SYSTEMS BIOLOGY

Single-cell connectomic analysis of adult mammalian lungs

Micha Sam Brickman Raredon[1,2,3], Taylor Sterling Adams[4], Yasir Suhail[5], Jonas Christian Schupp[4], Sergio Poli[6], Nir Neumark[4,7], Katherine L. Leiby[1,2,3], Allison Marie Greaney[1,2], Yifan Yuan[8], Corey Horien[3,9], George Linderman[3,10], Alexander J. Engler[1,2], Daniel J. Boffa[11], Yuval Kluger[7,10,12], Ivan O. Rosas[6], Andre Levchenko[1,5], Naftali Kaminski[4]*, Laura E. Niklason[1,2,8]*

Efforts to decipher chronic lung disease and to reconstitute functional lung tissue through regenerative medicine have been hampered by an incomplete understanding of cell-cell interactions governing tissue homeostasis. Because the structure of mammalian lungs is highly conserved at the histologic level, we hypothesized that there are evolutionarily conserved homeostatic mechanisms that keep the fine architecture of the lung in balance. We have leveraged single-cell RNA sequencing techniques to identify conserved patterns of cell-cell cross-talk in adult mammalian lungs, analyzing mouse, rat, pig, and human pulmonary tissues. Specific stereotyped functional roles for each cell type in the distal lung are observed, with alveolar type I cells having a major role in the regulation of tissue homeostasis. This paper provides a systems-level portrait of signaling between alveolar cell populations. These methods may be applicable to other organs, providing a roadmap for identifying key pathways governing pathophysiology and informing regenerative efforts.

INTRODUCTION

Cell-cell signaling via soluble and insoluble cues is known to be crucial to the regulation of cell niches and the emergence of tissue properties (*1*, *2*). However, signals that govern tissue homeostasis are challenging to capture in vivo because of their high complexity and sensitivity to the extracellular environment (*3–5*). Recent advances in single-cell RNA sequencing (scRNAseq) have allowed unprecedented analysis of retina, hypothalamus, cortex, intestine, blood, heart, lung, and other organs (*6–18*). An ambitious international plan is currently underway to consolidate single-cell data for every tissue in the human body (*19*, *20*). To date, however, there has been little analysis across species to identify signaling patterns that might fundamentally govern tissue development and homeostasis. Lung tissue structure and function in mammals is highly conserved, with stereotypical sizes of alveoli, blood-air diffusion distances, and geometries of airway branching (*21*). To determine whether stereotyped patterns of cell-cell communication accompany these conserved aspects of lung structure and function, we generated and analyzed scRNAseq data from lung tissue across four mammalian species (mouse, rat, pig, and human). By mapping our data against the FANTOM5 database of mammalian protein-protein interactions, we captured a complex interplay of molecular factors contributing to distal lung organization and function (*14*, *22*). Through species compari-

sons, we were able to quantify conserved and nonconserved patterns of homeostatic signaling in adult mammalian lung tissue. Leveraging graph theory–based connectomic tools, we demonstrate substantial evolutionary conservation in overarching cell-cell signaling network topologies. Further, we synthesize findings across species to illustrate cell type–specific niches and to identify key signaling roles for specific cell types. The resulting map presents a data-driven, systems-level portrait of the alveolus. Specific findings that can be explored with this map include extensive mesenchymal-epithelial interactions, macrophage-matrix engagement, and mechanistic underpinnings of alveolar type I (ATI) cell influence on microvascular endothelia and parenchymal homeostasis.

RESULTS

Single-cell profiling of lung creates a cross-species pulmonary atlas

To obtain single-cell data from healthy adult lung tissues, we isolated lungs from mice, rats, pigs, and human donors and dissociated them using common enzymatic digestions (see Materials and Methods). Samples were sequenced using 10× genomics in a 1-individual:1-batch ratio to allow deconvolution of any individual outliers. A complete summary of all batches can be found in table S1. The final dataset passing quality control (Fig. 1 and fig. S1) contained 17,867 healthy human cells (14 individuals, mixed gender, age 21 to 88), 11,452 cells from pig (2 individuals, male/male, 3.5 months old), 10,953 cells from rat (2 individuals, male/female, 9 weeks old), and 7744 cells from mouse (2 individuals, male/female, 9 weeks old). Cells had a mean read depth of ~88,000 reads per cell. Mean mitochondrial fraction across the entire dataset was 2.09%, indicating excellent transcriptional preservation. Batch-specific distribution for transcript count, gene count, and mitochondrial read fraction can be found in fig. S2 (A to C). Each species was clustered independently using a canonical correlation analysis (CCA), implemented in Seurat, to align the batches (fig. S1). Clustering in Seurat resolved approximately two-dozen well-demarcated clusters from each mammalian species,

[1]Department of Biomedical Engineering, Yale University, New Haven, CT 06511, USA. [2]Vascular Biology and Therapeutics, Yale University, New Haven, CT 06520, USA. [3]Medical Scientist Training Program, Yale School of Medicine, New Haven, CT 06510, USA. [4]Section of Pulmonary, Critical Care and Sleep Medicine, Yale University, New Haven, CT 06520, USA. [5]Yale Systems Biology Institute, Yale University, West Haven, CT 06516, USA. [6]Pulmonary and Critical Care Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. [7]Interdepartmental Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06510, USA. [8]Department of Anesthesiology, Yale University, New Haven, CT 06510, USA. [9]Interdepartmental Neuroscience Program, Yale University, New Haven, CT 06510, USA. [10]Applied Mathematics Program, Yale University, New Haven, CT 06511, USA. [11]Thoracic Surgery, Yale School of Medicine, New Haven, CT 06510, USA. [12]Department of Pathology, Yale University, New Haven, CT 06520, USA.
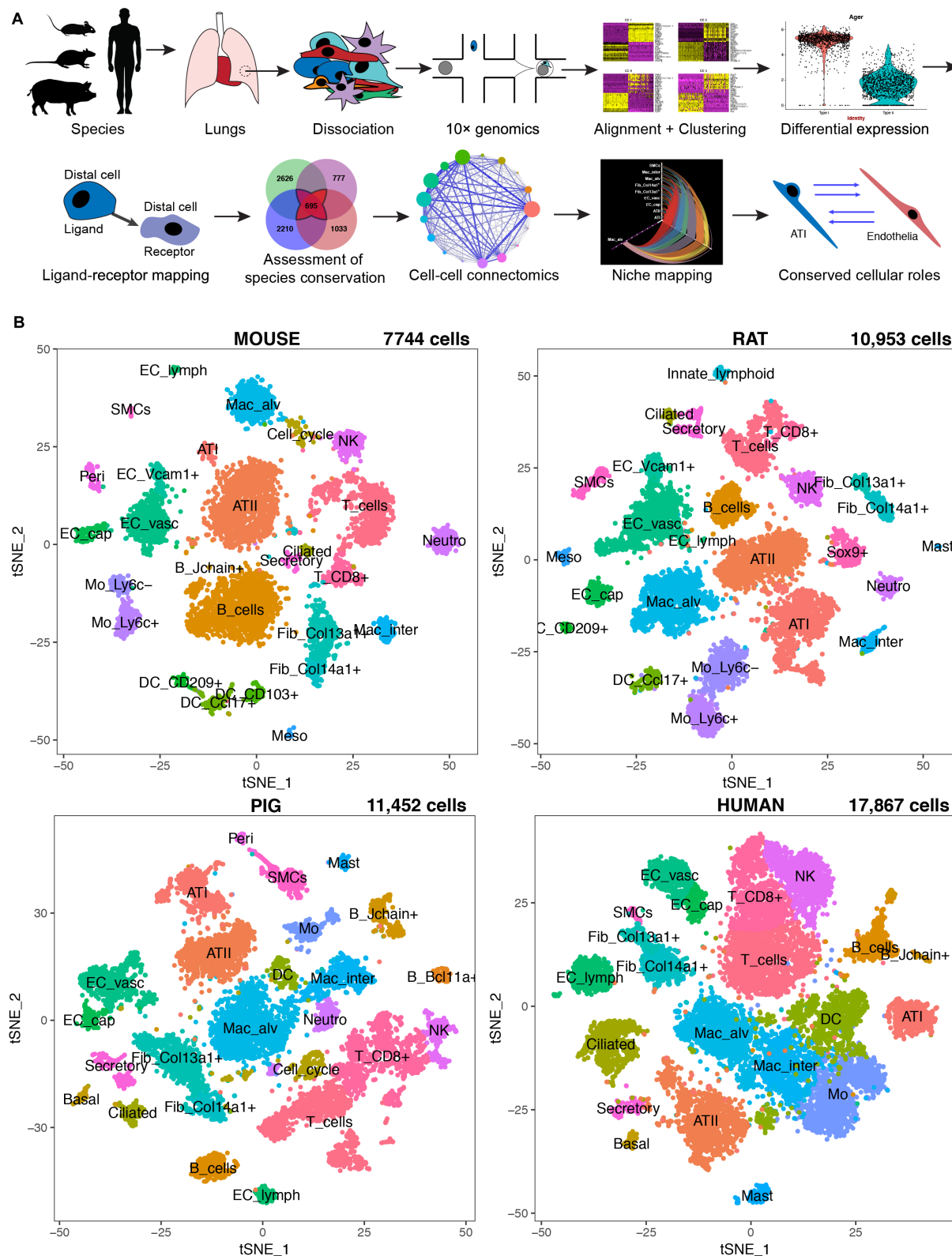*Corresponding author. Email: naftali.kaminski@yale.edu (N.K.); laura.niklason@yale.edu (L.E.N.)

