Arch Rheumatol 2021;36(4):548-559
doi: 10.46497/ArchRheumatol.2021.8625

ORIGINAL ARTICLE

Screening and identification of potential biomarkers and therapeutic targets for systemic sclerosis-associated interstitial lung disease

Biqing Huang[1,2], Jing Li[1], Jiuliang Zhao[1]

[1]*Department of Rheumatology and Clinical Immunology, Peking Union Medical College Hospital, Chinese Academy of Medical Science & Peking Union Medical College Hospital, Beijing, China*
[2]*Graduate School of Peking Union Medical College, Beijing, China*

ABSTRACT

Objectives: This study aims to analyze gene expression in lung tissue and lung fibroblasts of patients with systemic sclerosis-associated interstitial lung disease (SSc-ILD) to identify potential biomarkers and therapeutic targets and to examine its possible role in the pathogenesis of SSc-ILD.

Patients and methods: We obtained datasets from Gene Expression Omnibus (GEO) database, and used Robust Rank Aggregation to calculate the co-expressed differentially-expressed-genes (DEGs) in three chips, then analyzed the function, signaling pathways and the protein-protein interaction network of the DEGs. Finally, we verified the DEGs related to SSc-ILD by three databases of Comparative Toxicogenomics Database (CTD), GENE, and DisGeNET, respectively.

Results: There were 16 co-expressed DEGs related to SSc-ILD in three GEO series, of which six genes were upregulated, and 10 genes were downregulated. The CTD included 29,936 genes related to SSc, and the GENE and DisGeNET databases had 429 genes related to SSc.

Conclusion: The results of gene differential expression analysis suggest that interleukin-6, chemokine ligand 2, intercellular adhesion molecule 1, tumor necrosis factor alpha-induced protein 3, pentraxin 3, and cartilage oligomeric matrix protein may be implicated in the pathogenesis of SSc-ILD and are expected to be potential biomarkers and therapeutic targets for SSc-ILD.

Keywords: Biomarkers, IL-6, interstitial, lung disease, pentraxin 3 , scleroderma, systemic, therapeutic targets, TNFAIP3.

Systemic sclerosis (SSc) is an autoimmune disease characterized by fibrosis and vascular damage of skin and internal organ clinically.[1] The pathogenesis of SSc is still unclear and is thought to involve genetic factors, epigenetic modifications and environmental exposure.[2] As a heterogeneous connective tissue disease, pathogenesis of SSc is related to gender (female-to-male ratio: 3-14:1) and race (higher incidence in European population), with an incidence rate of 50 to 300 cases/million.[3] Involvement of lung, particularly interstitial lung disease (ILD), is the leading cause of death in patients with SSc.[4] Understanding the key molecular mechanisms behind disease manifestations is essential for early diagnosis of SSc-ILD and the development of targeted therapy.

Robust rank aggregation (RRA) is a method which has been used to integrate multiple Gene Expression Omnibus (GEO) series, and achieved good results.[5] In this study, we aimed to explore potential biomarkers and therapeutic targets of SSc-ILD through bioinformatics analysis and to provide new directions for early diagnosis of SSc-ILD and development of therapeutic drugs

Received: December 05, 2020 ***Accepted:*** February 18, 2021 ***Published online:*** June 24, 2021

Correspondence: Jing Li, MD. Department of Rheumatology and Clinical Immunology, Peking Union Medical College Hospital, Chinese Academy of Medical Science & Peking Union Medical College Hospital, 100730 Beijing, China.

Tel: 86-10-6915-5483 e-mail: lijing6515@pumch.cn

Citation:

Huang B, Li J, Zhao J. Screening and identification of potential biomarkers and therapeutic targets for systemic sclerosis-associated interstitial lung disease. Arch Rheumatol 2021;36(4):548-559.

based on RRA to screen out the key genes from the differentially expressed genes (DEGs) in three GEO series (GSE76808, GSE81292, GSE40839).

PATIENTS AND METHODS

Data attainment and processing

Gene expression data of SSc-ILD were obtained from the NCBI-GEO database (https://www.ncbi.nlm.nih.gov/geo). Three GEO series (GSE76808, GSE81292, GSE40839) were chosen in our study with the following inclusion criteria: keywords were "systemic sclerosis", "interstitial lung disease", and "human"; samples were from SSc-ILD patients and healthy controls. The "GEOquery" package of the R programming language was used to obtain the series of matrix files and related annotations of three GEO series (GSE76808, GSE81292, GSE40839) which were related to SSc-ILD. According to the annotation files provided by the GPL571, GPL18991 and GPL96, the expression matrix with probe IDs were converted into gene symbols, probes without a corresponding gene symbol were deleted, and the average value was calculated as the final expression value for genes corresponding to more than one probe. The original data were, then, manipulated with the process of background adjustment, quantile normalization, logarithmic transformation and summarization by using the "limma"[6] package of R language. Then, we used "Normalize Between Array" function to normalize the quantiles within and between arrays of all samples. Table 1 shows the detailed information about the three GEO series.

Identification of the DEGs

We used the empirical Bayes method in the "limma" package to analyze DEGs. Genes with adjusted p value of <0.05 and a |log2 fold change| of >1 were considered DEGs. We used "ggplot2" package of R language to visualize DEGs volcano map, and "pheatmap" package to draw cluster heat map of DEGs. The DEGs selected from three GEO series were sorted by RRA to screen out the key genes of differential expression.

Functional analysis of the DEGs

We used the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (https://david.ncifcrf.gov) to explore the biological functions of DEGs in SSc-ILD. The gene ontology (GO)[7] for DEGs was performed to assign the genes and their products into biological process (BP), cellular component (CC), and molecular function (MF). The Kyoto encyclopedia of Genes and genomes (KEGG)[8] pathways enrichment analysis was utilized to identify functional and metabolic pathways of DEGs. In this study, GO terms with a p value of <0.05 and KEGG pathways with a p value of <0.05 were considered statistically significant.

