



OPEN

Enrichr in silico analysis of MS-based extracted candidate proteomic biomarkers highlights pathogenic pathways in systemic sclerosis

Paraskevi P. Chairta, Paschalis Nicolaou & Kyproula Christodoulou✉

Systemic sclerosis (SSc) is a rheumatic disease characterised by vasculopathy, inflammation and fibrosis. Its aetiopathogenesis is still unknown, and the pathways/mechanisms of the disease are not clarified. This study aimed to perform in silico analysis of the already Mass Spectrometry (MS)-based discovered biomarkers of SSc to extract possible pathways/mechanisms implicated in the disease. We recorded all published candidate MS-based found biomarkers related to SSc. We then selected a number of the candidate biomarkers using specific criteria and performed pathway and cellular component analyses using Enrichr. We used PANTHER and STRING to assess the biological processes and the interactions of the recorded proteins, respectively. Pathway analysis extracted several pathways that are associated with the three different stages of SSc pathogenesis. Some of these pathways are also related to other diseases, including autoimmune diseases. We observe that these biomarkers are located in several cellular components and implicated in many biological processes. STRING analysis showed that some proteins interact, creating significant clusters, while others do not display any evidence of an interaction. All these data highlight the complexity of SSc, and further investigation of the extracted pathways/biological processes and interactions may help study the disease from a different angle.

Abbreviations

AID	Autoimmune disease
dcSSc	Diffused cutaneous SSc
ECM	Extracellular matrix
FC	Fold change
HLA	Human leukocyte antigen
ILD	Interstitial lung disease
lcSSc	Limited cutaneous SSc
LC	Liquid chromatography
MHC	Major histocompatibility complex
MS	Mass spectrometry
MALDI-TOF	Matrix-assisted laser desorption/ionisation time-time-of-flight
MRM	Multiple reaction monitoring
PPARs	Peroxisome proliferator-activated receptors
PANTHER	Protein analysis through evolutionary relationships
PAH	Pulmonary arterial hypertension
ROS	Reactive oxygen species
RA	Rheumatoid arthritis
SNPs	Single nucleotide polymorphisms
SELDI-TOF	Surface-enhanced laser desorption/ionisation time-of-flight
SLE	Systemic lupus erythematosus

Neurogenetics Department, The Cyprus Institute of Neurology and Genetics, 2371 Nicosia, Cyprus. ✉email: roula@cing.ac.cy

SSc	Systemic sclerosis
MS/MS	Tandem MS
2DE DIGE	Two-dimensional difference gel electrophoresis
2DE	Two-dimensional gel electrophoresis
UPS	Ubiquitin–proteasome system

Systemic sclerosis (SSc) is a chronic, heterogeneous, multisystemic, connective tissue autoimmune disease (AID). Vasculopathy, immune system abnormalities–inflammation and excessive deposition of collagen–fibrosis are characteristic features of SSc¹. Despite intensive research, the aetiopathogenesis of the disease remains poorly understood. An interaction between genetic, epigenetic alterations and environmental factors is suggested^{2,3}. Remarkably, SSc, like other AIDs, occurs more frequently in females than in males and its incidence and prevalence vary according to the geographical region and study⁴.

SSc is divided into two subgroups; limited cutaneous SSc (lcSSc) and diffused cutaneous SSc (dcSSc), based on clinical manifestations. In lcSSc, skin thickening is usually limited to the face, ankles, wrists and distal to knees and elbows. DcSSc involves skin thickening proximal to the knees and elbows and damage of internal organs (e.g. lungs and renal)⁵. Interstitial lung disease (ILD) and pulmonary arterial hypertension (PAH) are the most common types of pulmonary involvements in SSc. They are likely life-threatening as they account for 60% of death in the disease⁶.

Besides genetics and environmental factors, proteome and pathways dysregulation may play a crucial role in the pathogenesis of SSc. Proteins are considered the structural and functional molecules of the living organism and are responsible for the cell's phenotype. Proteomics is the only way to study protein–protein interaction, the mechanism/pathways of disease, and the environmental effect at the level of the organism⁷. In addition, it is essential for biomarkers discovery; disease-specific proteins that can be used and distinguish a pathological from a non-pathological condition. These biomarkers have several applications. We can use them for better classification, early prognosis, more accurate diagnosis and therapeutic targeting of the disease. In addition, many of the candidate biomarkers might also be interconnected and help extract critical physiopathological pathways⁸.

Early-stage proteomic studies employed the two-dimensional gel electrophoresis (2DE) application. In recent years, newly developed proteomics methodologies based on MS enabled more robust studies⁹. MS, a high-throughput analytical method, allows the detection, identification and quantification of proteins in many samples such as biopsies, saliva and plasma¹⁰. In addition, various studies use different MS proteomic techniques such as matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) and surface-enhanced laser desorption/ionisation time-of-flight (SELDI-TOF)⁸.

In this article, we gather published candidate MS-based discovered biomarkers of SSc and then further analyse a number of these molecules selected based on specific criteria through in silico online tools for the first time to extract possible pathways involved in the development of the disease.

Methods

MS-based discovered biomarkers recording. All candidate MS-based proteomic biomarkers for SSc discovered until 2021 were collected/extracted through two extensive systematic reviews^{8,11}. We recorded all biomarkers and grouped them based on general SSc, SSc subtypes; lcSSc and dcSSc and lung involvements (PF and ILD) related to SSc.

Quality assessment of the MS-based studies. The quality of data reported through recorded studies was evaluated according to the eight criteria recommended by Mischak et al¹² for biomarker identification and qualification in clinical proteomics. These criteria are the following: justification of the clinical question, description of the subjects, samples, methodology and statistical assessment, validation of the results, acknowledgement of limitations and authors contribution statement.

In silico analysis. Selection of proteins for in silico analysis. The selection of proteins for in silico analysis was performed using two different approaches.

The first selection of proteins for in silico analysis was performed based on three combined criteria; (1) proteins that were extracted through full text research articles, and (2) proteins that were extracted through the articles that fulfilled the first five quality assessment criteria [(1) identification and description of clinical question, (2) assessment of subjects description, (3) sampling description, (4) presentation of experimental methodology and (5) statistical approach] and (3) proteins with p -value ≤ 0.05 & Fold Change (FC) ≥ 1.5 or ≤ 0.67 or expressed only in disease or control samples.

The second selection of proteins for in silico analysis was performed including proteins that fulfilled the above criteria and were also validated with additional methods.

We performed in silico analysis in general SSc, dcSSc, lcSSc and lung involvements groups.

Enrichr. We used the Enrichr Web Server (<https://maayanlab.cloud/Enrichr/>)¹³ to classify the proteins/genes based on the following ontologies: (a) KEGG 2021 Human^{14–16} and (b) GO Cellular Component 2021. We performed separate pathway analyses for the upregulated and downregulated proteins.

PANTHER. We analysed all selected proteins using the Protein Analysis Through Evolutionary Relationships (PANTHER)—Gene List Analysis tool (<http://www.pantherdb.org/>), Version 17.0 (released 2020–02–22)¹⁷. This tool classified/sub-classified proteins (and their genes) into PANTHER GO-Slim Biological Processes.

Venn diagram. We produced the Venn diagrams using the InteractiVenn tool (<http://www.interactivenn.net/>) to easily visualise common and unique pathways through different analyses and subtypes¹⁸.

