RESOURCE ARTICLE

Quantitative single-cell interactomes in normal and virus-infected mouse lungs

Margo P. Cain1,2, Belinda J. Hernandez1 and Jichao Chen1,*

ABSTRACT

Mammalian organs consist of diverse, intermixed cell types that signal to each other via ligand-receptor interactions – an interactome – to ensure development, homeostasis and injury-repair. Dissecting such intercellular interactions is facilitated by rapidly growing single-cell RNA sequencing (scRNA-seq) data; however, existing computational methods are often not readily adaptable by bench scientists without advanced programming skills. Here, we describe a quantitative intuitive algorithm, coupled with an optimized experimental protocol, to construct and compare interactomes in control and Sendai virus-infected mouse lungs. A minimum of 90 cells per cell type compensates for the known gene dropout issue in scRNA-seq and achieves comparable sensitivity to bulk RNA sequencing. Cell lineage normalization after cell sorting allows cost-efficient representation of cell types of interest. A numeric representation of ligand-receptor interactions identifies, as outliers, known and potentially new interactions as well as changes upon viral infection. Our experimental and computational approaches can be generalized to other organs and human samples.

KEY WORDS: scRNA-seq, Ligand-receptor interaction, Bioinformatics, Lung viral injury

INTRODUCTION

In multicellular mammalian organs, specialized cell types, such as beta cells in the pancreas and cardiomyocytes in the heart, perform organ-specific functions in coordination with generic cell types, such as the omnipresent endothelial and immune cells. Such coordination occurs at the microscopic level, such that many organs can be viewed as basic units repeated millions of times, as exemplified by alveoli in the lung, glomeruli in the kidney and hair follicles in the skin. Besides cooperating for physiological functions, these multi-cell type units are also integrated niches for the known gene dropout issue in scRNA-seq. In addition, current algorithms and outputs are not readily adopted by bench scientists without advanced computational skills.

RESULTS

scRNA-seq has cell type resolution with sensitivity comparable to that of bulk RNA-seq

One challenge to construct an interactome using scRNA-seq is the so-called gene dropout issue, where only a few thousand genes are detected in a given cell due to technical inefficiency, compared to the 20,000-30,000 genes expected and obtained by bulk RNA-seq of typical mammalian cells (Hicks et al., 2018; Kharchenko et al., 2014). However, it is important to recognize that bulk RNA-seq does not detect all genes in every cell of the sample, and the apparent high coverage is the result of summing over tens of
thousands of cells. We thus hypothesized that, by combining cells of a cell type readily identifiable in scRNA-seq, we could achieve comparable sensitivity to bulk RNA-seq of the same purified cell type. To test this, we used as a standard our published bulk RNA-seq data of fluorescence-activated cell sorting (FACS)-purified alveolar type 1 (AT1) and alveolar type 2 (AT2) cells (Little et al., 2019), and evaluated scRNA-seq gene dropouts as a function of expression level (Fig. 1A; Fig. S1A). We found that the dropout rate, defined as

![Diagram A](image)

![Diagram B](image)

![Diagram C](image)

![Diagram D](image)

![Diagram E](image)

![Diagram F](image)

![Diagram G](image)

