1 **Integrative Analysis of Spatial Transcriptome with Single-cell Transcriptome and Single-cell**
2 **Epigenome in Mouse Lungs after Immunization**

3

4 Zhongli Xu, MDc PhDc[1,2], Xinjun Wang, PhDc [3], Li Fan, PhD[4], Fujing Wang, PhD[4], Jiebiao Wang,
5 PhD[3], Wei Chen, PhD[1,3,$], Kong Chen, PhD[4,$]

6

7 [1]Department of Pediatrics, University of Pittsburgh, PA, USA. [2]School of Medicine, Tsinghua
8 University, Beijing, China. [3]Department of Biostatistics, University of Pittsburgh, PA, USA.
9 [4]Department of Medicine, University of Pittsburgh, PA, USA.

10

11 [$]**Corresponding Author:**   Kong Chen, PhD

12   Department of Medicine

13   University of Pittsburgh

14   3459 Fifth Avenue, Pittsburgh, PA 15213

15   Phone: 412.692.2118; Fax: 412.692.2260

16   Email: koc5@pitt.edu

17

18   Wei Chen, PhD

19   Department of Pediatrics

20   University of Pittsburgh

21   UPMC Children's Hospital of Pittsburgh

22   4401 Penn Ave, Pittsburgh, PA 15224

23   Phone: 412.692.6241

24   Email: wec47@pitt.edu

25 **ABSTRACT**

26 Immunological memory is key to productive adaptive immunity. An unbiased, high throughput
27 gene expression profiling of tissue-resident memory T cells at their precise anatomical locations
28 within the lung is fundamental to understanding lung immunity, but such spatial information
29 has yet to be characterized. In this study, using a well-established *Klebsiella pneumoniae*
30 infection model, we performed an integrative analysis of spatial transcriptome with single-cell
31 RNA-seq and single-cell ATAC-seq on lung cells from mice after immunization using the 10x
32 Genomics Chromium and Visium platform. We employed several deconvolution algorithms and
33 established an optimized deconvolution pipeline to accurately decipher specific cell-type
34 composition by anatomic location. We identified and located 12 major cell types by scRNA-seq
35 and spatial transcriptomic analysis. Integrating scATAC-seq data from the same cells processed
36 in parallel with scRNA-seq, we found epigenomic profiles provide more robust cell type
37 identification, especially for lineage-specific T helper cells. When combining all three data
38 modalities, we observed a dynamic change in the location of T helper cells as well as their
39 corresponding chemokines for chemotaxis. Furthermore, cell-cell communication analysis of
40 spatial transcriptome provided evidence of lineage-specific T helper cells receiving designated
41 cytokine signaling. In summary, our first-in-class study demonstrated the power of multi-omics
42 analysis to uncover intrinsic spatial- and cell-type-dependent molecular mechanisms of lung
43 immunity. Our data provides a rich research resource of single cell multi-omics data as a
44 reference for understanding spatial dynamics of lung immunization.

45

46 **INTRODUCTION**

47 Immunological memory, consisting of B cell and T cell memory, is a key characteristic of
48 adaptive immunity upon encountering with pathogen invasion. Typically, memory T cells can be
49 classified into two categories: effector memory and central memory T cells. Effector memory T
50 cells are able to produce effector cytokines and have cytotoxic activity against pathogen
51 infected cells [1]. Tissue-resident memory T cells (TRM) have more recently been defined as a
52 new subset, predominantly residing in mucosal tissues, barrier surfaces, and other non-
53 lymphoid organs but are also present in lymphoid sites [2]. TRM are antigen experienced and are
54 capable of rapidly responding to re-exposure to cognate antigen. TRM in the lung have been
55 demonstrated to exhibit robust protective function against constant viral and bacterial
56 challenge of the lungs and respiratory tract.

57 *Klebsiella pneumoniae* is an important cause of community-acquired pneumonia. In 2011, the
58 U.S. National Institutes of Health Clinical Center experienced an outbreak of carbapenem-
59 resistant *K. pneumoniae* that affected 18 patients, 11 of whom died. Thus, in addition to
60 antimicrobial stewardship and hospital hygiene measures, there is a critical need for the
61 development of novel therapeutic approaches to prevent and/or treat antibiotic resistant
62 infections. Our previous work has demonstrated that *K. pneumoniae* specific Th17 cells are
63 induced by immunization with whole bacterial lysate [3].These memory Th17 cells are both
64 required and sufficient to provide serotype/antibody independent protection against a variety
65 of strains of *K. pneumonia* including the recently described multidrug resistant New Delhi
66 metallo-beta-lactamase strain.

67 Tissue localization of these TRM cells has been investigated thoroughly in the past and
68 comparisons between mouse and human TRM have been characterized extensively using the
69 well-established flow cytometric and transcriptomic approaches [4]. However, an unbiased, high
70 throughput gene expression profiling of TRM residing in various anatomical location within the
71 lung, such as airway vs parenchyma, has not been possible using conventional flow cytometry,
72 transcriptomic approaches, and even the most recently developed single-cell sequencing
73 technology. These methodologies are limited as anatomic location specific information is lost
74 after single-cell suspension is acquired.

75 Single-cell RNA-seq (scRNA-seq) has been gradually applied to study immune cells and immune
76 responses in mouse lungs [5–8]. Few recent studies utilize single-cell ATAC-seq (scATAC-seq) to
77 measure the chromatin architecture of immune cells in mouse lungs [9]. Although these
78 technologies provide rich information in understanding cell heterogeneity and biology
79 information of mouse lung, spatial information of single cell is lost in the process. Spatial
80 transcriptomics (ST) is a recently developed technology and has the ability to map
81 transcriptional signatures to distinct anatomical regions. To date, it has rarely been used in
82 understanding lung tissues. In this first-in-class study, we employed a commercially available ST
83 platform to investigate the spatial topography of gene expression of mouse lungs after
84 immunization. To overcome the resolution limitation of current ST technology, we also applied
85 state-of-art single-cell transcriptomics and single-cell epigenomics to jointly study the spatial
86 localization, transcriptome, and epigenome of T cells induced by this immunization. The

87 integrative analysis of three types of omics data provides an unprecedented and
88 comprehensive way to examine the genomic basis and dynamics of lung immunization.

89

90 **RESULTS**

91 **Cataloging 12 Lung Cell Types using scRNA-seq Data**

92 In this study, we inoculated mice with heat-killed *K. pneumoniae* and assigned them into two
93 groups, the immunized and the re-challenged groups. The immunized group was inoculated
94 twice on day 0 and day 7, whereas the re-challenged group was additionally inoculated on day
95 13. On day 14, both groups of mice were sacrificed for tissue harvesting. Slices A1 and A2 were
96 sectioned from the fresh frozen lung of the immunized mouse, whereas slices A3 and A4 were
97 from the re-challenged mouse. In a separate cohort, lung tissue from the re-challenged mouse
98 were harvested and single-cell suspension was obtained after enzymatic digestion. The four
99 lung slices were used for spatial transcriptomics, and the single-cell suspension from the re-
100 challenged mouse were subjected to scRNA-seq and scATAC-seq analyses. scRNA-seq data were
101 integrated with scATAC-seq data via label transfer and were used as a reference to deconvolute
102 spatial transcriptomics data (**Figures 1A**).

103 We generated scRNA-seq data from 3,337 cells collected from the re-challenged mouse. Graph-
104 based clustering identified ten clusters (**Figures S1A and S1C**). Cluster 1, which has an extremely
105 high percentage of mitochondrial genes (**Figures S1B**), was excluded from downstream
106 analyses. Cluster 6 consisted of several sporadic sub-clusters and was re-clustered into four
107 clusters (**Figures S1D and S1E**). 2,834 cells with good quality were retained for the downstream
108 analyses.

109 We carried out fine cluster annotation according to canonical markers and identified 12 lung
110 cell types (**Figures 1B**), including alveolar epithelial cells, club cells, fibroblasts, endothelial cells,
111 monocytes, macrophages, dendritic cells, neutrophils, B cells, T cells, NK cells, and erythrocytes.
112 Selected canonical markers for each cell type were shown in the dot plot (**Figures 1C**, *e.g.*, *Sftpb*
113 for alveolar epithelial cells, *Scgb1a1* for club cells, and *Cd3d* for T cells). The top five (by log2-
114 Fold Change) markers for each cell type were visualized using a heatmap (**Figures 1D**).

115

116 **Robust Cell-type Decomposition (RCTD) of Spatial Transcriptome using scRNA-seq Data**

117 Although spatial transcriptomics provides additional spatial information, its resolution has not
118 reached to single-cell level. Therefore, the expression profile of each spot in spatial
119 transcriptomics is from a mixture of a few cells, typically one to ten. To better interpret spatial
120 transcriptome data, it is vital to determine the proportions of different cell types within each
121 spot. Using our finely annotated scRNA-seq data as a reference, we carried out deconvolution
122 for each spot in the four slices using the robust cell-type decomposition (RCTD) method [10]. The
123 deconvolution results for slice A3 were shown as proportions of 12 cell types across slice A3
124 (**Figures S2A**).

125 To assess the robustness of this deconvolution method, we inspected whether well-
126 characterized cell types were colocalized with the expression of their canonical markers, as well

127 as corresponding histological structures. In slice A3, spots with a high proportion of club cells
128 were around the histological airways and colocalized with spots with increased expression of
129 *Scgb1a1* (**Figures 2Ai-iii**).

130 To examine the robustness and impact of reference panel on deconvolution, another public
131 scRNA-seq data (GEO: GSE119228) [5], including 20 annotated mouse lung cell types (**Figures
132 2Bi**), were also used as an independent reference to deconvolute spatial transcriptome for the
133 four slices. The deconvolution results for slice A3, using public scRNA-seq reference, were
134 shown (**Figures S2B**). In slice A3, proportions of T cells deconvoluted using in-house or public
135 scRNA-seq references were quite similar (**Figures 2Bii and 2Biii**). Deconvolution results can also
136 be represented using proportion matrices, whose rows indicate spots and columns indicate cell
137 types. To quantify the similarity between the deconvolution results using the two independent
138 references, Pearson correlation coefficients between the columns of the two proportion
139 matrices were calculated and visualized using correlation heatmaps (**Figures 2Ci-iv**). Although
140 we cannot perfectly match all the 12 cell types from our in-house data with the 20 cell types
141 identified in this public dataset, we found the proportions of some adaptive immune cells (*e.g.,*
142 T cells and B cells) and innate immune cells (*e.g.*, neutrophils and NK cells) were highly
143 correlated with those deconvoluted using the other reference in all four slices. Other cell types
144 were also highly correlated with their related cell types deconvoluted using the other reference
145 (*e.g.*, club cells in in-house data with ciliated cells in public data, alveolar epithelial cells in in-
146 house data with AT2 in public data). These analyses validated the robustness of the robust cell-
147 type decomposition (RCTD) method.

148

149 **Integrative Analysis of scRNA-seq and scATAC-seq Data Enabling the Identification of Th17**
150 **and Th1 Cells**

151 In parallel with scRNA-seq, we also generated scATAC-seq data from 4,908 cells collected from
152 the re-challenged mouse. After excluding cells with low quality, 4,794 cells were retained for
153 downstream analyses. Leveraging our finely annotated scRNA-seq data, we identified the 12
154 lung cell types in scATAC-seq data via label transfer (**Figures 3A**). Proportions of the 12 cell
155 types in scRNA-seq and scATAC-seq data were quite similar (**Figures 3B**), showing a good
156 biological agreement between two types of data.

157 Because we are interested in T cells in this study, we carried out re-clustering for T cells in
158 scATAC-seq data and identified four subtypes of T cells (**Figures 3C**), including Th17 cells, Th1
159 cells, other T cells 1, and other T cells 2. The top 20 (by log2-Fold Change) markers (gene activity
160 scores) for each subtype of T cells were visualized using a heatmap (**Figures 3D**). Canonical
161 transcription factors for subtypes of T cells were shown on the heatmap (*e.g.*, *Rorc* and *Rora* for
162 Th17 cells, *Tbx21* for Th1 cells). *Cd8b1* was a marker for other T cells 1, suggesting they could be
163 cytotoxic T cells.