Identification and verification of Hub genes

The protein-protein interaction (PPI) data of DEGs were downloaded from the database of the Search Tool for the Retrieval of Interacting Genes (STRING) (https://string-db.org), and the PPI network was established and visualized by the Cytoscape[9] software. According to PPI, the degree of connection and the interaction of each node in the network were used to detect the Hub gene. To further verify the key genes, a plugin of Cytoscape, MCODE,[10] was used to find out several functional modules based on the MCODE score, which represented the degree of association between DEGs. Subsequently, we also used another plugin of Cytoscape, Cell Hubba,[11] to explore important nodes in the PPI network through several topological algorithms such as network degree, betweenness, and proximity

Table 1. Information of three GEO series related to SSc-ILD

GEO series	Platform information	Sample source	Sample size (disease: health)	Issued time
GSE76808	GPL571	Lung tissue	14: 3	2017
GSE81292	GPL18991	Lung tissue	15: 5	2016
GSE40839	GPL96	Lung fibroblast	8: 10	2013

GEO: Gene expression omnibus; SSc-ILD: Systemic sclerosis associated interstitial lung disease.

centrality. We selected the first 14 genes identified by each topology algorithm, and found the shared genes as the most important central genes in the network in three ways (degree level in PPI network, MCODE score in cellular Hubba, and topology algorithm).[12]

CTD database verified SSc related DEGs

The Comparative Toxicogenomics Database (CTD) (https://ctdbase.org) is an important public resource database based on literature,[13] which is used to search for the interactions among chemical substances, genes, phenotypes, diseases and environments. In the database, inference scores were calculated based on the original documents to show the relationship between genes and diseases. In this study, we used the CTD database to analyze the association between the Hub genes and SSc, and to identify the relationship between them based on the inference score ranking.

Screening out SSc-related biomarkers and therapeutic targets

To get relevant targets of SSc, we searched through the DisGeNET database (https://www.disgenet.org/home) and the GENE database (https://www.ncbi.nlm.nih.gov/gene) with "systemic sclerosis" as the keyword, and the species was limited to humans.

RESULTS

Methods flowchart

Figure 1 showed the workflow of identification, functional analysis, and verification of DEGs in SSc-ILD.

DEGs in the SSc-ILD

Differential gene expression analysis was performed on three GEO series related to SSc-ILD and healthy controls, and visualized by volcano map and heat map, as shown in Figure 2. The results indicated that there were 599 DEGs in GSE40839, of which 230 were upregulated and 369 were downregulated; 686 DEGs in GSE76808, of which 290 were upregulated and 396 were downregulated; and 408 DEGs in GSE81292, of which 145 were upregulated and 263 were downregulated. The co-expressed DEGs of three datasets were sorted by RRA. The results showed that there were 16 co-expressed DEGs in those three datasets, of which six genes (C-X-C motif chemokine ligand 13 *[CXCL13]*, immunoglobulin heavy constant mu *[IGHM]*, DNA topoisomerase 2-alpha *[TOP2A]*, cartilage oligomeric matrix protein *[COMP]*, centrosomal protein 55 *[CEP55]*, and cyclin B1*[CCNB1]*) were upregulated, and 10 genes (interleukin-6 [IL-6], FosB proto-oncogene, AP-1 transcription factor

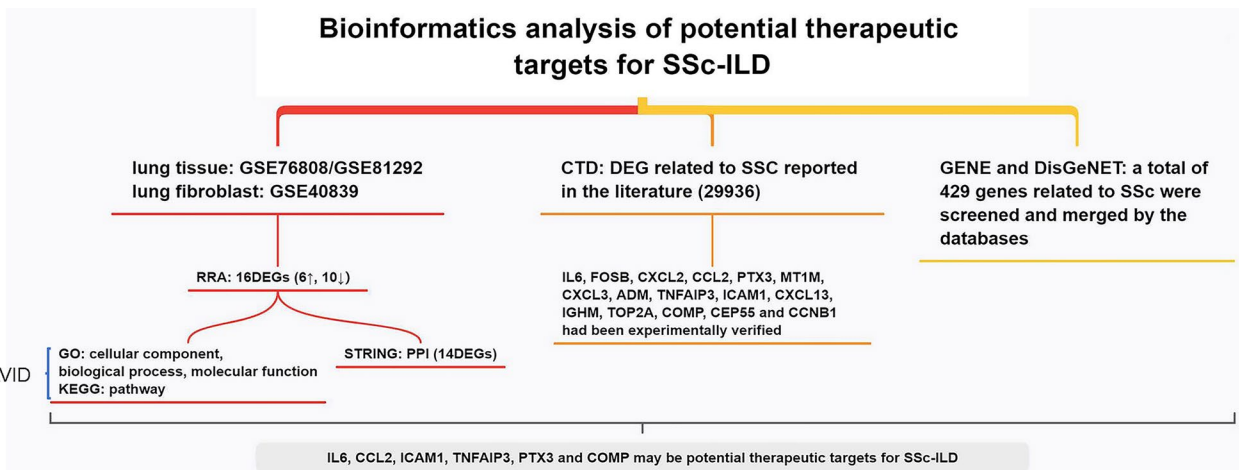

Figure 1. Overall study design. We obtained the datasets of SSs-ILD and healthy controls from GEO database, used RRA to calculate the co-expressed DEGs in three gene expressin chips (GSE76808, GSE1292, GSE40839). And we used DAVID to analyze function and signaling pathways of DEGs, and STRING to analyze the PPI network of DEGs. Finally, we verified the DEGs related to SSc-ILD by three databases of CTD, GENE DisGeNET, respectively.

SSc-ILD: Systemic sclerosis associated interstitial lung disease; GEO: Gene expression omnibus; RRA: Robust nark aggregation; DEGs: Differentially expressed genes; DAVID: Database for annotation, visualization, and integrated discovery; STRING: Search tool for the retrieval of interacting genes; PPI: Protein-protein enteraction; CTD: Comparative toxicogenomics database.

