Study of the mechanisms leading to radiation-induced pulmonary fibrosis through a single cell RNA sequencing approach

Juliette Soulier

Sorbonne Université

École doctorale 515 – Complexité du Vivant

Institut Curie UMR3347-U1021 / Équipe de recherche LONDONO

Study of the mechanisms leading to radiation-induced pulmonary fibrosis through a single cell RNA sequencing approach

Etude des mécanismes qui provoquent la fibrose pulmonaire post irradiation par une approche scRNAseq

Par Juliette Soulier

Thèse de doctorat de Biologie

Dirigée par Arturo Londoño-Vallejo et Charles Fouillade

Présentée et soutenue publiquement le 22 octobre 2024

Devant un jury composé de :

Stéphane Le Crom, Professeur, Président du jury et représentant de Sorbonne Université

François Paris, Directeur de recherche, Rapporteur

Rob Coppes, Professeur, Rapporteur

Laura Cantini, Chargée de recherche, Examinatrice

Arturo Londoño, Directeur de recherche, Directeur de thèse

Charles Fouillade, Chercheur, Co-directeur de thèse

ACKNOWLEDGEMENTS

First, I would like to thank my supervisors, Charles Fouillade and Arturo Londoño for trusting me with this project for the last almost five years. It has been very fulfilling and the experience of a lifetime, and you made it possible. You helped me through difficult times, keeping me motivated and giving me opportunities to attend and present my work during congresses, and I am very grateful for that. Arturo, thank you for being always so enthusiastic to all the results I produced, giving me a lot of ideas on what to investigate next. Charles, thank you for your numerous advice on how to proceed with the analysis, for patiently guiding me on the art of research and project management. Thank you both for starting my scientific career and for reinforcing my love for research.

Then, I would like to thank Pr Rob Coppes and Dr François Paris for accepting to review my thesis manuscript, as well as the other members of my thesis jury: Dr Laura Cantini, Pr Stéphane Le Crom, Charles Fouillade and Arturo Londoño-Vallejo. Thank you all for taking time to participate in the apotheosis of my thesis journey. I am also thankful to the members of my thesis committee: Dr Laura Cantini, Dr Kerstin Meyer and Dr Nicolas Servant for the great feedback, advice and encouragement they provided me during the last few years.

This work would not have been possible without the several great collaborations that took place during the last few years. I would like to thank the members of the LUSTRA project: Thomas Walter, Florian Mueller, Thomas Defard, Florian Massip and Christian Weber. We had great meetings, developing the autoFISH technique, and you also gave me very interesting feedback on my work. And the LUSTRA dinners afterward were also greatly appreciated. I also would like to thank the different platforms of Curie Institute, especially the single cell platform and the cytometry platform. I am very grateful for the Cochin Hospital and the Institut du Thorax that provided the human lung samples.

I did my M2 internship, my year as a study engineer and the beginning of my PhD in Arturo's team in Paris. I would like to thank them for welcoming me, especially during these

difficult times of COVID19. Especially, a huge thank you to Sandra Curras Alonso, who were there for me from the beginning, and who, during almost three years, guided me, taught me the basis of single cell RNA sequencing data analysis, reassured me, helped me. I am very lucky to have had you as my mentor.

Then at the beginning of my second year of PhD I followed Arturo to Orsay, and again this new team welcomed me with open arms. For a few months, I shared an office with Lucie, Margaux and Anouk, and I really appreciated our good discussions and laughs! I also had very great moments with the lunch team: Jessica, Delphine, Pierre Marie, Nathalie, Manon and Floriane in addition to my three office partners, so thank you for this moment every day that I was looking forward to. I also had three great co-PhD students during these years: Hugo and Ayan, who started one year after the beginning of my thesis and who I contributed to train at the beginning, and Maxime, who helped me with the writing of the methods of this work. It was great to work with you, and I enjoyed the last few months sharing an office with Ayan! It was also a pleasure working with Sophie, I enjoyed a lot learning more about the physical aspect of irradiation. Finally, I would like to acknowledge the work of Vincent Favaudon, a great scientist whose work is of great importance for the field of radiobiology. Overall, I am very lucky that I had the opportunity to work with such an amazing team, both on a scientific and personal level, with a lot of interesting lab meetings and very fun aperos.

I am very thankful to a great structure that was created recently at Curie Institute: the bioinformatics Hub. This grouping of Curie Institute's bioinformaticians and biostatisticians is a wonderful initiative that allowed to share knowledge and skills to solve problems and advance the progress of science together. Therefore, I would like to thank the members of the Hub, and especially Nicolas Servant, the head of the bioinformatic Hub, and Andrianteranagna Mamy Jean De Dieu, that welcomed me in the Hub. You helped me a lot, and I hope that I was able to also provide help to some of the members of the Hub.

I would also like to thank a few people that made a lot of moments and procedure much easier, and that were essential for conducting my research project: Marie France Lavigne in Paris, Veronique Piveux and Sandrine Bourgeois in Orsay. You helped me with all

the administrative procedures, ordre de mission, travel arrangements for conferences, the sometimes difficult communication with Sorbonne University, and countless other things.

Different structures made my thesis possible, and I would like to thank them. Sorbonne University funded my thesis, and the doctoral school ED515 Complexité du Vivant provided a great support and scientific environment, through events such as the Boris Ephrussi Days. Finaly, Curie Institute hosted my work and provided an amazing environment for research, with skilled platforms, conferences given by scientists from all over the world, the training unit, and talented researcher always willing to share their knowledge and participate in scientific discussions.

Je ne serais pas là sans les nombreuses personnes qui m'ont accompagné tout au long de ces années. Je tiens tout d'abord à remercier Florence Guedj et Sébastien Lahaye qui ont marqué mes années d'école primaire et m'ont fait découvrir les sciences et leur côté ludique. En classe préparatoire, j'ai rencontré des professeurs passionnés qui m'ont permis d'acquérir une compréhension inédite dans les domaines de la biologie, de la géologie, des mathématiques, de la physique et de la chimie. Je suis très reconnaissante d'avoir ces bases pluridisciplinaires, elles m'ont aidée à mieux comprendre de nombreux sujets, aussi bien au cours de ma vie professionnelle que personnelle. Durant ces deux années de travail intense, j'ai appris une méthode de travail qui me sera utile toute ma vie. J'ai également découvert le domaine de l'informatique et du code, que j'apprécie encore beaucoup aujourd'hui et qui fait maintenant partie de ma vie professionnelle. Je tiens donc à remercier en particulier Mme Ribet, Mme Metz, Mme Chevalier et Mr Percie du Sert.

I also would like to thank Pr Guinilla Karlsson Hedestam and Sharesta Khoenkhoen that welcomed me in the Karolinska Institut in Stockholm for my M1 internship and guided me through my first research project.

Je voudrais également remercier mes amis qui m’ont accompagnée pendant toutes ces années. Marion, ma première amie depuis la petite section de maternelle. Merci d’être toujours mon amie depuis toutes ces années, merci de tes petits messages de temps en temps qui me font toujours immensément plaisir, et merci d’être toujours là pour m’écouter quand

j'en ai besoin. Merci à Lisa, avec qui j'ai traversé la prépa et qui est maintenant une de mes meilleures amies et partenaire de Donjons et Dragons. Merci également à mes amis d'école d'ingénieur, et tout particulièrement à Kilian, qui restera toujours mon coloc ! Merci pour toutes ces soirées et weekends jeux de société et jeux de rôle, et merci de nous accueillir régulièrement chez toi pour un bon bol d'air en Haute Savoie. De même, merci à Bastian et Yann avec qui je passe régulièrement des soirées mémorables à occire des monstres et autres machins mutants. Plus récemment, merci à Amélie et Adrien de m'avoir régulièrement accueillie pour des soirées jeux toujours très sympa, et de s'être occupés de mes petits compagnons à écailles.

Merci à mes petits compagnons à poils. Nyora a égayé ma solitude lors des couvres feux et télétravail du COVID, avec ses câlins et coups de folies, et Smaug nous a rejoint un an plus tard pour ajouter son aura de chat heureux et mignon à notre petite famille. Leur amour vaut bien les nombreuses nuits écourtées par leurs jeux et demandes de câlins !

Un immense merci à ma famille, qui m'accompagne et m'encourage depuis toujours. Merci à tata Claire qui m'a fait découvrir le monde de la recherche lors de mon stage de 3ème à l'INRA, et qui m'a accueillie pendant ce temps chez elle et tonton Bruno. Merci à mes cousines Alice et Léna, avec qui je passe toujours de supers moment, que ce soient pour les fêtes de fin d'années et autres anniversaires, lors des vacances à Asnelles chez Mamée. Merci à Maminou, toujours là pour moi en cas de besoin, avec qui j'ai passé de supers vacances à faire tout plein d'activités, et que j'aime fort. Un merci tout spécial à Nanou et Grand père, qui ont toujours été si fiers de moi, et qui auraient été tellement heureux de voir l'accomplissement de toutes ces années de travail. Merci à tonton Olivier, toujours à faire des blagues et faire rire tout le monde, et avec qui j'ai la joie de partager tous mes anniversaires. Merci à mon cousin Aurélien pour tous ces étés à la Tranche, à étudier toute la faune qui pouvait passer par le jardin, jouer dans les vagues et faire enrager notre grand-mère.

Merci à Théo, qui restera toujours mon petit frère. Même si nous avons pu nous taper sur le nerf toutes ces années, je t'aime très fort, et je serai toujours là pour toi comme tu es là pour moi. Merci à mon papa, qui m'a répété que l'informatique était importante jusqu'à que je m'en rende compte et l'intègre à mon parcours, merci de m'avoir soutenue toutes ces

années. Je t'aime papa. Merci à Valérie qui a toujours plein de conseils à me prodiguer pour gérer mes problèmes de santé. Merci à ma maman d'avoir toujours su me remotiver dans les moments difficiles, qui m'a toujours dit que j'étais capable d'aller au bout de mes rêves, je n'en serai pas là sans toi. Tu es si forte, si aimante et attentionnée, et je t'aimerais et t'admirerai toujours. Merci à Philippe, qui sait toujours comment me faire marcher, mais qui est aussi toujours là quand j'en ai besoin.

Et finalement merci à Pierre, mon amoureux, mon partenaire. Merci à ta famille de m'avoir accueillie à bras ouverts ! Merci à Béatrice et Georges qui prennent soin de moi comme de leur fille, à sa grand-mère toujours adorable avec moi, et à Antoine avec qui j'ai tout de suite connecté et avec qui je suis toujours ravie de passer du temps. Merci Pierre d'être toujours là pour moi, dans les bons moments comme dans les moins bons, merci de toujours me comprendre, de me faire rire, de prendre soin de moi. Je ne peux pas rêver de meilleur compagnon pour partager ma vie.

Merci à tous, sans vous ce travail n'aurait pas été possible.

TABLE OF CONTENTS

ABBREVIATIONS

aCap = aerocytes

AM = alveolar macrophages

ARDS = acute respiratory distress syndrome

ASM = airway smooth muscle cells

Artery EC = artery endothelial cells

AT0 = type 0 pneumocytes

AT1 = type I pneumocytes

AT2 = type II pneumocytes

BALF = broncho-alveolar lavage fluid

BSA = bovine serum albumin

CBCT = cone beam computed tomography

cDC = conventional dendritic cells

COPD = chronic obstructive pulmonary disease

CTGF = connective tissue growth factor

DAMP = damage-associated molecular pattern

DC = dendritic cells

DEG = differentially expressed genes

DPBS = Dulbecco's phosphate buffered saline

ECM = extracellular matrix

EDTA = ethylenediaminetetraacetic acid

EMT = epithelial to mesenchymal transition

FACS = fluorescence-activated cell sorting

FBS = fetal bovine serum

FDR = false discovery rate

FPRI = fibrose pulmonaire radio-induite

gCap = general capillaries

GEM = gel beads-in-emulsion

GO-BP = gene ontology – biological processes

GRN = gene regulatory network

GSEA = gene set enrichment analysis

h = hour

HCA = human cell atlas

HLCA = human lung cell atlas

HTO = hashtag oligo

IR = irradiated

IL = interleukin

ILD = **i**nterstitial lung disease

IM = interstitial macrophages

IMRT = intensity-modulated radiotherapy

IPF = idiopathic pulmonary fibrosis

LPRI = lésion pulmonaire radio-induite

LPS = Lipopolysaccharide

Lymphatic EC = lymphatic endothelial cells

M = month

MCA = mouse cell atlas

MMP = matrix metalloproteinase

moDC = non-conventional monocyte-derived dendritic cells

NAC = N-acetylcysteine

NK cells = natural killer cells

NK T cells = natural killer T cells

NSCLC = non-small cell lung cancer

OXPHOS = oxidative phosphorylation

PBS = phosphate-buffered saline

PCA = principal component analysis

PCR = polymerase chain reaction

pDC = plasmacytoid dendritic cells

PDGF = platelet-derived growth factor

PMMA = poly methyl methacrylate

PCR = polymerase chain reaction

RBC = red blood cell

RILI = radiation-induced lung injury

RIPF = radio-induced pulmonary fibrosis

ROS = reactive oxygen species

RP = radiation pneumonitis

RT = reverse transcription

SARRP = small animal radiation research platform

SASP = senescence-associated secretory phenotype

SBRT = stereotactic body radiation therapy

SCLC = small cell lung cancer

scRNAseq = single cell RNA sequencing

SD = standard deviation

SMC = smooth muscle cells

SRS = stereotactic radiation surgery

TLR = toll-like receptor

TNF = tumor necrosis factor

UMI = unique molecular identifiers

TGFβ = transforming growth factor beta

tSNE = t-distributed stochastic neighbor embedding

UMAP = uniform manifold approximation and projection

UMI = unique molecular identifier

V = version

VEGF = vascular endothelial growth factor

Vein EC = vein endothelial cell

VSM = vascular smooth muscle cells

PUBLICATIONS

Juliette Soulier, Sandra Curras-Alonso, Maxime Dubail, Hugo Laporte, Ayan Mallick, Chloé Lafouasse, Marine Lefèvre, Mylène Bohec, Pierre Verrelle, Nicolas Girard, Agathe Segui-Givelet, Arturo Londoño Vallejo, Charles Fouillade. 2024 'Non-tumor lung irradiation during radiotherapy triggers a pro-angiogenic signaling'. *Manuscript in preparation*.

