

Altered AP-1, RUNX and EGR chromatin dynamics drive fibrotic lung disease

Eleanor Valenzi^{1,2}, Minxue Jia^{*3}, Peter Gerges^{*4}, Jingyu Fan⁴, Tracy Tabib^{2,4}, Rithika Behara², Yuechen Zhou², John Sembrat¹, Jishnu Das⁴, Panayiotis V Benos^{3,5}, Harinder Singh⁴, Robert Lafyatis²

¹Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, University of Pittsburgh

²Division of Rheumatology and Clinical Immunology, University of Pittsburgh

³Department of Computational and Systems Biology, University of Pittsburgh

⁴Center for Systems Immunology and Department of Immunology, University of Pittsburgh

⁵Department of Epidemiology, University of Florida

*These authors contributed equally to the work.

Abstract

Pulmonary fibrosis, including systemic sclerosis-associated interstitial lung disease (SSc-ILD), involves myofibroblasts and SPP1^{hi} macrophages as drivers of fibrosis. Single-cell RNA sequencing has delineated fibroblast and macrophages transcriptomes, but limited insight into transcriptional control of profibrotic gene programs. To address this challenge, we analyzed multiomic snATAC/snRNA-seq on explanted SSc-ILD and donor control lungs. The neural network tool ChromBPNet inferred increased TF binding at single base pair resolution to profibrotic genes, including CTHRC1 and ADAM12, in fibroblasts and SPP1 and CCL18 in macrophages. The novel algorithm HALO confirmed AP-1, RUNX, and EGR TF activity controlling profibrotic gene programs and established TF-regulatory element-gene networks. This TF action atlas provides comprehensive insights into the transcriptional regulation of fibroblasts and macrophages in healthy and fibrotic human lungs.

Introduction

Fibrosis is a lethal feature in numerous diseases. Myofibroblasts drive fibrosis due to their excessive production of extracellular matrix, resulting in tissue stiffness and organ dysfunction. While necessary for normal physiologic wound healing, dysregulated myofibroblasts contribute to fibrosis in pulmonary, renal, dermal, hepatic, and cardiac tissues. Macrophages, particularly a subset marked by SPP1, are implicated in activating myofibroblasts and infiltrate fibrotic tissues, including the lungs, liver¹, skeletal muscle² and heart³⁻⁵. Deleting or silencing *SPP1*, encoding Osteopontin, ameliorates bleomycin induced lung fibrosis⁶⁻⁸. Single cell RNA-sequencing (scRNA-seq) meta-analyses reveal that SPP1^{hi} macrophages share similar transcriptional states across organs affected by fibrosis^{1,9}.

Systemic sclerosis is a prototypic fibrotic disease, with fibrosis affecting multiple organs. Systemic sclerosis-associated interstitial lung disease (SSc-ILD) is the leading cause of disease-associated mortality in patients with SSc¹⁰. Despite current therapies targeting immune dysregulation, vasculopathy, and fibrosis, limited progress has been made in ameliorating disease-associated mortality or improving quality of life. Myofibroblasts and SPP1^{hi} macrophages are prominent cell populations and are associated with progressive lung disease¹¹. Thus, a deeper understanding of the transcriptional programs regulating these cells' profibrotic phenotypes is critical for elucidating the molecular mechanisms underlying disease and informing new therapeutic strategies.

ScRNA-seq has delineated the transcriptomes of various cell types in the healthy and fibrotic lung¹²⁻¹⁴, revealing perturbed cell states and altered cell-cell communication networks in disease^{15,16,17}. Despite generating biologically relevant inferences of altered cell phenotypes and signaling pathways, these studies are limited in their ability to decipher how gene regulatory networks are perturbed, leading to the altered cellular states present in fibrosis. Gene expression is regulated in a cell type- and tissue-specific manner by transcription factors (TFs) interacting with *cis*-regulatory elements (CREs) in the genome. Understanding changes in TF dynamics at CREs in key profibrotic cell populations, namely myofibroblasts and SPP1^{hi} macrophages, can provide direct insights into disrupted regulatory networks in fibrosis.

Single-nucleus ATAC-seq (snATAC-seq) permits the identification of open (accessible) chromatin regions (OCRs), that represent candidate CREs, in diverse cell populations including from diseased tissues. ChromBPNet, a convolutional neural network model, predicts at single-base-pair resolution how TF action contributes to chromatin accessibility¹⁸. Perturbations in ChromBPNet contribution scores of TF motifs infer changes in TF dynamics in the context of disease. Using multiomic snATAC-seq/snRNA-seq on cells isolated from human SSc-ILD and healthy lungs, we identified differentially accessible chromatin regions and TF binding dynamics, allowing us to infer gene-regulatory networks in fibrotic SSc-ILD cell populations. Focusing on fibroblasts and SPP1^{hi} macrophages, we identified altered binding of AP-1, RUNX and EGR TFs in SSc-ILD fibroblasts, as well as basic helix-loop-helix(bHLH) ZIP and AP-1 TFs in SSc-ILD macrophages involving CREs associated with genes promoting fibrosis. These studies provide *in vivo* insights into the altered TF dynamics driving the profibrotic states of these cell populations.

Results

Epigenomic markers of healthy and SSc-ILD lung cell populations

We performed paired scRNA-seq and snATAC-seq on 3 SSc-ILD and 2 control lung samples, and multiome snRNA-seq/snATAC-seq on nuclei from 6 SSc-ILD and 7 control explants. Subpleural tissue was collected at lung transplant in patients with SSc-ILD, and from organ donors without preexisting lung disease. Sixteen samples were partially enriched for mesenchymal cells. Histology showed usual interstitial pneumonia in 8 SSc-ILD samples and more predominant vascular disease in 1 (Supp Fig 1, Supp Table 1). After filtering, 67,446 nuclei (average 3,747 per donor) were included for analyses. Normalization, batch correction, and dimensionality reduction were performed using Seurat, Signac, and Harmony^{19,20}. SnATAC data alone was utilized for initial clustering after delineating OCRs/peaks using MACS2. We identified

35 distinct clusters based on gene expression and promoter accessibility (Figure 1A-B, Supp Fig 2). All major pulmonary cell types were identified with 446,087 OCRs among 67,446 nuclei. Nuclei clustered by cell type and were well integrated by chemistry, disease state, and individual sample after batch correction (Figure 1C-D, Supp Fig 3A-B). Low quality clusters were excluded from further analyses.

OCRs were annotated to the nearest gene and region (Figure 1E), with 20.32% of OCRs occurring within 3 kb of transcription start sites, and most OCRs in distal intergenic or intronic regions.²¹ Cell clusters were distinguished by differentially expressed genes (DEGs) and differentially accessible chromatin regions (DARs), with cluster 9 and 31 fibroblasts showing the most unique DARs (Figure 1G). DEGs and DARs were identified between SSc and control for each population (Figure 1F), with fibroblast, SPP1^{hi} macrophage, mixed epithelial, and arterial endothelial cells having the largest number of significant DARs, reflecting significant chromatin accessibility changes in these cell types with fibrosis.

Next, we identified cell type-specific TFs by gene expression and motif enrichment. TF motifs were identified using chromVAR²² (Supp Fig 4 A, B), revealing known regulators such as ETS1 in NK and T cells, SPI1 in macrophages, and GATA2 in mast cells, validating the analysis. Enriched fibroblast motifs included AP-1, Tal-related (including TWIST1 and HAND2), and TEAD TF family motifs. While TF motif activity and gene expression showed overall limited correlation, 23 TFs had significant positive correlations, including TWIST1 and HAND2 in fibroblasts, SPI1 in macrophages, and SOX18 in endothelial cells (Supp Fig 4C). We also evaluated enriched motifs for control lungs only, with similar results (Supp Fig 5).

Altered Chromatin accessibility of fibrotic mesenchymal cell populations in SSc-ILD

Given the significant chromatin accessibility changes between SSc and control fibroblasts, we subclustered fibroblast, smooth muscle, and pericyte populations (n=13,766 nuclei; Figure 2A). We identified 10 fibroblast populations, including 4 adventitial 2 alveolar, a CXCL2^{hi} fibroblast, and 3 myofibroblast-related clusters (myofibroblast, inflammatory myofibroblast, and activated myofibroblast). Myofibroblasts, inflammatory myofibroblasts, and activated myofibroblasts were expanded in SSc-ILD, along with smooth muscle and pericytes (Figure 2B,C, Supp Fig 6A), consistent with prior scRNA-seq analyses¹⁴, and reflective of both severe ILD and the group 3 pulmonary hypertension present in these patients.

We compared SSc-ILD fibroblasts to controls, SSc-ILD myofibroblasts vs SSc non-myogenic fibroblasts (including the adventitial, alveolar, and CXCL2^{hi} populations), and activated myofibroblasts to other fibroblasts to robustly investigate the DEGs in these fibroblast population (Supp Fig. 6C and 7). Activated myofibroblasts exhibited the highest expression of myofibroblast hallmark genes including *CTHRC1*, *POSTN*, *COMP*, and *COL8A1* (Fig 2D, Supp Fig 6C) and were associated with profibrotic pathways (Supp Fig 6D,E). Forty-two DEGs were upregulated in common in SSc fibroblasts, SSc myofibroblasts, and activated myofibroblasts, including *POSTN*, *TWIST1*, and *COL8A1* (Supp Fig 7).

In fibroblast clusters, 6,002 DARs were more accessible in SSc vs 6,693 in control fibroblasts (Figure 2E, Supp Table 2) with Cell migration and Collagen metabolic processes amongst the top upregulated pathways (Figure 2F). Activated myofibroblasts showed significant overlap in DARs with inflammatory myofibroblasts and myofibroblasts, suggesting similar epigenetic regulation (Supp Fig. 8A). Of 1,024 DARs more accessible in both SSc myofibroblasts and activated myofibroblasts (Supp Fig. 8D), 96 were annotated to upregulated genes in comparing SSc-ILD myofibroblasts vs non-myogenic fibroblasts, including *COMP*, *COL8A1*, and *TWIST1*. Many shared DARs occurred in genes previously related to fibrosis such as *COL8A1*, *FBN2* and *STAT3* (Figure 2G).

Mapping TF motifs in OCRs of SSc vs control fibroblasts revealed that AP-1, RUNX, and EGR families had the highest positive average difference motif activity scores, while C/EBP motifs showed more accessibility in control fibroblast OCRs (Figure 2H). Since multiple related

TFs often share nearly equivalent motifs, enriched TFs with adjusted p-value <0.05 and absolute average difference of 0.5 or greater were ranked based on the top score for each TF family.

Enhanced binding of AP-1, RUNX and EGR transcription factors by fibrotic fibroblasts

To investigate chromatin accessibility dynamics in fibroblasts, we employed ChromBPNet, a deep learning-based tool trained using DNA sequences and their corresponding ATAC-seq DNA cleavage chromatin accessibility patterns, that generates base pair resolution contribution scores after removing Tn5 enzyme cutting bias¹⁸. After removing bias, corrected signal profiles for each cell state were used to generate single-base pair resolution contribution scores, with de novo motif discovery by TF-MoDISco^{23,24} and annotation by JASPAR TF motifs²⁵. Analysis peaks were called using ChromBPNet contribution score tracks, with aligned genomic coordinates for a given TF motif used to generate cluster-specific aggregate contribution scores (Figure 3A). The seqlets inferred by ChromBPNet as showing TF binding were consistent with known binding for several families of TFs and showed differences in seqlet numbers between control and SSc-ILD fibroblasts for AP-1, RUNX, EGR, and C/EBP families (Figure 3B).

