Published in final edited form as:

Arthritis Rheumatol. 2015 November ; 67(11): 3004–3015. doi:10.1002/art.39287.

A Longitudinal Biomarker for the Extent of Skin Disease in Patients with Diffuse Cutaneous Systemic Sclerosis

Lisa M. Rice, BS, Jessica Ziemek, MPH, Eric A. Stratton, MTS, MPH, Sarah R. McLaughlin, MPH, Cristina M. Padilla, MS, Allison L. Mathes, PhD, Romy B. Christmann, MD, PhD, Giuseppina Stifano, MD, Jeffrey L. Browning, PhD, Michael L. Whitfield, PhD[†], Robert F. Spiera, MD[‡], Jessica K. Gordon, MD[‡], Robert W. Simms, MD, Yuqing Zhang, MPH, DSc, and Robert Lafyatis, MD

Boston University School of Medicine, Boston, MA

[†]Geisel School of Medicine at Dartmouth University Medical School, Hanover, NH

[‡]Hospital for Special Surgery, New York, NY

Abstract

Objective—The goal of this study was to define a pharmacodynamic biomarker based on gene expression in skin that would provide a biological measure of disease extent in patients with diffuse cutaneous systemic sclerosis (dcSSc) and that could be used to monitor skin disease longitudinally.

Methods—Skin biopsies taken from a cohort of dcSSc patients that included longitudinal samples were analyzed by microarray. Expression of genes correlating with the modified Rodnan skin score (MRSS) were examined by nanostring for change over time, and a generalized estimating equation used to define and validate longitudinal, pharmacodynamic biomarkers composed of multiple genes.

Results—Microarray analysis of genes parsed to include only genes correlating with the MRSS revealed prominent clusters of profibrotic/TGF β -regulated, IFN-regulated/proteasome, macrophage and vascular marker genes. Using genes changing longitudinally with the MRSS, two multigene, pharmacodynamic biomarkers were defined. The first was defined mathematically, applying a generalized estimating equation to longitudinal samples. This modeling method selected cross-sectional THBS1 and longitudinal THBS1 and MS4A4A genes. The second model was based on a weighted selection of genes, including additional genes with statistically significant change over time: CTGF, CD163, CCL2 and WIF1. Biomarker levels calculated using both models correlated highly with the MRSS in an independent validation dataset.

Conclusion—Skin gene expression can be used effectively to monitor SSc skin disease change over time. We have implemented these relatively simple models on a nanostring platform permitting highly reproducible assays that can be applied directly to samples from patients or collected as part of clinical trials.

INTRODUCTION

The pathogenesis of systemic sclerosis (SSc) remains largely unknown. Controlled clinical trials for the organs most commonly affected by fibrosis, the skin and lungs, have been mostly negative except for a modest benefit provided by cyclophosphamide, and more dramatic benefit but also significant treatment-related mortality by immunoablation and stem cell transplant. The accessibility of skin to simple and repetitive biopsy makes it a particularly attractive organ for understanding pathogenesis and the effect of therapeutics, since it is directly involved by the disease process and can be easily sampled longitudinally, before and after therapeutic interventions.

Expression of select genes assessed from a skin biopsy taken in the mid-forearm correlate with the degree of skin disease in patients with diffuse cutaneous systemic sclerosis (dcSSc), correlating with the modified Rodnan skin score (MRSS) [1]. We previously reported a cross-sectional relationship between expression levels of four genes and the MRSS using multiple linear regression [1]. The product of this equation correlated with the MRSS and, although not derived from an examination of longitudinal data, was found to change over time in parallel with the MRSS in a small number of patients. Two of the genes in this four-gene biomarker, cartilage oligomeric protein (COMP) and thrombospondin-1 (THBS1), are known to be highly regulated by transforming growth factor- β (TGF β) whereas the other two genes, interferon-induced protein 44 (IFI44), and sialoadhesin (SIGLEC1) are known to be highly regulated by interferon (IFN). Genes were tested for this biomarker on the basis of their known regulation by these two cytokines, but relatively few other genes were tested.

Gene expression profiling permits a more systematic approach to identification of pharmacodynamic biomarkers in SSc skin. The closer integration of clinical and biological outcomes also provides a much more powerful methodology for identifying pathogenic pathways than either in isolation. For example, it is now clear that many genes show altered expression in SSc compared to healthy skin. However, many of these alterations in gene expression are likely not central to disease pathogenesis, and the sheer number of them – in the thousands – makes it difficult to assess their role in pathogenesis. Here we utilize a novel approach to study gene expression in SSc skin, selecting genes on the basis of their relation to the MRSS prior to utilizing more standard approaches to interpreting high throughput gene expression, particularly clustering. Using this approach we can more clearly distinguish pathogenic mechanisms based on gene clusters, as well as select candidate pharmacodynamic gene biomarkers, confirming the validity of our earlier approach but also permitting generation of a more robust longitudinal pharmacodynamic biomarker for the extent of SSc skin disease.

