Journal of Inflammation Research

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Identification of Novel Biomarkers and Potential Therapeutic Targets for Systemic Sclerosis: An Integrated Analysis of Plasma Proteome-Wide Mendelian Randomization and Transcriptome

Hanchao Li[1], Qian Li[1], Xiaoxin Chen[2], Lingfei Mo[1], Yulu Wang[2], Xinyi Liu[1,2], Xiaohao Wang[3], Zechao Qu[3], Jing Wang[1], Yuanyuan Li[1]

[1]Department of Rheumatology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, People's Republic of China; [2]Health Science Center, Xi'an Jiaotong University, Xi'an, People's Republic of China; [3]Department of Spine Surgery, Hong Hui Hospital, Xi'an Jiaotong University, Xi'an, People's Republic of China

Correspondence: Yuanyuan Li; Jing Wang, Department of Rheumatology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, 710061, People's Republic of China, Email wudui220@163.com; kidip@163.com

Journal of Inflammation Research downloaded from https://www.dovepress.com/
For personal use only.

Background: Systemic sclerosis (SSc) is a complex autoimmune connective tissue disease. This study aimed to identify novel biomarkers for SSc through an integrated analysis of plasma proteome-wide Mendelian randomization (MR) and transcriptome, as well as to explore the potential mechanisms.

Methods: The data used were obtained from public databases. Initially, key plasma proteins causally associated with SSc were identified through a two-sample MR analysis. Subsequently, based on the key diseases related to both key plasma proteins (genes) and key drugs targeting these proteins (genes), phenotype scanning was conducted to predict potential adverse side effects of key plasma proteins (genes). Single-cell RNA sequencing (scRNA-seq) analysis was performed to identify key cell types in GSE138669 dataset. Differentially expressed genes (DEGs) within key cell types in SSc were intersected with genes encoding key plasma proteins to obtain candidate biomarkers, whose functions were subsequently explored. By analyzing candidate biomarker expression in GSE138669 and GSE181549 datasets, the biomarkers were identified. Further exploration included regulatory network, cellular heterogeneity, and cell trajectory analyses.

Results: Initially, 106 plasma proteins (corresponding to 104 genes) were identified. It was revealed that targeting 12 key plasma proteins (like CD40LG) for treating SSc might lead to side effects related to specific key diseases (like mesothelioma). After recognizing epithelial cells and fibroblasts as key cell types, 8 candidate biomarkers associated with pathways like "proteasome" were identified. Notably, CCL19 and LOXL2 were identified as biomarkers, which exhibited elevated expression in SSc. Regulatory elements such as FOXL1 and hsa-miR-5001-5p were predicted to target biomarkers. Remarkably, differentiation stages of key cell type with heterogeneity and the biomarker expression patterns across these stages might be associated with SSc progression.

Conclusion: CCL19 and LOXL2 were identified as novel biomarkers for SSc, providing insights into the exploration of the disease's pathogenesis and the development of new therapeutic targets.

Keywords: systemic sclerosis, mendelian randomization, plasma proteome, biomarker, single-cell RNA sequencing

Background

Systemic sclerosis (SSc), also known as scleroderma, is a complex autoimmune connective tissue disease characterized by progressive fibrosis of the skin and internal organs, microvascular dysfunction, and dysregulated immune responses.[1] The global incidence of SSc is estimated to range from 1.4 to 8.6 cases per 100,000 person-years.[2] Although relatively rare, SSc has the highest mortality rate among connective tissue diseases due to the lack of effective treatments.[1] The five-year survival rate of SSc is 74.9%, but it drops sharply to 40% in patients with visceral involvement.[3] Early

Received: 23 September 2025
Accepted: 3 December 2025
Published: 17 December 2025

© 2025 Li et al. This work is published and licensed by Dove Medical Press Limited. The full terms of this license are available at https://www.dovepress.com/terms.php and incorporate the Creative Commons Attribution – Non Commercial (unported, v4.0) License (http://creativecommons.org/licenses/by-nc/4.0/). By accessing the work you hereby accept the Terms. Non-commercial uses of the work are permitted without any further permission from Dove Medical Press Limited, provided the work is properly attributed. For permission for commercial use of this work, please see paragraphs 4.2 and 5 of our Terms (https://www.dovepress.com/terms.php).

diagnosis has the potential to improve patient outcomes.[4] However, the absence of reliable biomarkers to predict disease progression poses a significant challenge in determining the optimal timing for pharmacological intervention in early-stage SSc.[5] New diagnostic biomarkers are continuously being explored, including ultrasonography[6] as well as various profibrotic factors, interleukins, and chemokines.[7] Despite these advances, current markers are still inadequate for meeting the clinical requirement of accurate and early disease diagnosis.[5] Moreover, despite advancements in anti-fibrotic therapies, effective treatment options for SSc remain limited. Emerging therapeutic agents, such as tocilizumab and nintedanib, have demonstrated only modest efficacy in clinical trials.[1] Therefore, the continued exploration of reliable biomarkers and novel therapeutic targets are essential for improving treatment outcomes in SSc.

Plasma proteins play a crucial role in maintaining normal physiological functions and disease pathogenesis, serving as major sources of disease biomarkers and therapeutic targets.[8] One major source of plasma proteins is their appearance in blood circulation due to cellular leakage or active secretion.[9] Fibroblasts and epithelial cells are key target cells in the process of skin fibrosis. They secrete various chemokines and cytokines, which regulate the migration, localization, and function of immune cells within skin tissues, thereby inducing local inflammation and immune responses.[10] Additionally, they produce large amounts of extracellular matrix (ECM), promoting ECM deposition and directly leading to tissue fibrosis.[11] The detection of secreted proteins derived from fibroblasts and epithelial cells in plasma theoretically provides potential biomarkers and therapeutic targets for assessing fibrotic diseases. For example, hyaluronic acid and type III procollagen N-terminal peptide (PIIINP) have already been applied clinically in the diagnosis and evaluation of fibrosis in organs such as the liver and lung.[12,13]

Genome-wide association studies (GWAS) have identified thousands of protein quantitative trait loci (pQTLs) associated with plasma protein levels.[8] Mendelian randomization (MR), which leverages the random allocation of genetic variants at conception and their associations with disease risk factors, provides a powerful framework for inferring causal relationships between human phenotypes. Using pQTL and GWAS data, two-sample MR analysis can elucidate causal links between plasma proteins and diseases, thereby facilitating the identification of potential therapeutic targets. Single-cell RNA sequencing (scRNA-seq) enables transcriptomic analysis at the single-cell level, allowing for the identification of cellular heterogeneity and dynamic changes within complex tissues. MR analyses have already reported potential drug targets for various diseases, including diabetes,[14] alzheimer's disease,[15] cancer[16] etc. However, to date, no studies have utilized MR analysis in combination with human plasma proteomics to explore potential drug targets for SSc. By integrating Mendelian randomization of plasma proteins in SSc patients with single-cell transcriptomic profiling of skin tissues, it is possible to identify key plasma proteins derived from fibroblasts and epithelial cells.

In this study, we conducted a large-scale plasma proteome MR analysis using plasma pQTLs as instrumental variables and identified key plasma proteins associated with SSc. Differentially expressed genes (DEGs) in fibroblasts and epithelial cells were further analyzed using single-cell RNA sequencing (scRNA-seq) data from SSc skin samples. The intersection of key plasma proteins and DEGs in fibroblasts and epithelial cells was taken to identify candidate genes. Subsequently, the expression levels of these candidate genes were validated in another scRNA-seq dataset and transcriptomic database to identify critical pathogenic molecules in SSc. This study aims to identify novel SSc biomarkers and therapeutic targets while providing new insights into the pathogenesis of SSc.

Methods

Data Acquisition

For the two-sample Mendelian randomization (MR) analysis, SSc and plasma proteins were selected as the outcome and exposure factors, respectively. The trait IDs of plasma protein were obtained by searching for "prot-a" in the Integrative Epidemiology Unit (IEU) open genome-wide association study (GWAS) (IEU OpenGWAS) database (https://gwas.mrcieu.ac.uk/). The outcome data (finn-b-M13_SYSTSLCE) were derived from the same database through searching for "systemic sclerosis". This dataset included 213,447 samples (SSc: control = 302: 213,145) and 16,380,454 single-nucleotide polymorphisms (SNPs) from individuals of European descent.

For further analyses, two RNA sequencing (RNA-seq) datasets were also included in this study. GSE138669 (platform: GPL18573) and GSE181549 (platform: GPL13497) datasets were obtained from the Gene Expression

Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). Specifically, GSE138669 was a single-cell RNA-seq (scRNA-seq) dataset containing 12 SSc biopsy skin samples and 10 healthy biopsy skin samples from individuals. GSE181549, a bulk RNA-seq dataset, consisted of 113 SSc biopsy skin samples (1st biopsy) and 44 healthy biopsy skin samples from individuals. Notably, in this study, the SSc and healthy samples were categorized into SSc and control groups, respectively.

Flowchart of Study Design

A flowchart illustrating the overall study design and analytical workflow was constructed (Figure S1). In brief, we first used MR to identify key plasma proteins (genes) potentially causally associated with SSc. We then analyzed scRNA-seq data (GSE138669) to determine genes aberrantly expressed in fibroblasts and epithelial cells in SSc skin tissue. The intersection of these two gene sets was further validated across independent datasets (GSE181549) to assess consistency of expression patterns. Molecules showing consistent dysregulation were retained as the final candidates for this study. Detailed analytical procedures are provided in the subsequent sections.

Identification of Key Plasma Proteins Through Two-Sample MR Analysis

The MR analysis was employed for exploring the causal relationship between plasma proteins and SSc. The SNPs, plasma proteins, and SSc were utilized as instrumental variables (IVs), exposure factors, and outcome in MR analysis, respectively. This analysis was conducted based on 3 core assumptions: (1) SNPs exhibited significant correlations with plasma proteins; (2) SNPs influenced SSc only through their effects on plasma proteins; (3) SNPs were independent of any other confounding factors. The extract_instruments function in the TwoSampleMR package (v 0.6.6)[17] was used to read GWAS data for protein quantitative trait loci (pQTL) of plasma proteins and screen SNPs significantly correlated with plasma proteins ($p < 5\times10^{-8}$). The genome-wide significance threshold of $p < 5 \times 10^{-8}$ was applied to ensure strong genetic association with plasma protein levels, thereby satisfying the relevance assumption of Mendelian randomization and minimizing weak instrument bias.[14] To ensure the independence of genetic instruments, Only the SNPs without linkage disequilibrium (LD) were selected (clump = TRUE, $R^2 = 0.001$, window size = 500kb, nSNPs > 2).[14] For SSc data (finn-b-M13_SYSTSLCE), the harmonise_data function was applied to harmonize effect alleles and effect sizes, and the SNPs, plasma proteins, and SSc were matched. To avoid potential weak instrument bias, SNPs with F-statistics below 10 were excluded from the analysis.[14] To perform MR analysis, the mr function was utilized to combine 5 algorithms (MR Egger,[18] Weighted median,[19] Inverse variance weighted (IVW),[20] Simple mode[21] and Weighted mode[22]). The IVW algorithm was used as the primary approach to assess statistical significance ($p < 0.05$). Plasma proteins with odds ratios (ORs) greater than 1 were classified as risk factors for SSc, whereas those with ORs less than 1 were classified as protective factors for SSc. Afterwards, based on SNP effects on SSc and plasma proteins, the mr_scatter_plot function was applied to assess correlations between plasma proteins and SSc, generating scatter plots. Forest plots were constructed using the mr_forest_plot function to evaluate effects of plasma protein SNPs on SSc. Besides, through the mr_funnel_plot function, random-effects analysis was performed, generating funnel plots.

Subsequently, to assess the reliability of MR results, sensitivity analysis was performed using various tests. Initially, to identify heterogeneity, the mr_heterogeneity function was used to conduct Cochran's Q test ($p > 0.05$). Notably, significant heterogeneity ($p < 0.05$) prompted the use of a random-effects IVW model for MR analysis, while in the absence of heterogeneity ($p > 0.05$), a fixed-effects IVW model was used. Then, for detecting horizontal pleiotropy, the mr_pleiotropy_test function was used to conduct horizontal pleiotropy test ($p > 0.05$), ensuring that SNPs did not affect SSc through other pathways. MR-Egger regression and MR-PRESSO tests were conducted to further investigate pleiotropic effects and detect outlier SNPs. Moreover, leave-one-out (LOO) analysis was performed via the mr_leaveoneout function, which iteratively omitted individual SNPs to verify the stability of the causal estimates. Eventually, through the Steiger_filtering function, MR-Steiger directionality test was applied to eliminate the prospect of reverse causation. The evidence of a one-way causal association was strengthened when the correct causal direction was confirmed (TRUE) ($p < 0.05$). Importantly, the plasma proteins causally associated with SSc that passed all these tests were defined as key plasma proteins.

This study was conducted and reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization (STROBE-MR) guidelines.

Key Plasma Protein (Gene)-Key Disease-Key Drug Network Construction

Initially, key drugs targeting key plasma proteins (genes) were identified based on the Drug Signatures database (DSigDB) (https://dsigdb.tanlab.org/) ($p < 0.01$). Moreover, the DisGeNET database (http://www.disgenet.org/) was applied to explore diseases associated with key plasma proteins (genes). Remarkably, the key plasma protein (gene)-key drug and the key plasma protein (gene)-disease networks were established via the Cytoscape software (v 3.7.2).[23] Furthermore, diseases targeted by key drugs were predicted by the Comparative Toxicogenomics Database (CTD) (http://ctdbase.org/).

Importantly, after matching the diseases targeted by key drugs and the key plasma protein (gene)-disease network, the shared diseases were identified and defined as the key diseases, which were potentially affected by key plasma proteins (genes) and key drugs. Likewise, Cytoscape software (v 3.7.2)[23] was utilized to visualize the key plasma protein (gene)-key disease-key drug network.

Evaluation of Potential Adverse Side Effects Through Phenotype Scanning

Through the TwoSampleMR package (v 0.6.6),[17] phenotype scanning (using MR analysis) was employed to assess the potential adverse side effects of targeting key plasma proteins (genes) in SSc treatment. Specifically, key diseases were selected as outcomes, excluding those without corresponding trait IDs in the GWAS database. Key plasma proteins (genes) were considered as exposure factors, and the causal effects of these exposure factors on the outcomes were evaluated. Likewise, an OR greater than 1 indicated that the key plasma protein (gene) acted as a risk factor for corresponding key disease, while an OR less than 1 suggested a protective effect. If a key plasma protein (gene) was simultaneously identified as a risk factor or protective factor in both phenotype scanning and two-sample MR analysis, it was considered a suitable drug target for SSc. However, if the key plasma protein (gene) was identified as a different factor in these 2 analyses, this indicated that targeting this key plasma protein (gene) for SSc treatment could lead to adverse effects related to corresponding key disease. In such cases, the SSc treatment would likely result in unintended effects that were opposite to the intended therapeutic outcomes.

