












# An interactive murine single-cell atlas of the lung responses to radiation injury

Received: 29 July 2022

Accepted: 17 April 2023

Published online: 28 April 2023

 Check for updates

Sandra Curras-Alonso <sup>1,2</sup>, Juliette Soulier <sup>1,2</sup>, Thomas Defard <sup>3,4,5,6</sup>,  
Christian Weber<sup>6</sup>, Sophie Heinrich <sup>2</sup>, Hugo Laporte<sup>1,2</sup>, Sophie Leboucher <sup>7</sup>,  
Sonia Lameiras<sup>8</sup>, Marie Dutreix <sup>2</sup>, Vincent Favaudon<sup>2</sup>, Florian Massip <sup>3,4,5</sup>,  
Thomas Walter<sup>3,4,5</sup>, Florian Mueller <sup>6</sup>, José-Arturo Londoño-Vallejo <sup>1,2</sup> ✉ &  
Charles Fouillade <sup>2</sup> ✉

Radiation Induced Lung Injury (RILI) is one of the main limiting factors of thorax irradiation, which can induce acute pneumonitis as well as pulmonary fibrosis, the latter being a life-threatening condition. The order of cellular and molecular events in the progression towards fibrosis is key to the pathogenesis of the disease, yet their coordination in space and time remains largely unexplored. Here, we present an interactive murine single cell atlas of the lung response to irradiation, generated from C57BL6/J female mice. This tool opens the door for exploration of the spatio-temporal dynamics of the mechanisms that lead to radiation-induced pulmonary fibrosis. It depicts with unprecedented detail cell type-specific radiation-induced responses associated with either lung regeneration or the failure thereof. A better understanding of the mechanisms leading to lung fibrosis will help finding new therapeutic options that could improve patients' quality of life.

Radiation therapy is one of the main therapeutic options used to treat thoracic cancers. Nevertheless, the lung is a sensitive organ to ionizing radiation (IR) which makes it the main dose-limiting organ in the thorax<sup>1,2</sup>. Radiation-induced lung injury (RILI) results in both early and late molecular and cellular toxicities, some of them irreversible, impacting the quality of life of patients. IR induces DNA damage and oxidative stress, followed by inflammation and tissue reorganization, which may be resolved into regeneration and regain of organ function. However, depending on the total dose as well as patient sensitivity, radio-induced pneumonitis may evolve toward radiation-induced pulmonary fibrosis (RIPF), characterized by fibroblast and myofibroblast proliferation, excessive extracellular matrix (ECM) deposition, bronchiolization, and honeycomb cyst formation, leading to disruption of gas exchange and progressive organ failure<sup>3,4</sup>.

The mechanisms that determine success or failure of tissue regeneration after irradiation are not well understood. At the cellular level, pro-fibrogenic doses of irradiation trigger apoptosis, senescence, cytokine secretion, as well as cell transitions likely affecting multiple cell compartments. Single-cell technologies have provided powerful tools to untangle the complex cellular heterogeneity of organs such as the lung, an exceptional challenge in the respiratory field. In the past years, the Lung Biological Network of the Human Cell Atlas (HCA) has emerged with the objective of establishing a complete atlas of the healthy human lung, collecting data on molecular phenotypes of different cell types, cell transitions as well as their location all along the airways. In addition, LungMAP has gathered scRNA-seq datasets to provide a collaborative and open-access comprehensive molecular atlas of the healthy developing lung<sup>5</sup>. In parallel, other consortia have gathered efforts to build the Lung Disease Cell Atlas

<sup>1</sup>Institut Curie, CNRS UMR 3244, Sorbonne Université, PSL University, 75005 Paris, France. <sup>2</sup>Institut Curie, Inserm U1021-CNRS UMR 3347, University Paris-Saclay, PSL University, Centre Universitaire, 91405 Orsay Cedex, France. <sup>3</sup>Centre for Computational Biology (CBIO), Mines Paris, PSL University, 75006 Paris, France. <sup>4</sup>Institut Curie, PSL University, 75005 Paris, France. <sup>5</sup>INSERM, U900, 75005 Paris, France. <sup>6</sup>Imaging and Modeling Unit, Institut Pasteur, Université Paris Cité, Paris, France. <sup>7</sup>Institut Curie, CNRS UMR 3348, University Paris-Saclay, PSL University, Centre Universitaire, Orsay, France. <sup>8</sup>Institut Curie Genomics of Excellence (ICGex) Platform, Paris, France. ✉e-mail: [arturo.londono@curie.fr](mailto:arturo.londono@curie.fr); [charles.fouillade@curie.fr](mailto:charles.fouillade@curie.fr)

(e.g., the IPF cell atlas<sup>6,7</sup>, the COVID cell atlas<sup>8</sup>, the COPD cell atlas<sup>9</sup>) to better describe lung cell and tissue responses associated with disease. Such efforts have allowed to build an integrated cell atlas of the human lung in health and disease<sup>10</sup>. Nevertheless, to date, no single cell-based efforts have been aimed at studying RIPF. Understandably, progresses in that direction face major obstacles, in particular, due to very limited access to specimens from irradiated human lungs. In this context, well-characterized mouse models of RIPF, which recapitulate major features of the disease progression in the human (namely, an early period of inflammation, irreversibly followed by fibrosis and death) carry high information value.

The purpose of this work is to provide a whole organ single-cell atlas spanning the evolution over time towards pulmonary fibrosis, from the early response and the inflammatory phase to the end-point fibrotic process. This communication illustrates the potential of the information provided by this atlas by focusing on representative cell types from the different cell compartments: epithelial (AT2 cells), mesenchymal (fibroblasts and myofibroblasts), myeloid (AMs and IMs), and endothelial (aCap and gCap) cells. We have combined single-cell RNA sequencing (scRNA-seq) with single-molecule fluorescence in situ hybridization (smFISH) to confirm specific findings and have analyzed using computational tools the dynamics in cell-cell communications to pinpoint biologically meaningful interactions that might be involved in the evolution of RIPF.

## Results

### Cellular composition dynamics after lung irradiation

To study the impact of radiation therapy in the lung, we performed scRNA-seq using the 10x Chromium Controller V3 technology of dissociated lungs from non-irradiated mice (control) and from mice 1, 2, 3, 4, and 5 months after fibrogenic (17 Gy) and non-fibrogenic (10 Gy) doses of IR<sup>11</sup>. We used 5 age-matched, non-IR mice as control, 5 mice after 10 Gy thorax IR (one mouse per time point from 1 to 5 months) and 10 mice after 17 Gy thorax IR (two mice per time point from 1 to 5 months) (Fig. 1a). Each mouse lung was enzymatically and mechanically dissociated into a single cell suspension that was then loaded into the 10x microfluidic system.

A total of 102,869 cells were obtained, from which 22,378 belonged to the healthy mice, 26,360 to the IR<sub>10Gy</sub> mice and 54,131 to the IR<sub>17Gy</sub> mice. UMAP visualization after merging the total dataset or per condition displayed similar distributions of cells into clusters, both in the control and in the two IR conditions over time, which indicated the absence of obvious batch effects due to sample processing or sequencing, thus underlining the reproducibility of the single cell experimental procedure and analysis (Fig. 1b, Supplementary Fig. 1a–c).

We used cell type-specific markers from recently published scRNA-seq datasets<sup>12–14</sup> to annotate major populations. This resulted in the identification of 18 main cell types (Fig. 1b, c): non-immune cells, which comprise 4 epithelial cell populations (AT2 cells, AT1 cells, club cells, and ciliated cells), 3 mesenchymal cell clusters (fibroblasts, smooth muscle cells -SMC- and mesotheliocytes), 1 endothelial cell -EC- cluster and several clusters of immune cells, which include 7 myeloid compartments (monocytes, alveolar macrophages -AM-, interstitial macrophages -IM-, dendritic cells -DC-, plasmacytoid dendritic cells -pDC-, neutrophils and basophils) and 3 lymphoid compartments (T cells, natural killer -NK- cells and B cells). We also identified subpopulations of proliferating DC, AM, and T cells.

As a first approach to the study of RIPF, we sought to detect changes in cell proportions affecting the different compartments after irradiation. These results, illustrated in Fig. 1 and Supplementary Fig. 1, pointed to changes in the proportions of immune cells, which tended to increase after 10 Gy or 17 Gy irradiation, e.g., alveolar macrophages (Supplementary Fig. 1d), and of epithelial cells, which tended to progressively decrease, in particular after 17 Gy irradiation (Fig. 1d;

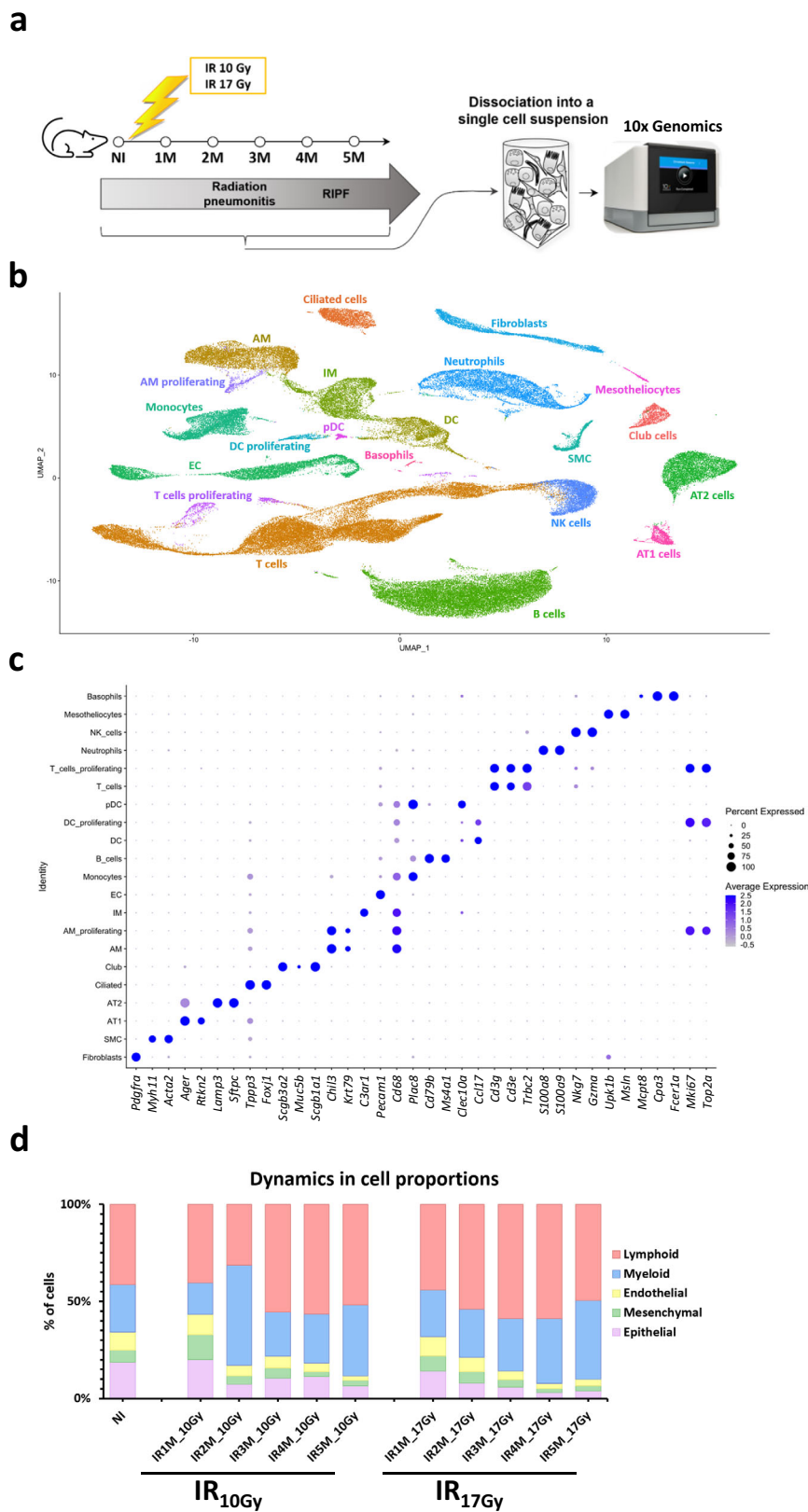
Supplementary Fig. 1e). Similar observations were made concerning the endothelial compartment (Fig. 1d; Supplementary Fig. 1f). However interesting, extreme caution should be exerted when interpreting such results as they can be easily biased due, for instance, to the relative efficiency in tissue dissociation (in particular under fibrotic conditions). Also, given the limits imposed by the technology in the number of cells analyzed per sample, changes in the numbers of one cell compartment will necessarily affect the calculated proportions in other compartments. Taking these considerations into account, in the following sections we coupled scRNA-seq analyses with smFISH experiments to obtain independent confirmation of changes in cell proportions affecting AT2 cells, macrophages, endothelial cells, and fibroblasts.

