SUPPLEMENTARY METHODS

Study Design and Patient Recruitment

This study was a single-centre, prospective observational study comprising of four distinct participant cohorts: early dcSSc (<5 years disease duration), established dcSSc (>5 years disease duration), lcSSc and healthy volunteers. Recruitment for each cohort occurred in parallel and ceased when the recruitment target for that cohort was reached (supplementary Figure 1).

This study received ethical approval from the NHS Research and Ethics Committee (REC number 6398). The work was performed within the strict General Data Protection Regulations (GDPR) compliant framework of UCL. All participants in this study provided informed consent for their participation, and for the use of their clinical data and samples for research purposes.

Patients with SSc were diagnosed in accordance with the 2013 American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) classification criteria [32] and described as diffuse or limited in terms of skin distribution according to LeRoy 1988 [33]. Clinical information for each patient was recorded prospectively at each visit, and stored on a secure network. MRSS was assessed at the time of sample collection by a well-trained observer with consistent scoring techniques. Only 2 observers were involved in recruitment, and both have been trained in MRSS.

All patients had routine blood tests done including serum and plasma parameters and autoantibody immunoblot.

The early dcSSc cohort were reviewed every 3 months over a 12-month period.

Sample collection

For everyone recruited into the study, serum, plasma and PAXtube blood, suction blister fluid and skin biopsy samples were obtained.

Skin Biopsy technique

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

Paired 4mm skin biopsies were obtained from the forearm of subjects. 1 biopsy was stored in 10% formalin and processed for histological purposes, while the other was stored in RNA*later*™ and stored at -80°C.

Serum protein analysis

Markers of collagen synthesis and degradation were analysed in the serum using Nordic Bioscience's biomarker assays (https://www.nordicbioscience.com/). Ten variables were analysed including: Pro-C3, Pro-C6, C3M, C6M, C4M2, VICM, P4NP7S, PINP, and C1M.

Serum levels of B-natriuretic peptide (BNP), TIMP-1 (Tissue inhibitor of metalloproteinases 1), PIIINP (collagen III synthesis), and hyaluronic acid (HA), components of the Enhanced Liver Fibrosis (ELF) test, were all analysed by The Doctors Laboratory (https://www.tdlpathology.com/).

Serum levels of monocyte chemoattractant protein-1 (MCP-1), oncostatin M (OSM), interleukin 6 (IL-6), and C-reactive protein (CRP) were determined by the GSK laboratories in Stevenage using MSD (Meso Scale Discovery) immunoassays.

RNA sequencing

RNA expression analysis was performed on skin samples stored in RNA*later*™ and blood samples collected and stored in PAXtubes. This was carried out by Epistem Ltd. RNA was isolated in 2 batches per tissue type, and RNAseq was run in batches of 20 samples on the Illumina NextSeq 550 (https://www.epistem.co.uk/).

Statistical Analysis

Statistical analysis was carried out using the software R (version3.6). Kruskal-Wallis with post-hoc Mann-Whitney U test or ANOVA with Tukey post-hoc analysis was used where appropriate. Where the level of response variable was below the threshold of detectability, the lowest detectable level was assigned. Spearman's rank correlation was used to assess correlations. The prospective cohort

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

were assigned the status of “improver”, “progressor” or “stable” based on change in MRSS of greater than or equal to 4 points, AND ≥ 20% change from baseline at the 12-month timepoint.

For the RNAseq results, normalised FPKM (fragments per kilobase of transcript per million) values were obtained using rlog() function within DESeq2 of Rsoftware. Differential gene expression was measured with the Bioconductor limma software, and cluster analysis was performed using the Rpackages ggplot2, heatmap.plus and edgeR. Significantly differentially expressed genes were selected as median FPKM ≥ 1 AND fold change (FC) ≥ 1.5 or ≤ 0.68 AND adjusted p ≤ 0.05 (FDR, Benjamini-Hockberg correction). Where more than 2 groups were analysed, ANOVA was performed, and differentially expressed genes selected with median FPKM ≥ 1 AND adjusted p ≤ 0.05.

GSEA was performed using GSEA 4.0.3[34, 35]. KEGG® pathways and Hallmark® pathways were chosen for pathway analysis. Significantly differentially expressed pathways were selected based on FDR <0.1. Where large numbers of genes remained following selection via these parameters, more stringent FDR thresholds were chosen to allow for improved visual representation on the heat map. Single sample GSEA (ssGSEA)[36] was performed against the significantly differentially expressed pathways identified using GSEA.

DcSSc disease signatures were obtained from GSE95065 [37], E-MEXP-1214, GSE58095 [13], GSE9285 [14], GSE125362 [38] and GSE76886 [22]. Affymetrix data were normalized using justRMA in affy R package. For the two-channel Agilent data, normexp background correction, within-array loess normalization and between-array quantile normalization were performed using limma. For Illumina BeadChip samples, *neqc* from limma was used in background correction and quantile normalization. Finally, differential expression was performed using linear model fit in limma. DcSSc gene signatures were identified by selecting genes with adjusted p < 0.1 and FC ≥ ± 1.5 in ≥ 3 comparisons out of 8 comparisons (comprised of the 6 listed studies). GSVA R package was used to calculate the enrichment scores of each signature.

Patient and Public involvement

The study was designed without public involvement. All patients and healthy volunteers gave informed consent for participation. The results of this work will be disseminated via relevant patient conferences.

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