UNIVERSITÀ DEGLI STUDI DI MILANO

CORSO DI DOTTORATO IN RICERCA CLINICA

XXXIV CICLO

TRANSCRIPTOMIC AND PROTEOMIC PROFILING OF PRECLINICAL SYSTEMIC SCLEROSIS

MED/09 MED/16

Chiara Bellocchi

Matricola R12212

TUTOR

Prof. Nicola Montano

SUPERVISOR

Dr. Lorenzo Beretta

COORDINATORE DEL DOTTORATO

Prof. Emilio Berti e Prof. Massimo Del Fabbro

A.A. 2020/2021

Abstract

Background: Systemic sclerosis (SSc) is a rare systemic autoimmune disease with a high morbidity and mortality. In the preclinical phase of the disease, patients present exclusively Raynaud phenomenon (RP) with a nailfold video-capillaroscopy positive for scleroderma pattern and/or the presence of specific autoantibodies. Research investigations on preclinical systemic sclerosis (Pre-SSc) patients are rare due to the exceptionality and the difficulty to intercept this specific subgroup of subjects that in five years will progress into a definite SSc diagnosis in about 50% of cases.

Objectives: to perform a comprehensive investigation of preclinical SSc at gene expression, proteomic and clinical level with the aim to identify specific early-stage disease biomarkers as well as progression biomarkers.

Methods: gene expression analysis through RNAseq technique was performed on whole blood samples of 35 Pre-SSc and 16 matched healthy controls (HC) collected at baseline and after four years. Gene expression module changes analysis was performed comparing evolving Pre-SSc, stable Pre-SSc and HC at baseline and follow-up. A proteomic analysis through SOMAscan technology was assessed on serum at baseline of a subgroup of 16 Pre-SSc (8 evolving Pre-SSc vs 8 stable Pre-SSc) and 8 HC selected on a clinical basis, matching for age, gender and autoantibody profile. Finally, proteins emerged from SOMAscan analaysis to be predictive of progression were validated on a validation cohort of 50 subjects with preclinical features whose serum aliquots were available at baseline.

Results: out of 35 Pre-SSc, 15 (42.9%) progressed toward a definite SSc after four years. The presence of gastro-esophageal reflux and a shorter time of RP were associated with a shorter time of progression.

RNA expression change analysis of evolving Pre-SSc vs stable Pre-SSc identified 73 genes with a corrected p value ≤ 0.05. At baseline, Pre-SSc had type I IFN gene expression modules (M 3.4, M 5.12, M 1.2) increased in comparison to HC. NK gene expression modules were decreased in evolving Pre-SSc over time.

Out of 286 proteins assessed by SOMAscan, 10 proteins were able to predict at baseline progressors from not progressors. Evolving Pre-SSc showed increased levels of NKp30, Endostatin, bFGF, ECM1, FGF18, Fibronectin 1.3, PAFAH1B2, FABP and decreased levels of PHI and Ubiquitin1.

High levels of endostatin and reduced serum levels of PAFAH1B2 were confirmed with ELISAs in the validation cohort to correlate with a shorter time to progression.

Conclusion: a type I IFN signature distinguished preclinical SSc from HC and a reduced NK signature was associated to SSc progression. Proteins linked to pathways of fibrosis, extracellular matrix organization, positive regulation of cell proliferation, angiogenesis, signal transduction were discovered to predict disease progression. Moreover, endostatin emerged as a biomarker worthy of future mechanistic investigations.

Abstract

Introduzione: La sclerosi sistemica (SSc) è una malattia rara autoimmune sistemica con un'elevata morbilità e mortalità. Nella fase preclinica della malattia, i pazienti sono caratterizzati unicamente dal fenomeno di Raynaud (FR) e dalla presenza di autoanticorpi SSc specifici e di una video-capillaroscopia periungueale positiva per scleroderma pattern. Studi di ricerca sui pazienti con sclerosi sistemica in forma preclinica (Pre-SSc) sono ad oggi numericamente rari a causa della difficoltà ad intercettare questo specifico sottogruppo di soggetti che in cinque anni progredirà verso una diagnosi definitiva di SSc in circa il 50% dei casi.

Obiettivi: effettuare uno studio approfondito dei pazienti con SSc in fase preclinica valutandone l'espressione genica, il profilo proteomico e le caratteristiche cliniche per poter identificare possibili biomarcatori specifici dello stadio precoce e della progressione di malattia.

Metodi: una analisi di espressione genica mediante metodica RNAseq è stata eseguita su campioni di sangue intero di 35 Pre-SSc e 16 controlli sani (HC) raccolti al basale e dopo quattro anni. L'analisi modulare dei cambiamenti di espressione genica ha confrontato i Pre-SSc in progressione (evolving Pre-SSc), con i pazienti che hanno mantenuto un profilo preclinico stabile nel tempo (stable Pre-SSc) e con i controlli sani. Un'analisi di proteomica tramite tecnologia SOMAscan è stata effettuata al basale sul siero di un sottogruppo di 16 Pre-SSc (8 evolving Pre-SSc vs 8 stable Pre-SSc) e 8 HC selezionati su base clinica per corrispondenza di età, genere e profilo autoanticorpale. Infine, le proteine emerse dall'analisi SOMAscan essere predittive della progressione di malattia, sono state successivamente convalidate in una coorte di validazione composta da 50 soggetti con SSc di tipo preclinico le cui aliquote di siero erano disponibili al basale.

Risultati: dopo quattro anni, 15 Pre-SSc su 35 (il 42,9%) sono progrediti in una diagnosi di SSc definitiva. La presenza di reflusso gastro-esofageo e un tempo più breve di FR sono risultati essere associati ad un tempo di progressione più veloce.

A livello trascrizionale, il confronto dell'espressione genica (analisi pre-post) nei pazienti evolving Pre-SSc vs gli stable Pre-SSc ha identificato 73 geni differenziali con p corretto $\leq 0{,}05$. Al basale, i pazienti Pre-SSc hanno mostrato moduli di espressione genica dell'IFN di tipo I (M 3.4, M 5.12, M 1.2) aumentati rispetto ai controlli sani. I moduli di espressione genica NK (natural killers) sono risultati essere ridotti nei pazienti evolving Pre-SSc sia al basale che al follow-up.

A livello proteomico, su 286 proteine valutate tramite SOMAscan, 10 proteine sono state in grado di predire al basale i progressori dai non progressori. Gli evolving Pre-SSc hanno mostrato livelli incrementati di NKp30, endostatina, bFGF, ECM1, FGF18, fibronectina 1.3, PAFAH1B2, FABP e livelli ridotti di PHI e ubiquitina1. Tramite metodica ELISA, alti livelli di endostatina e ridotti livelli sierici di PAFAH1B2 sono stati confermati nella coorte di validazione e si sono correlati con un più breve tempo di progressione.

Conclusioni: la signature dell'IFN di tipo I ha distinto i pazienti con SSc preclinica dai controlli sani e una signature NK ridotta si è associata alla progressione di malattia. È stato scoperto che proteine legate a pathways di fibrosi, organizzazione della matrice extracellulare, regolazione positiva della proliferazione cellulare, angiogenesi e trasduzione del segnale sono in grado di predire la progressione di malattia. L'endostatina è inoltre emersa come un biomarcatore valido per future indagini meccanicistiche.

Dedicata a tutti i pazienti,
che credono nella ricerca
sperando in un futuro migliore.

Indice

Introduction

Systemic sclerosis clinical aspects

Systemic sclerosis (SSc) is a systemic autoimmune disease in which microvascular impairment, fibrosis and autoantibodies production are interconnected events that lead to a progressive multi organ damage.

SSc is a considered a rare disease with a prevalence that varies from 150 to 400 cases per million and a female to male ratio of about 8:2 [1]. SSc has also a great morbidity and mortality as shown by a standardized mortality ratio of 3.5 [2]. Raynaud phenomenon (RP) is the first sign of systemic sclerosis and it is due to an uncontrolled vasospasm mainly triggered by cold temperatures. Although the specific mechanisms of the first events in SSc pathogenesis are still unclear, the continuous release of autoantigens by injured endothelial cells could further induce autoimmunity and chronic inflammation [3]. Moreover, chronic ischemia persistently activates fibroblasts, resulting in irreversible organ fibrosis [4]. In more advanced disease stages, fibrosis tends to involve more skin, lungs and the gastrointestinal tract although any organ can potentially be affected.

Fibrosis of the skin is present in the dermal layer and is associated with adipose tissue reabsorption. Fibrotic skin has a centripetal diffusion, starting from the fingers and progressively involving hands, forearms and arms. Based on skin involvement, a limited cutaneous form (lcSSc) (with skin fibrosis distal to elbows / knees with or without facial skin fibrosis) and a diffuse cutaneous form (dcSSc) (with skin fibrosis present both distally and proximally to the elbows / knees with contribution of face and trunk) are recognized [5].

Besides skin, SSc can affect multiple organs. First of all, the vasculopathy that affects principally the small vessels causing Raynaud phenomenon, an uncontrolled vasospasm, leads to ischemia of tissues especially distally with appearance of digital ulcers and pitting scars of the fingertips. Telangiectasias which are a postcapillary venules dilation, are observed on hands, face, lips and oral mucosa of scleroderma patients [6].

A pulmonary involvement is mainly represented by two forms: interstitial lung disease (ILD) and pulmonary arterial hypertension (PAH). ILD is in about the 50% of SSc, mainly in the dcSSc subset and mostly represented by the non-specific interstitial pneumonia form which can complicate into honeycombing patterns. It is also one the main causes of death in SSc [7,8]. PAH is characterized by chronic blood vessel damage, endothelial injury and inflammatory molecules production in a feedback loop that lead to pulmonary arterial

wall remodeling. The prevalence of PAH in SSc ranges from 8 to 12% (diagnosis based on right heart catheterization) and PAH is a major cause of mortality in SSc with a lower survival if compared to idiopathic PAH [9].

Wall fibrosis, muscle atrophy and consequent dysmotility of GI tract affect about a 90% of SSc patients [10]. More frequently, the upper GI tract is the one involved, manifesting lower esophageal sphincter dysfunction and altered peristalsis that both lead to gastroesophageal reflux that when chronic, can end into Barrett's syndrome [11]. Gastroparesis due to the "Watermelon stomach" or gastric vascular ectasia is a rarer manifestation. Finally, dysmotility of the intestinal tract, is associated with constipation/pseudo-obstruction or on the opposite, with bacterial overgrowth and diarrhea that can cause important malabsorption [12,13].

About a 10-15% of patients, especially dcSSc, can develop a clinical acute emergency, the scleroderma renal crisis. This complication due to kidney vascular remodeling and obliteration for thrombosis and microangiopathic hemolysis, is associated with rapid kidney function impairment and arterial hypertension [14].

Also, heart can be affected in about 45-60% SSc (higher prevalence in dcSSc subset) with variable manifestations that ranges from conduction abnormalities, coronary artery disease, pericarditis and myocardial myositis or fibrosis [15–17].

Arthritic manifestations are uncommon but possible in SSc with often involvement of tendons. Acro-osteolysis that is a bone reabsorption of distal phalanges and muscle myositis are a possible manifestation as well [18,19].

In SSc several specific and mutually exclusive auto-antibodies have been recognized such as anticentromere, anti-topoisomerase I (anti Scl-70), anti-RNA polymerase 3, antifibrillarin, anti-Th/To, anti-PM-Scl. The auto-antibodies pattern is correlated with clinical features but it is still not clear if it has a direct pathogenetic role in SSc development [20].

Classification criteria

The 2013 American college of Rheumatology/European league defined classification criteria for systemic sclerosis [21]. A total score of 9 is sufficient to classify the patients for a definite form of SSc.

According to LeRoy and Medsger criteria, a group of patients at a pre-clinical stage of disease are defined as early SSc if characterized by uniquely presenting Raynaud's phenomenon, SSc-specific autoantibodies and/or typical abnormalities at nailfold videocapillaroscopy that, following the ACR/EULAR 2013 criteria, imply a max total score of 8 [22].

ACR/EULAR criteria for the classification of systemic sclerosis [21]

Items	Sub-items	Score
Skin thickening of the fingers of both hands extending proximal to the metacarpophalangeal joints (sufficient criterion)		9
Skin thickening of the fingers^ (only count the highest score)	Puffy fingers	2
	Sclerodactyly of the fingers (distal to MCP but proximal to the PIPs)	4
Fingertip lesions^ (only count the highest score)	Digital Tip Ulcers	2
	Finger Tip Pitting Scars	3
Telangiectasia		2
Abnormal nailfold capillaries		2
Pulmonary arterial hypertension and/or Interstitial lung Disease* (*Maximum score is 2)	PAH ILD	2
Raynaud's phenomenon		3
Scleroderma related antibodies** (any of anti-centromere, anti-topoisomerase l)	Anti-centromere Anti-topoisomerase l	3
[anti-Sd 70], anti-RNA polymerase III) (**Maximum score is 3)	Anti-RNA polymerase III	

Patients having a total score of 9 or more are being classified as having definite systemic sclerosis.
^Add the maximum weight (score) in each category to calculate the total score.

Pre-clinical systemic sclerosis

As described before, pre-clinical SSc patients present solely RP, a nailfold video-capillaroscopy positive for a "scleroderma pattern", and positive SSc specific auto-antibodies, that thus do not reach the score of 9 expected

from the 2013 ACR/EULAR criteria [21,22]. This group of patients with very early signs of SSc but no other fibrotic clinical features, are definable as Pre-SSc patients (pre-clinical SSc, also named in literature as early SSc). Patients with preclinical systemic sclerosis have about a 50% risk of disease progression into a definite/established fibrotic form within 5 years of diagnosis [23]. SSc is a progressive disease and in a limited number of patients it is observable the passage from the pre-clinical phase to the initial developing of swollen fingers (scleroderma or puffy fingers) still without involvement of the rest of the skin and internal organs. The European Scleroderma Trial and Research (EUSTAR), identified this group characterized by RP, puffy swollen digits, scleroderma pattern at capillaroscopy and/or positive scleroderma specific antibodies as patients with very early diagnosis of SSc (VEDOSS) [24].

The latency of the progression from preclinical systemic sclerosis (Pre-SSc) to an established diagnosis (fulfilling completely the 2013 ACR/EULAR criteria) seem to correlate with the auto-antibody profile shown by the patients [25,26].

Currently, due to the difficulty (low number of recruitable patients) in performing an early diagnosis of SSc, research studies focused on pre-SSc patients are numerically limited. Of great interest, the studies performed on few pre-SSc cohorts so far found out the presence of immunological alterations since the very early stages of the disease suggesting that further promising investigations should be performed on this subgroup.

Pathogenesis and biomarkers: what it is known

SSc pathogenesis is still unclear and complex mechanisms are interconnected leading to SSc development. A consensus of a genetic susceptibility triggered by the exposure to environmental factors, is shared as the main initial step behind autoimmune diseases development, SSc included. Microvasculopathy with endothelial cell damage, excessive fibrosis and immune dysregulation with autoimmunity are the main processes in SSc [27]. Studies on gene, RNA expression, epigenetic and protein levels have been performed in SSc to find biomarkers for disease diagnosis, disease activity assessment and therapeutic targets.