STRING 11.5. We used the Search Tool for Retrieval of Interacting Genes/Proteins—STRING Version: 11.5 (<http://string-db.org/>)¹⁹ to assess the interactions of selected genes/proteins. We set the confidence level to 0.700 (high). We pulled interactions based on database, experiments, and co-expression.

Results

Gathering of data from previous MS-based proteomics studies. Up to date, 29 MS-based proteomics studies were performed on SSc^{10,20–47}. Out of these studies, 25 were published as full text research articles. Two systematic reviews by Balanescu P et al.^{8,11} record these studies, where a total of 655 candidate proteomic biomarkers were identified (Additional file 1). Twenty-eight out of these biomarkers were further validated using additional methods.

In most studies, proteins were separated with 2DE or Two-dimensional difference gel electrophoresis (2DE DIGE) or Liquid Chromatography (LC). Proteins were identified with MALDI-TOF, SELDI-TOF or Tandem MS (MS/MS) techniques. Multiple Reaction Monitoring (MRM) and older-standard methods (IF, WB, IHC and ELISA) were used to validate these proteins.

Quality assessment of the MS-based studies. All full text research articles included identification and description of clinical question (25/25) as well as sampling description (25/25). In the most of the studies, authors described the assessment of the subjects (24/25) and statistical evaluation (23/25) as well as presented the experimental methodology (24/25). However, validation of the results (14/25), acknowledge limitations (7/25) and authors contribution (10/25) are lacked from several manuscripts. All these details are shown in additional file 2.

In silico analysis of the 1st selection of proteins. We performed a pathway analysis of the significantly upregulated ($p \leq 0.05$ & $FC \geq 1.5$) and downregulated ($p \leq 0.05$ & $FC \leq 0.67$) selected proteomic biomarkers. We investigated the different subtypes separately; general SSc (Additional file 3A), dcSSc (Additional file 3B), lcSSc (Additional file 3C) and lung involvements (Additional file 3D) groups. We extracted some common significant pathways from the upregulated and downregulated proteins of different groups (Fig. 1). In addition, we found a significant association of some of the extracted pathways with the three different stages of SSc pathogenesis. Among these pathways are the complement and coagulation cascades, TGF-beta signalling pathway, Antigen processing and

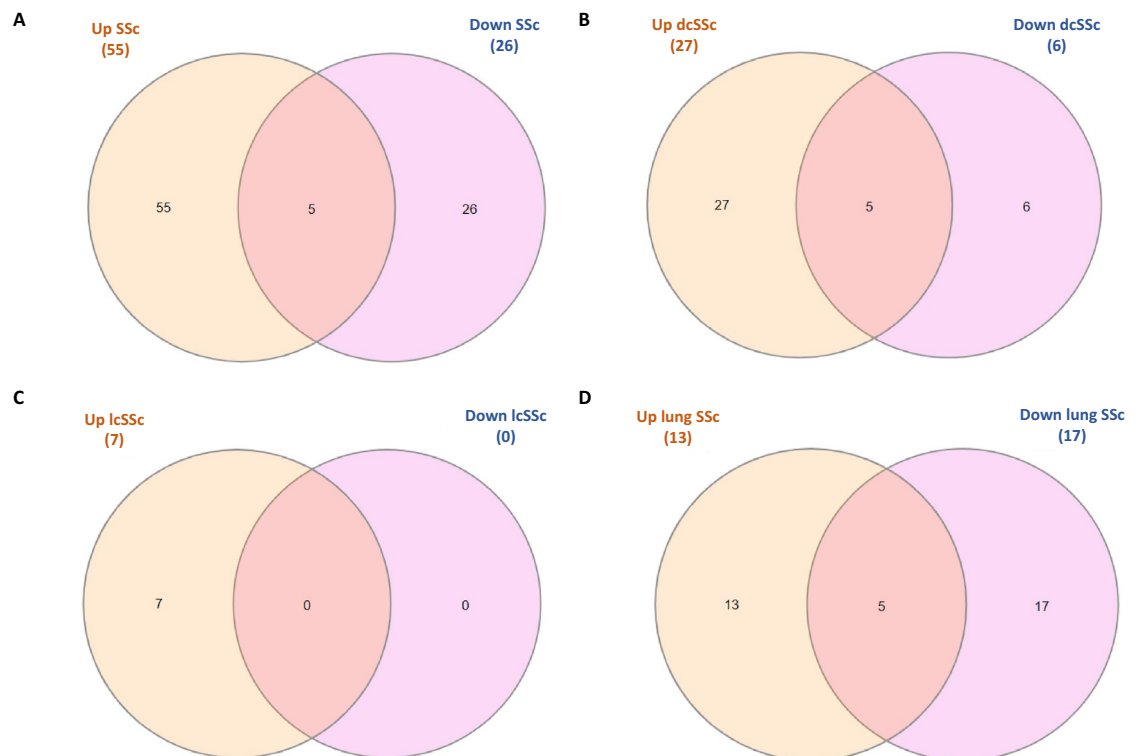


Figure 1. Venn diagrams indicate the number of common and unique significant pathways. These Venn diagrams indicate the number of common and unique significant pathways ($p < 0.05$) extracted through EnrichR KEGG 2021 from the significantly upregulated ($p \leq 0.05$ & $FC \geq 1.5$) and downregulated ($p \leq 0.05$ & $FC \leq 0.67$) proteins in (A) SSc patients, (B) dcSSc patients, (C) lcSSc patients and (D) lung disease related to SSc compared to controls. The overlapping areas represent the common pathways.

presentation, PPAR signalling pathway, Focal adhesion, IL-17 signalling pathway, ECM-receptor interaction, Chemokine signalling pathway, Fc gamma R-mediated phagocytosis, Platelet activation, Phagosome, Oxidative phosphorylation and Leukocyte transendothelial migration.

We used Enrichr for the GO Cellular Component 2021 analysis (Fig. 2A–D, Additional file 4). In all subgroups except lcSSc, proteins are involved in more than 100 cellular components, and more than 50 out of these components have a significant p -value ($p \leq 0.05$). Collagen-containing extracellular matrix (GO:0062023) and focal adhesion (GO:0005925) are among these significant components.

We performed PANTHER analysis to classify the proteins/genes into different biological processes (Fig. 3, Additional file 5). Dysregulated proteins of general SSc, dcSSc, lcSSc and lung involvement subgroups are implicated in 12, 12, 9 and 12 biological processes with additional sub-processes, respectively. Interestingly, 9 main biological processes are shared by all subgroups.

We used the STRING 11.5 database to assess whether the selected SSc candidate proteomic biomarkers have an interaction between them (Fig. 4A–D). Some clusters, as well as direct and indirect interactions, were observed in different subgroups. This analysis showed that some genes/proteins have a high confidence score of protein–protein interaction networks (e.g. PF4-CXCL8). In contrast, some others do not present any pieces of evidence for interaction (e.g. ICAM1, ABCC12) (Fig. 4A). In addition, some of the proteins interact together as they are in the same family (e.g. KRT1, KRT5 and KRT14) (Fig. 4B).

In silico analysis of the 2nd selection of proteins. All *in silico* analyses were performed on a smaller group of 21 candidate validated biomarkers for SSc. Enrichr KEGG 2021 Human, Enrichr GO Cellular Component 2021 and PANTHER biological processes analyses results are shown in Additional file 6 (A–C). STRING analysis results of these proteins are shown in Additional file 7.