Ascertainment of Key Cell Type and Candidate Biomarkers Through scRNA-Seq Analysis

An extensive investigation was conducted on the scRNA-seq data from the GSE138669 dataset to explore the potential of key plasma proteins (genes) as biomarkers for SSc. Initially, the cell type profiles of SSc and control samples were explored. Specifically, the Seurat package (v 5.0.1)[24] was applied to process the scRNA-seq data. Rigorous quality control criteria were applied to retain high-quality cells (200 < detected genes (nFeature) < 4500, mitochondrial gene content (percent.mito) < 10%, 500 < total gene count (nCount_RNA) < 20,000). Following this, the data were normalized via the NormalizeData function, and the FindVariableFeatures function was employed to identify genes with high variation coefficients across cells. The top 2000 variable genes were retained for further analyses. RunPCA function was applied to perform principal component analysis (PCA), through which the distribution of samples was assessed and potential batch effects were examined. Then, an Elbow Plot was generated to visualize the proportion of variance explained by each principal component. The statistical significance of each principal component was assessed via JackStraw function and visualized by JackStrawPlot function. Principal components with $p < 0.05$ were selected for subsequent analyses. Distinct cell clusters were recognized through uniform manifold approximation and projection (UMAP) clustering (resolution = 0.1). Thereafter, the FindAllMarkers function was used to identify the positive marker genes for each cluster (min.pct = 0.6, only.pos = TRUE, $\log_2$fold change (FC) threshold = 0.5), with the Wilcoxon rank-sum test applied for differential analysis ($p < 0.05$). Cell cluster annotation was performed by referencing marker genes from the literature[25,26] and the CellMarker database (http://xteam.xbio.top/CellMarker/), with the marker gene expression illustrated for each cell type via Dotplot function. After performing differential expression analysis with FindAllMarkers

function, the cell types exhibiting significantly differential expression of specific key plasma proteins (genes) between SSc and control groups were defined as key cell types (|average log_2FC| > 0.5, $p < 0.05$). The union of differentially expressed genes (DEGs) within each key cell type between groups was intersected with the genes encoding key plasma proteins through ggvenn package (v 0.1.9),[27] yielding candidate biomarkers.

Functional Analysis of Key Cell Type and Candidate Biomarkers

Within the GSE138669 dataset, ReactomeGSA package (v 1.12.0)[28] was utilized to perform a functional enrichment analysis on key cell types across all samples, enhancing the understanding of their biological roles.

Additionally, in order to explore signaling pathways in which candidate biomarkers were involved in SSc, gene set enrichment analysis (GSEA) was conducted. Specifically, based on all samples in the GSE138669 dataset, the psych package (v 2.2.9)[29] was applied to conduct Spearman correlation analysis between candidate biomarkers and other genes, and the genes were sorted by correlation coefficients (cor) in descending order. Then, the clusterProfiler package (v 4.2.2)[30] was utilized to conduct GSEA on each candidate biomarker (|normalized enrichment score (NES)| > 1, adjusted $p < 0.05$). The reference gene set "c2.cp.kegg.v7.5.1.symbols.gmt" obtained from Molecular Signatures Database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb) was used.

Recognition and Expression Analysis of Biomarkers

Further screening was conducted on the candidate biomarkers. Specifically, candidate biomarker expression differences between SSc and control groups in GSE138669 and GSE181549 datasets were analyzed through Wilcoxon rank-sum test ($p < 0.05$). Candidate biomarkers showing consistent expression trends and significant differences between groups across datasets were ultimately identified as biomarkers.

Moreover, within the GSE138669 dataset, the distribution and expression of biomarkers in different key cell types were illustrated, and the expression differences between SSc and control groups were analyzed via Wilcoxon rank-sum test ($p < 0.05$).

Regulatory Network Exploration of Biomarkers

To understand the molecular regulatory mechanisms of biomarkers in SSc, the JASPAR database (http://jaspar.genereg.net/) from the NetworkAnalyst platform was employed to predict transcription factors (TFs) targeting biomarkers. Additionally, microRNAs (miRNAs) targeting biomarkers were predicted in TargetScan database (https://www.targetscan.org/). Ultimately, Cytoscape software (v 3.7.2) was used to visualize the TF-mRNA and miRNA-mRNA regulatory networks.

Cellular Heterogeneity and Trajectory Analyses of Key Cell Types

With the purpose of evaluating the heterogeneity of key cell types, using the previously mentioned methods, the key cell types were further clustered into distinct subclusters through UMAP dimensionality reduction (resolution = 0.07) in GSE138669 dataset. Similarly, cell annotation was performed using the FindAllMarkers function (min.pct = 0.5, log_2FC threshold = 0.25, test.use = auc), with reference to marker genes from the literature[31–33] and the CellMarker database. This process facilitated the identification of subtypes within the key cell types. The marker gene expression across different subtypes was illustrated. Next, a trajectory analysis of these subtypes was conducted via the Monocle2 package (v 2.26.0),[34] enabling the simulation of differentiation processes within the key cell types. Furthermore, the branched expression analysis modeling (BEAM) method was applied to investigate the dynamics of biomarker expression during the differentiation of the key cell types.

Human Tissues and Mice Tissues

Skin samples from five SSc patients and age- and sex-matched healthy controls were collected from the Department of Pathology, First Affiliated Hospital of Xi'an Jiaotong University. All patients met the 2013 American College of Rheumatology/European League Against Rheumatism classification criteria for SSc.[35]

Skin tissues from SSc model mice utilized in this study were obtained from previous, unpublished research. Briefly, 15 male ICR mice (4–6 weeks old, weighing 25–30 g) were numbered and randomly assigned to one of three groups

using a random number table: the NC (normal control) group, the BLM (bleomycin) model group, and the HClO (hypochlorous acid) model group. All mice were housed in the same SPF-level animal facility, with cages from all experimental groups placed in adjacent positions on the same rack to ensure identical environmental conditions. The mice were housed with a maximum of 3 animals per cage, and the cage was considered the experimental unit for husbandry purposes. All experimental procedures, including injections, were performed sequentially according to the mouse identification number to minimize order effects. Mice in the BLM and HClO groups received subcutaneous injections of bleomycin (BLM, 2 mg/mL, 100 μL every other day) and hypochlorous acid solution (HClO, containing 9% active chlorine, 100 μL every other day), respectively. The NC group received equal volumes of PBS on the same schedule. The group allocation was not performed in a blinded manner.

After 28 days of modeling, mice were deeply anesthetized via inhalation of isoflurane and then euthanized by cervical dislocation. Skin tissues were collected for analysis. Histopathological examination via Hematoxylin and Eosin (H&E) staining and Masson's trichrome staining revealed a significant increase in dermal thickness in both model groups compared to the NC group, confirming the successful establishment of the SSc models. All animal samples were included in the analysis. The skin tissue sections and skin-derived cDNA from the prior study were used for the current investigation.

Humane endpoints were strictly defined and implemented in this study. The monitored indicators included: (1) Weight Loss: Sustained reduction of body weight exceeding 20% compared to baseline; (2) Behavioral Signs: Severe lethargy, inability to access food or water, or prolonged absence of exploratory behavior; (3) Physical Condition: Signs of persistent pain (e.g., hunched posture), ulcerative skin lesions, or difficulty breathing. These parameters were assessed daily throughout the 28-day modeling period. Any animal meeting the predefined criteria for humane endpoints was euthanized immediately to minimize suffering.

This animal study was conducted in strict accordance with the ARRIVE guidelines.

Immunohistochemistry (IHC)

IHC staining was performed as previously described,[36] using anti-CCL19 antibodies (ER63961, Huabio Biotechnology Co., Ltd., China) and anti-LOXL2 antibodies (ET1706-11, Huabio Biotechnology Co., Ltd., China) at a 1:200 dilution. Quantitative analysis was conducted using ImageJ (National Institutes of Health) with the IHC Toolbox (https://imagej.net/ij/plugins/ihc-toolbox) plugin.

RNA Extraction and qRT-PCR

Mouse skin RNA extraction, reverse transcription, and qRT-PCR procedures can be referenced from previous studies.[37] Briefly, total RNA was extracted using Trizol (Invitrogen; San Diego, CA, USA), and cDNA was synthesized using a reverse transcription kit (Takara Biotechnology; Dalian, China). The primers were shown in Table S1. Relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method and normalized to *Actb*.

Statistical Analysis

R language (v 4.1.0)[38] was employed to conduct bioinformatics analyses. Besides, Wilcoxon rank-sum test was employed in this study to assess differences between groups, setting the significance threshold at $p < 0.05$. Differences in LOXL2 and CCL19 levels in humans and mice were analyzed using Student's *t*-tests (for comparisons between two groups) or one-way analysis of variance (ANOVA, for comparisons among three groups) with the SPSS software (version 26.0; IBM Corp., Armonk, NY, USA). A p-value < 0.05 was considered statistically significant.

Results

Acquisition of Key Plasma Proteins Causally Associated with SSc

Initially, 752 SNPs were retained for the two-sample MR analysis. Based on the IVW algorithm, 109 plasma proteins were identified to have causal relationships with SSc. Specifically, 64 plasma proteins (corresponding to 62 genes) emerged as risk factors related to increased SSc risk ($p < 0.05$, $OR > 1$). Conversely, 45 plasma proteins (corresponding

to 45 genes) were identified as protective factors, which were related to reduced SSc risk ($p < 0.05$, OR < 1) (Table 1). Additionally, scatter plots showed that the identified risk factors were generally positively correlated with SSc, while the protective factors were generally inversely correlated with SSc (Figure S2). Notably, the data points in the scatter plots were mostly centered around 0, indicating non-existent direct associations between SNPs and SSc, with negligible impacts from confounding factors. The robustness of the IVW models in demonstrating significant effects was affirmed by forest plots (Figure S3). Additionally, funnel plots depicted in Figure S4 supported the adherence of the MR analysis results to Mendel's second law of independent assortment.

Furthermore, sensitivity analysis was carried out to further evaluate the reliability of MR analysis results. In detail, Cochran's Q test revealed that 6 plasma proteins exhibited significant heterogeneity ($p < 0.05$). In contrast, the other 103 plasma proteins showed no heterogeneity ($p > 0.05$). Therefore, a random-effects IVW model was used for MR analysis (Table S2). Subsequently, horizontal pleiotropy test showed that 107 plasma proteins exhibited no evidence of horizontal pleiotropy ($p > 0.05$), which were included in downstream analysis (Table S3). Remarkably, LOO analysis confirmed stability of MR analysis results, as the effects of SNPs on SSc remained consistent even after the removal of individual SNPs (Figure S5). Moreover, the MR-Steiger directionality test indicated that the causal directions from 106 plasma proteins to SSc were confirmed to be correct ($p < 0.05$) (Table S4). Consequently, these 106 plasma proteins (corresponding to 104 genes) causally associated with SSc were ultimately named as key plasma proteins (genes) (Table 2 and Figure S1).

Complex Key Plasma Protein (Gene)-Key Disease-Key Drug Network

A total of 34 key drugs and 521 diseases related to key plasma proteins (genes) were predicted. Notably, the constructed key plasma protein (gene)-key drug network contained 154 interaction pairs (such as CD74-arsenite) involving 55 key plasma proteins (genes) (Figure 1A). Additionally, the key plasma protein (gene)-disease network included 703 interaction pairs (such as GFAP-depression) involving 521 diseases and 68 key plasma proteins (genes) (Figure 1B). Moreover, 4939 diseases targeted by key drugs were predicted. Subsequently, 114 key diseases potentially affected by key plasma proteins (genes) and key drugs were identified, and a key plasma protein (gene)-key disease-key drug network was established. This network comprised 5221 interaction pairs (such as GFAP-depression) involving 44 key plasma proteins (genes), key drugs (corresponding to 97 names in the CTD database), and 114 key diseases. These networks elucidated the complex relationships between key plasma proteins (genes), drugs, and diseases, providing insights into potential therapeutic strategies and biomarkers for various conditions.

Potential Adverse Side Effects of Targeting Key Plasma Proteins in SSc Treatment

For evaluating the potential adverse side effects of targeting key plasma proteins (genes) in SSc treatment, the causal relationships between 56 key diseases (with corresponding trait IDs in the GWAS database) and key plasma proteins (genes) were investigated through phenotype scanning. It was revealed that targeting CD40LG, CD74, CEACAM21, COL9A2, EDAR, ELL2, ESR1, GRIA4, IL1RN, LAG3, LOXL2, and NELL1 as potential drug targets for SSc could lead to specific key disease-related side effects due to reversed causal relationships (Figure 2A–L). These inverse relationships suggested that these targets, while beneficial in treating SSc, might inadvertently induce effects contrary to the intended therapeutic outcomes. For example, targeting CD40LG in SSc may trigger mesothelioma-related adverse effects, as CD40LG is a risk factor for SSc but a protective factor for mesothelioma (Figure 2A). Moreover, CD74 (Figure 2B), ESR1 (Figure 2G), GRIA4 (Figure 2H), and NELL1 (Figure 2L) serve as protective factors in Sjögren's syndrome, indicating that therapies against these molecules may worsen the disease. Likewise, LOXL2 is protective in diffuse brain injury, and its inhibition may aggravate the condition (Figure 2K). These findings highlighted the critical importance of considering potential side effects of targeted therapies in the treatment of SSc, especially in relation to their impact on other associated diseases.

Recognition of Key Cell Types and Candidate Biomarkers

Within the GSE138669 dataset, the potential of key plasma proteins (genes) as biomarkers for SSc was further explored at the single-cell level. Initially, key cell types involved in SSc progression were explored. Specifically, through filtration

Table 1 Plasma Proteins with Causal Relationships with SSc

id.exposure	Outcome	Method	nsnp	b	se	p	OR	OR_lci95	OR_uci95
prot-a-1898	Systemic sclerosis || id:finn-b-M13_SYSTSLCE	IVW	29	−0.358	0.106	0.001	0.699	0.568	0.861
prot-a-1942			11	−0.238	0.108	0.028	0.788	0.637	0.974
prot-a-805			10	0.213	0.097	0.027	1.238	1.024	1.496
prot-a-2030			24	0.233	0.103	0.024	1.263	1.032	1.546
prot-a-541			8	0.492	0.248	0.048	1.635	1.005	2.661
prot-a-149			15	0.363	0.149	0.015	1.437	1.073	1.924
prot-a-2452			13	0.505	0.148	0.001	1.657	1.239	2.217
prot-a-91			11	0.203	0.086	0.019	1.225	1.034	1.451
prot-a-2522			11	−0.386	0.150	0.010	0.680	0.506	0.912
prot-a-1842			19	0.247	0.109	0.024	1.280	1.033	1.586
prot-a-1915			13	0.366	0.182	0.044	1.442	1.010	2.059
prot-a-888			15	−0.321	0.139	0.021	0.726	0.553	0.953
prot-a-503			9	0.567	0.216	0.009	1.762	1.153	2.693
prot-a-1774			10	−0.457	0.218	0.036	0.633	0.413	0.970
prot-a-3122			16	−0.328	0.160	0.041	0.720	0.526	0.986
prot-a-1777			11	0.525	0.204	0.010	1.690	1.132	2.523
prot-a-447			15	0.365	0.169	0.031	1.441	1.033	2.008
prot-a-2071			15	0.196	0.096	0.042	1.216	1.007	1.469
prot-a-3027			5	−0.685	0.346	0.048	0.504	0.256	0.994
prot-a-1359			15	−0.298	0.138	0.031	0.742	0.566	0.974
prot-a-1839			14	−0.225	0.101	0.026	0.799	0.655	0.974
prot-a-1200			10	0.594	0.229	0.009	1.812	1.156	2.840
prot-a-1023			15	−0.419	0.205	0.041	0.658	0.440	0.983
prot-a-2658			16	0.367	0.145	0.011	1.444	1.087	1.919
prot-a-1261			12	0.310	0.136	0.023	1.364	1.044	1.782
prot-a-2038			21	−0.410	0.151	0.007	0.664	0.494	0.892
prot-a-2946			16	0.471	0.196	0.016	1.601	1.090	2.351
prot-a-1504			16	0.361	0.147	0.014	1.434	1.075	1.914
prot-a-1224			13	−0.383	0.178	0.031	0.681	0.481	0.965
prot-a-1275			18	−0.573	0.164	0.000	0.564	0.408	0.778
prot-a-2401			9	−0.464	0.220	0.035	0.629	0.408	0.968
prot-a-1862			17	0.354	0.168	0.035	1.425	1.026	1.980
prot-a-3157			14	−0.352	0.175	0.045	0.703	0.499	0.992
prot-a-1682			8	0.518	0.231	0.025	1.679	1.067	2.643
prot-a-1245			14	0.356	0.159	0.025	1.428	1.047	1.949
prot-a-1309			13	−0.470	0.187	0.012	0.625	0.433	0.901