Fig. 1. See next page for legend.
the percentage of genes detected by scRNA-seq but not scRNA-seq, decreased with increasing expression levels, as measured by fragments per kilobase of transcript per million mapped reads (FPKM) from bulk RNA-seq. For example, only 4% of genes at 10 FPKM were missed in scRNA-seq of 345 AT1 cells, and this percentage was even lower (0.6%) for AT2 cells where 1252 cells were sequenced. Technical or biological noise likely dominated for low FPKM genes, as 61 and 54 genes were detected by scRNA-seq but not bulk RNA-seq for AT1 and AT2 cells, respectively. The dropout rate is the percentage of genes not detected by scRNA-seq for a given range of bulk RNA-seq expression values. scRNA-seq data are from postnatal day (P)7 (GSE144769); bulk RNA-seq data are from P5 for AT1 cells and P8 for AT2 cells, as published (Little et al., 2019). (B) Consistency between AT1 cell (left) and AT2 cell (right) gene expression from scRNA-seq and bulk RNA-seq gene expression of the corresponding cell type. The scRNA-seq equivalent of the FPKM value in bulk RNA-seq, which sums the expression in all cells of the sample, is considered the sum of normalized unique molecular identifiers (UMIs; normalized by the sequencing depth of each cell) over all cells in the corresponding cell type. Spearman’s rank correlation coefficient. (C) scRNA-seq gene dropout rates of endothelial cell (EC) subpopulations at P7, with the number of cells sequenced in parentheses, as a function of bulk RNA-seq expression (FPKM) of FACSPurified total ECs at P7, as published (Vila Ellis et al., 2020). (D) scRNA-seq average gene expression for the 4 EC subpopulations, color/symbol coded as in C, as a function of bulk RNA-seq gene expression of total ECs. Over a wide range of FPKM values, subpopulation-specific genes are readily identifiable in scRNA-seq. Examples are marked with the same subpopulation symbols but in black. Thy1 is enriched in lymphatic ECs; Cd24a, Car4 and Igfbp7 are enriched in Car4 ECs; Plvap is depleted in Car4 ECs; Eln is enriched in Plvap ECs. One outlying, non-differential gene (Malat1, Table S1) is removed to decompresse the plot. (E) Table showing the first 10 genes for the first 10 cells to illustrate how average and total gene expression values are calculated. Non-zero values are highlighted in red and tallied to obtain the number of cumulative unique genes. (F) The number of genes detected, as tallied in E, plateaus as a function of the number of cells sequenced for a given cell type (Plvap ECs of adult lungs). Cells are sorted in ascending (light blue) or descending (dark blue) order in E by the number of genes detected in the cell. The dashed gray line marks 90 cells. (G) The number of genes detected for indicated cell types with the number of cells sequenced in parentheses. The dashed gray line marks 90 cells. E-G use scRNA-seq data from the adult lung (GSE144678).

The varying performance of scRNA-seq seemingly as a function of cell number prompted a systematic evaluation: we computationally sampled the Plvap EC population (Fig. 1E) and found that the number of detected genes rapidly approached the technical limit, such that ~90 cells were necessary to detect 10,000 genes (Fig. 1F). The same 90-cell cutoff held true when comparing cell types of different cell numbers (Fig. 1G) and therefore was used to guide the optimization of our scRNA-seq wet-lab protocol as described below.

Overall, our analyses showed that scRNA-seq performs as well as bulk RNA-seq in detecting and quantifying genes when computationally combining enough cells for a given cell type, and outperforms bulk RNA-seq in identifying cell type heterogeneity and associated marker genes.

**An optimized normalized lung scRNA-seq protocol**

The theoretical minimal number of cells, as determined above, could be difficult to obtain in practice, owing to highly skewed proportions of the dozens of cell types in a typical mammalian organ. For example, published lung scRNA-seq datasets had limited representation of endothelial and mesenchymal cells and significant variations in cell proportions across experiments (Fig. 2A). We reasoned that consistent and sufficient sampling of major lung cell types could be achieved by first purifying and then sequencing equal proportion of the 4 cell lineages – epithelial, endothelial, immune and mesenchymal lineages – which were generally considered non-interconvertible. As a benchmark, we determined via immunostaining the in vivo average proportions of the 4 listed cell lineages as 26%, 38%, 17% and 19% – a skewed and variable distribution that warranted consideration in experimental design (Fig. 2B). We then identified 3 cell surface markers that robustly distinguished the 4 lineages in FACS and, in comparison with our immunostaining results, introduced biases presumably due to varying efficiency in dissociating cells of different lineages (Fig. 2C). To reduce the cost of scRNA-seq, we remixed and sequenced equal numbers of cells from the purified 4 lineages after taking into account lineage-specific difference in cell viability (Fig. 2C).

This cell-lineage-level normalization was a cost-effective trade-off between non-selective whole-lung scRNA-seq and in-depth albeit narrow-focused cell type-specific scRNA-seq. Proportions of cell lineages and individual cell types within a lineage could be retrieved by analyzing FACS and scRNA-seq data, respectively (Fig. 2D). Our method routinely captured 18 lung cell types in a sufficient number to construct the interactome.