Fig. 1. Cross-species pulmonary scRNAseq dataset. (**A**) Lungs from mice, rats, pigs, and human donors were dissociated, barcoded using 10× genomics, and sequenced on Illumina sequencers. Cells were clustered following genome alignment and dimensionality reduction. Markers were identified via differential expression and correlated to literature and the Human Protein Atlas (*24*) and in-house immunostaining. Comparative ligand-receptor analysis reveals stereotyped cellular roles in tissue homeostasis. (**B**) A comparable, heterogeneous community of cells was profiled in all four species. Species-conserved cell types include type I, type II, ciliated, and secretory epithelium; capillary, vascular, and lymphatic endothelium; B cells, T cells, and natural killer (NK) cells; SMCs; Col13[+] and Col14[+] fibroblasts; and two distinct subsets of macrophages.

with species-variable distribution of cell types (Fig. 1B and fig. S2D). Figure 2 shows a limited marker atlas for all cell types, allowing ready comparison across species. Figure S3 shows a more in-depth marker atlas, across species, organized by cell class. Diverse epithelial populations were captured including ATI and ATII cells and ciliated, secretory, and basal epithelium. We identified four major stromal populations that were conserved across species, namely, Col13a1[+] fibroblasts, Col14a1[+] fibroblasts, and smooth muscle cells (SMCs) (elsewhere referred to as lipofibroblasts, interstitial fibroblasts, and myofibroblasts) (18, 23), as well as mesothelial stromal cells. Because the separation of cell types within the pulmonary mesenchyme is difficult, fibroblasts, pericytes, and SMCs were subset from each species and were then reclustered independently to better explore cross-species clustering and analogous labeling (see fig. S4). As there is a significant shared character between the two populations of fibroblast on both scRNAseq and histology, we were unable to identify definitively segregated locations for each population based on histology. This task is fundamentally complicated given the sparsity of positive and specific cluster markers for the Col14a1[+] population.

On the basis of histologic data from the Human Protein Atlas (24) and in-house immunostaining in mouse, rat, and human (selected examples are shown in Fig. 2, B to D), we identified two highly distinct categories of blood endothelium in the lung, which we assigned on the basis of histologic location as gas-exchanging capillary endothelium and mixed vascular endothelium, with some representation in the microvascular region. In addition, there is a cluster of lymphatic endothelium (Fig. 1B). Further, we profiled a wide variety of immune cells that are resident in lung, including B cells, T cells, natural killer cells, monocytes, dendritic cells, neutrophils, and multiple populations of macrophages. To quantitatively compare cell clusters across species and to assure analogous labeling, we converted nonhuman gene names to human homologs if possible (see Materials and Methods). Jaccard indices were then calculated for each cluster-cluster comparison across all species. A representation of analogous labeling of cell types across species is shown in fig. S5. These results and cell identity assignments align well with the recent scRNAseq literature (18, 23).

Comparative ligand-receptor connectomics identifies conserved patterns of extracellular signaling

To explore cross-species similarity in global signaling topologies, we mapped all possible vectors of ligand-receptor signaling in each species and then compared them. Following cell cluster identification, cluster-wise mean expression was calculated for all known and literature-supported ligands and cell surface receptors present in the FANTOM5 database (Fig. 3A) (25). When two cell clusters expressed a cognate ligand-receptor pair in greater than 5% of cells, an edge connecting the two clusters was created. Edge weight for network analysis was calculated as the sum of the two scaled expression values (see Supplementary Text for reasoning). The signaling molecules that were considered included growth factors, cytokines, chemotactic proteins, hormones, secreted matrix proteins, and cell-cell adhesion molecules. Each ligand-receptor pair was manually classified into 1 of 38 signaling families on the basis of established literature (see data file S1). Some of these ligands were pan-expressed in most cell types at low levels, such as ICAM-1. Others, such as VEGFA, were markedly elevated in certain cell clusters, suggesting biologically relevant producers or receivers (Fig. 3, B and C). The resulting mapping created a weighted, directed, and multiplexed network profiling potential cell-cell interactions within the adult lung, which quantitatively rewards outlying producers and receivers for each signaling mechanism (Fig. 3D).

To quantitatively compare signaling networks across the four species, we calculated network centrality metrics for the limited collection of cells found across all species. Centrality rankings were created on the basis of the sum of all edge weights. Kleinberg hub and authority scores [an indicator of community influence (hub) and reception (authority)] and general network degree rankings (an indicator of bidirectional connectivity) were both found to be highly conserved across species. Mesenchymal cells generally displayed the highest level of connectivity with other cell types across all four species (Fig. 3D), largely due to their high production of matrix proteins that can interact with integrins on many cell types. In contrast, lymphoid (B and T) cells generally displayed the lowest level of connectivity, which was surprising given their immune character, but in line with their transient presence within parenchymal tissue. Unexpectedly, ATI cells were consistently identified as a major signaling hub at a level approaching that of mesenchymal cells. Initial investigations revealed that this high centrality of ATI cells was backed by their expression of key growth factor genes including BDNF, SEMA3E, VEGFA, and PDGFA, as well as multiple WNT ligand genes (violin plots demonstrating these species-conserved patterns are shown in fig. S6). The expression of these diverse growth factors suggests that ATI cells, in addition to their role in gas exchange, may play a previously underappreciated role in regulating other cell types in the distal lung by their ability to produce soluble signals within the adult mammalian alveolus.

To directly quantify the conservation of the ranking of nodal signaling across the four species studied, we calculated node degree and Spearman correlation coefficients for node rank for multiple cell centrality rankings (using the human dataset as a reference) over increasing expression thresholds (Fig. 3, E and F). In this test, a ρ value of −1 corresponded to an inverse rank correlation, 0 corresponded to a complete lack of correlation, and 1 represented a perfect rank correlation. The relative rankings of cell-specific network centrality were found to be highly correlated to human across species (ρ > 0.75), demonstrating species-conserved patterns of signaling topology. Further, this finding held over increasingly strict expression-level cutoffs (Fig. 3F), demonstrating that species conservation of global signaling architecture is not a spurious finding due to arbitrary thresholding of data. These findings collectively suggested that although cells in the different species are not transcriptionally identical, systems-level patterns of signaling and connectivity were still highly conserved in adult mammalian lung.

Recipient cell–specific networks illustrate niche character

To explore alveolar cell niches, further analysis focused on species-conserved interactions between physically near alveolar cell types (Fig. 4A). Consideration of histologic proximity is essential in studies of cell-cell communication, given that many signaling molecules, such as Ephrins and WNTs, require short distances for successful signal transduction (26–28). For the purpose of this and all further analysis, species-conserved edges were defined as those found in human and at least one other species, with a cross-species cell-cell interaction rank P value (see Materials and Methods) of <0.05. Edge weights were averaged across species to yield mean weights. This master edge list of species-conserved signaling modalities can be interactively explored in our online resource (lungconnectome.net).
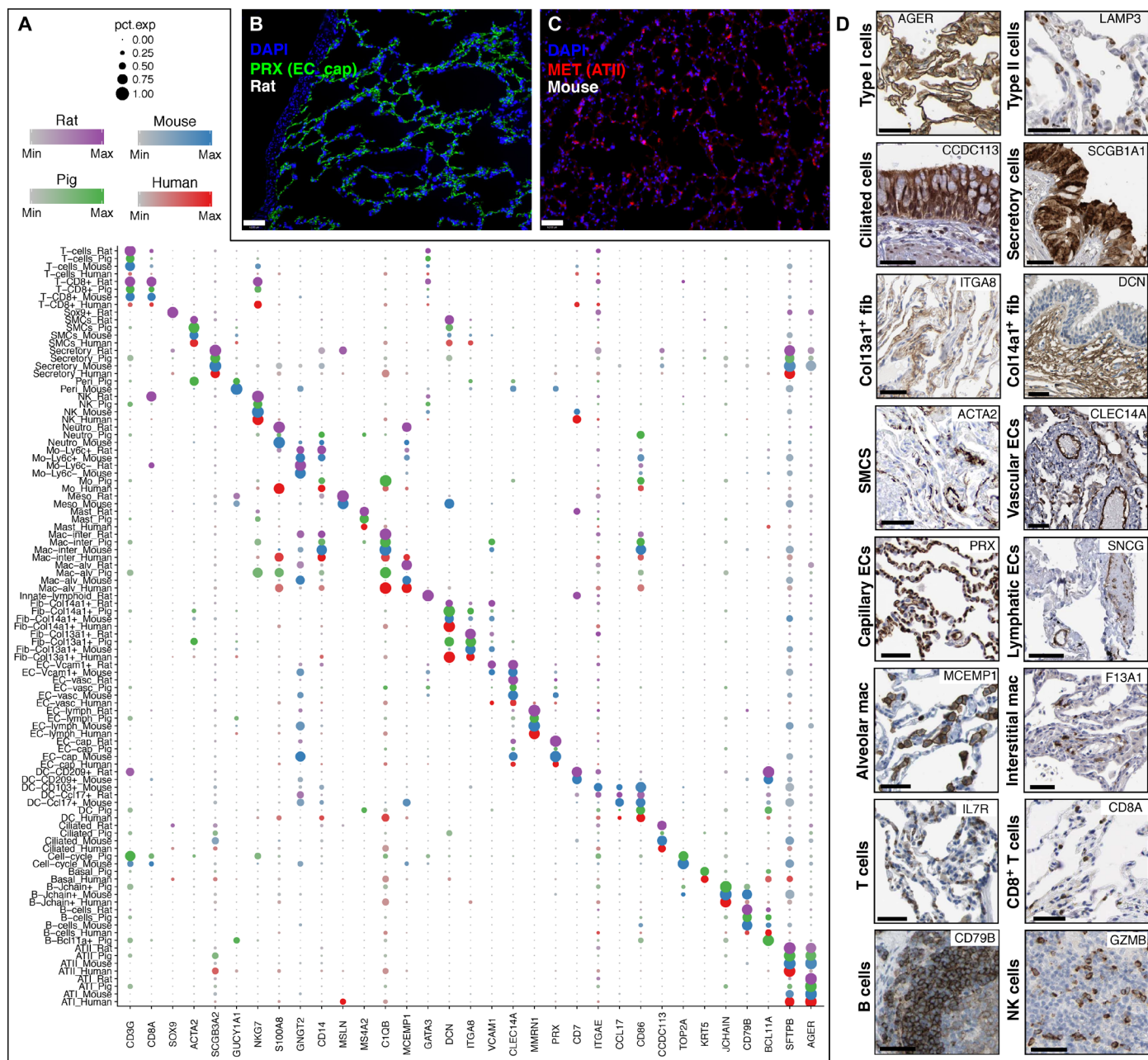
Fig. 2. Cross-species cell atlas representation. (**A**) The species-split DotPlot shows cross-species comparison of selected top markers for all cell populations profiled across all species. Color saturation indicates the strength of expression in positive cells, while dot size reflects the percentage of each cell cluster expressing the gene. Each color represents one species. Certain cell types were only profiled in certain species, such as basal cells marked by KRT5 (pig and human) and Sox9+ epithelium (rat). ITGA8/Itga8 reliably marks the Col13a1+ fibroblast population across species. Histology for highly expressed specific markers of conserved cell types, from the Human Protein Atlas [immunoperoxidase images (**D**)] and from in-house immunostaining [immunofluorescent images (**B** and **C**)], shows positive physical locations of identified cell type–specific markers. Markers shown are those that were present in the human dataset having homologs in the nonhuman species. Scale bars, 50 μm (D) and 62 μm (B and C).