### Transcriptomic dynamics in AT2 cells point to transdifferentiation in response to fibrogenic irradiation

AT2 cells play an important role in the lung as they secrete surfactants that maintain surface tension and prevent alveolus collapse. They are also key elements in lung homeostasis because of their stem cell capacity and ability to differentiate into AT1 cells, the latter being responsible for air exchange. In agreement with the detected decrease in the proportion of epithelial cells, the total number of AT2 cells per sample was decreased in all samples after IR. AT2 cells represented 12.6% (SD 3.92%) in the control mice. This proportion decreases after IR<sub>10Gy</sub> but it is more progressive and pronounced after IR<sub>17Gy</sub> (Fig. 2a, Supplementary Fig. 2a). We validated this observation in smFISH experiments where NI, IR<sub>10Gy</sub>, and IR<sub>17Gy</sub> lung tissue sections were probed against *Lamp3*, a robust and specific marker for AT2 cells (Supplementary Fig. 2b). As shown in (Fig. 2b, c, Supplementary Fig. 2c), there is indeed a significant decrease in the AT2 cell proportion in the 17 Gy condition after 5 months. Interestingly, the volume of the AT2 cells, which was estimated based on the distribution of *Lamp3* smFISH signal in the cytoplasm of these cells (see Methods), showed a significant increase in this particular condition, while it was not altered after IR<sub>5M10Gy</sub> (Fig. 2c), suggesting possible significant changes in the genetic program exclusively associated with pro-fibrogenic IR.

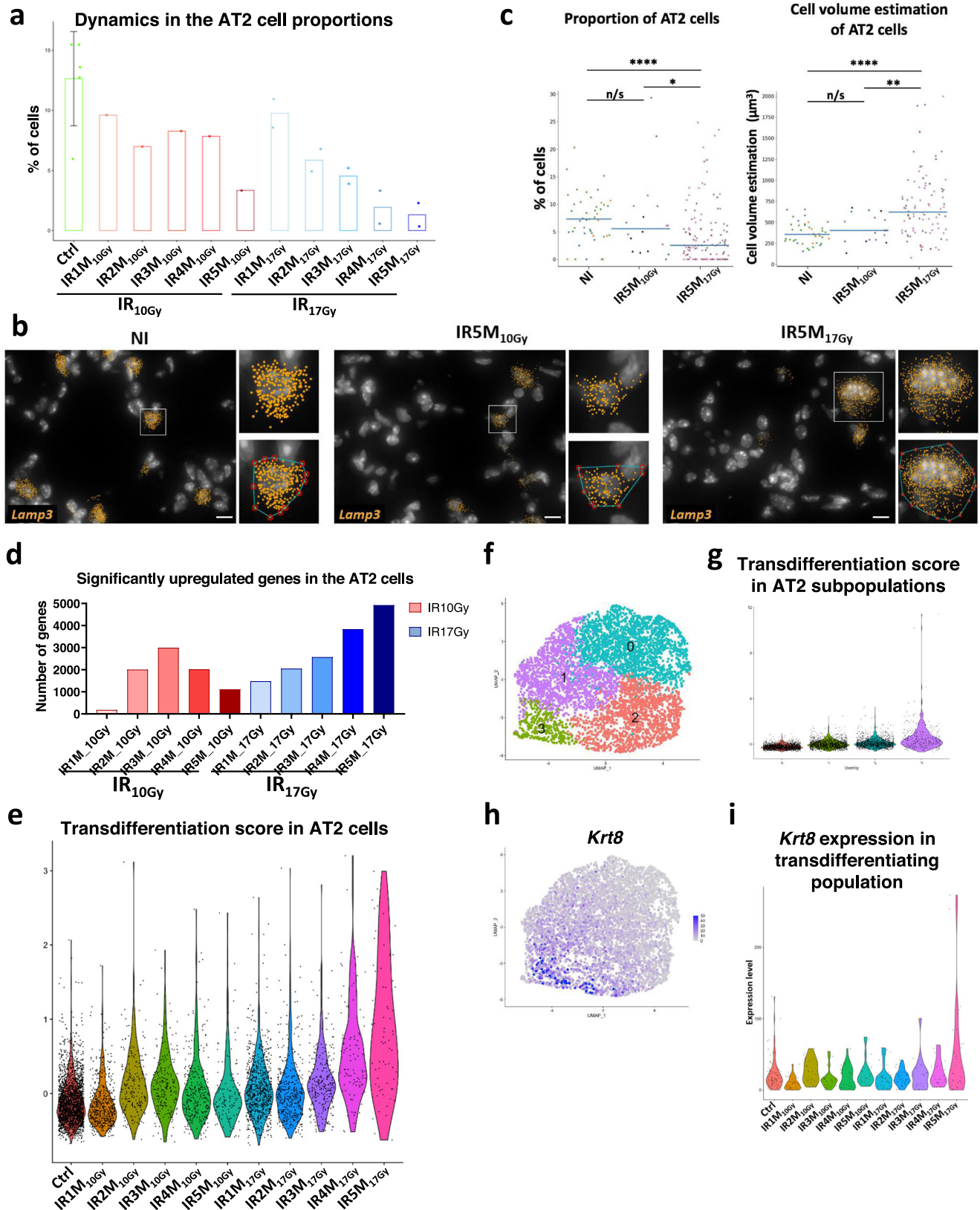
The distribution of the irradiated samples within the AT2 cells cluster did not entirely overlap with the control samples (Supplementary Fig. 2a), suggesting IR-induced transcriptional changes. Differentially expressed genes (DEG) analysis revealed qualitative and quantitative differences in the transcriptional response to IR<sub>17Gy</sub> and to IR<sub>10Gy</sub> (Fig. 2d). After IR<sub>17Gy</sub>, the number of upregulated genes increased steadily up to 5M. In comparison, the response to IR<sub>10Gy</sub> was a transient one, reaching a peak after 3 months post-IR and decreasing afterwards (Fig. 2d). Amongst the genes differentially expressed by irradiated AT2 cells there were genes typically involved in Epithelial-Mesenchymal transition (EMT) (e.g., *Serpine2*, *Tgfb2*, *Col6a2*, *Col4a1*, *Col4a2*, *Colgalt1*, *Egfl6*)<sup>15</sup>. To better characterize EMT dynamics in AT2 cells after IR, we selected 31 EMT-associated genes that were significantly upregulated after IR<sub>5M17Gy</sub> and followed their expression from 1 to 5 months after 10 and 17 Gy IR (Supplementary Fig. 2h). Interestingly, while all these genes were only slightly and transiently upregulated after IR<sub>10Gy</sub>, they only showed strong upregulation at 4 and 5 months after IR<sub>17Gy</sub> (Supplementary Fig. 2h–j), that is, during the fibrogenic phase. Whether EMT in AT2 cells directly contributes to RIPF or constitutes a secondary response to the pro-fibrotic micro-environment remains to be determined.

Another look at the list of deregulated genes after a pro-fibrogenic dose revealed a number of genes known to be specific markers of AT1 cells (e.g., *Akap5*, *Aqp5*). This observation suggested that, similar to what has been described in response to other types of injuries<sup>16</sup>, AT2 cells may undergo transdifferentiation towards AT1 in response to IR. To further explore this possibility, we analyzed the expression of the 30 genes composing the AT2 > AT1 transdifferentiation signature. Strikingly, 1M after 17 Gy irradiation, 30% of these genes are



**Fig. 1 | Radiation-induced lung injury cell atlas.** **a** Scheme of the experimental set up. Mice are irradiated with a dose of 10 Gy or 17 Gy and monitored by CT-scan to follow the development of pulmonary fibrosis. Mice are sacrificed at 1, 2, 3, 4, and 5 months after IR, and lungs are enzymatically and mechanically dissociated into a single cell suspension before being loaded in the 10x Chromium Controller System (image provided by 10x Genomics). **b** UMAP visualization of 102,869 cells from 20

different samples (5 NI; 5 IR<sub>10Gy</sub>, one per time point; 10 IR<sub>17Gy</sub>, two per time point) annotated by cell type. **c** Dot plot of the expression of the markers used for cell type identification. **d** Dynamics in cell proportions of the endothelial, mesenchymal, epithelial, lymphoid, and myeloid cells across the NI and IR conditions at the different time points and doses.



upregulated, a proportion that increases progressively over the following months, reaching 63.3% at 5M after IR<sub>17Gy</sub> (Supplementary Fig. 2d). This observation lends support to the assumption that a process of AT2 > AT1 transdifferentiation occurs progressively after IR<sub>17Gy</sub> reaching its strongest expression when fibrosis is already declared (Fig. 2e, Supplementary Fig. 2d). In contrast, when a

non-fibrogenic dose is given (IR<sub>10Gy</sub>), such transdifferentiation appears to be limited both in strength (26.7% of the genes) and time, reaching its maximum at 3 months after IR<sub>10Gy</sub> (Fig. 2e, Supplementary Fig. 2d).

More precisely, we observed a particular subpopulation of AT2 cells that showed the highest transdifferentiation score, the cluster 3 (Fig. 2f,g). This cluster was characterized by the expression of *Krt8*

**Fig. 2 | Cellular and molecular changes in the AT2 cells during RILI reveal a transdifferentiation profile after fibrogenic doses of IR.** **a** Dynamics in the proportion of the AT2 cells in the NI ( $n = 5$ ) and at the different time points after IR<sub>10Gy</sub> ( $n = 1$ ) and IR<sub>17Gy</sub> ( $n = 2$ ). Error bar refers to the standard deviation of the data. **b** Automatic *Lamp3* mRNA (orange) detection with Big-FISH in NI, IR5M<sub>10Gy</sub>, and IR5M<sub>17Gy</sub> lung tissue sections. Inset top panel shows an AT2 cell; inset bottom panel shows the convex hull of a cluster of mRNA spots. Scale bars, 10  $\mu$ m. **c** Quantification and cell volume estimation of the *Lamp3*<sup>+</sup> cells in NI, IR5M<sub>10Gy</sub>, and IR5M<sub>17Gy</sub> lung tissue sections. To compare two groups, the *P* value was computed with the Mann–Whitney–Wilcoxon test (two-sided test) from scipy ( $n/s$ , adjusted *p* value >0.05; \*, adjusted *p* value <0.05; \*\*, adjusted *p* value <0.01; \*\*\*, adjusted

*p* value <0.001; \*\*\*\*, adjusted *p* value <0.0001). Each dot represents one analyzed image. Each color per time point represents a different biological replicate (NI  $n = 3$ ; IR5M<sub>10Gy</sub>  $n = 3$ ; IR5M<sub>17Gy</sub>  $n = 5$ ). **d** Dynamics in the significantly upregulated genes in the AT2 cells compared to the NI samples at the different time points after IR<sub>10Gy</sub> and IR<sub>17Gy</sub>. **e** Violin plot showing the single cell score calculated based on the transdifferentiation expressed genes in the AT2 cells. **f** UMAP visualization of the different AT2 cell subpopulations. **g** Violin plot showing the single cell score calculated based on the transdifferentiation expressed genes in the different AT2 cell subpopulations. **h** UMAP visualization of the expression of *Krt8*. **i** Violin plot of *Krt8* expression in the AT2 cells cluster three in the NI samples and at the different time points after IR<sub>10Gy</sub> and IR<sub>17Gy</sub>.

(Fig. 2h), which was mainly upregulated 5 months after 17 Gy irradiation (Fig. 2i). Previous studies have demonstrated the existence of an alveolar epithelial *Krt8*<sup>+</sup> transitional stem cell state that derives from activated AT2 cells and differentiates into AT1 cells<sup>17</sup>, suggesting that the cluster 3 may transdifferentiate towards AT1 cells. To address this hypothesis, we performed a trajectory analysis using AT2 and AT1 cells which shows the connection between these two epithelial populations of the alveolus (Supplementary Fig. 2e). Moreover, the pseudotime analysis using the transdifferentiation-associated genes to order the cells suggests that AT2 cells differentiate towards AT1 cells (Supplementary Fig. 2f). To identify specific regulatory factors, we ran SCENIC on the AT2 subset (Supplementary Fig. 2g). This regulatory network analysis showed the activation of four transcription factors at 4 and 5 months post irradiation: *Stat1*, *Stat3*, *Irf7*, and *Xbp1*. Interestingly, *Stat1* and *Stat3* have previously been shown to be activated during the development of idiopathic pulmonary fibrosis<sup>18</sup>.

Overall, these results indicate that there is a progressive loss of AT2 cells only after a fibrogenic irradiation dose and that this loss is associated with progressive and profound changes in the transcriptome landscape pointing to physiopathogenic transdifferentiation processes.