prot-a-1079			10	0.525	0.205	0.011	1.690	1.130	2.526
prot-a-654			14	0.346	0.131	0.008	1.414	1.094	1.827
prot-a-2929			18	−0.324	0.152	0.033	0.723	0.537	0.974
prot-a-2277			13	0.237	0.121	0.050	1.267	1.000	1.605
prot-a-140			12	0.402	0.202	0.047	1.495	1.006	2.220
prot-a-1093			16	−0.550	0.259	0.034	0.577	0.347	0.958
prot-a-1785			18	−0.315	0.105	0.003	0.730	0.595	0.896
prot-a-2148			15	0.377	0.178	0.034	1.458	1.028	2.067
prot-a-1231			15	−0.450	0.141	0.001	0.638	0.484	0.840
prot-a-459			9	0.534	0.259	0.040	1.705	1.026	2.834
prot-a-316			19	−0.270	0.126	0.032	0.763	0.596	0.977
prot-a-2692			19	−0.219	0.105	0.037	0.803	0.654	0.987
prot-a-2749			12	−0.406	0.183	0.026	0.666	0.466	0.953
prot-a-2396			13	0.339	0.128	0.008	1.404	1.093	1.804
prot-a-315			20	−0.349	0.140	0.013	0.705	0.536	0.929
prot-a-1979			18	0.355	0.166	0.033	1.427	1.030	1.977
prot-a-1693			10	0.418	0.203	0.040	1.519	1.019	2.262
prot-a-971			10	0.184	0.089	0.039	1.202	1.009	1.431
prot-a-835			15	0.545	0.183	0.003	1.724	1.204	2.470
prot-a-1039			18	0.357	0.164	0.029	1.430	1.036	1.972
prot-a-2588			11	0.378	0.171	0.027	1.459	1.044	2.039
prot-a-911			19	−0.344	0.148	0.020	0.709	0.530	0.948
prot-a-3123			37	−0.175	0.056	0.002	0.839	0.752	0.937
prot-a-2283			14	0.489	0.210	0.020	1.631	1.082	2.460
prot-a-3117			12	0.460	0.168	0.006	1.584	1.140	2.201
prot-a-951			13	−0.389	0.183	0.034	0.678	0.473	0.971
prot-a-3275			8	−0.548	0.266	0.040	0.578	0.343	0.974
prot-a-3031			19	0.260	0.108	0.016	1.297	1.050	1.601
prot-a-1762			10	0.538	0.198	0.006	1.712	1.162	2.521
prot-a-2704			12	0.449	0.209	0.032	1.566	1.039	2.362
prot-a-1218			12	−0.442	0.205	0.031	0.643	0.430	0.960
prot-a-2746			14	0.389	0.177	0.028	1.475	1.043	2.086
prot-a-1537			10	0.487	0.200	0.015	1.627	1.100	2.406
prot-a-1325			12	−0.588	0.225	0.009	0.556	0.357	0.863
prot-a-395			18	0.522	0.195	0.008	1.685	1.149	2.470
prot-a-157			21	0.308	0.153	0.044	1.361	1.008	1.837
prot-a-2859			8	0.512	0.250	0.041	1.668	1.021	2.724
prot-a-355			12	0.471	0.214	0.028	1.601	1.053	2.434
prot-a-557			12	0.490	0.166	0.003	1.632	1.178	2.261

(*Continued*)

Table 1 (Continued).

id.exposure	Outcome	Method	nsnp	b	se	p	OR	OR_lci95	OR_uci95
prot-a-2109			12	−0.812	0.292	0.005	0.444	0.251	0.787
prot-a-2982			14	−0.263	0.113	0.020	0.768	0.616	0.959
prot-a-2915			16	−0.429	0.172	0.013	0.651	0.465	0.912
prot-a-650			16	−0.342	0.170	0.043	0.710	0.509	0.990
prot-a-991			9	0.666	0.223	0.003	1.946	1.257	3.015
prot-a-187			13	0.340	0.165	0.040	1.405	1.016	1.943
prot-a-1987			7	−0.586	0.251	0.020	0.556	0.340	0.911
prot-a-1536			10	0.519	0.248	0.037	1.680	1.032	2.734
prot-a-2025			19	0.249	0.083	0.003	1.283	1.091	1.508
prot-a-523			11	0.412	0.185	0.026	1.510	1.050	2.171
prot-a-642			12	0.374	0.149	0.012	1.454	1.085	1.948
prot-a-454			9	−0.496	0.222	0.025	0.609	0.394	0.940
prot-a-626			11	0.377	0.169	0.025	1.458	1.048	2.029
prot-a-1052			9	0.599	0.239	0.012	1.820	1.139	2.908
prot-a-3109			13	0.246	0.091	0.007	1.279	1.070	1.528
prot-a-2823			12	0.476	0.167	0.004	1.609	1.160	2.233
prot-a-1176			13	−0.434	0.195	0.026	0.648	0.442	0.950
prot-a-666			11	0.407	0.198	0.040	1.502	1.018	2.217
prot-a-1424			11	0.661	0.224	0.003	1.937	1.248	3.006
prot-a-875			8	0.568	0.256	0.026	1.765	1.069	2.913
prot-a-724			16	−0.313	0.152	0.040	0.731	0.542	0.986
prot-a-1975			18	0.339	0.152	0.026	1.404	1.041	1.892
prot-a-2026			6	−0.561	0.246	0.023	0.570	0.352	0.925
prot-a-1767			11	−0.529	0.256	0.039	0.589	0.357	0.973
prot-a-2566			16	0.411	0.167	0.014	1.508	1.087	2.091
prot-a-932			23	0.380	0.127	0.003	1.463	1.140	1.876
prot-a-1976			13	0.628	0.279	0.024	1.874	1.085	3.237
prot-a-721			9	−0.582	0.246	0.018	0.559	0.345	0.904
prot-a-2326			13	−0.509	0.197	0.010	0.601	0.408	0.885
prot-a-1572			14	−0.402	0.200	0.045	0.669	0.451	0.991
prot-a-1973			7	−0.641	0.321	0.046	0.527	0.281	0.989
prot-a-470			18	−0.408	0.204	0.046	0.665	0.446	0.992
prot-a-1894			8	0.674	0.306	0.028	1.961	1.076	3.573
prot-a-1662			16	0.421	0.174	0.015	1.523	1.084	2.141

Table 2 The List of Key Plasma Proteins

id.exposure	Corresponding Protein	Corresponding Gene
prot-a-1023	Protein FAM134B	FAM134B
prot-a-1039	Protein FAM189A2	FAM189A2
prot-a-1052	MIP18 family protein FAM96A	FAM96A
prot-a-1079	Fc receptor-like protein 1	FCRL1
prot-a-1093	Fibroblast growth factor 23	FGF23
prot-a-1176	Glyceraldehyde-3-phosphate dehydrogenase, testis-specific	GAPDHS
prot-a-1200	Glial fibrillary acidic protein	GFAP
prot-a-1218	Lactoylglutathione lyase	GLO1
prot-a-1224	Glucocorticoid modulatory element-binding protein 2	GMEB2
prot-a-1231	Glucosamine 6-phosphate N-acetyltransferase	GNPNAT1
prot-a-1245	Glypican-1	GPC1
prot-a-1261	Protein GPR107	GPR107
prot-a-1275	Glutamate receptor 4	GRIA4
prot-a-1309	Histone acetyltransferase type B catalytic subunit	HAT1
prot-a-1325	Protein HEG homolog 1	HEG1
prot-a-1359	Hepatocyte nuclear factor 1-alpha	HNF1A
prot-a-140	DCC-interacting protein 13-alpha	APPL1
prot-a-1424	Interferon alpha-7	IFNA7
prot-a-149	Rho GTPase-activating protein 30	ARHGAP30
prot-a-1504	Interleukin-1 receptor antagonist protein	IL1RN
prot-a-1536	Interleukin-5 receptor subunit alpha	IL5RA
prot-a-1537	Interleukin-5 receptor subunit alpha	IL5RA
prot-a-157	Rho guanine nucleotide exchange factor 25	ARHGEF25
prot-a-1572	Iron-sulfur cluster assembly enzyme ISCU, mitochondrial	ISCU
prot-a-1662	Kallikrein-4	KLK4
prot-a-1682	Keratin, type I cytoskeletal 17	KRT17
prot-a-1693	Lymphocyte activation gene 3 protein	LAG3
prot-a-1762	Lysyl oxidase homolog 2	LOXL2
prot-a-1767	Leucine-rich repeat and calponin homology domain-containing protein 4	LRCH4
prot-a-1774	Lymphoid-restricted membrane protein	LRMP
prot-a-1777	Low-density lipoprotein receptor-related protein 1, soluble	LRP1
prot-a-1785	Leucine-rich repeat-containing protein 19	LRRC19
prot-a-1839	MANSC domain-containing protein 4	MANSC4
prot-a-1842	Dual specificity mitogen-activated protein kinase kinase 4	MAP2K4
prot-a-1862	Methyl-CpG-binding domain protein 4	MBD4
prot-a-1894	Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase B	MGAT4B
prot-a-1898	MHC class I polypeptide-related sequence B	MICB
prot-a-1915	Matrix metalloproteinase-16	MMP16
prot-a-1942	39S ribosomal protein L33, mitochondrial	MRPL33
prot-a-1973	MAX-interacting protein 1	MXI1
prot-a-1975	Matrix-remodeling-associated protein 8	MXRA8
prot-a-1976	Matrix-remodeling-associated protein 8	MXRA8
prot-a-1979	Myc target protein 1	MYCT1
prot-a-1987	Myeloid zinc finger 1	MZF1
prot-a-2025	NADH dehydrogenase [ubiquinone] iron-sulfur protein 4, mitochondrial	NDUFS4
prot-a-2026	NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial	NDUFV2
prot-a-2030	Protein kinase C-binding protein NELL1	NELL1
prot-a-2038	Neuralized-like protein 4	NEURL4
prot-a-2071	Epididymal secretory protein E1	NPC2
prot-a-2109	Vesicle-fusing ATPase	NSF
prot-a-2148	Opalin	OPALIN

(*Continued*)

Table 2 (Continued).

id.exposure	Corresponding Protein	Corresponding Gene
prot-a-2277	Pirin	PIR
prot-a-2283	Group 10 secretory phospholipase A2	PLA2G10
prot-a-2326	Corticotropin	CRHBP
prot-a-2396	Myeloblastin	PRTN3
prot-a-2401	Pregnancy-specific beta-1-glycoprotein 2	PSG2
prot-a-2452	Poly(U)-binding-splicing factor PUF60	PUF60
prot-a-2522	Replication initiator 1	REPIN1
prot-a-2566	E3 ubiquitin-protein ligase RNF149	RNF149
prot-a-2588	Dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit 1	RPN1
prot-a-2658	Signal peptidase complex catalytic subunit SEC11C	SEC11C
prot-a-2692	Kallistatin	SERPINA4
prot-a-2704	Histone-lysine N-methyltransferase SETMAR	SETMAR
prot-a-2746	Urea transporter 1	SLC14A1
prot-a-2749	Mitochondrial glutamate carrier 2	SLC25A18
prot-a-2823	Kunitz-type protease inhibitor 1	SPINT1
prot-a-2859	Sia-alpha-2,3-Gal-beta-1,4-GlcNAc-R:alpha 2,8-sialyltransferase	ST8SIA3
prot-a-2915	Synaptotagmin-5	SYT5
prot-a-2929	Tax1-binding protein 3	TAX1BP3
prot-a-2946	Teneurin-4	TENM4
prot-a-2982	Toll/interleukin-1 receptor domain-containing adapter protein	TIRAP
prot-a-3027	Transmembrane and ubiquitin-like domain-containing protein 2	TMUB2
prot-a-3031	Tumor necrosis factor-inducible gene 6 protein	TNFAIP6
prot-a-3109	Thiosulfate sulfurtransferase	TSTD3
prot-a-3117	Tuftelin	TUFT1
prot-a-3122	Thioredoxin domain-containing protein 11	TXNDC11
prot-a-315	Complement C4	C4A
prot-a-3157	UDP-glucuronosyltransferase 1-8	UGT1A8
prot-a-3275	Zinc finger protein 75D	ZNF75D
prot-a-355	Calpain-3	CAPN3
prot-a-395	C-C motif chemokine 19	CCL19
prot-a-447	CD40 ligand	CD40LG
prot-a-454	Lymphocyte function-associated antigen 3	CD58
prot-a-459	HLA class II histocompatibility antigen gamma chain	CD74
prot-a-470	T-cell surface protein tactile	CD96
prot-a-503	Carcinoembryonic antigen-related cell adhesion molecule 21	CEACAM21
prot-a-523	Complement factor I	CFI
prot-a-541	Chitinase-3-like protein 2	CHI3L2
prot-a-557	Carbohydrate sulfotransferase 5	CHST5
prot-a-626	Collagen alpha-2(IX) chain	COL9A2
prot-a-642	Carboxypeptidase A2	CPA2
prot-a-650	Cellular retinoic acid-binding protein 1	CRABP1
prot-a-654	Cyclic AMP-responsive element-binding protein 3-like protein 4	CREB3L4
prot-a-666	Cytokine receptor-like factor 1:Cardiotrophin-like cytokine factor 1 Complex	CRLF1
prot-a-721	Cathepsin F	CTSF
prot-a-724	Cathepsin H	CTSH
prot-a-805	Protein DEPP	DEPP1
prot-a-835	DnaJ homolog subfamily A member 4	DNAJA4
prot-a-875	Dual specificity protein phosphatase 16	DUSP16
prot-a-888	Tumor necrosis factor receptor superfamily member EDAR	EDAR
prot-a-91	Anaphase-promoting complex subunit 7	ANAPC7

(*Continued*)

Table 2 (Continued).

id.exposure	Corresponding Protein	Corresponding Gene
prot-a-911	EH domain-binding protein 1	EHBP1L1
prot-a-932	RNA polymerase II elongation factor ELL	ELL2
prot-a-951	Ectonucleoside triphosphate diphosphohydrolase 1	ENTPD1
prot-a-971	Endoplasmic reticulum aminopeptidase 1	ERAP1
prot-a-991	Estrogen receptor	ESR1

of scRNA-seq data, 53,111 cells and 22,445 genes were retained for further analyses (Figure S6A and B). Variations in these genes were depicted in Figure S6C, highlighting the 2000 highly variable genes. PCA indicated that the cells from different samples were mixed, with no batch effect present (Figure S6D). After selecting the top 30 principal components for UMAP downscaling according to the Elbow Plot and PCA replacement test ($p < 0.05$), 15 distinct cell clusters were determined (Figure 3A–C). Annotation of cell clusters to 12 types was further refined using marker genes (Table S5), including epithelial cells, fibroblasts, keratinocytes, smooth muscle cells, endothelial cells, innate lymphoid cells, macrophages, eccrine cells, Schwann cells, melanocytes, mast cells, and myeloid-like cells (Figure 3D). The SSc group showed a similar cell clustering pattern to the control group, but there were differences in the distribution of specific cell types (such as fibroblasts) (Figure 3E). Marker gene expression across cell types was illustrated in Figure 3F. Remarkably, epithelial cells and fibroblasts were identified as key cell types, which exhibited significant differential expression of specific key plasma proteins (genes) between SSc and control groups ($p < 0.05$). In detail, the epithelial cells in SSc group exhibited 1000 DEGs, including 634 upregulated and 366 downregulated ($p < 0.05$) (Figure 3G). Similarly, there were 998 DEGs within fibroblasts in the SSc group, including 873 upregulated and 125 downregulated genes ($p < 0.05$) (Figure 3H). Importantly, the union of these DEGs was intersected with 104 genes encoding key plasma proteins, resulting in 8 candidate biomarkers, including ANAPC7, CCL19, CTSH, ESR1, GPC1, KRT17, LOXL2, and MGAT4B (Figures 3I and S1).

