Description of Datasets

Discovery Datasets

UCSF1 Cohort (Skin; GSE9285): Milano *et al.* (1) profiled forearm and back skin biopsies from adult patients with diffuse scleroderma (dSSc), limited scleroderma (lSSc), morphea, eosinophilic fasciitis as well as from healthy controls. For our discovery analysis, we removed morphea and eosinophilic fasciitis samples.

Boston Cohort (Skin; GSE32413): Pendergrass *et al.* (2) profiled 72 forearm and back biopsies from adult dSSc patients enrolled in an open-label trial of rituximab, from patients not treated with rituximab, and from healthy controls. We used all biopsies in our discovery analysis.

Validation Datasets

Houston Cohort (Skin; GSE58095): Assassi *et al.* (3) profiled adult SSc patients enrolled in the GENISOS cohort or an open label imatinib study. This dataset only included baseline pre-treatment samples. We used all biopsies in our discovery analysis.

HSS Cohort (Skin; GSE65405): Gordon *et al.* (4) profiled forearm biopsies from adult SSc patients enrolled in an open-label, pilot clinical trial of nilotinib. We included all biopsies as validation in our analysis.

Northwestern Cohort (Skin): Hinchcliff *et al.* profiled forearm and back biopsies from patients with dSSc as well as from healthy control patients. SSc patients were treated with mycophenolate mofetil. Each SSc patient was profiled at up to 5 time points: baseline, 6-, 12-, 24-, and 36-months. Patients with other diseases (morphea, rheumatoid arthritis, plasma cell dyscrasia-associated scleroderma, and mixed connective tissue disease) were removed from the analysis.

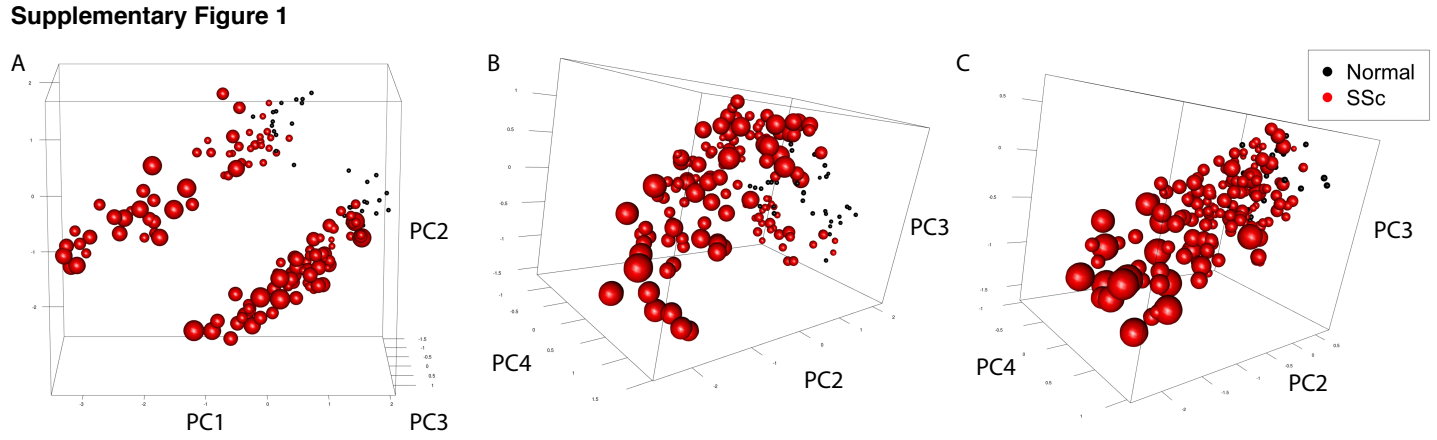
Stanford Cohort (Skin): Fiorentino *et al.* profiled forearm skin biopsies obtained from adult dSSc or lSSc patients as well as from healthy controls. All biopsies were included in our validation analysis.

UCSF2 Cohort (Skin): Boin *et al.* profiled forearm skin biopsies obtained from adult dSSc or lSSc patients enrolled into the UCSF Scleroderma Center cohort as well as from healthy controls. All biopsies were included in our validation analysis.