### Fibrogenic irradiation triggers a strong ECM genes response in fibroblasts

Lung fibroblasts are crucial for maintaining the integrity of the alveolar structure and play key roles in the response to injury through proliferation and remodeling of surrounding tissue. Specific analysis of the fibroblast compartment from all lung samples allowed us to distinguish 3 different sub-populations (Fig. 3a): two different matrix fibroblasts, one *Coll3a1*-positive (*Coll3a1*<sup>+</sup>, *Tcf21*<sup>+</sup>) and one *Coll4a1*-positive (*Coll4a1*<sup>+</sup>, *Pi16*<sup>+</sup>, *Meg3*<sup>+</sup>)<sup>13</sup>, and one sub-population of myofibroblasts (*Hhip*<sup>+</sup>, *Cdh11*<sup>+</sup>, *Pdgfrb*<sup>+</sup>)<sup>13,19,20</sup> (Supplementary Fig. 3a). Surprisingly, most fibroblasts obtained 5 months after fibrogenic irradiation clustered with the latter (Fig. 3a, b) and represented over 80% of all fibroblasts in these samples (Fig. 3c). The analysis of the dynamics affecting fibroblasts subpopulations after irradiation indicated that the proportion of myofibroblasts increased dramatically 4 and 5 months after 17 Gy irradiation in detriment of both types of matrix fibroblasts (Fig. 3c). This observation was unexpected since lungs appeared to be quite fibrotic when processed, especially after 5 M post-17Gy irradiation. To corroborate this information, we performed smFISH using a probe to detect *Pdfgra*, a general marker for fibroblasts (Supplementary Fig. 3b). As shown in Fig. 3d, e, the number of fibroblasts detected per field of view is significantly higher 5 months after irradiation which is in striking contrast to the numbers inferred from the scRNA-seq analyses. This observation suggests a bias in the cell composition after the dissociation of injured lungs, in particular under fibrogenic conditions. On the other hand, a smFISH analysis using probes to detect *Hhip*, a specific marker for myofibroblasts, also indicated a higher proportion of these cells detected per field of view 5 months after fibrogenic irradiation (Fig. 3d, e). Finally, a small number of fibroblasts appeared to co-express both *Pdfgra* and

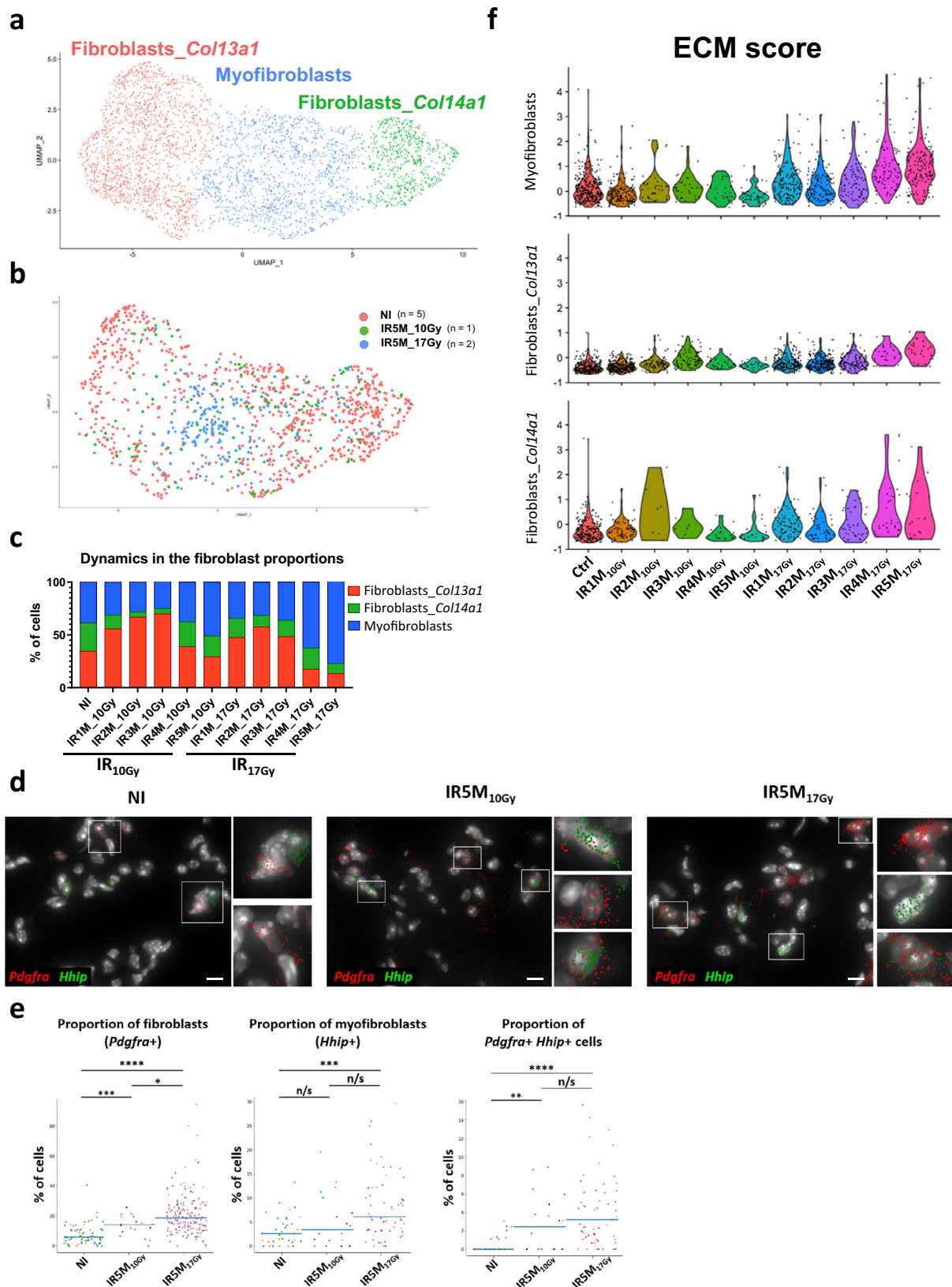
*Hhip*, and this number appeared to increase 5 months after irradiation (0.4% in NI to 2.8% IR5M<sub>10Gy</sub> and to 4.2% IR5M<sub>17Gy</sub>).

Despite the detected bias in cell composition at late time points revealed by smFISH, we attempted a fibroblast DEG analysis between samples. This analysis indicates that the number of differentially upregulated genes increases steadily after 17 Gy irradiation (Supplementary Fig. 3c). As predicted, the transcriptomic response of fibroblasts to irradiation comprises genes known to impact ECM deposition, which were significantly upregulated in all three compartments (Fig. 3f, Supplementary Fig. 3d). None of these ECM genes were significantly upregulated 4 and 5 M after IR<sub>10Gy</sub>, supporting the notion that this is part of the toxic response of fibroblasts specifically associated to lung fibrosis.

### Specific macrophage compartments display either pro-inflammatory or profibrotic profiles after IR17Gy

Analysis of the 11,678 macrophages from NI, IR<sub>10Gy</sub> and IR<sub>17Gy</sub> samples at different time points identified two main macrophages populations (Fig. 4a, b): alveolar macrophages (AM), characterized by the expression of *Chil3* and *Plet1* (Supplementary Fig. 4a) and interstitial macrophages (IM), which shows high levels of expression of *C1qa* and *C1qb* (Supplementary Fig. 4a). Further analysis showed that IMs could be subdivided in three different subsets; the first subset is characterized by the expression of *Folr2*, *Ccl8* and *Cd163* (called here IM\_C1), the second subset expresses *H2-Dma*, *Zmynd15* or *Cd63* (IM\_C2), and the third subset shows high levels of expression of *Ccr2* and *S100a4* (IM\_C3) (Fig. 4a, Supplementary Fig. 4a). Similarly, AMs were distributed into 2 different subpopulations: one characterized by the expression of *Krt19*, *Fabp1* and *Krt79* (here called AM\_C1) and the other one characterized by high levels of *Chil3*, *Wfdc21* and *Ctsd* (AM\_C2) (Fig. 4a, Supplementary Fig. 4a). Macrophages from all samples were not homogeneously distributed in these subpopulations. For instance, the AM\_C2 subpopulation was particularly enriched in AMs obtained 5 months after 17 Gy irradiation (Fig. 4b). This alveolar macrophage subpopulation appeared almost exclusively after irradiation and was already strongly reinforced at 4 months after 17 Gy (Fig. 4c) reaching 72–78% of the total AMs. The changes that affected the proportions of IMs after irradiation were more subtle, although a relative increase in IM\_C3 is noticeable 4 and 5 months after 17 Gy irradiation (from 11.4% in the NI to 38.6% in the IR5M) (Fig. 4c). Furthermore, an increase in the numbers of both AMs and IMs was confirmed in the lungs by smFISH 5 months post-17Gy while this is not the case after 10 Gy irradiation (Fig. 4d, e, Supplementary Fig. 4c).

DEG analysis of IMs showed an upregulation of markers known to define the macrophage M1 activated state (e.g., *Ccr2*, *Stat1*)<sup>21–23</sup>, as well as high levels of proinflammatory cytokines and chemokines (e.g., *Nfkb1*) mainly after IR<sub>17Gy</sub>, and to a lesser extent after IR<sub>10Gy</sub> (Fig. 4f). Moreover, the upregulation of these inflammatory genes seemed to specifically occur within the IM\_C3 cluster (Fig. 4f, h). These results suggest the emergence of a pro-inflammatory M1 population of IM during the fibrogenic phase at 4 and 5 M after IR<sub>17Gy</sub> (Fig. 4f). Similar DEG analysis of the AMs showed the upregulation of genes related to a profibrotic response (e.g., *Lpl*), as well as other markers known to be



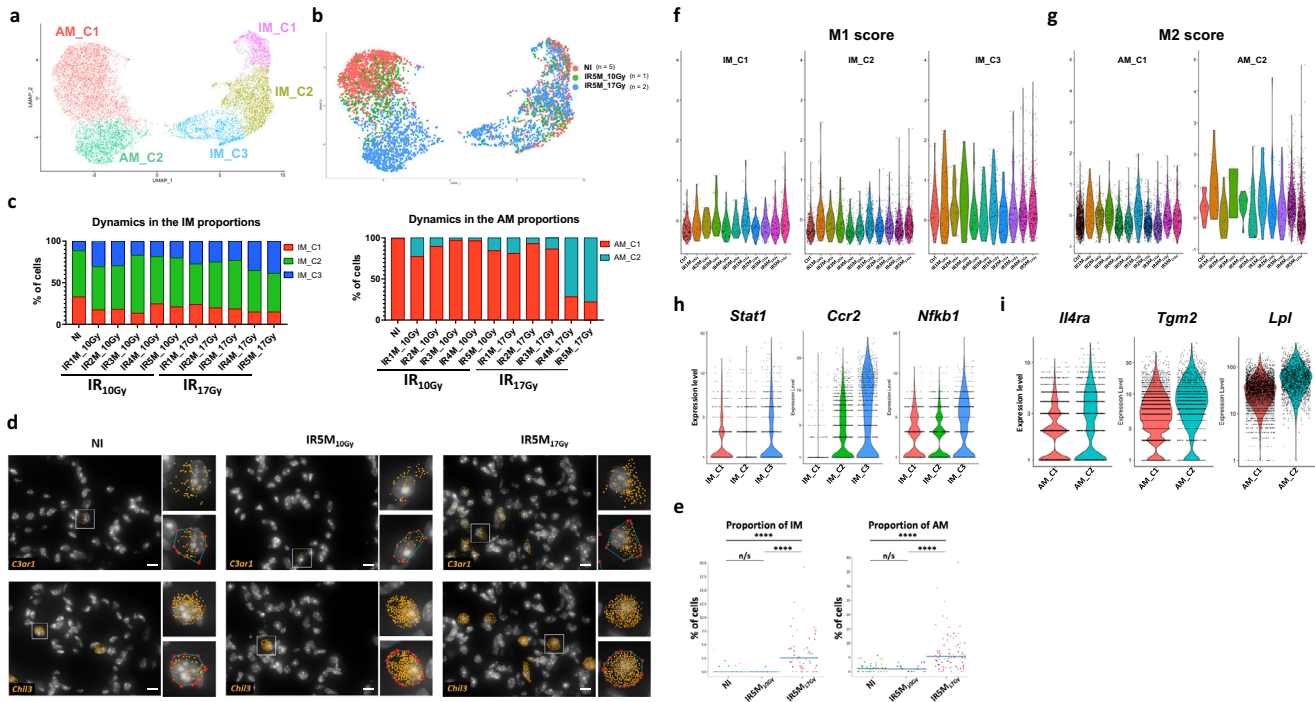
characteristic of the macrophage M2 activated phenotype (e.g., *Tgm2*)<sup>21</sup>, along with Th2 cytokines receptors (e.g., *Il4ra*) (Fig. 4g, i) and genes that have been related to foam cells (e.g., *Cd63*, *Nr1h3*, *Abcg1*) (Supplementary Fig. 4d). These profibrotic markers were over-expressed in the AM\_C2 subpopulation, which was specifically enriched after IR, and became predominant 4-5 months after 17 Gy

irradiation, suggesting particular toxic role conditions under profibrotic conditions.