The most frequently identified de novo seqlet in SSc and control fibroblast peaks matched the consensus AP-1 binding site. Unique and overlapping AP-1 binding sites and their computed contribution score products were identified in both control and SSc-ILD fibroblasts (Figure 3C). We then clustered seqlets and visualized the average contribution weight matrix for each cluster (Figure 3D, 3E, Supp Fig 9A,B). Although scores were higher in nearly every cluster for SSc fibroblasts, cluster 7 had distinctly higher scores in SSc, while cluster 9 had higher scores in control fibroblasts. The AP-1 cluster 7 seqlets were annotated to nearest genes by CHIPseeker, followed by pathway analysis by Gene Ontology (Figure 3F). AP-1 binding was notably stronger in genes like *CDH8*, *TNC*, and *ITGB3*, involved in Cell junction organization (Figure 3G). The computational process of ChromBPNet in refining the pseudobulk ATAC-seq data is shown in analysis of *ADAMI2*, a metalloprotease whose expression is increased in lung (Figure 2D) and skin myofibroblasts²⁶, correlates with the severity of skin disease²⁷, and predicts response to anti-TGF β therapy in SSc skin²⁸. A strong TF footprint for AP-1 is identified in the *ADAMI2* intron 11 in SSc fibroblasts (Figure 3H). Similarly, a duplicate AP-1 site was detected upstream of the transcription start site of the activated myofibroblast marker *CTHRC1*, a significant marker of pathogenic fibroblasts in IPF and murine fibrosis^{13,29} (Figure 3I).

RUNX, EGR, C/EBP, and TEAD motifs were analyzed in an analogous manner. RUNX footprints were present in a small number of seqlets, yet markedly increased in SSc fibroblasts (Supp Fig 10A,11) with enriched pathways linked to genes involved in Multicellular Organism Development and Anatomical Structure Morphogenesis (cluster 9), and Regulation of Signal Transduction and Anatomical Structure Development (cluster 10; Supp Fig 10B-E). Other genes linked to lung fibrosis with upregulated expression and enhanced RUNX binding in SSc fibroblasts included *FHL2*, *PAPPA* and *ECM2*, as well as *CTTNB2*, a gene previously linked to dermal fibrosis (Supp Fig 11C-F).

EGR clusters 1, 2, 9, and 10 had the greatest contribution score increases in SSc fibroblasts, and were associated with Actin-filament Based Processes and Cell Morphogenesis (Supp Fig 10F-H, Supp Fig 11G,H). EGR footprinting was present in intron 2 of *ADAMI2* (Supp Fig 10I), which as mentioned above is also regulated by AP-1 (Figure 3H). TEAD footprints were more evenly distributed to CREs in SSc and control fibroblasts (Supp Fig 12A,B). TEAD cluster 8 contained higher contribution scores in SSc fibroblasts, and was associated with Regulation of Actin-Filament Based Processes and Cardiac Muscle Tissue Growth, consistent with the TEAD TFs known role in muscle differentiation (Supp Fig 12C,D)³⁰. C/EBP-binding seqlets were decreased in SSc fibroblasts, most strikingly in cluster 8 (Supp Fig 9C,D).

Altered Chromatin accessibility of macrophage populations in SSc-ILD

As in prior scRNA-seq studies, we identified macrophage clusters implicated in driving fibrosis in SSc-ILD and IPF by increased expression of *SPP1* and other marker genes (Figures 1A,

4A)^{14,31}. We combined the SPP1^{hi} macrophages to examine DARs between SSc-ILD and control lungs, finding 5,194 DARs more accessible in SSc SPP1^{hi} macrophages, and 4,133 in control (Figure 4B, Supp Table 3). Upregulated pathways included Cell-matrix Adhesion, PDGFR Signaling, and ERK1/2 Cascade (Figure 4C). Top DARs in SSc SPP1^{hi} macrophages were annotated to *CHIT3*, *PSD3*, and *PRC* (Figure 4E). ChromVAR identified bHLH-ZIP (USF1, USF2, MXI1, MITF, TFEB), EGR and NFAT TFs as highly enriched SSc, while CTCF motifs were enriched in control SPP1^{hi} macrophages (Figure 4D).

ChromBPNet analysis predicted bias-corrected TF binding at single base pair resolution from the pseudobulk SPP1^{hi} macrophages. TF-MoDISco indicated that Ets-related binding sites were the most common, followed by C/EBP, AP-1, and CTCF (Figure 5A). AP-1, bHLH-ZIP, and EGR motifs were increased in SSc SPP1^{hi} macrophage seqlets compared to control. In examining AP-1 contribution scores genome wide, cluster 7 had distinctly high scores in SSc (Figure 5B, Supp Fig 16 A,B) and was linked to pathways including Cell-matrix Adhesion and Cell Migration (Figure 5C,D). Cluster 8 had high scores in controls and was enriched for ERBB and JAK-STAT signaling (Figure 5E,F). bHLH-ZIP footprinting for the motif examined (MA0093.3) was mildly increased for unique SSc seqlets (Figure 5G, Supp Fig 16 C,D), with cluster 9 containing higher scores in SSc and linked to Metalloendopeptidase Activity and Autophagy (Figure 5I,J). In SSc, higher bHLH-ZIP contribution scores were present in the *CCL18* promoter, a chemokine ligand influencing collagen production by fibroblasts and highly upregulated by SPP1^{hi} macrophages³².

Gene regulatory network construction

To further examine the dysregulated gene regulatory networks of fibroblasts and SPP1^{hi} macrophages, we constructed TF-CRE-gene linkages from our multiome RNA/ATAC-seq data (Figure 6A). To better account for the sparsity of single-nucleus genomic data and to utilize multi-omics representations delineating cellular states, we first inferred latent representations using hierarchical causal modeling for single cell multi-omics data (HALO) and constructed metacells of paired RNA-seq and ATAC-seq from our multiome chemistry dataset (n=6 SSc, 7 Control samples)³³. Metacells are small groups of cells that are more granular than clusters, with each metacell representing a distinct cellular state^{34,35}. We then utilized our larger independent scRNA-seq dataset (n=17 SSc, 13 control lungs)³⁶ to identify DEGs and CREs (both promoters and more distal elements) occurring within 250 kilobases of the DEGs narrowed to DARs for the comparison of interest. Enriched TF motifs within these DARs were identified through motif matching, followed by TF-CRE-gene linkages.

For the fibroblast nuclei we first re-clustered the multiome samples based on HALO representations, followed by cell annotation, construction of metacells, and identification of cell states enriched in SSc fibroblasts using neighborhood-level differential abundance testing (Figure 6B,C, Supp Fig 18A)³⁷. RNA velocity was applied to assess cellular dynamics of the myofibroblast population³⁸. Velocity projections indicated that myofibroblasts arise from both alveolar and adventitial fibroblast populations (Figure 6D). In comparing SSc vs control fibroblasts, AP-1, EGR, YY1, and ETS-related family TFs had the largest positive average difference in motif activity (Figure 6E). The TF ZNF148 had the greatest number of combined proximal and distal CRE links (551) to DEGs, with IRF1 and KLF4 also having greater than 500 links. For this comparison, 65 TFs showed statistically significant differences in both gene expression and motif activity, 28 had discordant directionality for these differential values while 4 TFs were downregulated with negative motif activity differences, and 33 TFs were upregulated with positive motif activity differences (Supp Table 4). The AP-1 TFs JUND and JDP2 had a top 5 positive change for both measures. Members of the AP-1 family with significantly enriched motif activity (JUND, JUNB, JDP2, FOSL1) were linked to 280 unique DEGs (243 upregulated, 37 downregulated) for the SSc vs control fibroblasts, including the upregulated genes *CTHRC1*, *ADAM12*, *IGFBP2*, *SPP1* and *TGFBI*. Enriched EGR TFs EGR1, EGR2, and EGR3 were linked

to 320 unique DEGs (273 upregulated, 46 downregulated) including the upregulated genes *ADAM12*, *TMEM158*, *IGFBP2*, and *SPPI*.

Myeloid cells were examined in a similar manner. SPP1^{hi} macrophages were increased in SSc, while non-classical monocytes were increased in control samples (Figure 6F,G, Supp Fig 18D). Velocity projections predicted that SPP1^{hi} macrophages arise from both the FABP4^{hi} macrophage and monocyte populations, although a larger transition from monocytes to SPP1^{hi} macrophages was indicated in fibrotic lungs. (Figure 6H, Supp Fig 18F). In examining motif activity for this model, EGR, AP-1, bHLH-ZIP, and Ets-related TFs had the largest positive average difference in motif activity (Figure 6I). The TF ZNF148 had the greatest number of combined proximal and distal CRE links (146) to DEGs. For this comparison, 29 TFs showed statistically significant differences in both gene expression and motif activity, 12 had discordant directionality for these differential values while 7 TFs were downregulated with negative motif activity difference, and 10 TFs were upregulated with positive motif activity difference (Supp Table 5). TFEB and MXI1 had a top 5 positive change for both measures. Members of the bHLH-ZIP family with significantly enriched motif activity (MITF, MXI1, and TFEB) were linked to 64 unique DEGs (53 upregulated, 11 downregulated) for the SSc vs control SPP1^{hi} macrophage population, including the upregulated genes *PAQR5*, *MREG*, *SEL1L3*, *RTN4R*, and *APOE*. Enriched AP-1 TFs FOSL1 and JDP2 were similarly linked to 69 unique DEGs (54 upregulated, 15 downregulated) including the upregulated genes *CHIT1*, *CHI3L1*, *MMP7*, *DNASE2B*, *RTN4R*, and *CDH23*.

Discussion

Here, we delineate dynamic chromatin changes and altered TF binding in fibrotic lung disease. Our study provides a comprehensive map of TF regulation in the human lung and focuses on primary TFs regulating the transition of fibroblasts and macrophages to profibrotic states^{36,39}. Our study generated the largest multiomic lung cell atlas of both control and fibrotic ILD. Applying ChromVAR we inferred the TFs regulating cell states in healthy and fibrotic lungs. Applying the deep learning model ChromBPNet, we detect TF binding at single nucleotide resolution, confirming known TF motifs and inferring the altered binding in profibrotic fibroblasts and macrophages in an unbiased manner. HALO metacells enabled robust trajectory analysis and linked TF and CREs to regulated genes, supporting the TF binding indicated by ChromBPNet.

In SSc fibroblasts, we observed enriched binding to AP-1 TF motifs and high contribution score footprints of AP-1 binding in multiple fibrosis-related genes, such as collagen. AP-1 TFs, comprised of homo- and hetero-dimers of Jun, Fos, Atf, and Maf, are implicated in fibrosis^{40,41}. Fra-2 transgenic mice develop skin fibrosis, ILD, and vascular disease⁴². AP-1 is upregulated in a TGF- β -dependent manner in SSc dermal fibrosis, with reduction in collagen synthesis *in vitro* and prevention of dermal fibrosis *in vivo* following treatment with a selective AP-1 inhibitor⁴³. Our data suggest AP-1 regulates key myofibroblast marker genes, including *CTHRC1* and *ADAM12*. *CTHRC1* has been identified as a myofibroblast marker in fibrotic lung diseases ranging from IPF to COVID-19^{13,44}. Its deletion either aggregates or ameliorate fibrosis in lungs and liver respectively. Deletion of *ADAM12* blocks the development of profibrotic progenitors⁴⁵.