Figure 4 (B and C) shows two hive plot representations of cellular niche character. The total set of alveolar niches is shown in fig. S7. These plots allow visualization of the relative contribution of each cell cluster and, from these clusters, each signaling family to the niche for one particular cell type. Alveolar macrophages (Fig. 4B) express a number of cell-cell adhesion receptors and, as such, are in a position to interact with a broad variety of cells. Capillary endothelia (Fig. 4C), on the other hand, strongly express receptors such as KDR, FLT1, and NRP1, and their niche plot therefore shows a large contribution from VEGF family signaling. Preliminary niche explorations revealed extensive mesenchymal-epithelial cross-talk and high production of vascular growth factors and morphogens by ATI cells.
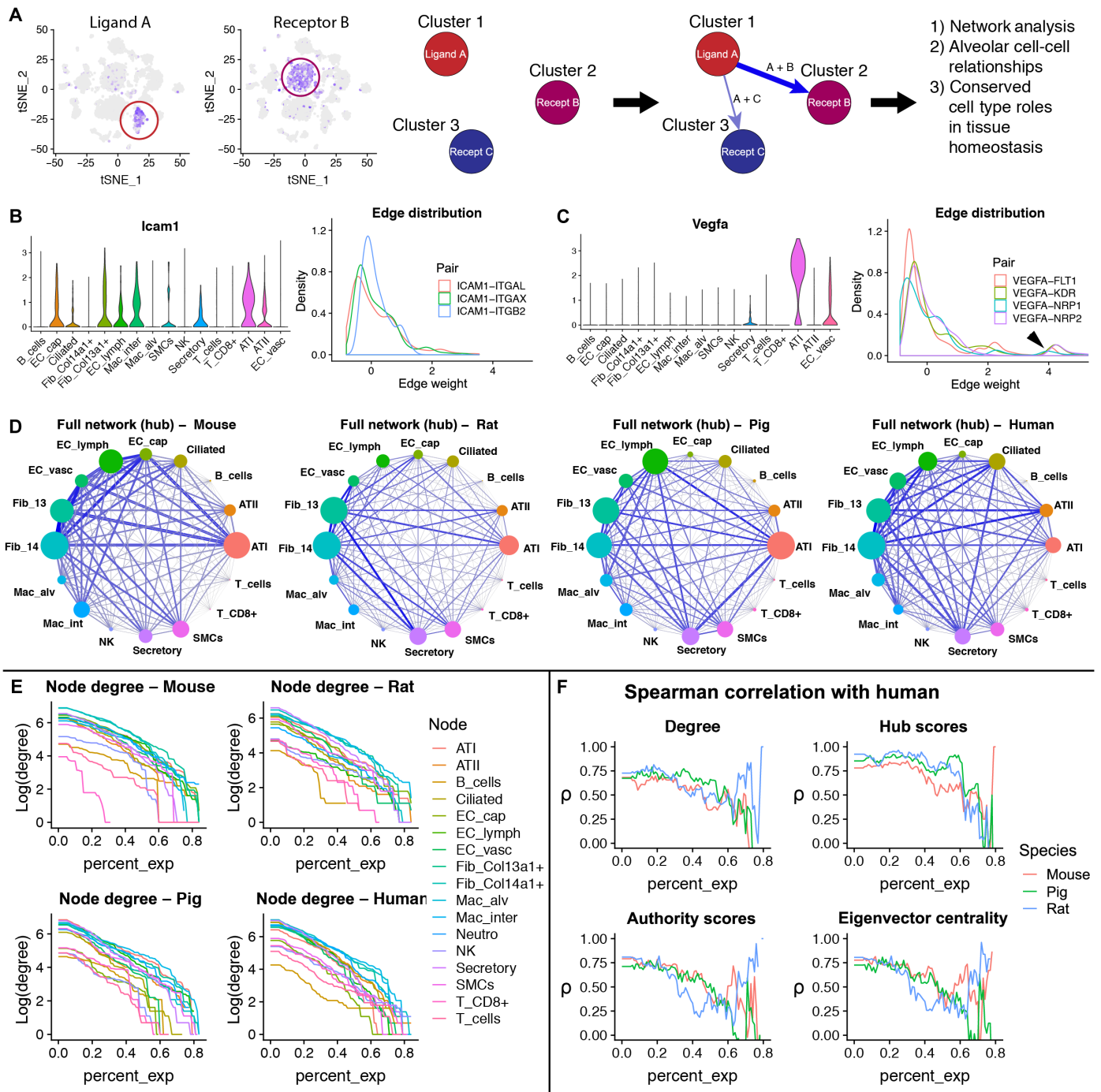
Fig. 3. Global network comparison reveals conserved signaling patterns underlying adult lung homeostasis. (**A**) Schematic of data processing. Gene expression was averaged by cluster and mapped against the FANTOM5 ligand-receptor database. When two nodes expressed a cognate ligand-receptor pair, a weighted edge was created (see Supplementary Text). (**B**) Promiscuous expression across cell types (ICAM1) causes a low distribution of edge weights. (**C**) Cell type–specific expression (VEGFA) causes a higher edge weight distribution (black arrows). (**D**) Global connectomes across the four species, showing the sum of edge weights between conserved nodes. Vertex (colored cell node) size is proportional to the Kleinberg hub score, while the thickness of the edges is proportional to the sum of the weights between two nodes. Similarities in overall signaling structure between the cell nodes are readily observable. (**E**) Comparison of degree rankings and the effect of thresholding. Stromal cells have the highest degrees, while lymphoid cells have the lowest. The x axis ("percent_exp") is percent of cluster expressing the marker, and the y axis is the degree of each node plotted on a logarithmic scale. (**F**) Quantitative cross-species correlations. Plots of Spearman correlation coefficients ("ρ"), with human as a reference, of the rankings of nodal centrality metrics over increasing thresholds of the fraction of cells in a node that expresses the ligand and receptor. Note that the correlation coefficients are relatively high (>0.75) and remain so up to ~40% expression, meaning that 40% of the cells in the node express either the ligand or the receptor. This demonstrates that node-node relationships are robust to thresholding. This suggests a high degree of evolutionary conservation in pulmonary connectomic structure, because all four species are very highly correlated.
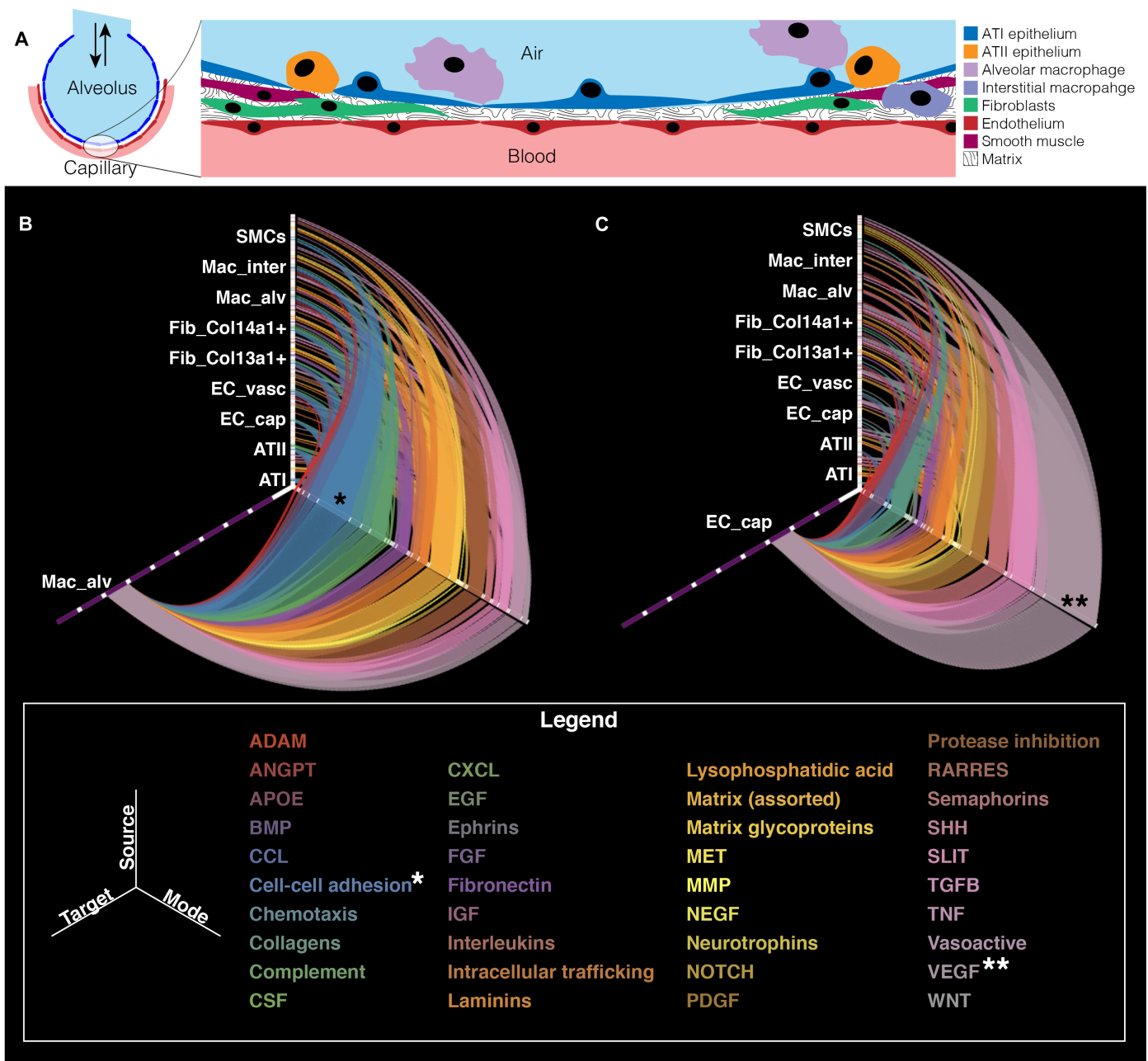
Fig. 4. Niche character visualizations from species-conserved signaling. (**A**) Cell classes in close physical proximity in alveolar lung. Downstream analysis was limited to nine cell types spatially registering to the alveolar septal wall. All mapped ligand-receptor pairs were classified into 1 of 38 signaling families (see legend). Hive plot illustrations of niche character are presented here for (**B**) alveolar macrophages and (**C**) capillary endothelium (see fig. S7 for all other niche visualizations). Alveolar macrophages display high connectivity and express distinctly high levels of cell-cell adhesion molecules (*). Capillary endothelium, in contrast, expresses numerous receptors for VEGF family signaling (**).