On a gene level, SSc has a low concordance rate in monozygotic twins (4.2%) also if compared with other autoimmune diseases [28]. Genome wide association studies (GWAS) are genome analysis able to individuate the presence of genetic variants that correlate to a disease. From GWAS and immunochip studies,

SSc emerged to be characterized by multiple genetic variances of human leucocyte antigen (HLA) region associated genes, in particular class II HLA polymorphisms are highly significantly associated with SSc [29–31]. Several studies show polymorphisms among non-HLA genes that predispose to SSc susceptibility; polymorphisms of type I interferon (IFN) regulatory factors (*IRF5, IRF4, IRF7*) as well as variants of genes expressed by T lymphocytes such as *STAT4, CD247, IL12* or by B cells such as *BANK1, BLK* as well as associated to vasculopathy such as *DDX6* variant or to fibrotic pathways such as *CSK* (c-Src tyrosine kinase) and *CAV1* (Caveolin 1) [32–39]. The development of advanced technologies, the next-generation sequencing technologies (whole genome sequencing, whole exome sequencing (WES) and whole transcriptome sequencing also known as RNA sequencing (RNA-seq)) allowed more detailed and untargeted analyses enabling to find rare variants or to analyze the whole transcriptome (while previous microarrays for example, needed to know the sequence to study *a priori*). The first WES analysis performed on SSc cohort show a missense rare variant (rs55687265) as a potential causal variant in *ATP8B4*, a gene associated with increased risk of SSc; these data were not confirmed by further replication studies [40]. To summarize, from an in-silico analysis of overall SSc associated genes proposed in [41], five pathways emerge to be involved: T cell activation, NF-kB, Type I IFN, immune system process and B cell activation.

On a gene expression level, from RNA-seq and microarray analyses we know that an interferon type I signature is prominent in SSc on peripheral blood cells as well as pathways of type I IFN, Toll-like receptor cascade, p53 protein function, platelet degranulation and activation [42–45]. Studies performed on skin of SSc patients show a gene expression heterogeneity, with inflammatory and fibrotic signatures as well as "healthy controls-like" molecular subsets [46–48]. Early dcSSc skin reveals both innate and adaptive inflammatory presence with CD8 and CD4 T cell, B cell, NK and macrophage signatures. In a multi cohort study, SSc skin show a 415 gene signature correlating longitudinally with disease severity [49]. A metanalysis on skin gene expression changes in response to treatments reveals that changes occur in inflammatory signature together with the improvement of clinical features in response to five different therapies [50]. Studies on other affected tissues, such as lungs and esophagus, are rare because of difficulties in obtaining samples biopsies [51,52].

Epigenetic modifications modulate directly the gene expression and epigenetic is considerable a link between genetics and the environmental factors [53]. Epigenetics mechanisms include DNA methylation,

histone modifications, microRNAs and long noncoding RNAs (lncRNAs). In SSc, DNA methylation process with altered hypomethylation is present in fibroblasts, whole blood and T cells with a predominance in type I IFN related genes [54–57]. Histone post translational modifications (acetylation and methylation) are demonstrated in SSc fibroblasts and CD4 T cells as well [58,59]. MicroRNA are small non-coding RNAs (about 22 nucleotides) that regulate the post-transcriptional expression of protein-coding genes [60]. MicroRNA expression changes have recently been investigated in SSc. MiR-618 is increased in pDCs leading to a decrease of IRF8 and consequent release of INFα upon TLR9 [61]. In fibroblasts, miR-29a is overexpressed [62]. Regarding circulating miRNA, 21 miRNAs discriminate SSc from healthy controls and SLE patients and miR-150 is lower in SSc than healthy controls [63,64]. Finally, the role of long non-coding RNA (lncRNA) is under investigation. A recent study shows 676 lncRNA distinguishing SSc skin to healthy controls, mostly of them being antisense lncRNAs [29,65]. Many other studies on epigenetic in SSc are currently ongoing and interest on X-linked epigenetic modifications in SSc is growing [66].

The study of extracellular vesicles (EVs) that are membrane structures secreted and easily isolable from bodily fluids, containing among others, miRNAs, proteins, lipids or others molecules is promising in SSc. EVs are classified based on their biophysical properties [67,68]. Endothelial cells-derived EVs are higher in SSc patients than healthy controls and inversely correlated with digital ulcers [69].

Studies on a large-scale serum protein level are numerically limited. Technologies used until now are antibody-based or mass spectrometry-based analyses that are not able to discriminate all variants of a proteins or consent to study a limited number of proteins chosen a priori (up to hundreds) with costs and timing limits. Another issue is also the much higher broad range concentrations of proteins in fluids [70]. Recent promising technologies as the aptamer-based, can detect different concentrations and a larger numbers of molecules (up to thousands) [71]. In SSc, studies on limited number of serum/plasma proteins chosen a priori identified the presence of increased expression of adhesion molecules, platelet and endothelial/vascular markers, adipokines as well as several interleukins/chemokines (IL-6, IL-8 and others) [72,73]. A study based on aptamer technology in dcSSc, identified as upstream regulators tumor necrosis factor-α (TNF-α), interferon-γ (IFN-γ), transforming growth factor-β (TGF-β), and interleukin 13 (IL-13) [74]. From a recent comprehensive analysis on a panel of 230 proteins performed on baseline dcSSc enrolled in the Scleroderma: Cyclophosphamide Or

Transplantation Trial in comparison matched unaffected controls, ninety proteins are differentially expressed in dcSSc and eighteen correlate with modified Rodnan skin score (mRSS), a measure of skin thickness used as primary or secondary outcome measure in clinical trials. Proteins of interest correlate with clinical features such as Carbohydrate antigen 15.3 with ILD and EGFR (epidermal growth factor receptor) inversely with mRSS and proteins related to type I IFN signature are present in dcSSc [73].

What it is known in pre-clinical SSc stages

At a gene expression level, a type I IFN signature is already detected in PBMCs of Pre-SSc patients as shown by a study performed on 19 EarlySSc patients, 12 subjects with primary Raynaud, 7 patients with definitive form of SSc and 21 lcSSc compared to 30 healthy controls. The type I IFN signature is present well before the appearance of cutaneous fibrosis [75]. This observation supports the hypothesis that the pathogenetic drivers of definite SSc are already present in the early/pre-clinical stage of disease. At the best of knowledge, no other investigations of gene expression in pre-clinical SSc patients have been performed yet.

Regarding the molecular characterization of Pre-SSc at a protein level, very few studies focused their analysis on a small group of proteins chosen a priori. According to a study on 24 EarlySSc patients, 48 definite SSc patients, and 24 osteoarthritis/fibromyalgia patients, endothelial, T-cell and fibroblast activation and a marked increase of IL-33 levels are present in SSc preclinical patients [76]. A more recent study identified significantly elevated serum levels of CXCL10, CXCL11, TNFR2 and CHI3L1 in EarlySSc and notably, CXCL10 and CXCL11 are Type I IFN inducible proteins in SSc as previously demonstrated by [77,78]. Moreover, baseline levels of CXCL10 and TNFR2 correlated with a shorter clinical time of progression from the pre-clinical to the fibrotic stage of SSc. Further studies on the protein profile of Pre-SSc are needed, moreover proteomics studies or studies using aptamer technologies have not been yet performed in the pre-clinical stage of SSc patients.

New gene expression studies and molecular characterization of Pre-SSc would be of great interest and importance to understand and intercept the pathology before its definitive evolution.

Project Aims

Considering the opportunity time-window between Raynaud's Phenomenon appearance and the definitive organ involvement, early diagnosis of SSc is of extreme importance both from a clinical and a biomolecular point of view. The present project aims to perform at the same time a clinical characterization of pre-clinical SSc patients along with a gene expression and a proteomic analysis in order to identify specific early-stage disease biomarkers.

Specifically, at a gene expression level, confirmation of changes in the type I IFN pathways and the discovery of other signatures in the pre-clinical stages of SSc may be of interest to develop new therapies or to retrospectively corroborate the role of widely used therapies in SSc, but whose preventive significance is currently unknown.

The identification of biomarkers eventually correlated with the progression into a definitive SSc subset is also explored. Moreover, it is also an aim of the present project to perform in this specific rare subset of patients with pre-clinical SSc a more comprehensive proteomic analysis of a broader panel of proteins not chosen a priori using a new aptamer technology.

Methods

Patients and Methods

Clinical features and blood samples of 35 patients at a pre-clinical stage (Pre-SSc) were collected consecutively at baseline and re-collected after four years. Pre-SSc patients were defined according to LeRoy and Medsger criteria for the classification of early systemic sclerosis [22] that is the presence of Raynaud Phenomenon plus SSc‐specific autoantibodies and/or SSc‐specific nailfold videocapillaroscopic changes without any other sign of definite SSc and/or fibrosis [21] thus at high risk of developing SSc [23]. Clinical and laboratory data at 4 years were available to determine whether Pre-SSc patients had progressed to definite SSc according to the ACR/EULAR criteria [21] or whether they had remained stable at a pre-clinical stage. Patients with clinical progression were then classified as definite SSc in absence of fibrotic skin disease, or diffuse cutaneous SSc (dcSSc) or limited cutaneous SSc (lcSSc) accordingly to the extent of skin fibrosis [22]. Blood samples of 16 age- and ethnicity matched healthy controls were also collected. Controls had no history of systemic autoimmune diseases or concomitant relevant diseases such as diabetes, cancer or infectious diseases and were not relatives of the Pre-SSc patients enrolled.

A cohort of 50 Pre-SSc whose serum samples were consecutively collected at baseline was available for a subsequent validation analysis. As already described above for the 35 Pre-SSc, the validation cohort was characterized by patients with preclinical features of SSc in absence of fibrotic skin disease at baseline and whose clinical prospective data were available to determine who had progressed and who did not. Data and samples collection was approved by the local Ethical Committee Comitato Etico Milano Area 2; patients provided signed informed consent. Gene expression and proteomic analyses were conducted in collaboration with the University of Texas Health Science Center at Houston (UTH).

RNA sequencing

Whole blood samples were collected in PAX gene tubes and stored at -80° C. RNA was extracted according to Manufacturer's protocol and quality was assessed by Agilent 2100 Bioanalyzer (Agilent Technologies). Globin genes were depleted using GLOBIN Clear kit. mRNA was enriched from total RNA using oligo(dt)

beads (NEB Next Ultra II RNA Kit following the poly(A) enrichment workflow). The mRNA was subsequently fragmented randomly in fragmentation buffer, and reverse transcribed to cDNA. The cDNAs were converted to double stranded cDNAs, then subjected to end-repair, A-tailing, and adapter ligation, size selection and PCR enrichment. Library concentration was first quantified using a Qubit 2.0 fluorometer (Life Technologies), and then diluted to 1 ng/ul before checking insert size on an Agilent 2100 and quantifying to greater accuracy by quantitative PCR (q-PCR) (library activity >2nM). Libraries were pooled into Novaseq6000 machines according to molarity and expected data volume. A paired-end 150 bp sequencing strategy was used to generate an average of 88 million reads per sample. Transcript data were normalized with DESeq2. Transcripts were filtered with count-per-million (CPM) keeping those genes who have at least 2 samples with 'cpm greater than 1'. A total of 17746 transcripts out of 60641 in raw count file passed the filtering criteria. Transcripts were considered as differentially expressed if $p<0.05$ and Log2 fold change >0.2 (or less than -0.2). An unpaired analysis was used for comparison of group of subject samples at baseline and a paired analysis was performed for longitudinal samples.

Modular analysis

Modular analysis using 62 curated whole blood modules was conducted using the original repertoire analysis described in [79] and in [80]. At a population level the analysis summarizes each module by recording the percentage of statistically up and down regulated genes within the corresponding module in comparison to the reference group [81]. A similar approach can be done at the sample level. The resulting proportions are plotted using circles and color-coded red for up and blue for down regulation. In addition, a gene set analysis was conducted using the QuSAGE algorithm for the modular analysis of differentially expressed genes [82]. QuSAGE tests whether the average log2 fold change of a gene set is different from zero. The method correctly adjusts for gene-to-gene correlations within a gene set and provides an easy interpretable metric for the magnitude of differential regulation. A threshold value of Log2 fold change >0.2and $p<0.05$ was used to identify differentially expressed module.

Differential expression pre-post analysis and validation

Based on [83] effect estimates were calculated via ANCOVA models using the change in RNA expression values (post values minus pre values corrected for baseline values) as outcomes ("ANCOVA change" method). A 10.000-fold permutation strategy (min P strategy) was used to account for multiple testing and to control the family-wise error rate [84].

SOMAscan analysis

To study circulating proteins (proteomics), samples from a subgroup of patients and controls were analyzed via the SomaScan® assay. Specifically, out of the overall 35 Pre-SSc, sixteen patients (8 stable Pre-SSc versus 8 evolving Pre-SSc) and 8 matched healthy controls were initially selected on a clinical basis, matching for age, gender and autoantibody profile and aliquots of sera were stored at -80° C. SOMAscan assay was performed as described in [85]. All samples were clarified by centrifugation before use. These samples were screened using an aptamer-based screening platform pioneered by Somalogic. This assay uses aptamer–protein interactions to detect proteins within a sample. In the assay, aptamer-coated streptavidin beads are first added to the sample to allow the aptamers to bind to the proteins. Next, the bound proteins are biotinylated, and the aptamer–protein complexes are cleaved from the streptavidin beads. These aptamer–protein complexes are then conjugated to a second streptavidin bead, and aptamers are separated from the proteins. The aptamers are then collected from the sample and quantitated by hybridization to a DNA microarray. The final output is the relative fluorescence unit (RFU) for each protein; these RFU values were then normalized and statistically analyzed. The limit of detection (LOD) of the aptamer-based scan was determined by spiking proteins into buffer before the assay. The limits of quantitation (LOQ) were established along with the LOD, and the median lower LOQ value is approximately 3-fold higher than the LOD.

ELISA validation analysis

The role of the 10 selected proteins detected with the SOMAscan and significantly associated with discriminating progressors vs non-progressor was then verified via ELISAs in a cohort of 50 Pre-SSc with available baseline serum and prospective data.

Serum levels of NKp30, Endostatin, bFGF, ECM1, FGF18, PHI, FN1.3, Ubiquitin+1, PAFAH-beta subunit, FABP were detected through ELISA assay accordingly to the manufacturer instructions. The ELISA assay, employs the quantitative sandwich enzyme immunoassay technique. A monoclonal antibody specific for human analytes has been pre-coated onto a microplate. Standards and samples are pipetted into the wells and any analyte present is bound by the immobilized antibody. After washing away any unbound substances, an enzyme-linked polyclonal antibody for human specific analyte is added to the wells. Following a wash to remove any unbound antibody-enzyme reagent, a substrate solution is added to the wells and color develops in proportion to the amount of specific analyte bound in the initial step. The color development is stopped and the intensity of the color is measured by spectrophotometer.

Statistical analysis

Results of SOMAscan sample assay were expressed as relative fluorescence units (RFU) which are proportional to the amount of target protein in the sample.

Principal component analysis (PCA) was used to identify outliers on the basis of RFU values. Overall, out of the 16 Pre-SSc selected, three cases were excluded and thus 7 progressors (evolving Pre-SSc) and 6 non-progressors (stable Pre-SSc) completed the analysis.

Univariate analysis was performed via prediction models in proteins with RFU > 1.5-fold or < 1/1.5-fold HCs values; overall 286 proteins were selected for the analysis. Bootstrap aggregating was used to determine the accuracy of predictions (progression vs non-progression) of categorized baseline protein values whose empirical p-values were determined via a 10.000-fold step-down permutation approach. Ten proteins were significantly associated with disease progression at the 0.05 threshold. Gene Ontologies (GO terms) of selected proteins with corrected $p<0.05$, were aggregated to determine the relevance of biological processes in patients at risk of evolution.

