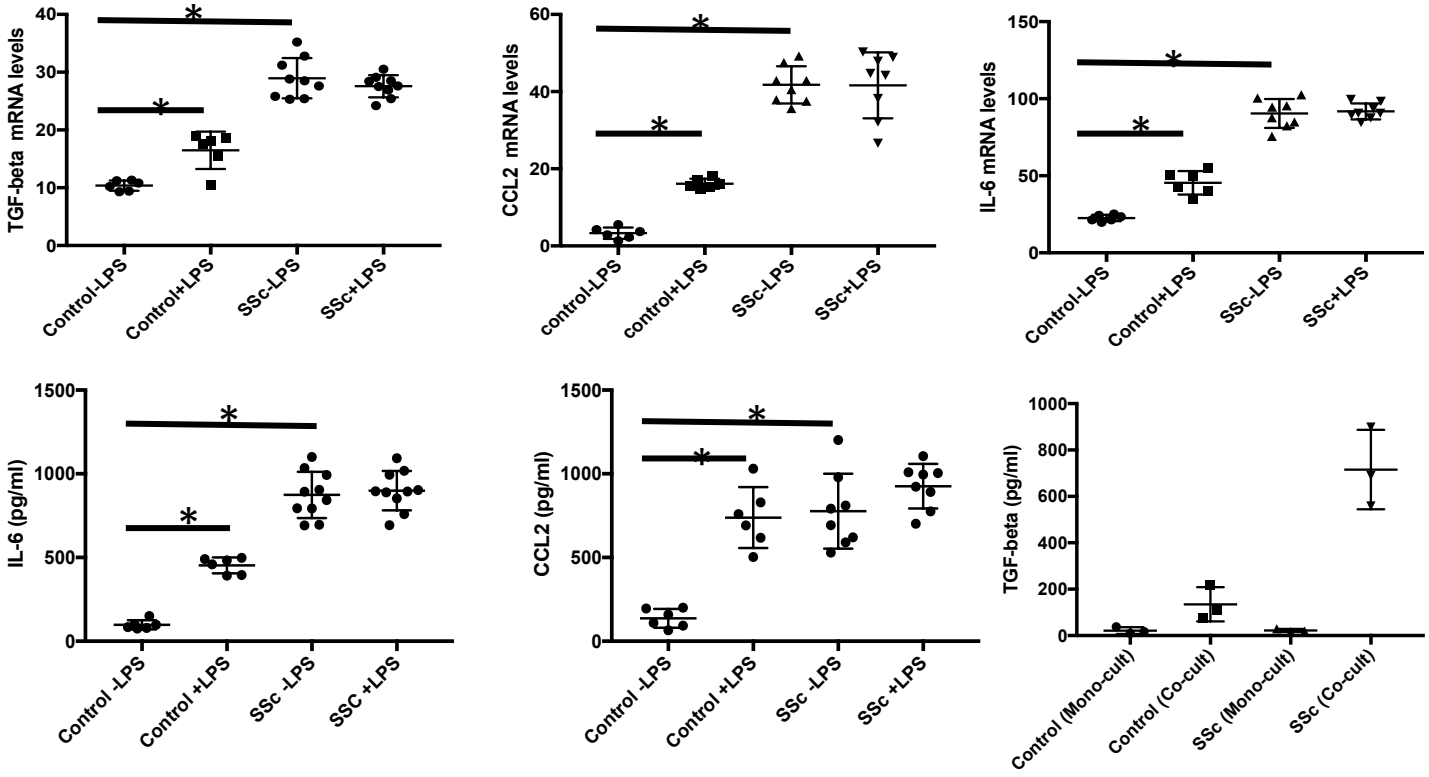


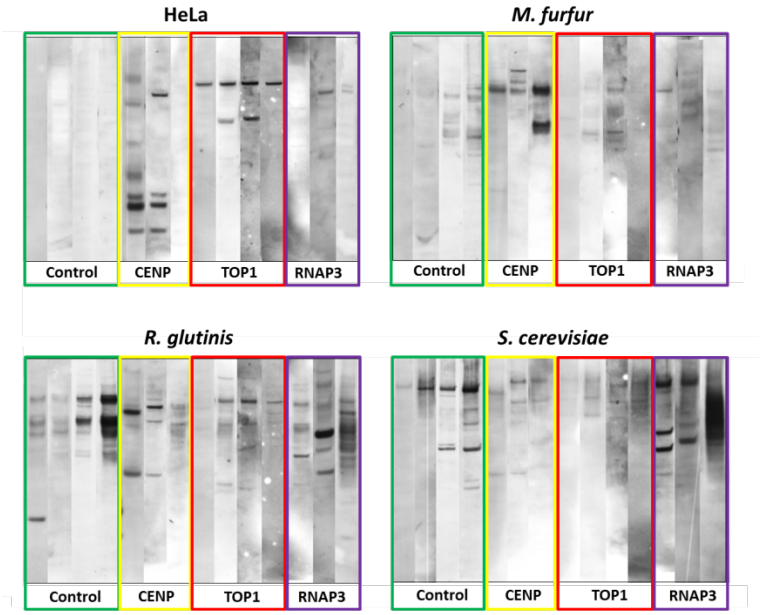
Figure 11. **SSc macrophages produce elevated levels of pro-fibrotic mediators.** SSc patient or control monocytes were differentiated in autologous plasma for 5 days and stimulated or not with 10 ng/ml LPS. RNA was extracted and sups were collected from cell cultures. Cytokine levels were analyzed by qRT-PCR for analysis of mRNA expression (A) or by ELISA for secreted protein production (2B). n=7 controls and 8 SSc patients, $p<0.05$

Task 5: Apply protocols to characterize the inflammatory infiltrate in the skin of SSc patients. We have optimized a protocol for the digestion of human skin that consists of Collagenase IV and DNAse I.

Milestone 3: Determine if SSc patients have a specific immune response against *R. glutinis* that is different from healthy controls and if this response can drive fibrosis.

Task 1: Test patient sera for cross-reactivity against *R. glutinis* antigens. In these experiments, we set out to test the hypothesis that autoantibody reactivity observed in SSc could recognize the same proteins in fungi, indicating that autoantibodies may have originated in response to fungal infection. Western blots were performed using R. glutinis, Malassezia furfur, Saccharomyces cerevisiae, and HeLa whole cell lysates (to test cross-reactivity with humans) and probed with sera collected from both healthy controls and SSc patients representing the three major autoantibody groups (Controls, CENP, TOP1, and RNAP3). Clear differences in cross-reactivity were evident between patient subsets. SSc patients showed a pattern of cross reactivity against R. glutinis lysates that was distinct from that observed in healthy controls. Among clinical autoantibody groups, a band consistent with the presence of TOP1 was seen in 3 of 4 TOP1 patients against R. glutinis and HeLa cells (Figure 7) this band was not observed in either M. furfur or S. cerevisiae, suggesting the possibility of cross-reactivity between R. glutinis and human TOP1 (Figure 7). Specific cross reactivity was also observed in CENP and RNAP3 patients; the identity of these proteins is being investigated.

Figure 11. Western blots using SSc and control Sera



Task 2. Identify the cross-reacting proteins by mass spectrometry. Serum-immunoprecipitation of *R. glutinis* and HeLa cell whole cell lysates followed by mass spectrometry was performed to identify immunoreactive proteins associated with *R. glutinis*. As part of this investigation, we performed a high-throughput analysis of all autoantibodies present in SSc sera, revealing novel targets and associated processes associated with SSc, which is now in press (Johnson, et al. *Arthritis Research and Therapy*, 2016).

Serum-immunoprecipitation of *R. glutinis* whole cell lysates followed by mass spectrometry revealed considerable reactivity in both SSc patients and healthy controls; however, identification of target peptides was not possible due to the absence of a sufficiently well-annotated *R. glutinis* proteome. Cross reactivity of autoantibodies against human and fungal proteins has been confirmed in *S. cerevisiae*, with autoantibodies against human TOP1 strongly cross-reactive to the equivalent protein in *S. cerevisiae*.

Task 3: Use isolated PBMCs and isolated monocytes to examine the cytokines secreted and changes in gene expression when cells are exposed to *R. glutinis* or other putative micro / mycobiome triggers.

We have shown that PBMCs and macrophages from SSc patients (monocytes isolated from peripheral blood of SSc patients that are differentiated in autologous sera) are activated under basal conditions, as these cells secrete pro-and anti-inflammatory mediators in the absence of exogenous activation (Figure 11). Therefore, additional stimulation with LPS does not appear to further activate the cells. These activated cells produce a wide range of pro-fibrotic molecules that have been implicated in SSc, including IL-6 and TGFbeta. We believe these are the major drivers of fibrosis in SSc.

Task 4: Determine if chronic exposure to *R. glutinis* or other micro / mycobiome components stimulate a fibrotic response in a mouse model of SSc.

> *Despite attempts to carry out this task, we encountered technical problems and were unable to complete this as planned.*

CONCLUSION:

We completed the majority of tasks associated with each milestone during the course of our grant. This has resulted in several publications, including one that is in the final stages of review (Johnson et al. see appendix). Analyses of the data have revealed widespread dysbiosis in SSc skin. We have identified activated DCs and MØs as the key cell types driving the inflammatory signature, a phenotype consistent with the presence of a mycobiome trigger. We have generated the scIGVHD mouse model, optimized our cell isolation procedures and methods for isolating macrophages from these mice and experiments are underway to demonstrate the key role of MØs as fibrotic drivers. We have analyzed the cross-reactivity of autoantibodies with human and fungal components. A paper reporting the autoantibody cross-reactivity to human proteins in HeLa cells has been published [14].

KEY RESEARCH ACCOMPLISHMENTS

- RNA-seq followed by metagenomic analyses for microbiome components reveals dysbiosis as a common feature of SSc skin, which increases with disease duration.
- Our work has shown that the innate immune system (macrophages and DCs) are likely drivers of SSc and that these cells respond to microbiome components.
- We have successfully established the sclGVHD model in the laboratory and are examining the immune drivers and their response to microbiome components.
- Cross reactivity with microbiome component is observed and there are data to suggest basal activation of immune cells in SSc.

The next reporting period:

N/A

4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

The major impact of this project is that we are demonstrating a novel paradigm for the initiation of SSc. This has the potential to dramatically change the way we think about SSc and the role of the innate immune system in driving disease.

What was the impact on other disciplines?

This study impacts areas of genomics, metagenomics, microbiology, innate immunity, and autoimmunity. The methods we demonstrate and develop here will affect all of these fields. In particular, this study begins to develop methods for both systems biology and metagenomic sequencing analyses that can be used in other rare diseases.

What was the impact on technology transfer?

Technical demands associated with this project necessitated the development of a novel method to isolate DNA, RNA, and miRNAs from a single skin biopsy. This method has been submitted as a disclosure to our technology transfer office (TTO). We pursued this with our TTO office and it was determined the method was too similar to an existing protocol and therefore could not be patented.

What was the impact on society beyond science and technology?

Systemic sclerosis (SSc) is a disease that has no FDA approved therapies and often has a very poor prognosis. If our metagenomic results are confirmed, this will provide not only a better understanding of the molecular processes driving disease pathogenesis, but also identify alternative strategies, such as anti-fungal treatment, as a possible treatment for SSc.

5. CHANGES/PROBLEMS

None to report.

6. PRODUCTS:

None at this time.

Oral Presentations: (Chronological Order)

Presentations for Michael L. Whitfield, PhD

11/18 "*Systems biology, machine learning and bioinformatics approaches to understand pathogenesis and direct treatment of systemic sclerosis*" Boehringer Ingelheim, Ridgefield CT

11/18 "*An update on molecular subsets, the microbiome, and clinical trials in systemic sclerosis*" Dartmouth-Hitchcock Medical Center, Division of Rheumatology, Lebanon NH

10/18 "*Machine Learning Classification of Peripheral Blood Gene Expression Identifies a Subset of Patients with Systemic Sclerosis Most Likely to Show Clinical Improvement in Response to Hematopoietic Stem Cell Transplant*" American College of Rheumatology Annual Meeting, Chicago, IL

10/18 "*Translation Genomics and Bioinformatics Core: Applications of Machine Learning, Epigenetics and Intrinsic Molecular Subsets in Systemic Sclerosis*" P50 Advisory Board Meeting, Pittsburgh, PA

9/18 "*Molecular response to abatacept in SSc skin biopsies from the ASSET trial*" Bristol Myers Squib, Princeton NJ.

5/18 "*Molecular patient stratification and quantitative measurement of therapeutic responses in SSc clinical trials*" Scleroderma Clinical Trials Consortium Meeting, Philadelphia, PA

5/18 "*Distribution of African American Patients in the SSc Molecular Subsets*" GRASP Consortium Meeting, NIH, Bethesda MD

3/18 "*Immune-fibrotic interactions in SSc: lessons from data integration, multi-tissue genomics, and 3D tissue models*" Scleroderma Research Foundation Annual Workshop, San Francisco CA

2/18 "*Multi-Organ RNA-Sequencing Of Patients With Systemic Sclerosis (SSc) Finds That Intrinsic Subsets Are Conserved Across Organ Systems*" Scleroderma World Congress, Bordeaux, France

---

12/17 "*Understanding pathogenesis of systemic sclerosis and related fibrotic conditions through functional genomic networks*" Rheumatology Grand Rounds, Hospital for Special Surgery, New York, NY

12/17 "*Systemic Sclerosis Phase 2 Study: Gene Expression Patterns effect of Lenabasum on gene in skin biopsies from systemic sclerosis patients in the JBT101-SSc-001 study*" Corbus Pharmaceuticals Anabasum phase 3 Investigator Meeting, Chicago, IL

10/17 "*Systems Biology of scaring and fibrosis*" ScarFree Foundation, London UK

9/17 "*Molecular stratification and pathway evaluation in systemic sclerosis*" Third Rock Ventures, Boston MA

5/17 Organizer and Chair, "*Big Data in the Life Sciences Symposium*". Burroughs Wellcome Training Grant, Dartmouth College Hanover NH

5/17 "*RNA-sequencing to assess the molecular response to abatacept in SSc skin biopsies from the ASSET trial*" NIH ACE Investigators meeting, Bethesda MD

3/17 "*Molecular Characterization of Multi-Organ System Involvement in Patients with Systemic Sclerosis*" Scleroderma Research Foundation Annual Workshop, San Francisco, CA

3/17 "*Systemic Sclerosis Phase 2 Study: Gene Expression Patterns EFFECT OF ANABASUM ON GENE EXPRESSION IN SKIN BIOPSIES FROM SYSTEMIC SCLEROSIS PATIENTS IN THE JBT101-SSc-001 STUDY*" Corbus Research Day, NY, NY

2/17 "*Systems Biology in SSc*", UCB Fibrosis Summit, London, UK

12/16 "*Systems biology and bioinformatics approaches to understand complex autoimmune diseases*" University of Michigan, Computational Medicine and Bioinformatics Seminar Series.

8/16 "*Personalized medicine in SSc*". Scleroderma Foundation New Hampshire Patient Support Group, Dartmouth-Hitchcock Medical Center, Hanover NH

6/16 "*Systems biology and bioinformatics approaches to evaluating disease mechanisms and therapeutic trials in autoimmunity and fibrosis*" University of North Carolina at Chapel Hill, Chapel Hill, NC

6/16 "*Burroughs Wellcome Training Program: Big Data in the Life Sciences*" Burroughs Wellcome Fund, Durham NC

5/16 "*Systems biology and bioinformatics approaches to evaluating disease mechanisms and therapeutic trials in autoimmunity and fibrosis*" University of Michigan, Rheumatology Grand Rounds, Ann Arbor MI

4/16 "*The genome and scleroderma's social network*" Scleroderma Foundation New England, Patient Education Seminar, Boston MA

3/16 "*Multi-organ systems biology reveals a common immune-fibrotic axis in systemic sclerosis, pulmonary fibrosis and pulmonary arterial hypertension*" Scleroderma Research Foundation Annual Workshop, San Francisco, CA

2/16 "*Systems biology and bioinformatics approaches to understand complex autoimmune diseases*" University College London, London UK

2/16 "*Integrative, multi-organ systems biology of systemic sclerosis reveals a macrophage signature associated with disease severity in multiple end-target tissues*" Scleroderma World Congress, Lisbon Portugal

2/16 "*Systems biology and bioinformatics approaches to understand complex autoimmune diseases*" Pittsburgh School of Medicine, Pittsburgh, PA

2/16 "*Advances in understanding pathogenesis and treatment in systemic sclerosis*" Rheumatology Grand Rounds, Dartmouth-Hitchcock Medical Center.

12/15 "*Big Data in the Life Sciences*" North Carolina Central University, Durham, NC

12/15 "*Big Data in the Life Sciences*" North Carolina State University, Raleigh, NC

12/15 "*Big Data in the Life Sciences*" University of North Carolina, Chapel Hill, NC

11/15 "*Multi-tissue genomic networks and systems biology in systemic sclerosis*". Scleroderma Foundation Workshop, ACR Annual meeting, San Francisco CA

8/15 *"Systems Biology in Systemic Sclerosis."* Session Chair and topic introduction. Scleroderma Basic Science Workshop, Cambridge UK.

6/15 "*Defining overlapping pathology between SSc patients and commonly used mouse models of disease*" Actelion, Basel Switzerland. (*cancelled due to illness*)

6/15 "*Genomic and Proteomic Quantification of the Heterogeneity of SSc: Implications for Pathogenesis and Treatment*". EULAR. Rome, Italy. (*cancelled due to illness*)

6/15 "*Genomics, Bioinformatics and Systems Biology for Precision Medicine in Systemic Sclerosis*". NIH CORT (P50) Advisory Committee meeting. Boston University Medical Center, Boston MA

4/15 "*Genomics, Bioinformatics and Systems Biology for Precision Medicine in Systemic Sclerosis*". SScores (NIH P30) Advisory Committee meeting. Boston University Medical Center, Boston MA.

3/15 "*A macrophage-associated inflammatory signature is found in all SSc tissues and associated with more severe disease*" Scleroderma Research Foundation Workshop on Scleroderma, San Francisco, CA

3/15 "*Molecular stratification and drug response for SSc clinical trials*" Pfizer, Cambridge, MA.

2/15 "*Enabling Precision Medicine in SSc Clinical Trials*" Discussion leader and presenter, NIAMS roundtable discussion on Scleroderma: Advancing Potential Drugs to Patient Care

2/15 "*Linking autoimmune systemic sclerosis and cancer: disease stratification, co-expression networks and genetic polymorphisms*" Cancer Mechanisms Program, Norris Cotton Cancer Center.

1/14 "*Mechanisms of Systemic Sclerosis (Scleroderma) pathogenesis by systems level genomic analyses*" Genomic Medicine Grand Rounds, Geisel School of Medicine.

12/14 "*Untangling molecular changes in SSc clinical trials: Gene expression subsets, response signatures and pathway changes*" ASSET Investigator Meeting. University of Michigan, MI

11/14 "*Identification of the Microbiome As a Potential Trigger of Systemic Sclerosis By Metagenomic RNA-Sequencing of Skin Biopsies*" ACR Basic Research Conference Boston, MA.

Abstracts and Presentations: (Chronological Order)

1. Michael E. Johnson, Zhenghui Li, Michelle T. Dimon, Tammara A. Wood, Robert Lafyatis, Sarah T. Arron, **Michael L. Whitfield.** Identification of the microbiome as a potential trigger of systemic sclerosis by metagenomic RNA-sequencing of skin biopsies. American College of Rheumatology Annual Meeting, 2014

2. Zhenghui Li, Eleni Marmarelis, Kun Qu, Lionel Brooks, Patricia A. Pioli, Howard Y. Chang, Robert Lafyatis, and **Michael L. Whitfield**. RNA-seq and miR-seq analysis of SSc skin across intrinsic gene expression subsets shows differential expression of non-coding RNAs regulating SSc gene expression. American College of Rheumatology Annual Meeting, 2014

3. Espinoza M, Mehta BK, Wang Y, Hoffmann A, Carns MA, Kosarek N, Wood TA, Hinchcliff M, Whitfield ML. Characterization of the Esophageal Microbiome in Patients with Systemic Sclerosis (SSc) [abstract]. *Arthritis Rheumatol.* 2018; 70 (suppl 10)

Manuscripts:

The following manuscripts from Dr. Whitfield's lab have relevance to this proposal. Publications #1, #5, #11, and #12 directly derives from work performed to accomplish the aims of this proposal.

1. Arron ST*, Dimon MT, Li Z, Johnson ME, Wood T., Feeney L, Angeles JG, Lafyatis R, **Whitfield ML***. High *Rhodotorula* sequences in skin transcriptome of patients with diffuse systemic sclerosis. J. Invest Derm. 2014, Mar 7. doi: 10.1038/jid.2014.127.
2. Johnson ME, Mahoney JM, Marmarelis E, Sargent JR, Wu MR, Spotts K, Hinchcliff M, **Whitfield ML***. Experimentally-derived fibroblast gene signatures identify molecular pathways associated with distinct subsets of systemic sclerosis patients in three independent cohorts. PLoS One. 2015 Jan 21;10(1):e0114017. doi: 10.1371/journal.pone.0114017. eCollection 2015.

3. Mahoney JM, Taroni J, Martyanov V, Wood TA, Greene CS, Pioli PA, Hinchcliff M, **Whitfield ML***. Systems level analysis of systemic sclerosis shows a network of immune and profibrotic pathways connected with genetic polymorphisms. PLoS Comput Biol. 2015 Jan 8;11(1):e1004005. doi: 10.1371/journal.pcbi.1004005. eCollection 2015 Jan.
4. Taroni JN, Martyanov V, Wood TA, Choe S, Huang CC, Hirano I, Yang GY, Brenner D, Jung B, Carns M, Podluski S, Chang RW, Varga J, **Whitfield ML**, Hinchcliff M. Genome-wide gene expression analysis of systemic sclerosis esophageal biopsies identifies disease-specific molecular subsets. Arthritis Res Ther. 2015 Jul 29;17:194. doi: 10.1186/s13075-015-0695-1
5. Michael E. Johnson, Andrew V. Grassetti, Jaclyn N. Taroni, Shawn M. Lyons, Devin Schweppe, Jessica K. Gordon, Robert F. Spiera, Robert Lafyatis, Paul J. Anderson, Scott A. Gerber, **Michael L. Whitfield**. Stress Granules and RNA Processing Bodies are Novel Autoantibody Targets in Systemic Sclerosis. Arthritis Res Ther. 2016 Jan 22;18:27. doi: 10.1186/s13075-016-0914-4.
6. Sargent JL, Li Z, Aliprantis AO, Greenblatt M, Lemaire R, Wu MH, Wei J, Taroni J, Harris A, Long KB, Burgwin C, Artlett CM, Blankenhorn EP, Lafyatis R, Varga J, Clark SH, **Whitfield ML**. Identification of Optimal Mouse Models of Systemic Sclerosis by Interspecies Comparative Genomics. Arthritis Rheumatol. 2016 Aug;68(8):2003-15. doi: 10.1002/art.39658
7. Martyanov V, **Whitfield ML**. Molecular stratification and precision medicine in systemic sclerosis from genomic and proteomic data. Curr Opin Rheumatol. 2016 Jan;28(1):83-8. doi: 10.1097/BOR.0000000000000237. Review.
8. Taroni JN, Greene CS, Martyanov V, Wood TA, Christmann RB, Farber HW, Lafyatis RA, Denton CP, Hinchcliff ME, Pioli PA, Mahoney JM, **Whitfield ML**. A novel multi-network approach reveals tissue-specific cellular modulators of fibrosis in systemic sclerosis. Genome Med. 2017 Mar 23;9(1):27. doi: 10.1186/s13073-017-0417-1. PMID: 28330499
9. Taroni JN, Martyanov V, Mahoney JM, **Whitfield ML**. A Functional Genomic Meta-Analysis of Clinical Trials in Systemic Sclerosis: Toward Precision Medicine and Combination Therapy. J Invest Dermatol. 2017 May;137(5):1033-1041. doi: 10.1016/j.jid.2016.12.007. Epub 2016 Dec 21. PMID: 28011145
10. Hinchcliff M*, Toledo DM*, Taroni JN, Wood TA, Franks JM, Ball MS, Hoffmann A, Amin SM, Tan AU, Tom K, Nesbeth Y, Lee J, Ma M, Aren K, Carns MA, Pioli PA, **Whitfield ML**. Mycophenolate mofetil treatment of systemic sclerosis reduces myeloid cell numbers and attenuates the inflammatory gene signature in skin. J Invest Dermatol. 2018 Jan.
11. Johnson ME, Pioli PA, Whitfield ML. Gene expression profiling offers insights into the role of innate immune signaling in SSc. Semin Immunopathol. 2015 Sep;37(5):501-9. doi: 10.1007/s00281-015-0512-6. Epub 2015 Jul 30. Review. PMID: 26223504
12. Johnson ME, Franks JM, Cai G, Mehta BK, Wood TA, Archambault K, Pioli PA, Simms RW, Orzechowski N, Arron S, Whitfield ML. Microbiome dysbiosis is associated with disease duration and increased inflammatory gene expression in systemic sclerosis skin. Arthritis Res Ther. 2019 Feb 6;21(1):49. doi: 10.1186/s13075-019-1816-z.
13. Cai G, Franks JM, and **Whitfield ML**. Precision weighted mRNA abundance in RNA-seq differential expression analysis. *Submitted*
14. Bhandari R, Ball MS, Popovich D, Martyanov V, Carns M, Arroyo E, Aren K, Hinchcliff M, **Whitfield ML**, and **Pioli PA** Pro-fibrotic Activation of Human Macrophages in Systemic Sclerosis. *In preparation.*

The following additional papers were published by Drs. Whitfield, Lafyatis and Arron during the funding period.

1. Long KB, Li Z, Burgwin C, Cho SG, Martyanov V, Sassi-Gaha S, Earl J, Eutsey R, Ahmed A, Ehrlich GD, Artlett CM, **Whitfield ML**, Blankenhorn EP *. The Tsk2/+ mouse fibrotic phenotype is due to a gain-of-function mutation in the PIIINP segment of the *Col3a1* gene. J. Invest Derm. 2014, Oct 20. doi: 10.1038/jid.2014.455.

2. Iwamoto N, Vettori S, Maurer B, Brock M, Jüngel A, Calcagni M, Gay RE, **Whitfield ML**, Distler J.H.W, Gay S, Distler O*. Downregulation of miR-193b in systemic sclerosis regulates the proliferative vasculopathy by urokinase-type plasminogen activator expression. Ann Rheum Dis. 2014 Nov 10. pii: annrheumdis-2014-205326. doi: 10.1136/annrheumdis-2014-205326. [Epub ahead of print]

3. Marangoni RG, Korman B, Wei J, Wood TA, **Whitfield ML,** Scherer PE, Tourtellotte WG and Varga J*. Myofibroblasts in Cutaneous Fibrosis Originate from Intradermal Adipocytes. Arthritis Rheumatol. 2015 Apr;67(4):1062-73. doi: 10.1002/art.38990.

4. Chakravarty EF, Martyanov V, Fiorentino D, Wood TA, Haddon DJ, Jarrell JA, Utz PJ, Genovese MC, **Whitfield ML**, Chung L. A Pilot Randomized Placebo-Controlled study of Abatacept for the Treatment of Diffuse Cutaneous Systemic Sclerosis. Arthritis Research & Therapy, Arthritis Res Ther. 2015 Jun 13;17(1):159.

5. Fresolimumab treatment decreases biomarkers and improves clinical symptoms in systemic sclerosis patients. Rice LM, Padilla CM, McLaughlin SR, Mathes A, Ziemek J, Goummih S, Nakerakanti S, York M, Farina G, **Whitfield ML**, Spiera RF, Christmann RB, Gordon JK, Weinberg J, Simms RW, Lafyatis R. J. Clin. Invest. 2015 Jun 22. pii: 77958. doi: 10.1172/JCI77958

6. Lisa M. Rice, Jessica Ziemack, Eric Stratton, Sarah Mclauglin, Cristina Padilla, Allison Mathes, Romy Christmann, Giuseppina Stifano, Jeff Browning, **Michael L. Whitfield**, Robert Spiera, Jessica Gordon, Robert Simms, Yuqing Zhang, Robert Lafyatis. A longitudinal biomarker for the extent of skin disease in patients with diffuse cutaneous systemic sclerosis. Arthritis Rheumatol. 2015 Nov;67(11):3004-15. doi: 10.1002/art.39287.

7. Gordon JK, Martyanov V, Wood TA, Spiera RF, **Whitfield ML.** Nilotinib (Tasigna™) in the Treatment of Early Diffuse Systemic Sclerosis: An Open-Label, Pilot Clinical Trial. Arthritis Res Ther. 2015 Aug 18;17:213. doi: 10.1186/s13075-015-0721-3.

8. Brooks L, Lyons SM, Mahoney JM, Welch JD, Liu Z, Marzluff WF, and **Whitfield ML**. A multi-protein occupancy map of the histone mRNP. RNA. 2015 Nov;21(11):1943-65. doi: 10.1261/rna.053389.115. Epub 2015 Sep 16.

9. Lyons SM, Cunningham CH, Welch JD, Groh B, Guo AY, Wei B, **Whitfield ML**, Xiong Y, Marzluff WF. A subset of replication-dependent histone mRNAs are expressed as polyadenylated RNAs in terminally differentiated tissues. Nucleic Acids Res. 2016 Jul 8. pii: gkw620.

10. Franks JM, Cai G, **Whitfield ML**. Feature Specific Quantile Normalization Enables Cross-Platform Classification of Molecular Subtypes using Gene Expression Data. Bioinformatics. 2018 Jan 17 https://doi.org/10.1093/bioinformatics/bty026

11. Franks JM, Martyanov V, Cai G, Wood TA, and **Whitfield ML**. Novel Machine Learning Classifier Accurately Predicts Intrinsic Molecular Subsets for Patients with Systemic Sclerosis. *submitted*

12. Cai G, Franks JM, Mehta B, Nesbeth YN, Hinchcliff ME, **Whitfield ML**. A novel Skin Disease Severity Score (SDSS) as a surrogate for MRSS. *In preparation*

13. Mehta BK, Franks J, Cai G, Toledo D, Wood TA, Archambault KA, Kosarek N, Kolstad K, Stark M, Valenzuela A, Fiorentino D, Fernandez-Becker N, Becker L, Nguyen L, Clarke J, Boin F, Wolters P, Chung L and **Whitfield ML**. Multi-Organ RNA-Sequencing Of Patients With Systemic Sclerosis (SSc) Finds That Intrinsic Subsets Are Conserved Across Organ Systems, *In preparation*.