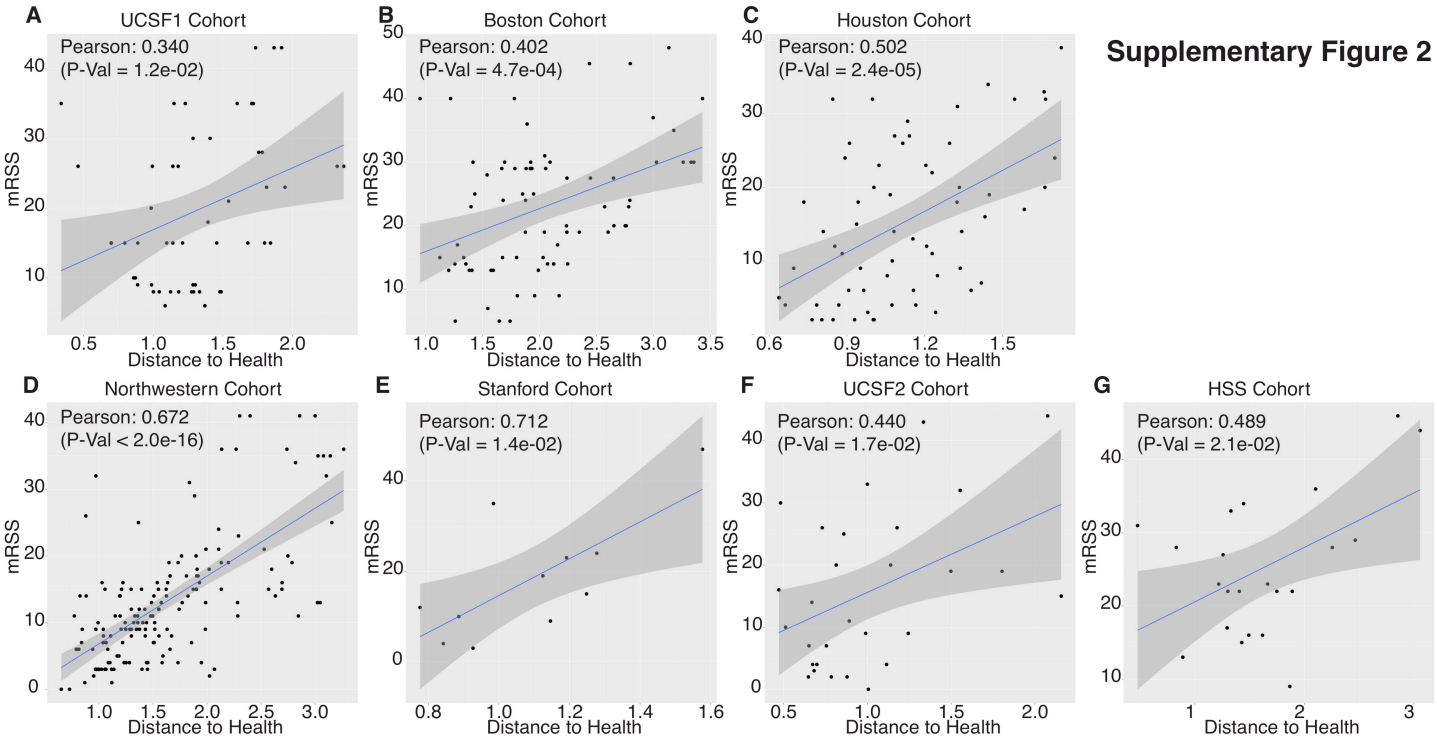
References

1. Milano, A. *et al.* Molecular Subsets in the Gene Expression Signatures of Scleroderma Skin. *PLoS ONE* **3,** e2696 (2008).
2. Pendergrass, S. A. *et al.* Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. *Journal of Investigative Dermatology* **132,** 1363–1373 (2012).
3. Assassi, S. *et al.* Dissecting the Heterogeneity of Skin Gene Expression Patterns in Systemic Sclerosis. *Arthritis Rheumatol* (2015). doi:10.1002/art.39289
4. Gordon, Jessica K., et al. Nilotinib (Tasigna™) in the treatment of early diffuse systemic sclerosis: an open-label, pilot clinical trial. *Arthritis research & therapy* 17.1 (2015): 1-14.