MATERIALS AND METHODS

Patients

All skin samples and clinical data were collected under approval by Institutional Review Boards, most as part of a biomarker protocol at the Boston University Scleroderma Center, and the validation RNAs as part of a clinical trial of nilotinib at the Hospital for Special Surgery (Trial Registration: NCT01166139, [2]). Skin was biopsied on the dorsal surface of

the mid-forearm, immediately placed into RNAlater and stored at -80°C until RNA purification.

Statistical analysis

Gene expression was transformed ($\log_2$) to improve normality. Pearson's correlations were calculated to examine the relationships between MRSS and gene expression. Simple regression between MRSS and change in gene expression was visualized using the R package "corrgrams" (R 3.0.1) and ordered using principal component analysis of the correlation matrix [3 4]. Inter-Rater Reliability of gene expression data was assessed (R 3.0.1 package "irr") using a two-way absolute model for intra-class correlation (ICC) [5]. Generalized estimating equations (GEE) were used to examine the relation of gene expression at baseline and its change over time to the MRSS [6].

RNA purification

For RNA purification skin was minced and homogenized with a Polytron homogenizer, purified using the RNeasy Mini Kit (Qiagen), and the concentration of total RNA measured (Nanodrop 1000; ThermoScientific).

Microarray analyses

200 ng of RNA was analyzed by microarray using standard protocols on Affymetrix U133A 2.0 arrays by the Boston University Microarray Core. Microarray gene expression data was analyzed after selecting genes correlating with the MRSS ($r>0.4$), using Cluster 3.0 [7]. This parsed dataset (1,770 genes out of 22,278 genes) was analyzed by unsupervised clustering of genes and samples using average correlation and uncentered linkage. Clusters were visualized using Treeview 1.1.6 [8].

Nanostring analysis

For nanostring analysis 5uL of each RNA (20 ng/uL) to be analyzed was added to tubes in 12 strip tubes. 20uL Reporter code set diluted in buffer code set solution (NanoString Technologies), and 5uL of the Capture Probe Set (NanoString Technologies) was then added to each tube. The solution was inverted to mix, centrifuged and placed in a thermocycler (MJ Mini; Bio-rad) overnight at 65°C . Samples were removed from the thermocycler and processed in the nCounter PrepStation (NanoString Technologies). Once complete samples were transferred to the Digital Analyzer (NanoString Technologies). Expression of genes was normalized to housekeeping genes (Nanostring nCounter).

Initial analysis of pharmacodynamic biomarkers

For initial exploratory studies, expression of representative genes in each cluster on microarray showing high correlations with the MRSS were tested in additional skin samples ("Discovery cohorts") from patients with dcSSc using a series of custom-designed Exploratory nanostrings. Gene expression based on these studies was correlated with the MRSS (these results are summarized in Supplementary Table 1).

Genes on Exploratory nanostrings showing expression that correlated highly and consistently with the MRSS were selected for further biomarker development and modeling

by testing in a third custom-designed nanostring (SSc nanostring 3.0) that included 17 putative biomarker genes (Supplemental Table 2). Expression of these genes was normalized to the geometric mean of 18 housekeeping genes.

Biomarker model development using generalized estimating equation

Generalized estimating equations (GEE) were used to examine the relationship between MRSS and gene expression in longitudinal biopsies. As previously described this equation can be written as: $Y_{ij} = \beta_0 + \beta_b x_{i0} + \beta_w (x_{i(j-1)} - x_{i0}) + \gamma CF_{i(j-1)} + e_{ij}$ [6]. This model estimates the cross-sectional association of the risk factor and the outcome variable with (β_b) between individuals, as well as the longitudinal association within subjects between change in risk factor and change in outcome variable with (β_w) within an individual. In this study: Y is the observed measurement of the outcome (MRSS); x is the gene of interest; and CF represents potential confounders. A longitudinal model of change was developed using the GENMOD procedure of SAS, University Edition (Statistical Analysis System Institute Inc); compound symmetry was used for all analyses. Beta-coefficients were derived from the original dataset, which included 15 patients that had at least one follow up visit with the same rheumatologist where a skin was biopsied. These beta-coefficients were held constant and tested in the validation data set, which included 10 patients whose skin was biopsied longitudinally at the Hospital for Special Surgery.

Calculation of scores using pharmacodynamic models of MRSS

2-Gene SSc Model and Weighted Model skin scores were calculated in a group of skin RNAs collected for a phase-I clinical trial of nilotinib. To calculate the 2-Gene SSc score of these samples the beta-coefficients derived from the original data set were used and gene expression of THBS1 and MS4a4a entered into the final correlation equation. For the Weighted Model, the Z-score of each gene included in the model was calculated on the basis of the average and standard deviation of all SSc samples. These scores were added together and 15 added to maintain a positive value for all samples and more closely approximate values seen for clinical MRSS.