Crucial Functional Pathways of Key Cell Types and Candidate Biomarkers

Within GSE138669 dataset, the functional roles of the identified key cell types were further explored. It was found that epithelial cells and fibroblasts were associated with biological pathways such as "sterols are 12-hydroxylated by CYP8B1" and "Biogenic amines are oxidatively deaminated to aldehydes by MAOA and MAOB", highlighting the pivotal roles of these cell types in SSc (Figure 4A).

Moreover, the relevant pathways of candidate biomarkers were also examined. ANAPC7, CCL19, CTSH, ESR1, GPC1, KRT17, LOXL2, and MGAT4B were enriched in 32, 41, 59, 54, 58, 59, 49, and 92 distinct pathways, respectively (adjusted $p < 0.05$). It was noted that most of the candidate biomarkers were associated with "proteasome", "phosphatidylinositol signaling system", "glycosaminoglycan degradation", and "galactose metabolism". The top 5 significant pathways were illustrated in Figure 4B–I. These findings suggested that candidate biomarkers might contribute to SSc progression by modulating these pathways.

Identification and Expression Profiles of Biomarkers

Candidate biomarkers were subjected to further screening within the GSE138669 and GSE181549 datasets. It was found that among candidate biomarkers, CCL19 and LOXL2 were consistently and significantly upregulated in the SSc group across both datasets ($p < 0.001$) (Figure 5A and B). This significant differential expression confirmed their diagnostic value. Therefore, CCL19 and LOXL2 were ultimately selected as biomarkers for SSc.

Within the GSE138669 dataset, CCL19 and LOXL2 were both distributed in key cell types (epithelial cells and fibroblasts) (Figure 5C and D). Furthermore, CCL19 expression in both key cell types and LOXL2 expression in fibroblasts exhibited significant differences between SSc and control groups ($p < 0.01$) (Figure 5E and F).

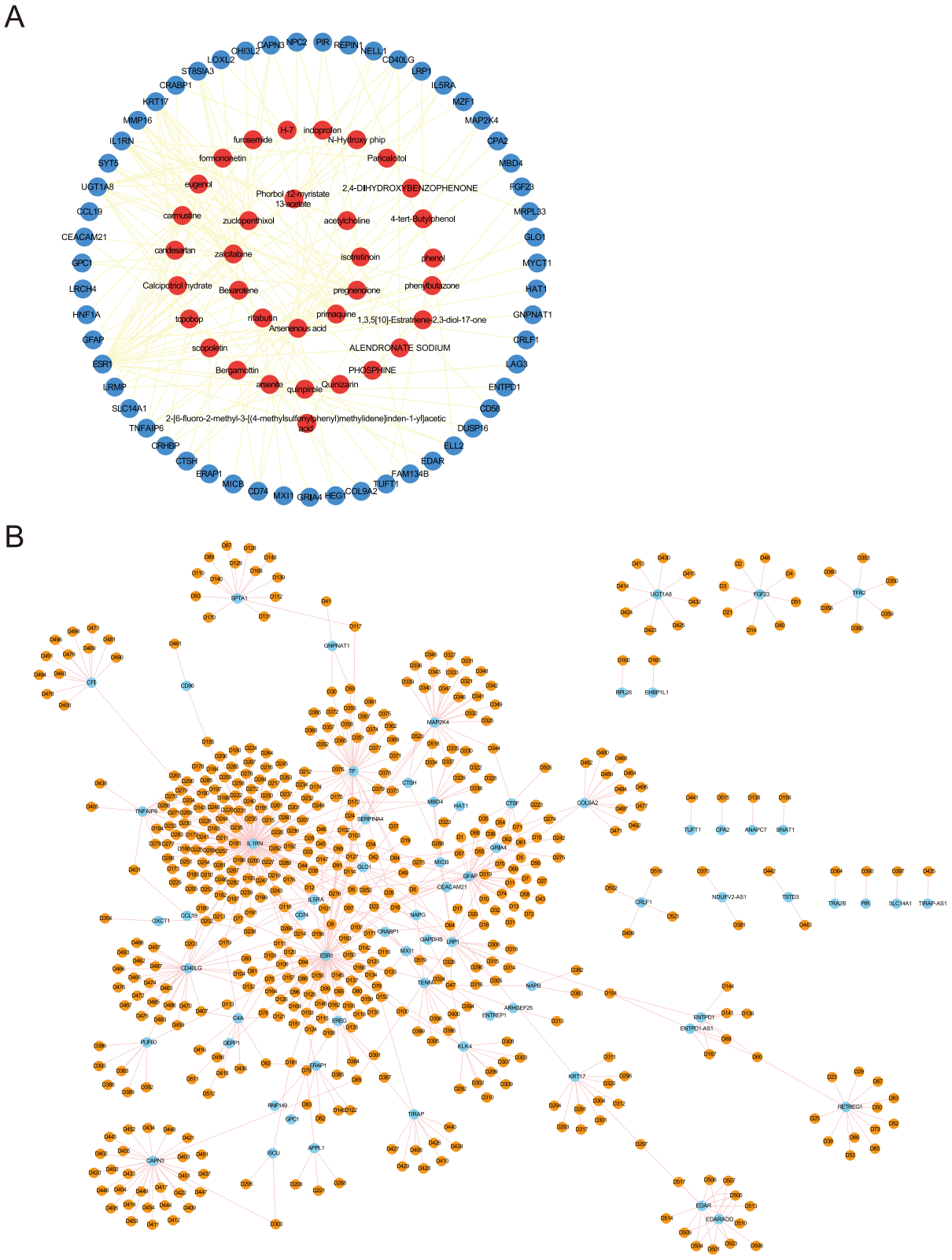

Figure 1 Network analysis of key plasma proteins (genes) with drugs and diseases. (**A**) Interaction network between key plasma proteins (genes) and potential therapeutic drugs. (**B**) Network showing the associations between key plasma proteins (genes) and diseases.

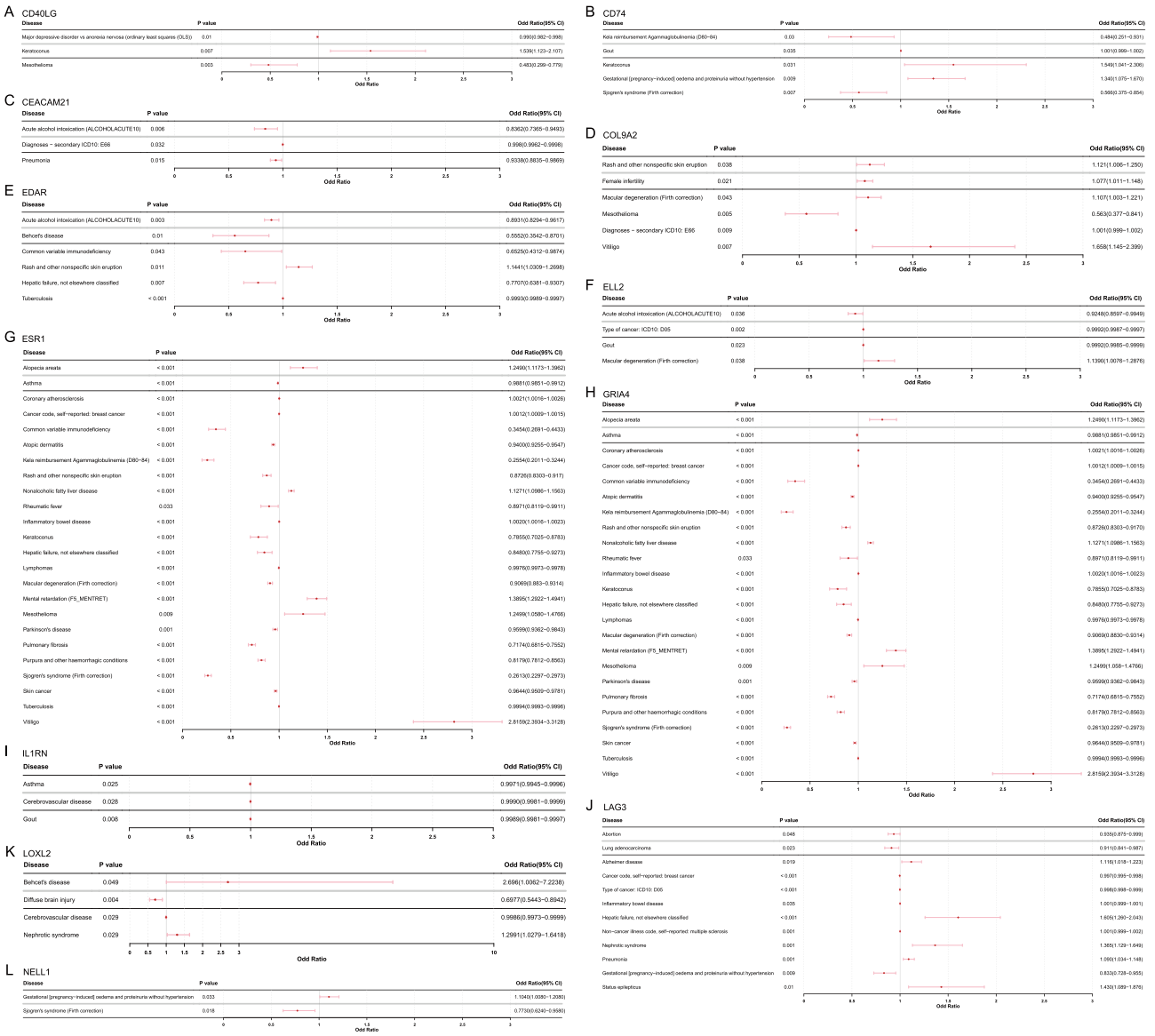

Figure 2 Causal relationships between key plasma proteins and diseases. Through causal inference analysis, the following key plasma proteins may exhibit potential adverse effects when targeted therapeutically: CD40LG (**A**), CD74 (**B**), CEACAM21 (**C**), COL9A2 (**D**), EDAR (**E**), ELL2 (**F**), ESR1 (**G**), GRIA4 (**H**), IL1RN (**I**), LAG3 (**J**), LOXL2 (**K**), and NELL1 (**L**).

Abbreviations: CD40LG, CD40 ligand; CEACAM21, carcinoembryonic antigen-related cell adhesion molecule 21; COL9A2, collagen type IX alpha 2 chain; EDAR, ectodysplasin A receptor; ELL2, elongation factor for RNA polymerase II 2; ESR1, estrogen receptor 1; GRIA4, glutamate ionotropic receptor AMPA type subunit 4; IL1RN, interleukin 1 receptor antagonist; LAG3, lymphocyte activating 3; LOXL2, lysyl oxidase-like 2; NELL1, neural EGFL like 1.

The expression levels of biomarkers in the skin of SSc patients and two SSc mouse models were further validated. In the skin tissues of SSc patients and healthy controls, IHC showed that CCL19 was primarily expressed in certain keratinocytes and stromal cells within the dermis, while LOXL2 was mainly distributed in the connective tissue of the dermal layer (Figure 5G). In mouse skin tissues, IHC revealed that the expression patterns of CCL19 and LOXL2 were largely consistent with those observed in human skin tissues (Figure 5G). The proportion of positive staining areas for CCL19 and LOXL2 was significantly higher in the skin of SSc patients compared to the control group (Figure 5H). Similarly, the mRNA expression levels of CCL19 and LOXL2 were elevated in the skin of mice induced by bleomycin and HClO compared to the NC group (Figures 5I and S1).

https://doi.org/10.2147/JIR.S561740

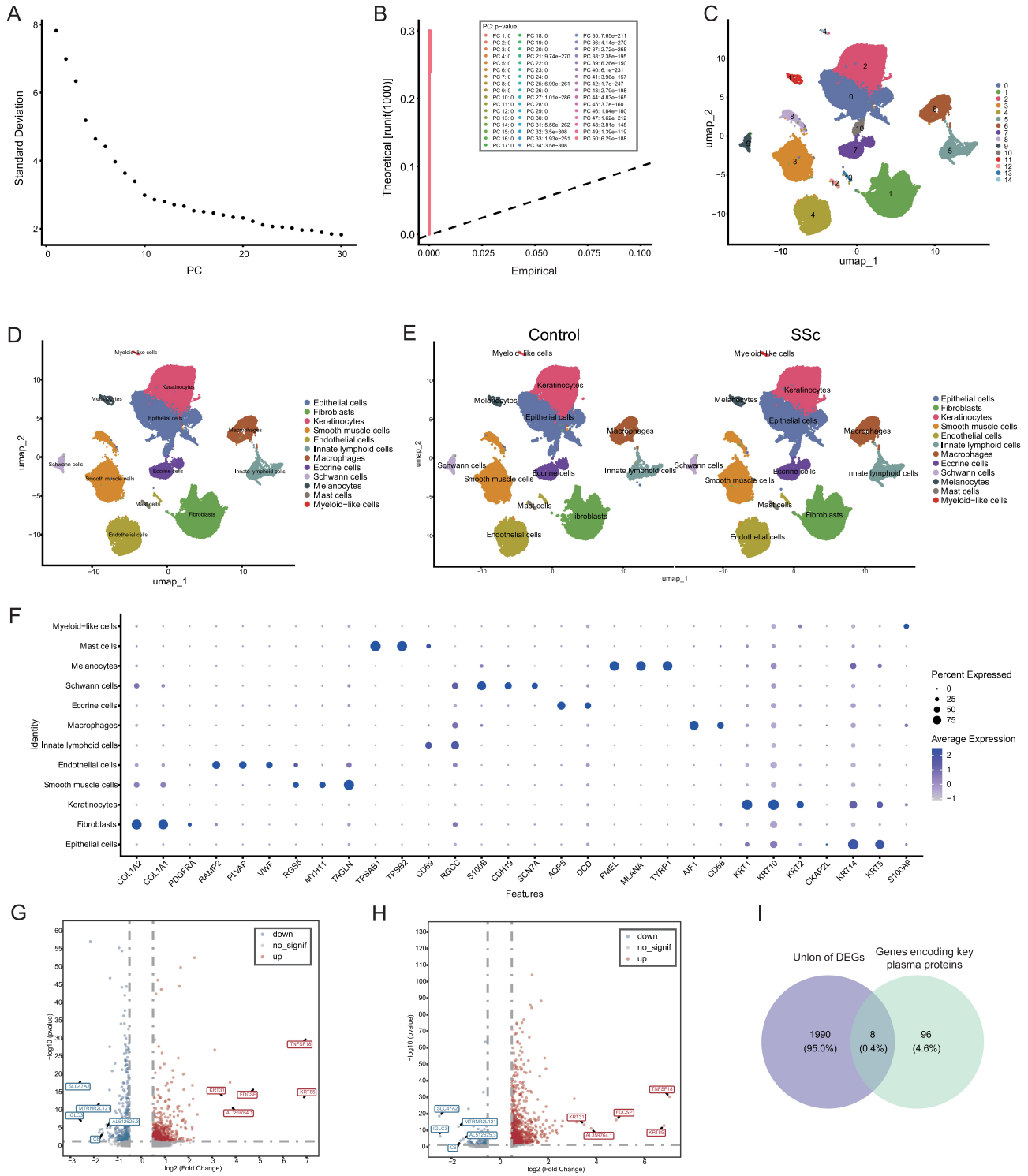

Figure 3 Single-cell RNA sequencing analysis of SSc skin tissue and identification of candidate biomarkers. (**A**) Elbow plot demonstrating the selection of 30 PCs for downstream analysis. (**B**) PCA Replacement Test evaluating the significance of each PC. (**C**) Dimensionality reduction and clustering using the top 30 PCs, visualized via UMAP, revealing 15 cell clusters. (**D**) Identification of 12 distinct cell types based on cluster-specific markers. (**E**) Clustering analysis stratified by control and SSc groups. (**F**) Marker gene expression levels visualized across clusters. (**G** and **H**) DEGs identified in epithelial cells (**G**) and fibroblasts (**H**). (**I**) Intersection of DEGs from epithelial cells and fibroblasts overlapped with plasma protein-associated genes to identify eight candidate biomarkers.