Supplementary Figure Legends

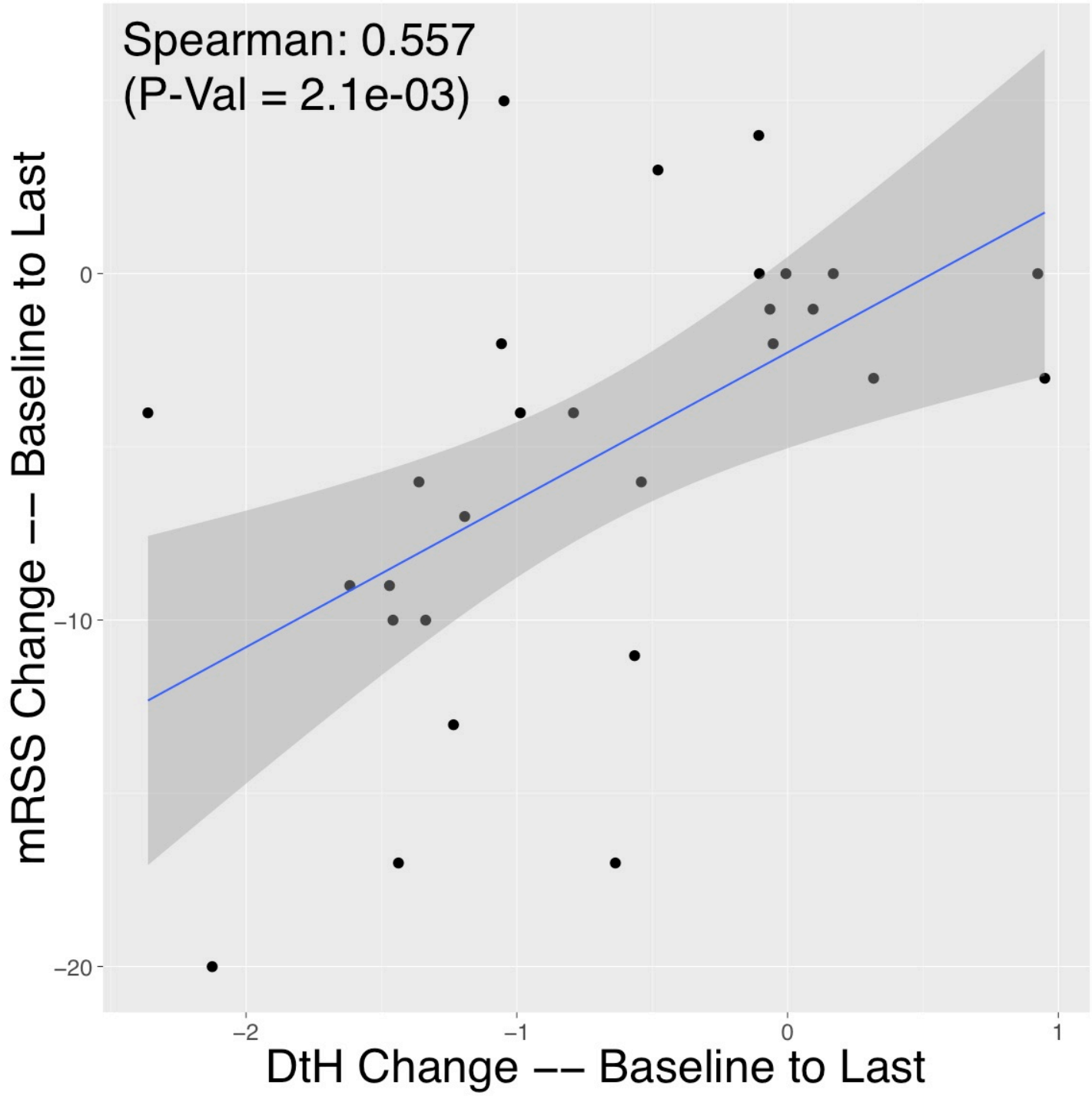


Supplementary Figure 1: PCA of the discovery cohort. We performed the PCA using the 415 genes in the SSc signature. Each sphere represents a skin biopsy (SSc - red, healthy control – black). Sphere radius is proportional to patient mRSS at the time of biopsy. See Supplementary Table 1 for number of samples. (A) Discovery Cohorts plotted in the first three PCs demonstrating batch effects between them. (B) Discovery cohorts and (C) Northwestern cohort plotted using PC2, PC3 and PC4.