Degrees obtained that are supported by this award

Dr. Zhenghui Li, who worked on the microbiome project, completed his Ph.D during year 2 of funding. He has received direct support from this grant. Dr. Jaclyn Taroni received her PhD in 2015 and also contributed to the data analyses outlined in this grant. She was supported by an NIH T32 training grant during this time. This grant supported Dr. Michael Johnson during his Post-doctoral Fellowship. Although no degree was awarded, it supported his further professional training.

Development of cell lines, tissue or serum repositories

None

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

None

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: For collaborative awards, independent reports are required from BOTH the Initiating PI and the Collaborating/Partnering PI. A duplicative report is acceptable; however, tasks shall be clearly marked with the responsible PI and research site. A report shall be submitted to https://ers.amedd.army.mil for each unique award.

An identical final progress report will be sent from Dr. Arron

9. REFERENCES

1. Whitfield, M.L., et al., *Systemic and cell type-specific gene expression patterns in scleroderma skin.* Proc Natl Acad Sci U S A, 2003. **100**(21): p. 12319-24.
2. Milano, A., et al., *Molecular subsets in the gene expression signatures of scleroderma skin.* PLoS ONE, 2008. **3**(7): p. e2696.
3. Pendergrass, S.A., et al., *Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies.* J Invest Dermatol, 2012. **132**(5): p. 1363-73.
4. Hinchcliff, M., et al., *Molecular signatures in skin associated with clinical improvement during mycophenolate treatment in systemic sclerosis.* J Invest Dermatol, 2013. **133**(8): p. 1979-89.
5. Taroni, J.N., et al., *Molecular characterization of systemic sclerosis esophageal pathology identifies inflammatory and proliferative signatures.* Arthritis Res Ther, 2015. **17**: p. 194.
6. Arron, S.T., et al., *High Rhodotorula sequences in skin transcriptome of patients with diffuse systemic sclerosis.* J Invest Dermatol, 2014. **134**(8): p. 2138-45.
7. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
8. Franks, J.M., G. Cai, and M.L. Whitfield, *Feature Specific Quantile Normalization Enables Cross-Platform Classification of Molecular Subtypes using Gene Expression Data.* Bioinformatics, 2018.
9. Johnson, M.E., et al., *Experimentally-derived fibroblast gene signatures identify molecular pathways associated with distinct subsets of systemic sclerosis patients in three independent cohorts.* PLoS One, 2015. **10**(1): p. e0114017.
10. Hinchcliff, M., et al., *Molecular Signatures in Skin Associated with Clinical Improvement during Mycophenolate Treatment in Systemic Sclerosis.* J Invest Dermatol, 2013.
11. Dimon, M.T., et al., *IMSA: Integrated Metagenomic Sequence Analysis for Identification of Exogenous Reads in a Host Genomic Background.* PLoS ONE, 2013. **8**(5): p. e64546.
12. Greenblatt, M.B., et al., *Interspecies comparison of human and murine scleroderma reveals IL-13 and CCL2 as disease subset-specific targets.* Am J Pathol, 2012. **180**(3): p. 1080-94.
13. Zhang, Y., et al., *Murine sclerodermatous graft-versus-host disease, a model for human scleroderma: cutaneous cytokines, chemokines, and immune cell activation.* J Immunol, 2002. **168**(6): p. 3088-98.

14. Johnson, M.E., et al., *Stress granules and RNA processing bodies are novel autoantibody targets in systemic sclerosis.* Arthritis Res Ther, 2016. **18**(1): p. 27.

10. APPENDIX

1. Arron ST, Dimon MT, Li Z, Johnson ME, Wood T., Feeney L, Angeles JG, Lafyatis R, Whitfield ML. High Rhodotorula sequences in skin transcriptome of patients with diffuse systemic sclerosis. J. Invest Derm. 2014, Mar 7. doi: 10.1038/jid.2014.127.

2. Michael E. Johnson, Andrew V. Grassetti, Jaclyn N. Taroni, Shawn M. Lyons, Devin Schweppe, Jessica K. Gordon, Robert F. Spiera, Robert Lafyatis, Paul J. Anderson, Scott A. Gerber, Michael L. Whitfield. Stress Granules and RNA Processing Bodies are Novel Autoantibody Targets in Systemic Sclerosis. Arthritis Res Ther. 2016 Jan 22;18:27. doi: 10.1186/s13075-016-0914-4.

3. Johnson ME, Pioli PA, Whitfield ML. Gene expression profiling offers insights into the role of innate immune signaling in SSc. Semin Immunopathol. 2015 Sep;37(5):501-9. doi: 10.1007/s00281-015-0512-6. Epub 2015 Jul 30. Review. PMID: 26223504

4. Johnson ME, Franks JM, Cai G, Mehta BK, Wood TA, Archambault K, Pioli PA, Simms RW, Orzechowski N, Arron S, Whitfield ML. Microbiome dysbiosis is associated with disease duration and increased inflammatory gene expression in systemic sclerosis skin. Arthritis Res Ther. 2019 Feb 6;21(1):49. doi: 10.1186/s13075-019-1816-z.

Appendix

1. Arron ST, Dimon MT, Li Z, Johnson ME, Wood T., Feeney L, Angeles JG, Lafyatis R, Whitfield ML. High Rhodotorula sequences in skin transcriptome of patients with diffuse systemic sclerosis. J. Invest Derm. 2014, Mar 7. doi: 10.1038/jid.2014.127.

2. Michael E. Johnson, Andrew V. Grassetti, Jaclyn N. Taroni, Shawn M. Lyons, Devin Schweppe, Jessica K. Gordon, Robert F. Spiera, Robert Lafyatis, Paul J. Anderson, Scott A. Gerber, Michael L. Whitfield. Stress Granules and RNA Processing Bodies are Novel Autoantibody Targets in Systemic Sclerosis. Arthritis Res Ther. 2016 Jan 22;18:27. doi: 10.1186/s13075-016-0914-4.

3. Johnson ME, Pioli PA, Whitfield ML. Gene expression profiling offers insights into the role of innate immune signaling in SSc. Semin Immunopathol. 2015 Sep;37(5):501-9. doi: 10.1007/s00281-015-0512-6. Epub 2015 Jul 30. Review. PMID: 26223504

4. Johnson ME, Franks JM, Cai G, Mehta BK, Wood TA, Archambault K, Pioli PA, Simms RW, Orzechowski N, Arron S, Whitfield ML. Microbiome dysbiosis is associated with disease duration and increased inflammatory gene expression in systemic sclerosis skin. Arthritis Res Ther. 2019 Feb 6;21(1):49. doi: 10.1186/s13075-019-1816-z.

HHS Public Access

Author manuscript

J Invest Dermatol. Author manuscript; available in PMC 2015 February 01.

Published in final edited form as:

J Invest Dermatol. 2014 August ; 134(8): 2138–2145. doi:10.1038/jid.2014.127.

High *Rhodotorula* sequences in skin transcriptome of patients with diffuse systemic sclerosis

Sarah T. Arron[1,‡], **Michelle T. Dimon**[1], **Zhenghui Li**, **Michael E. Johnson**[2], **Tammara Wood**, **Luzviminda Feeney**[1], **Jorge Gil Angeles**[1], **Robert Lafyatis**[3], and **Michael L. Whitfield**[2,‡]

[1]Department of Dermatology, University of California, San Francisco, San Francisco, CA

[2]Department of Genetics, Dartmouth Geisel School of Medicine, Hanover, NH

[3]Department of Medicine, Boston University School of Medicine, Boston, MA

Abstract

Previous studies have suggested a role for pathogens as a trigger of systemic sclerosis (SSc), though neither a pathogen nor a mechanism of pathogenesis is known. Here we show enrichment of *Rhodotorula* sequences in the skin of patients with early, diffuse SSc compared to normal controls. RNA-seq was performed on four SSc and four controls, to a depth of 200 million reads per patient. Data were analyzed to quantify the non-human sequence reads in each sample. We found little difference between bacterial microbiome and viral read counts, but found a significant difference between the read counts for a mycobiome component, *R. glutinis*. Normal samples contained almost no detected *R. glutinis* or other *Rhodotorula* sequence reads (mean score 0.021 for *R. glutinis*, 0.024 for all *Rhodotorula*). In contrast, SSc samples had a mean score of 5.039 for *R. glutinis* (5.232 for *Rhodotorula*). We were able to assemble the D1–D2 hypervariable region of the 28S rRNA of *R. glutinis* from each of the SSc samples. Taken together, these results suggest *R. glutinis* may be present in the skin of early SSc patients at higher levels than normal skin, raising the possibility that it may be triggering the inflammatory response found in SSc.

Introduction

Systemic Sclerosis (SSc) is a rare and poorly understood systemic autoimmune disease that results in skin fibrosis and severe internal organ involvement. There is a limited understanding of its pathophysiology and there is little data to indicate what may trigger the disease. One in three patients dies within 10 years of diagnosis (Steen and Medsger 2007); there are no validated diagnostic markers and no curative treatments.

‡To whom correspondence should be addressed: Sarah T. Arron, Department of Dermatology, University of California, San Francisco, 1701 Divisadero Street, Box 0316, San Francisco, CA 94143-0316, tel: (415) 353-7839, fax: (415) 353-7838, arrons@derm.ucsf.edu and Michael L. Whitfield, Ph.D., Department of Genetics, Geisel School of Medicine at Dartmouth, 7400 Remsen, Hanover, NH 03755, phone 603.650.1109, fax 603.650.1188, michael.L.whitfield@dartmouth.edu.

Conflict of Interest

Dr. Whitfield has filed patents for gene expression biomarkers in systemic sclerosis and is a scientific founder of Celdara Medical LLC. The remaining authors declare no conflict of interest.

We have demonstrated gene expression based subsets within SSc patients (Whitfield et al. 2003; Milano et al. 2008; Chung et al. 2009; Sargent et al. 2009; Pendergrass et al. 2012) by analysis of skin biopsies in three independent cohorts (Milano et al. 2008; Pendergrass et al. 2012; Hinchcliff et al. 2013). Using genome-wide and bioinformatic-driven strategies (Sargent et al. 2009; Greenblatt et al. 2012), SSc patients may now be divided into pathway-centric subsets. These are the inflammatory, fibroproliferative, limited and normal-like subsets (Milano et al. 2008; Pendergrass et al. 2012; Hinchcliff et al. 2013). The inflammatory subset of patients is characterized by infiltrating immune cells that include T and B lymphocytes and macrophages (Milano et al. 2008; Greenblatt et al. 2012; Pendergrass et al. 2012; Hinchcliff et al. 2013). We have recently shown that two major pathways driving fibrosis in the inflammatory subset of patients are the profibrotic IL-13 and IL-4 pathways, which signal through a shared receptor IL-4RA (Greenblatt et al. 2012). We have also demonstrated that SSc patients that map to the inflammatory subset show improvement while taking a commonly used SSc therapeutic agent, mycophenolate mofetile (MMF), while the patients in the fibroproliferative subset do not show any clinical improvement (Hinchcliff et al. 2013). We have shown that the gene expression subsets are stable over periods of 6 – 12 months (Pendergrass et al. 2012) although recent meta-analysis of all published datasets suggest the groups may be long lived, but interconnected (Mahoney, Johnson, Whitfield, *Submitted*). In such a longitudinal model, the inflammatory group may be a key point in the initiation of disease since most genetic changes that have been associated with SSc risk occur in the immune system, suggesting that genetics along with some environmental trigger is an initiating event in SSc.

Identification of such environment triggers for most systemic autoimmune diseases has been elusive despite the significant health burden these diseases impose worldwide. Links have been suggested between SSc and cytomegalovirus, parvovirus B19, Epstein-Barr virus, endogenous retroviruses, and Chlamydia (Hamamdzic et al. 2002; Grossman et al. 2011), but these reports have not been substantiated. Hypotheses include molecular mimicry, in which homology between pathogen and self-peptides results in cross-activation of autoreactive lymphocytes; chronic inflammation and endothelial cell damage; and microbial superantigens activating immune response in the absence of cognate antigen (Grossman et al. 2011).

High-throughput sequencing technologies allow analysis of both host and pathogen-expressed sequences and genomes to identify exogenous viral, bacterial or fungal triggers of disease (the metagenome). Metagenomic analyses allow an unbiased assessment of all microorganisms in a complex disease sample. Here we present a comprehensive characterization of the metagenome in the skin of early, active SSc patients that show an inflammatory gene expression signature. Analysis of the full metagenome (including the microbiome, fungal mycobiome, and viral sequences) in patients with these earliest signs of disease, and mapping host and metagenomic sequences has identified a common environmental fungus, *Rhodotorula glutinis* as over-represented in SSc skin. Our preliminary studies suggest that disease pathogenesis may include a common environmental trigger that we hypothesize elicits immune activation in a permissive host genetic background.

Results

Patient characteristics

Lesional skin was obtained from the forearm of four patients with early, diffuse SSc within 6 months of first onset of non-Raynaud's symptoms (Table 1). All patients were in the inflammatory intrinsic subset. Two patients were untreated and two had received low-dose immunosuppression. Control skin was obtained from the forearm of healthy patients.

Initial metagenomic analysis shows fungal reads in SSc samples

We performed quality filtering and human sequence filtering using the human genome (hg19). Over 99% of the total readset was derived from human or nonhuman primate in both SSc and control samples. On average, 4×10^5 reads remained per sample after host filtering (Table 2). IMSA mapped reads to the NCBI non-redundant nucleotide (nt) database and generated taxonomy reports. In this analysis, each taxonomic level is given a score based on the number of reads aligning to sequences in that taxonomic category, where reads with multiple best alignments generate partial scores for each category with an alignment. Figure 1 demonstrates the breakdown of read scores in the dataset by taxonomic division. The microbial reads had a significantly different distribution between SSc and normal samples, with significantly more reads in SSc samples mapping to the plant and fungal division. At the top level, IMSA uses the GenBank divisions for an overview of metagenomic results. In this organization, plant and fungal sequences are combined (Ouellette and Boguski 1997). There were no significant differences in the scores derived from bacteria or viruses between SSc and control.

Figure 2 shows a heat map of taxonomy scores for bacterial, viral and fungal genera. For this analysis and all subsequent analyses, only reads with a single best alignment were retained to prevent noise from reads aligning across multiple species. Unsupervised clustering discriminated between normal and SSc samples, with the cluster driven by fungal genera in the *Basidiomycota* phylum (Figure 2). Common skin colonizers *Streptococcus*, *Propionibacterium* and *Malassezia* were represented across all samples, while SSc samples were enriched in genera of order *Sporidiobolales*, including *Rhodotorula*, *Rhodosporidium*, and *Sporobolomyces*.

SSc samples contain significantly more reads derived from Rhodotorula species

To determine the source of these differences between SSc and normal samples, TaxMaps were generated to visualize the scores of the taxonomic categories inside the plant and fungal division (Figure 3). Average TaxMaps for the normal samples showed *Malassezia globosa* and *Trimorphomyces papilionaceus* as the only species with an average normalized score above 0.05 (Figure 3A). By contrast, the SSc samples had more diverse fungal sequences, with *M. globosa*, *T. papilionaceus* but also *Bullera sakaeratica*, *Leucosporidium sp AY30*, *Rhodotorula hordea*, *Rhodotorula glutinis* and *Rhodotorula mucilaginosa* all showing average scores above 0.05 (supplemental information contains each individual TaxMap with all scores above 0.01). The most striking difference between SSc and normal TaxMaps is the large number of *R. glutinis* and *R. mucilaginosa* reads (Figure 3B). Normal samples averaged a total score of 0.55 for the entire plant and fungal division with no

species having an average score over 0.11. In the SSc samples, *R. glutinis* had an average score of 5.04, while the closely related *R. mucilaginosa* had an average score of 0.20.

SSc samples had a 252-fold increase in *R. glutinis* score per million total reads (normal mean score= 0.021, 95% CI −0.01–0.05, SSc mean score 5.039, 95% CI 2.97–7.11, Wilcoxon rank-sum p=0.01) (Figure 3C).

Assembly of *Rhodotorula* contigs indicates a species closest to *Rhodotorula glutinis*

Next we assembled longer *Rhodotorula* contigs from each SSc sample individually. We used PRICE for the assembly as it is designed to assemble paired-end reads in a complex metagenomic dataset into contigs (Ruby et al. 2013). We seeded contig assembly with *Rhodotorula* reads. The four normal skin samples had insufficient numbers of *Rhodotorula* reads for contig assembly. Each SSc sample generated at least one contig which aligned to *R. glutinis* 28S rRNA. The D1–D2 hypervariable region at the 5′ end of 28S rRNA of *R. glutinis* was covered in each sample. As sequence for this region is available for a wide variety of *Rhodotorula* fungal species, we used this area for further phylogenetic analysis. A multiple sequence alignment was performed using this region of 28S from our four SSc samples as well as selected sequences from NCBI for *Rhodotorula* and related fungal species. This alignment was used to create a phylogentic tree (Figure 4). The general structure of this tree is quite similar to other published phylogenetic trees for Rhodotorula (Biswas et al. 2001). The four sequences from the SSc samples cluster together, with sequences from *R. mucilaginosa*, *R. glutinis*, and *R. graminis*.

In addition, we aligned the original read sets against *R. glutinis* 28S rRNA (NCBI record FJ345357) and viewed the resulting alignment in IGV (Thorvaldsdóttir et al. 2013) (Figure 5). The alignment shows many more *Rhodotorula* reads in the SSc samples. In addition, the reads aligning in the SSc samples have fewer sequence differences from the NCBI record, suggesting their source is more similar to *R. glutinis* than the reads aligning from the normal samples. In addition, the normal samples have six bases in the region from 900–1050 bp that show variability within each normal sample, suggesting multiple *Rhodotorula* species may present.

PRICE was able to assemble additional contigs whose best alignment was to *Rhodotorula* genes, though the longest contig in each sample was a sequence whose best alignment in the nt database was to *R. glutinis* 28S rRNA (FJ345357). Given that rRNA is present at much higher quantities than other transcripts, this is not unsurprising. Other genes assembled were likewise genes expected to be present at high levels, such as *R. glutinis* 18S rRNA (HQ420261.1), *R. mucilaginosa* 18S rRNA (X84326.1) and *R. taiwanensis* RS1 complete mitochondrial genome (HF558455.1) as identified by best hit in the nt database. *R. taiwanensis* is a recently identified, novel *Rhodotorula* species closely related to *R. mucilaginosa* and *R. glutinis var dairenensis* whose entire mitochondrial genome was recently sequenced (Huang et al. 2011) (Zhao et al. 2013). It is one of the few *Rhodotorula* mitochondrial genomes in the database and likely indicative *R. glutinis* mitochondrial sequences, which are absent from the database.

Discussion

Previous studies have suggested a role for pathogens as a trigger of SSc, though neither the pathogen nor the mechanism of pathogenesis is known (Grossman et al. 2011). Here we present RNA-seq data on the microbial species present in skin samples from four patients presenting with early, diffuse SSc and four normal patients. Human reads were filtered from the read sets and the resulting reads were aligned to the NCBI nt database to quantify the non-human species present as inferred by the number of reads aligning to each species. While the quantity of most microbial species showed no difference between normal and SSc samples, *R. glutinis* levels were significantly higher in SSc samples. While normal patients had almost no detectable *R. glutinis* sequences (mean score 0.02 per million reads), patients with SSc had a mean *R. glutinis* score of 5.04 per million reads. Further, we demonstrate that a 28S rRNA sequence most similar to *R. glutinis* can be assembled from each of the SSc samples. Additional studies are needed to definitively determine which species of *Rhodotorula* are present in these samples, although at the level of sequence available from our RNA-seq assemblies, the species appears to be most similar to *R. glutinis* (NCBI record FJ345357) though there appears to also be similarity to *R. mucilaginosa*. This is not surprising since these two species are closely related.

Rhodotorula are environmental yeast found in soil, air, lake and seawater as well as peanuts, fruit juices, crustaceans and mollusks (reviewed in (Wirth and Goldani 2012)). *Rhodotorula* can also be found on plastic shower curtains, toothbrushes, humidifiers and dishwashers (Alvarez-Fernández et al. 1998) (Zalar et al. 2011). *Rhodotorula* are an opportunistic pathogen, particularly as a cause of central venous catheter and peritoneal dialysis-associated fungemia (Tuon and Costa 2008). Disseminated fungemia can occur in immunocompetent and immunocompromised hosts (Tuon and Costa 2008). *Rhodotorula* has also been reported in localized infection of the skin and lung of humans and animals (Alvarez-Fernández et al. 1998), (Kayman et al. 2013) (Monga and Garg 1980). *R. glutinis* cell wall preparations can stimulate macrophage activation *in vitro*, suggesting that this yeast might drive pulmonary inflammation (Sorenson et al. 1998), and hypersensitivity pneumonitis has been reported with inhaled *Rhodotorula* (Alvarez-Fernández et al. 1998). A recent rat model of disseminated *R. mucilaginosa* infection revealed involvement of the lungs, liver and spleen with a granulomatous inflammatory reaction (Wirth and Goldani 2012). *Rhodotorula*-associated peritoneal fibrosis has been reported in patients with dialysis-associated fungemia (Eisenberg et al. 1983), suggesting that inflammation-driven fibrosis in the skin or lung is a potential consequence of infection by this fungal species.

These data also raise the hypothesis that *Rhodotorula* colonization is a consequence of skin disease, rather than a trigger of systemic sclerosis. This cross-sectional study will inform future longitudinal studies of the mycobiome of patients with systemic sclerosis. Latrogenic immunosuppression may also predispose patients to fungal infections regardless of the underlying disease; however two of the SSc patients in this study had not been treated with immunosuppressive agents. Future studies may examine the mycobiome in other fibrotic skin diseases and in latrogenic immunodeficiency.

All biopsies in this study were taken from the forearm. We used control biopsies from healthy patients rather than clinically unaffected skin from SSc patients as previous studies have demonstrated molecular changes in the unaffected skin of SSc (Whitfield et al. 2003; Milano et al. 2008). Recent studies have shown that the mycobiome of the skin varies by site (Findley et al. 2013); future research in this area will require a survey of lesional and clinically unaffected skin from a variety of body sites. Future studies will also be needed to determine which layer of the skin is colonized by Rhodotorula.

One possible explanation of these results is lab contamination in the RNA-seq, however we believe this is unlikely. The four SSc samples were collected on different dates. In at least one case, a normal was collected at the same time, under the same conditions as the SSc sample (the normal was the spouse of the SSc patient). Samples were treated identically from the point of collection onwards. It is difficult to imagine a scenario where the SSc samples could have become contaminated in the lab without the normal samples being similarly affected.

Future research will be required to test this hypothesis and fulfill Hill's epidemiologic criteria for causal association. It is crucial to demonstrate that pathogen exposure precedes development of SSc, which will require prospective studies. The modern genomics view of Koch's postulates stipulates that fewer copies of pathogen nucleic acid exist in normal tissue, consistent with our data. Longitudinal studies will be needed to demonstrate that pathogen load correlates with disease severity and resolution or relapse. Finally, molecular and cellular research is needed to determine how this pathogen triggers inflammation and fibrosis in SSc. Preclinical animal models will allow *in vivo* research on the effect of pathogen on disease.

Materials & Methods

Sample collection

All study participants gave written, informed consent under a Boston University Medical Center Institutional Review Board approved protocol. The study conformed to the Declaration of Helsinki Principals. Single 4 mm punch biopsies were obtained from lesional forearm skin of four patients with early, diffuse SSc within 6 months of first onset of non-Raynaud's symptoms, and normal forearm skin of four controls without disease in a design similar to our original microarray studies (Whitfield et al. 2003). Tissue was stored in RNAlater at −80° C until processed. Samples were processed at Dartmouth Geisel School of Medicine under a protocol approved by the Committee for the Protection of Human Subjects (CPHS) at the Geisel School of Medicine.

RNA-seq sample preparation and sequence alignment

Total RNA was extracted from skin biopsies using QIAgen RNeasy Plus Mini kit. A modified protocol was used to isolate both large (>200 nt) and small (<200 nt) RNA fraction. Only the large RNA fraction was used for this study. Ribosomal RNA was depleted from the large RNA fraction using Invitrogen Ribominus kit. RNA-Seq library was synthesized by NuGen Ovation RNA-Seq System v2 using cDNA prepared by random

hexamer priming. Libraries were multiplexed and sequenced on an Illumina HiSeq 2000 platform and 187–242 million 50 bp paired-end reads were obtained per sample.

Metagenomic analysis was performed with the Integrated Metagenomic Sequence Analysis (IMSA) package (Dimon et al. 2013). The samples were also analyzed by DNA microarray and assigned to the intrinsic gene expression subset (Milano et al. 2008). All four of the SSc patients were assigned to the inflammatory intrinsic subset defined by correlation to centroids. RNA-seq data from these eight patients are available at NCBI GEO at accession number GSEXXXXX (*In process, number will be added in proof*).

Filtering low quality and human reads

Human reads were filtered from the RNA-seq read sets using IMSA (Dimon et al. 2013). Reads were quality filtered to remove any reads with more than 3 bases with a quality score below 15. Next, human reads were removed by progressively more stringent alignments to the human genome (hg19), first with bowtie, then with blat, followed by blast. The final blast alignment removed all reads aligning to the human genome with an E-value of 1xe-8 or better.

Initial Taxonomic Analysis

Once the human reads had been filtered from the dataset, the reads were aligned to NCBI's nt database using blast (E-value <= 1e-15). Scores were generated at every level of NCBI's taxonomy to classify the resulting sequences. Reads aligning to a single classification would add a score of one to that classification. For reads that aligned to multiple species with equal scores, each species was given a partial score (i.e. if a read aligned to two species equally, each would get a score of 0.5 from the read; if the read aligned to three species equally, each would receive 0.33 from the read).

For further analysis, only reads with a single best alignment were retained to avoid spurious hits. In addition, scores were normalized to the millions of reads in the initial read set. Cluster analysis was done using only bacterial, fungal and viral genera with a score above 0.01 in at least one sample. Unsupervised clustering was performed using Cluster 3.0 (Eisen et al. 1998) and visualized using TreeView (Saldanha 2004).

The TaxMap bubble diagrams were created by IMSA, visualized using GraphViz (Gansner and North 2000). TaxMaps show the score at each taxonomic level, in this case for the unique reads only. An average TaxMap was created for Normal and SSc samples by considering only the nodes present in all four samples of the given type, calculating the mean value across the four samples, then only displaying nodes with a score above 0.05 to make the graphs more readable. Individual unfiltered TaxMaps for each sample can be found in the supplemental info for both Plant/Fungal reads and for Bacterial reads.

To align reads to *R. glutinis* 28S rRNA (NCBI FJ345357), we used Bowtie (Langmead et al. 2009) to align the full RNA-seq dataset against the NCBI record. Results were converted to a sorted BAM file and visualized in IGV (Thorvaldsdóttir et al. 2013).

Assembled sequences

For each SSc sample, longer *Rhodotorula* sequences were assembled using PRICE (Ruby et al. 2013). Normal samples did not contain enough *Rhodotorula* reads for assembly. The assembly was run for each sample individually, seeded with all the reads aligning uniquely to *Rhodotorula* (NCBI tax id 5533) (specific flags: -nc 40 -mol 40 -tol 20 -mpi 90 -target 90 1 1 1).

To determine the source of contigs assembled by PRICE, a BLAST alignment to NCBI's nt database was performed using the online BLAST website. Every SSc sample contained a sequence aligning to the 5′ end of the 28S transcript of *Rhodotorula glutinis*. In specific, every sequence covered bases 786–1147 of NCBI record FJ345357 (Khot et al. 2009), spanning the D1–D2 hypervariable region of the 28S rRNA. Multiple sequence alignment was performed in MUSCLE (Edgar 2004). A phylogenetic tree was created using Phylogeny.fr, which uses MUSCLE for the sequence alignment then constructs a maximum likelihood tree using PhyML (Dereeper et al. 2008).

Acknowledgments

This work was supported by grants from the Scleroderma Research Foundation (SRF) to MLW and NIH P50 AR060780 to RL and MLW. ZL received support from NIH R01 AR061384 (MLW) and DoD CDMRP PR100338P1 (MLW). TW also received support from NIH P30 AR061271 to RL and MLW. MEJ also received support from a SYNERGY pilot grant from Geisel School of Medicine at Dartmouth.

Abbreviations

SSc	systemic sclerosis, scleroderma
IMSA	Integrated Metagenomic Sequence Analysis

References

Alvarez-Fernández JA, Quirce S, Calleja JL, et al. Hypersensitivity pneumonitis due to an ultrasonic humidifier. Allergy. 1998; 53:210–2. [PubMed: 9534923]

Biswas SK, Yokoyama K, Nishimura K, et al. Molecular phylogenetics of the genus Rhodotorula and related basidiomycetous yeasts inferred from the mitochondrial cytochrome b gene. Int J Syst Evol Microbiol. 2001; 51:1191–9. [PubMed: 11411687]

Chung L, Fiorentino DF, Benbarak MJ, et al. Molecular framework for response to imatinib mesylate in systemic sclerosis. Arthritis Rheum. 2009; 60:584–91. [PubMed: 19180499]

Dereeper A, Guignon V, Blanc G, et al. Phylogeny.fr: robust phylogenetic analysis for the nonspecialist. Nucleic Acids Res. 2008; 36:W465–469. [PubMed: 18424797]

Dimon MT, Wood HM, Rabbitts PH, et al. IMSA: integrated metagenomic sequence analysis for identification of exogenous reads in a host genomic background. PLoS ONE. 2013; 8:e64546. [PubMed: 23717627]

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004; 32:1792–1797. [PubMed: 15034147]

Eisen MB, Spellman PT, Brown PO, et al. Cluster analysis and display of genome-wide expression patterns. PNAS. 1998; 95:14863–8. [PubMed: 9843981]

Eisenberg ES, Alpert BE, Weiss RA, et al. Rhodotorula rubra peritonitis in patients undergoing continuous ambulatory peritoneal dialysis. Am J Med. 1983; 75:349–52. [PubMed: 6881189]

Findley K, Oh J, Yang J, et al. Topographic diversity of fungal and bacterial communities in human skin. Nature. 2013; 498:367–70. [PubMed: 23698366]

Gansner ER, North SC. An open graph visualization system and its applications to software engineering. SOFTWARE - PRACTICE AND EXPERIENCE. 2000; 30:1203–33.