The significant redistribution of accessible AP-1 sites across the genome, and the identification of clusters of seqlets with AP-1 motifs showing no change or increased binding in control fibroblasts highlights the complex action of this TF family. Several mechanisms are likely in play. Specific AP-1 TFs may increase accessibility, while others decrease accessibility in SSc-ILD fibroblasts. For example, increased JunB replacing Jun regulates an enhancer upstream of COL1A2⁴⁰. Although binding of other TFs may influence binding to AP-1 sites locally or distally, our data remarkably did not indicate binding of TFs immediately adjacent to the AP-1 sites, which appear rather as islands of enhancer binding. Distal TF interactions, DNA loop formation, and/or epigenetic changes at the level of histone or DNA modifications appear more likely explanations. Determining the molecular mechanisms driving AP-1 binding differences between healthy and

SSc-ILD fibroblasts will clarify determinants of AP-1 action, the resulting gene dysregulation, and strategies for intervention.

Although RUNX binding was inferred in fewer seqlets than many TFs, its footprinting was more consistently increased in SSc and appeared in few control fibroblast OCRs. RUNX proteins bind DNA after complexing with their coactivator core-binding factor β -subunit(CBFB)⁴⁶. Inhibition of RUNX1 *in vitro* blocks myofibroblast differentiation⁴⁷, RUNX footprinting is enriched in SSc dermal fibroblasts⁴⁸ and RUNX1 has also been implicated in myofibroblast differentiation in scRNA-seq studies of SSc skin²⁶. Only one of 16 genes highlighted in cultured SSc dermal fibroblast ATAC-seq analyses was identified in our analysis: CTTNBP2⁴⁸. This may reflect differences between lung and skin myofibroblasts or the *in vitro* versus *in vivo* analysis enabled by snATAC-seq. Although pathway analysis of genes proximal to RUNX seqlets showing increased binding in SSc-ILD did not suggest traditional fibrosis pathways, seqlets associated with genes showing increased expression in SSc-ILD myofibroblasts included genes implicated in lung fibrosis. Notably, FHL2 is one of 12 genes overexpressed in both rat and human lung fibrosis⁴⁹. It binds to focal adhesions and transmits mechanical signals to regulate gene expression⁵⁰ and its pharmacologic inhibition by Arbutin ameliorates rat lung fibrosis⁵¹. ECM2 and PAPPA are markers of IPF and IPF severity⁵². Senescent cells secrete increased PAPP-A, suggesting a role in IPF^{52,53}. ECM2 deletion inhibits skeletal muscle-derived satellite cell differentiation⁵⁴, suggesting a possible role in myofibroblast differentiation. Further studies will be needed to clarify how RUNX proteins control myofibroblast differentiation. However, the availability of small molecule inhibitors for both RUNX-DNA binding and CBFB-RUNX binding⁵⁵⁻⁵⁷, as well as its significant post-transcriptional regulation by ubiquitin-proteasome degradation⁵⁸, offers a variety of approaches to inhibiting RUNX activity.

EGR1 is implicated in fibrotic signals from IL-13 and TGF- β ^{59,60}. EGR1 deletion inhibits wound healing and myofibroblast differentiation, while overexpression enhances wound repair⁶¹ and upregulates genes associated with wound healing and fibrosis⁶². EGR1 deletion also protects against renal fibrosis⁶³, but aggravates experimental liver fibrosis⁶⁴. EGR2 and EGR3 both mediate TGF- β dependent signals including collagen expression and myofibroblast differentiation but regulate distinct genes^{65,66}. Previous *in vitro* studies indicate SSc dermal fibroblasts have enhanced EGR1 accessibility⁶⁶, suggesting EGR1 may play parallel roles in myofibroblast differentiation in the skin and lungs.

In SPP1^{hi} macrophages, we identified enrichment of bHLH-ZIP, AP-1, and EGR motifs in SSc compared to controls, as well as high contribution scores of these TFs. The bHLH-ZIP TFs demonstrated the most significant motif enrichment, and the largest increase in SSc seqlets when identified *de novo* by ChromBPNNet and TF-ModISco, suggesting a distinct regulatory program in SPP1^{hi} macrophages compared to fibroblasts. The bHLH-ZIP TFs include USF1 and USF2, as well as MiT/TFE members MITF, TFEB, MXI1, and TFE3⁶⁷. Because these TFs share the same consensus motif, we cannot clearly distinguish which is most important. TFEB is a master regulator of lysosomal biogenesis and autophagy⁶⁸, while members also act in mitophagy, lipid catabolism, and tumor-associated macrophage polarization⁶⁹⁻⁷¹. They link mitochondrial and lysosomal functions, both implicated in SSc^{72,73}. We previously identified TFEB as a putative regulator of pro-fibrotic macrophages in SSc-ILD³⁶. USF1 is linked to aging-related macrophage function decline⁷⁴ and liver fibrosis⁷⁵. The enriched duplicate bHLH-ZIP footprint in proximal CCL18 is notable, as elevated CCL18 is associated with SSc-ILD severity and performs well as a diagnostic biomarker of SSc-ILD^{76,77}. CCL18 is decreased in the serum and skin biopsies of patients with SSc following IL-6 inhibition⁷⁸, and is downregulated by JAK inhibitors⁷⁹. CCL18 increases collagen production by lung fibroblasts⁸⁰, and genetic variability in *CCL18* affects gene expression and mortality in IPF⁸¹. bHLH-ZIP TFs may represent a key influence on the feedback loop between fibrotic macrophages and fibroblasts, warranting further investigation.

To our knowledge, this is the first large snATAC-sequencing study of adult healthy and fibrotic lungs, providing a rich resource on chromatin accessibility in the human lung. We utilized

deep learning models to uncover intricate sequence rules governing TF binding, providing insights into transcriptional programs regulated by CREs in fibroblasts and macrophages. As our study used explanted tissue from patients with severe SSc-ILD undergoing lung transplants, it may not reflect changes in mild disease or the broader SSc-ILD population. As OCRs were annotated by least base-pairs distance, not all annotations may reflect the regulated gene, as CREs can regulate distant genes via long-range interactions. While our study focused on OCRs and TF binding sites, additional non-coding regulatory mechanisms including long non-coding RNAs⁸², microRNAs⁸³, and transposable elements⁸⁴ may also influence gene expression.

In summary, utilizing multiomic single-cell approaches in advanced SSc-ILD and control lungs, we identified cell populations with significant changes in chromatin accessibility, altered TF activity in SSc-ILD fibroblasts and SPP1^{hi} macrophages, and the resulting TF-CRE-gene regulatory networks. The snATAC-seq data maps TFs determining cell populations in healthy and fibrotic lungs, highlighting cis-regulatory changes responsible for the diseased phenotypes of profibrotic macrophages and myofibroblasts. Our study lays the groundwork for functional validation of these differential CREs, their target genes, and the TFs driving them. This knowledge could guide the development of targeted therapeutics aimed at modulating TF or their coregulators to mitigate progressive fibrotic lung disease.

Methods

snATAC-seq/multiome Analysis

The University of Pittsburgh Institutional Review Board approved procedures involving human samples. Explanted subpleural peripheral lung tissue was digested to single-cell suspensions using Liberase DL (Roche) and bovine pancreas DNase I (Sigma-Aldrich) as previously described¹⁵. Samples were processed immediately upon receiving from the operating room. Some samples were partially enriched for fibroblasts using magnetic-activated depletion of CD45, CD31, CD163, and/or CD326 positive cells, with 50/50 loading of the enriched and non-enriched cell suspension at the time of running snATAC-seq as detailed in Supplementary Table 1. To increase data quality dead cell removal (STEMCELL, Annexin V kit) and CD66 depletion were added to sample processing for all digestions as well starting with sample SC421. For the snATAC-seq only samples, single-cell suspensions were split, with a portion used for performing scRNA-seq as previously described¹⁴, and the remaining suspension used for nuclei generation and snATAC-seq as previously described⁸⁵. For the multiome samples, transposition of nuclei and library preparation were carried out per the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression kit protocol (10x Genomics Catalog no: 1000283). Libraries were sequenced using the Illumina NovaSeq 6000 platform.

CellRanger ATAC pipeline (v2.0.0) and CellRanger Arc pipeline (v2.0.2) with h38 human genome reference were used for initial read processing, library quality control, peak calling and cell calling. Samples were filtered for: minimum features>200 and <150,000, peak region fragments>200 and < 150,000, nucleosome signal<10. For the snATAC-seq samples, the R package Signac (v1.9.0)¹⁹ was used for merging samples, unsupervised clustering, and downstream analyses. A unified peak set was created and quantified in each dataset, followed by new fragment objects, Seurat objects, and merging of the 5 samples. Batch correction at the individual sample level was performed using the R package harmony (v0.1.1)²⁰. Multiome chemistry samples were aggregated using CellRanger Arc 'aggr' for creation of a unified fragment file, followed by disaggregation, filtering, peak calling, and clustering by weighted nearest neighbor analysis. The multiome object peaks were quantified within the snATAC-seq object, followed by merging of these objects. Term frequency-inverse document frequency (TF-IDF) normalization and singular value decomposition (SVD) were performed on the DNA accessibility data using Signac's 'RunTFIDF' and 'RunSVD' functions, followed by removal of low-quality nuclei clusters. Peak calling was performed by cluster using MACS2⁸⁶, with this peak set used for downstream analysis and subclustering of the mesenchymal populations. For manual annotation of principal cell types, the UMAP projection of nuclei was queried for clusters showing

accessibility of known cell type-specific genes, as well as by gene expression profiles. DARs were calculated by logistic regression test with number of peak region fragments as the latent variable and requiring peak accessibility in 5 percent or more of nuclei in at least one population. Wilcoxon rank sum test with Bonferroni FDR correction was used for DEG testing. Signac was used for display of pseudo-bulk coverage tracks of Tn5 integration for designated cell populations.

For re-clustering of mesenchymal clusters, we subsetted the following clusters: 0,9,17,22,23,31,33, and 34, and performed normalization, dimensionality reduction, batch effect removal, clustering and cell type annotation as described above. Low-quality clusters and clusters expressing gene markers for 2 or more cell types were excluded from downstream analysis. Gene expression data was log-normalized and scaled using Seurat's 'NormalizeData' and 'ScaleData' functions. Principal component analysis (PCA) was performed with Seurat's 'RunPCA' function using variable features determined with Seurat's 'FindVariableFeatures' function. Motif activity scores were calculated using chromVAR²². To calculate differential motif accessibility between populations a logistic regression test was used with number of peak region fragments and MACS2 peak fragments as latent variables. Chromatin regions were annotated by CHIPseeker²¹, with pathway enrichment performed by GeneOntology enrichment analysis^{87,88}, utilizing genes or annotated genes with adjusted p-value<0.05 and absolute log₂ fold-change greater than 0.5.

Differentially expressed genes utilized in the construction of DEG-CRE-TF linkages for SSc versus control fibroblast and SSc versus control SPP1hi macrophages were obtained from our previously published dataset of 17 SSc-ILD and 13 control lung scRNA-seq samples³⁶.

