**Supplementary Figure 1. FACS gating strategies for human lung and peripheral blood cells. a,** Sequential FACS data and sorting gates (red) for dissociated human lung cells from subject sample D1b (plate B001223) following MACS depletion of highly abundant immune ($CD45^+$) and endothelial ($CD31^+$) cells. The final sort (right) was of viable single cells from the lung epithelial ($EPCAM^+CD45^-$), immune ($CD45^+EPCAM^-$), and stromal/endothelial ($EPCAM^-CD45^-$) compartments into 384-well plates for SS2 scRNAseq. Plots are representative of FACS repeated on 3 subjects used for scRNAseq. **b,** Sequential FACS data and sorting gates (red) for white blood cells isolated on a Ficoll gradient of matched subject peripheral blood (subject 1, plate BP1). Viable, single $CD235a^-$ (non-RBC) cells were captured without additional gating (panel 4), or further sorted as CD8 T ($CD8^+$; panel 8), CD4 T ($CD4^+$; panel 7), B ($CD19^+CD3^-$; panel 6), NK ($CD19^-CD3^-CD56^+CD14^-$; panel 9), or $CD14^+$ monocytes ($CD19^-CD3^-CD56^-CD14^+$; panel 9) for SS2 scRNAseq. Contours, 5% increments in cell density. Sorting on blood cells was done only for subject 1.

**Supplementary Table 1. Canonical cell types (45) in the human lung and their abundances, markers, and available expression data.** a, numbers of each type were calculated with their abundances and the total number of lung cells (estimated by comparing volume of lungs to the whole body). b, Canonical markers were obtained from referenced expression data or commonly used markers in the literature. c, Expression profiles captured immediately following tissue dissociation are considered primary. d, Alveoli were assumed to occupy ~90% of the total lung volume for all estimations. e, Inferred from mean relative abundance in proximal, medial and distal airway epithelium. f, Calculated by stereology g, Resin casts showed similar surface area of arteries and veins. h, Vascular smooth muscle is estimated to be slightly more abundant than airway smooth muscle. i, abundance of a more general cell type was split evenly. j, inferred from impression of light or electron microscopy. k, inferred from histological abundance in non-perfused healthy tissue. l, inferred from abundance among immune cells with FACS. m, Calculated using microfluidic capture.

**Supplementary Table 2. Human lung cell cluster identities, abundances, and locations.** Cell numbers are stratified by type, subject, and sequencing technology. Cell abbreviations indicated are used throughout.

**Supplementary Table 3. Surface markers used to isolate canonical immune cell types in bulk mRNA sequencing.** See Methods for details on antibodies.

**Supplementary Table 4. Enriched markers found in each cluster, with transcription factors, receptors/ligands, and disease associated genes annotated.** Includes all enriched genes for each type (p-val > 0.05, MAST, downsampled cells per group to 100 for SS2 clusters or 500 for 10x clusters). Abbreviations: avg_logFC, the natural log of the average fold change between the cell type and other cell types in its tissue compartment; pct_in_cluster, percentage of cells within the cluster that express the gene; pct_out_cluster, percentage of cells outside cluster that express the gene; p_val_adj, p-value with Bonferroni correction applied; TF, transcription factor; OMIM, Online Mendelian Inheritance in Man; GWAS, genome wide association study.

**Supplementary Table 5. P-value and scores of each CellPhoneDB Receptor-Ligand interaction from each cluster.** Expression scores are given for significant (p > 0.1,

CellPhoneDB statistical framework) pairwise interactions between all SS2 cells (cell numbers given in Supplementary Table 2). Table includes all 1,085 receptor-ligand pairs and all pairwise cell type combinations. Abbreviations: partner_a, cellphonedb ID for the first interaction partner protein; partner_b, cellphonedb ID for the second interaction partner protein; source, reference from cellphonedb; secreted, whether the ligand is secreted or membrane bound; is_integrin, whether the interaction includes an integrin complex.

**Supplementary Table 6. Mouse lung cell cluster identities and their abundances in each dataset.** Cell numbers are stratified by type, mouse, and sequencing technology. Mouse-specific molecular markers are indicated.

**Supplementary Table 7. Genes specific to mouse and human in each cluster and lung wide.** Includes all differentially expressed genes (p-val > 0.05, MAST) for all SS2 cells in indicated mouse and human clusters (human and mouse cell numbers given in Supplementary Tables 2 and 6, respectively). Abbreviations: avg_logFC, the natural log of the average fold change between the mouse and human cell type indicated; pct_mouse, percentage of mouse cells within the cluster that express the gene; pct_human, percentage of human cells within the cluster that express the gene; p_val_adj, p-value with Bonferroni correction applied; enriched, gene is enriched in cluster in mouse or human.

**Supplementary Table 8. Evolutionary changes in cellular patterns of lung gene expression between mouse and human SS2 cells.** Numbers of genes and percentages are mean (±SEM) from using a median expression cutoff minus 0 to 2 standard deviations (0.25 increments) and a median percent cutoff plus 0 to 2 standard deviations (0.25 increments). Human and mouse cell numbers given in Supplementary Tables 2 and 6, respectively

**Supplementary Table 9. Evolutionary and functional classes of genes. Includes genes conserved between mouse and human.** Abbreviations: Evo type, evolutionary scenario gene falls into, with NA for genes not expressed; Gene class, type of gene (receptor, ligand, enzyme, transcription factor, etc); conserved clusters, number of homologous clusters gene where gene is expressed in both mouse and human; Human specific clusters, number of homologous clusters where the gene is only expressed in human; Mouse specific clusters, above for mouse.