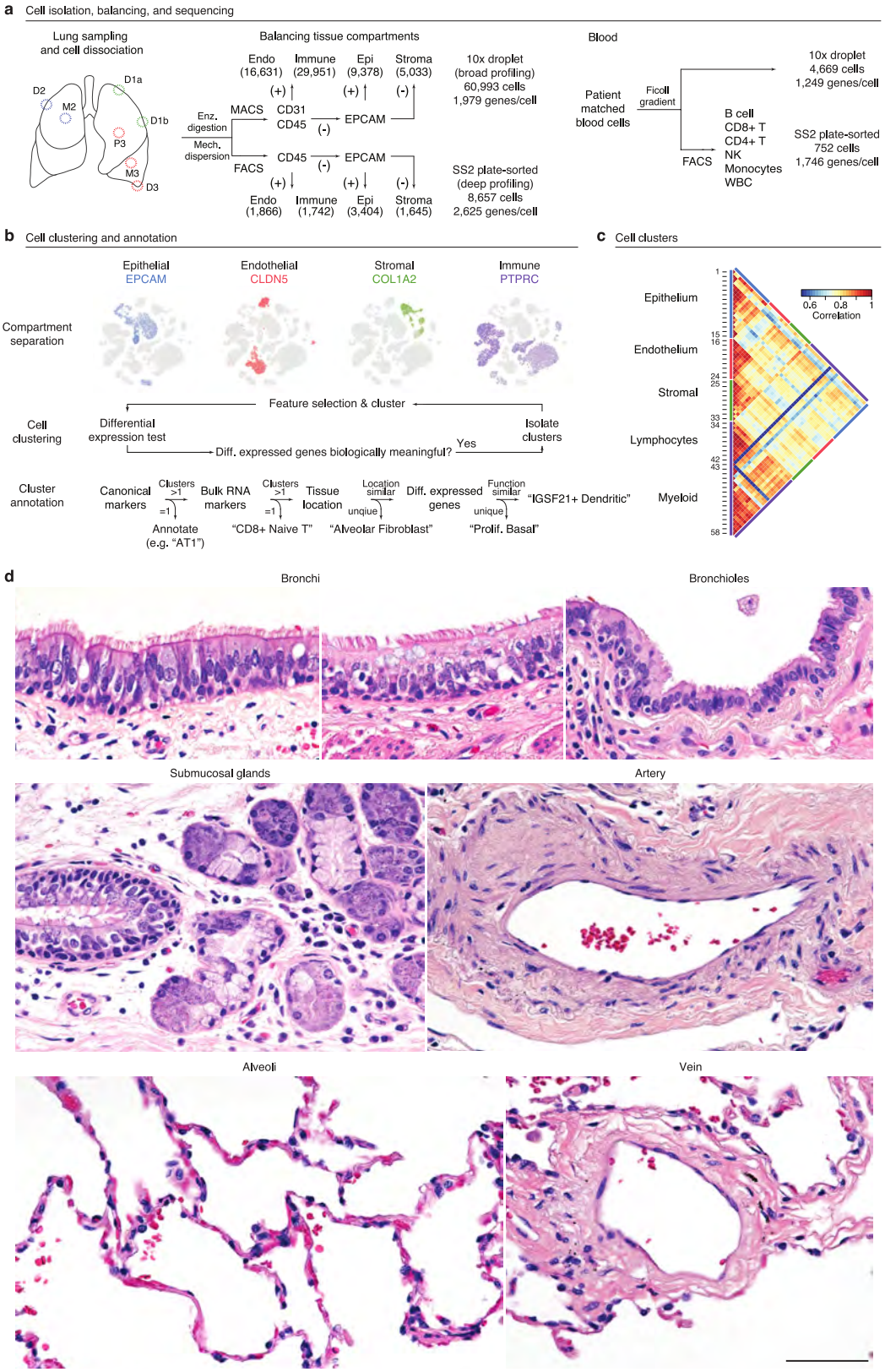
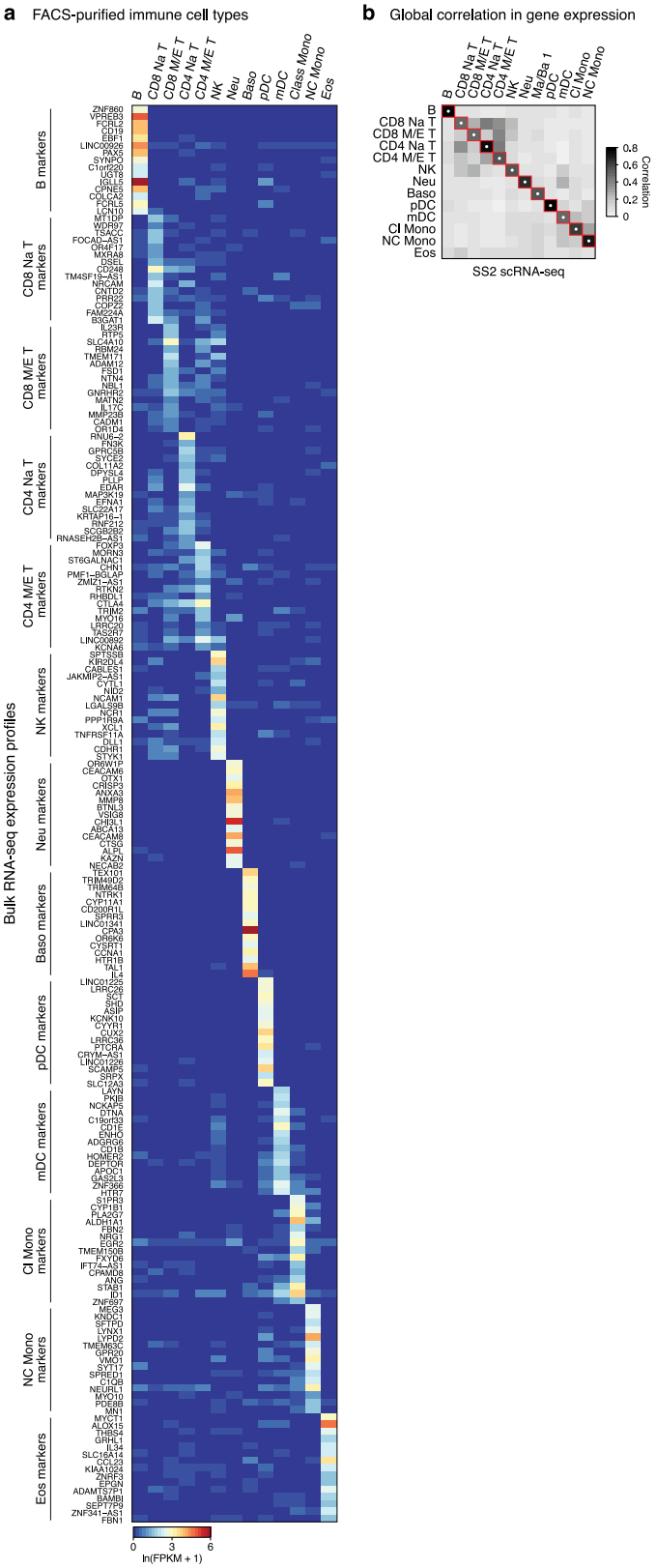


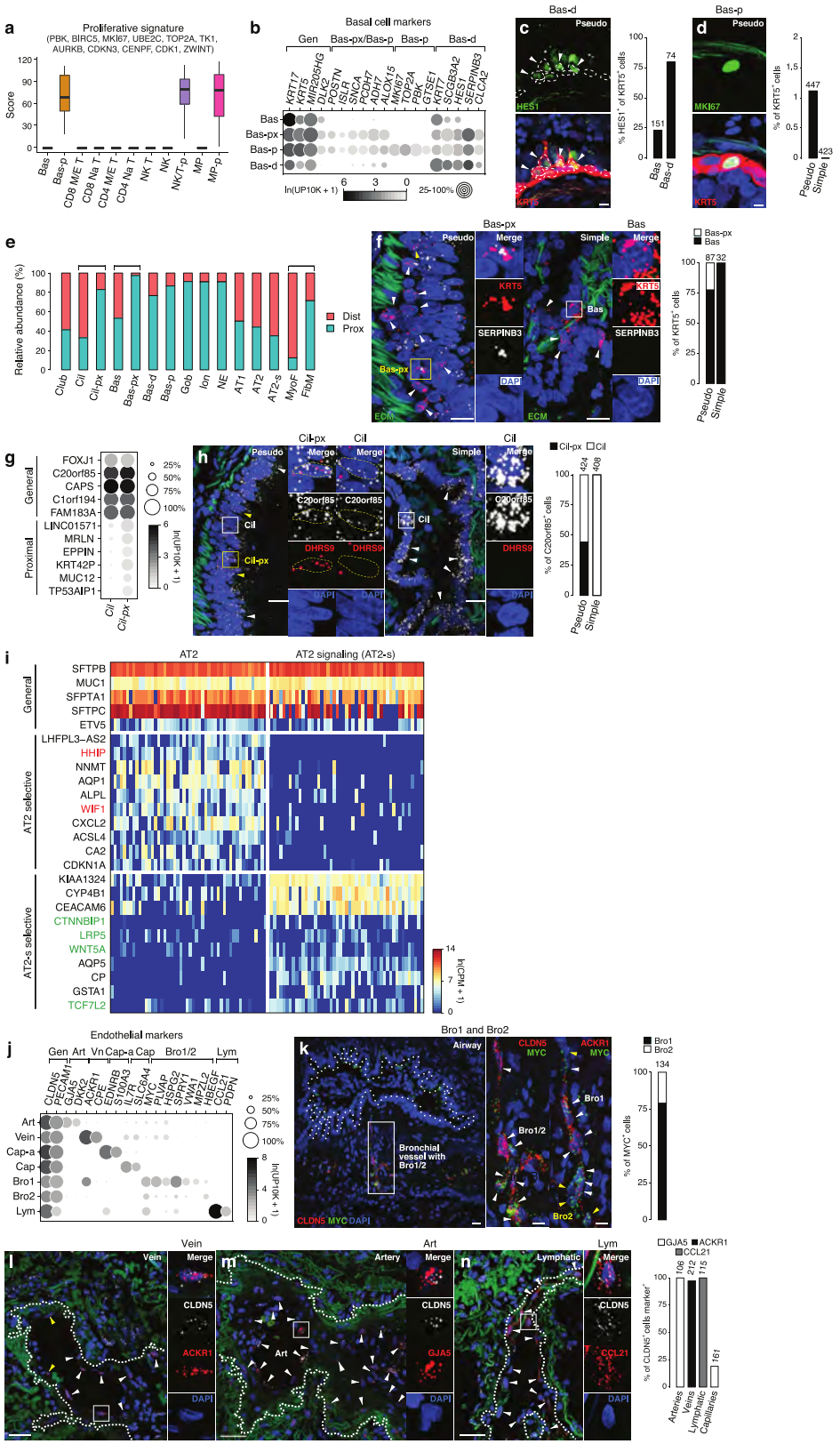
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Strategy for scRNA-seq and annotation of human lung and blood cells. a, Workflow for capture and mRNA sequencing of single cells from the healthy unaffected regions indicated (D, distal; M, medial; P, proximal lung tissue; see **d**) of fresh, surgically resected lungs with focal tumours from three participants (1, 2 and 3) and their matched peripheral blood. Cell representation was balanced among the major tissue compartments (endothelial, immune, epithelial and stroma) by magnetic and fluorescence activated cell sorting (MACS and FACS) using antibodies for the indicated surface markers (CD31, CD45, EPCAM). Cell capture and scRNA-seq was done using 10x droplet technology or SS2 analysis of plate-sorted cells. Number of profiled cells from each compartment are shown in parentheses. For blood, immune cells were isolated on a high density Ficoll gradient, and unsorted cells profiled by 10x and sorted cells (using canonical markers for the indicated immune populations) by SS2. Total cell number (all three participants) and median number of expressed genes per cell are indicated for each method. **b**, Cell clustering and annotation pipeline. Cell expression profiles were computationally clustered by nearest-neighbour relationships and clusters were then separated into tissue compartments based on expression of compartment-specific markers (*EPCAM* (blue), *CLDN5* (red), *COL1A2* (green), and *PTPRC* (purple)), as shown for *t*-distributed stochastic neighbour embedding (*t*-SNE) plot of lung and blood cell expression profiles obtained by 10x from participant 3. Cells from each tissue compartment were then iteratively re-clustered until differentially-expressed genes driving clustering were no longer biologically meaningful. Cell cluster annotation was based on expression of canonical marker genes from the literature, markers found through RNA sequencing of purified cell populations (bulk RNA markers), ascertained tissue location, and inferred molecular function from differentially-expressed genes. **c**, Heat map of pairwise Pearson correlations of the average expression profile of each cluster in the combined 10x dataset plus SS2 analysis of neutrophils. *n* values are in Supplementary Table 2. Tissue compartment and identification number of each of the 58 clusters are indicated. For more details on statistics and reproducibility, see Methods. **d**, Representative micrographs of donor lungs from formalin-fixed, paraffin-embedded sections stained with haematoxylin and eosin showing bronchi, bronchioles, submucosal glands, arteries, veins and alveoli near regions used for scRNA-seq. Staining repeated on at least five sections (encompassing different anatomical regions) from each participant used for scRNA-seq. Scale bar, 100 μm.