Abbreviations: PCA, principal component analysis; PC, principal component; UMAP, Uniform Manifold Approximation and Projection; DEGs, differentially expressed genes; SSc, systemic sclerosis.

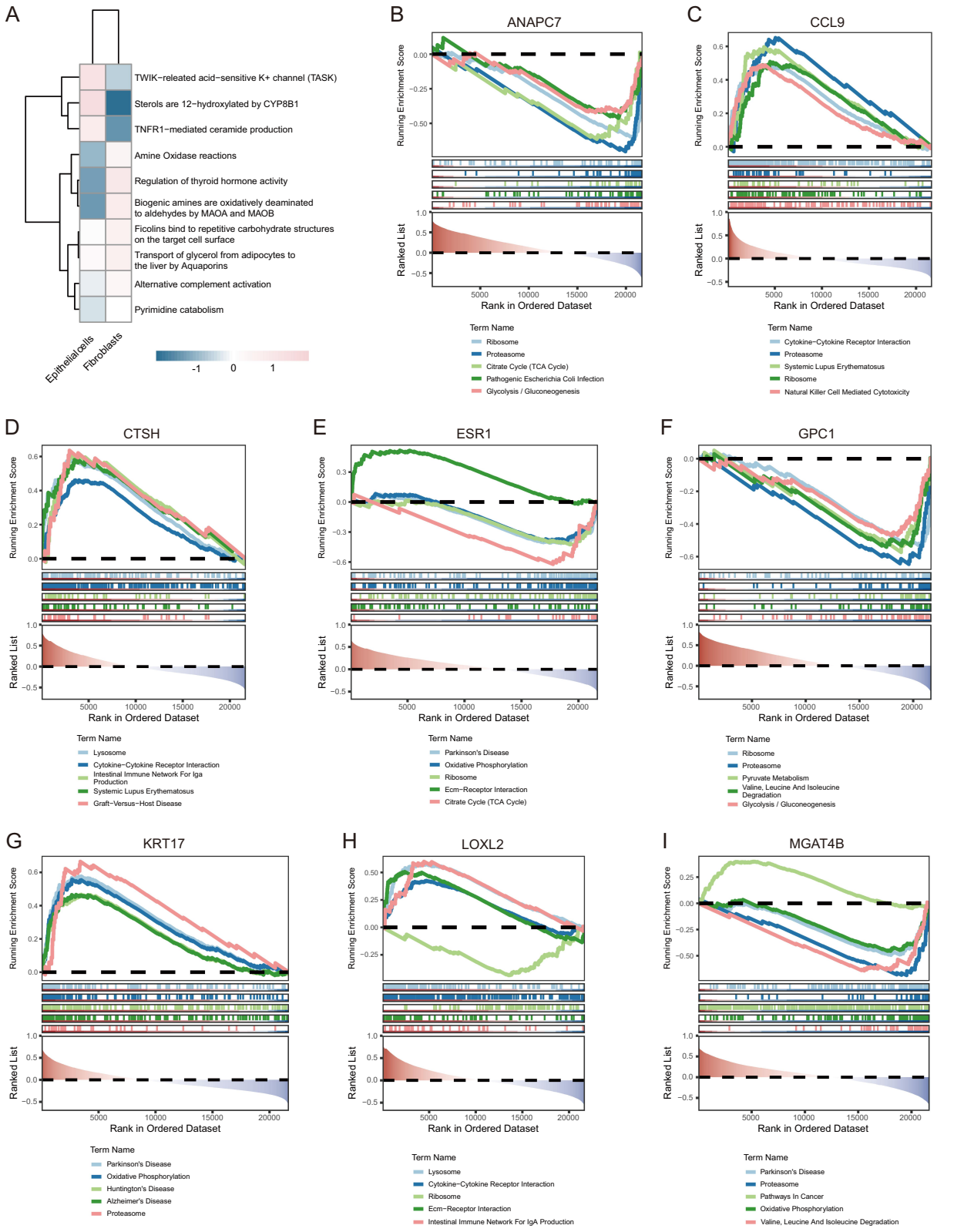

Figure 4 Functional pathway analysis of key cell types and candidate biomarkers. (**A**) Comparison of the top 10 significantly enriched biological pathways in epithelial cells and fibroblasts between SSc patients and controls. (**B–I**) Top five enriched pathways from single-gene GSEA analysis for ANAPC7 (**B**), CCL19 (**C**), CTSH (**D**), ESR1 (**E**), GPC1 (**F**), KRT17 (**G**), LOXL2 (**H**), and MGAT4B (**I**).

Abbreviations: GSEA, gene set enrichment analysis; SSc, systemic sclerosis; ANAPC7: Anaphase Promoting Complex Subunit 7; CCL19: C-C Motif Chemokine Ligand 19; CTSH: Cathepsin H; ESR1: Estrogen Receptor 1; GPC1: Glypican 1; KRT17: Keratin 17; LOXL2: Lysyl Oxidase Like 2; MGAT4B: Alpha-1,3-Mannosyl-Glycoprotein 4-Beta-N-Acetylglucosaminyltransferase B.

https://doi.org/10.2147/JIR.S561740

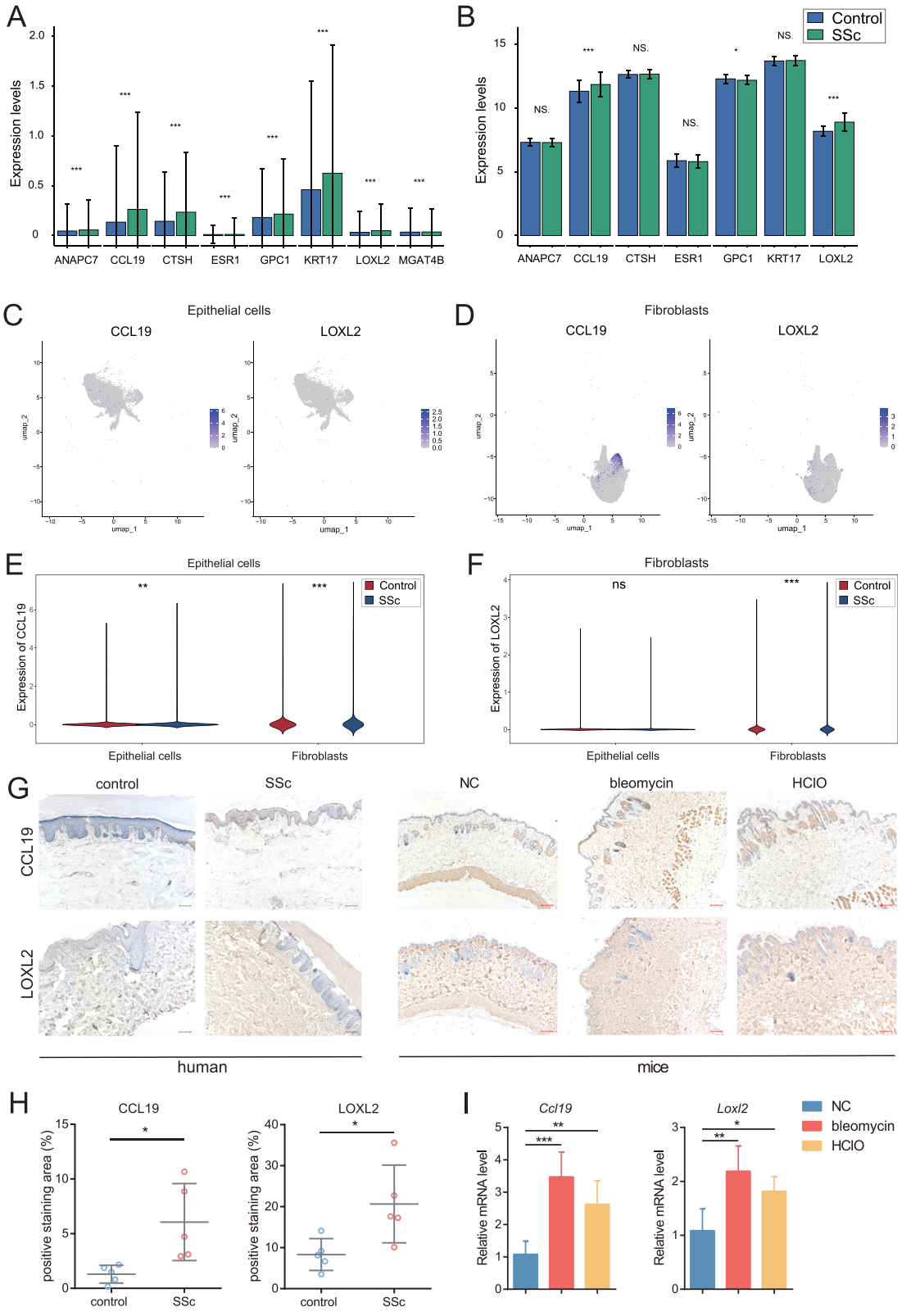

Figure 5 Biomarker validation and expression profiling. (**A** and **B**) Expression levels of eight candidate biomarkers in GSE138669 (**A**) and GSE181549 (**B**) datasets. (**C** and **D**) Distribution of CCL19 and LOXL2 in epithelial cells (**C**) and fibroblasts (**D**). (**E** and **F**) Differential expression analysis of CCL19 and LOXL2 in epithelial cells (**E**) and fibroblasts (**F**). (**G**) IHC for CCL19 and LOXL2 expression in SSc skin tissues and bleomycin- and HClO-induced murine SSc models (n = 5 per group, bar = 100μm). (**H**) Semi-quantitative analysis of IHC results from SSc and control skin tissues. (**I**) The relative mRNA level of *Ccl19* and *Loxl2* in skin tissue from NC, bleomycin-, and HClO-induced mice models (n = 5 in each group). (**P* < 0.05, ***P* < 0.01, ****P* < 0.001).

Abbreviations: IHC, immunohistochemistry; SSc, systemic sclerosis; CCL19: C-C Motif Chemokine Ligand 19; LOXL2: Lysyl Oxidase Like 2.

Elaborate Molecular Regulatory Networks of Biomarkers

Upstream regulatory elements of biomarkers were predicted. The constructed TF-mRNA and miRNA-mRNA regulatory network comprised 13 TFs and 74 miRNAs targeting specific biomarkers (Figure 6A and B). Specifically, CCL19 was targeted by 11 TFs and 14 miRNAs. Additionally, LOXL2 was targeted by 3 TFs and 62 miRNAs. It was worth mentioning that CCL19 and LOXL2 were co-regulated by 1 TF (FOXL1) and 2 miRNAs (hsa-miR-5001-5p and hsa-miR-4498). Identifying these regulatory elements was crucial for elucidating the mechanisms underlying the biomarkers' roles in SSc.

Cellular Heterogeneity and Differentiation Trajectories of Epithelial Cells and Fibroblasts

After UMAP analysis, the epithelial cells and fibroblasts were further clustered into 8 and 4 subclusters, respectively (Figure 7A and B). Using marker genes (Tables S6 and 7), they were assigned to different cell subtypes (Figure 7C and D). In detail, epithelial cells were annotated into 5 subtypes, including basal epithelial cells (BE), other epithelial cells (OE), collecting duct principal cells (CD), luminal epithelial cells (LE), and type B intercalated cell (TB). Additionally, fibroblasts were annotated into 4 subtypes, including inflammatory fibroblasts, matrix metalloproteinase (MMP) fibroblasts, stress fibroblasts, and collagen fibroblasts. The marker gene expression across different cell subtypes was illustrated in Figure 7E and F. Collectively, these findings emphasized the heterogeneity of epithelial cells and fibroblasts, which provided critical insights for understanding their roles in SSc.

Subsequently, the differentiation trajectories of epithelial cells and fibroblasts over time were tracked and presented in Figure 8A and B, with darker colors representing earlier stages of differentiation. Epithelial cells and fibroblasts displayed 3 and 11 distinct differentiation states, respectively (Figure 8C and D). Most epithelial cells were predominantly in the BE differentiation subtype, while fibroblasts were mainly found in the inflammatory fibroblast and MMP fibroblast subtypes (Figure 8E and F). In both SSc and control groups, epithelial cells were primarily in the early to mid-differentiation stages, while fibroblasts were mostly in the late-differentiation stages (Figure 8G and H). Remarkably, the expression patterns of biomarkers during key cell type differentiation were also explored. It was found that the expression of CCL19 slightly increased during the mid-differentiation stage of epithelial cells, before returning to baseline levels. The expression of LOXL2 did not change during the differentiation of epithelial cells (Figure 8I). Additionally, with

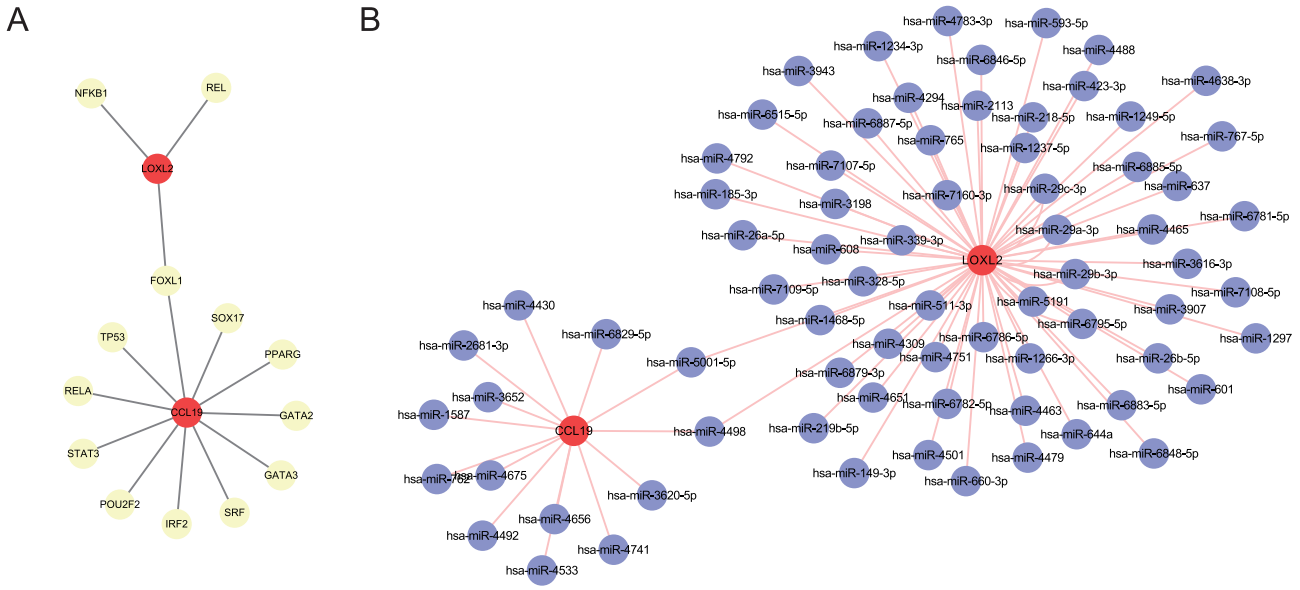

Figure 6 Regulatory networks associated with CCL19 and LOXL2. (**A**) Transcription factor network analysis of CCL19 and LOXL2. (**B**) miRNA-target interaction network of CCL19 and LOXL2.
Abbreviations: CCL19: C-C Motif Chemokine Ligand 19; LOXL2: Lysyl Oxidase Like 2; miRNA, microRNA.