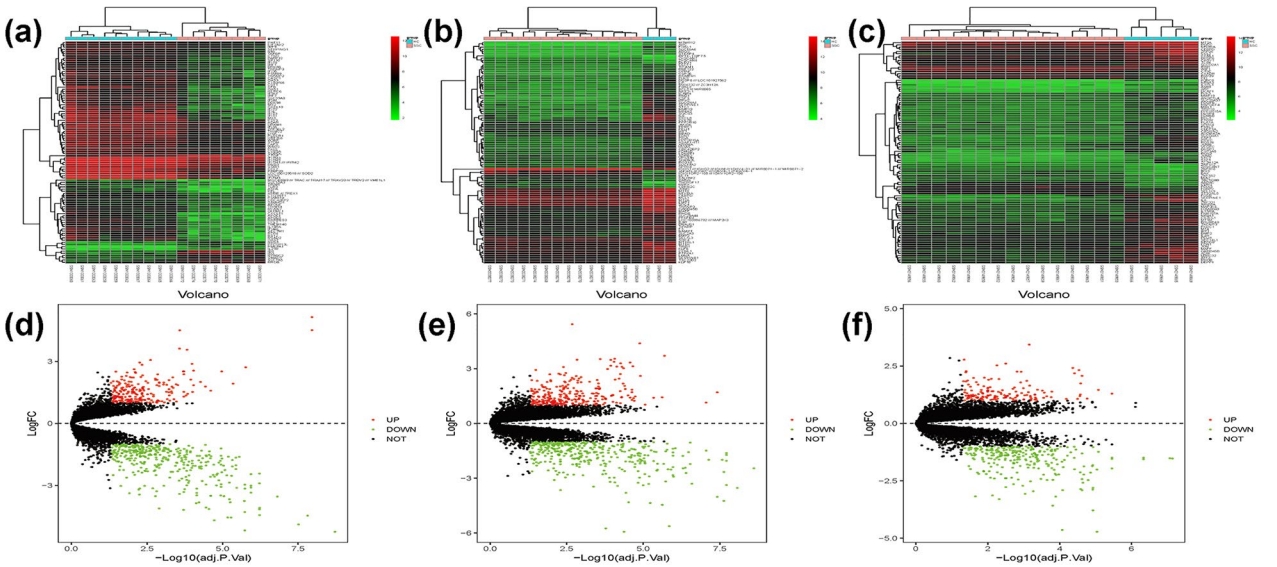

Figure 2. The heat map and volcano map of DEGs in three GEO series of SSc-ILD and healthy controls. **(a-c)** Heat map displayed results of GSE40839, GSE76808, and GSE81292 respectively. **(d-f)** volcano map displayed results of GSE40839, GSE76808, and GSE81292 respectively. The red dots, green dots, and black dots respectively represented genes with up-regulated expression, down-regulated expression and no differential expression when SSc-ILD was compared to healthy controls. DEGs were set to be adjusted p-value <0.05 and |log2 fold change| >1.

DEGs: Differentially expressed genes; SSc-ILD: Systemic sclerosis associated interstitial lung disease.

subunit *[FOSB]*, C-X-C motif chemokine ligand 12 *[CXCL12]*, C-C motif chemokine ligand 2 *[CCL2]*, pentraxin 3 *[PTX3]*, metallothionein 1M *[MT1M]*, C-X-C motif chemokine ligand 3 *[CXCL3]*, adrenomedullin *[ADM]*, tumor necrosis factor alpha-induced protein 3 *[TNFAIP3]*, and intercellular adhesion molecule 1 *[ICAM1]*) were downregulated. The detailed information is shown in Figure 3.

Functional enrichment analysis of DEGs

As shown in Figure 4, six upregulated DEGs and 10 downregulated DEGs were functionally annotated, which were mainly divided into BP, CC, and MF. In BP category, the upregulated genes were mainly related to the process of the mitotic metaphase plate congression, and the downregulated genes were mainly about inflammatory response, cellular response to lipopolysaccharide, and immune response, etc. In CC category, the upregulated genes were mainly related to the component of the extracellular matrix and centriole, and the downregulated genes were mainly about the component of the extracellular matrix. In MF group, the upregulated genes were mainly related to the function of heparin binding, and the downregulated genes were mainly about the function of chemokine activity and chemokine receptor binding. Besides, the signaling-pathway enrichment analysis of DEGs showed that these 16 DEGs were mainly related to tumor necrosis factor signaling pathway, cytokine-cytokine receptor interaction pathway, malaria infection, NOD-like receptor signaling pathway, and chemokine signaling pathway, etc.

Construction of PPI network and selection of Hub gene

Based on the STRING and used the Cytoscape software (version 3.7.2),[14] a PPI network of 16 DEGs was constructed. The network contained 14 nodes, 27 edges, and two functional modules, as shown in Figure 5. The hub gene was identified by the Cell Hubba of the Cytoscape software. The top 10 central nodes with the highest connectivity were IL-6, *CCL2, ICAM1, CXCL3, CXCL2, TNFAIP3, CXCL13, PTX3, ADM,* and *CEP55.*

Validation of Hub gene in CTD database

The CTD included 29936 DEGs related to SSc. Table 2 shows the results of the co-expressed

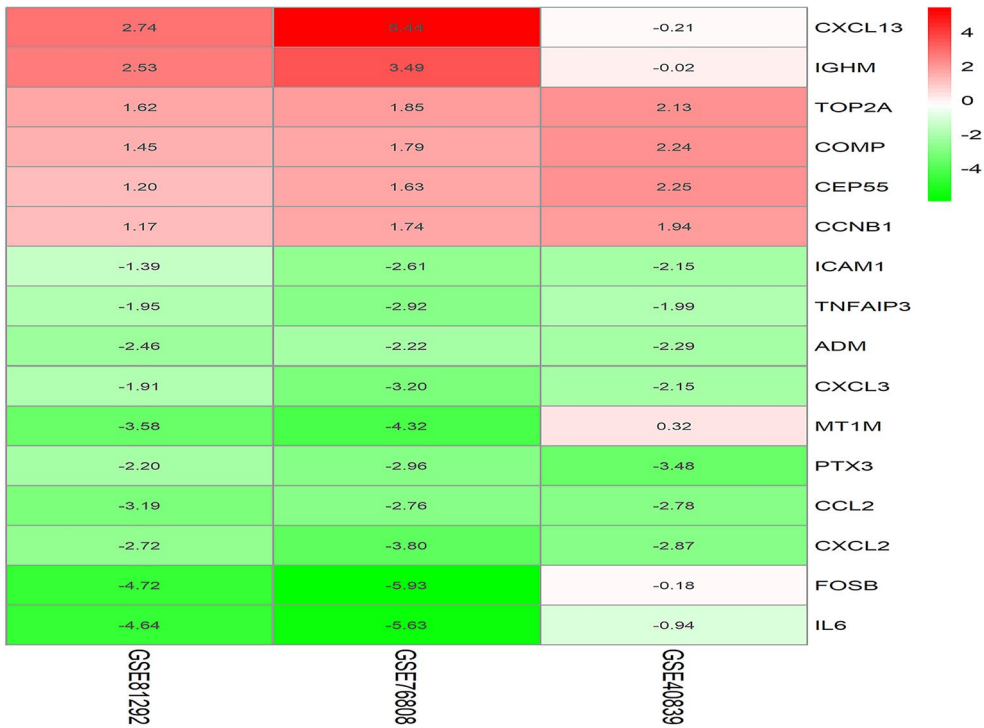

Figure 3. The c-expressed DEGs in three GEO series sorted by RRA. The red color represented up-regulated genes and the green color represented down-regulated genes.