**Numeric representation of ligand-receptor interaction**

As ligand-receptor interaction was directional – consisting of ligand-expressing signaling cells and receptor-expressing receiving cells – we evaluated each cell type in our scRNA-seq for its potential as a ligand-expressing cell when paired with each of all cell types, including itself in the case of autocrine interaction (Fig. 3A; Table S2). For each of these directional cell type pairs, we used a scatterplot to visualize all 2356 ligand-receptor pairs, such that a data point off both axes indicated the presence of the

---

**Fig. 1. Cell type-level scRNA-seq data compare favorably to bulk RNA-seq data.** (A) scRNA-seq gene dropout rates of AT1 and AT2 cells, with the number of cells sequenced in parentheses, as a function of bulk RNA-seq expression (FPKM) of the corresponding FACS-purified cell types. The dropout rate is the percentage of genes not detected by scRNA-seq for a given range of bulk RNA-seq expression values. scRNA-seq data are from postnatal day (P)7 (GSE144769); bulk RNA-seq data are from P5 for AT1 cells and P8 for AT2 cells, as published (Little et al., 2019). (B) Consistency between AT1 cell (left) and AT2 cell (right) gene expression from scRNA-seq and bulk RNA-seq gene expression of the corresponding cell type. The scRNA-seq equivalent of the FPKM value in bulk RNA-seq, which sums the expression in all cells of the sample, is considered the sum of normalized unique molecular identifiers (UMIs; normalized by the sequencing depth of each cell) over all cells in the corresponding cell type. Spearman’s rank correlation coefficient. (C) scRNA-seq gene dropout rates of endothelial cell (EC) subpopulations at P7, with the number of cells sequenced in parentheses, as a function of bulk RNA-seq expression (FPKM) of FACSPurified total ECs at P7, as published (Vila Ellis et al., 2020). (D) scRNA-seq average gene expression for the 4 EC subpopulations, color/symbol coded as in C, as a function of bulk RNA-seq gene expression of total ECs. Over a wide range of FPKM values, subpopulation-specific genes are readily identifiable in scRNA-seq. Examples are marked with the same subpopulation symbols but in black. Thy1 is enriched in lymphatic ECs; Cd24a, Car4 and Igfbp7 are enriched in Car4 ECs; Plvap is depleted in Car4 ECs; Eln is enriched in Plvap ECs. One outlying, non-differential gene (Malat1, Table S1) is removed to decompresse the plot. (E) Table showing the first 10 genes for the first 10 cells to illustrate how average and total gene expression values are calculated. Non-zero values are highlighted in red and tallied to obtain the number of cumulative unique genes. (F) The number of genes detected, as tallied in E, plateaus as a function of the number of cells sequenced for a given cell type (Plvap ECs of adult lungs). Cells are sorted in ascending (light blue) or descending (dark blue) order in E by the number of genes detected in the cell. The dashed gray line marks 90 cells. (G) The number of genes detected for indicated cell types with the number of cells sequenced in parentheses. The dashed gray line marks 90 cells. E-G use scRNA-seq data from the adult lung (GSE144678).
corresponding ligand and receptor, as exemplified by the expected Vegfa-Kdr expression in the AT1 cell-Car4 EC pair (Vila Ellis et al., 2020; Yang et al., 2016) (Fig. 3A). In these scatterplots, user-defined horizontal and vertical thresholds could be used to tally all ligand-receptor pairs present in specific cell type pairs—an approach commonly employed in the literature but at the expense of available quantitative expression values (Camp et al., 2017; Cohen et al., 2018; Skelly et al., 2018). Furthermore, the number of ligand-receptor pairs was not necessarily a valid predictor of functional interactions, and a single threshold was unlikely to suit all ligands and receptors expressed at varied levels.

To overcome these limitations, we sought a numeric representation of ligand-receptor interaction, which we reasoned should positively correlate with the expression values of both the ligand and the receptor. Inspired by the principle of chemical equilibrium where the product—the ligand-receptor complex in this case, which should determine the signaling output—equals the product of the substrates divided by the equilibrium constant, we chose to represent the interaction strength by multiplying the average scRNA-seq expression values of the ligand and the receptor in the involved cell types (Fig. 3B). By assuming comparable protein translation and delivery among cell types, this...
numeric representation allowed us to leverage 324 cell type pairs for a given ligand-receptor pair and readily captured potential interactions as outliers. This was demonstrated using heatmaps for the known interactions between AT1 cells and Car4 ECs via Vegfa and Kdr, and between ECs and pericytes via Pdgfb and Pdgfrb (Fig. 3C). These visually outlying interactions were quantitatively defined as those outside 3 times the interquartile range (Tables S3 and S4).

