Supplementary Tables and Figures

Supplementary Figure 1: Overview of study design. Schematic showing the timing and types of samples collected

Supplementary Figure 2: Analysis of markers of subtypes of collagen synthesis and correlation with MRSS. A) Median serum Pro-C3 across the BIOPSY cohort (TUKEY post-hoc p values <0.05 included). B) Correlation of Pro-C3 and MRSS. C) Median serum Pro-C6 across the BIOPSY cohort (TUKEY post-hoc p values <0.05 included). D) Correlation between Pro-C6 and MRSS in all scleroderma subsets. E) Median C3 fibrotic index across the BIOPSY cohort (TUKEY post-hoc p values <0.05 included). F) Correlation between C3 fibrotic index and MRSS in all scleroderma subsets. G) Median C6 fibrotic index across the BIOPSY cohort (TUKEY post-hoc p values <0.05 included). H) Correlation between C6 fibrotic index and MRSS. I) Correlation between ELF test and MRSS J) Correlation between TIMP1 and MRSS. K) Correlation between PIIINP and MRSS. L) Correlation between HA and MRSS.

Supplementary Figure 3: A) Serum analytes over time represented as group proportion change from baseline, proportion change from baseline by autoantibody status, and proportion change from baseline by skin status. Pro-C3, C3M, Pro-C6, C6M, Pro-C4 and C4M represented. ANOVA performed at each time point. * p<0.05, ** p<0.01, *** p<0.001. B) Serum analytes over time represented as group proportion change from baseline, proportion change from baseline by autoantibody status, and proportion change from baseline by skin status. Pro-C1, C3 fibrotic index, C6 fibrotic index, C4 turnover, MCP-1, and IL-6 represented. ANOVA performed at each time point. * p<0.05, ** p<0.01, *** p<0.001.

Supplementary figure 4: A) and B) PCA and hierarchical clustering from skin differentially expressed genes comparing HC (turquoise) and early dcSSc (red). Hierarchical clustering includes 491significantly different genes (FDR <0.001). C&D) PCA and hierarchical clustering from differentially expressed genes from skin of early dcSSc subgroup at baseline comparing autoantibody subtypes. 384 genes included; p<0.01. (red = ARA, green= ATA, blue = other). E and F) PCA and hierarchical clustering from blood samples of early dcSSc patients at baseline comparing autoantibody subtypes. 222 genes included; p<0.02.

Supplementary Figure 5: PCA plots of longitudinal analysis of gene expression from early dcSSc patients. A) PCA plot from gene expression of early dcSSc skin samples of longitudinal paired samples showing baseline and 12m, and B) baseline and 3m. C) PCA plot from gene expression of early dcSSc blood samples of longitudinal paired samples showing baseline and 12m, and D) baseline and 3m. Notable overlap of gene expression seen between paired samples.

Supplementary Figure 6: Subgroup analysis of single sample GSEA of all significantly differentially expressed KEGG pathways A) ssGSEA of all significantly differentially expressed KEGG pathways comparing early dcSSc (red) and HC (turquoise). B) ssGSEA of all significantly expressed KEGG pathways between early dcSSc ARA and HC (red), and early dcSSc ATA and HC (green). C) sssGSEA of all Hallmark pathways significantly elevated in early dcSSc ATA (red) and HC and early dcSSc ATA (green) with HC.

Supplementary Figure 7: Schematic of our main results, providing a potential understanding between clinical phenotypes by autoantibody subtype, and molecular differences by autoantibody subtype. The Venn diagram to demonstrate number of common Hallmark pathways between ARA

Clark KEN, *et al. Ann Rheum Dis* 2021; 80:1584–1593. doi: 10.1136/annrheumdis-2021-220402

against HC (red) and ATA against HC (green), as well as number of pathways specific to each comparison.

Supplementary Figure 8

High definition figure 4C: Unsupervised hierarchical clustering of all baseline BIOPSY cohort skin samples based on 731 differentially expressed genes (FDR <0.001). Disease subtype indicated by colour bar (red=early dcSSc, green = established dcSSc, purple = lcSSc, turquoise = HC).

Supplementary Figure 9

High definition figure 4D: Hierarchical clustering based on 61 significantly differing gene expressions (FDR <0.1) from skin comparing ARA (red) and ATA (green) positive early dcSSc.

Supplementary Figure 10

Hierarchical clustering of single sample GSEA (ssGSEA) utilising significantly differentially expressed pathways. A) ssGSEA of significantly differentially expressed KEGG pathways across whole SSc spectrum and healthy controls (colour bar). SSc patient subgroups highlighted with early dcSSc (red), established dcSSc (green), lcSSc (purple), and HC (turquoise).

Supplementary Figure 11

High definition Figure 6C) Cleveland dot plot demonstrating the normalised enrichment score for Hallmark pathways in ARA compared to HC (red) and ATA compared to HC (green), and significance of pathway

Supplementary Table 1: Median concentration of serum analytes across each disease subset, and p value based on ANOVA of log transformed results. Significant p values highlighted in bold

Supplementary Table 2: Correlation between serum proteins and baseline MRSS of all SSc patients. Significant p values highlighted in bold (<0.05)

Supplementary Table 3: List of 191 differentially expressed genes (FDR <0.15) between early dcSSc ARA patients, and early dcSSc ATA patients (see supplementary EXCEL spreadsheet)

Clark KEN, *et al. Ann Rheum Dis* 2021; 80:1584–1593. doi: 10.1136/annrheumdis-2021-220402