

Molecular basis for clinical diversity between autoantibody subsets in diffuse cutaneous systemic sclerosis

Authors:

Kristina E.N. Clark¹, Corrado Campochiaro¹, Eszter Csomor², Adam Taylor², Katherine Nevin², Nicholas Galwey², Mary A. Morse², Jennifer Singh², Yee Voan Teo², Voon H. Ong¹, Emma Derrett-Smith¹, Nicolas Wisniacki², Shaun M Flint², Christopher P. Denton^{1*}

Affiliations

¹UCL Centre for Rheumatology and Connective Tissue Diseases, London, U.K.

²GlaxoSmithKline, Stevenage, U.K.

***Corresponding author:**

Christopher Denton PhD FRCP
Professor of Experimental Rheumatology
UCL Centre for Rheumatology and Connective Tissue Diseases
2nd Floor – UCL Medical School Building
Royal Free Campus
Rowland Hill Street
London NW3 2PF,
U.K.

Tel +44 (0) 20 7794 0432
c.denton@ucl.ac.uk

Abstract

Objectives: Clinical heterogeneity is a cardinal feature of systemic sclerosis (SSc). Hallmark SSc autoantibodies are central to diagnosis and associate with distinct patterns of skin and organ-based complications. Understanding molecular differences between patients will benefit clinical practice and research and give insight into pathogenesis of the disease. We aimed to improve understanding of the molecular differences between key diffuse cutaneous SSc subgroups as defined by their SSC specific autoantibodies

Methods: We have used high dimensional transcriptional and proteomic analysis of blood and skin in a well characterised cohort of SSc (n=52) and healthy controls (n=16) to understand the molecular basis of clinical diversity in SSc and explore differences between the hallmark ANA reactivities.

Results: Our data define a molecular spectrum of SSc based upon skin gene expression and serum protein analysis, reflecting recognised clinical subgroups. Moreover, we show that anti-topoisomerase-1 (ATA) and anti-RNA polymerase III (ARA) specificities associate with remarkably different longitudinal change in serum protein markers of fibrosis, and divergent gene expression profiles. Overlapping and distinct disease processes are defined using individual patient pathway analysis.

Conclusions: Our findings provide insight into clinical diversity and imply pathogenetic differences between ANA based subgroups. This supports stratification of SSc cases by ANA antibody subtype in clinical trials and may explain different outcomes across ANA subgroups in trials targeting specific pathogenic mechanisms.

Keywords systemic sclerosis, transcriptomics, autoantibody,

INTRODUCTION

Systemic sclerosis (SSc) patients are characterised by antinuclear autoantibodies (ANA), including anti-topoisomerase-1 (ATA, Scl-70), anti-centromere (ACA) or anti-RNA polymerase III (ARA)[1].

Different major organ-based complications link with ANA autoantibodies. For example, ATA is associated with significant interstitial lung fibrosis [1, 2], while ARA carries a ten-fold increased risk of scleroderma renal crisis [3]. These strong associations with specific disease manifestations suggests that there are pathobiological differences beyond ANA underlying diverse clinical outcomes.

Skin and blood are readily accessible to compare gene and protein expression in SSc subgroups to better understand molecular correlates of clinical phenotypes. Skin analysis may be especially informative to understand differences between ANA subgroups because skin changes over time have been linked to ANA reactivities. ARA generally has a higher peak skin score than ATA in early dcSSc but faster improvement whereas ATA may show slower regression [4, 5].

With the objective of understanding the molecular basis for heterogeneity in SSc, we undertook a detailed longitudinal analysis of skin and blood samples from a cohort of early stage dcSSc followed over 12 months. This included measurement of serum proteins reflecting pathogenesis or extracellular matrix turnover and with genome wide assessment of gene expression. To put our findings in the broader context we also studied late stage dcSSc and lcSSc and have compared our findings with matched healthy control subjects. We have specifically tested the hypothesis that hallmark ANA autoantibodies specific to SSc associate with different patterns of gene expression and proteins reflecting fundamental differences in pathogenesis in dcSSc. Our results strongly suggest that ANA specificity defines distinct biologic subgroups within SSc with implications for case selection for clinical trials and therapeutic strategies in clinical practice.

METHODS

This was a single centre, prospective observational study comprising of four distinct participant cohorts: early dcSSc (<5 years duration), established dcSSc (> 5 years duration), lcSSc and healthy volunteers (HC). Blood samples for serum, plasma and in PAXtubes were collected with concomitant 4mm skin biopsies in RNALater

The early dcSSc cohort were reviewed every 3 months for a 12 month period, with blood sample collection and clinical assessment at each visit, and a 4mm skin biopsy at baseline, month 3 and month 12.

Serum was analysed for soluble markers associated with collagen synthesis and degradation, and fibrosis, including the constituents of the Enhanced Liver Fibrosis (ELF) test. RNA expression analysis was performed on skin and whole blood.

Additional methodology is described in the supplementary material.

Statistical analysis:

The prospective cohort were assigned the status of “improver”, “progressor” or “stable” based on change in MRSS of greater than or equal to 4 points, AND $\geq 20\%$ change from baseline at the 12-month timepoint. For soluble markers, ANOVA with Tukey post-hoc analysis or Kruskal Wallis with post hoc Mann Whitney U test were used. The Benjamini Hochberg False Discovery Rate was used for multiple comparisons. All statistical analysis was performed using the software R. GSEA was used for pathway analysis.

Patient involvement

Patients and healthy volunteers provided informed consent and attended visits as part of routine care or for purposes of research sampling.

RESULTS

Patient characteristics

The BIOPSY (**B**IOlogical **P**henotyping of diffuse **S**ystemic sclerosis) data set was generated to provide a platform for the integrated analysis of skin and blood samples, together with detailed clinical phenotyping (supplementary figure 1). The study recruited 52 SSc patients (21 early dcSSc, 15 established dcSSc, 16 lcSSc) and 16 gender-matched HC to the early dcSSc cohort. 36 (69%) of the SSc patients are female. Baseline characteristics are summarised in Table 1. Mean disease duration in the early dcSSc cohort was 24 months (sd 12 months). ANA frequency in BIOPSY reflected the overall dcSSc population: ATA n= 14 (27%), anti-ARA n= 12 (23%) and other n=26 (50%), which is aligned with those of other recent large SSc cohorts [1, 5].

One patient died during the study period from cardiac complications. These cases were managed in line with current treatment guidelines in the UK [6]. As expected, all patients with early dcSSc were on immunosuppression by 12 months; most often (85%) including mycophenolate mofetil (MMF). Median baseline skin score (MRSS) for early dcSSc was 18 (IQR 19). At a group level MRSS peak was 21 (22) at 3 months and fell to 16 (14.25) at 12 months (Figure 1). The median MRSS for the established dcSSc patients was 10 (6.5) and for the lcSSc was 4 (1.25). Lower skin scores were seen in subjects with more established disease of greater than 36 months duration, and in cases of early disease with less than 20 months duration. There was no significant relationship between disease duration and baseline MRSS ($r=0.133$, $p=0.575$).

For around half of the BIOPSY cohort MRSS was clinically stable over 12 months. The remaining cases split between those that are significantly worsening (n=4) and those that show clinically significant improvement (n=5). Prospective dcSSc cases were divided into the three most recognised ANA based subgroups, namely ARA, ATA or “other” for the purposes of analysis (which includes ANA positive, ENA (extractable nuclear antigen) negative, or alternative ENAs). Group level change in MRSS for

the ANA subgroups is shown in Figure 1. There was equal distribution of autoantibody subsets (specifically ATA and ARA) in each of the skin trajectory cohorts.

	early dcSSc at baseline	early dcSSc at 12months	Established dcSSc	LcSSc	HC
Total (n)	21	20	15	16	16
F (%)	12 (57.1)	11 (55)	12 (80)	12 (75)	9 (56.3)
Age (yrs)	52 (23-75)	54 (24-76)	56.9 (24-73)	52.5 (27-85)	43.3 (28-81)
Disease duration (yrs)	1.75 (0.5-4.9)	2.6 (1.5-5.9)	13 (5-20.8)	9 (3.5-30.4)	
MRSS	18 (7-39)	16(2-38)	10 (2-36)	4 (2-12)	
Antibody					
ATA (%)	8 (38.1)	7 (35)	4 (26.7)	2 (12.5)	
ARA (%)	6 (28.6)	6 (30)	6 (40)	0	
ACA (%)	0	0	0	10 (62.5)	
ANA neg (%)	2 (9.5)	2 (10)	1 (6.7)	1 (6.3)	
Other (%)	5 (23.8)	5 (25)	7 (46.7)	3 (18.8)	
Organ involvement					
Lung (%)	6 (28.6)	7 (35)	8 (53.3)	0	
Muscle (%)	6 (28.6)	6 (30)	1 (6.7)	0	
Kidney (%)	3 (14.3)	4 (20)	1 (6.7)	0	
PAH (%)	1 (4.8)	1 (5)	1 (6.7)	0	
Cardiac (%)	3 (14.3)	3 (15)	1 (6.7)	1 (6.3)	
GI (%)	3 (14.3)	3(15)	4 (26.7)	1 (6.3)	
Overlap conditions					
RA (%)	1 (4.8)	3 (15)	0	1 (6.3)	
PM/DM (%)	6 (28.6)	6 (30)	3 (20)	0	
Sjogrens (%)	0	0	1 (6.7)	2 (12.5)	
Immunosuppression					
MMF (%)	9 (42.9)	17 (85)	9 (60)	0	
MTX (%)	7 (33.3)	5 (25)	2 (13.3)	3 (18.8)	
HCQ (%)	5 (23.8)	5 (25)	1 (6.7)	5 (31.3)	
Azathioprine (%)	1 (4.8)	0	0	0	
Tocilizumab (%)	1 (4.8)	3 (15)	0	0	
Cyclophosphamide (%)	1 (4.8)	0	0	0	
IvIG (%)	0	2 (10)	1 (6.7)	0	
Untreated (%)	3 (14.3)	0	5 (33.3)	9 (56.2)	

Table 1: Clinical and demographic features of the BIOPSY cohort. Results presented as median and range unless otherwise stated. Abbreviations: F (female), MRSS (modified Rodnan skin score), ATA (anti-

topoisomerase antibody), ARA (anti-RNA polymerase III antibody), ACA (anti-centromere antibody), ANA (antinuclear antibody), PAH (pulmonary arterial hypertension), GI (gastrointestinal), RA (rheumatoid arthritis), PM (polymyositis), DM (dermatomyositis), MMF (mycophenolate mofetil), MTX (methotrexate), HCQ (hydroxychloroquine), IvIG (intravenous immunoglobulin).

