**SUPPLEMENTAL METHODS**

**Subjects and tissue preparation**

The University of Pittsburgh Institutional Review Board approved procedures involving human samples. Normal control lung tissue was obtained from organ donors, after rejection of the lungs for transplant. SSc-ILD lung tissue was obtained from explanted lungs at the time of lung transplantation. Following explantation of the lungs, sections from the upper and lower lobe were dissected, placed in Perfadex, and arrived for processing within twenty minutes of removal from the patient. Sections were allocated for formalin fixation, preservation in RNA Later, and immediate digestion for scRNA-seq. For the 13 samples included in the primary analysis, tissue for scRNA-seq was diced then enzymatically digested in DMEM (Thermo Fisher Scientific) containing 0.7 mg/mL collagenase A (Roche) and 30 ug/mL DNAase I (Roche) for one hour while undergoing further mechanical dispersion with the gentleMACS OctoDissociator (Miltenyi Biotec). The resulting cell suspension was washed with PBS, filtered twice through a 70-micron cell strainer, and underwent RBC lysis. Cells were then resuspended in PBS containing 0.04% BSA. ScRNA-seq library preparation was performed using the 10X Genomics Chromium System and its associated V2 chemistry reagents per the manufacturer’s protocol.

**Single-cell RNA library preparation and sequencing**

Using the 10X Genomics Chromium System, cells were mixed with reverse transcription reagents, loaded into a Single-cell A chip, followed by 3’ gel beads and partitioning oil. Cells were separated into oil micro-droplet partitions containing a cell and gel-bead scaffold for an oligonucleotide composed of oligo-DT, 10X and UMI barcodes, and reverse transcription reagents as described.[1] Reverse Transcription was performed, the emulsion broken, and pooled fractions obtained using a recovery agent. cDNAs were amplified by 11 cycles of PCR (C1000, Bio-Rad), enzymatically sheared and DNA fragment ends were repaired, A-tailed and adaptors ligated. The library was quantified with the KAPA Universal Library Quantification Kit KK4824 (KAPA Biosystems) and evaluated for cDNA length on a bioanalyzer using a High Sensitivity DNA kit (Agilent). ScRNA-seq libraries for the 13 samples in the primary analysis were sequenced on an Illumina NextSeq-500 through the University of Pittsburgh Genomics Core Sequencing Facility. Raw sequencing reads were examined by quality metrics and mapped to human reference genome GRCh38 using the Cell Ranger pipeline (10X Genomics). To ensure PCR amplified transcripts were counted only once, only single UMIs were counted for gene expression level.[2] In this way, cell x UMI count matrices were generated for downstream analysis.

**CITE-Seq and Cell Hashing with 10X Genomics single cell RNA-seq 3 Prime v3 chemistry**

For the three lung samples used for CITE-seq, after allocating sections for formalin fixation, the remaining tissue was diced then enzymatically digested per below. Single cell gene expression (GEX) with associated cell hashing (HTO) and CITE-seq (ADT) libraries were generated using 10X Genomics’ 3 Prime v3 chemistry and BioLegend TotalSeq-A antibodies. Gene expression libraries were constructed according to 10X Genomics’ manufacturer protocol while HTO and ADT libraries were generated using a modified version of the protocol developed by the Technology Innovation lab at the New York Genome Center[3, 4]. In brief, lungs were diced and digested in 3 mL of 1mg/mL Liberase DL (Roche) supplemented with 200 U/mL DNase I, Type IV (Sigma-Aldrich) for 35 minutes at 37C. For scleroderma lungs, upper and lower lobe sections were separately digested. Dissociated/digested tissue was then washed with cold PBS containing 2%FBS and filtered through 70 micron strainers twice. Approximately 1 million cells were then resuspended in 100 microliters of cold PBS containing 2% FBS. Ten microliters of Fc receptor block (TruStain FcX, BioLegend, USA) was added and cells were incubated for 20 minutes to block nonspecific antibody binding. Subsequently, cells were incubated with 0.5 μg of barcoded antibodies for 30 min at 4 °C. For scleroderma lungs, the upper and lower lobe sections were also incubated with separate TotalSeq-A human cell hashing antibodies. Cells were washed 3× with cold PBS containing 0.04% BSA (centrifuged at 400g 5 min at 4 °C). After the final wash, cells were resuspended in a final volume of 600 microliters of PBS containing 0.04% BSA, filtered through a 40-micron strainer, and counted. The 10x single-cell run was performed according to the manufacturer's instructions (10x Genomics, USA) with approximately 6,400 cells loaded per lung section to obtain a yield of ∼4,000 cells per lung section. Upper and lower lobe lung cells were pooled into the same well for a total of ~12,800 cells loaded. The cDNA was amplified for eleven cycles and following amplification products were subsequently size separated with SPRIselect (Beckman Coulter, USA) into <300 nt fragments containing antibody-derived tags (ADTs) and hashtag-oligos (HTOs) and >300 nt fragments containing cDNAs derived from cellular mRNA. The ADTs and HTOs were amplified for ten and twelve additional cycles, respectively, using specific primers that added P5 and P7 sequences. The gene expression library was constructed according to the manufacturer’s instructions. For sequencing, libraries were pooled at a ratio of 85% GEX to 10% ADT to 5% HTO. The pooled libraries were sequenced twice: once on an Illumina NovaSeq 6000 instrument through the UPMC Genome Center and again using an Illumina NextSeq-500 through the University of Pittsburgh Genomics Core Sequencing Facility. Raw data was demultiplexed using Cell Ranger 3.0.2’s mkfastq function and aligned to 10X Genomics’ human reference genome GRCh38 3.0.0 using cellranger count. Further analysis was performed in R using Seurat 2.3.4.