Discussion

SSc is a rheumatic AID with complex aetiopathogenesis and heterogeneous clinical manifestations. Despite that SSc is characterised by vasculopathy, immune system abnormalities and fibrosis, the molecules implicated in these processes are not fully elucidated. Several “omics” studies have been performed on SSc; however, the pathways and the mechanisms involved in the disease are still unclear^{48,49}. Therefore, this study aimed to use a number of selected candidate proteomic biomarkers recorded to date for SSc to extract the central pathways implicated in the pathogenesis of SSc.

We used the Enrichr Web Server to perform pathway analysis. Several extracted pathways are associated with immunity, as immune system abnormalities and inflammation are prominent features of SSc pathogenesis. Previous studies on SSc have already reported some of the pathways and molecules we pulled through this analysis.

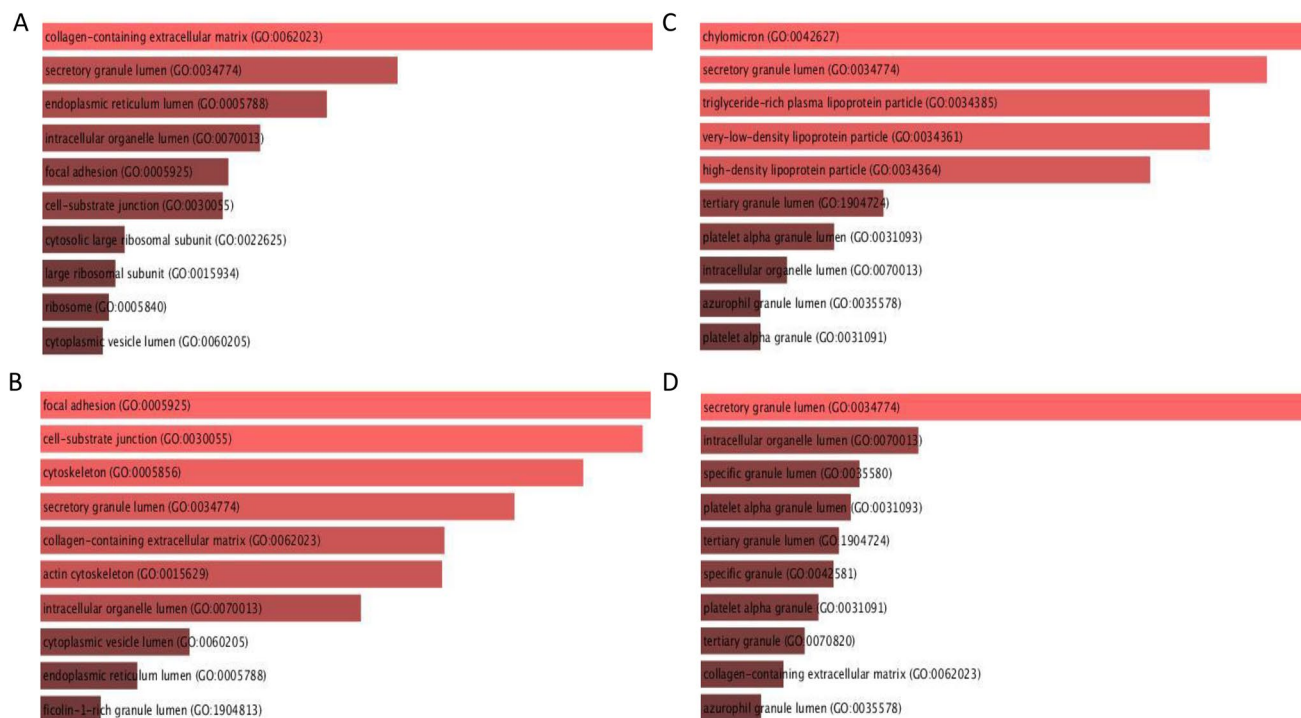


Figure 2. Top 10 significant Enrichr GO Cellular Component 2021. Top 10 significant Enrichr GO Cellular Component 2021 of significantly differentially expressed proteins ($p \leq 0.05$) with $FC \geq 1.5$ or ≤ 0.67 in (A) SSc patients, (B) dcSSc patients, (C) lcSSc patients and (D) lung disease related to SSc compared to controls. The brightness and the length of the bar represent the significance of the category/ term.

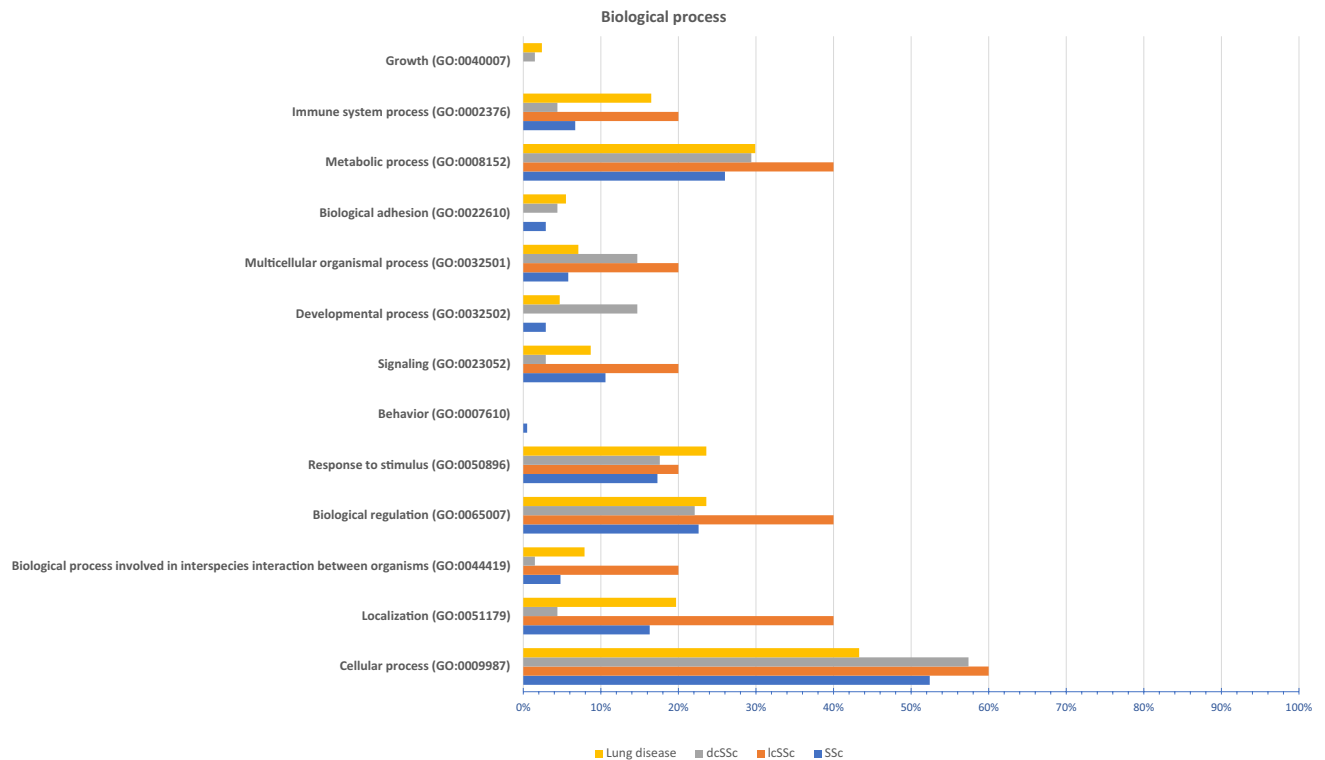


Figure 3. PANTHER biological process analysis. The bar chart shows the biological processes associated with the significantly dysregulated genes/proteins ($p \leq 0.05$ & $FC \geq 1.5$ or ≤ 0.67) in different subtypes; SSc, dcSSc, lcSSc and lung disease associated with SSc.