The estimated time-to-evolution according to the Turnbull method (survival analysis for interval-censored data) was performed for the two proteins that resulted confirmed at ELISA (Endostatin and PAFAH1B2 (Platelet-activating factor acetylhydrolase IB subunit beta)).

Results

Demographic data

Among the 35 Pre-SSc subjects included in the study, 15 (42.9%) presented signs of progression toward a definite SSc at 4 years. Baseline demographic characteristics of Pre-SSc patients, stratified by 4-year evolution outcome, are presented in **Table 1**. At baseline, patients who did progress (evolving Pre-SSc) and who did not (stable Pre-SSc) presented a comparable age, gender, auto-antibody profile, Raynaud's phenomenon duration and pulmonary function test percentage of predicted values; as per inclusion criteria patients did not have any skin involvement including puffy fingers; upper GI tract symptoms, mainly gastroesophageal reflux disease, were referred in a small percentage. No patient was treated with steroids, immunosuppressants or biologicals at the time of baseline evaluation and thereafter; evolving Pre-SSc and stable Pre-SSc were treated with antiplatelets agents and calcium channels blockers in a comparable percentage.

Features of progression at 4-years included puffy fingers in 60% of cases, skin fibrosis in 26.7% and teleangectasia, alone or in combination with the above, in 33%. Progression was not associated in this cohort with significant differences for mostly of clinical features, including disease duration considered as RP appearance (mean of both group equal to 13 years at follow-up) and FVC %. See **Table 2**.

Table 1: Baseline clinical characteristics of Pre-SSc patients along with stratification according to the 4-years evolution outcome

Features	All patients (35)	Stable Pre-SSc (20)	Evolving Pre-SSc (15)	HC (16)
Age, mean years (SD)	56,9 (13.3)	58.4 (11.6)	54.9 (15.4)	55.3 (12.6)
Gender, Female n (%)	30 (85.7)	16 (80)	14 (93.3)	15
Ethnicity Caucasian n (%)	35 (100)	20 (100)	15 (100)	16 (100)
ANA n (%)	33 (94.3)	18 (90)	15 (100)	N.A.
ACA n (%)	22 (62.9)	12 (60)	10 (66.7)	N.A.
Anti-Scl70	7 (20)	3 (15)	4 (26.7)	N.A.
Other	4 (11.4)	3 (15)	1 (6.7)	N.A.
RP years duration mean (SD)	9.3 (7.9)	9.6 (8.4)	8.8 (7.5)	N.A.
FVC (%) mean (SD)	116.1 (17.5)	115.9 (19.6)	116.3 (14.8)	N.A.
DLCO (%) mean (SD)	86.7 (18.3)	86.8 (19.9)	86.7 (16.8)	N.A.
Evolution n (%)	0	0	0	N.A.
SSc clinical features				
None	18 (51.4)	11 (55)	7 (46.7)	N.A.
Skin n (%)	0	0	0	N.A.
Upper GI n (%)	10 (28.6)	6 (30)	4 (26.7)	N.A.
Teleangectasia n (%)	0	0	0	N.A.
Cardioaspirin n (%)	30 (85.7)	20 (100)	10 (66.6)	N.A.
CCB n (%)	25 (71.4)	16 (80)	9 (60)	N.A.

Abbreviations: FU: follow-up; ANA: anti-nuclear antibodies; ACA: anti-centromere antibodies; RP: Raynaud Phenomenon; FVC: forced vital capacity; DLCO: diffusing capacity of the lung for carbon monoxide; N: number; GI: gastro-intestinal; CCB: calcium channel blockers; NA: not applicable; SD: Standard deviation.

Table 2: Subjects clinical features at follow-up

Features	Stable Pre-SSc 1 (20)	Evolving Pre-SSc (15)
Age, mean years (SD)	62.3 (11.6)	59,1 (15.4)
Gender, Female n (%)	16 (80)	14 (93.3)
Ethnicity Caucasian n (%)	20 (100)	15 (100)
ANA n (%)	18 (90)	15 (100)
ACA n (%)	12 (60)	10 (66.7)
Anti-Scl70	3 (15)	4 (26.7)
Other	3 (15)	1 (6.7)
RP years duration mean (SD)	13.5 (8.6)	13 (7.6)
FVC (%) mean (SD)	118 (17.6)	117.6 (17.2)
DLCO (%) mean (SD)	87 (23.4)	78.8 (14.6)
Evolution n (%)	0	15 (100)
SSc clinical features		
None	9 (45)	0
Skin n (%)	0	13 (86.7)
Only Puffy fingers	0	9 (60)
lcSSc features	0	4 (26.7)
dcSSc	0	0 (0)
Upper GI n (%)	6 (30)	9 (60)
Teleangectasia n (%)	1 (5)	5 (33.3)
Cardioaspirin n (%)	15 (75)	13 (86.7)
CCB n (%)	16 (80)	12 (80)

Abbreviations: ANA: anti-nuclear antibodies; ACA: anti-centromere antibodies; RP: Raynaud Phenomenon; FVC: forced vital capacity; DLCO: diffusing capacity of the lung for carbon monoxide; N: number; GI: gastro-intestinal; CCB: calcium channel blockers; NA: not applicable; SD: Standard deviation

RNASeq gene expression data

Peripheral blood cell global gene expression profile of overall baseline Pre-SSc samples to matched healthy controls revealed 589 differentially expression genes (DEGs) with FDR<0.05. A comparison between stable Pre-SSc and evolving Pre-SSc subsets at baseline did not show any significant difference (FDR<0.05).

Analysis of change (post-values minus pre-values corrected for baseline values) in RNA expression values identified 73 genes with a corrected p value ≤0.05. The full list of significant genes is reported in **Table 3** along with the change in post - pre expression values and the directionality of change (positive values indicate an increase in transcription while low negative values a decrease)**.** Genes related to collagen type VI such as

COL6A2, the TNF receptor superfamily TNFRSF18, granzyme proteins GZMB, leukotriene receptor LTB4R2, and CXCR6 are among the list.

Table 3: RNA expression change (pre-post) of evolving Pre-SSc and stable Pre-SSc

Gene name	Protein Name	p value*	Change in evolving Pre-SSc §	Change in stable Pre-SSc §	Direction Favours
AC010326.5	N.A.	0.02	0.2179	0.0197	Evo. Pre-SSc
NODAL	Nodal homolog	0.02	0.2195	0.0287	Evo. Pre-SSc
FAM169B	Protein FAM169B	0.02	0.1214	-0.1521	Evo. Pre-SSc
UAP1	UDP-N-acetylhexosamine pyrophosphorylase	0.02	-0.09	0.1289	Stable Pre-SSc
AL137009.2	N.A.	0.02	0.2616	0.0666	Evo. Pre-SSc
COL6A2	Collagen alpha-2(VI) chain	0.02	-0.165	0.0359	Stable Pre-SSc
AC040162.1	N.A.	0.02	0.1863	0.0442	Evo. Pre-SSc
SAP30-DT	N.A.	0.02	-0.2348	0.1701	Stable Pre-SSc
AVIL	Advillin	0.02	0.1637	0.0816	Evo. Pre-SSc
TNFRSF18	Tumor necrosis factor receptor superfamily member 18	0.02	-0.1282	0.0397	Stable Pre-SSc
FGFBP2	Fibroblast growth factor-binding protein 2	0.02	-0.1511	0.0632	Stable Pre-SSc
PPP2R2B	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B beta isoform	0.03	-0.0106	0.0908	Stable Pre-SSc
MARF1	Meiosis regulator and mRNA stability factor 1	0.03	0.1238	-0.0495	Evo. Pre-SSc
GZMB	Granzyme B	0.03	-0.2022	0.0509	Stable Pre-SSc
MATK	Megakaryocyte-associated tyrosine-protein kinase	0.03	-0.1927	0.0391	Stable Pre-SSc
DUSP2	Dual specificity protein phosphatase 2	0.04	-0.2153	0.055	Stable Pre-SSc
PRF1	Perforin-1	0.04	-0.1019	0.0413	Stable Pre-SSc
AC103706.1	N.A.	0.05	-0.1885	0.0425	Stable Pre-SSc
AC011346.1	N.A.	0.05	0.3334	-0.0841	Evo. Pre-SSc
KLHL18	Kelch-like protein 18	0.05	0.222	0.0083	Evo. Pre-SSc
NAP1L4P1	N.A.	0.05	0.4158	0.1554	Evo. Pre-SSc
LINC01547	N.A.	0.05	0.1566	0.075	Evo. Pre-SSc
NCOA4	Nuclear receptor coactivator 4	0.05	0.2043	0.0452	Evo. Pre-SSc
AC009962.1	N.A.	0.05	0.4363	0.0313	Evo. Pre-SSc
AC020765.4	N.A.	0.05	0.1758	0.0333	Evo. Pre-SSc
KLRF1	Killer cell lectin-like receptor subfamily F member 1	0.05	-0.0578	0.1268	Stable Pre-SSc
HVCN1	Voltage-gated hydrogen channel 1	0.05	0.1382	-0.0322	Evo. Pre-SSc
CLIC3	chloride intracellular channel protein 3	0.05	-0.1727	0.0312	Stable Pre-SSc

AL121753.1	N.A.	0.05	0.23	-0.0838	Evo. Pre-SSc
MICAL2	[F-actin]-monooxygenase MICAL2	0.05	0.2119	0.0235	Evo. Pre-SSc
AL133268.1	N.A.	0.05	0.3593	0.1676	Evo. Pre-SSc
HENMT1	Small RNA 2'-O-methyltransferase	0.05	-0.1135	0.1442	Stable Pre-SSc
NKG7	Natural killer cell protein 7	0.05	-0.1514	0.0163	Stable Pre-SSc
SLC31A1	High affinity copper uptake protein 1	0.05	0.2143	-0.0508	Evo. Pre-SSc
USP4	Ubiquitin carboxyl-terminal hydrolase 4	0.05	0.179	0.0564	Evo. Pre-SSc
AC092718.4	N.A.	0.05	-0.3892	-0.0112	Stable Pre-SSc
TRDC	T cell receptor delta constant	0.05	-0.0736	0.0247	Stable Pre-SSc
HOPX	Homeodomain-only protein	0.05	-0.0683	0.151	Stable Pre-SSc
RGS9	Regulator of G-protein signaling 9	0.05	-0.1044	0.0363	Stable Pre-SSc
AL031733.2	N.A.	0.05	-0.0047	0.0848	Stable Pre-SSc
AC010285.3	N.A.	0.05	0.3875	0.1247	Evo. Pre-SSc
PLAAT3	Phospholipase A and acyltransferase 3	0.05	-0.0592	0.0427	Stable Pre-SSc
RAB38	Ras-related protein Rab-38	0.05	-0.2426	0.0763	Stable Pre-SSc
MIR4453HG	N.A.	0.05	0.1278	0.0245	Evo. Pre-SSc
LGR6	Leucine-rich repeat-containing G-protein coupled receptor 6	0.05	-0.0751	0.0567	Stable Pre-SSc
RRM2P3	N.A.	0.05	0.2359	0.0077	Evo. Pre-SSc
PPM1H	Protein phosphatase 1H	0.05	0.186	0.0347	Evo. Pre-SSc
AC097382.3	N.A.	0.05	-0.0838	0.1019	Stable Pre-SSc
NCALD	Neurocalcin-delta	0.05	-0.0679	0.0671	Stable Pre-SSc
AL135818.1	N.A.	0.05	-0.1597	-0.0654	Stable Pre-SSc
AC092535.5	N.A.	0.05	-0.1043	0.0607	Stable Pre-SSc
AL137792.1	N.A.	0.05	0.2206	-0.0464	Evo. Pre-SSc
FSD1	Fibronectin type III and SPRY domain-containing protein 1	0.05	-0.0854	0.0757	Stable Pre-SSc
LRRC43	Leucine-rich repeat-containing protein 43	0.05	-0.1593	0.1545	Stable Pre-SSc
ID2	DNA-binding protein inhibitor ID-2	0.05	-0.0502	0.1008	Stable Pre-SSc
NHLH1	Helix-loop-helix protein 1	0.05	0.2717	-0.0457	Evo. Pre-SSc
AC022706.1	N.A.	0.05	-0.1551	0.0351	Stable Pre-SSc
TTC3P1	N.A.	0.05	0.1902	-0.0192	Evo. Pre-SSc
ARMC12	Armadillo repeat-containing protein 12	0.05	-0.3018	0.1106	Stable Pre-SSc
LTB4R2	Leukotriene B4 receptor 2	0.05	0.0831	0.0299	Evo. Pre-SSc
LIM2	Lens fiber membrane intrinsic protein	0.05	-0.0598	0.0687	Stable Pre-SSc
PLAAT5	Phospholipase A and acyltransferase 5	0.05	-0.2276	0.0476	Stable Pre-SSc
SH2B3	Lymphocyte-specific adapter protein Lnk	0.05	0.1717	0.009	Evo. Pre-SSc
PDCD6IP	Programmed cell death 6-interacting protein	0.05	0.2392	0.1306	Evo. Pre-SSc
AC009299.3	N.A.	0.05	0.3744	0.1299	Evo. Pre-SSc
GAB1	Growth factor receptor bound protein 2-associated protein 1	0.05	0.2915	0.0831	Evo. Pre-SSc

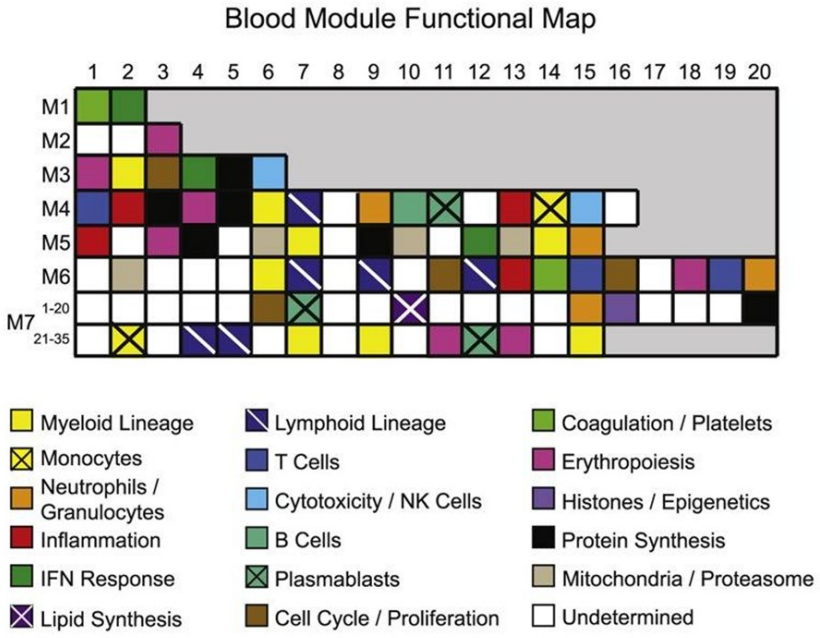
AL683842.1	N.A.	0.05	0.1767	0.0304	Evo. Pre-SSc
AL355490.2	N.A.	0.05	0.1922	-0.0209	Evo. Pre-SSc
GZMK	Granzyme K	0.05	-0.0545	0.0443	Stable Pre-SSc
CXCR6	C-X-C chemokine receptor type 6	0.05	-0.1165	0.0221	Stable Pre-SSc
NUP50	Nuclear pore complex protein Nup50	0.05	0.1958	0.1112	Evo. Pre-SSc
TMEM164	Transmembrane protein 164	0.05	0.1676	0.0028	Evo. Pre-SSc
GNLY	Granulysin	0.05	-0.1472	0.012	Stable Pre-SSc

Abbreviations. N.A. not applicable; Evo, evolving; * ANCOVA permutation p value. § Positive values indicate an increase in transcription while low negative values a decrease

Identification of differentially expressed modules

Sixty-two gene expression modules observed in whole blood across a variety of inflammatory and infectious diseases were investigated and a biological function was assigned to each module based on the function of genes present in that module. Some of the modules remained undetermined. See **Figure 1**.