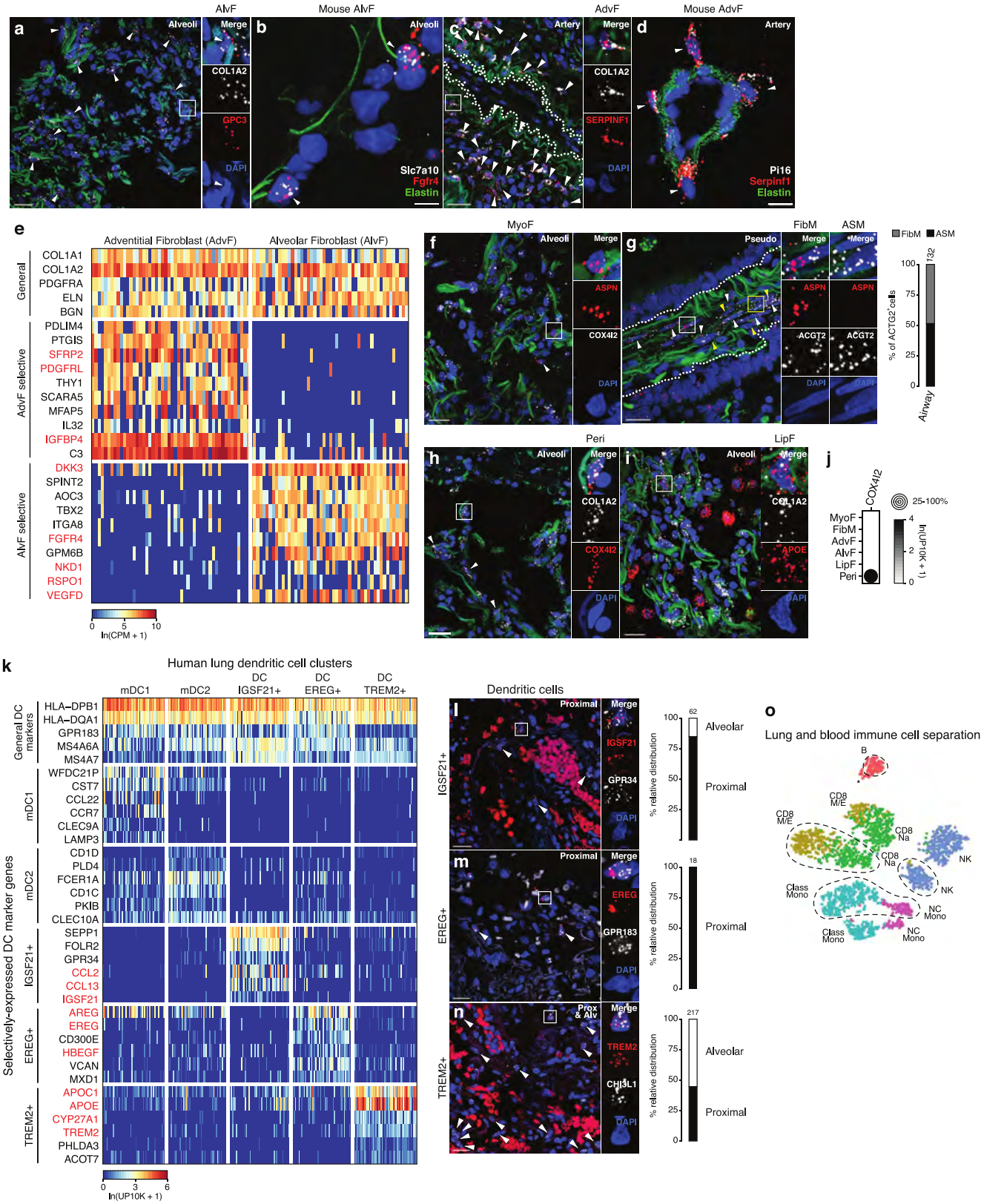
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Selectively expressed RNA markers of human immune cell types from bulk mRNA sequencing of FACS-purified immune cells. a, Heat map of RNA expression of the most selectively-expressed genes from bulk mRNA sequencing of the indicated FACS-sorted immune populations (Supplementary Table 3). This dataset provided RNA markers for human immune cell populations that have been classically defined by their cell surface markers. **b**, Heat map of pairwise Pearson correlation scores between the average expression profiles of the immune cell types indicated that were obtained from bulk mRNA sequencing (BulkSeq, **a**) to the average scRNA-seq profiles of human blood immune cells in the SS2 dataset annotated by canonical markers and enriched RNA markers from the bulk RNA-seq analysis. The highest correlation in overall gene expression (white dot) of each annotated immune cell cluster in the SS2 dataset (columns) was to the bulk RNA-seq of the same FACS-purified immune population (rows), supporting the scRNA-seq immune cluster annotations (red squares). Cell numbers are in Supplementary Table 2. For more details on statistics and reproducibility, see Methods.



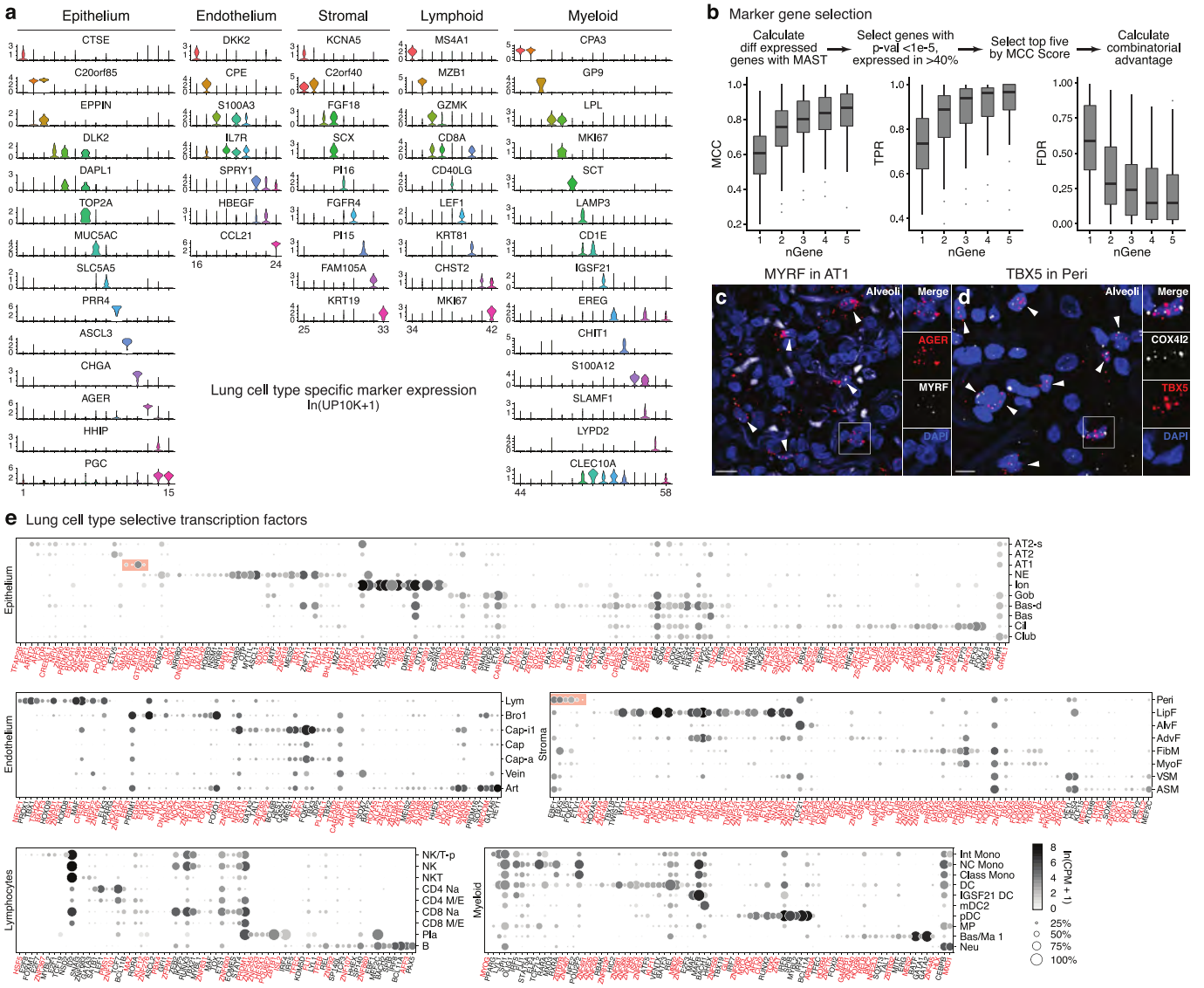
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Expression differences and localization of lung cell states and canonical epithelial and endothelial subtypes. a, Proliferative signature score (based on expression of indicated genes in cells from 10x dataset; cell numbers are in Supplementary Table 2) of each cluster of basal cells, T and natural killer cells, and macrophages. Three clusters had high scores: proliferating basal cells (Bas-p), proliferating natural killer/T cells (NK/T-p), and proliferating macrophages. **b**, Dot plot of mean level of expression (dot intensity, grey scale) of indicated basal cell markers and percent of cells in population with detected expression (dot size) for 10x dataset. Note partial overlap of markers among different basal populations. **c**, Immunostaining of adult human pseudostratified airway for differentiation marker HES1 (green) in basal cells (marked by KRT5, red) with DAPI (nuclear) counter stain (blue). Scale bars, 10 µm. Note apical processes extending from HES1$^+$ basal cells (arrowheads) indicating migration away from basal lamina as they differentiate. Other HES1$^+$ cells have turned off basal marker KRT5. Dashed outlines, basal cell nuclei. Quantification shows fraction of basal cells (cuboidal KRT5$^+$ cells on basement membrane) and differentiating basal (Bas-d) cells (KRT5$^+$ cells with apical processes) that were HES1$^+$. *n* denotes KRT5$^+$ cells scored in sections of two human lungs with staining repeated on four participants. **d**, Immunostaining of adult human pseudostratified airway for proliferation marker MKI67 (green) in basal cells (marked by KRT5, red) with DAPI counter stain (blue). Scale bars, 5 µm. Quantification shows abundance of proliferating (MKI67-expressing) basal cells in pseudostratified (pseudo) and simple epithelial airways; *n* denotes KRT5$^+$ cells scored in sections of two human lungs with staining repeated on four participants. **e**, Relative abundance of epithelial and stromal cell types in scRNA-seq analysis of human lung samples obtained from proximal (blue; 10x cells from P3) and distal (red; 10x cells from D1a, D1b, D2, D3) lung sites. In addition to the expected proximal enrichment of some airway cell types (goblet cells, ionocytes, neuroendocrine cells) and distal enrichment of alveolar cell types (AT1, AT2, AT2-signalling, myofibroblasts), note three bracketed pairs of related cell types (ciliated and proximal ciliated; basal and proximal basal (Bas-px) cells; myofibroblasts and fibromyocytes) with one of them proximally enriched. Relative enrichment values are provisional because they can be influenced by efficiency of collection during cell dissociation and isolation. Cell number for proximal cells are (from left to right): 357, 275, 73, 175, 153, 191, 39, 145, 57, 24, 20, 10, 328, 1,505, 235, 25 and 70; and for distal cells are: 537, 806, 15, 197, 4, 58, 6, 14, 336, 0, 2, 1, 467, 2,095, 434, 198 and 28. **f**, RNAscope smFISH and quantification for general basal marker *KRT5* (red) and proximal basal cell marker *SERPINB3* (white) with DAPI counter stain (blue) and ECM autofluorescence (green) on proximal, pseudostratified bronchi and distal, simple bronchioles. Scale bars, 20 µm (inset, 10 µm). Note enrichment of proximal basal cells (*KRT5 SERPINB3* double positive, yellow arrowhead and box) enrichment at base of pseudostratified airways. *SERPINB3* was not detected in simple airways, indicating that basal cells (but not proximal basal cells) are present there. Staining repeated on two participants. **g**, Dot plot of expression in ciliated and proximal ciliated cells of canonical (general) ciliated cell markers and specific proximal ciliated markers (in 10x dataset). **h**, smFISH and quantification of human pseudostratified epithelial (left) and simple epithelial (right) airways for general ciliated marker *C20orf85* (white) and proximal ciliated marker *DHRS9* (red) with DAPI counterstain (blue) and ECM autofluorescence (green). Note restriction of proximal ciliated cells to pseudostratified airways. Scale bars, 10 µm. Staining repeated on two particpants. **i**, Heat map of expression of representative general AT2, AT2 selective, and AT2-signalling selective marker genes in AT2 and AT2-signalling human lung cells (SS2 data). AT2 selective markers include negative regulators of Hedgehog and Wnt signalling pathways (for example, *HHIP* and *WIF1*, highlighted red) and AT2-signalling selective markers include Wnt ligands, receptors and transcription factors (for example, *WNT5A*, *LRP5* and *TFC7L2* highlighted green). Values shown are ln(CPM + 1) for 50 randomly selected cells in each cluster (SS2 data). **j**, Dot plot of expression of endothelial markers (10x dataset). **k**, Micrograph (low magnification, left) of bronchial vessel (boxed region) showing vessel location near airway (dotted outline). smFISH for general endothelial marker *CLDN5* (red, centre), bronchial vessel-specific markers *MYC* (green) and Bro1-specific marker *ACKR1* (red, right) on serial sections of bronchial vessel cells (arrowheads), co-stained for DAPI (blue). Scale bar, 10 µm. Quantification shows relative abundance of Bro1 and Bro2 cells. Staining repeated on two participants. **l**–**n**, smFISH and quantification of vessel types indicated (dotted outlines) showing vein marker *ACKR1* (red; **l**), artery marker *GJA5* (red; **m**), lymphatic marker *CCL21* (red; **n**), and general endothelial marker *CLDN5* with DAPI counter stain (blue) and ECM autofluorescence (green). Scale bars, 50 µm (**l**), 30 µm (**m**) and 40 µm (**n**). Staining repeated on two participants. For more details on statistics and reproducibility, see Methods.