There was no significant difference between group level skin score change between different immunosuppressive treatments, or between those that were already on immunosuppression and those started during the first 3 months of follow up.

Differential longitudinal change in serum protein markers between ANA subgroups

Baseline serum protein marker analysis

At baseline, markers of collagen synthesis discriminated early dcSSc from HCs (supplementary table 1 and supplementary figure 2). Composite fibrotic indices (C3 fibrotic index and C6 fibrotic index) did not outperform the markers of protein synthesis. The ELF composite score discriminated early dcSSc from HCs and was driven largely by PIIINP and TIMP1 (figure 2).

Longitudinal serum protein marker analysis in the early dcSSc cohort

Longitudinal changes in serum proteins over 12 months in early dcSSc in serum proteins explored differences based on skin score trajectory and ANA defined subgroups.

Only ProC1 showed association longitudinally with skin progression (Figure 3). There were consistent and remarkable differences in the change in serum proteins between the major ANA based subgroups (Figure 3 and supplementary Figure 3). This was most evident for ELF, and the 3 constituents (PIIINP, HA, TIMP1) and pro-C1, where there is a linear increase overall for both ARA and “other” groups whereas ATA shows decline over time from baseline values.

Integrated transcriptional analysis of skin

Baseline transcriptional analysis of skin

To better understand the molecular basis for longitudinal clinical and serum protein differences between subgroups of SSc, a detailed analysis of global gene expression was undertaken across the BIOPSY cohort for skin and blood RNA.

There was clear differentiation between early dcSSc and HC by principal component analysis (PCA), and unsupervised clustering of significantly differentiated genes (731 genes, FDR<0.001) on baseline samples (Figure 4B and 4C), with established dcSSc and lcSSc having a more similar transcriptional phenotype in skin.

A focused analysis of early dcSSc and HC baseline skin biopsy samples identified 491 differentially expressed genes (FC \geq 1.5 and FDR< 0.001) that separated these subpopulations and indicating a very distinct molecular signature shared by most cases of early dcSSc (supplementary Figure 4A and 4B).

Next, we explored differences in skin gene expression within the early dcSSc patients based upon ANA status. PCA and unsupervised cluster analysis of differentially expressed genes (n=384, p<0.01, FDR 0.4) clearly separating ARA and ATA patients (supplementary Figure 4C and 4D) with “other ANA” patients being intermediate between ARA and ATA in some cases.

Analysis of ARA and ATA early dcSSc patients revealed 61 differentially expressed genes at baseline (FC \geq 1.5, FDR<0.1) that fully differentiated these ANA subgroups (Figure 4D and supplementary table 3). These include genes previously associated with fibrosis and SSc showing significant difference between autoantibodies within the early dcSSc subgroup, and across the whole SSc spectrum. Examples include INHBA (inhibin subunit beta A), IL6ST (interleukin 6 signal transducer), APLN (Apelin), and C6 (complement 6) (Figure 5D-G).

Similar analysis was performed on whole blood baseline samples, although we could not identify any genes that would directly differentiate ARA+ and ATA+ cases (supplementary figure 4E and 4F).

Longitudinal transcriptional analysis of paired early dcSSc samples at 3 months and 12months

Longitudinal sampling of the early dcSSc cases at 3 and 12 months showed stability of the gene expression profiles in skin and blood over time (supplementary Figure 5) suggesting that gene expression-based classification is a robust assessment that changes relatively little at a global level over 12 months.

Systemic sclerosis specific gene expression in skin shows relevant changes across the disease spectrum

To compare our findings with previously reported SSc associated gene expression signatures in skin we used a robust SSc associated composite signature of SSc specific genes identified from publicly available gene expression datasets for whole skin[7-11]. Our analysis replicated this SSc associated signature across different time points for the BIOPSY cohort of early dcSSc and showed consistent relevant differences across the BIOPSY cohort for both upregulated and downregulated SSc signature scores (Figure 4A). Both the upregulated and downregulated genes of the SSc signature showed differences from healthy controls for all SSc subgroups. The global differences reflected a spectrum of the disease, with greatest difference observed in the baseline early dcSSc and least in established dcSSc and lcSSc. Notably, the signature became attenuated at later time points in the longitudinal cohort, and in late stage dcSSc and lcSSc, in contrast to the relative stability of overall gene expression in BIOPSY for individual patients. This suggests that the global expression signature of SSc reflects stage and severity of skin disease. Overall, the composite disease associated signature analysis provides strong external validation of our cohort compared to other datasets although likely to be less informative about patient level MRSS change than our analysis of the prospectively collected and rigorously phenotyped BIOPSY dataset.

Differences in gene expression for ARA and ATA positive dcSSc compared with healthy controls

To explore similarities and differences between gene expression profiles for the two major ANA antibody subtypes of early dcSSc, we compared the baseline differences between ARA and ATA subgroups and HC in skin. In skin 664 and 903 differentially expressed genes were identified

between ATA versus HC and ARA versus HC respectively, with only 386 transcripts shared between the two disease subpopulations (Figure 5A-C). This further suggests meaningful differences between gene expression profiles in the skin of the two ANA based subgroups.

Patient-level pathway analysis differentiates autoantibody subsets

To better understand the functional significance of differentially expressed genes in skin at baseline, GSEA was used for individualised pathway analysis across the BIOPSY cohort (Figure 6A).

The comparison of differentially expressed Hallmark® pathways for ATA and ARA versus healthy controls for skin suggested overlapping differential pathway expression, with clear differences between the two major ANA subgroups as well as overlap (figure 6B-D and supplementary figure 6B). None of the gene sets for the parallel whole blood analysis passed the threshold for difference on GSEA. Overlapping pathways using Hallmark® are linked to aspects of SSc pathobiology that are likely to be shared across dcSSc cases. These pathways include allograft rejection, inflammation, IL6 signalling, TGFbeta signalling, angiogenesis, and complement as well as upregulation of interferon alpha response. Conversely estrogen response and MYC targets are increased for ATA positive skin but downregulated in ARA compared to HC, while adipogenesis, UV response down, and androgen are increased in ARA, but downregulated in ATA. These data provide insight into differences that could be highly relevant to the clinical, biomarker and gene expression features of these ANA based subgroups.

DISCUSSION

In the present study we have used the intrinsic clinical diversity across the SSc spectrum to help interpret molecular phenotypes and elucidate differences in potential transcriptional drivers in different stages and subsets of disease. This has important implications for both clinical practice and research, especially early-stage drug trials that will necessarily include relatively small numbers of

patients, and risk being confounded by clinical and molecular imbalance between treatment arms. By demonstrating for the first time clear differences in serum proteins and skin gene expression between ANA subgroups of early dcSSc our findings begin to explain how ANA reactivities are such strong predictors of clinical outcome and internal organ involvement [1].

The results of serum protein analysis provide an anchor for our findings. We show that serum markers that have been validated as cross sectional markers of skin fibrosis [8] have remarkably different trajectories of change between ANA subgroups, specifically the two dominant ANA reactivities, ATA and ARA.. Given our findings, despite the well-established correlation of the ELF test with MRSS and FVC [12], interpretation on a group level in early dcSSc with a mixed ANA profile, and especially over time, may be misleading. Unlike previous work on circulating markers of collagen turnover [13, 14], we did not identify clear differences between markers of collagen degradation (C1M, C4M, C6M) between disease subgroups [15].

Taken together, whole skin gene expression analysis differentiates stage and subset of SSc and gives robust insight into the differential gene expression between SSc and HC. Differential gene expression resulted in complete separation of early dcSSc and HC [similar to Skaug et al [16]], with limited and established dcSSc also forming moderately distinctive subgroups. As previously reported [8, 9], we observed relative stability in gene expression profiles over 3 months and 12 months. Skin transcriptomic differences between ATA and ARA patients with early dcSSc are especially relevant in the context of the contrasting longitudinal changes in serum markers of fibrosis observed in the BIOPSY study. This implies fundamental differences in skin biology and possibly pathogenic mechanism between ARA and ATA subgroups. This is supported by a recently published analysis of data from the GENOSIS cohort, which suggests distinct gene expression differences between major ANA reactivities [17].

Our data suggest that a relatively small number of transcripts clearly separate ARA and ATA skin gene expression. Many of these genes have already been identified to show altered expression in

SSc (IL6ST (gp130), APLN, C6 [18-21]), or other fibrotic diseases (INHBA [22]). We found shared signatures across these autoantibody subsets, as well as differences likely contribute to the distinctive clinical phenotype of these autoantibody profiles.

The fact that there were no differentiating transcriptomes in the blood between ARA+ and ATA+ patients, suggests that these key differences are important in the skin pathology, and clinical diversity of skin disease notable in these autoantibodies.

Hallmark ANA associated differences may offer insight into diversity in outcome and treatment response within early dcSSc, including clinical trials. Some recent studies have analysed intrinsic subset gene sets, which found patients who responded to MMF or abatacept (a CTLA4-Ig fusion protein) tended to be in the inflammatory subset [11, 23] whereas those who responded to dasatinib (a tyrosine kinase inhibitor with anti-fibrotic potential) were in the fibroproliferative group [24]. However, these studies did not look at the differential response to targeted therapies based on antibody subtype. It is possible that the intrinsic gene subsets [9] are differentially represented between hallmark ANA subgroups in early stage SSc and that future classification approaches incorporating both molecular and serological aspects may provide further opportunities for case stratification.

However, molecular differences between ATA and ARA identified in the present study may have relevance to treatment response for skin or internal organ disease in SSc based on other recent trial data. For example, subgroup analysis of recent Phase 2 and Phase 3 trials of tocilizumab in dcSSc suggest treatment benefit was much more marked in ATA positive patients, where prevention of decline in lung function on tocilizumab was highly significant in ATA positive subjects but not statistically significant in ATA negative [25-27]. In contrast, the RISE-SSc trial of riociguat showed a major benefit preventing MRSS progression in the ARA subgroup and no benefit for the ATA subgroup [28]. Finally, the large SENSICIS trial of nintedanib showed a numerically greater

preservation of lung function in ATA negative compared with ATA positive cases. This is notable because the ATA negative group also demonstrated numerically greater improvement in MRSS [29]. These are consistent with our hypothesis that ANA subgroups may respond differently to therapies targeting specific pathogenic mechanisms in the skin and lung.

These clinical associations raise the possibility that some of the SSc specific autoantibodies may have a direct role in pathogenesis, and that it may differ between ARA and ATA. The strongest evidence is for ATA, where ATA immune complexes (ICs) have greater effect on the IFN mRNA signature in fibroblasts compared to ARA-ICs and controls [30, 31], a key cell type mediating skin fibrosis in SSc, and contributing to the heterogeneity seen in SSc.