164 To confirm the identity of Th17 and Th1 cells, further analyses were performed. For Th17 cells,
165 there were much more peaks in genomic regions around *Il17a* and *Rorc*, compared with other
166 subtypes of T cells (**Figures 3Ei and 3Eii**). In contrast, the peaks from Th1 cells dominated
167 genomic regions around *Ifng* and *Tbx21* (**Figures 3Eiii and 3Eiv**). Motif footprinting analysis

168 further confirmed RORC was dominated by Th17 cells, and TBX21 was dominated by Th1 cells
169 (**Figures S3Ai-ii and S3Bi-ii**).

170 Although subtypes of T cells were almost indistinguishable only using scRNA-seq data, we
171 managed to identify the four subtypes of T cells via label transfer from scATAC-seq data to
172 scRNA-seq data (**Figures 3F**). To confirm the identity of Th17 and Th1 cells in scRNA-seq data,
173 further analyses were performed. Selected canonical markers for Th17 and Th1 cells were
174 shown in the dot plot (**Figures 3G**, *e.g.*, *Rora* for Th17 cells, *Tbx21* for Th1 cells). The top 20 (by
175 log2-Fold Change) markers for each subtype of T cells were visualized using a heatmap (**Figures
176 S3C**). Canonical markers for subtypes of T cells can be seen in the heatmap (*e.g.*, *Rora* for Th17
177 cells, *Ifng* for Th1 cells, and *Cd8b1* for other T cells 1). We also conducted gene regulatory
178 network analysis using SCENIC [11,12] and identified RORA as a cell-type specific regulator for Th17
179 cells in scRNA-seq data (**Figures S3D**). By integrative analysis of scRNA-seq and scATAC-seq
180 data, 15 lung cell types (including four subtypes of T cells) were identified in scRNA-seq data
181 (**Figures 3H**), with the potential to be deconvoluted in spatial transcriptome.

182

183 **Dynamic Changes of Cell Locations upon *K. pneumoniae* Re-challenge**

184 Using the updated scRNA-seq data with final 15 cell types, we carried out deconvolution again
185 for each spot in the four slices (**Figures S4**). The proportions of 15 cell types across four slices
186 were summarized by a box plot (**Figures 4A**). We performed *t*-tests to compare the differences
187 between the immunized (slice A1 and A2) and the re-challenged mouse (slice A3 and A4), with
188 significant differences found for each cell type. Generally, there were more monocytes,
189 macrophages, dendritic cells, neutrophils, Th1 cells, and NK cells in the re-challenged mouse. In
190 contrast, the immunized mouse had more B cells, Th17 cells, and other T cells 1 (**Figures 4A and
191 S4**).

192 After pooling spots from the slices for the immunized (slice A1 and A2) or the re-challenged
193 mouse (slice A3 and A4), Pearson correlation coefficients between the columns of the
194 proportion matrices were calculated. Significant same-spot co-occurrence of different cell types
195 was found in the immunized and the re-challenged mouse (**Figures 4B**). For the immunized
196 mouse, Th17 cells, monocytes, and endothelial cells tended to appear in the same spots.
197 Dendritic cells, B cells, fibroblasts, and other T cells 1 were also close to one another. For the re-
198 challenged mouse, other T cells 1, B cells, Th17 cells, Th1 cells, and dendritic cells often
199 appeared in the same spots. While Th1 cells, NK cells, and myeloid cells were possibly close to
200 one another *in vivo.* These observations remained unchanged with each slice analyzed
201 separately (**Figures S5C**).

202 Localization and segmentation of airway and blood vessels are important in our analysis. We
203 defined airways and blood vessels according to the proportion of club cells and the histological
204 blood vessels (**Figures S5A and S5B**). After excluding spots within the blood vessels and spots
205 whose distances to the airways longer than 1,000 micrometers, we calculated weighted
206 distances to the airways for Th17 and Th1 cells in four slices, according to the formula (**Figures
207 4D**). Th17 cells were found closer to the airways than Th1 cells in the re-challenged mouse,
208 whereas Th1 cells were closer to the airways in the immunized mouse (**Figures 4C**). This

209   conclusion was independent of the definition of airways and blood vessels, since it remained
210   unchanged even if a set of cut-offs were used to define these structures (**Figures S5D**).

211   To find the spatial distribution patterns of immune cells, natural spline regression was
212   performed to fit the non-linear relationship between the proportions of immune cells and the
213   distances to the airways. Generally, B cells and other T cells 1 were proximal to the airways in
214   the immunized mouse, whereas neutrophils were distal to the airways in the re-challenged
215   mouse (**Figures 4Ei-iv**). The same analysis for all 15 cell types was also performed (**Figures S5Ei-**
216   **iv**). We also performed natural spline regression between the expression of genes and the
217   distances to the airways. 3,655 distance-associated genes (FDR-adjusted P-value < 0.05 in both
218   slices) were identified in the immunized mouse (**Table S1**), while 3,407 were identified in the re-
219   challenged mouse (**Table S2**). Gene Ontology (GO) enrichment analysis was performed for
220   these distance-associated genes [13–15]. 1,293 significantly (FDR-adjusted P-value < 0.05) over-
221   represented biological processes were identified for the immunized mouse, and 1,294 for the
222   re-challenged mouse. The top 50 (hierarchically sorted by fold-enrichment) over-represented
223   biological processes for the re-challenged mouse included many immune responses (*e.g.*,
224   proliferation, differentiation, activation, aggregation, adhesion, and chemotaxis of immune
225   cells). In contrast, few immune responses were over-represented in the immunized mouse
226   (**Table 1**). *Ccl20*, a top distance-associated gene, was highly expressed around the airways in the
227   re-challenged mouse, instead of the immunized mouse (**Figures 4F**). Since CCL20 is capable of
228   binding to CCR6, a chemokine receptor expressed on Th17 cells, this finding possibly explains
229   why Th17 cells were closer to the airways than Th1 cells in the re-challenged mouse.

230

Biological Differences upon *K. pneumoniae* Re-challenge

232   Upon *K. pneumoniae* re-challenge, the spatial transcriptome was tremendously changed
233   (**Figures S6A**), while little batch effects were found between the two slices from the same
234   mouse. The spots we defined as airways according to the proportion of club cells were also
235   shown in a UMAP plot (**Figures S6B and S5A**), with significant biological differences found
236   between the immunized and the re-challenged mouse.

237   Differential expression (DE) analysis was performed to compare the airways in the re-
238   challenged mouse versus the immunized mouse, with 2,071 significantly (FDR-adjusted P-value
239   < 0.05) differentially expressed genes (DEGs) identified (**Table S3**). To overcome the limitations
240   of the analysis based on different thresholds of selected DEGs, we performed Gene Set
241   Enrichment Analysis (GSEA) using the unfiltered, ranked gene list (including 16,937 genes), and
242   found 379 significantly (FDR-adjusted P-value < 0.05) enriched GO biological processes. The top
243   30 (by normalized enrichment score) up-regulated and down-regulated pathways were shown
244   by a lollipop plot (**Figures 5A**). Compared with the immunized mouse, most up-regulated
245   pathways in the airways of the re-challenged mouse were related to immune responses (*e.g.*,
246   migration and chemotaxis of immune cells, and response to stimuli), whereas most down-
247   regulated pathways were related to catabolic and metabolic processes, as well as ciliary
248   functions. AW112010, a top DE gene up-regulated in the airways upon re-challenge, was highly
249   expressed in the airways of the re-challenged mouse (**Figures 5B**). AW112010 has also been
250   reported capable of promoting the differentiation of inflammatory T cells [16]. *Cbr2*, a top DE

251 gene down-regulated in the airways upon re-challenge, was highly expressed in the airways of
252 the immunized mouse (**Figures 5C**), and may function in the metabolism of endogenous
253 carbonyl compounds [17] and alveolar epithelial cell plasticity [18].

254

255 **Spatial Transcriptomics Showing the Potential to Analyze Cell-cell Communication**

256 In order to perform cell-cell communication networks analysis, cell-type enriched spots were
257 identified according to their ranks of proportions of cell types (**Figures S6D**), with little
258 difference found between the two slices from the same mouse (**Figures S6C**). Fibroblasts,
259 endothelial cells, and erythrocytes were not included in the analysis due to the difficulty of
260 interpreting the interactions with these cells.

261 Cell-cell communication networks among the cell-type enriched spots were inferred in each
262 slice using CellChat (**Figures S6E**)[19]. To compare the differences of communication patterns
263 between the re-challenged and the immunized mouse, differential interaction strength
264 between cell-type enriched spots was shown by a heatmap (**Figures 6A**). The communication
265 among myeloid cells, Th1 cells, and Th17 cells was increased in the re-challenged mouse, also
266 revealed by our same-spot co-occurrence analysis. For the immunized mouse, dendritic cells
267 could be the major sender of communication signals, while B cells and other T cells 1 were
268 essential receivers of the signals.

269 The conserved and context-specific signaling pathways were identified and visualized by bar
270 plots (**Figures 6B**). TNF pathway was only turned on in the re-challenged mouse, while CXCL and
271 CCL pathways were increased. We also confirmed some well-known signaling pathways in the
272 cell-cell communication analysis of spatial transcriptome, outperforming the cell-cell
273 communication analysis of scRNA-seq data. The performance of this analysis could be
274 improved, as the resolution of spatial transcriptomics evolves. We found Th17 enriched spots in
275 the re-challenged mouse were the receivers of the IL6 signaling pathway (**Figures 6Ci**), given the
276 receptor gene *Il6ra* was highly expressed in Th17 enriched spots from slice A3 and slice A4
277 (**Figures 6Cii**). Leveraging scATAC-seq data, we confirmed macrophages and dendritic cells were
278 the sources of the IL6 signaling (**Figures 6Ciii**), and Th17 cells were the targets (**Figures 6Civ**).
279 For the TGF-β signaling pathway, we also found Th17 enriched spots in the re-challenged
280 mouse were its receivers (**Figures 6Di**), with the receptor genes *Tgfbr1*, *Tgfbr2*, and *Acvr1b*
281 found highly expressed in Th17 enriched spots from slice A3 and slice A4 (**Figures 6Dii**). By
282 checking peaks in genomic regions around these receptor genes, Th17 cells were confirmed to
283 have the capacity of expressing the receptors for the TGF-β signaling (**Figures 6Diii-v**). These
284 data suggested that spatial transcriptomics data can be extrapolated and utilized to perform
285 cell-cell communication networks analysis.

286

287 **DISCUSSION**

288 As the development of single-cell sequencing technologies, scRNA-seq has been gradually
289 applied to study immune cells and immune responses in mouse lungs [5–8]. In many cases, T
290 helper cells and cytotoxic T cells can be identified in these scRNA-seq studies, but not their

291   subtypes. Few studies additionally utilize scATAC-seq to measure the chromatin architecture of
292   immune cells in mouse lungs [9], which allows the identification of subtypes of T helper cells. As
293   the birth of spatial transcriptomics, this cutting-edge technology is also adopted by this field to
294   study mouse lungs infected by influenza [20].

295   To our knowledge, this is the first study to integrate spatial transcriptome with single-cell
296   transcriptome and single-cell epigenome in mouse lungs. A key advantage of our study lies in
297   the capture of intact anatomical structures of mouse lungs, providing location specific
298   information and preserving information from cells prone to damage. Integrating with single-cell
299   multi-omics profiled from matched tissue, sub-single-cell resolution were further achieved in
300   our spatial analysis of cells residing in lung, especially TRM. In our study, we first identified 12
301   lung cell types in scRNA-seq data, covering major epithelial, mesenchymal, and immune cells.
302   Using our finely annotated scRNA-seq data as a reference, we deconvoluted spatial
303   transcriptome of the four slices and inferred proportions of cell types for each spot. We
304   established an optimized deconvolution pipeline to accurately decipher specific cell-type
305   composition at sub-single-cell resolution, by checking the correlation of deconvolution results
306   using two independent references and the colocalization of canonical markers and histological
307   structures. Integrating scATAC-seq data from the same cells processed side by side with scRNA-
308   seq, we found epigenomic profiles provide more robust cell type identification, especially for
309   lineage-specific T helper cells, and further identified four subtypes of T cells. Combining all
310   three data modalities, we mapped the 15 lung cell types to histological structures for the four
311   slices at sub-single-cell resolution.