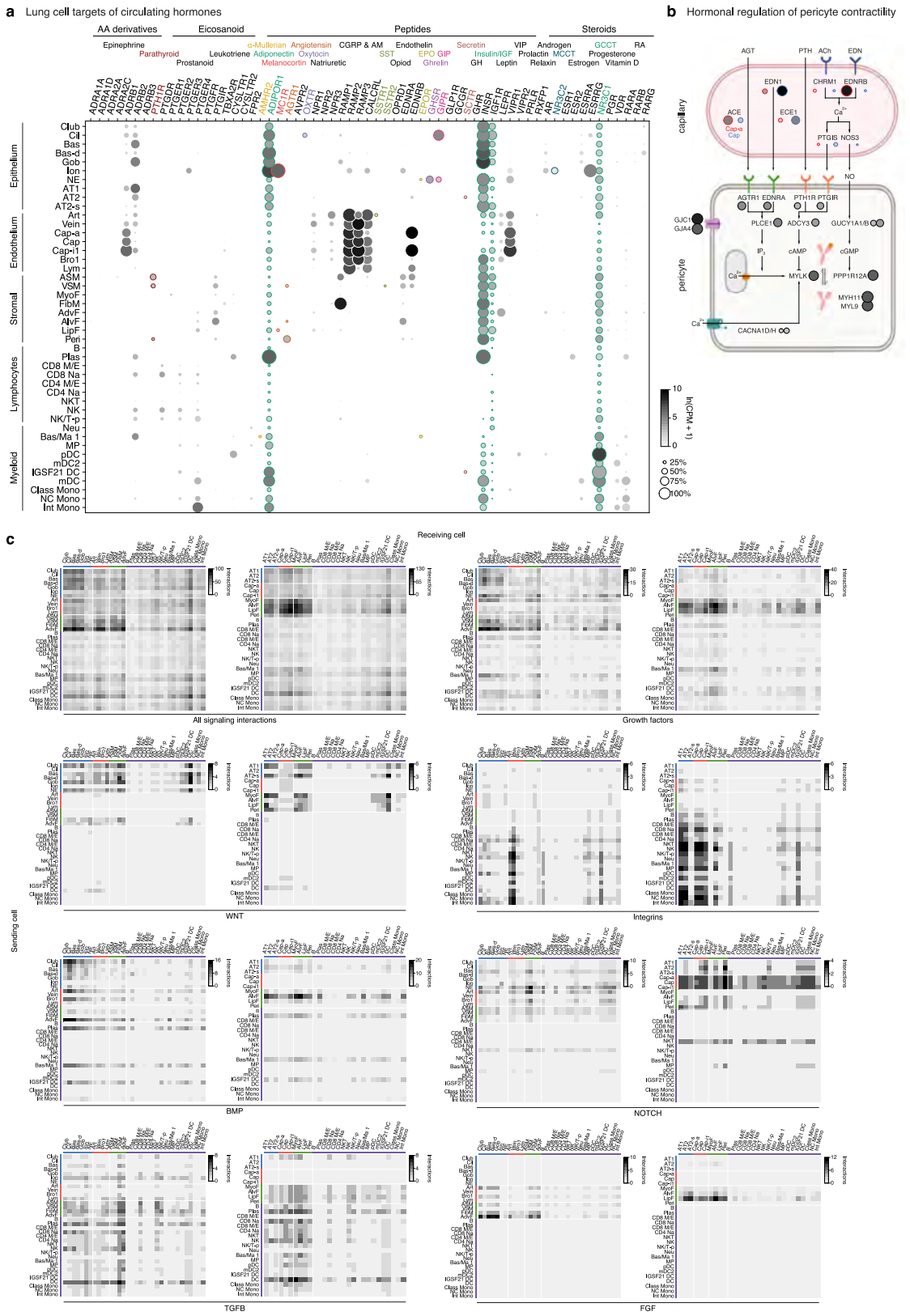


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Markers and lung localization of stromal and dendritic subtypes. a–**d**, smFISH for RNA of indicated marker genes of alveolar fibroblasts (**a**, **b**) and adventitial fibroblasts (**c**, **d**) in adult human (**a**, **c**) and mouse (**b**, **e**) alveolar (**a**, **b**) and pulmonary artery (**c**, **d**) sections. ECM autofluorescence (green; **a**, **c**) to show blood vessels; Elastin (green, **b**, **d**); DAPI counterstain (blue, all panels). Staining repeated on two human participants or three mice. **a**, smFISH probes: general fibroblast marker *COL1A2* (white) and alveolar fibroblast-selective marker *GPC3* (red). Arrowheads denote alveolar fibroblasts. Inset, close-up of boxed region showing merged (top) and split channels of an alveolar fibroblast. Scale bars, 20 µm (inset 60 µm). **b**, smFISH probes: alveolar fibroblast-selective markers *Slc7a10* (white) and *Frfr4* (red). Elastin (green) shows alveolar entrance ring. Arrowheads denote alveolar fibroblasts. Scale bar, 5 µm. **c**, smFISH probes: general fibroblast marker *COL1A2* (white) and adventitial fibroblast -selective marker *SERPINF1* (red). Adventitial fibroblasts (some indicated by arrowheads) localize around blood vessels (ECM, green). Inset, close-up of boxed region showing merged (top) and split channels of an adventitial fibroblast. Dashed line denotes the artery boundary. Scale bars, 30 µm (inset 90 µm). **d**, smFISH probes: adventitial fibroblast-selective markers *Pi16* (white) and *Serpinf1* (red). Adventitial fibroblasts (arrowheads) surround artery (marked by elastin, green). Scale bar, 10 µm. **e**, Heat map of expression of representative general, adventitial-selective, and alveolar-selective fibroblast markers in 50 randomly selected cells from adventitial (left) and alveolar (right) fibroblast clusters (SS2 dataset). Note specialization (highlighted red) in growth factors (AdvF: *PDGFRL*, *IGFBP4*; AlvF: *FGFR4*, *VEGFD*) and morphogen (AdvF: *SFRP2*; AlvF: *NKD1*, *DKK3*) signalling or regulation. **f**, **g**, smFISH and quantification of cell abundance in human alveolar (**f**) and pseudostratified epithelial airway (**g**) sections probed for myofibroblast and fibromyocyte marker *ASPN* (red), and for fibromyocyte and airway smooth muscle markers COX4I2 (white; **f**) and *ACTG2* (white; **g**). ECM autofluorescence, green; DAPI counterstain, blue. Inset (**f**), boxed region showing close-up of merged (top) and split channels of *ASPN*$^{+}$ *COX4I2*$^{-}$ myofibroblast. Myofibroblasts and fibromyocytes (see below) probably make up remaining cells in Fig. 1f quantification. Inset (**g**), boxed regions showing close-up of merged (top) and split channels of fibromyocyte (white box) and airway smooth muscle (yellow box) cells. Fibromyocytes (white arrowheads) and airway smooth muscle (yellow arrowheads) are intermingled in wall of pseudostratified airway (dotted outline). Staining repeated on two participants. **h**, **i**, smFISH of human alveolar sections probed for general stromal marker *COL1A2* (white), pericyte marker *COX4I2* (red; **h**), lipofibroblast marker *APOE* (red; **i**). ECM autofluorescence, green; DAPI counterstain, blue. Inset (**h**), boxed region showing close-up of pericyte. Inset (**i**), boxed region showing close-up of *COL1A2 APOE* double-positive LipF. LipF cells are intermingled among other stromal cells (single-positive *COL1A2*) and macrophages (single-positive *APOE*). Quantification in Fig. 1f. Scale bars, 20 µm. Staining repeated on two participants . **j**, Dot plot of *COX4I2* expression in alveolar stromal cell types (10x dataset). **k**, Heat map of expression of dendritic cell marker genes in 50 randomly selected cells from indicated dendritic cell clusters (human blood and lung 10x datasets). Cells in all clusters express general dendritic markers including antigen presenting genes but each cluster also has its own selective markers. Red highlighted markers distinguishing the newly identified dendritic cell clusters (IGSF21^{+}, EREG^{+}, TREM2^{+}) suggest different roles in asthma (IGSF21^{+}), growth factor regulation (EREG^{+}), and lipid handling (TREM2^{+}). **l**–**n**, smFISH of adult human lung proximal and alveolar (Alv) sections as indicated probed for IGSF21^{+} dendritic cell markers *IGSF21* (red) and *GPR34* (white) (**l**), EREG^{+} dendritic cell marker *EREG* (red) and general dendritic cell marker *GPR183* (white) (**m**), and TREM2^{+} dendritic cell markers *TREM2* (red) and *CHI3L1* (white) (**n**). DAPI counterstain, blue. Non-punctate signal in red channel (**l**, **n**) is erythrocyte autofluorescence. Insets, boxed regions showing merged and split channels of close-up of single dendritic cell of indicated type. Scale bars, 20 µm. Arrowheads denote double-positive cells. Quantification shows distribution of each dendritic type; note IGSF21^{+} and EREG^{+} dendritic cells show strong proximal enrichment. Staining repeated on two participants. **o**, *t*-SNE of expression profile clusters of monocytes and B, T and natural killer cells (10x dataset, participant 1, 2,622 cells). Note separate cell clusters of each immune cell type isolated from lung (no outline) and blood (dashed outline). Asterisk denotes small number of B cells isolated from the lung that cluster next to blood B cells. For more details on statistics and reproducibility, see Methods.