**Combined analysis of multiple samples**

To minimize batch effects in combining multiple samples for integrated analysis, an individual object was created for each sample then aligned using Seurat’s RunMultiCCA function.[5] Canonical correlation vectors were chosen for downstream analysis, canonical correlation analysis (CCA) subspaces aligned, and a single integrated analysis performed for tSNE visualization and clustering. The CCA method for combining multiple samples was compared with a principal component analysis (PCA) of this dataset as well. For the original clustering of all lung cells, PCA resulted in significantly more separation by sample, in particular amongst cell types with larger numbers of cells. However, when reclustering the combined SSc-ILD and control mesenchymal cells, CCA analysis separated cell types less accurately, likely due to the large differences in numbers of certain cell types present in SSc-ILD versus control samples. PCA analysis resulted in more accurate separation of cell types and was thus used for the included mesenchymal cell analysis. After clustering and visualization with tSNE, cell populations were identified through examination of gene markers in the associated transcriptome. Cell phase was predicted using CellCycleScoring in Seurat.

When reclustering limited cell types, a distinct group of cells emerged with only low-expression of a smaller number of genes and low average UMIs compared to the other clusters. Many of these cells also had low expression of markers of multiple cell types and many mitochondrial genes. These low expression clusters most likely represented poor quality cells that are a mix of empty droplets, damaged, and dying cells, and were removed from further analysis.[6]

**Gene ontology analysis**

Functional enrichment with Gene Ontology biological processes was performed using DAVID Bioinformatics Resources functional annotation version 6.8.[7, 8] All differentially expressed genes with adjusted p-value less than 0.05 and absolute value average log2 fold change greater

**Immunofluorescence imaging methods**

 Serial sections of formalin-fixed, paraffin-embedded human lung explants were deparaffinized and rehydrated for antibody staining. Slides were placed in citrate buffer pH 6, steamed for 20 minutes and cooled 20 minutes at room temperature for antigen retrieval before washing in phosphate buffered saline. For immunofluorescent staining (IF), the Tyramide SuperBoost kit (Invitrogen, USA) was used to amplify signals in co-stained tissues as per manufacturers protocol. Monoclonal mouse anti-SMA (1:1000; M0851; Clone14A; Dako Denmark AS, Denmark) was labeled with Alexa Fluor 488 and polyclonal rabbit anti-CTHRC1 (1:200 ab85739; Abcam, USA) was labeled with Alexa Fluor 568. All slides were counterstained with nuclear stain Hoechst 33342 (Invitrogen, USA). Masson’s trichrome staining was performed on serial sections by the University of Pittsburgh McGowan Institute Histology Core.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sample ID | Myofibroblasts (% of Total Fibroblasts) | ProliferatingMyofibroblasts(% of Total Fibroblasts) | MFAP5 Fibroblasts(% of Total Fibroblasts) | SPINT2 Fibroblasts (% of Total Fibroblasts) | Pericytes | Smooth Muscle |
| CONTROL 1 | 1 (3.45) | 1 (3.45) | 21 (72.41) | 6 (20.69) | 0 | 1 |
| CONTROL 2 | 2 (4.26) | 1 (2.13) | 7 (14.89) | 37 (78.72) | 12 | 25 |
| CONTROL 3 | 6 (11.11) | 0 (0) | 28 (51.85) | 20 (37.04) | 1 | 8 |
| CONTROL 4 | 41 (15.02) | 1 (0.37) | 96 (35.16) | 135 (49.45) | 3 | 14 |
| CONTROL 5 | 8 (10.81) | 2 (2.70) | 35 (47.30) | 29 (39.19) | 8 | 6 |
| SSC 1 | 33 (47.83) | 2 (2.90) | 29 (42.03) | 5 (7.25) | 21 | 75 |
| SSC 2 | 25 (40.32) | 0 (0) | 35 (56.45) | 2 (3.23) | 9 | 25 |
| SSC 3 | 145 (81.92) | 28 (15.82) | 4 (2.26) | 0 (0) | 76 | 2 |
| SSC 4 | 101 (71.63) | 22 (15.61) | 17 (12.06) | 1 (0.71) | 120 | 36 |
| SSC 5 | 230 (92.74) | 9 (3.63) | 8 (3.23) | 1 (0.40) | 57 | 51 |
| SSC 6 | 57 (98.28) | 0 (0) | 1 (1.72) | 0 (0) | 10 | 40 |
| SSC 7 | 18 (17.48) | 0 (0) | 18 (17.48) | 67 (65.05) | 49 | 71 |
| SSC 8 | 14 (13.86) | 0 (0 ) | 78 (77.23) | 9 (8.91) | 16 | 60 |