Sandra Curras-Alonso, **Juliette Soulier**, Thomas Defard, Christian Weber, Sophie Heinrich, Hugo Laporte, Sophie Leboucher, et al. 2023. 'An Interactive Murine Single-Cell Atlas of the Lung Responses to Radiation Injury'. *Nature Communications* 14 (1): 2445. https://doi.org/10.1038/s41467-023-38134-z.

Sandra Curras-Alonso, **Juliette Soulier**, Thomas Walter, Florian Mueller, Arturo Londoño-Vallejo, and Charles Fouillade. 2021. 'Spatial Transcriptomics for Respiratory Research and Medicine'. *European Respiratory Journal* 58 (1). https://doi.org/10.1183/13993003.04314-2020.

Thomas Defard, Hugo Laporte, Mallick Ayan, **Juliette Soulier**, Sandra Curras-Alonso, Christian Weber, Florian Massip, et al. 2024. 'A Point Cloud Segmentation Framework for Image-Based Spatial Transcriptomics'. *Communications Biology* 7 (1): 1–13. https://doi.org/10.1038/s42003-024-06480-3.

RÉSUMÉ ÉTENDU EN FRANÇAIS

Le poumon est un organe complexe composé de plusieurs populations cellulaires spécialisées interagissant entre elles pour remplir sa fonction principale : effectuer les échanges gazeux entre l'air extérieur et le sang afin d'absorber l'oxygène qui sera apporté aux différents tissus et organes du corps, et évacuer les déchets transportés par le sang, composés principalement de dioxyde de carbone. Ces échanges gazeux sont réalisés par les unités fonctionnelles des poumons : les alvéoles. Ces dernières sont composées de cellules épithéliales, endothéliales, mésenchymateuses et immunitaires qui jouent toutes un rôle crucial dans la fonction pulmonaire. Ces différentes populations sont organisées de manière spécifique, ce qui permet une surface maximale pour les échanges gazeux, une protection contre les menaces extérieures, et des interactions intercellulaires entre les différentes populations. Cependant, en cas de blessure, agression ou maladie, cette organisation peut être perturbée et la fonction pulmonaire peut être compromise. Différents mécanismes de réparation et de régénération des blessures existent pour que le poumon revienne à un état d'homéostasie.

Les radiations peuvent causer de telles lésions aux poumons. La réponse à une lésion pulmonaire radio-induite (LPRI) débute par une étape inflammatoire, avec une implication importante des cellules immunitaires, notamment des macrophages. Si la réparation de la blessure échoue, la lésion peut alors évoluer vers une fibrose pulmonaire radio-induite (FPRI) avec accumulation de tissus cicatriciels, composés principalement de fibroblastes, de myofibroblastes et de matrice extracellulaire, avec une destruction de la structure des alvéoles. Bien que les principaux événements du développement de la fibrose pulmonaire radio-induite aient été décrits, les événements cellulaires et moléculaires détaillés survenant au cours de la progression de la maladie restent inconnus. Une meilleure compréhension de ces processus permettrait de développer des stratégies ou des traitements pour prévenir ou ralentir le développement de la FPRI.

Dans ce but, nous avons ici quatre objectifs principaux : 1) identifier les principales altérations cellulaires et moléculaires et leur temporalité affectant trois populations cruciales

du poumon : les cellules épithéliales, les cellules endothéliales et les macrophages, 2) identifier les différences dans la réponse du poumon lorsqu'ils sont exposés à une dose d'irradiation fibrogène ou non fibrogène, 3) étudier la réponse du poumon humain à la radiothérapie et la comparer aux évènements observés chez la souris, 4) développer et donner accès à des atlas et bases de données des réponses des tissus pulmonaires de la souris et de l'homme aux lésions radiologiques à l'échelle de la cellule unique.

Pour étudier la temporalité des événements conduisant à la fibrose pulmonaire et l'effet de différentes doses d'irradiation, nous avons utilisé un modèle murin d'irradiation thoracique complète, à dose fibrogène de 13Gy ou non fibrogène de 10Gy. De plus, nous avons obtenu des échantillons de lobectomies de six patients souffrant d'une tumeur de Pancoast. Ces patients ont subi une radiothérapie néoadjuvante six à huit semaines avant l'intervention chirurgicale. Pour chaque patient, un échantillon a été obtenu d'une région irradiée du poumon et un témoin d'une région non irradiée du poumon. Ces échantillons ont été analysés à l'aide de la technologie de séquençage d'ARN de cellule unique (scRNA seq), un outil permettant une meilleure compréhension de la diversité cellulaire et de la variabilité du transcriptome des échantillons analysés. Nous avons par la suite utilisé des techniques de bio-informatique pour analyser les données générées et identifier les différents processus induits par l'irradiation des tissus sains des poumons.

Dans un premier temps, nous avons observé une importante activité inflammatoire durant les premiers points de temps post irradiation, avec une production accrue de cytokines par plusieurs populations immunitaires, Les macrophages semblent jouer un rôle important à différents moments de la réponse à l'irradiation. Juste après l'irradiation, des macrophages interstitiels sont recrutés depuis les monocytes circulants. Au cours des mois suivant la lésion, les macrophages alvéolaires développent un profil M2 pro fibrotique, et les macrophages interstitiels un profil M1 pro inflammatoire. De plus, nous observons une modification du transcriptome des macrophages alvéolaires après une irradiation pro- fibrogène vers une augmentation du métabolique lipidique.

Nous avons ensuite étudié la régénération des populations épithéliales à la suite des dommages causés par l'irradiation. En effet, l'épithélium pulmonaire (et plus précisément les pneumocytes de type I et II) est particulièrement vulnérable et susceptible à la destruction

ou à l'apoptose en cas d'infection ou d'agression. Nous avons identifié un premier processus de réparation durant les premières heures suivant l'irradiation via une dédifférenciation des pneumocytes de types II en pneumocytes de type 0, qui ont la capacité de se multiplier et se redifférencier en pneumocytes de type I ou II. Lors des points de temps plus tardifs de l'irradiation, nous observons également un processus direct de transdifférenciation des pneumocytes de type II en pneumocytes de type I. Cependant, une forte dose d'irradiation peut impacter de manière plus durable le profil des pneumocytes de type II : en effet, après une irradiation fibrogenic, différents éléments indiquent une possible transition épithéliale à mésenchymateuse de cette population. Les cellules épithéliales montrent également des signes de senescence en réaction à l'irradiation.

Pour finir, nous avons étudié les cellules endothéliales et les processus de réparation vasculaire après irradiation. Les lésions vasculaires en sont une conséquence majeure. L'irradiation provoque la mort des cellules endothéliales et la désorganisation du réseau vasculaire, comme cela a été démontré à la fois chez des patients humains et chez des modèles murins. Le remplacement de ces cellules et capillaires endothéliaux endommagés est principalement déclenché par une signalisation pro-angiogénique ainsi que via la spécialisation des cellules endothéliales en deux états différents : les cellules « tip » et les cellules « stalk ». Les cellules « tip » développent des filipodes et peuvent migrer, suivant un gradient extracellulaire de VEGFA. Ces cellules occupent une position dominante dans la formation du nouveau vaisseau et sont suivies par les cellules « stalk » qui forment les parois du vaisseau. Après irradiation, nous observons une augmentation de la proportion de cellules d'identité « tip » chez les gCap, la population progénitrice des cellules endothéliales capillaires. Ces cellules « tip » présentent également un phénotype migratoire accru. Nous avons de plus montré une conservation de ces processus chez les patients humains ayant subi une radiothérapie. Enfin, les cellules endothéliales et particulièrement les aérocytes présentent de manière similaire aux cellules épithéliales des signes de sénescence induite par l'irradiation.

Dans l'ensemble, nous sommes convaincus que ces travaux ont considérablement amélioré notre compréhension des lésions pulmonaires induites par les radiations, jetant ainsi de solides bases pour de nouvelles recherches sur de nombreuses questions non résolues. L'utilisation du séquençage d'ARN de cellule unique a fourni un outil puissant pour

élucider les diverses réponses cellulaires tissulaires normales et les interactions intercellulaires initiées par les lésions pulmonaires induites par les radiations. De plus, nos résultats mettent en évidence de nouvelles cibles thérapeutiques potentielles et ouvrent la voie à une utilisation innovante de médicaments pour lutter contre cet effet secondaire mortel de la radiothérapie.

CHAPTER I
Introduction

CHAPTER I – Introduction

1. Physiology of the respiratory system

a) Structure of the lung and the alveoli

The lung is the organ responsible for the gas exchanges between the outside environment and the blood, with the absorption of oxygen and the release of waste gas, such as carbon dioxide.

The upper airways are composed of several structures. The nose and the mouth are the entry and exit point for the air into the respiratory system. The nasal hairs are also the first filter for the pathogens and particles inhaled with the outside air. The sinuses heat and humidify the air before its entry into the lower airways. The pharynx and the larynx allow the air to go between the upper and lower airway and protect the lower airway from contamination by food or liquids.

The lower respiratory airway starts with the lower part of the larynx and is then constituted by several structures with diverse functions. The trachea is an elastic and cartilaginous pipe with a diameter of 1.5 to 2 cm and a length of 10 to 12 cm that starts by the larynx and ends with the two main bronchi. The trachea transports the air to the lungs and also heats it and humidifies it. The two main bronchi are the first division of the respiratory systems and bring the air into the left and the right lungs. They are supported by cartilage and then further divide into smaller bronchi that bring the air to the different parts of the lung. The last subdivision of the bronchi gives rise to the bronchioles, smaller tubes of less than a millimeter in diameter that are not supported anymore by cartilage. The end of the bronchioles, called the respiratory bronchioles, are the first structure that performs gas exchanges (Albertine 2016). They can also change in diameter in order to reduce or increase airflow through processes called bronchodilatation and bronchoconstriction. The terminal bronchioles secrete a substance called surfactant that allows the bronchioles to expand during inhalation and prevents them from collapsing during expiration. The bronchioles end by the alveoli, the functional unit of the lungs that allows most of the gas exchanges (**FIGURE 1.1,** (Ahmad and Balkhyour 2020)). The alveoli are composed of a unique layer of different cells that forms a small air sac that allows the gas exchange, with cells producing surfactant,

immune cells that protect the body against outside threats, and cells performing the gas exchange (**FIGURE 1.2,** (Curras Alonso 2021)). The alveoli are highly vascularized: they are coated with capillaries vessels. On average, human lungs contain 300 to 500 million alveoli, for a gas exchange surface of 100 to 150 m^2.

These terminal structures are also supported by other tissues. The diaphragm is the main respiratory muscle that contracts and relaxes to bring the air inside and out of the lungs. The ribs form a protective cage surrounding the lungs and the heart, protecting it from outside mechanical threats.