ChromBPNet Analysis

Pseudobulk snATAC sequencing file generation: Fragment file containing the position of Tn5 integration site along with the cell barcodes for the relevant cell subpopulations were extracted using Signac. Single-nucleus ATAC sequencing BAM files for fibroblasts and SPP1hi macrophages were subsetted from the entire single-cell Multiome and ATAC-seq datasets using Sinto's 'filterbarcodes -b' function on the fragment file (<https://timoast.github.io/sinto>). Peak files were generated on the bam file using MACS2.

ChromBPNet takes the binary alignment matrix (BAM) files as input along with the peaks called from the pseudobulk BAM file and generates three important outputs. The first is the learned bias pattern from the inaccessible chromatin regions. Tn5 cleavage patterns in these regions are caused by enzyme cutting DNA sequence preferences. The second is the bias-corrected chromatin accessibility tracks that are generated after removing the learned bias from the accessibility signals. The third is, the contribution score which assigns an importance score for each base pair in the open chromatin regions. The score prioritizes base pairs that are predicted to contribute to the accessibility of the region where they are located and in designated cellular context.

Pseudo-bulk ATAC-seq profiles for SSc and control fibroblasts and corresponding SSP1 hi macrophages were used to train bias models on non-accessible genomic regions which were frozen and utilized for training secondary models on accessible (peak) genomic regions. After removing bias, corrected signal profiles for each of the four cell states were used to generate single-base pair resolution contribution scores. TF-MoDISco^{23,24} enabled de novo motif discovery from a sampling of 50,000 seqlets from each cell state which were annotated with JASPAR TF motifs. To complement TF-MoDISco, analysis peaks (30bp maximum) were called using the ChromBPNet contribution score tracks. Sequence contribution scores were visualized by dynseq tracks in the UCSC genome browser⁸⁹. The contribution score peaks were intersected with JASPAR motif coordinates²⁵ to enable comprehensive analyses of known TF motifs with high contribution scores occurring within accessible chromatin regions. Homer and Deeptool were used to scan for specific TFs motif in the genome using its PWM. The aligned genomic coordinates for a given TF motif were used to generate cluster-specific aggregate contribution scores. K-Means sub clustering enabled genomic regions to be identified with dynamic aggregate contribution

scores in control vs. SSc fibroblasts and in control vs. SSc SPP1^{hi} macrophages. These regions were associated with nearby genes using ChIPseeker.

Multi-Omics METAcell Construction

We utilized HALO, a variational autoencoder model³³, to infer latent representations from scRNA-seq and snATAC-seq data. Using the representations, a neighborhood graph was constructed using Scanpy (v1.9.3) to facilitate graph clustering and UMAP embedding visualization⁹⁰. Given the sparse nature of single-cell data, we aggregated individual cells exhibiting similar states into METAcells using COEM framework⁹¹. These METAcells were constructed based on cell-cell similarities inferred from the latent representations, which accounts for potential time-lag events between RNA-seq and ATAC-seq data.

Construction of TF-CRE-Gene Linkages

To identify cis-regulatory elements (CREs) from single-cell multiome data, we adopted the nonlinear regression-based framework, DIRECT-NET⁹². Specifically, we employed the nonlinear predictive model XGBoost to regress the expression level of a gene against the accessibility values of candidate distal CREs (defined as peaks within 250kbp upstream and downstream of the transcription start site). CREs were then selected based on the importance scores derived from the XGBoost model. CREs with the importance scores higher than the maximum of the median of importance scores and 0.001 were included for TF regulatory network construction.

To construct a disease-associated TF-gene regulatory network, we first identified differentially expressed genes (DEGs) ($\log_{2}FC > 0.5$ and $p_{adj} < 0.05$) across conditions (control vs. disease, or comparisons among cell states). We then identified differentially accessible CREs ($\log_{2}FC > 0.1$ and $p_{adj} < 0.05$) of these DEGs, ensuring that these differentially accessible CREs also had high importance scores. TFs in the differentially accessible CREs and promoter regions were identified using the JASPAR2020 motif database and the motifmatchr function in the ChromVAR package^{22,93}. Finally, we constructed a TF-gene regulatory network by integrating CRE-TF with CRE-gene relationships. The resultant network includes two types of regulatory linkages: TF-CRE-gene and TF-promoter-gene. If a SSc GWAS locus is located at CREs of DEGs, this information was overlaid onto the network [GWAS input: https://www.ebi.ac.uk/gwas/efotraits/EFO_0000717]. The network was visualized using Cytoscape software⁹⁴.

For RNA velocity trajectory analysis, the spliced and un-spliced RNA count matrices were generated using Cell Ranger output BAM files with Velocity CLI(v0.17.17)³⁸. These matrices were then fed into scVelo (v 0.3.1)⁹⁵ to compute RNA velocity with 'Dynamical' mode.

We evaluated differences in cell abundances associated with disease using the Milo framework for differential abundance testing³⁷, with its Python implementation, Milopy (<https://github.com/emdann/milopy>). Milo analysis was performed using the standard workflow, the KNN graph was generated from the latent representation of scRNA-seq and snATAC-seq data.

Acknowledgements

The authors would like to acknowledge the Center for Organ Recovery & Education (CORE) as well as the organ donors and their families for the generous donation of tissues used in this study. This research was supported in part by the University of Pittsburgh Center for Research Computing, RRID:SCR_022735, through the resources provided. Specifically, this work used the HTC cluster, which is supported by NIH award number S10OD028483.

Financial Support

Support for the studies was provided by NIH NIAMS P50AR080612 to RL and HS, DOD SL210018 to RL and HS, NIH NHLBI K08 HL161258 to EV, and from the Pulmonary Fibrosis Foundation, Francis Family Foundation, and National Scleroderma Foundation to EV.

Conflict of Interest

RL reports grants from Bristol Meyers Squibb, Formation, Moderna, Regeneron and Pfizer, EV reports grants from Boehringer Ingelheim. RL served or serves as a consultant with Abbvie, Mediar, Bristol Meyers Squibb, Formation, Thirona Bio, Sanofi, Boehringer Ingelheim, Merck, Genentech/Roche, EMD Serono, Morphic, Third Rock Ventures, Bain Capital and Zag Bio. RL sits on independent data safety monitoring committees for Advarra/GSK and Genentech. RL holds stock in Thirona Bio Inc and is president and holds stock in Modumac Therapeutics Inc.

Main Figures

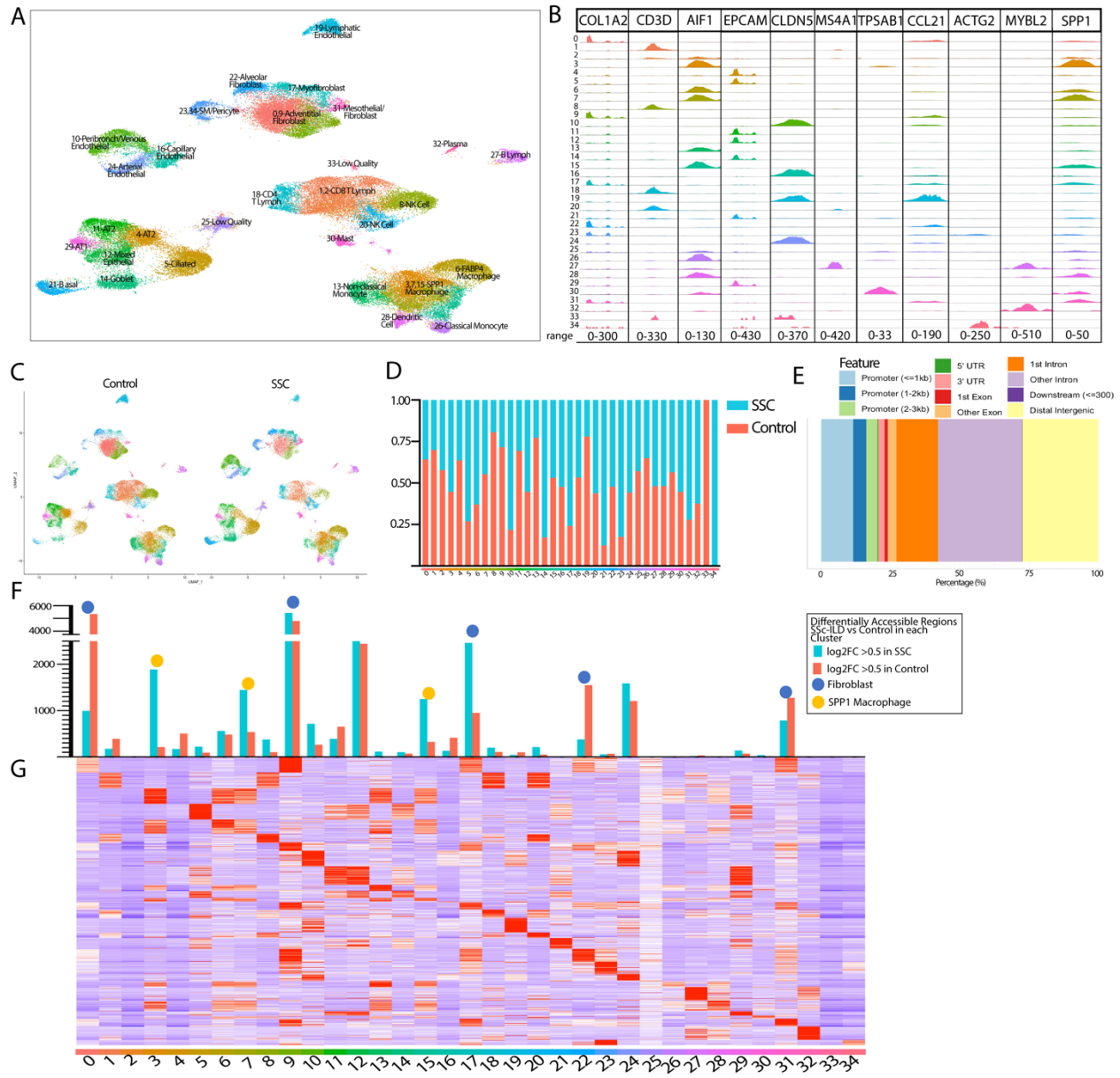


Figure 1.

A- Clustering of the 9 SSc and 9 Control samples after filtering, includes 67, 446 nuclei.