**Table 2. Mesenchymal Populations Cell Numbers and Proportions by Sample.**

Number of cells present from each sample in each mesenchymal cell population. For myofibroblasts, proliferating myofibroblasts, *MFA5*hi fibroblasts and *SPINT2*hi fibroblasts the proportion of total fibroblasts from each sample that the particular subpopulation comprises is listed in parentheses.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample** | **Sex** | **Age** | **Tissue Type** | **Pathology** | **mPAP (mmHg)** | **PVR (Woods units)** | **Immunosuppression** |
| Control 6 |  |  | Control | No abnormal tissue |  |  |  |
| SSC 9 | Female | 41 | SSc upper lobe | Unclassifiable chronic fibrosing interstitial pneumonia with prominent chronic lymphocytic inflammation | 49 | Not measured | azathioprine, hydroxychloroquine, prednisone |
| SSC 10 | Female | 41 | SSc lower lobe | Unclassifiable chronic fibrosing interstitial pneumonia with prominent chronic lymphocytic inflammation | 49 | Not measured | azathioprine, hydroxychloroquine, prednisone |
| SSC 11 | Male | 33 | SSc upper lobe | Predominately UIP with sections of NSIP | Not measured | Not measured | mycophenolate |
| SSC 12 | Male | 33 | SSc lower lobe | UIP | Not measured | Not measured | mycophenolate |

**Table 3. Characteristics of Patient Samples in Replication Cohort.**

Demographics, pathological review of adjacent tissue, and clinical characteristics of the patient samples included. Mean pulmonary artery pressure (mPAP) and pulmonary vascular resistance measurements (PVR) are from the last right heart catheterization preceding lung transplantation. Immunosuppression listed includes the medications received in the 90 days preceding lung transplantation.

|  |  |
| --- | --- |
| Sample ID | GEO Sample |
| CONTROL 1 | SC45 |
| CONTROL 2 | SC56 |
| CONTROL 3 | SC59 |
| CONTROL 4 | SC156 |
| CONTROL 5 | SC155 |
| CONTROL 6 | SC277 |
| SSC 1 | SC52 |
| SSC 2 | SC51 |
| SSC 3 | SC64 |
| SSC 4 | SC63 |
| SSC 5 | SC109 |
| SSC 6 | SC108 |
| SSC 7 | SC136 |
| SSC 8 | SC135 |
| SSC 9/10 | SC281 (HTO 3&4) |
| SSC 11/12 | SC284 (HTO 5&6) |

**Table 4.** Key for matching sample IDs from manuscript to sample designations in GEO

**References**

1. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017 Jan 16; 8:14049.

2. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods. 2014 Feb; 11(2):163-166.

3. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods. 2017 Sep; 14(9):865-868.

4. Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM, 3rd, et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol. 2018 Dec 19; 19(1):224.

5. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018 Jun; 36(5):411-420.

6. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al. Classification of low quality cells from single-cell RNA-seq data. Genome Biol. 2016 Feb 17; 17:29.

7. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009 Jan; 37(1):1-13.

8. Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, et al. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res. 2007 Jul; 35(Web Server issue):W169-175.