Greenblatt MB, Sargent JL, Farina G, et al. Interspecies Comparison of Human and Murine Scleroderma Reveals IL-13 and CCL2 as Disease Subset-Specific Targets. Am J Pathol. 2012

Grossman C, Dovrish Z, Shoenfeld Y, et al. Do infections facilitate the emergence of systemic sclerosis? Autoimmun Rev. 2011; 10:244–7. [PubMed: 20863912]

Hamamdzic D, Kasman LM, LeRoy EC. The role of infectious agents in the pathogenesis of systemic sclerosis. Curr Opin Rheumatol. 2002; 14:694–8. [PubMed: 12410093]

Hinchcliff ME, Huang CC, Wood TA, et al. Molecular Signatures in Skin Associated with Clinical Improvement During Mycophenolate Treatment in Systemic Sclerosis. J Invest Dermatol. 2013 In Press.

Huang C-H, Lee F-L, Tien C-J, et al. Rhodotorula taiwanensis sp. nov., a novel yeast species from a plant in Taiwan. Antonie Van Leeuwenhoek. 2011; 99:297–302. [PubMed: 20680683]

Kayman T, Sarigüzel FM, Koç AN, et al. Etiological agents of superficial mycoses in Kayseri, Turkey. J Eur Acad Dermatol Venereol. 2013; 27:842–5. [PubMed: 22672104]

Khot PD, Ko DL, Fredricks DN. Sequencing and analysis of fungal rRNA operons for development of broad-range fungal PCR assays. Appl Environ Microbiol. 2009; 75:1559–65. [PubMed: 19139223]

Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]

Milano A, Pendergrass SA, Sargent JL, et al. Molecular subsets in the gene expression signatures of scleroderma skin. PLoS ONE. 2008; 3:e2696. [PubMed: 18648520]

Monga DP, Garg DN. Ovine pulmonary infection caused by Rhodotorula rubra. Mykosen. 1980; 23:208–11. [PubMed: 7402216]

Ouellette BF, Boguski MS. Database divisions and homology search files: a guide for the perplexed. Genome Res. 1997; 7:952–5. [PubMed: 9331365]

Pendergrass SA, Lemaire R, Francis IP, et al. Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. J Invest Dermatol. 2012; 132:1363–73. [PubMed: 22318389]

Ruby JG, Bellare P, Derisi JL. PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. G3 (Bethesda). 2013; 3:865–80. [PubMed: 23550143]

Saldanha AJ. Java Treeview—extensible visualization of microarray data. Bioinformatics. 2004; 20:3246–8. [PubMed: 15180930]

Sargent JL, Milano A, Bhattacharyya S, et al. A TGFbeta-responsive gene signature is associated with a subset of diffuse scleroderma with increased disease severity. J Invest Dermatol. 2009; 130:694–705. [PubMed: 19812599]

Sorenson WG, Shahan TA, Simpson J. Cell wall preparations from environmental yeasts: effect on alveolar macrophage function in vitro. Ann Agric Environ Med. 1998; 5:65–71. [PubMed: 9852493]

Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis, 1972–2002. Ann Rheum Dis. 2007; 66:940–4. [PubMed: 17329309]

Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14:178–92. [PubMed: 22517427]

Tuon FF, Costa SF. Rhodotorula infection. A systematic review of 128 cases from literature. Rev Iberoam Micol. 2008; 25:135–40. [PubMed: 18785780]

Whitfield ML, Finlay DR, Murray JI, et al. Systemic and cell type-specific gene expression patterns in scleroderma skin. Proc Natl Acad Sci USA. 2003; 100:12319–24. [PubMed: 14530402]

Wirth F, Goldani LZ. Experimental Rhodotorulosis infection in rats. APMIS. 2012; 120:231–5. [PubMed: 22339681]

Zalar P, Novak M, de Hoog GS, et al. Dishwashers--a man-made ecological niche accommodating human opportunistic fungal pathogens. Fungal Biol. 2011; 115:997–1007. [PubMed: 21944212]

Zhao XQ, Aizawa T, Schneider J, et al. Complete mitochondrial genome of the aluminum-tolerant fungus Rhodotorula taiwanensis RS1 and comparative analysis of Basidiomycota mitochondrial genomes. Microbiologyopen. 2013; 2:308–17. [PubMed: 23427135]

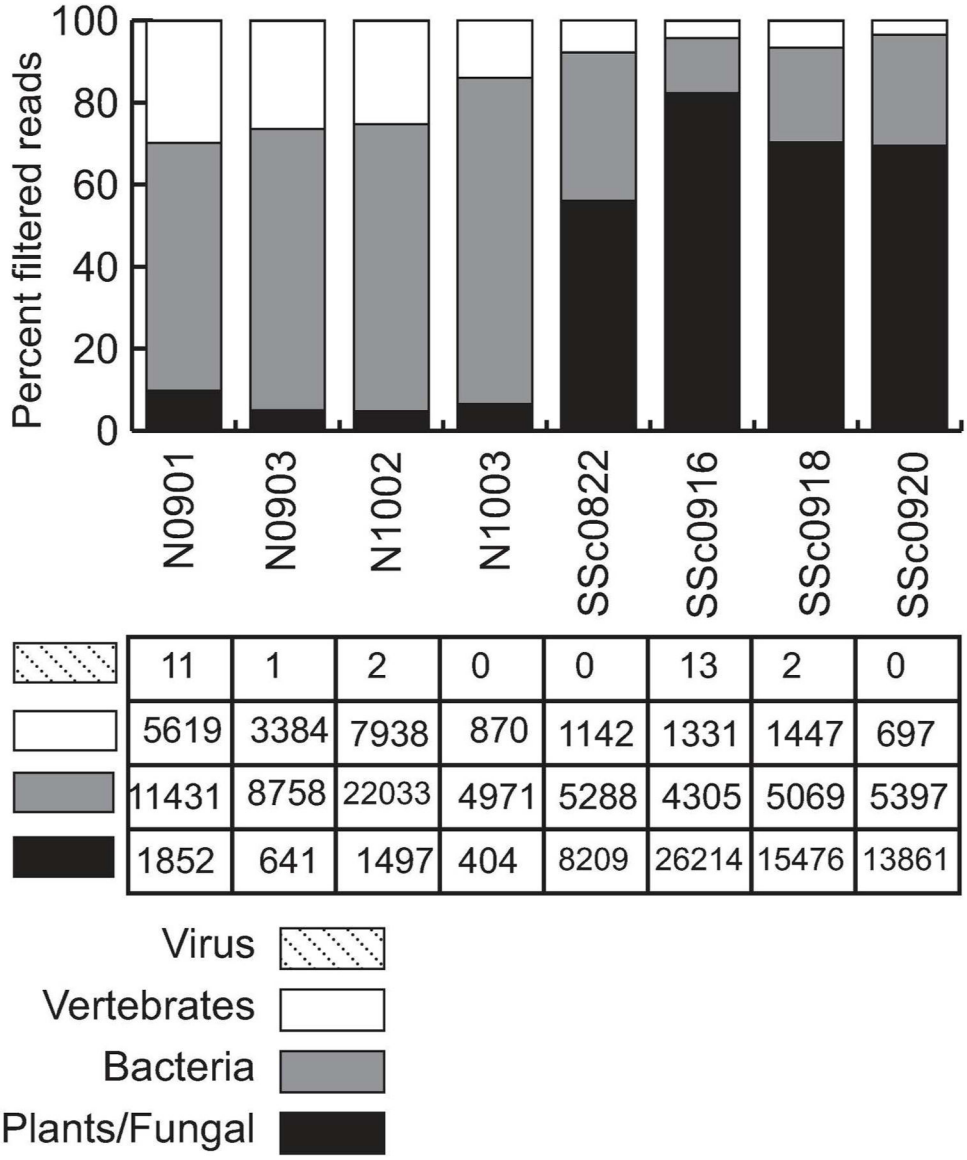


Figure 1. IMSA analysis reveals plant/fungal sequences in SSc samples

Division-level breakdown of the reads remaining after filtering human reads. NCBI divisions group plants and fungal sequences together. Normal samples begin with “N” while SSc samples begin with “SSc”. Numbers shown in the table below are the IMSA score for each division per million input reads, indicating the relative abundance as a proportion of total reads. The most striking difference between normal and SSc samples is the abundance of plant/fungal reads in SSc samples, with an associated reduction in other divisions. Vertebrate reads (primarily human reads not filtered by IMSA due to mismatches to the human genome) are about a quarter of the non-plant/fungal reads. The other three-quarters of the non-plant/fungal reads are bacteria, the amount of which varies by sample but does not have a clear difference in abundance between normal and SSc samples.

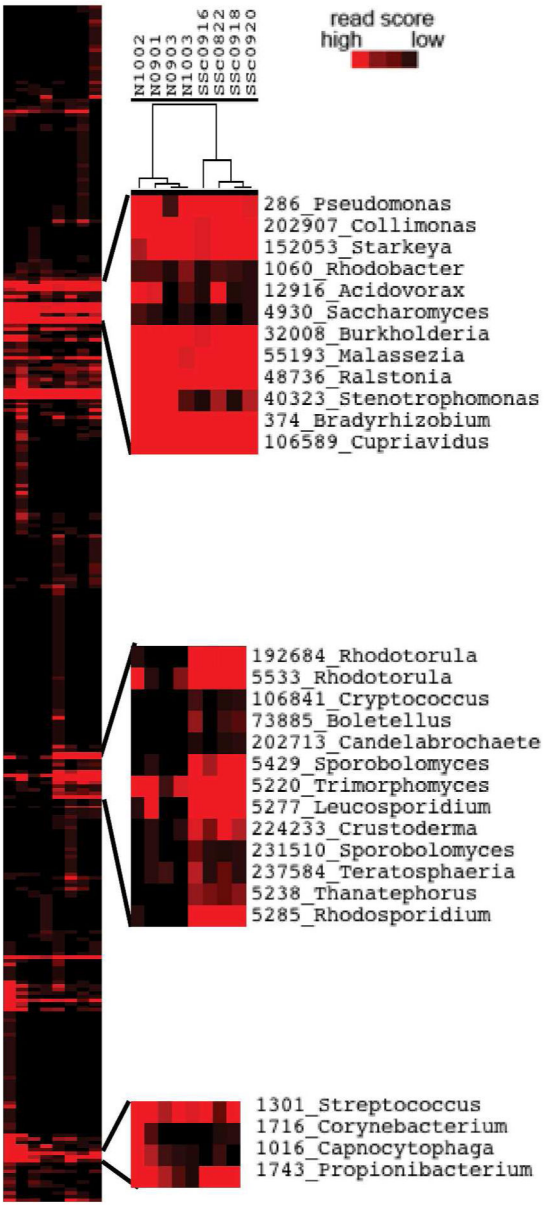


Figure 2. Unsupervised clustering of genera scores

All bacterial, fungal and virual genera with a score above 0.01 in at least 1 sample were clustered by IMSA score, normalized per million reads in the original read set. The tree shows normal samples cluster together on the left while SSc samples cluster together on the right. The top and bottom call-outs show normal skin flora expressed in both normal and SSc samples. The center call-out shows fungal genera expressed predominantly in SSc samples.

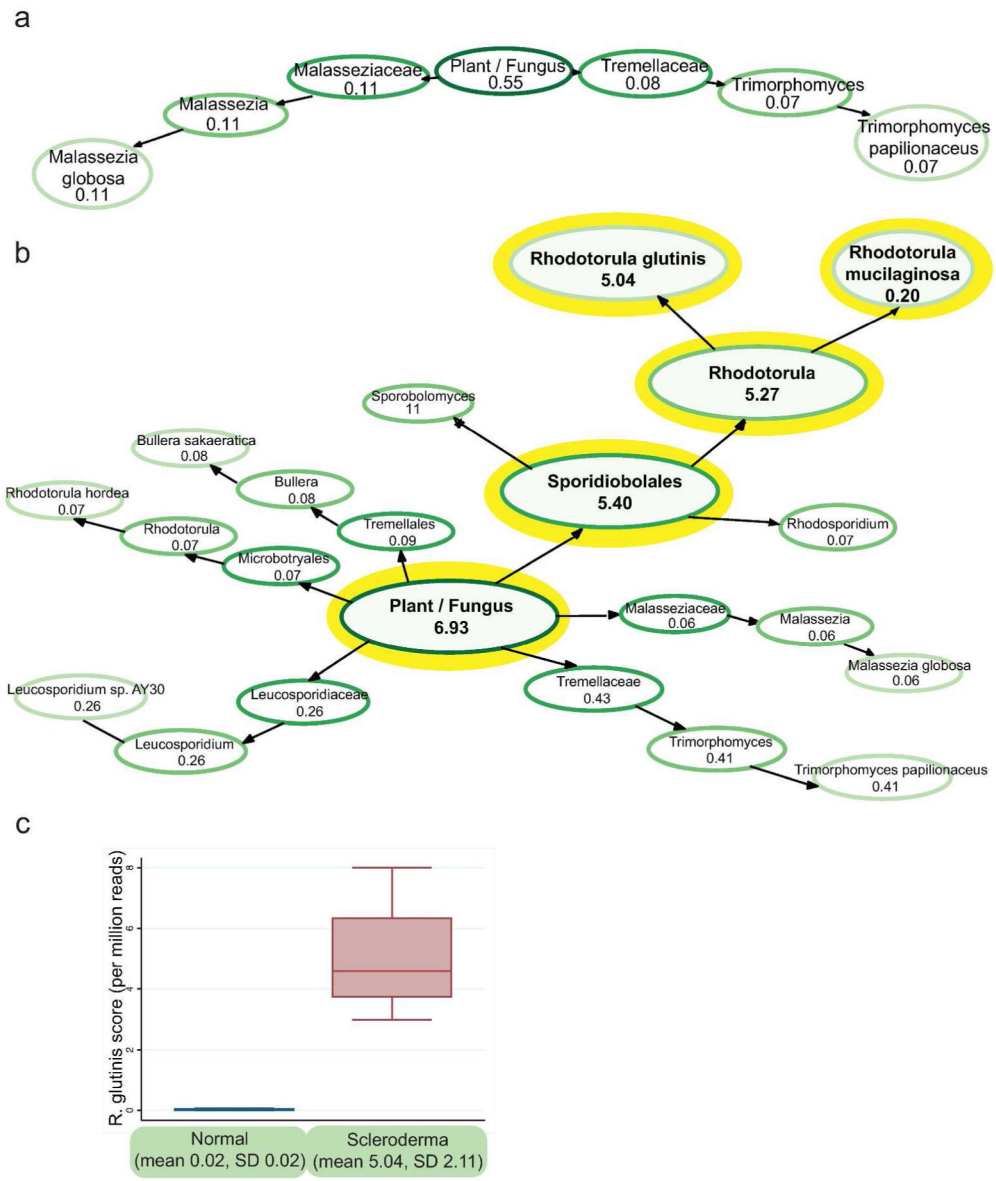


Figure 3. *Rhodotorula glutinis* in Normal and SSc skin samples

a. TaxMap visualization of plant/fungal reads shows few fungal species in normal skin. TaxMaps show the IMSA score for each level of the taxonomic hierarchy, allowing quick visualization of the metagenome of a sample. The score shown is the average score for the normal samples counting only reads with a single best alignment, normalized per million input reads. Only nodes with a value above 0.05 are shown to make the figure easier to read. b. TaxMap visualization demonstrates *R. glutinis* as the dominant species in SSc skin. c. Reads with a single best alignment to *R glutinis* are present at 252-fold higher frequency in SSc skin than normal skin.

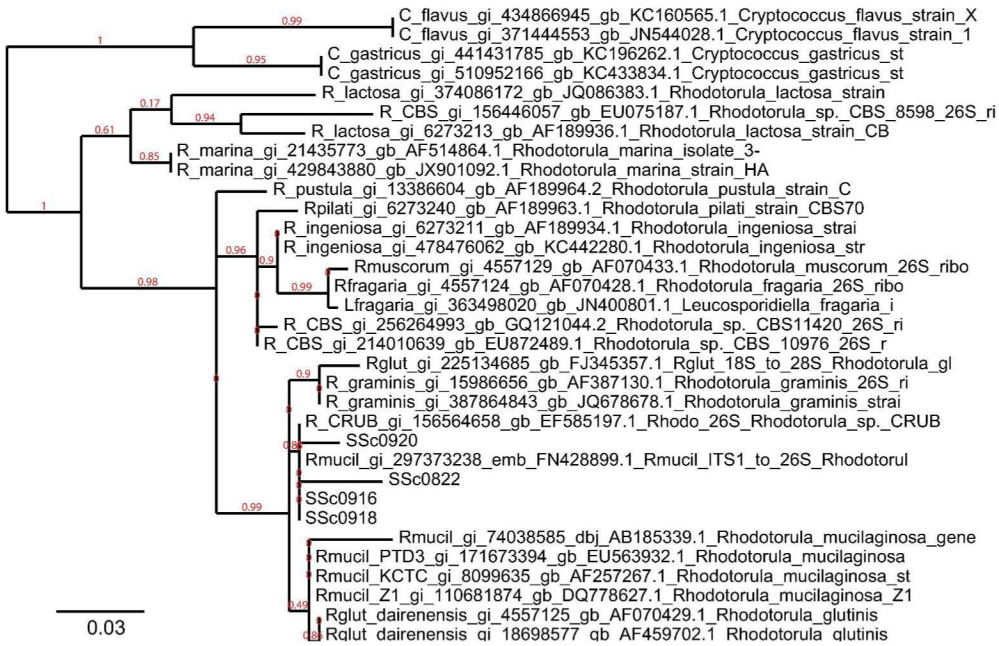


Figure 4. Phylogenetic tree of assembled 28S rRNA sequences

Phylogenetic tree of 28S rRNA sequences from selected NCBI *Rhodotorula* sequences. The maximum likelihood tree was constructed with PhyML and rendered with TreeDyn. Species names and accession numbers are given for sequences downloaded from GenBank. The four SSc samples are grouped with *R. mucilaginosa*, close to sequences from *R. glutinis* and *R. graminis*.

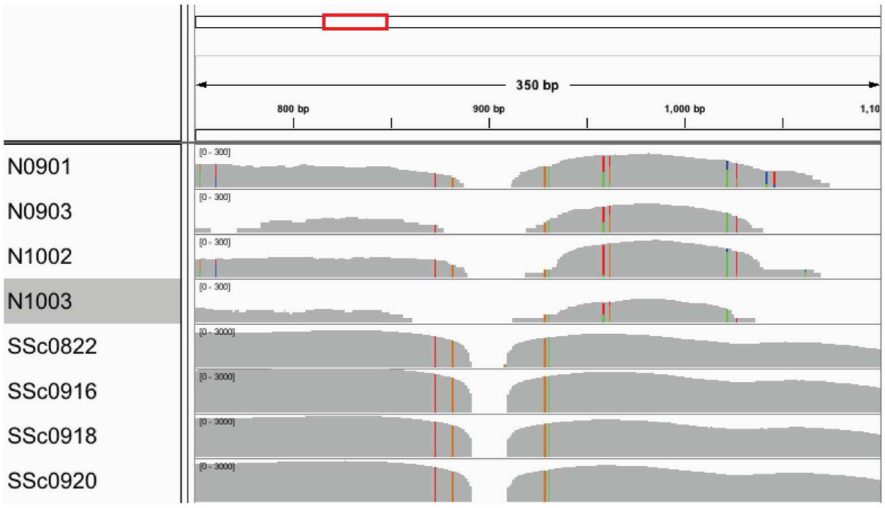


Figure 5. IGV visualization of original read set to *R. glutinis* 28S rRNA

Aligning raw reads to *R. glutinis* 28S rRNA sequence (FJ345357) shows many reads aligning in the SSc samples but much fewer reads in the normal samples. The position shown is from 750–1100 in the sequence, which is the end of ITS2 and the first 400 bases of 28S rRNA. The gray histogram shows the depth of coverage at each base along the sequence; note that the axis is 10-fold higher for SSc samples. Colored bars show areas where the aligned reads differ from the reference sequence.

Table 1

Clinical features of SSc patients

Healthy controls included two white males and two white females, ages 24, 27, and unknown.

	sex	race	age	diagnosis	disease duration (months)	MRSS	immunosuppression
SSc0882	M	W	60	dcSSc	7	38	none
SSc0916	F	W	56	dcSSc	36	44	none
SSc0918	M	W	50	dcSSc	3	15	prednisone
SSc0920	F	W	65	dcSSc	6	34	mycophenolate

MRSS: modified Rodnan skin score, M: Male, F: Female, W: White, dcSSc: diffuse cutaneous systemic sclerosis.

Table 2

Sequence read counts and IMSA scores

Rhodotorula scores are normalized per million reads in the initial readset.

Sample	Initial Fastq	High Quality Reads	Reads After Human Filter	*Rhodotorula* Score (normalized)	*R. glutinis* score (normalized)
N0901	188,871,690	157,580,342	498,774	0.005	0.005
N0903	241,690,526	186,982,854	391,620	0.000	0.000
N1002	199,655,946	170,008,890	935,514	0.075	0.065
N1003	192,679,262	110,673,772	222,532	0.016	0.016
SSc0882	203,708,052	146,919,248	304,502	3.102	2.985
SSc0916	228,994,824	155,180,738	301,846	8.253	8.000
SSc0918	243,159,154	171,379,530	337,658	4.684	4.495
SSc0920	188,011,530	126,597,588	256,664	4.888	4.675

Johnson et al. Arthritis Research & Therapy (2016) 18:27
DOI 10.1186/s13075-016-0914-4

Arthritis Research & Therapy

RESEARCH ARTICLE Open Access

Stress granules and RNA processing bodies are novel autoantibody targets in systemic sclerosis

Michael E. Johnson[1], Andrew V. Grassetti[1], Jaclyn N. Taroni[1], Shawn M. Lyons[2], Devin Schweppe[1], Jessica K. Gordon[3], Robert F. Spiera[3], Robert Lafyatis[4], Paul J. Anderson[2], Scott A. Gerber[1] and Michael L. Whitfield[1,5*]

Abstract

Background: Autoantibody profiles represent important patient stratification markers in systemic sclerosis (SSc). Here, we performed serum-immunoprecipitations with patient antibodies followed by mass spectrometry (LC-MS/MS) to obtain an unbiased view of all possible autoantibody targets and their associated molecular complexes recognized by SSc.

Methods: HeLa whole cell lysates were immunoprecipitated (IP) using sera of patients with SSc clinically positive for autoantibodies against RNA polymerase III (RNAP3), topoisomerase 1 (TOP1), and centromere proteins (CENP). IP eluates were then analyzed by LC-MS/MS to identify novel proteins and complexes targeted in SSc. Target proteins were examined using a functional interaction network to identify major macromolecular complexes, with direct targets validated by IP-Western blots and immunofluorescence.

Results: A wide range of peptides were detected across patients in each clinical autoantibody group. Each group contained peptides representing a broad spectrum of proteins in large macromolecular complexes, with significant overlap between groups. Network analyses revealed significant enrichment for proteins in RNA processing bodies (PB) and cytosolic stress granules (SG) across all SSc subtypes, which were confirmed by both Western blot and immunofluorescence.

Conclusions: While strong reactivity was observed against major SSc autoantigens, such as RNAP3 and TOP1, there was overlap between groups with widespread reactivity seen against multiple proteins. Identification of PB and SG as major targets of the humoral immune response represents a novel SSc autoantigen and suggests a model in which a combination of chronic and acute cellular stresses result in aberrant cell death, leading to autoantibody generation directed against macromolecular nucleic acid-protein complexes.

Keywords: Systemic sclerosis, Scleroderma, Autoantibody, RNA processing bodies, Stress granules

Background

Systemic sclerosis (SSc) is a rare systemic autoimmune disease of unknown etiology characterized by skin fibrosis, internal organ involvement, vascular abnormalities, and autoantibody production. Patients are broadly classified as having either limited (lSSc) or diffuse (dSSc) disease based primarily upon the extent of skin involvement and autoantibody profiles. While a wide array of autoantibodies have been described for SSc, only a small number of these targets are used for clinical diagnosis and stratification. Autoantibodies targeting RNA polymerase III (RNAP3), topoisomerase 1 (TOP1; commonly referred to as Scl70), and centromere proteins (CENP) represent the three the most common, clinically measured autoantibodies observed in SSc [1, 2]. Other autoantibodies, including fibrillarin (U3RNP), Pm/Scl, Ku, U1RNP, U11/U12, and Th/To have also been described [1, 3] but are not routinely measured for clinical subtyping.

* Correspondence: michael.l.whitfield@dartmouth.edu
[1]Department of Genetics, Geisel School of Medicine at Dartmouth, Hanover, NH, USA
[5]Dartmouth Medical School, Hinman Box 7400, Hanover, NH 03755, USA
Full list of author information is available at the end of the article

While the processes underlying autoantibody production in SSc remain poorly understood, the presence of certain autoantibodies is strongly predictive of clinical outcomes [1–3]. TOP1 and RNAP3 autoantibodies are almost exclusively seen in dSSc, while CENP, Th/To, and U1RNP antibodies are more commonly associated with lSSc [1, 3]. U3RNP autoantibodies are not associated with either clinical subset, and are often found in conjunction with other autoantibodies, including both TOP1 and CENP [3]. Certain antibodies, such as TOP1 and U11/12, have been shown to be predictive of poorer overall prognosis, including increased likelihood of pulmonary fibrosis [4] and cardiac involvement, while RNAP3 autoantibodies have recently been linked to co-occurrence of SSc with cancer [5].

Despite the importance of autoantibodies in SSc, the vast majority of target identification and phenotypic screening has been performed using methods targeting only a single autoantibody, with little ability to detect novel or low abundance autoantibodies. Furthermore, these methods fail to address the possibility of co-occurrence of multiple autoantibodies within a patient, which may have important clinical implications. Autoantigen microarrays have proven successful for screening large numbers of autoantibodies in parallel, however target identification is limited to those antigens produced and printed on the antigen microarrays [6]. To address these limitations, we performed immunoprecipitations (IP) of HeLa whole cell lysates using sera from RNAP3-, CENP-, and TOP1-positive patients, as well as healthy controls, followed by mass spectrometry (LC-MS/MS) to provide an unbiased assessment of all autoantibodies present in these SSc patients. This method provides a better view of the full range of autoantibodies present in SSc, including both novel and established targets, and provides insights into the general processes underlying autoantibody production.

Methods

Clinical samples

Patient serum was obtained from Boston University Medical School, Boston (BUMC), MA, USA and the Hospital for Special Surgery (HSS), New York, NY, USA. All relevant study protocols were approved by the Dartmouth College committee for the protection of human subjects, and the internal review boards of both BUMC and HSS. Informed consent was obtained from all patients prior to sample collection. Patients were diagnosed with either dSSc or limited SSc, as determined using the 1980 American College of Rheumatology classification criteria. Detection of major autoantibody reactivities was performed using standard clinical assays.

Human cell lysates

HeLa cells were cultured in DMEM supplemented with 10 % fetal bovine serum (FBS) (v/v) and 100 IU/mL penicillin-streptomycin. Cells were grown to approximately 80 % confluence, harvested in IP lysis buffer (150 mM NaCl, 50 mM Tris pH 7.5, 1 mM $MgCl_2$, 1 mM EDTA, 0.5 % Triton X-100, 2.5 mM β-mercaptoethanol, 1 mM sodium molybdate, 1 mM sodium fluoride, 1 mM sodium tartrate, 1 mM dithiothreitol (DTT), and protease inhibitors (Roche, Indianapolis, IN, USA)), lysed by passage through a pre-chilled high-gauge syringe, and centrifuged for 15 minutes to pellet debris. Lysates were then clarified by incubating for 4 h at 4 °C on a rotating platform. Protein concentrations were quantified using a standard bicinchoninic acid (BCA) protein assay kit (Thermo Scientific, Waltham, MA, USA).

Serum immunoprecipitation

Patient serum was cross-linked to Protein G Dynabeads (Invitrogen, St. Louis, MO, USA) prior to IP. First, 100 μL serum (approximately 1 mg IgG) was added to 50 μL Protein G beads and incubated for 5 h at 4 °C. Samples were then washed in PBS, equilibrated in crosslinking buffer (50 mM HEPES, pH 8.2), and cross-linked to Protein G beads by the addition of 20 mM dimethyl pimelimidate and 300 mM HEPES (DMP) solution for 10 minutes at room temperature (repeated three times). The crosslinking reaction was then terminated by the addition of 50 mM ammonium bicarbonate, and the resulting antibody bead mixture added to 500 μL cell lysate (diluted to 4 mg/mL in IP lysis buffer). Samples were incubated overnight at 4 °C on a rotating platform, washed in cold IP lysis buffer, and eluted in a buffer containing 2 % SDS, 75 mM NaCl, 50 mM Tris pH 8.1, and 20 % glycerol at 65 °C for 5 minutes. Eluates were reduced by the addition of 0.1 M dithiothreitol (DTT) (to a final concentration 5 mM), and incubated at 80 °C for 5 minutes. Samples were then resolved by SDS-PAGE, split into high (>60 kDa) and low (<60 kDa) molecular weight fractions and analyzed by mass spectrometry.

Mass spectrometry

Proteins contained in Coomassie-stained gel regions were digested overnight with trypsin (1:200 w/v) at 37 °C. Following digestion, peptides were extracted from the gels, dried, and analyzed by nanoscale LC-MS/MS. LC-MS/MS analyses were performed on either LTQ Orbitrap Classic or Orbitrap Fusion LC-MS/MS platforms. LTQ Orbitrap Classic analyses were conducted as described previously [7].