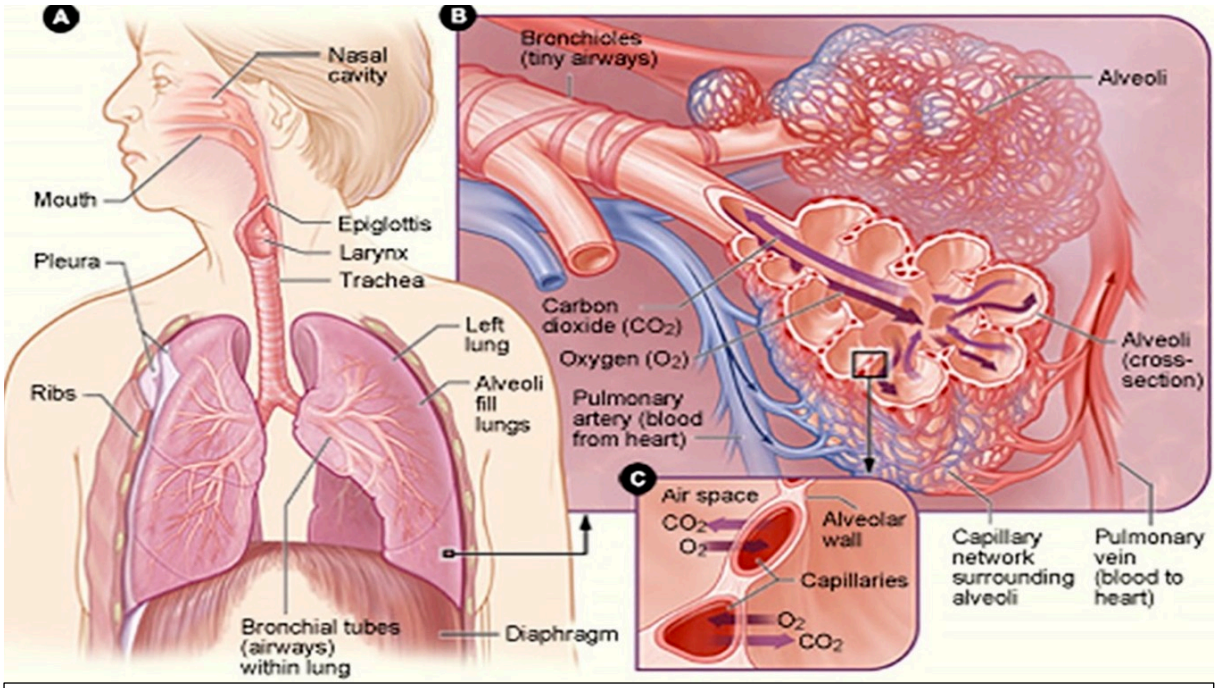


FIGURE 1.1. schematic representation of the human airway and the alveoli
(Ahmad et Balkhyour 2020)

b) Cellular composition

The lungs are composed of various types of cells that participate in its function: epithelial cells, endothelial cells, mesothelial cells, mesenchymal cells, myeloid cells and lymphoid cells. Most of them are resident cells that stay in the lung, but circulating immune cells also interact with the lung cell populations. In physiological conditions, the lung is a lowly proliferating organ, with a low cell turnover (I. Y. R. Adamson 1985). However, when the lungs

sustain an injury, it is capable of regeneration. There is not a single stem cell niche that is able to produce all the different cell types of the lung, but rather some cell types that are able to proliferate, de-differentiate or re-differentiate (Kotton and Fine 2008).

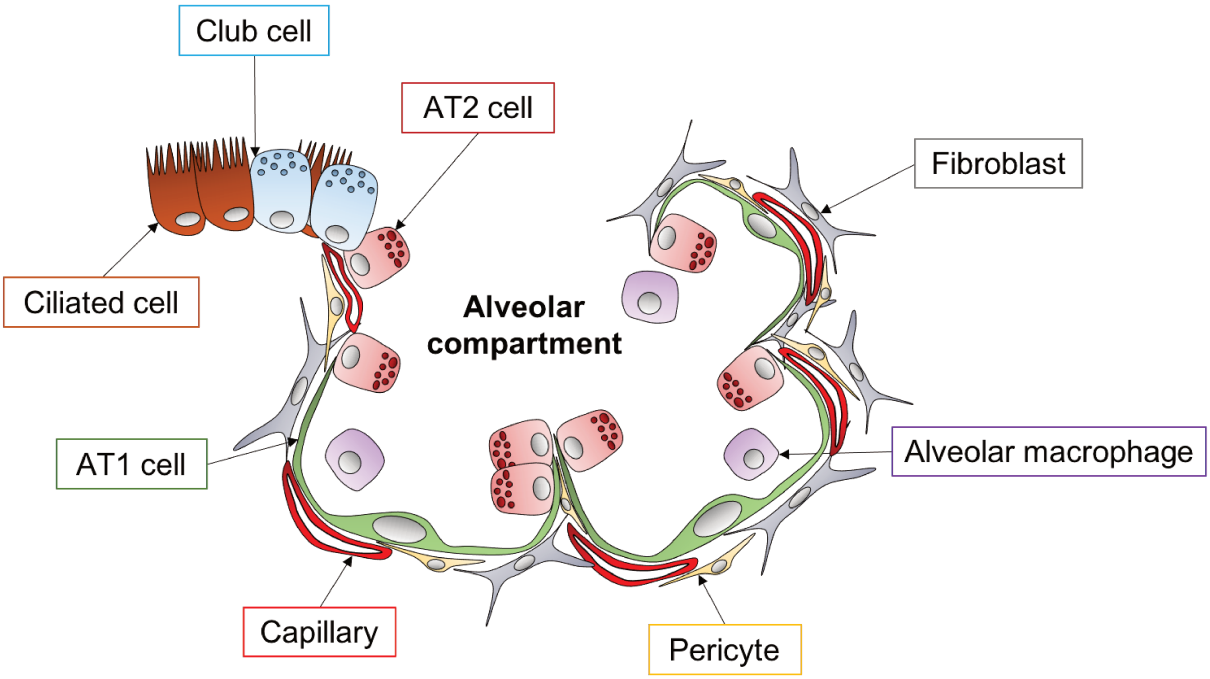


FIGURE 1.2. schematic representation of the alveolar structure.
The bronchioalveolar duct junction is mainly composed with ciliated cells and club cells and it leads to the alveolar region. The alveolar compartment is lined with two types of specialized epithelial cells, the type I pneumocytes (AT1) and type II pneumocytes (AT2) cells. The capillary endothelial cells are close to the AT1 cells, to form the respiratory membrane. Fibroblasts and pericytes are found in the stroma region. The alveolar macrophages are localized in the alveolar space (Alonso 2021).

i. Epithelial cells

The epithelial cells are in charge of gas exchanges, constitute the alveoli and are the most abundant cells of the lungs. The type I pneumocytes (AT1 cells) cover 95% of the alveoli surface (Fujino et al. 2011). They are long, flat cells that allow the passage of the gases between the air and the blood. They are unable to proliferate or replicate. Most of the rest of the alveoli is composed of type II pneumocytes (AT2 cells), smaller cubelike cells. This population presents several functions: first, they can proliferate and differentiate into AT1 cells, thus they are responsible for the regeneration of both pneumocyte populations in case of injury or disease. They also secrete surfactant that prevents the collapse of the alveoli and facilitates the gas exchanges (Castranova et al. 1988). Other epithelial cells, found in the

terminal bronchioles, perform a support function. The goblet cells are the main secretory cells of the lung, producing mucus. This is a substance mainly composed of mucin glycoproteins and it has surfactant properties as well as antibiotic properties and it collects small outside particles that managed to enter the lungs (Ma, Rubin, and Voynow 2018). The ciliated cells, as their name suggests, are cells whose apical end is covered with cilia which with their movement can move the mucus within bronchioles, and evacuate the debris and pathogens trapped in it. Finally, the basal cells have the ability to proliferate and differentiate into ciliated or goblet cells, to replace them in case of injury, for example (Tilley et al. 2015). Single cell RNA sequencing allowed great progresses in the identification and molecular characterization of the cellular diversity of the different organs in different organisms, notably thanks to global initiative like the mouse cell atlas (MCA) (R. Wang et al. 2023) or the human cell atlas (HCA) (Regev et al. 2017). The different lung epithelial subsets can be identified using their characterized profiles and their expression of described characteristic markers (**ANNEXE I, table 5.1**).

Alveolar epithelial type I (AT1) and type II (AT2) cells are fundamental for the function of the lungs. However, AT1 cells are fragile and susceptible to destruction or apoptosis upon infection or aggression (Kuwano 2007). Both new AT1 and AT2 cells can be produced in physiological or pathological conditions to replace missing cells or repair injured areas. AT2 cells, besides their function as surfactant producer cells, have the ability to de-differentiate into cells called type 0 pneumocytes (AT0 cells). These AT0 cells are then able to proliferate and differentiate into new AT1 or AT2 cells (Kadur Lakshminarasimha Murthy et al. 2022). This regeneration process allows the maintaining of the alveoli structure and the gas exchange function of the lungs. The other types of epithelial cells are also able to regenerate. Basal cells can proliferate and re-differentiate into club cells and ciliated cells. Furthermore, club cells also have self-renewal abilities and can re-differentiate into ciliated and basal cells (F. Chen and Fine 2016; Alysandratos, Herriges, and Kotton 2021).

Other cell populations may appear under pathological conditions. This is the case of the aberrant basaloid cells, that have been described in lungs of patients with idiopathic pulmonary fibrosis or chronic obstructive pulmonary disease. These cells express only some of the basal cell markers (Adams et al. 2020) but they overexpress genes involved in epithelial to mesenchymal transition, cytokine response, senescence, thus they might play a role in

injury repair (Selman and Pardo 2020). Other studies describe that these aberrant cells can also be characterized by the expression of the KRT8 markers, and that they might then enter different states: either senescence or transition to AT1 cells (F. Wang et al. 2023).

A direct type II to type I pneumocyte trans-differentiation process has also been described after a LPS injury. After the injury, AT2 cells proliferate to replenish the pneumocyte pool, then a subset of these cells goes into cell cycle arrest and trans-differentiate. This cell cycle arrest has been shown to be induced by TGF-β signaling, but it is then inactivated in order to proceed with the trans-differentiation to AT1 cells (Riemondy et al. 2019).

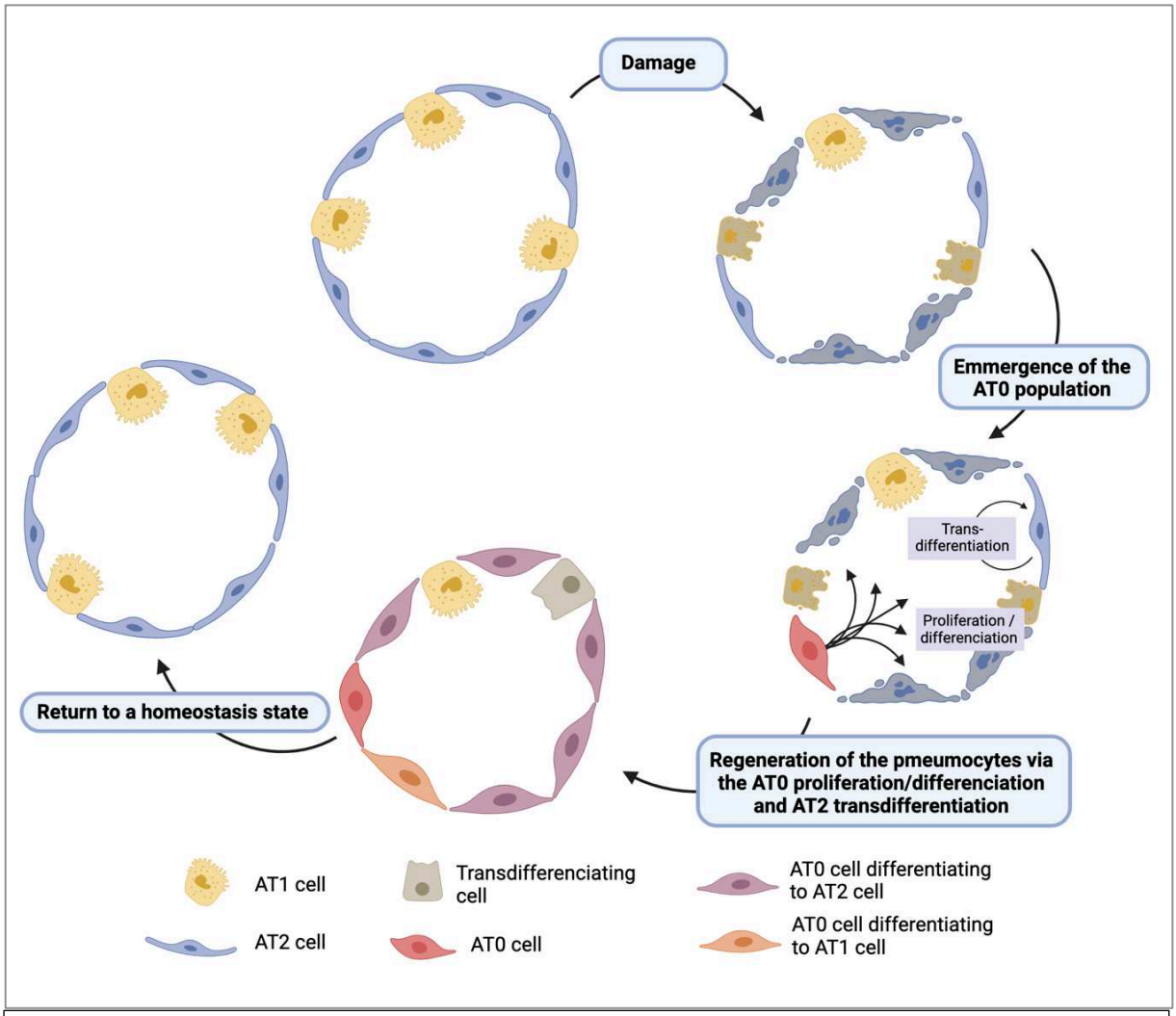


FIGURE 1.3. schematic representation of the regeneration of the lung epithelial cells

ii. Mesenchymal cells

The mesenchymal cells are responsible for the structure of the lungs, forming the connective tissue and secreting extracellular matrix (ECM).