Histological assessment of fibrosis and inflammation

Hematoxylin and Eosin stained sections were scored by a blinded observer (RL) for the degree of fibrosis scoring hyalinized collagen, using a 10-point scale, similarly to how described previously [9 10].

RESULTS

Microarray gene expression was evaluated in skin biopsies from 15 patients with dcSSc and four healthy controls (Supplemental 3). All but three dcSSc patients were within 3 years of their first non-Raynaud's symptom, one of these having limited disease for 8 years before transitioning to diffuse disease approximately 18 months prior to biopsy.

Genes making up the “four-gene biomarker” fall into two distinct clusters

Genes showing expression that correlated positively with the MRSS (Pearson $r = 0.4$; 1,770 of 22,278 genes, Supplemental Table 4) were analyzed by unsupervised clustering for both

genes and subjects. Preselecting genes correlating positively with the MRSS gave easily recognizable clusters, showing related functions and plausible biological significance.

Three of the four genes included in the original four-gene biomarker [1], COMP, THBS1 and SIGLEC1, were included in the clustering analysis (all correlating with the MRSS with $r = 0.4$). The fourth, IFI44, was not, showing a relatively low correlation with the MRSS. COMP and THBS1 clustered very closely with the MRSS (MRSS scores were added to the dataset before clustering), as did a significant group of other genes easily recognized for their regulation by TGF β , as extracellular matrix proteins, or as genes previously described upregulated in SSc skin (selected genes in this cluster are shown in Figure 1, upper panel, the complete cluster of genes is shown in Supplemental Figure 1). In particular, THBS1 and COMP clustered with matrix proteins: collagen VIII alpha 2, collagen X alpha 1, collagen XI alpha 1, and fibronectin, as well as secreted frizzled related protein 4 and wingless-type MMTV integration site family member 2 (Wnt2), Wnt-related genes increased in SSc skin [11]. Notably, the interleukin-4 receptor gene, which correlates with the MRSS and forms part of the IL-4 and IL-13 receptor, also clustered with these genes [12]. Expression of genes in this cluster correlated relatively highly with the MRSS, most ranging between $r=0.6$ to 0.7 (Figure 1).

SIGLEC1 was found in a cluster with a group of other IFN regulated genes (select genes shown in Figure 1 and full cluster shown in Supplemental Figure 2), including guanylate binding protein, interferon induced transmembrane proteins 1, 2, and 3, phospholipid scramblase 1, interferon induced protein 35, and MHC class II DR beta. Additionally this cluster included genes encoding proteins mediating intracellular signals by interferon: STAT1 and interferon regulatory factor 7, and mediating intracellular signals from toll-like receptors: myeloid differentiation primary response gene 88. IFN and COMP/THBS1 clusters showed increased expression in many of the same patients, but some patients with a high skin score only showed increased gene expression in the IFN cluster, such as patients SD5-10, SD6-4 and SD10-22, consistent with our previous report [1]. The correlation of genes in this IFN cluster with the MRSS was lower than genes found in the COMP/THBS1 cluster, generally in the range of $r = 0.45$ to 0.65 .

Another profibrotic gene cluster shows multiple TGF β -regulated genes

Another cluster immediately adjacent to the COMP/THBS1 cluster contained many additional recognizable TGF β -regulated genes (Figure 2A, and complete cluster shown in Supplemental Figure 3). Several of the genes in this cluster have been found upregulated in SSc, such as serpin peptidase inhibitor [1], connective tissue growth factor [13 14], fibronectin, collagen, type IV [15], and insulin-like growth factor binding protein 3 [16]. Also found in this cluster are lysyl-oxidase, type 2 (LOXL2), a gene promoting collagen cross-linking and recently implicated in pathological fibrosis [17], and WNT1 inducible signaling protein (WISP1), a Wnt inhibitor upregulated in SSc [11]. The correlation of these genes with the MRSS was relatively high, similar to that seen in the COMP/THBS1 cluster.

A cluster of genes associated with MHC class I, proteasome and antigen processing

Immediately adjacent to the cluster of IFN-regulated genes described above, another cluster of genes was easily recognizable containing almost exclusively genes associated with MHC I and antigen processing (MHCI/Proteasome cluster, Figure 2B). MHC class I molecule upregulation in this cluster was accompanied by increased expression of tapasin, proteasome 26S subunit, non-ATPase, 9 (PSMD9/LMP2), PSMD10 (LMP10) and proteasome activator subunit 2 beta (PA28b), all known to be strongly upregulated by IFN γ [18], but also by type I IFNs [19–21]. Several patients are discrepant between this and the SIGLEC1/IFN cluster (Figure 1), for example patient SD05-4, showing higher expression in the MHCI/proteasome cluster and SD10-20, showing higher expression in the SIGLEC1/IFN cluster.