DEGs: Differentially expressed genes; GEO: Gene expression omnibus; RRA: Robust nark aggregation.

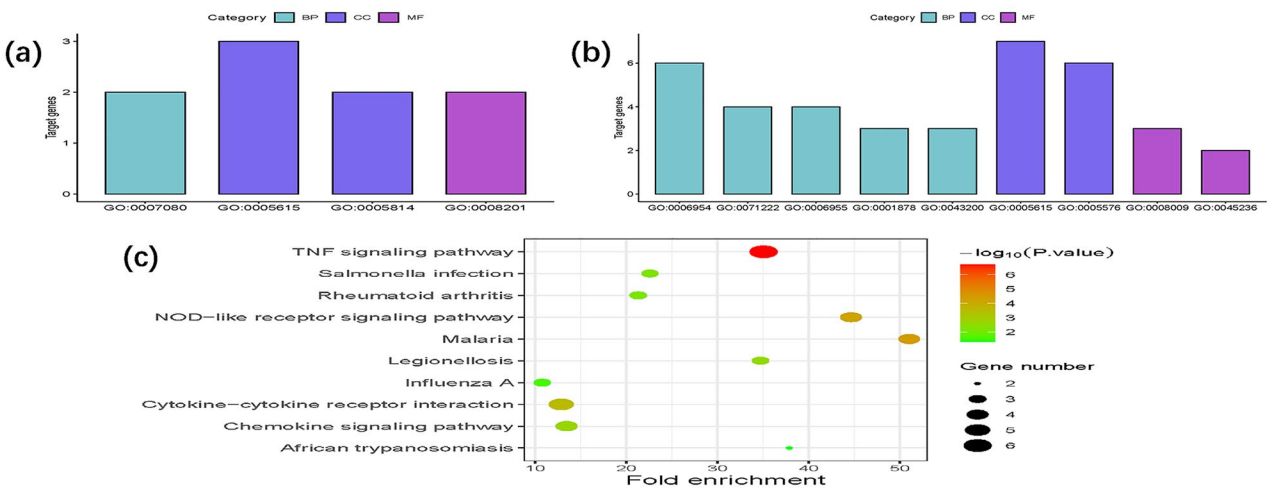

Figure 4. Functional enrichment analysis of DEGs. **(a)** Showed the functional annotations of up-regulated genes, which were mainly related to cell components; **(b)** showed the functional annotations of down-regulated genes, which were mainly related to biological processes; **(c)** showed the KEGG pathway enrichment analysis of DEGs, which were mainly related to the tumor necrosis factor signaling pathway cytokine-cytokine receptor interaction and malaria infection. In this study, GO terms with $p<0.05$ and KEGG pathway with $p<0.05$ were set to be significant.

GO: Gene ontology; KEGG: Kyoto encyclopedia of genes and genomes.

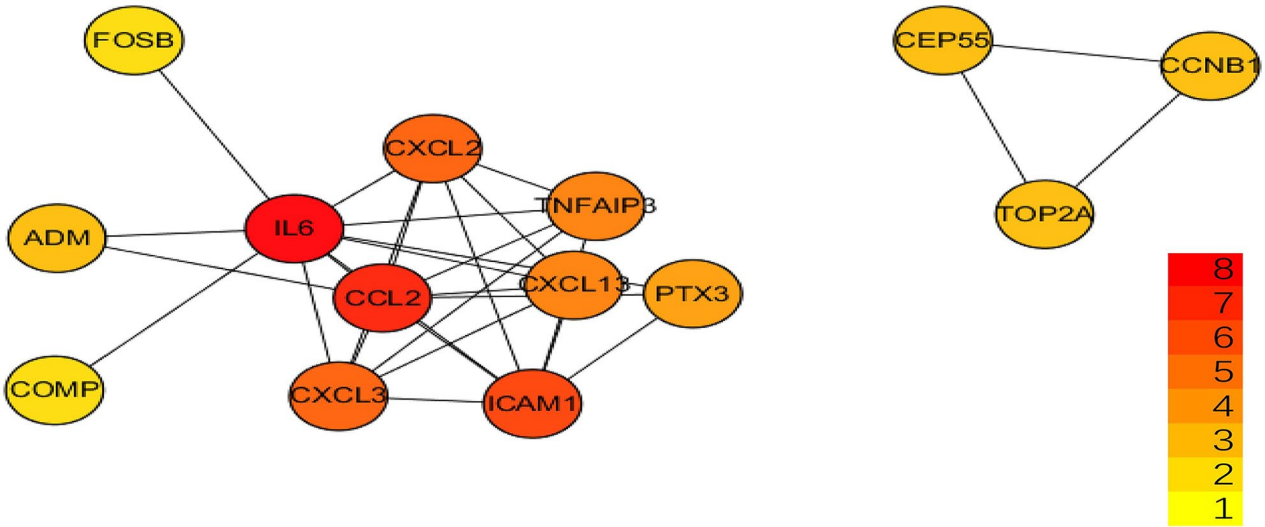

Figure 5. Visual representation of the PPI network of DEGs. The color intensity of each node was positively related to the degree of the node.

PPI: Protein-protein enteraction; DEGs: Differentially expressed genes.

DEGs selected by RRA in three GEO series related to SSc-ILD in the CTD database, and the correlations between each DEG and SSc are listed.

GENE and DisGeNET databases to verify the Hub gene

There were 424 genes related to SSc in the GENE database, including six DEGs *(IL-6, CCL2,*

Table 2. Correlation between 16 DEGs and SSc in CTD Database

Gene	Gene full name	Inference score	Reference score
IL6	Interleukin-6	37.5	47
CCL2	C-C motif chemokine 2	33.11	15
ICAM1	Intercellular adhesion molecule 1	27.56	13
CXCL2	C-X-C motif chemokine 2	25.43	10
CXCL3	C-X-C motif chemokine 3	17.25	6
FOSB	Protein fosB	11.99	6
CCNB1	G2/mitotic-specific cyclin-B1	17.12	6
ADM	Adrenomedullin	15.12	5
TNFAIP3	Tumor necrosis factor alpha-induced protein 3	14.7	5
TOP2A	DNA topoisomerase 2-alpha	12.97	5
CEP55	Centrosomal protein of 55 kDa	14.46	4
CXCL13	C-X-C motif chemokine 13	6.3	4
COMP	Cartilage oligomeric matrix protein	3.8	3
PTX3	Pentraxin-related protein PTX3	5.38	2
IGHM	Immunoglobulin heavy constant Mu	2.95	1
MT1M	Metallothionein-1M	2.59	1

DEGs: Differentially expressed genes; SSc: Systemic sclerosis; CTD: Comparative Toxicogenomics Database.

