Supplementary information
for

Single Nucleus RNASeq Profiling of Mouse Lung: Reduced Dissociation Bias and Improved Rare Cell Type Detection Compared with Single Cell RNASeq

Jeffrey R Koenitzer, Haojia Wu, Jeffrey J Atkinson, Steven L Brody, Benjamin D Humphreys

Benjamin D. Humphreys
Division of Nephrology
Washington University School of Medicine
660 S. Euclid Ave., CB 8129
St Louis, MO 63110
humphreysbd@wustl.edu

This PDF file includes: Figures E1-E12, Supplemental Methods

Figure E1

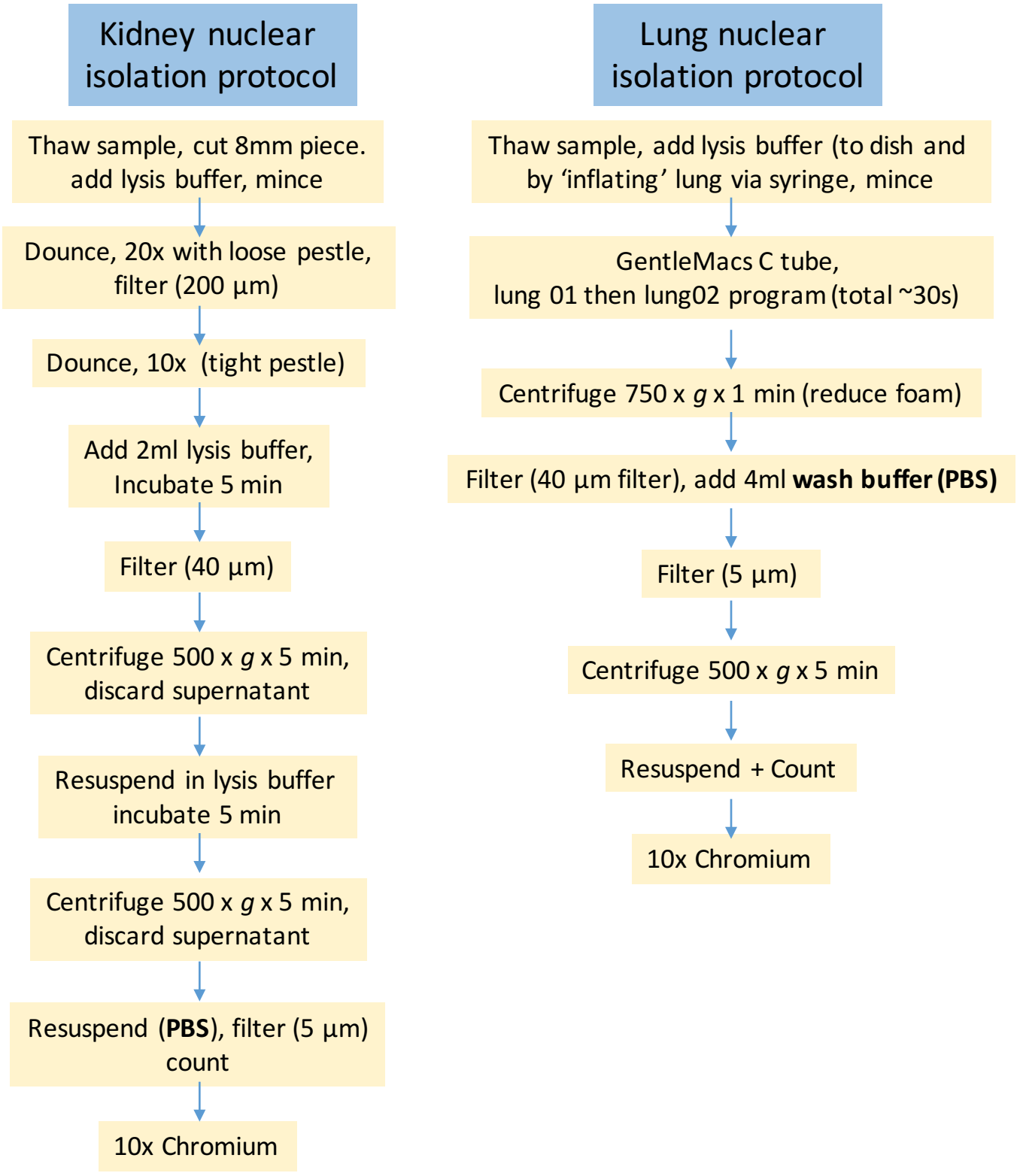

Figure E1. Comparison of kidney and lung single nucleus isolation protocols. (Left) Kidney isolation involves Dounce homogenization and the majority is performed in the presence of lysis buffer, with multiple incubations and centrifugations. (Right) the protocol for lung nuclear isolation employs GentleMacs dissociation and lysis buffer is quickly diluted with saline wash buffer. The protocol is shorter, with one 5 minute centrifugation and no incubation steps. All steps performed at 4°C or on ice, all buffers contain RNase inhibitors.