The complement system is the central mechanism of antibody-mediated immunity and part of the innate immune system. The primary function of this system is to recognize and eliminate foreign pathogens either through the stimulation of phagocytosis or direct killing⁵⁰. In our study, we identified 17 dysregulated (upregulated/downregulated) molecules implicated in this pathway. Interestingly, these molecules are involved in both the coagulation and the complement (alternative, classical and lectin pathways) pathway cascades. Previous studies on SSc have also reported dysregulated proteins implicated in this pathway. Senaldi et al.⁵¹ report that C3d and ratios of C3d:C3 and Ba:factor B were higher in patients with dcSSc than controls, while C4d:C4, C4d, C3d:C3 and C3d levels were higher in patients with lcSSc compared to controls. These results show that activation of the complement system occurs in patients with SSc and might be associated with the clinical severity of the disease⁵¹. In a more recent study, Landi et al.²³ observed decreased complement C3a levels in patients with PF compared to healthy smoking controls. In addition, pathway analysis showed that the dysregulated proteins in that study are involved in pathways associated with disease pathogenesis (alternative complement, classical complement, lectin induced complement and blood coagulation pathway)²³. Ludwicka-Bradley et al.⁵² also report that imbalance of the coagulation cascade, especially release of the thrombin molecule, is associated with the earliest SSc events following tissue injury and is mainly observed in SSc patients with lung disease.

Antigen processing and presentation is an additional pathway extracted through our analysis (7 molecules) associated with SSc and all AIDs. This pathway is responsible for processing and presenting the foreign antigen to the appropriate immune cells through the Major Histocompatibility Complex (MHC) mechanism⁵³. Acosta-Herrera et al.⁵⁴ have reported the association of this pathway with SSc. They performed sequential conditional analyses of the human leukocyte antigen (HLA) region. They conclude on seven amino acids, nine classical alleles and nine Single Nucleotide Polymorphisms (SNPs) associated with the disease and its subtypes.

Chemokine signalling and IL-17 signalling are two additional pathways extracted from our analysis and associated with immune system regulation. The chemokine signalling pathway was extracted with a significant p -value only from the general SSc downregulated proteins (4 molecules). Chemokines, also known as chemotactic cytokines, regulate the recruitment and migration of immune cells. These molecules were found to be associated with several rheumatic diseases, including SSc, systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA)⁵⁵. For instance, Van Bon et al.²⁸ reported that the plasma level of CXCL8 is higher in SSc patients than in healthy controls, while PPBP is lower. However, Scambi et al.²⁶ reported that the levels of PPBP is higher in serum of dcSSc and lcSSc patients compared to healthy controls.

IL-17 family is a group of six inflammatory cytokines that are associated with cancer progression and AIDs. Despite that IL-17 signalling pathway is a cytokines pathway that includes several activators and inhibitors, previous studies reported that it might be implicated in fibrosis of SSc pathogenesis⁵⁶. However, its exact role on the disease has not been elucidated. Some studies reported that IL-17A expression is higher in serum of SSc patients, and others reported it lower in SSc patients than in healthy controls⁵⁶. In addition, Wakhlu et al.⁵⁷ showed that high levels of IL-6, IL-17A and TGF β 1 molecules in serum of SSc patients are associated with skin

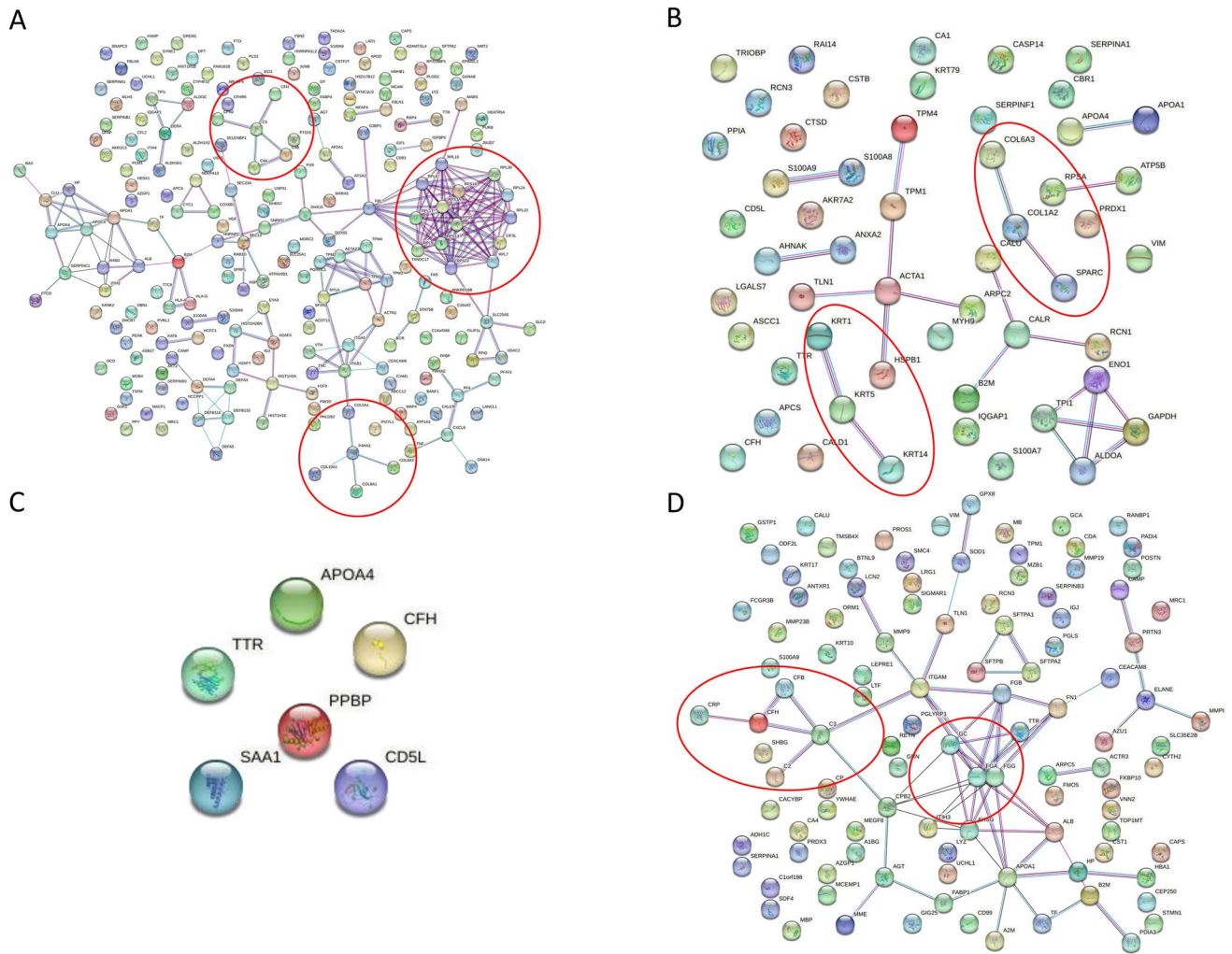


Figure 4. STRING (v. 11.5) analysis of the dysregulated genes/proteins. This figure reveals functional interactions between the dysregulated genes/proteins ($p \leq 0.05$ & $FC \geq 1.5$ or ≤ 0.67) in different subtypes; (A) SSs, (B) dcSSc, (C) lcSSc and (D) lung disease associated with SSs. Blue, pink and black coloured lines indicate databases, co-expression, and experimental evidence, respectively. Red cycles indicate significant clusters.

fibrosis and ILD. This pathway was significantly extracted from three out of four subgroups in our analysis, but 7 different molecules than the above are implicated. Although many studies reported dysregulated chemokine and interleukin molecules associated with SSs, chemokine and IL-17 signalling pathways are extracted through pathway analysis using SSs dysregulated proteins for the first time in our study.