https://doi.org/10.2147/JIR.S561740

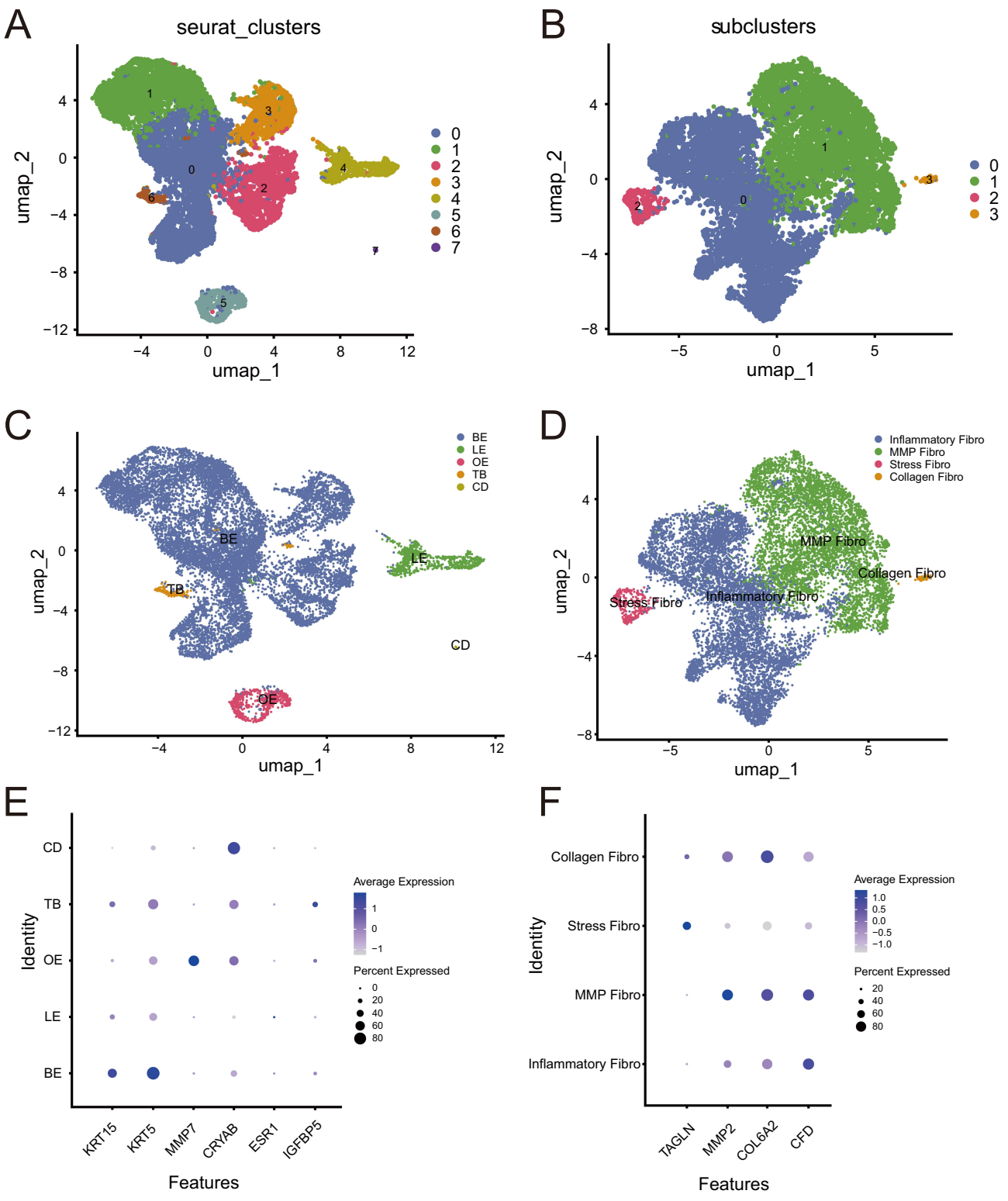

Figure 7 Cellular heterogeneity analysis of epithelial cells and fibroblasts. (**A** and **B**) UMAP plots showing further clustering of epithelial cells (**A**) and fibroblasts (**B**). (**C** and **D**) Annotation of cell subtypes based on canonical markers for epithelial cells (**C**) and fibroblasts (**D**). (**E** and **F**) Visualization of marker gene expression levels corresponding to (**C** and **D**), respectively.

Abbreviation: UMAP, Uniform Manifold Approximation and Projection.

fibroblast differentiation, CCL19 expression generally showed a decreasing trend, with a slight increase during the late differentiation stage, while LOXL2 expression remained largely unchanged (Figure 8J). In summary, the key cell type differentiation stages and the expression patterns of biomarkers across these stages might be associated with SSc progression.

Discussion

In recent years, the search for novel biomarker and therapeutic targets has been a major focus in SSc. Multiple cytokines and chemokines have been recognized as potential biomarkers or therapeutic targets for SSc.[39,40] However, most of these targets have been identified through correlation-based studies, which do not establish causality. In contrast, this study employs a proteome-wide MR analysis using large-scale pQTL data, allowing for more precise identification of key pathogenic proteins. A total of 106 key plasma proteins were identified through MR, while 1998 differentially expressed genes (DEGs) in fibroblasts and epithelial cells were identified from scRNA-seq datasets. By intersecting the key plasma proteins with the DEGs, we identified eight candidate proteins (genes): MGAT4B, KRT17, ANAPC7, GPC1, CTSH,

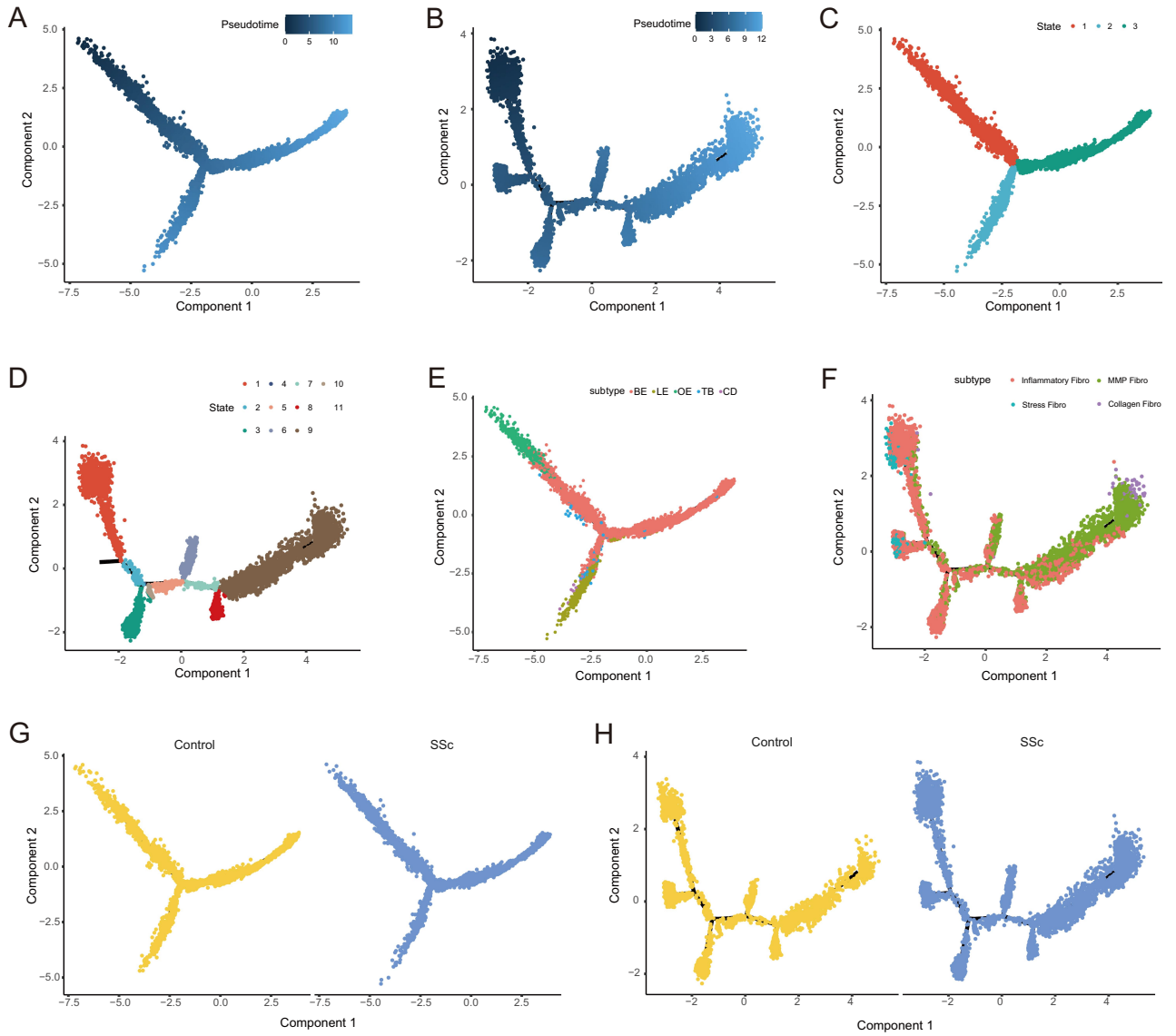

Figure 8 Continued.

https://doi.org/10.2147/JIR.S561740

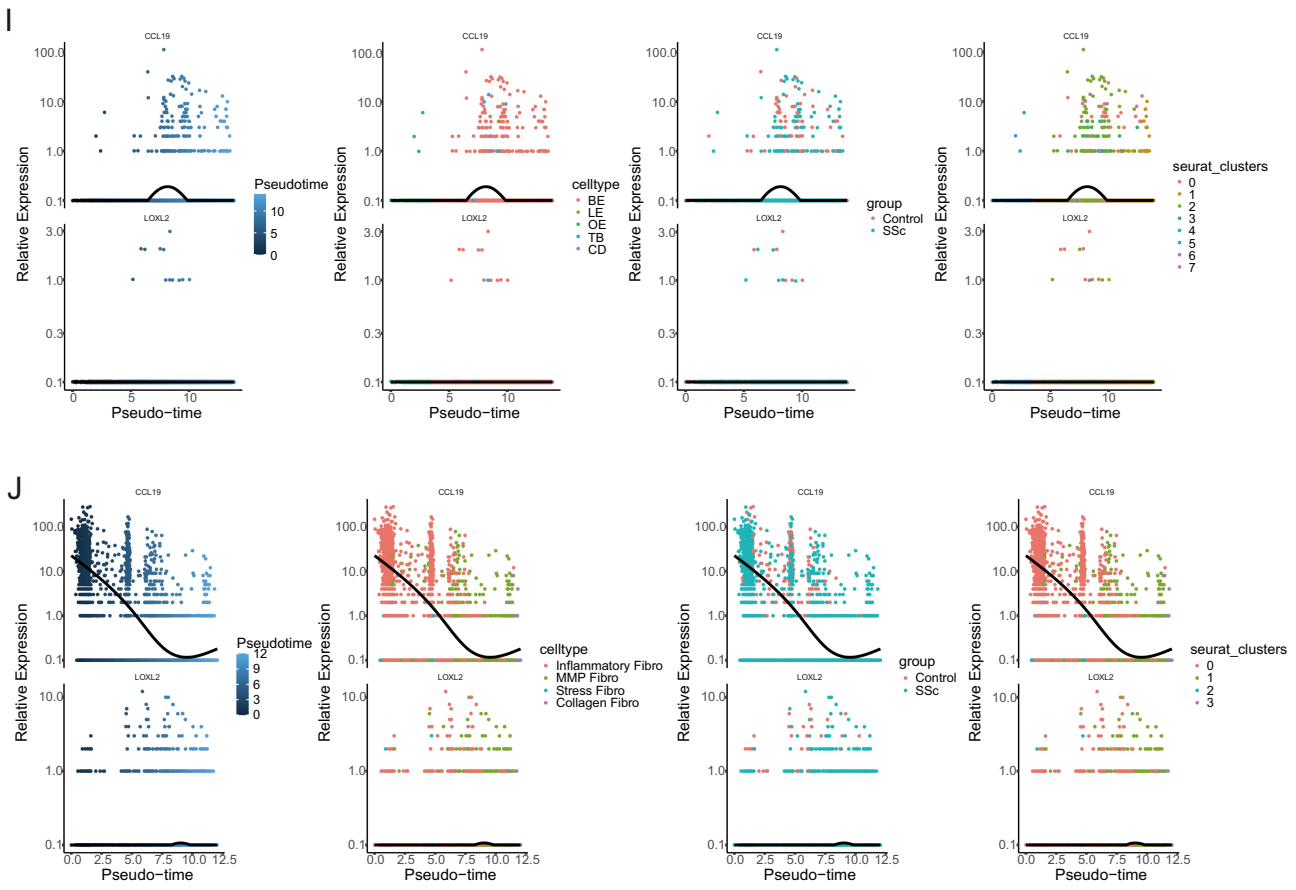

Figure 8 Differentiation trajectories and dynamic expression of biomarkers in epithelial cells and fibroblasts. (**A** and **B**) Pseudotime trajectory analysis reveals differentiation paths of epithelial cells (**A**) and fibroblasts (**B**). (**C** and **D**) Distinct differentiation states in epithelial cells (**C**) and fibroblasts (**D**). (**E** and **F**) Temporal progression of cell subclusters in epithelial cells (**E**) and fibroblasts (**F**). (**G** and **H**) Pseudotime comparisons between SSc and control samples in epithelial cells (**G**) and fibroblasts (**H**). (**I** and **J**) Dynamic expression of CCL19 and LOXL2 across differentiation states of epithelial cells (**I**) and fibroblasts (**J**).
Abbreviations: SSc, systemic sclerosis; CCL19: C-C Motif Chemokine Ligand 19; LOXL2: Lysyl Oxidase Like 2.

LOXL2, CCL19, and ESR1. Subsequent validation in both scRNA-seq datasets and bulk RNA-seq transcriptomic datasets confirmed that only CCL19 and LOXL2 were consistently upregulated. Furthermore, IHC and qRT-PCR confirmed the elevated expression of CCL19 and LOXL2 in the skin tissues of SSc patients and SSc model mice. Additionally, we analyzed drugs associated with the 106 key plasma proteins, explored the diseases linked to these drugs, and examined the direct disease associations of these plasma proteins. The expression dynamics of *CCL19* and *LOXL2* in fibroblasts and epithelial cells were also examined using pseudotime trajectory analysis. In summary, CCL19 and LOXL2 were ultimately identified as potential biomarkers and therapeutic targets for SSc.