Strikingly, this specific transcriptional change of AM from AM\_C1 to AM\_C2 is accompanied by morphological changes detected in smFISH: the estimated volume of the AMs is increased 5 months after IR<sub>17Gy</sub> (from 459  $\mu\text{m}^3$  to 925  $\mu\text{m}^3$ ), while it remains constant after

**Fig. 3 | Myfibroblasts contribute to the ECM deposition after IR<sub>17Gy</sub>.** **a** UMAP visualization of 3488 cells from the different fibroblast subpopulations annotated by cell type. **b** UMAP visualization of NI (*n* = 5), IR5M<sub>10Gy</sub> (*n* = 1) and IR5M<sub>17Gy</sub> (*n* = 2) fibroblasts annotated by time point. **c** Dynamics in the proportion of the fibroblast subpopulations at the different time points after IR<sub>10Gy</sub> and IR<sub>17Gy</sub>. **d** Automatic *Pdgfra* (red) and *Hhip* (green) mRNA detection with Big-FISH in NI, IR5M<sub>10Gy</sub>, and IR5M<sub>17Gy</sub> lung tissue sections. Scale bars, 10 μm. **e** Quantification of the *Pdgfra*<sup>+</sup>, *Hhip*<sup>+</sup> and *Pdgfra*<sup>+</sup> *Hhip*<sup>+</sup> cells in the NI, IR5M<sub>10Gy</sub> and IR5M<sub>17Gy</sub> lung tissue sections.

To compare two groups, the *P* value was computed with the Mann–Whitney–Wilcoxon test (two-sided test) from *scipy* (*n/s*, adjusted *p* value >0.05; \*, adjusted *p* value <0.05; \*\*, adjusted *p* value <0.01; \*\*\*, adjusted *p* value <0.001; \*\*\*\*, adjusted *p* value <0.0001). Each dot represents one analyzed image. Each color per time point represents a different biological replicate (NI *n* = 3; IR5M<sub>10Gy</sub> *n* = 3; IR5M<sub>17Gy</sub> *n* = 5). **f** Violin plot showing the single cell score calculated based on the ECM expressed genes in the myfibroblasts, fibroblasts *Coll3a1*, and fibroblasts *Coll4a1*.



**Fig. 4 | Proinflammatory and profibrotic profile of alveolar and interstitial macrophages after fibrogenic doses of IR.** **a** UMAP visualization of 11,678 cells from the different IM and AM subpopulations annotated by cell type. **b** UMAP visualization of NI (*n* = 5), IR5M<sub>10Gy</sub> (*n* = 1) and IR5M<sub>17Gy</sub> (*n* = 2) IM and AM annotated by time point. **c** Dynamics in the proportion of the IM and AM subpopulations at the different time points after IR<sub>10Gy</sub> and IR<sub>17Gy</sub>. **d** Automatic *C3ar1* (orange) and *Chil3* (orange) mRNA detection with Big-FISH in NI, IR5M<sub>10Gy</sub> and IR5M<sub>17Gy</sub> lung tissue sections. Scale bars, 10 μm. **e** Quantification of the *C3ar1*<sup>+</sup> and *Chil3*<sup>+</sup> cells in the NI, IR5M<sub>10Gy</sub>, and IR5M<sub>17Gy</sub> lung tissue sections. To compare two groups, the *P* value was computed with the Mann–Whitney–Wilcoxon test (two-sided test) from *scipy*

(*n/s*, adjusted *p* value >0.05; \*, adjusted *p* value <0.05; \*\*, adjusted *p* value <0.01; \*\*\*, adjusted *p* value <0.001; \*\*\*\*, adjusted *p* value <0.0001). Each dot represents one analyzed image. Each color per time point represents a different biological replicate (NI *n* = 3; IR5M<sub>10Gy</sub> *n* = 3; IR5M<sub>17Gy</sub> *n* = 5). **f** Violin plot showing the single cell score calculated based on the M1 signature in the different IM subpopulations. **g** Violin plot showing the single cell score calculated based on the M2 signature in the different AM subpopulations. **h** Violin plots of M1 genes expression in the different IM subpopulations. **i** Violin plots of M2 genes expression in the different AM subpopulations.

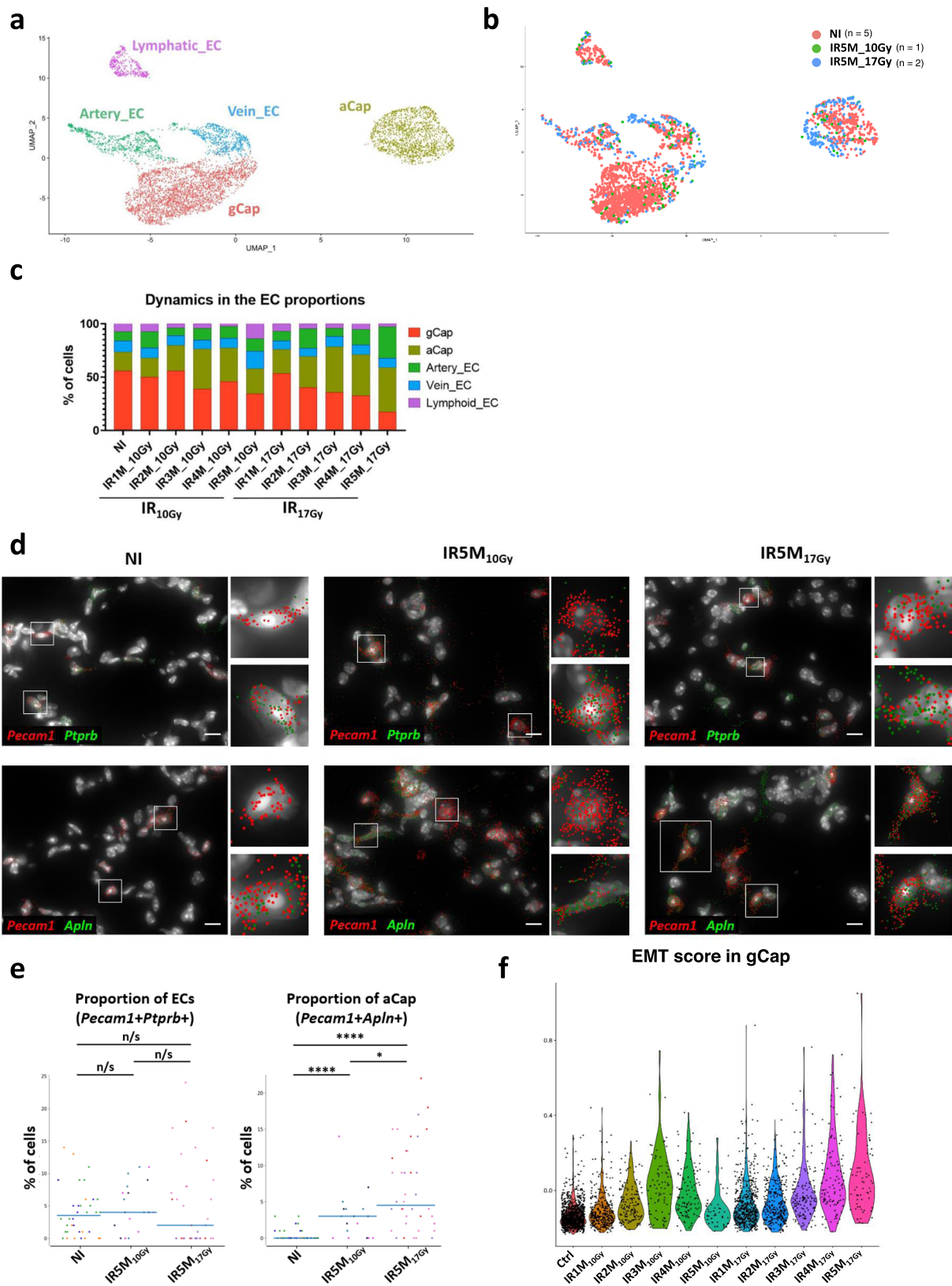
IR5M<sub>10Gy</sub> (Supplementary Fig. 4d). Moreover, the spatial distribution of AMs showed major changes in IR5M<sub>17Gy</sub>. While in the NI control and IR5M<sub>10Gy</sub> lungs individual AMs are localized within the alveolar compartment, enlarged AMs are gathered in clusters in the fibrotic tissue 5 months after IR<sub>17Gy</sub> (Supplementary Fig. 4d).

**Endothelial cells (ECs) undergo strong transcriptomic changes specifically after fibrogenic doses of IR**

Precise annotation of 6,482 ECs from the NI, IR<sub>10Gy</sub>, and IR<sub>17Gy</sub> samples led to the identification of 5 main compartments: lymphatic ECs, artery ECs, vein ECs, gCap and aCap (Fig. 5a, b). These compartments were defined by the expression of markers already described in the literature (Supplementary Fig. 5a): *Efnb2* and *Fbln5* for artery ECs; *Nr2f2* and *Vuf* for vein ECs; *Mmrn1*, *Fxyd6*, and *Fgl2* for lymphatic ECs, gCaps were defined by the expression of *Ptprb* and *Gpihbp1*, and aCap by the expression of *Fibin*, *Car4*, *Apln*, *Tmcc2*, and *Prx*. Examination of the evolution in the proportions of the different endothelial populations after IR revealed a progressive decrease in the proportion of gCap over

time after IR<sub>17Gy</sub> (from 55,8% in the NI to 17,3% at IR5M), together with an increase in the proportion of aCap (from 17,9% to 41,6%) (Fig. 5c). Interestingly, this event is not as pronounced in mice irradiated at IR<sub>10Gy</sub>.

Changes in cell proportions were validated by smFISH experiments that combined *Pecam1*, a canonical marker for endothelial cells, with either *Apln*, for the identification of aCap, or with *Ptprb*, which is expressed in all ECs except aCap and lymphatic ECs (Supplementary Fig. 5a, b). On the one hand, co-staining of NI and IR5M samples did not show any significant change in the proportion of *Pecam1*<sup>+</sup> *Ptprb*<sup>+</sup> cells 5 months after 10 Gy and 17 Gy IR (Fig. 5d, e). On the other hand, we could observe an increase in the proportion of aCap (*Pecam1*<sup>+</sup> *Apln*<sup>+</sup>) 5 months after IR, both at 10 and 17 Gy (Fig. 5d, e). Moreover, immunohistochemistry of NI and IR5M<sub>17Gy</sub> lung tissue samples with an *Apln* antibody confirmed this increase in aCap cells after radiation injury (Supplementary Fig. 5c). Therefore, these results confirm the increase in the aCap population observed by scRNA-seq.



DEG analysis of irradiated capillary ECs versus control revealed a very strong gene deregulation induced at 5 M after 17 Gy while the 5 M response induced by 10 Gy was quite limited (Supplementary Fig. 5d). When we looked for signatures, the 17 Gy irradiation response in gCap was associated with a large deregulation of genes involved in endothelial to mesenchymal transition (EndoMT) (e.g.,

*Col4a2*, *Col4a1*). We generated an EndoMT single-cell score based on the 200 genes of the EMT signature in the GSEA and observed a progressive increase of the score in the months following IR<sub>17Gy</sub>, reaching its maximum 5 M post-IR (Fig. 5f). On the contrary, after IR<sub>10Gy</sub> there was an initial increase up to 3 M post-IR and a consecutive decrease back to the NI levels (Fig. 5f). The EndoMT



**Fig. 5 | Characterization of the ECs after radiation injury.** **a** UMAP visualization of 6482 cells from the different EC subpopulations annotated by cell type. **b** UMAP visualization of NI ( $n = 5$ ), IR5M<sub>10Gy</sub> ( $n = 1$ ) and IR5M<sub>17Gy</sub> ( $n = 2$ ) ECs annotated by time point. **c** Dynamics in the proportion of the EC subpopulations at the different time points after IR<sub>10Gy</sub> and IR<sub>17Gy</sub>. **d** Automatic *Pecam1* (red), *Ptprb* (green), and *Apln* (green) mRNA detection with Big-FISH in NI, IR5M<sub>10Gy</sub>, and IR5M<sub>17Gy</sub> lung tissue sections. Scale bars, 10  $\mu\text{m}$ . **e** Quantification of the *Pecam1* + *Ptprb* + and *Pecam1* + *Apln* + cells in the NI, IR5M<sub>10Gy</sub> and IR5M<sub>17Gy</sub> lung tissue sections. To

compare two groups, the  $P$  value was computed with the Mann–Whitney–Wilcoxon test (two-sided test) from *scipy* ( $n/s$ , adjusted  $p$  value  $>0.05$ ; \*, adjusted  $p$  value  $<0.05$ ; \*\*, adjusted  $p$  value  $<0.01$ ; \*\*\*, adjusted  $p$  value  $<0.001$ ; \*\*\*\*, adjusted  $p$  value  $<0.0001$ ). Each dot represents one analyzed image. Each color per time point represents a different biological replicate (NI  $n = 3$ ; IR5M<sub>10Gy</sub>  $n = 3$ ; IR5M<sub>17Gy</sub>  $n = 3$ ). **f** Violin plot showing the single cell score calculated based on the EMT signature from the GSEA in the gCap at the different time points after IR<sub>10Gy</sub> and IR<sub>17Gy</sub>.

signature appears to be more robust in the gCap compartment than in the aCap (Supplementary Fig. 5e, f).