In physiological conditions, fibroblasts are the most abundant mesenchymal cells. They sustain the structure of the alveoli and secrete extracellular matrix, which is also crucial for intercellular communication (White 2015). Different sub-populations of fibroblasts have been described according to their location in the lung and their proximity to certain structures. The peri bronchial fibroblasts are located in the walls of the bronchi and other conducting airways, the adventitial fibroblasts are surrounding the bronchi vascular bundles, and the alveolar fibroblasts are located near the alveoli (Tsukui et al. 2020) (**FIGURE 1.4**).

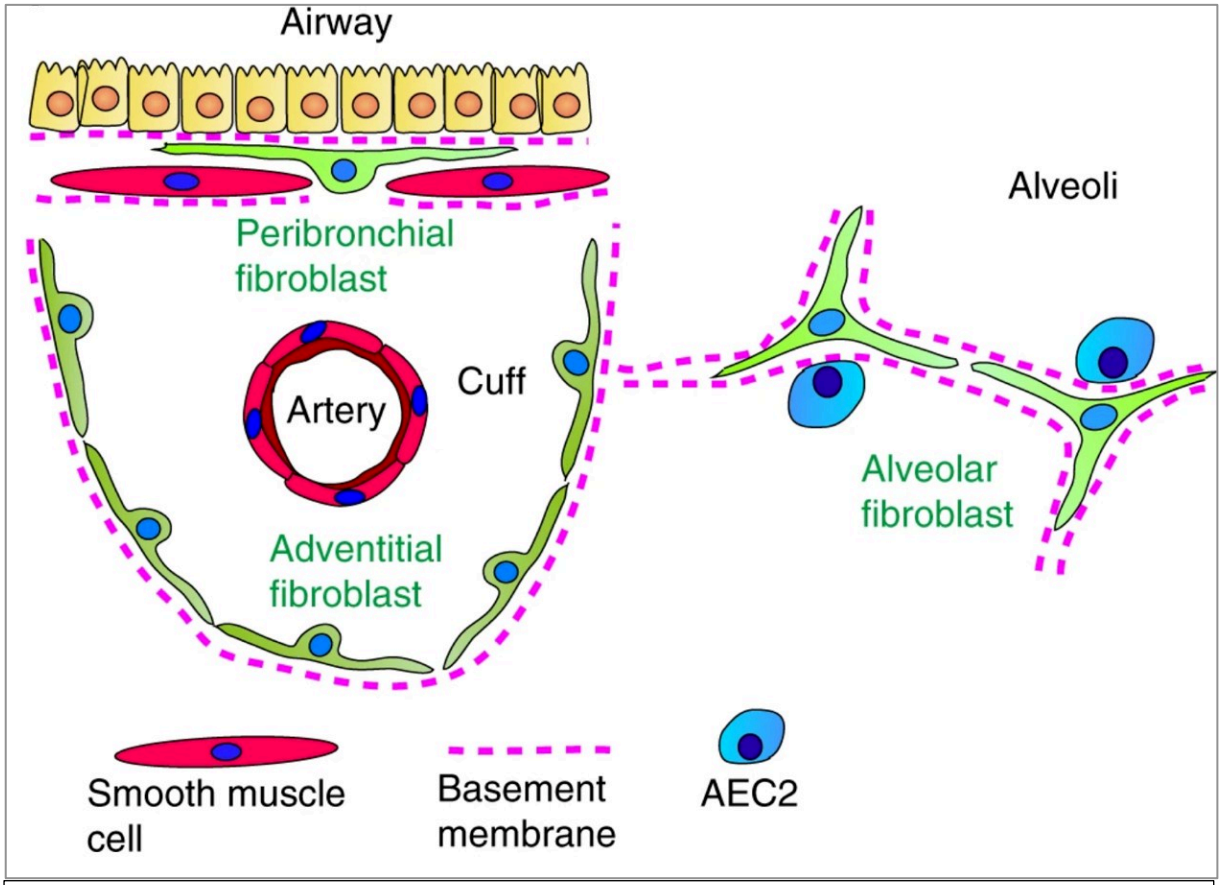


FIGURE 1.4. schematic representation of the mesenchymal lung cells localization (Tsukui et al. 2020).

There is a subset of fibroblasts that have a high lipid content that seems to have an important role in AT2 cells homeostasis (Habiel and Hogaboam 2017). Pericytes are a less

abundant population located near the small vessels of the lungs and provide them with mechanical and biochemical support (Garrison et al. 2023). Myofibroblasts are almost absent during physiological conditions and can derive from fibroblasts or pericytes after injury. They are involved in wound healing mainly through ECM deposition and restoration of the tissue barrier integrity (Hung 2020). Furthermore, after injury, myofibroblasts are thought to be able to de-differentiate back to lipofibroblasts (El Agha et al. 2017).

Smooth muscle cells (SMC) are cells with contractile properties that control the diameter of the different bronchi, bronchioles and vessels, thus controlling the airflow and blood pressure (Q. Gu and Lee 2006). They can be divided into airway smooth muscle cells (ASM) and vascular smooth muscle cells (VSM). ASM seems to have more rapid contraction velocity than VSM and be more implicated in different pathologies like asthma (Fernandes et al. 2004).

Mesenchymal cells play a major role in the lung response to injury and regeneration. Upon acute lung injury, fibroblasts differentiate into activated myofibroblasts. Different signaling pathways are activated, like the TGF-β, mTOR or EGFR. An increased amount of collagen and ECM is produced to repair the damaged tissues. However, if an excessive amount of it is produced it can lead to pathologies like idiopathic pulmonary fibrosis (IPF) (Whitsett et al. 2019). The different lung mesenchymal subsets can be identified using their characterized profiles and the expression of described characteristic markers (**ANNEXE I, table 5.2**).

iii. Endothelial cells

Endothelial cells are the cells forming the walls of capillaries that bring the blood in contact with the alveoli to perform the gas exchanges through the AT1 cells. Capillary endothelial cells can be subdivided into several subtypes performing distinct functions. The aerocytes (aCap) are the endothelial cells responsible for the gas exchange. They are large cells with ramifications in close contact to AT1 cells to present the maximum surface for gas exchange. They are thought to not be able to proliferate or self-renew. The other type of capillary endothelial cells is the general capillaries (gCap), they also perform gas exchange but have also been predicted to be able to proliferate for self-renewal and differentiation to aCap (Gillich et al. 2020). Other small vessels are in charge of transporting the blood between the

lung capillaries and the main circulation: the artery endothelial cells (artery EC) form the vessels that carry the unoxygenated blood from the circulation to the capillaries in contact to the alveoli, and then the vein endothelial cells (vein EC) form the vessels that carry the oxygenated blood back to the circulation (Trimm and Red-Horse 2023). Finally, the lymphatic endothelial cells (lymphatic EC) form the lymphatic vessels that carry lymph out of the lungs (Lorusso et al. 2015).

During a lung injury, the lung vessels can also be damaged, especially the smaller capillaries vessels. gCap cells have been described to be able to proliferate and produce new aCap and gCap cells (Gillich et al. 2020). The process of vessel regeneration is called angiogenesis. This process involves a complex signaling network involving the reception of vascular endothelial growth factor (VEGF) molecules (Abhinand et al. 2016). Angiogenesis has been shown to involve the differentiation of the endothelial cells into two cell states: tip cells and stalk cells.

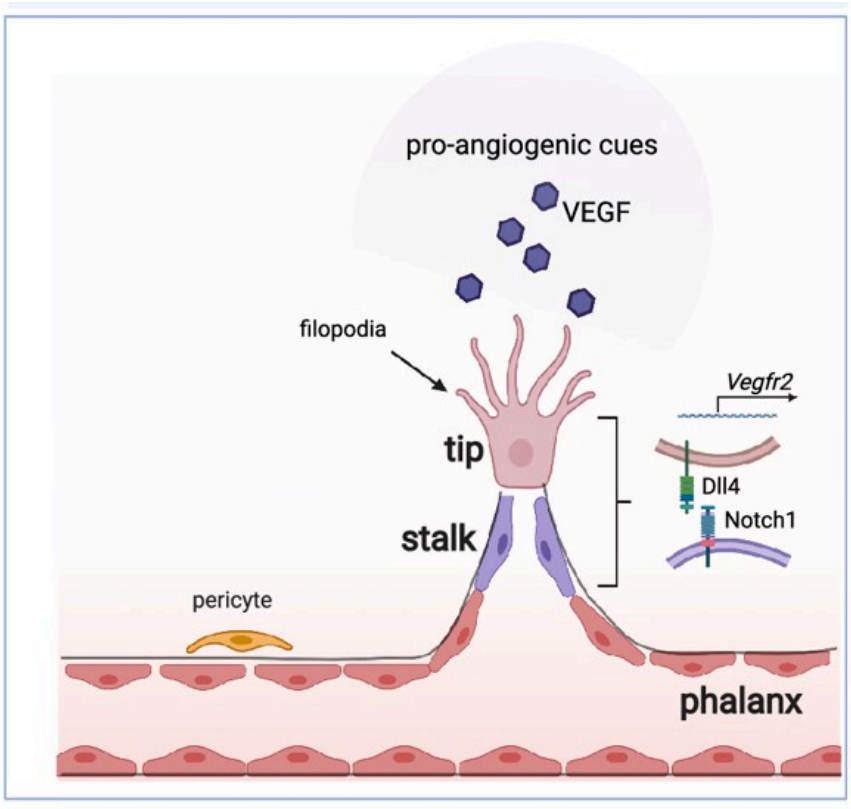


FIGURE 1.5. schematic representation of the tip and stalk cells involvement in angiogenesis (Pasut et al. 2021)

Tip cells express the VEGF receptor KDR, therefore are able to receive the VEGF signaling. Tip cells develop filipodia and can migrate, following an extracellular gradient of VEGFA and occupy a leading position in the new vessel formation. They are followed by the stalk cells that divide to form the walls of the growing new vessel (Jakobsson et al. 2010). Tip cells receive the VEGF signaling, which triggers the expression of several genes including DLL4. DLL4 acts then as a ligand for the NOTCH receptor expressed by the stalk cells. In stalk cells, NOTCH activation inhibits the expression VEGFR2 and activates the expression of VEGFR1, maintaining the stalk cell identity (W. Chen et al. 2019; Pasut et al. 2021; Hellström, Phng, and Gerhardt 2007) (**FIGURE 1.5**).

The different lung endothelial subsets can be identified using their characterized profiles and the expression of described characteristic markers (**ANNEXE I, table 5.3**).

iv. Mesothelial cells

The lung mesothelial cells are part of the pleura, a layer of cells that envelop the lungs, and they have a role of protection (H. Batra and Antony 2015). The mesothelial cells can be identified with the expression of Msln, Wt1 and Upk3b in mice or MSLN, UPK3B and WT1 in human (Travaglini et al. 2020).

v. Myeloid cells

The myeloid cells are the first responder in the lung, they are crucial for the innate immune response and the initiation of the adaptative response (Cook and MacDonald 2016; Zaynagetdinov et al. 2013). Some myeloid cells of the lung are stably present in the tissue, so they are called resident cells, while others are circulating in the blood and travel through the lungs within the vessels. Monocytes are immune cells from the innate immune system that can be circulating or tissue resident (Rodero et al. 2015). They have phagocytosis properties and can proliferate and differentiate into interstitial macrophages under physiological conditions, and into alveolar macrophages under pathological conditions (F. Li et al. 2022). Monocytes can be distinguished between two subtypes: classical monocytes and non-classical monocytes. Classical monocytes account for 90 percent of monocytes, the rest of the monocytes being non-classical monocytes and intermediate monocytes sharing features of both categories. Non classical monocytes derive from classical monocytes and are considered

the most mature form of monocytes. These two monocyte populations have different roles in inflammation and injury response. Classical monocytes have been shown to be highly inflammatory and secrete pro-inflammatory cytokines, whereas non classical monocytes secrete more anti-inflammatory cytokines and are involved in repair processes (Anbazhagan et al. 2014).

Interstitial macrophages (IM) are located in the parenchyma of the lung. They are phagocytic cells, have an antigen-presenting activity and have an important immunoregulatory role (Schyns, Bureau, and Marichal 2018). In mice, we can distinguish two subsets of interstitial macrophages based on the structures nearby: the nerve-associated IM located near nerve bundles, and vessel-associated IM, located close to blood vessels (Ural et al. 2020).