Network mapping and gene-level findings support stereotyped cellular roles

An online resource was created to allow interactive exploration of signaling topology by signaling family (lungconnectome.net). This tool allows ready identification of biologically relevant patterns, some of which we will discuss here. ATII cells display strong macrophage cross-talk based on CSF2, complement, GAS6, OSM, and APOE, aligning with known roles for ATII-macrophage cross-talk in surfactant handling and immune regulation (*29–31*). Sonic hedgehog

(SHH), known to be crucial for epithelial guidance of mesenchymal cells during branching morphogenesis (*32–34*), emerges in the cross-species dataset as exclusively linking epithelial populations and mesenchymal cells. Notch ligand DLL4, a major regulator of vascular development (*35*), is produced strongly by endothelial cells in the adult and can be sensed by a variety of cells, including SMCs and fibroblasts. Ephrins, spatial guidance proteins known for their role in tissue patterning (*36*), are expressed highly by endothelial populations and appear to be sensed preferentially by epithelia.

Last, the morphogen SLIT2 seems to specifically mediate stromal-epithelial and stromal-endothelial interactions.

Gene-level exploration of the species-conserved connectome revealed that Col13a1+ fibroblasts display elevated transcription of the genes HGF and FGF10. These ligands are known potent regulators of epithelial phenotype and interact with receptors encoded by MET and FGFR2 (respectively), both of which were expressed by ATII cells. Both interactions are known or hypothesized to be important regulators of the ATII niche (37–39). The Fib-HGF-MET-ATII signaling vector, including colocalization of ligand and receptor, was confirmed histologically at the protein level (fig. S8). Ligand (HGF) expression correlated with PDGFRA expression by fibroblasts, receptor (MET) expression correlated with SPC expression by ATII cells, and ligand and receptor expression colocalized in the alveolar region of lung.

ATI cells dominate the semaphorin and VEGF alveolar signaling across species

A major finding from this analysis was that ATI cells dominated the production of semaphorins, a class of spatial guidance molecule likely received by pulmonary endothelium. ATI cells transcribe very high levels of SEMA3E, sensed by vascular-specific PLXND1. Semaphorin action on this receptor triggers a known potent repulsive cue for endothelial cells and is essential for well-regulated microvascular development (40, 41). This leads to the somewhat unexpected implication that ATI cells may be responsible for establishing a spatially defined repulsive boundary for microvascular endothelial cells, providing them with a narrowly defined interalveolar spatial niche in which they are allowed to undergo morphogenesis and maintenance (Fig. 5, A and B).

ATI production of SEMA3E and VEGFA was histologically confirmed at the protein level (Fig. 5B). The species-conserved networks for each of these signaling families are shown in Fig. 5C. To quantify cellular roles within these networks, we performed a network centrality analysis (see Materials and Methods). Figure 5D shows the result of this analysis. In all four species, ATI cells claim the highest cumulative outgoing edge weight for SEMA and VEGF family signaling and display dominant Kleinberg hub scores. Conversely, in all four species, capillary endothelia display the highest, or near-highest, cumulative incoming edge weight and dominant Kleinberg authority scores. These findings show that the ATI-VEGF-EC and ATI-SEMA-EC signaling vectors are strongly conserved across species. This suggested an evolutionarily conserved pattern of epithelial-endothelial regulation.

Alveolar network analysis identifies signaling modalities dominated by specific cell types

The network centrality analysis technique was then extended to all signaling modalities and all cell types in the conserved alveolar connectome. The results of this comprehensive network centrality analysis are shown in Fig. 6 and fig. S9. To aid visualization, we grouped each signaling family according to which cell class was found to dominate ligand production within that network. Epithelia were found to dominate networks based on CSF, EGF, SHH, VEGF, and WNT family ligand-receptor interactions. This places epithelium in a powerful position to regulate multiple classes of cells, including macrophages, epithelia (autocrine), mesenchyme, and endothelia (Fig. 6B, top). Macrophages, in contrast, are the top-scoring nodes in networks based on ADAM, CCL, interleukin, and MMP family

signaling mechanisms, placing them in an excellent position to regulate mesenchyme and to influence immune behavior (Fig. 6B, bottom). Endothelia dominate networks based on BMP, Ephrin, NOTCH, and TNF family ligand-receptor pairs, showing high degrees of autocrine, EC-ATI, and EC-mesenchyme connectivity (fig. S9B, top). Mesenchyme predictably dominates many networks based on matrix-integrin interactions, as well as networks based on ANGPT, FGF, IGF, SLIT, and TGFB family signaling (fig. S9B, bottom). These findings collectively paint a network-level portrait of species-conserved interalveolar signaling delegation in pulmonary homeostasis.

DISCUSSION
Summary
It is currently thought that chronic lung diseases, such as pulmonary fibrosis and emphysema, result, at least in part, from errant or misguided cellular attempts to restore a homeostatic milieu following injury. Most pharmacological therapies—targeting an individual signaling axis—effectively focus on correcting a single edge of a much larger homeostatic network, attempting to swing tissue dynamics back to normal. However, if we are to make progress in the deciphering of chronic disease and the engineering of regenerative tissues, it is essential that we map and devise ways to comprehend systems-level interactions and multicellular interplay.

Taken holistically, the network-level map presented here suggests the existence of evolutionarily conserved roles for each cell type and cell class profiled in adult mammalian lung. ATI cells, displaying high expression of endothelial-guiding semaphorins and VEGF family cues, are in a key position to inform and regulate homeostatic endothelial morphogenesis. This conclusion is consistent with the extremely close physical proximity of these two cell types, which can be less than 100 nm apart in the pulmonary alveolus. It also aligns with the enormous importance of ATI cells in the maintenance of lung microvascular barrier and effective gas exchange. As maintaining barrier function is arguably the foremost design criterion in lung tissue homeostasis, this analysis suggests that healthy ATI cells may be essential for this process.

Macrophages are in an excellent position to sense the tissue microenvironment, expressing receptors for cell-cell adhesion cues and chemotactic ligands. Furthermore, as major potential recipients of complement family (largely produced by ATII cells) and CCL family (largely produced by interstitial macrophages) information, macrophages are predictably well suited to respond to immune threat. This is well in line with literature suggesting macrophages as key "monitors" of tissue microenvironment (42).

Mesenchymal cells predictably produce a great variety of matrix proteins. They also, however, appear to occupy key roles in the regulation of endothelial phenotype (through ANGPT family signaling), epithelial phenotype (via FGF and MET family signaling), and parenchymal spatial coordination (reflected by the SLIT family network) (43). Endothelial cells conversely dominate BMP family signaling, which gives them an avenue of cross-communication with epithelial cells—including ATI cells. Further, endothelia express vasoactive cues, which likely give them influence over associated mesenchymal populations. Taken holistically, our analysis unearths lineage-specific functional roles in pulmonary homeostasis and points to a fine balance of cell type interdependency involving all cell lineages.

These selected findings represent only some of what can be deduced from this dataset. The systems-level analysis we have
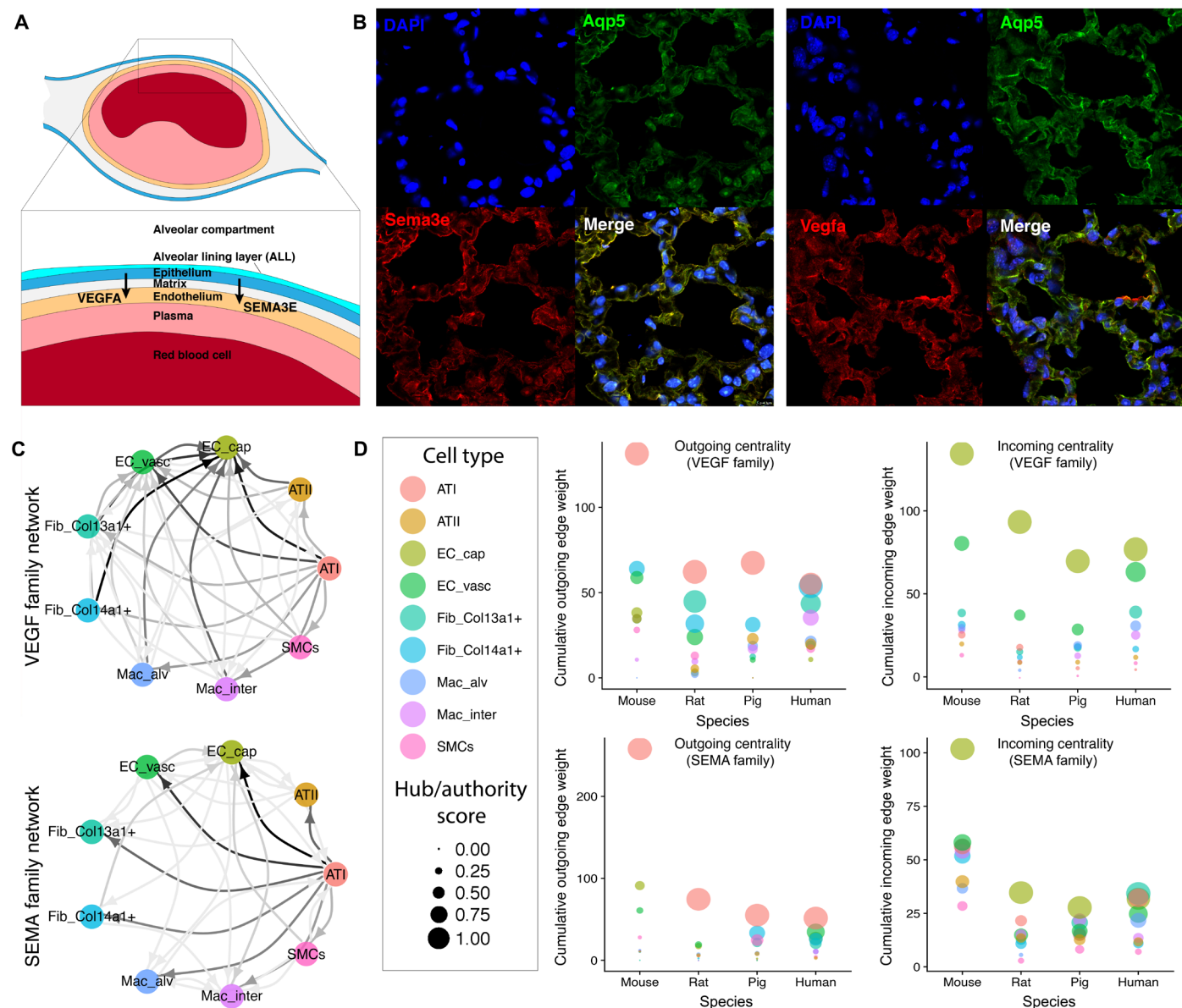
Fig. 5. ATI cells dominate VEGF and SEMA family niche networks in all species. (**A**) Capillary endothelia and ATI cells are in extremely close proximity in the pulmonary alveolus, often separated by less than 100 nm of shared basement membrane. (**B**) ATI cells, marked by Aqp5 in green, express high levels of Sema3e and Vegfa at the protein level, correlating with scRNAseq findings. (**C**) Ligand-receptor mapping shows that endothelial populations express high levels of receptors for both Vegfa and Sema3e, allowing receipt of spatial and growth cue information from ATI cells. (**D**) Network centrality analysis shows that ATI cells have dominant hub scores and top cumulative outgoing edge weight, while capillary endothelia have dominant authority scores and top cumulative incoming edge weight, in both the VEGF and SEMA family signaling networks.