ICAM1, TNFAIP3, PTX3 and COMP), calculated by RRA from three GEO series in this study. There were 11 genes related to SSc in the DisGeNET database verified experimentally, including Era Like 12S Mitochondrial RRNA Chaperone 1 *(ERAL1)*, fibrillarin *(FBL)*, lymphotoxin alpha *(LTA)*, RNA-binding region-containing protein 3 *(RNPC3)*, WD Repeat Domain 11 *(WDR11)*, endothelin 1 *(EDN1)*, estrogen receptor 1 *(ESR1)*, Human leukocyte antigen *DQB1 (HLA-DQB1)*, *MMP1, TNF*, and *TOP1*. Among them, according to the condition of DEGs set in this study, *EDN1* was differentially expressed in GSE76808 and GSE81292, but not differentially expressed in GSE40839. The *TOP1* was differentially expressed in GSE76808, but not differentially expressed in GSE81292 and GSE40839. The remained genes were not differentially expressed in those three GEO series.

DISCUSSION

In this study, we used the RRA algorithm to screen out the DEGs of the three gene expression profiles of SSc-ILD, combined with CTD, GENE and DisGeNET databases, and finally identified six potential key genes of SSc-ILD, namely *IL-6, CCL2,* and *CCL2*. The *ICAM1, TNFAIP3, COMP,* and *PTX3* were expected to serve as biomarkers and therapeutic targets for SSc-ILD.

Interleukin-6 was a classic pro-inflammatory cytokine and was considered an important protein in the pathogenesis of SSc.[15] Previous studies showed that serum and skin IL-6 levels in SSc patients were elevated, and IL-6 levels were related to the severity of SSc.[15] Similarly, De Lauretis et al.[16] found that serum IL-6 was a predictive marker for early decline of pulmonary function and mortality in SSc-ILD. In another study, serum IL-6 levels were still significantly increased in advanced SSc-ILD.[17] Serum IL-6 levels can be also used as an evaluation index for the efficacy of cyclophosphamide in the treatment of SSc-ILD.[18] Studies have confirmed that IL-6 is involved in the pathogenesis of SSc.[19] Interleukin 6 is implicated in the process of pulmonary fibrosis by promoting the upregulation of collagen-related gene expression, promoting the proliferation and differentiation of fibroblasts and myofibroblasts, inhibiting T cell apoptosis, and regulating the balance of Th17 cells and Treg cells.[19,20] In an animal model of SSc, knocking out the IL-6 gene significantly reduced lung inflammation and collagen deposition in mice.[15] In addition, there were several clinical trials using IL-6 receptor monoclonal antibody tocilizumab (TCZ) to treat SSc.[21-25] One of the Phase III clinical trials revealed that TCZ could improve lung function in patients with early and active SSc-ILD.[21] Tocilizumab also showed good efficacy and safety in the treatment of juvenile SSc and significantly improved the lung function of patients.[22] It was an immunoglobulin G antibody that could bind to IL-6 receptor,[26] thereby blocking the IL-6/STAT3/Smad3 axis trans-signaling pathway, which may reduce lung inflammation and fibrosis.[27]

The *CCL2* was a chemokine produced by endothelial cells, monocytes, type II alveolar cells and other cells,[20] and is the main chemokine and activator of monocytes and T cells.[28] It was involved in fibroblast stimulation, myofibroblast differentiation, T cell transport, and Th2 cell phenotypic polarization[20] and, through CCR2, upregulated the expression of transforming growth factor-β (TGF-β), which may stimulate fibroblasts to produce collagen.[28] Both human and animal model studies have shown that *CCL2* promoted fibrosis through a series of mechanisms involving inflammation, angiogenesis, and accumulation of myofibroblasts.[29] Wu et al.[30] reported that the serum *CCL2* level was significantly increased in SSc patients and was related to the presence and severity of ILD, and had a prognostic value in evaluating the lung function and survival rate of SSc-ILD. In addition, the levels of *CCL2* messenger ribonucleic acid (mRNA) and protein in bronchoalveolar lavage (BAL) of SSc-ILD patients were significantly higher than those in the normal control group, and the level of *CCL2* in BAL fluid was negatively correlated with lung function parameters, and positively correlated with computed tomography (CT) scores.[31] There was currently a Phase II clinical trial using *CCL2* monoclonal antibody (carlumab) to treat idiopathic pulmonary fibrosis (IPF); however, the results showed that carlumab could not bring benefits to patients with IPF and might even worsen the lung function of patients; patients receiving carlumab treatment actually had higher total *CCL2* and free *CCL2* in their serum than patients in the placebo

group.[32] It suggested that the overall blockade of *CCL2* might be harmful, as it might play a role in other important anti-fibrotic pathways. A recent research showed that one target of *CCL2* was CCR21+CD41+T cells. These T cells function similarly to regulatory T cells and were found to exert anti-fibrosis effects in animal models.[33] Taken together, in the complex environment of fibrosis, interrupting a single receptor-ligand interaction may not be enough to overcome the activation of multiple competitive and compensatory pathways. Therefore, it is necessary to further study the production and function of *CCL2* and other CCR2 ligands to target *CCL2* as an anti-fibrosis treatment strategy.

The *ICAM1* is an adhesion molecule induced by IL-1β and tumor necrosis factor-α (TNF-α) and other pro-inflammatory cytokines,[34] and expressed in a variety of cell types, including leukocytes, epithelial cells, endothelial cells and fibroblasts cell.[35] Studies have shown that serum *ICAM1* of SSc patients, compared to healthy controls, increased at the time of the initial diagnosis, and negatively correlated with vital capacity significantly.[36] Delle Sedie et al.[37] reported that, regardless of pulmonary complications, serum *ICAM1* levels in SSc increased, and were positively correlated with inflammation markers such as C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR), indicating that *ICAM1* might be a useful biomarker reflecting the inflammatory state of SSc. The SSc animal model showed that the concentration of *ICAM1* in serum and lung tissue homogenate increased and was closely related to the degree of fibrosis,[38] suggesting that *ICAM1* might be involved in the pathogenesis of SSc-ILD.