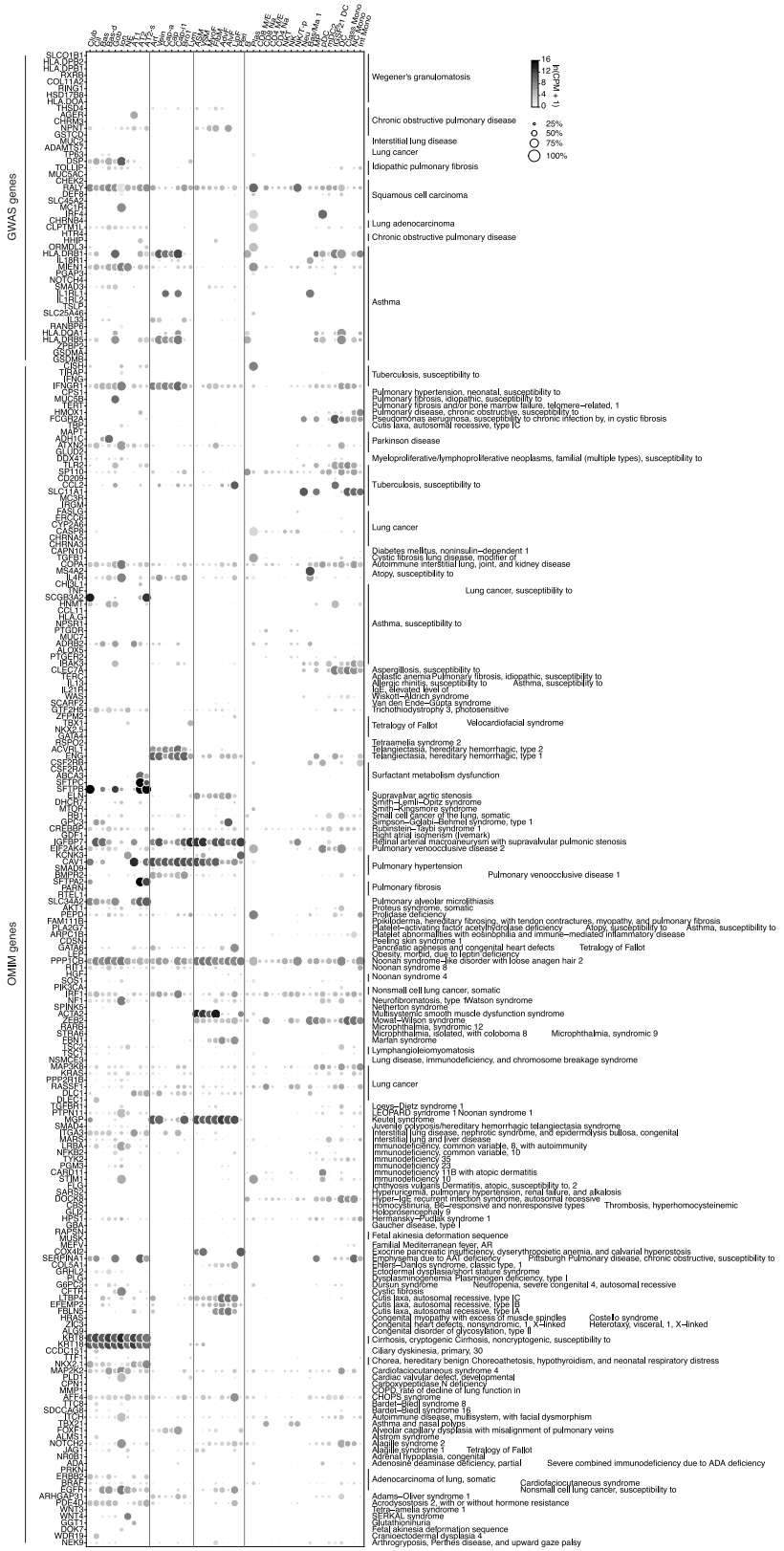


Extended Data Fig. 5 | Markers and transcription factors that distinguish human lung cell types. a, Violin plots of expression levels (ln(UP10K + 1)) of the most sensitive and specific markers (gene symbols) for each human lung cell type in its tissue compartment (10x dataset). Cell numbers given in Supplementary Table 2. **b**, Scheme for selecting the most sensitive and specific marker genes for each cell type using Matthews correlation coefficient (MCC). Box-and-whisker plots below show MCCs, true positive rates (TPR), and false discovery rates (FDR) for each cell type ($n = 58$) using indicated number (nGene) of the most sensitive and specific markers (10x dataset). Note all measures saturate at approximately 2–4 genes, hence simultaneous in situ probing of a human lung for the approximately 100–200 optimal markers would assign identity to nearly every cell. **c**, Alveolar section of human lung probed by smFISH for AT1 marker *AGER* and transcription factor *MYRF*. *MYRF* is selectively expressed in AT1 cells (arrowheads; 97% of *MYRF*$^+$ cells were *AGER*$^+$, $n = 250$ scored cells). Inset, boxed region showing merged and split channels of AT1 cell. Scale bar, 10 µm. Staining repeated on two participants. **d**, Alveolar section of human lung probed by smFISH for pericyte marker *COX4I2* and transcription factor *TBX5*. *TBX5* is enriched in pericytes (arrowheads, 92% of *TBX5*$^+$ cells were *COX4I2*$^+$, $n = 250$). Inset, boxed region showing merged and split channels of pericyte. Scale bar, 5 µm. Staining repeated on two participants. **e**, Dot plot of expression of enriched transcription factors in each lung cell type (SS2 dataset). Red text, genes not previously associated with the cell type. Red shading, transcription factors including *MYRF* that are highly enriched in AT1 cells, and *TBX5* and others highly enriched in pericytes. For more details on statistics and reproducibility, see Methods.

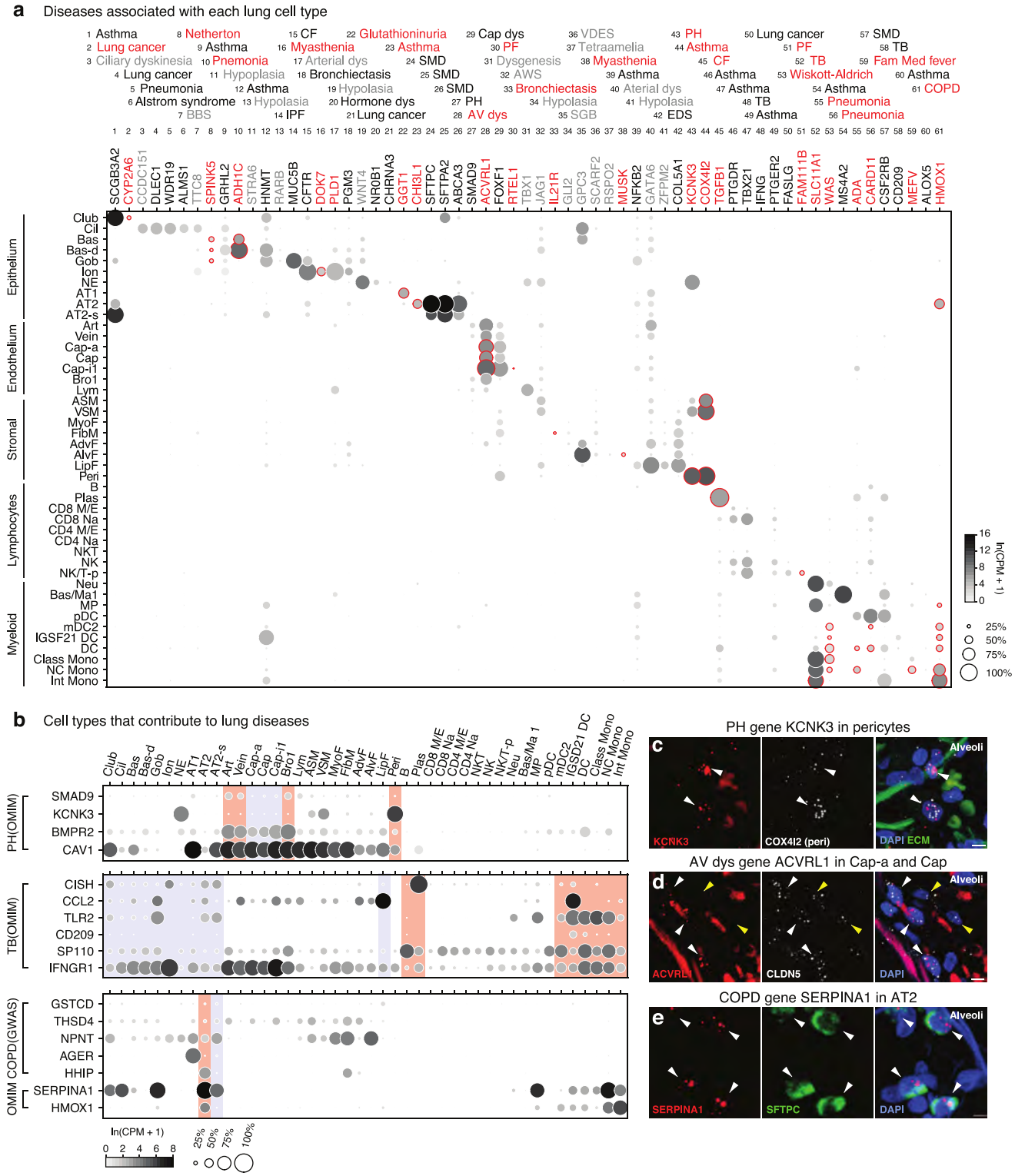


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Lung cell targets of circulating hormones and local signals. a, Dot plot of hormone receptor gene expression in lung cells (SS2 dataset). Type and name of cognate hormones for each receptor are shown at top. Teal, broadly-expressed receptors in lung; other colours, selectively-expressed receptors (<3 lung cell types). Small coloured dots next to cell type names show selectively targeted cell types. AA, amino acid; AM, adrenomedullin; CGRP, calcitonin gene-related peptide; EPO, erythropoietin; GCCT, glucocorticoid; GH, growth hormone; GIP, gastric inhibitory peptide; IGF, insulin-like growth factor; MCCT, mineralocorticoid; RA, retinoic acid; SST, somatostatin. **b**, Schematic of inferred pericyte cell contractility pathway and its regulation by circulating hormones (AGT, PTH) and capillary expressed signals (EDN, NO). Dots show expression of indicated pathway genes: values at left (outlined red) in each pair of dots in capillary diagram (top) show expression in Cap-a cells (aerocytes) and at right (outlined blue) show expression in general Cap cells (SS2 dataset). Note most signal genes are preferentially expressed in Cap relative to Cap-a cells. **c**, Heat maps showing number of interactions predicted by CellPhoneDB software between human lung cell types located in proximal lung regions (left panel in each pair) and distal regions (right) based on expression patterns of ligand genes ('sending cell') and their cognate receptor genes ('receiving cell') (SS2 dataset). The pair of heat maps at the top left show values for all predicted signalling interactions ('all interactions'), and other pairs show values for the indicated types of signals (growth factors, cytokines, integrins, WNT, Notch, BMP, FGF and TFGβ). Predicted interactions between cell types range from 0 (lymphocyte signalling to neutrophils) to 136 (AdvF signalling to Cap-i1). Note expected relationships, such as immune cells expressing integrins to interact with endothelial cells and having higher levels of cytokine signalling relative to their global signalling, and unexpected relationships, such as fibroblasts expressing most growth factors and lack of Notch signalling originating from immune cells. For more details on statistics and reproducibility, see Methods.