### Cell-cell interaction analysis underlines the importance of the collagen pathway in the evolution of RIPF

To predict which cellular alterations could be more directly implicated in the development of RIPF, we investigated the evolution of potential cell-cell communications in response to irradiation. We used CellChat<sup>24</sup> to pinpoint, based on the scRNA-seq data and the expression of ligands and receptors, cell types that could be interacting at a specific moment. First, we determined the number of potential existing interactions in the different conditions. This approach indicated an increase in cell-cell communications after IR, this increase being higher after 17 Gy than 10 Gy (Supplementary Fig. 6a). The analysis of the state of communication between major cellular compartments showed that the interactions between mesenchymal cells and endothelial cells increased over time, in particular at the latest time points (3 M, 4 M, and 5 M) after IR<sub>17Gy</sub> compared to NI (Supplementary Fig. 6b) and to IR<sub>10Gy</sub> (Fig. 6a).

Further refinement of this analysis suggested an increased communication between, on one side, fibroblasts *Col14a1* and myofibroblasts (acting as sources), and, on the other side, gCap cells (being the target) (Fig. 6b). Supplementary Fig. 6c illustrates the relative force of the registered changes affecting interactions between different cell compartments. Pathway communication analysis from fibroblasts *Col14a1* and myofibroblasts to gCap identified multiple pathways that were increased 5 months after IR<sub>17Gy</sub> compared to IR<sub>10Gy</sub> (Fig. 6c) (e.g., *Collagen*, *Fn1*, *Angptl*, *Vegf*). We further focused on the study of the collagen pathway as it shows a progressive increase in strength after 3 M in both mesenchymal cell types (Supplementary Fig. 6d). Next, we identified specific pairs of ligands and receptors that were increased after IR<sub>17Gy</sub> and absent in NI and IR<sub>10Gy</sub> at the latest time points after IR. This analysis revealed the ligands *Col1a1* and *Col1a2* to be upregulated in fibroblasts *Col14a1* and myofibroblasts and the receptor *Itga3* to be upregulated in the gCap (Fig. 6d, Supplementary Fig. 6e). These analyses support the role of the collagen pathway, more concretely through *Col1a1-Itga3* and *Col1a2-Itga3* interactions connecting mesenchymal cells (matrix fibroblasts and myofibroblasts) to capillary endothelial cells (gCap) during the development of RIPF (Fig. 6e).

### An interactive web-based interface to study lung responses to irradiation

This work describes some of the key features occurring after radiation injury in the lung at the cellular and molecular level. Nevertheless, due to a large amount of generated data, not all the cell types were described, nor all the molecular alterations that occur during fibrogenesis were studied. For this reason, we have built a dedicated website that is accessible to the scientific community, so that anyone can explore our murine single-cell atlas of the lung response to radiation injury and use it for their own research (Supplementary Fig. 7). This open-access website ([https://lustra.shinyapps.io/Murine\\_RIPF\\_Atlas/](https://lustra.shinyapps.io/Murine_RIPF_Atlas/)), built using the R package ShinyCell<sup>25</sup>, allows the investigation of all the different lung cell populations, as well as the changes in gene expression after the different time points and doses of IR.

The website offers a UMAP visualization of the data in which the different metadata parameters can be represented, e.g., the main cell compartments (epithelial, myeloid, mesenchymal, lymphoid, and endothelial cells), the 21 described cell types (Fig. 7a) as well as 31 precise sub-cell types, the different conditions (Ctrl, IR1M<sub>10Gy</sub>, IR2M<sub>10Gy</sub>, IR3M<sub>10Gy</sub>, IR4M<sub>10Gy</sub>, IR5M<sub>10Gy</sub>, IR1M<sub>17Gy</sub>, IR2M<sub>17Gy</sub>, IR3M<sub>17Gy</sub>, IR4M<sub>17Gy</sub>, IR5M<sub>17Gy</sub>), the different time points (Ctrl, 1M, 2M, 3M, 4M, 5M) and the dose of IR (NI, 10 Gy, 17 Gy). Moreover, this first *Metadata vs GeneExpr* tab allows the users to visualize both cell metadata and gene expression side-by-side on low-dimensional representations (Fig. 7b), which allows a direct search of the expression of different genes in the different cell compartments. Moreover, the visualization of two cell metadata or two gene expressions side by side on low dimensional representations is also possible with the *Metadata vs Metadata* and *GeneExpr vs GeneExpr* tabs respectively.

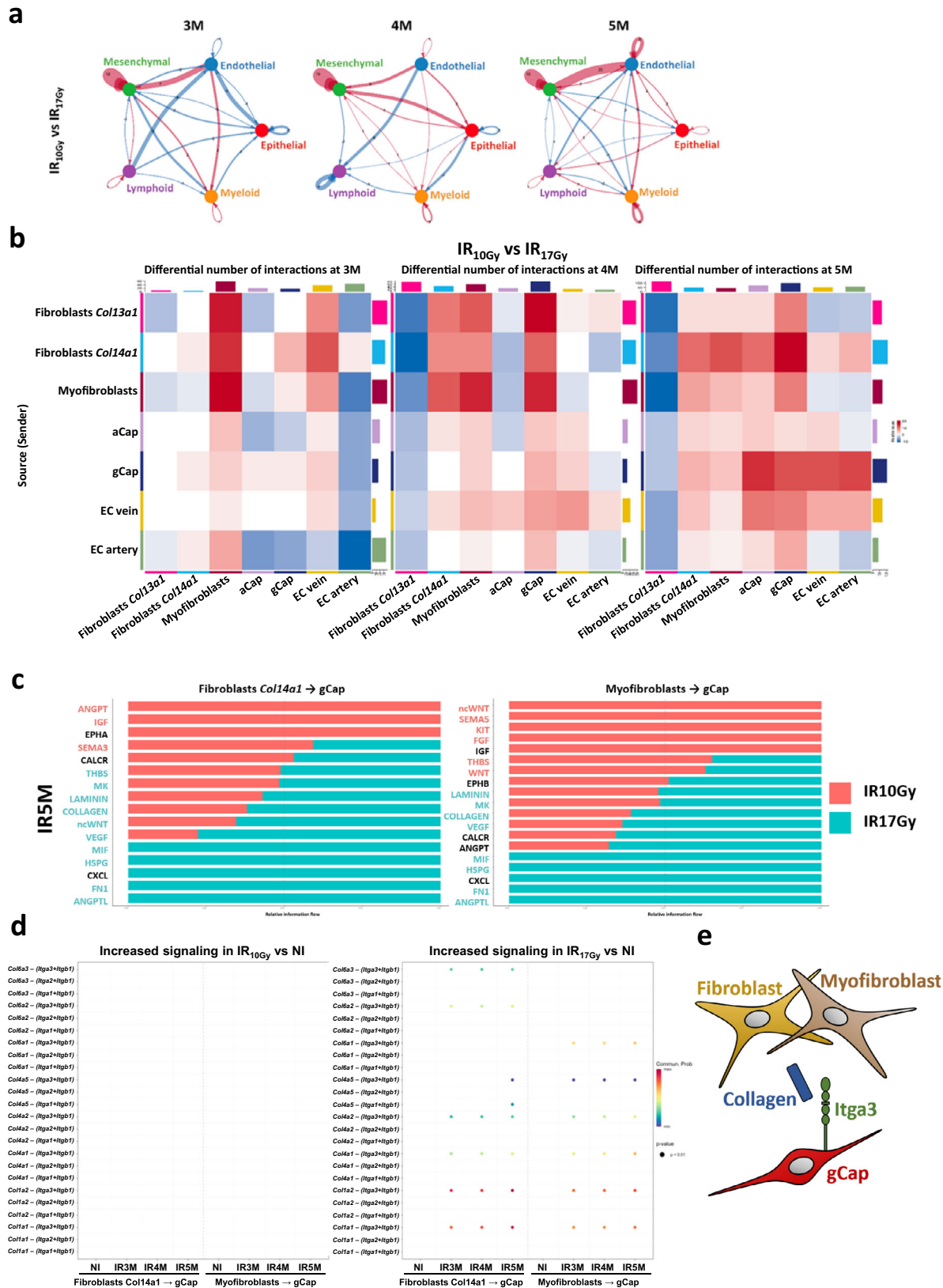
This site also allows a straightforward exploration of the differences in cell proportions using the *Proportion plot* tab, in which the proportions of the main cell compartments, the different cell types (Fig. 7c), and sub cell types along the different conditions (i.e., different time points and doses of IR) can be examined. Another interesting feature that this tool offers is the visualization of the co-expression of two different genes in a single UMAP representation—*Gene coexpression* tab—(Fig. 7d), providing also the number and percentage of cells that express both, none, or only one of the genes. In addition, four different visualization methods are available to study the gene expression: Violin plot (Fig. 7e), boxplot, bubble plot (Fig. 7f), and heatmap (*Violinplot/Boxplot* tab and *Bubbleplot/Heatmap* tab), which allow the comparison of the expression of genes along the different conditions, thus, providing insightful information about radiation-induced fibrosis. These multiple options allow a wide range of methods for data visualization that can be adapted to the user's needs.

In conclusion, this web-based interface presents a rich dataset that permits the scientific community to investigate the response of the lung to radiation injury for their features of interest with no need of previous bioinformatic knowledge. Raw data is also available at GSE211713 for deeper analysis.

## Discussion

We have built a murine single-cell atlas illustrating the early and late responses of the lung to radiation. The dataset contains more than 100,000 cells from 20 different lungs from non-irradiated as well as IR<sub>10Gy</sub> and IR<sub>17Gy</sub> mice. Cell type annotations based on published studies identify the 21 main lung cell populations already described using scRNA-seq approaches<sup>26–28</sup>. This study revealed, in each lung population, the progressive transcriptional changes occurring in the months following radiation injury from the acute inflammatory phase to the development of pulmonary fibrosis after exposure to fibrogenic dose of radiation. The comparison of the molecular alterations induced by a non-fibrogenic (IR<sub>10Gy</sub>) versus a fibrogenic dose (IR<sub>17Gy</sub>) allowed to uncover the physiopathological features of radiation injury in the lung.

This time-course single-cell RNA-seq analysis reproduced the classical features of pulmonary fibrosis such as the loss of AT2 cells, the expansion of myofibroblasts as well as the accumulation of foamy macrophages fostering a pro-inflammatory environment<sup>29–32</sup>.



In particular, the pro-fibrotic AM subcluster that we identified under fibrogenic conditions (AM\_C2 in our data) may share similarities with a transitional profibrotic macrophage subgroup found in diseased mice using the bleomycin-induced-lung injury model<sup>33</sup>. These transitional macrophages showed an expression profile intermediate between monocyte-derived macrophages and AM that are localized in the

fibrotic niche. Moreover, the pro-fibrotic AM expresses genes that have been related to a “foam” phenotype, which corresponds to lipid-laden polarized M2 macrophages that accumulate oxidized phospholipids. These foamy macrophages have been found in different mouse models of lung-induced injury as well as in patients with fibrotic lung diseases<sup>30</sup>. Interestingly, our spatial smFISH analysis in lung tissue after

**Fig. 6 | Cell-cell interaction analysis between the different lung cell populations after different time points and doses of IR.** **a** Circle plot showing the differential number of interactions between IR<sub>10Gy</sub> and IR<sub>17Gy</sub> in the main cellular compartments at 3 M, 4 M, and 5 M post-IR: mesenchymal, endothelial, epithelia, myeloid and lymphoid. Red (or blue) colored edges represent increased (or decreased) signaling in the IR<sub>17Gy</sub> compared to the IR<sub>10Gy</sub>. **b** Heatmap showing the differential number of interactions between IR<sub>10Gy</sub> and IR<sub>17Gy</sub> in the endothelial and mesenchymal subpopulations at 3 M, 4 M, and 5 M post-IR. Red (or blue) represents increased (or decreased) signaling in the IR<sub>17Gy</sub> compared to the IR<sub>10Gy</sub>. The top-colored bar plot represents the sum of column of values displayed in the heatmap

(incoming signaling). The right-colored bar plot represents the sum of row of values (outgoing signaling). **c** Bar graph showing significant signaling pathways ranked based on differences in the overall information flow within the inferred networks between IR5M<sub>10Gy</sub> and IR5M<sub>17Gy</sub> from the Fibroblasts *Col14a1* and Myofibroblasts (sources) to the gCap (targets). The top signaling pathways colored red are enriched after IR5M<sub>10Gy</sub>, and the ones colored green were enriched after IR5M<sub>17Gy</sub>. **d** Increased signaling ligand-receptor pairs of the Collagen pathway in IR<sub>10Gy</sub> and IR<sub>17Gy</sub> compared to NI at 3 M, 4 M, and 5 M after IR. **e** Schematic drawing of the intercellular communication between fibroblasts and myofibroblasts with the gCap through the collagen pathway.

radiation injury indicates that these foamy macrophages tend to aggregate in fibrotic foci with an increased cell volume and a high level of endogenous fluorescence. These observations corroborate with data from other groups showing that clusters of M2 macrophages are detected after IR, contribute to fibrosis development, and are controlled by adenosine signaling<sup>34,35</sup>.