For Orbitrap Fusion analyses, samples were loaded onto an EASY-nLC 1000 Liquid Chromatograph (Thermo Scientific, Waltham, MA, USA) and separated by reverse-phase high pressure liquid chromatography (RP-HPLC)

using an approximately 36-cm column with a 100-μM inner diameter packed with 3 μm 120 Å C_{18} particles (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). The resultant peptide eluate was directed into an Orbitrap Fusion Tribrid Mass Spectrometer operating in a data-dependent sequencing acquisition mode across a 30-minute reverse-phase gradient (6 % acetonitrile, 0.1 % formic acid to 30 % acetonitrile, 0.1 % formic acid) at 350 nL/min flow rate. The Orbitrap Fusion was operated with an Orbitrap MS1 scan at 120 K resolution, followed by Orbitrap MS2 scans of higher energy collision-induced dissociation (HCD) fragment ions (30 % HCD energy) at 15 K resolution using a maximum cycle type of 2 s, precursor ion dynamic exclusion window of 15 s, +2, +3, and +4 precursor ions selected for LC-MS/MS, and maximum ion injection times of 100 ms (MS1) and 50 ms (MS2). The resulting tandem mass spectra were data-searched using the COMET search engine [8] against a *Homo sapiens* proteome database (source: Uniprot; download date: 2 July 2013) with a precursor ion tolerance of +/- 1 Da [9] and a fragment ion tolerance of 0.02 Th. Peptide spectra matches (PSMs) were filtered to <1 % false discovery rate using the target decoy strategy [10], and reported.

IP-western blots

Anti-UPF1 antibody was kindly provided by Dr. Lynne Maquat (University of Rochester Medical Center, Rochester, NY, USA). Antibodies to MOV10 and CAPRIN1 were purchased from Proteintech (Chicago, IL, USA); antibodies to G3BP1 and USP10 were purchased from Santa Cruz Biotechnology (Santa Cruz, CA, USA). Serum immunoprecipitation of HeLa lysates was performed as described above; 50 % of each eluate (15 μL) was then run on a 10 % bis-tris precast gel (Life Technologies, Carlsbad, CA, USA). HeLa whole cell lysate (100 μg) was used as a positive control; no loading control was performed due to the absence of viable targets present in all IP eluates. Western blots were then run following standard protocols, and visualized using Western Lightning ECL Pro or Ultra substrate (Perkin Elmer Inc., Waltham, MA, USA), as necessary.

Data analysis

Non-redundant peptide hits, defined as mass spectra mapping exclusively to a given peptide fragment, were used for all downstream analyses. Pairwise comparisons between samples were performed by Fisher's exact test using the Bonferroni correction for multiple hypothesis testing. Venn diagrams were generated using VENNY [11]. Network analysis was performed using the Genome-scale Integrated Analysis of gene Networks in Tissues (GIANT; http://giant.princeton.edu/) global network [12] and visualized using Cytoscape [13]. Communities in the network were detected using fast-greedy modularity as implemented in igraph. Functional annotation of individual communities was performed using g:Profiler [14]. Semiquantitative enrichment of SSc-associated autoantibodies was determined using a binary assessment of autoantibody presence or absence in a sample. Preferential enrichment in SSc was defined as all proteins detected in >50 % of all patient samples at a frequency >1.5-fold relative to controls. Enrichment of biological processes and cellular components was determined using g:Profiler using the g:SCS threshold correction for multiple hypothesis testing and a functional category size ≤500 genes. Hierarchical clustering was performed using Cluster 3.0 [15], and visualized using Java TreeView [16].

Immunofluorescence

The day prior to the experiment, 10^5 U2OS cells were seeded onto 11 mm glass coverslips and allowed to attach overnight at 37 °C/5 % CO_2 in DMEM containing 10 % FBS (Gibco). Cells were treated with 100 μM sodium (meta)arsenite (Sigma Aldrich) for 1 h to induce the formation of stress granules and then with 4 % paraformaldehyde solution at room temperature for 15 - minutes followed by blocking and permeabilization with 5 % normal horse serum, 0.1 % digitonin in Tris-buffered saline. Staining was performed with anti-eIF3b (Santa Cruz), anti-SK1-Hedls (Santa Cruz), and patient sera for 1 h at room temperature. Secondary antibodies (anti-goat-Cy3, anti-mouse-Cy2, and anti-human-Cy5) were purchased from Jackson Laboratories and incubated at room temperature for 1 h. Conventional fluorescence microscopy was performed using a microscope (model Elipse E800, Nikon, Tokyo, Japan) with epifluorescence optics with a digital camera (model CCD-SPOT RT; Diagnostic Instruments, Sterling Heights, MI). Images were compiled using Adobe Photoshop software (CS6; Adobe Systems, San Jose, CA).

Results

Identification of proteins cross-reacting to serum antibodies

Immunoprecipitations (IP) of HeLa whole cell lysates were performed using sera obtained from 13 SSc patients and 4 healthy controls. HeLa cells were chosen based upon their consistent, high level of expression of a broad range of proteins from the human genome [17].

SSc patients were divided into three groups, TOP1, RNAP3, and CENP, as measured in a reference laboratory; clinical data for each patient are shown in Table 1. These groups were chosen based upon their relative frequency, and their importance in clinical diagnosis. Immunoprecipitated proteins were analyzed by LC-MS/MS, and the resulting spectra aligned to the reference human proteome (UCSC version hg19). Data are presented in two ways; first to identify the total number of

Table 1 Clinical information for patients involved in this study

Sample	Group	Age (years)	Sex	Race	Disease type	ILD/PAH	Disease duration (years)	ANA pattern	ANA titer	MRSS
SSc 1	TOP1	36	F	White	Diffuse	mild ILD	2.5		1:320	43
SSc 132	TOP1	49	F	White	Diffuse	No		Homogeneous	1:640	27
SSc 218	TOP1	55	F	White	Diffuse	ILD		Homogeneous/nucleolar	1:2560	18
SSc 208	TOP1	64	M	White	Diffuse	No		Nucleolar	1:1280	37
SSc 5	RNAP3	53	M	White	Diffuse	No	0.75	Speckled	1:80	36
SSc 7	RNAP3	45	F	Black	Diffuse	No	0.5	Speckled	1:80	27
SSc 10	RNAP3	52	M	White	Diffuse	No	0.5		0	22
SSc 18	RNAP3	69	F	White	Diffuse	ILD	0.5	Nucleolar	1:160	44
SSc 159	CENP	54	F	Mixed	Limited	No	7	Centromere	1:1280	2
SSc 177	CENP	64	F	White	Limited	No	15	Discrete speckled	4+	
SSc 194	CENP	66	F	White	Limited	No	18	Discrete speckled	4+	6
SSc 238	CENP	53	F	White	Limited	No	6	Centromere	1:640	5
SSc 226	CENP	55	F	Asian	Diffuse	No		Centromere	1:1280	6
HC 162	Control	24	M	White						
HC 400	Control	21	M	White						
HC 117	Control		M							
HC 118	Control		M							

Blank cells indicate information not available at the time of sample collection. *ILD* interstitial lung disease, *PAH* pulmonary arterial hypertension, *ANA* anti-nuclear antibody, *MRSS* modified Rodnan skin score, *SSc* systemic sclerosis, *M* male, *F* female

peptides that could be aligned to each protein (total hits), and second to identify all non-redundant peptides that mapped exclusively to a given protein (non-redundant hits). A complete list of all data can be found in Additional file 1: Table S1.

Exclusivity and co-occurrence of SSc autoantibodies

We observed a high degree of reproducibility between patients within their respective autoantibody groups (TOP1, RNAP3, and CENP; Fig. 1). The greatest degree of overlap between peptides was observed among RNAP3 patients (Fig. 1a and c), with 420 proteins (54.1 %) detected in all four patients (Fig. 1c). The remaining groups exhibited significant overlap in three of four (TOP1) and four of five (CENP) patients, respectively (Fig. 1a), along with a single outlier that showed either higher (SSc 208; TOP1) or lower (SSc 226; CENP) total peptide hits relative to other samples in these groups. Within TOP1, 111 proteins (14.2 %) were detected in all four patients (Fig. 1d), while CENP exhibited 48 proteins (10.5 %) common to all patients (Fig. 1c). The least overlap was seen in healthy controls, with only 40 proteins (7.6 %; Fig. 1b) common across individuals.

Across all samples, 283 proteins (25.0 %) were detected in at least one patient in each of the four autoantibody groups (Fig. 1e, Additional file 2: Table S2). Some of these proteins likely represent background signals (serum albumin (ALB), β-tubulin (TUBB), and ribosomal proteins), while others are considered specific to SSc despite trace level detection in controls. For example, multiple SSc autoantibody targets, including Ku (XRCC5 and XRCC6), Ro52/TRIM21, and nucleophosmin/B23 (NPM1) were present in this set of proteins. In contrast, 87 proteins (7.7 %) were detected in all three SSc groups, but were absent in controls (Fig. 1e; Additional file 2: Table S2). Functional analyses of these proteins revealed strong enrichment of proteins involved in oxidative stress responses and nucleic acid processing (Additional file 3: Table S3B).

Of the 1,130 non-redundant proteins identified, 473 (41.8 %) were unique to a given autoantibody group (Fig. 1f); however, the vast majority of these proteins were exclusive to a single patient, with only 111 (23.5 %) detected in two or more patients. These results suggest a wide range of autoantibody responses within each of the clinical autoantibody groups beyond what has already been described.

Among the major autoantibody groups, immunoprecipitation of RNAP3 was exclusive to the RNAP3 group, with no RNAP3 peptides detected in any of the other samples (Table 2). In contrast, TOP1 peptides were consistently highest among TOP1 patients, but were also detected at low levels in all four RNAP3$^+$ patients, and in two controls (Table 2). As these patients were negative for TOP1 autoantibodies by clinical testing, these results indicate a higher degree of sensitivity for our IP/MS protocol compared to standard ELISA-based methods used clinically. In contrast, CENP was only

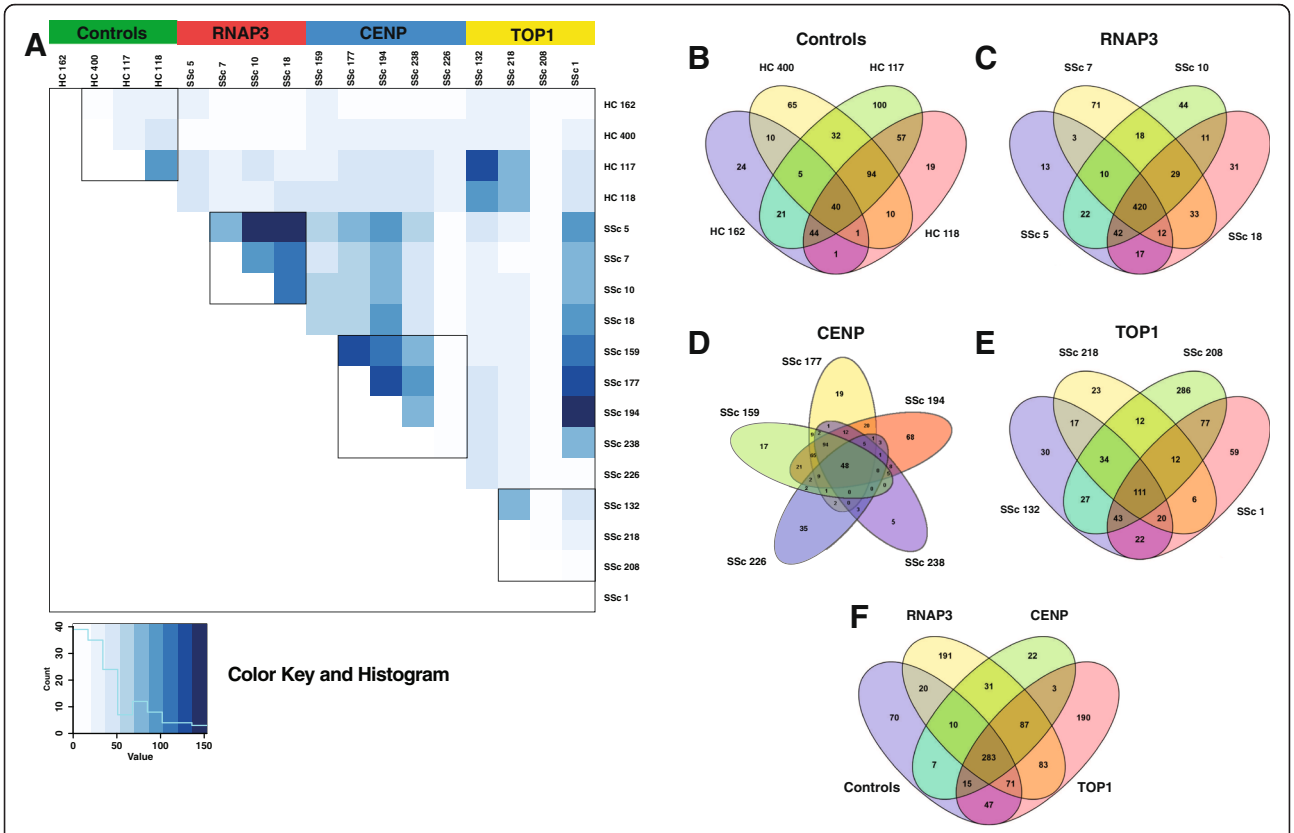


Fig. 1 Overview of mass spectrometry results. **a** Correlation matrix of non-redundant protein hits for all patients and controls. Comparisons were performed using Fisher's exact test with the Bonferroni correction. *Black boxes* indicate intra-group comparisons for each of the four clinically defined groups. *Green* controls; *red* RNA polymerase III (*RNAP3*); *blue* centromere protein (*CENP*); *yellow* topoisomerase I (*TOP1*). **b-f** *Venn diagrams* depict overlap in non-redundant peptide hits within and between groups. **b** healthy controls, **c** RNAP3, **d** CENP, **e** TOP1, and **f** overlap between groups

detected at low levels in the CENP group, likely because it remained bound to the tightly packed centromere complex of chromatin.

Other known SSc autoantigens were also detected. RuvBL [18] was strongly detected in all SSc samples, while virtually absent in controls. Ku and Su, along with a wide array of anti-tRNA synthetases [19] were routinely detected in both the RNAP3 and TOP1 subsets, but were only weakly present in the CENP and control groups (Table 2).

Several autoantigens previously implicated in SSc were found at low, background levels in both SSc and control samples. Ro52/TRIM21 [20] and nucleophosmin/B23 [21] were widely detected across all four groups, suggesting widespread reactivity to these proteins in SSc and controls. We did not find evidence in SSc of enrichment of Pm/Scl autoantibodies, which target exosome components EXOSC1-10 [22]. Peptides for these proteins were absent in the CENP group, but were detected at low levels in other subsets, including controls. Autoantigens not detected here include many of the URNPs, PDGFR, matrix metalloproteinases, tissue plasminogen activator, and vascular receptor antibodies (Table 2).

Functional clustering of identified proteins

To identify functional interactions among autoantigens, all 763 non-redundant protein hits were submitted as a query to the GIANT global average network. This approach included both SSc-specific targets and those detected at background levels in controls, to better understand the full range of autoreactive proteins and complexes. Nine distinct communities were identified within the resulting network, in which each gene is represented by a node, and two genes share an edge if they are predicted to functionally interact (Additional file 4: Figure S1). Analysis of each of these communities by g:Profiler revealed functional enrichment for a wide range of biological processes associated with important disease processes and components (Additional file 4: Figure S1). Community 1 is dominated by ribosomal proteins, eukaryotic initiation factor 3 (eIF3) subunits, and includes the SSc autoantibody target nucleophosmin/B23. Communities 2 and 8 show strong enrichment for Gene Ontology (GO) terms mRNA processing, ribonucleoprotein complex, and cytosolic stress granule. Community 2 is dominated primarily by DEAD box

Table 2 SSc-associated autoantibodies observed in this study

Alias	Associated proteins	Disease subset	Clinical associations	Prevalence in this dataset (avg/freq)				Reference
				Control (n = 4)	RNAP3 (n = 4)	TOP1 (n = 4)	CENP (n = 5)	
Major autoantibodies								
RNA Pol III	POLR3A	dSSc	Renal crisis, cancer	-	+++ (28/4)	-	-	Graf, et al. 2012 [1]; Mehra, et al. 2013 [3]
Scl70	TOP1	dSSc	Poor prognosis, internal organ involvement, and proteinuria	+ (3/2)	+ (4/4)	+++ (19/4)	-	Mehra, et al. 2013 [3]
Centromere	**CENPB**, CENPH	lSSc/CREST	PAH, ILD	-	-	-	+ (1/2)	Mehra, et al. 2013 [3]
Other SSc autoantibodies present in our dataset								
Endothelial Cell	TUBB, **VCL**, LMNA, RPLP0	SSc	PAH	+ (1/1)	++ (6/4)	+ (4/2)	+ (0/1)	Dib, et al. 2012 [34], Naniwa, et al. 2007 [35]
Fibroblast	ENO1, G6PD, **HSPA1A**, HSPA1B, VIM	SSc	PAH	+ (3/4)	+++ (12/4)	++ (5/3)	++ (8/5)	Terrier, et al. 2008 [36], Terrier, et al. 2009 [37]
Histone	H1FX, HIST1H1B, **HIST1H4A**	SSc	PF, internal organ involvement, decreased survival	+ (1/1)	+ (3/3)	+ (1/1)	-	Mehra, et al. 2013 [3]
B23	NPM1	dSSc, CENP- lSSc	PAH	+ (4/4)	++ (7/4)	++ (5/4)	++ (6/5)	Mehra, et al. 2013 [3]
Ku	**XRCC5**, XRCC6	lSSc	Myositis	+ (3/3)	+++ (12/4)	++ (8/4)	+ (2/3)	Graf, et al. 2012 [1]; Mehra, et al. 2013 [3]
Su	AGO2	SSc, PM/Scl	Unknown	-	+ (1/2)	+ (3/1)	-	Satoh, et al. 2013 [38]
Mitochondrial (M2)	**DLD**, PDHB	lSSc	Strong association with primary biliary cirrhosis	+ (1/1)	+ (2/3)	+ (1/1)	-	Mehra, et al. 2013 [13]
Pm/Scl	EXOSC1-**10**	SSc	PF, digital ulcers; decreased risk of PAH and GI symptoms	+ (2/2)	++ (5/3)	+ (2/2)	-	Mehra, et al. 2013 [13]
hnRNPs	**HNRNPA1**-3, HNRNPL	SSc	Common in SARDs	+ (0/1)	++ (7/4)	+ (3/4)	+ (2/4)	Siapka, et al. 2007 [39]
U1	**SNRNPA**, SPRNP70	SSc	Co-occurrence with SS-A/SS-B, PAH, overlap syndrome	-	+ (2/4)	+ (1/2)	+ (0/1)	Graf, et al. 2012 [1]; Mehra, et al. 2013 [3]
U5	SNRNP200	SSc, PM/Scl	Unknown	++ (6/3)	++ (9/4)	++ (8/3)	+ (1/2)	Kubo, et al. 2002 [40]
RO52/TRIM21	TRIM21	SSc	ILD, other autoimmune diseases	++ (6/3)	+++ (12/4)	++ (6/4)	++ (8/4)	Mehra, et al. 2013 [3]
RuvB	**RUVBL1**, RUVBL2	dSSc	Common in SARDs, older age at onset, male sex	+ (1/1)	++ (7/4)	+ (3/4)	+ (2/4)	Kaji, et al. 2014 [18]
Annexin V	ANXA5	dSSc, CENP- lSSc	Digital ischemia	+ (2/2)	++ (7/4)	+ (4/3)	+ (3/4)	Mehra, et al. 2013 [3]
SS-B/LA	SS-A, **SS-B**	SSc	ILD, other autoimmune diseases	-	+ (3/4)	+ (2/2)	+ (0/1)	Mehra, et al. 2013 [3]
Peroxiredoxin	PRDX1	SSc	Disease duration, PF, cardiac involvement, TOP1+ patients	+ (2/4)	++ (8/4)	+ (3/3)	+ (4/4)	Mehra, et al. 2013 [3]
hUBF/NOP90	UBTF	lSSc	Mild organ involvement, favorable prognosis	-	+ (1/2)	-	-	Mehra, et al. 2013 [3]
Th/To	POP1	lSSc	PF, renal crisis, poor prognosis, myositis, PAH	+ (1/2)	+ (1/1)	+ (3/3)	-	Graf, et al. 2012 [1]; Mehra, et al. 2013 [3]

Table 2 SSc-associated autoantibodies observed in this study *(Continued)*

PL-12	AARS	SSc, PM/DM	ILD without myositis	-	-	+ (1/1)	+ (1/1)	Hamaguchi, et al. 2013 [19]
OJ	IARS	SSc, PM/DM	ILD without myositis	+ (1/1)	-	+ (3/3)	-	Hamaguchi, et al. 2013 [19]
EJ	GARS	SSc, PM/DM	ILD, myositis	-	+ (3/4)	+ (2/2)	+ (0/1)	Hamaguchi, et al. 2013 [19]
Jo-1	HARS	SSc, PM/DM	ILD, myositis	-	+ (1/4)	-	-	Hamaguchi, et al. 2013 [19]
PL-7	TARS	SSc, PM/DM	ILD, myositis	+ (2/2)	++ (8/4)	++ (6/4)	+ (2/4)	Hamaguchi, et al. 2013 [19]
Ha	YARS	SSc, PM/DM	Interstitial pneumonia	-	+ (0/1)	-	-	Hashish, et al 2005 [41]
Zo	**FARSA**, FARSB	SSc, PM/Scl	Anti-synthetase syndrome	-	+ (2/4)	-	-	Betteridge, et al. 2007 [42]
SSc autoantibodies not detected in our dataset								
Fibrillarin	U3RNP	dSSc	More frequent in blacks; severe disease, poor prognosis	-	-	-	-	Mehra, et al. 2013 [3]
U11/U12 RNP	SNRNP35	SSc	Lung fibrosis, gastrointestinal involvement	-	-	-	-	Mimori, 1999 [43]
PDGFR	PDGFR	SSc	Unknown	-	-	-	-	Svegliati Baroni, et al. 2006 [44]
MMP	MMP family	dSSc	Skin, lung, and vascular fibrosis	-	-	-	-	Mehra, et al. 2013 [3]
tPA	PLAT	lSSc	PAH	-	-	-	-	Mehra, et al. 2013 [3]
IFI16	IFI16	lSSc	Common in SARDs	-	-	-	-	Mehra, et al. 2013 [3]
Fibrillin 1	FBN1	dSSc	Choctaw and Japanese patients; absent in Caucasians	-	-	-	-	Mehra, et al. 2013 [3]
Vascular Receptors	AGTR2, EDN1	SSc	TOP1+ patients, renal crisis	-	-	-	-	Mehra, et al. 2013 [3]
ATF2	ATF2	SSc	Longer disease duration, decreased lung function	-	-	-	-	Mehra, et al. 2013 [3]

Data are presented as the average of all peptide hits across each autoantibody group, followed by the frequency of peptide detection within the group. For autoantibodies known to target more than one protein or subunit, data for a single representative protein are shown, with the specific protein highlighted in bold. Associated proteins indicate specific protein targets identified in this study; among autoantibodies not identified here, the most common targets are listed. Symbols: -, +, ++, and +++ indicate an average of 0, 1–4, 5–9, and ≥10 peptide hits per group, respectively

SSc systemic sclerosis, *lSSc* limited cutaneous SSc, *dSSc* diffuse cutaneous SSc, *PAH* pulmonary arterial hypertension; *ILD* interstitial lung disease; *CREST* CREST syndrome (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia); *PM/Scl* polymyositis/scleroderma, *PM/DM* polymyositis/dermatomyositis

helicases proteins, while community 8 contains a diverse array of proteins including multiple SSc autoantibodies, including TOP1, SSB, Pm/Scl proteins, URNPs, and HNRNPs, and numerous serine/arginine-rich splicing factors. Community 3 consists primarily of aminoacyl tRNA synthetases, a cluster often targeted in autoimmune diseases [19, 23]. Communities 4, 5, and 9 are strongly associated with a variety of GO processes known to play a major role in SSc, including wound healing, IFN signaling, and response to oxidative stress. Major proteins include CD44, HLAs, myosins, and filamin proteins in community 4 and tricarboxylic acid cycle proteins in community 5. Community 9 contains multiple protein disulfide isomerases and peroxiredoxins, protein folding enzymes such as calnexin (CANX) and calreticulin (CALR), and the major collagen processing enzyme prolyl 4-hydroxylase beta (P4HB). Community 6 contains multiple annexin and 14-3-3 proteins; enriched GO processes include ribonucleoprotein complex assembly, mitochondrial transport, RNA processing, and anchoring junction. Community 7 associated with GO terms include cell cycle, RNA polymerase III complex, DNA-PK-Ku complex, and antigen processing and presentation. Community 7 includes several SSc autoantibody targets including Ku proteins XRCC5 and 6, RUVBL1 and 2, RNA polymerase I and II subunits, multiple proteasomal subunits, and T-complex proteins.

Preferential detection of autoantibodies in SSc

Subsequent comparisons between groups were performed in a semiquantitative manner based on the presence or absence of a given protein in an immunoprecipitant, with quantitative analyses limited to comparisons within an individual sample. To identify biological processes and cellular components differentially targeted in SSc, with minimal to no background detection in controls, we examined all proteins detected in >50 % of SSc samples at a frequency >1.5-fold relative to controls, resulting in a list of 137 differentially detected proteins (Fig. 2; Additional file 2: Table S2). Enriched biological processes included *ncRNA metabolic process*, response to oxygen radical, and triglyceride-rich lipoprotein particle remodeling. Preferentially targeted cellular components include cytosolic stress granule, lipid-protein complex, pigment granule, and anchoring junction; molecular functions include antioxidant activity and mRNA binding (Additional file 3: Table S3C).

RNA processing centers are major targets of SSc autoantibodies

The strong enrichment for GO terms associated with mRNA processing and stress response, as well as the identification of cytosolic stress granule as an enriched cellular component, led us to further investigate the role of stress granules (SG) and RNA processing bodies (PB) in the autoantibody response of SSc. SG and PB represent distinct, non-membranous cytoplasmic entities, which arise in response to different cellular stresses including oxidative stress, hypoxia, viral infection, unfolded proteins, and amino acid deprivation [24]. These structures exist in constant flux, driven by the availability of constituent mRNPs, regulating the fate of untranslated mRNAs in response to translational arrest [25]. While SG are generally absent under normal conditions, PB are constitutively present at low levels due to their role as microRNA processing centers. Both structures have been shown to arise in response to cellular stresses, including oxidative stress, ischemia, and cancer [26], all of which are known to be important in SSc pathogenesis [5, 27].

In addition to the 137 differentially detected proteins described above, a wide range of PB/SG constituents were readily detected across most SSc samples (Additional file 5: Table S4). Substantial reactivity was seen against PB components such as UPF1 and MOV10, and SG proteins FXR1 and FXR2, G3BP1 and G3BP2, and USP10. Only background levels of reactivity were seen in healthy controls.

Validation of PB/SG antibodies in SSc

In order to validate the differential abundance of PB/SG proteins identified by LC-MS/MS, HeLa whole cell lysates were immunoprecipitated using antibodies from each patient as described in the LC-MS/MS analyses. Western blots were performed by resolving equal volumes of IP eluates by SDS-PAGE and transferring to nitrocellulose. Blots were then probed with antibodies targeting PB/SG proteins UPF1, MOV10, CAPRIN1, G3BP1, and USP10. These targets were chosen based upon their high level of detection across SSc groups, as determined based on our LC-MS/MS data. Strong reactivity was seen against all five proteins in SSc with only trace levels detected in controls (Fig. 3a), indicating widespread immune responses against these protein complexes.

Further validation was performed using immunofluorescence (IF) staining of U2OS cells maintained under conditions of oxidative stress to induce PB/SG formation. Cells were probed with patient sera in combination with PB and SG markers SK1-Hedls and eIF3b, respectively. Co-localization between patient sera and PB/SG markers was observed in six of nine SSc patients, with at least one positive sample in each of the three autoantibody groups; no staining was seen for any of the three healthy controls (Fig. 3b). These results are consistent with that seen on LC-MS/MS, particularly among RNAP3 patients, who exhibited the strongest and most consistent autoantibody responses across both methods. Taken together, these data strongly implicate PB/SG as novel targets of SSc autoantibody responses.