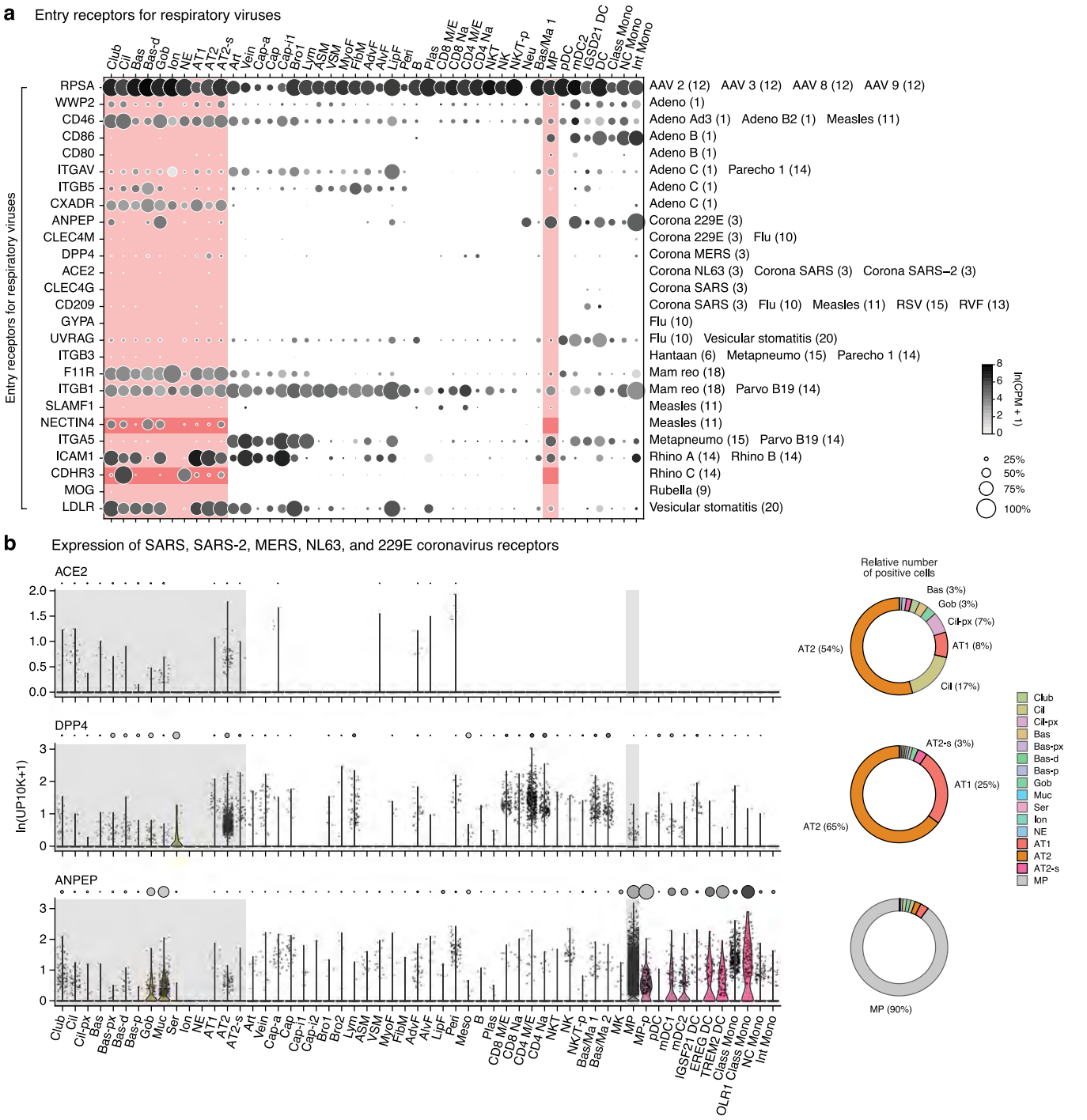


Extended Data Fig. 7 | Lung cell expression patterns of genes implicated in lung disease. Dot plots of expression (in SS2 dataset) of 233 lung disease genes curated from genome-wide association studies (GWAS; genome-wide association genes $\geq 10^{-20}$ significance) and Online Mendelian Inheritance in Man (OMIM). For more details on statistics and reproducibility, see Methods.

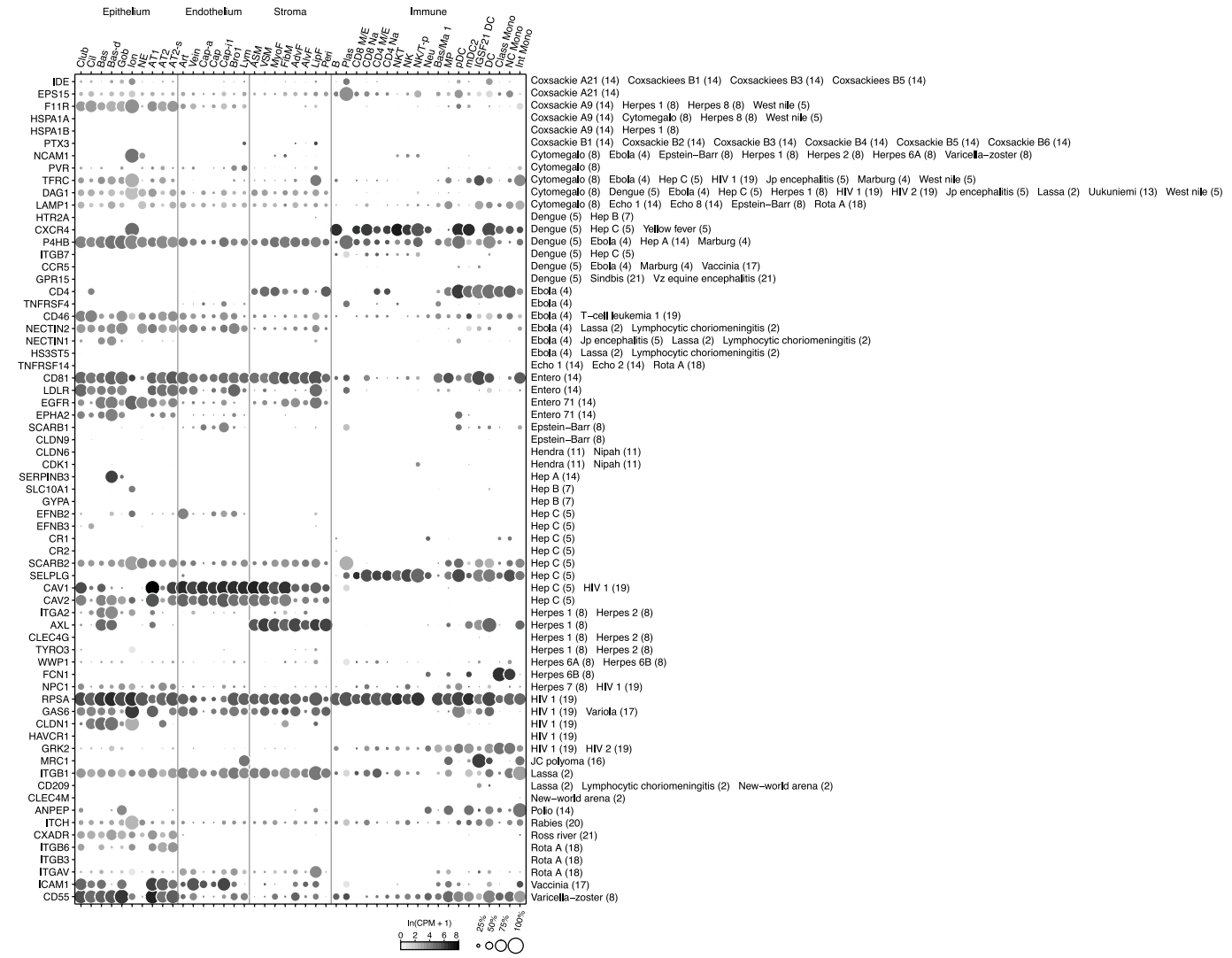


Extended Data Fig. 8 | See next page for caption.

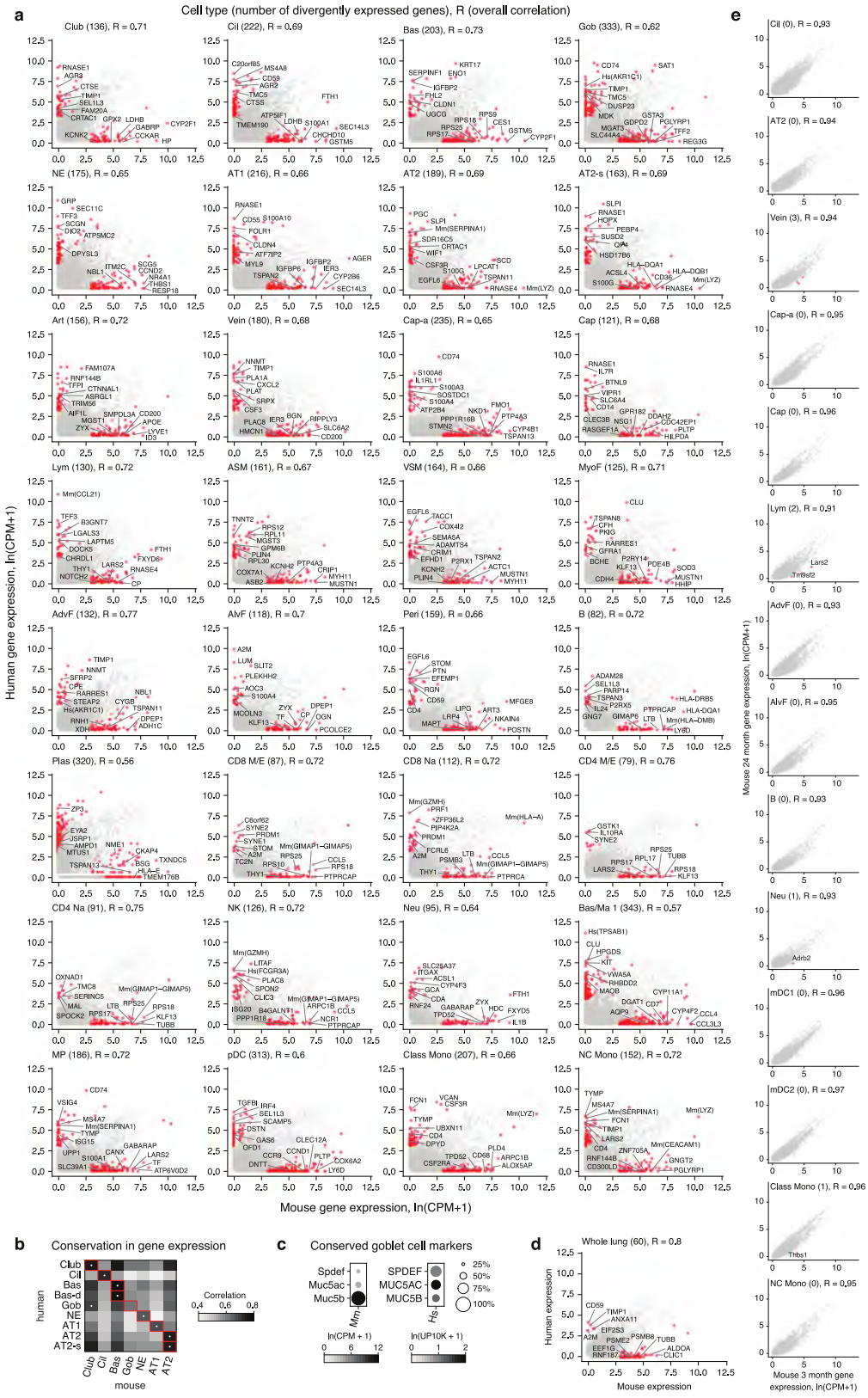
Extended Data Fig. 8 | Mapping cellular origins of lung disease by cell-selective expression of disease genes. a, Dot plots of expression of lung disease genes (numbered, associated disease shown above) enriched in specific lung cell types (SS2 datasets). Red, novel cell type association of gene or disease; grey, diseases with developmental phenotype. AWS, Alagille–Watson syndrome; BBS, Bardet–Biedl syndrome; CF, cystic fibrosis; Dys, dysplasia; EDS, Ehlers-Danlos syndrome; Fam Med, familial Mediterranean; IPF, idiopathic pulmonary fibrosis; PH, pulmonary hypertension; SGB, Simpson-Golabi-Behmel; SM, smooth muscle; SMD, surfactant metabolism dysfunction; TB, tuberculosis; VDES, Van den Ende-Gupta syndrome. **b**, Dot plot of expression (SS2 dataset) of all genes implicated in pulmonary hypertension, tuberculosis and COPD or emphysema (OMIM, Mendelian disease genes from OMIM database; GWAS, genome-wide association genes $\geq 10^{-20}$ significance). Note canonical AT2 cells (red shading) express all and AT2-signalling cells (blue shading) express most. **c**, smFISH of alveolar section of adult human lung probed for in pulmonary hypertension disease gene *KCNK3* (red) and pericyte marker *COX4I2* (white) with DAPI counterstain (blue) and ECM autoflourescence (green). Note pericyte-specific expression (arrowheads, 91% of *COX4I2*$^+$ pericytes were *KCNK3*$^+$, $n = 77$). Scale bar, 5 µm. Cell numbers for each type given in Supplementary Table 2. **d**, smFISH of alveolar section of adult human lung probed for atrioventricular (AV) dysplasia gene *ACVRL1* (red), endothelial marker *CLDN5* (white) with DAPI counterstain. Note *ACVRL1 CLDN5* double-positive capillaries (white arrowheads, 70% of *CLDN5*$^+$ capillaries were *ACVRL1*$^+$, $n = 102$) and some *CLDN5* single-positive capillaries (yellow arrowheads). Scale bar, 5 µm. **e**, smFISH of alveolar section of adult human lung probed for COPD or emphysema gene *SERPINA1* and AT2 marker *SFTPC*, and DAPI. Note AT2-specific expression (arrowheads; 93% of AT2 cells were *SERPINA1*$^+$, $n = 176$). Scale bar, 5 µm. For more details on statistics and reproducibility, see Methods.