CCL19 is a crucial chemokine highly expressed in stromal and dendritic cells within lymphoid organs, with moderate expression in various other tissues.[41] Its primary functional receptor, CCR7, is predominantly found on T cells, B cells, and lymphoid tissues.[41] Consequently, CCL19 plays a key role in coordinating immune cell homing and localization, contributing to the initiation of inflammatory responses and the maintenance of immune tolerance, thereby preserving immune system homeostasis and functionality.[42] CCL19 is one of the primary chemokines for the recruitment of macrophages and T cells,[43] both of which have been identified as key players in SSc pathogenesis.[44,45] Notably, recent studies have demonstrated that $CCL19^{+}$ fibroblasts play a crucial role in forming tertiary lymphoid structures (TLS) in tumors,[46] hidradenitis suppurativa,[47] and oral lichen planus,[48] emphasizing their importance in local immune microenvironments. Currently, research on CCL19 in SSc is limited. *Mathes A.L.*[43] reported elevated CCL19 expression in SSc skin tissues, which was associated with markers of vascular inflammation and macrophage recruitment. Immunofluorescence analysis revealed colocalization of CCL19 with the macrophage marker CD163 in SSc skin

tissues,[43] suggesting a role for CCL19 in the chemotaxis and localization of macrophages. Additionally, two studies have explored the relationship between CCL19 and interferon (IFN) in SSc. It was observed that SSc patients with higher serum IFN-alpha-2a levels also exhibited increased serum CCL19 levels.[49] Moreover, as an IFN-stimulated gene, CCL19 has been identified as a predictor of clinical progression in limited cutaneous SSc.[50]

In this study, CCL19 was highlighted due to its elevated expression in SSc skin fibroblasts and epithelial cells. However, as previously mentioned, these two cell types are not considered the primary theoretical sources of CCL19. Therefore, we cannot confirm whether the elevated circulating CCL19 originates from the secretion of skin fibroblasts and epithelial cells rather than from other lymphoid organs in SSc patients. Nevertheless, this does not diminish their significance in SSc pathogenesis. Our findings indicate an upregulation of CCL19 in SSc skin fibroblasts, suggesting the presence of CCL19+ fibroblasts. However, the role of these fibroblasts in SSc skin tissues, particularly their potential involvement in TLS formation, remains unexplored. Currently, no studies have specifically investigated the function of $CCL19^{+}$ epithelial cells in SSc. One study reported that fibroblasts from SSc patients activate the IFN pathway in keratinocytes (the primary epithelial cells in skin tissue) via exosomes, leading to the upregulation of multiple interferon-stimulated genes. Interestingly, keratinocytes alone were found to be unresponsive to IFN-α stimulation. Although this study did not assess CCL19 expression in keratinocytes, given that CCL19 is an interferon-stimulated gene and is highly relevant to SSc, we speculate that $CCL19^{+}$ epithelial cells may be induced by fibroblast-mediated regulation. Further research is needed to elucidate the specific pathogenic mechanisms of $CCL19^{+}$ fibroblasts and epithelial cells in SSc and to explore the therapeutic potential of targeting CCL19 in SSc treatment.

LOXL2 is a member of the LOX family and functions as a secreted, copper-dependent amine oxidase.[51] Fibrosis is characterized by the pathological accumulation of ECM, ultimately leading to tissue stiffening and structural and functional impairment. The primary biological role of the LOX family is to facilitate covalent cross-linking between ECM components, particularly collagen and elastin, thereby reducing ECM degradation and ultimately contributing to fibrosis.[51] Additionally, LOXL2 has been implicated in fibrosis through its ability to activate fibroblasts.[52] Research on LOXL2 in fibrosis has been most extensive in liver and lung fibrosis. Early studies suggested that LOXL2 could be a promising therapeutic target for fibrosis reversal in the liver.[51] Consequently, a humanized monoclonal antibody targeting LOXL2 (SIMTUZUMAB, developed by Gilead Sciences SA) was tested in multiple clinical trials for liver and lung fibrosis. However, these trials ultimately failed.[51,52] The failure of SIMTUZUMAB was not attributed to the LOXL2 target itself but rather to the insufficient activity of the antibody.[51,52] Moreover, solely inhibiting LOXL2 may lead to compensatory upregulation of other LOX family members, potentially affecting therapeutic efficacy. For example, LOXL2 knockdown in lung tissue has been shown to result in increased LOX expression. Furthermore, LOXL2 is primarily localized in dense connective tissues, and there is no clear evidence that monoclonal antibody therapies can efficiently target and bind to LOXL2 within these tissues.[52] To overcome these limitations, novel small-molecule inhibitors targeting LOXL2 and the LOX family have been explored in recent years, demonstrating anti-fibrotic effects in preclinical models of liver and lung fibrosis.[52] Studies on LOXL2 in SSc are also limited. It has been reported that TGF-β1 and insulin-like growth factor-II can upregulate LOXL2 expression in dermal fibroblasts.[53,54] Moreover, the use of a pan-lysyl oxidase inhibitor has been shown to ameliorate skin and lung fibrosis in an SSc model,[55] suggesting that LOXL2 may serve as a potential therapeutic target for SSc.

The TF-mRNA regulatory network analysis in this study revealed that the two key biomarkers, CCL19 and LOXL2, are commonly regulated by the transcription factor FOXL1. Evidence from previous studies has provided important insights into the role of FOXL1 in fibrosis disease. *Miyashita* et al reported that FOXL1 is highly expressed in lung fibroblasts, where it is associated with DNA hypomethylation and super-enhancer formation, thereby regulating multiple fibroblast functions, including cell proliferation and collagen contractility.[56] More importantly, FOXL1 expression was markedly elevated in the lung tissues of patients with idiopathic pulmonary fibrosis, suggesting its potential role in fibrotic pathology.[56] Beyond fibrosis regulation, FOXL1 also exhibits notable anti-inflammatory properties. It has been shown that FOXL1 carried by bone marrow mesenchymal stem cell–derived exosomes suppresses high glucose–induced inflammation, apoptosis, and oxidative stress by activating the METTL3/ATXN2L pathway.[57] Although FOXL1 has not been reported in SSc, those findings suggest it may promote fibrosis by regulating fibroblast function while also influencing inflammation resolution. FOXL1 expression in lung fibroblasts is markedly higher than in fibroblasts from

other tissues,[56] and given that pulmonary fibrosis is a common organ complication of SSc, highlights a potential role of FOXL1 in SSc-related lung involvement.

The 2013 American College of Rheumatology (ACR)/European Alliance of Associations for Rheumatology (EULAR) criteria remain the standard for classifying SSc,[35] but the only serological markers included—scleroderma-related antibodies such as anti–topoisomerase I and anti–RNA polymerase III—carry limited diagnostic weight, underscoring the need for more reliable molecular indicators. CCL19 and LOXL2 may provide such indicators, as our study demonstrates their elevated expression in both plasma and skin tissue. Their dual detectability offers advantages for routine laboratory assessment as well as pathological evaluation. Beyond their diagnostic potential, CCL19 and LOXL2 also show promise as therapeutic targets. Currently, the clinical management of SSc-related skin fibrosis relies primarily on immunomodulatory therapies, such as methotrexate and rituximab, while effective treatments—particularly antifibrotic agents—remain lacking.[58] Both CCL19 and LOXL2 were identified through Mendelian randomization, supporting a causal role in SSc pathogenesis. LOXL2 is of particular therapeutic interest, as small-molecule LOXL2 inhibitors have already been developed and have shown benefit in liver and lung fibrosis.[52] These findings support not only the feasibility of targeting LOXL2 to treat cutaneous fibrosis in SSc but also the potential for therapeutic benefit in SSc-associated pulmonary and hepatic fibrosis. Furthermore, although no CCL19-targeted therapies have yet entered clinical development, early studies investigating its principal receptor, CCR7, have emerged,[59] indicating that this pathway remains a promising direction for further exploration.

Currently, only a limited number of MR studies have been conducted in SSc. These studies have explored the potential pathogenic roles of body fat distribution,[60] telomere length,[61] circulating levels of cytokines,[61] gut microbiome and plasma metabolome,[62] and basal metabolic rate[63] in SSc. Additionally, two MR studies have analyzed the relationships between SSc and lung cancer,[64] thyroid diseases, and appendicitis.[65] Among those studies, only one study explored the causal relationship between IL-5 and SCGF-β in circulating proteins and SSc,[61] while the other studies did not focus on potential therapeutic protein targets for SSc. In contrast, our study is the first to perform an MR analysis on the entire circulating proteome, providing a more comprehensive exploration of key plasma proteins with causal relationships in SSc. Given the critical role of fibroblasts and epithelial cells in SSc, we integrated scRNA-seq data, ultimately identifying only two key molecules, CCL19 and LOXL2. However, this does not imply that the remaining 104 key plasma proteins are unimportant. Besides fibroblasts and epithelial cells, immune cells such as macrophages, T cells, and B cells are also essential in SSc pathogenesis. Further studies incorporating these immune cells to refine the selection of key pathogenic plasma proteins remain an important avenue for future research. Notably, all 106 key plasma proteins were subjected to protein (gene)–key disease–key drug network analyses and phenotype scanning, providing a comprehensive map of their potential therapeutic relevance and off-target effects. This integrated approach not only facilitates the identification of candidate molecules for drug development but also highlights possible adverse outcomes, offering a valuable framework for prioritizing targets in translational and precision medicine studies. The network-based analyses highlight the extensive and interconnected relationships among key plasma proteins, diseases, and pharmacological agents. The identification of over 5000 interaction pairs within the integrated protein–disease–drug network underscores the biological and therapeutic complexity underlying systemic autoimmune disorders. These multilayered connections suggest that many of the key plasma proteins may exert pleiotropic effects across diverse disease pathways, implying both therapeutic potential and the risk of unintended consequences when these molecules are targeted. Moreover, phenotype scanning further identified potential adverse effects associated with several candidate molecules. For example, our results suggest that targeting CD40L (CD40LG) may increase the risk of mesothelioma. This prediction aligns with previous findings, as activation of CD40—the receptor for CD40L—has shown promising antitumor activity in clinical studies,[66] and gene delivery of CD40LG has demonstrated protective effects in murine mesothelioma models.[67] Conversely, inhibition of the CD40L pathway may exacerbate the progression of multiple malignancies, including mesothelioma. Potential adverse effects associated with targeting LOXL2 were also identified. LOXL2 appears to act as a protective factor in diffuse brain injury, indicating that therapeutic inhibition of LOXL2 may exacerbate this condition. It is important to emphasize, however, that the phenotype scanning only assessed the effects of these molecules on 56 key diseases; thus, these findings do not represent the full spectrum of possible therapeutic adverse

reactions. The identified potential risks require further experimental validation, and additional adverse effects that could not be captured by our approach may exist, warranting further investigation in future studies.

Our study has several limitations. First, the GWAS data for both exposure and outcome were derived from European populations. Although restricting the study population can reduce population stratification bias, it also limits the generalizability of our findings to other populations. Moreover, only 302 SSc patients were included in this study, which is a relatively small sample size for an MR analysis. This limitation is largely due to the inherently low prevalence of SSc worldwide. Addressing these constraints will require future studies with larger and more diverse SSc cohorts to validate and extend our findings. Second, as mentioned above, we identified 106 key plasma proteins potentially involved in SSc pathogenesis. For subsequent analyses, we focused only on those showing differential expression in fibroblasts and epithelial cells. Since these two cell types act as effector cells in the fibrotic process, our selected therapeutic targets are likely biased toward molecules involved in ECM production and regulation. However, SSc is fundamentally an immune dysregulation disorder, and investigating differentially expressed genes in immune cells—particularly adaptive immune cells—may reveal additional targets directly related to disease mechanisms. Finally, we only validated the expression of CCL19 and LOXL2 in a limited number of clinical samples, and further studies with larger sample sizes are needed to provide more robust evidence. Future research should address these limitations. First, integrating multi-omics data could offer a more comprehensive understanding of the disease, aiding in the identification of novel biomarkers and therapeutic targets. Further mechanistic studies on CCL19 and LOXL2, including their interactions with other proteins and signaling pathways, are crucial for developing novel therapeutic strategies. In ongoing experiments, we are evaluating LOXL2-targeted interventions in SSc animal models and assessing whether IFN drives the upregulation of CCL19 in SSc skin fibroblasts and epithelial cells. We hope that these follow-up studies will further clarify the biological roles of CCL19 and LOXL2 in SSc.

Conclusion

In this study, we identified 106 proteins associated with SSc risk using MR, representing the first comprehensive plasma-proteome MR analysis in SSc. Further analysis incorporating scRNA-seq datasets and an RNA-seq dataset from skin tissue allowed us to confirm CCL19 and LOXL2 as key molecules involved in SSc pathogenesis through fibroblasts and epithelial cells. Notably, we report for the first time their potential roles as molecular markers and therapeutic targets in SSc-related skin fibrosis.

Abbreviations

SSc, Systemic Sclerosis; GWAS, Genome-wide Association Study; pQTL, Protein Quantitative Trait Locus; MR, Mendelian Randomization; ECM, Extracellular Matrix; scRNA-seq, Single-cell RNA Sequencing; DEGs, Differentially Expressed Genes; IEU, Integrative Epidemiology Unit; SNPs, Single-nucleotide Polymorphisms; GEO, Gene Expression Omnibus; IVs, Instrumental Variables; IVW, Inverse variance weighted; OR, Odds Ratio; LOO, Leave-one-out; DSigDB, Drug Signatures Database; CTD, Comparative Toxicogenomics Database; PCA, Principal Component Analysis; UMAP, Uniform Manifold Approximation and Projection; FC, Fold Change; GSEA, Gene Set Enrichment Analysis; NES, Normalized Enrichment Score; MSigDB, Molecular Signatures Database; TFs, Transcription Factors; miRNAs, MicroRNAs; BEAM, Branched Expression Analysis Modeling; BLM, Bleomycin; HClO, Hypochlorous Acid; PBS, Phosphate-buffered Saline; IHC, Immunohistochemistry; qRT-PCR, Quantitative reverse Transcription Polymerase Chain Reaction; ANOVA, Analysis of Variance; CD74, Cluster of Differentiation 74; GFAP, Glial fibrillary Acidic Protein; CD40LG, CD40 Ligand; CEACAM21, Carcinoembryonic Antigen-related Cell Adhesion Molecule 21; COL9A2, Collagen Type IX Alpha 2 Chain; EDAR, Ectodysplasin A Receptor; ELL2, Elongation Factor for RNA Polymerase II 2; ESR1, Estrogen Receptor 1; GRIA4, Glutamate Ionotropic Receptor AMPA Type Subunit 4; IL1RN, Interleukin 1 Receptor Antagonist; LAG3, Lymphocyte-activation Gene 3; NELL1, Neural EGFL Like 1; BE, Basal Epithelial Cells; OE, Other Epithelial Cells; CD, Collecting Duct principal Cells; LE, Luminal Epithelial Cells; TB, Type B intercalated Cell; MMP, Matrix Metalloproteinase; CCR7, C-C Chemokine Receptor Type 7; IFN, Interferon; TLS, Tertiary Lymphoid Structures; TGF-β1, Transforming Growth Factor Beta 1.

Data Sharing Statement

All datasets were obtained from public databases, with detailed sources specified in the manuscript. R code and other results are available from the corresponding author upon reasonable request.

Ethics Approval and Informed Consent

The clinical and animal samples used in this study were approved by the Ethics Committee of the First Affiliated Hospital of Xi'an Jiaotong University (Approval No.: LLSBPJ-2023-205) and the Animal Experiment Ethics Committee of Xi'an Jiaotong University (Approval No.: XJTUAE2023-305), respectively. These ethical approvals permit further secondary analyses of the archived pathological tissue sections and mouse specimens. All pathological tissues were collected for clinical purposes, and all patients provided written informed consent for the use of their pathological samples in this study.

Consent for Publication

This study utilized only publicly available, de-identified data from the GWAS and GEO. The original studies obtained informed consent from all participants, and ethical approval for data sharing and secondary analysis was granted by their respective institutional review boards. Therefore, no additional consent for publication was required for this analysis. The use of skin tissue specimens was approved by the ethics committee, and informed consent was obtained.

Acknowledgments

I would like to express my heartfelt gratitude to Jie Liu (Meta Platforms, Inc., USA) for his invaluable guidance on the R programming throughout this study.