B- Identification of cell clusters by ATAC pseudobulk in promoter regions of known cell type identifying genes

C- Clustering divided by Control versus SSc origin

D- Proportion of cells in each cluster originating from SSc versus Control lungs

E- Annotation of peak regions for lung clustering. Peaks called per cluster by macs2 with annotation by ChIPseeker

F- Number of statistically significant (adj p-val < 0.05) differentially accessible regions when comparing SSc versus Control nuclei within each cluster

G- Top differentially accessible regions for each cluster

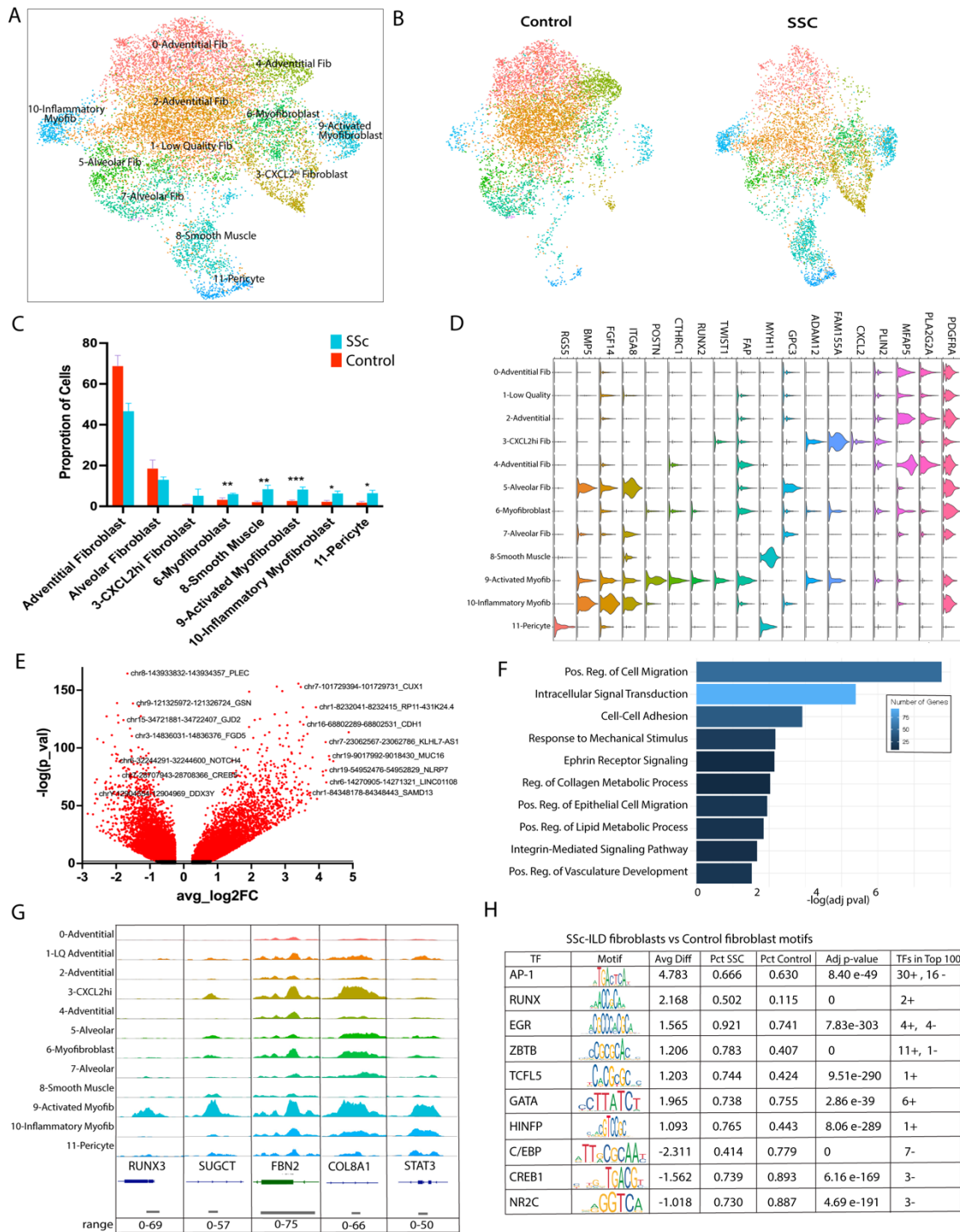


Figure 2.

A- Subclustering of mesenchymal nuclei

B- Mesenchymal nuclei divided by Control versus SSc

C- Proportion of mesenchymal nuclei consisting of each cell type at the individual sample level, comparing SSc vs Control. Bar represents the mean with error bar the standard error of the mean. *=0.0106-0.0142, **=p=0.0028-0.0056, ***=p=0.0005;

D- Gene expression of select cell type identifiers.

E- SSc Fibroblasts vs Control Fibroblasts DARs with labeling of select DARs and the nearest annotated gene

F- Pathway enrichment of OCRs more accessible in SSc vs Control fibroblasts

G- ATAC pseudobulk by mesenchymal cluster of 5 regions with top DARs in activated myofibroblasts

H- Top enriched TF family motifs in SSc-ILD fibroblasts vs Control fibroblasts, ranked by absolute value of average difference and separated into positively and negatively enriched. Final column notes number of TFs in the same class/family present in the top 100, as well as the direction of their enrichment (positive or negative)

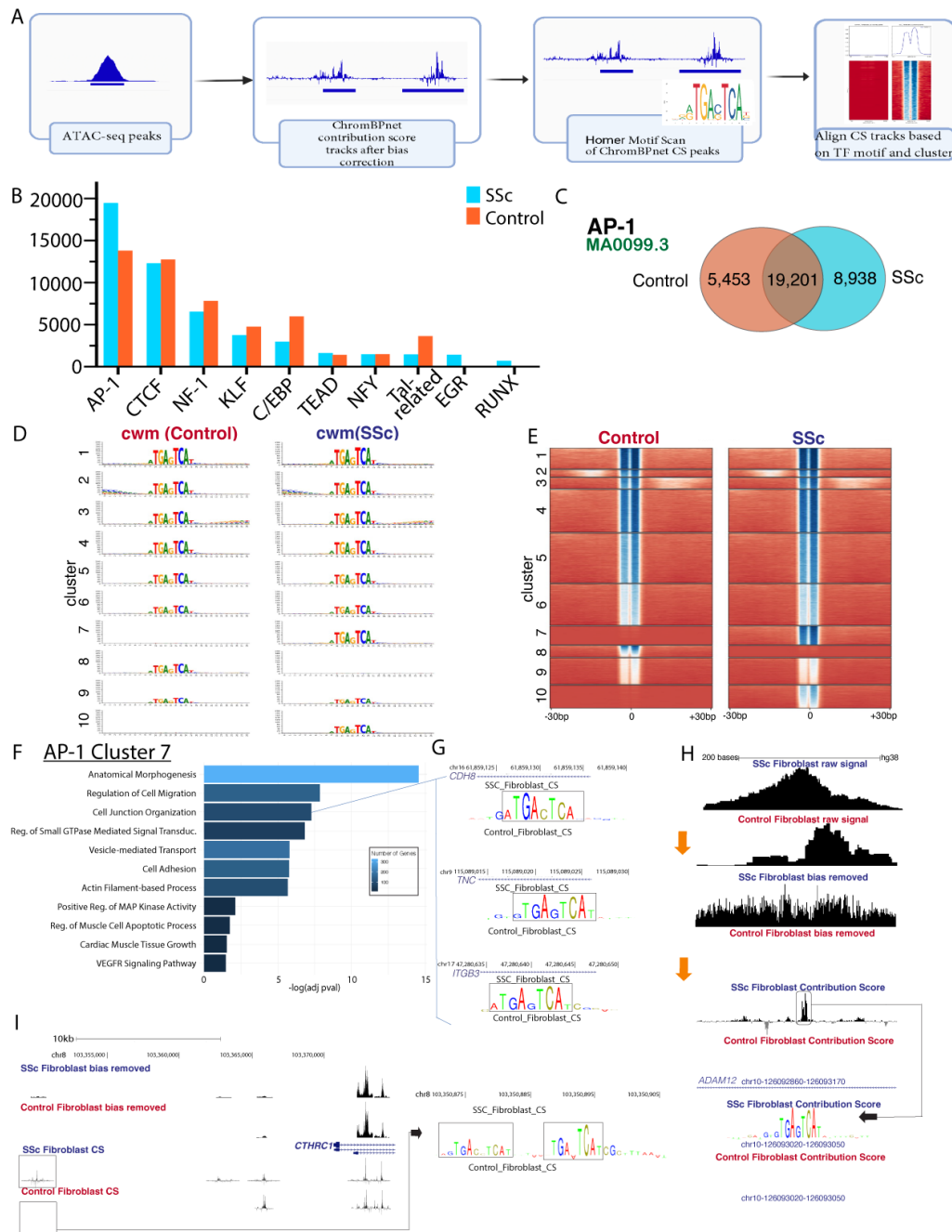


Figure 3. ChromBPNet analysis of 6880 control fibroblast nuclei and 5039 SSc fibroblast nuclei.

A- Graphic representation of analysis pipeline for ATAC pseudobulk ChromBPNet analysis

B- Number of seqlets detected by transcription factor family (largest number for each family depicted) in SSc and control fibroblasts by TF-MoDISco genome sampling.

C- Number of AP-1 seqlets (as detected for the MA0099.3 FOS::JUN matrix profile) unique to and shared by SSc and control fibroblasts.

D- Contribution weighted matrices for SSc and control fibroblasts by cluster

E- Heatmap of AP-1 contribution score dot product for each seqlet/chromatin region divided by SSc and control fibroblasts into 10 clusters. Darker blue indicates higher contribution score.

F- Pathways enriched by gene ontology for the AP-1 cluster 7 chromatin regions.

G- Contribution score track depictions for SSc and control fibroblasts of 3 chromatin regions (annotated to *CDH8*, *TNC*, and *ITGB3*) enriched in the cell junction organization pathway, from AP-1 cluster 7.

H- ChromBPNet analysis of *ADAM12* intronic region with AP-1 footprinting for SSc and control fibroblast pseudobulk ATAC data. Multiple levels from pseudobulk raw signal, bias removed signal, and contribution score tracks depicted.

I- Pseudobulk bias removed signal and contribution score tracks for *CTHRC1* region with duplicate AP-1 footprinting in SSc fibroblasts.

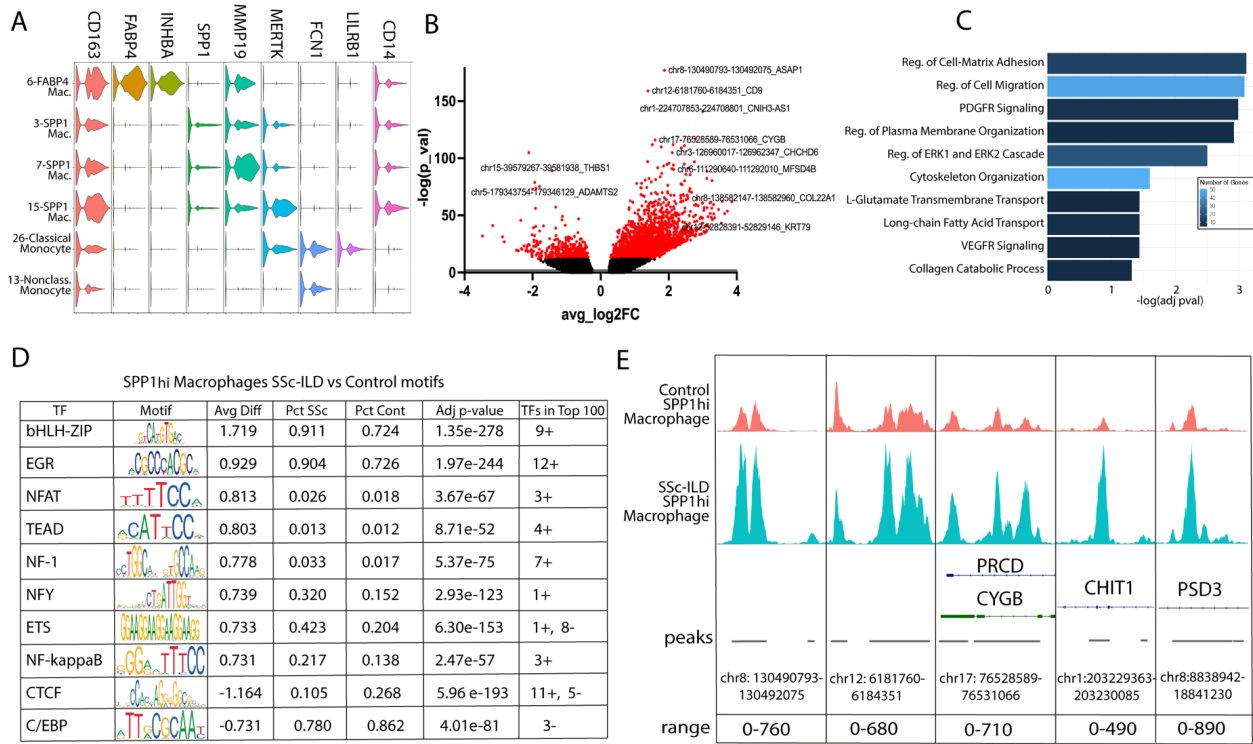


Figure 4.