Figure E2

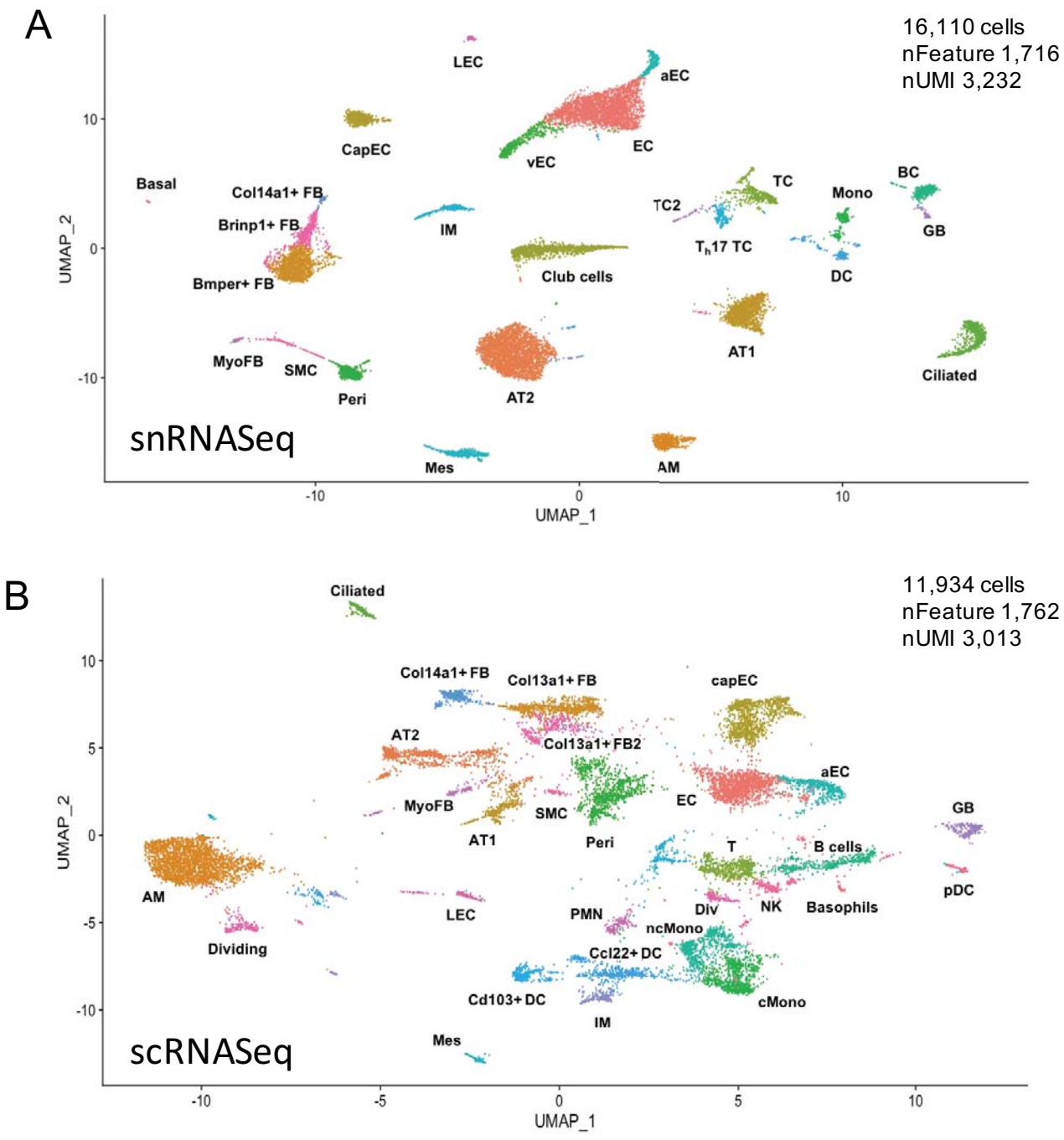

Figure E2. Annotated uMAP plots of snRNASeq versus scRNASeq. (A) Combined snRNASeq data from two replicates, with 26 cell types. (B) Single cell RNASeq data from one mouse, with 29 cell types. Similar gene and UMI detection rates per cell were observed. EC, endothelial cells; aEC, arterial endothelial cells; LEC, lymphatic endothelial cells; AT1, alveolar type 1 epithelial cells; AT2, alveolar type 2 epithelial cells; BC, B cells; GB, germinal B cells; TC, T cells; DC, dendritic cells; FB, fibroblasts; MyoFB, myofibroblasts; AM, alveolar macrophages; IM, interstitial macrophages; Peri, pericytes; cMono, classical monocytes; ncMono, nonclassical monocytes; NK, natural killer cells; Mes, mesothelial cells; SMC, smooth muscle cells. pDC, plasmacytoid dendritic cells

Figure E3

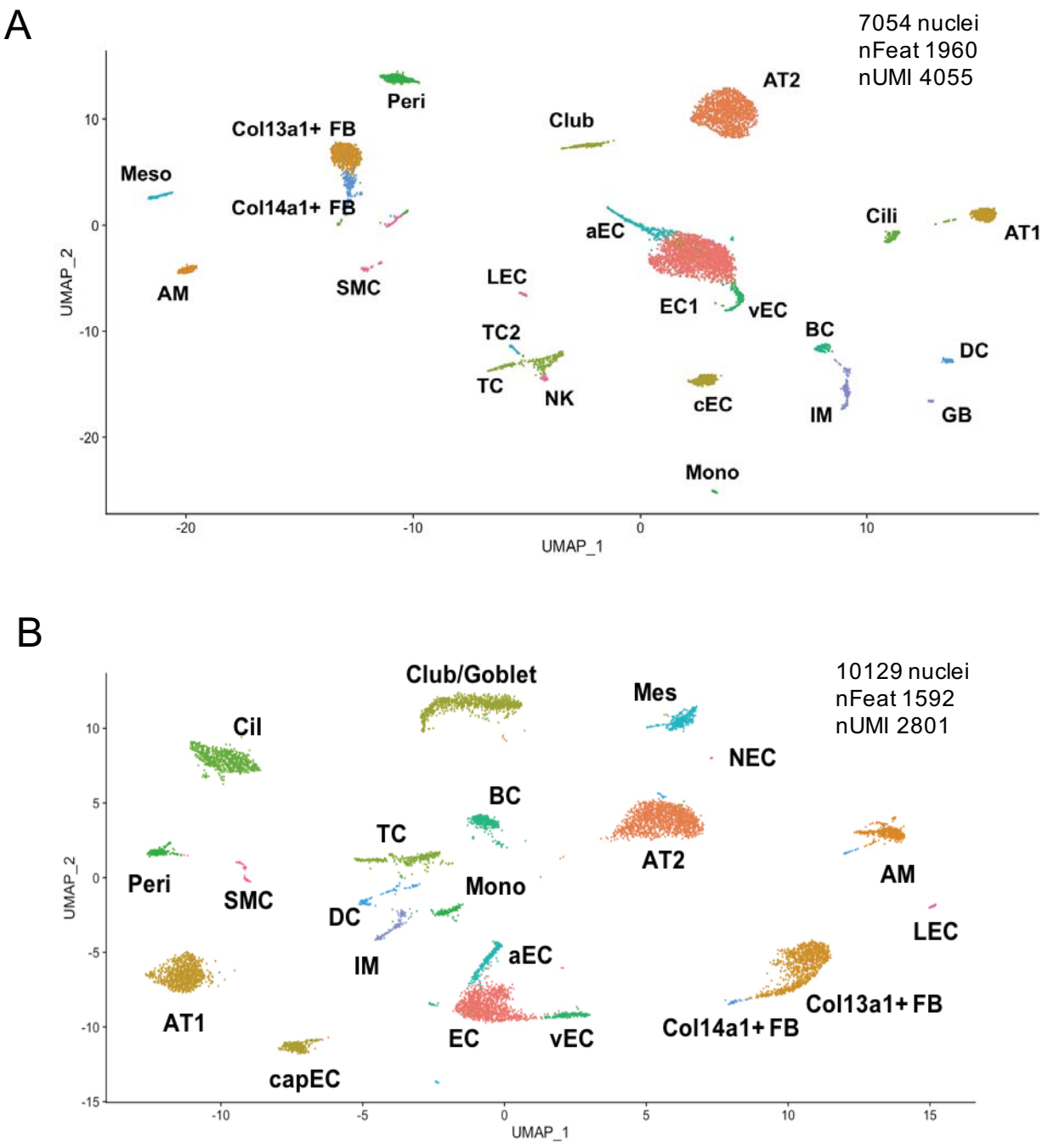

Figure E3. Annotated uMAPs for individual snRNASeq studies on control mouse lung. Overall similar gene and cell type detection between experiments, with the first (A) improved in terms of genes and UMI detected per nucleus. Neuroendocrine cells were identified in one of two runs (B). EC, endothelial cells; LEC, lymphatic endothelial cells; AT1, alveolar type 1 epithelial cells; AT2, alveolar type 2 epithelial cells; BC, B cells; GB, germinal B cells; TC, T cells; DC, dendritic cells; FB, fibroblasts; AM, alveolar macrophages; IM, interstitial macrophages; Peri, pericytes; Mono, monocytes; NK, natural killer cells; Mes, mesothelial cells; SMC, smooth muscle cells; NEC, neuroendocrine cells; nFeat, number of features per nucleus; nUMI, number of UMI per nucleus.