performed seeks to answer fundamental questions about the essential structure of homeostatic signaling. However, there are myriad other analyses that can be performed on this cross-species dataset, including translationally relevant investigations into cell type analogs, homologous activation of transcriptional programs, and species-specific findings not explored here.

Drawbacks and limitations

There are some non-negligible limitations to the techniques used here. First and foremost, current single-cell techniques lose spatial registration, and careful inferences must be made regarding the

histologic location of profiled cells. We have attempted to do this with care. Second, the connectomic method used, which relies on cluster-wise averaging of single-cell z scores, is not necessarily optimal for all cell signaling questions. However, this definition of edge weight is well suited to network-level quantification of cellular roles, particularly considering that this study involved cross-species comparisons.

Here, we aimed to identify cellular connectomics that are conserved across species in the adult mammalian lung. While we have tried as much as possible to sample appropriately comparable ages across species, this task was difficult, given that rodents reach sexual
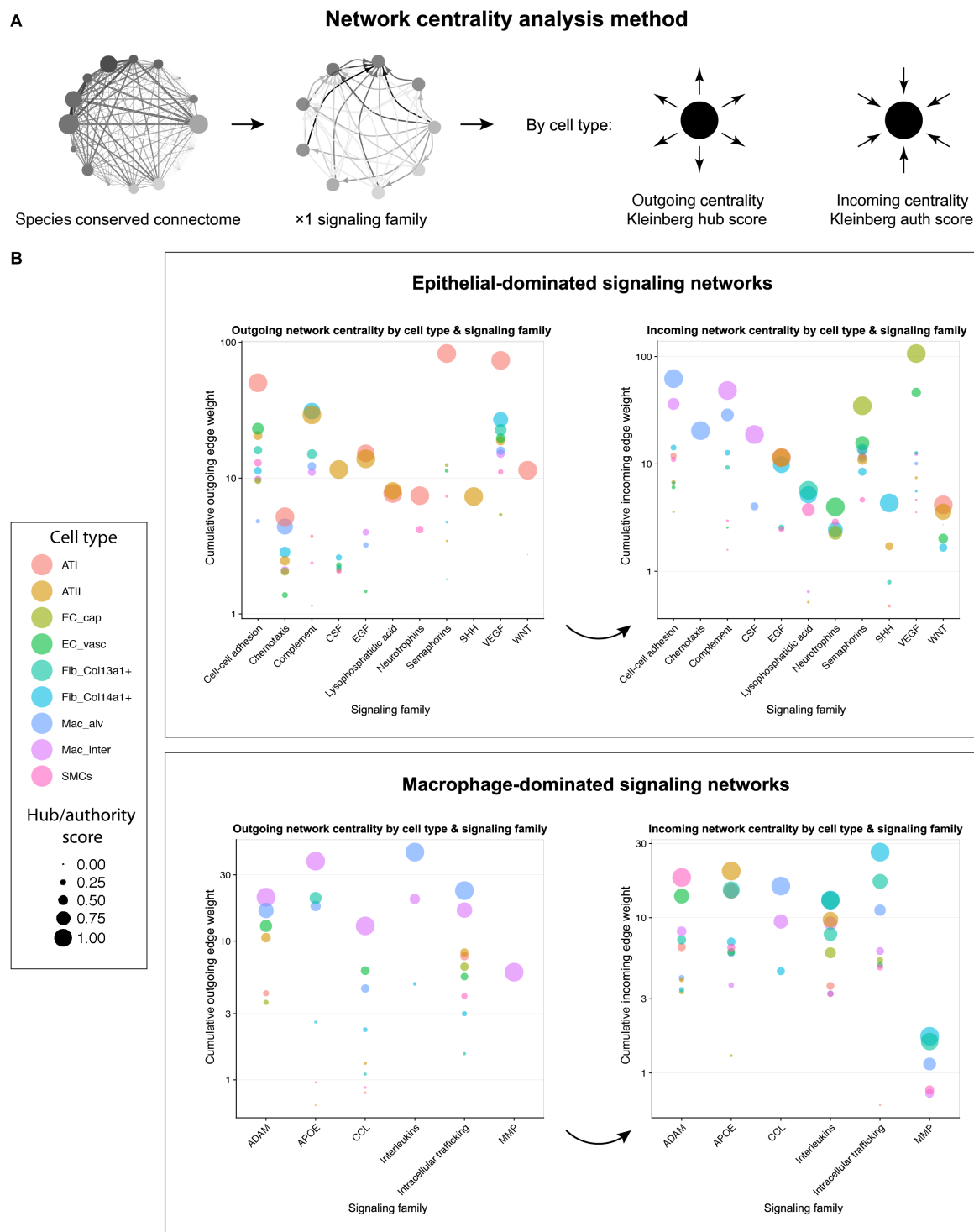
Fig. 6. Network centrality analysis categorizes each signaling network by dominant producers and receivers. (**A**) The species-conserved connectome is first subset to a single signaling family. Outgoing centrality, defined as the cumulative outgoing edge weight and the Kleinberg network hub score, and incoming centrality, defined as the cumulative incoming edge weight and the Kleinberg network authority score, are then calculated for each node. (**B**) Identification of signaling family networks dominated by epithelial cells (top) and macrophage populations (bottom). The left side of each panel shows outgoing centrality metrics (cumulative outgoing edge weight for each cell, Kleinberg hub scores), while the right side shows incoming centrality metrics (cumulative incoming edge weight for each cell, Kleinberg authority scores). These trends suggest species-conserved cellular roles in pulmonary tissue and allow identification of dominant producers and receivers of each class of signal in homeostatic lung tissue. Additional results detailing endothelia- and mesenchyme-dominated networks are shown in fig. S9. Note here the ATI dominance of cell-cell adhesion networks acting predominantly on alveolar macrophages versus ATII dominance of CSF networks acting predominantly on interstitial macrophages. Macrophages predictably dominate networks based on immunoregulatory cues such as CCL and interleukin family ligand-receptor pairs.

maturity very early in their development and do not generally stop growing before the end of their life span. This makes finding comparable ages across species difficult, because humans and pigs at sexual maturity generally cease growing. The human subjects profiled naturally cover a broad age range. Further, rodents were raised in barrier facilities, while pigs were farm-raised and human data naturally come directly from patients. This means that tissues obtained from different species will have different immune exposures and histories. Thus, we feel that the common cellular connectomic findings across species may represent conserved homeostatic interactions that may underlie the function of the alveolus, a finding that will require significant additional validations.

Conclusion

Understanding homeostatic tissue signaling is a daunting task. The fundamental power of the techniques described here is the ability to simultaneously measure interlocking signaling milieus in a single, whole tissue, a task that was impossible before the advent of single-cell technologies. By synthesizing information from multiple species, we have crafted a consensus vision of some of the signals regulating mammalian lung. We have performed a tailored meta-analysis that reveals an intriguing and explorable map of alveolar homeostasis. We present this map to the greater lung and systems biology community as a resource for further exploration. We hope that it may act as a starting demonstration for how to comprehensively understand intercellular coordination in complex tissues.

MATERIALS AND METHODS

Experimental design

The goal of this study was to profile single cells from the lungs of four species of mammal in as comparable a fashion as possible. We chose mice, rats, pigs, and human donors to cover the full range of generally used translational animal models. The intent was to devise a single method of single-cell dissociation and sequencing, which could be applied to tissues from each species. The following protocol and enzymatic concentrations were optimized to provide the highest cell yield (approximately 5 to 10% of cells in the lung) without sacrificing cell viability (approximately 80 to 85% viable). This was largely achieved, although tissues from the larger mammals required slightly longer enzymatic digestion times. Total in-house preparation time from tissue acquisition to library capture was between 1.5 and 2.0 hours for all samples. To increase the statistical power of the human dataset, an additional 12 batches were sequenced from cryopreserved dissociations from 11 separate individuals. These samples were dissociated via a different method by a collaborating laboratory (see the next two sections); the CCA described here allows excellent batch alignment between these two distinct approaches to dissociation and sample handling.

Cell dissociation (Yale)

For rodent samples, animals were selected between 7 and 9 weeks old, anesthetized with ketamine (75 mg/kg) and xylazine (5 mg/kg), and injected intraperitoneally with sodium heparin (400 U/kg). The abdomen was opened in sterile fashion, and the xiphoid process was lifted to reveal the underside of the diaphragm. The diaphragm was punctured and retracted to reveal the chest cavity. The rib cage was lifted, and an incision was made axially to reveal the trachea. The trachea was separated from the esophagus, opened using fine scissors,

cannulated with a barbed fitting, and secured with 4-0 silk. The right ventricle was opened with scissors, and the pulmonary artery (PA) was cannulated with an 18-gauge feeding needle and secured with 4-0 monofilament polypropylene. An incision was made in the left ventricle, and the lungs were perfused via the PA with 100 ml of phosphate-buffered saline (PBS) containing sodium heparin (100 U/ml) and sodium nitroprusside (0.01 mg/ml) with concurrent airway ventilation. The heart-lung block was then removed from the chest to be perfused with enzyme. Pig samples were harvested from heparinized farm-raised pigs undergoing nonpulmonary terminal procedures. Samples were cut from distal regions of the lung and placed in ice-cold saline. A single lobule weighing approximately 2.5 g with associated large airway was dissected out and cannulated via a conducting airway for enzyme perfusion. Human samples were acquired from healthy tissues excised from the distal margins of lobectomies and/or tissues rejected from transplant. The human tissue pieces received were approximately 1 cm^3 and did not contain airways or vessels of sufficient size to cannulate.

