**METHODS**

**Single-cell RNA-sequencing.** SSc-ILD (n=13) and healthy control (HC, n=6) lung tissues obtained from the lung transplant service at University of Pittsburgh Medical Center. Control lungs were obtained from lungs rejected for transplant but otherwise healthy. Tissue were enzymatically digested and processed for single cell RNA-sequencing (scRNA-seq), using 3’ v2 and v3 kits (10X Genomics) as previously described (16). Demographic data on the SSc-ILD and HC patient samples is described in the supplement (Table S1). This research is approved by IRB 19100326-003.

**Transcriptomic data analysis.** Gene expression data derived from scRNA-seq were normalized and batch corrected using the Harmony function in Seurat before unsupervised UMAP clustering into general lymphocyte populations. These populations were further subclustered into specific lymphocyte subpopulations. To identify these subclusters, differentially expressed genes (DEGs) were determined for each group (Table S2). Using DEG, each subcluster of lymphoid cells was first identified through extensive literature review, pathway analysis, and use of online tools, including Human Enrichr, Human Atlas, and Gene Ontology.

Average gene expression by each lymphoid cluster was determined for both SSc-ILD and HC cohorts (Table S3). Gene Expression data for all samples can be found on Gene Expression Omnibus, GSE128169 and GSE212109. Seurat calculated MAST and Bonferroni corrections, providing corrected p-values by comparing one cluster to the other populations and by comparing SSc-ILD gene expressions in each cluster to those of HC. Absolute cell numbers and proportion of cells in each cluster were compared between SSc-ILD and HC, and statistical significance determined using Kruskal-Wallis test. Pearson’s R was used to see if a statistically significant correlation is observed between the age of the samples and the proportion of cluster 8 NK cells and cluster 5 CD8+ tissue resident memory T cells.

**Pathway Analysis & Pseudotime Analysis.** Pathway analysis was performed on DEGs of each subcluster derived from scRNA-seq analysis utilizing Ingenuity Pathway Analysis (QIAGEN Inc) and GO Enrichment Analysis. Pseudotime analysis was performed utilizing R package, Monocle 3, to characterize the trajectory within CD8+ T cell and NK cell subpopulations. The scRNA-seq lymphoid dataset was normalized and dimensionality reduced using UMAP. Then, utilizing approximate graph abstraction, Monocle 3, partitioned cells in these clusters to derive the differentiation trajectory of one cluster to another (17).

**Connectome Analysis.** R package connectome analysis was utilized to visualize patterns ligand-receptor interactions in single-cell datasets (18). Interaction between ligand and their receptors were explored between the lymphoid cell populations as well as between the other cell types represented in the lung, including endothelial, epithelial, and fibroblast cells. Single-cell data was first normalized and interactions detected using Connectome, limiting analysis to only those clusters with >75 cells captured. The data was then filtered to include only edges (interactions) with ligand and receptor z-scores above 0.25 and with both the ligand and receptor expressed in >10% of the cells in each cluster, respectively. CircosPlots were generated to visualize all edges originating cell subpopulation of interest as well as all edges (interaction) received from various cell clusters onto cell subpopulation of interest.

**Immunohistochemistry with Granzyme B and CD56**
The paraffin-embedded lung slides underwent dewaxing and antigen retrieval in citrate buffer pH 6.0 (ThermoFisher, Cat #: 005000). Subsequently, blocking steps were performed using 3% H2O2 and 10% goat serum (ThermoFisher, Cat #: 50062Z). The target antibody, either anti-granzyme B (Abcam, ab134933, 1:50) or anti-CD56 (ThermoFisher, 07-5603, 1:200), was added for overnight incubation at 4°C. For both single and double staining, the ThermoFisher Tyramide SuperBoost™ kit (Alexa Fluor 488 or 568) was used. An Olympus FluoView 3000 confocal microscope was used for imaging the slides.

**Cytokine treatment of pulmonary fibroblasts.** Primary cultures of fibroblasts from two patients with SSc-ILD lung, cultured in DMEM supplemented with 10% FBS, were serum starved for 24 hours and then treated with PDGF-DD (R&D Systems), PDGF-AB (R&D Systems), or left untreated for 48 hours. Cells were lysed in RNAzol, RNA prepared (Qiagen) and RT-PCR was carried out using COL1A1: Hs00164004\_m1; COL10A1: Hs00166657\_m1; ACTA2: Hs00426835\_g1 primer sets(ThermoFisher).

**CITE-seq.** For CITE-seq experiments, cryopreserved lung cells selected on CD45/EPCAM/CD31+ MicroBeads (Miltenyi) from 2 normal lung, 2 SSc-ILD lungs

were thawed, washed, resuspended in 2%FBS PBS, and incubated with 10 microliters of Human Fc Block /100 microliter cell suspension and incubated for 10 minutes

Cell suspension were then incubated with TotalSeq-C human hashtags for 30 minutes at 4 degrees (1 microliter per sample), washed 3 times, filtered through a 40 micron strainer, counted and Pooled (125,000 live cells per sample). The pool was centrifuged and resuspended with 10 microliters of FcBlock (Biolegend, Human TruStain FcX™) for 10 minutes. Cells were then incubated 30 minutes with Custom TotalSeq-C CITE-Seq panel (Biolegend, cocktail 99813), washed and 35000 cells were mixed with reverse transcription reagents and loaded into Chromium instrument, aiming to recover 20000 total cells (5000 per sample).

The gene expression and Feature Barcode libraries were then constructed according to the manufacturer's directions (10X Genomics User Guide CG000330 and Chromium Single Cell 5' Reagent Kits User Guide (v2 Chemistry Dual Index) with Feature Barcoding technology for Cell Surface Protein and Immune Receptor Mapping. Cellranger-7.1.0 was used to align FASTQ files to references. Gene expression data was aligned to 10X Genomics' human reference: refdata-gex-GRCh38-2020-A, see also (1).

1. Valenzi E, Bulik M, Tabib T, Morse C, Sembrat J, Trejo Bittar H, et al. Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease. Ann Rheum Dis. 2019;78(10):1379-87.