Figure E4

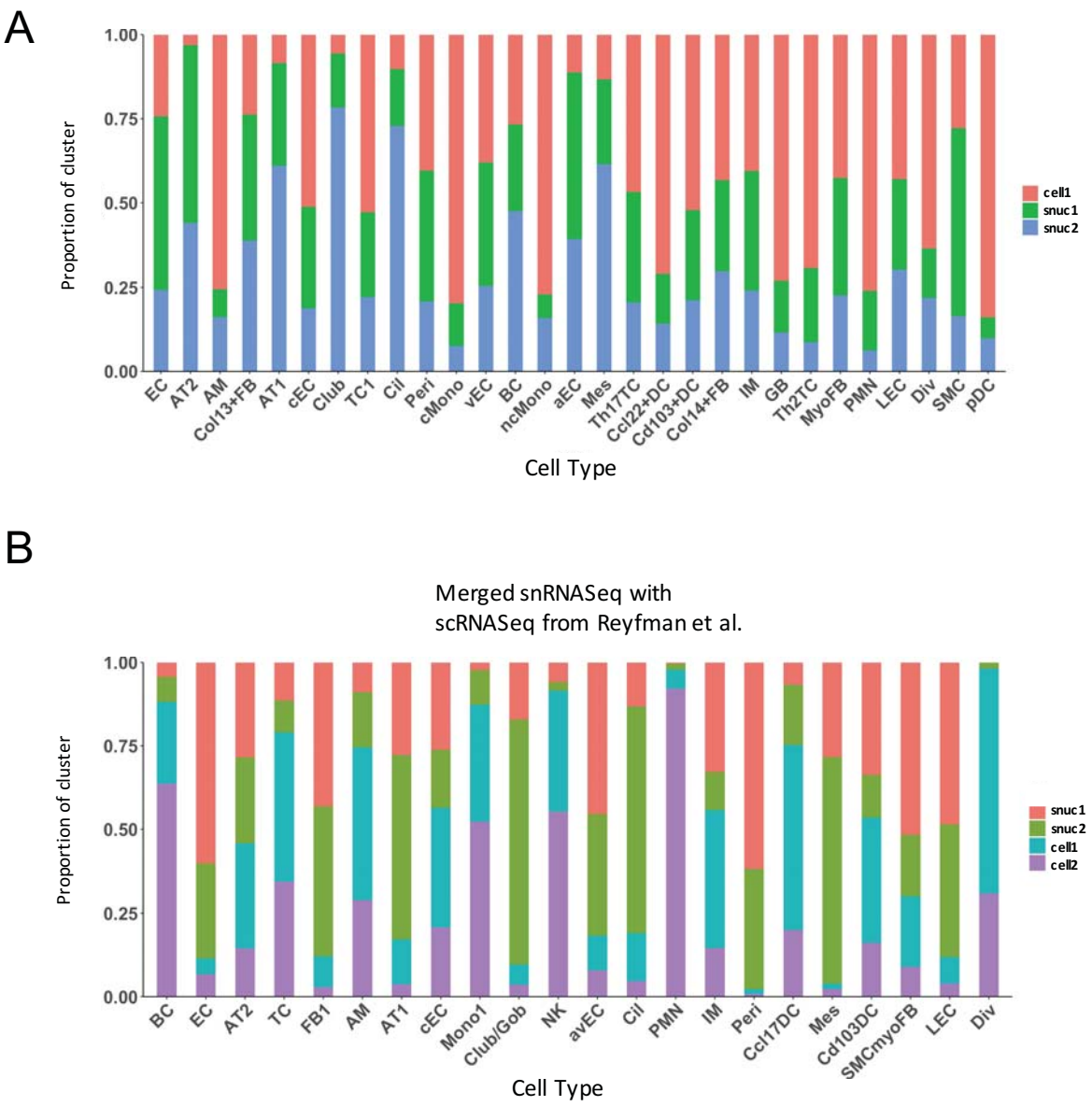

Figure E4. Cell type ratios in snRNASeq vs scRNASeq. Objects were randomly downsampled to contain equal cell numbers in each replicate. (A) Cell type distributions from combined snRNASeq and scRNASeq data, by replicate. (B) Cell type distributions by replicate after combination of snRNASeq datasets with two scRNASeq datasets from Reyfman et al. EC, endothelial cells; AT2, alveolar type 2 epithelial cells; AM, alveolar macrophages; FB, fibroblasts; AT1, alveolar type 1 epithelial cells; cEC, capillary endothelial cells; TC, T cells; Cil, ciliated cells; Peri, pericytes; cMono, classical monocytes; vEC, venous endothelial cells; BC, B cells; ncMono, nonclassical monocytes; aEC, arterial endothelial cells; Mes, mesothelial cells; DC, dendritic cells; IM, interstitial macrophages; GB, germinal B cells; MyoFB, myofibroblasts; PMN, neutrophils; LEC, lymphatic endothelial cells; Div, dividing cells; SMC, smooth muscle cells; pDC, plasmacytoid dendritic cells; avEC, arterial and venous endothelial cells; NK, natural killer cells.