Author Contributions

YL and QL: Conceptualization, Supervision. HL, XL, XW, ZQ and YL: Methodology, Formal analysis, Data curation. HL, QL and JW: Writing – original draft. XC, YW and LM: Investigation, Resources. All authors participated in Writing – review & editing. All authors gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

The study was supported by the National Natural Science Foundation of China (No.82371811) and the Shaanxi Province Natural Science Foundation (No. 2023-JC-YB-754 and No. 2023-JC-YB-774).

Disclosure

The authors declare that they have no competing interests.

References

1. Volkmann ER, Andreasson K, Smith V. Systemic sclerosis. *Lancet*. 2023;401(10373):304–318. doi:10.1016/S0140-6736(22)01692-0
2. Good SD, Lee JY, Johnson RE, Volkmann ER. A scoping review of the epidemiology of systemic sclerosis and its organ manifestations: 2018–2024. *Curr Opin Rheumatol*. 2025;37(2):103–112. doi:10.1097/BOR.0000000000001063
3. Bournia VK, Fragoulis GE, Mitrou P, et al. All-cause mortality in systemic rheumatic diseases under treatment compared with the general population, 2015-2019. *RMD Open*. 2021;7(3):e001694. doi:10.1136/rmdopen-2021-001694
4. Jerjen R, Nikpour M, Krieg T, Denton CP, Saracino AM. Systemic sclerosis in adults. Part I: clinical features and pathogenesis. *J Am Acad Dermatol*. 2022;87(5):937–954. doi:10.1016/j.jaad.2021.10.065
5. Bellando-Randone S, Matucci-Cerinic M. Very early systemic sclerosis. *Best Pract Res Clin Rheumatol*. 2019;33(4):101428. doi:10.1016/j.berh.2019.101428
6. Hughes M, Bruni C, Cuomo G, et al. The role of ultrasound in systemic sclerosis: on the cutting edge to foster clinical and research advancement. *J Scleroderma Relat Disord*. 2020;6(2):123–132. doi:10.1177/2397198320970394
7. Di Maggio G, Confalonieri P, Salton F, et al. Biomarkers in systemic sclerosis: an overview. *Curr Issues Mol Biol*. 2023;45(10):7775–7802. doi:10.3390/cimb45100490
8. Zheng J, Haberland V, Baird D, et al. Phenome-wide mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet*. 2020;52(10):1122–1131. doi:10.1038/s41588-020-0682-6

9. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics*. 2002;1(11):845–867. doi:10.1074/mcp.R200007-MCP200
10. Worrell JC, O'Reilly S. Bi-directional communication: conversations between fibroblasts and immune cells in systemic sclerosis. *J Autoimmun*. 2020;113:102526.
11. Rosendahl AH, Schönborn K, Krieg T. Pathophysiology of systemic sclerosis (scleroderma). *Kaohsiung J Med Sci*. 2022;38(3):187–195. doi:10.1002/kjm2.12505
12. Kim J, Seki E. Hyaluronan in liver fibrosis: basic mechanisms, clinical implications, and therapeutic targets. *Hepatol Commun*. 2023;7(4). doi:10.1097/HC9.0000000000000083
13. Yu D, Yin G, Lei J, et al. The correlation between serum levels of laminin, type IV collagen, type III procollagen N-terminal peptide and hyaluronic acid with the progression of post-COVID-19 pulmonary fibrosis. *Front Cell Develop Biol*. 2024;12:1382244.
14. Yuan S, Xu F, Li X, et al. Plasma proteins and onset of type 2 diabetes and diabetic complications: proteome-wide Mendelian randomization and colocalization analyses. *Cell Rep Med*. 2023;4(9):101174. doi:10.1016/j.xcrm.2023.101174
15. Ou YN, Yang YX, Deng YT, et al. Identification of novel drug targets for Alzheimer's disease by integrating genetics and proteomes from brain and blood. *Mol Psychiatry*. 2021;26(10):6065–6073. doi:10.1038/s41380-021-01251-6
16. Zhang L, Xiong Y, Zhang J, Feng Y, Xu A. Systematic proteome-wide mendelian randomization using the human plasma proteome to identify therapeutic targets for lung adenocarcinoma. *J Transl Med*. 2024;22(1):330.
17. Li J, Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLOS Genet*. 2017;13(11):e1007081.
18. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512–525. doi:10.1093/ije/dyv080
19. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304–314. doi:10.1002/gepi.21965
20. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG. Using published data in mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol*. 2015;30(7):543–552. doi:10.1007/s10654-015-0011-z
21. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife*. 2018;7:e34408.
22. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. 2017;46(6):1985–1998. doi:10.1093/ije/dyx102
23. Liu P, Xu H, Shi Y, Deng L, Chen X, Xu L. Potential molecular mechanisms of plantain in the treatment of gout and hyperuricemia based on network pharmacology. *Evid Based Complement Alternat Med*. 2020;2020(1). doi:10.1155/2020/3023127
24. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnol*. 2015;33(5):495–502. doi:10.1038/nbt.3192
25. Yin H, Distler O, Shen L, et al. Endothelial response to type i interferon contributes to vasculopathy and fibrosis and predicts disease progression of systemic sclerosis. *Arthritis Rheumatol*. 2023;76(1):78–91. doi:10.1002/art.42662
26. Hukara A, Bonazza GA, Tabib T, et al. Elevated Fcy receptor expression augments pro-inflammatory macrophage phagocytosis in systemic sclerosis and associated rheumatic diseases. *Rheumatology*. 2024.
27. Zheng Y, Gao W, Zhang Q, et al. Ferroptosis and autophagy-related genes in the pathogenesis of ischemic cardiomyopathy. *Front Cardiovasc Med*. 2022;9:906753.
28. Griss J, Viteri G, Sidiropoulos K, Nguyen V, Fabregat A, Hermjakob H. ReactomeGSA - efficient multi-omics comparative pathway analysis. *Mol Cell Proteomics*. 2020;19(12):2115–2125. doi:10.1074/mcp.TIR120.002155
29. Saidmamatov O, Jammatov J, Sousa C, Barros R, Vasconcelos O, Rodrigues P. Translation and adaptation of the adult Developmental Coordination Disorder/Dyspraxia Checklist (ADC) into Asian Uzbekistan. *Sports*. 2023;11(7):135. doi:10.3390/sports11070135
30. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation*. 2021;2(3):100141. doi:10.1016/j.xinn.2021.100141
31. Uhlitz F, Bischoff P, Peidli S, et al. Mitogen-activated protein kinase activity drives cell trajectories in colorectal cancer. *EMBO Mol Med*. 2021;13(10). doi:10.15252/emmm.202114123
32. Kong L, Pokatayev V, Lefkovith A, et al. The landscape of immune dysregulation in Crohn's disease revealed through single-cell transcriptomic profiling in the ileum and colon. *Immunity*. 2023;56(2):444–458.e445. doi:10.1016/j.immuni.2023.01.002
33. Cords L, Tietscher S, Anzeneder T, et al. Cancer-associated fibroblast classification in single-cell and spatial proteomics data. *Nat Commun*. 2023;14(1). doi:10.1038/s41467-023-39762-1
34. Joo EH, Kim S, Park D, et al. Migratory tumor cells cooperate with cancer associated fibroblasts in hormone receptor-positive and HER2-negative breast cancer. *Int J Mol Sci*. 2024;25(11):5876. doi:10.3390/ijms25115876
35. van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. *Ann Rheum Dis*. 2013;72(11):1747–1755. doi:10.1136/annrheumdis-2013-204424
36. Li H, Li Q, Hao Z, et al. A recombinant IL-1beta vaccine attenuates bleomycin-induced pulmonary fibrosis in mice. *Vaccine*. 2024;42(18):3774–3788. doi:10.1016/j.vaccine.2024.04.091
37. Li H, Ju B, Luo J, et al. Type I interferon-stimulated genes predict clinical response to belimumab in systemic lupus erythematosus. *Eur J Pharmacol*. 2025;987:177204. doi:10.1016/j.ejphar.2024.177204
38. Staples TL. Expansion and evolution of the R programming language. *Royal Soc Open Sci*. 2023;10(4). doi:10.1098/rsos.221550
39. O'Reilly S. Emerging therapeutic targets in systemic sclerosis. *J Mol Med*. 2024;102(4):465–478. doi:10.1007/s00109-024-02424-w
40. Muruganandam M, Ariza-Hutchinson A, Patel RA, Sibbitt Jr WL Jr. Biomarkers in the pathogenesis, diagnosis, and treatment of systemic sclerosis. *J Inflamm Res*. 2023;16:4633–4660. doi:10.2147/JIR.S379815
41. Gowhari Shabgah A, Al-Obaidi ZMJ, Sulaiman Rahman H, et al. Does CCL19 act as a double-edged sword in cancer development? *Clin Exp Immunol*. 2022;207(2):164–175. doi:10.1093/cei/uxab039
42. Moschovakis GL, Bubke A, Friedrichsen M, et al. The chemokine receptor CCR7 is a promising target for rheumatoid arthritis therapy. *Cell Mol Immunol*. 2019;16(10):791–799. doi:10.1038/s41423-018-0056-5

43. Mathes AL, Christmann RB, Stifano G, et al. Global chemokine expression in systemic sclerosis (SSc): CCL19 expression correlates with vascular inflammation in SSc skin. *Ann Rheum Dis*. 2014;73(10):1864–1872. doi:10.1136/annrheumdis-2012-202814
44. Campitiello R, Soldano S, Gotelli E, et al. The intervention of macrophages in progressive fibrosis characterizing systemic sclerosis: a systematic review. *Autoimmun Rev*. 2024;23(10):103637. doi:10.1016/j.autrev.2024.103637
45. Zhang M, Zhang S. T Cells in Fibrosis and Fibrotic Diseases. *Front Immunol*. 2020;11:1142. doi:10.3389/fimmu.2020.01142
46. Onder L, Papadopoulou C, Lutge A, et al. Fibroblastic reticular cells generate protective intratumoral T cell environments in lung cancer. *Cell*. 2025;188(2):430–446e420. doi:10.1016/j.cell.2024.10.042
47. Yu -W-W, Barrett JNP, Tong J, et al. Skin immune-mesenchymal interplay within tertiary lymphoid structures promotes autoimmune pathogenesis in hidradenitis suppurativa. *Immunity*. 2024;57(12):2827–2842.e2825. doi:10.1016/j.immuni.2024.11.010
48. Lai YR, Pan L, Jiang XK, et al. CCL19 + fibroblasts promote tertiary lymphoid structure in oral lichen planus: a retrospective study. *Oral Dis*. 2025;31(7):2191–2205. doi:10.1111/odi.15266
49. Gerges E, Cauvet A, Schwarz M, Avouac J, Allanore Y. Association of serum interferon alpha-2a levels with disease severity and prognosis in systemic sclerosis. *Rheumatology*. 2024.
50. Di Donato S, Ross R, Karanth R, et al. Serum type I interferon score for prediction of clinically meaningful disease progression in limited cutaneous systemic sclerosis. *Arthritis Rheumatol*. 2025;77(7):929–941. doi:10.1002/art.43120
51. Puente A, Fortea JI, Cabezas J, et al. LOXL2-A new target in antifibrogenic therapy? *Int J Mol Sci*. 2019;20(7):1634. doi:10.3390/ijms20071634
52. Chen W, Yang A, Jia J, Popov YV, Schuppan D, You H. Lysyl Oxidase (LOX) family members: rationale and their potential as therapeutic targets for liver fibrosis. *Hepatology*. 2020;72(2):729–741. doi:10.1002/hep.31236
53. Waldrep KM, Rodgers JI, Garrett SM, Wolf BJ, Feghali-Bostwick CA. The role of SOX9 in IGF-II-mediated pulmonary fibrosis. *Int J Mol Sci*. 2023;24(14):11234. doi:10.3390/ijms241411234
54. Semkova ME, Hsuan JJ. TGFbeta-1 induced cross-linking of the extracellular matrix of primary human dermal fibroblasts. *Int J Mol Sci*. 2021;22(3):984. doi:10.3390/ijms22030984
55. Yao Y, Findlay A, Stolp J, Rayner B, Ask K, Jarolimek W. Pan-lysyl oxidase inhibitor PXS-5505 ameliorates multiple-organ fibrosis by inhibiting collagen crosslinks in rodent models of systemic sclerosis. *Int J Mol Sci*. 2022;23(10):5533. doi:10.3390/ijms23105533
56. Miyashita N, Horie M, Suzuki HI, et al. FOXL1 regulates lung fibroblast function via multiple mechanisms. *Am J Respir Cell Mol Biol*. 2020;63(6):831–842. doi:10.1165/rcmb.2019-0396OC
57. Niu C, Dong D, Cui L, Dong Y, Wang W. Exosomal FOXL1 from bone marrow mesenchymal stem cells activates the METTL3/ATXN2L pathway to ameliorate high glucose-induced human retinal microvascular endothelial cell injury. *Diabetol Metab Syndr*. 2025;17(1). doi:10.1186/s13098-025-01804-7
58. Del Galdo F, Lescoat A, Conaghan PG, et al. EULAR recommendations for the treatment of systemic sclerosis: 2023 update. *Ann Rheumatic Dis*. 2025;84(1):29–40. doi:10.1136/ard-2024-226430
59. Leser FS, Júnyor F, Pagnoncelli IB, et al. CCL21-CCR7 blockade prevents neuroinflammation and degeneration in Parkinson's disease models. *J Neuroinflammation*. 2025;22(1). doi:10.1186/s12974-024-03318-x
60. Villanueva-Martin G, Acosta-Herrera M, Kerick M, et al. The effect of body fat distribution on systemic sclerosis. *J Clin Med*. 2022;11(20):6014. doi:10.3390/jcm11206014
61. Jiang Z, Yao X, Lan W, et al. Associations of the circulating levels of cytokines with risk of systemic sclerosis: a bidirectional mendelian randomized study. *Front Immunol*. 2024;15.
62. Xie S, Meng Q, Wang L. The effect of gut microbiome and plasma metabolome on systemic sclerosis: a bidirectional two-sample mendelian randomization study. *Front Microbiol*. 2024;15:1427195. doi:10.3389/fmicb.2024.1427195
63. Wu T, Wang Y, Xia Y, et al. Causal association of basal metabolic rate on systemic sclerosis: a bidirectional mendelian randomization study. *Arch Dermatol Res*. 2024;316(8). doi:10.1007/s00403-024-03248-x
64. Peng H, Wu X, Wen Y, et al. Association between systemic sclerosis and risk of lung cancer: results from a pool of cohort studies and mendelian randomization analysis. *Autoimmunity Rev*. 2020;19(10):102633. doi:10.1016/j.autrev.2020.102633
65. Han Z, Han P, Wang F, et al. Negative causal exploration of systemic sclerosis: a Mendelian randomization analysis. *Sci Rep*. 2024;14(1).
66. McVey JC, Beatty GL. Facts and hopes of CD40 agonists in cancer immunotherapy. *Clin Cancer Res*. 2025;31(11):2079–2087. doi:10.1158/1078-0432.CCR-24-1660
67. Friedlander PL, Delaune CL, Abadie JM, et al. Efficacy of CD40 ligand gene therapy in malignant mesothelioma. *Am J Respir Cell Mol Biol*. 2003;29(3):321–330. doi:10.1165/rcmb.2002-0226OC

Journal of Inflammation Research

Dovepress
Taylor & Francis Group

Publish your work in this journal

The Journal of Inflammation Research is an international, peer-reviewed open-access journal that welcomes laboratory and clinical findings on the molecular basis, cell biology and pharmacology of inflammation including original research, reviews, symposium reports, hypothesis formation and commentaries on: acute/chronic inflammation; mediators of inflammation; cellular processes; molecular mechanisms; pharmacology and novel anti-inflammatory drugs; clinical conditions involving inflammation. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: https://www.dovepress.com/journal-of-inflammation-research-journal