Concomitantly to the bronchiolization and the destruction of alveoli observed during fibrosis development, the progressive disappearance of AT2 is associated with profound changes in the transcriptome of remaining AT2 cells. Apart from an increased expression of genes associated with EMT, we identified, among the AT2 population after IR, an upregulation of AT1 marker genes, suggesting that AT2 cells are differentiating into AT1 cells. Detailed analysis confirmed an enrichment in Krt8+ transitional AT2 cells, known to support alveolar regeneration<sup>17</sup>. The fact that AT2 cells are known to give rise to AT1 cells<sup>36</sup> and that, after lipopolysaccharide (LPS)-induced lung injury, AT2 cells activate a similar transdifferentiation program<sup>16</sup> indicate that a common regenerative process is activated after IR to restore normal alveolar architecture and lung function. Considering that the transdifferentiation program is turned off in the months following exposure to a nonfibrogenic dose of radiation whereas, after a fibrogenic dose, AT2 cells maintain and reinforce this transdifferentiation program, it is tempting to speculate that failure to differentiate into functional AT1 cells leads to AT2 cells exhaustion, contributing to alveoli destruction and fibrosis development. However, more in-depth analysis of AT2 after radiation injury is required to demonstrate a direct physiopathological implication in fibrosis development after IR.

Endothelial cells are particularly affected by radiation<sup>37</sup>. Our single-cell RNA-seq analysis showed that endothelial cells activate an EndoMT program in the months following exposure to radiation. It is known that EndoMT participates in many human fibrotic disorders<sup>38</sup> and a previous study identified hypoxia and TGF $\beta$  as EndoMT inducers after radiation injury<sup>39</sup>. It has been suggested by Hashimoto and colleagues that endothelial cells can give rise to myofibroblasts in the mouse model of pulmonary fibrosis induced by bleomycin<sup>40</sup>. Interestingly, cell-cell communication analysis after radiation injury pointed to increased signaling between endothelial cells and stromal cells, particularly myofibroblasts. These results support the idea that radiation triggers EndoMT phenotype coupled to increased interactions with myofibroblasts, suggesting that endothelial cells may contribute as a source in the expansion of myofibroblasts during the development of RIFP.

In summary, this study describes the transcriptional changes occurring at the single-cell level in the lung in the months following radiation injury. This large and comprehensive dataset provides a starting resource to explore and characterize molecular mechanisms underlying the different physiological and pathological steps occurring after IR (e.g., from acute wound healing to fibrosis development). To facilitate data visualization and interrogation, a user-friendly web-based interface is accessible to the scientific community without the need for specific computational skills.

## Methods

### Mice and ethics statement

Studies were performed in accordance with the recommendations of the European Community (2010/63/UE) for the care and use of laboratory animals. Experimental procedures were specifically approved by the ethics committee of the Institut Curie CEEA-IC #118 (Authorization number APAFIS#5479-201605271 0291841 given by the National Authority) in compliance with the international guidelines. Females C57BL/6J mice purchased from Charles River Laboratories at the age of 6 weeks were housed in Institut Curie animal facilities.

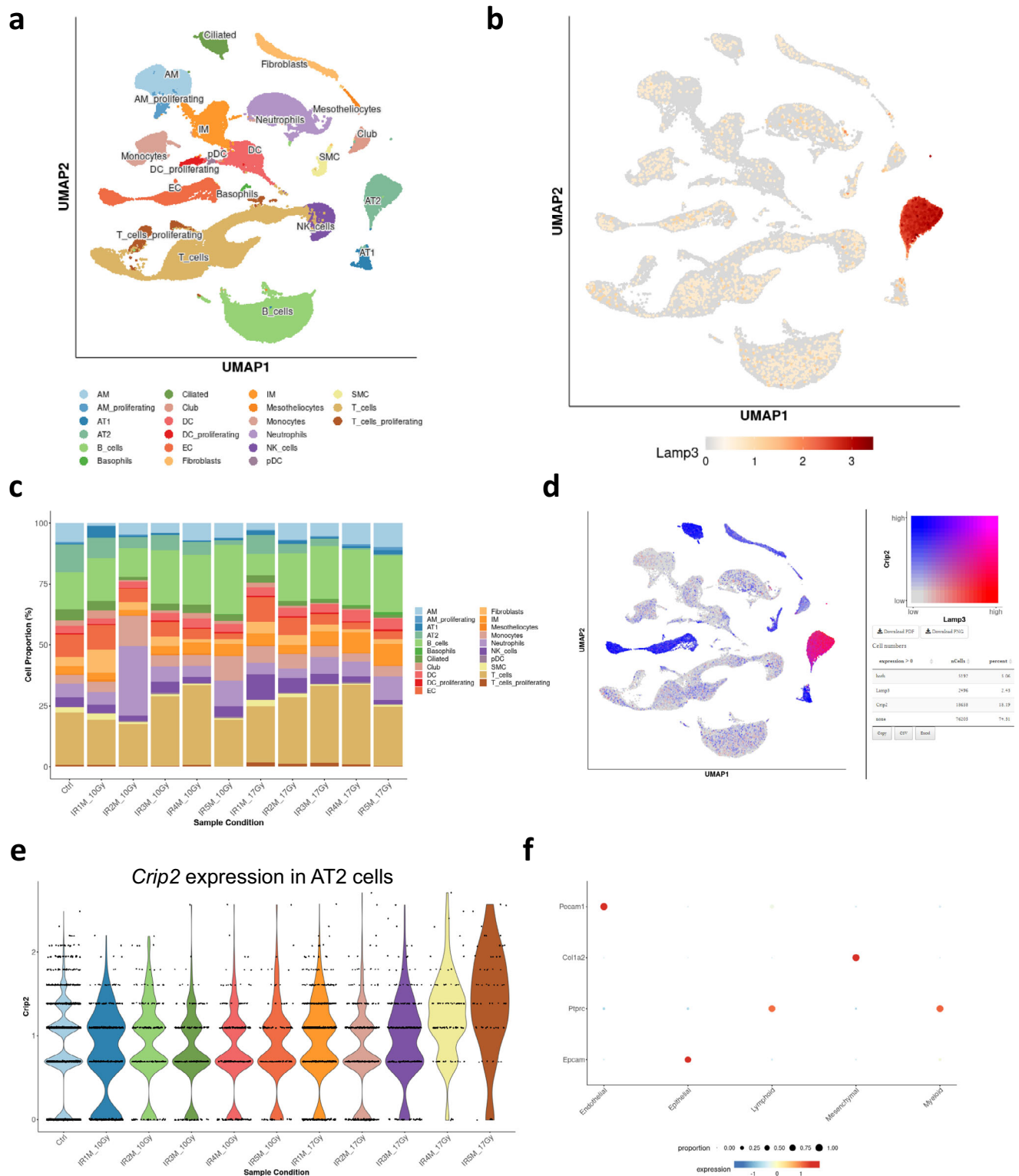
### Radiation injury

Here, the classical C57BL/6J female mouse model of lung radiation toxicities were used<sup>11</sup>. Collimation, time-resolved fluence measurement, chemical dosimetry, depth-dose distribution, anesthesia of the mouse, mouse immobilization, and irradiation of mouse thorax were carried out as previously described<sup>41,42</sup>. Briefly, mice were anesthetized with a nose cone using 2.5% isoflurane in air, without adjunction of oxygen, immobilized in a dorsal position, and set in a vertical position at 500 mm of the electron source. Then, mice were exposed to bilateral thorax irradiation with a dose of 10 or 17 Gy at the age of 10–12 weeks using the 4.5-MeV linear electron accelerator facility (Kinatron).

In order to determine the level of pulmonary fibrosis in mice, the lung was imaged with three-dimensional X-ray on the cone beam computed tomography (CBCT) module of the Small Animal Radiation Research Platform (SARRP, Xstrahl). 1.5–2% isoflurane was used to anesthetize the mice, which were maintained on a PMMA vertical stand in the vertical upright position. The 3D reconstruction of the images was calculated from 1,440 projections using the integrated software Murislice (XStrahl). Then, ImageJ/FIJI (ImageJ, NIH, Bethesda, MD) was used to reconstruct the slices. This analysis allowed to determine the level of fibrosis of the mice. Further details can be found in ref. 42.

### Lung tissue dissociation

Mice were killed by cervical dislocation and the ribcage was opened to clear the trachea. Mouse trachea was perfused with 1.5 ml of 50 U/ml dispase (Serlabo, WO-LS02100; Sigma Corning, DLW354235) using a 20 G needle, followed by 0.5 ml of 1% agarose (Invitrogen, 15510-027) to block the exit of the dispase. Lungs were resected, minced with a scalpel into small pieces, and added into 3 ml of 1 $\times$  DPBS MgCl<sub>2</sub><sup>+</sup> and CaCl<sub>2</sub><sup>+</sup> (Gibco, 14040-091). Then 320  $\mu$ l of 25 U/ml elastase (Worthington, LS002292) were added and the suspension was homogenized and incubated for 30 min at 37 °C with orbital shaking. Enzymatic activity was inhibited with 5 ml of PF10 (1 $\times$  DPBS containing 10% fetal bovine serum (FBS)) and 20  $\mu$ l of 0.5 M EDTA pH 8 (Invitrogen, AM9260G). Cell suspension was filtered through 100  $\mu$ m nylon cell strainer (Fisher Scientific, 22363549), which was rinsed with 5 ml of PF10. This was followed by 37.5  $\mu$ l of 10 mg/ml DNase I (Sigma, D4527-40KU) treatment and incubation on ice for 3 min. Cell suspension was filtered again through a 40  $\mu$ m nylon cell strainer (Fisher Scientific, 087711) and 5 ml of PF10 were added to rinse it. Samples were centrifuged for 6 min at 150 g and 4 °C, pellet was resuspended in red blood cell (RBC) lysis buffer (Roche,



**Fig. 7 | Web-based interface for the murine single-cell atlas of the lung response to radiation injury. a** UMAP visualization of 102,869 cells from 20 different samples (5 NI; 5 IR<sub>10Gy</sub>, one per time point; 10 IR<sub>17Gy</sub>, two per time point) annotated by cell type. **b** UMAP visualization of the expression of *Lamp3*. **c** Dynamics in cell proportions of the main cell types across the NI and IR

conditions at the different time points and doses. **d** UMAP visualization of the expression of *Lamp3* (red), *Crip2* (blue), and the co-expression of both (pink). **e** Violin plot of *Crip2* expression in the AT2 cells in the NI samples and at the different time points after IR<sub>10Gy</sub> and IR<sub>17Gy</sub>. **f** DotPlot of the expression of the marker genes used to identify the main 5 cell compartments.

11814389001) and incubated for 90 s at room temperature (RT) before stopping the lysis with 6 ml of PF10. Then, 500 µl of pure FBS were placed at the bottom of the sample, prior to a final centrifugation for 6 min at 150 × g and 4 °C. The pellet was resuspended

in 1 ml of 1× DPBS containing 0.02% bovine serum albumin (BSA) (Sigma, D4527-40KU), and cell counting was done in a Malassez. Finally, concentration of the samples was adjusted to 1 million cells/ml in 1× DPBS containing 0.02% BSA.

### Droplet based scRNA-seq (10x GENOMICS)

Single-cell 3'-RNA-Seq samples were prepared using single cell V3.1 reagent kit and loaded in the Chromium Controller according to standard manufacturer protocol (10x Genomics, PN-120237) to capture 6,000 cells. Briefly, dissociated lung single cells are encapsulated in nanodroplets (GEMs) using a microfluidic device. These GEMs are generated combining barcoded single cell 3' V3.1 gel beads, a master mix that contains the reverse transcription (RT) reagents, the single cells, and partitioning oil onto the Chromium Next GEM Chip. After cell lysis, RNAs are captured on the gel beads coated with oligos containing an oligo-dTTT, unique molecular identifiers (UMIs) and a specific barcode.