312   Our data provides further insights into dynamic changes of cell locations upon *K. pneumoniae*
313   re-challenge. We found Th17 cells were closer to the airways than Th1 cells in the re-challenged
314   mouse, whereas Th1 cells were closer to the airways in the immunized mouse without re-
315   challenge. This could be explained by the increased expression of *Ccl20* around the airways in
316   the re-challenged mouse. *Ccl20*, a top distance-associated gene identified by our spatial
317   analysis, has the ability to attract *Ccr6* expressing cells. The CCL20/CCR6 axis has been shown to
318   play crucial roles in recruiting Th17 cells in many organs as well as various disease settings [21,22].
319   We discovered different spatial distribution patterns of immune cells in the lungs of the two
320   mice, finding B cells were proximal to the airways in the immunized mouse, whereas
321   neutrophils were distal to the airways in the re-challenged mouse. We also identified thousands
322   of distance-associated genes for the two mice by natural spline regression, confirming immune
323   responses related genes were over-represented in the re-challenged mouse. These spatial
324   analyses were only possible when location specific information was captured by spatial
325   transcriptomics, highlighting the value of our study compared with conventional flow
326   cytometric and transcriptomic approaches.

327   Comparing the biological differences in the airways upon *K. pneumoniae* re-challenge, we found
328   pathways related to migration and chemotaxis of immune cells, and response to stimuli were
329   up-regulated in the re-challenged mouse, whereas pathways related to catabolic and metabolic
330   processes, and ciliary functions were up-regulated in the immunized mouse. We also
331   performed cell-cell communication analysis of spatial transcriptome, providing evidence of
332   lineage-specific T helper cells receiving designated cytokine signaling. Our study shows the
333   potential to perform DE analysis between specific cells or regions across the slices and analyze

334 cell-cell communication using spatial transcriptomics. The performance of these analyses could
335 be improved as spatial transcriptomics advances towards single-cell resolution.

336 Several limitations were recognized in our study. First, we had a small sample size for our
337 scRNA-seq and scATAC-seq data as the purpose of generating these data was to provide a
338 reference for deconvolution, instead of carrying out a census of lung cell types. Second, there is
339 no ground truth to evaluate deconvolution results, although we optimized our deconvolution
340 pipeline by assessing its robustness. This issue could be eventually resolved as the resolution of
341 the spatial transcriptomics technology improves. Third, our method to define airways and blood
342 vessels may not be fully accurate and can be further improved by both high-resolution
343 histological images and spatial transcriptome. Last, more biological replicates and experimental
344 validations may be needed to extend this study. For example, the cellular source of CCL20
345 around the airways in the re-challenged mouse could be determined by immunofluorescence.

346 In summary, we presented a comprehensive single-cell multi-omics study on immunized mouse
347 lungs and generated novel hypotheses for understanding underlying biological mechanism.
348 Recently, spatial analysis has been made compatible with Formalin-Fixed Paraffin-Embedded
349 (FFPE) tissue specimens. It is foreseeable that a massive amount of data will be generated from
350 historically preserved samples. Our spatial transcriptomics data processing pipeline provides a
351 timely solution to these analyses and contributes to advance the field of lung biology and
352 respiratory medicine.

353

354 **MATERIALS AND METHODS**

355 **Mouse models**

356 All mice used in this study were wildtype and purchased from Jackson Lab (Cat# 000664).
357 Animals were maintained in pathogen-free conditions in the animal facility at the University of
358 Pittsburgh Medical Center. All experiments were approved by the University of Pittsburgh
359 Institutional Animal Care and Use Committee.

360

361 **In vivo inflammation induction**

362 6-8 weeks old C57BL/6 mice were immunized with heat-killed *K. Pneumoniae* (ATCC-43816) as
363 previously described [3]. Briefly, mice were injected with heat-killed *K. Pneumoniae* twice (Day0
364 and Day7) or three times (Day0, Day7, and Day13) intranasally and sacrificed on Day14. Lungs
365 were removed and digested by Collagenase/DNase to obtain single-cell suspension.
366 Mononuclear cells after red blood cell lysis and filtration with a 40 µM cell strainer were
367 subjected to single cell RNA-seq (scRNA-seq) and single cell ATAC-seq (scATAC-seq) library prep
368 following the protocols by 10x Genomics using the Chromium controller (10x Genomics). To
369 yield sufficient IL-17A producing cells and reduce doublets formation, we targeted 3,000-5,000
370 cells/nuclei for recovery. Libraries were QC'ed on an Agilent TapeStation and sequenced on an
371 Illumina Novaseq.

372

Spatial Transcriptomics Experiment

We conducted ST experiment using 10X Genomics Visium platform.

Tissue harvesting: Mouse lungs were harvested, and the left lobes were inflated with 1mL mixture of 50% sterile PBS/ 50%Tissue-Tek OCT compound (SAKURA FINETEK) followed by frozen in alcohol bath on dry ice. OCT blocks were stored in -80C until further processing.

ST library prep: OCT blocks were sectioned at 10µm in thickness, 6.5mm X 6.5mm in size, attached to the Visium slides, then stained with hematoxylin and eosin following 10x Genomics Visium fresh frozen tissue processing protocol. H&E Images were taken by a fluorescence and tile scanning microscope (Olympus Fluoview 1000) then the slides underwent tissue removal and library generation per 10x Genomics demonstrated protocol.

Raw Sequencing Data Processing

The sequenced scRNA-seq library was processed and aligned to mm10 mouse reference genome using Cell Ranger software (version 3.1.0) from 10x Genomics, with unique molecular identifier (UMI) counts summarized for each barcode. To distinguish cells from the background, cell calling was performed on the full raw UMI count matrix, with the filtered UMI count matrix generated (31,053 genes x 3,337 cells).

The sequenced scATAC-seq library was processed and aligned to mm10 mouse reference genome using Cell Ranger ATAC software (version 1.1.0) from 10x Genomics, with fragments and peak counts summarized for each barcode. To distinguish cells from the background, cell calling was performed on the full raw peak count matrix, with the filtered peak count matrix generated (84,317 peaks x 4,908 cells).

Each sequenced spatial transcriptomics library was processed and aligned to mm10 mouse reference genome using Space Ranger software (version 1.2.2) from 10x Genomics, with UMI counts summarized for each spot. To distinguish tissue overlaying spots from the background, tissue overlaying spots were detected according to the images. And only barcodes associated with these tissue overlaying spots were retained, with the filtered UMI count matrices generated. We also manually excluded spots not covered by tissue but not detected by Space Ranger and further filter the UMI count matrices (slice A1: 32,285 genes x 3,689 spots; slice A2: 32,285 genes x 2,840 spots; slice A3: 32,285 genes x 3,950 spots; slice A2: 32,285 genes x 3,765 spots).

scRNA-seq Data Analysis

After imported into R, the filtered UMI count matrix was analyzed using the R package Seurat (version 4.0.1)[23]. The percentage of mitochondrial genes per cell was calculated for further check of the quality of cells. Regularized negative binomial regression (SCTransform)[24] was used to normalize UMI count data, with the removal of confounding effects from mitochondrial mapping percentage. To improve the speed of the normalization, glmGamPoi [25] was invoked in the procedure. 3,000 highly variable genes were identified and used in principal component

11

412 analysis to reduce dimensionality. We determined to use the first 50 principal components in
413 clustering analysis according to the elbow plot. Uniform Manifold Approximation and
414 Projection (UMAP) dimensionality reduction [26] was performed with the first 50 principal
415 components as input to visualize cells. Using the Shared Nearest-neighbor (SNN) graph as input,
416 cells were then clustered using the original Louvain algorithm with resolution = 0.2.

417 Cluster 1 was marked as low-quality cells and excluded from downstream analysis (2,834 cells
418 retained) because its median percentage of mitochondrial genes was 87.7%, whereas those of
419 all other clusters were lower than 10.5%. Markers for each cluster were identified using a
420 Wilcoxon Rank Sum test with only.pos = TRUE, min.pct = 0.25, and logfc.threshold = 0.25.
421 According to the UMAP plot, cluster 6 was found to consist of several sporadic sub-clusters. We
422 isolated cells within cluster 6 and repeated the procedures from normalization to
423 dimensionality reduction. These cells were then clustered using the original Louvain algorithm
424 with resolution = 0.6. Markers were also identified as described above.

425

426 **Characterization of 12 Lung Cell Types in scRNA-seq Data**

427 Fine cluster annotation was performed for the retained 12 clusters in scRNA-seq data according
428 to canonical markers: alveolar epithelial cells (*Sftpa1*, *Sftpb*, *Sftpc*, and *Sftpd*), club cells
429 (*Scgb1a1*, *Muc5b*, *Scgb3a1*, and *Scgb3a2*), fibroblasts (*Col3a1*, *Col1a2*, *Col1a1*, and *Mfap4*),
430 endothelial cells (*Cdh5*, *Mcam*, *Vcam1*, and *Pecam1*), monocytes (*Cd14* and *Itgam*),
431 macrophages (*Itgax*, *Cd68*, *Mrc1*, and *Marco*), dendritic cells (*Aif1*, *H2-DMb1*, *H2-Eb1*, and *H2-*
432 *Aa*), neutrophils (*Gsr*, *Pglyrp1*, *S100a8*, and *Ly6g*), B cells (*Ms4a1*, *Cd79a*, *Igkc*, and *Cd19*), T cells
433 (*Cd3d*, *Cd3e*, and *Cd3g*), NK cells (*Ncr1* and *Nkg7*), and erythrocytes (*Hbb-bt* and *Hba-a2*).

434

435 **Spatial Transcriptomics Data Analysis**

436 After imported into R, the filtered UMI count matrix was analyzed using the R package Seurat
437 (version 4.0.1)[23]. Regularized negative binomial regression (SCTransform)[24] was used to
438 normalize UMI count matrices, and glmGamPoi [25] was invoked in the procedure to improve the
439 speed of the normalization. Four matrices from the four slices were merged to analyze them
440 together. 3,000 highly variable genes were identified in each matrix, and the union set of them
441 was set as highly variable genes for the merged data and used in principal component analysis
442 to reduce dimensionality. We determined to use the first 30 principal components in clustering
443 analysis according to the elbow plot. Uniform Manifold Approximation and Projection (UMAP)
444 dimensionality reduction [26] was performed with the first 30 principal components as input to
445 visualize spots.

446

447 **Deconvolution of Spatial Transcriptome**

448 Deconvolution for each spot in the four slices was performed using the robust cell-type
449 decomposition (RCTD) method [10]. Before running the R package RCTD (version 1.2.0), scRNA-
450 seq data used as a reference were processed, with gene expression matrix, the annotation for
451 each cell, and the total UMI count for each cell extracted and saved in the RDS object. Spatial

452   transcriptome for the four slices was also processed, with spot location matrices, gene
453   expression matrices, and the total UMI count for each spot extracted and saved in the RDS
454   object.

455   RCTD objects were created for each slice from the processed RDS objects, with max_cores = 24,
456   test_mode = F, and CELL_MIN_INSTANCE = 6. RCTD pipeline was run on the RCTD objects, with
457   doublet_mode = full. The deconvolution results were matrices of cell type weights for each
458   spot. The cell type weights were normalized to make the sum of cell type weights in each spot
459   equal to 1. Proportion matrices, whose rows indicate spots and columns indicate cell types,
460   were then created and stored in the analyzed spatial transcriptome for loading.