A- Gene expression of myeloid cell type markers

B- SSc SPP1^{hi} Macrophages vs Control SPP1^{hi} Macrophages differentially accessible regions with labeling of select DARs and the nearest annotated gene.

C- Pathway enrichment analysis of chromatin regions more accessible in SSc vs Control SPP1^{hi} macrophages.

D-TF Motifs enriched in SSc SPP1^{hi} Macrophages vs Control SPP1^{hi} Macrophages, ranked by absolute value of average difference and separated into positively and negatively enriched. Final column notes number of TFs in the same class/family present in the top 100.

E- Plot of pseudobulk enriched peaks differentially accessible in SSc SPP1^{hi} Macrophages

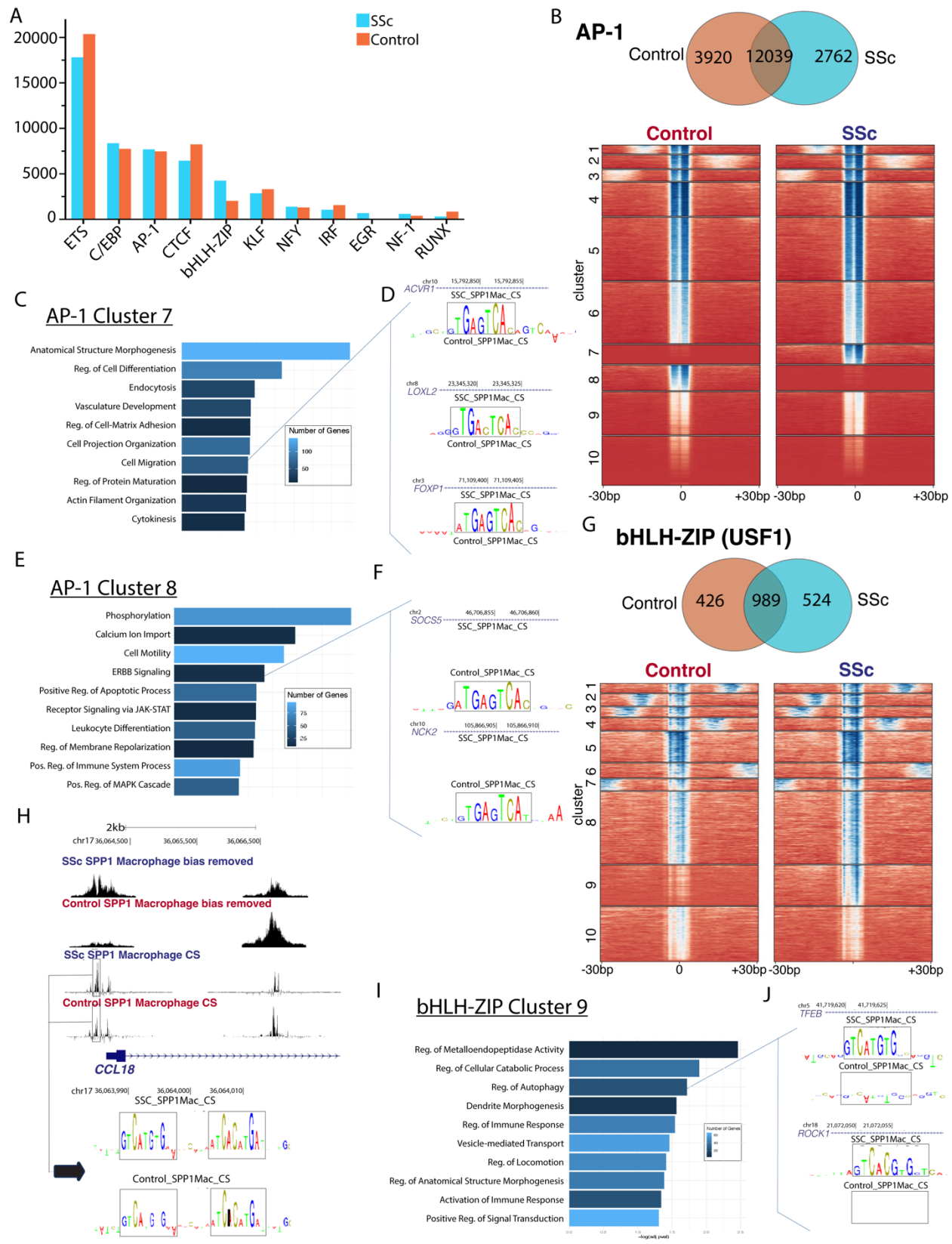


Figure 5. SPP1^{hi} Macrophage ChromBPNet Analysis of 3968 control nuclei and 3902 SSc nuclei.

A- Number of seqlets detected by transcription factor family (largest number for each family depicted) in SSc and control SPP1^{hi} Macrophages by TF-ModISco genome sampling.

B- Number of AP-1 seqlets unique to and shared by SSc and control SPP1^{hi} Macrophages. Heatmap of AP-1 contribution score dot product for each seqlet/chromatin region divided by SSc and control SPP1^{hi} Macrophages into 10 clusters. Darker blue indicates higher contribution score.

C- Pathways enriched by gene ontology for the AP-1 cluster 7 chromatin regions.

D- Contribution score track depictions for SSc and control SPP1^{hi} Macrophages of 3 chromatin regions (annotated to *ACVRI*, *LOXL2* and *FOXPI*) enriched in the cell migration pathway, from AP-1 cluster 7.

E- Pathways enriched by gene ontology for the AP-1 cluster 8 chromatin regions.

F- Contribution score track depictions for SSc and control SPP1^{hi} Macrophages of 2 chromatin regions (annotated to *SOCS5* and *NCK2*) enriched in the ERBB signaling pathway, from AP-1 cluster 8.

G- Number of bHLH-ZIP seqlets unique to and shared by SSc and control SPP1^{hi} Macrophages. Heatmap of bHLH-ZIP contribution score dot product for each seqlet/chromatin region divided by SSc and control SPP1^{hi} Macrophages into 10 clusters. Darker blue indicates higher contribution score.

H- Pseudobulk bias removed signal and contribution score tracks for *CCL18* region with duplicate bHLH-ZIP footprinting in SSc and control SPP1^{hi} Macrophages.

I- Pathways enriched by gene ontology for the bHLH-ZIP cluster 9 chromatin regions.

J- Contribution score track depictions for SSc and control SPP1^{hi} Macrophages of 2 chromatin regions (annotated to *TFEB* and *ROCK1*) enriched in the regulation of autophagy pathway, from bHLH-ZIP cluster 9.

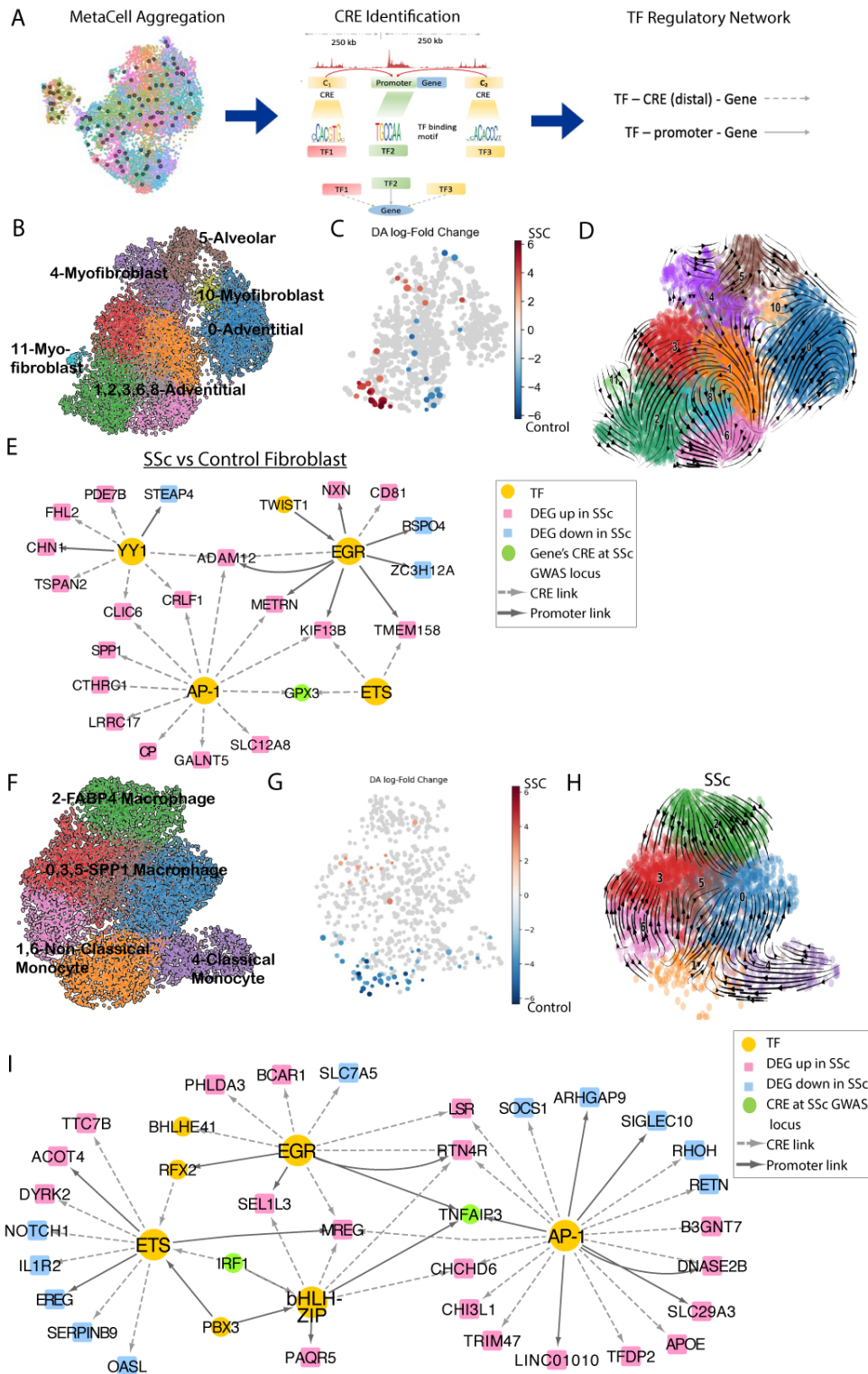


Figure 6.