Incubation of the GEMs produces barcoded, full-length cDNA from poly-A mRNA. After reverse transcription, GEMs are broken and cDNAs are purified with silane magnetic beads. Then, barcoded full-length cDNA is amplified by PCR to generate enough material for library construction. Amplified cDNA is purified again, and cDNA quality control was assessed by capillary electrophoresis (Bioanalyzer, Agilent) before the preparation of the libraries.

Finally, libraries are prepared using a fixed proportion of the total cDNA. Enzymatic fragmentation and size selection are used to optimize the cDNA amplicon size. During the GEM incubation, the read 1 primer sequence is added to the molecules. At this step, P5, P7, a sample index, and the read 2 primer sequence are added via End Repair, A-tailing, Adaptor Ligation, and PCR. This way, the final libraries contain the P5 and P7 primers used in Illumina bridge amplification. Finally, libraries were sequenced on a NovaSeq sequencer (Illumina). Each measurement comes from independent NI, IR<sub>10Gy</sub>, or IR<sub>17Gy</sub> mice.

### scRNA-seq data analysis

First, raw sequencing was processed using the 3.0.2 Cell ranger pipeline (10x Genomics). Briefly, Illumina sequencing files (bcl2) were demultiplexed and mapped onto the mm10 reference genome. This allows the creation of a count matrix table for each of the samples. Then, the count matrices were individually loaded in R (4.0.5) and analyzed using Seurat package v4.0.1<sup>43</sup>.

In order to remove the contamination from the “soup” of cell-free RNAs, we used SoupX<sup>44</sup>. This tool is able to remove ambient RNA contamination from droplet-based scRNA-seq experiments. Briefly, first SoupX estimates the mRNA expression profile from empty droplets. Then it estimates the contamination fraction and the fraction of UMIs originating from the background in each cell. Finally, it corrects the expression of each cell using the ambient mRNA expression profile and estimated contamination.

SoupX matrices were imported in Seurat, and a Seurat object was created for each of the samples with the function CreateSeuratObject. Samples were merged and the object was normalized using the SCTransform normalization function, a normalization and variance stabilization method using regularized negative binomial regression<sup>45</sup>. Then, PCA was performed and the first 20 PCs were selected (based on inspection of PC elbow plot) as input for RunUMAP function for dimension reduction and visualization with the Uniform Manifold Approximation and Projection (UMAP) dimensional reduction technique<sup>46</sup>. Finally, cells were clustered with the functions FindNeighbors and FindClusters, based on the default Seurat parameters with resolution parameter set to 0.8. In a first step, we performed general QC: gene counts matrices were filtered and cells with nCounts <200 and nFeatures >6000 RNA molecules sequenced, as well as percentage of mitochondrial genes >15% were removed. In a second step, each cell population was cleaned in more details by removing the cells that presented a high percentage of mitochondrial genes, a low nCounts, or that expressed markers characteristic of other cell types, which suggests the presence of doublets. All quality controls are present on the web interface.

Then, cell-type marker genes for each cluster were identified using the function FindAllMarkers of Seurat. Cell clusters were named based on cell type-specific markers from recently published scRNA-seq datasets<sup>12–14</sup>. For each of the analyzed clusters, we performed differential expression analysis using the function FindMarkers with the MAST package (two-sided test), which identifies differentially expressed genes between two groups of cells using a hurdle model tailored to scRNA-seq data<sup>47</sup>. We studied the genes with a logFC threshold of >1 or <-1 and an adjusted p-value threshold of <0.05 using the Gene Set Enrichment Analysis (GSEA) computational method, which defines if a set of genes shows statistically significant, concordant differences between two states.

Violin plots and Heatmaps were generated with Seurat for specific significant differentially expressed genes. To calculate the single cell data score, first a dataset corresponding to a cell state or biological process is chosen. Then, the table of expression of these genes by each cell is extracted. For each gene, the expression data is centered and normalized. Finally, for each cell, we computed the mean of the centered normalized gene expression. This gave us an expression score for each cell which was then plotted with UMAP. The trajectory analysis of the AT1 and AT2 cells was performed using Monocle 3<sup>48</sup>. The pseudotime analysis of the AT1 and AT2 cells was done with Monocle 2<sup>49</sup>, where the marker genes of the AT2 cluster 3 were used to order the cells. The gene regulatory network analysis was constructed on the AT2 using SCENIC<sup>50</sup>. Lastly, all cell-cell interactions analysis were done using CellChat<sup>24</sup>.

### Mouse lung tissue processing for smFISH

Mice were injected with a mix of 100 mg/kg ketamine and 10 mg/kg xylazine and we waited for the mice to be fully asleep. Then, the ribcage was opened, the heart was perfused with 10 ml of cold 1× PBS pH 7.4 (Invitrogen, AM9624) and it was resected just after the perfusion was finished. Then, the mouse trachea was perfused with cold 4% paraformaldehyde (PFA) (Euromedex, 15714-S) until the lungs were fully expanded with no air left inside. The trachea was closed with a thread to avoid PFA leakage. The lungs were resected out of the ribcage and kept in a falcon with cold 4% PFA overnight (o/n) under rotation at 4 °C. After fixation, the 5 lobes were separated and kept individually in cold 1× PBS containing 30% sucrose (Sigma, S7903) during 6 h under rotation at 4 °C. Lobes were rinsed in cold 1× PBS and pre-embedded in cold 50% optimal cutting temperature (OCT) compound (VWR, 411243) diluted in 1× PBS during 30 min under rotation at 4 °C. Finally, each lobe was embedded in square embedding molds (VWR, POLS18646ACODE45) containing OCT, frozen in dry ice during 20 min and stored at -80 °C.

### smFISH probes design and preparation

The design of smFISH probes was performed with the R-package Oligostan<sup>51</sup>. For a given target mRNA, Oligostan outputs a list containing all the potential probe sequences fulfilling these requirements: length between 26 nt and 32 nt, score around  $\Delta G_{37^\circ\text{C}}$  value of 90%, minimal distance between probes of 2 nt, GC percentage between 40 and 60%, 5 different criteria for probe composition (nucleotide composition in A < 28%, no AAAA stacks, C nucleotide composition between 22 and 28 %, no CCCC stacks in any six consecutive nucleotides in the first 12 positions and no 4 nonconsecutive C in any 6 consecutive nucleotides in the first 12 positions). To each RNA-specific sequence, a shared readout sequence (FLAP Y, 27 nt) was added, and probes with the highest scores were selected. To obtain a good smFISH signal, around 30 probes per RNA were used (*Lamp3* (31), *Chil3* (32), *C3ar1* (32), *Pdgfra* (32), *Hhip* (32), *Pecam1* (32), *Ptprb* (32), *Apln* (32), *Fibin* (28), *Prx* (32), *Tmcc2* (29))<sup>51</sup>.

For visualization, we used secondary probes carrying two fluorophores (either Cy3 or Cy5) on both ends through 5' and 3' amino modifications: FLAP Y-Cy3 (/5Cy3/AA TGC ATG TCG ACG AGG TCC

GAG TGT AA/3Cy3Sp/) and FLAP Y-Cy5 (/5Cy5/AA TGC ATG TCG ACG AGG TCC GAG TGT AA/3Cy5Sp/).

smFISH primary probes and secondary probes (fluorescent FLAP probes) were produced and bought from Integrated DNA Technologies (IDT). Primary probes are delivered frozen in 96-well plates at a final concentration of 100  $\mu$ M in Tris-EDTA pH 8.0 (TE) buffer. An equimolar mixture of the primary probes was prepared per set of probes and diluted to 20  $\mu$ M in TE buffer pH 8 (Invitrogen, AM9849). Secondary fluorescent FLAP probes are delivered lyophilized. They were resuspended in TE buffer at a final concentration of 100  $\mu$ M. Primary and secondary probe stocks were stored at  $-20^{\circ}\text{C}$ .

### smFISH on lung tissue sections

OCT-embedded mouse lung lobes were cut into 16  $\mu$ m tissue sections in a cryostat (Leica CM 1950) and mounted in previously cleaned and coated coverslips. Briefly, extensive cleaning was achieved by washing the coverslips (Menzel-Glaser, 20  $\times$  20, #1) three times in ethanol (VWR, 20821.310), acetone (Honeywell, 32201), and water. Then, they were sonicated in 1M KOH (Honeywell, 06005) in H<sub>2</sub>O, rinsed with water and dried in an oven at 70  $^{\circ}\text{C}$  for 10 min. For tissue attachment, coverslips were then coated with 2% (3-aminopropyl)triethoxysilane (APTS) (Sigma, A3648-100ML) in H<sub>2</sub>O for 2 min and rinsed with water twice. Finally, the coating was activated in the oven at 70  $^{\circ}\text{C}$  for 60 min. Coverslips with lung tissue sections were placed into 6 well plates, fixed with cold 4% PFA for 15 min at 4  $^{\circ}\text{C}$ , washed twice with cold 1 $\times$  PBS, and kept in 70% ethanol at 4  $^{\circ}\text{C}$  o/n.

smFISH probes were generated by pre-hybridizing the primary probes with the secondary FLAP Y (FLAP Y-Cy3 or FLAP Y-Cy5) probes in a PCR machine. Briefly, a mix containing 40 pmol of the primary probes, 50 pmol of the FLAP Y probe, 1 $\times$  NEBuffer 3 (BioLabs, B7003S) and ultra-pure water was incubated in a thermocycler for 3 min at 85  $^{\circ}\text{C}$ , 3 min at 65  $^{\circ}\text{C}$  and 5 min at 25  $^{\circ}\text{C}$ . The hybridization mix was prepared by mixing 2  $\mu$ l of the smFISH probes and 98  $\mu$ l of hybridization buffer (100 mg/ml dextran sulfate (Sigma, D8906-10G) and 10% deionized formamide (Invitrogen, AM9342) in 2 $\times$  SSC (Invitrogen, AM9763)) per sample. Hybridization buffer can be prepared in advance and frozen in aliquots. These aliquots were thaw and heated to 100  $^{\circ}\text{C}$  for 5 min, and let cool to RT. In case of co-staining, 96  $\mu$ l of hybridization buffer was mixed with 2  $\mu$ l of primary probes hybridized to FLAP Y-Cy3 and 2  $\mu$ l of primary probes hybridized to FLAP Y-Cy5.

For the hybridization of the smFISH probes in the tissue, first tissue sections were re-hydrated twice with washing buffer I (WBI) (2 $\times$  SSC) for 3 min at RT, followed by a last incubation in washing buffer II (WBII) (10% deionized formamide in 2 $\times$  SSC) for 3 min at RT. The hybridization of the smFISH probes in the lung tissue sections was carried out in a hybridization chamber. Here, the tissue sections were placed upside-down onto a 100  $\mu$ l droplet of the hybridization mix and incubated at 37  $^{\circ}\text{C}$  o/n.

To finish, tissue sections were washed three times: first in pre-heated WBII at 37  $^{\circ}\text{C}$  for 30 min, second in WBII containing 0.05  $\mu$ g/ $\mu$ l DAPI (Sigma, D9542-1MG) at 37  $^{\circ}\text{C}$  for 30 min, and third in 1 $\times$  PBS for 5 min at RT. Samples were mounted in SlowFade Diamonds (Invitrogen #S36963) and stored at 4  $^{\circ}\text{C}$  until image acquisition.

### Image acquisition

For each sample, 54 three-dimensional image stacks were captured with an interval of 0.3  $\mu$ m (total of 16.2  $\mu$ m) on a widefield microscope (Upright Widefield Apotome Zeiss) equipped with a 63  $\times$  1.4 numerical aperture (NA) objective and a CCD camera (CoolSNAP HQ2) and controlled with ZEN microscope software (ZEISS Microscopy). Three lasers were used to excite DAPI (excitation time 40 ms) and smFISH probes labeled with Cy3 (excitation time 300 ms) and Cy5 (excitation time 300 ms). When acquiring tiles with a 5  $\times$  5 FOV, 16 three-dimensional image stacks were captured with an interval of 0.3  $\mu$ m (total of 4.8  $\mu$ m).

### Image analysis

Image analysis and result generation are performed automatically with a provided command line script performing the steps described next.

Nuclei were automatically segmented with Cellpose<sup>52</sup> and RNA detection was performed with FISH quant v.2<sup>53</sup>. Points detected in several channels at the same location were considered to be the result of auto-fluorescence and therefore removed.