461   In total, three scRNA-seq references were used in deconvolution, which were in-house scRNA-
462   seq data with 12 cell types, in-house scRNA-seq data with 15 cell types (including four subtypes
463   of T cells), and public scRNA-seq data with 20 cell types.

464   Pearson correlation coefficients between the columns (indicating different cell types) of the
465   two proportion matrices deconvoluted using in-house (12 cell types) and public (20 cell types)
466   references were calculated in each slice. Correlation r matrices were hierarchically clustered
467   and visualized in heatmaps.

468

469   **scATAC-seq Data Analysis**

470   After imported into R, the filtered peak count matrix was analyzed using the R package Signac
471   (version 1.1.1)[27]. Gene annotations were extracted from Ensembl release 79 of the mm10
472   mouse reference genome. Nucleosome signal score and Transcriptional Start Site (TSS)
473   enrichment score per cell were calculated for further check of the quality of cells. Cells whose
474   fraction of fragments in peaks > 15, ratio of reads in genomic blacklist regions < 0.05,
475   nucleosome signal score < 4, and TSS enrichment score > 2 were retained for downstream
476   analysis. Latent Semantic Indexing (LSI)[28], which is combined steps of Term Frequency–Inverse
477   Document Frequency (TF-IDF) followed by Singular Value Decomposition (SVD), was used to
478   normalize and reduce the dimensionality of peak count data, with all the peaks selected as
479   variable features. We found the first LSI component was highly correlated with sequencing
480   depth and determined to use the second to the fortieth LSI components in non-linear
481   dimensionality reduction. Uniform Manifold Approximation and Projection (UMAP)
482   dimensionality reduction [26] was performed with the second to the fortieth LSI components as
483   input to visualize cells.

484

485   **Label Transferring from scRNA-seq Data to scATAC-seq Data**

486   A gene activity matrix was created in scATAC-seq data by counting the number of fragments
487   mapping to promoter or gene body regions of all protein-coding genes for each cell.
488   Regularized negative binomial regression (SCTransform)[24] was used to normalize the gene
489   activity matrix. Transfer anchors were identified by canonical correlation analysis between the
490   normalized gene activity matrix in scATAC-seq data and the normalized gene expression matrix
491   in scRNA-seq data [29]. Annotations were then transferred from scRNA-seq to scATAC-seq data

13

492   with the second to the fortieth LSI components in scATAC-seq data used for weighting anchors.
493   Canonical markers were checked in the gene activity matrix for each predicted cell type. The
494   proportions of 12 cell types were also compared between scRNA-seq and scATAC-seq data.

495

496   **Characterization of Four Subtypes of T cells in scATAC-seq Data**

497   We isolated predicted T cells in scATAC-seq data and repeated the procedures from
498   normalization to non-linear dimensionality reduction. Using the Shared Nearest-neighbor (SNN)
499   graph leveraging the second to the fortieth LSI components as input, T cells were then clustered
500   into four subtypes using the Smart Local Movement (SLM) algorithm with resolution = 0.2.

501   Markers for each subtype were identified in the gene activity matrix using a Wilcoxon Rank Sum
502   test with only.pos = TRUE, min.pct = 0.25, and logfc.threshold = 0.25. Differential accessible
503   analysis was performed between the possible Th17 and Th1 cells using logistic regression, with
504   fraction of fragments in peaks set as latent variable and min.pct = 0.1. Peaks in genomic regions
505   around (including all the differentially accessible regions or 10,000 bps apart from the gene
506   bodies) canonical markers for Th17 and Th1 cells were visualized. Motif footprinting analysis
507   was also performed to provide supportive evidence. Taken together, four subtypes of T cells
508   were annotated according to canonical markers: Th17 cells (*Il17a*, *Rorc*, and *Rora*), Th1 cells
509   (*Ifng* and *Tbx21*), other T cells 1, and other T cells 2.

510

511   **Label Transferring from scATAC-seq Data to scRNA-seq Data**

512   We isolated annotated T cells in scRNA-seq data and repeated the procedures from
513   normalization to non-linear dimensionality reduction. Transfer anchors were identified by
514   canonical correlation analysis between the normalized gene activity matrix of T cells in scATAC-
515   seq data and the normalized gene expression matrix of T cells in scRNA-seq data [29]. Annotations
516   were then transferred from T cells in scATAC-seq to T cells in scRNA-seq data with the first 50
517   principal components in T cells from scRNA-seq data used for weighting anchors.

518

519   **Characterization of Four Subtypes of T cells in scATAC-seq Data**

520   Canonical markers were checked for the predicted subtypes of T cells in scRNA-seq data (*Rorc*,
521   *Rora*, *Ccr6*, and *Ccr4* for Th17 cells; *Ifng*, *Tbx21*, *Cxcr3*, *Ccr5*, and *Il12rb2* for Th1 cells). Markers
522   for each predicted subtype of T cells were identified using a Wilcoxon Rank Sum test with
523   only.pos = TRUE, min.pct = 0.25, and logfc.threshold = 0.25.

524   To detect active transcription factor (TF) modules, the R package SCENIC (version 1.2.4)[11,12] was
525   used to analyze the annotated scRNA-seq data, including 15 cell types. We downloaded the
526   RcisTarget database containing transcription factor motif scores for gene promoter and around
527   Transcription Start Site (TSS) for mm10 mouse reference genome from
528   (https://resources.aertslab.org/cistarget/databases/mus_musculus/mm10/refseq_r80/mc9nr/g
529   ene_based/). The gene expression matrix was filtered according to default settings, and 9392
530   genes were retained and used to compute a gene-gene correlation matrix. Co-expression

14

531   module detection was performed using the GENIE3 algorithm based on random forest.
532   Transcription factor network analysis was performed to detect co-expression modules enriched
533   for target genes of each candidate TF from the RcisTarget database, with regulons identified.
534   The R package AUCell (version 1.12.0) was used to compute a score for each TF module in each
535   cell.

536   Regulon Specificity Score (RSS) was calculated for regulons in each cell type according to Area
537   Under the Curve (AUC) of regulons. Cell-type specific regulators were then identified, and those
538   for the four subtypes of T cells were visualized to provide supportive evidence (Rora_extended
539   (15g) for Th17 cells).

540

541   **Same-spot Co-occurrence Analysis**

542   Pearson correlation coefficients between the columns (indicating different cell types) of the
543   proportion matrices were calculated in each slice to broadly assess spatial cell type co-
544   occurrence in the same spot. Spots from the slices for the immunized (slice A1 and A2) or the
545   re-challenged mouse (slice A3 and A4) were pooled together, respectively, and Pearson
546   correlation coefficients were also calculated for them. Correlation r matrices were hierarchically
547   clustered and visualized in heatmaps.

548

549   **Defining Airways and Blood Vessels in Four Slices**

550   Airways were defined according to the proportion of club cells in four slices. We manually set
551   the thresholds in each slice to match the selected spots with the histological airways. Spots
552   whose proportion of club cells higher than the thresholds were defined as airways (20% for
553   slice A1, 20% for slice A2, 10% for slice A3, and 10% for slice A4).

554   Blood vessels were defined according to the histological blood vessels. We created a training
555   set using manual annotation of histological structures in the image of slice A1 and trained a
556   random trees pixel classifier using QuPath (version 0.2.3)[30] with downsample = 16. The
557   probability of blood vessels was predicted for each pixel in the image of four slices using the
558   trained classifier. If the probability of blood vessels in the spot corresponding pixel was higher
559   than 0.5, the spot would be defined as blood vessels.

560   Alternatively, the proportion of club cells and the expression of *Mgp* were used to define
561   airways and blood vessels with different cut-offs, from 90th quantile to 95th quantile. Spots
562   whose proportion or expression higher than the selected quantile were defined as airways or
563   blood vessels.

564

565   **Spatial Transcriptomics Spot Distance-based Analyses**

566   For all distance-based analyses, spots defined as blood vessels and spots whose distances to the
567   airways longer than 1,000 micrometers were excluded. Because cells within the blood vessels
568   were different from those in the lung parenchyma, and the analyses including cells extremely
569   distal to the airways were not stable, given the capture area of each slice was only 6.5 x 6.5

15

570 mm^2. Only a few spots' distances to the airways were longer than 1,000 micrometers (5.89% for
571 slice A1, 5.75% for slice A2, 15.00% for slice A3, and 11.43% for slice A4).

572 Weighted distances to the airways for Th17 and Th1 cells were calculated in four slices
573 according to the formula, allowing for each spot's distance to the nearest airway and the
574 proportions of Th17 and Th1 cells in each spot.

575 To find the spatial distribution patterns of cell types, natural spline regression (with three
576 degrees of freedom) was performed to fit the non-linear relationship between the proportions
577 of cell types and the distances to the airways.

578 To identify distance-associated genes, we constructed natural splines (with three degrees of
579 freedom) for the distances to the airways in each slice and created design matrices. We then
580 fitted linear models for each gene in normalized gene expression matrices, with the R package
581 limma (version 3.46.0)[31] invoked to speed up the procedure. P-values for each spline in each
582 slice were corrected for multiple testing using Benjamini-Hochberg correction. Genes whose
583 FDR-adjusted P-values < 0.05 for at least one spline were considered significant in that slice.
584 And genes significant in both slices for the immunized (slice A1 and A2) or the re-challenged
585 mouse (slice A3 and A4) were defined as distance-associated genes.

586

Gene Ontology (GO) Enrichment Analysis

588 Gene ontology enrichment analysis was performed on the distance-associated genes identified
589 in the immunized and re-challenged mouse using PANTHER Classification System (version
590 16.0)[13–15]. GO biological process complete was used as annotation dataset, and the analysis was
591 performed using Fisher's exact test, with FDR adjustment for multiple testing. Over-represented
592 pathways were then hierarchically clustered.

593

Differential Expression Analysis

595 Differential expression analysis was performed to compare the airways in the re-challenged
596 mouse versus the immunized mouse using the R package MAST (version 1.16.0)[32]. MAST
597 procedure was invoked in Seurat FindMarkers function (test.use = MAST) with logfc.threshold =
598 0, min.pct = 0, in order to obtain an unfiltered gene list. The gene list was then ranked by log2-
599 fold change (L2FC). Genes whose L2FC equal to 0 were excluded due to their ranks were not
600 available.

601

Gene Set Enrichment Analysis (GSEA)

603 To overcome the limitations of the analysis based on manually selected DEGs, Gene Set
604 Enrichment Analysis (GSEA)[33] was performed on the unfiltered, ranked gene list (including
605 16,937 genes) using the R package clusterProfiler (version 3.18.1)[34]. The parameters for the
606 clusterProfiler gseGO function were set as ont = BP, keyType = SYMBOL, nPerm = 10,000,
607 minGSSize = 3, maxGSSize = 800, pvalueCutoff = 0.05, OrgDb = org.Mm.eg.db, pAdjustMethod =

608 fdr. The top 30 (by normalized enrichment score) up-regulated and down-regulated pathways
609 were then visualized by a lollipop plot.

610

Defining Cell-type Enriched Spot in Spatial Transcriptomics Data

612 The expression profile of each spot in spatial transcriptomics is a mixture of a few cells, and it is
613 irrational to annotate a spot with a cell type directly. To perform cell-cell communication
614 analysis of spatial transcriptome, we annotated spots as cell-type enriched spots according to
615 their proportions of cell types.

616 The mean proportions of cell types were available for each slice, as well as the number of spots
617 for each slice. The mean proportions could be interpreted as the expected proportions of cell
618 types in each spot. The product of the mean proportions and the number of spots could be
619 interpreted as how many spots could represent each cell type on average. The spots with the
620 highest proportions were selected according to the product and defined as cell-type enriched
621 spots. Spots defined as cell-type enriched spots for multiple cell types were then excluded.

622 For example, the mean proportion of club cells was 6.39% in slice A1, and the number of spots
623 was 3,689 in slice A1. Thus, the top 236 spots with the highest proportion of club cells were
624 defined as club cell enriched spots. A spot would be excluded if the spot was defined as a club
625 cell enriched spot and a Th17 enriched spot simultaneously.