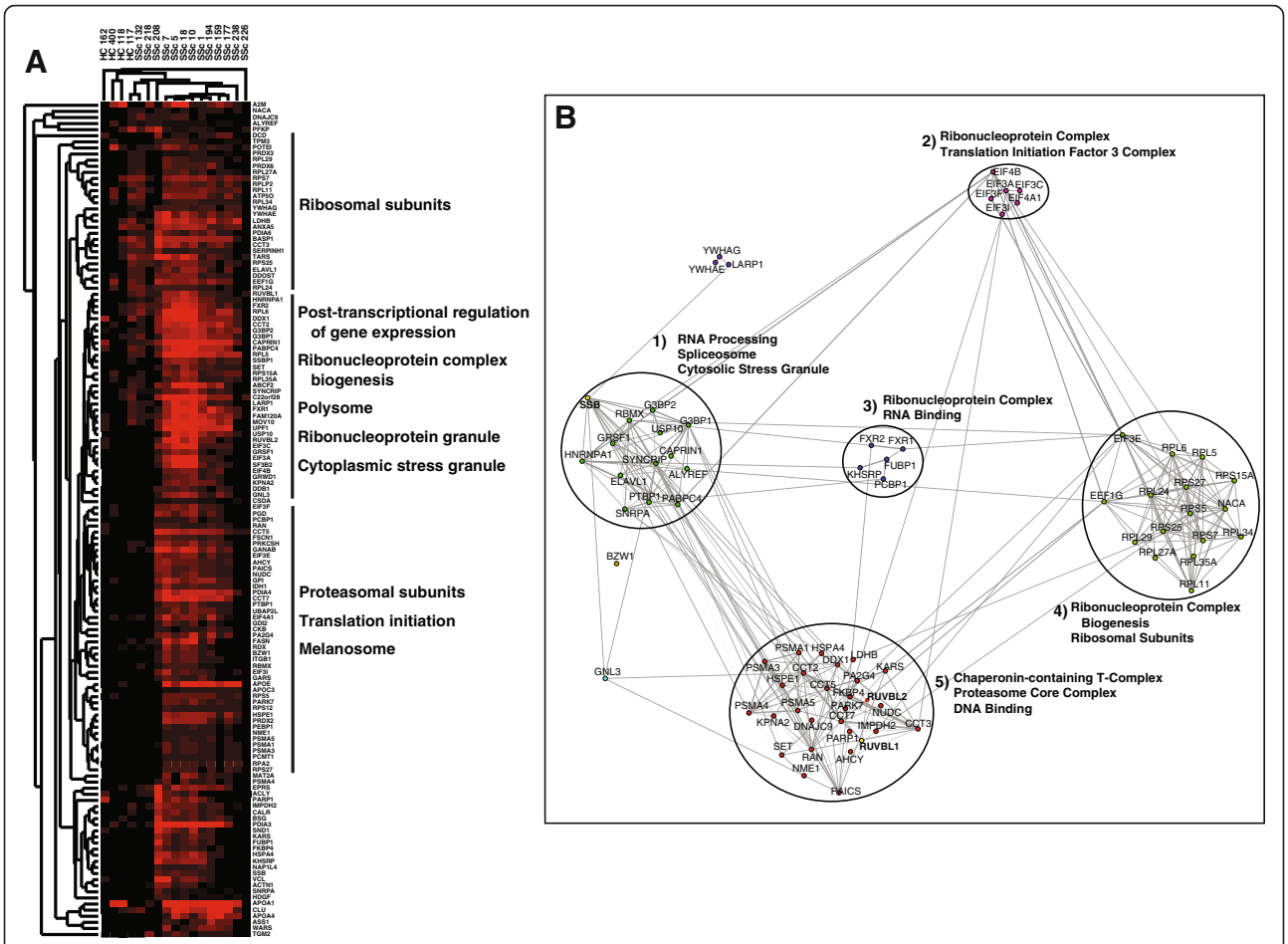


Fig. 2 Proteins differentially detected in systemic sclerosis (*SSc*). Semiquantitative enrichment of SSc-associated autoantibodies was determined using a binary assessment of autoantibody presence or absence in a sample. Preferential enrichment in SSc was defined as all proteins detected in >50 % of all patient samples at a frequency >1.5-fold relative to controls. **a** Heat map of proteins differentially detected in SSc. **b** Network analysis of differentially detected proteins. Community detection was performed using the GIANT global network; functional annotation was performed using gProfiler

Discussion

Autoantibodies have long been used in the diagnosis of SSc, with different autoantibodies predictive of clinical outcomes, including interstitial lung disease, pulmonary arterial hypertension, and skin involvement. While a wide array of SSc-associated autoantibodies have been described, diagnoses are often performed based upon the presence or absence of reactivity against three proteins: RNAP3, TOP1, and CENP. The data presented here suggest a much broader autoantibody response, which is reflective of underlying disease pathologies. Strong subset-specific reactivity was evident against both RNAP3 and TOP1, with no RNAP3 peptides detected in any of the other groups; however, all four RNAP3 patients exhibited modest reactivity against TOP1, indicating a degree of overlap between these two autoantibody groups. When peptides recovered are extended beyond the three major targets, we found substantial overlap across the three major SSc groups. We found peptides from the autoantigens of RuvBL1/2, which appear to act as general markers of SSc, with consistent detection across all SSc groups, with almost no reactivity seen in controls. In contrast, some common SSc autoantigens such as B23 and Ro52/TRIM21 were recovered in virtually all samples, and in controls, indicating an important degree of baseline reactivity against some of the more common autoantibody targets.

In this proof-of-concept study, we do not attempt to address the clinical implications of the autoantibody responses described here due to the limited number of patients analyzed, nor are we able to speculate on the presence or absence of these autoantibodies in other SSc autoantibody groups. Our depth in this study comes from the number of potential antigens analyzed, which cover the full proteome. Future studies examining a much larger cohort of SSc patients, along with representatives of other

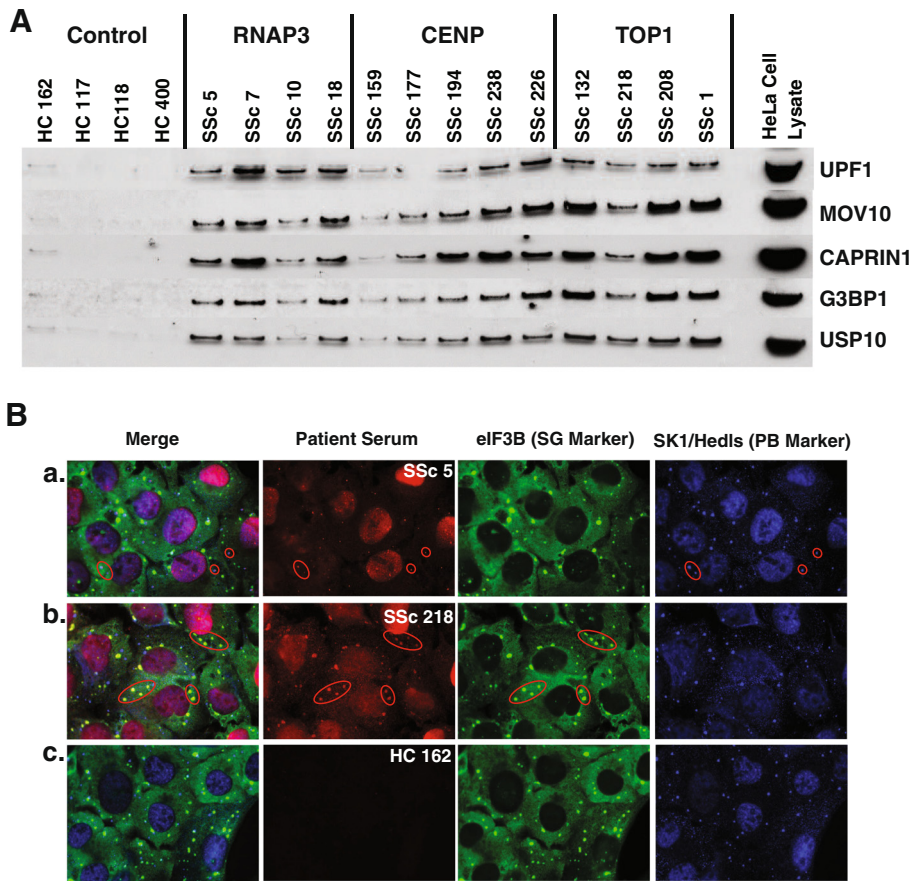


Fig. 3 Validation of RNA processing bodies (*PB*)/stress granules (*SG*) as a target of the SSc autoimmune response. **a** HeLa cell lysates were immunoprecipitated using patient sera, resolved by SDS-PAGE, and probed with antibodies targeting known PB and SG proteins; HeLa whole cell lysate was used as a control. **b** Immunofluorescence was performed in U2OS cells treated with sodium (meta)arsenite to induce the formation of SG. Cells were then fixed with 4 % paraformaldehyde and permeabilized with 5 % normal horse serum and 0.1 % digitonin in Tris-buffered saline. Staining was performed with anti-eIF3b (SG marker), anti-SK1-Hedls (PB marker), and patient sera. Representative images depicting co-localization between patient sera and SG/PB markers are shown, with sites of co-localization circled in *red*. *RNAP* RNA polymerase, *CENP* centromere protein *TOPI* topoisomerase I

autoimmune diseases, will be necessary to determine the clinical value of these potential autoantibodies.

This is not the first study to suggest the presence of multiple autoantibodies in SSc. Immunoassays performed by Op De Beeck, et al. revealed the presence of multiple autoantibodies in a small subset of SSc patients [28]. A similar analysis by Graf et al. using the EUROLINE immunoassay revealed the presence of multiple autoantibodies in 11 % of patients [1].

Autoantibodies against extracellular immune signaling receptors and extracellular matrix proteins were conspicuously absent in these data; this includes the absence of numerous autoantibodies previously implicated in SSc pathogenesis, such as anti-fibrillin 1, anti-MMP, and anti-PDGFR [29]. Additional analyses in other cell types, such as fibroblasts or endothelial cells, and cells maintained under physiologically relevant growth conditions, such as immune activation or oxidative stress, may be useful for identifying other proteins and complexes which may play a role in disease pathogenesis.

In addition to identifying novel autoantibody targets, the unbiased nature of mass spectrometry provides additional insights into the processes potentially underlying autoimmunity. The preferential detection of proteins associated with RNA processing and oxidative stress as a general feature of SSc autoantibodies may be indicative of their origins. Combined with the consistent targeting of PB/SG described here spanning all SSc patients, these data suggest a basic model in which disease-specific pathologies give rise to specific autoantibodies. Strong induction of SGs is observed in response to cellular stresses, including oxidative stress and ischemia, two well-established phenomena in SSc [27]. SG/PB are also readily induced in response to the tumor microenvironment, consistent with recent evidence linking RNAP3-positive SSc and cancer [5, 30]. Combined with evidence

linking transforming growth factor (TGF)-β signaling with an increase in PB formation [31], many of the major processes underlying SSc pathogenesis appear broadly consistent with an immune response against cells undergoing a stress response. PBs are also known to associate with other cytoplasmic structures, such as U bodies [32], which house an number of well-established SSc autoantibody targets, including U1, U5, and U11/U12. Taken together, these data suggest a model in which autoantibodies arise as a secondary phenotype in response to SSc-related processes already underway.

Some evidence suggesting a link between PBs and autoimmunity has been described previously. Bhanji, et al. observed reactivity against a number of PB-associated proteins, including GE1/*Hedls*, GW182, and Ago2 in a range of autoimmune diseases, including SSc [33]. Among these autoantigens, only Ago2 was also identified in this analysis, suggesting a persistent, yet diverse response against this complex. A similar analysis of SG-associated proteins has not been performed; however given the reactivity to PB proteins seen in other autoimmune diseases, reactivity in other autoimmune diseases is possible. A detailed analysis using both mass spectrometry and other methods will be necessary to understand the degree to which autoantibodies to SG/PBs are seen in other autoimmune diseases, and the clinical implications of these findings.

This work has several limitations. First, we cannot eliminate the possibility that some proteins found in our mass spectrometry data result from co-IP of multiprotein complexes by a single autoantibody; however, we were able to confirm the presence of multiple PB/SG autoantibodies by other means (Fig. 3). We also cannot rule out the possibility that some targets were missed due to their being sequestered into tightly packed molecular complexes associated with chromatin. For example, the presence of CENP autoantibodies within these samples had been established using clinical methods, indicating its absence in our mass spectrometry data is likely a result of its sequestration into large macromolecular complexes with limited solubility. A lack of age- and gender-matched controls likely underestimates the degree of baseline reactivity seen in unaffected controls. Finally, the small number of patient samples used in this study prevents any clinical interpretation, and the variability in the number of peptides recovered between experiments limits direct quantitative comparisons between autoantibody groups.

Conclusions

The data presented here provide evidence of diverse immune reactivities in SSc targeting a wide array of protein complexes. Among these complexes, autoantibodies targeting PB/SG were consistently identified across both clinical SSc subsets and major autoantibody groups, suggesting a potential novel autoantibody target. Taken together, these data suggest immune responses to proteins involved in cellular stress may be a common mechanism for autoantibody generation.

Additional files

Additional file 1: Table S1. Complete list of peptides identified in this analysis. *TP* number of total peptides mapping to a protein, *UP* number of unique peptides mapping to a protein, *UM* number of non-redundant peptides mapping exclusively to a protein, *MW* molecular weight, *Length* protein length in amino acids. (XLS 639 kb)

Additional file 2: Table S2. Gene lists used in these analyses. (XLSX 17 kb)

Additional file 3: Table S3. Systemic sclerosis (SSc)-specific enrichment of processes and components. Proteins differentially detected in SSc were analyzed using gProfiler. Statistically significant processes and components are shown. A) Peptides detected at any level across all four groups. B) Peptides identified in all SSc groups, but absent in controls. C) Analysis of 137 proteins differentially detected in SSc. *BP* biological process, *CC* cellular component, *MF* molecular function, *ke* KEGG pathway, *re* REACTOME pathway. (ODS 70 kb)

Additional file 4: Figure S1. Network analysis of systemic sclerosis (SSc) autoantigens. All 763 non-redundant peptide hits identified in two or more patients were analyzed using the Genome-scale Integrated Analysis of gene Networks in Tissues (GIANT) global network to identify functionally associated protein networks. Analysis of community function was performed using gProfiler. SSc-associated autoantibodies are highlighted in *yellow*. (EPS 1902 kb)

Additional file 5: Table S4. Processing body and stress granule proteins identified in this analysis. *Proteins with multiple subunits. Data indicate non-redundant peptide hits. (ODS 20 kb)

Abbreviations

CENP: centromere protein; DMEM: Dulbecco's modified Eagle's medium; dSSc: diffuse systemic sclerosis; ELISA: enzyme-linked immunosorbent assay; FBS: fetal bovine serum; GIANT: Genome-scale Integrated Analysis of gene Networks in Tissues; IFN: interferon; kDa: kiloDalton; LC-MS/MS: liquid chromatography tandem-mass spectrometry; IP: immunoprecipitation; lSSc: limited systemic sclerosis; PB: RNA processing bodies; PBS: phosphate-buffered saline; RNAP3: RNA polymerase III; SG: stress granules; SSc: systemic sclerosis; TOP1: topoisomerase I.

Competing interests

Dr. Whitfield has received royalties for patents regarding gene expression biomarkers in Scleroderma and is a scientific founder of Celdara Medical LLC. Dr. Lafyatis has received both grants and consulting fees from Genzyme/Sanofi, Shire, Regeneron, Biogen, BMS, Inception, Precision Dermatology, PRISM, UCB, Pfizer and Roche/Genentech; he received consulting fees from Lycera, Novartis, Celgene, Amira, Celdara, Celltex, Dart Therapeutics, Idera, Intermune, Medimmune, Promedior, Zwitter, Actelion, EMD Serono, Akros, Extera, Reneo, Scholar Rock, and HGS. No authors have any non-financial conflicts of interest to report.

Authors' contributions

MEJ conceived of the study, performed experiments, analyzed data, and wrote the manuscript. AVG performed mass spectrometry and helped to revise the manuscript. JNT performed data analysis and helped to revise the manuscript. SML performed immunofluorescence experiments, and helped to revise the manuscript. DS performed mass spectrometry, and revised the manuscript. JKG, RFS, and RL provided clinical samples, and revised the manuscript. PJA designed experiments, provided technical assistance, and revised the manuscript. SAG and MLW conceived of the study, participated in its design, and helped to revise the manuscript. All authors read and approved the final manuscript.

Acknowledgements
This work was supported by grants from the NIH National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) Center of Research Translation (P50 AR060780 to MLW and RL), the Department of Defense (PR130908 to MLW), and a SYNERGY grant from the Geisel School of Medicine at Dartmouth (to MLW). JNT is supported in part by a grant from the National Institute of General Medical Sciences (NIGMS; T32GM008704). SAG is supported by grants R01-CA155260 and S10-OD016212 from the NIH. JKG is supported by a Kellen Foundation Clinician Scientist Development Award from the Hospital for Special Surgery. Written informed consent was obtained from all participants for publication of their individual details in this manuscript. All consent forms are held by the respective authors' institutions and are available for review by the Editor-in-Chief.

Author details
[1]Department of Genetics, Geisel School of Medicine at Dartmouth, Hanover, NH, USA. [2]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, MA, USA. [3]Department of Rheumatology, Hospital for Special Surgery, New York, NY, USA. [4]Boston University School of Medicine, Boston, MA, USA. [5]Dartmouth Medical School, Hinman Box 7400, Hanover, NH 03755, USA.

Received: 6 October 2015 Accepted: 3 January 2016
Published online: 22 January 2016

References

1. Graf SW, Hakendorf P, Lester S, Patterson K, Walker JG, Smith MD, et al. South Australian Scleroderma Register: autoantibodies as predictive biomarkers of phenotype and outcome. Int J Rheum Dis. 2012;15:102–9.
2. Steen VD. Autoantibodies in systemic sclerosis. Semin Arthritis Rheum. 2005; 35:35–42.
3. Mehra S, Walker J, Patterson K, Fritzler MJ. Autoantibodies in systemic sclerosis. Autoimmun Rev. 2013;12:340–54.
4. Fertig N, Domsic RT, Rodriguez-Reyna T, Kuwana M, Lucas M, Medsger TA, et al. Anti–U11/U12 RNP antibodies in systemic sclerosis: A new serologic marker associated with pulmonary fibrosis. Arthritis Care Res. 2009;61:958–65.
5. Joseph CG, Darrah E, Shah AA, Skora AD, Casciola-Rosen LA, Wigley FM, et al. Association of the autoimmune disease scleroderma with an immunologic response to cancer. Science. 2014;343:152–7.
6. Robinson WH, DiGennaro C, Hueber W, Haab BB, Kamachi M, Dean EJ, et al. Autoantigen microarrays for multiplex characterization of autoantibody responses. Nat Med. 2002;8:295–301.
7. Yore MM, Kettenbach AN, Sporn MB, Gerber SA, Liby KT. Proteomic analysis shows synthetic oleanane triterpenoid binds to mTOR. PLoS ONE. 2011;6:e22862.
8. Eng JK, Jahan TA, Hoopmann MR. Comet: An open-source MS/MS sequence database search tool. Proteomics. 2013;13:22–4.
9. Hsieh EJ, Hoopmann MR, MacLean B, MacCoss MJ. Comparison of database search strategies for high precursor mass accuracy MS/MS data. J Proteome Res. 2009;9:1138–43.
10. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods. 2007;4:207–14.
11. Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. 2007.
12. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, et al. Zaslavsky E. Understanding multicellular function and disease with human tissue-specific networks. Nat Genet. 2015;47:569–76.
13. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.
14. Reimand J, Arak T, Vilo J. g: Profiler—a web server for functional interpretation of gene lists (2011 update). Nucleic Acids Res. 2011;39:W307–15.
15. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. PNAS. 1998;95:14863–8.
16. Saldanha AJ. Java Treeview—extensible visualization of microarray data. Bioinformatics. 2004;20:3246–8.
17. Novoradovskaya N, Perou C, Whitfield M, Basehore S, Pesich R, Aprelikova O, et al.: Universal human, mouse and rat reference RNA as standards for microarray experiments. In: Mol Bio Cell: 2002: Amer Soc Cell Biology 8120 Woodmont Ave, STE 750, Bethesda, MD 20814-2755 USA; 2002: 241A-241A.
18. Kaji K, Fertig N, Medsger TA, Satoh T, Hoshino K, Hamaguchi Y, et al. Autoantibodies to RuvBL1 and RuvBL2: A novel systemic sclerosis-related antibody associated with diffuse cutaneous and skeletal muscle involvement. Arthr Care Res. 2014;66:575–84.
19. Hamaguchi Y, Fujimoto M, Matsushita T, Kaji K, Komura K, Hasegawa M, et al. Common and Distinct Clinical Features in Adult Patients with Anti-Aminoacyl-tRNA Synthetase Antibodies: Heterogeneity within the Syndrome. PLoS ONE. 2013;8:e60442.
20. Fujimoto M, Shimozuma M, Yazawa N, Kubo M, Ihn H, Sato S, et al. Prevalence and clinical relevance of 52-kDa and 60-kDa Ro/SS-A autoantibodies in Japanese patients with systemic sclerosis. Ann Rheum Dis. 1997;56:667–70.
21. Ulanet DB, Wigley FM, Gelber AC, Rosen A. Autoantibodies against B23, a nucleolar phosphoprotein, occur in scleroderma and are associated with pulmonary hypertension. Arth Care Res. 2003;49:85–92.
22. Brouwer R, Vree Egberts WTM, Hengstman GJD, Raijmakers R, van Engelen BGM, Peter Seelig H, et al. Autoantibodies directed to novel components of the PM/Scl complex, the human exosome. Arth Res. 2002;4:134–8.
23. Lega J-C, Fabien N, Reynaud Q, Durieu I, Durupt S, Dutertre M, et al. The clinical phenotype associated with myositis-specific and associated autoantibodies: A meta-analysis revisiting the so-called antisynthetase syndrome. Autoimmun Rev. 2014;13:883–91.
24. Kedersha N, Ivanov P, Anderson P. Stress granules and cell signaling: more than just a passing phase? Trends Biochem Sci. 2013;38:494–506.
25. Kedersha N, Anderson P. Regulation of translation by stress granules and processing bodies. Prog Mol Biol Transl Sci. 2009;90:155–185.
26. Anderson P, Kedersha N. Stress granules: the Tao of RNA triage. Trends Biochem Sci. 2008;33:141–50.
27. Katsumoto TR, Whitfield ML, Connolly MK. The pathogenesis of systemic sclerosis. Annu Rev Pathol-Mech. 2011;6:509–37.
28. Op De Beéck K, Vermeersch P, Verschueren P, Westhovens R, Mariën G, Blockmans D, et al. Antinuclear antibody detection by automated multiplex immunoassay in untreated patients at the time of diagnosis. Autoimmun Rev. 2012;12:137–43.
29. Chung L, Utz P. Antibodies in scleroderma: Direct pathogenicity and phenotypic associations. Curr Rheumatol Rep. 2004;6:156–63.
30. Anderson P, Kedersha N, Ivanov P. Stress granules, P-bodies and cancer. Biochim Biophys Acta. 2015, 1849, 861–870.
31. Blanco FF, Sanduja S, Deane NG, Blackshear PJ, Dixon DA. Transforming growth factor β regulates P-body formation through induction of the mRNA decay factor tristetraprolin. Mol Cell Bio. 2014;34:180–95.
32. Liu J-L, Gall JG. U bodies are cytoplasmic structures that contain uridine-rich small nuclear ribonucleoproteins and associate with P bodies. PNAS. 2007; 104:11655–9.
33. Bhanji RA, Eystathioy T, Chan EKL, Bloch DB, Fritzler MJ. Clinical and serological features of patients with autoantibodies to GW/P bodies. Clin Immunol. 2007;125:247–56.
34. Dib H, Tamby MC, Bussone G, Regent A, Berezné A, Lafine C, Broussard C, Simonneau G, Guillevin L, Witko-Sarsat V, et al. Targets of anti-endothelial cell antibodies in pulmonary hypertension and scleroderma. Eur Respir J. 2012;39(6):1405–1414.
35. Naniwa T, Sugiura Y, Banno S, Yoshinouchi T, Matsumoto Y, Ueda R. Ribosomal P protein P0 as a candidate for the target antigen of anti-endothelial cell antibodies in mixed connective tissue. Clin Exp Rheumatol. 2007;25:593–598.
36. Terrier B, Tamby MC, Camoin L, Guilpain P, Broussard C, Bussone G, et al. Identification of target antigens of antifibroblast antibodies in pulmonary arterial hypertension. Am J Resp Crit Care Med. 2008;177(10):1128–1134.
37. Terrier B, Tamby MC, Camoin L, Guilpain P, Bérézné A, Tamas N, et al. Anti-fibroblast antibodies from systemic sclerosis patients bind to α-enolase and are associated with interstitial lung disease. Ann Rheum Dis. 2009.
38. Satoh M, Chan JY, Ceribelli A, del-Mercado MV, Chan EK. Autoantibodies to Argonaute 2 (Su antigen). In: Ten Years of Progress in GW/P Body Research. Springer. 2013:45-59.
39. Siapka S, Patrinou-Georgoula M, Vlachoyiannopoulos PG, Guialis A. Multiple specificities of autoantibodies against hnRNP A/B proteins in systemic rheumatic diseases and hnRNP L as an associated novel autoantigen. Autoimmunity. 2007;40(3):223–233.
40. Kubo M, Ihn H, Kuwana M, Asano Y, Tamaki T, Yamane K, Tamaki K. Anti-U5 snRNP antibody as a possible serological marker for scleroderma–polymyositis overlap. Rheumatol. 2002;41(5):531–534.

41. Hashish L, Trieu E, Sadanandan P, Targoff I. Identification of autoantibodies to tyrosyl-tRNA synthetase in dermatomyositis with features consistent with anti-synthetase syndrome. In: Arthritis Rheum. 2005; 2005: S312-S312
42. Betteridge Z, Gunawardena H, North J, Slinn J, McHugh N. Anti-synthetase syndrome: a new autoantibody to phenylalanyl transfer RNA synthetase (anti-Zo) associated with polymyositis and interstitial pneumonia. Rheumatol. 2007;46(6):1005–1008.
43. Mimori T. Autoantibodies in Connective Tissue Diseases. Clinical Significance and Analysis of Target Autoantigens. Internal Med. 1999;38(7):523–532.
44. Svegliati BS, Santillo M, Bevilacqua F, Luchetti M, Spadoni T, Mancini M, et al. Stimulatory autoantibodies to the PDGF receptor in systemic sclerosis. N Engl J Med. 2006;354(25):2667–2676.

HHS Public Access

Author manuscript

Semin Immunopathol. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:

Semin Immunopathol. 2015 September ; 37(5): 501–509. doi:10.1007/s00281-015-0512-6.

Gene expression profiling offers insights into the role of innate immune signaling in SSc

Michael E. Johnson[1], **Patricia A. Pioli**[2,3], and **Michael L. Whitfield**[1]

Patricia A. Pioli: patricia.a.pioli@dartmouth.edu; Michael L. Whitfield: michael.whitfield@dartmouth.edu

[1]Department of Genetics, Geisel School of Medicine at Dartmouth, 7400 Remsen, Hanover, NH 03755, USA

[2]Department of Obstetrics and Gynecology, Geisel School of Medicine at Dartmouth, One Medical Center Drive, Lebanon, NH 03756, USA

[3]Department of Microbiology and Immunology, Geisel School of Medicine at Dartmouth, One Medical Center Drive, Lebanon, NH 03756, USA

Abstract

Systemic sclerosis (SSc) is characterized by inflammation, vascular dysfunction, and ultimately fibrosis. Progress in understanding disease pathogenesis and developing effective disease treatments has been hampered by an incomplete understanding of SSc heterogeneity. To clarify this, we have used genomic approaches to identify distinct patient subsets based on gene expression patterns in SSc skin and other end-target organs. Here, we review what is known about the gene expression-based subsets in SSc, currently defined as the inflammatory, fibroproliferative, limited, and normal-like subsets. The inflammatory subset of patients is characterized by infiltrating immune cells that include T cells, macrophages, and possibly dendritic cells, although little is known about the mediators these cells secrete and the pathways that govern cell activation. Prior studies have suggested a role for pathogens as a trigger of immune responses in SSc, and recent data have identified viral and mycobiome components as potential environmental triggers. We present a model based on analyses of gene expression data and a review of the literature, which suggests that the gene expression subsets observed in patients possibly represent distinct, interconnected molecular states of disease, to which an innate immune response is central that results in the generation of clinical disease.

Introduction

Systemic sclerosis (SSc) is a clinically heterogeneous autoimmune disease characterized by fibrosis of the skin and internal organs, vascular abnormalities, and persistent immune activation. While the etiology of SSc remains poorly understood, the earliest clinical symptoms are primarily associated with the vascular system, characterized by vasospastic episodes, referred to as Raynaud's phenomenon. However, despite similarities in early symptoms, substantial heterogeneity exists between patients with respect to disease

Correspondence to: Michael L. Whitfield, michael.whitfield@dartmouth.edu.

progression and the organs affected, hindering our understanding of pathophysiology and complicating the interpretation of clinical trials.

To overcome issues of clinical heterogeneity inherent in SSc, high-throughput gene expression has been used to better understand the pathways and processes that drive the disease. These analyses can be used to identify the major cell types driving pathogenesis in a complex tissue sample as well as to define reproducible gene expression profiles indicative of different forms of disease. Three independent skin biopsy datasets have been generated that identify reproducible gene expression subtypes characteristic of different forms or states of SSc [1–4]. Recent gene expression analyses of SSc lung biopsies provide an important addition to these efforts, expanding our understanding of disease-associated gene expression to a second end-target tissue [5, 6]. Here, we provide an update regarding insights into SSc pathogenesis using gene expression profiling, with an emphasis on the role of the innate immune system as a potential initiator and driver of disease pathology.

Identification of gene expression-based intrinsic subsets of disease

The initial gene expression studies in SSc skin biopsies focused on small cohorts and the identification of differences between SSc and healthy controls [7, 8]. These studies revealed both inflammatory and fibrotic gene expression signatures that characterized diseased tissue. One of these studies [7] showed a surprising result, which was the nearly identical disease-specific patterns of gene expression in biopsies taken from lesional forearm and non-lesional back skin of an SSc patient. This suggested that aberrant gene expression could be found even in unaffected tissues, highlighting the truly systemic nature of the disease [7]. A second result from these studies was that fibroblasts grown in culture do not accurately recapitulate the aberrant gene expression observed in SSc skin, with similar results seen using paired skin biopsies and fibroblast cultures in Gardner et al. [8], as well as non-paired samples in Whitfield et al. [7].

Clinical heterogeneity is a major factor confounding our understanding of SSc. Early studies examining heterogeneity in tumors were able to demonstrate the existence of reproducible gene expression subsets within a given tumor type [9–13]. Using the same approach to understand the variability in end-target tissues affected by SSc was therefore logical; however, this type of analysis had not been previously used for autoimmune diseases. The first reported study of gene expression heterogeneity in SSc by Milano et al. [1] identified four “intrinsic” gene expression subsets among patients with SSc. These included a fibroproliferative subset, which exhibited strong induction of proliferation genes, an inflammatory subset characterized by robust upregulation of genes associated with both innate and adaptive immune responses, a limited subset centered on a cluster of clinically limited patients, and a normal-like subset, which consisted of both healthy controls and a subset of patients with both limited and diffuse SSc. This study provided a proof-of-concept that heterogeneity within SSc clinical groups could be measured using genome-wide molecular profiling.

An analysis of two subsequent cohorts of patients confirmed and expanded these observations. Pendergrass et al. [2] reproduced the fibroproliferative, inflammatory, and

normal-like subsets, driven by similar proliferation and immune-related signals; the limited SSc subset was not found due to the absence of lSSc patients in this cohort. Pendergrass et al. included serial biopsies from an investigator-initiated trial of rituximab, which showed no clinical efficacy [14] but demonstrated stable gene expression within patients over the time period analyzed (6–12 months).