Figure E5

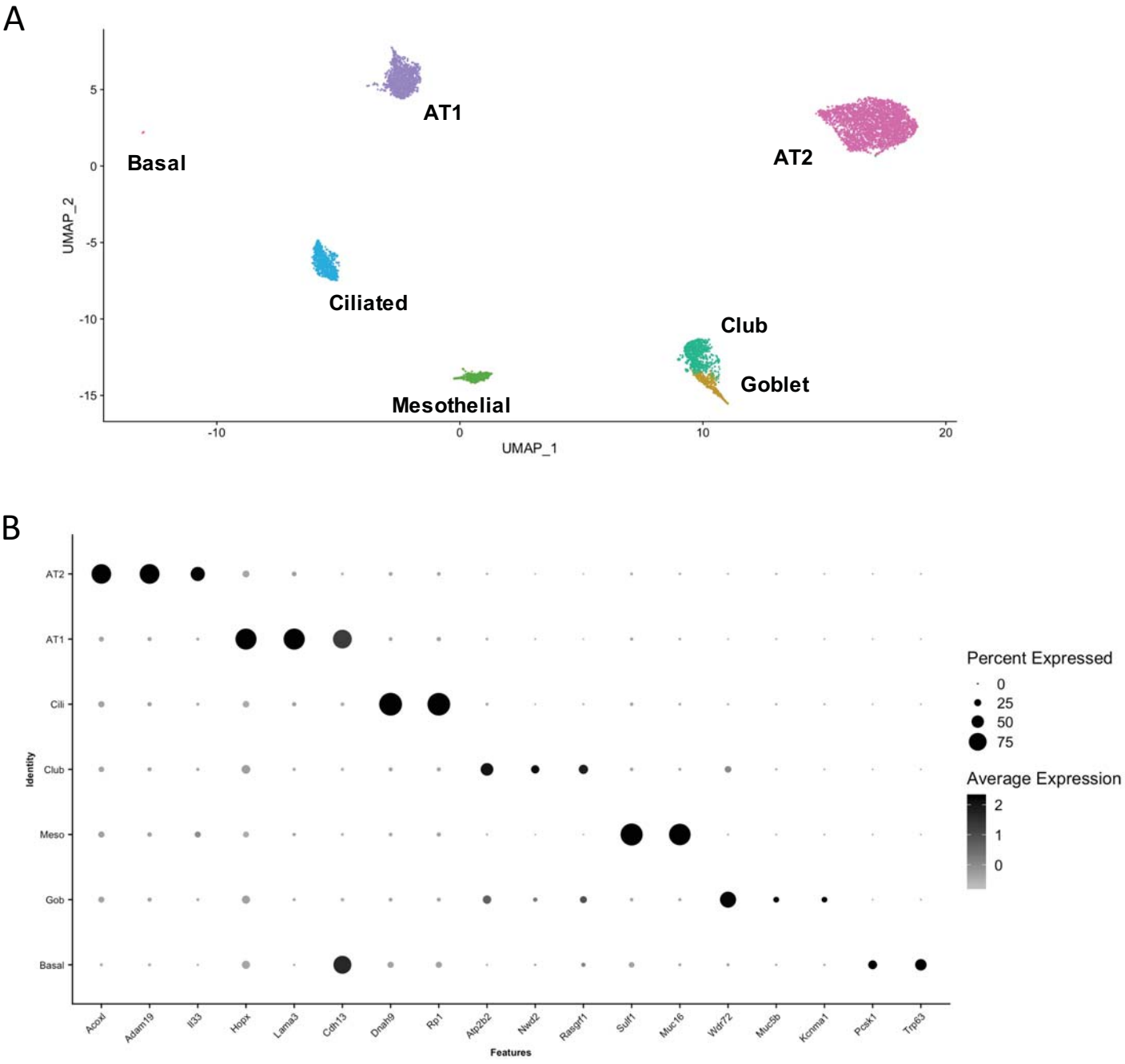

Figure E5. Epithelial subpopulations in snRNASeq data. (A) uMAP of epithelial cell types from combined snRNASeq data including basal and goblet cells. (B) Marker genes for each of the epithelial clusters. AT1, alveolar type 1; AT2, alveolar type 2; Cili, ciliated; Meso, mesothelial; Gob, goblet.

Figure E6

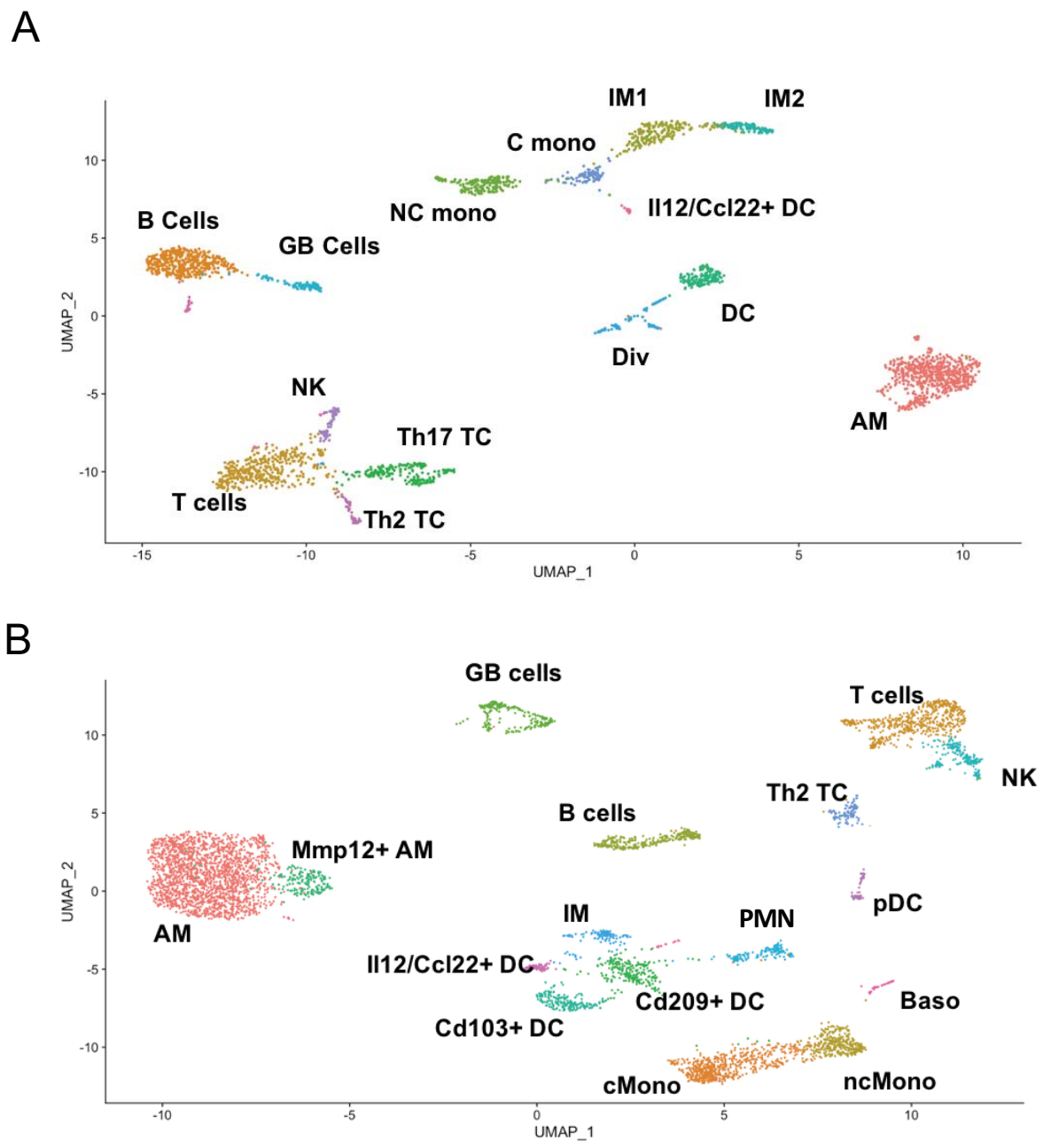

Figure E6. Immune cell sub-clustering in snRNASeq and scRNASeq. (A) uMAP of immune populations from combined single nucleus data, with 14 cell types including subpopulations of interstitial macrophages and Th17 T cells. (B) Immune populations from scRNASeq, with improved resolution of dendritic cell subtypes, neutrophils, and basophils. AM, alveolar macrophages; IM, interstitial macrophages; DC, dendritic cells; cMono, classical monocytes; ncMono, nonclassical monocytes; Baso, basophils; PMN, neutrophils; pDC, plasmacytoid dendritic cells; NK, NK cells; GB, germinal B cells; TC, T cells; Th2, T helper 2; Th17, T helper 17.