Ubiquitination is a covalent post-translational modification of a target protein with ubiquitin, which has a profound impact on the stability and activity of its substrate, thereby regulating the immune system at the molecular and cellular levels.[39] The TNFAIP3-encoded A20 is a ubiquitin editing enzyme that negatively regulates the immune response by terminating the activation of NF-κB.[40] A20 interacted with downstream substrates of immune receptors, including Toll-like receptors, nucleotide-bound oligomerization domain receptors, lymphocyte receptors and cytokine receptors, and are the key regulators of inflammatory signaling pathways. Due to its pleiotropic function as a ubiquitin binding protein, deubiquitinase and ubiquitin ligase, and its extensive role in various signaling pathways, abnormal levels of A20 were related to many diseases, such as rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), and SSc.[39] There was a close genetic linkage between A20 and SSc, as a number of cohort studies show that single nucleotide polymorphisms (SNPs) about A20 are closely related to the susceptibility of SSc,[41-44] and the genome-wide association studies and immunochip studies conducted in recent years have also confirmed that the gene variation of TNFAIP3 is related to SSc.[45] Another study observed the expression of A20 by culturing human skin fibroblasts *in vitro*.[46] This study also observed the effect of A20 overexpression or small interfering RNA (siRNA)-mediated A20 knockdown in human skin fibroblasts on the fibrotic responses induced by TGF-β. The results showed that TGF-β induced a continuous downregulation of A20 in normal fibroblasts. A20 overexpression could stop TGF-β induced stimulation of collagen gene expression and myofibroblast transformation, and interfere with typical Smad signaling transduction and Smad dependent transcription reaction.[46] However, siRNA-mediated knockdown of A20 gene could enhance the magnitude of the fibrotic response induced by TGF-β. These results suggest that A20 may negatively regulate the intensity of fibrosis reaction. The SSc-related A20 gene mutations may cause damage to expression or function of A20, and direct inhibition of A20 by TGF-β in fibrotic environment may play an important role in maintaining of the fibrotic response. Upregulating the expression of A20 by drugs to enhance the inhibitory effect of the A20 signaling pathway may be a novel therapeutic strategy.

Bone *COMP* is a pentameric molecule which was first found in cartilage and later in fibrotic tissues of tendons, skin, and lungs. Elevated levels of *COMP* are related to fibrotic status, including SSc.[47] Hesselstrand et al.[48] found that serum *COMP* level was related to the severity of SSc fibrosis, and could be used as a predictor of mortality in SSc patients. The *COMP* promoted the differentiation of skin fibroblasts and lung epithelial cells into myofibroblasts by enhancing the activity of TGF-β, thereby increasing the synthesis of extracellular matrix. The TGF-β

could also promote the expression of *COMP*, form a positive feedback system in the process of fibrosis, and prolong the fibrotic cycle.[49] Animal and cellular experiments on *COMP* showed that it could increase the expression of type I collagen and recombine collagen fibers to harden the extracellular matrix (ECM), which corresponded to the decrease of forced vital capacity in the lungs; fibroblasts without *COMP* did not deposit type I and XII collagen in ECM. Instead, these collagen proteins were retained in the endoplasmic reticulum, suggesting that *COMP* was necessary for their secretion. The lack of *COMP* had a negative effect on collagen assembly and inactivation of fibrosis.[50] Taken together, these findings suggest that *COMP* plays a major role in fibrosis by abnormally increasing type I collagen synthesis and matrix reorganization and the elevated levels of *COMP* can be used as a marker and potential therapeutic target for a variety of fibrotic diseases, including SSc-ILD.

The *PTX3* is an evolutionarily conserved soluble pattern recognition receptor produced by macrophages, dendritic cells, fibroblasts, activated endothelial cells, smooth muscle cells and other cells.[51] It plays a key role in innate immune response and inflammation, as well as tissue damage and remodeling.[51] It has a variety of functions: it may inhibit the elimination of apoptotic substances, produce abundant autoantibodies, amplify the immune response, enhance complement activation, and play a pathogenic role in autoimmune diseases such as RA, SLE, and SSc.[51] The pleiotropic effects of *PTX3* on inflammation and fibrosis, as well as its inhibitory effect on neovascularization, indicate that it is an interesting candidate mediator in the pathogenesis of SSc.[52] In a clinical observation study, the plasma *PTX3* concentration of SSc patients showed a tendency to increase, compared to the control group, although the difference was not statistically significant.[53] Similarly, Ilgen et al.[54] found in a cohort study that there was no significant difference in the distribution of serum *PTX3* between SSc and healthy control. However, patients with diffuse SSc and lung involvement had lower serum *PTX3* levels, compared to patients with localized SSc and no lung involvement. It was also observed in another *in vitro* study that patients with newly diagnosed SSc had significantly higher serum *PTX3* levels than those from healthy donors and cyclophosphamide-treated SSc patients.[55] The inconsistent results between these studies can be attributed to the different severity of the disease in studied populations. Silencing the expression of *PTX3* gene by siRNA could restore the ability of cultured SSc microvascular endothelial cells to form capillary-like blood vessels and promote vascularization.[56] These findings indicate that the constitutive production of *PTX3* in the lesion is related to the pathogenesis of SSc.

Systemic sclerosis is a rare inflammatory disease, often associated with ILD, leading to death. The latest advances in early diagnosis and treatment of SSc-ILD indicate that there is an urgent need for biomarkers to assess the overall risk of death, as well as the possibility of disease progression and treatment response. To date, there is no easy-to-use biomarkers to assess the possibility of ILD progression in SSc. This study obtained data from GEO and conducted research based on bioinformatics analysis methods. It was found that IL-6, *CCL2*, *ICAM1*, *TNFAIP3*, *COMP*, and *PTX3* were related to SSc-ILD, providing a direction for finding novel biomarkers and therapeutic targets for SSc-ILD, and is helpful to promote the research progress of pathogenesis, disease markers, and related treatments for SSc-ILD. However, most of the results still needed molecular experiments such as qPCR and western blotting, as well as cell function and animal model experiments to further verification. Therefore, this experiment provided guidance and reference for follow-up research.

Due to the relatively small number of samples in this experiment, the final results of the experiment may be slightly different from the actual results and, thus, further prospective cohort trials are needed to identify biomarkers of clinical value and potential therapeutic targets for SSc-ILD. In addition, samples from lung tissue could only explain the lung involvement of SSc-ILD, but not the systemic inflammatory state of SSc-ILD, and when they were used as biomarkers of disease activity, they might only indicate the lung involvement of SSc-ILD.