Leukocyte transendothelial migration is also an essential pathway. CD4 + T lymphocytes transendothelial migration is enhanced in SSs, and the migrating cells present an activated phase⁵⁸. In a previous study, Xu et al.⁵⁹ performed KEGG pathway analysis of ITGA5, ITGB2, and ITGB5 that were found to be upregulated in the skin of SSs patients compared to control and reported that these proteins are implicated in the leukocyte transendothelial migration pathway. In our study, this pathway was significantly extracted from general SSs (upregulated proteins) and lung disease related to SSs (downregulated proteins) subgroups (ITGB1;ACTN1;ICAM1;ITGAM;MMP9). Our results show that ITGAM and MMP9 proteins are mainly dysregulated for leukocyte transendothelial migration, indicating that they might promote lung disease related to SSs by affecting this procedure.

Interestingly, phagosome, as well as Fc gamma R-mediated phagocytosis pathways, were also extracted in our analysis. Phagocytosis is a fundamental process for innate immune response where the specialised phagocytes such as macrophages, monocytes, dendritic cells, and neutrophils recognise, engulf and eliminate the pathogenic microorganisms⁶⁰. This process has a crucial role in maintaining tissue homeostasis by removing apoptotic cells⁶¹. Receptors of phagocytes are categorized into opsonic and non-opsonic. The non-opsonic receptor directly recognises substances on the phagocytic target surface, while the opsonic receptor identifies host-derived proteins bound to the foreign particle. Complement and Fc are the most essential and well-characterised opsonic phagocytic receptors⁶¹. Previous studies showed that defects in the clearance of apoptotic cells might lead to several diseases, including AIDS⁶². Davis et al.⁶³ showed that anti-FcγR autoantibodies are detected in the serum of SSs patients but not in controls. In addition, Kadono et al.⁶⁴ in a more recent study, report that the levels of anti-FcγRIIB/C autoantibodies in the serum of SSs patients are significantly higher than in healthy controls.

These findings support the idea that phagocytosis is implicated in the pathogenesis of SSc, but further studies are needed to elucidate its exact role.

Peroxisomes are essential organelles that regulate lipid homeostasis. They also interact with some pathogens to escape the immune response, turn off/on pivotal cellular immune signalling and control the interaction between host and pathogen⁶⁵. Peroxisome proliferator-activated receptors (PPARs); PPAR- α , - β/δ , and γ are nuclear receptors that regulate several processes, including inflammation, differentiation and cell proliferation⁶⁶. Previous studies prove that PPAR- γ is implicated in the pathogenesis of SSc, especially aberrant expression of PPAR- γ contributes to fibrosis and vascular remodelling stages⁶⁷. Shiwen et al.⁶⁸ showed reduced expression of PPAR- γ in fibroblasts of patients with dcSSc compared to healthy controls. In addition, Marangoni et al.⁶⁹ reported that PPAR- γ is also related with SSc at the genetic level, as PPAR- γ intronic SNP (rs10865710) is associated with SSc susceptibility. Interestingly, in our study, PPARs signalling pathway was extracted, but the implicated proteins are included in all three signals and not only in the PPAR- γ part.

TGF- β is a pathway interconnected with PPAR signalling and significantly extracted from our analysis (3 molecules in general SSc analysis). Wei et al.⁷⁰ suggested that PPAR- γ expression in SSc is suppressed by excessive activity of TGF- β and may contribute to the abnormal activation of fibroblast and fibrogenesis. This signalling pathway regulates several cellular processes, including extracellular matrix (ECM) synthesis, cell differentiation, apoptosis and cell growth⁷¹. Increased levels of TGF- β 1 in tissue and serum of patients with SSc suggest that it might be implicated in fibrosis. In addition, studies showed that TGF- β signalling could be a mediator of experimental PAH and blocking of this signalling was shown to reduce the severity of PAH⁷¹. Dantas et al. assessed the levels of the active TGF- β 1 isoform in skin biopsies and serum from patients with SSc and healthy controls. They reported that the serum level of TGF- β 1 was significantly higher in patients compared to controls. At the same time, no significant difference was observed in the mRNA expression of TGF β 1 in skin biopsies of patients with SSc compared to controls. They also noted that patients with SSc who have high levels of TGF- β 1 in serum are more likely to have dcSSc, DU, SSc (Anti-Topoisomerase) ATA +, higher Rodnan skin score and lung fibrosis. These results suggest that this molecule might be a marker for vascular and fibrotic manifestations of SSc⁷².

The ECM-receptor interaction is also included in the essential signalling pathways associated with SSc, especially in fibrosis. Several previous studies identified dysregulated proteins implicated in this pathway and suggest that it plays a crucial role in developing the disease^{59,73}. Xu et al.⁵⁹ showed that protein and mRNA levels of three integrins (ITGB5, ITGB2 and ITGA5) are upregulated in SSc patients than in controls. Pathway analysis showed that these proteins/genes are implicated in leukocyte transendothelial migration, focal adhesion, ECM-receptor interaction and ECM turnover. This pathway was extracted in our study, not only from integrin proteins but also from the collagen family proteins. Focal adhesion and cell adhesion molecules pathways were also extracted from our analysis. Adhesion molecules are proteins that mediate cell-to-ECM as well as cell-to-cell interaction. Based on their function, these molecules are located on the surface of different cells and implicated in several processes, including T-Cell Mediated Immunity and leukocyte migration and fibrogenesis^{74,75}. Although these two pathways were not discussed in depth in previous studies, Hasegawa et al.⁷⁶ report that serum adhesion molecules can be used as predictive biomarkers for early SSc. In addition, in our previous study, we report that the ICAM1 protein is significantly over-expressed in affected than unaffected skin biopsies of patients with SSc¹⁰.