Taken together, our findings support the overarching hypothesis that there are distinct but overlapping pathogenic processes linking immunity and fibrosis in skin in all dcSSc, especially the interplay between adipocyte function, immunity, and fibrosis. Thus, in ARA positive cases, local connective tissue/adipocyte biology may be key to the severity and progression of skin change, and this may be independent of immune cell drive. In contrast, ATA positive dcSSc may reflect more persistent or refractory immune cell driven skin fibrosis that is less dependent on local factors and adipocyte biology. In addition, these observations may fit with novel mechanisms proposed by Wernig et al [32] linking fibrosis to failed elimination of myofibroblasts. It is plausible that this mechanism is more relevant in ARA positive cases of dcSSc than ATA.

There are notable strengths to this study. Firstly, this is a well characterised cohort of patients, prospectively collected with only 2 assessors performing MRSS (minimising interobserver variability). We present a real-life treated cohort of dcSSc patients who, as would be expected, developed complications during the study period, and had medications changed. By including a broad spectrum of SSc patients, we can interpret any findings in the context of all patients with SSc.

There are also limitations. Being a single centre study requiring significant time commitment of subjects meant that it comprised a relatively small cohort of patients. Within the prospective cohort

of patients, there are only small numbers of progressors or improvers, so these findings should be interpreted with caution. There were also some missing samples, due to patient refusal, or technical difficulties. Although we have speculated about treatment effects, this was an observational study, unable to formally compare treatments between patients.

In conclusion, BIOPSY provides a template for translational research that can integrate clinical observation and modern integrative molecular methods. In this way we have been able to better understand biological differences between subsets of SSc, and the relationship between skin disease, autoantibody subgroup and candidate molecular markers. Our results have implications for clinical practice, trial design and basic science studies of SSc.

Acknowledgments and affiliations:

Funding

This work was funded by a research grant to UCL from GlaxoSmithKline, and Medical Research Council UK grant MR/T001631/1 (fellowship to KENC)

Competing Interests

EC, AT, KN, NG, MAM, JS, YVT, NW, SF are employees of GlaxoSmithKline

CD reports personal fees from GlaxoSmithKline, Galapagos, Boehringer Ingelheim, Roche, CSL Behring, Corbus Horizon, and Arxx Therapeutics outside the submitted work

Other authors declare that they have no competing interests.

References

1. Nihtyanova SI, Sari A, Harvey JC, Leslie A, Derrett-Smith EC, Fonseca C, et al. Using autoantibodies and cutaneous subset to develop outcome-based disease classification in systemic sclerosis. *Arthritis & rheumatology (Hoboken, NJ)*. 2019 Nov 4.
2. Nihtyanova SI, Schreiber BE, Ong VH, Rosenberg D, Moinzadeh P, Coghlan JG, et al. Prediction of pulmonary complications and long-term survival in systemic sclerosis. *Arthritis & rheumatology (Hoboken, NJ)*. 2014 Jun; 66(6):1625-1635.
3. Penn H, Howie AJ, Kingdon EJ, Bunn CC, Stratton RJ, Black CM, et al. Scleroderma renal crisis: patient characteristics and long-term outcomes. *QJM : monthly journal of the Association of Physicians*. 2007 Aug; 100(8):485-494.
4. Nihtyanova SI, Denton CP. Autoantibodies as predictive tools in systemic sclerosis. *Nature reviews Rheumatology*. 2010 Feb; 6(2):112-116.
5. Nihtyanova SI, Denton CP. Pathogenesis of systemic sclerosis associated interstitial lung disease. *Journal of scleroderma and related disorders*. 2020; 5(2_suppl):6-16.

6. Denton CP, Hughes M, Gak N, Vila J, Buch MH, Chakravarty K, et al. BSR and BHPH guideline for the treatment of systemic sclerosis. *Rheumatology (Oxford, England)*. 2016 Oct; 55(10):1906-1910.
7. Makino K, Makino T, Stawski L, Mantero JC, Lafyatis R, Simms R, et al. Blockade of PDGF Receptors by Crenolanib Has Therapeutic Effect in Patient Fibroblasts and in Preclinical Models of Systemic Sclerosis. *The Journal of investigative dermatology*. 2017 Aug; 137(8):1671-1681.
8. Assassi S, Swindell WR, Wu M, Tan FD, Khanna D, Furst DE, et al. Dissecting the heterogeneity of skin gene expression patterns in systemic sclerosis. *Arthritis & rheumatology (Hoboken, NJ)*. 2015 Nov; 67(11):3016-3026.
9. Milano A, Pendergrass SA, Sargent JL, George LK, McCalmont TH, Connolly MK, et al. Molecular subsets in the gene expression signatures of scleroderma skin. *PloS one*. 2008 Jul 16; 3(7):e2696.
10. Franks JM, Martyanov V, Cai G, Wang Y, Li Z, Wood TA, et al. A Machine Learning Classifier for Assigning Individual Patients With Systemic Sclerosis to Intrinsic Molecular Subsets. *Arthritis & rheumatology (Hoboken, NJ)*. 2019 Oct; 71(10):1701-1710.
11. Hinchcliff M, Huang CC, Wood TA, Matthew Mahoney J, Martyanov V, Bhattacharyya S, et al. Molecular signatures in skin associated with clinical improvement during mycophenolate treatment in systemic sclerosis. *The Journal of investigative dermatology*. 2013 Aug; 133(8):1979-1989.
12. Abignano G, Blagojevic J, Bissell LA, Dumitru RB, Eng S, Allanore Y, et al. European multicentre study validates enhanced liver fibrosis test as biomarker of fibrosis in systemic sclerosis. *Rheumatology (Oxford, England)*. 2019 Feb 1; 58(2):254-259.
13. Juhl P, Vinderslev Iversen L, Karlsmark T, Asser Karsdal M, Bay-Jensen AC, Mogensen M, et al. Association of metabolites reflecting type III and VI collagen formation with modified Rodnan skin score in systemic sclerosis - a cross-sectional study. *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals*. 2019 Jun; 24(4):373-378.
14. Pernille Juhl A-CB-J, Roger Hesselstrand, Anne Sofie Siebuhr, Dirk Wuttge. SEROLOGICAL ASSESSMENT OF THE FIBROTIC INDEX IN SCLEROSIS: A CROSS SECTIONAL STUDY. *Annals of the Rheumatic Diseases*. 2019; 78(Suppl 2):845.
15. Dobrota R JS, Juhl P, Maurer B, Wildi L, Bay-Jensen AC, Karsdal M, Herrick A, Distler J, Allanore Y, Hoffman-Vold A, Siebuhr AS, Distler O. Circulating collagen neo-epitopes and their role in the prediction of fibrosis in systemic sclerosis - a multicentre cohort study. *Lancet Rheumatol*. 2020.
16. Skaug B, Khanna D, Swindell WR, Hinchcliff ME, Frech TM, Steen VD, et al. Global skin gene expression analysis of early diffuse cutaneous systemic sclerosis shows a prominent innate and adaptive inflammatory profile. *Ann Rheum Dis*. 2019 Nov 25.
17. Inamo J. Association of differentially expressed genes and autoantibody type in patients with systemic sclerosis. *Rheumatology (Oxford, England)*. 2020 Sep 10.
18. Taniguchi T, Asano Y, Fukasawa T, Yoshizaki A, Sato S. Critical contribution of the interleukin-6/signal transducer and activator of transcription 3 axis to vasculopathy associated with systemic sclerosis. *The Journal of dermatology*. 2017 Aug; 44(8):967-971.
19. Taniguchi T, Miyagawa T, Tamaki Z, Nakamura K, Yamashita T, Saigusa R, et al. A possible implication of reduced levels of LIF, LIFR, and gp130 in vasculopathy related to systemic sclerosis. *Archives of dermatological research*. 2017 Dec; 309(10):833-842.
20. Apostolidis SA, Stifano G, Tabib T, Rice LM, Morse CM, Kahaleh B, et al. Single Cell RNA Sequencing Identifies HSPG2 and APLNR as Markers of Endothelial Cell Injury in Systemic Sclerosis Skin. *Frontiers in immunology*. 2018; 9:2191.
21. Benbassat C, Schlesinger M, Luderschmidt C, Valentini G, Tirri G, Shoenfeld Y. The complement system and systemic sclerosis. *Immunologic research*. 1993; 12(3):312-316.
22. Wietecha MS, Pensalfini M, Cangkruma M, Müller B, Jin J, Brinckmann J, et al. Activin-mediated alterations of the fibroblast transcriptome and matrisome control the biomechanical properties of skin wounds. *Nature communications*. 2020 May 25; 11(1):2604.

23. Chakravarty EF, Martyanov V, Fiorentino D, Wood TA, Haddon DJ, Jarrell JA, et al. Gene expression changes reflect clinical response in a placebo-controlled randomized trial of abatacept in patients with diffuse cutaneous systemic sclerosis. *Arthritis research & therapy*. 2015 Jun 13; 17:159.
24. Martyanov V, Kim GJ, Hayes W, Du S, Ganguly BJ, Sy O, et al. Novel lung imaging biomarkers and skin gene expression subsetting in dasatinib treatment of systemic sclerosis-associated interstitial lung disease. *PloS one*. 2017; 12(11):e0187580.
25. Khanna D, Denton CP, Jahreis A, van Laar JM, Frech TM, Anderson ME, et al. Safety and efficacy of subcutaneous tocilizumab in adults with systemic sclerosis (faSScinata): a phase 2, randomised, controlled trial. *Lancet (London, England)*. 2016 Jun 25; 387(10038):2630-2640.
26. Huang S, Ye W, Lin C, Jahreis A, Denton C. Relationship of Baseline Measures to the Change in the Forced Vital Capacity in a Phase 3 Randomized Controlled Trial of Tocilizumab for the Treatment of Early Systemic Sclerosis. *B16 ILD THERAPY II:A2750-A2750*.
27. Khanna D, Lin CJF, Furst DE, Goldin J, Kim G, Kuwana M, et al. Tocilizumab in systemic sclerosis: a randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet Respiratory medicine*. 2020 Oct; 8(10):963-974.
28. Khanna D, Allanore Y, Denton CP, Kuwana M, Matucci-Cerinic M, Pope JE, et al. Riociguat in patients with early diffuse cutaneous systemic sclerosis (RISE-SSc): randomised, double-blind, placebo-controlled multicentre trial. *Ann Rheum Dis*. 2020 May; 79(5):618-625.
29. Mayes M HK, Gahlemann M, Fischer A, Raghu G, Girard M, Alves M, Stowasser S, Distler J, Matucci-Cerinic M, Volkman E, Kuwana M, Distler O. . Effect of Anti-Topoisomerase I Antibody Status on Decline in Lung Function in Patients with Systemic Sclerosis-Associated Interstitial Lung Disease: Data from the SENSICIS Trial [abstract]. *Arthritis Rheumatol* 2019; 71 (suppl 10).
30. Raschi E, Chighizola CB, Cesana L, Privitera D, Ingegnoli F, Mastaglio C, et al. Immune complexes containing scleroderma-specific autoantibodies induce a profibrotic and proinflammatory phenotype in skin fibroblasts. *Arthritis research & therapy*. 2018 Aug 29; 20(1):187.
31. Raschi E, Privitera D, Bodio C, Lonati PA, Borghi MO, Ingegnoli F, et al. Scleroderma-specific autoantibodies embedded in immune complexes mediate endothelial damage: an early event in the pathogenesis of systemic sclerosis. *Arthritis research & therapy*. 2020 Nov 9; 22(1):265.
32. Lerbs T, Cui L, King ME, Chai T, Muscat C, Chung L, et al. CD47 prevents the elimination of diseased fibroblasts in scleroderma. *JCI insight*. 2020 Aug 20; 5(16).
33. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. *Arthritis Rheum*. 2013 Nov; 65(11):2737-2747.
34. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA, Jr., et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *The Journal of rheumatology*. 1988 Feb; 15(2):202-205.
35. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. 2005; 102(43):15545-15550.
36. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics*. 2003 2003/07/01; 34(3):267-273.
37. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nature genetics*. 2006 May; 38(5):500-501.