Dulbecco's modified Eagle's medium (DMEM) containing collagenase/dispase (1 mg/ml; Roche), elastase (3 U/ml; Worthington), and deoxyribonuclease (DNase) (20 U/ml; Worthington) preheated to 37°C was used to dissociate all tissues. For the rodent tissues, enzyme was instilled into the trachea 5× to 10× and also perfused through the vasculature (as the PA was still accessible). For pig tissue, enzyme was only instilled into the trachea 5× to 10×, and for human tissues, the enzyme mixture was delivered to the lung parenchyma repeatedly via a 25-gauge needle. In all cases, care was made to fully recruit and inflate the alveolar regions of the lung and to have the tissue fully saturated with enzyme. Following inflation and/or perfusion with enzyme, nonlung tissue was removed, including the heart and tracheas in the rodents. The remaining tissue and effluent were placed on a rocker at 37°C and 60 rpm until the tissue was soft enough to be dissociated, 20 min for the rodent tissues and 40 min for the large mammals. At this point, the lungs were pushed through a wire sieve using a weighing spatula until only dense connective tissue remained. The strainer was then rinsed with 20 ml of ice-cold DMEM containing 10% fetal bovine serum (FBS), 1% penicillin/streptomycin, 1% amphotericin, and 0.1% gentamicin. The tissue solution was spun down for 5 min at 300g, the supernatant was discarded, and the pellet was resuspended in red cell lysis buffer at a 1:1 ratio with pellet volume. After 60 s at room temperature, the buffer was diluted with 0.01% bovine serum albumin (BSA) in PBS (0.1 mg/ml) and spun down a second time. If necessary, this step was performed a second time. The pellet was resuspended again in 0.01% BSA in PBS and filtered through a 70-μm filter. The cells were then spun for 3 min at 300g, resuspended in 0.01% BSA in PBS, and passed twice through a 40-μm filter. The resulting cell suspension was counted, assessed for viability, and serially diluted to the desired concentration for single-cell library preparation.

Cell dissociation (Brigham and Women's Hospital)

Human lung explants were obtained from rejected donor lung organs that underwent lung transplantation at the Brigham and Women's Hospital or donor organs provided by the National Disease Research Interchange (NDRI). The study protocol was approved by the Partners Healthcare Institutional Board Review (protocol #2011P002419). Lungs were sliced and washed with cold sterile PBS two times. Visible airway structures, vessels, blood clots,

and mucin were removed. Tissue was minced mechanically into small pieces (<5 mm) and then incubated for 45 min in 37°C within the digestion medium, which consisted of elastase (30 U/ml; Elastin Products Company, Owensville, MO), DNase I (0.2 mg/ml; Sigma-Aldrich, St. Louis, MO), liberase (0.3 mg/ml; Roche, Basel, Switzerland), and 1% penicillin/streptomycin diluted in DMEM/F12 medium (Lonza, Basel, Switzerland). Digested tissue was filtered using a metal strainer (Sigma-Aldrich). Unfiltered tissue was incubated a second time in digestion medium for 30 min, followed by repeat filtration. Filtered tissue was then mixed with FBS to obtain a concentration of 10% FBS, and the mixture was incubated at 4°C. Suspension was then centrifuged at 300g at 4°C for 10 min. The pellet was resuspended in red cell lysis buffer (VWR, Radnor, PA) for 3 min in 37°C and centrifuged again. The pellet was resuspended in DMEM/F12 medium and filtered using a 100-µm strainer (Fisher Scientific, Waltham, MA). Freezing medium (10% FBS and 10% dimethyl sulfoxide in DMEM/F12) was then added to the filtrate. Cell suspensions were aliquoted and stored in liquid nitrogen for downstream applications.

10× library preparation, sequencing, and alignment
scRNAseq libraries were generated with the Chromium Single Cell 3′ v2 assay (10× Genomics). Samples were diluted for an expected cell recovery population of 5000 cells per lane. Libraries were sequenced on the HiSeq4000 platform (Illumina) to a depth of 160 to 220 million reads per library. Read 1 sequencing was 26 base pairs (bp) long, and read 2 sequencing was 98 bp long. FASTQ read 2 sequences were trimmed of the reverse complement sequence of the template switch oligo, as well as for poly(A) contaminants. Paired reads with a trimmed read 2 less than 30 bp long were discarded. Downstream processing was conducted with zUMIs v0.4 (44) with the default parameters; the top 10,000 cell barcodes ranked by reads were output for each sample. Ensembl GRCm38 release 92, Ensembl Rnor6.0 release 92, Ensembl Sscrofa11.1 release 92, and Ensembl GRCh38 release 91 were used to align the mouse, rat, pig, and human data, respectively.

Data filtration, normalization, and scaling
Cells were accepted with greater than 500 and fewer than 10,000 unique mapped genes. Analysis was performed on cells with less than 5% mitochondrial mapping for the mouse, rat, and pig samples and less than 8% for the human samples. Expression values were normalized to 10,000 transcripts per cell. Scaled values were calculated using Seurat v2.0's ScaleData function with default settings, with regression of nUMI and percent mitochondrial transcripts.

Summary statistics
Information regarding the library summary numbers was derived from the alignment summary file provided by the alignment program STAR v2.6. The average number of reads per cell in each sample was calculated by dividing the number of trimmed reads by the total number of cells profiled in each sample.

Clustering
Samples within each species were aligned and clustered independently. First, variable genes were determined for each sample using Seurat's FindVariableGenes function with default parameters. CCA, as implemented in Seurat v2.0's AlignSubspace function, was then used to align samples within each species using only genes that were found the top 1000 variable genes in at least two samples per species. Clusters

in the resulting subspace were then identified using FindClusters and mapped using RunTSNE. Alignment between species samples can be seen in fig. S1.

Integration of human samples across dissociation methods
For the human dataset, scRNAseq data were integrated from both dissociation methods described above. Briefly, the top 12 cryopreserved samples were selected on the basis of their sampling of alveolar cell types (ATI and ATII cells), as determined by the standard Seurat clustering pipeline. The identified samples were then downsampled to no greater than 1000 cells per identity to facilitate computational efficiency. A CCA was then performed between the fresh and cryopreserved human samples.

Marker determination
Differential expression for clusters was performed using the FindAllMarkers function in Seurat, using the Wilcoxon rank test. Markers were calculated for the cells in each cluster against the cells within all the other clusters within a single-species object. Markers for cell type identification were calculated before homolog conversion.

Homolog conversion
To facilitate cross-species comparisons, including all connectomic work, we converted nonhuman genes to human homologs when possible. First, a table was created using BioMart of all possible species-species homologs (45). Then, if a 1:1 match existed between a nonhuman gene and a human gene, the nonhuman gene name was converted. Second, if the given gene had multiple potential human homologs but one of those homologs was simply the uppercase of the nonhuman gene in question, the gene name was converted to that specific human homolog. All other remaining genes maintained their original names. This approach allowed appropriate cross-species comparisons without eliminating relevant cross-species differences.

Jaccard index plots
The plots in fig. S5 were made by modifications of functions from the "matchSCore" package in R (46). That is, Jaccard indices were calculated comparing all markers for each node with an average log fold change of >1, in every species, to those of each node in every other. Homolog conversion (see above) was performed before testing. The heat map representation of these data was made using the "complexHeatmap" package.

Human Protein Atlas mapping
The immunoperoxidase images in Fig. 2 were mined from the Human Protein Atlas per their Creative Commons license (24). Antibodies were limited to those that had been validated with at least an "approved" immunohistochemistry score and had simultaneous staining for literature-supported positive controls. Specific antibodies used are tabulated in data file S2.

Immunostaining and imaging
In brief, human and rodent tissues were inflation-fixed with 10% formalin for 3 hours, transferred in 70% ethanol, and embedded in paraffin. Deparaffinized tissue sections were soaked in antigen retrieval solution [0.05% Tween 20, 1 mM EDTA, and 10 mM tris (pH 9)] at 75°C for 20 min, followed by blocking (10% FBS and

0.2% Triton X-100) at room temperature for 20 min. Lung sections were then stained with primary antibodies (concentrations are listed in data file S2) overnight at 4°C, followed by incubation with secondary antibodies (concentrations are listed in data file S2) at room temperature for 1 hour and 4′,6-diamidino-2-phenylindole (DAPI) nuclear stain (1 min). Samples were imaged with confocal microscopy (Leica SP5). Detailed staining protocol can be found in the study of Yuan et al. (47).

Connectome definition

Average scaled gene expression values were calculated for each cluster on a gene-wise basis across the whole sample using Seurat's AverageExpression function. These cluster-average expression values were then mapped against the FANTOM5 database of known and literature-supported ligand-receptor pairs to create directed edges with various attributes. Ligands or receptors with no cognate molecules expressed in the dataset were excluded. For each potential ligand-receptor interaction present within the dataset, a directed edge was created connecting any two clusters expressing the cognate molecules at >5% of each cluster. The weight attribute of this edge was defined as the sum of the average scaled expression values of cognate molecules (see below for reasoning).

Weight definition

Extensive discussion has occurred among the authors regarding weight definition for this analysis. Although the most obvious approach is to add or multiply the normalized expression values of cognate molecules in the respective clusters of each edge, such an approach has a number of problems that were considered for this study. First, to do so would obscure the relevance of signaling molecules transcribed at low levels but which are highly specific to certain clusters. Second, variable transcript quality and capture rate make comparing normalized expression levels inappropriate across batches. We therefore defined the weight of an edge as the sum of the z scores (scaled expression) of cognate molecules in the respective clusters of each edge. This approach rewards outliers within a given batch, at the cost of penalizing molecules that are pan-expressed across the entire cell population. This bears on interpretation of the results in this study, because a strong score reflects the observation that a node tended to be an outlying producer of molecules within the set being tested, some of which may be expressed at very low levels (such as WNTs).