In a third independent study, Hinchcliff et al. [3] examined the effects of mycophenylate mofetil (MMF), a potent immunosuppressant commonly used to treat SSc. Intrinsic subset assignments were again reproducible in patients for as long as 24 months, indicative of long-term stability in terms of gene expression. Furthermore, four patients classified in the inflammatory gene expression subset demonstrated clear clinical improvement with MMF treatment. In contrast, three patients classified as either fibroproliferative or normal-like failed to improve, consistent with the use of intrinsic subset assignment as a predictor of treatment outcomes.

Two other recent investigator-initiated clinical trials for abatacept [15] and nilotinib [16] have also shown subset-specific responses to therapy. Chakravarty et al. [15] found that patients who improve while on abatacept therapy map to the inflammatory gene expression subset. Comparison of pre- and post-treatment biopsies revealed a decrease in CD28 co-stimulatory signaling, which is the molecular target of abatacept. This pathway change was only observed in patients who improved during treatment, with no changes seen in non-responders or placebo controls. These data suggest that patients in the inflammatory subset are the most likely to improve with abatacept therapy. Another study by Gordon et al. [16] showed that patients who improve with nilotinib, a tyrosine kinase inhibitor (TKI), have high expression of genes associated with increased transforming growth factor β (TGF-β) and platelet-derived growth factor (PDGF) signaling at baseline, suggesting a mechanistic link between pathway activation and clinical improvement. Those who failed to improve did not show high-level expression of these genes. This study argues for targeting patients with TGF-β/PDGF pathway activation with TKIs.

Most recently, the inflammatory and normal-like gene expression subsets have been reproduced in skin biopsies derived from an independent cohort and analyzed by an independent group of investigators; however, they were unable to reproduce the fibroproliferative subset identified in prior studies. An important result from this study that provides insight into the etiology of the subsets was the observation that normal-like patients have the longest disease duration and likely represent end-stage, inactive disease [17].

Meta-analyses of gene expression in skin and genetic polymorphisms

To better identify the genes and processes common across three published SSc datasets [1–3], Mahoney et al. [4] developed a data mining procedure termed mutual information consensus clustering (MICC) to identify co-expression modules conserved across datasets. As co-expressed genes tend to be functionally related, this method permits identification of processes central to the pathology of SSc. Two groups of genes (consensus gene clusters) were consistently identified. A direct analysis of gene-gene interactions within and between the two consensus gene clusters was performed using the IMP gene-gene interaction

Bayesian network in association [18] with 41 SSc-associated polymorphic genes identified by genome-wide association studies (GWAS) [19]. Genes were shown to cluster into five distinct communities, each associated with a distinct biological process important to SSc pathogenesis. Three groups of genes that were strongly associated with the inflammatory subset showed enrichment for processes associated with response to interferon signaling, B cell receptor signaling, adaptive immune processes, monocyte chemotaxis, and M2 macrophage activation. Another group, consisting of genes from both the inflammatory and fibroproliferative clusters, exhibited enrichment for genes associated with TGF-β and PDGF signaling as well as extracellular matrix (ECM) remodeling. These findings suggest that these processes, in part, connect the inflammatory response to ECM deposition. Finally, a set of genes with increased expression in fibroproliferative patients was strongly associated with cell proliferation. An important result from this study was that 30 of 41 SSc-associated polymorphic genes were associated with inflammatory genes and processes or formed bridges between inflammatory processes and the ECM. These results implicate a mechanistic link between genetic risk factors and important disease processes in the network. A second implication from this study is that the intrinsic gene expression subsets are likely mechanistically interconnected with the inflammatory signature, leading to activation of ECM deposition and TGF-β and PDGF signaling, which ultimately results in the proliferative response. These data suggest that the intrinsic subsets are long-lived but interconnected, although capturing the transition from one subset to another experimentally has been difficult.

Expansion of gene expression profiling beyond skin

Gene expression profiles have been shown to be broadly consistent between lesional and non-lesional skin within a given patient [7]; however, these phenotypes are quickly lost in culture, limiting the use of in vitro techniques for the study of SSc [8]. While this phenomenon suggests the disease is systemic, it begs the question of whether or not the subsets observed in the skin are found in other end-target tissues. In this regard, both fibrotic and immune processes have been observed in two separate studies of SSc lung disease. Gene expression has been analyzed in late-stage lung samples from SSc patients that included individuals with both pulmonary arterial hypertension (PAH) and interstitial lung disease (ILD) [5]. These studies revealed increased expression of genes involved in fibrosis in SSc-associated ILD, including type I and type III collagen, IGFBPs, MMP-7, CTGF, osteopontin, and tissue inhibitors of metalloproteases 1 (TIMP-1). SSc-PAH lungs shared functional groups with idiopathic PAH lung samples, showing enrichment for interferon, IL-4, IL-17, and antigen presentation signaling. SSc-associated PAH also showed increased expression of inflammatory genes, including chemokines CCL2, CXCL10, and CX3CL1. A recent study by Christmann et al. used open lung biopsies of patients with SSc-related ILD and healthy controls to study early lung pathogenesis in SSc [6]. A total of 21 patients with either dSSc (n= 11) or lSSc (n=10) was included in this study, representing both early- and late-stage disease. Significantly, upregulated processes included collagen expression, TGF-β signaling, IFN signaling, and M2 macrophage activation, broadly consistent with gene activation clusters in Mahoney et al. Such widespread concordance between the skin and lung implicates major immune-related processes as drivers of pathology in SSc.

Moreover, gene expression analysis of esophageal biopsies of patients with SSc has shown that the intrinsic subsets can be found in a tissue other than the skin [20]. Both the inflammatory and fibroproliferative subsets were identified in SSc esophageal biopsies, suggesting that intrinsic subset gene expression may be a consistent feature of SSc tissues and may represent pathophysiological states of disease. Efforts are currently underway to perform a multi-tissue network analysis for common gene expression found in all SSc end-target tissues, which should provide insight into common mechanisms of pathogenesis (Taroni, Mahoney, Pioli, and Whitfield, *In preparation*).

Pathway activation events underlying the intrinsic gene expression subsets in skin

Given the success of MMF and abatacept as treatments for patients in the inflammatory subset [3, 15] and the suggestion that TKIs may benefit patients with TGF-β and PDGF activation [16, 21], a greater emphasis has been placed on understanding the pathways that underlie each of the four intrinsic subsets with the ultimate goal of identifying therapeutic targets. To this end, two studies comparing in vitro-derived fibroblast gene expression signatures for TGF-β and IL-4/IL-13 against the Milano et al. skin biopsy dataset revealed strong correlations with the fibroproliferative and inflammatory subsets, respectively, indicating a potential role for each of these pathways in disease pathology [22, 23].

A more expansive follow-up study compared 13 gene signatures experimentally derived in dermal fibroblasts against the three skin biopsy datasets described above, along with additional 82 arrays associated with the existing Hinchcliff et al. dataset [24]. Hierarchical clustering of this expanded Milano-Pendergrass-Hinchcliff (MPH) dataset recreated all four intrinsic subsets and provided a sharper distinction between the limited and normal-like subsets, based on a strong expression of genes associated with lipid signaling and oxidative reduction found in the normal-like group that is absent in limited patients. Alignment of pathway-specific gene signatures showed enrichment in specific patient subsets. The strongest association was seen between genes activated by PDGF in fibroblasts and the fibroproliferative subset (Fig. 1). TGF-β showed a correlation with a subset of fibroproliferative patients and, interestingly, also showed a strong correlation with patients in the inflammatory subset (Fig. 1b). Consistent with the results of Sargent et al. [22], TGF-β was strongly correlated with disease severity [24].

Several different pathways showed enrichment in the inflammatory patient subset, including genes induced by sphingosine-1-phosphate (S1P), IL-4, tumor necrosis factor α (TNF-α), lipopolysaccharide (LPS), and polyinosinic/ polycytidylic acid (poly(I-C); Fig. 1). Interferon α (IFN-α) signaling was also elevated in this subset, although correlation with this pathway failed to reach statistical significance in the analysis. A common theme linking each of the immune activation pathways associated with the inflammatory subset is their convergence on NF-κB. Engagement of Toll-like receptors (TLRs) by their respective ligands initiates a cascade of signal transduction that culminates in activation of NF-κB, resulting in the production of acute phase cytokines, including TNF-α and IFN-α. TGF-β is also known to activate NF-κB via induction of TGF-β-associated kinase 1 (TAK1), suggesting a common theme linking these signals. Indeed, TLR signaling has been implicated directly in the

persistent activation of fibroblasts in SSc. In this model, tissue injury leads to the generation and release of damage-associated molecular patterns (DAMPs), which include matrix components (such as hyaluronic acid), nucleic acids, and immune complexes. Binding of DAMPs to TLRs signals fibroblasts to activate tissue repair processes and collagen production. Prolonged damage results in sustained fibroblast activation and unchecked fibrogenesis [25]. From these data and the literature, we have put forth hypotheses regarding the pathogenesis of SSc based on the assumption that the intrinsic gene expression subsets may be longitudinally connected (Fig. 2).

Potential innate immune triggers in SSc

In addition to mounting gene expression data, several additional lines of evidence point to a role for innate immune activation in the pathogenesis of SSc. Notably, aberrant responses to TLR stimulation have been noted in SSc patients [26, 27], and inflammasome activation has been implicated as a mediator of fibrosis in murine models of SSc [28]. Because these mechanisms also mediate immune defense against microbial invasion, it is possible that SSc is triggered by pathogenic encounter. In this regard, metagenomic analysis of skin biopsies from patients with early dSSc demonstrates significant overrepresentation of transcripts from the environmental fungus *Rhodotorula glutinis* compared with normal controls [29]. Notably, each of the patients identified in this study mapped to the inflammatory intrinsic gene expression subset, suggesting a TLR-mediated event may be responsible for the subsequent induction of NF-κB activation observed in this patient cohort. In further support of a potential role for a microbial trigger in SSc, Epstein-Barr viral nucleic acids and proteins have been detected in both lesional and non-lesional skin of dSSc patients, and EBV infection of SSc fibroblasts induced TLR activation [26].

Antigen-presenting cells as a driver of pathology in SSc

Evidence of macrophage activation in SSc skin [4, 23] and lung [6, 30], coupled with increased expression of genes associated with monocyte/macrophage recruitment and differentiation in SSc [4, 31], suggests a role for antigen-presenting cells (APCs) in disease pathogenesis. Because macrophages have significant plasticity and are subject to modulation by local microenvironmental factors, in vivo macrophage activation spans a broad spectrum of polarization states; classically activated (M1) and alternatively activated (M2) macrophages form the extreme ends of this activation spectrum [32]. Gene expression and molecular profiling studies suggest macrophages are skewed toward an M2 activation state in SSc, as demonstrated by enhanced expression of CCL18, CD163, and IL-10R [4, 6, 33]. As recent studies have shown, M2 polarization may be elicited with a variety of stimuli, including immune complexes, IL-4, IL-13, IL-10, and TGF-β [34]. Significantly, immune complexes have been detected in SSc patients [35, 36]; elevated expression of IL-4, IL-13, and IL-10 has been identified in SSc patient sera [37]; and numerous studies have implicated TGF-β in the pathogenesis of SSc [22, 38, 39]. Therefore, it is possible that these stimuli contribute to the phenotypic and functional activation state characteristic of SSc macrophages. M2 macrophages produce a number of pro-fibrotic factors including TGF-β, CCL2, and CCL18, which are known to play important roles in SSc pathogenesis [34]. These cells are also potent sources of TGF-β [40], which mediates T cell activation [41], and

IL-10 [40], which may further enhance M2 activation. As such, it is likely that these cells play a key role in the initiation of fibrosis in SSc and are responsible for the activation of inflammation and adaptive immune responses observed in SSc.

The importance of macrophages in the pathogenesis of SSc was demonstrated by a comparative gene expression profiling study that used the sclGVHD mouse as a model of human inflammatory intrinsic patient subset [23]. This work showed IL-13 pathway activation in both sclGVHD and the inflammatory intrinsic subset and identified CCL2 as a downstream target of the IL-13 signaling pathway. In addition to its pro-fibrotic properties, NF-κB-regulated CCL2 is a potent M2 macrophage activator and chemotactic factor. Significantly, blockade of CCL2 protected sclGVHD mice from clinical and pathological disease, suggesting that inhibition of macrophage recruitment and activation may be useful therapeutically.

In addition to macrophages, plasmacytoid dendritic cells (pDCs) have also been implicated as important mediators of SSc disease progression. As key regulators of innate and adaptive immunity, pDCs circulate in the peripheral blood, present antigen, and secrete copious type I interferons (mainly IFN-α and IFN-β) in response to stimulation [42]. Type I interferons modulate immune cell differentiation and proliferation as well as inflammatory cytokine production. Increased expression of type I interferon genes has been detected in the peripheral blood and sera of SSc patients [43, 44], and pDCs have been identified as the major source of type I interferons in these patients [44, 45]. Distinct from their role in interferon production, proteomic analysis of pDCs isolated from patients with SSc demonstrated that these cells secreted elevated levels of CXCL4 and other chemokines, which correlated directly with disease severity [46]. Furthermore, expression of CXCL4 was highest coincident with early disease and diminished over time, suggesting a role in the initiation of disease. Given the role of pDCs as the primary source of IFN-α, coupled with the strong IFN signaling responses seen in the skin and lung, these data implicate pDCs as early drivers of SSc pathogenesis.

Effects of APCs on adaptive immunity

Gene expression profiles of the skin and lung from SSc patients have shown increased expression of chemokines and chemokine receptor genes that are associated with recruitment of T helper type 2 (T_H2) cells [4, 47]. Consistent with the potent induction of M2 macrophages and pDCs early in disease, the earliest adaptive immune responses in SSc appear dominantly T_H2-skewed, driven by a combination of IL-4, IL-13, and TGF-β [34]. Indeed, these cytokines are present in the skin, sera, and bronchoalveolar fluid of SSc patients [22, 23, 37, 48, 49], as are both M2 macrophages [50, 51] and T_H2 cells [52], highlighting the importance of these factors in SSc pathogenesis. Intriguingly, the spontaneous rate of CD8 T cell apoptosis is elevated in SSc compared with controls [53], which may account for the enhanced CD4:CD8 ratio in SSc.

However, despite the strong T_H2-like immune response observed early in the disease, these signals diminish over time. The graft-versus-host disease model used to evaluate the effects of IL-13 on fibrosis examined mice at 2 and 5 weeks post-splenocyte transfer and revealed a

strong IL-13 signature at 2 weeks, which was attenuated at 5 weeks [23]. This suggests that initial activation of CCL2 and other fibrotic genes occurs through IL-13 signaling but is likely sustained through the activity of other signaling pathways and cell types. In this regard, mounting evidence suggests a role for T_H17 cells in the pathogenesis of SSc, with clear differences between diffuse and limited disease [54–58]. This progression is observed in many other autoimmune diseases, including multiple sclerosis, systemic lupus erythematosus, psoriasis, neuromyelitis optica, Crohn's disease, inflammatory bowel disease, and rheumatoid arthritis, all of which exhibit a strong T_H17-like bias [59–62].

Under normal conditions, type I IFNs are potent inhibitors of T_H17 activity [57]. However, the documented decrease in IFN-α-producing pDCs [46] and subsequent time-dependent attenuation in IFN-associated signaling [24] suggests a mechanism by which T cells may shift from a T_H2- to T_H17-dominant phenotype with increased disease progression. Such a shift is likely mediated by cytokines (especially IL-6) that are produced by myeloid and/or T_H2 cells. In this model, myeloid-derived TGF-β, in combination with IL-6, suppresses T_{reg} production and promotes differentiation of T_H17 cells [63]. The strong TGF-β and TNF-α gene expression signatures observed in the inflammatory subset, in conjunction with pervasive inflammatory infiltrates, are consistent with a T_H17-like immune response [64]. In support of this hypothesis, immunohistochemical analysis of SSc skin and gene expression profiling of PBMCs from SSc patients indicated expansion of circulating T_H17 cells and increased infiltration of IL-17^+ cells in SSc skin. Furthermore, increased frequencies of activated T_H17 cells have been noted in all SSc subtypes, and intracellular expression of TGF-β, which induces T_H17 differentiation, is specifically elevated in patients with late disease.

Cumulative data suggest a progressive model of SSc disease pathogenesis

Collectively, the analyses presented here suggest a progressive model of SSc pathogenesis (Fig. 2). Our experimental data have shown that a patient's intrinsic subset assignment is relatively stable in serial biopsies taken over the course of 6, 12, and sometimes even 24 months [2, 3], but our computational network analyses and meta-analyses of multiple patient cohorts have suggested that the intrinsic subsets are long-lived but likely mechanistically interconnected [4, 24]. Capturing patients transitioning between the inflammatory, fibroproliferative, or normal-like gene expression subsets in longitudinal studies is required to truly prove they are interconnected, and this has been hampered by the slow and variable disease progression, combined with the limited resolution one can obtain with serial biopsies from patients. A recent study by Assassi et al. [17] extends preliminary results shown by Pendergrass et al. [2] and confirms that patients in the normal-like subset have the longest disease duration, indicating it is likely the late, inactive form of the disease.

In the progressive model (Fig. 2), we propose that a disease trigger, in the presence of a permissive genetic background, initiates an innate immune response through the activation of TLRs, which culminates in NF-κB activation, marking the inflammatory subset as the likely initiation point for most SSc patients. We believe that the intensity and duration of patients' active immune responses likely have some bearing on their prognosis and overall outcomes. These early responses are likely mediated in part through pDCs and

macrophages, which induce the expression of T_H2-like cytokines, along with fibrotic mediators, such as TGF-β. This early T_H2 bias results in persistent M2 macrophage activation, which further exacerbates the fibrotic phenotype. Over time, the gradual decrease in pDC involvement results in a loss of IFN-α signaling, causing a transition to a more T_H17-like disease and the suppression of T_{reg} function. This persistent inflammation further perpetuates the chronic fibrosis phenotype driven by TGF-β, which then stimulates the production of pro-fibrotic PDGF. Eventual resolution of inflammation allows for downregulation of innate immune responses and a transition into the fibroproliferative subset. Continued proliferation is supported through differentiation of resident adipocytes into fibroblasts, resulting in a persistent replicative phenotype, in combination with a decrease in lipid signaling [65]. Finally, exhaustion of the adipocyte layer may result in the loss of the proliferative signature, ultimately resulting in transition to a more quiescent form of the disease, consistent with the normal-like subset of patients.

References

1. Milano A, et al. Molecular subsets in the gene expression signatures of scleroderma skin. PLoS One. 2008; 3:e2696. [PubMed: 18648520]
2. 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; 132:1363–1373. [PubMed: 22318389]
3. Hinchcliff M, et al. Molecular signatures in skin associated with clinical improvement during mycophenolate treatment in systemic sclerosis. J Invest Dermatol. 2013; 133:1979–1989. [PubMed: 23677167]
4. 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:e1004005. [PubMed: 25569146]
5. Hsu E, Shi H, Jordan RM, Lyons-Weiler J, Pilewski JM, Feghali-Bostwick CA. Lung tissues in patients with systemic sclerosis have gene expression patterns unique to pulmonary fibrosis and pulmonary hypertension. Arthritis Rheum. 2011; 63:783–794. [PubMed: 21360508]
6. Christmann RB, et al. Association of interferon- and transforming growth factor beta-regulated genes and macrophage activation with systemic sclerosis-related progressive lung fibrosis. Arthritis Rheumatol. 2014; 66:714–725. [PubMed: 24574232]
7. Whitfield ML, et al. Systemic and cell type-specific gene expression patterns in scleroderma skin. Proc Natl Acad Sci U S A. 2003; 100:12319–12324. [PubMed: 14530402]
8. Gardner H, et al. Gene profiling of scleroderma skin reveals robust signatures of disease that are imperfectly reflected in the transcript profiles of explanted fibroblasts. Arthritis Rheum. 2006; 54:1961–1973. [PubMed: 16736506]
9. Perou CM, et al. Molecular portraits of human breast tumours. Nature. 2000; 406:747–752. [PubMed: 10963602]
10. Alizadeh AA, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000; 403:503–511. [PubMed: 10676951]
11. Garber ME, et al. Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci U S A. 2001; 98:13784–13789. [PubMed: 11707590]
12. Bhattacharjee A, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A. 2001; 98:13790–13795. [PubMed: 11707567]
13. Sorlie T, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003; 100:8418–8423. [PubMed: 12829800]
14. Lafyatis R, et al. B cell depletion with rituximab in patients with diffuse cutaneous systemic sclerosis. Arthritis Rheum. 2009; 60:578–583. [PubMed: 19180481]

15. Chakravarty EF, et al. A pilot randomized placebo-controlled study of abatacept for the treatment of diffuse cutaneous systemic sclerosis. Arthritis Res Ther. 2015 In press.
16. Gordon JK, et al. Nilotinib in the treatment of early diffuse systemic sclerosis: an open-label, pilot clinical trial. Arthritis Res Ther. 2015 Submitted.
17. Assassi S, et al. Dissecting the heterogeneity of skin gene expression patterns in systemic sclerosis. Arthritis Rheum. 2015 In Press.
18. Wong AK, Park CY, Greene CS, Bongo LA, Guan Y, Troyanskaya OG. IMP: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks. Nucleic Acids Res. 2012; 40:W484–W490. [PubMed: 22684505]
19. Assassi S, Radstake TR, Mayes MD, Martin J. Genetics of scleroderma: implications for personalized medicine? BMC Med. 2013; 11:9. [PubMed: 23311619]
20. Taroni J, et al. Genome-wide gene expression analysis of systemic sclerosis esophageal biopsies identifies disease-specific molecular subsets. Arthritis Res Ther. 2015 In press.
21. Chung L, et al. Molecular framework for response to imatinib mesylate in systemic sclerosis. Arthritis Rheum. 2009; 60:584–591. [PubMed: 19180499]
22. Sargent JL, et al. A TGFbeta-responsive gene signature is associated with a subset of diffuse scleroderma with increased disease severity. J Invest Dermatol. 2009; 130:694–705. [PubMed: 19812599]
23. Greenblatt MB, et al. Interspecies comparison of human and murine scleroderma reveals IL-13 and CCL2 as disease subset-specific targets. Am J Pathol. 2012; 180:1080–1094. [PubMed: 22245215]
24. Johnson ME, et al. Experimentally-derived fibroblast gene signatures identify molecular pathways associated with distinct subsets of systemic sclerosis patients in three independent cohorts. PLoS One. 2015; 10:e0114017. [PubMed: 25607805]
25. Bhattacharyya S, Wei J, Tourtellotte WG, Hinchcliff M, Gottardi CG, Varga J. Fibrosis in systemic sclerosis: common and unique pathobiology. Fibrogenesis Tissue Repair. 2012; 5:S18. [PubMed: 23259815]
26. Farina A, et al. Epstein-Barr virus infection induces aberrant TLR activation pathway and fibroblast-myofibroblast conversion in scleroderma. J Invest Dermatol. 2014; 134:954–964. [PubMed: 24129067]
27. van Bon L, et al. Distinct evolution of TLR-mediated dendritic cell cytokine secretion in patients with limited and diffuse cutaneous systemic sclerosis. Ann Rheum Dis. 2010; 69:1539–1547. [PubMed: 20498209]
28. Artlett CM, Sassi-Gaha S, Rieger JL, Boesteanu AC, Feghali-Bostwick CA, Katsikis PD. The inflammasome activating caspase 1 mediates fibrosis and myofibroblast differentiation in systemic sclerosis. Arthritis Rheum. 2011; 63:3563–3574. [PubMed: 21792841]
29. Arron ST, et al. High Rhodotorula sequences in skin transcriptome of patients with diffuse systemic sclerosis. J Invest Dermatol. 2014; 134:2138–2145. [PubMed: 24608988]
30. Christmann RB, et al. Interferon and alternative activation of monocyte/macrophages in systemic sclerosis-associated pulmonary arterial hypertension. Arthritis Rheum. 2011; 63:1718–1728. [PubMed: 21425123]
31. Mathes AL, et al. Global chemokine expression in systemic sclerosis (SSc): CCL19 expression correlates with vascular inflammation in SSc skin. Ann Rheum Dis. 2014; 73:1864–1872. [PubMed: 23873879]
32. Mosser DM, Edwards JP. Exploring the full spectrum of macrophage activation. Nat Rev Immunol. 2008; 8:958–969. [PubMed: 19029990]
33. Higashi-Kuwata N, et al. Characterization of monocyte/ macrophage subsets in the skin and peripheral blood derived from patients with systemic sclerosis. Arthritis Res Ther. 2010; 12:R128. [PubMed: 20602758]
34. Atamas SP, White B. Cytokine regulation of pulmonary fibrosis in scleroderma. Cytokine Growth Factor Rev. 2003; 14:537–550. [PubMed: 14563355]
35. Chen ZY, et al. Immune complexes and antinuclear, antinucleolar, and anticentromere antibodies in scleroderma. J Am Acad Dermatol. 1984; 11:461–467. [PubMed: 6237134]

36. Seibold JR, Medsger TA Jr, Winkelstein A, Kelly RH, Rodnan GP. Immune complexes in progressive systemic sclerosis (scleroderma). Arthritis Rheum. 1982; 25:1167–1173. [PubMed: 6753851]
37. Hasegawa M, Fujimoto M, Kikuchi K, Takehara K. Elevated serum levels of interleukin 4 (IL-4), IL-10, and IL-13 in patients with systemic sclerosis. J Rheumatol. 1997; 24:328–332. [PubMed: 9034992]
38. Leask A, Abraham DJ. TGF-beta signaling and the fibrotic response. Faseb J. 2004; 18:816–827. [PubMed: 15117886]
39. Varga J, Pasche B. Transforming growth factor beta as a therapeutic target in systemic sclerosis. Nat Rev Rheumatol. 2009; 5:200–206. [PubMed: 19337284]
40. Gordon S. Alternative activation of macrophages. Nat Rev Immunol. 2003; 3:23–35. [PubMed: 12511873]
41. Li MO, Sanjabi S, Flavell RA. Transforming growth factor-beta controls development, homeostasis, and tolerance of T cells by regulatory T cell-dependent and -independent mechanisms. Immunity. 2006; 25:455–471. [PubMed: 16973386]
42. Reizis B, Bunin A, Ghosh HS, Lewis KL, Sisirak V. Plasmacytoid dendritic cells: recent progress and open questions. Annu Rev Immunol. 2011; 29:163–183. [PubMed: 21219184]
43. Assassi S, et al. Systemic sclerosis and lupus: points in an interferon-mediated continuum. Arthritis Rheum. 2010; 62:589–598. [PubMed: 20112391]
44. Eloranta ML, et al. Type I interferon system activation and association with disease manifestations in systemic sclerosis. Ann Rheum Dis. 2010; 69:1396–1402. [PubMed: 20472592]
45. Kim D, et al. Induction of interferon-alpha by scleroderma sera containing autoantibodies to topoisomerase I: association of higher interferon-alpha activity with lung fibrosis. Arthritis Rheum. 2008; 58:2163–2173. [PubMed: 18576347]
46. van Bon L, et al. Proteome-wide analysis and CXCL4 as a biomarker in systemic sclerosis. NEJM. 2014; 370:433–443. [PubMed: 24350901]
47. Luzina IG, Atamas SP, Wise R, Wigley FM, Xiao HQ, White B. Gene expression in bronchoalveolar lavage cells from scleroderma patients. Am J Respir Cell Mol Biol. 2002; 26:549–557. [PubMed: 11970906]
48. Falanga V, Gerhardt CO, Dasch JR, Takehara K, Ksander GA. Skin distribution and differential expression of transforming growth factor β1 and β2. J Dermatol Sci. 1992; 3:131–136. [PubMed: 1498091]
49. Sfikakis PP, McCune BK, Tsokos M, Aroni K, Vayiopoulos G, Tsokos GC. Immunohistological demonstration of transforming growth factor-β isoforms in the skin of patients with systemic sclerosis. Clin Immunol Immunopathol. 1993; 69:199–204. [PubMed: 8403557]
50. Higashi-Kuwata N, Makino T, Inoue Y, Takeya M, Ihn H. Alternatively activated macrophages (M2 macrophages) in the skin of patient with localized scleroderma. Exp Dermatol. 2009; 18:727–729. [PubMed: 19320738]
51. Pechkovsky DV, et al. Alternatively activated alveolar macrophages in pulmonary fibrosis—mediator production and intracellular signal transduction. Clin Immunol. 2010; 137:89–101. [PubMed: 20674506]
52. Truchetet M-E, Brembilla NC, Montanari E, Allanore Y, Chizzolini C. Increased frequency of circulating Th22 in addition to Th17 and Th2 lymphocytes in systemic sclerosis: association with interstitial lung disease. Arth Res Ther. 2011; 13:R166. [PubMed: 21996293]
53. Kessel A, et al. Increased CD8+ T cell apoptosis in scleroderma is associated with low levels of NF-kappa B. J Clin Immunol. 2004; 24:30–36. [PubMed: 14997031]
54. Mathian A, et al. Activated and resting regulatory T cell exhaustion concurs with high levels of interleukin-22 expression in systemic sclerosis lesions. Ann Rheum Dis. 2012; 71:1227–1234. [PubMed: 22696687]
55. Murata M, et al. Clinical association of serum interleukin-17 levels in systemic sclerosis: is systemic sclerosis a Th17 disease? J Dermatol Sci. 2008; 50:240–242. [PubMed: 18329249]
56. Radstake TR, et al. The pronounced Th17 profile in systemic sclerosis (SSc) together with intracellular expression of TGFβ and IFNγ distinguishes SSc phenotypes. PLoS One. 2009; 4:e5903. [PubMed: 19536281]

57. Rodríguez-Reyna TS, et al. Th17 peripheral cells are increased in diffuse cutaneous systemic sclerosis compared with limited illness: a cross-sectional study. Rheumatol Int. 2012; 32:2653–2660. [PubMed: 21789610]

58. Deleuran B, Abraham DJ. Possible implication of the effector CD4+ T-cell subpopulation TH17 in the pathogenesis of systemic scleroderma. Nat Clin Pract Rheum. 2007; 3:682–683.