Platelet activation and oxidative phosphorylation are also crucial pathways that contribute to all stages of SSc pathogenesis: vasculopathy, autoimmunity and fibrosis and they were extracted from our analysis⁷⁷⁻⁷⁹. Agache et al.⁸⁰ reported that platelet activation molecules such as P-selectin are associated with disease severity and activity. It was also reported that platelets of SSc patients have a high response to collagen, adenosine diphosphate, adrenaline and 5-hydroxytryptamine, which are correlated with different processes⁷⁷. Even though oxidative phosphorylation is essential for energy production in the organism, an abnormal function of this mechanism might lead to a pathological condition through the overproduction of Reactive Oxygen Species (ROS). For example, high levels of ROS were observed in the skin fibroblast of patients with SSc compared to healthy controls⁷⁹. In addition, oxidative stress, which is the imbalance and accumulation of ROS, may cause endothelial cells activation and damage/dysfunction⁷⁸. All of these pathways discussed above were extracted with a significant *p*-value either in one or more of our 1st selection analyses and might be implicated in the pathogenesis of SSc. Although in the 2nd selection the number of validated proteins was relatively small, complement and coagulation cascades and IL-17 signalling pathway were also extracted through these analyses with a significant *p*-value. Interestingly, Wnt signalling pathway, that is associated with SSc, was extracted through the 2nd selection analyses only with a significant *p*-value (Additional file 6A). Wnt signalling is an essential pathway that is implicated in several processes such as organogenesis, cell migration and cell fate during the embryonic development⁸¹. Wei et al.⁸² used genome-wide expression datasets and reported a decreased expression of Wif1 and Dkk2 (WNT antagonists) and an increased expression of Lef1 (Wnt target) and Frizzled2 (Wnt receptor) in dcSSc compared to healthy controls skin biopsies. In addition, an immunohistochemical study showed an elevated expression of nuclear β -catenin (transcriptional coactivator) in these skin biopsies. These evidences suggest a hyperactivation of Wnt signalling in patients with SSc and support that this pathway plays a key role in the pathogenesis of fibrosis in SSc⁸².

Enrichr analyses were also performed to classify the proteins based on their cellular components. These analyses showed that the recorded SSc associated proteins are located in many cellular components, some of them in more than one component (Additional file 4). The variety of cellular components confirms the complexity and heterogeneity of the disease. In addition, nine cellular components were extracted with significant *p*-value and were common in all analyses; general SSc, dcSSc, lcSSc and lung disease related-SSc. In all analyses except from lcSSc, the collagen-containing extracellular matrix (GO:0062023), related to SSc fibrosis, was extracted in the top-10 cellular components. Interestingly, collagen-containing extracellular matrix (GO:0062023) was also extracted with a significant *p*-value in the general SSc group of the 2nd selection analysis (Additional file 6B).

Biological process analysis is carried out to show the specific objective that an organism is programmed to achieve. Our 1st analysis of altered proteins indicated that these proteins are involved in 13 different biological

processes and many sub-processes (Additional file 5). The percentage of genes classified in each biological process is varied based on the SSc subgroup. Cellular homeostasis that was extracted from our analysis was also extracted from DAVID in a previous study by Corallo et al.³⁶. However, our analysis is more enriched as different databases give a different outcome. Biological process analysis of the 2nd selection of validated proteins showed that they are involved in 11 different biological processes and some sub-processes (Additional file 6C).

STRING analysis was performed to assess whether the recorded proteins are interconnected and create clusters. Direct and indirect interactions via databases, experiments and co-expression with a high confidence score (0.700) were extracted and are presented in Fig. 4A–D. Different interactions and clusters are observed in different subgroups. For example, in the analysis of general SSc (Fig. 4A), the cluster with the highest score contains several ribosomal proteins. Clusters that contain the complement system and collagen proteins were also observed. In the analysis of dcSSc (Fig. 4B), the collagen and keratin protein clusters associated with fibrosis were extracted. In addition, analysis of lung disease associated with SSc extracted the complement system and fibrinogen proteins cluster, related with the pathogenesis of SSc. Although some proteins show interactions between them with a high score, many do not display any pieces of evidence of an interaction. However, STRING analysis of the 2nd selection of proteins does not show any essential interaction between them (Additional file 7).

Conclusions

This study shows that many proteins with different functions are implicated in the pathogenesis of SSc. Pathway analysis confirms the heterogeneity of the disease as the recorded dysregulated proteins are involved in approximately 240 different pathways. Different pathways and molecules are involved in various stages of SSc pathogenesis and different SSc subtypes. Some of the extracted pathways are also related to other diseases, including AIDs, indicating that they share common mechanisms. Some of the significantly extracted pathways were associated with SSc in previous studies, but some of them were not studied in depth. Cellular component and biological process analyses confirmed the complexity of the disease as the recorded molecules are implicated in 13 different biological processes and located in more than 230 cellular components. All these analyses that were performed for the first time for all the MS-based SSc proteomic biomarkers in our study are essential for producing new data associated with the disease pathogenesis, for the identification/determination of the biological mechanisms and molecular interactions that lead to the disease, and for developing novel aspects for further studies.

Data availability

The datasets generated and/or analysed during the current study are available through Balanescu P et al.^{8,11}. This data can be found in Additional file 1.

Received: 12 January 2022; Accepted: 30 January 2023

Published online: 02 February 2023

References

- Sierra-Sepúlveda, A. et al. Systemic sclerosis pathogenesis and emerging therapies, beyond the fibroblast. *Biomed. Res. Int.* **2019**, 4569826 (2019).
- Chairta, P. et al. Genetic susceptibility to systemic sclerosis in the Greek-Cypriot population: A pilot study. *Genet. Test Mol. Biomarkers* **24**, 309–317 (2020).
- Tsou, P.-S., Varga, J. & O'Reilly, S. Advances in epigenetics in systemic sclerosis: Molecular mechanisms and therapeutic potential. *Nat. Rev. Rheumatol.* **17**, 596–607 (2021).
- Bairkdar, M. et al. Incidence and prevalence of systemic sclerosis globally: A comprehensive systematic review and meta-analysis. *Rheumatology* **60**, 3121–3133 (2021).
- Barnes, J. & Mayes, M. D. Epidemiology of systemic sclerosis: Incidence, prevalence, survival, risk factors, malignancy, and environmental triggers. *Curr. Opin. Rheumatol.* **24**, 165–170 (2012).
- Fischer, A., Patel, N. M. & Volkman, E. R. Interstitial lung disease in systemic sclerosis: Focus on early detection and intervention. *Open Access Rheumatol. Res. Rev.* **11**, 283–307 (2019).
- Graves, P. R. & Haystead, T. A. J. Molecular biologist's guide to proteomics. *Microbiol. Mol. Biol. Rev.* **66**, 39–63 (2002).
- Bălănescu, P., Lădaru, A., Bălănescu, E., Băicuș, C. & Dan, G. A. Systemic sclerosis biomarkers discovered using mass-spectrometry-based proteomics: A systematic review. *Biomarkers Biochem. Indic. Expo. Response Susceptibility Chem.* **19**, 345–355 (2014).
- You, S.-A. & Wang, Q. K. Proteomics with two-dimensional gel electrophoresis and mass spectrometry analysis in cardiovascular research. *Methods Mol. Med.* **129**, 15–26 (2006).
- Chairta, P. et al. Comparative analysis of affected and unaffected areas of systemic sclerosis skin biopsies by high-throughput proteomic approaches. *Arthritis Res. Ther.* **22**, 107 (2020).
- Bălănescu, P., Bălănescu, A., Bălănescu, E. & Băicuș, C. Candidate proteomic biomarkers in systemic sclerosis discovered using mass-spectrometry: An update of a systematic review (2014–2020). *Rom. J. Intern. Med.* **59**, 101–111 (2021).
- Mischak, H. et al. Recommendations for biomarker identification and qualification in clinical proteomics. *Sci. Transl. Med.* **2**, 46ps42 (2010).
- Kuleshov, M. V. et al. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
- Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
- Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).
- Mi, H. et al. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49**, D394–403 (2021).
- Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P. & Minghim, R. InteractiVenn: A web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinform.* **16**, 169 (2015).
- Szklarczyk, D. et al. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).
- Aden, N. et al. Proteomic analysis of scleroderma lesional skin reveals activated wound healing phenotype of epidermal cell layer. *Rheumatology (Oxford)* **47**, 1754–1760 (2008).