Figure E7

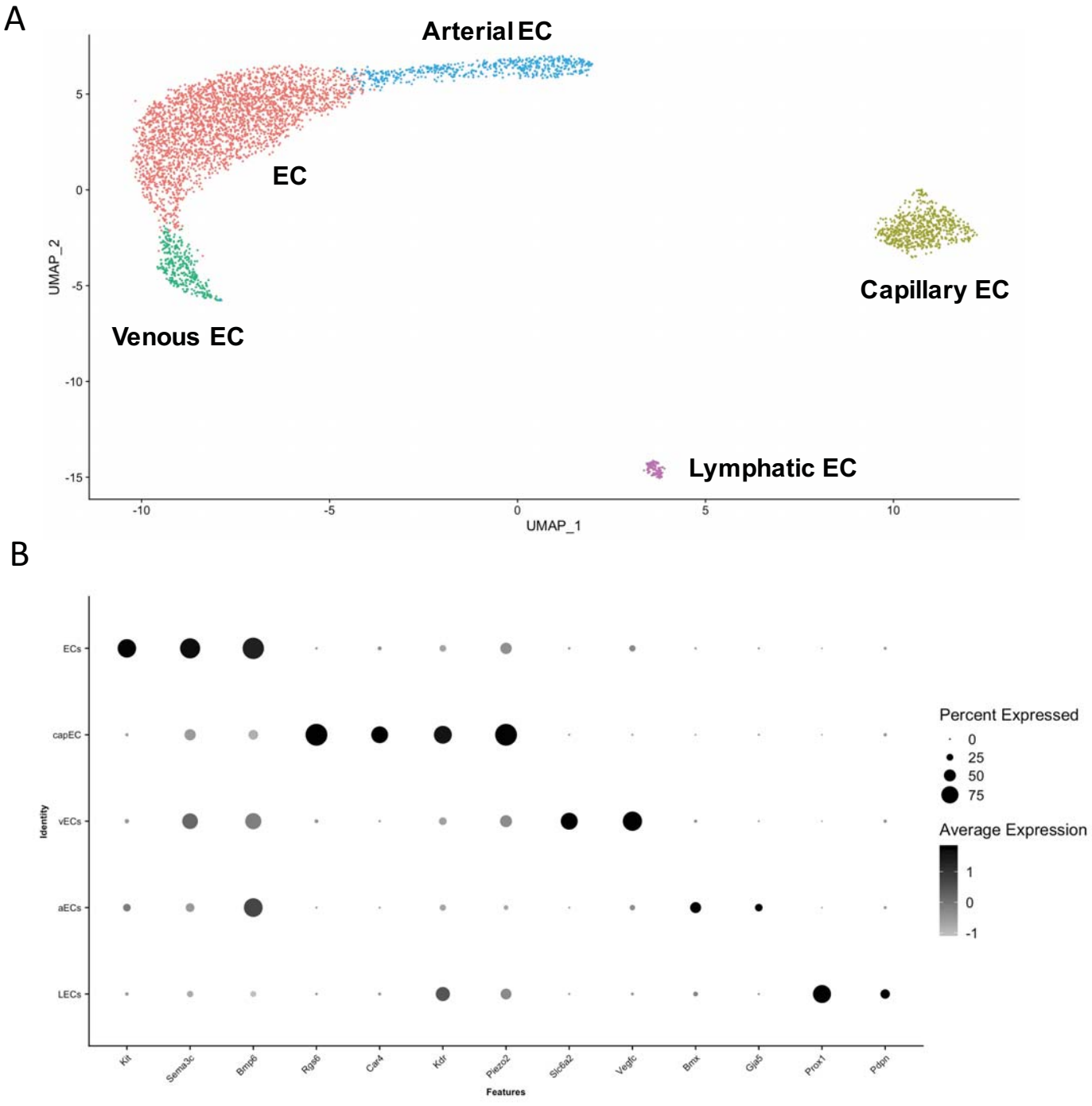

Figure E7. Endothelial subpopulations in snRNASeq data. (A) uMAP of endothelial cell types from combined snRNASeq data including distinct arterial, venous, lymphatic, and capillary endothelial cells (EC). (B) Dot plot of marker genes for subtypes of endothelial cells. capEC, capillary endothelial cells; vECs, venous endothelial cells; aECs, arterial endothelial cells; LECs, lymphatic endothelial cells.

Figure E8

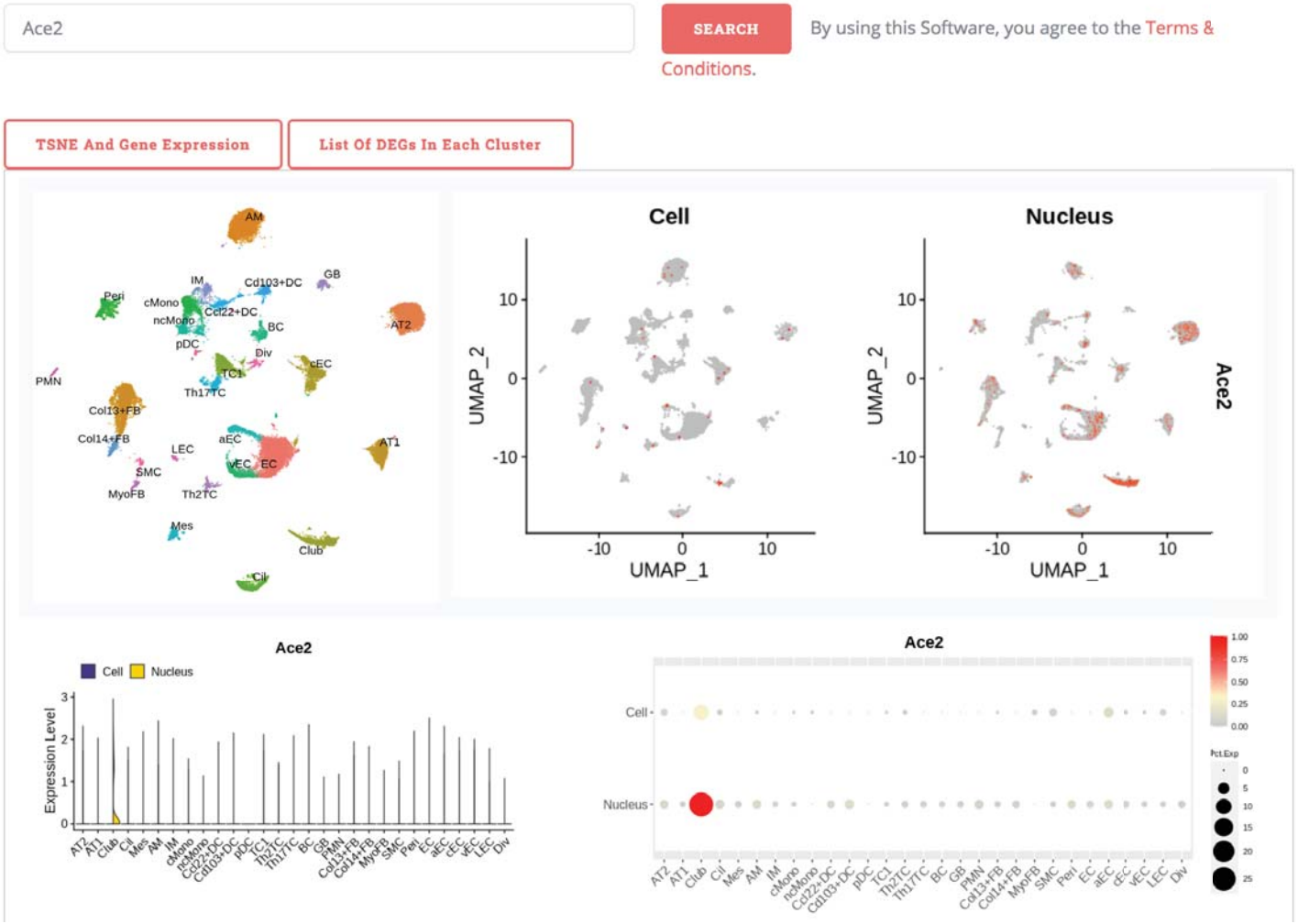

Figure E8. Sample of KIT website output. Combined scRNASeq and snRNASeq data from control mouse lung is available online and searchable at humphreyslab.com/SingleCell. Output is split by cell and nucleus expression for each cell type, and displayed in uMAP, violin plot, and dot plot formats.