Identify interactions altered upon viral infection

Next, we extended our numeric interactome analysis to compare lungs at baseline and upon perturbation. We chose a Sendai virus infection model (Goldblatt et al., 2020), instead of genetic models, because infection was expected to induce global changes involving multiple cell types and their associated cell interactions, necessitating a quantitative genomic analysis as our interactome method. At 2 weeks after infection, when the virus had largely been cleared and the lung was undergoing repair (Goldblatt et al., 2020; Holtzman et al., 2005), scRNA-seq detected infection-induced proliferation of ECs and AT2 cells, as well as aberrant appearance of Trp63-expressing basal-like cells (Fig. 4A,B), reminiscent of pods or lineage-negative epithelial progenitors observed upon severe H1N1 virus infection (Kumar et al., 2011; Vaughan et al., 2015; Zuo et al., 2015).

For each directional cell type pair, we compared control and infected lungs by plotting the corresponding numeric interactions of individual ligand-receptor pairs, such that data points above or below the diagonal line represented enhanced or diminished interactions, respectively, whereas those on the y- or x-axis represented de novo or lost interactions, respectively. One example was an infection-induced interaction between myeloid cells and T cells via Cxcl9 and Cxcr3 (Fig. 4C,D; Table S5). This was supported by CXCR3 immunostaining, showing its upregulation in clustered T cells possibly as a result of chemotaxis (Groom and Luster, 2011; Weng et al., 1998) (Fig. 4E).

To quantify infection-induced changes in interactions, we resorted to the aforementioned concept of outliers by pooling the differences in a given ligand-receptor interaction across 324 cell type pairs and used the same 3 times interquartile range cutoffs (Table S5). We found that 3.6% (27,611 out of 763,344) interactions were outliers, involving 448 and 433 unique ligands and receptors in 289 cell type pairs and averaging 85 outlying interactions per cell type pair (Table S5).
Fig. 4. See next page for legend.
Integrate interactomes with signaling pathway analysis

To corroborate our interactomes and capitalize on the transcriptomic information available from scRNA-seq, we sought to analyze the activities of signaling pathways and integrate them with associated ligand-receptor pairs. We chose signaling pathways annotated by Gene Ontology (GO), instead of Ingenuity Pathway Analysis (IPA), because of its public availability and inclusiveness. Although these databases did not take into account specific biological contexts, limiting their use in general, we reasoned that by averaging over a sufficient number of bona fide, context-independent pathway members, a pathway signature might be still detectible, as assumed in gene set enrichment analysis (Subramanian et al., 2005).

We considered all genes in each GO pathway as a metagene and generated the corresponding GO pathway (i.e. metagene) score for each cell, which was then averaged over all cells of a given cell type (Fig. 5A; Table S6). Using the same outlier concept, albeit with a less stringent cutoff of 1.5 times the interquartile range because of the limited number of cell types available (18 cell types, instead of 324 cell type pairs), we compared GO pathway scores across 18 cell types, as well as their changes upon infection, and identified outlying pathways (Tables S7 and S8). We identified 4% (276 out of 6300) outlying pathways at baseline and 5% (321 out of 6300) outlying alterations upon infection. To integrate with our interactomes, for a given outlying ligand-receptor pair, we required the receptor to be a member of an outlying signaling pathway.

Fig. 4. Interactome analysis identifies outlying changes in interaction upon Sendai virus infection. (A) Uniform Manifold Approximation and Projection (UMAP) plots (left) of scRNA-seq of control and Sendai virus-infected lungs at 2 weeks post-infection, showing 18 cell types, as identified by their marker genes in a dot plot (right). (B) Feature plots showing increased proliferation in AT2 cells and ECs (Mki67; arrowheads), as well as basal-like cells (Tsp63; arrow) upon infection. (C) Schematic illustrating comparison of control and Sendai virus-infected lungs for each directional cell type pair. Autocrine interactions are boxed in red along the diagonal. A hypothetical cell type pair (D to B) illustrates the locations of possible changes in interactions, as demonstrated in a real example showing 2 color-coded interactions. Cell types are abbreviated as in Fig. 2D. The unit UMI² indicates the product of ligand and receptor expression. (D) Dot plots showing upregulation of Cxcl9 and Cxcr3 upon infection. (E) Confocal images of immunostained control and Sendai virus-infected lungs showing upregulation of CXCR3 in T cells (CD3E), especially when clustered possibly as a result of chemotaxis via CXCR3 (arrowheads). The asterisk indicates non-specific extracellular matrix staining. Scale bar: 10 µm.