In conclusion, IL-6, *CCL2*, *ICAM1*, *TNFAIP3*, *COMP*, and *PTX3* may be involved in the pathogenesis of SSc/SSc-ILD and have potential roles as SSc-ILD biomarkers and therapeutic targets. Currently, there are anti-IL-6R monoclonal

antibodies and anti-*CCL2* monoclonal antibodies used to treat SSc; however, whether they are effective in the treatment of SSc-ILD still needs to be investigated in further clinical trials. The use of IL-6, *CCL2*, *ICAM1*, TNFAIP3, *COMP*, and *PTX3* as biomarkers for SSc-ILD or the development of drugs targeting these genes for the treatment of SSc-ILD also requires further experimental studies to explore their specific molecular mechanisms.

Declaration of conflicting interests

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The authors received no financial support for the research and/or authorship of this article.

REFERENCES

1. Jin J, Chou C, Lima M, Zhou D, Zhou X. Systemic sclerosis is a complex disease associated mainly with immune regulatory and inflammatory genes. Open Rheumatol J 2014;8:29-42.
2. Volkmann ER, Varga J. Emerging targets of disease-modifying therapy for systemic sclerosis. Nat Rev Rheumatol 2019;15:208-24.
3. Gabrielli A, Avvedimento EV, Krieg T. Scleroderma. N Engl J Med 2009;360:1989-2003.
4. Distler O, Assassi S, Cottin V, Cutolo M, Danoff SK, Denton CP, et al. Predictors of progression in systemic sclerosis patients with interstitial lung disease. Eur Respir J 2020;55:1902026.
5. Võsa U, Kolde R, Vilo J, Metspalu A, Annilo T. Comprehensive meta-analysis of microRNA expression using a robust rank aggregation approach. Methods Mol Biol 2014;1182:361-73.
6. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
7. Gene Ontology Consortium. Gene ontology consortium: Going forward. Nucleic Acids Res 2015;43:D1049-56.
8. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 2017;45:D353-D361.
9. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape string app: Network analysis and visualization of proteomics data. J Proteome Res 2019;18:623-32.
10. Chen S, Yang D, Lei C, Li Y, Sun X, Chen M, et al. Identification of crucial genes in abdominal aortic aneurysm by WGCNA. PeerJ 2019;7:e7873.
11. Nahálková J. The molecular mechanisms associated with PIN7, a protein-protein interaction network of seven pleiotropic proteins. J Theor Biol 2020;487:110124.
12. Sun D, Wan X, Pan BB, Sun Q, Ji XB, Zhang F, et al. Bioinformatics analysis of genes and pathways of CD11b+/Ly6C intermediate macrophages after renal ischemia-reperfusion injury. Curr Med Sci 2018;38:70-7.
13. Luo J, Li H, Liu Z, Li C, Wang R, Fang J, et al. Integrative analyses of gene expression profile reveal potential crucial roles of mitotic cell cycle and microtubule cytoskeleton in pulmonary artery hypertension. BMC Med Genomics 2020;13:86.
14. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498-504.
15. Brown M, O'Reilly S. The immunopathogenesis of fibrosis in systemic sclerosis. Clin Exp Immunol 2019;195:310-21.
16. De Lauretis A, Sestini P, Pantelidis P, Hoyles R, Hansell DM, Goh NS, et al. Serum interleukin 6 is predictive of early functional decline and mortality in interstitial lung disease associated with systemic sclerosis. J Rheumatol 2013;40:435-46.
17. Renaud L, da Silveira WA, Takamura N, Hardiman G, Feghali-Bostwick C. Prominence of IL6, IGF, TLR, and bioenergetics pathway perturbation in lung tissues of scleroderma patients with pulmonary fibrosis. Front Immunol 2020;11:383.
18. Numajiri H, Yoshizaki A, Fukasawa T, Ebata S, Nakamura K, Yamashita T, et al. Rapid alteration of serum interleukin-6 levels may predict the reactivity of i.v. cyclophosphamide pulse therapy in systemic sclerosis-associated interstitial lung disease. J Dermatol 2018;45:1221-4.
19. O'Reilly S, Cant R, Ciechomska M, van Laar JM. Interleukin-6: A new therapeutic target in systemic sclerosis? Clin Transl Immunology 2013;2:e4.
20. Bonhomme O, André B, Gester F, de Seny D, Moermans C, Struman I, et al. Biomarkers in systemic sclerosis-associated interstitial lung disease: review of the literature. Rheumatology (Oxford) 2019;58:1534-46.
21. Khanna D, Lin CJF, Furst DE, Goldin J, Kim G, Kuwana M, et al. Tocilizumab in systemic sclerosis: A randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Respir Med 2020;8:963-74.
22. Adrovic A, Yildiz M, Haslak F, Koker O, Aliyeva A, Sahin S, et al. Tocilizumab therapy in juvenile systemic sclerosis: A retrospective single centre pilot study. Rheumatol Int 2021;41:121-8.
23. Shima Y, Kawaguchi Y, Kuwana M. Add-on tocilizumab versus conventional treatment for systemic sclerosis