Extended Data Fig. 9 | Lung cell expression patterns of respiratory virus receptors. a, Dot plot showing expression in human lung cell types of entry receptors (indicated at left) for respiratory viruses (indicated at right, numbers indicate viral families) (SS2 dataset). Red shading, cell types inhaled viruses could directly access (epithelial cells and macrophages); darker red shading shows expression values for measles receptor *NECTIN4* and rhinovirus C receptor *CDHR3*. **b**, Violin plots (left) and dot plots (immediately above violin plots) showing expression of coronavirus receptors *ACE2*, *DPP4*, and *ANPEP* in lung cell types (10x dataset, cell numbers given in Supplementary Table 2). Grey shading, cell types inhaled viruses can directly access. Doughnut plots (right) showing relative number of receptor-expressing cells of cell types viruses can directly access (shaded grey in **a**), normalized by their abundance values from Supplementary Table 1 (and refined by the relative abundance values in Fig. 1 and Extended Data Figs. 3 and 4). Note prevalence of AT2 alveolar cells for *ACE2*, receptor for SARS-CoV and SARS-CoV-2, and for *DPP4*, receptor for MERS-CoV, in contrast to prevalence of macrophages for *ANPEP*, receptor for common cold causing coronavirus 229E. For more details on statistics and reproducibility, see Methods.

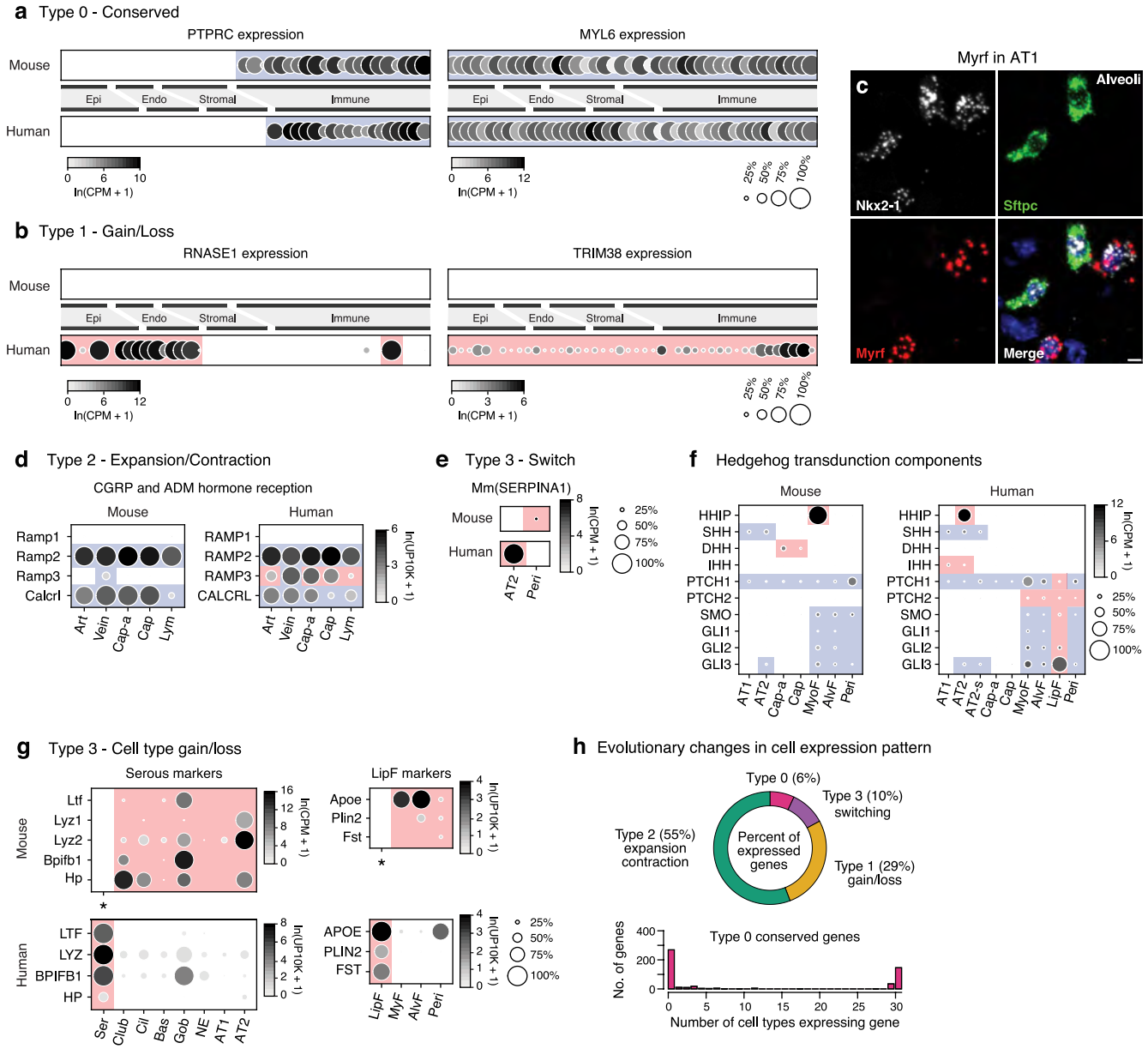


Extended Data Fig. 10 | Lung cell expression patterns of non-respiratory virus receptors. Dot plot of expression of entry receptors for non-respiratory viruses in human lung cell types (compare with Extended Data Fig. 9a showing expression of receptors for respiratory viruses). For more details on statistics and reproducibility, see Methods.



Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | Comparison of mouse and human gene expression profiles in homologous lung cell types and across age. a, Scatter plots showing median expression levels (ln(CPM + 1)) in indicated cell types of each expressed human gene and mouse orthologue (mouse and human SS2 datasets, human and mouse cell numbers given in Supplementary Tables 2 and 6, respectively). Note tens to hundreds of genes that show a 20-fold or greater expression difference (and $P < 0.05$, MAST) between species (red dots, gene names indicated for some and total number given above). Basophil/mast cell 1 (Bas/Ma 1) cells have the most differentially expressed genes (343), and $CD4^+$ M/E T cells have the least (79). Pearson correlation scores (*R* values) between the average mouse and human gene expression profiles for each cell type are indicated. 'Mm()' and 'Hs()' denotes genes in which duplications between mouse and human were collapsed to HomologyID. **b**, Heat map showing global transcriptome Pearson correlation between indicated human and mouse epithelial cells (SS2 dataset, human and mouse cell numbers given in Supplementary Tables 2 and 6, respectively). Red outline denotes homologous cell types based on classical markers described in Supplementary Table 6. White dot denotes human-to-mouse correlation. **c**, Dot plot of expression of canonical goblet cell markers *MUC5B* and *MUC5AC* and transcription factor *SPDEF* in mouse (left) and human (right) goblet cells. **d**, Scatter plot showing average expression levels (dots) across all cells ('pseudo-bulk' lung expression) of each expressed human gene and mouse orthologue (mouse and human SS2 datasets). Scale, ln(CPM + 1). Pearson correlation (*R* values) between the average mouse and human gene expression profiles are indicated. **e**, Scatter plots comparing median expression levels (ln(CPM + 1)) in indicated mouse lung cell types of each expressed gene at age 3 months (*x* axis) and 24 months (*y* axis) in SS2 datasets from Tabula Muris Senis[40] (cell numbers given in Supplementary Table 6). Pearson correlation scores between average gene expression profile for each cell type at each age are indicated (*R* values), along with number of genes (red dots) showing 20-fold or greater expression difference (and $P < 0.05$, MAST) between ages. Names of some genes are given next to the corresponding red dot. For more details on statistics and reproducibility, please see Methods.



Extended Data Fig. 12 | Patterns of conserved and divergent gene expression across human and mouse lung cell types. a, Dot plots of *PTPRC* and *MYL6* expression in mouse and human lung cell types (SS2 datasets) showing two examples of conserved (type 0) expression pattern. Blue shading, homologous cell types with conserved expression. **b**, Dot plots showing gain of expression (type 1 change) in several human cell types of *RNASE1* (left) and all human cell types of *TRIM38* (right). Red shading, cell types with divergent (gained) expression. **c**, Alveolar section of adult mouse lung probed by smFISH for general alveolar epithelial marker *Nkx2-1*, AT2 marker *Sftpc*, and transcription factor *Myrf*. Note *Myrf* is selectively expressed in mouse AT1 cells (*Nkx2-1*$^{+}$ *Sftpc*$^{-}$ cells), as it is in humans (Extended Data Fig. 6c). Scale bar, 5 µm. Staining repeated on three mice. **d**, Dot plots of expression of CGRP and ADM hormone receptor genes showing expansion of expression (type 2 change) in human endothelial cells (10x data sets). **e**, Dot plots of expression of emphysema-associated gene *SERPINA1* showing switched expression (type 3 change) from mouse pericytes (top) to human AT2 cells (bottom) (SS2 datasets). **f**, Dot plots comparing expression and conservation of HHIP with those of other Hedgehog pathway genes including ligands (SHH, DHH, IHH), receptors (PTCH1, PTCH2, SMO), and transducers (GLI1, GLI2, GLI3) (SS2 datasets). **g**, Dot plots of expression of serous cell markers *LTF*, *LYZ*, *BPIFBP1* and *HP* showing switched expression (type 3 change) from mouse airway epithelial cells to human serous cells, which mice lack (asterisk). Dot plots of expression of lipid handling genes *APOE*, *PLIN2* and *FST* show switched expression (type 3 change) from mouse alveolar stromal cells to human lipofibroblasts, which mice lack (asterisks). 'Mm()' or 'Hs()', genes in which duplications between mouse and human were collapsed to HomologyIDs (10x and SS2 datasets). **h**, Pie chart of fraction of expressed genes in lung showing each of the four types of evolutionary changes in cellular expression patterns from mouse to human. Histogram below shows number of lung cell types that the 602 genes with perfectly conserved cellular expression patterns (type 0) are expressed in; note that almost all are expressed in either a single cell type (67%) or nearly all cell types (33%). For more details on statistics and reproducibility, see Methods.