Fig. 5. Integration of interactome and pathway scores to prioritize interactions. (A) Schematic illustrating the integration of outlying interactions (left) and outlying pathways (right) occurring within the same receiving cell type. Outliers are based on comparison of an array of n² (324 in this study) cell type pairs for interactions and a column of n (18 in this study) cell types for pathways. Common outliers in interactions/pathways or their changes have the receptor gene included in the pathway gene list in the same receiving cell type. (B) Dot plots and heatmaps illustrating a real example of Bdnf-Ntrk2 being both an outlying interaction and an outlying pathway in ECs (double arrowhead). The BDNF pathway is also outlying for the Wnt5a fibroblast population upon infection (asterisk) – likely driven by increased Bdnf expression (dot plot), which is not identified as an outlying ligand-receptor interaction as expected from our algorithm, demonstrating the utility of the integrated analysis. The cell types for the columns are not fully labeled but are in the same order as those for the rows, as in Fig. 3B.
pathway in the same cell type as the receiving one in the interactome (Fig. 5A). This integration led to 0.3% (2323 out of 763,344) outlying interactions at baseline and 0.7% (5516 out of 763,344) outlying alterations upon infection that were also supported by the corresponding pathway activation (Tables S9 and S10). One example was infection-enhanced Bdnf-Ntrk2 signaling from AT1 cells to ECs, largely driven by Ntrk2 upregulation (Fig. 5B). It was tempting to speculate that the supposedly angiogenic role of this signaling (Dalton et al., 2015; Kermani and Hempstead, 2007) contributed to the infection-induced EC proliferation (Fig. 4B).

**DISCUSSION**

Meaningful interpretation of rapidly growing single-cell data depends on biological insights of bench scientists, who are often discouraged by the complexity of existing computational tools. As bench scientists ourselves, we describe here optimized experimental and computational methods to construct single-cell interactomes in lungs at baseline and upon infection. Our approaches are intuitive and readily adaptable to study other organs and species by those equipped with a consumer-grade computer and an intermediate level of R programming skills, but without formal education in computer and statistical sciences. Our interactome improves upon existing ones (Camp et al., 2017; Cohen et al., 2018; Kumar et al., 2018; Raredon et al., 2019; Skelly et al., 2018; Vento-Tormo et al., 2018; Vieira Braga et al., 2019) in the following 3 aspects.

First, we systematically evaluate and then capitalize on our observations that, at the cell type level, scRNA-seq has comparable sensitivity to genetic driver-based bulk RNA-seq and yet is more robust in identifying and purifying a given cell type (Fig. 1). Our cell type-level analysis bypasses the gene dropout issue inherent to scRNA-seq and imputation methods that are still under development (Gong et al., 2018; Hicks et al., 2018; Huang et al., 2018; Kharchenko et al., 2014; van Dijk et al., 2018). Experimentally, our normalization method is guided by computational assessment of the minimal cell number needed and captures all major lung cell types without inhibitory cost or resorting to non-commercial platforms. A potential concern about cell type-level analysis is failure to capture cellular heterogeneity; this, however, can be alleviated by careful cell type identification. Furthermore, a cell type that is not sufficiently distinct or abundant may not be reliably analyzed even with individual-cell-level algorithms, as evidenced by the significant cell outlying interactions at baseline and 0.7% outlying alterations upon infection that were also supported by the corresponding pathway activation (Tables S9 and S10). One example was infection-enhanced Bdnf-Ntrk2 signaling from AT1 cells to ECs, largely driven by Ntrk2 upregulation (Fig. 5B). It was tempting to speculate that the supposedly angiogenic role of this signaling (Dalton et al., 2015; Kermani and Hempstead, 2007) contributed to the infection-induced EC proliferation (Fig. 4B).

Experimentally, our normalization method is guided by computational assessment of the minimal cell number needed and captures all major lung cell types without inhibitory cost or resorting to non-commercial platforms. A potential concern about cell type-level analysis is failure to capture cellular heterogeneity; this, however, can be alleviated by careful cell type identification. Furthermore, a cell type that is not sufficiently distinct or abundant may not be reliably analyzed even with individual-cell-level algorithms, as evidenced by the significant cell outlying interactions at baseline and 0.7% outlying alterations upon infection that were also supported by the corresponding pathway activation (Tables S9 and S10). One example was infection-enhanced Bdnf-Ntrk2 signaling from AT1 cells to ECs, largely driven by Ntrk2 upregulation (Fig. 5B). It was tempting to speculate that the supposedly angiogenic role of this signaling (Dalton et al., 2015; Kermani and Hempstead, 2007) contributed to the infection-induced EC proliferation (Fig. 4B).