Key messages

What is already known about this subject?

- Linking skin and protein expression to clinical differences between subgroups in systemic sclerosis (SSc) has been challenging.
- The hallmark antinuclear autoantibodies (ANA) used to diagnose SSc also predict clinically important differences in skin and internal organ disease.

What does this study add?

- This study uses clinical and ANA heterogeneity across a well characterised broad SSc cohort to better understand the molecular architecture of early diffuse cutaneous SSc.
- We demonstrate for the first-time striking differences in longitudinal patterns of serum protein markers between ANA subgroups in SSc.
- High dimensional analysis of skin gene expression with patient level pathway analysis suggests biological basis for differences between ANA based subgroups.

How might this impact on clinical practice or future developments?

- Defining the molecular basis for clinical diversity gives insight into SSc disease biology relevant to clinical practice and trial design.

Figure Legends

Figure 1: Baseline and longitudinal change in MRSS in the BIOPSY cohort A) Median MRSS across the BIOPSY cohort (TUKEY post-hoc p values <0.05 included) of early dcSSc (n=21), established dcSSc (n=15), lcSSc (n=16), HC (n=16). B) Mean MRSS and SEM of early dcSSc during prospective follow up. C) Mean MRSS and SEM based on skin status from baseline to 12 months. D) Mean MRSS and SEM by autoantibody subset in early dcSSc cohort.

Figure 2: Baseline scores for ELF and constituents in BIOPSY cohort A-D) ELF test and constituents at baseline (TUKEY post-hoc p values included). A) Enhanced liver fibrosis test (ELF) at baseline. B) Tissue inhibitor of metalloproteinases-1 (TIMP-1). C) Hyaluronic acid (HA). D) Type III Procollagen peptide (PIIINP).

Figure 3: Longitudinal analysis of circulating proteins by group and autoantibody subset in the BIOPSY cohort. Results of ELF test and constituents and mean change (+/-SEM) as fraction over time on a group level, and by antibody subtype. A) Enhanced liver fibrosis test (ELF). B) Hyaluronic acid (HA). C) Tissue inhibitor of metalloproteinases-1 (TIMP-1). D) Type III Procollagen peptide (PIIINP). * p<0.05, ** p<0.01, *** p<0.001

Figure 4: Whole skin transcriptomic analysis for the BIOPSY cohort differentiates autoantibody subsets. A) Overall view of enrichment scores of genes upregulated in dcSSc and downregulated in dcSSc across BIOPSY cohort and time points. The SSc specific composite signature is derived from multiple publicly available datasets and reflects those genes that are consistently up or downregulated in SSc skin biopsies. * p < 0.05, ** p < 0.001. B) and C) PCA and 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). D) Hierarchical clustering based on 61 significantly differing gene expressions (FDR <0.1) from skin comparing ARA (red) and ATA (green) positive early dcSSc.

Figure 5: Results from analysis of gene expression of early dcSSc autoantibody subgroups compared to HC. A) Venn diagram of number of significantly differently expressed genes in skin in ARA compared to HC, and ATA compared to HC and those significantly differently expressed in both. B) Venn diagram to show number of significantly differently expressed genes in blood by autoantibody comparisons. C) Table showing top 20 significantly differently expressed genes (FDR<0.05) with highest fold change between autoantibody and healthy control, and corrected p value found in skin. D-G) Scatter plots of select genes associated with fibrosis. Genes selected from supplementary table 3, colour defined by autoantibody state. Mean value of each disease subgroup and autoantibody state indicated by "X". TUKEY post hoc p value across disease subgroups included (*p<0.05, **p<0.01, ***p<0.001). Gene expression expressed in log format D) INHBA E) IL6ST F) APLN, and G) C6.

Figure 6: Pathway analysis for differentially expressed pathways across the BIOPSY cohort. 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). B) ssGSEA of all Hallmark pathways significantly elevated in early dcSSc ARA (red) with HC and early dcSSc ATA (green) with HCC

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.

List of supplementary materials:

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.

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

Supplementary Figure 1: Overview of study design.

Supplementary Figure 2: Baseline markers of collagen C3 and C6 synthesis, their fibrotic index, and correlation with MRSS

Supplementary Figure 3: Analysis of markers of collagen synthesis and degradation, and key soluble markers, over 12 months

Supplementary figure 4: Differentially expressed genes from early dcSSc and HC in skin and blood

Supplementary Figure 5: PCA plots of longitudinal analysis of gene expression from early dcSSc

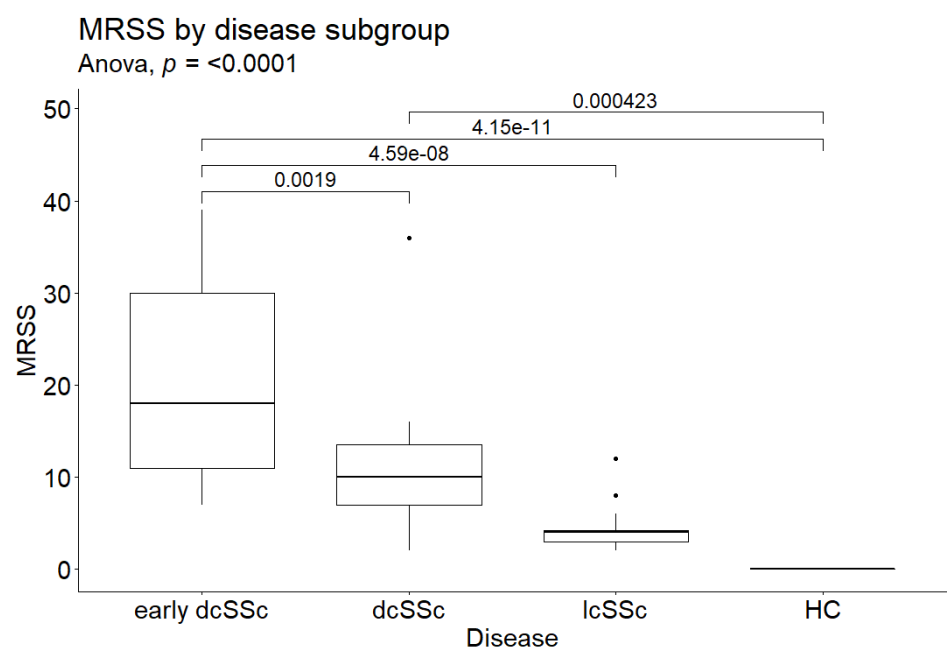
Supplementary Figure 6: subgroup analysis of single sample GSEA of all significantly differentially expressed pathways

Supplementary Figure 7: Schematic representation linking skin score trajectory to pathogenic mechanisms that drive candidate molecular markers in skin and blood samples