Macrophage-associated genes overlap with IFN-regulated genes

Another relatively easily identifiable cluster of genes included macrophage markers, most distinctly CD163 and macrophage scavenger receptor 1 (selected genes, Figure 3A; and complete cluster, Supplemental Figure 4). Immediately adjacent, another cluster showed macrophage and IFN γ related genes (Figure 3B and supplemental Figure 5). We have recently shown that two genes found in this cluster, IL-13 receptor alpha 1 (IL13RA1) and interferon gamma receptor 1 are found primarily on circulating monocytes and are associated with limited cutaneous SSc patients who have pulmonary arterial hypertension [22]. Included in this cluster are also genes found associated with IFN: gamma-associated protein 16 and phospholipid scramblase 1, emphasizing the close relationship between this upregulated macrophage marker and IFN-regulated genes. Immediately adjacent to this cluster was another cluster that contains chemokine (C-C motif) ligand 2 (CCL2, Figure 3B). These results are thus consistent with our recent observations that IL13RA1 and CCL2 correlate highly with the MRSS, showing $r = 0.711$ and 0.792 , respectively [12].

Vascular marker genes cluster together

A cluster of genes associated with endothelium was also easily identifiable (Figure 3C, full cluster shown in Supplemental Figure 6). Upregulated von Willebrand factor is a marker of vascular disease in SSc sera [23]. Other well-known markers or products of endothelial cells were also found in this cluster: P-selectin, angiopoietin-2, intracellular adhesion molecule 2, plexin D1 and junctional adhesion molecule 2. The correlation of these markers of endothelial cells with the MRSS was lower than seen in several of the other clusters, typically between 0.45 and 0.5.

Development of nanostring based skin biomarker

We tested expression of genes correlating most highly with the MRSS in each of the families described above using several “Exploratory” nanostring constructs containing TGF β or interferon-regulated genes, and genes found in the MHC class I/proteasome, Macrophage and Vascular clusters. The nanostring platform permitted multiplex analyses and showed higher reproducibility than RT-PCR (data not shown). On repeat testing in skin biopsies from more patients, some of the genes analyzed continued to correlate highly with the MRSS, whereas others correlated less highly on repeat testing and were dropped from serial nanostring constructs (see Supplemental Methods and Supplemental Table 1).

On the basis of these results we constructed a more refined nanostring (SSc nanostring 3.0, Supplemental Table 2) and tested a much larger array of skin biopsies including both biopsies from patients with longer standing disease and patients with longitudinal biopsies (Supplemental Table 3). Many of the genes identified in our microarray and exploratory nanostring studies again correlated highly with the MRSS (Figure 4). Surprisingly, the disease duration did not significantly affect the correlations for most genes analyzed, and the correlations actually trended higher in biopsies from patients of longer disease duration for many genes (Figure 4). WIF1 correlated negatively with the MRSS, although this correlation was less robust in biopsies from patients with longer disease duration.

We compared change in gene expression with MRSS and other clinical features and depicted these by a correlation matrix, which displays patterns of linear dependence among variables (Figure 5A). There was high linear dependence between TGF β -regulated genes: THBS1 and CTGF and between macrophage related genes: CD163 and MS4A4A.

Statistically modeling a longitudinal, pharmacodynamic biomarker for SSc skin disease

In order to define a skin gene biomarker that could be used in clinical and clinical trial settings, we examined both the relationship of gene expression to the MRSS cross-sectionally between patients, and longitudinally within individual patients. These data showed that several genes correlated more significantly with changes in the MRSS longitudinally, than with the MRSS cross-sectionally (Figure 5B, compare longitudinal and cross-sectional p-values, see Supplemental Table 5 for beta coefficients and confidence intervals, see Supplemental Table 6 for changes in MRSS and associated changes in gene expression).

Using all available nanostring data, we developed two models for assessing changes in MRSS over time. We first developed a model based purely on mathematical criteria using a generalized estimating equation. This model estimates the association cross-sectionally between the risk factor (baseline gene expression) and the outcome variable (MRSS) with one beta coefficient, as well as the association longitudinally with a second beta coefficient. Each available gene was tested individually for significance of cross-sectional and longitudinal beta coefficients (Supplemental Table 5). CCL2, CD163, COMP, CTGF, IGFBP3, IL13RA1, MS4A4A, THBS1 and WIF1 had highly significant longitudinal beta coefficients ($p < 0.015$) and were then modeled in combinations. Comparing models, THBS1 and MS4A4A combined to give the best longitudinal regression equation:

$$\text{MRSS} = -27.6844 + [4.46(\text{baseline THBS1})] + [5.31(\Delta\text{MS4A4A}) + 4.96(\Delta\text{THBS1})].$$

Thus, the beta coefficient for THBS1 at baseline estimates the cross-sectional association between that biomarker and MRSS between individuals, whereas the beta coefficient for change of THBS1 and MS4A4A represent longitudinal associations between change in each biomarker and the change in MRSS within an individual. We refer to this as the 2-gene SSc skin biomarker or 2GSSc.