**Materials and Methods**

**Sendai virus infection**

Sendai virus (parainfluenza type 1) strain 52 (ATCC #VR-105, RRID: SCR_001672) was expanded in primary rhesus monkey kidney cells (Cell Pro Labs #103-175). Wild-type 8-week-old C57Bl/6 mice obtained from The Jackson Laboratory were anesthetized with isoflurane, suspended by the maxillary incisors on a dosing board with a 60° incline, and infected with a sub-lethal dose of 2.1×10^6 plaque-forming units (pfu) of virus in 40 μl phosphate buffered saline (PBS) through oropharyngeal aspiration. For immunostaining, both male and female mice were used. For scRNA-seq, a pair of male mice was used. Investigators were not blind to the group allocation. No power calculation was used to determine the sample size. All animal experiments were approved by the Institutional Animal Care and Use Committee at Texas A&M Health Science Center Institute of Biosciences and Technology and MD Anderson Cancer Center.

**Section immunostaining and cell counting**

For cell lineage counting by immunostaining, lungs were inflated and harvested as previously described (Yang et al., 2016). Briefly, mice were anesthetized using Avertin (T45402, Sigma-Aldrich) and the lungs perfused with PBS. The trachea was cannulated and the lungs were inflated to full capacity with 0.5% paraformaldehyde (PFA; P6148, Sigma-Aldrich) in PBS at 25 cm H2O pressure. The harvested lungs were immersion-fixed in 0.5% PFA at room temperature for 4-6 h and washed overnight in PBS at 4°C. Section immunostaining was performed following published protocols with minor modifications (Alainis et al., 2014; Chang et al., 2013). Fixed lungs were dissected and cryoprotected overnight at 4°C in 20% sucrose in PBS containing 10% optimal cutting temperature compound (OCT; 4583, Tissue-Tek). The lobes were then frozen in OCT-filled embedding molds on a dry ice and 100% ethanol slurry then kept at −80°C until sectioned. Sections were obtained at 10 μm thickness and then blocked in PBS with 0.3% Triton X-100 and 5% normal donkey serum (017-000-121, Jackson ImmunoResearch). Blocked sections were incubated in a humidified chamber at 4°C overnight with the following primary antibodies diluted in PBS containing 0.3% Triton X-100: anti-GFP (chicken, 1:5000, Abcam, AB13970), anti-EGR (rabbit, 1:5000, Abcam, AB92513), anti-CD45 (rat, 1:2000, eBioscience, 14-0451-81), anti-CD3E (Amersham, 1:250, BioLegend, 100301) and anti-CXCR3 (rat, 1:250, R&D Systems, MAB1685). The sections were submerged in PBS for 30 min and incubated with 4’,6-diamidino-2-phenylindole (DAPI) and secondary antibodies (Jackson ImmunoResearch) diluted in PBS with 0.3% Triton X-100 at room temperature for 1 h. The sections were washed
again in PBS then mounted with Aqua-Poly/Mount (18606, Polysciences) and imaged on a Nikon A1plus confocal microscope or an Olympus FV1000 confocal microscope. For cell lineage counting, 3 images with airways located in the center were taken from two Rosa<sup>Smo<sup>GFP</sup></sup>+/<sup>-</sup>; Shh<sup>Cre</sup>+/<sup>-</sup> lungs, in which all epithelial cells were labeled with GFP (Little et al., 2019) to allow co-staining for endothelial and immune cells. Endothelial cells (ERG<sup>+<sup></sup></sup> nuclei) and alveolar epithelial cells (GFP<sup>+</sup> nuclei) excluding the airways were counted using Fiji’s ‘Find Maxima’ function. Airway epithelial cells were counted by drawing a region of interest around airways and using Fiji to Find Maxima for DAPI. Immune cells (CD45<sup>+</sup>) were counted by drawing a region of interest around the expected percentage, 25%, and lineage-specific viability correction determined the percentages of each lineage. These percentages were compared to the number of cells assigned to each cell type by the total number of cells in the sample. The number of unique genes detected in each cell type was counted. Dot plots were used to visualize the marker genes identifying each cell type.