21. Giusti, L. *et al.* Specific proteins identified in whole saliva from patients with diffuse systemic sclerosis. *J. Rheumatol.* **34**, 2063–2069 (2007).
22. Guerranti, R. *et al.* Serum proteome of patients with systemic sclerosis: Molecular analysis of expression and prevalence of hapto-globin alpha chain isoforms. *Int. J. Immunopathol. Pharmacol.* **23**, 901–909 (2010).
23. Landi, C. *et al.* Towards a functional proteomics approach to the comprehension of idiopathic pulmonary fibrosis, sarcoidosis, systemic sclerosis and pulmonary Langerhans cell histiocytosis. *J. Proteomics* **83**, 60–75 (2013).
24. Larsen, K. *et al.* Functional and phenotypical comparison of myofibroblasts derived from biopsies and bronchoalveolar lavage in mild asthma and scleroderma. *Respir. Res.* **7**, 11 (2006).
25. Rottoli, P. *et al.* Cytokine profile and proteome analysis in bronchoalveolar lavage of patients with sarcoidosis, pulmonary fibrosis associated with systemic sclerosis and idiopathic pulmonary fibrosis. *Proteomics* **5**, 1423–1430 (2005).
26. Scambi, C. *et al.* Comparative proteomic analysis of serum from patients with systemic sclerosis and sclerodermatous GVHD. Evidence of defective function of factor H. *PLoS ONE* **5**, e12162 (2010).
27. Shirahama, R., Miyazaki, Y., Okamoto, T., Inase, N. & Yoshizawa, Y. Proteome analysis of bronchoalveolar lavage fluid in lung fibrosis associated with systemic sclerosis. *Allergol. Int.* **59**, 409–415 (2010).
28. van Bon, L. *et al.* Proteomic analysis of plasma identifies the Toll-like receptor agonists S100A8/A9 as a novel possible marker for systemic sclerosis phenotype. *Ann. Rheum. Dis.* **73**, 1585–1589 (2014).
29. van Bon, L. *et al.* Proteomic analysis of Systemic Sclerosis serum identifies the toll-like receptor agonists S100A8/A9 as a novel possible pathogenic marker. *J. Transl. Med.* **8**, P67 (2010).
30. van Bon, L. *et al.* Proteome-wide analysis and CXCL4 as a biomarker in systemic sclerosis. *N Engl J Med.* **370**, 433–443 (2014).
31. Baldini, C. *et al.* Proteomic analysis of saliva: A unique tool to distinguish primary Sjögren's syndrome from secondary Sjögren's syndrome and other sicca syndromes. *Arthritis Res. Ther.* **13**, R194–R194 (2011).
32. Xiang, Y. *et al.* Comprehensive investigation of disease-specific short peptides in sera from patients with systemic sclerosis: Complement C3f-des-arginine, detected predominantly in systemic sclerosis sera, enhances proliferation of vascular endothelial cells. *Arthritis Rheum.* **56**, 2018–2030 (2007).
33. TR, R., L, V.B., J, B. CXCL4, a novel marker for systemic sclerosis that clearly correlates with the clinical hallmarks and pathological events. *Arthritis Rheum.* **2010**; 62.
34. Ryu, C. *et al.* Bioactive plasma mitochondrial DNA is associated with disease progression in scleroderma-associated interstitial lung disease. *Arthritis Rheumatol.* **72**, 1905–1915 (2020).
35. Landi, C. *et al.* Bronchoalveolar lavage proteomic analysis in pulmonary fibrosis associated with systemic sclerosis: S100A6 and 14-3-3 ϵ as potential biomarkers. *Rheumatology (Oxford)* **58**, 165–178 (2019).
36. Corallo, C. *et al.* Proteomic investigation of dermal fibroblasts isolated from affected and unaffected skin samples from patients with limited cutaneous systemic sclerosis: 2 distinct entities?. *J. Rheumatol.* **44**, 40–48 (2017).
37. Mullenbrock, S. *et al.* Systems analysis of transcriptomic and proteomic profiles identifies novel regulation of fibrotic programs by miRNAs in pulmonary fibrosis fibroblasts. *Genes* **9**, 588 (2018).
38. Chaigne, B. *et al.* Proteomic analysis of human scleroderma fibroblasts response to transforming growth factor- β . *Proteomics Clin. Appl.* **13**, e1800069 (2019).
39. Dumit, V. I. *et al.* Altered MCM protein levels and autophagic flux in aged and systemic sclerosis dermal fibroblasts. *J. Invest. Dermatol.* **134**, 2321–2330 (2014).
40. Schiller, H. B. *et al.* Deep proteome profiling reveals common prevalence of MZB1-positive plasma B cells in human lung and skin fibrosis. *Am. J. Respir. Crit. Care Med.* **196**, 1298–1310 (2017).
41. Bargagli, E. *et al.* Calgranulin B (S100A9) levels in bronchoalveolar lavage fluid of patients with interstitial lung diseases. *Inflammation* **31**, 351–354 (2008).
42. Bogatkevich, G. S., Ludwicka-Bradley, A., Singleton, C. B., Bethard, J. R. & Silver, R. M. Proteomic analysis of CTGF-activated lung fibroblasts: Identification of IQGAP1 as a key player in lung fibroblast migration. *Am. J. Physiol. Lung Cell Mol. Physiol.* **295**, L603–L611 (2008).
43. Chiang, T. M. & Postlethwaite, A. E. Increase in phosphatidylinositol-3 kinase activity by nitrotyrosylation of lysates of platelets from patients with systemic sclerosis. *Biochim. Biophys. Acta.* **1760**, 32–37 (2006).
44. Coral-Alvarado, P. *et al.* Using proteomic analysis for studying the skin fibroblast protein profile in systemic sclerosis (INTECH Open Access Publisher, London, 2012).
45. De Santis, M. *et al.* β -Thymosins and interstitial lung disease: Study of a scleroderma cohort with a one-year follow-up. *Respir. Res.* **12**, 22 (2011).
46. Del Galdo, F., Shaw, M. A. & Jimenez, S. A. Proteomic analysis identification of a pattern of shared alterations in the secretome of dermal fibroblasts from systemic sclerosis and nephrogenic systemic fibrosis. *Am. J. Pathol.* **177**, 1638–1646 (2010).
47. Fietta, A. *et al.* Analysis of bronchoalveolar lavage fluid proteome from systemic sclerosis patients with or without functional, clinical and radiological signs of lung fibrosis. *Arthritis Res. Ther.* **8**, R160 (2006).
48. Pattanaik, D., Brown, M., Postlethwaite, B. C. & Postlethwaite, A. E. Pathogenesis of systemic sclerosis. *Front. Immunol.* **6**, 272 (2015).
49. Mehta, B. K., Espinoza, M. E., Hinchcliff, M. & Whitfield, M. L. Molecular “omic” signatures in systemic sclerosis. *Eur. J. Rheumatol.* **7**, S173–S180 (2020).
50. Chen, M., Daha, M. R. & Kallenberg, C. G. M. The complement system in systemic autoimmune disease. *J. Autoimmun.* **34**, J276–J286 (2010).
51. Senaldi, G., Lupoli, S., Vergani, D. & Black, C. M. Activation of the complement system in systemic sclerosis relationship to clinical severity. *Arthritis Rheum.* **32**, 1262–1267 (1989).
52. Ludwicka-Bradley, A., Silver, R. M. & Bogatkevich, G. S. Coagulation and autoimmunity in scleroderma interstitial lung disease. *Semin. Arthritis Rheum.* **41**, 212–222 (2011).
53. Riedhammer, C. & Weissert, R. Antigen presentation, autoantigens, and immune regulation in multiple sclerosis and other autoimmune diseases. *Front. Immunol.* **6**, 322 (2015).
54. Acosta-Herrera, M. *et al.* Comprehensive analysis of the major histocompatibility complex in systemic sclerosis identifies differential HLA associations by clinical and serological subtypes. *Ann. Rheum. Dis.* **80**, 1040–1047 (2021).
55. Miyabe, Y., Lian, J., Miyabe, C. & Luster, A. D. Chemokines in rheumatic diseases: Pathogenic role and therapeutic implications. *Nat. Rev. Rheumatol.* **15**, 731–746 (2019).
56. Nakashima, T. *et al.* Impaired IL-17 signaling pathway contributes to the increased collagen expression in scleroderma fibroblasts. *J. Immunol.* **188**, 3573–3583 (2012).
57. Wakhlu, A. *et al.* Serum interleukin-6, interleukin-17A, and transforming growth factor beta are raised in systemic sclerosis with interstitial lung disease. *Indian J. Rheumatol.* **13**, 107–112 (2018).
58. Stummvoll, G. H. *et al.* Increased transendothelial migration of scleroderma lymphocytes. *Ann. Rheum. Dis.* **63**, 569–574 (2004).
59. Xu, D., Li, T., Wang, R. & Mu, R. Expression and pathogenic analysis of integrin family genes in systemic sclerosis. *Front. Med.* **8**, 1083 (2021).
60. Pauwels, A.-M., Trost, M., Beyaert, R. & Hoffmann, E. Patterns, receptors, and signals: Regulation of phagosome maturation. *Trends Immunol.* **38**, 407–422 (2017).
61. Rosales, C. & Uribe-Querol, E. Phagocytosis: A fundamental process in immunity. *Biomed. Res. Int.* **2017**, 9042851 (2017).