nature research

Corresponding author(s): Mark Krasnow and Stephen Quake

Last updated by author(s): Nov 5, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Smart-Seq2 cell capture: Sony SH800S Cell Sorter software v1.8, BD Diva software v8. Smart-Seq2 cDNA synthesis and Library preparation: Mosquito Liquid Handler software, MANTIS Liquid Handler Control software. Library quality control: Tapestation or Fragment analyzer software. Microscopy: Zen Black
Data analysis	Smart-Seq2 sequence de-multiplexing: bcl2fastq v2.19.0.316. Smart-Seq2 alignment and gene counting: skewer v0.2.2, STAR v2.6.1d. 10x sequence de-multiplexing, alignment, and UMI counting: CellRanger v2.0.1 Filtering, clustering, and annotating cells: Seurat v2.3, R v3.6.3 Microscopy: Imaris v9.2.0, Fiji v2 Flow cytometry: Flowjo v10.5.3 Cell interactions: CellPhoneDB v2.1.2, python v3.7.2 The code for demultiplexing counts/UMI tables, clustering, annotation, downstream analyses, and obtaining source data/generating figures that include single cell expression data is available on GitHub (https://github.com/krasnowlab/HLCA).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Counts/UMI tables, cellular metadata, Seurat objects, and scanpy objects are available on Synapse (https://www.synapse.org/#!Synapse:syn21041850). The data can be explored in a browser using cellxgene at https://hlca.ds.czbiohub.org/. Human sequencing data is available by data access agreement on the European

Genome-phenome Archive (EGA) under accession EGAS00001004344. Use of human sequencing data is restricted to not for profit research only and requires approval or a waiver from requesting investigator's institutional review board. Mouse sequencing data is available on the National Institute of Health's Sequence Read Archive (SRA) under BioProject accession PRJNA632939. Source data behind immunostaining or smFISH quantification (Figure 1; Extended Data Figures 3 and 4) are available within the manuscript files.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Single cells were isolated from peripheral blood or surgically resected, unaffected lung tissue from patients undergoing pulmonary lobectomies for focal tumors. We sampled two distal regions from Patient 1, one distal and one medial region from Patient 2, and a distal, medial, and proximal region from Patient 3. Sample size was not predetermined. Cell atlas completeness was assessed by the percentage of canonical lung cell types captured, which was greater than 90% after 3 patients.
Data exclusions	Poor quality cells were excluded from clustering using pre-established thresholds for the number of reads (Smart-Seq2 <50000) or UMIs (10x <1000) and genes (<500) detected. Doublets were excluded by manual inspection following clustering, looking for evidence of coherent expression profiles from two cell types.
Replication	Flow cytometer, single molecule in situ hybridization, and immunohistochemistry experiments were consistent across patient samples or mice. >80% molecular types were observed in more than one patient either by sequencing or follow up experiments.
Randomization	Samples from human and mouse were all considered controls and were not randomized.
Blinding	Patients were de-identified and assigned an identifier. Researchers were aware of basic demographic information and relevant medical history, clinicians were aware of each patient's full history.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Magnetically conjucated antibodies: anti-human CD45 (Miltenyi 130-045-801, 1:50 dilution), human EPCAM (Miltenyi 130-061-101, 1:50) and human CD31 (Miltenyi 130-091-935, 1:100).

Fluorescently conjugated antibodies for lung single cell sorting, used at the manufacturers recommended concentration: anti-human CD45 (Biolegend 304006, clone HI30, https://www.biolegend.com/en-us/products/fitc-anti-human-cd45-antibody-707) and EPCAM (eBioscience 25-9326-42, clone 1B7, https://www.thermofisher.com/antibody/product/CD326-EpCAM-Antibody-clone-1B7-Monoclonal/25-9326-42).

Fluorescently conjugated antibodies for immune single cell sorting, used at 1:20 dilution: anti-human CD3 (BD 563548, clone UCHT1, https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/th-1-cells/surface-markers/human/buv395-mouse-anti-human-cd3-ucht1-also-known-as-ucht-1-ucht-1/p/563548), CD4 (BD 340443, clone SK3, https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/th-1-cells/surface-markers/human/apc-mouse-anti-human-cd4-sk3-also-known-as-leu3a/p/340443), CD8 (BD 340692, clone SK1, https://www.bdbiosciences.com/us/applications/clinical/blood-cell-disorders/asr-reagents/cd8-fitc-sk1/p/340692), CD14 (BD 557831, clone MPP9, https://www.bdbiosciences.com/eu/applications/research/stem-cell-research/hematopoietic-stem-cell-markers/human/negative-

markers/apc-cy7-mouse-anti-human-cd14-mp9-also-known-as-mp-9/p/557831), CD19 (Biolegend 302234, clone HIB19, https://www.biolegend.com/en-us/products/brilliant-violet-421-anti-human-cd19-antibody-7144), CD47 (BD 563761, clone B6H12, https://www.bdbiosciences.com/eu/applications/research/stem-cell-research/cancer-research/human/bv711-mouse-anti-human-cd47-b6h12/p/563761), CD56 (BD 555516, clone B159, https://www.bdbiosciences.com/us/applications/research/stem-cell-research/hematopoietic-stem-cell-markers/human/negative-markers/pe-mouse-anti-human-cd56-b159/p/555516), and CD235a (BD 559944, clone GA-R2, https://www.bdbiosciences.com/us/reagents/research/antibodies-buffers/immunology-reagents/anti-human-antibodies/cell-surface-antigens/pe-cy5-mouse-anti-human-cd235a-ga-r2-hir2/p/559944)

Fluorescent antibodies for immune bulk cell sorting, used at 1:20 dilution: anti-human CD16 (BD 558122, clone 3G8, https://www.bdbiosciences.com/us/applications/research/stem-cell-research/cancer-research/human/pacific-blue-mouse-anti-human-cd16-3g8/p/558122), CD123 (BD 560826, clone 7G3, https://www.bdbiosciences.com/us/applications/research/b-cell-research/surface-markers/human/pe-cy7-mouse-anti-human-cd123-7g3/p/560826), CCR3 (R&D FAB155F, clone 61828, https://www.rndsystems.com/products/human-ccr3-fluorescein-conjugated-antibody-61828_fab155f), ITGB7 (BD 551082, clone FIB504, https://www.bdbiosciences.com/us/reagents/research/antibodies-buffers/immunology-reagents/anti-human-antibodies/cell-surface-antigens/apc-rat-anti-integrin-7-fib504/p/551082), CD3 (BD 555341, clone HIT3a, https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/th-1-cells/surface-markers/human/pe-cy5-mouse-anti-human-cd3-hit3a/p/555341), CD14 (Invitrogen MHCD1406, clone TuK4, https://www.thermofisher.com/antibody/product/CD14-Antibody-clone-TuK4-Monoclonal/MHCD1406), CD19 (BD 555414, clone HIB19, https://www.bdbiosciences.com/eu/applications/research/clinical-research/oncology-research/blood-cell-disorders/surface-markers/human/pe-cy5-mouse-anti-human-cd19-hib19/p/555414), and CD56 (BD 555517, clone B159, https://www.bdbiosciences.com/us/applications/research/stem-cell-research/hematopoietic-stem-cell-markers/human/negative-markers/pe-cy5-mouse-anti-human-cd56-b159/p/555517) (“basophils, neutrophils and eosinophils”); anti-human CD16 (BD 558122, see above), CD14 (BD 347497, MPP9, https://www.bdbiosciences.com/us/applications/research/stem-cell-research/hematopoietic-stem-cell-markers/human/negative-markers/pe-mouse-anti-human-cd14-mp9-also-known-as-mp-9/p/347497), CD4 (BD 340443, see above), CD3 (BD 555341, see above), CD8 (BD 555368, clone RPA-T8, https://www.bdbiosciences.com/us/reagents/research/antibodies-buffers/immunology-reagents/anti-human-antibodies/cell-surface-antigens/pe-cy5-mouse-anti-human-cd8-rpa-t8/p/555368), CD19 (BD 555414, see above), and CD56 (BD 555517, see above) (“classical and nonclassical monocytes”); anti-human CD16 (BD 558122, see above), CD1c (Miltenyi Biotec 130-098-007, clone AD5-8E7, discontinued, new product https://www.miltenyibiotec.com/US-en/products/cd1c-bdca-1-antibody-anti-human-ad5-8e7.html), CD11c (BD 340544, clone S-HCL-3, https://www.bdbiosciences.com/us/reagents/research/clinical-research---ruo-gmp/single-color-antibodies/apc-mouse-anti-human-cd11c-s-hcl-3/p/340544), CCR3 (R&D FAB155F, see above), CD123 (BD 560826, see above), HLA-DR (BD 335796, clone L243, https://www.bdbiosciences.com/us/applications/research/stem-cell-research/mesenchymal-stem-cell-markers-bone-marrow/human/negative-markers/apc-cytrade7-mouse-anti-human-hla-dr-l243/p/335796), CD3 (BD 555341, see above), CD4 (BD 555348, clone RPA-T4, https://www.bdbiosciences.com/eu/applications/research/t-cell-immunology/th-1-cells/surface-markers/human/pe-cy5-mouse-anti-human-cd4-rpa-t4/p/555348), CD8 (BD 555368, see above), CD14 (Invitrogen MHCD1406, see above), CD19 (BD 555414, see above), and CD56 (BD 555517) (“pDCs, mDCs, CD16+ DCs”); anti-human IgM/IgD (BD 555778, clone IA6-2, https://www.bdbiosciences.com/us/applications/research/b-cell-research/immunoglobulins/human/fitc-mouse-anti-human-igd-ia6-2-also-known-as--ia6-2/p/555778), CD19 (BD 557835, clone SJ25C1, https://www.bdbiosciences.com/eu/applications/research/clinical-research/oncology-research/blood-cell-disorders/surface-markers/human/pe-cy7-mouse-anti-human-cd19-sj25c1-also-known-as-sj25-c1/p/557835), CD27 (BD 558664, clone M-T271, https://www.bdbiosciences.com/us/applications/research/clinical-research/oncology-research/blood-cell-disorders/surface-markers/human/apc-mouse-anti-human-cd27-m-t271/p/558664), CD20 (BD 335794, clone L27, https://www.bdbiosciences.com/us/applications/research/stem-cell-research/hematopoietic-stem-cell-markers/human/negative-markers/apc-cytrade7-mouse-anti-human-cd20-l27/p/335794), CD3 (BD 555341, see above), CD4 (BD 555348, see above), CD14 (Invitrogen MHCD1406, see above), and CD56 (BD 555517, see above) (“B cells”); anti-human CD16 (BD 558122, see above), CD57 (BD 347393, clone HNK-1, https://www.bdbiosciences.com/us/applications/research/t-cell-immunology/t-follicular-helper-tfh-cells/surface-markers/human/fitc-mouse-anti-human-cd57-hnk-1/p/347393), CD56 (BD 557747, clone B159, https://www.bdbiosciences.com/eu/applications/research/stem-cell-research/hematopoietic-stem-cell-markers/human/negative-markers/pe-cy7-mouse-anti-human-cd56-b159/p/557747), CD3 (BD 555341, see above), CD4 (BD 555348, see above), CD14 (Invitrogen MHCD1406, see above), and CD19 (BD 555414, see above) (“NK cells”); and anti-human CD45RA (Biolegend 304118, clone IV N906, https://www.biolegend.com/en-us/products/pacific-blue-anti-human-cd45ra-antibody-3339), CCR7 (R&D FAB197F, clone 150503, https://www.rndsystems.com/products/human-ccr7-fluorescein-conjugated-antibody-150503_fab197f), CD62L (BD 555544, clone DREG-56, https://www.bdbiosciences.com/eu/applications/research/t-cell-immunology/regulatory-t-cells/surface-markers/human/pe-mouse-anti-human-cd62l-dreg-56/p/555544), CD45RO (BD Pharmingen 560608, clone UCHL1, https://www.bdbiosciences.com/us/applications/research/b-cell-research/surface-markers/human/pe-cy7-mouse-anti-human-cd45ro-uchl1/p/560608), CD4 (BD 340443, see above), CD8 (BD 340584, clone SK1, https://www.bdbiosciences.com/us/reagents/research/clinical-research---ruo-gmp/single-color-antibodies/apc-mouse-anti-human-cd8-sk1/p/340584), CD11b (BD 555389, clone ICFR44, https://www.bdbiosciences.com/us/applications/research/stem-cell-research/mesenchymal-stem-cell-markers-bone-marrow/human/negative-markers/pe-cy5-mouse-anti-human-cd11b-icrf44-also-known-as-44/p/555389), CD14 (Invitrogen MHCD1406, see above), CD19 (BD 555414, see above), CD56 (BD 555517, see above) (“T cells”).