Alveolar macrophages (AM) also have phagocytosis capabilities, are located in the alveoli and are the first responders in the lungs against the threats coming with the outside air. They can be pro-inflammatory or anti-inflammatory and have a role in the clearing of the infection by secreting factors through processes of efferocytosis. In physiological conditions the AM compartment can self-renew, but in case of their depletion monocytes can also differentiate into AM (Hou et al. 2021).

Dendritic cells (DC) are phagocytic and antigen-presenting cells that are mainly present in the tissues surrounding the blood vessels and pleura (Lambrecht et al. 2001). DC can be divided into three categories: the conventional DC (cDC), plasmacytoid DC (pDC) and non-conventional monocyte-derived DC (moDC) (Cook and MacDonald 2016).

Neutrophils are very abundant in the blood stream, but they also exist as resident cells in certain organs, including the lung. Lung resident neutrophils are part of the innate immune response through phagocytosis activity. They also regulate inflammation to protect the lungs from injury or infection (Bae et al. 2022). Basophils are another agent of the immune response and can induce inflammation in the lung (Schwartz, Eberle, and Voehringer 2016).

The resident myeloid cells compartment can also be damaged or destroyed upon injury. In physiological conditions, IM are produced by the differentiation of circulating monocytes, and resident AM can proliferate and self-renew. However, during injury if there

is an important destruction of the AM compartment, circulating monocytes can also differentiate into new recruited AM (Shi et al. 2021) (**FIGURE 1.6**).

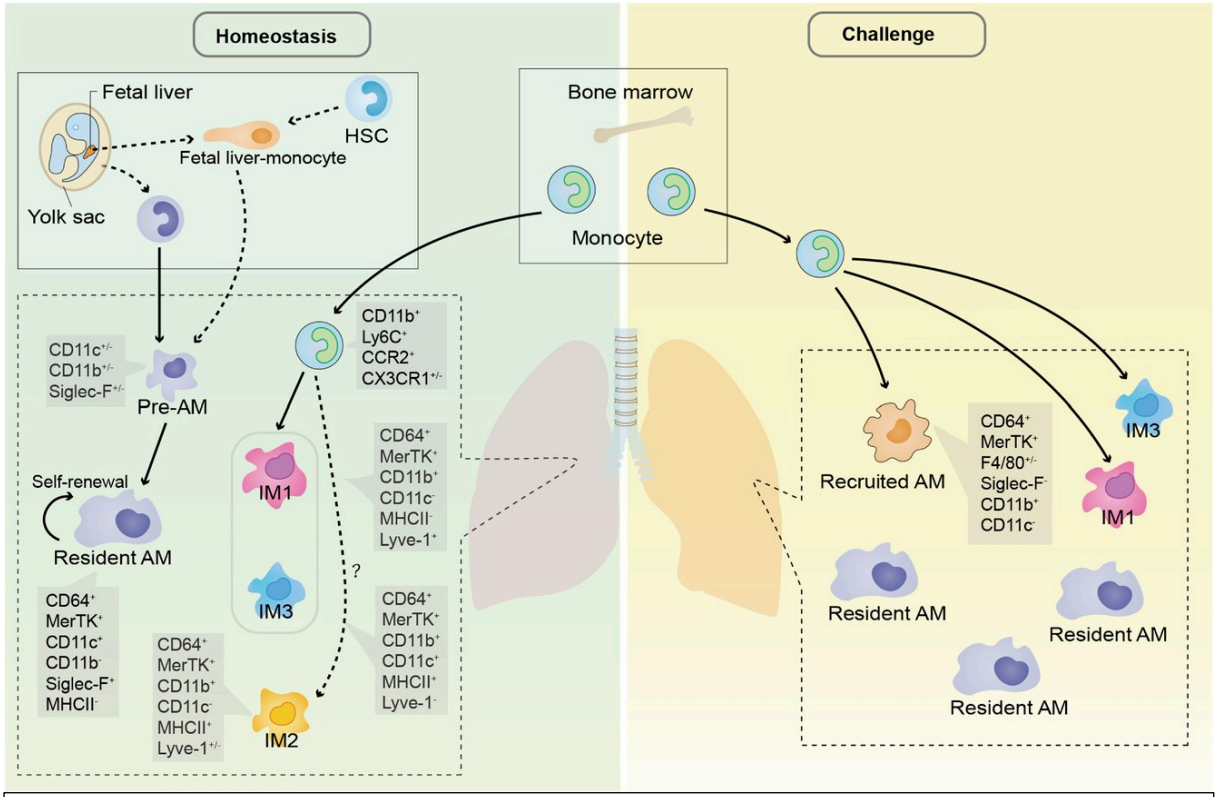


FIGURE 1.6. schematic representation of the origin of AM and IM subsets in mice. Resident AMs, derived from embryo (yolk sac and/or fetal liver), are capable of self-replicating during homeostasis and challenge in lung. During steady state, the maintenance of AM pool rarely needs the contribution of bone marrow -derived monocytes, but in the circumstance of inflammation, monocytes are strongly recruited to areas of inflammatory alveoli and differentiated into recruited monocyte-derived AMs (Shi et al. 2021).

The different lung myeloid subsets can be identified using their characterized profiles and the expression of described characteristic markers (**ANNEXE I, table 5.4**).

vi. Lymphoid cells

The lymphoid cells are part of the adaptative immune response and play an important role in the fight against pathogens or threats and in the processes of inflammation. During lung diseases, T cells are essential for the initiation and the maintaining of inflammation and the recruitment of other agents of the immune response (Cosio, Majo, and Cosio 2002). Indeed, T cells, as well as other immune cell types, can secrete chemokines that have a chemoattractant effect on neutrophils, monocytes or macrophages (Barnes 2016). There are two main populations of T cells: CD4+ helper T cells and CD8+ cytotoxic T cells (Xiong and

Bosselut 2012). The CD4+ T cells express MHC-II genes, they play a support function in the immune response by activating cells from the innate immune system through cytokine expression (Luckheeram et al. 2012). The CD8+ T cells are able, upon contact with an antigen, to activate, proliferate, recognize virus-infected cells, bacteria or cancer cells, and kill them (N. Zhang and Bevan 2011).

Natural killer cells (NK cells) are mainly involved in the innate immune response (Hervier et al. 2019), but a subset of lymphoid cells share characteristics of both NK cells and T cells, called natural killer T cells (NK T cells), and are a bridge between innate and adaptative immune cells (Hodge and Hodge 2019).

The B cells are part of the adaptative immune system and can differentiate into memory cells for a faster response to infections. Some B cell subsets are also involved in the resolution of inflammation (Polverino et al. 2016). We can also distinguish the plasma cells, which are terminally differentiated antibody-secreting B cells (Kunkel and Butcher 2003). The different lung lymphoid subsets can be identified using their characterized profiles and the expression of described characteristic markers (**ANNEXE I, table 5.5**).

2. Pathology of the respiratory system

a) Chronic respiratory diseases

In 2017, more than half a million people worldwide were living with a chronic respiratory disease and it was the third cause of death, with an 18% increase since 1990 ('Prevalence and Attributable Health Burden of Chronic Respiratory Diseases, 1990–2017: A Systematic Analysis for the Global Burden of Disease Study 2017' 2020). Patient living with a chronic respiratory disease can present a shortness of breath, chronic cough, lack of energy and endurance. These symptoms are disabling and affect the quality of life. Furthermore, these patients are more vulnerable to respiratory infections, such as the flu or the COVID 19 virus (Z. He, Zhong, and Guan 2022). The most common chronic respiratory diseases are chronic obstructive pulmonary disease (COPD), asthma, interstitial lung disease (ILD) and idiopathic pulmonary fibrosis (IPF).

i. Chronic obstructive pulmonary disease

COPD is a chronic respiratory disease characterized by the obstruction of the airflow form the lungs. While the exposure to chemicals like cigarette smoke or the presence of infections are important factors, the risk of developing COPD is also strongly influenced by the expression of some genetic variants (Silverman 2020). COPD arises from the abnormal resolution of airway inflammation caused by long-term chemical exposure. Once this process starts, it continues even if the exposure to the chemical stops. Different phenomena are known to be part of the physiopathology of COPD. Oxidative stress elevates during the development of COPD, mainly due to chemical exposure and oxidant production by immune cells during inflammation. Indeed, the number of immune cells in the lung is increased in patients with COPD, particularly for the neutrophils. When activated, these cells produce cytokines and reactive oxygen species (ROS) which contribute to inflammation. Furthermore, pulmonary macrophages also become more abundant and release several pro-inflammatory molecules, such as H_2O_2, superoxide anion, proteases and growth factors and chemokines, in addition to the recruitment of leukocytes. These changes are linked with structural changes, such as the loss of parenchyma, endothelial and epithelial cells, and an increase of smooth muscle cells (Bourdin et al. 2009). Furthermore, the inflammation and oxidation are self-sustaining, inducing the secretion of more mucus with an increased expression of the Muc5b

and Mu5ac genes and the aggregation of more neutrophils, a phenomenon that has been described to help the reduction of the inflammation by degrading cytokines and chemokines (Hahn et al. 2019). Neutrophils also express matrix metalloproteinase (MMP) genes and elastase, leading to remodeling of the airway. The activated macrophages also lead to the production of cytokines like tumor necrosis factor (TNF)α and different types of interleukins. In particular, the secretion of interleukin-6 can induce production of oxygens radicals and elastase, leading to permeabilization and destruction of the lung tissue (Guo et al. 2022).

Different molecular targets and drugs are under investigation along with clinical trials to try to slow down the development of COPD or cure it. As oxidative stress is one of the main mechanisms in the development of COPD, several clinical trials have tested N-acetylcysteine (NAC) to reduce oxidative stress in the lungs of COPD patients. The different clinical trials have shown that NAC can help improve the airway function, but the results are very variable between patients and clinical trials. Like oxidative stress, several cytokines are involved in the development of COPD and are under investigation as a potential druggable target. However, these promising targets have shown limited results during the different studies and clinical trials. Different pathways inhibitors are being tested, like TNFα inhibitors, interleukins inhibitors, MMP inhibitors (Guo et al. 2022). So far, the most efficient target seems to be NLRP3. The NRLP3 inflammasome is a macromolecular complex sensing cell stress or danger signals and initiating inflammation and has been shown to be a critical regulator of inflammation in different lung diseases (Sayan and Mossman 2016). Different in vitro and in vivo experiments in mice show that the inhibition of NLRP3 leads to decrease of inflammation without any side effect (Guo et al. 2022).

ii. Asthma

Asthma is the most common chronic disease in developed countries, affecting up to 10% of adults (Barnes 2017). It is characterized by a lung chronic inflammatory disorder resulting in recurrent episodes of wheezing, chest tightness or coughing. The most common type of asthma is allergen induced. Non-allergic asthma can be caused by upper airway infection, gastroesophageal reflux, airborne irritant exposure, cold air or intense emotion (Padem and Saltoun 2019). Two lung cell types are known to be the main elements involved in the physiopathology of asthma: epithelial cells and smooth muscle cells. Upon exposure to

contaminants or asthma triggers, epithelial cells secrete cytokines and interleukins, which activate various immune cells, such as dendritic cells, basophils or T helper 2 cells. Furthermore, nerves are also activated by this signaling (Erle and Sheppard 2014).

Several genetic factors have been shown to predispose to allergic asthma: 60 to 80 percent of asthma susceptibility can be explained by genetic variations. Furthermore, dysregulation of the immune barrier is a crucial factor in the triggering of asthma. A leaky epithelium leads to faster infiltration of outside particles in the lungs, thus can trigger a more intense immune reaction. This is one of the reasons why the incidence of asthma has increased during the last decades: during that period, the presence of chemical and physical threats polluting the air has increased: microorganisms, diesel, cigarette smoke, nanoparticles, microplastics or other pollutants. All these elements have been shown to damage the lung epithelium barrier (Komlósi et al. 2022).

iii. Interstitial lung diseases

Interstitial lung diseases represent a group of more than 200 diseases (Demedts et al. 2001). They are characterized by a restrictive ventilatory defect due to a reduced distensibility of the lung parenchyma (Chetta, Marangio, and Olivieri 2004). ILD patients present inflammation of the lung tissues, and some also develop fibrosis decreasing the efficiency of the gas exchanges (Wijsenbeek, Suzuki, and Maher 2022). Various factors are known to increase the susceptibility to developing an ILD. Some patients present genetic predispositions (Devine and Garcia 2012), but external factors also play an important role. ILD can be caused by infection, certain drugs, radiation, cigarette smoke or chemical exposure. However, for the majority of the patients the cause remains unknown (Raghu, Nyberg, and Morgan 2004). ILD cellular and molecular mechanisms are still not fully understood, however some key mechanisms have been described. The chronic damage to the lungs leads to AT2 cells presenting an increased activity and proliferation in order to repair the structure of the alveoli, however failure to repair the injury leads to pulmonary fibrosis (Antoniou et al. 2014; Guler and Corte 2021).

iv. Idiopathic pulmonary fibrosis

Idiopathic pulmonary fibrosis is a severe form of ILD. Like many others respiratory diseases, the main risk factor is exposure to airborne chemicals like cigarette smoke, dust, fumes, fibers or other pollutants and particles. Infections, and in particular viral infections are also an important risks factor for IPF (Phan et al. 2021).