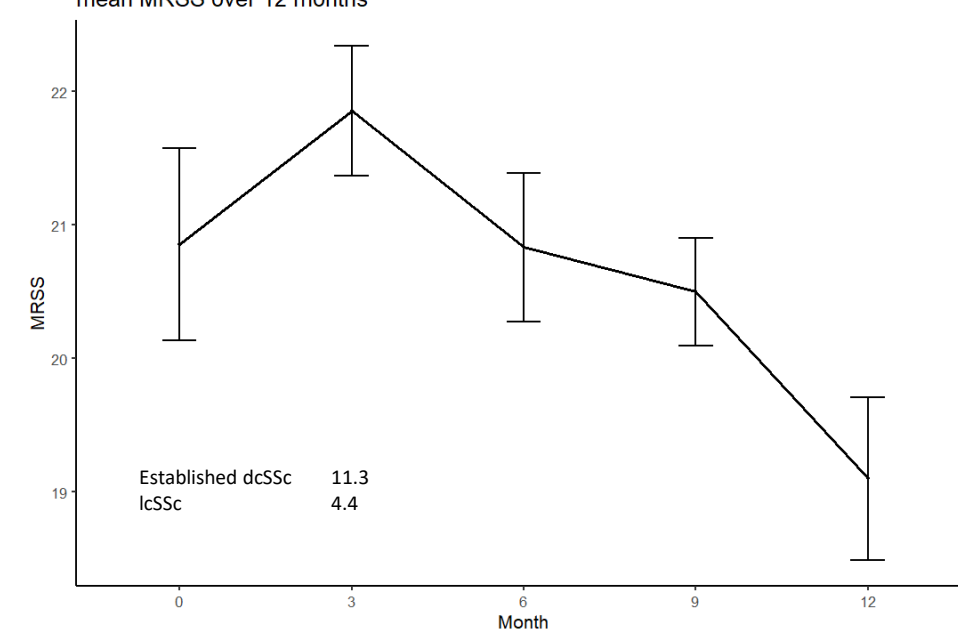
Figure 1

Baseline and longitudinal change in MRSS in the BIOPSY cohort

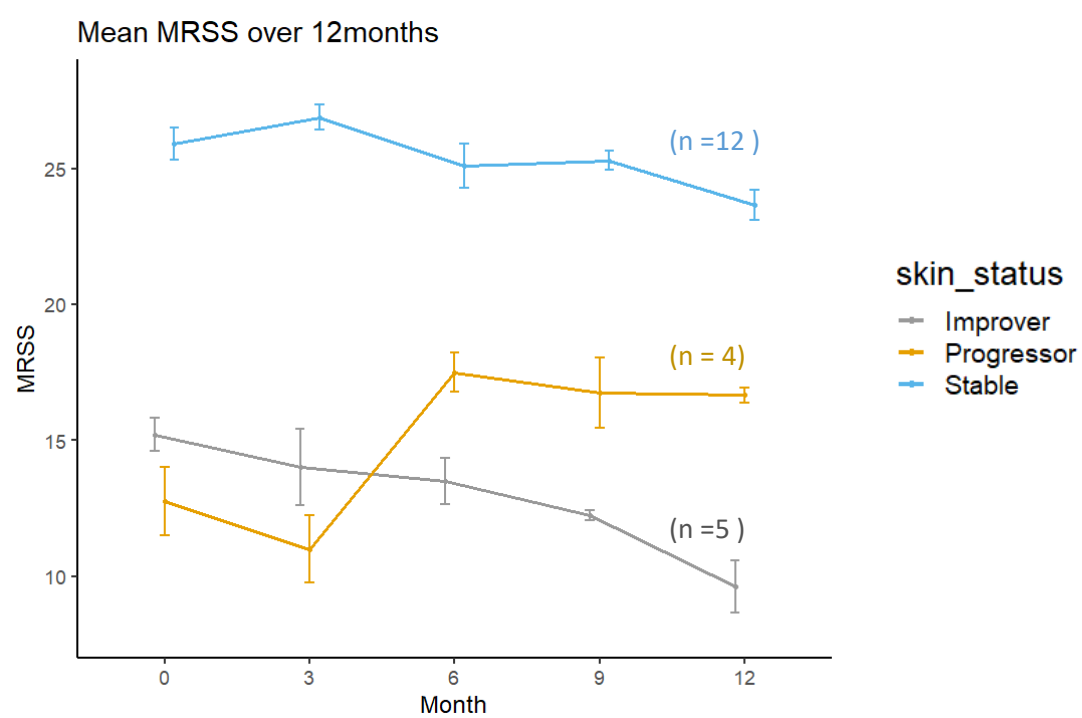
A



B



C



D

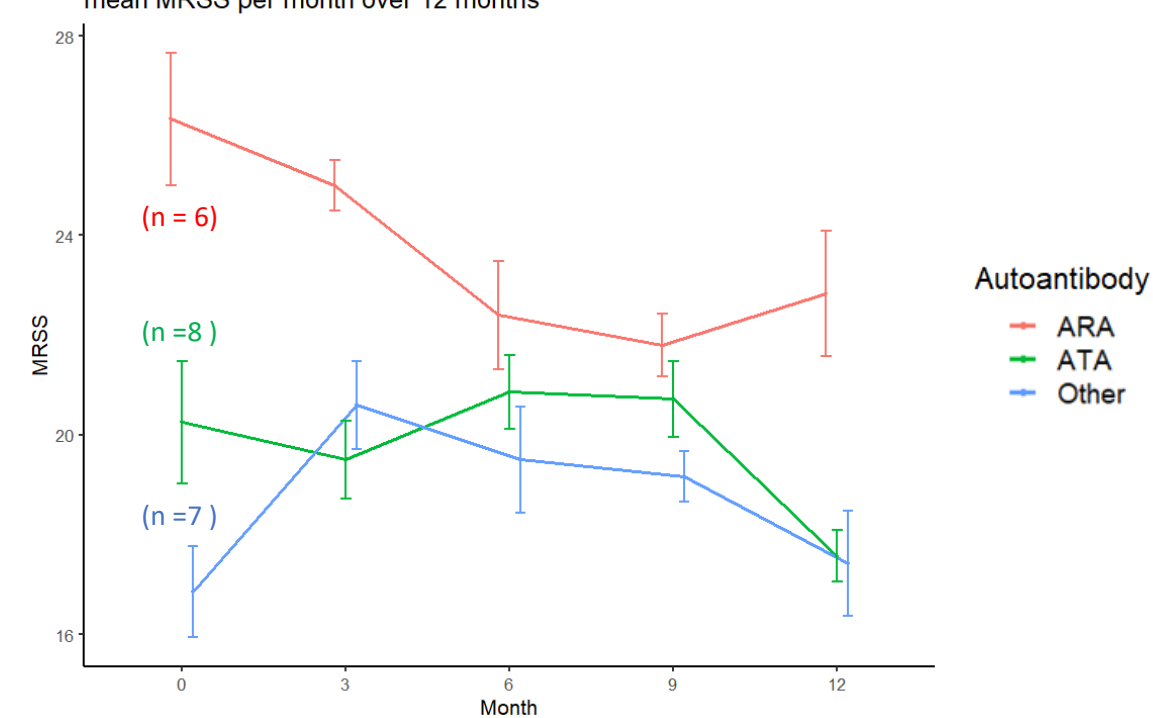
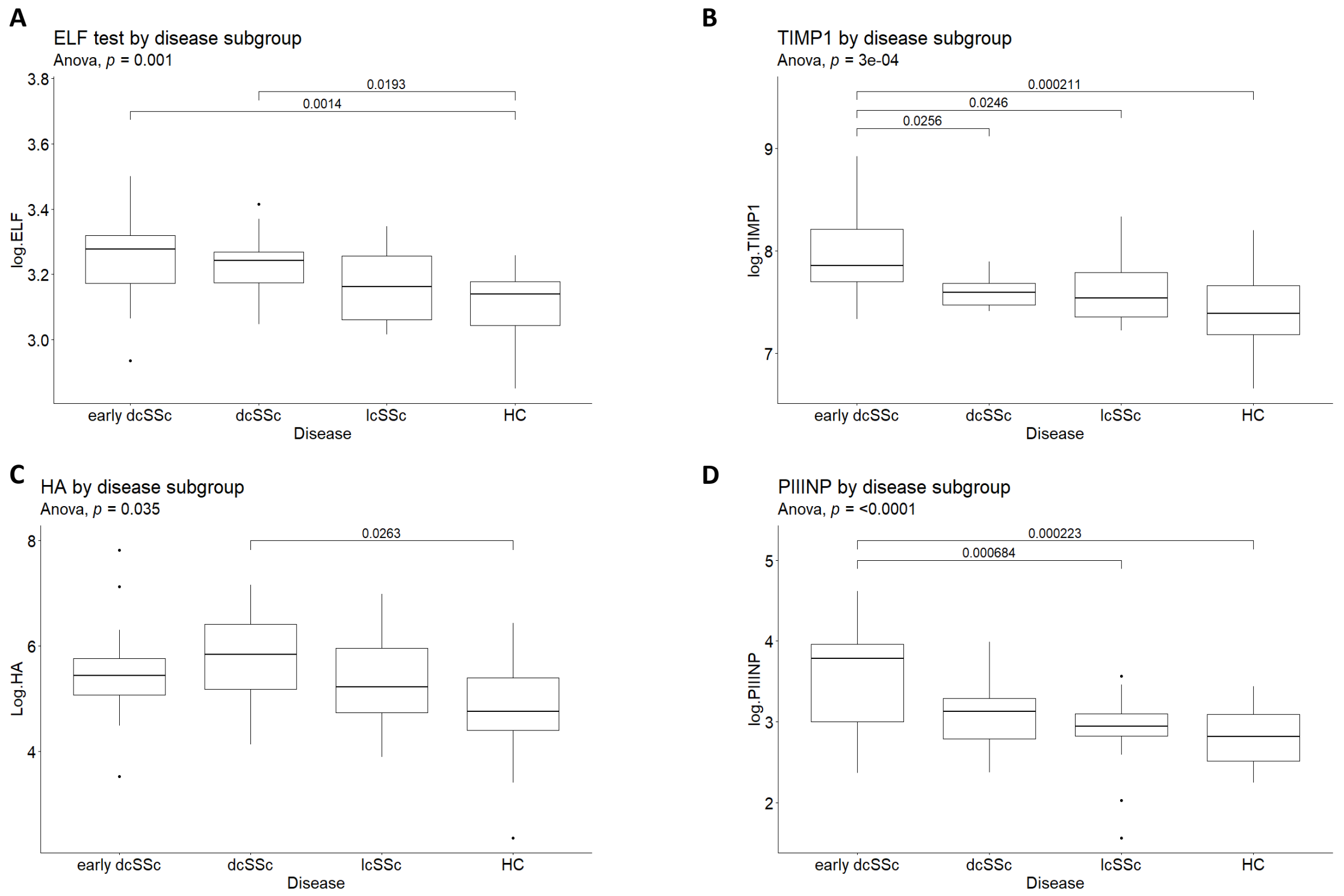


Figure 2

Baseline scores for ELF and constituents in BIOPSY cohort

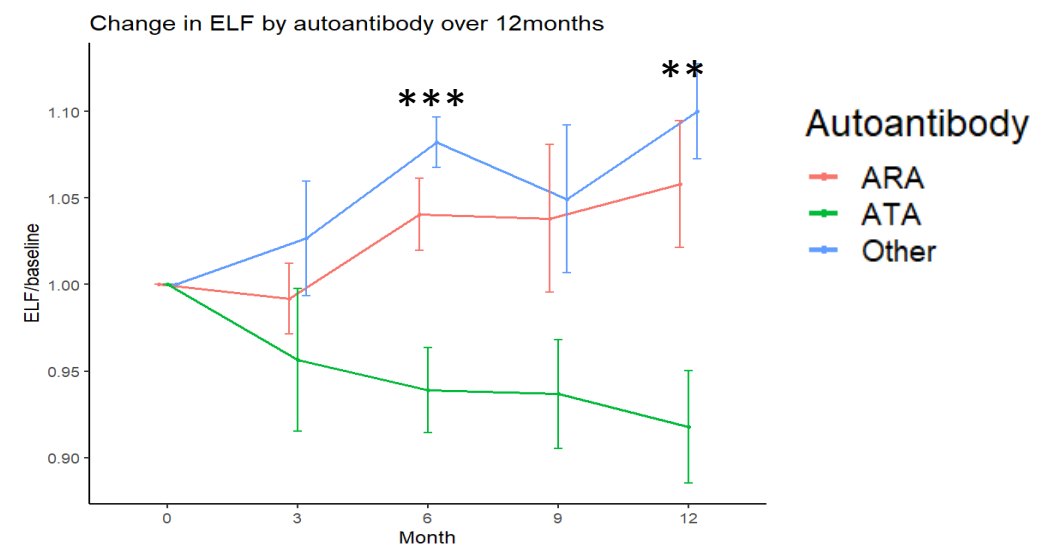
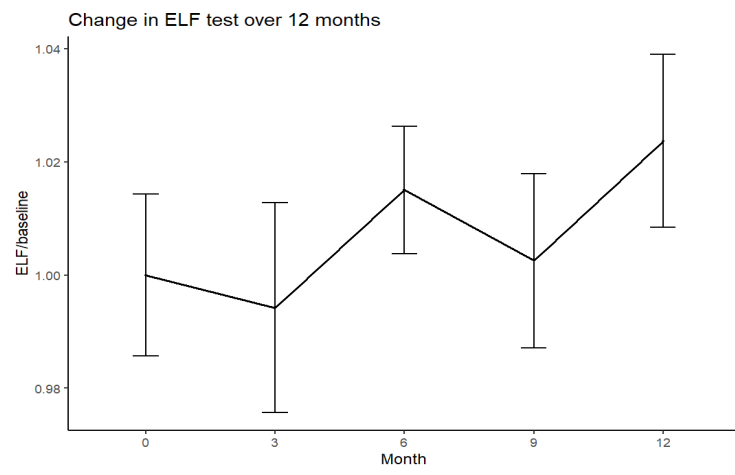