Supplementary Figure 2: Distance to Health (DTH) is significantly correlated with mRSS across all datasets. Correlation plots comparing mRSS with DTH for all of the datasets. (A-B) Discovery cohorts (N = 54 and 72). (C-G) Validation cohorts (N = 66, 161, 13, 29, 22). Each dot represents a patient sample. The x-axis represents a DTH for a patient skin biopsy. The y-axis represents mRSS at the time the biopsy was taken. The blue line is the line of best fit and the grey region represents its 95% CI.

Supplementary Figure 3

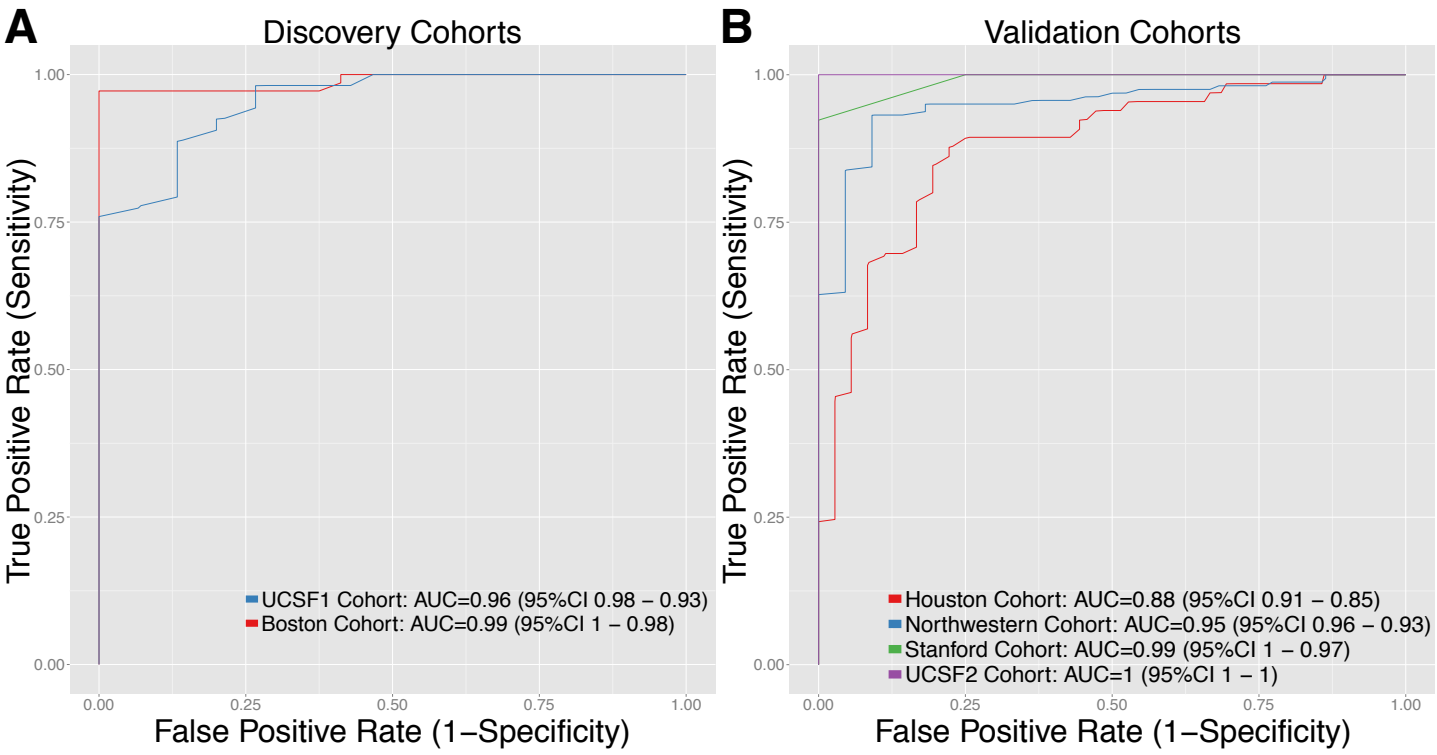


Supplementary Figure 3: Longitudinal changes in DTH correlate with changes in mRSS. Correlation plot comparing changes in DTH with contemporaneous changes in mRSS in the Northwestern Cohort. Each point