After diagnosis, the median survival is three to five years. On the cellular level, this disease is characterized by the accumulation of mesenchymal cells (fibroblasts and myofibroblasts) and an increase in their production of ECM. The activation of these cells occurs in foci and leads to the formation of scar tissues in different parts of the lung. Furthermore, other cell types participate in the development of fibrosis during IPF. There is an accumulation of alveolar macrophages and neutrophils that release cytokines, interleukins, recruit other immune cells like monocytes, secrete profibrotic agents like platelet-derived growth factor (PDGF) or transforming growth factor beta (TGFβ). Non-immune cells are also involved in this disease. AT1 cells are damaged during the development of IPF but are not properly replaced by AT2 cells differentiating to new AT1 cells. Instead, fibroblasts migrate to these damaged areas, destroying the alveolar structure. Blood vessels are also damaged early in the disease, before the development of fibrosis. It is mainly the capillary cells that are affected with a loss of barrier function, a decrease of angiogenic factor and an increase of anti-angiogenic factors (Bagnato and Harari 2015).

b) Lung cancer

i. Epidemiology and risk factors

Lung cancer is the leading cause of cancer-related death worldwide (GBD 2017 Disease and Injury Incidence and Prevalence Collaborators 2018). In 2018, more than two million new cases of lung cancer were diagnosed. Even if there has been a lot of progress in cancer treatment possibilities and efficiency during the last decades, the mortality of lung cancer is still high: the five-year survival is on average 25 percent (Ettinger et al. 2010). However, when detected early, the outcome improves drastically: the five-year survival rises to 70-85 percent. Therefore, the current effort is put on prevention, education of public and primary health care providers, and early screening (Myers and Wallen 2024).

The incidence of lung cancer varies greatly between regions of the world. In countries where there is a high incidence of cigarette smoking among the population, there is an average 20-fold increase in the incidence of lung cancer, and more than 80% of lung cancer cases in the developed countries are thought to be attributable to cigarette smoking. Other environmental risk factors include cannabis smocking, asbestos exposure, radon exposure, air pollution, exposure to arsenic, or inflammation and damages caused by infection. COPD patients are also at a higher risk of developing lung cancer. Furthermore, men are twice as likely to suffer from lung cancer than woman. Certain ethnicities present also a higher risk, like African American, or a lower risk, like Hispanics. Finally, several genes or chromosomal regions have been shown to be implicated in predisposition of certain patients and families to develop lung cancer (Thandra et al. 2021).

Lung cancer can be divided into two categories according to the appearance of the cells with a histopathology analysis: cells from small cell lung cancer (SCLC) are flatter and smaller than cells in non-small cell lung cancer (NSCLC).

ii. Non-small cell lung cancer

Non-small cell lung cancer is the most frequent form of lung cancer, accounting for 85 percent of the cases (Ettinger et al. 2010). Two main subtypes of NSCLC exist: adenocarcinoma and squamous cell carcinoma.

Adenocarcinoma accounts for 40 percent of the primary lung cancer, it is the most frequent subtype (Hutchinson et al. 2019). Its diagnosis requires the presence of neoplastic gland formation, pneumocyte marker expression in the form of TTF-1 with or without napsin, or intracytoplasmic mucin (Clark and Alsubait 2024). It has its origin in distal epithelium and grows following the epithelial wall (P. Hao et al. 2020). Different hypotheses have been made about the cell population of origin, but the definitive answer remains unclear. Probably, adenocarcinoma has the ability to start from different progenitor cell populations, and that can be influenced by the microenvironment or oncogenic drivers (Seguin, Durandy, and Feral 2022). After the primary development in the distal epithelium, lung adenocarcinoma can spread to the pleura, diaphragm, pericardium, bronchi, and then in more distant locations like vessels, lymph nodes, or other lobes.

Squamous cell carcinoma is the other most common type of NSCLC, and it is diagnosed when there is the presence of keratin producing tumor cells, with histochemistry analysis which shows expression of p40, p63, CK5 or desmoglein (Clark and Alsubait 2024). In the development of squamous cell carcinoma, we can also observe intracellular bridging and squamous pearl structures (W. D. Travis 2011). Approximately half of squamous cell carcinoma arise from the central region of the lungs, and the other half originates from the periphery (Sung, Cho, and Lee 2020).

iii. Small cell lung cancer

Small-cell lung cancer accounts for 15 percent of the cases of lung cancer. It is more aggressive than NSCLC, with patients presenting a poorer prognosis. It presents a high proliferative rate and a high rate of early metastasis, which are more frequently found in the lung lymph nodes, other parts of the lung, but also the liver, brain, adrenal glands or circulating in the blood vessels. Genomic analysis of the tumors showed a high mutational burden, with inactivation of the tumor suppressing genes TP53 and RB1 in most cases. We can also observe that 20 percent of the SCLC cases present a combination of SCLC cells and NSCLC cells, but the presence of these NSCLC cells does not seem to affect the survival rate of the patients.

iv. The particularity of the Pancoast tumor

The first case of Pancoast tumors was described in 1838 by Hare (Hare 1838). Then in 1924 Henry K. Pancoast reported several cases of patients with "*shadow of growth in extreme left apex and destruction of posterior portions of left second and third ribs and adjacent transverse processes*" (PANCOAST 1924).

These tumors are also known as superior sulcus tumors and are characterized by their localization at the apex of the lung, invading the parietal pleura. They can also invade muscles, ribs, thoracic vertebral bodies, vessels and other surrounding structures, rending them very complex to treat (Panagopoulos et al. 2014).

Most Pancoast tumors can be classified as NSCLC and affects mainly men over sixty years old with a history of heavy smocking. They represent three to five percent of all lung

cancers, thus they are a relatively rare subset. However, due to the localization and the tissue implication of the tumor, it is a very challenging disease with high mortality: the median survival time of patients is only 2,6 years.

c) Treatments possibilities of the pathologies of the respiratory system

The pathologies of the respiratory can be treated in different ways. First, the different lung cancers are usually treated with a combination of treatment modalities: surgery, chemotherapy, immunotherapy and radiotherapy.

i. Surgery

For most of the types of lung cancer, surgery is the first treatment modality if the cancer is diagnosed at an early stage (before the spread of metastasis). The operation consists of the resection of the tumor and the surrounding margins in order to remove a maximum of the tumor mass (Hoy, Lynch, and Beck 2019). However, the latest stages with metastasis are often consider unresectable and are treated with other modalities.

ii. Chemotherapy

Chemotherapy is one of the main treatment modalities used for lung cancer, alone or combined with other therapeutic options. One of the main chemotherapeutic agents used to this day is Cisplatin, but other molecules like Etoposide, Mitomycin, Ifosfamide are used, according to the patient situation (Bernhardt and Jalal 2016; Artal Cortés, Calera Urquizu, and Hernando Cubero 2015). However, numerous tumors develop resistance to chemotherapeutic agents, requiring the use of other molecules or another treatment modality (Kim 2016).

iii. Immunotherapy

Recently, immunotherapy approaches have been used to treat lung cancer and have shown great promises. Specifically, anti-PD1 antibodies have been reported to increase the proportion of good patient outcome in both first line and second line treatment of tumors presenting more than fifty percent of PDL1 positive cells (accounting for approximately 30% of the tumors) (Giroux Leprieur et al. 2017).

iv. Targeted treatment options

The cancer is usually tested for different mutations and, in combination with different treatments, personalized treatment can be implemented according to the tumor's genetic profile (Myers and Wallen 2024) For example, in some situations anti-angiogenic agents can be used (Ang, Tan, and Soo 2015).

v. Radiotherapy

Radiotherapy can be used to treat cancer in various situations. When the cancer is detected at an early stage, radiotherapy is often used after surgery (in combination with chemotherapy or not) to kill the potential remaining cancer cells. For more advanced stages of cancer, when the surgery is not an option, radiotherapy can be used to reduce the size of the tumor, therefore alleviating the symptoms and potentially prolongating the life of the patients (Choi et al. 2001; Vinod and Hau 2020). Similarly, the recurrences are usually treated with radiotherapy, possibly in combination with chemotherapy. In the case of the Pancoast tumor, due to the complexity of the localization of the tumor, a unique protocol has been implemented. The patient is treated with neoadjuvant chemotherapy and radiotherapy prior to the surgery, with an average total radiation dose of 60Gy, which is a high dose radiation, to reduce the complexity of the tumor. The surgery then consists most of the time in a lobectomy, sometimes a pneumonectomy (Kwong et al. 2005). From the point of view of research, Pancoast tumors offer a unique opportunity to collect non-tumoral irradiated human lung tissues.

Different particles are used to deliver radiotherapy: photons, electrons or protons. The most frequently used are proton and photon, and they can be delivered through different modalities. The conventional irradiation method uses x-ray films to guide the irradiation (Pfeiffer et al. 2002). For a more precise targeting of the tumor, it is now possible to use a CT-scanner to guide the therapy. This method is called 3D conformal radiotherapy (Schlegel 2006). Another advanced irradiation method is the intensity-modulated radiotherapy (IMRT) that is using 3D mapping with CT scan to delimit the tumor and then with several beams delivers a maximum intensity of irradiation to the tumor while the surrounding tissues only receive a low dose of irradiation (Taylor and Powell 2004). Stereotactic Radiation is another

method that uses 3D CT scanning to deliver precise high radiation dose to the tumor. It can be delivered through stereotactic body radiation therapy (SBRT) or stereotactic radiation surgery (SRS) (C. W. Song et al. 2021). Another possible way to deliver the radiation directly to the tumor is the use of brachytherapy (or curie therapy). It consists of putting the radiation source directly in or next to the tumor through surgery (Chargari et al. 2019).

As electrons release their energy close to the surface of the treated tissue, it is often used to treat superficial tumors, such as skin cancer, but it can also be used intraoperatively to treat directly the tumor (Petersen et al. 2002).

While radiotherapy is a very efficient tool to control tumor growth and kill cancer cells, part of the normal tissue also receives a significant dose of irradiation, and some patients develop normal tissues toxicities to irradiation.

vi. Treatment of early normal tissue toxicity to irradiation

In the first days/weeks following the radiotherapy, or even during the treatment, some patients develop signs of normal tissue radiation toxicity, also named radio induced lung injury (RILI): shortness of breath, coughing, chest pain and other chest or respiratory symptoms. When these first symptoms appear, the radiotherapy is reduced or suppressed, antibiotic can be given in case of infection, the patient is also given anti-inflammatory and drugs to treat the symptoms. If the condition of the patient worsens, the radiotherapy treatment is stopped, glucocorticoids and antibiotics are administered and according to the difficulties breathing the patient can be hospitalized with respiratory support (Yan et al. 2022).

vii. Treatment of late normal tissue toxicity to irradiation

In a small percentage of patients, the symptoms persist and worsen through the weeks/months following the radiotherapy, and scare tissues appear in the lungs. This late, chronic and irreversible state of radiation toxicity is called radio induced pulmonary fibrosis (RIPF). At the state of pulmonary fibrosis, no improvement can be expected, and the life expectancy of the patient is from a few months to a few years (Mehta 2005). However, several treatments are being investigated that could alleviate or prevent the development of RIPF.

Most of these treatments are similar to the ones used for treatment of IPF. There is clinical evidence that the administration of statins helps slow the progression of IPF. Different mechanisms have been proposed to explain this effect, the main theory being the inhibition of the fibroblast to myofibroblast differentiation (Dolivo et al. 2023). Senolytics shows also some potential in the treatment of IPF (Merkt et al. 2020). A first clinical study with 14 patients showed the feasibility of the senolytic treatment of IPF patients and some encouraging first results (Justice et al. 2019). Other clinical trials are ongoing, like the trial GKT137831 that targets specifically senescent myofibroblasts (Duncan 2023). However, these treatments are targeting IPF patients, so it might show different outcomes for RIPF patients. Numerous preclinical studies have been conducted targeting specifically the prevention and treatment of RIPF. For example, in mice, inhibition of HIF-1α before irradiation inhibits the development of RIPF (Nam et al. 2021). Others promising druggable targets are the ROS, the cytokines and chemokines, TGF-β, PDGF or CTGF (H. Jin et al. 2020).