Using the gene expression data to calculate the biomarker predicted skin score showed that it correlated highly with the clinical MRSS (Figure 6, $R=0.743$, $p=0.0001$). To further validate this model in an independent dataset we calculated the predicted skin score in skin RNAs collected from another institution (comparison of patient clinical features see Supplemental Table 7). Again the 2GSSc skin biomarker score correlated highly with the observed MRSS (Figure 6, $R=0.810$, $p=0.0001$). We examined reproducibility of this technology by examining gene expression in concurrent adjacent biopsies in 5 patients. These analyses showed very high intraclass correlations for THBS1 (0.966) and MS4A4A (0.951).

Weighted modeling a longitudinal, pharmacodynamic biomarker for SSc skin disease

The purely statistical 2GSSc skin biomarker included only 2/9 genes we had identified as changing longitudinally with $p<0.015$ in GEE equation. Therefore, we also constructed a Weighted Model that would include more genes. Genes were selected that clustered separately but correlated significantly, longitudinally with the MRSS (Figure 4). Using Z-scores to quantify differences in levels of expression (see Supplemental methods), we constructed the following expression:

$$\text{MRSS} = (\text{THBS1} + \text{CTGF} + \text{CCL2} + \text{CD163} + \text{MS4A4A} - \text{WIF}) + 15.$$

Calculating the predicted MRSS based on the dataset used to derive this Weighted Model (WSSc skin biomarker) is thus similar to calculations of IFN signature score used in systemic lupus erythematosus [24]. Scores derived using this model also correlated highly with the clinical MRSS (Figure 6, $R=0.688$, $p=0.0002$). We further validated this model in an independent dataset where the WSSc skin biomarker also correlated highly with the observed MRSS (Figure 6, $R=0.794$, $p=0.0001$).

2GSSc and WSSc skin biomarkers correlate with histological features of skin

To provide an additional anchor to the 2GSSc and WSSc skin biomarkers, we examined the relationship between biomarker expression, and fibrosis and inflammation assessed histologically on biopsies taken in parallel from adjacent skin on the same day. These results showed that the MRSS correlated moderately well with both skin fibrosis and inflammation (Figure 6E). The 2GSSc and WSSc skin biomarkers also correlated moderately with fibrosis and inflammation, almost as strongly as the MRSS. Surprisingly, macrophage markers CD163 and MS4A4A correlated only weakly with histological inflammation, suggesting that the cells most associated with the MRSS are not the same quantitatively as the perivascular inflammatory cells typically seen by routine histology.

DISCUSSION

The main goal in our studies of biomarkers has been to identify a pharmacodynamic biomarker that could be effectively implemented into clinical trials to provide supplementary information to the MRSS and other skin outcomes regarding drug efficacy. For a biomarker to fulfill this role it needs to change with change in a validated clinical outcome, in this case the MRSS. The current study confirms that THBS1 alone is a very strong pharmacodynamic biomarker, and presently the most reliable single pharmacodynamic biomarker based on the

consistent observation of its high correlation with the MRSS in multiple datasets and, in particular, its performance in a recently completed trial of anti-TGF β , fresolimumab [25].

TGF β regulates THBS1 mRNA expression in vitro and it clusters tightly with other TGF β -regulated genes. TGF β activation and its effect on dermal cells likely represent a step directly in the pathway leading to fibrosis and resulting clinical manifestations. This feature is key for performance of THBS1 as a pharmacodynamic biomarker, since biomarkers that are not directly part of the pathogenic pathway might be affected by ineffective therapeutic interventions. In addition, the effect of TGF β on THBS1 most likely represents a step relatively distal in the pathogenic process, discussed further below. Detecting events distal in pathogenesis is also important for a pharmacodynamic biomarker, as effective therapeutics targeting pathogenic events downstream from the biomarker will remain undetected by the biomarker.

We propose to continue to validate two pharmacodynamic biomarker constructs in ongoing and planned clinical trials. Our calculated biomarker, 2GSSc biomarker score, is constructed solely on mathematical considerations. MS4A4A is a macrophage marker closely related structurally to CD20 seen on B cells. It clusters with other macrophage markers but modeling shows it as a more robust longitudinal biomarker than other macrophage markers, and the addition of other genes to the model does not add statistical significance. We propose a second model, WSSc biomarker score, to include a broader array of genes that correlate with changes in the MRSS. This may be better from the standpoint of not relying on only two measures and also because it captures more of the pathways and gene clusters that might drive pathogenesis. Both of these composite biomarkers validated an independent cohort of samples collected from a different institution.