and cytokine analysis to identify an endotype to tocilizumab therapy. Mod Rheumatol 2019;29:134-9.
24. Zacay G, Levy Y. Outcomes of patients with systemic sclerosis treated with tocilizumab: Case series and review of the literature. Best Pract Res Clin Rheumatol 2018;32:563-71.
25. Fernández-Codina A, Fernández-Fernández J, Fernández-Pantiga A. Use of tocilizumab in systemic sclerosis: A brief literature review. Rev Clin Esp 2018;218:298-304.
26. Shima Y. The benefits and prospects of interleukin-6 inhibitor on systemic sclerosis. Mod Rheumatol 2019;29:294-301.
27. Epstein Shochet G, Brook E, Bardenstein-Wald B, Shitrit D. TGF-β pathway activation by idiopathic pulmonary fibrosis (IPF) fibroblast derived soluble factors is mediated by IL-6 trans-signaling. Respir Res 2020;21:56.
28. Hasegawa M. Biomarkers in systemic sclerosis: Their potential to predict clinical courses. J Dermatol 2016;43:29-38.
29. Osafo-Addo AD, Herzog EL. CCL2 and T cells in pulmonary fibrosis: An old player gets a new role. Thorax 2017;72:967-8.
30. Wu M, Baron M, Pedroza C, Salazar GA, Ying J, Charles J, et al. CCL2 in the circulation predicts long-term progression of interstitial lung disease in patients with early systemic sclerosis: Data from two independent cohorts. Arthritis Rheumatol 2017;69:1871-8.
31. Matsushita T, Takehara K. An update on biomarker discovery and use in systemic sclerosis. Expert Rev Mol Diagn 2017;17:823-33.
32. Raghu G, Martinez FJ, Brown KK, Costabel U, Cottin V, Wells AU, et al. CC-chemokine ligand 2 inhibition in idiopathic pulmonary fibrosis: A phase 2 trial of carlumab. Eur Respir J 2015;46:1740-50.
33. Milger K, Yu Y, Brudy E, Irmler M, Skapenko A, Mayinger M, et al. Pulmonary CCR2+CD4+ T cells are immune regulatory and attenuate lung fibrosis development. Thorax 2017;72:1007-20.
34. Thakkar V, Patterson KA, Stevens W, Wilson M, Roddy J, Sahhar J, et al. Increased serum levels of adhesion molecules ICAM-1 and VCAM-1 in systemic sclerosis are not specific for pulmonary manifestations. Clin Rheumatol 2018;37:1563-71.
35. Hopkins AM, Baird AW, Nusrat A. ICAM-1: Targeted docking for exogenous as well as endogenous ligands. Adv Drug Deliv Rev 2004;56:763-78.
36. Hasegawa M, Asano Y, Endo H, Fujimoto M, Goto D, Ihn H, et al. Serum adhesion molecule levels as prognostic markers in patients with early systemic sclerosis: A multicentre, prospective, observational study. PLoS One 2014;9:e88150.
37. Delle Sedie A, Riente L, Maggiorini L, Pratesi F, Tavoni A, Migliorini P, et al. Potential biomarkers in patients with systemic sclerosis. Int J Rheum Dis 2018;21:261-5.
38. Fernandez IE, Amarie OV, Mutze K, Königshoff M, Yildirim AÖ, Eickelberg O. Systematic phenotyping and correlation of biomarkers with lung function and histology in lung fibrosis. Am J Physiol Lung Cell Mol Physiol 2016;310:L919-27.
39. Mooney EC, Sahingur SE. The Ubiquitin System and A20: Implications in health and disease. J Dent Res 2021;100:10-20.
40. Martens A, van Loo G. A20 at the crossroads of cell death, inflammation, and autoimmunity. Cold Spring Harb Perspect Biol 2020;12:a036418.
41. Dieudé P, Guedj M, Wipff J, Ruiz B, Riemekasten G, Matucci-Cerinic M, et al. Association of the TNFAIP3 rs5029939 variant with systemic sclerosis in the European Caucasian population. Ann Rheum Dis 2010;69:1958-64.
42. Koumakis E, Giraud M, Dieudé P, Cohignac V, Cuomo G, Airò P, et al. Brief report: Candidate gene study in systemic sclerosis identifies a rare and functional variant of the TNFAIP3 locus as a risk factor for polyautoimmunity. Arthritis Rheum 2012;64:2746-52.
43. Terao C, Ohmura K, Kawaguchi Y, Nishimoto T, Kawasaki A, Takehara K, et al. PLD4 as a novel susceptibility gene for systemic sclerosis in a Japanese population. Arthritis Rheum 2013;65:472-80.
44. Wei P, Yang Y, Guo X, Hei N, Lai S, Assassi S, et al. Identification of an association of TNFAIP3 polymorphisms with matrix metalloproteinase expression in fibroblasts in an integrative study of systemic sclerosis-associated genetic and environmental factors. Arthritis Rheumatol 2016;68:749-60.
45. Orvain C, Assassi S, Avouac J, Allanore Y. Systemic sclerosis pathogenesis: Contribution of recent advances in genetics. Curr Opin Rheumatol 2020;32:505-14.
46. Bhattacharyya S, Wang W, Graham LV, Varga J. A20 suppresses canonical Smad-dependent fibroblast activation: Novel function for an endogenous inflammatory modulator. Arthritis Res Ther 2016;18:216.
47. Andréasson K, Jönsson G, Hesselstrand R, Norrgren H. Persistent elevation of fibrosis biomarker cartilage oligomeric matrix protein following hepatitis C virus eradication. World J Hepatol 2019;11:330-4.
48. Hesselstrand R, Wildt M, Bozovic G, Andersson-Sjöland A, Andréasson K, Scheja A, et al. Biomarkers from bronchoalveolar lavage fluid in systemic sclerosis patients with interstitial lung disease relate to severity of lung fibrosis. Respir Med 2013;107:1079-86.
49. Posey KL, Coustry F, Hecht JT. Cartilage oligomeric matrix protein: COMPopathies and beyond. Matrix Biol 2018;71-72:161-73.
50. Schulz JN, Nüchel J, Niehoff A, Bloch W, Schönborn K, Hayashi S, et al. COMP-assisted collagen secretion--a novel intracellular function required for fibrosis. J Cell Sci 2016;129:706-16.

51. Wu Q, Cao F, Tao J, Li X, Zheng SG, Pan HF. Pentraxin 3: A promising therapeutic target for autoimmune diseases. Autoimmun Rev 2020;19:102584.
52. Shirai Y, Okazaki Y, Inoue Y, Tamura Y, Yasuoka H, Takeuchi T, et al. Elevated levels of pentraxin 3 in systemic sclerosis: Associations with vascular manifestations and defective vasculogenesis. Arthritis Rheumatol 2015;67:498-507.
53. Kawashiri SY, Nishino A, Igawa T, Takatani A, Shimizu T, Umeda M, et al. Prediction of organ involvement in systemic sclerosis by serum biomarkers and peripheral endothelial function. Clin Exp Rheumatol 2018;36 Suppl 113:102-8.
54. İlgen U, Yayla ME, Düzgün N. Low serum fibroblast growth factor 2 levels not accompanied by increased serum pentraxin 3 levels in patients with systemic sclerosis. Clin Rheumatol 2017;36:367-72.
55. Borghini A, Manetti M, Nacci F, Bellando-Randone S, Guiducci S, Matucci-Cerinic M, et al. Systemic sclerosis sera impair angiogenic performance of dermal microvascular endothelial cells: Therapeutic implications of cyclophosphamide. PLoS One 2015;10:e0130166.
56. Margheri F, Serratì S, Lapucci A, Chillà A, Bazzichi L, Bombardieri S, et al. Modulation of the angiogenic phenotype of normal and systemic sclerosis endothelial cells by gain-loss of function of pentraxin 3 and matrix metalloproteinase 12. Arthritis Rheum 2010;62:2488-98.