Figure 1

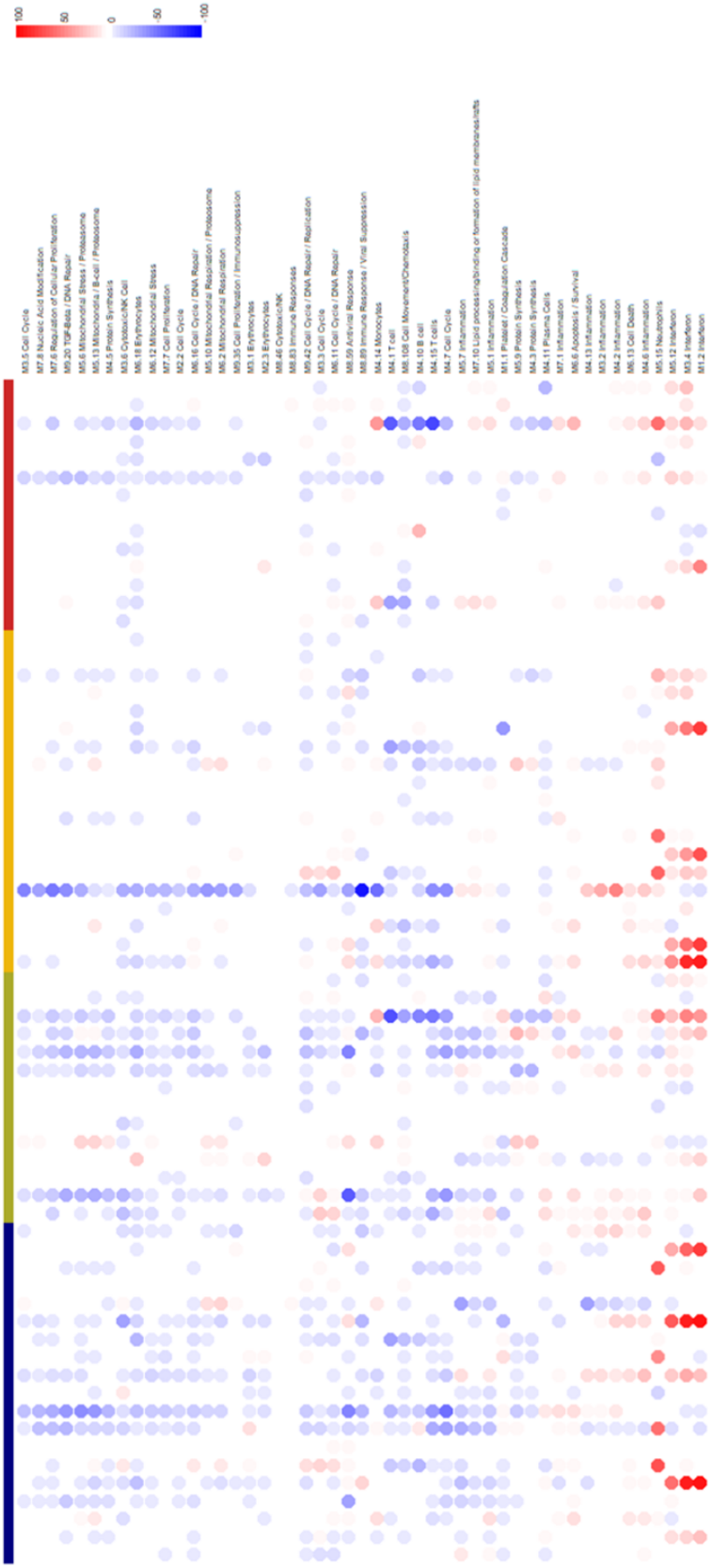


Annotation of modules based on known biological function of genes included in a given module

Module map of differentially expressed modules in baseline and follow-up patient's samples compared to healthy controls is presented in **Figure 2**. An increase of the IFN type I modules is observable almost only in stable Pre-SSc patients at baseline and in those subjects who did not evolve at follow-up in comparison to

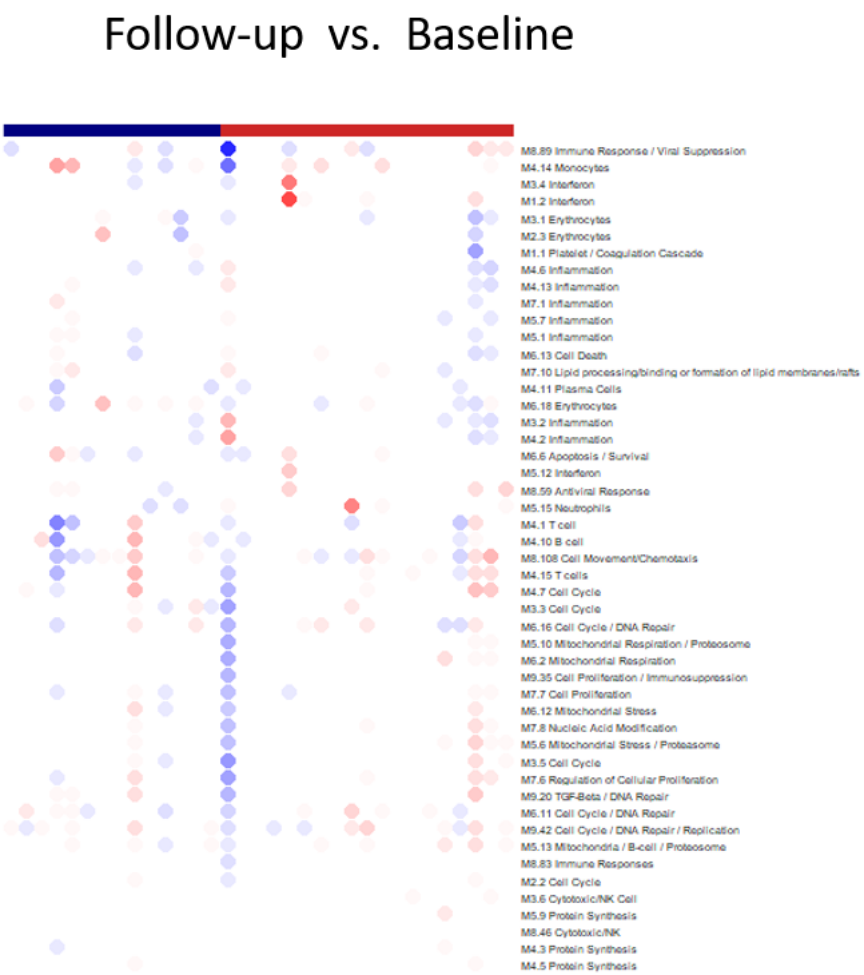
healthy controls. Module maps of differentially expressed modules of follow-up vs baseline patients as well as follow-up patients compared to healthy controls are represented in **Figure 3a and 3b.**

Figure 2



Rows and columns represent modules and subject samples respectively. The top bar represents the subject groups according to the classification at follow-up and time of sampling. Blue: stable Pre-SSc at baseline, Orange: stable Pre-SSc at follow-up, Green: evolving Pre-SSc patients at baseline; Red: evolving Pre-SSc patients at follow-up.

Figure 3 a)



Rows and columns represent modules and subject samples respectively. The top bar represents the subject groups. Blue: patients who did progress to definite SSc at follow-up (evolving Pre-SSc), Red: patients who did not progressed and still presented preclinical features at follow-up (stable Pre-SSc).

Figure 3 b)

Rows and columns represent modules and subject samples respectively. The top bar represents the subject groups. Blue: patients who did progress to definite SSc at follow-up (evolving Pre-SSc), Red: patients who did not progressed and still presented preclinical features at follow-up (stable Pre-SSc).

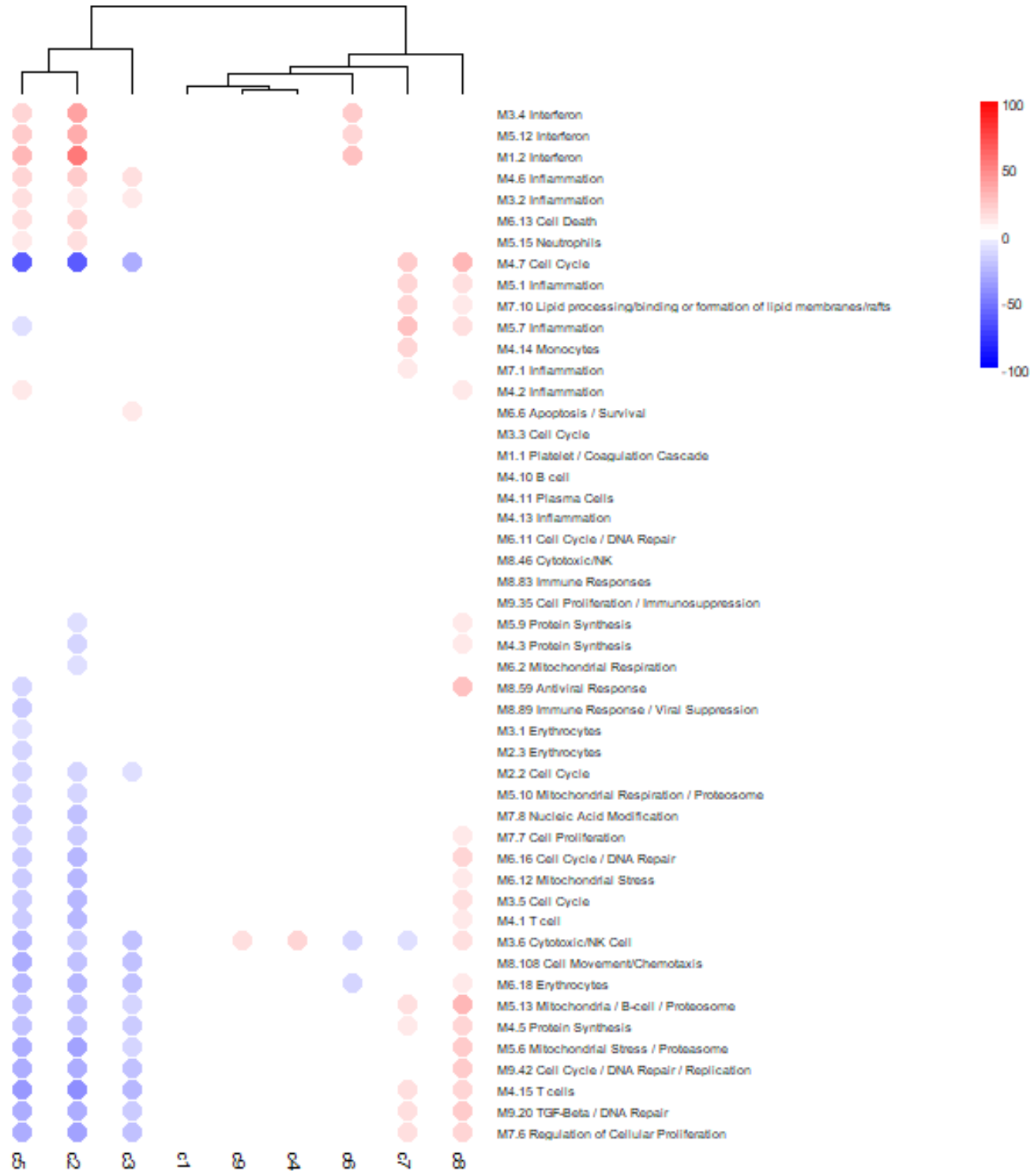
Nine comparisons were performed as summarized in **Table 4.** In brief, differentially expressed gene (DEG) and differentially modules expression of stable Pre-SSc and evolving Pre-SSc both at baseline and follow-up were compared to healthy controls, or baseline vs follow-up within the same subset groups (eg. Stable Pre-SSc baseline vs stable Pre-SSc follow-up) or comparing the two disease subset groups (eg. Stable Pre-SSc baseline vs evolving Pre-SSc baseline). Exception for C5 (comparison number 5), in the other 8 comparisons, although raw p values showed differentially expressed genes, DEG did not show statistical significance if adjusted for FDR or Bonferroni. Modular analysis showed an increase of the IFN modules (M) M3.4, M5.12 and M1.2 in comparisons C2, C5 and C6 and a decrease of the Cytotoxic/NK module M3.6 in C3, C2, C5 with its increase in C4, C8 e C9. (**see Figure 4**).

Table 4 : the nine comparisons (C1 to C9) between subjects groups (Stable Pre-SSc, Evolving Pre-SSc and healthy controls)

Comparisons		**Raw**	**FDR**	**Bonf**
C1	Stable Pre-SSc baseline vs. Evolving Pre-SSc baseline	354	0	0
C2	Stable Pre-SSc baseline vs. HC	4857	0	0
C3	Evolving Pre-SSc baseline vs. HC	2911	0	0
C4	Stable Pre-SSc follow-up vs. evolving Pre-SSc follow-up	554	0	0
C5	Stable Pre-SSc + evolving Pre-SSc baseline vs. HC	4641	589	0
C6	Stable Pre-SSc + evolving Pre-SSc follow-up vs. HC	642	0	0
C7	Evolving Pre-SSc follow-up vs. evolving Pre-SSc baseline	3148	0	0
C8	Stable Pre-SSc follow-up vs. stable Pre-SSc baseline	2442	0	0
C9	Stable Pre-SSc follow-up + baseline vs. evolving Pre-SSc follow-up + baseline	392	0	0

Abbreviations: Raw: raw p values; FDR: false discovery rate; Bonf: Bonferroni.