Node definition

The definition of nodes for graph theory analyses in biological systems is a nontrivial problem (48). In certain fields, the choice of node is fairly straightforward (choosing cities as nodes in studies of airline traffic, choosing individual people as nodes in studies of social networks, etc.). In other cases, node designation is not clear. In neuroscience research, for example, the question of what constitutes a node is open for debate [see (49, 50) for a discussion]. Furthermore, the ability to define nodes rests on the assumptions that it is possible to do so in the system under study, it is biologically plausible, and it is useful scientifically (48). The definition of edges is also nontrivial—similar issues affecting node designation also affect edge designation [e.g., (48–50)]. These issues are not merely theoretical, because there are numerous examples showing that network-level properties can change depending on how nodes and edges are defined (51–53).

Hence, it rests with the researcher to define these elements of the graph in a way that is driven by the data at hand yet is constrained by known biological principles. These designations then allow the analysis of network-level properties. In time, better sampling techniques will afford higher-resolution datasets, and definitions of nodes and edges will likely change as a consequence. Nevertheless, network-level analyses of these "coarse data" still provide a tractable way to lay a foundation for future connectomic work in the lung.

Global network analysis

The global networks created by ligand-receptor mapping are highly multimodal, with $\sim 0.25 \times 10^6$ edges divided between $\sim 1 \times 10^3$ specific ligand-receptor pairwise mechanisms (modalities). For many mechanisms, edges were promiscuous, connecting most nodes in the graph together combinatorially due to similar levels of gene expression in many cell types. For the global network comparisons in Fig. 3, we were predominantly interested in outlier connections where ligand and receptor expression were differentially expressed above the mean of the global community, and therefore, we limited our dataset to only those edges where the cluster average z score was positive for both ligand and receptor. Weights of edges were defined as the sum of the z scores for ligand and receptor (see Supplementary Text). Global node-node affinity, when used, was defined as the sum of the weights of all edges connecting pairs of nodes. Nodal degree, hub, authority, and eigenvector centrality were calculated using the "igraph" package in R (54). For Fig. 3C, hub scores were calculated using this global node-node global affinity. For Fig. 3 (D and E), all edges above the percentage threshold were incorporated into calculations of degree and centrality. Thresholding graphs were created by iteratively subsetting the edge list for each species to include only those edges above a given percentage expression and then recalculating degree distributions, nodal centrality rankings, and Spearman correlations for each threshold. Spearman correlations were calculated by aligning node order across species and comparing centrality metrics using the human as a reference. Network graphs were created using the igraph package in R.

Definition of species-conserved connectome

Species-conserved edges, for the purposes of this analysis, were defined as those found above the aforementioned 5% ligand and receptor expression threshold in human and in at least one other species. Statistically significant species-conversed edges were then defined as those with a Fisher's cross-species combined P value of <0.05. Single-species P values were generated as described in the "Statistical analysis" section.

Distal niche definition and hive plot generation

For distal niche exploration, the master dataset was filtered to include only edges between cell types likely to be in close proximity based on histologic findings, namely, ATI and ATII cells, capillary and vascular cells, fibroblast populations, SMCs, and macrophages. The "niche" for a given cell type was defined as all statistically significant species-conserved edges landing on that cell type from potentially proximal cell types. Hive plots were created using HiveGraph 1.0 (http://wodaklab.org/hivegraph/) created by the Wodak laboratory (55).

Network centrality analysis

Network centrality analysis was applied either to individual species datasets (Fig. 5) or to the combined species-conserved connectome

(Fig. 6 and fig. S9). In each case, the method was as follows. First, the connectome was subset to yield only edges within the signaling family of interest. This connectome was then converted to a network object in igraph. Cumulative outgoing and incoming edge weights were then tallied for each node within the network. Kleinberg hub and authority scores were also calculated for each node. ggplot2 (56) was used to plot the resulting multidimensional data structures, which contain cross-referenced information for incoming and outgoing network centrality, by cell type, by signaling family, and for each species (if applicable).

Online resource

The interactive webtool presenting the full dataset can be found at lungconnectome.net. The website was developed using R version 3.5.2 using the following packages: Shiny, dplyr, ggplot2, igraph, and visNetwork.

Software

RNAseq alignment was performed with STAR v2.6 (57). Downstream computation was performed in R. Single-cell analysis, population characterization, and plotting were performed with functions within the Seurat package from the Satija laboratory (58).

Statistical analysis

Statistically significant edges were defined as follows. Each interaction (characterized by the secreting cell type, ligand, receptor, and the receiving cell type) was assigned a score calculated as the sum of the scaled ligand and receptor expression (the edge weight). A P value for each interaction was calculated using a one-sided Wilcoxon rank sum test separately for each species. There were no assumptions made for the statistical distribution of the ligand-receptor product score, and the P value only considered the score in relation to the scores of other interactions in the same species. Last, a P value for the significance of each interaction type across all species was calculated by combining the respective species-specific P values using Fisher's method. Doing so allows both the fusion of statistical significance evidence from multiple species to boost the overall statistical significance for that interaction and the appropriate scale of the P value if the particular ligand or receptor is missing from one of the species.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/12/eaaw3851/DC1

Table S1. Summary data for all individuals profiled.
Fig. S1. Batch alignment across individuals.
Fig. S2. Dataset quality control and cell type distributions.
Fig. S3. Cross-species marker atlas organized by cell class.
Fig. S4. Cross-species comparative mesenchymal subclustering.
Fig. S5. matchSCore plots and quantification of cross-species homolog conversion effects.
Fig. S6. Selected secreted signals found to be associated with ATI cells across species.
Fig. S7. Quantitative hive plots illustrating niche character for all nine alveolar cell types.
Fig. S8. Protein-level validation of the Fib-HGF-MET-ATII signaling axis.
Fig. S9. Network centrality analysis for endothelial- and mesenchymal-dominated networks.
Data file S1. Excel spreadsheet detailing the signaling family categorizations for each ligand-receptor pair mapped within this dataset.
Data file S2. Excel spreadsheet detailing reagents and antibodies used.

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. X. Zhou, R. A. Franklin, M. Adler, J. B. Jacox, W. Bailis, J. A. Shyer, R. A. Flavell, A. Mayo, U. Alon, R. Medzhitov, Circuit design features of a stable two-cell system. *Cell* **172**, 744–757.e17 (2018).

2. C. Bonnans, J. Chou, Z. Werb, Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* **15**, 786–801 (2014).

3. J. D. Humphrey, E. R. Dufresne, M. A. Schwartz, Mechanotransduction and extracellular matrix homeostasis. *Nat. Rev. Mol. Cell Biol.* **15**, 802–812 (2014).

4. A. J. Miller, J. R. Spence, In vitro models to study human lung development, disease and homeostasis. *Physiology* **32**, 246–260 (2017).

5. I. Amit, D. R. Winter, S. Jung, The role of the local environment and epigenetics in shaping macrophage identity and their effect on tissue homeostasis. *Nat. Immunol.* **17**, 18–25 (2016).

6. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).

7. A. Zeisel, A. B. Muñoz-Manchado, S. Codeluppi, P. Lönnerberg, G. La Manno, A. Juréus, S. Marques, H. Munguba, L. He, C. Betsholtz, C. Rolny, G. Castelo-Branco, J. Hjerling-Leffler, S. Linnarsson, Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).

8. B. Tasic, V. Menon, T. N. Nguyen, T. K. Kim, T. Jarsky, Z. Yao, B. Levi, L. T. Gray, S. A. Sorensen, T. Dolbeare, D. Bertagnolli, J. Goldy, N. Shapovalova, S. Parry, C. Lee, K. Smith, A. Bernard, L. Madisen, S. M. Sunkin, M. Hawrylycz, C. Koch, H. Zeng, Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).

9. M. Baron, A. Veres, S. L. Wolock, A. L. Faust, R. Gaujoux, A. Vetere, J. H. Ryu, B. K. Wagner, S. S. Shen-Orr, A. M. Klein, D. A. Melton, I. Yanai, A single-cell transcriptomic map of the human and mouse pancreas reveals Inter- and intra-cell population structure. *Cell Syst.* **3**, 346–360.e4 (2016).

10. J. N. Campbell, E. Z. Macosko, H. Fenselau, T. H. Pers, A. Lyubetskaya, D. Tenen, M. Goldman, A. M. J. Verstegen, J. M. Resch, S. A. McCarroll, E. D. Rosen, B. B. Lowell, L. T. Tsai, A molecular census of arcuate hypothalamus and median eminence cell types. *Nat. Neurosci.* **20**, 484–496 (2017).

11. A. L. Haber, M. Biton, N. Rogel, R. H. Herbst, K. Shekhar, C. Smillie, G. Burgin, T. M. Delorey, M. R. Howitt, Y. Katz, I. Tirosh, S. Beyaz, D. Dionne, M. Zhang, R. Raychowdhury, W. S. Garrett, O. Rozenblatt-Rosen, H. N. Shi, O. Yilmaz, R. J. Xavier, A. Regev, A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).

12. A.-C. Villani, R. Satija, G. Reynolds, S. Sarkizova, K. Shekhar, J. Fletcher, M. Griesbeck, A. Butler, S. Zheng, S. Lazo, L. Jardine, D. Dixon, E. Stephenson, E. Nilsson, I. Grundberg, D. McDonald, A. Filby, W. Li, P. L. de Jager, O. Rozenblatt-Rosen, A. A. Lane, M. Haniffa, A. Regev, N. Hacohen, Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017).

13. X. Han, R. Wang, Y. Zhou, L. Fei, H. Sun, S. Lai, A. Saadatpour, Z. Zhou, H. Chen, F. Ye, D. Huang, Y. Xu, W. Huang, M. Jiang, X. Jiang, J. Mao, Y. Chen, C. Lu, J. Xie, Q. Fang, Y. Wang, R. Yue, T. Li, H. Huang, S. H. Orkin, G.-C. Yuan, M. Chen, G. Guo, Mapping the mouse cell Atlas by Microwell-Seq. *Cell* **172**, 1091–1107.e17 (2018).

14. D. A. Skelly, G. T. Squiers, M. A. McLellan, M. T. Bolisetty, P. Robson, N. A. Rosenthal, A. R. Pinto, Single-cell transcriptional profiling reveals cellular diversity and intercommunication in the mouse heart. *Cell Rep.* **22**, 600–610 (2018).

15. B. Treutlein, D. G. Brownfield, A. R. Wu, N. F. Neff, G. L. Mantalas, F. H. Espinoza, T. J. Desai, M. A. Krasnow, S. R. Quake, Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).