**Cell dissociation and labeling**
Mice were anesthetized using Avertin injected intraperitoneally, the chest cavity was exposed and the lungs were perfused with PBS injected through the right ventricle. After clearance of circulating cells, the lungs were removed into PBS and finely minced using forceps. The lung tissue was digested in Leibovitz’s L-15 medium (Gibco, 21083-027) with 2 mg/ml collagenase type B (Worthington, CLS-1, LS004197), 2 mg/ml elastase (Worthington, ESL, LS002294) and 0.5 mg/ml DNase I (Worthington, D, LS002007) for 30 min at 37°C heat block. Digestion was stopped by addition of fetal bovine serum (FBS; Invitrogen, 10082-139) to a final concentration of 20%. The samples were moved to ice and all remaining steps were performed in a cold room. The tissue was triturated at 15 min into digestion and also after digestion was quenched. The resulting cell suspension was filtered through a 70 μm cell strainer (Falcon, 352350), centrifuged at 1500 g for 1 min and the pellet resuspended in red blood cell lysis buffer (15 mM NH<sub>4</sub>Cl, 12 mM NaHCO<sub>3</sub>, 0.1 mM EDTA, pH 8.0) for 3 min. The cells were pelleted again at 5000 rpm for 1 min and washed with Leibovitz’s L-15 medium with 10% FBS, then filtered into a 5 ml tube with a cell strainer cap (Falcon, 352235). The cells were incubated with the following conjugated antibodies: CD45-PE/Cy7 (BioLegend, 103114), ICAM2-A647 (Life tech, A15452), E-cadherin-A488 (Invitrogen, 53-3249-82) at a concentration of 1:250 for 30 min. The cells were centrifuged at 1500 g for 1 min, washed in Leibovitz’s L-15 medium with 10% FBS for 5 min and finally resuspended in Leibovitz’s L-15 medium with 10% FBS and filtered through a strainer cap. The sample was incubated with SYTOX Blue (Invitrogen, S34857) and sorted on a BD FACSaria Fusion Cell Sorter. The cells were collected in a volume of 250 μl PBS with 10% FBS for each collection tube.

**FACS sorting and normalization**
After exclusion of dead cells and doublets, cells were gated into the 4 cell lineage populations using a serial gating strategy. All CD45<sup>+</sup> cells were collected as the immune population; CD45<sup>-</sup> cells exhibiting ICAM2<sup>+</sup> signal were collected as endothelial cells; CD45<sup>-</sup>, ICAM2<sup>-</sup>, but E-cadherin<sup>+</sup> cells were collected as epithelial cells; cells negative for all markers were collected as mesenchymal cells. To determine cell lineage population-specific reduction in viability after sorting, an equal number of cells from each lineage were mixed and, after adding fresh SYTOX Blue, re-sorted to tally the number of unique genes detected in each cell type. The number of unique genes detected in each cell type was counted. Dot plots were used to visualize the marker genes identifying each cell type.

**scRNA-seq data analysis**

**Data generation**
The normalized whole-lung samples were processed through the Chromium Single Cell Gene Expression Solution Platform (10× Genomics) using the Chromium Single Cell 3′ Library and Gel Bead Kit (v2, rev D). Library sequencing was performed on an Illumina NextSeq500 using a 26X124 sequencing run format with 8 bp index (Read 1). The single-cell reads were aligned against the mm10 mouse reference genome (provided by 10× Genomics), counted and aggregated using the Cell Ranger pipeline (version 3.0, 10× Genomics). Raw data for the P7 wild-type lung and control and 2 weeks post-Sendai virus infection lungs were deposited at NCBI Gene Expression Omnibus (GEO) under accession numbers GSE144769 and GSE144678. R codes for figures and supplementary tables are included in the Supplementary Data.

**Cell type identification**
Data processing was performed on a personal computer equipped with an Intel Core i5-6300U and 16.0 GB RAM using the R package Seurat (version 3.1.2) (Butler et al., 2018). Unless specified, default parameters were used. Single-cell count data used for all downstream calculations was normalized to the number of reads per cell using Seurat’s normalization function (normalized UMI). Data were scaled before undergoing dimensionality reduction using the RunPCA Seurat function. Cells were visualized on a projection map using the RunUMAP function. Feature plots for known lung cell type markers were used to identify the cell types clustered using the Louvain algorithm implemented in the FindClusters function after cells were embedded in a graph using FindNeighbors. The resolution parameter was adjusted to appropriately cluster the cells based on the known cellular gene markers. The cell type distributions were calculated by dividing the number of cells assigned to each cell type by the total number of cells in the sample. The number of unique genes detected in each cell type was counted. Dot plots were used to visualize the marker genes identifying each cell type.