59. Axtell RC, Raman C. Janus-like effects of type I interferon in autoimmune diseases. Immunol Rev. 2012; 248:23–35. [PubMed: 22725952]

60. Brand S. Crohn's disease: Th1, Th17 or both? The change of a paradigm: new immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn's disease. Gut. 2009; 58:1152–1167. [PubMed: 19592695]

61. Camporeale A. IL-6, IL-17, and STAT3: a holy trinity in auto-immunity? Front Biosci. 2012; 17:2306–2326.

62. Kimura A, Kishimoto T. Th17 cells in inflammation. Int Immunopharmacol. 2011; 11:319–322. [PubMed: 21035432]

63. Zheng SG. Regulatory T cells versus Th17: differentiation of Th17 versus Treg, are they mutually exclusive? IL-17, IL-22 and their producing cells: role in inflammation and autoimmunity: springer. 2013:91–107.

64. Steinman L. A brief history of TH17, the first major revision in the TH1/TH2 hypothesis of T cell-mediated tissue damage. Nat Med. 2007; 13:139–145. [PubMed: 17290272]

65. Marangoni RG, et al. Myofibroblasts in murine cutaneous fibrosis originate from adiponectin-positive intradermal progenitors. Arthritis Rheumatol. 2015; 67:1062–1073. [PubMed: 25504959]

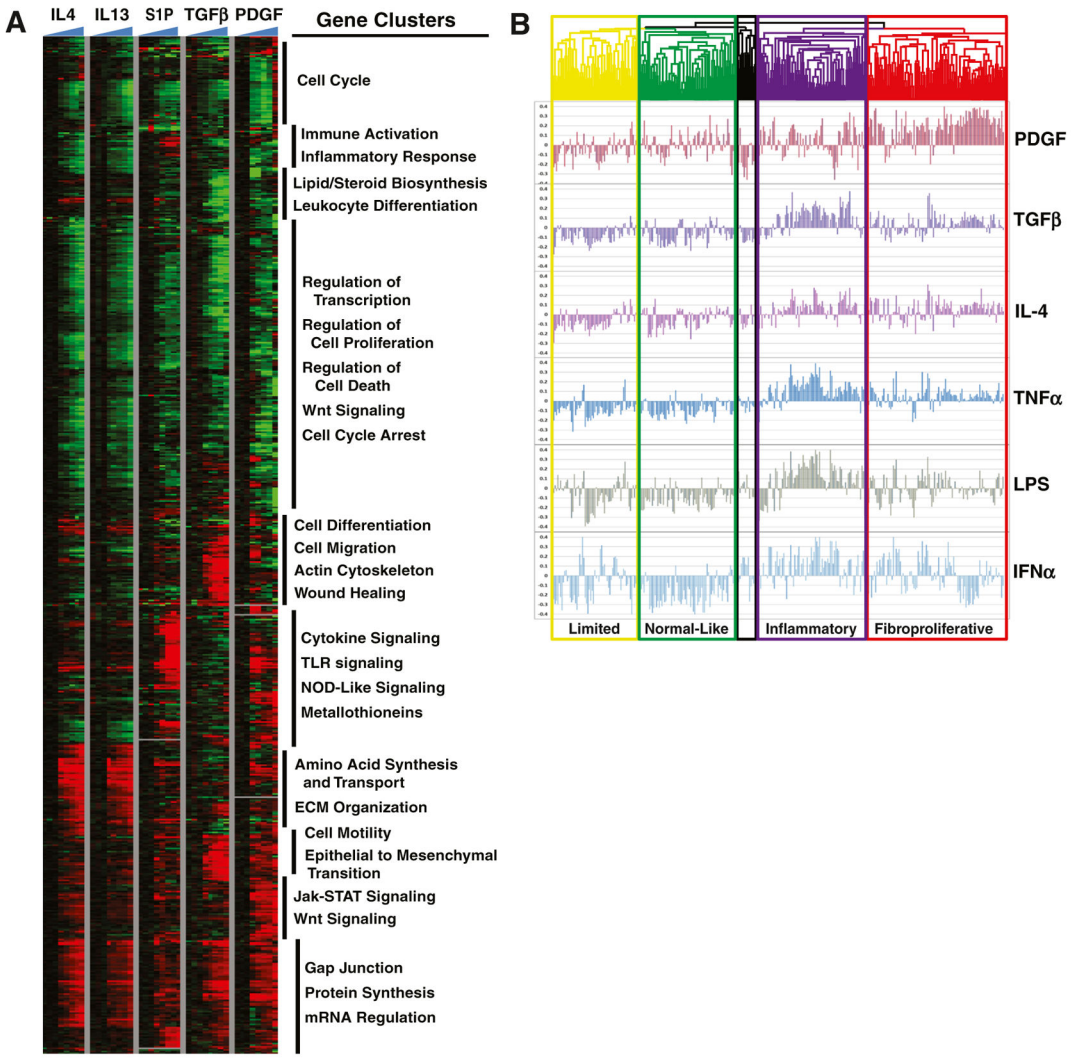


Fig. 1.

Pathway activation signatures show differential expression across the intrinsic gene expression subsets. **a** Normal and SSc dermal fibroblasts were treated with different pro-fibrotic and immune mediators that have been implicated in SSc. A subset of pathways and fibroblast time courses is shown. **b** Hierarchical clustering was performed on 329 microarray hybridizations from 287 unique biopsies representing 111 patients: 70 dSSc, 10 lSSc, 26 healthy controls, 4 morphea, and 1 eosinophilic fasciitis from three independent data sets [1–3], as published in [24]. The array tree is color coded to indicate intrinsic subset designations (*yellow* = limited, *green* = normal-like, *purple* = inflammatory, *red* = fibroproliferative, and *black* = unassigned). Pearson correlation coefficients were calculated between each pathway and a sample and plotted. Adapted from Johnson et al. PLoS One (2015) with permission [24]

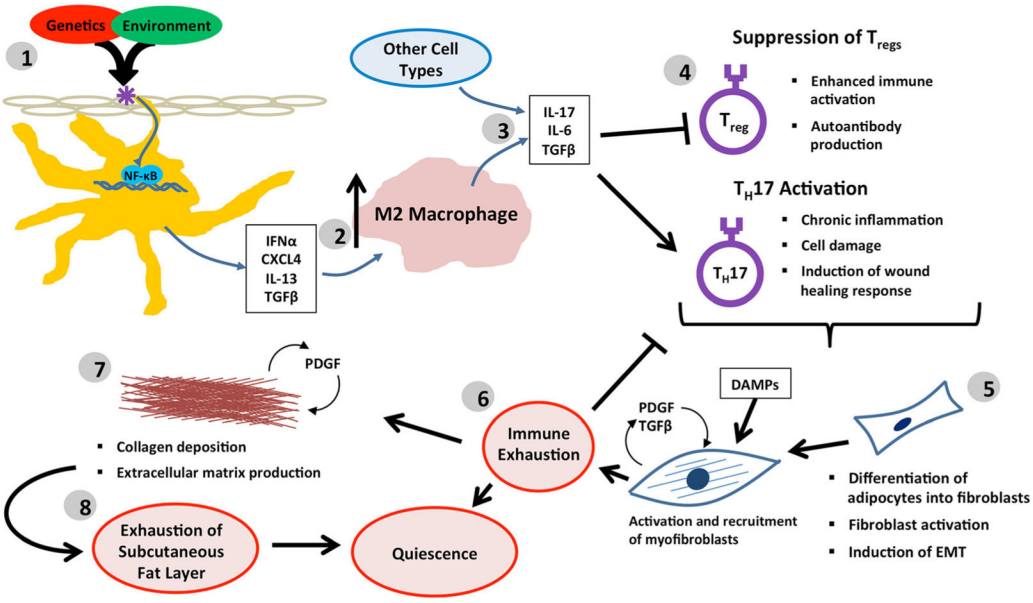


Fig. 2.

Progressive model of SSc pathogenesis. *1* In this model, SSc pathogenesis is initiated by a disease trigger in a permissive genetic background, resulting in an innate immune response signaling through NF-κB. These early responses may be mediated in part through pDCs and macrophages, which induce the expression of T_H2-like cytokines, along with fibrotic mediators, such as TGF-β. *2* This early T_H2 bias results in persistent M2 macrophage activation, which further exacerbates the fibrotic phenotype. *3* Over time, the gradual decrease in pDC involvement results in a loss of IFN-α signaling, *4* resulting in transition to a more T_H17-like disease and the suppression of T_{reg} function. *5* This persistent inflammation further perpetuates the chronic fibrosis phenotype driven by TGF-β, which then stimulates production of pro-fibrotic PDGF. *6* Eventual resolution of inflammation allows for downregulation of innate immune responses. *7* Continued proliferation is supported through differentiation of resident adipocytes into fibroblasts, resulting in a persistent replicative phenotype, in combination with a decrease in lipid signaling. *8* Exhaustion of the adipocyte layer results in the loss of proliferating cells, ultimately resulting in transition to a more quiescent form of the disease

Arthritis Research & Therapy

RESEARCH ARTICLE Open Access

Microbiome dysbiosis is associated with disease duration and increased inflammatory gene expression in systemic sclerosis skin

Michael E. Johnson[1], Jennifer M. Franks[1,2], Guoshuai Cai[1,8], Bhaven K. Mehta[1], Tammara A. Wood[1], Kimberly Archambault[1], Patricia A. Pioli[3], Robert W. Simms[7], Nicole Orzechowski[5], Sarah Arron[6] and Michael L. Whitfield[1,2,4*]

Abstract

Background: Infectious agents have long been postulated to be disease triggers for systemic sclerosis (SSc), but a definitive link has not been found. Metagenomic analyses of high-throughput data allows for the unbiased identification of potential microbiome pathogens in skin biopsies of SSc patients and allows insight into the relationship with host gene expression.

Methods: We examined skin biopsies from a diverse cohort of 23 SSc patients (including lesional forearm and non-lesional back samples) by RNA-seq. Metagenomic filtering and annotation was performed using the Integrated Metagenomic Sequencing Analysis (IMSA). Associations between microbiome composition and gene expression were analyzed using single-sample gene set enrichment analysis (ssGSEA).

Results: We find the skin of SSc patients exhibits substantial changes in microbial composition relative to controls, characterized by sharp decreases in lipophilic taxa, such as *Propionibacterium*, combined with increases in a wide range of gram-negative taxa, including *Burkholderia*, *Citrobacter*, and *Vibrio*.

Conclusions: Microbiome dysbiosis is associated with disease duration and increased inflammatory gene expression. These data provide a comprehensive portrait of the SSc skin microbiome and its association with local gene expression, which mirrors the molecular changes in lesional skin.

Keywords: Microbiome, Systemic sclerosis, Scleroderma, Metagenomics, RNA-sequencing

Introduction

Systemic sclerosis (SSc) is a progressive autoimmune disease that results in inflammation, fibrosis, and dysfunction of multiple organ systems including the skin, lungs, gastrointestinal tract, and blood vessels. Recent advances have provided significant insight into the molecular and immunologic changes characteristic of SSc patients [1–5]; however, the underlying mechanisms that initiate and perpetuate disease pathologies remain poorly understood.

Evidence for dysbiosis as a source of disease pathology is well-documented in inflammatory skin conditions, such as psoriasis, where patients exhibit significant increases in both *Propionibacterium* and *Staphylococcus* on lesional skin [6]. In atopic dermatitis (AD), patients exhibit temporal shifts in skin microbiome composition, with microbiome diversity decreasing during disease flares, characterized by significant increases in *Staphylococcus* levels, followed by increased diversity thereafter [7]. These patterns of dysbiosis suggest a mechanism by which relative changes in the abundance of specific taxa directly influence disease pathology [7].

* Correspondence: Michael.L.Whitfield@Dartmouth.edu
[1]Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Hanover, NH, USA
[2]Program in Quantitative Biomedical Sciences, Geisel School of Medicine at Dartmouth, Hanover, NH, USA
Full list of author information is available at the end of the article

A wide array of potential etiologic agents have been proposed for SSc, including viruses, bacteria, and fungi. Viruses such as cytomegalovirus (CMV), parvovirus B19, Epstein-Barr virus (EBV), and endogenous retroviruses have all been postulated as potential triggers of SSc [8–10]. EBV transcripts have been reported in lesional skin of SSc patients [11]. Among bacteria, *Helicobacter pylori* has been implicated in the etiology and progression of numerous autoimmune diseases, though its role in SSc remains controversial with studies both confirming and refuting such an association [10, 12, 13]. The most recent addition to the list of potential etiologic agents is the fungus *Rhodotorula glutinis*, a ubiquitous environmental contaminant and occasional skin commensal, which was found to be strongly associated with lesional skin of early-stage, untreated diffuse SSc patients [14].

Here, we examined skin biopsies from a diverse cohort of SSc patients and healthy controls by RNA-sequencing (RNA-seq) to obtain an unbiased assessment of the SSc microbiome and its relationship with patient gene expression. We show a reproducible shift in microbiome composition characterized by decreases in lipophilic taxa, along with increases in a variety of gram-negative bacteria that mirror local changes in inflammatory gene expression. These changes are closely tied to underlying gene expression associated with lipid signaling and immune activation. Genus-level taxonomic changes were associated with disease duration and the inflammatory intrinsic gene expression subset. Together, these data demonstrate that the skin microbiome composition in SSc mirrors molecular pathogenesis.

Methods

Patient selection

Study participants provided written, informed consent prior to sample collection in accordance with the Declaration of Helsinki Protocol and the Institutional Review Boards of Boston University Medical Center, Boston, MA, Dartmouth-Hitchcock Medical Center, Lebanon, NH, and the Hospital for Special Surgery, New York, NY. All patients met the American College of Rheumatology classification criteria for SSc [15], with further classification as either diffuse [16] (dSSc) or limited [17] (lSSc) disease. SSc patients with disease duration < 2.5 years were classified as "early stage" and patients with disease duration > 8 years were classified as "late stage" for this analysis.

Biopsy processing and RNA-seq

Lesional forearm and, for a subset of patients, non-lesional back skin was collected by punch biopsy (4 mm) from 15 SSc patients and 6 healthy volunteers. An additional 8 baseline samples collected as part of a Nilotinib clinical trial [18] were also included in this analysis. Following collection, samples were immediately placed in RNA*Later* (Life Technologies, Carlsbad, CA) at 4 °C overnight, followed by − 80 °C until needed. Tissue homogenization was performed using the Qiagen TissueLyser II (Qiagen, Gaithersburg, MD). RNA extraction was performed using the Qiagen RNeasy Fibrous Tissue Mini Kit run on the QIAcube (Qiagen). RNA concentration and RNA integrity were assessed using the Agilent 4200 TapeStation (Agilent, Santa Clara, CA). RNA-seq libraries were generated from 100 ng total RNA prepared using the Illumina TruSeq Stranded Total RNA Library Prep Kit with Ribo-Gold rRNA depletion (Illumina, San Diego, CA). Samples were then multiplexed and sequenced on an Illumina NextSeq 500 sequencer, producing an average of 80–100 million 75-bp paired-end reads per sample.

Human gene expression analyses

Raw sequencing reads were aligned to the human genome (hg19) using STAR aligner [19] and expressed as fragments per million mapped reads (FPKM). Designation of intrinsic molecular subsets for SSc patients was performed using a gene-specific normalization method to render RNA-seq values distributions similar to microarray so that supervised machine learning algorithms can be applied regardless of the platform used to generate data, as described [20]. Normalized RNA-seq data were classified using a support vector machine trained using a merged and curated dataset composed of samples from GSE9285, GSE32413, and GSE45485. To visualize results, the probe ID list from Johnson et al. [4] was collapsed on gene ID. This gene list was compared against normalized FPKM values for all 36 RNA-seq samples, resulting in a total of 1010 overlapping genes; a full list of all genes and normalized expression values is shown in Additional file 1: Table S1. Data were then hierarchically clustered using Cluster 3.0 [21] and visualized using Java TreeView [22].

Metagenomic filtering and microbiome annotation

Metagenomic filtering and microbiome annotation was run using the Integrated Metagenomic Sequencing Analysis (IMSA) software package [23] and compared against the National Center for Biotechnology Information (NCBI) non-redundant nucleotide database (minimum significance = 1×10^{-15}), followed by a secondary BLAST alignment against the NCBI viral genome repository (minimum significance = 1×10^{-5}). To limit inclusion of spurious hits, sample annotation was limited to sequences mapping to five or fewer species, with ties split equally across species. Outputs were then filtered based on taxonomy to include only archaea, bacteria, fungi, and viruses. Normalization of taxonomic outputs was performed by rounding down to the nearest integer and rarefying to the level of the depth of the lowest sample using the Quantitative Insights Into Molecular

Ecology (QIIME) platform [24]. Batch effects associated with library preparation were removed by median centering across taxa. Statistical analyses were performed using Statistical Package for the Social Sciences (SPSS) software (IBM, version 23); additional analyses, including corrections for multiple hypothesis testing using the method of Benjamini & Hochberg [25], were performed in R.

Pathway activation and microbiome abundance

Single-sample gene set enrichment analysis (ssGSEA) [26] was run as a module in GenePattern, using relevant KEGG pathways as the query gene sets. A correlation matrix was then generated by calculating Pearson's correlations for all combinations of ssGSEA values and genus-level abundance across all patients. Data were then hierarchically clustered using Cluster 3.0 and visualized using Java TreeView.

Results

Patient characteristics

Lesional forearm skin biopsies were collected from 23 SSc patients; seven patients also provided biopsies of non-lesional back skin. Forearm skin biopsies were also obtained from 6 age- and gender-matched healthy controls. Samples were collected from three independent clinical centers and included both clinically limited (lSSc) and diffuse (dSSc) disease, with disease duration ranging from 0 to 35 years. The patient population consisted primarily of early-stage patients (disease duration ≤ 2 years), though a handful of very late-stage patients (disease duration > 10 years) were also included to assess microbiome changes over time. Clinical information on these patients is summarized in Table 1; a full breakdown of patient clinical information is presented in Additional file 2: Table S2. Assessments of skin involvement were determined based on overall modified Rodnan skin score (mRSS), as local scores were not available for all patients. No significant differences in age, sex, or race were evident between SSc and controls ($p > 0.05$ for all).

Table 1 Summary clinical information

	Control subjects ($N = 6$)	SSc patients ($N = 23$)
Age, median (range) years	53 (25–67)	53 (27–77)
Sex, *N* (%) female	4 (67%)	19 (83%)
Race, *N* (%) Caucasian	5 (83%)	20 (87%)
SSc subtype, *N* (%) diffuse	NA	15 (65%)
MRSS, median (range)	NA	16 (0–44)
Disease duration from first non-Raynaud's, median (range) years	NA	1.0 (0–35)
ILD/PAH, *N* (%)	NA	8 (35%)
ANA primary pattern, *N* (%) patients		
Homogenous	NA	1 (4%)
Nucleolar	NA	5 (22%)
Speckled	NA	6 (26%)
Centromere	NA	2 (9%)
SSc-specific antibodies, *N* (%)		
Anti-centromere	NA	3 (13%)
Scl-70	NA	3 (13%)
RNA polymerase III	NA	5 (22%)
Current therapies, *N* (%)	NA	17 (74%)
Prior therapies, *N* (%)		
Amlodipine	NA	4 (17%)
Methotrexate	NA	4 (17%)
Prednisone	NA	3 (13%)

Abbreviations: SSc, systemic sclerosis; ANA, anti-nuclear antibodies; MRSS, modified Rodnan skin score; ILD, interstitial lung disease; PAH, pulmonary arterial hypertension; NA, not applicable
Current and prior therapies include all treatments observed in three or more patients

Sequencing and annotation

RNA-seq was performed on 36 skin biopsies, from 29 unique patients, resulting in an average of 83 million reads per sample (range 51,278,817–112,643,430). Raw sequencing reads were aligned to the human genome (hg19) using STAR aligner [19], and the expression level of each gene was expressed as fragments per million mapped reads (FPKM). Intrinsic gene expression subset designations were determined based on support vector machine classification using normalized FPKM values [20]. Hierarchical clustering using the gene list from Johnson et al. [4], resulting in a total of 1010 overlapping genes, revealed distinct molecular subsets of disease, characterized by strong immune activation, lipid signaling, and proliferation signals, consistent with previous publications [1–3] (Fig. 1a; Additional file 1: Table S1). Together, these data suggest our patient cohort is representative of the four major intrinsic gene expression subsets of SSc. Additionally, we find that forearm and back samples largely tend to cluster together, consistent with previous analyses (Fig. 1b) [27].

Filtering of human sequence reads and microbiome annotation was performed using Integrated Metagenomic Sequence Analysis (IMSA) [23], yielding an average of 18,794 informative hits, defined as sequences mapping to five or fewer species, per skin biopsy (range 3098–74,429) across 1870 genera. To adjust for library-specific effects, all data were rarefied to the level of the lowest sample, followed by median centering of each genus by library preparation batch. This approach substantially reduced batch effects associated with library preparation, enabling direct comparisons of sample outputs across patients.

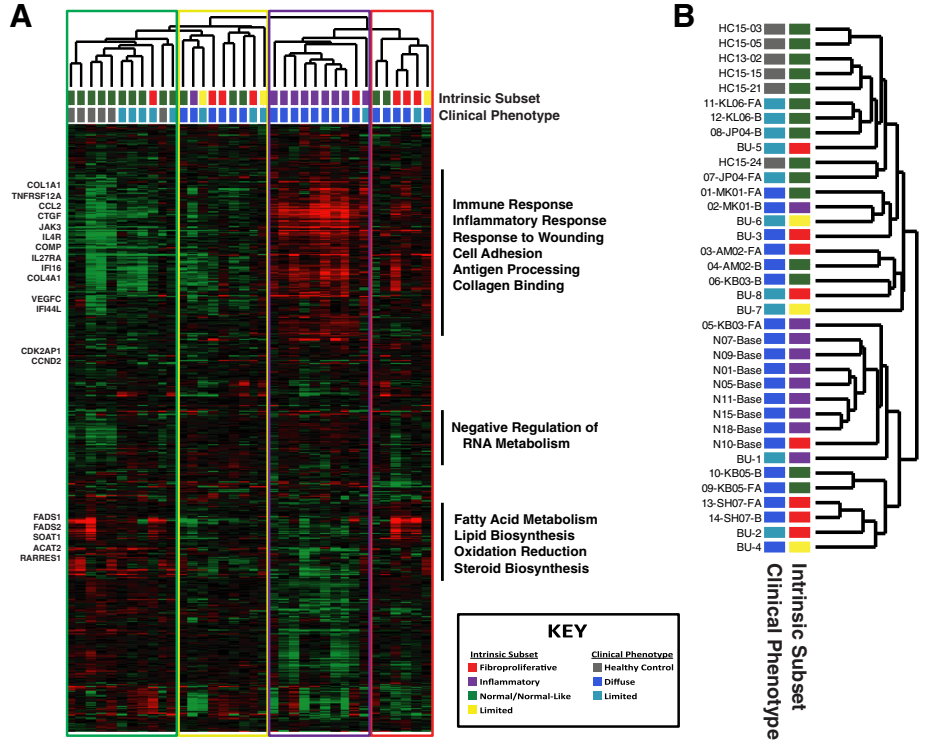


Fig. 1 Intrinsic subset analysis of RNA-seq reads from SSc skin. **a** Assignment of intrinsic molecular subsets for SSc patients was performed using a support vector machine (SVM) developed for the purpose (Franks et al. In Press). Displayed are the 1010 genes from Johnson et al. [4] collapsed on gene ID and extracted from the normalized FPKM values for all 36 RNA-seq samples. Hierarchical clustering revealed distinct molecular subsets of disease, consistent with previous publications [1–3]. The sample dendrogram is colored to indicate intrinsic subset designations: normal-like (green), limited (yellow), inflammatory (purple), proliferative (red). **b** Hash marks indicate SSc clinical diagnosis associated with each sample. Black bars indicate genes that clustered together hierarchically; the most significantly overrepresented GO terms are listed

Antimicrobial gene expression is suppressed in SSc lesional skin

Antimicrobial peptides (AMPs), including cathelicidin (CAMP/LL-37), α-defensins, and β-defensins, are an essential component of epithelial barrier defenses. To assess the role of AMPs in SSc, we compared gene expression levels between SSc and controls, as well as between lesional forearm and non-lesional back skin. Among the major AMPs, dermcidin (DCD) is highly expressed across samples, regardless of disease type, while other major AMPs, including cathecidin (CAMP) and the α-defensins, were virtually undetected, with no difference in expression between SSc and controls. In contrast, β-defensin 1 (DEFB1), an AMP produced by epithelial cells, is expressed across all samples; however, these levels are significantly lower is SSc lesional skin compared to healthy controls ($p < 0.001$ by unpaired *t*-test), as well as in lesional forearm compared to non-lesional back skin ($p = 0.007$ by paired *t*-test) (Additional file 3: Table S3). Similar results were also seen between SSC lesional skin and healthy controls in a previous SSc skin RNA-seq dataset (Additional file 3: Table S3), suggesting a potential mechanism underlying microbiome differences in SSc patients.

Microbiome genus-level differences are correlated with SSc clinical phenotypes

SSc patients exhibited large changes in microbiome composition relative to controls, characterized by decreases in lipophilic taxa, such as *Propionibacterium* and *Staphylococcus*, combined with increases in a wide range of Gram-negative bacteria, including *Burkholderia*, *Citrobacter*, and *Vibrio* ($p < 0.05$ for all; Fig. 2a; Additional file 4: Table S4). These differences were not associated with clinical subtype, with limited and diffuse disease exhibiting broadly similar abundances of major taxa (Fig. 2a). Decreases were also observed in the fungus *Malassezia* relative to controls, with the greatest decrease occurring in dSSc patients.

Associations between disease duration and genus-level abundance were also evident, with significant ($p < 0.05$) or near-significant ($p < 0.10$) differences in 6 of the top 21 genera, including *Propionibacterium*, *Salmonella*, and *Enterobacter* (Fig. 2b; Additional file 4: Table S4). Relative decreases in *Propionibacterium* were evident for both early- and late-stage patients, relative to controls. We observed differential directions for the relative abundance of *Salmonella* including significant increases for

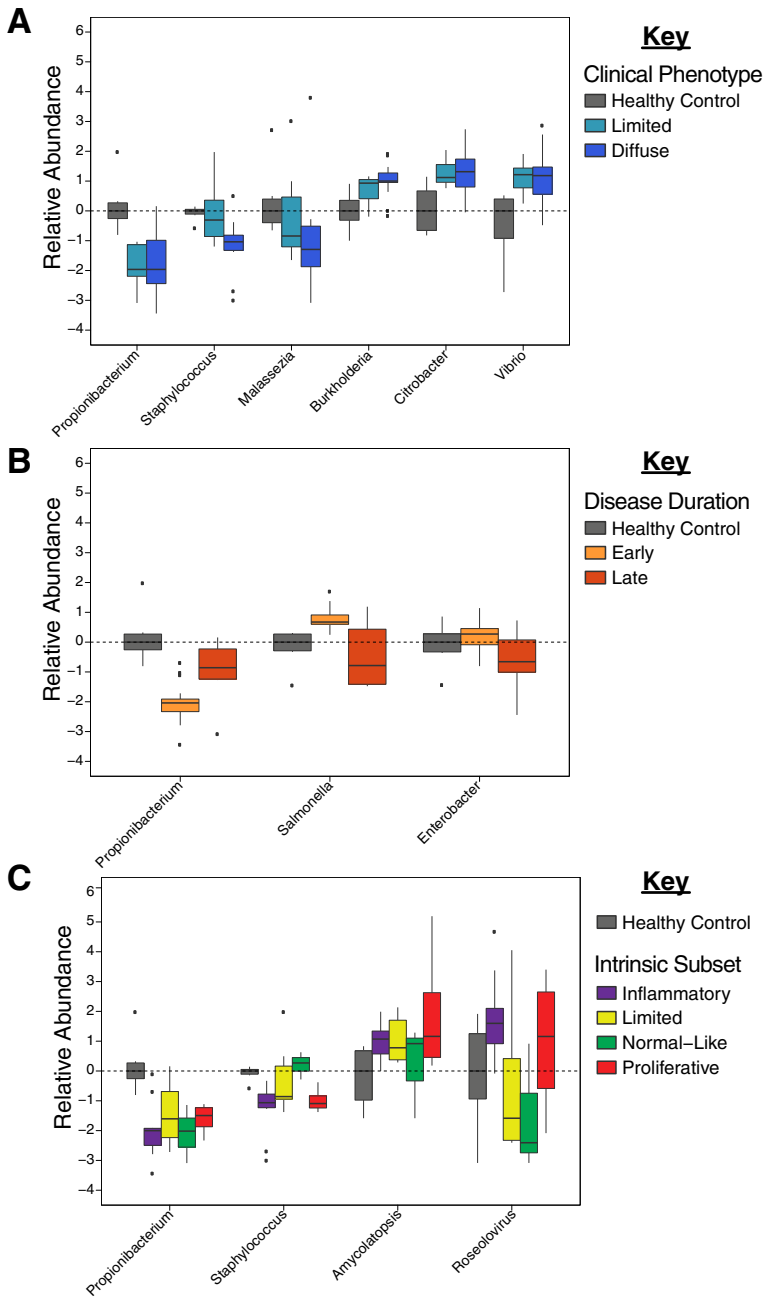


Fig. 2 Differential abundance of major skin taxa. SSc lesional skin exhibits significant changes in microbiome composition, relative to controls. Differential abundance of select genera, relative to controls, based on **a** clinical subtype, **b** disease duration (early, < 5 years; late, > 5 years), and **c** intrinsic molecular subset [1]

early-stage patients and reduced abundance in late-stage patients. Comparisons between the four intrinsic molecular subsets of disease revealed modest differences associated with the normal-like and inflammatory subsets, with normal-like patients broadly mimicking differences seen between SSc and controls, while the inflammatory group was characterized by decreased *Staphylococcus* and increased *Roseolovirus*, relative to other subsets (Fig. 2c). The absence of more acute genus-level distinctions between subsets is likely the result of high levels of some genera in both inflammatory and proliferative patients (Fig. 2c), thereby limiting the diagnostic value of any single genus. Other clinical cofactors, including sex and autoantibody status, were not statistically different between groups. A full list of comparisons for each clinical cofactor is shown in Additional file 4: Table S4.