626

Cell-cell Communication Analysis of Spatial Transcriptome

628 Cell-cell communication analysis was performed using CellChat [19]. Before running the R package
629 CellChat (version 1.1.0), spatial transcriptome for the four slices was processed, with
630 normalized gene expression matrices, the annotations for cell-type enriched spots extracted
631 and saved in the CellChat object. The processed data from both slices for the immunized (slice
632 A1 and A2) or the re-challenged mouse were also pooled together and saved in the CellChat
633 objects to compare the differences between the two mice.

634 A manually curated database of literature-supported ligand-receptor interactions in mouse was
635 loaded for the analysis. And all ligand-receptor interactions, including paracrine/autocrine
636 signaling interactions, extracellular matrix (ECM)-receptor interactions, and cell-cell contact
637 interactions, were included in the analysis.

638 Over-expressed ligands or receptors were first identified for each type of cell-type enriched
639 spots in each CellChat object. Over-expressed ligand-receptor interactions were also identified
640 if either ligand or receptor was over-expressed. Then, we computed the communication
641 probability and inferred cellular communication network according to default settings. The
642 communication probability at signaling pathway level was computed by summarizing the
643 communication probabilities of all ligands-receptors interactions associated with each signaling
644 pathway. The aggregated cell-cell communication network was calculated by counting the
645 number of links or summarizing the communication probability and visualized in four slices
646 using heatmaps.

647   To figure out in which type of cell-type enriched spots interactions significantly changed in the
648   re-challenged mouse versus the immunized mouse, differential interaction strength was
649   identified and visualized using a heatmap. The conserved and context-specific signaling
650   pathways were identified by comparing the information flow for each signaling pathway, which
651   was defined by the total weights in the network. Selected signaling pathways were visualized by
652   circle plots using CellChat netVisual_aggregate function.

653

660

661   **Author contributions**

662   Z.X., W.C., and K.C. conceived the project and designed the experiments. L.F. and F.W.
663   performed scRNA-seq, scATAC-seq, and ST experiments. Z.X. and X.W. performed scRNA-seq,
664   scATAC-seq, and ST data analysis. J.W. provided guidance for deconvolution of ST data. Z.X.,
665   W.C., and K.C. wrote the manuscript with input from all authors. W.C. and K.C. supervised the
666   work.

667

668   **Compliance with ethics guidelines**

669   None of authors have any conflict of interest to report. All animal protocols and procedures
670   were reviewed and approved by the University of Pittsburgh Institutional Animal Care and Use
671   Committee.

672

673   **Data and materials availability**

674   Raw and processed data of scRNA-seq, scATAC-seq and ST will be deposited to Gene Expression
675   Omnibus (GEO) upon acceptance of the paper. Code and scripts necessary to repeat analyses in
676   this manuscript are available upon request.

677

REFERENCES

1. Masopust, D., Vezys, V., Marzo, A. L. & Lefrançois, L. Preferential localization of effector memory cells in nonlymphoid tissue. *Science* **291,** 2413–2417 (2001).

2. Turner, D. L. & Farber, D. L. Mucosal resident memory CD4 T cells in protection and immunopathology. *Front. Immunol.* **5,** 331 (2014).

3. Chen, K. *et al.* Th17 cells mediate clade-specific, serotype-independent mucosal immunity. *Immunity* **35,** 997–1009 (2011).

4. Szabo, P. A., Miron, M. & Farber, D. L. Location, location, location: Tissue resident memory T cells in mice and humans. *Sci. Immunol.* **4,** (2019).

5. Cohen, M. *et al.* Lung Single-Cell Signaling Interaction Map Reveals Basophil Role in Macrophage Imprinting. *Cell* **175,** 1031–1044.e18 (2018).

6. Ballesteros, I. *et al.* Co-option of Neutrophil Fates by Tissue Environments. *Cell* **183,** 1282–1297.e18 (2020).

7. Kurihara, C. *et al.* Crosstalk between nonclassical monocytes and alveolar macrophages mediates transplant ischemia-reperfusion injury through classical monocyte recruitment. *JCI Insight* **6,** (2021).

8. Wang, L. *et al.* Single-cell transcriptomic analysis reveals the immune landscape of lung in steroid-resistant asthma exacerbation. *Proc. Natl. Acad. Sci. USA* **118,** (2021).

9. Delacher, M. *et al.* Single-cell chromatin accessibility landscape identifies tissue repair program in human regulatory T cells. *Immunity* **54,** 702–720.e17 (2021).

10. Cable, D. M. *et al.* Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* (2021). doi:10.1038/s41587-021-00830-w

11. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14,** 1083–1086 (2017).

12. Van de Sande, B. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15,** 2247–2276 (2020).

13. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).

14. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* **49,** D325–D334 (2021).

15. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47,** D419–D426 (2019).

16. Yang, X., Bam, M., Becker, W., Nagarkatti, P. S. & Nagarkatti, M. Long Noncoding RNA AW112010 Promotes the Differentiation of Inflammatory T Cells by Suppressing IL-10 Expression through Histone Demethylation. *J. Immunol.* **205,** 987–993 (2020).

17. Nakanishi, M., Deyashiki, Y., Ohshima, K. & Hara, A. Cloning, expression and tissue

715 distribution of mouse tetrameric carbonyl reductase. Identity with an adipocyte 27-kDa
716 protein. *Eur. J. Biochem.* **228,** 381–387 (1995).

717 18. Mutze, K., Vierkotten, S., Milosevic, J., Eickelberg, O. & Königshoff, M. Enolase 1 (ENO1)
718 and protein disulfide-isomerase associated 3 (PDIA3) regulate Wnt/β-catenin-driven trans-
719 differentiation of murine alveolar epithelial cells. *Dis. Model. Mech.* **8,** 877–890 (2015).

720 19. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.*
721 **12,** 1088 (2021).

722 20. Boyd, D. F. *et al.* Exuberant fibroblast activity compromises lung function via ADAMTS4.
723 *Nature* **587,** 466–471 (2020).

724 21. Wang, C., Kang, S. G., Lee, J., Sun, Z. & Kim, C. H. The roles of CCR6 in migration of Th17
725 cells and regulation of effector T-cell balance in the gut. *Mucosal Immunol.* **2,** 173–183
726 (2009).

727 22. Turner, J.-E. *et al.* CCR6 recruits regulatory T cells and Th17 cells to the kidney in
728 glomerulonephritis. *J. Am. Soc. Nephrol.* **21,** 974–985 (2010).

729 23. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* (2021).
730 doi:10.1016/j.cell.2021.04.048

731 24. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq
732 data using regularized negative binomial regression. *Genome Biol.* **20,** 296 (2019).

733 25. Ahlmann-Eltze, C. & Huber, W. glmGamPoi: fitting Gamma-Poisson generalized linear
734 models on single cell count data. *Bioinformatics* **36,** 5701–5702 (2021).

735 26. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection
736 for Dimension Reduction. *arXiv* (2018).

737 27. Stuart, T., Srivastava, A., Lareau, C. & Satija, R. Multimodal single-cell chromatin analysis
738 with Signac. *BioRxiv* (2020). doi:10.1101/2020.11.09.373613

739 28. Cusanovich, D. A. *et al.* Multiplex single cell profiling of chromatin accessibility by
740 combinatorial cellular indexing. *Science* **348,** 910–914 (2015).

741 29. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177,** 1888–1902.e21
742 (2019).

743 30. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci.*
744 *Rep.* **7,** 16878 (2017).

745 31. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and
746 microarray studies. *Nucleic Acids Res.* **43,** e47 (2015).

747 32. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes
748 and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16,** 278
749 (2015).

750 33. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for
751 interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102,** 15545–15550

752   (2005).

753   34. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing
754   biological themes among gene clusters. *OMICS* **16,** 284–287 (2012).
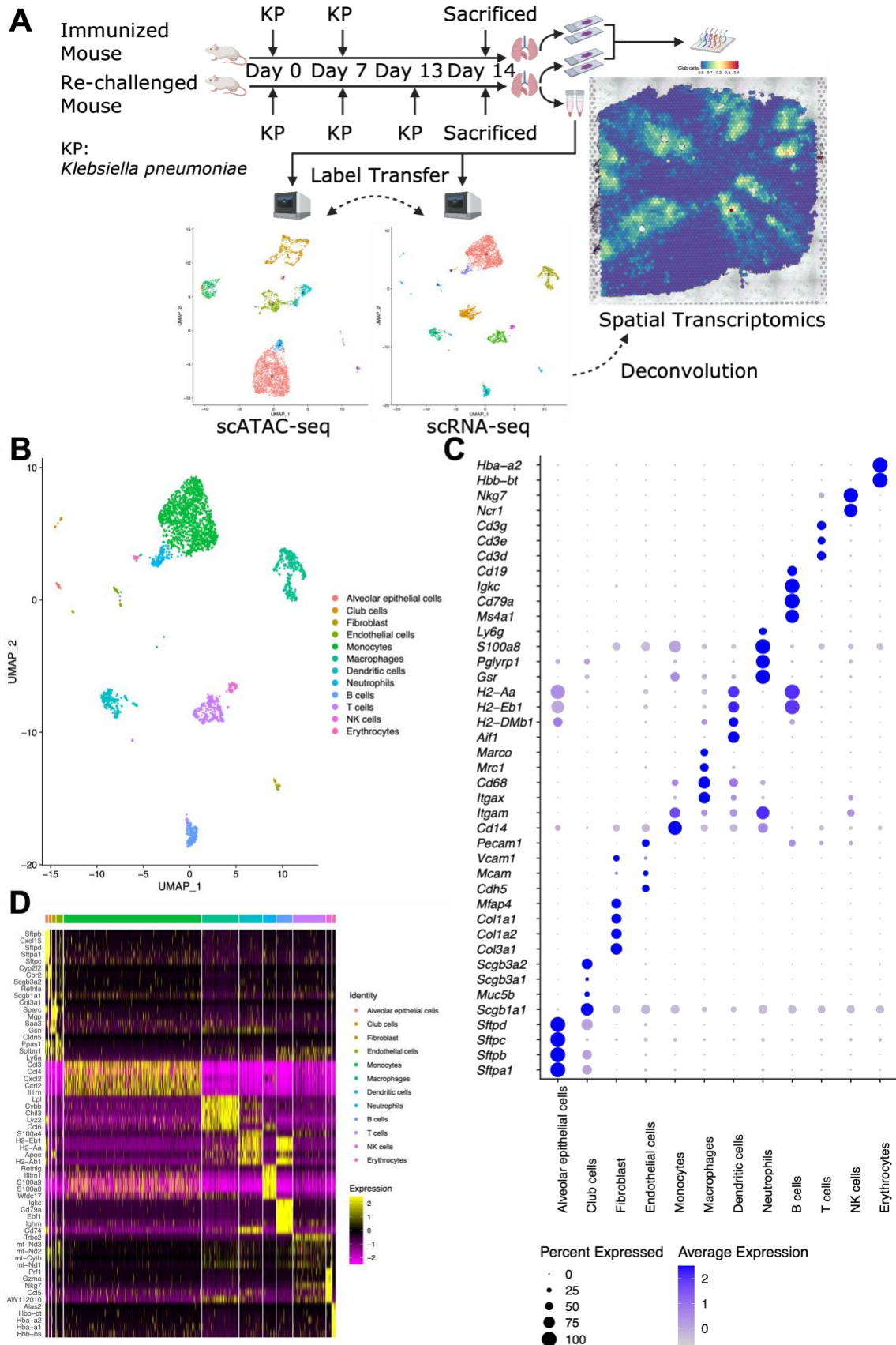
Figure 1. Generation of multi-omics datasets of mice lungs after immunization

(A) Overview of study design. scRNA-seq data work as a bridge to link scATAC-seq and spatial transcriptomics data.

(B) UMAP plot of 12 lung cell types identified in scRNA-seq data, with manual annotation according to canonical markers.

(C) Dot plot showing selected canonical markers for each cell type.

(D) Heatmap showing top five (by log2-Fold Change) markers for each cell type.