ELF score and components

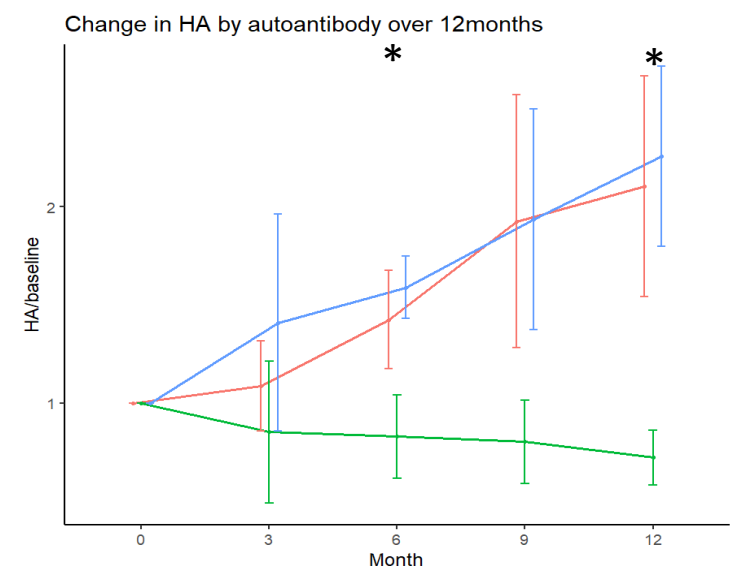
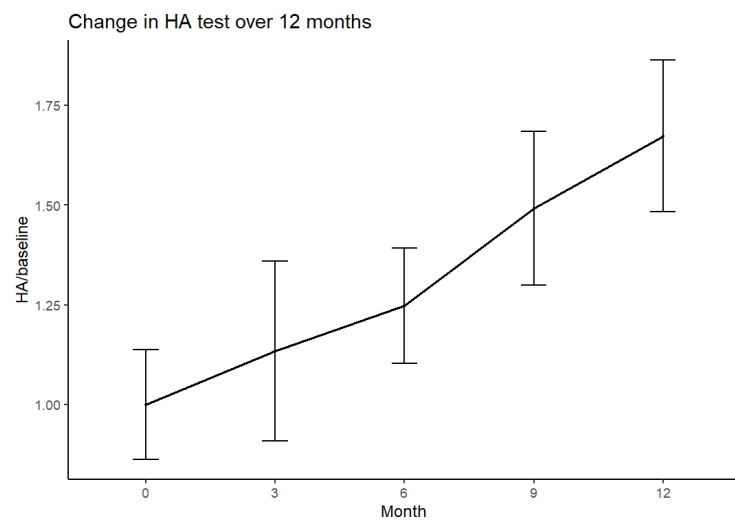
Whole Group

By Autoantibody

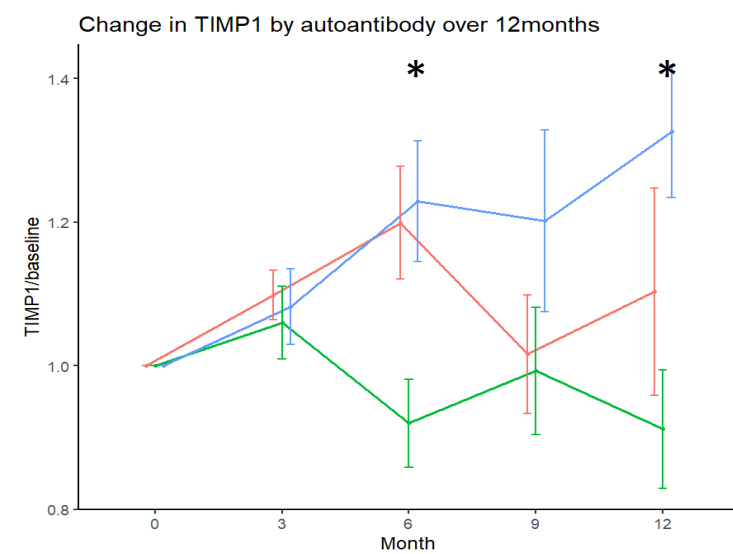
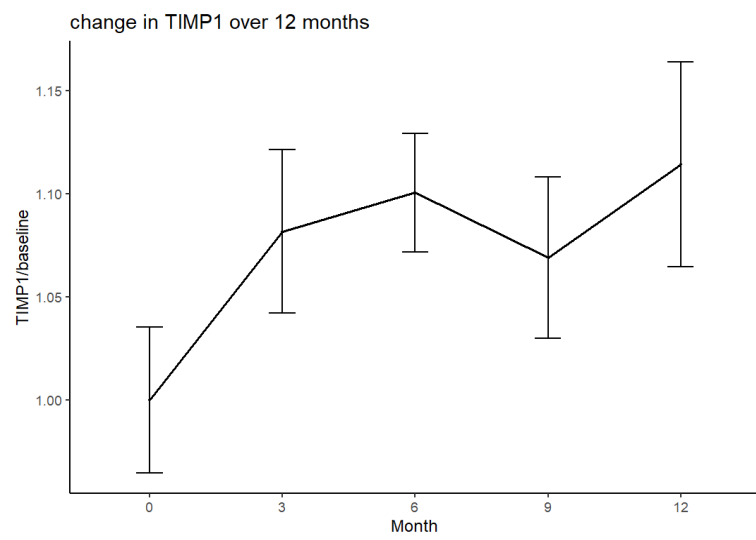
A



B



C



D

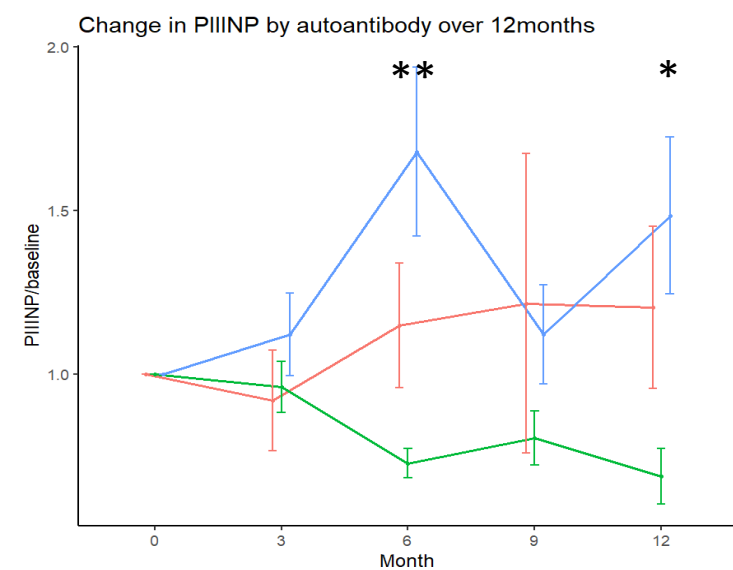
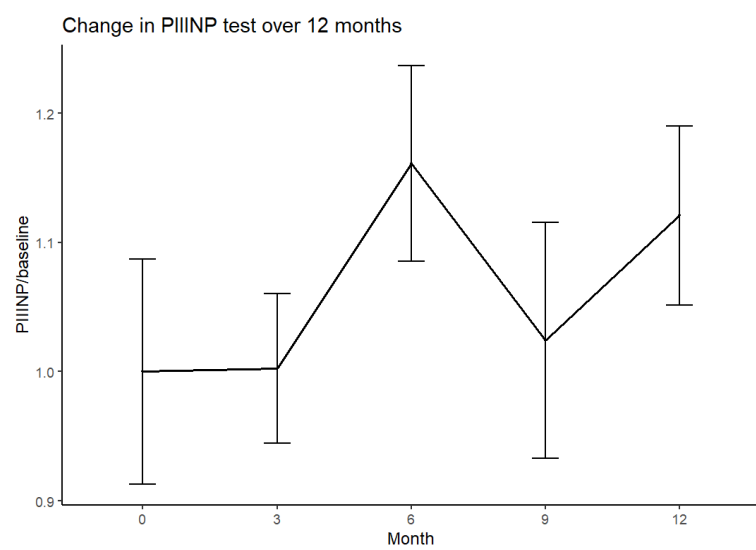


Figure 4 Whole skin transcriptomic analysis for the BIOPSY cohort differentiates autoantibody subsets