16. J.-H. Lee, T. Tammela, M. Hofree, J. Choi, N. D. Marjanovic, S. Han, D. Canner, K. Wu, M. Paschini, D. H. Bhang, T. Jacks, A. Regev, C. F. Kim, Anatomically and functionally distinct lung mesenchymal populations marked by Lgr5 and Lgr6. *Cell* **170**, 1149–1163.e12 (2017).

17. J. A. Zepp, W. J. Zacharias, D. B. Frank, C. A. Cavanaugh, S. Zhou, M. P. Morley, E. E. Morrisey, Distinct mesenchymal lineages and niches promote epithelial self-renewal and myofibrogenesis in the lung. *Cell* **170**, 1134–1148.e10 (2017).

18. T. Xie, Y. Wang, N. Deng, G. Huang, F. Taghavifar, Y. Geng, N. Liu, V. Kulur, C. Yao, P. Chen, Z. Liu, B. Stripp, J. Tang, J. Liang, P. W. Noble, D. Jiang, Single-cell deconvolution of fibroblast heterogeneity in mouse pulmonary fibrosis. *Cell Rep.* **22**, 3625–3640 (2018).

19. A. Regev, S. A. Teichmann, E. S. Lander, I. Amit, C. Benoist, E. Birney, B. Bodenmiller, P. Campbell, P. Carninci, M. Clatworthy, H. Clevers, B. Deplancke, I. Dunham, J. Eberwine, R. Eils, W. Enard, A. Farmer, L. Fugger, B. Göttgens, N. Hacohen, M. Haniffa, M. Hemberg, S. Kim, P. Klenerman, A. Kriegstein, E. Lein, S. Linnarsson, J. Lundeberg, P. Majumder, J. C. Marioni, M. Merad, M. Mhlanga, M. Nawijn, M. Netea, G. Nolan, D. Pe'er, A. Phillipakis, C. P. Ponting, S. Quake, W. Reik, O. Rozenblatt-Rosen, J. Sanes, R. Satija, T. N. Schumacher, A. Shalek, E. Shapiro, P. Sharma, J. W. Shin, O. Stegle, M. Stratton, M. J. T. Stubbington, A. van Oudenaarden, A. Wagner, F. Watt, J. Weissman, B. Wold, R. Xavier, N. Yosef; Human Cell Atlas Meeting Participants, The human cell Atlas. bioRxiv 121202 [**Preprint**]. 8 May 2017. https://doi.org/10.1101/121202.

20. H. B. Schiller, D. T. Montoro, L. M. Simon, E. L. Rawlins, K. B. Meyer, M. Strunz, F. A. Vieira Braga, W. Timens, G. H. Koppelman, G. R. S. Budinger, J. K. Burgess, A. Waghray, M. van den Berge, F. J. Theis, A. Regev, N. Kaminski, J. Rajagopal, S. A. Teichmann, A. V. Misharin, M. C. Nawijn, The Human Lung Cell Atlas: A high-resolution reference map of the human lung in health and disease. *Am. J. Respir. Cell Mol. Biol.* **61**, 31–41 (2019).

21. P. Gehr, D. K. Mwangi, A. Ammann, G. M. O. Maloiy, C. Richard Taylor, E. R. Weibel, Design of the mammalian respiratory system. V. Scaling morphometric pulmonary diffusing capacity to body mass: Wild and domestic mammals. *Respir. Physiol.* **44**, 61–86 (1981).

22. J. X. Zhou, R. Taramelli, E. Pedrini, T. Knijnenburg, S. Huang, Extracting intercellular signaling network of cancer tissues using ligand-receptor expression patterns from whole-tumor and single-cell transcriptomes. *Sci. Rep.* **7**, 8815 (2017).

23. I. Angelidis, L. M. Simon, I. E. Fernandez, M. Strunz, C. H. Mayr, F. R. Greiffo, G. Tsitsiridis, E. Graf, T.-M. Strom, O. Eickelberg, M. Mann, F. J. Theis, H. B. Schiller, An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. bioRxiv 351353 [**Preprint**]. 20 June 2018. https://doi.org/10.1101/351353.

24. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

25. J. A. Ramilowski, T. Goldberg, J. Harshbarger, E. Kloppmann, M. Lizio, V. P. Satagopam, M. Itoh, H. Kawaji, P. Carninci, B. Rost, A. R. R. Forrest, A draft network of ligand–receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866 (2015).

26. J.-P. Himanen, N. Saha, D. B. Nikolov, Cell–cell signaling via Eph receptors and ephrins. *Curr. Opin. Cell Biol.* **19**, 534–542 (2007).

27. R. Nusse, C. Fuerer, W. Ching, K. Harnish, C. Logan, A. Zeng, D. Ten Berge, Y. Kalani, in *Cold Spring Harbor symposia on quantitative biology* (Cold Spring Harbor Laboratory Press, 2008), vol. 73, pp. 59–66.

28. H. Clevers, K. M. Loh, R. Nusse, An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. *Science* **346**, 1248012 (2014).

29. K. Westphalen, G. A. Gusarova, M. N. Islam, M. Subramanian, T. S. Cohen, A. S. Prince, J. Bhattacharya, Sessile alveolar macrophages communicate with alveolar epithelium to modulate immunity. *Nature* **506**, 503–506 (2014).

30. J. A. Whitsett, T. Alenghat, Respiratory epithelial cells orchestrate pulmonary innate immunity. *Nat. Immunol.* **16**, 27–35 (2015).

31. J. A. Whitsett, S. E. Wert, T. E. Weaver, Alveolar surfactant homeostasis and the pathogenesis of pulmonary disease. *Annu. Rev. Med.* **61**, 105–119 (2010).

32. S. Bellusci, Y. Furuta, M. G. Rush, R. Henderson, G. Winnier, B. L. Hogan, Involvement of Sonic hedgehog (Shh) in mouse embryonic lung growth and morphogenesis. *Development* **124**, 53–63 (1997).

33. C. V. Pepicelli, P. M. Lewis, A. P. McMahon, Sonic hedgehog regulates branching morphogenesis in the mammalian lung. *Curr. Biol.* **8**, 1083–1086 (1998).

34. A. C. White, J. Xu, Y. Yin, C. Smith, G. Schmid, D. M. Ornitz, FGF9 and SHH signaling coordinate lung growth and development through regulation of distinct mesenchymal domains. *Development* **133**, 1507–1517 (2006).

35. T. Iso, Y. Hamamori, L. Kedes, Notch signaling in vascular development. *Arterioscler. Thromb. Vasc. Biol.* **23**, 543–553 (2003).

36. I. V. Beamish, L. Hinck, T. E. Kennedy, Making connections: Guidance cues and receptors at nonneural cell–cell junctions. *Cold Spring Harb. Perspect. Biol.* **10**, a029165 (2017).

37. R. J. Mason, C. C. Leslie, K. McCormick-Shannon, R. R. Deterding, T. Nakamura, J. S. Rubin, J. M. Shannon, Hepatocyte growth factor is a growth factor for rat alveolar type II cells. *Am. J. Respir. Cell Mol. Biol.* **11**, 561–567 (1994).

38. Y. Ito, K. Correll, J. A. Schiel, J. H. Finigan, R. Prekeris, R. J. Mason, Lung fibroblasts accelerate wound closure in human alveolar epithelial cells through hepatocyte growth factor/c-Met signaling. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **307**, L94–L105 (2014).

39. T. Volckaert, S. De Langhe, Lung epithelial stem cells and their niches: Fgf10 takes center stage. *Fibrogenesis Tissue Repair* **7**, 8 (2014).

40. H. Aghajanian, C. Choi, V. C. Ho, M. Gupta, M. K. Singh, J. A. Epstein, Semaphorin 3d and semaphorin 3e direct endothelial motility through distinct molecular signaling pathways. *J. Biol. Chem.* **289**, 17971–17979 (2014).

41. M. J. Hamm, B. C. Kirchmaier, W. Herzog, Sema3d controls collective endothelial cell migration by distinct mechanisms via Nrp1 and PlxnD1. *J. Cell Biol.* **215**, 415–430 (2016).

42. Y. Okabe, R. Medzhitov, Tissue biology perspective on macrophages. *Nat. Immunol.* **17**, 9–17 (2016).

43. H. Blockus, A. Chédotal, Slit-robo signaling. *Development* **143**, 3037–3044 (2016).

44. S. Parekh, C. Ziegenhain, B. Vieth, W. Enard, I. Hellmann, zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. *Gigascience* **7**, giy059 (2018).

45. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184 (2009).

46. E. Mereu, G. Iacono, A. Guillaumet-Adkins, C. Moutinho, G. Lunazzi, C. P. Santos, I. Miguel-Escalada, J. Ferrer, F. X. Real, I. Gut, H. Heyn, matchSCore: Matching single-cell phenotypes across tools and experiments. bioRxiv 314831 [**Preprint**]. 7 May 2018. https://doi.org/10.1101/314831.

47. Y. Yuan, A. J. Engler, M. S. Raredon, A. le, P. Baevova, M. C. Yoder, L. E. Niklason, Epac agonist improves barrier function in iPSC-derived endothelial colony forming cells for whole organ tissue engineering. *Biomaterials* **200**, 25–34 (2019).

48. C. T. Butts, Revisiting the foundations of network analysis. *Science* **325**, 414–416 (2009).

49. M. Rubinov, O. Sporns, Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).

50. A. Fornito, A. Zalesky, M. Breakspear, Graph analysis of the human connectome: Promise, progress, and pitfalls. *Neuroimage* **80**, 426–444 (2013).

51. A. Zalesky, A. Fornito, I. H. Harding, L. Cocchi, M. Yücel, C. Pantelis, E. T. Bullmore, Whole-brain anatomical networks: Does the choice of nodes matter? *Neuroimage* **50**, 970–983 (2010).

52. A. Fornito, A. Zalesky, E. Bullmore, Network scaling effects in graph analytic studies of human resting-state fMRI data. *Front. Syst. Neurosci.* **4**, 22 (2010).

53. M. Bastiani, N. J. Shah, R. Goebel, A. Roebroeck, Human cortical connectome reconstruction from diffusion weighted MRI: The effect of tractography algorithm. *Neuroimage* **62**, 1732–1749 (2012).

54. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJ. Complex Syst.* **1695**, 1–9 (2006).

55. M. Krzywinski, I. Birol, S. J. Jones, M. A. Marra, Hive plots—Rational approach to visualizing networks. *Brief. Bioinform.* **13**, 627–644 (2012).

56. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).

57. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

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