There are limitations to these biomarkers. Serial change in forced vital capacity, measures of vascular severity (such as digital ulcers, pulmonary arterial hypertension), and other clinical outcomes might be associated with different gene expression patterns. Although recent data argue that MRSS often peaks within 18 months, our data indicate that the correlation between the biomarkers and the skin score extends over 36 months or even longer. These observations suggest that, although skin scores begin to abate, there may be continuing disease activity, manifest as continuing alterations in skin gene expression.

Parsing out and clustering only genes whose expression correlates with the MRSS provided some novel insights into pathogenesis. Most prominently and correlating most strongly with the MRSS, two clusters show profibrotic genes, most known to be regulated by TGF β and many previously implicated in SSc pathogenesis: thrombospondin1, cartilage oligomeric protein 1, connective tissue growth factor, cadherin 11, fibronectin and collagens type IV, VIII, X, and XI. Since matrix proteins directly contribute to skin thickening and these genes correlate most highly with the skin score, we speculate that regulation of these genes is very close to the final step in SSc pathogenesis.

Two clusters of genes are highly enriched in IFN-regulated and/or proteasomal genes. Although the relative roles of type I and type II IFNs in SSc remains enigmatic, it is notable that proteasomal genes cluster discretely, suggesting the possibility that both IFN types

might be activating gene expression and involved in pathogenesis. IFN regulated genes also appear in one of two prominent clusters of macrophage regulated genes. The other cluster contains well -known macrophage markers CD14, also a prognostic biomarker for progressive skin disease [26], and CD163.

Perhaps most exciting in this clustering paradigm is the appearance of genes found expressed mainly by endothelial cells. We hypothesize that these genes are markers of vascular injury, since von Willebrand factor is upregulated in many diseases associated with vascular injury [27 28]. Many studies and clinical observations suggest that autoinflammatory pathways trigger vascular injury and fibrosis. The observed clusters highlight macrophage driven and vascular injury pathways but do not clarify primacy. The lack of prominent T cell signature(s) is notable, but suggests that T cell influences occur early and transiently in SSc or are not a major factor in the skin. Together these results provide a strong paradigm for SSc pathogenesis, indicating the altered gene expression in inflammatory and vascular cells drives connective tissue gene expression and fibrosis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Tammara Wood for technical assistance and Kate Brennan for help with manuscript preparation.

Funding Information: This work was supported by National Institutes of Health, Boston University Medical Center CTSI: UL1-TR000157; National Institute of Arthritis Musculoskeletal and Skin Disease grants: Scleroderma Core Centers (5P30AR061271), Scleroderma Center of Research Translation (1P50AR060780) and 2R01AR051089 to RL; the BUMC Microarray Core; a Clinician Scientist Development Award through the Kellen Foundation at Hospital for Special Surgery to JG; and Scleroderma Research Foundation and Dr. Ralph and Marian Falk Medical Research Trust grants to MLW. The nilotinib clinical trial was also supported by an investigator-initiated grant from Novartis and by the Rudolph Rupert Scleroderma Program at Hospital for Special Surgery.

Financial Support: Robert Spiera has received grants Novartis, Roche/Genentech, Human Genome Sciences/ GlaxoSmithKline, Actelion, United Therapeutics, Chemocentryx and Bristol Myers Squibb; and consulting fees from Boehringer Ingelheim and Alexion. Jessica Gordon has received grants from Novartis. Jeffrey Browning received consulting fees from Pfizer. Michael Whitfield has conducted financial activities with Celdara Medical LLC. Robert Simms received grants from Actelion, Celgene, Reata, Bayer, Genentech and Intermune; and both grants and consulting fees from Cytospor Therapeutics and Gilead. Robert Lafyatis has received both grants and consulting fees from Genzyme/Sanofi, Shire, Regeneron, Biogen, Bristol Myers Squibb, Inception, Precision Dermatology, PRISM, UCB, Precision Dermatology, Pfizer and Roche/Genentech. He has 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 Human Genome Sciences. Boston University, Lisa Rice, Giuseppina Stifano and Robert Lafyatis have submitted a provisional patent on results of this study.

References

1. Farina G, Lafyatis D, Lemaire R, et al. A four-gene biomarker predicts skin disease in patients with diffuse cutaneous systemic sclerosis. *Arthritis and rheumatism*. 2010; 62(2):580–8. [published Online First: Epub Date]. DOI: 10.1002/art.27220 [PubMed: 20112379]
2. Gordon JK, Magr C, Udeh U, et al. Nilotinib (Tasigna™) In The Treatment Of Early Diffuse Systemic Sclerosis: A Single Group, Open Label Pilot Clinical Trial – One Year Results. *American College of Rheumatology Meeting Abstracts*. 2013:709.
3. Friendly M. Corrggrams: Exploratory Displays for Correlation Matrices. *The American Statistician*. 2002; 56:316–24.