**Bulk RNA-seq versus scRNA-seq comparison**
To calculate the dropout rate of genes in the scRNA-seq data, normalized UMI averaged across all cells for a given cell type was compared to gene expression (FPKM) from the corresponding bulk RNA-seq data (AT1 and AT2 cells from GSE129583; endothelial cells from GSE124324). For genes within specified bins of bulk RNA-seq expression values, the fraction of genes with zero expression in scRNA-seq was calculated and plotted.

**Analysis of the number of genes expressed per cell type**
The number of genes detected by each cell type of the cell was counted. The cells were ordered both by increasing and by decreasing the number of genes expressed. Then, as one cell was added after another, the number of unique genes detected in that cell not detected in the previous cells was tallied, resulting in the cumulative number of unique genes detected for that cell type.

**Interactome analysis**
A list of ligand-receptor pair genes from a published human database (Ramilowski et al., 2015) was matched to the gene names present in the Seurat object. Out of 708 ligand genes from the human database, 647 gene names matched in Seurat, and out of 691 receptor gene names, 646 matched. The normalized ligand and receptor gene count data were accessed from the Seurat object, averaged over all the cells of each cell type and plotted for each cell type pair. For each ligand-receptor pair for each cell type pair, we multiplied the average ligand and receptor gene expression. The heatmap visualizing these values was created using ggplot2 (version 3.2.1). To identify outlying interactions, we computed the interquartile range (IQR) across all cell type pairs for each ligand-receptor interaction and used a cutoff of 3 times IQR.

**GO pathway score**
The 350 GO pathway gene lists, publicly available (GO Consortium; containing the phrase ‘SIGNALING_PATHWAY’ in the ‘Biological Process’ category), were used to compute pathway scores for each cell type using the AddModuleScore function in Seurat. Outlier pathways were identified for each cell type by applying a 1.5 times IQR cutoff across the scores for all cells for each pathway. To identify possible pathway activation as an outcome of a signaling interaction, the gene list for each outlying pathway was used to match the receptor genes of outlying interactions for the same cell type as the signal-receiving cell.

**Acknowledgements**
We thank Odemaris Narvaez del Pilar for generating the scRNA-seq data of the P7 mouse lung.

**Competing interests**
The authors declare no competing or financial interests.
human lungs identifies novel cell states in health and in asthma. *Nat. Med.* **25**, 1153-1163. doi:10.1038/s41591-019-0468-5

Vila Ellis, L., Cain, M. P., Hutchison, V., Fiodby, P., Crandall, E. D., Borok, Z., Zhou, B., Ostrin, E. J., Wythe, J. D. and Chen, J. (2020). Epithelial vegfa specifies a distinct endothelial population in the mouse lung. *Dev. Cell* **52**, 817-830.e6. doi:10.1016/j.devcel.2020.01.009

Weng, Y., Siciliano, S. J., Waldburger, K. E., Sirotina-Meisher, A., Staruch, M. J., Daugherty, B. L., Gould, S. L., Springer, M. S. and DeMartino, J. A. (1998). Binding and functional properties of recombinant and endogenous CXCR3 chemokine receptors. *J. Biol. Chem.* **273**, 18288-18291. doi:10.1074/jbc.273.29.18288

Yang, J., Hernandez, B. J., Martinez Alanis, D., Narvaez del Pilar, O., Vila-Ellis, L., Akiyama, H., Evans, S. E., Ostrin, E. J. and Chen, J. (2016). The development and plasticity of alveolar type 1 cells. *Development* **143**, 54-65. doi:10.1242/dev.130005

Yuan, T., Volckaert, T., Chanda, D., Thannickal, V. J. and De Langhe, S. P. (2018). Fgf10 signaling in lung development, homeostasis, disease, and repair after injury. *Front. Genet.* **9**, 418. doi:10.3389/fgene.2018.00418

Zuo, W., Zhang, T., Wu, D. Z., Guan, S. P., Liew, A. A., Yamamoto, Y., Wang, X., Lim, S. J., Vincent, M., Lessard, M. et al. (2015). p63^+Krt5^+ distal airway stem cells are essential for lung regeneration. *Nature* **517**, 616-620. doi:10.1038/nature13903