A- Graphic representation of analysis pipeline for multiome mesenchymal and myeloid subsets: 1) Metacell creation using paired ATAC-seq and RNA-seq data to decrease sparsity, 2) Cis-regulatory elements identified by regressing the expression level of a gene of interest (DEGs for the same population comparisons) using accessibility values of peaks located within 250kbp of the transcription start site. Enriched TFs then identified in the differentially accessible CREs. 3)TF regulatory networks constructed by integrating both CRE-gene and CRE-TF relationships.

B- Multiome samples fibroblast clustering (smooth muscle, pericytes, and doublet nuclei excluded from this umap)

C- Visualization of Milo differential abundance testing results: each point represents a neighborhood, and points are colored according to the log fold change (logFC) in cell abundance between disease and control cells. Red indicates neighborhood enriched in SSc condition and blue indicates neighborhood enriched control condition

D- RNA velocity of multiome fibroblasts demonstrating suggested differentiation of myofibroblasts from both alveolar and adventitial subpopulations

E- Gene regulatory networks for top enriched TFs (ranked by adj p-value) in SSc fibroblasts vs Control fibroblasts. DEGs for this comparison were generated from a larger scRNA-seq dataset of 17 SSc and 13 control lungs. Only DEGs with $\text{abslog}_2\text{FC} > 1$ are depicted.

F- Multiome samples myeloid clustering with cell type identifications

G-Visualization of Milo differential abundance testing results. Red indicates neighborhood enriched in SSc condition and blue indicates neighborhood enriched control condition

H-RNA velocity of multiome SSc myeloid cells

I-Gene regulatory networks for top enriched TFs (ranked by adj p-value) in SSc SPP1^{hi} macrophages vs Control SPP1^{hi} macrophages. DEGs for this comparison were generated from a larger scRNA-seq dataset of 17 SSc and 13 control lungs. Only DEGs with $\text{abslog}_2\text{FC} > 1$ are depicted.

References

- 1 Fabre, T. *et al.* Identification of a broadly fibrogenic macrophage subset induced by type 3 inflammation. *Sci Immunol* **8**, eadd8945, doi:10.1126/sciimmunol.add8945 (2023).
- 2 Coulis, G. *et al.* Single-cell and spatial transcriptomics identify a macrophage population associated with skeletal muscle fibrosis. *bioRxiv*, doi:10.1101/2023.04.18.537253 (2023).
- 3 Bevan, L. *et al.* Specific macrophage populations promote both cardiac scar deposition and subsequent resolution in adult zebrafish. *Cardiovasc Res* **116**, 1357-1371, doi:10.1093/cvr/cvz221 (2020).
- 4 Rotem, I. *et al.* Osteopontin promotes infarct repair. *Basic Res Cardiol* **117**, 51, doi:10.1007/s00395-022-00957-0 (2022).
- 5 Shirakawa, K. *et al.* MerTK Expression and ERK Activation Are Essential for the Functional Maturation of Osteopontin-Producing Reparative Macrophages After Myocardial Infarction. *J Am Heart Assoc* **9**, e017071, doi:10.1161/JAHA.120.017071 (2020).
- 6 Berman, J. S. *et al.* Altered bleomycin-induced lung fibrosis in osteopontin-deficient mice. *Am J Physiol Lung Cell Mol Physiol* **286**, L1311-1318, doi:10.1152/ajplung.00394.2003 (2004).
- 7 Hatipoglu, O. F. *et al.* Osteopontin silencing attenuates bleomycin-induced murine pulmonary fibrosis by regulating epithelial-mesenchymal transition. *Biomed Pharmacother* **139**, 111633, doi:10.1016/j.biopha.2021.111633 (2021).
- 8 Ji, J. *et al.* Increased expression of OPN contributes to idiopathic pulmonary fibrosis and indicates a poor prognosis. *J Transl Med* **21**, 640, doi:10.1186/s12967-023-04279-0 (2023).
- 9 Ouyang, J. F. *et al.* Systems level identification of a matrisome-associated macrophage polarisation state in multi-organ fibrosis. *Elife* **12**, doi:10.7554/eLife.85530 (2023).
- 10 Steen, V. D. & Medsger, T. A. Changes in causes of death in systemic sclerosis, 1972-2002. *Ann Rheum Dis* **66**, 940-944, doi:10.1136/ard.2006.066068 (2007).
- 11 Christmann, R. B. *et al.* Association of Interferon- and transforming growth factor beta-regulated genes and macrophage activation with systemic sclerosis-related progressive lung fibrosis. *Arthritis Rheumatol* **66**, 714-725, doi:10.1002/art.38288 (2014).
- 12 Adams, T. S. *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* **6**, eaba1983, doi:10.1126/sciadv.aba1983 (2020).
- 13 Habermann, A. C. *et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv* **6**, eaba1972, doi:10.1126/sciadv.aba1972 (2020).
- 14 Valenzi, E. *et al.* Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease. *Ann Rheum Dis*, doi:10.1136/annrheumdis-2018-214865 (2019).
- 15 Valenzi, E. *et al.* Disparate Interferon Signaling and Shared Aberrant Basaloid Cells in Single-Cell Profiling of Idiopathic Pulmonary Fibrosis and Systemic Sclerosis-Associated Interstitial Lung Disease. *Front Immunol* **12**, 595811, doi:10.3389/fimmu.2021.595811 (2021).
- 16 Padilla, C. M. *et al.* Increased CD8+ tissue resident memory T cells, regulatory T cells, and activated natural killer cells in systemic sclerosis lungs. *Rheumatology (Oxford)*, doi:10.1093/rheumatology/kead273 (2023).
- 17 Kurche, J. S., Stancil, I. T., Michalski, J. E., Yang, I. V. & Schwartz, D. A. Dysregulated Cell-Cell Communication Characterizes Pulmonary Fibrosis. *Cells* **11**, doi:10.3390/cells11203319 (2022).

- 18 Brennan, K. J. *et al.* Chromatin accessibility in the Drosophila embryo is determined by transcription factor pioneering and enhancer activation. *Dev Cell* **58**, 1898-1916 e1899, doi:10.1016/j.devcel.2023.07.007 (2023).
- 19 Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333-1341, doi:10.1038/s41592-021-01282-5 (2021).
- 20 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296, doi:10.1038/s41592-019-0619-0 (2019).
- 21 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).
- 22 Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978, doi:10.1038/nmeth.4401 (2017).
- 23 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
- 24 Avsec, Z. *et al.* Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat Genet* **53**, 354-366, doi:10.1038/s41588-021-00782-6 (2021).
- 25 Castro-Mondragon, J. A. *et al.* JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **50**, D165-D173, doi:10.1093/nar/gkab1113 (2022).
- 26 Tabib, T. *et al.* Myofibroblast transcriptome indicates SFRP2(hi) fibroblast progenitors in systemic sclerosis skin. *Nat Commun* **12**, 4384, doi:10.1038/s41467-021-24607-6 (2021).
- 27 Rice, L. M. *et al.* A longitudinal biomarker for the extent of skin disease in patients with diffuse cutaneous systemic sclerosis. *Arthritis Rheumatol* **67**, 3004-3015, doi:10.1002/art.39287 (2015).
- 28 Rice, L. M. *et al.* Fresolimumab treatment decreases biomarkers and improves clinical symptoms in systemic sclerosis patients. *J Clin Invest* **125**, 2795-2807, doi:10.1172/JCI77958 (2015).
- 29 Tsukui, T. *et al.* Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat Commun* **11**, 1920, doi:10.1038/s41467-020-15647-5 (2020).
- 30 Joshi, S. *et al.* TEAD transcription factors are required for normal primary myoblast differentiation in vitro and muscle regeneration in vivo. *PLoS Genet* **13**, e1006600, doi:10.1371/journal.pgen.1006600 (2017).
- 31 Morse, C. *et al.* Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur Respir J*, doi:10.1183/13993003.02441-2018 (2019).
- 32 Hohne, K. *et al.* Pro-Fibrotic Effects of CCL18 on Human Lung Fibroblasts Are Mediated via CCR6. *Cells* **13**, doi:10.3390/cells13030238 (2024).
- 33 Mao, H. *et al.* HALO: Hierarchical Causal Modeling for Single Cell Multi-Omics Data. *bioRxiv*, doi: <https://doi.org/10.1101/2022.10.17.512602> (2024).
- 34 Persad, S. *et al.* SEACells infers transcriptional and epigenomic cellular states from single-cell genomics data. *Nat Biotechnol* **41**, 1746-1757, doi:10.1038/s41587-023-01716-9 (2023).
- 35 Baran, Y. *et al.* MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol* **20**, 206, doi:10.1186/s13059-019-1812-2 (2019).

- 36 Papazoglou, A. *et al.* Epigenetic Regulation of Profibrotic Macrophages in Systemic Sclerosis-Associated Interstitial Lung Disease. *Arthritis Rheumatol* **74**, 2003-2014, doi:10.1002/art.42286 (2022).
- 37 Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat Biotechnol* **40**, 245-253, doi:10.1038/s41587-021-01033-z (2022).
- 38 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498, doi:10.1038/s41586-018-0414-6 (2018).
- 39 Gao, X. *et al.* Osteopontin Links Myeloid Activation and Disease Progression in Systemic Sclerosis. *Cell Rep Med* **1**, 100140, doi:10.1016/j.xcrm.2020.100140 (2020).
- 40 Ponticos, M. *et al.* Failed degradation of JunB contributes to overproduction of type I collagen and development of dermal fibrosis in patients with systemic sclerosis. *Arthritis Rheumatol* **67**, 243-253, doi:10.1002/art.38897 (2015).
- 41 Schonhaler, H. B., Guinea-Viniegra, J. & Wagner, E. F. Targeting inflammation by modulating the Jun/AP-1 pathway. *Ann Rheum Dis* **70 Suppl 1**, i109-112, doi:10.1136/ard.2010.140533 (2011).
- 42 Reich, N. *et al.* The transcription factor Fra-2 regulates the production of extracellular matrix in systemic sclerosis. *Arthritis Rheum* **62**, 280-290, doi:10.1002/art.25056 (2010).
- 43 Avouac, J. *et al.* Inhibition of activator protein 1 signaling abrogates transforming growth factor beta-mediated activation of fibroblasts and prevents experimental fibrosis. *Arthritis Rheum* **64**, 1642-1652, doi:10.1002/art.33501 (2012).
- 44 Melms, J. C. *et al.* Author Correction: A molecular single-cell lung atlas of lethal COVID-19. *Nature* **598**, E2, doi:10.1038/s41586-021-03921-5 (2021).
- 45 Dulauroy, S., Di Carlo, S. E., Langa, F., Eberl, G. & Peduto, L. Lineage tracing and genetic ablation of ADAM12(+) perivascular cells identify a major source of profibrotic cells during acute tissue injury. *Nat Med* **18**, 1262-1270, doi:10.1038/nm.2848 (2012).
- 46 Chuang, L. S., Ito, K. & Ito, Y. RUNX family: Regulation and diversification of roles through interacting proteins. *Int J Cancer* **132**, 1260-1271, doi:10.1002/ijc.27964 (2013).
- 47 Dubey, S., Dubey, P. K., Umeshappa, C. S., Ghebre, Y. T. & Krishnamurthy, P. Inhibition of RUNX1 blocks the differentiation of lung fibroblasts to myofibroblasts. *J Cell Physiol* **237**, 2169-2182, doi:10.1002/jcp.30684 (2022).
- 48 Tsou, P. S., Palisoc, P. J., Ali, M., Khanna, D. & Sawalha, A. H. Genome-Wide Reduction in Chromatin Accessibility and Unique Transcription Factor Footprints in Endothelial Cells and Fibroblasts in Scleroderma Skin. *Arthritis Rheumatol* **73**, 1501-1513, doi:10.1002/art.41694 (2021).
- 49 Bauer, Y. *et al.* A novel genomic signature with translational significance for human idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* **52**, 217-231, doi:10.1165/rcmb.2013-0310OC (2015).
- 50 Fujimoto, Y. & Nakazawa, N. The roles of FHL2 as a mechanotransducer for cellular functions in the mechanical environment. *Front Cell Dev Biol* **12**, 1431690, doi:10.3389/fcell.2024.1431690 (2024).
- 51 Shi, M., Cui, H., Shi, J. & Mei, Y. Silencing FHL2 inhibits bleomycin-induced pulmonary fibrosis through the TGF-beta1/Smad signaling pathway. *Exp Cell Res* **423**, 113470, doi:10.1016/j.yexcr.2023.113470 (2023).
- 52 Bale, L. K. *et al.* Pregnancy-associated plasma protein-A (PAPP-A) is a key component of an interactive cellular mechanism promoting pulmonary fibrosis. *J Cell Physiol* **237**, 2220-2229, doi:10.1002/jcp.30687 (2022).