4. Wright, K. Corrgram: Plot a correlogram. R package version 1.6. 2014. <http://CRAN.R-project.org/package=corrgram>
5. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1996; 1:30–46.
6. Zhang Y, Zhang B, Wise B, et al. Statistical approaches to evaluating the effect of risk factors on the pain of knee osteoarthritis in longitudinal studies. *Current opinion in rheumatology*. 2009; 21(5): 513–9. [published Online First: Epub Date]. DOI: 10.1097/BOR.0b013e32832ed69d [PubMed: 19584728]
7. de Hoon MJ, Imoto S, Nolan J, et al. Open source clustering software. *Bioinformatics*. 2004; 20(9): 1453–4. [published Online First: Epub Date]. DOI: 10.1093/bioinformatics/bth078 [PubMed: 14871861]
8. Saldanha AJ. Java Treeview—extensible visualization of microarray data. *Bioinformatics*. 2004; 20(17):3246–8. [published Online First: Epub Date]. DOI: 10.1093/bioinformatics/bth349 [PubMed: 15180930]
9. Kissin EY, Merkel PA, Lafyatis R. Myofibroblasts and hyalinized collagen as markers of skin disease in systemic sclerosis. *Arthritis and rheumatism*. 2006; 54(11):3655–60. [published Online First: Epub Date]. DOI: 10.1002/art.22186 [PubMed: 17075814]
10. Mathes AL, Christmann RB, Stifano G, et al. Global chemokine expression in systemic sclerosis (SSc): CCL19 expression correlates with vascular inflammation in SSc skin. *Annals of the rheumatic diseases*. 2014; 73(10):1864–72. [published Online First: Epub Date]. DOI: 10.1136/annrheumdis-2012-202814 [PubMed: 23873879]
11. Bayle J, Fitch J, Jacobsen K, et al. Increased expression of Wnt2 and SFRP4 in Tsk mouse skin: role of Wnt signaling in altered dermal fibrillin deposition and systemic sclerosis. *The Journal of investigative dermatology*. 2008; 128(4):871–81. [PubMed: 17943183]
12. 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. *The American journal of pathology*. 2012; 180(3):1080–94. [published Online First: Epub Date]. DOI: 10.1016/j.ajpath.2011.11.024 [PubMed: 22245215]
13. Igarashi A, Nashiro K, Kikuchi K, et al. Significant correlation between connective tissue growth factor gene expression and skin sclerosis in tissue sections from patients with systemic sclerosis. *The Journal of investigative dermatology*. 1995; 105(2):280–4. [PubMed: 7636314]
14. Kikuchi K, Kadono T, Ihn H, et al. Growth regulation in scleroderma fibroblasts: increased response to transforming growth factor-beta 1. *The Journal of investigative dermatology*. 1995; 105(1):128–32. [PubMed: 7615966]
15. Gerstmeier H, Gabrielli A, Meurer M, et al. Levels of type IV collagen and laminin fragments in serum from patients with progressive systemic sclerosis. *The Journal of rheumatology*. 1988; 15(6):969–72. [PubMed: 3418647]
16. Brissett M, Veraldi KL, Pilewski JM, et al. Localized expression of tenascin in systemic sclerosis-associated lung fibrosis and its regulation by IGF binding protein (IGFBP)-3. *Arthritis and rheumatism*.
17. Barry-Hamilton V, Spangler R, Marshall D, et al. Allosteric inhibition of lysyl oxidase-like-2 impedes the development of a pathologic microenvironment. *Nature medicine*. 16(9):1009–17.
18. Strehl B, Seifert U, Kruger E, et al. Interferon-gamma, the functional plasticity of the ubiquitin-proteasome system, and MHC class I antigen processing. *Immunological reviews*. 2005; 207:19–30. [PubMed: 16181324]
19. Seifert U, Bialy LP, Ebstein F, et al. Immunoproteasomes preserve protein homeostasis upon interferon-induced oxidative stress. *Cell*. 2010; 142(4):613–24. [PubMed: 20723761]
20. Nandi D, Woodward E, Ginsburg DB, et al. Intermediates in the formation of mouse 20S proteasomes: implications for the assembly of precursor beta subunits. *The EMBO journal*. 1997; 16(17):5363–75. [PubMed: 9311996]
21. Abarca-Heidemann K, Friederichs S, Klamp T, et al. Regulation of the expression of mouse TAP-associated glycoprotein (tapasin) by cytokines. *Immunology letters*. 2002; 83(3):197–207. [PubMed: 12095710]