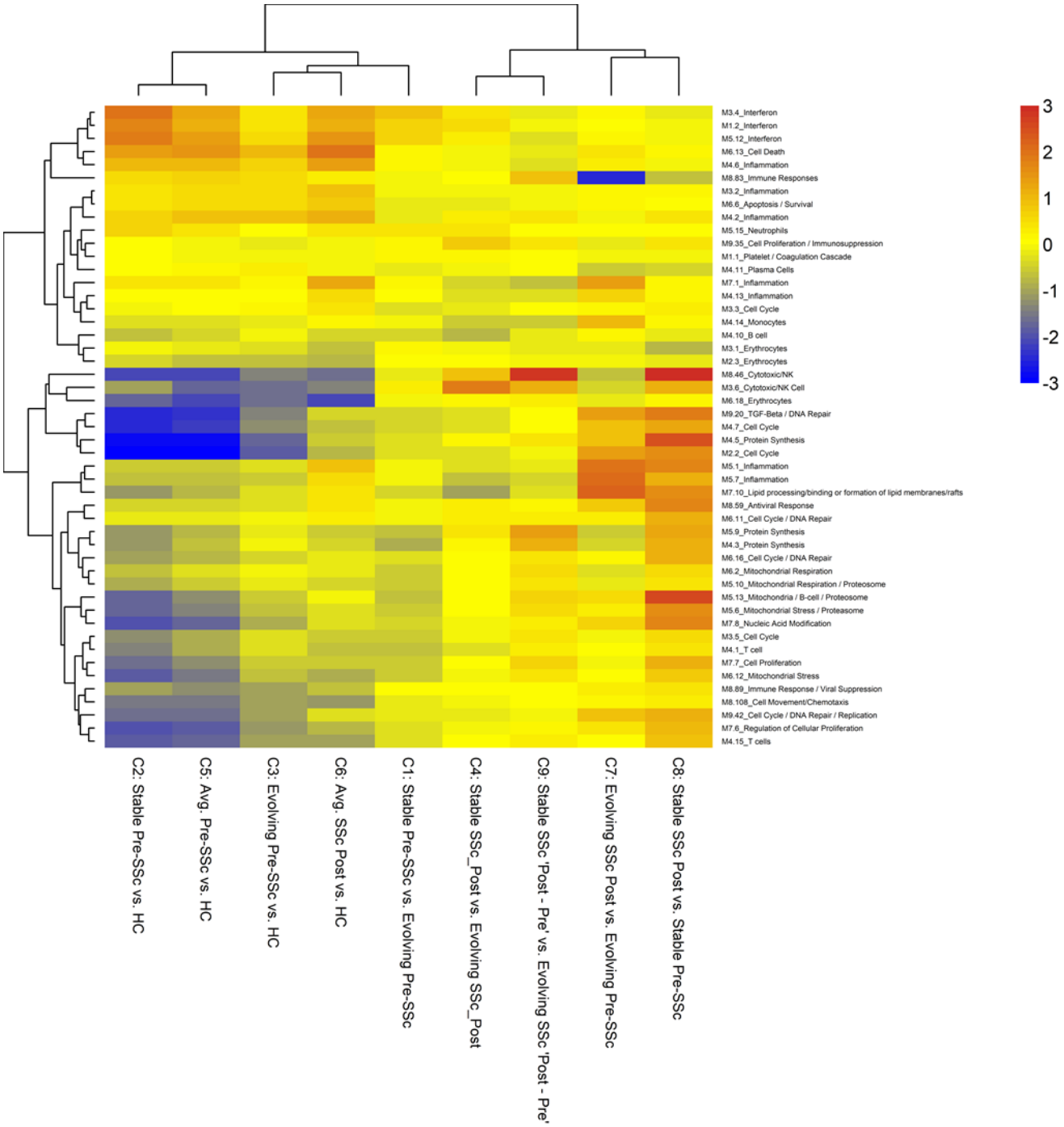
Figure 4



Rows represent modules and columns represent comparisons (C1 to C9).

Comparable results regarding the Cytotoxic/NK module M3.6 and M8.46 emerged also from the QuSage analysis as shown in the heatmap of **Figure 5**

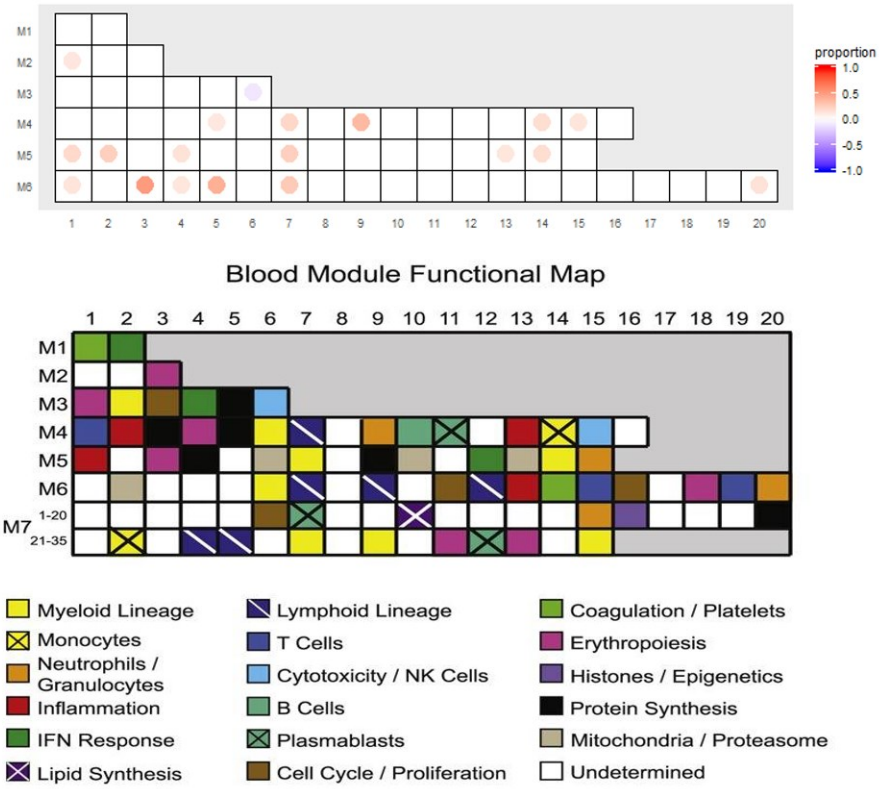
Figure 5



QuSage Heatmap of -log10pvalue *sign(logFC).Rows represent modules and columns represent comparisons (C1 to C9).

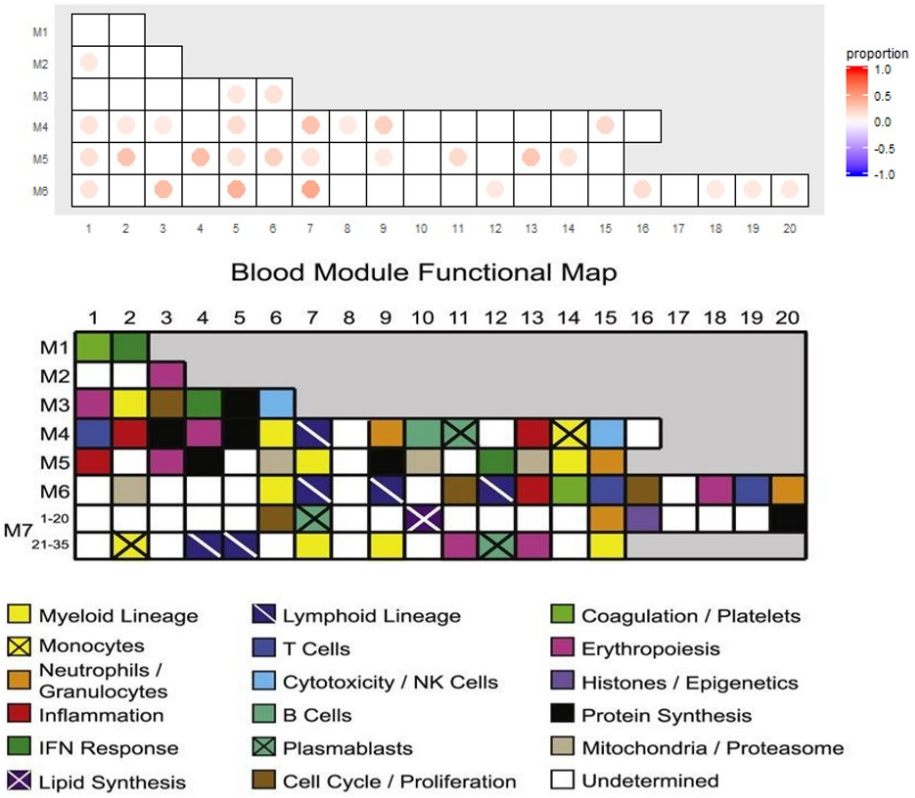
Modular analysis for C7 (evolving Pre-SSc follow-up vs. evolving Pre-SSc baseline) and for C8 (stable Pre-SSc follow-up vs. stable Pre-SSc baseline) are shown in **Figure 6 and 7.** From modular analysis of C7, evolving Pre-SSc patients were characterized by a reduced cytotoxic/NK module (M 3.6) and increased neutrophils/granulocyte module M 4.9 and M 6.20 as well as by other modules undetermined (**Figure 6**). From modular analysis of C8, stable Pre-SSc patients kept an increased lymphoid lineage (M 4.7 and M 6.7) and a weakly increased module M 5.13 mitochondria/proteasome as well as other undetermined modules, while cytotoxic/NK module (M 3.6) and neutrophils/granulocyte module M 4.9 were slightly increased (**Figure 7**).

Figure 6



Evolving Pre-SSc followup vs. evolving Pre-SSc baseline

Figure 7



Stable Pre-SSc follow-up vs. stable Pre-SSc baseline

SOMAscan analysis

Out of 16 Pre-SSc selected for protein level assessment, three outliers were excluded from the analysis as described in the methods section. On the remaining subjects (n=13), after 4 years of observation, 7 (53.8%) did evolve and 6 did not evolve. Non progressors and progressors were similar regarding baseline characteristics (ACA+ in 66% vs 57%; FVC % 105 [97-102] vs 110 [109-115]; DLCO %, 92 [87-105] vs 85 [82-101]). The 13 subjects were compared to HC matched for age and gender (**Table 5)**. In SOMAscan assay 286 proteins (those with RFU > 1.5-fold or < 1/1.5-fold HCs values) were selected for the analysis. Of these, 10 proteins were significantly associated at baseline with disease progression at the 0.05 threshold (see **Figure 8**).

Their relative RFUs are represented via heatmaps in **Figure 9**, **left panel**. The clustering well separated patients that will progress from those who will not (stable Pre-SSc). GO term analysis revealed that patients at risk for progression shared several biological processes related to fibrosis, vascular function and angiogenesis and that these were upregulated compared to non-progressors (**Figure 9**, **right panel**).

Table 5: demographic and autoantibody profile of Pre-SSc vs healthy controls analyzed with SOMAscan assay

Features	Pre-SSc (13)	HC (8)
Age, mean years (SD)	53.5 (6.3)	55.8 (4.1)
Gender, Female n (%)	10 (76.9)	7 (87.5)
Ethnicity, n (%)		
Caucasian	13 (100)	8 (100)
Autoantibodies n (%)		
ANA	13 (100)	N.A.
ACA	8 (61.5)	N.A.
ANA nu	2 (15.4)	N.A.
Anti-Scl70	3 (23.1)	N.A.
Time of observation yrs (SD)	4 (0.6)	N.A.

Abbreviations: ANA: anti-nuclear antibodies; ACA: anti-centromere antibodies; N: number; NA: not applicable; SD: Standard deviation

Figure 8

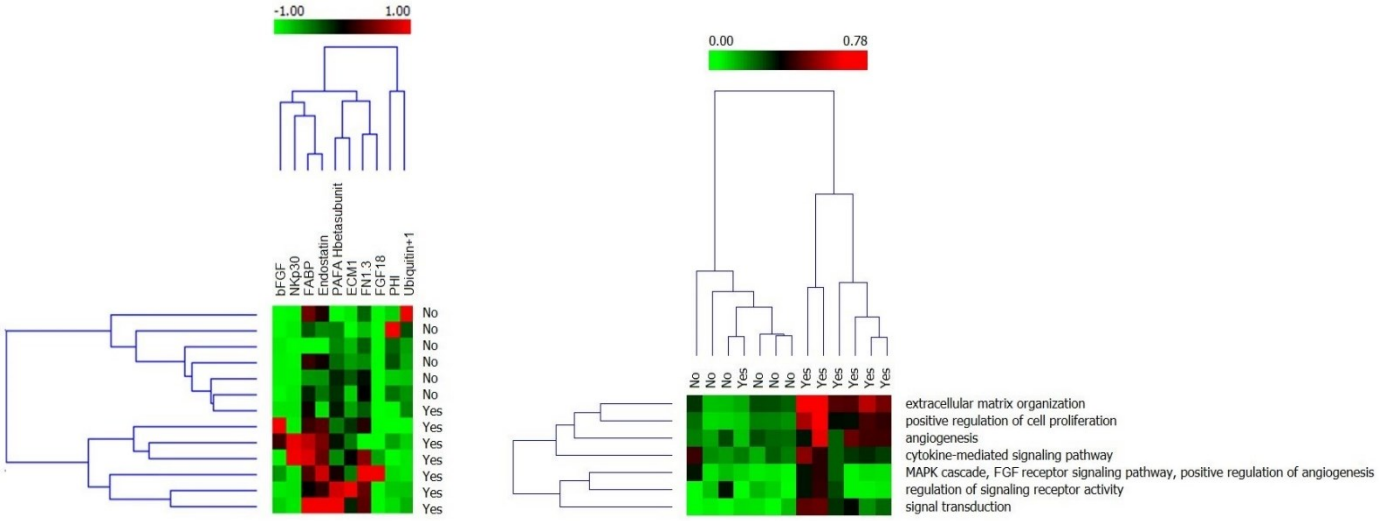
Target	Gene	RFU* Non progressors (n=6)	RFU* Progressors (n=7)	Risk threshold*	Accuracy**	Perm P**
NKp30	NCR3	458,9 (439,7 - 483)	909,8 (541,4 - 1749,8)	≥ 512,434	0.835	0.0071
Endostatin	COL18A1	22628,7 (20458,3 - 25445,5)	30969 (28044,7 - 33112,1)	≥ 27514,39	0.835	0.0075
bFGF	FGF2	1965,3 (1811,6 - 2174,5)	6861,4 (3590,6 - 13495,8)	≥ 2791,214	0.83	0.0075
ECM1	ECM1	5892,1 (5317,5 - 6558,2)	8591,8 (7539 - 10011,6)	≥ 7222,982	0.766	0.0248
FGF 18	FGF18	344,6 (327,1 - 367)	796,2 (393,1 - 3761,6)	≥ 369,586	0.755	0.0324
PHI	GPI	665,5 (547 - 871,2)	390 (349,2 - 449,4)	≤ 469,331	0.749	0.0345
Fibronectin 1.3 (FN1.3)	FN1	1510,8 (1264,9 - 1790,1)	2543,9 (1818,1 - 2949,9)	≥ 1921,266	0.740	0.0402
Ubiquitin +1	RPS27A	3467,7 (2805,1 - 5077,7)	1947,5 (1607,1 - 2356,8)	≤ 2455,237	0.740	0.0402
PAF AHβ subunit (PABF)	PAFAH1B2	3692,6 (3344,8 - 4014,2)	4751,2 (4114,8 - 5588,7)	≥ 4101,777	0.734	0.0459
FABP	FABP1	3679,5 (2870,6 - 4466,3)	5280,5 (4731,5 - 6006,1)	≥ 4675,742	0.732	0.0467

*Mean of values after 100-resampling with replacement (medians, interquartile ranges)

**Accuracy of resampled thresholds in out-of bag samples (bagging accuracy); Family-wise corrected p values after 10.000 permutations

Ten proteins were significantly associated with disease progression at baseline

Figure 9



On the left panel: heatmap representation of the ten proteins that significantly predict at baseline the disease progression Yes: evolving Pre-SSc; No: stable Pre-SSc. On the right panel: heatmap representation of pathways that significantly predict at baseline the disease progression.

ELISA confirmation analysis in validation cohort of Pre-SSc

A validation cohort of fifty patients with Pre-SSc features at baseline and whose clinical prospective data were available to determine who had progressed and who did not was selected. Out of 50, 20 (40%) progressed to definitive SSc after a mean of 863 days (2.4 years). Baseline characteristics of validation cohort patients are summarized in **Table 6**. After progression, the evolving Pre-SSc (n=20) were characterized by skin fibrosis (100%), teleangectasia (30%) and upper GI tract symptoms (50%) (**Table 7**). Out of the 10 proteins found at SOMAscan, this analysis found in evolving Pre-SSc vs stable Pre-SSc, increased levels of Endostatin (mean 152 ng/ml vs 92.73 ng/ml p= 0.009).