3. Physiopathology of normal tissue radiation toxicities

a) Models of fibrosis and lung injury in mice

i. Thoracic irradiation model

To study RIPF, it is interesting to use the model of mouse thorax irradiation. Indeed, mice are a well-studied organism that share a lot of characteristics with humans for the development of RILI and RIPF. After thoracic irradiation, the lungs of mice go through an acute inflammation phase that evolves after four to five months post irradiation into an irreversible pulmonary fibrosis, leading to respiratory failure (E. L. Travis 1980). The use of the mouse model allows us to follow and study the different stages of RILI and RIPF, in order to understand the cellular and molecular events that lead to the development of RIPF.

ii. Bleomycin injury model

Bleomycin is a chemotherapeutic drug that can induce pulmonary fibrosis in experimental animals, making it a widely used model to study the mechanisms of lung fibrosis and evaluate potential therapies (Gul et al. 2023). It causes DNA damage, leading to lung injury, inflammation, and subsequent interstitial fibrosis (Miura et al. 2022). The bleomycin model exhibits histopathological changes similar to those seen in human idiopathic pulmonary fibrosis, including decrease in the proportion of AT1 cells, proliferation of AT2 cells, vascular endothelial damage, alveolar infiltration, collagen deposition, and interstitial and pleural fibrosis (Della Latta et al. 2015). While a single intratracheal dose of bleomycin in C57BL/6 mice can induce acute inflammatory changes, it often does not lead to the development of chronic fibrosis (Miura et al. 2022). Furthermore, contrary to RIPF, bleomycin induced fibrosis in mice is reversible. Spontaneous regression of the fibrosis typically occurs 3-4 weeks after the initial intratracheal administration of bleomycin (S. Zhang et al. 2023).

iii. Lipopolysaccharide injury model

Lipopolysaccharide (LPS)-induced lung injury is one of the most commonly used rodent models for acute respiratory distress syndrome (ARDS) and mimics the neutrophilic inflammatory response observed in ARDS patients (Asti et al. 2000). LPS has the ability to induce the release of numerous inflammatory mediators, including TNF-α, IL-1β, IL-6, NO and

superoxide anions, and can activate the Toll-like receptor 4 (TLR4) signaling pathway, leading to the activation of NF-κB and MAPK pathways. This results in the accumulation of inflammatory cells like neutrophils and macrophages in the lungs, increased vascular permeability, and pulmonary edema (Hu et al. 2017). Additionally, the LPS-induced lung injury model has been used to study long-term pulmonary functional outcomes and changes in angiogenic signaling pathways involving VEGF up to 4 weeks after the initial injury (Tsikis et al. 2022).

b) Biological events implicated in RILI and leading to RIPF

The aim of radiotherapy is to impact and cause death of the cancer cells while sparing the surrounding healthy tissues. However, there is always toxicity to a certain extent in the non-tumoral tissues. After lung radiotherapy, 5 to 20 percent of the patients will develop radiation pneumonitis (RP), an early form of radiation-induced lung injury (RILI) that can develop hours to days after radiation treatment. The damage responsible for this pathology has been described to be caused by DNA damage and the generation of ROS. This can cause inflammation and cell death, destruction of the alveolar structure and the vascular integrity, the development of a hypoxic environment. This damage can either resolve and heal after the end of the radiotherapy treatment or worsen and evolve into RIPF during the few months to years after lung irradiation. RIPF is characterized by the destruction of the alveolar structure of large portions of the lung and the development of scar tissues compose of fibroblasts, myofibroblasts and ECM deposition, that thicken the walls of the lungs and limits the gas exchanges. The development of RILI and RIPF can be followed by CT scanning the lungs of the patients and looking for changes in density of the lung parenchyma (Giuranno et al. 2019).

i. Initial inflammatory response of the lung to a radiation injury

Radiotherapy causes direct cytotoxic effects to the normal lung tissue, leading to the death of alveolar epithelial and endothelial cells. This disrupts the alveolar barrier function and increases vascular permeability, allowing the influx of inflammatory cells and fluid into the alveolar spaces. Cell death also induces a sterile inflammatory response. Within hours to days after radiation exposure, there is vascular congestion, leukocyte infiltration, and intra-alveolar edema. This early phase is characterized by the release of pro-inflammatory

cytokines and mediators such as TNFα, interleukin (IL)-1, and IL-6. These cytokines contribute to the recruitment and activation of immune cells, including neutrophils and macrophages, to the site of injury.

Neutrophils are among the first responders to tissue injury, including radiation-induced damage to the lung. Upon activation, they migrate to the site of injury guided by chemokines and adhesion molecules. Then they release ROS, proteases, and pro-inflammatory cytokines, contributing to tissue damage and inflammation. Excessive neutrophil activation can exacerbate lung injury and promote the development of fibrosis.

Radiation exposure causes the release of damage-associated molecular patterns (DAMPs) such as HMGB1, cytosolic DNA, and ATP from damaged lung cells. These DAMPs are recognized by toll-like receptors (TLRs) on macrophages and dendritic cells, leading to their activation and the production of pro-inflammatory cytokines like TNFα, IL-1, and IL-6 (Dar, Henson, and Shiao 2019; de Andrade Carvalho and Villar 2018). Macrophages are key regulators of the inflammatory response and tissue repair processes following radiation-induced lung injury. Both resident alveolar macrophages and recruited monocyte-derived macrophages contribute to tissue homeostasis and immune surveillance in the lung. Upon exposure to radiation-induced damage signals, macrophages become activated and phagocytose cellular debris, apoptotic cells, and pathogens. Different populations of macrophages play different roles in the reaction to radiation injury. Activated macrophages secrete pro-inflammatory cytokines (e.g., TNFα, IL-1, IL-6) and chemokines, recruiting other immune cells to the site of injury. Other macrophage populations also play a role in resolving inflammation and promoting tissue repair by producing anti-inflammatory mediators and growth factors. Globally, macrophage accumulation and activation contribute to the development of hypoxia, stimulating the production of additional inflammatory, profibrogenic, and proangiogenic cytokines that can sustain the non-healing tissue response (Arroyo-Hernández et al. 2021; Hill 2005).

Other immune cells play a role in the response to radiation injury and inflammation of the lungs. T cells and B cells are important components of the adaptive immune response to radiation-induced lung injury. T cells are involved in regulating inflammation, tissue repair, and fibrosis. Subsets of T cells, such as regulatory T cells, can suppress excessive inflammation

and promote tissue healing. Dysregulation of lymphocyte responses can influence the severity and progression of lung injury and fibrosis (Wirsdörfer and Jendrossek 2016). Radiation can also directly activate NK cells, enhancing their cytotoxic function against stressed or damaged lung cells. This NK cell activation contributes to the initial inflammatory response following radiation exposure (Boopathi, Den, and Thangavel 2023).

This initial inflammatory phase, known as radiation-induced pneumonitis, typically occurs within 1-3 weeks of radiation exposure and is characterized by clinical symptoms like shortness of breath, dry cough, fever, and chest pain (Hill 2005).

ii. Regeneration of the endothelial compartment

The lung endothelium is highly radiosensitive, and endothelial cell death is an important initiating event during the early stage of RILI. Furthermore, like for other cell populations, irradiation causes DNA damage, oxidative stress, and mitochondrial dysfunction in endothelial cells, leading to even more apoptosis. Radiation exposure upregulates adhesion molecules on endothelial cells, promoting increased interaction with inflammatory cells and their trafficking into the lung tissue (Wijerathne et al. 2021).

Irradiation also impairs the barrier function of endothelial cells by disrupting tight and adherent junctions, resulting in increased vascular permeability and extravasation of proteins into the lung interstitium. As a consequence, endothelial cell dysfunction impairs angiogenesis and vascular remodeling required for proper wound healing in irradiated lung tissue (Korpela and Liu 2014).

Animal studies have shown that targeted protection of lung endothelial cells can mitigate radiation pneumonitis and fibrosis, highlighting the central role of endothelial damage in RILI pathogenesis (Rannou et al. 2015).

iii. Regeneration of the epithelial compartment

Radiation-induced lung injury (RILI) has a significant impact on the alveolar epithelial cells. Radiation exposure causes direct DNA damage and the generation of reactive oxygen species, which can lead to apoptosis and death of AT1 and AT2 cells. This disrupts the epithelial-endothelial barrier function and compromises the integrity of the lung tissue (Yan

et al. 2022). Dysfunction of the epithelial barrier allows for increased permeability, leading to leakage of fluid, proteins, and inflammatory cells into the lung interstitium and alveolar spaces. It contributes to the development of pulmonary edema and inflammation. The release of pro-inflammatory cytokines induces the recruitment and activation of immune cells, such as neutrophils and macrophages, amplifying the inflammatory response in the lung.

Furthermore, while the epithelial cells show some ability to recover over time, the damaged induced by radiation-induced injury to the AT2 cells in particular can prevent full regeneration as they are the main source of new AT1 and AT2 cells (Kadur Lakshminarasimha Murthy et al. 2022).

Radiation exposure can also induce epithelial to mesenchymal transition (EMT) in lung epithelial cells, contributing to tissue remodeling, fibrosis, and metastasis in certain cases. Transitioned cells lose epithelial markers and gain mesenchymal markers, facilitating their migration and invasion into the surrounding tissue (Almeida et al. 2013).

iv. Proliferation of mesenchymal cells and accumulation of extracellular matrix

Radiation-induced lung injury can have significant effects on mesenchymal cells, which play key roles in maintaining lung structure and function, as well as in the response to tissue injury and repair. Radiation exposure can stimulate lung fibroblasts to become activated and differentiate into myofibroblasts. Myofibroblasts play a central role in the development of fibrosis, a pathological process characterized by excessive deposition of collagen and other ECM proteins, leading to tissue scarring and remodeling. Activated fibroblasts/myofibroblasts contribute to the fibrotic response by proliferating, migrating to the site of injury, and secreting profibrotic factors such as TGF-β and connective tissue growth factor (CTGF). Radiation-induced injury alters the composition and organization of the lung extracellular matrix, which provides structural support to the lung parenchyma and regulates cellular behavior. Mesenchymal cells, including fibroblasts, myofibroblasts, and pericytes, participate in remodeling the extracellular matrix by synthesizing and degrading matrix proteins such as collagen, elastin, and fibronectin. Dysregulated ECM remodeling can

contribute to tissue stiffness, impaired gas exchange, and aberrant wound healing processes in the irradiated lung.

Mesenchymal cells can also modulate the inflammatory response to radiation-induced lung injury by secreting cytokines, chemokines, and other immunomodulatory factors. Radiation-exposed mesenchymal cells may exhibit altered expression profiles of inflammatory mediators, influencing the recruitment and activation of immune cells and contributing to the perpetuation of inflammation and tissue damage (Y. Huang et al. 2017).

v. Failure of the healing processes and irreversible fibrosis

The initial stage of RILI and the events described above can evolve in two ways. If the damage is not too extensive, the healing processes are successful, the inflammation is cleared, and the lungs return to a homeostasis situation. However, if the dose of irradiation is too high or the tissues are too radiosensitive, the inflammation can persist, and the conditions evolve to RIPF.

The late stage of radiation-induced pulmonary fibrosis involves a dysregulated wound healing process leading to excessive deposition of extracellular matrix proteins like collagen by activated myofibroblasts. Overall, a persistent inflammatory response can be detected in the lungs affected by RIPF, with recruitment of inflammatory cells like macrophages and lymphocytes to the irradiated lung tissue. These populations contribute to the release of pro-fibrotic cytokines and growth factors like TGF-β by inflammatory cells, which stimulate fibroblast proliferation and differentiation into myofibroblasts. These cells produce excessive amounts of extracellular matrix proteins, especially collagen, leading to fibrosis and architectural distortion of the lung parenchyma. Furthermore, an impaired re-epithelialization of damaged alveolar epithelial cells contributes to the failed resolution of the wound healing process and vascular injury and remodeling also play a role by causing hypoxia and promoting fibroblast activation (Jarzebska et al. 2021; Hanania et al. 2019).

The end result of this process is the replacement of normal lung parenchyma with fibrous scar tissue, leading to loss of lung volume, impaired gas exchange, and potential complications like pulmonary hypertension and respiratory failure. The fibrotic changes are irreversible (Y. W. Choi et al. 2004).

Different elements are used to diagnose RIPF. The patients present a persistent dry cough, shortness of breath, and chest pain, typically developing 6-12 months after completing radiation therapy to the thoracic region (Y. W. Choi et al. 2004). When a radiotherapy-treated patient presents some of these symptoms, an assessment is done with imagery. A CT scan a the chest of a patient suffering from RIPF can show a volume loss of the irradiated lung parenchyma, parenchymal opacities, traction bronchiectasis, and honeycombing (a specific radiological pattern seen in advanced pulmonary fibrosis, characterized by the presence of clustered cystic air spaces with thick, fibrous walls, typically distributed in the peripheral and subpleural regions of the lungs) (Jarzebska et al. 2021). Other affections also need to be ruled out in order to confirm the diagnosis of RIPF, like other causes of pulmonary fibrosis, tuberculosis (if lung apices irradiated) or tumor recurrence.

Initially, these different events were studied with classical experimental methods, like immunostaining or *in vitro* experiments. However, these techniques fail to recapitulate the complexity of the condition, as the lung is a complex organ composed of multiple cell populations presenting a coordinated response to aggressions or diseases. During the past years, modern tools have been developed and allow a more complex and detailed understanding of the lung response to irradiation. Single cell RNA sequencing is particularly powerful to study the numerous lung population, their complex changes in transcriptome after irradiation and their interactions with each other.