represents a patient. The x-axis represents the patient's change in DTH from their baseline skin biopsy to their last biopsy. The yC axis represents the change in mRSS from baseline visit to the last visit for each patient. The blue line is the line of best fit and the grey region represents its 95% CI. N = 28.

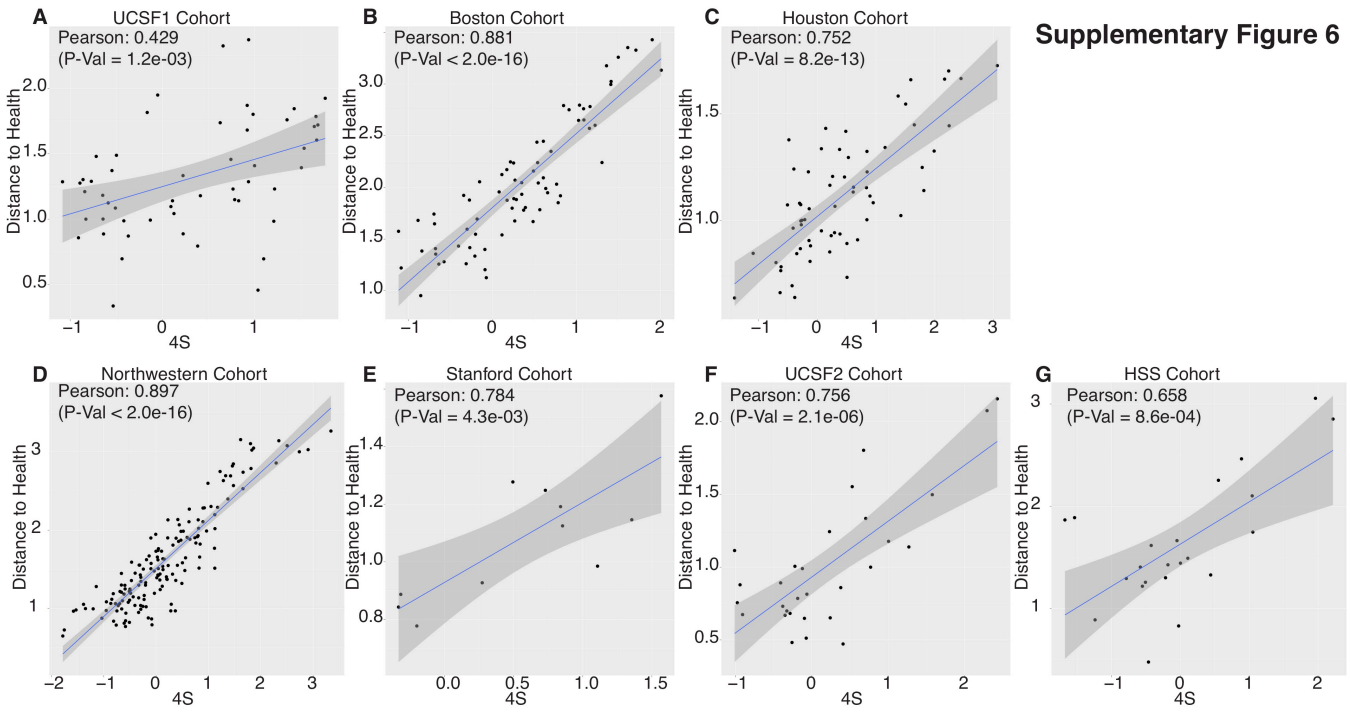
Supplementary Figure 4 (separate file): PCA of each patient in the Northwestern Cohort: Each sphere represents a skin biopsy from the Northwestern Cohort, where the size of a sphere size is proportional to mRSS and color indicates its assigned intrinsic subset (black, healthy control; green, normal-like; blue, limited; red, fibroproliferative; purple, inflammatory). The green region represents the health bubble. The golden