Custom written Python scripts were used to determine the cell-type of a nucleus. Importantly, marker genes were chosen such that each cell type is identified with only one marker gene. First, RNAs of each gene were clustered with the algorithms OPTICS and DBSCAN<sup>54</sup> in their implementation in scikit-learn (<https://scikit-learn.org/>). We used the following functions and parameters:

```
sklearn.cluster.OPTICS is used to determine the core
points and ordering. Parameters are min_sample=4 and min_
cluster_size=4.
sklearn.cluster.cluster_optics_dbscan is used for the
clustering, with the parameters obtained in the previous step. The
parameter eps was manually adjusted for each gene to account for
differences in the expression levels. Other parameters are left to
their default value.
```

The next step consisted in assigning the detected RNA clusters to the individual cells (or nuclei, as there is no cytoplasmic marker). For this, we calculated the convex hull for each individual cell with Delaunay tessellation (using spatial.Delaunay from scipy). In order to assign a convex hull to a nucleus, we required that there was at least an overlap of  $K\%$ , where  $K$  was set individually for different cell types (20 to 55%) to account for different cell morphologies. Non-assigned convex hulls were removed from the analysis (this typically happened when nuclei were not present in the image). The proportion of positive cells for a specific cell type marker in an image was calculated as the number of positive nuclei of this marker over the total number of nuclei segmented in the image. To compare two groups (ex NI vs IR5M), the  $P$  value was computed with the Mann-Whitney-Wilcoxon test (two-sided test) from scipy ( $*P < 0.05$ ;  $**P < 0.01$ ;  $***P < 0.001$ ,  $****P < 0.0001$ ). Cell volume was estimated as the volume of the convex hull associated to a nucleus (Fig. 2b). If a point cloud contained several nuclei (for instance for neighboring cells of the same type), the average cell volume per nuclei was calculated. For the violin plots, each independent NI, IR<sub>10Gy</sub>, or IR<sub>17Gy</sub> mice are represented with different colors, and dots of the same color represent different analyzed images from the same mouse sample.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The scRNA-seq datasets generated during the current study have been deposited in the Gene Expression Omnibus (GEO) repository, with the accession code [GSE211713](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE211713). Processed data can be explored through an interactive web interface ([https://lustra.shinyapps.io/Murine\\_RIPF\\_Atlas/](https://lustra.shinyapps.io/Murine_RIPF_Atlas/)). Source data are provided with this paper.

### Code availability

We make the full image analysis Python code available on the following Github page [https://github.com/tdefa/cell\\_type\\_calling\\_2channels](https://github.com/tdefa/cell_type_calling_2channels) and on Zenodo <https://zenodo.org/record/7360791> (<https://doi.org/10.5281/zenodo.7360791>). Other codes are available from the authors upon kind request.

### References

1. Coggle, J. E., Lambert, B. E. & Moores, S. R. Radiation effects in the lung. *Environ. Health Perspect.* **70**, 261–291 (1986).

2. Morgan, G. W., Pharm, B. & Breit, S. N. Radiation and the lung: a reevaluation of the mechanisms mediating pulmonary injury. *Int. J. Radiat. Oncol. Biol. Phys.* **31**, 361–369 (1995).
3. Wynn, T. A. Integrating mechanisms of pulmonary fibrosis. *J. Exp. Med.* **208**, 1339–1350 (2011).
4. Chen, Z., Wu, Z. & Ning, W. Advances in molecular mechanisms and treatment of radiation-induced pulmonary fibrosis. *Transl. Oncol.* **12**, 162–169 (2018).
5. Krzyzanowski, M. C., Levy, J., Page, G. P., Gaddis, N. C. & Clark, R. F. Using semantic web technologies to power LungMAP, a molecular data repository. In: Proceedings of The International Workshop on Semantic Big Data 1–6 (ACM). <https://doi.org/10.1145/3066911.3066916> (2017).
6. Neumark, N., Cosme, C., Rose, K.-A. & Kaminski, N. The idiopathic pulmonary fibrosis cell atlas. *Am. J. Physiol.-Lung Cell. Mol. Physiol.* **319**, L887–L892 (2020).
7. Adams, T. S. et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* **6** [www.ipfcellatlas.com](http://www.ipfcellatlas.com) (2020).
8. Unterman, A. et al. Single-cell multi-omics reveals dysynchrony of the innate and adaptive immune system in progressive COVID-19. *Nat. Commun.* **13**, 440 (2022).
9. Sauler, M. et al. Characterization of the COPD alveolar niche using single-cell RNA sequencing. *Nat. Commun.* **13**, 494 (2022).
10. Sikkema, L. et al. An integrated cell atlas of the human lung in health and disease. Preprint at <https://doi.org/10.1101/2022.03.10.483747> (2022).
11. Travis, E. L. The sequence of histological changes in mouse lungs after single doses of x-rays. *Int. J. Radiat. Oncol. Biol. Phys.* **6**, 345–347 (1980).
12. Reyfman, P. A. et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1517–1536 (2019).
13. Xie, T. et al. Single-cell deconvolution of fibroblast heterogeneity in mouse pulmonary fibrosis. *Cell Rep.* **22**, 3625–3640 (2018).
14. Travaglini, K. et al. A molecular cell atlas of the human lung from single cell RNA sequencing. <https://doi.org/10.1101/742320> (2019).
15. Salton, F., Volpe, M. C. & Confalonieri, M. Epithelial–mesenchymal transition in the pathogenesis of idiopathic pulmonary fibrosis. *Medicina (Kaunas)* **55**, 83 (2019).
16. Riemondy, K. A. et al. Single cell RNA sequencing identifies TGF $\beta$  as a key regenerative cue following LPS-induced lung injury. *JCI Insight*. **5**, e123637 (2019).
17. Strunz, M. et al. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat. Commun.* **11**, 3559 (2020).
18. Prête, C. M., Yao, E., O’Donoghue, R. J. J., Mutsaers, S. E. & Knight, D. A. *Stat3. Proc. Am. Thorac. Soc.* **9**, 177–182 (2012).
19. Lao, T. et al. Hhip haploinsufficiency sensitizes mice to age-related emphysema. *Proc. Natl Acad. Sci. USA* **113**, E4681–E4687 (2016).
20. Hsia, L.-T. et al. Myofibroblasts are distinguished from activated skin fibroblasts by the expression of AOC3 and other associated markers. *Proc. Natl Acad. Sci. USA* **113**, E2162–E2171 (2016).
21. Martinez, F. O. & Gordon, S. The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000Prime Rep.* **6**, 13 (2014).
22. Francis, M. et al. Editor’s highlight: CCR2 regulates inflammatory cell accumulation in the lung and tissue injury following ozone exposure. *Toxicol. Sci.* **155**, 474–484 (2017).
23. Arora, S., Dev, K., Agarwal, B., Das, P. & Syed, M. A. Macrophages: their role, activation and polarization in pulmonary diseases. *Immunobiology* **223**, 383–396 (2018).
24. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
25. Ouyang, J. F., Kamaraj, U. S., Cao, E. Y. & Rackham, O. J. L. ShinyCell: Simple and sharable visualisation of single-cell gene expression data. <http://biorxiv.org/lookup/doi/10.1101/2020.10.25.354100>. <https://doi.org/10.1101/2020.10.25.354100> (2020).
26. Raredon, M. S. B. et al. Single-cell connectomic analysis of adult mammalian lungs. <http://advances.sciencemag.org/> (2019).
27. Angelidis, I. et al. An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. *Nat. Commun.* **10**, 963 (2019).
28. Zilionis, R. et al. Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species. *Immunity* **50**, 1317–1334.e10 (2019).
29. Traver, G. et al. Loss of Nrf2 promotes alveolar type 2 cell loss in irradiated, fibrotic lung. *Free Radic. Biol. Med.* **112**, 578–586 (2017).
30. Romero, F. et al. A pneumocyte-macrophage paracrine lipid axis drives the lung toward fibrosis. *Am. J. Respir. Cell Mol. Biol.* **53**, 74–86 (2015).
31. Kendall, R. T. & Feghali-Bostwick, C. A. Fibroblasts in fibrosis: novel roles and mediators. *Front. Pharmacol.* **5**, 123 (2014).
32. Chen, Q. & Liu, Y. Heterogeneous groups of alveolar type II cells in lung homeostasis and repair. *Am. J. Physiol. Cell Physiol.* **319**, C991–C996 (2020).
33. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
34. Wirsdorfer, F. et al. Extracellular adenosine production by ecto-50-nucleotidase (CD73) enhances radiation-induced lung fibrosis. *Cancer Res.* **76**, 3045–3056 (2016).
35. De Leve, S. et al. Loss of CD73 prevents accumulation of alternatively activated macrophages and the formation of prefibrotic macrophage clusters in irradiated lungs. *FASEB J.* **31**, 2869–2880 (2017).
36. Barkauskas, C. E. et al. Type 2 alveolar cells are stem cells in adult lung. *J. Clin. Investig.* **123**, 3025–3036 (2013).
37. Baselet, B., Sonveaux, P., Baatout, S. & Aerts, A. Pathological effects of ionizing radiation: endothelial activation and dysfunction. *Cell. Mol. Life Sci.* **76**, 699–728 (Birkhauser Verlag AG, 2019).
38. Piera-Velazquez, S., Mendoza, F. A. & Jimenez, S. A. Endothelial to mesenchymal transition (EndoMT) in the pathogenesis of human fibrotic diseases. *J. Clin. Med.* **5**, E45 (2016).
39. Choi, S. H. et al. A hypoxia-induced vascular endothelial-to-mesenchymal transition in development of radiation-induced pulmonary fibrosis. *Clin. Cancer Res.* **21**, 3716–3726 (2015).
40. Hashimoto, N. et al. Endothelial-mesenchymal transition in bleomycin-induced pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* **43**, 161–172 (2010).
41. Favaudon, V. et al. Ultrahigh dose-rate FLASH irradiation increases the differential response between normal and tumor tissue in mice. *Sci. Transl. Med.* **6**, 245ra93–245ra93 (2014).
42. Fouillade, C. et al. FLASH irradiation spares lung progenitor cells and limits the incidence of radio-induced senescence. *Clin. Cancer Res.* **26**, 1497–1506 (2020).
43. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
44. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience* **9**, g1aa151 (2020).
45. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
46. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. <https://arxiv.org/abs/1802.03426> (2020).

47. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
48. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
49. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
50. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
51. Tsanov, N. et al. SmiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res.* **44**, e165 (2016).
52. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
53. Imbert, A. et al. FISH-quant v2: a scalable and modular tool for smFISH image analysis. *RNA* **28**, 786–795 (2022).
54. Ester, M., Kriegel, H.-P. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 226–231 (1996).

## Acknowledgements

This work has received financial support through an Agence Nationale de la Recherche (ANR) grant (Lustra to A.L.-V., F.Mu., T.W., and C.F.). It has also received support from EDF (to A.L.-V. and C.F.), La Ligue Contre Le cancer (to A.L.-V. and S.C.A.), the Institut National du Cancer (INCa) (to A.L.-V.), the Institut Curie ICGEx granting program (to A.L.-V.) and the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 666003 (to S.C.A.). Furthermore, this work was supported by the French government under the management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute to T.W.). F.Mu. and C.W. acknowledge funding by Institut Pasteur. The authors greatly acknowledge the Multimodal Imaging Center-Light Microscopy Facility of the Institut Curie (CNRS UMS2016/InermUS43/Institut Curie/Université Paris-Saclay), as well as the Cell and Tissue Imaging Platform—PICT-IBISA (member of France-Bioimaging—ANR-10-INBS-04) of the U934/UMR3215 of Institut Curie for help with light microscopy. The authors wish to thank Christophe Alberti, Elodie Belloir, Cédric Lantoine, and Virginie Dangles-Marie from the animal core facility of Institut Curie. The contribution of the bioinformatics core facility at U900 Inserm-Institut Curie is gratefully acknowledged. High-throughput sequencing was performed by the ICGEx NGS platform of the Institut Curie supported by the grants ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France Génomique Consortium) from the Agence Nationale de la Recherche ("Investissements d'Avenir" program), by the ITMO-Cancer Aviesan (Plan Cancer III) and by the SiRIC-Curie program (SiRIC Grant INCa-DGOS-465 and

INCa-DGOS-Inserm\_12554). Data management, quality control, and primary analysis were performed by the Bioinformatics platform of the Institut Curie.

## Author contributions

S.C.A., C.F., C.W., H.L., S.Le, and S.La performed the experimental studies. S.C.A., J.S., C.F., T.D., T.W., and F.Mu. carried out the computational analysis of scRNA-seq data and smFISH images. S.H. built the web interface. S.C.A., J.S., T.D., T.W., F.Mu, J.A.L.V., and C.F. participated in the writing of the paper. J.A.L.V. and C.F. supervised the work. S.C.A., J.S., T.D., C.W., S.H., H.L., S.Le, S.La, M.D., V.F., F.Ma, T.W., F.Mu, J.A.L.V., and C.F. contributed to data discussion.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-38134-z>.

**Correspondence** and requests for materials should be addressed to José-Arturo Londoño-Vallejo or Charles Fouillade.

**Peer review information** *Nature Communications* thanks Ting Xie and Ping-kun Zhou for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023