Figure 2. Validation of the robustness of deconvolution method for spatial transcriptomics
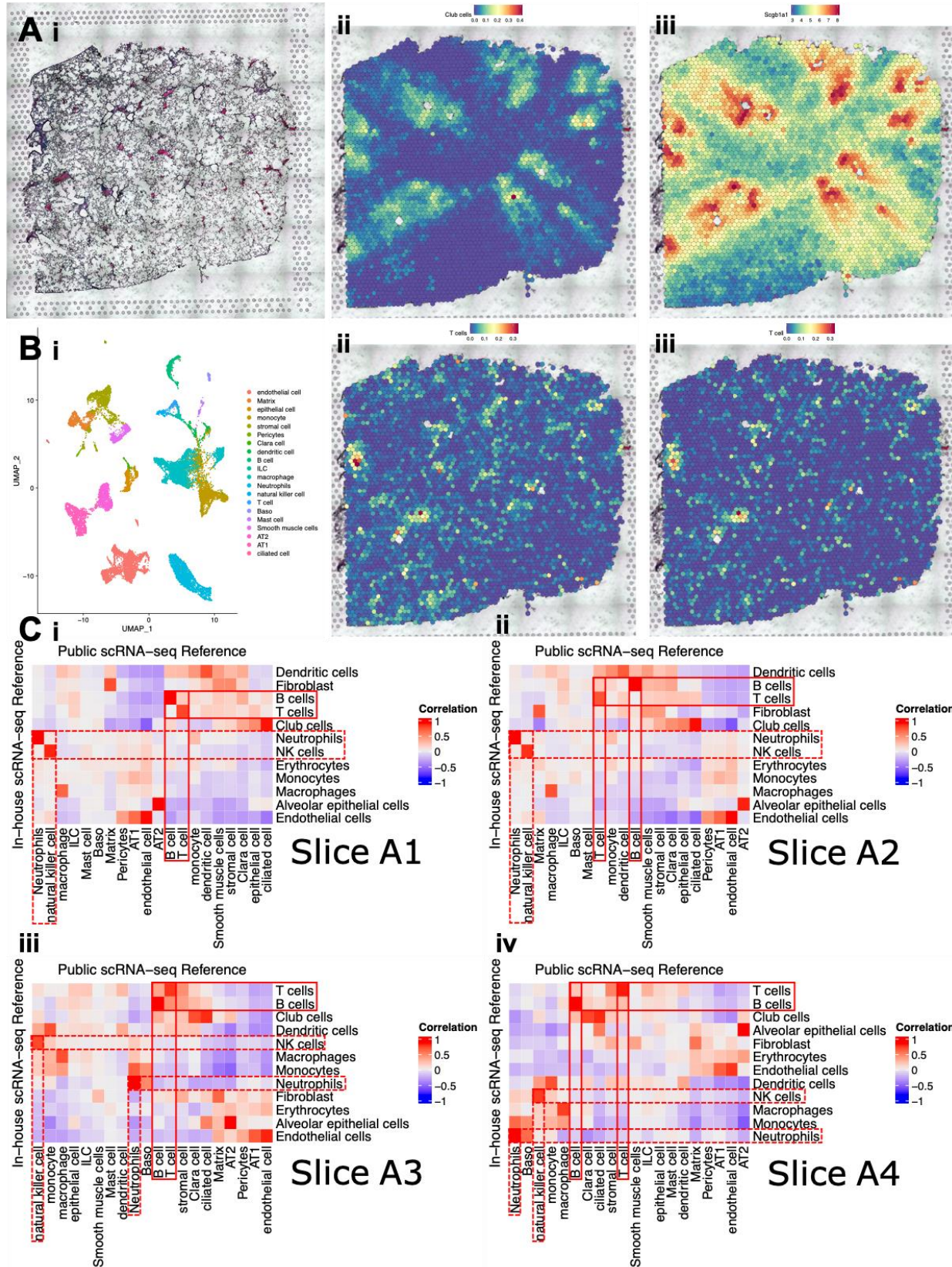
(A) Proportion of club cells colocalizing with the expression of *Scgb1a1* and the histological airways. (i) Histology of slice A3 showing the location of the airways. (ii) Proportion of club cells across slice A3, deconvoluted using in-house scRNA-seq data. (iii) Expression of *Scgb1a1*, a canonical marker for club cells, across slice A3.

(B) Proportions of T cells deconvoluted using the two independent scRNA-seq references were quite similar. (i) UMAP plot of 20 lung cell types identified in Cohen *et al.*'s public scRNA-seq data (GSE119228). (ii) Proportion of T cells across slice A3, deconvoluted using in-house scRNA-seq data. (iii) Proportion of T cells across slice A3, deconvoluted using Cohen *et al*.'s public scRNA-seq data.

(C)(i-iv) Correlation heatmap visualizing the proportions of cell types deconvoluted using in-house (in rows, 12 types) and public (in columns, 20 types) scRNA-seq data were highly correlated in slice A1-A4. Pearson's r values were indicated by the color bars. Red boxes with solid lines highlighted selected adaptive immune cells, T cells and B cells. Red boxes with dashed lines highlighted selected innate immune cells, neutrophils and NK cells.
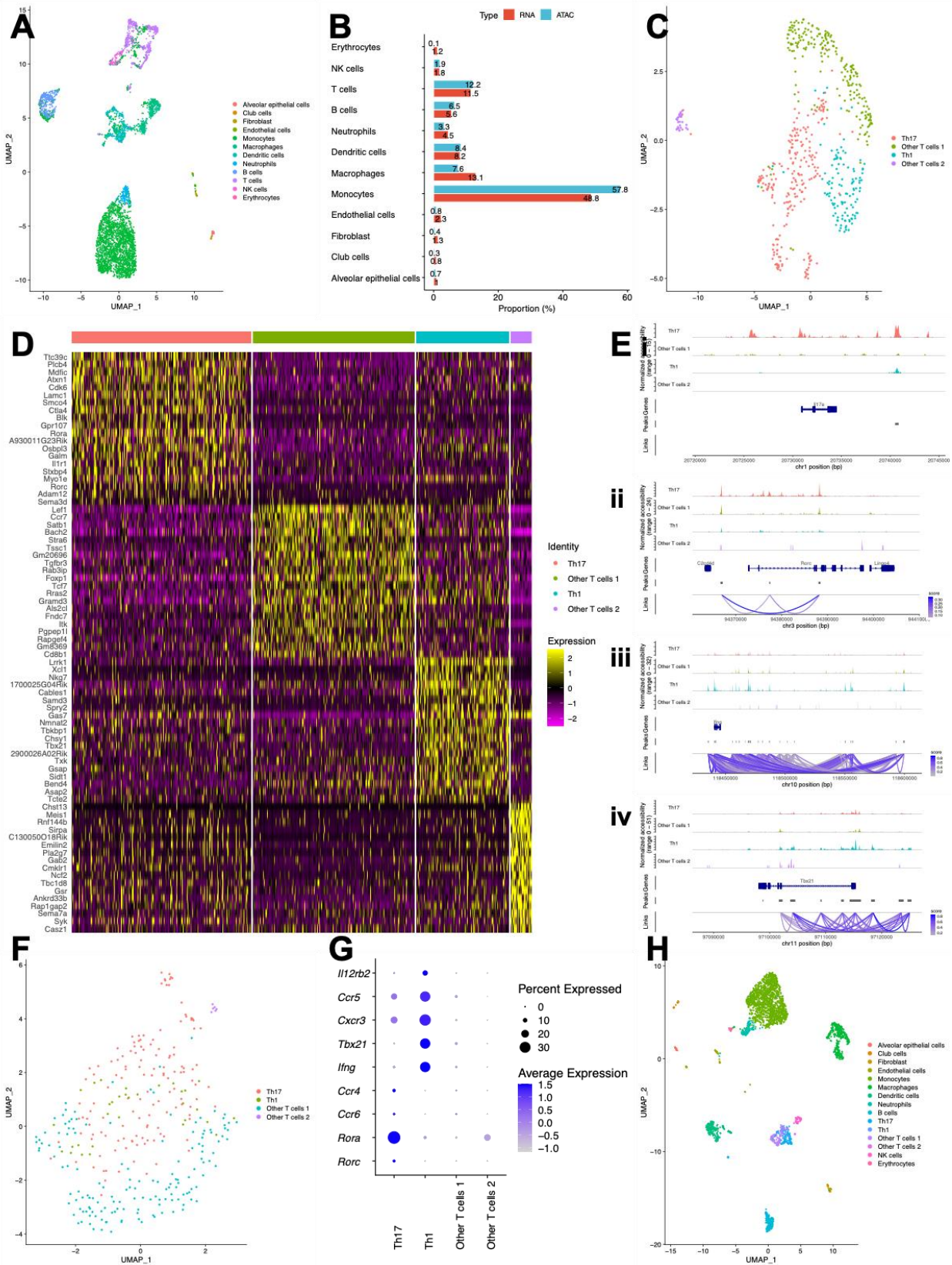
Figure 3. Identification of subtypes of T cells by integrating scRNA-seq and scATAC-seq data

(A) UMAP plot of 12 lung cell types in scATAC-seq data identified via label transfer.

(B) Bar plot showing proportions of 12 cell types in scRNA-seq and scATAC-seq data were quite similar.

(C) UMAP plot of four subtypes of T cells in scATAC-seq data.

(D) Heatmap showing top 20 (by log2-Fold Change) markers (gene activity scores calculated using peaks) for each subtype of T cells in scATAC-seq data.

(E)(i-iv) Peaks in genomic regions around *Il17a*, *Rorc*, *Ifng*, and *Tbx21*, canonical markers for Th17 and Th1 cells.

(F) UMAP plot of four subtypes of T cells in scRNA-seq data identified via label transfer.

(H) Dot plot showing selected canonical markers for Th17 and Th1 cells in scRNA-seq data.

(G) UMAP plot of 15 lung cell types (including four subtypes of T cells) in scRNA-seq data.

$$Weighted\ Distance = \frac{\sum_{i=1}^{n} Proportion_i \times Distance_{min,i}}{\sum_{i=1}^{n} Proportion_i}$$
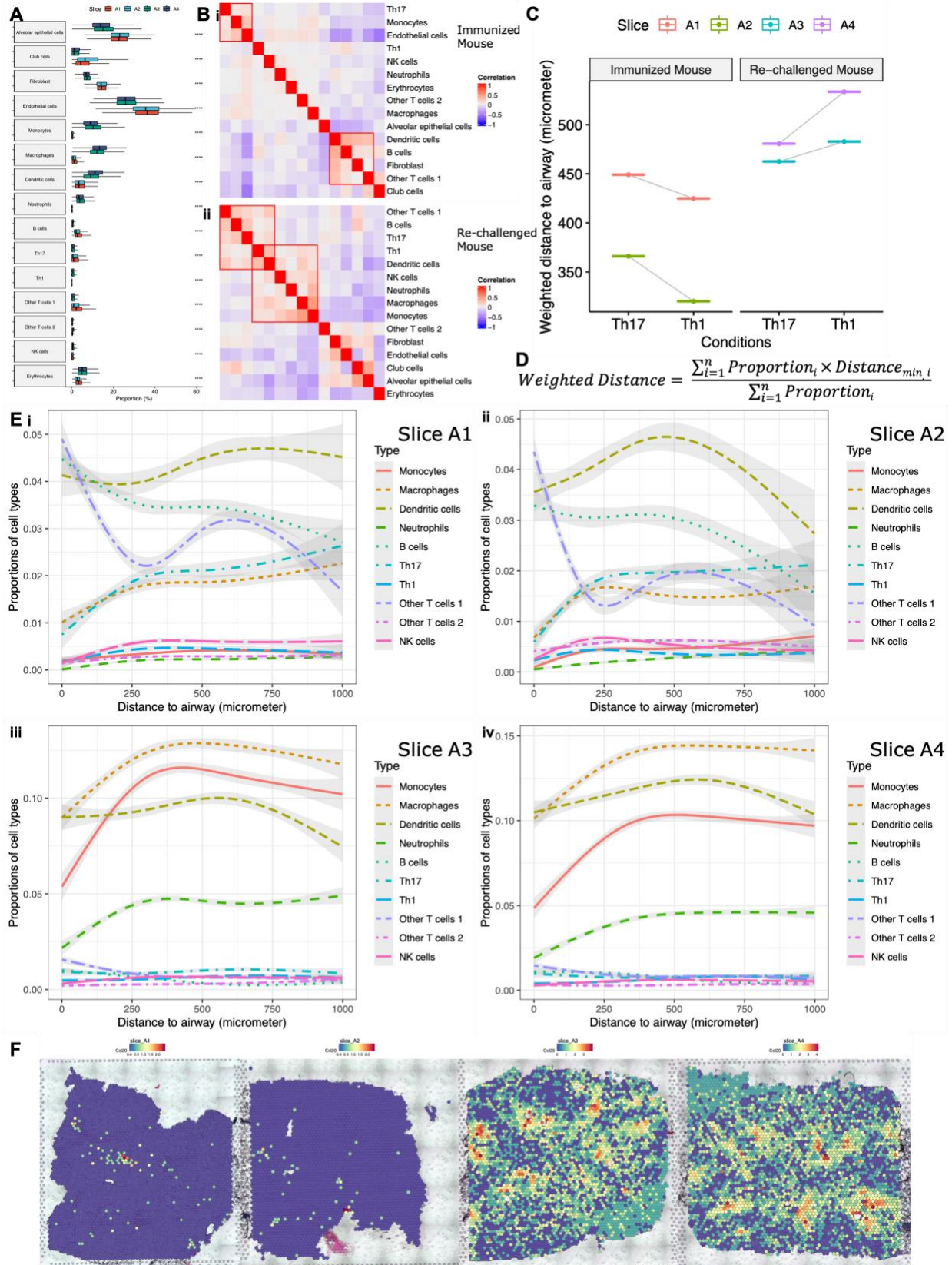
Figure 4. Spatial analyses of mice lungs after immunization