Figure E9

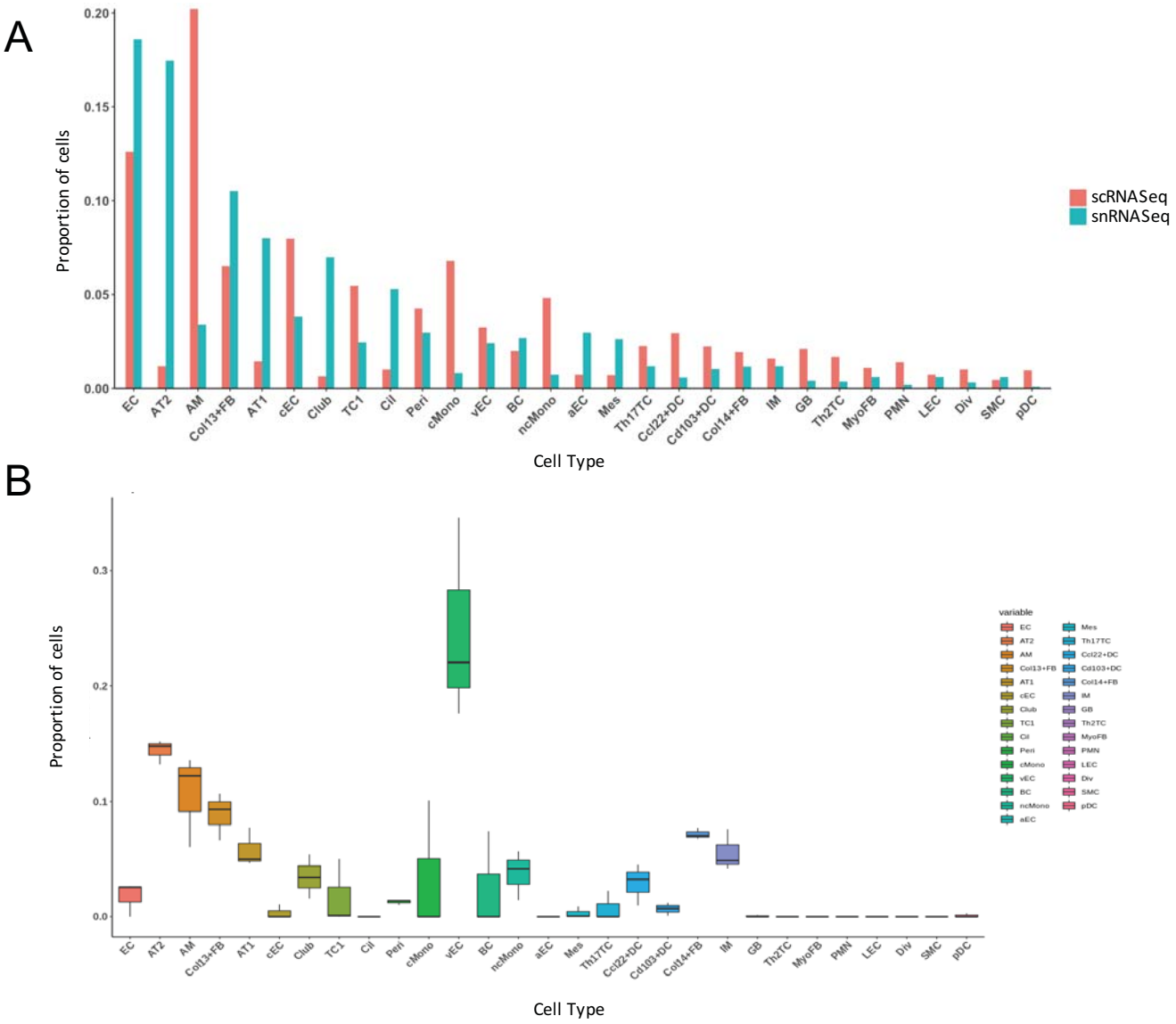

Figure E9. Bulk RNASeq deconvolution and comparison to cell type ratios in snRNASeq and scRNASeq. (A) Bar chart of cell types from snRNASeq and scRNASeq data, as proportions of total cells. (B) Box plot of cell type proportions from deconvolution of bulk RNASeq data, inferred from markers in snRNASeq+scRNASeq. EC, endothelial cells; AT2, alveolar type 2 epithelial cells; AM, alveolar macrophages; FB, fibroblasts; AT1, alveolar type 1 epithelial cells; cEC, capillary endothelial cells; TC, T cells; Cil, ciliated cells; Peri, pericytes; cMono, classical monocytes; vEC, venous endothelial cells; BC, B cells; ncMono, nonclassical monocytes; aEC, arterial endothelial cells; Mes, mesothelial cells; DC, dendritic cells; IM, interstitial macrophages; GB, germinal B cells; MyoFB, myofibroblasts; PMN, neutrophils; LEC, lymphatic endothelial cells; Div, dividing cells; SMC, smooth muscle cells; pDC, plasmacytoid dendritic cells.

Figure E10

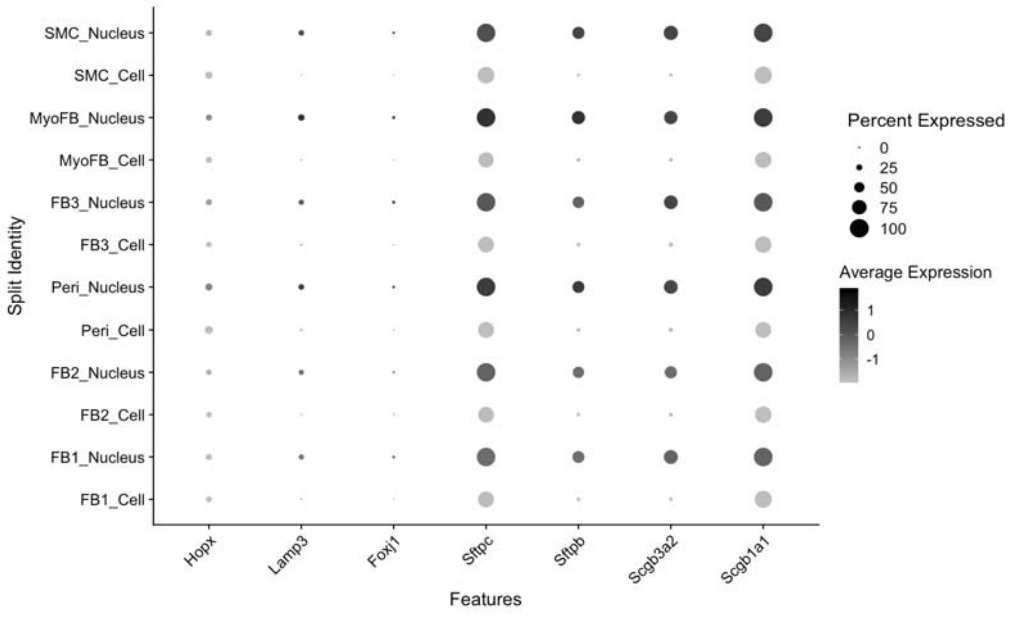

Figure E10. Contamination of mesenchymal cell transcriptomes by epithelial genes in snRNASeq versus scRNASeq. Dot plot showing diffuse detection of alveolar and airway epithelial genes, particularly club cell and alveolar type 2 cell markers, in single nucleus data more than single cell data. SMC, smooth muscle cells; MyoFB, myofibroblasts; FB, fibroblasts; Peri, pericytes.