- 53 Qian, W. *et al.* Complex Involvement of the Extracellular Matrix, Immune Effect, and Lipid Metabolism in the Development of Idiopathic Pulmonary Fibrosis. *Front Mol Biosci* **8**, 800747, doi:10.3389/fmolb.2021.800747 (2021).
- 54 Liu, C., Tong, H., Li, S. & Yan, Y. Effect of ECM2 expression on bovine skeletal muscle-derived satellite cell differentiation. *Cell Biol Int* **42**, 525-532, doi:10.1002/cbin.10927 (2018).
- 55 Kim, M. S., Gernapudi, R., Choi, E. Y., Lapidus, R. G. & Passaniti, A. Characterization of CADD522, a small molecule that inhibits RUNX2-DNA binding and exhibits antitumor activity. *Oncotarget* **8**, 70916-70940, doi:10.18632/oncotarget.20200 (2017).
- 56 Illendula, A. *et al.* Small Molecule Inhibitor of CBFbeta-RUNX Binding for RUNX Transcription Factor Driven Cancers. *EBioMedicine* **8**, 117-131, doi:10.1016/j.ebiom.2016.04.032 (2016).
- 57 Hirose, Y. *et al.* Chb-M', an Inhibitor of the RUNX Family Binding to DNA, Induces Apoptosis in p53-Mutated Non-Small Cell Lung Cancer and Inhibits Tumor Growth and Repopulation In Vivo. *J Med Chem* **67**, 9165-9172, doi:10.1021/acs.jmedchem.4c00315 (2024).
- 58 Gomathi, K., Akshaya, N., Srinaath, N., Moorthi, A. & Selvamurugan, N. Regulation of Runx2 by post-translational modifications in osteoblast differentiation. *Life Sci* **245**, 117389, doi:10.1016/j.lfs.2020.117389 (2020).
- 59 Chen, S. J. *et al.* The early-immediate gene EGR-1 is induced by transforming growth factor-beta and mediates stimulation of collagen gene expression. *J Biol Chem* **281**, 21183-21197, doi:10.1074/jbc.M603270200 (2006).
- 60 Cho, S. J. *et al.* Role of early growth response-1 (Egr-1) in interleukin-13-induced inflammation and remodeling. *J Biol Chem* **281**, 8161-8168, doi:10.1074/jbc.M506770200 (2006).
- 61 Wu, M. *et al.* Essential roles for early growth response transcription factor Egr-1 in tissue fibrosis and wound healing. *Am J Pathol* **175**, 1041-1055, doi:10.2353/ajpath.2009.090241 (2009).
- 62 Bhattacharyya, S. *et al.* Egr-1 induces a profibrotic injury/repair gene program associated with systemic sclerosis. *PLoS One* **6**, e23082, doi:10.1371/journal.pone.0023082 (2011).
- 63 Ho, L. C. *et al.* Egr-1 deficiency protects from renal inflammation and fibrosis. *J Mol Med (Berl)* **94**, 933-942, doi:10.1007/s00109-016-1403-6 (2016).
- 64 Pritchard, M. T. & Nagy, L. E. Hepatic fibrosis is enhanced and accompanied by robust oval cell activation after chronic carbon tetrachloride administration to Egr-1-deficient mice. *Am J Pathol* **176**, 2743-2752, doi:10.2353/ajpath.2010.091186 (2010).
- 65 Fang, F. *et al.* The early growth response gene Egr2 (Alias Krox20) is a novel transcriptional target of transforming growth factor-beta that is up-regulated in systemic sclerosis and mediates profibrotic responses. *Am J Pathol* **178**, 2077-2090, doi:10.1016/j.ajpath.2011.01.035 (2011).
- 66 Fang, F. *et al.* Early growth response 3 (Egr-3) is induced by transforming growth factor-beta and regulates fibrogenic responses. *Am J Pathol* **183**, 1197-1208, doi:10.1016/j.ajpath.2013.06.016 (2013).
- 67 Huang, C. *et al.* Tetramerization of upstream stimulating factor USF2 requires the elongated bent leucine zipper of the bHLH-LZ domain. *J Biol Chem* **299**, 105240, doi:10.1016/j.jbc.2023.105240 (2023).
- 68 Agostini, F. *et al.* The Regulation of MiTF/TFE Transcription Factors Across Model Organisms: from Brain Physiology to Implication for Neurodegeneration. *Mol Neurobiol* **59**, 5000-5023, doi:10.1007/s12035-022-02895-3 (2022).

- 69 Park, K. *et al.* Lysosomal Ca(2+)-mediated TFEB activation modulates mitophagy and functional adaptation of pancreatic beta-cells to metabolic stress. *Nat Commun* **13**, 1300, doi:10.1038/s41467-022-28874-9 (2022).
- 70 Yu, S., Wang, Z., Ding, L. & Yang, L. The regulation of TFEB in lipid homeostasis of non-alcoholic fatty liver disease: Molecular mechanism and promising therapeutic targets. *Life Sci* **246**, 117418, doi:10.1016/j.lfs.2020.117418 (2020).
- 71 Larionova, I., Kazakova, E., Patysheva, M. & Kzhyshkowska, J. Transcriptional, Epigenetic and Metabolic Programming of Tumor-Associated Macrophages. *Cancers (Basel)* **12**, doi:10.3390/cancers12061411 (2020).
- 72 Mori, T., Tamura, N., Waguri, S. & Yamamoto, T. Autophagy is involved in the sclerotic phase of systemic sclerosis. *Fukushima J Med Sci* **66**, 17-24, doi:10.5387/fms.2019-28 (2020).
- 73 Cantanhede, I. G. *et al.* Exploring metabolism in scleroderma reveals opportunities for pharmacological intervention for therapy in fibrosis. *Front Immunol* **13**, 1004949, doi:10.3389/fimmu.2022.1004949 (2022).
- 74 Moss, C. E. *et al.* Aging-related defects in macrophage function are driven by MYC and USF1 transcriptional programs. *Cell Rep* **43**, 114073, doi:10.1016/j.celrep.2024.114073 (2024).
- 75 Dai, W. *et al.* Identification of LBH and SPP1 involved in hepatic stellate cell activation during liver fibrogenesis. *Hum Cell* **36**, 1054-1067, doi:10.1007/s13577-023-00889-4 (2023).
- 76 Prasse, A. *et al.* CCL18 as an indicator of pulmonary fibrotic activity in idiopathic interstitial pneumonias and systemic sclerosis. *Arthritis Rheum* **56**, 1685-1693, doi:10.1002/art.22559 (2007).
- 77 Gyorfi, A. H. *et al.* Performance of serum biomarkers reflective of different pathogenic processes in systemic sclerosis-associated interstitial lung disease. *Rheumatology (Oxford)* **63**, 962-969, doi:10.1093/rheumatology/kead332 (2024).
- 78 Khanna, D. *et al.* Safety and efficacy of subcutaneous tocilizumab in adults with systemic sclerosis (faSScinate): a phase 2, randomised, controlled trial. *Lancet* **387**, 2630-2640, doi:10.1016/S0140-6736(16)00232-4 (2016).
- 79 Lescoat, A. *et al.* Combined anti-fibrotic and anti-inflammatory properties of JAK-inhibitors on macrophages in vitro and in vivo: Perspectives for scleroderma-associated interstitial lung disease. *Biochem Pharmacol* **178**, 114103, doi:10.1016/j.bcp.2020.114103 (2020).
- 80 Atamas, S. P. *et al.* Pulmonary and activation-regulated chemokine stimulates collagen production in lung fibroblasts. *Am J Respir Cell Mol Biol* **29**, 743-749, doi:10.1165/rcmb.2003-0078OC (2003).
- 81 Wiertz, I. A. *et al.* Genetic Variation in CCL18 Gene Influences CCL18 Expression and Correlates with Survival in Idiopathic Pulmonary Fibrosis: Part A. *J Clin Med* **9**, doi:10.3390/jcm9061940 (2020).
- 82 Bridges, M. C., Daulagala, A. C. & Kourtidis, A. LNCcation: lncRNA localization and function. *J Cell Biol* **220**, doi:10.1083/jcb.202009045 (2021).
- 83 O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front Endocrinol (Lausanne)* **9**, 402, doi:10.3389/fendo.2018.00402 (2018).
- 84 Fueyo, R., Judd, J., Feschotte, C. & Wysocka, J. Roles of transposable elements in the regulation of mammalian transcription. *Nat Rev Mol Cell Biol* **23**, 481-497, doi:10.1038/s41580-022-00457-y (2022).

- 85 Valenzi, E. *et al.* Single-Nucleus Chromatin Accessibility Identifies a Critical Role for TWIST1 in IPF Myofibroblast Activity. *Eur Respir J*, doi:10.1183/13993003.00474-2022 (2023).
- 86 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
- 87 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).
- 88 Gene Ontology, C. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, doi:10.1093/genetics/iyad031 (2023).
- 89 Nair, S. *et al.* The dynseq browser track shows context-specific features at nucleotide resolution. *Nat Genet* **54**, 1581-1583, doi:10.1038/s41588-022-01194-w (2022).
- 90 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15, doi:10.1186/s13059-017-1382-0 (2018).
- 91 Mao, H., Jia, M., Dou, J. X., Zhang, H. & Benos, P. V. COEM: Cross-Modal Embedding for MetaCell Identification. *arXiv*, doi:<https://doi.org/10.48550/arXiv.2207.07734> (2022).
- 92 Zhang, L., Zhang, J. & Nie, Q. DIRECT-NET: An efficient method to discover cis-regulatory elements and construct regulatory networks from single-cell multiomics data. *Sci Adv* **8**, eabl7393, doi:10.1126/sciadv.abl7393 (2022).
- 93 Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92, doi:10.1093/nar/gkz1001 (2020).
- 94 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
- 95 Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408-1414, doi:10.1038/s41587-020-0591-3 (2020).