sphere indicates the centroid of healthy controls. Numbers in brackets represent the mRSS score at a given time point for a given patient. Each figure (N = 183 for each) displays the same set of samples with different patient trajectory highlighted.

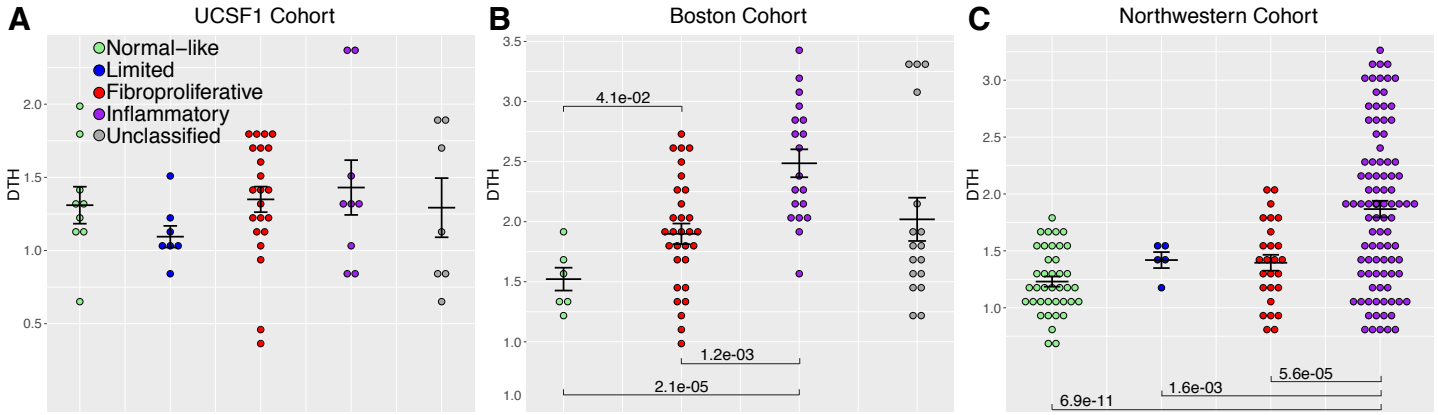
Supplementary Figure 5



Supplementary Figure 5: 4S accurately distinguishes healthy and diseased biopsies. (A) ROCs for distinguishing SSc patients from healthy controls in (A) the discovery and (B) validation cohorts. We excluded the HSS Cohort because it lacked healthy control biopsies. See Supplementary Table 1 for the number of case and control samples in each group.



Supplementary Figure 6: 4S and DTH are high correlated across all cohorts. Correlation plots comparing 4S with DTH for all of the datasets. (A-B) Discovery cohorts. (C-G) Validation cohorts. Each dot represents a patient skin biopsy. The X-axis represents a DTH for a patient biopsy. The Y-axis represents mRSS at the time of biopsy. The blue line is the line of best fit and the grey region represents its 95% CI.



Supplementary Figure 7: Comparison of DTH across intrinsic subsets. Beeswarm dot plots (mean ± SEM) comparing DTH across intrinsic subsets in (A) the UCSF1 Cohort, (B) the Boston Cohort, and (C) the Northwestern Cohort. Bars between groups represent FDR corrected p-values (q values) from Student's unpaired, two-sided T Tests of 4S between the intrinsic subsets. Bars only shown when q value < 5%. Unclassified samples were not assigned to any intrinsic subsets in the original publications. (A) N = 9, 7, 22, 9, and 7; (B) N = 6, 0, 30, 19, and 17; (C) N = 39, 5, 26, 91, and 0 for normal-like, limited, fibroproliferative, inflammatory, and unclassified, respectively.