(A) Box plot showing proportions of 15 cell types across four slices were different. To compare the differences between the immunized (slice A1 and A2) and the re-challenged mouse (slice A3 and A4), t-tests were performed for each cell type. ****: p-value < 1×10^{-4}.

(B)(i-ii) Correlation heatmap visualizing the same-spot co-occurrence of 15 cell types in the immunized and the re-challenged mouse. ST spots from the slices for the immunized (slice A1 and A2) or the re-challenged mouse (slice A3 and A4) were pooled together, respectively. Red boxes highlighted cell types tending to appear in the same spots, that is to say, possibly close to one another *in vivo*.

(C) Weighted distances to the airways for Th17 and Th1 cells in four slices, after excluding spots within the blood vessels and spots whose distances to the airways longer than 1,000 micrometers.

(D) Formula defining weighted distance, allowing for each spot's distance to the nearest airway and the proportions of Th17 and Th1 cells in each spot.

(E)(i-iv) Proportions of immune cells over distance to the airways showing the spatial distribution of immune cells in four slices, after excluding spots within the blood vessels and spots whose distances to the airways longer than 1,000 micrometers. The curves were obtained from natural spline (with three degrees of freedom) regression.

(F) Expression of *Ccl20*, a top distance-associated gene, across four slices.
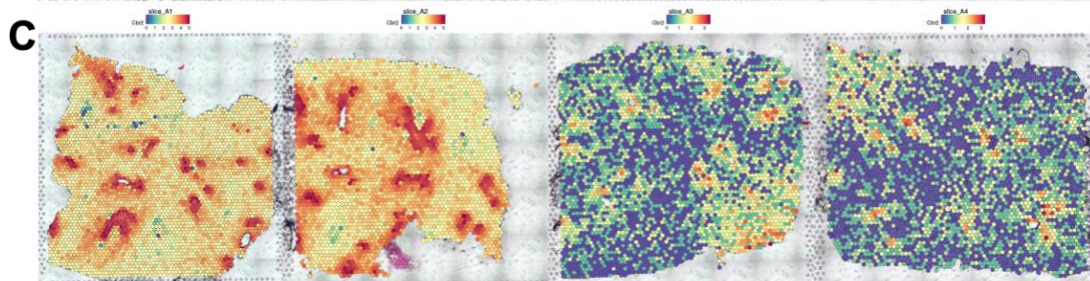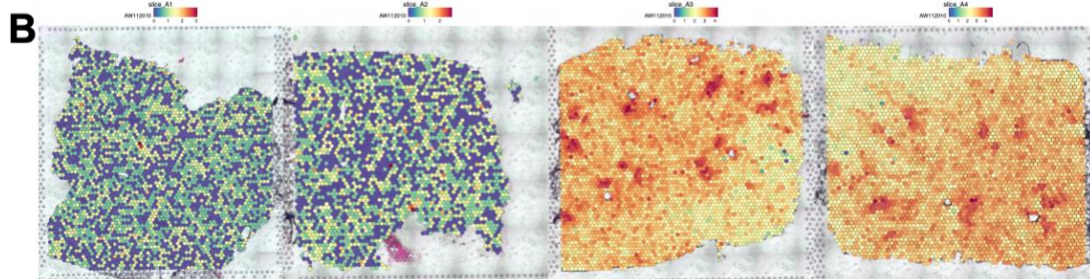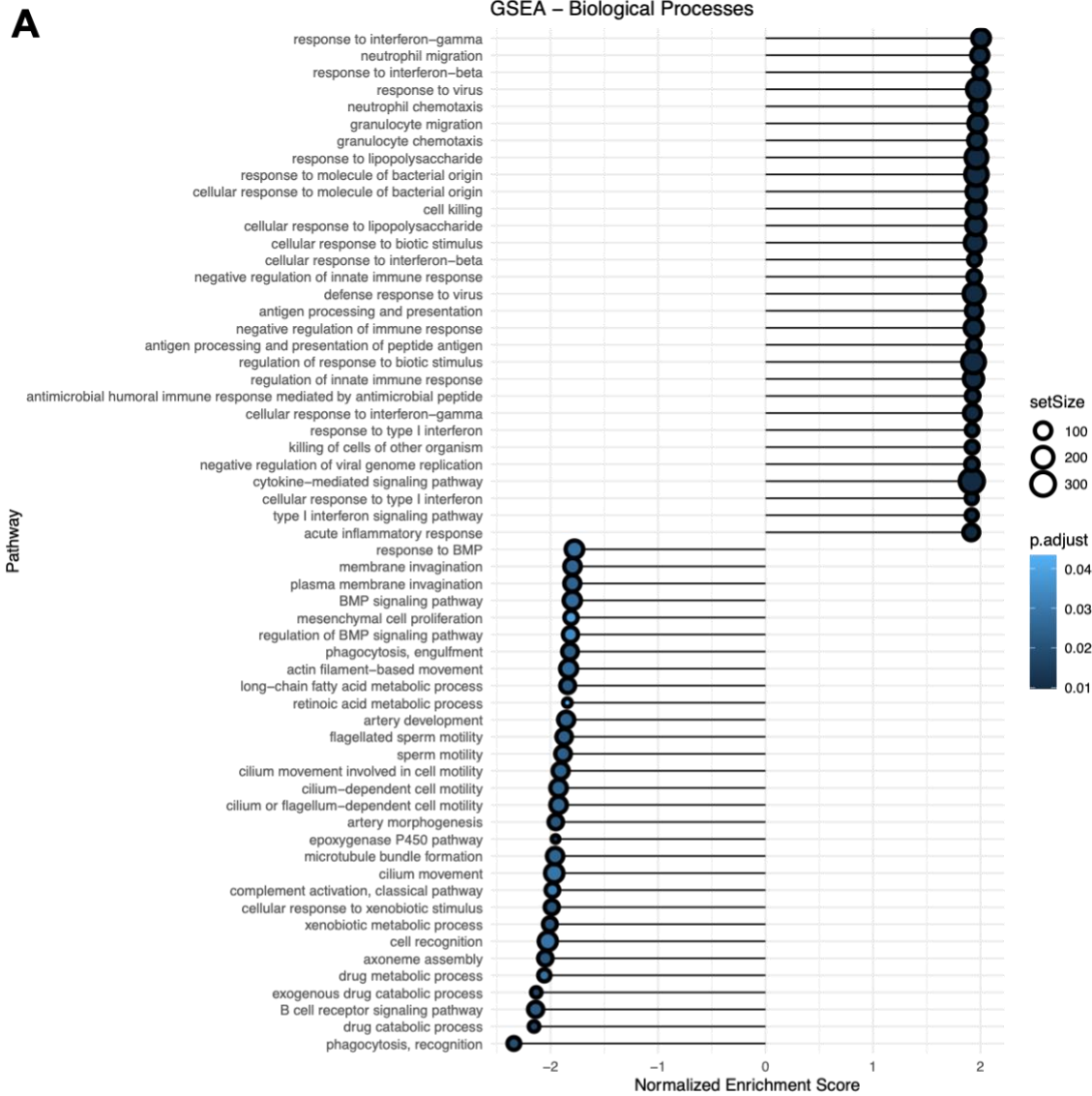
Figure 5. Differential expression analysis of spots annotated as airways in the immunized and the re-challenged mouse

(A) Gene Set Enrichment Analysis (GSEA) of all the 16,937 genes available in the DE analysis of the airways.

(B) Expression of AW112010, a top DE gene up-regulated in the airways upon re-challenge, across four slices.

(C) Expression of *Cbr2*, a top DE gene down-regulated in the airways upon re-challenge, across four slices.
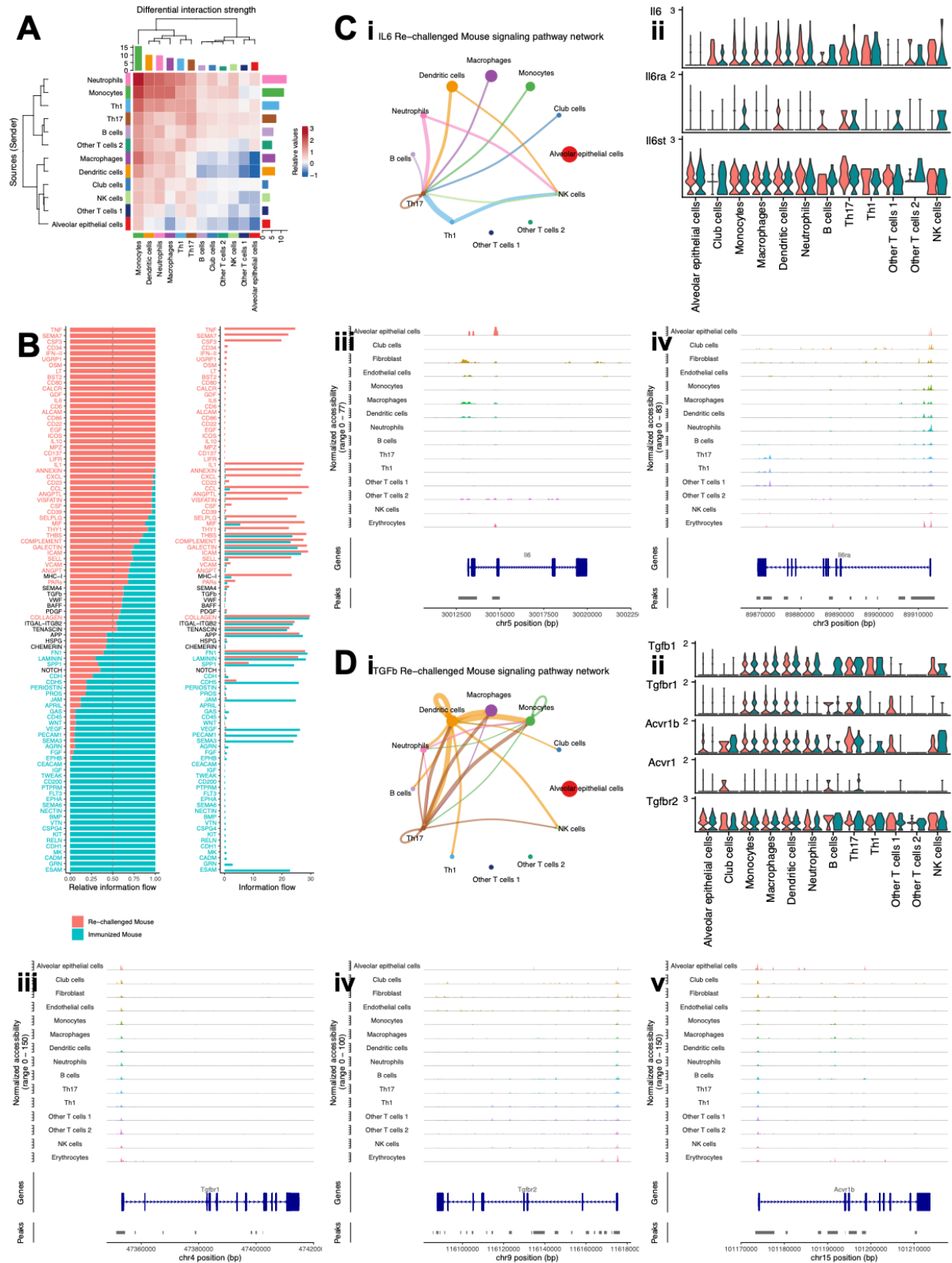
Figure 6. Cell-cell communication among cell-type enriched spots