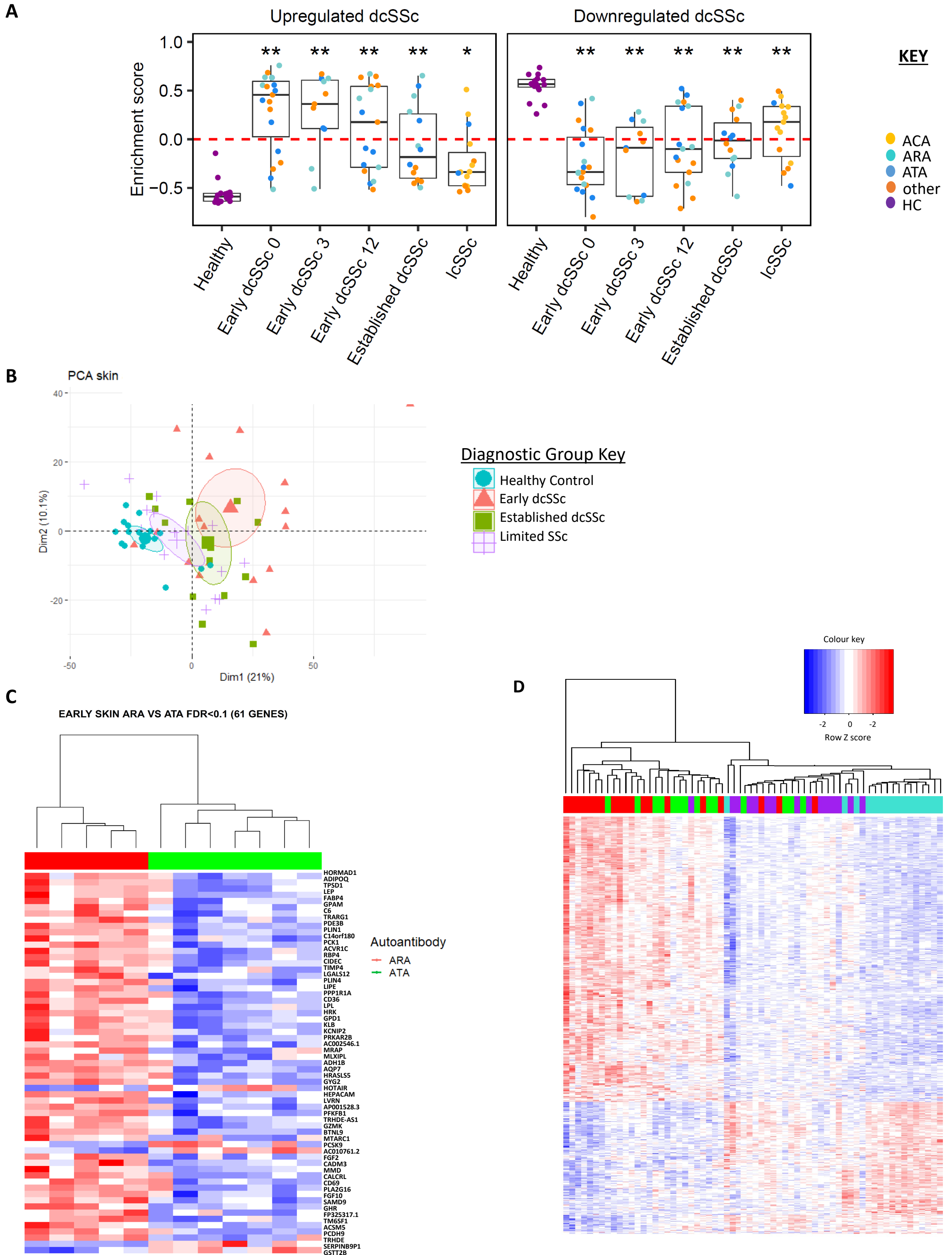
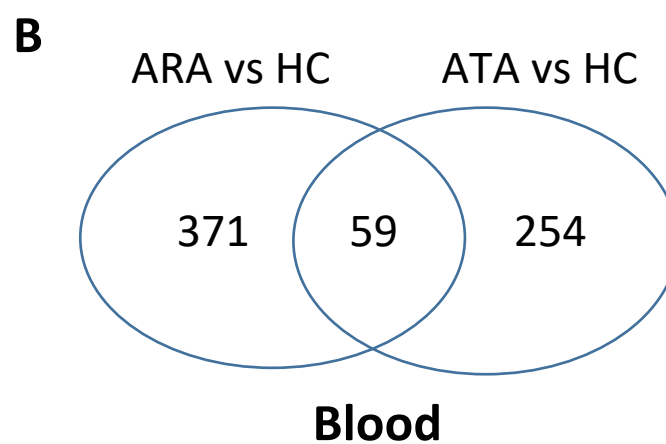
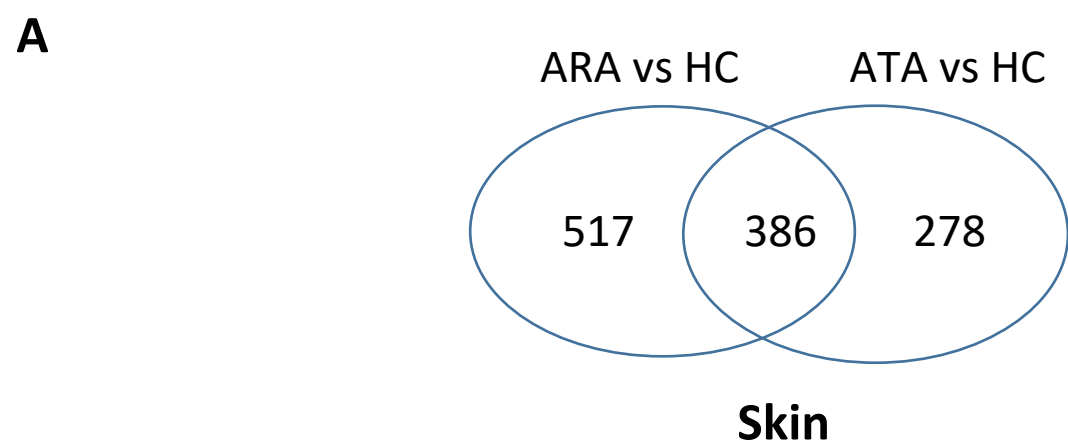




Figure 5 Results from analysis of gene expression of early dcSSc autoantibody subgroups compared to HC



C

SKIN - upregulated genes										
Upregulated in ARA vs HC			Upregulated in both ARA vs HC and ATA vs HC					Upregulated in ATA vs HC		
Gene	F.C	FDR	FDR	FC	Gene	FC	FDR	Gene	F.C	FDR
PLA2G2D	5.789	0.000	0.000	6.313	CXCL9	1.845	0.001	MT-TC	6.013	0.013
CXCL13	4.357	0.000	0.000	5.686	TNFSF18	2.063	0.000	PRR33	5.893	0.021
CXCL10	4.088	0.000	0.000	5.288	SFRP4	4.243	0.000	MTND1P23	5.885	0.023
SAA1	3.457	0.001	0.000	5.034	COMP	4.018	0.000	CICP14	2.565	0.012
C6	3.255	0.000	0.000	4.930	COL4A4	3.939	0.000	C22orf23	2.374	0.024
PRR9	2.659	0.036	0.001	4.815	LTF	2.716	0.012	CD177	2.277	0.009
HAPLN1	2.607	0.001	0.000	4.509	ADAMDEC1	2.595	0.000	APLN	2.214	0.002
PCK1	2.461	0.005	0.000	4.426	ADAM12	3.303	0.000	SPINK7	2.128	0.008
HORMAD1	2.449	0.020	0.000	4.362	COL11A1	3.176	0.000	MT-ND6	2.107	0.017
INHBA	2.438	0.000	0.000	4.249	SULF1	2.759	0.000	PERM1	2.105	0.030
C14orf180	2.435	0.004	0.000	4.242	THBS4	2.861	0.001	GP1BA	1.991	0.000
CXCL11	2.384	0.001	0.001	4.082	SERPINE1	2.193	0.009	CRMP1	1.946	0.004
IDO1	2.381	0.000	0.000	3.994	HP	1.999	0.022	TMPRSS4	1.925	0.006
ADIPOQ	2.298	0.002	0.000	3.694	SCG2	2.868	0.000	GSTT2B	1.918	0.031
DGKI	2.241	0.000	0.000	2.972	CPXM1	3.684	0.000	MTCO1P12	1.915	0.027
PLA2G2A	2.218	0.007	0.000	2.609	LAMP5	3.673	0.000	COL1A1	1.915	0.002
TIMP4	2.178	0.001	0.000	3.618	TNFSF4	2.117	0.000	APLNR	1.901	0.001
MMP9	2.134	0.001	0.000	3.568	TIMD4	1.707	0.003	CYTL1	1.894	0.000
CIDEC	2.123	0.002	0.000	3.477	PENK	1.931	0.000	GPX2	1.873	0.003
SLAMF8	2.109	0.000	0.000	3.387	COL8A1	2.303	0.000	GPR68	1.864	0.000