62. Abdolmaleki, F. *et al.* The role of efferocytosis in autoimmune diseases. *Front Immunol.* **9**, 1645 (2018).
63. Davis, K., Boros, P., Keltz, M., Unkeless, J. C. & Fleischmajer, R. Circulating Fcγ receptor-specific autoantibodies in localized and systemic scleroderma. *J. Am. Acad. Dermatol.* **33**, 612–616 (1995).
64. Kadono, T., Tomita, M., Tamaki, Z., Sato, S. & Asano, Y. Serum levels of anti-Fcγ receptor IIB/C antibodies are increased in patients with systemic sclerosis. *J. Dermatol.* **41**, 1009–1012 (2014).
65. Di Cara, F. Peroxisomes in host defense. *PLoS Pathog.* **16**, e1008636 (2020).
66. Peters, J. M., Shah, Y. M. & Gonzalez, F. J. The role of peroxisome proliferator-activated receptors in carcinogenesis and chemoprevention. *Nat. Rev. Cancer* **12**, 181–195 (2012).
67. Wei, J., Bhattacharyya, S. & Varga, J. Peroxisome proliferator-activated receptor γ: Innate protection from excessive fibrogenesis and potential therapeutic target in systemic sclerosis. *Curr. Opin. Rheumatol.* **22**, 671–676 (2010).
68. Shi-wen, X. *et al.* Rosiglitazone alleviates the persistent fibrotic phenotype of lesional skin scleroderma fibroblasts. *Rheumatology (Oxford)* **49**, 259–263 (2010).
69. Marangoni, R. G. *et al.* A candidate gene study reveals association between a variant of the peroxisome proliferator-activated receptor gamma (PPAR-γ) gene and systemic sclerosis. *Arthritis Res. Ther.* **17**, 128 (2015).
70. Wei, J. *et al.* PPARγ downregulation by TGFβ in fibroblast and impaired expression and function in systemic sclerosis: A novel mechanism for progressive fibrogenesis. *PLoS ONE* **5**, e13778 (2010).
71. Ayers, N. B., Sun, C.-M. & Chen, S.-Y. Transforming growth factor-β signaling in systemic sclerosis. *J Biomed Res.* **32**, 3–12 (2018).
72. Dantas, A. T. *et al.* Reassessing the role of the active TGF-β1 as a biomarker in systemic sclerosis: Association of serum levels with clinical manifestations. *Dis. Markers* **2016**, 6064830 (2016).
73. Feng, D. & Gerarduzzi, C. Emerging roles of matricellular proteins in systemic sclerosis. *Int. J. Mol. Sci.* **21**, 4776 (2020).
74. Ren, G., Roberts, A. I. & Shi, Y. Adhesion molecules: Key players in Mesenchymal stem cell-mediated immunosuppression. *Cell Adh. Migr.* **5**(1), 20–22 (2011).
75. Mackay, C.R., von Andrian, U.H. 2014 Chapter 22 - Adhesion Molecules and Chemoattractants in Autoimmunity. In Rose NR, Mackay (Eds.) IRBT-TAD (5th Edn). Boston: Academic Press; London, pp. 297–308.
76. Hasegawa, M. *et al.* Serum adhesion molecule levels as prognostic markers in patients with early systemic sclerosis: A multicentre, prospective, observational study. *PLoS ONE* **9**, e88150 (2014).
77. Ramirez, G. A. *et al.* The role of platelets in the pathogenesis of systemic sclerosis. *Front Immunol.* **3**, 160 (2012).
78. Abdulle, A. E., Diercks, G. F. H., Feelisch, M., Mulder, D. J. & van Goor, H. The role of oxidative stress in the development of systemic sclerosis related vasculopathy. *Front. Physiol.* **9**, 1177 (2018).
79. Doridot, L., Jeljeli, M., Chêne, C. & Batteux, F. Implication of oxidative stress in the pathogenesis of systemic sclerosis via inflammation, autoimmunity and fibrosis. *Redox Biol.* **25**, 101122 (2019).
80. Agache, I., Rădoi, M. & Duca, L. Platelet activation in patients with systemic scleroderma—pattern and significance. *Rom. J. Intern. Med.* **45**, 183–191 (2007).
81. Komiya, Y. & Habas, R. Wnt signal transduction pathways. *Organogenesis* **4**, 68–75 (2008).
82. Wei, J. *et al.* Wnt/β-catenin signaling is hyperactivated in systemic sclerosis and induces Smad-dependent fibrotic responses in mesenchymal cells. *Arthritis Rheum.* **64**, 2734–2745 (2012).

Acknowledgements

The authors would like to thank the Cyprus Institute of Neurology and Genetics for the support.

Author contributions

P.P.C., P.N. and K.C. conceived the study. P.P.C. designed the in silico analysis. P.P.C. carried out the in silico analysis. K.C. supervised. P.P.C. drafted the manuscript. P.N. and K.C. reviewed and edited the manuscript. All authors revised and approved the final manuscript.

Funding

This research received no external funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-29054-5>.

Correspondence and requests for materials should be addressed to K.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023