Table 6: Baseline clinical characteristics of Pre-SSc validation cohort according to evolution outcomes

Features	Overall pre-SSc (50)	Stable SSc (30)	Evolving SSc (20)
Age, mean years (SD)	55.9 (14.04)	56.8 (13.18)	54.55 (15.15)
Gender, Female n (%)	44 (88)	26 (86.7)	18 (90)
Ethnicity Caucasian n (%)	50 (100)	30 (100)	20 (100)
ANA n (%)	47 (94)	28 (93)	19 (95)
ACA n (%)	32 (64)	17 (57)	15 (75)
Anti-Scl70 (%)	8 (16)	3 (10)	5 (25)
Other	20 (40)	11 (36.67)	9 (45)
RP months duration mean (SD)	128.44 (115.75)	154.3 (120.78)	90 (95.41)
FVC (%) mean (SD)	115.6 (16.04)	116.73 (17.8)	112.55 (12.81)
DLCO (%) mean (SD)	84.18 (17.2)	86.93 (16.99)	80.05 (16.65)
Evolution n (%)	0 (0)	0 (0)	0 (0)
Skin n (%)	0 (0)	0	0 (0)
Upper GI n (%)	19 (38)	9 (30)	10 (50)
Teleangectasia n (%)	0 (0)	0	0 (0)
Cardioaspirin n (%)	42 (84)	26 (87)	16 (80)
CCB n (%)	36 (72)	21 (7)	15 (75)

Abbreviations: ANA: anti-nuclear antibodies; ACA: anti-centromere antibodies; RP: Raynaud Phenomenon; FVC: forced vital capacity; DLCO: diffusing capacity of the lung for carbon monoxide; N: number; GI: gastro-intestinal; CCB: calcium channel blockers; NA: not applicable; SD: Standard deviation

Table 7: Subjects clinical features at follow-up

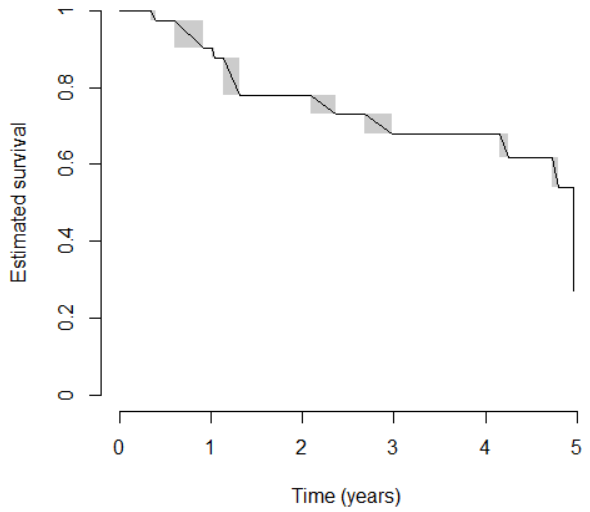
Features	Stable Pre-SSc 1 (30)	Evolving Pre-SSc (20)
Evolution n (%)	0 (0)	20 (100)
Skin n (%)	0 (0)	20 (100)
Only Puffy Fingers	0 (0)	13 (65)
lcSSc	0 (0)	7 (35)
dcSSc	0 (0)	0 (0)
Upper GI n (%)	9 (30)	10 (50)
Teleangectasia n (%)	0	6 (30)
Cardioaspirin n (%)	26 (87)	16 (80)
CCB n (%)	21 (7)	15 (75)

Abbreviations: lcSSc: limited cutaneous systemic sclerosis; dcSSc: diffuse cutaneous systemic sclerosis; N: number; GI: gastro-intestinal; CCB: calcium channel blockers; NA: not applicable; SD: Standard deviation

Time to evolution analysis

The estimated time-to-evolution according to the Turnbull method (survival analysis for interval-censored data) is shown below in **Figure 10**

Figure 10



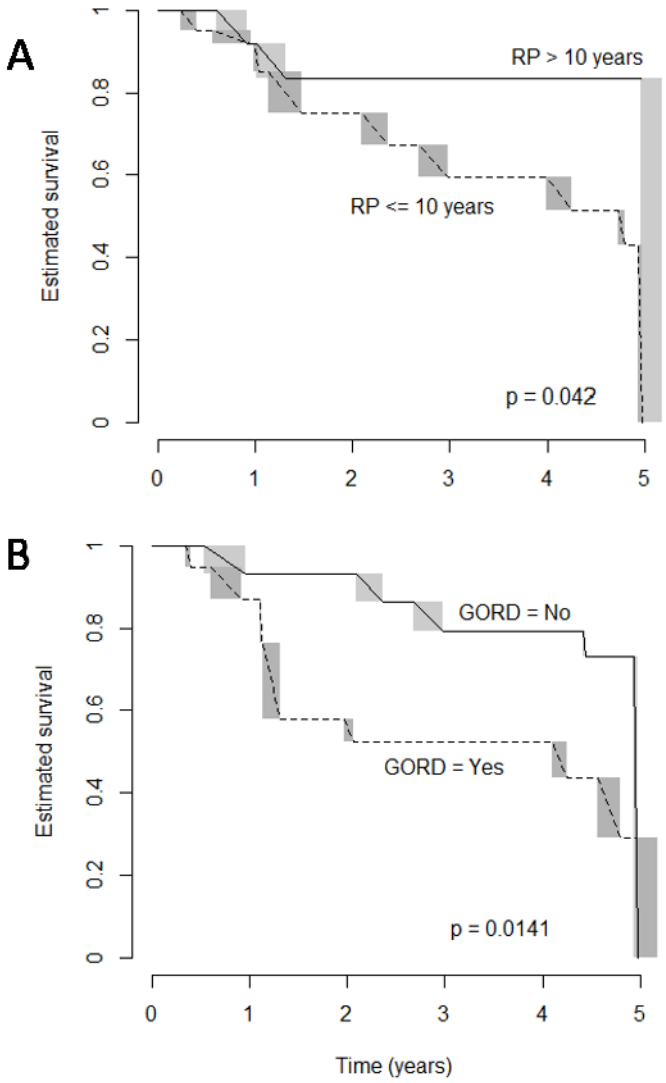
Patients with Raynaud's duration shorter than 10 years (p = 0.0425) at the time of evaluation or with reflux disease (p = 0.014) had shorter times to progression while none of the other baseline clinical characteristics was associated with time-to-evolution (see **Figure 11 A and 11 B**). Continuous variables were categorized after cutpoint estimation on right-censored samples. See **Table 8**.

Table 8: clinical characteristics association with time to evolution

Variable	P value
RP > 120 months	0.0425
GORD	0.0141
ACA	0.2368
Anti-Scl70	0.1332
Age ≥ 55 years	0.4531
Sex	0.9635
DLCO > 80% predicted	0.5826
Use of aspirin	0.734
Use of CCB	0.4769

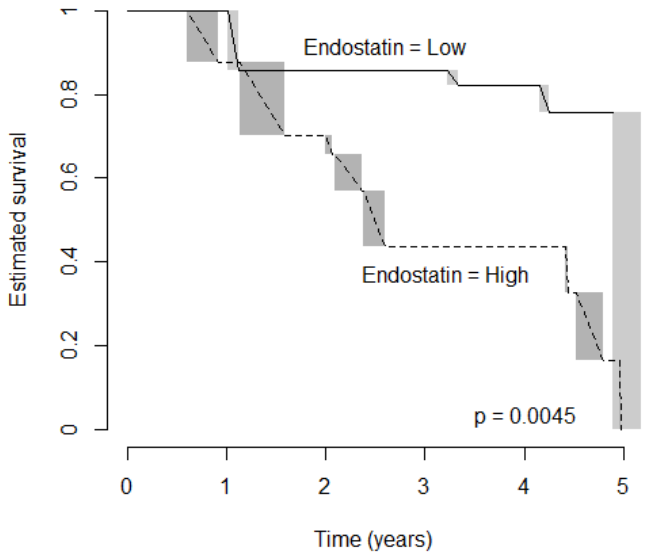
Abbreviations: RP: Raynaud Phenomenon; ACA: anti-centromere antibodies; DLCO: diffusing capacity of the lung for carbon monoxide; GORD: Gastro-oesophageal reflux disease; CCB: calcium channel blockers;

Figure 11


A
Estimated survival
0
0.2
0.4
0.6
0.8
1
RP > 10 years
RP <= 10 years
p = 0.042
0
1
2
3
4
5
B
Estimated survival
0
0.2
0.4
0.6
0.8
1
GORD = No
GORD = Yes
p = 0.0141
0
1
2
3
4
5
Time (years)

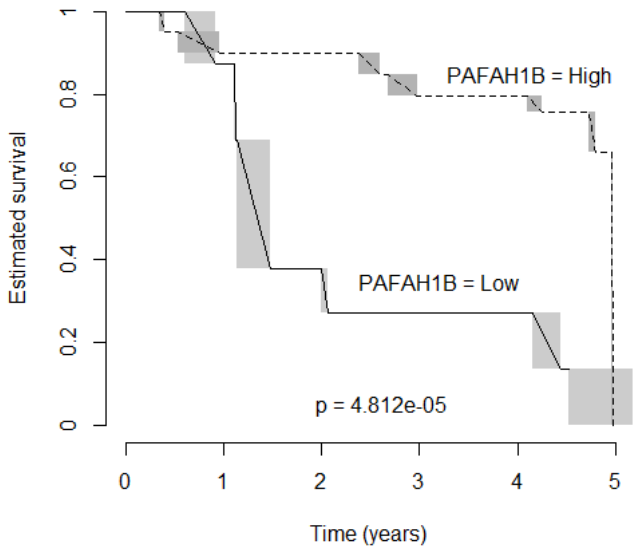
Patients with high endostatin levels (≥ 117 pg/mL) had shorter times to evolution (**Figure 12**)

Figure 12



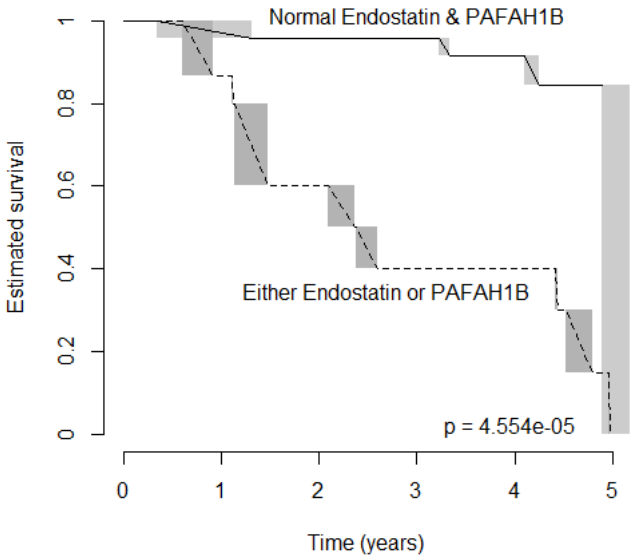
Patients with PAFAH1B2 (Platelet-activating factor acetylhydrolase IB subunit beta) < 1.68 pg/mL levels had shorter times to progression (**Figure 13**)

Figure 13



Patients with neither high Endostatin nor low PAFAH1B2 has much longer times to progression compared to patients with at least one of the two risk factors for progression (p = 4.554e-05) **Figure 14**.

Figure 14



Discussion

The present project explored SSc patients at a preclinical stage of disease performing a comprehensive investigation on a clinical, transcriptomic and proteomic level. From the modular analysis, transcripts signatures emerged at characterizing preclinical patients prone to disease progression and distinguishing Pre-SSc from healthy controls. Proteins able to predict disease progression were individuated as well.

At a clinical level, the present study confirmed previous observations of esteemed Pre-SSc clinical progression into a definite SSc form in about 50% of patients in 5 years. Indeed, in 4 years of follow up we observed a progression of 42.9 % of Pre-SSc. As previously assessed from the Canadian group that followed-up for >10 years 586 consecutive patients presenting RP, SSc is a progressive disease in which capillaroscopy abnormalities and/or specific autoantibodies associated to RP, independently predict progression. When both capillaroscopy abnormalities and specific autoantibodies were present, authors found out that 65.9 % subjects progress into a definite SSc in 5 years, 72.7% in 10 years and 79.5% at follow-up (about 20 years) [23]. In our cohort of Pre-SSc, clinical features at baseline such as FVC or DLCO were not significantly differential in evolving Pre-SSc versus stable Pre-SSc comparison, while the presence of gastro-esophageal reflux and a shorter RP time (considered as disease duration) were significantly associated with a shorter time of progression as assessed from the survival analysis in patients of our validation cohort. Surprisingly, Pre-SSc with Scl-70 positivity progressed with a small discrepancy between the two groups (4 in evolving Pre-SSc vs 3 in stable-Pre SSc). This same observation of a lack of significant difference in progression between Pre-SSc ACA positive and Anti Scl-70 positive is also described in [26]; the small sample size could be an explanation for both the cases.

At a transcript level of investigation, in both stable Pre-SSc and evolving Pre-SSc a type I IFN signature was identified at baseline in comparison to controls. Indeed, interferon module M 3.4, M 5.12, M 1.2 were significantly increased in the baseline modular analysis, confirming prior observations of a type I IFN signature in systemic sclerosis patients as well as in pre-clinical SSc [42,43,75]. However, a comparison between stable Pre-SSc and evolving Pre-SSc did not show, as instead intuitively expected, a significant difference of the IFN modules within these two groups both in the baseline and in the pre-post comparison analyses. In line with this premise, at baseline, comparisons between stable Pre-SSc and evolving Pre-SSc did

not show any significant differential module or signature (both in modular and QuSAGE analysis), testifying that potential biomarkers of progression could will not be identifiable at the very beginning of pre-clinical stages. It has to be noticed that in general all the signatures assessed showed a weak signal probably due to the very minimal clinical discrepancies between two groups of patients that, even when classified as progressors, presented the very first few features of definite SSc.

In comparison to healthy controls at baseline, overall Pre-SSc showed a decrease of several modules (Erythrocytes M 3.1 and M 2.3, Cytotoxic/NK module M3.6 and M8.46, TGF-beta/DNA repair M 9.20, T cells M 4.1 and M 4.15, regulation of cellular proliferation M 7.6, Cell cycle M 3.5 and M 2.2). Regarding the Cytotoxic/NK modules, a decreased NK signature was kept also by the evolving Pre-SSc as shown by the pre-post analysis while on the opposite, stable Pre-SSc showed a slightly increased NK signature. In line with our results that associate a decreased NK signature with SSc appearance, a reduced cytotoxic/NK module was observed at baseline in the whole blood of the SCOT trial dcSSc participants and subsequently an increase of this same signature was assessed after autologous hematopoietic stem-cell transplantation (HSCT) (that indeed had a normalizing effect to the dcSSc molecular signatures) [80]. Moreover, immunophenotyping performed in two SSc cohorts show reduced NK circulating levels in SSc compared to healthy controls [45]. Interestingly as shown in [86], CD56+ cells NK isolated from patients at different stages of SSc and stimulated with TLR1/2, present an intermediate secretion pattern of pro-inflammatory cytokines in Pre-SSc confirming a role of innate immunity in the pre-fibrotic stages of SSc.

In the pre-post modular analysis of evolving Pre-SSc, together with the decrease of cytotoxic/NK module, an increase of neutrophils/granulocyte module M 4.9 and M 6.20 was observed. Also, this finding support what already seen in [80] where an increased neutrophil module is observed in dcSSc vs controls before HSCT. On the other side, in stable Pre-SSc pre-post modular analysis, the neutrophils/granulocyte module M 4.9 was slightly increased as well as the cytotoxic/NK module (M 3.6) as described before.