22. Pendergrass SA, Hayes E, Farina G, et al. Limited systemic sclerosis patients with pulmonary arterial hypertension show biomarkers of inflammation and vascular injury. *PloS one*. 2010; 5(8):e12106. [published Online First: Epub Date]. doi: 10.1371/journal.pone.0012106 [PubMed: 20808962]
23. Kahaleh MB, Osborn I, LeRoy EC. Increased factor VIII/von Willebrand factor antigen and von Willebrand factor activity in scleroderma and in Raynaud's phenomenon. *Annals of internal medicine*. 1981; 94(4 pt 1):482–4. [PubMed: 6782927]
24. Kirou KA, Lee C, George S, et al. Coordinate overexpression of interferon-alpha-induced genes in systemic lupus erythematosus. *Arthritis and rheumatism*. 2004; 50(12):3958–67. [published Online First: Epub Date]. DOI: 10.1002/art.20798 [PubMed: 15593221]
25. Rice LM, Padilla CM, McLaughlin S, et al. Decreased biomarker and clinical improvement in systemic sclerosis after fresolimumab. *Journal of Clinical Investigation*. 2015 in press.
26. Stifano G, Affandi AJ, Mathes AL, et al. Chronic Toll-like receptor 4 stimulation in skin induces inflammation, macrophage activation, transforming growth factor beta signature gene expression, and fibrosis. *Arthritis research & therapy*. 2014; 16(4):R136. [published Online First: Epub Date]. doi: 10.1186/ar4598 [PubMed: 24984848]
27. Montoro-Garcia S, Shantsila E, Lip GY. Potential value of targeting von Willebrand factor in atherosclerotic cardiovascular disease. *Expert opinion on therapeutic targets*. 2014; 18(1):43–53. [published Online First: Epub Date]. DOI: 10.1517/14728222.2013.840585 [PubMed: 24073984]
28. Wiseman S, Marlborough F, Doubal F, et al. Blood markers of coagulation, fibrinolysis, endothelial dysfunction and inflammation in lacunar stroke versus non-lacunar stroke and non-stroke: systematic review and meta-analysis. *Cerebrovascular diseases*. 2014; 37(1):64–75. [published Online First: Epub Date]. DOI: 10.1159/000356789 [PubMed: 24401164]

Figure 1. Genes clustering with components of the four-gene biomarker

Genes selected from genes clustering with COMP/THBS1 (top panel, see Fig. 1s for all genes in this cluster) and SIGLEC1 (bottom panel, see Fig. 2s for all genes in this cluster) after unsupervised clustering of genes and subjects after selection of genes showing a correlation with the MRSS ≥ 0.4 . The MRSS added into the gene set before normalization and clustering appears in the heatmap of the COMP/THBS1 cluster, with the numerical MRSS values show in "SKIN SCORE" between the two clusters. Four-gene biomarker genes are highlighted in yellow (COMP, THBS1, and SIGLEC1). Subjects (25 SSc patients and 4 healthy controls) analyzed are shown at the top of the figure. Red indicates up-regulated expression, green indicates down-regulated expression.

Arthritis Rheumatol. Author manuscript; available in PMC 2017 July 21.

Figure 4. Relationship between expression of biomarker genes and the MRSS

Correlations (Pearson's) between nanostring mRNA skin gene expression and the MRSS in skin biopsies from patients with dcSSc (n=63). Values are stratified according to disease duration at time of biopsy; blue squares ≤24 months, pink squares ≤36 months, and open circles ≥37 months. Data are shown as R-values.

Figure 5. Comparisons of gene expression in longitudinal skin biopsies from dcSSc patients with MRSS

Panel A shows the pairwise correlation matrix of clinical characteristics and absolute change in mRNA gene expression from baseline. The color indicates direction of correlation (red= negative, blue= positive); the magnitude of the correlation is indicated by the percentage of the filled circle and intensity of shading. Variables are ordered using principal components analysis. Panel B shows the significance of association cross-sectionally and longitudinally of baseline gene expression and change in gene expression with MRSS values indicated in each graphic are p-values with beta coefficients and confidence intervals in Supplemental Table 5. Each color represents an individual patient followed over multiple visits.

Figure 6. Testing and validation of 2GSSc and Weighted Models

Graphs show correlations between 2GSSc biomarker score derived using the 2GSSc Model equation (panels A: $p < 0.0001$ and B: $p < 0.0001$) or the WSSc biomarker score using the Weighted Model equation (panels C: $p = 0.0002$ and D: $p < 0.0001$) and the clinically assessed MRSS. Values used to develop the models are from Boston Medical Center patients (panels A and C). Values used to validate the models are from Hospital for Special Surgery patients (panels B and D). R-values are indicated on each panel. Panel E. Histological features of skin were scored and compared to the MRSS, the 2GSSc biomarker score, the WSSc biomarker score, or genes making up the biomarker scores (THBS1, ADAM12, CCL2, CD163, CTGF, MS4A4A, and WIF1). Correlations between histological fibrosis and inflammation are charted in the corrgram with R-values as shown. The extent and shading of blue indicates the degree of positive correlations, the extent and shading of red the degree of negative correlations.