Primary antibodies used for immunohistochemistry: anti-proSP-C (rabbit, Chemicon AB3786, 1:250 dilution), HES1 (rabbit, Cell Signaling 11988S clone D6P2U, 1:100), MUC-1 (hamster, Thermo Scientific HM1630, clone MH1, 1:250), Ki67 (rat, DAKO M7249 clone MIB-1, 1:100), and Keratin-5 (chicken, Biolegend 905901, 1:100)

Validation

Antibodies used for magnetic cell separation were validated by Miltenyi Biotec from human tissue containing known target populations: CD45 (blood and bone marrow;https://www.miltenyibiotec.com/US-en/products/cd45-microbeads-human.html), EPCAM (lung adencarcinoma;https://www.miltenyibiotec.com/US-en/products/cd326-epcam-microbeads-human.html), and CD31 (foreskin;https://www.miltenyibiotec.com/US-en/products/cd31-microbead-kit-human.html). See indicated websites for specific details.

All antibodies for flow cytometry were validated against isotype controls in human cells by manufacturer for that application, see manufacturers' websites (noted above) for details.

Antibodies for immunohistochemistry were validated by their manufacturer in human tissue with a canonical staining pattern: pro-SP-C (lung;https://www.emdmillipore.com/US/en/product/Anti-Prosurfactant-Protein-C-proSP-C-Antibody,MM_NF-AB3786), HES1 (breast carcinoma;https://www.cellsignal.com/products/primary-antibodies/hes1-d6p2u-rabbit-mab/11988), MUC-1 (breast carcinoma;https://www.thermofisher.com/order/catalog/product/HM-1630-P1#/HM-1630-P1), Ki67 (tonsillar;https://www.agilent.com/en/product/immunohistochemistry/antibodies-controls/primary-antibodies/ki-67-antigen-(dako-omnis)-76239), Keratin 5 (skin;https://www.biolegend.com/en-us/products/keratin-5-polyclonal-chicken-antibody-purified-10957). See indicates websites for specific details.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	1) 2m old female B6,Axin2-CreER (https://www.jax.org/strain/018867, heterozygous) bred into B6,mTmG (https://www.jax.org/strain/007676, heterozygous) 2) 2m old female FVB,Tbx4-Cre (http://www.informatics.jax.org/allele/MGI:5635865, heterozygous) bred into B6,Ai6 (https://www.jax.org/strain/007906, heterozygous) 3) Tabula Muris Senis: C57J/B6 mice, male and female, ages 1m, 3m, 12m, 18m, 21m, 24m, and 30m. All animals were species Mus musculus.
Wild animals	Study did not involve wild animals.
Field-collected samples	Study did not involve samples collected from the field.
Ethics oversight	All mouse experiments followed applicable regulations and guidelines and were approved by the Institutional Animal Care and Use Committee at Stanford University (Protocol 9780).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Patient ages (in order) were 75, 46, and 51; sexes male, female, and male; all patients had normal pulmonary function tests and were otherwise healthy except for early stage focal tumors.
Recruitment	Patients were recruited and consented by C.S.K. and J.B.S after selecting surgical resection with curative intent as their treatment option for early-stage, focal tumors. Stanford Hospital's patient pool is not necessarily representative of the broader human population and there may be additional cell types and states found from future studies from other centers. While we selected subjects with no history of serious pulmonary disease (other than their focal lung tumors) and included both genders in our study, expression profiles from captured cell types will vary to some extent with age, ethnicity, gender, socioeconomic status, and health status.
Ethics oversight	Patient tissues were obtained under a protocol approved by Stanford University's Human Subjects Research Compliance Office (IRB 15166) and informed consent was obtained from each patient prior to surgery. All experiments followed applicable regulations and guidelines.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Tissue source Freshly resected lung tissue was procured intraoperatively from patients undergoing lobectomy for focal lung tumors. Normal lung tissues (~5 cm3) were obtained from uninvolved regions and annotated for the specific lung lobe and location along the airway or periphery. Pathological evaluation (by G.B.) confirmed normal histology of the profiled regions, except for areas of very mild emphysema in Patient 1. Patient 1 was a 75 year-old male with a remote history of smoking, diagnosed with early stage adenocarcinoma who underwent left upper lobe (LUL) lobectomy; two blocks of normal tissue were obtained from lung periphery ("Distal 1a and 1b").

Patient 2 was a 46 year-old male, non-smoker with a right middle lobe (RML) endobronchial carcinoid, who underwent surgical resection of the right upper and middle lobes; two blocks of tissue were selected from mid-bronchial region ("Medial 2") and periphery ("Distal 2") of right upper lobe (RUL). Patient 3 was a 51 year-old female, non-smoker with a LLL endobronchial typical carcinoid, who underwent LLL lobectomy; three tissue blocks were resected from the bronchus ("Proximal 3"), mid-bronchial ("Medial 2"), and periphery ("Distal 3") of the LLL. All tissues were received and immediately placed in cold phosphate buffered saline (PBS) and transported on ice directly to the research lab for single cell dissociation procedures. Peripheral blood was collected from patients 1 and 3 in EDTA tubes.

Lung cell processing and staining

Individual human lung samples were dissected, minced, and placed in digestion media (400 µg/ml Liberase DL (Sigma 5401127001) and 100 µg/ml elastase (Worthington LS006365) in RPMI (Gibco 72400120) in a gentleMACS c-tube (Miltenyi 130-096-334). Samples were partially dissociated by running 'm_lung_01' on a gentleMACS Dissociator (Miltenyi 130-093-235), incubated on a Nutator at 37°C for 30 minutes, and then dispersed to a single cell suspension by running 'm_lung_02'. Processing buffer (5% fetal bovine serum in PBS) and DNAse I (100 µg/ml, Worthington LS006344) were then added and the samples rocked at 37°C for 5 minutes. Samples were then placed at 4oC for the remainder of the protocol. Cells were filtered through a 100 µm filter, pelleted (300 x g, 5 minutes, 4°C), and resuspended in ACK red blood cell lysis buffer (Gibco A1049201) for 3 minutes, after which the buffer was inactivated by adding excess processing buffer. Cells were then filtered through a 70 µm strainer (Fisherbrand 22363548), pelleted again (300 x g, 5 minutes, 4°C), and resuspended in magnetic activated cell sorting (MACS) buffer (0.5% BSA, 2 mM EDTA in PBS) with Human FcR Blocking Reagent (Miltenyi 130-059-901) to block non-specific binding of antibodies (see below).

Immune and endothelial cells were overrepresented in our previous mouse single cell suspensions. To partially deplete these populations in our human samples, we stained cells isolated from lung with MACS microbeads conjugated to CD31 and CD45 (Miltenyi 130-045-801, 130-091-935) then passed them through an LS MACS column (Miltenyi, 130-042-401) on a MidiMACS Separator magnet (Miltenyi, 130-042-302). Cells retained on the column were designated "immune and endothelial enriched." The flow through cells were then split, with 80% immunostained for FACS (see below) and the remaining 20% stained with EPCAM microbeads (Miltenyi 130-061-101). EPCAM stained cells were passed through another LS column. Cells retained on the column were labeled "epithelial enriched", and cells that flowed through were designated "stromal".

Following negative selection against immune and endothelial cells by MACS, the remaining human lung cells were incubated with FcR Block (Becton Dickinson (BD) 564219) for 5 minutes and stained with directly conjugated anti-human CD45 (Biolegend 304006) and EPCAM (eBioscience 25-9326-42) antibodies on a Nutator for 30 minutes. Cells were then pelleted (300 x g, 5 minutes, 4°C), washed with FACS buffer three times, then incubated with cell viability marker Sytox blue (1:3000, ThermoFisher S34857).

Immune cell processing and staining

Immune cells, including granulocytes, were isolated from peripheral blood using a high density ficoll gradient56. Briefly, peripheral blood was diluted 10-fold with FACS buffer (2% FBS in PBS), carefully layered on an RT Ficoll gradient (Sigma HISTOPAQUE®-1119), and centrifuged at 400 x g for 30 minutes at room temperature. The buffy coat was carefully removed, diluted 5-fold with FACS buffer, pelleted (300 x g, 5 minutes, 4°C), and incubated in ice cold FACS buffer containing DNAse I (Worthington LS006344) for 10 minutes at 4°C. Clumps were separated by gentle pipetting to create a single cell suspension.

Immune cells from subject matched blood were incubated with FcR Block and Brilliant Violet buffer (BD 563794) for 20 minutes and then stained with directly conjugated anti-human CD3 (BD 563548), CD4 (BD 340443), CD8 (BD 340692), CD14 (BD 557831), CD19 (Biolegend 302234), CD47 (BD 563761), CD56 (BD 555516), and CD235a (BD 559944) antibodies for 30 minutes. Cells were pelleted (300 x g, 5 minutes, 4°C), washed with FACS buffer twice, and then incubated with the viability marker propidium iodide.

Immune bulk sort processing and staining

Immune cells for bulk mRNA sequencing were incubated with Fc Block for 20 minutes and then stained with one of six panels of directly conjugated antibodies for 30 minutes: anti-human CD16 (BD 558122), CD123 (BD 560826), CCR3 (R&D FAB155F), ITGB7 (BD 551082), CD3 (BD 555341), CD14 (Invitrogen MHCD1406), CD19 (BD 555414), and CD56 (BD 555517)

("basophils, neutrophils and eosinophils"); anti-human CD16 (BD 558122), CD14 (BD 347497), CD4 (BD 340443), CD3 (BD 555341), CD8 (BD 555368), CD19 (BD 555414), and CD56 (BD 555517) ("classical and nonclassical monocytes"); anti-human CD16 (BD 558122), CD1c (Miltenyi Biotec 130-098-007), CD11c (BD 340544), CCR3 (R&D FAB155F), CD123 (BD 560826), HLA-DR (BD 335796), CD3 (BD 555341), CD4 (BD 555348), CD8 (BD 555368), CD14 (Invitrogen MHCD1406), CD19 (BD 555414), and CD56 (BD 555517) ("pDCs, mDCs, CD16+ DCs"); anti-human IgM/IgD (BD 555778), CD19 (BD 557835), CD27 (BD 558664), CD20 (BD 335794), CD3 (BD 555341), CD4 (BD 555348), CD14 (Invitrogen MHCD1406), and CD56 (BD 555517) ("B cells"); anti-human CD16 (BD 558122), CD57 (BD 347393), CD56 (BD 557747), CD3 (BD 555341), CD4 (BD 555348), CD14 (Invitrogen MHCD1406), and CD19 (BD 555414) ("NK cells"); and anti-human CD45RA (Biolegend 304118), CCR7 (R&D FAB197F), CD62L (BD 555544), CD45RO (BD Pharmingen 560608), CD4 (BD 340443), CD8 (BD 340584), CD11b (BD 555389), CD14 (Invitrogen MHCD1406), CD19 (BD 555414), CD56 (BD 555517) ("T cells"). Cells were washed with FACS buffer twice, incubated with the viability marker propidium iodide.

Instrument

Lung samples: Sony SH800S, Blood samples: FACS Aria II

Software

Lung samples: Sony SH800S Cell Sorter software v1.8, Blood samples: BD Diva software v8

Cell population abundance

We sorted equal numbers of epithelial, immune, and endothelial/stromal cells with the gating strategy described below. The percentage of expression profiles from each tissue compartment is consistent with the sorting (30% endothelial/immune, 34% epithelial, 36% immune).

Gating strategy

Lung samples: Living single cells (Sytox blue-negative) were sorted into lysis plates based on three gates: EPCAM+CD45- (designated "epithelial"), EPCAM-CD45+ (designated "immune"), and EPCAM-CD45- (designated "endothelial or stromal"). Blood samples: Living (propidium iodide-negative) single, non-red blood (CD235a-) cells were sorted into lysis plates along with specific immune populations: B cells (CD19+CD3-), CD8+ T cells (CD8+), CD4+ T cells (CD4+), NK cells (CD19-CD3-CD56+CD14-), classical monocytes (CD19-CD3-CD56-CD14+). (See Extended Data Figure S1).

Please see Extended Data Table S3 for sorting strategy from bulk sorted immune populations.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.