The RNA expression change (pre-post) analysis of evolving Pre-SSc and stable Pre-SSc identified 73 genes with a corrected p value <=0.05. Among genes that favored evolving Pre-SSc, GAB1 encodes for a protein (Growth factor receptor bound protein 2-associated protein 1) that regulate cellular growth response, transformation and apoptosis and increased expression of GAB1 promotes inflammation and fibrosis in

systemic sclerosis [87]. Of interest, it is also SH2B3 (protein name Lymphocyte-specific adapter protein Lnk) that is associated as a determinant of cardiac inflammation and fibrosis as well as PPM1H (Protein phosphatase 1H) associated with interferon-α in systemic lupus erythematosus [88–90]. Among genes that favored stable pre-SSc, TNFRSF18 (Tumor necrosis factor receptor superfamily member 18) is thought to play a key role in immunological self-tolerance maintained by regulatory T cells and it has a role of cellular immunity in systemic sclerosis pathogenesis [91]; CXCR6 (C-X-C chemokine receptor type 6) belongs to the chemokines receptors family and it is expressed in T cells regulating T lymphocytes migration. An increased serum level of CXCR6 is found in SSc serum and in dermal endothelial cells likely promoting angiogenesis in SSc skin [92].

At a proteomic level, out of the 286 proteins assessed with SOMAscan, 10 proteins were able to predict at baseline evolving Pre-SSc from stable Pre-SSc. Progressors were characterized by increased levels of NKp30 (natural cytotoxicity receptor 3), Endostatin, bFGF (basic fibroblast growth factor), ECM1 (Extracellular matrix protein 1), FGF18 (Fibroblast growth factor 18), Fibronectin 1.3, PAFAH1B2 (Platelet-activating factor acetylhydrolase IB subunit alpha2) and FABP (fatty acid-binding protein) and by decreased levels of PHI (Glucose-6-Phosphate Isomerase) and Ubiquitin1. Pathways that emerged from GO terms deriving from these ten proteins were among others, extracellular matrix organization, positive regulation of cell proliferation, angiogenesis, signal transduction. bFGF is a fibrogenic protein previously found to be increased in skin of SSc [93]. Fibronectin is a glycoprotein of the ECM and a profibrotic role of both these two components in dermal SSc is well established [94–96]. Of relevance, high levels of endostatin were confirmed in the subsequently ELISA measurement in the validation cohort of baseline Pre-SSc. Moreover, from time to evolution analysis, high levels of endostatin and reduced levels PAFAH1B2 were related with a shorter time to progression. While knowledge on PAFAH1B2 in scleroderma do not emerge from a literature investigation, endostatin is a protein well known to be associated with SSc vascular manifestation (such PAH, digital ulcers and renal crisis) [97–99]. It is elevated in serum of SSc patients [100] and in baseline dcSSc investigated in [73], endostatin was upregulated both at a transcript and at a protein level, correlating positively with mRSS. Endostatin is also associated with skin involvement and fibrosis in SSc [101,102].

The present study presented some weaknesses such as the small number of subjects enrolled; this is due to the difficulty in intercepting patients at a preclinical stage in a rare autoimmune disease. Subjects of this

study had all a long disease duration, intended as a long time of RP since its first appearance although disease duration was homogeneous among the groups. A limited number of subjects were characterized by Scl-70 autoantibody positivity, although also this feature was well balanced between the groups.

The project has several strengths. To the best of knowledge, this is the first study to perform a RNAseq analysis as well as a proteomic analysis in Pre-SSc subjects. It is a study focused on Pre-SSc, aiming at investigating the very first biomolecular features in the shift between a pre-clinical to clinical phase of a disease. Due to the rarity of SSc and more than that of pre-SSc, this is the first study with a larger sample size of Pre-SSc with homogeneous clinical features.

As future perspectives, the present study could enlarge the interest in the early identification of subject at high risk to develop systemic sclerosis. Intercepting the passage from a preclinal to clinical stage of SSc, will allow to early identify patients with features of dcSSc thus at high risk of early ILD development, a major cause of mortality in SSc. The ability to start proper treatments in the earliest time possible, represent the possibility to stop the disease progression and definitive damage, truly changing the quality of life of scleroderma patients. Moreover, the investigation on the biomolecular features of preclinical SSc subjects, could identify biomarkers candidate for future drugs development. Lastly, future studies on larger cohorts of preclinical SSc subjects could provide answers to the potential protective role of early calcium channel blockers use in this disease.

In conclusion, the present study showed a type I IFN signature that distinguished subjects with pre-clinical SSc than healthy controls. A reduced NK signature was related to SSc progression as well as, from a clinical perspective, a shorter time of Raynaud and the presence of gastro-esophageal reflux. Proteins able to predict disease evolution were individuated, with pathways of fibrosis, extracellular matrix organization, positive regulation of cell proliferation, angiogenesis, signal transduction; among them endostatin is an interesting protein already assessed to be related to vascular and fibrotic features in well-established definite Systemic sclerosis, thus representing a biomarker worthy of future mechanistic investigations.

References

1 Barnes J, Mayes MD. Epidemiology of systemic sclerosis: incidence, prevalence, survival, risk factors, malignancy, and environmental triggers. *Curr Opin Rheumatol* 2012;**24**. doi:10.1097/BOR.0b013e32834ff2e8

2 Elhai M, Meune C, Avouac J, *et al.* Trends in mortality in patients with systemic sclerosis over 40 years: a systematic review and meta-analysis of cohort studies. *Rheumatology (Oxford)* 2012;**51**. doi:10.1093/rheumatology/ker269

3 Asano Y. The Pathogenesis of Systemic Sclerosis: An Understanding Based on a Common Pathologic Cascade across Multiple Organs and Additional Organ-Specific Pathologies. *J Clin Med* 2020;**9**. doi:10.3390/jcm9092687

4 Brown M, O'Reilly S. The immunopathogenesis of fibrosis in systemic sclerosis. *Clin Exp Immunol* 2019;**195**. doi:10.1111/cei.13238

5 LeRoy EC, Black C, Fleischmajer R, *et al.* Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol* 1988;**15**.

6 Pearson DR, Werth VP, Pappas-Taffer L. Systemic sclerosis: Current concepts of skin and systemic manifestations. *Clin Dermatol*;**36**. doi:10.1016/j.clindermatol.2018.04.004

7 Elhai M, Meune C, Avouac J, *et al.* Trends in Mortality in Patients with Systemic Sclerosis Over 40 Years: A Systematic Review and Meta-Analysis of Cohort Studies. *Rheumatology (Oxford)* 2012;**51**:1017–26. doi:10.1093/rheumatology/ker269

8 Volkmann ER, Fischer A. Update on Morbidity and Mortality in Systemic Sclerosis-Related Interstitial Lung Disease. *J scleroderma Relat Disord* 2021;**6**. doi:10.1177/2397198320915042

9 Morrisroe K, Stevens W, Sahhar J, *et al.* Epidemiology and disease characteristics of systemic sclerosis-related pulmonary arterial hypertension: results from a real-life screening programme. *Arthritis Res Ther* 2017;**19**. doi:10.1186/s13075-017-1250-z

10 Ebert EC. Gastric and enteric involvement in progressive systemic sclerosis. *J Clin Gastroenterol* 2008;**42**. doi:10.1097/MCG.0b013e318042d625

11 Denaxas K, Ladas SD, Karamanolis GP. Evaluation and management of esophageal manifestations in systemic sclerosis. *Ann Gastroenterol*;**31**. doi:10.20524/aog.2018.0228

12 Andréasson K, Alrawi Z, Persson A, *et al.* Intestinal dysbiosis is common in systemic sclerosis and associated with gastrointestinal and extraintestinal features of disease. *Arthritis Res Ther* 2016;**18**:278. doi:10.1186/s13075-016-1182-z

13 Sakkas LI, Simopoulou T, Daoussis D, *et al.* Intestinal Involvement in Systemic Sclerosis: A Clinical Review. *Dig Dis Sci* 2018;**63**. doi:10.1007/s10620-018-4977-8

14 Woodworth TG, Suliman YA, Li W, *et al.* Scleroderma renal crisis and renal involvement in systemic sclerosis. *Nat Rev Nephrol* 2018;**14**. doi:10.1038/nrneph.2017.183

15 Vacca A, Meune C, Gordon J, *et al.* Cardiac arrhythmias and conduction defects in systemic sclerosis. *Rheumatology (Oxford)* 2014;**53**. doi:10.1093/rheumatology/ket377

16 Komócsi A, Vorobcsuk A, Faludi R, *et al.* The impact of cardiopulmonary manifestations on the mortality of SSc: a systematic review and meta-analysis of observational studies. *Rheumatology (Oxford)* 2012;**51**. doi:10.1093/rheumatology/ker357

17 Rodríguez-Reyna TS, Morelos-Guzman M, Hernández-Reyes P, *et al.* Assessment of myocardial fibrosis and microvascular damage in systemic sclerosis by magnetic resonance imaging and coronary angiotomography. *Rheumatology (Oxford)* 2015;**54**. doi:10.1093/rheumatology/keu350

18 Morrisroe KB, Nikpour M, Proudman SM. Musculoskeletal Manifestations of Systemic Sclerosis. *Rheum Dis Clin North Am* 2015;**41**. doi:10.1016/j.rdc.2015.04.011

19 Sandler RD, Matucci-Cerinic M, Hughes M. Musculoskeletal hand involvement in systemic sclerosis. *Semin Arthritis Rheum* 2020;**50**. doi:10.1016/j.semarthrit.2019.11.003

20 Tartar DM, Chung L, Fiorentino DF. Clinical significance of autoantibodies in dermatomyositis and systemic sclerosis. *Clin Dermatol*;**36**. doi:10.1016/j.clindermatol.2018.04.008

21 van den Hoogen F, Khanna D, Fransen J, *et al.* 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. *Arthritis Rheum* 2013;**65**. doi:10.1002/art.38098

22 LeRoy EC, Medsger TA. Criteria for the classification of early systemic sclerosis. *J Rheumatol* 2001;**28**.

23 Koenig M, Joyal F, Fritzler MJ, *et al.* Autoantibodies and microvascular damage are independent predictive factors for the progression of Raynaud's phenomenon to systemic sclerosis: a twenty-year prospective study of 586 patients, with validation of proposed criteria for early systemic sclerosis. *Arthritis Rheum* 2008;**58**. doi:10.1002/art.24038

24 Minier T, Guiducci S, Bellando-Randone S, *et al.* Preliminary analysis of the very early diagnosis of systemic sclerosis (VEDOSS) EUSTAR multicentre study: evidence for puffy fingers as a pivotal sign for suspicion of systemic sclerosis. *Ann Rheum Dis* 2014;**73**. doi:10.1136/annrheumdis-2013-203716

25 Valentini G, Vettori S, Cuomo G, *et al.* Early systemic sclerosis: short-term disease evolution and

factors predicting the development of new manifestations of organ involvement. *Arthritis Res Ther* 2012;**14**. doi:10.1186/ar4019

26 Valentini G, Marcoccia A, Cuomo G, *et al.* Early systemic sclerosis: Analysis of the disease course in patients with marker autoantibody and/or capillaroscopic positivity. *Arthritis Care Res* 2014;**66**:1520–7. doi:10.1002/acr.22304

27 Varga J, Trojanowska M, Kuwana M. Pathogenesis of systemic sclerosis: recent insights of molecular and cellular mechanisms and therapeutic opportunities. *J Scleroderma Relat Disord* 2017;**2**. doi:10.5301/jsrd.5000249

28 Feghali-Bostwick C, Medsger TA, Wright TM. Analysis of systemic sclerosis in twins reveals low concordance for disease and high concordance for the presence of antinuclear antibodies. *Arthritis Rheum* 2003;**48**. doi:10.1002/art.11173

29 Angiolilli C, Marut W, van der Kroef M, *et al.* New insights into the genetics and epigenetics of systemic sclerosis. *Nat Rev Rheumatol* 2018;**14**. doi:10.1038/s41584-018-0099-0

30 Acosta-Herrera M, Kerick M, Lopéz-Isac E, *et al.* Comprehensive analysis of the major histocompatibility complex in systemic sclerosis identifies differential HLA associations by clinical and serological subtypes. *Ann Rheum Dis* Published Online First: 1 April 2021. doi:10.1136/annrheumdis-2021-219884

31 Chairta P, Nicolaou P, Christodoulou K. Genomic and genetic studies of systemic sclerosis: A systematic review. *Hum Immunol* 2017;**78**. doi:10.1016/j.humimm.2016.10.017

32 Xu Y, Wang W, Tian Y, *et al.* Polymorphisms in STAT4 and IRF5 increase the risk of systemic sclerosis: a meta-analysis. *Int J Dermatol* 2016;**55**. doi:10.1111/ijd.12839

33 López-Isac E, Martín J-E, Assassi S, *et al.* Brief Report: IRF4 Newly Identified as a Common Susceptibility Locus for Systemic Sclerosis and Rheumatoid Arthritis in a Cross-Disease Meta-Analysis of Genome-Wide Association Studies. *Arthritis Rheumatol (Hoboken, NJ)* 2016;**68**. doi:10.1002/art.39730

34 Carmona FD, Gutala R, Simeón CP, *et al.* Novel identification of the IRF7 region as an anticentromere autoantibody propensity locus in systemic sclerosis. *Ann Rheum Dis* 2012;**71**. doi:10.1136/annrheumdis-2011-200275

35 Radstake TRDJ, Gorlova O, Rueda B, *et al.* Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet* 2010;**42**. doi:10.1038/ng.565

36 Bossini-Castillo L, Martin J-E, Broen J, *et al.* A GWAS follow-up study reveals the association of the IL12RB2 gene with systemic sclerosis in Caucasian populations. *Hum Mol Genet* 2012;**21**. doi:10.1093/hmg/ddr522

37 Dieudé P, Wipff J, Guedj M, *et al.* BANK1 is a genetic risk factor for diffuse cutaneous systemic sclerosis and has additive effects with IRF5 and STAT4. *Arthritis Rheum* 2009;**60**. doi:10.1002/art.24885

38 López-Isac E, Acosta-Herrera M, Kerick M, *et al.* GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat Commun* 2019;**10**. doi:10.1038/s41467-019-12760-y

39 Manetti M, Allanore Y, Saad M, *et al.* Evidence for caveolin-1 as a new susceptibility gene regulating tissue fibrosis in systemic sclerosis. *Ann Rheum Dis* 2012;**71**. doi:10.1136/annrheumdis-2011-200986

40 Gao L, Emond MJ, Louie T, *et al.* Identification of Rare Variants in ATP8B4 as a Risk Factor for Systemic Sclerosis by Whole-Exome Sequencing. *Arthritis Rheumatol (Hoboken, NJ)* 2016;**68**. doi:10.1002/art.39449

41 Orvain C, Assassi S, Avouac J, *et al.* Systemic sclerosis pathogenesis: contribution of recent advances in genetics. *Curr Opin Rheumatol* 2020;**32**. doi:10.1097/BOR.0000000000000735

42 Tan FK, Zhou X, Mayes MD, *et al.* Signatures of differentially regulated interferon gene expression and vasculotrophism in the peripheral blood cells of systemic sclerosis patients. *Rheumatology (Oxford)* 2006;**45**:694–702. doi:10.1093/rheumatology/kei244