Figure E11

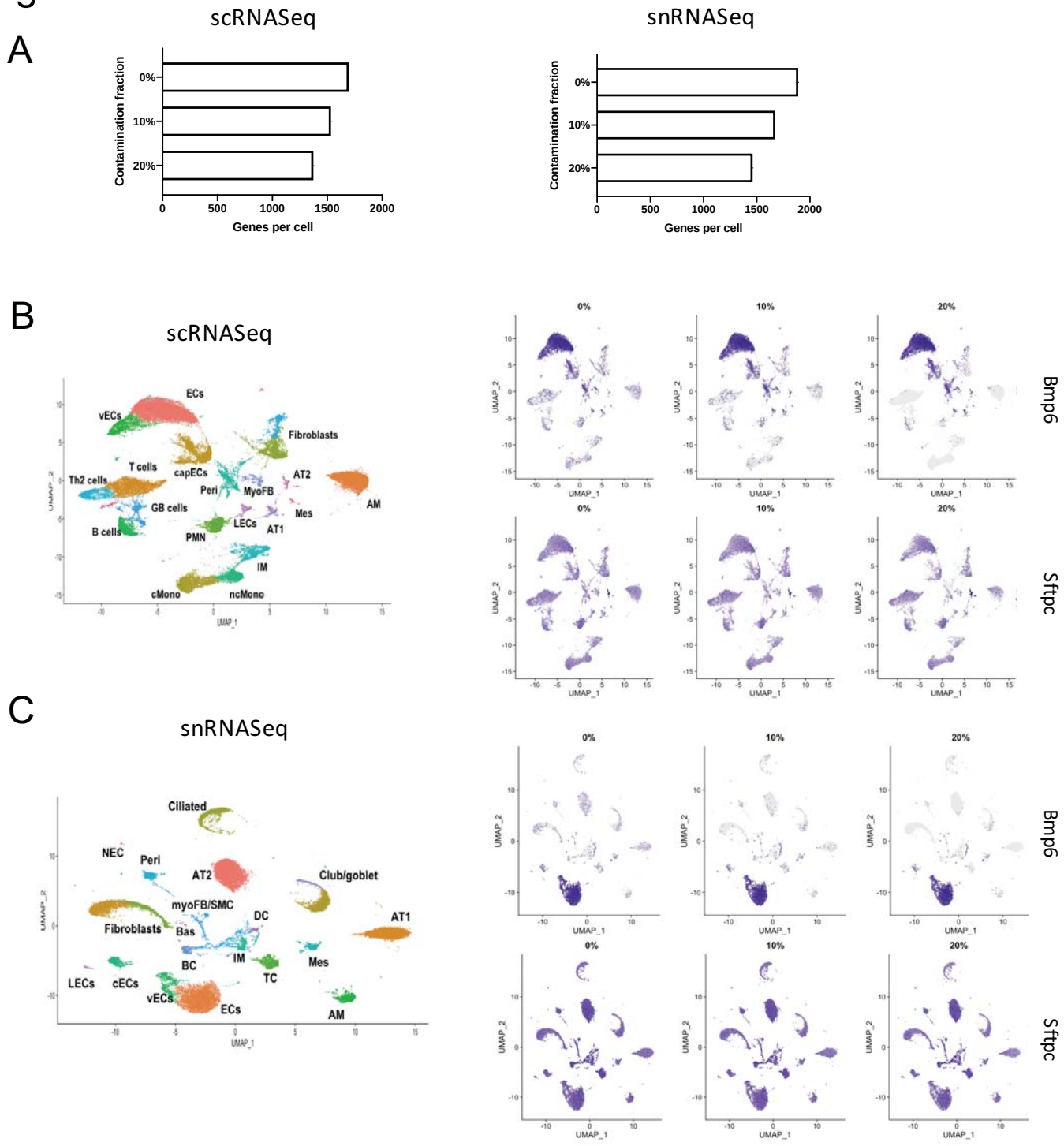

Figure E11. SoupX removal of contaminant background genes. (A) Bar graphs showing the overall reduction in gene detection with different supplied contamination fractions in SoupX. (B) Combined scRNASeq object after SoupX with supplied contamination fractions of 0, 10%, and 20% and effect on expression patterns of endothelial marker *Bmp6* and AT2 marker *Sftpc*. (C) Effect of supplied contamination fractions of 0-20% on the same genes in snRNASeq data. EC, endothelial cells; vECs, venous endothelial cells; cECs, capillary endothelial cells; LEC, lymphatic endothelial cells; BC, B cells; TC, T cells; AM, alveolar macrophages; IM, interstitial macrophages; DC, dendritic cells; AT1, alveolar type 1 cells; AT2, alveolar type 2 cells; Mes, mesothelial cells; Bas, basal cells; myoFB, myofibroblasts; Peri, pericytes; NEC, neuroendocrine cells; SMC, smooth muscle cells; PMN, neutrophils; cMono, classical monocytes; nMono, nonclassical monocytes.