(A) Heatmap showing differential interaction strength between cell-type enriched spots. Outgoing signals were shown in rows, while incoming signals were shown in columns. Increased (or decreased) signals in the re-challenged mouse compared to the immunized mouse were represented using red (or blue) in the color bar. The sum of values within the same column was summarized using the colored bar plot on the top. The sum of values within the same row was summarized using the colored bar plot on the right.

(B) Bar plots showing overall information flow of each signaling pathway. Relative information flow was shown in the stacked bar plot, while raw information flow was shown in the regular bar plot. Enriched signaling pathways were colored in red or cyan.

(C) Th17 enriched spots in the re-challenged mouse were the receivers of the IL6 signaling pathway. (i) Circle plot visualizing the inferred communication network of the IL6 signaling pathway in the re-challenged mouse. The network of the IL6 signaling pathway in the immunized mouse was not significant. Circle sizes represented the number of spots in each group. Edge colors were consistent with the senders of the signal (sources), and edge weights represented the interaction strength. (ii) Violin plot visualizing the expression of genes related to IL6 signaling pathway in cell-type enriched spots from the re-challenged mouse. Gene expression from slice A3 was colored in red, and that from slice A4 was colored in cyan. (iii-iv) Peaks in genomic regions around *Il6* and *Il6ra* in scATAC-seq data.

(D) Th17 enriched spots in the re-challenged mouse were the receivers of the TGF-β signaling pathway. (i) Circle plot visualizing the inferred communication network of the TGF-β signaling pathway in the re-challenged mouse. The network of the TGF-β signaling pathway in the immunized mouse was also significant but not shown. Circle sizes represented the number of spots in each group. Edge colors were consistent with the senders of the signal (sources), and edge weights represented the interaction strength. (ii) Violin plot visualizing the expression of genes related to TGF-β signaling pathway in cell-type enriched spots from the re-challenged mouse. Gene expression from slice A3 was colored in red, and that from slice A4 was colored in cyan. (iii-v) Peaks in genomic regions around *Tgfbr1*, *Tgfbr2*, and *Acvr1b* in scATAC-seq data.

Table 1 Top 50 (hierarchically sorted by fold-enrichment) over-represented biological processes for both mice

Immunized Mouse				Re-challenged Mouse			
GO Biological Process	Fold Enrichment	P-value	FDR-adjusted P-value	GO Biological Process	Fold Enrichment	P-value	FDR-adjusted P-value
protein localization to bicellular tight junction	6.22	1.36E-03	2.04E-02	protein localization to ciliary transition zone	6.64	4.12E-04	7.33E-03
inner dynein arm assembly	6.22	2.60E-05	6.53E-04	intraciliary retrograde transport	6.2	4.72E-06	1.36E-04
cranial ganglion development	6.22	3.06E-03	3.94E-02	inner dynein arm assembly	6.09	5.62E-05	1.23E-03
positive regulation of cell migration by vascular endothelial growth factor signali	5.53	1.01E-03	1.60E-02	hemoglobin biosynthetic process	5.81	1.61E-03	2.33E-02
peripheral nervous system myelin maintenance	5.53	1.01E-03	1.60E-02	protein localization to bicellular tight junction	5.69	3.77E-03	4.65E-02
intraciliary retrograde transport	5.39	3.27E-05	8.07E-04	negative regulation of skeletal muscle tissue development	5.69	3.77E-03	4.65E-02
outer dynein arm assembly	5.29	2.42E-06	7.67E-05	positive regulation of homotypic cell-cell adhesion	5.17	2.53E-03	3.37E-02
cerebrospinal fluid circulation	5.26	1.54E-04	3.18E-03	epithelial cilium movement involved in determination of left/right asymmetry	5.17	2.53E-03	3.37E-02
establishment of blood-brain barrier	5.18	3.35E-04	6.17E-03	positive regulation of integrin-mediated signaling pathway	5.17	2.53E-03	3.36E-02
endothelial tube morphogenesis	5.18	3.35E-04	6.17E-03	negative regulation of CD4-positive, alpha-beta T cell proliferation	5.17	2.53E-03	3.36E-02
glomerulus morphogenesis	5.09	7.32E-04	1.21E-02	negative regulation of actin nucleation	5.17	2.53E-03	3.36E-02
dichotomous subdivision of an epithelial terminal unit	5.09	7.32E-04	1.21E-02	antigen processing and presentation of exogenous peptide antigen via MHC class I	5.17	2.53E-03	3.36E-02
defense response to tumor cell	4.98	1.60E-03	2.34E-02	cerebrospinal fluid circulation	5.11	3.27E-04	6.00E-03
axoneme assembly	4.92	6.34E-19	8.26E-17	neutrophil activation involved in immune response	4.83	1.68E-03	2.41E-02
epithelial cilium movement involved in determination of left/right asymmetry	4.84	3.52E-03	4.43E-02	branching involved in labyrinthine layer morphogenesis	4.74	4.95E-04	8.50E-03
positive regulation of aspartic-type peptidase activity	4.84	3.52E-03	4.43E-02	defense response to tumor cell	4.65	3.83E-03	4.70E-02
surfactant homeostasis	4.84	3.73E-05	9.05E-04	outer dynein arm assembly	4.65	4.46E-05	1.00E-03
regulation of cilium beat frequency	4.56	3.65E-04	6.65E-03	positive regulation of superoxide anion generation	4.62	1.36E-05	3.51E-04
negative regulation of endothelial cell differentiation	4.52	2.43E-03	3.27E-02	regulation of Fc receptor mediated stimulatory signaling pathway	4.43	2.48E-03	3.34E-02
regulation of aspartic-type endopeptidase activity involved in amyloid precursor ʃ	4.52	2.43E-03	3.27E-02	collagen-activated tyrosine kinase receptor signaling pathway	4.43	2.48E-03	3.33E-02
glomerulus vasculature development	4.39	2.55E-04	4.92E-03	establishment of endothelial intestinal barrier	4.43	2.48E-03	3.33E-02
sperm axoneme assembly	4.38	4.46E-06	1.35E-04	negative regulation of leukocyte degranulation	4.43	2.48E-03	3.33E-02
positive regulation of protein localization to endosome	4.31	1.67E-03	2.42E-02	T cell migration	4.43	1.99E-05	4.83E-04
regulation of vasculogenesis	4.28	5.39E-04	9.26E-03	respiratory burst	4.27	1.61E-03	2.33E-02
cellular response to low-density lipoprotein particle stimulus	4.28	5.39E-04	9.25E-03	sperm axoneme assembly	4.18	1.89E-05	4.65E-04
venous blood vessel development	4.26	1.77E-04	3.57E-03	regulation of superoxide anion generation	4.18	1.89E-05	4.65E-04
membrane raft organization	4.24	5.84E-05	1.35E-03	positive regulation of neutrophil chemotaxis	4.18	1.89E-05	4.63E-04
regulation of establishment of endothelial barrier	4.15	3.71E-04	6.75E-03	negative regulation of transforming growth factor beta production	4.09	3.56E-03	4.46E-02
regulation of podosome assembly	4.15	3.71E-04	6.74E-03	positive regulation of vascular endothelial growth factor production	4.09	4.07E-05	9.28E-04
leukocyte aggregation	4.15	1.14E-03	1.76E-02	surfactant homeostasis	4.06	6.78E-04	1.11E-02
trachea morphogenesis	4.15	3.57E-03	4.48E-02	positive regulation of leukocyte adhesion to vascular endothelial cell	4.04	1.33E-04	2.69E-03
semaphorin-plexin signaling pathway involved in axon guidance	4.15	3.57E-03	4.48E-02	regulation of neutrophil activation	3.98	2.28E-03	3.14E-02
establishment of endothelial intestinal barrier	4.15	3.57E-03	4.48E-02	leukocyte aggregation	3.98	2.28E-03	3.13E-02
tongue morphogenesis	4.15	3.57E-03	4.47E-02	regulation of cilium beat frequency	3.98	2.28E-03	3.13E-02
collagen-activated signaling pathway	4	2.41E-03	3.25E-02	Arp2/3 complex-mediated actin nucleation	3.98	2.28E-03	3.13E-02
adhesion of symbiont to host	4	2.41E-03	3.24E-02	neutrophil chemotaxis	3.95	2.64E-11	1.82E-09
retina vasculature development in camera-type eye	3.93	5.30E-04	9.11E-03	positive regulation of monocyte chemotaxis	3.92	2.86E-04	5.31E-03
epithelial tube branching involved in lung morphogenesis	3.89	9.20E-06	2.54E-04	positive regulation of sprouting angiogenesis	3.92	2.86E-04	5.30E-03
lymphangiogenesis	3.89	1.63E-03	2.36E-02	regulation of killing of cells of other organism	3.91	1.47E-03	2.18E-02
negative regulation of stress fiber assembly	3.89	9.20E-06	2.53E-04	establishment of endothelial barrier	3.91	7.57E-06	2.07E-04
stress fiber assembly	3.8	1.10E-03	1.71E-02	positive regulation of granulocyte chemotaxis	3.89	3.73E-05	8.62E-04
transcytosis	3.79	2.45E-04	4.78E-03	positive regulation of myoblast fusion	3.84	9.45E-04	1.49E-02
branching involved in salivary gland morphogenesis	3.73	7.41E-04	1.22E-02	regulation of actin cytoskeleton reorganization	3.84	3.26E-06	9.69E-05
Arp2/3 complex-mediated actin nucleation	3.73	3.39E-03	4.31E-02	ribosomal small subunit assembly	3.84	9.45E-04	1.49E-02
vascular endothelial growth factor receptor signaling pathway	3.73	1.67E-04	3.40E-03	monocyte chemotaxis	3.82	1.60E-05	4.00E-04
antigen processing and presentation of exogenous peptide antigen	3.69	1.13E-04	2.41E-03	glutathione metabolic process	3.81	7.58E-10	4.17E-08
cellular response to fluid shear stress	3.66	2.27E-03	3.10E-02	positive regulation of T-helper cell differentiation	3.79	6.10E-04	1.02E-02
regulation of cell-cell adhesion mediated by cadherin	3.66	2.83E-06	8.93E-05	regulation of vasculogenesis	3.74	3.16E-03	4.05E-02
branching involved in blood vessel morphogenesis	3.64	2.83E-06	8.93E-05	microglial cell activation	3.74	3.35E-05	7.79E-04
ribosomal small subunit assembly	3.6	1.52E-03	2.25E-02	superoxide anion generation	3.74	3.16E-03	4.05E-02

Biological processes related to immune responses were highlighted in red.