43 York MR, Nagai T, Mangini AJ, *et al.* A macrophage marker, Siglec-1, is increased on circulating monocytes in patients with systemic sclerosis and induced by type I interferons and toll-like receptor agonists. *Arthritis Rheum* 2007;**56**:1010–20. doi:10.1002/art.22382

44 Assassi S, Mayes MD, Arnett FC, *et al.* Systemic sclerosis and lupus: points in an interferon-mediated continuum. *Arthritis Rheum* 2010;**62**. doi:10.1002/art.27224

45 Beretta L, Barturen G, Vigone B, *et al.* Genome-wide whole blood transcriptome profiling in a large European cohort of systemic sclerosis patients. *Ann Rheum Dis* 2020;**79**. doi:10.1136/annrheumdis-2020-217116

46 Whitfield ML, Finlay DR, Murray JI, *et al.* Systemic and cell type-specific gene expression patterns in scleroderma skin. *Proc Natl Acad Sci U S A* 2003;**100**. doi:10.1073/pnas.1635114100

47 Sargent JL, Milano A, Bhattacharyya S, *et al.* A TGFbeta-responsive gene signature is associated with a subset of diffuse scleroderma with increased disease severity. *J Invest Dermatol* 2010;**130**. doi:10.1038/jid.2009.318

48 Derrett-Smith EC, Martyanov V, Chighizola CB, *et al.* Limited cutaneous systemic sclerosis skin demonstrates distinct molecular subsets separated by a cardiovascular development gene expression signature. *Arthritis Res Ther* 2017;**19**. doi:10.1186/s13075-017-1360-7

49 Lofgren S, Hinchcliff M, Carns M, *et al.* Integrated, multicohort analysis of systemic sclerosis identifies robust transcriptional signature of disease severity. *JCI insight* 2016;**1**:e89073. doi:10.1172/jci.insight.89073

50 Taroni JN, Martyanov V, Mahoney JM, *et al.* A Functional Genomic Meta-Analysis of Clinical Trials in Systemic Sclerosis: Toward Precision Medicine and Combination Therapy. *J Invest Dermatol* 2017;**137**. doi:10.1016/j.jid.2016.12.007

51 Taroni JN, Martyanov V, Huang C-C, *et al.* Molecular characterization of systemic sclerosis esophageal pathology identifies inflammatory and proliferative signatures. *Arthritis Res Ther* 2015;**17**. doi:10.1186/s13075-015-0695-1

52 Taroni JN, Greene CS, Martyanov V, *et al.* A novel multi-network approach reveals tissue-specific cellular modulators of fibrosis in systemic sclerosis. *Genome Med* 2017;**9**. doi:10.1186/s13073-017-0417-1

53 Aslani S, Sobhani S, Gharibdoost F, *et al.* Epigenetics and pathogenesis of systemic sclerosis; the ins and outs. *Hum Immunol* 2018;**79**. doi:10.1016/j.humimm.2018.01.003

54 Altorok N, Tsou P-S, Coit P, *et al.* Genome-wide DNA methylation analysis in dermal fibroblasts from patients with diffuse and limited systemic sclerosis reveals common and subset-specific DNA methylation aberrancies. *Ann Rheum Dis* 2015;**74**. doi:10.1136/annrheumdis-2014-205303

55 Hattori M, Yokoyama Y, Hattori T, *et al.* Global DNA hypomethylation and hypoxia-induced expression of the ten eleven translocation (TET) family, TET1, in scleroderma fibroblasts. *Exp Dermatol* 2015;**24**. doi:10.1111/exd.12767

56 Matatiele P, Tikly M, Tarr G, *et al.* DNA methylation similarities in genes of black South Africans with systemic lupus erythematosus and systemic sclerosis. *J Biomed Sci* 2015;**22**. doi:10.1186/s12929-015-0142-2

57 Ding W, Pu W, Wang L, *et al.* Genome-Wide DNA Methylation Analysis in Systemic Sclerosis Reveals Hypomethylation of IFN-Associated Genes in CD4+ and CD8+ T Cells. *J Invest Dermatol* 2018;**138**. doi:10.1016/j.jid.2017.12.003

58 Noda S, Asano Y, Nishimura S, *et al.* Simultaneous downregulation of KLF5 and Fli1 is a key feature underlying systemic sclerosis. *Nat Commun* 2014;**5**. doi:10.1038/ncomms6797

59 Wang Q, Xiao Y, Shi Y, *et al.* Overexpression of JMJD3 may contribute to demethylation of H3K27me3 in CD4+ T cells from patients with systemic sclerosis. *Clin Immunol* 2015;**161**. doi:10.1016/j.clim.2015.03.006

60 Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;**116**. doi:10.1016/s0092-8674(04)00045-5

61 Rossato M, Affandi AJ, Thordardottir S, *et al.* Association of MicroRNA-618 Expression With Altered Frequency and Activation of Plasmacytoid Dendritic Cells in Patients With Systemic Sclerosis. *Arthritis Rheumatol (Hoboken, NJ)* 2017;**69**. doi:10.1002/art.40163

62 Jafarinejad-Farsangi S, Farazmand A, Mahmoudi M, *et al.* MicroRNA-29a induces apoptosis via increasing the Bax:Bcl-2 ratio in dermal fibroblasts of patients with systemic sclerosis. *Autoimmunity* 2015;**48**. doi:10.3109/08916934.2015.1030616

63 Steen SO, Iversen L V, Carlsen AL, *et al.* The circulating cell-free microRNA profile in systemic sclerosis is distinct from both healthy controls and systemic lupus erythematosus. *J Rheumatol* 2015;**42**. doi:10.3899/jrheum.140502

64 Honda N, Jinnin M, Kira-Etoh T, *et al.* miR-150 down-regulation contributes to the constitutive type I collagen overexpression in scleroderma dermal fibroblasts via the induction of integrin β3. *Am J Pathol* 2013;**182**. doi:10.1016/j.ajpath.2012.09.023

65 Messemaker TC, Chadli L, Cai G, *et al.* Antisense Long Non-Coding RNAs Are Deregulated in Skin Tissue of Patients with Systemic Sclerosis. *J Invest Dermatol* 2018;**138**:826–35. doi:10.1016/j.jid.2017.09.053

66 Fioretto BS, Rosa I, Romano E, *et al.* The contribution of epigenetics to the pathogenesis and gender dimorphism of systemic sclerosis: a comprehensive overview. *Ther Adv Musculoskelet Dis* 2020;**12**. doi:10.1177/1759720X20918456

67 Lötvall J, Hill AF, Hochberg F, *et al.* Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. *J Extracell vesicles* 2014;**3**. doi:10.3402/jev.v3.26913

68 Wermuth PJ, Piera-Velazquez S, Rosenbloom J, *et al.* Existing and novel biomarkers for precision medicine in systemic sclerosis. *Nat Rev Rheumatol* 2018;**14**. doi:10.1038/s41584-018-0021-9

69 Michalska-Jakubus M, Kowal-Bielecka O, Smith V, *et al.* Plasma endothelial microparticles reflect the extent of capillaroscopic alterations and correlate with the severity of skin involvement in systemic sclerosis. *Microvasc Res* 2017;**110**. doi:10.1016/j.mvr.2016.11.006

70 Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 2002;**1**. doi:10.1074/mcp.r200007-mcp200

71 Albaba D, Soomro S, Mohan C. Aptamer-Based Screens of Human Body Fluids for Biomarkers. *Microarrays (Basel, Switzerland)* 2015;**4**. doi:10.3390/microarrays4030424

72 Cossu M, Andracco R, Santaniello A, *et al.* Serum levels of vascular dysfunction markers reflect disease severity and stage in systemic sclerosis patients. *Rheumatology* 2016;**55**:1112–6. doi:10.1093/rheumatology/kew017

73 Bellocchi C, Ying J, Goldmuntz EA, *et al.* Large-Scale Characterization of Systemic Sclerosis Serum Protein Profile: Comparison to Peripheral Blood Cell Transcriptome and Correlations With Skin/Lung Fibrosis. *Arthritis Rheumatol (Hoboken, NJ)* 2021;**73**. doi:10.1002/art.41570

74 Rice LM, Mantero JC, Stifano G, *et al.* A Proteome-Derived Longitudinal Pharmacodynamic Biomarker for Diffuse Systemic Sclerosis Skin. *J Invest Dermatol* 2017;**137**. doi:10.1016/j.jid.2016.08.027

75 Brkic Z, van Bon L, Cossu M, *et al.* The interferon type I signature is present in systemic sclerosis before overt fibrosis and might contribute to its pathogenesis through high BAFF gene expression and high collagen synthesis. *Ann Rheum Dis* Published Online First: 14 September 2015. doi:10.1136/annrheumdis-2015-207392

76 Vettori S, Cuomo G, Iudici M, *et al.* Early Systemic Sclerosis: Serum Profiling of Factors Involved in Endothelial, T-cell, and Fibroblast Interplay is Marked by Elevated Interleukin-33 Levels. *J Clin Immunol* Published Online First: 2014. doi:10.1007/s10875-014-0037-0

77 Cossu M, van Bon L, Preti C, *et al.* Earliest Phase of Systemic Sclerosis Typified by Increased Levels of Inflammatory Proteins in the Serum. *Arthritis Rheumatol (Hoboken, NJ)* 2017;**69**. doi:10.1002/art.40243

78 Liu X, Mayes MD, Tan FK, *et al.* Correlation of interferon-inducible chemokine plasma levels with disease severity in systemic sclerosis. *Arthritis Rheum* 2013;**65**. doi:10.1002/art.37742

79 Chaussabel D, Baldwin N. Democratizing systems immunology with modular transcriptional repertoire analyses. *Nat Rev Immunol* 2014;**14**. doi:10.1038/nri3642

80 Assassi S, Wang X, Chen G, *et al.* Myeloablation followed by autologous stem cell transplantation normalises systemic sclerosis molecular signatures. *Ann Rheum Dis* 2019;**78**. doi:10.1136/annrheumdis-2019-215770

81 Banchereau R, Hong S, Cantarel B, *et al.* Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients. *Cell* 2016;**165**. doi:10.1016/j.cell.2016.03.008

82 Yaari G, Bolen CR, Thakar J, *et al.* Quantitative set analysis for gene expression: a method to quantify gene set differential expression including gene-gene correlations. *Nucleic Acids Res* 2013;**41**:e170–e170. doi:10.1093/nar/gkt660

83 O'Connell NS, Dai L, Jiang Y, *et al.* Methods for Analysis of Pre-Post Data in Clinical Research: A Comparison of Five Common Methods. *J Biom Biostat* 2017;**8**. doi:10.4172/2155-6180.1000334

84 Rempala GA, Yang Y. On Permutation Procedures for Strong Control in Multiple Testing with Gene Expression Data. *Stat Interface* 2013;**6**. doi:10.4310/SII.2013.v6.n1.a8

85 Stanley S, Vanarsa K, Soliman S, *et al.* Author Correction: Comprehensive aptamer-based screening identifies a spectrum of urinary biomarkers of lupus nephritis across ethnicities. *Nat Commun* 2020;**11**. doi:10.1038/s41467-020-16944-9

86 Cossu M, van Bon L, Nierkens S, *et al.* The magnitude of cytokine production by stimulated CD56+ cells is associated with early stages of systemic sclerosis. *Clin Immunol* 2016;**173**. doi:10.1016/j.clim.2016.09.004

87 Shi X, Liu Q, Zhao H, *et al.* Increased expression of GAB1 promotes inflammation and fibrosis in systemic sclerosis. *Exp Dermatol* 2019;**28**. doi:10.1111/exd.14033

88 Flister MJ, Hoffman MJ, Lemke A, *et al.* SH2B3 Is a Genetic Determinant of Cardiac Inflammation and Fibrosis. *Circ Cardiovasc Genet* 2015;**8**. doi:10.1161/CIRCGENETICS.114.000527

89 Ghodke-Puranik Y, Imgruet M, Dorschner JM, *et al.* Novel genetic associations with interferon in systemic lupus erythematosus identified by replication and fine-mapping of trait-stratified genome-wide screen. *Cytokine* 2020;**132**. doi:10.1016/j.cyto.2018.12.014

90 Kariuki SN, Franek BS, Kumar AA, *et al.* Trait-stratified genome-wide association study identifies novel and diverse genetic associations with serologic and cytokine phenotypes in systemic lupus erythematosus. *Arthritis Res Ther* 2010;**12**. doi:10.1186/ar3101

91 Beşliu AN, Bănică LM, Lonescu R, *et al.* Role of cellular immunity in systemic sclerosis pathogenesis: update on CD4+T cells population studies. *Roum Arch Microbiol Immunol*;**68**.

92 Rabquer BJ, Tsou P-S, Hou Y, *et al.* Dysregulated expression of MIG/CXCL9, IP-10/CXCL10 and CXCL16 and their receptors in systemic sclerosis. *Arthritis Res Ther* 2011;**13**. doi:10.1186/ar3242

93 Lawrence A, Khanna D, Misra R, *et al.* Increased expression of basic fibroblast growth factor in skin of patients with systemic sclerosis. *Dermatol Online J* 2006;**12**.

94 Carvalheiro T, Malvar Fernández B, Ottria A, *et al.* Extracellular SPARC cooperates with TGF-β signalling to induce pro-fibrotic activation of systemic sclerosis patient dermal fibroblasts. *Rheumatology (Oxford)* 2020;**59**. doi:10.1093/rheumatology/kez583

95 Karmouty-Quintana H, Molina JG, Philip K, *et al.* The Antifibrotic Effect of A2B Adenosine Receptor Antagonism in a Mouse Model of Dermal Fibrosis. *Arthritis Rheumatol (Hoboken, NJ)* 2018;**70**. doi:10.1002/art.40554

96 Cutolo M, Ruaro B, Montagna P, *et al.* Effects of selexipag and its active metabolite in contrasting the profibrotic myofibroblast activity in cultured scleroderma skin fibroblasts. *Arthritis Res Ther* 2018;**20**. doi:10.1186/s13075-018-1577-0

97 Hebbar M, Peyrat JP, Hornez L, *et al.* Increased concentrations of the circulating angiogenesis

inhibitor endostatin in patients with systemic sclerosis. *Arthritis Rheum* 2000;**43**:889–93. doi:10.1002/1529-0131(200004)43:4<889::AID-ANR21>3.0.CO;2-5

98 Gigante A, Margiotta D, Navarini L, *et al.* Serum level of endostatin and digital ulcers in systemic sclerosis patients. *Int Wound J* 2018;**15**. doi:10.1111/iwj.12882

99 Mecoli CA, Shah AA, Boin F, *et al.* The Utility of Plasma Vascular Biomarkers in Systemic Sclerosis: A Prospective Longitudinal Analysis. *Arthritis Rheumatol (Hoboken, NJ)* 2020;**72**. doi:10.1002/art.41265

100 Delle Sedie A, Riente L, Maggiorini L, *et al.* Potential biomarkers in patients with systemic sclerosis. *Int J Rheum Dis* 2018;**21**:261–5. doi:10.1111/1756-185X.13196

101 Farouk HM, Hamza SH, El Bakry SA, *et al.* Dysregulation of angiogenic homeostasis in systemic sclerosis. *Int J Rheum Dis* 2013;**16**:448–54. doi:10.1111/1756-185X.12130

102 Yamaguchi Y, Takihara T, Chambers RA, *et al.* A peptide derived from endostatin ameliorates organ fibrosis. *Sci Transl Med* 2012;**4**. doi:10.1126/scitranslmed.3003421