Figure E12

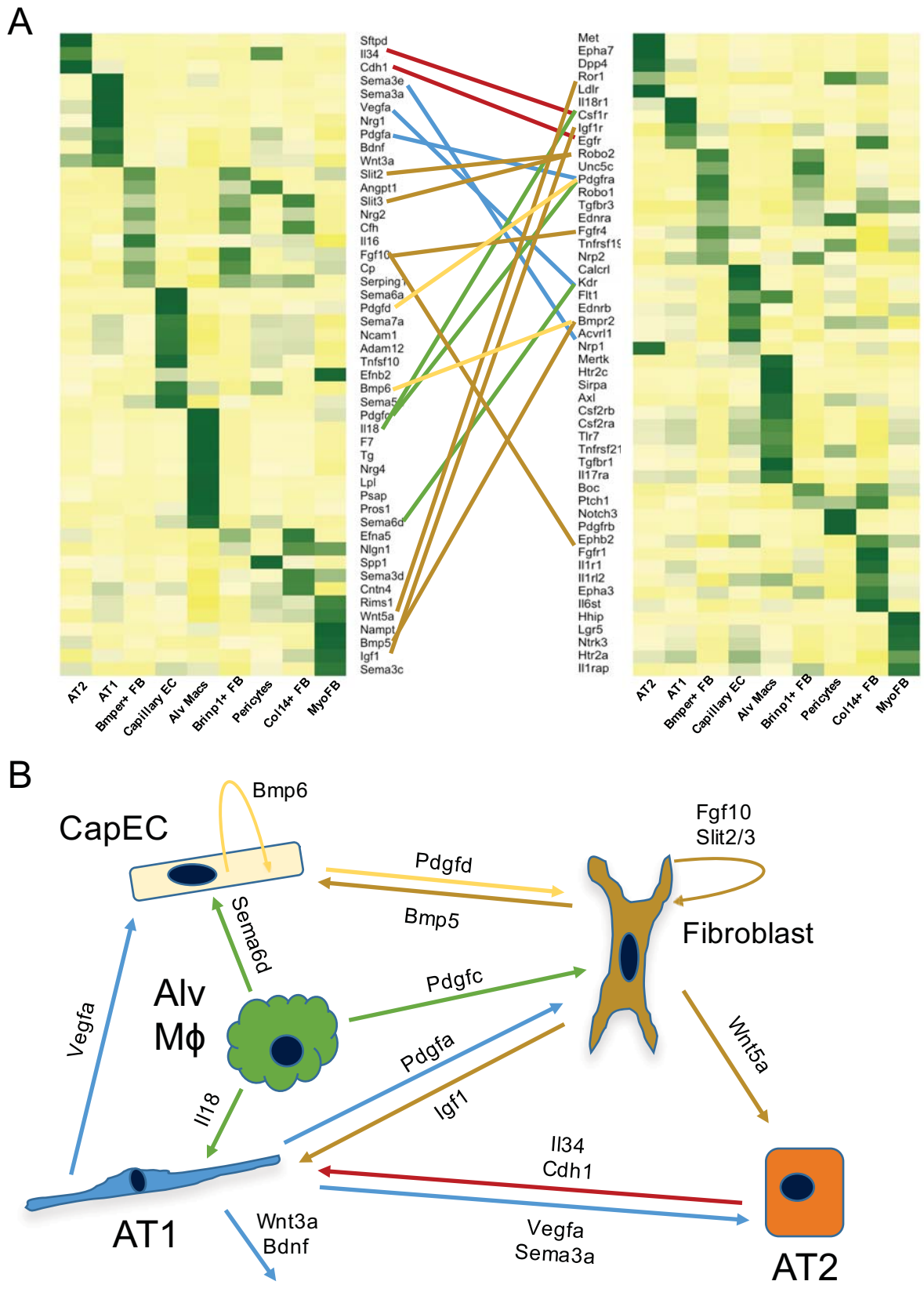

Figure E12. Ligand-receptor mapping in alveolar cell types. (A) Cell type-specific expression heatmaps of ligands and receptors illustrating potential networks of intercellular communication. (B) Alveolar signaling pathways suggested by the current snRNASeq data. AT1, alveolar type 1 cells; AT2, alveolar type 2 cells; FB, fibroblasts; MyoFB, myofibroblasts; EC, endothelial cells.

Supplemental Methods

Kidney dissociation protocol

Initially nuclear dissociation of mouse lung was performed using a previously described protocol for mouse kidney (1) with minor modifications. Briefly, frozen lung tissue was thawed on ice, cut to a ~7 mm piece, and placed in 1 mL ice-cold lysis buffer A (Nuclei EZ Lysis buffer (NUC-101, Sigma-Aldrich) supplemented with protease (Roche #589279100) and RNAse (Promega #N2615, Life Technologies #AM2696) inhibitors (200 U/ml RNasein Plus and 100 U/ml SUPERaseIN)), and minced to 1-2 mm pieces with scissors in a petri dish. Pieces were transferred to a Dounce homogenizer and an additional 1 mL lysis buffer A was added. After 20 strokes of homogenization with a loose pestle, the homogenate was passed through a 200 μm filter, followed by additional homogenization with a tight pestle (10 strokes) and transfer to a conical tube. An additional 2 mL lysis buffer A was added, followed by 5 min incubation and passage through a 40 μm strainer. The lysate was centrifuged at 500 x *g* for 5 min, and the supernatant discarded, with resuspension of the pellet in lysis buffer B (EZ Lysis buffer supplemented with protease inhibitor and RNase inhibitors (40 U/ml RNasein Plus and 20 U/ml SUPERaseIN). The sample was again incubated for 5 min followed by centrifugation at 500 x *g* for 5 min. The supernatant was discarded and the pellet resuspended in 1 mL resuspension buffer (PBS with 40 U/ml RNasein Plus), and the nuclei passed through a 5 μm filter. Nuclei were counted by hemocytometer and diluted to 10,000 per μm before proceeding to 10x Chromium.

Bioinformatics

Data were further processed using Seurat (2). For quality control, only genes expressed in >5 cells and cells expressing at least 200 genes were retained. No mitochondrial gene expression cutoff was used for single nucleus data, while cells expressing >10% mitochondrial genes were excluded from single cell analysis. Quality control cutoffs of 200 genes (minimum) and 2,500 genes (maximum) per cell were also used. Data were normalized and scaled, and the number of principal components estimated using *RunPCA* followed by *ElbowPlot*. Dimensionality reduction was performed using *RunUMAP*. Markers for cell clusters were identified using the *FindAllMarkers* function in *Seurat*, and cell types were annotated using canonical markers. To combine independent experiments, datasets were merged using the *Harmony* package to correct for batch effects. To further resolve cell subtypes (immune, endothelial, etc.), the *Subset* function in *Seurat* was used, with subsets of cells subjected to a second round of principal components identification and dimensionality reduction as above. Clusters of presumed doublets were identified and removed manually (relying on the presence of marker genes from multiple clusters and relatively high UMI counts).

Merging of cell and nucleus datasets

All cells (11,473) and nuclei (16,110) were merged into a single object using the *merge* function followed by batch correction with *Harmony* (3). PCA identification, clustering, and cell type annotation were then performed by the same workflow as above. Mesenchymal cells were isolated using *subset* for fibroblast/myofibroblast, pericyte, and smooth muscle cell clusters. For some analyses, *AverageExpression* was used to compare gene across clusters, otherwise, expression was profiled using the *DoHeatmap*, *Dotplot*, *Vlnplot*, and *FeaturePlot* functions within Seurat. For cluster composition comparisons, objects were randomly downsampled to

contain equal numbers of cells from each component (e.g. cells vs. nuclei or individual replicates).

Receptor-Ligand Analysis
To identify ligand-receptor interactions, we grouped cell types from the alveolar milieu and employed a curated ligand-receptor (LR) list with 2,557 LR pairs, as described (1). Receptors and ligands were selected on the basis of their differential expression in the selected subgroups of alveolar cells. For plotting, ligands and receptors with unique expression (q-val > 0.75) were selected, followed by manual removal of collagen/integrin pairs. The gplots package function *heatmap.2* was used for visualization.

Removal of contaminant genes with SoupX
We employed the *R* package *SoupX* to clean single cell and nucleus data of contaminant genes (4). Here, raw and filtered CellRanger outputs were used to create *SoupChannel* objects for single nucleus and cell data. Contamination fractions of 0%, 10% and 20% were assigned using *SetContaminationFraction*, and genes removed using *AdjustCounts*. The resulting count matrices were input back to Seurat, and the objects for each level of contamination merged using *Harmony*.

Deconvolution of bulk RNA-Seq data
To infer cell type ratios from bulk RNASeq data, the *R* package *bseq-sc* was used as previously described (5). Bulk RNASeq data for whole lung was obtained from Angelidis et al., GSE124872 (6), with the count table converted to an *ExpressionSet* object (aged mice were excluded for the purpose of our analysis). The combined single cell/nucleus object from Figure 1 was similarly converted to an *ExpressionSet* using the *SeuratToExpressionSet* command from the package *BisqueRNA*. The top 20 markers for each cluster were used for deconvolution, as in the tutorial for the *bseq-sc* package (https://shenorrlab.github.io/bseqsc/vignettes/bseq-sc.html).

References

1. Wu H, Kirita Y, Donnelly EL, Humphreys BD. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. J Am Soc Nephrol. 2019 Jan;30(1):23–32.

2. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018 May;36(5):411–20.

3. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019 Dec;16(12):1289–96.

4. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet based single-cell RNA sequencing data [Internet]. Bioinformatics; 2018 Apr [cited 2020 Feb 17]. Available from: http://biorxiv.org/lookup/doi/10.1101/303727

5. Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A, et al. A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. Cell Syst. 2016 26;3(4):346-360.e4.

6. Angelidis I, Simon LM, Fernandez IE, Strunz M, Mayr CH, Greiffo FR, et al. An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. Nat Commun. 2019 Feb 27;10(1):963.