Core microbiome by patient is predictive of clinical involvement

To identify changes in microbiome composition associated with clinical covariates, we calculated the number of taxa that accounted for 90% of the annotated reads across our entire dataset, which we collectively refer to as the SSc skin core microbiome. The SSc skin core microbiome was composed of 103 genera and included representatives from bacteria, fungi, and viruses. Organisms not included in the core microbiome were exclusively low abundance taxa found in only a small number of samples. Hierarchical clustering of the SSc skin core microbiome revealed patterns of microbial abundance closely mimicking that seen within an individual, characterized by clear differences between SSc and controls (Fig. 3a). Organisms of the SSc skin microbiome formed distinct branches within the dendrogram. Lipophilic commensals (*Malassezia*, *Propionibacterium*, and *Cutibacterium*) were the predominant genera in normal-like patients, Gram-negative bacteria (*Veillonella*, *Prevotella*, *Neisseria*, and *Actinomycetes*) were abundant in the limited and proliferative subsets, and viruses (*Roseolovirus* and *Cyprinivirus*) were highest in inflammatory patients (Fig. 3a). These patterns are consistent with the various environmental niches associated with each class of organisms and are suggestive of changes in skin morphology and immune activation associated with each subset.

SSc skin microbiome profiles were analyzed using principal component analysis (PCA) to identify the broad, population-based changes associated with clinical covariates. Lesional forearm and non-lesional back skin were not significantly different among SSc patients ($p = 0.097$; Fig. 3b; Additional file 5: Figure S1). Similarly, no significant differences were evident based on SSc clinical subtype ($p = 0.156$; Fig. 3c) or mRSS at the time of biopsy (Additional file 6: Figure S2). In contrast, microbiome profiles were strongly correlated with intrinsic subset, with the strongest differences seen in normal-like and inflammatory patients, indicative of a link between disease activity of microbial abundance ($p = 0.014$; Fig. 3a, d).

Microbiome composition is correlated with inflammatory pathway activation in SSc skin biopsies

Given the close association seen between clinical subtype and microbiome composition, we next sought to identify relationships between relevant molecular pathways and taxonomic abundance using single-sample gene set enrichment analysis (ssGSEA). ssGSEA analysis generates a single value quantifying the extent to which a given gene set is coordinately up- or downregulated in a sample. This analysis was repeated for all available KEGG pathways, generating a table of pathway activation scores for each patient sample (Additional file 7: Table S5). Using this data, we then used Pearson's correlations to compare each of these individual pathways against all genera in the SSc skin core microbiome (Additional file 8: Figure S3). The resulting correlation matrix allows for a direct comparison of gene expression and microbiome composition (Fig. 4a).

Hierarchical clustering of this dataset revealed strong associations between human gene expression and microbial abundance. Processes such as T cell, B cell, chemokine, and transforming growth factor beta (TGFβ) signaling in the absence of fatty acid signaling are strongly indicative of the inflammatory subset (Fig. 4a, cluster 1). Lower immune activation signals in combination with major fatty acid metabolism processes are commonly seen in the proliferative subset (Fig. 4a, cluster 2), while fatty acid signaling in the absence of immune activation is most commonly seen in normal-like patients (Fig. 4a, cluster 3).

Taxonomic abundance was strongly associated with the molecular processes of immune activation, lipid metabolism, cell proliferation, and Notch/Wnt signaling (Fig. 4a). Clustering of these processes was strongly correlated with differences in microbial abundance between SSc and controls, with statistically significant differences evident in 5 of 7 clusters (paired t-test, $p < 0.05$ for all; Fig. 4a, b). Among the most significant clusters was cluster 1, dominated by major lipophilic taxa, such as *Malassezia* and *Propionibacterium*, along with numerous Gram-positive Actinobacteria species (Additional file 9: Figure S4). These organisms were significantly more abundant in healthy controls ($p < 0.001$ for Actinobacteria and *Propionibacterium* by paired *t*-test) and exhibited strong positive correlations to lipid metabolism and cell proliferation KEGG pathways (Fig. 4a, b). In contrast, cluster 5 exhibits substantial increases in a wide range of Proteobacteria and other Gram-negative taxa in SSc patients ($p < 0.001$ by paired *t*-test) and is strongly correlated with KEGG immune activation pathways, including Toll-like receptor (TLR) and TGFβ signaling (Fig. 4a, d). Cluster 3 shows strong, positive correlations with immune activation, lipid metabolism, and Notch/Wnt signaling and is associated with statistically significant decreases in Bacteroidetes levels in SSc patients ($p = 0.028$ by paired *t*-test), combined with modest increases in Proteobacteria, relative to controls ($p = 0.085$ by paired *t*-test; Fig. 4a, e). These data demonstrate a strong association between underlying gene expression and the composition of the skin microbiome in SSc.

Discussion

Recent studies have provided significant insight into the immunologic and gene expression-based changes characteristic of SSc patients; however, the underlying mechanisms that initiate and perpetuate disease pathologies remain poorly understood. Analyses of the

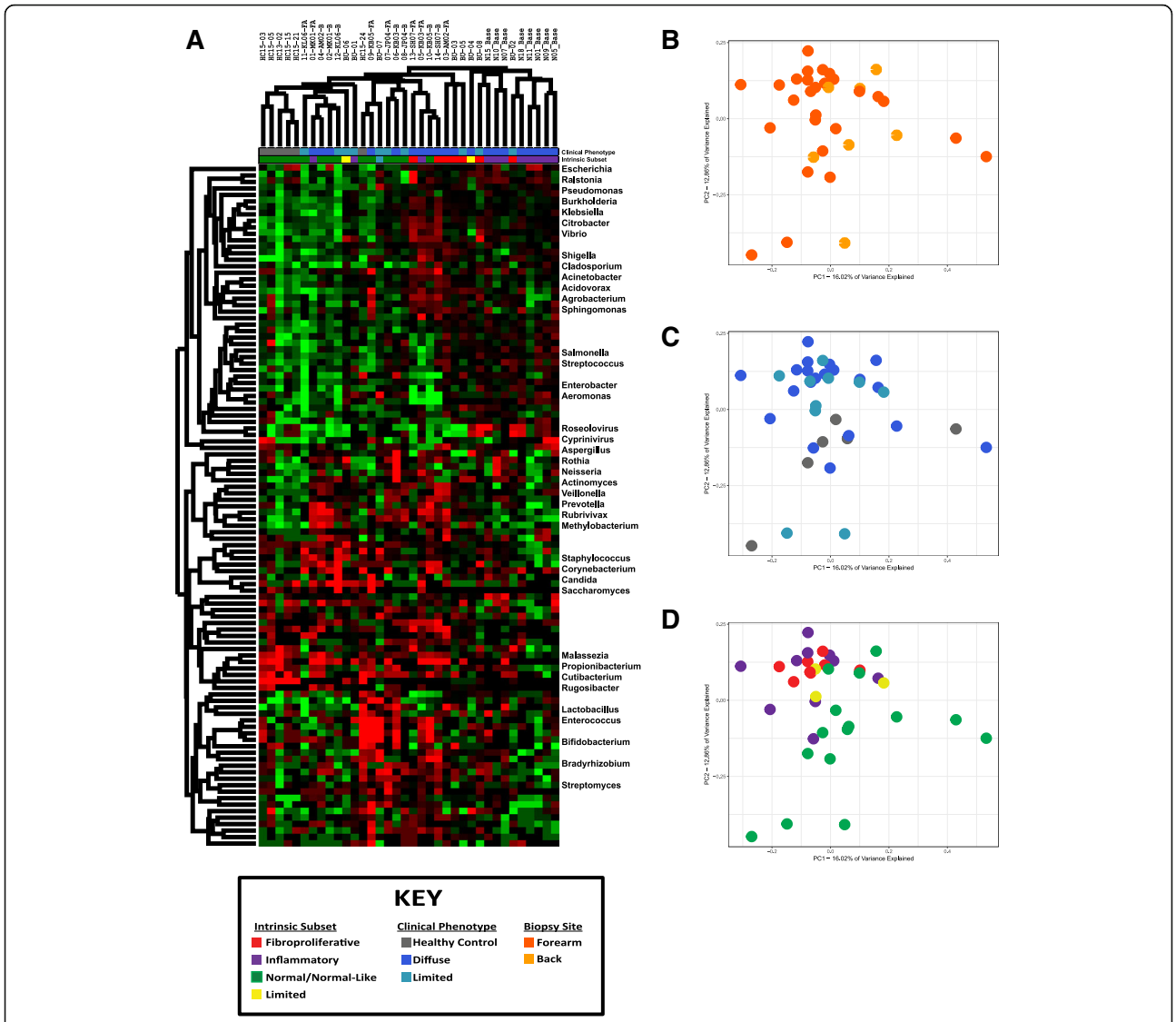


Fig. 3 Distribution of the SSc skin core microbiome. The distribution and relative abundance of the SSc skin core microbiome was calculated by rarefaction to the depth of the lowest sample, and filtering to retain the fewest taxa necessary to account for 90% of all reads, resulting in a total of 103 unique genera. Data were then $\log_2$-transformed and median centered by library preparation. **a** Hierarchical clustering of the core microbiome. Hash marks below the dendrogram indicate intrinsic subset designations and SSc clinical diagnosis for each sample. Principal component analysis of the core microbiome was performed to identify associations between microbiome composition and **b** biopsy location, **c** clinical diagnosis, and **d** intrinsic subset

skin [14] and gut [28] of SSc patients have revealed substantial changes in microbiome composition (dysbiosis), though the role of these organisms in disease pathology is not known. Here, we examined skin biopsies from a diverse cohort of SSc patients by RNA-seq, allowing for the unbiased metagenomic analysis of all potential pathogens, including bacteria, fungi, and viruses, as well as providing a platform from which to investigate the relationship between microbiome composition and underlying gene expression. Limitations of this study include the small number of samples and incomplete clinical data for some patients including a lack of local skin scores. The bimodal distribution of disease duration in our cohort is a confounding factor when assessing changes in microbiome composition over time.

Previous studies examining the composition of the bacterial microbiome in healthy individuals identified three basic environments, dry, moist, and sebaceous, which are reflected in the bacterial populations of these regions [29]. Sebaceous regions, such as the face and back, harbored large proportions of *Propionibacterium* and *Staphylococcus* species, while dry regions, such as the forearm, show a shift towards lower levels of

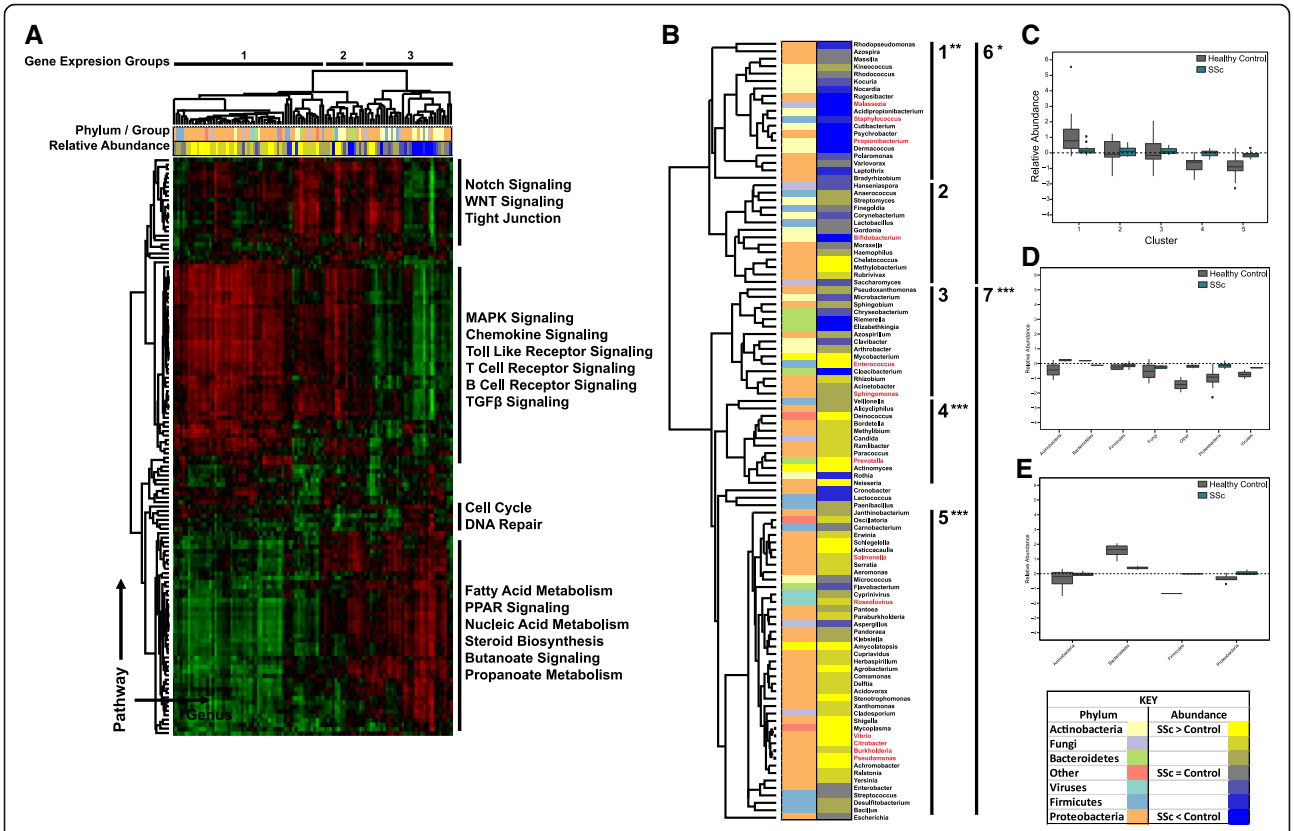


Fig. 4 Microbiome composition is associated with pathway activation in SSc skin. Single-sample gene set enrichment analysis (ssGSEA) was run against normalized FPKM values for all 36 patient samples, using curated KEGG pathways as the probe gene sets. A correlation matrix was then generated by calculating Pearson's correlations for all combinations of ssGSEA values and genus-level abundance across all patients. **a** Hierarchical clustering of the correlation matrix revealed strong associations between SSc-associated gene expression pathways and microbial composition. **b** Taxonomic clustering based on gene expression. Hash marks indicate phylum/group associated with each sample. Relative abundance indicates the degree to which each genus is differentially present in SSc patients, relative to controls with yellow indicating abundance is higher in SSc, while blue indicates abundance is higher in controls. Black bars indicate KEGG pathways that clustered together hierarchically, with representative pathways listed alongside each cluster (*$p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ by paired *t*-test). Clinically relevant genera are highlighted in red. **c** Relative abundance of all genera by taxonomic cluster. **d**, **e** Distribution of taxa is shown for cluster 5 (**d**) and cluster 3 (**e**)

Propionibacterium in combination with higher percentages of *Proteobacteria*, *Corynebacteria*, and *Flavobacteriales* [29]. When considered in this context, where the bacterial microbiome changes as a function of local lipid and moisture levels, the data presented here paint a picture of SSc as a more extreme version of this process, with disease-specific anatomical changes playing a major role in shaping microbiome composition. Lesional SSc forearm skin exhibited significant decreases in *Propionibacterium* and *Staphylococcus* levels relative to controls, along with increases in a wide range of Proteobacteria (Fig. 2), a continuation of the normal differences seen between sebaceous and dry regions of healthy controls. The underlying basis for these changes is also evident at the molecular level, with abundance of lipophilic taxa, such as *Propionibacterium* and *Malassezia*, strongly associated with fatty acid metabolism and lipid signaling pathways, while Proteobacteria were more elevated in patients with active immune signaling (Fig. 4). Microbiome profiles were not significantly different between lesional forearm and non-lesional back skin in this study, with paired samples clustering strongly based on patients, with minimal effect seen in terms of the anatomical site (Additional file 5: Figure S3). This observation further implicates the underlying gene expression as a major driver of microbiome composition, as disease-related changes in gene expression are consistent between SSc lesional forearm and non-lesional back, yet the microbiome profiles of these sites are strongly divergent in healthy controls.

From a mechanistic standpoint, changes in the SSc skin microbiome may be attributed to physical changes associated with fibrotic skin. Atrophy of both hair follicles and sebaceous glands is commonly seen in SSc patients [30], resulting in the loss of both an essential food source, as well as the physical niche where many of these species reside [31]. This loss of skin appendages leads to a weakening of the acid mantle, and a loss of skin barrier function [32].

The strong association between increased Gram-negative taxa, particularly Proteobacteria, and immune activation shown here suggests a potential link between the skin microbiome and immune activation. Analysis of host-microbiome interactions, particularly with the host immune system, is necessary to determine the extent to which these organisms are capable of exacerbating and perpetuating the inflammatory responses in SSc skin.

In a preliminary study of the skin microbiome, increased levels of *R. glutinis* were detected on lesional skin of four untreated, early-stage patients, with only background levels seen in controls, suggesting a potential link between disease etiology and the skin microbiome [14]. Unfortunately, a direct assessment regarding the etiologic nature of this organism was not possible here due to differences in the two patient cohorts, both in terms of prior treatment and disease duration. The majority of early-stage patients described here were not receiving immunosuppressive therapy at the time of biopsy, though all were receiving treatment for SSc-associated symptoms, including vascular symptoms and gastrointestinal reflux. In contrast, five of six late-stage patients (disease duration > 5 years) were untreated at the time of biopsy, consistent with the more quiescent nature of the disease in this population. *Rhodotorula* sequences were consistently detected in both SSc and controls, though these levels never rose above the background noise. Such an observation indicates that while *Rhodotorula* may be increased in very early disease, colonization does not persist over time.

Few viral pathogens were detected in our cohort, with no reads associated with EBV, parvovirus B19, or CMV identified in lesional skin. In contrast, we did consistently identify sequences associated with *Roseolovirus*, a genus which contains both human Herpesviruses (HHV) 6 and 7, which exhibited modest increases in inflammatory SSc patients. As EBV (HHV4), CMV (HHV5), and *Roseolovirus* are all members of the Herpesvirus family, detection of active viral transcription in inflammatory lesional skin does suggest a potential link between life-long latent viral infections and disease pathology, though further studies will be necessary to prove such an association.

Conclusions

The data presented here demonstrate a possible mechanistic link between SSc skin microbiome composition and disease pathology, with a loss of skin appendages and lipid signaling leading to decreases in lipophilic taxa, and a shift to a largely Gram-negative environment. Host-microbiome studies will be necessary to assess the extent to which the microbiome shapes SSc-associated gene expression and vice versa.

Additional files

Additional file 1: Table S1. Gene expression data. Normalized RNA-seq data were classified using a support vector machine trained using a merged and curated dataset composed of samples from GSE9285, GSE32413, and GSE45485. To visual results, the probe ID list from Johnson et al. [4] was collapsed on gene ID. This gene list was compared against normalized FPKM values for all 36 RNA-seq samples, resulting in a total of 1010 overlapping genes. Intrinsic subset assignments for individual genes are not possible based on the nature of gene expression in SSc patients. Each patient's intrinsic subset assignment was determined based on the collective co-expression of all 1010 genes in this dataset, with both high and low expression of individual genes important for determining subset distinctions. Furthermore, both high and low expression of individual genes often extends across multiple intrinsic subsets. This inherently prevents providing subset-specific calls for individual genes. (XLSX 418 kb)

Additional file 2: Table S2. Full clinical data for all patients included in this study. (XLSX 14 kb)

Additional file 3: Table S3. Antimicrobial gene expression in lesional and control skin. (XLSX 36 kb)

Additional file 4: Table S4. Differences in genus-level abundance by clinical covariate. Statistical analyses were performed comparing genus-level abundance between groups, presented as *p* values. Data were compared using the Mann-Whitney U test, corrected for multiple hypothesis testing using the method of Benjamini & Hochberg. Statistically significant differences ($p < 0.05$) are highlighted in yellow; differences significant to $p < 0.10$ are shown in pink. (XLSX 24 kb)

Additional file 5: Figure S1. Principal component analysis of lesional forearm samples based on mRSS. Principal component analysis of core microbiome profiles based on mRSS. Data were limited to SSc lesional forearm samples only. Patients were divided into quartiles based on mRSS score at the time of biopsy (low, < 5; medium, 6–15; high, 16–30; very high, > 30). (PPTX 77 kb)

Additional file 6: Figure S2. Principal component analysis of paired lesional forearm samples. Core microbiome profiles from seven paired forearm and back samples were analyzed by principal component analysis to assess the relationship between anatomical sites. Samples were color coded by A) anatomical site, and B) patient. (PPTX 91 kb)

Additional file 7: Table S5. Single-sample gene set enrichment analysis in SSc patients. Single-sample gene set enrichment analysis (ssGSEA) [26] was run as a module in GenePattern, using relevant KEGG pathways as the query gene sets. Raw pathway enrichment scores are shown for each sample in our dataset. (XLSX 102 kb)

Additional file 8: Figure S3. Comparing gene expression with taxonomic abundance. Single-sample gene set enrichment analysis (ssGSEA) provides a quantitative measurement, expressed as a single value, describing the extent to which a given gene set is coordinately up- or downregulated in a sample. A. To reduce the dimensionality of the data, the activation of a given KEGG pathway was assessed using ssGSEA, reducing a large set of functionally related genes to a single value for each patient. This process was repeated for all available KEGG pathways, generating a table of pathway activation scores for each patient sample (Additional file 2: Table S2). B. Pearson's correlations were then used to compare each set of pathway activation scores against the relative abundance of each genus in the SSc skin core microbiome. C. This process is repeated for each combination of KEGG pathway and genus, producing a correlation matrix. D. Data are then clustered hierarchically and visualized to identify patterns of gene expression, and its relationship to microbial abundance. (PPTX 113 kb)

Additional file 9: Figure S4. Differences in kingdom- and phylum-level abundance between groups. Kingdom- and phylum-level abundance is shown for all major gene expression clusters (Fig. 4a, b). Data are presented as log2 median-centered values for all forearm biopsies from SSc (blue) and controls (gray). (TIF 7050 kb)

Abbreviations

AD: Atopic dermatitis; CMV: Cytomegalovirus; dSSc: Diffuse systemic sclerosis; EBV: Epstein-Barr virus; FPKM: Fragments per million mapped reads;

IMSA: Integrated metagenomic sequencing analysis; lSSc: Limited systemic sclerosis; NCBI: National Center for Biotechnology Information; QIIME: Quantitative Insights Into Molecular Ecology; SPSS: Statistical Package for Social Sciences; SSc: Systemic sclerosis; ssGSEA: Single-sample gene set enrichment analysis; TGFβ: Transforming growth factor beta; TLR: Toll-like receptor

Acknowledgements
N/A

Funding
This work was supported by CDMRP Grant W81XWH-14-1-0224 from the Department of Defense to MLW and SA and a grant from the Scleroderma Research Foundation (MLW).

Availability of data and materials
The full RNA-seq data from this study will be made freely available from the NCBI Short Read Archive (SRA).

Authors' contributions
MEJ, JMF, PAP, SA, and MLW are responsible for the study concept and design. RWS, NO, and TAW are responsible for the sample acquisition. KA, TAW, BKM, and GC are responsible for the data preparation. MEJ, JMF, PAP, SA, and MLW are responsible for the data analysis and interpretation. MEJ, JMF, and MLW are responsible for the manuscript writing. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Study participants provided written, informed consent prior to sample collection in accordance with the Declaration of Helsinki Protocol and the Institutional Review Boards of Boston University Medical Center, Boston, MA, Dartmouth-Hitchcock Medical Center, Lebanon, NH, and the Hospital for Special Surgery, New York, NY.

Consent for publication
N/A

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Hanover, NH, USA. [2]Program in Quantitative Biomedical Sciences, Geisel School of Medicine at Dartmouth, Hanover, NH, USA. [3]Department of Microbiology and Immunology, Geisel School of Medicine at Dartmouth, Hanover, NH, USA. [4]Department of Biomedical Data Science, Program in Quantitative Biomedical Sciences, Geisel School of Medicine at Dartmouth, Hanover, NH, USA. [5]Division of Rheumatology, Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA. [6]Division of Dermatology, University of California, San Francisco, USA. [7]Division of Rheumatology, Arthritis Center, Boston University Medical Center, Boston, MA, USA. [8]Department of Environmental Health Science, University of South Carolina Arnold School of Public Health, Columbia, SC, USA.

Received: 18 July 2018 Accepted: 8 January 2019
Published online: 06 February 2019

References

1. Milano A, Pendergrass SA, Sargent JL, George LK, McCalmont TH, Connolly MK, Whitfield ML. Molecular subsets in the gene expression signatures of scleroderma skin. PLoS One. 2008;3(7):e2696.
2. 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):1363–73.
3. Hinchcliff M, Huang C-C, Wood TA, Mahoney JM, Martyanov V, Bhattacharyya S, Tamaki Z, Lee J, Carns M, Podlusky S. Molecular signatures in skin associated with clinical improvement during mycophenolate treatment in systemic sclerosis. J Invest Dermatol. 2013;133(8):1979–89.
4. Johnson M, Mahoney J, Taroni J, Sargent J, Marmarelis E, Wu M, Varga J, Hinchcliff M, Whitfield M. Experimentally-derived fibroblast gene signatures identify molecular pathways associated with distinct subsets of systemic sclerosis patients in three independent cohorts. PLoS One. 2015;10(1):e0114017.
5. Mahoney JM, Taroni J, Martyanov V, Wood TA, Greene CS, Pioli PA, Hinchcliff ME, Whitfield ML. 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.
6. Weyrich LS, Dixit S, Farrer AG, Cooper AJ. The skin microbiome: associations between altered microbial communities and disease. Australas J Dermatol. 2015;56(4):268–74.
7. Kong HH, Oh J, Deming C, Conlan S, Grice EA, Beatson MA, Nomicos E, Polley EC, Komarow HD, Murray PR. Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. Genome Res. 2012;22(5):850–9.
8. Grossman C, Dovrish Z, Shoenfeld Y, Amital H. Do infections facilitate the emergence of systemic sclerosis? Autoimmun Rev. 2011;10(5):244–7.
9. Hamamdzic D, Kasman LM, LeRoy EC. The role of infectious agents in the pathogenesis of systemic sclerosis. Curr Opin Rheumatol. 2002;14(6):694–8.
10. Radic M, Kaliterna DM, Radic J. Helicobacter pylori infection and systemic sclerosis-is there a link? Joint Bone Spine. 2011;78(4):337–40.
11. Farina A, Cirone M, York M, Lenna S, Padilla C, McLaughlin S, Faggioni A, Lafyatis R, Trojanowska M, Farina GA. Epstein-Barr virus infection induces aberrant TLR activation pathway and fibroblast-myofibroblast conversion in scleroderma. J Invest Dermatol. 2014;134(4):954–64.
12. Csiki Z, Gal I, Sebesi J, Szegedi G. Raynaud syndrome and eradication of helicobacter pylori. Orv Hetil. 2000;141(52):2827–9.
13. Danese S, Zoli A, Cremonini F, Gasbarrini A. High prevalence of Helicobacter pylori type I virulent strains in patients with systemic sclerosis. J Rheumatol. 2000;27(6):1568–9.
14. Arron ST, Dimon MT, Li Z, Johnson ME, Wood TA, Feeney L, Angeles JG, Lafyatis R, Whitfield ML. High Rhodotorula sequences in skin transcriptome of patients with diffuse systemic sclerosis. J Invest Dermatol. 2014;134(8):2138–45.
15. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Arthritis Rheum. 1980;23(5):581–90.
16. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger T Jr, Rowell N, Wollheim F. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol. 1988;15(2):202.
17. Mayes MD. Classification and epidemiology of scleroderma. Semin Cutan Med Surg. 1998;1(1):22–6.
18. Gordon JK, Martyanov V, Magro C, Wildman HF, Wood TA, Huang W-T, Crow MK, Whitfield ML, Spiera RF. Nilotinib (Tasigna™) in the treatment of early diffuse systemic sclerosis: an open-label, pilot clinical trial. Arthritis Res Ther. 2015;17(1):1.
19. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
20. Franks JM, Cai G, Whitfield ML. Feature specific quantile normalization enables cross-platform classification of molecular subtypes using gene expression data. Bioinformatics. 2018.
21. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. Bioinformatics. 2004;20(9):1453–4.
22. Saldanha AJ. Java Treeview—extensible visualization of microarray data. Bioinformatics. 2004;20(17):3246–8.
23. Dimon MT, Wood HM, Rabbitts PH, Arron ST. IMSA: integrated metagenomic sequence analysis for identification of exogenous reads in a host genomic background. PLoS One. 2013;8(5):e64546.
24. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335–6.
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol. 1995:289–300.
26. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009;462(7269):108–12.
27. Whitfield ML, Finlay DR, Murray JI, Troyanskaya OG, Chi JT, Pergamenschikov A, McCalmont TH, Brown PO, Botstein D, Connolly MK. Systemic and cell

type-specific gene expression patterns in scleroderma skin. Proc Natl Acad Sci U S A. 2003;100(21):12319–24.
28. Volkmann ER, Chang YL, Barroso N, Furst DE, Clements PJ, Gorn AH, Roth BE, Conklin JL, Getzug T, Borneman J. Association of Systemic Sclerosis With a Unique Colonic Microbial Consortium. Arthritis Rheum. 2016;68(6):1483–92.
29. Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, Bouffard GG, Blakesley RW, Murray PR, Green ED. Topographical and temporal diversity of the human skin microbiome. Science. 2009;324(5931):1190–2.
30. Calonje JE, Brenn T, Lazar AJ, McKee PH. Pathology of the skin: Elsevier Health Sciences; 2011.
31. Tanghetti EA. The role of inflammation in the pathology of acne. J Clin Aesthet Dermatol. 2013;6(9):27–35.
32. Schmid-Wendtner M-H, Korting HC. The pH of the skin surface and its impact on the barrier function. Skin Pharmacol Physiol. 2006;19(6):296–302.