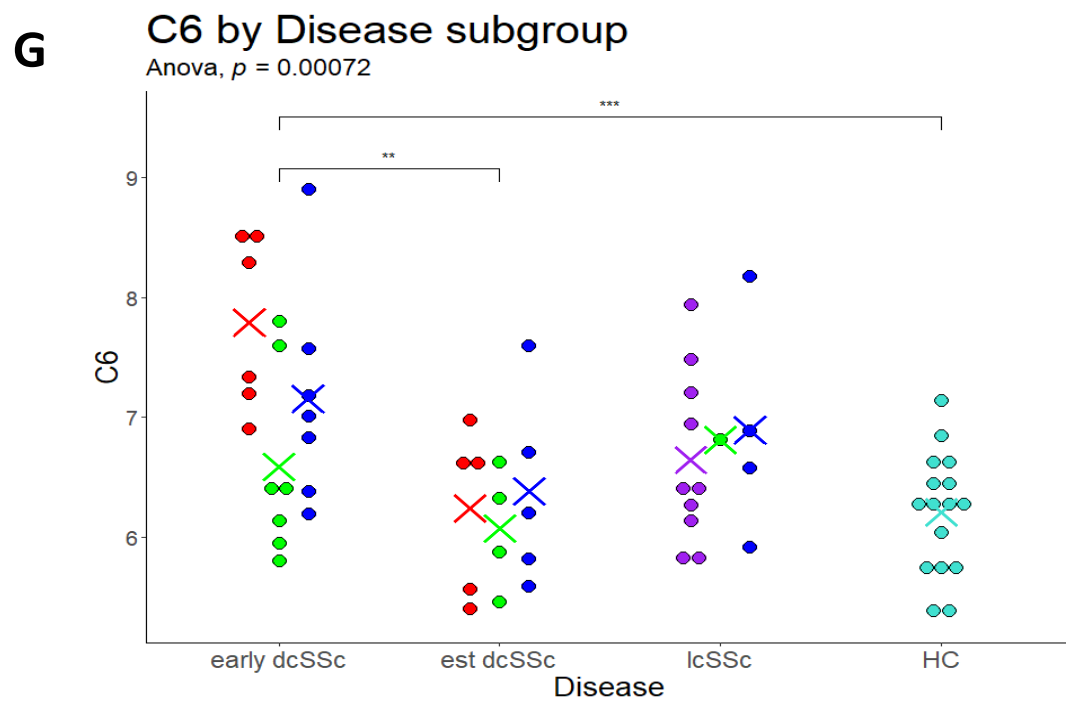
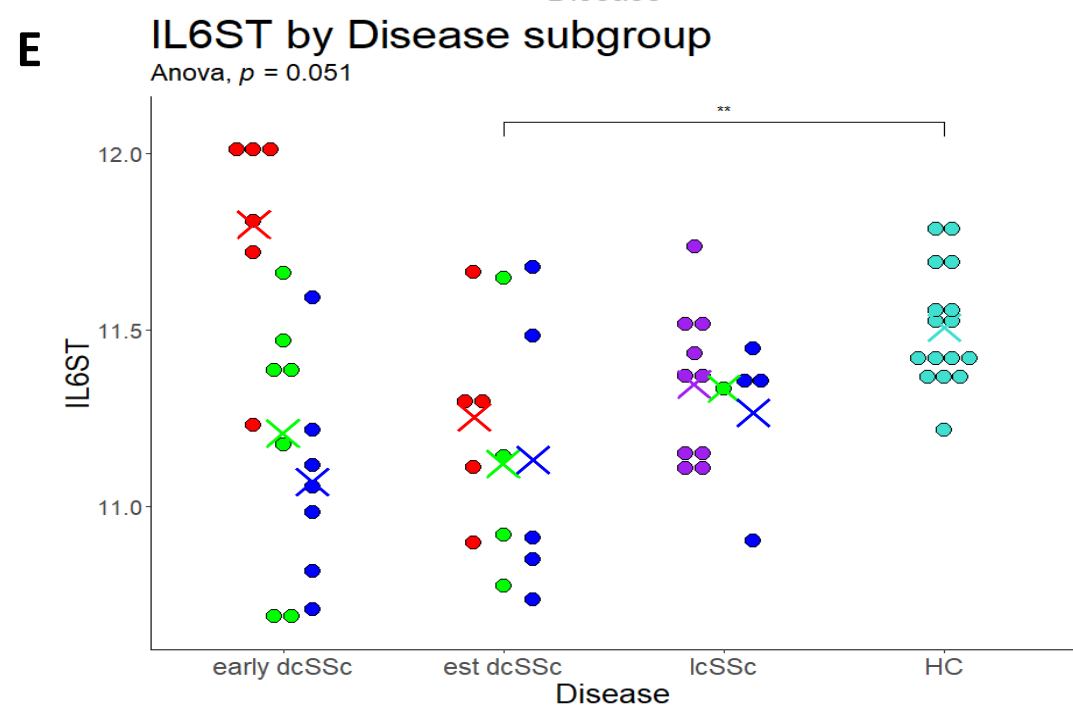
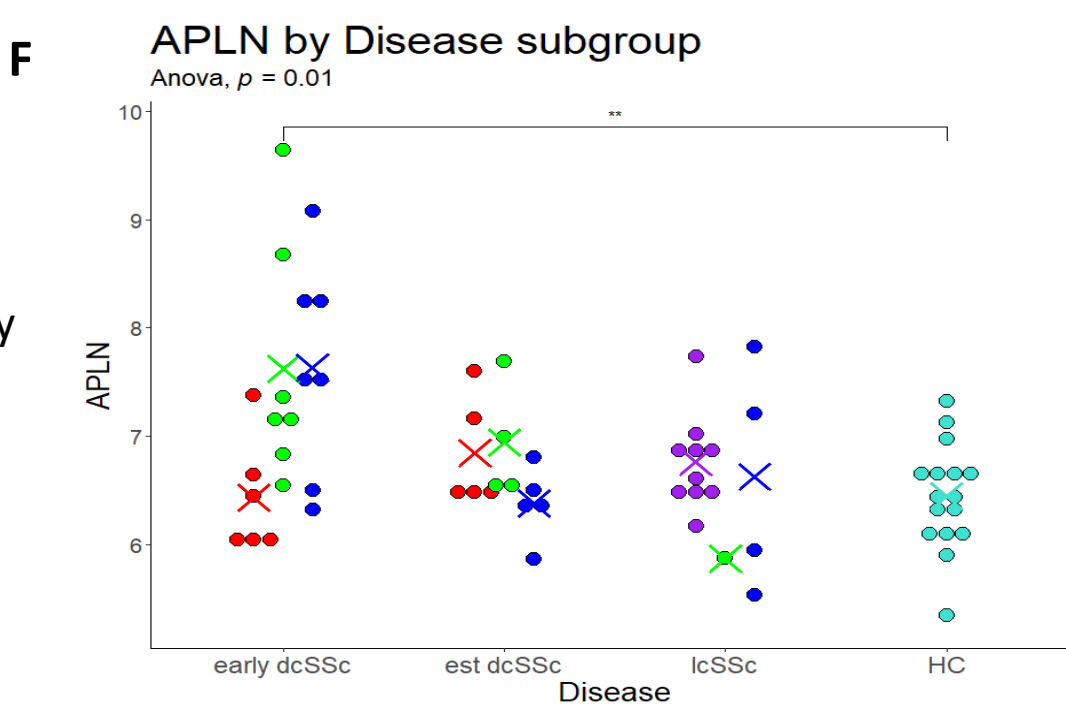
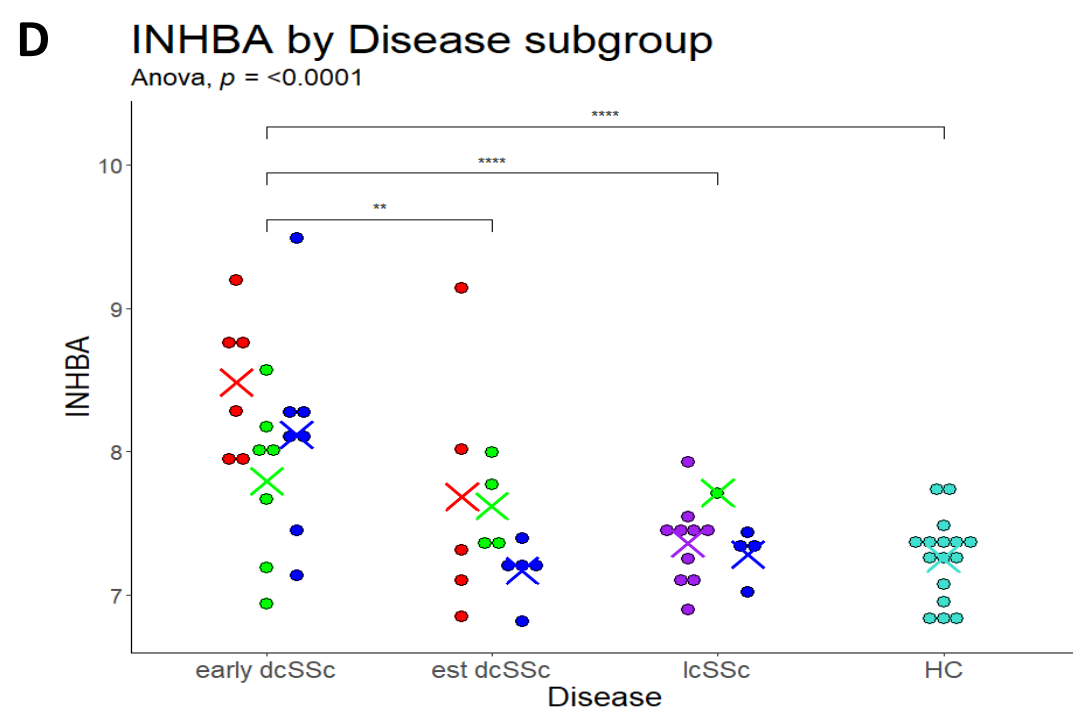
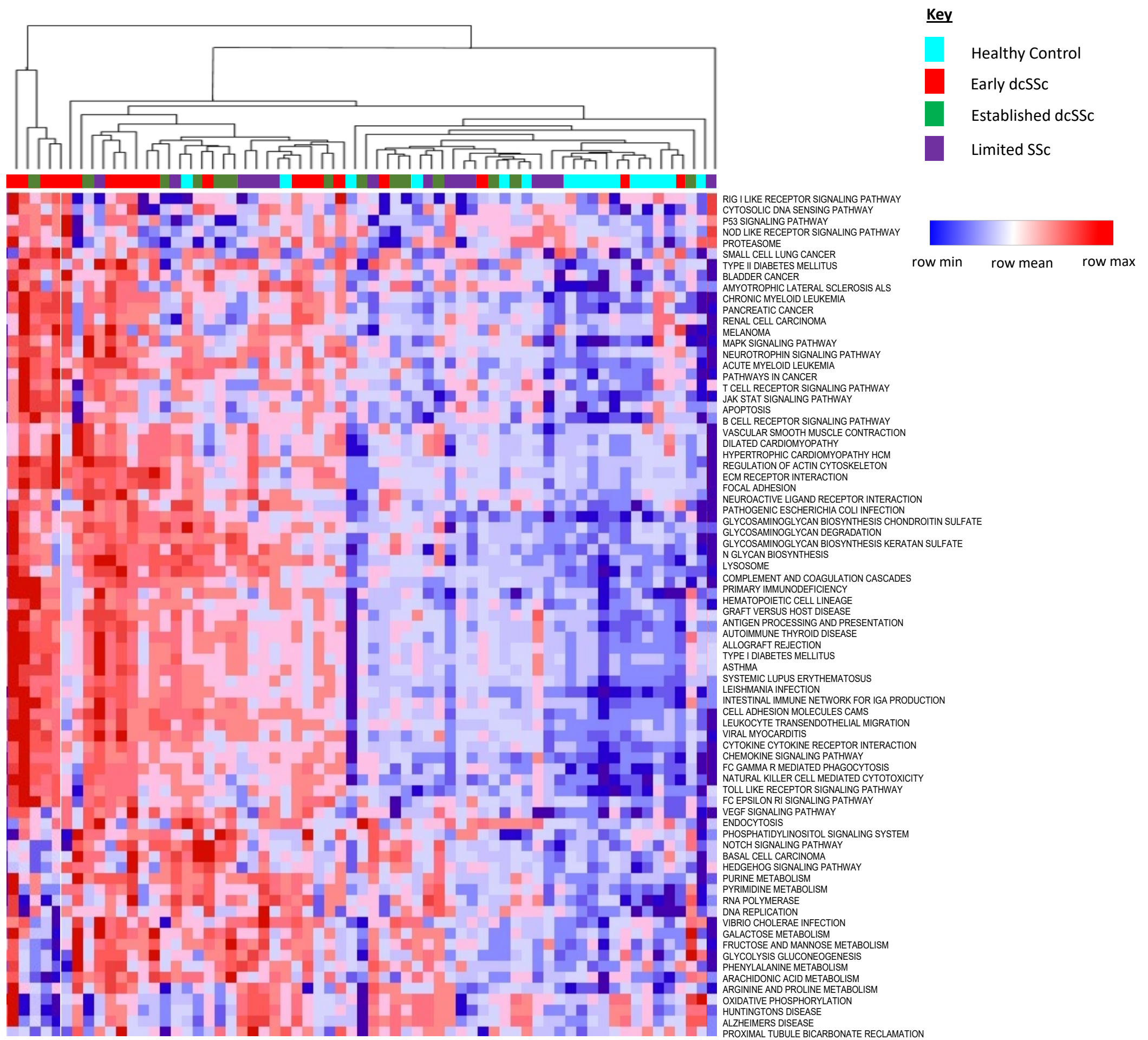


Figure 6

Pathway analysis for differentially expressed Pathways across the BIOPSY cohort

A



B

Normalised enrichment score for significant Hallmark pathways by autoantibody