4. Single cell RNA sequencing: experimental method and tools for analysis

a) Numerous single cell RNA sequencing experimental methods

Different technologies exist to perform single cell RNA sequencing analysis on samples. Generally, a scRNAseq experiment comprises two steps: single cell isolation and sequencing.

Single cell isolation can be performed using various techniques. The most commonly used is the 10X Chromium method, based on a fluidic droplet-based technique. Once the tissue is dissociated to as single cell suspension, cells are injected into a fluidic system and each cell is encapsulated with barcoded gel beads and enzyme using partitioning oil. Within the droplets, the cells are lysed, the barcodes attach to the 3' polyA end of the mRNA and then undergo reverse transcription to generate barcoded cDNA, with all the generated cDNA from one individual cell sharing a common barcode. The cDNA can then be pooled together to be sequenced. With this method, only the 3' end of the mRNA is sequenced and only mRNAs with a polyA end are sequenced. Another widely used method is Smart-seq2. After the creation of the single cell suspension, the cells are separated into the individual wells of a plate using a fluorescence-activated cell sorting (FACS) sorter or a specialized machine, with one cell per well. After that, the process is similar to the 10X Chromium method. However, in the case of Smart-seq2, the mRNAs are fully sequenced (Picelli et al. 2014). Other techniques that need less equipment are being developed. For instance, the start-up Scipio developed a gel-based method called Asteria. After the creation of the single cell suspension, barcode beads are put into the tube with the cells and each bead will attach to a single cell. The cells are then diluted into a gel solution that after jellification will physically isolate the cell-bead pairs from each other. The cells can then be lyzed and the mRNAs are captured by the nearby bead. They can then be recovered and undergo reverse transcription, amplified and sequenced.

b) Pre-processing: in silico removal of contaminating mRNA, doublets

ScRNAseq data analysis is a growing field with many new tools being developed continuously for all the different steps of the analysis workflow.

After the sequencing of the sample, the first step of the analysis of scRNAseq data is the demultiplexing of the data, the alignment and the transcript quantification. The demultiplexing of the data is the attribution of each RNA sequenced to its droplet, and each droplet sequenced to its sample of origin, using the hashtag oligos (HTO) that were sequenced with the RNA. The reads are then aligned to the chosen reference genome or transcriptome. Finally, the expression matrix can be created by counting the number of reads that share both a unique cell barcode and a unique molecular identifier (UMI), that are assigned to the same gene. The tools performing these different steps are usually grouped in ready-to-use pipelines. The pipeline most frequently used for the 10X generated data is Cell Ranger (Zheng et al. 2017).

Optional steps can then be performed for the pre-processing of the single cell RNA sequencing data. The most common one is the removal of contaminating mRNA. Indeed, when the single cell suspension is generated during the experimental phase, some cells die or are destroyed and release mRNA in the medium. These free mRNA molecules are then encapsulated with the cells in the droplets and are labelled as belonging to this cell. Tools have been developed to informatically remove these background reads that add noise to the data. SoupX (Young and Behjati 2020) is the most frequently used one. This method uses the content of empty droplets (as determined by cellranger through the EmptyDrops method (Lun et al. 2019)) that have been sequenced to characterize the profile of the mRNA contamination of a sample, then determine the percentage of contamination of the data by the mRNA soup, and finally correct the expression profile of each cell by removing the ambient mRNA contamination. The ambient mRNA contamination has been estimated to be approximately five percent.

During the encapsulation process, two cells are sometimes encapsulated in the same droplet. Therefore, they are labelled and sequenced as one unique cell. There are several ways to filter out these cells that can perturb the analysis. First, as two cells contain more mRNAs than one, a first filter can be applied to remove the cells with the highest number of mRNAs. Furthermore, the cells presenting markers for several populations in combination with a high number of mitochondrial genes are also excluded. The value of the threshold is very dependent on the type of tissue, as the number of transcripts considered “normal” can vary greatly between tissues. A threshold that can be used is mean + 2SD (standard

deviations), but it is important to verify that some populations of the experiment don't present a lower or higher transcriptional activity, like cancer cells, to avoid deleting them by accident (AlJanahi, Danielsen, and Dunbar 2018). Tools have also been developed to automatically remove the doublet cells, like DoubletFinder (McGinnis, Murrow, and Gartner 2019). It first generates simulated doublets by combining cells from different clusters and then uses these artificial doublets to identify the real ones and remove them.

c) Quality controls

Different types of quality controls must be performed on single cell RNA sequencing data in order to obtain reliable results during subsequent analyses. First, the remaining empty droplets are eliminated by filtering out the cells with a low number of transcripts (usually less than 20 transcripts). Then, we can also look at the percentage of mitochondrial mRNA per cell to assess the quality of the cells. Cells expressing a high number of mitochondrial mRNA are often damaged or dying cells, therefore they are removed (Q. Zhao et al. 2002). Similarly to the high number of transcripts threshold, "normal" percent of mitochondrial genes depend on the tissue studied and the population considered, so it is better to set not to stringent cutoff values and refine them further in the analysis (AlJanahi, Danielsen, and Dunbar 2018). Several methods have recently been developed to automatically perform these quality controls, like validDrops (Kavaliauskaite and Madsen 2023) or ddqc (Ayshwarya Subramanian et al. 2022). However, these algorithms can have difficulties with the heterogeneity of transcriptional profiles of complex tissues like the lung.

d) Visualization of single cell RNA sequencing data

A single cell RNA sequencing data object can be summarized to a table with a very high dimension (thousands of genes and thousands of cells). Therefore, dimension reduction techniques are needed to be able to visualize the data in a two-dimensional space. Different methods have been developed.

One of the first methods used is the t-distributed stochastic neighbor embedding (tSNE) method (van der Maaten and Hinton 2008). This method is based on the SNE method (Hinton and Roweis 2002) that converts the high-dimensional Euclidean distances between data points into conditional probabilities that represent similarities. Then t-SNE added

symmetry to the matrix and uses a t student distribution to calculate the similarity between two points (**FIGURE 1.7A**).

Another method was later developed: the uniform manifold approximation and projection (UMAP) method (McInnes, Healy, and Melville 2020). This algorithm works in two steps: first, it builds a k-neighbor graph and then computes the low-dimensional representation. It has been shown to allow a more accurate visualization of the structure of the data, together with a better separation of the biologically relevant clusters, with a faster computation time (Becht et al. 2019; Xiang et al. 2021) (**FIGURE 1.7B**). UMAP is currently the most widely used dimension reduction algorithm for single cell RNA seq analysis data visualization.

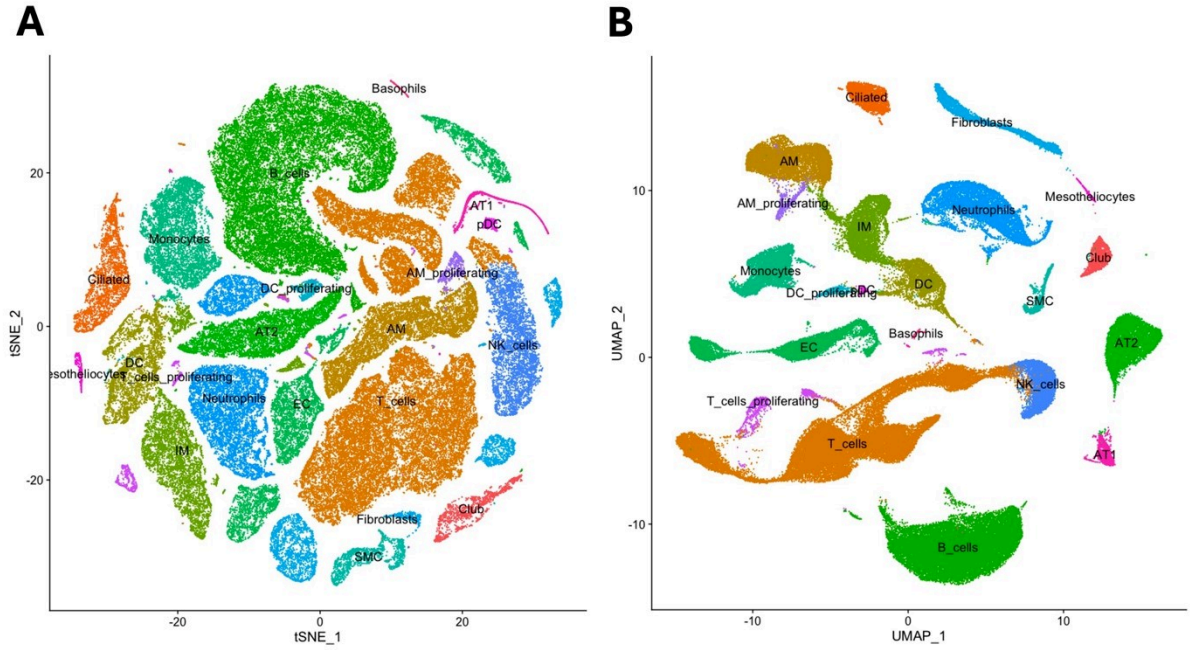


FIGURE 1.7. tSNE and UMAP visualization of the scRNA seq data from Curras-Alonso et al. 2023.
A: tSNE plot of the single cell RNA sequencing data from mice lungs annotated by cell population; B: UMAP plot of the single cell RNA sequencing data from mice lungs annotated by cell population.

e) Merging or integration of the single cell RNA sequencing data

When working with single cell RNA sequencing data, there is often more than one sample, corresponding to different experimental conditions and/or to biological replicates. Therefore, the different samples need to be gathered in a single object for further analysis.

There are two ways of grouping the samples into one object: either merging or integration. The merging of the data will simply put the samples together without modification, while the integration of the data will correct for an eventual batch effect. Batch effect is the difference between the data from two different experiments that is not due to biological variability, but rather to differences between the different experimental setups (like sample handling, preparation). However, the integration of different samples also masks some of the biological variability (Argelaguet et al. 2021). Therefore, it is important to integrate data only when the presence of a batch effect directly compromises the analysis of the data.

i. Determination of the presence or absence of batch effect

There are several ways to determine whether or not there is batch effect in a dataset. The simplest way is to merge the data and generate a UMAP plot of the data labeled with the experiment of origin. If the cells from different experiments (carried out under the same laboratory conditions) cluster and are mixed together, we can usually conclude that there is no batch effect. However, if the same cells from different experiments (but carried out under the same laboratory conditions) are separated, there is likely a strong batch effect in the dataset, and integration of the data is needed before performing analysis. However, with this method, if there are no biological replicates or if the biological replicates have been generated within the same experiment, it can be difficult to distinguish the biological variation and an eventual batch effect.

Therefore, methods have been developed to quantify batch effect and used to decide to merge or integrate the data. The kBET method checks if different random groups of cells present the same distribution of batch labels as the full dataset (Büttner et al. 2019). Another method, CellMixS, fits a linear model for each gene to determine the variability that can be explained by either the cell type or the batch (Lütge et al. 2021). These methods can help to make the decision to integrate or merge the different samples.

ii. Different integration methods

Different methods for integration of single cell RNA sequencing data exist. The most frequently used are the methods developed by Seurat in R (Stuart et al. 2019), and Harmony in Python (Korsunsky et al. 2019). The Seurat method first selects the most variable features

for each dataset (one dataset per experiment), then identifies anchors from the different datasets (anchors are two cells coming from different datasets that are predicted to come from the same cell state), filters, scores, and weighs these anchor correspondences and finally uses these anchors to produce a corrected data matrix. The Harmony algorithm iteratively learns a cell-specific linear correction function. For each iteration, the cells are assigned to a cluster (with favoring of cluster mixing datasets), then the centroid of each cluster is calculated, the cells for the cluster are corrected to cluster next to the centroid, and a new iteration is started with this new data matrix.

A comparative analysis of the different integration methods existing for single cell RNA sequencing data showed that both these methods are very efficient for batch removal, with an acceptable equilibrium with biological variability conservation (Luecken et al. 2022).

f) Annotation: manual, automatic (different methods)

The annotation of the data is a very critical step in the analysis. Indeed, an incorrect identification of the different populations will decrease the quality of all the following steps of the analysis. The first method to annotate the cell populations is to look at the expression of the known markers in the different cell clusters. A cluster expressing markers specific of a given population will be assigned to this given identity. This method is effective but can be time consuming and subject to the interpretation of the person performing the analysis.

That is why automatic annotations tools have been developed. These tools rely on different methods and require different inputs (Pasquini et al. 2021).

Some automatic annotation tools take as a reference a list of markers characteristic of the populations that need to be identified. They then use a scoring system to assign the different cell identities. ScCATCH (Shao et al. 2020) and SCSA (Y. Cao, Wang, and Peng 2020) are two of the tools that allow such analysis in R. Other algorithms for automatic cell type annotation based on list of markers exists. For example, CellAssign relies on a probabilistic Bayesian model (A. W. Zhang et al. 2019). However, the quality of the output of these methods relies on the quality of the marker list and on the expression of these markers in the pathological conditions. The expression of some genes can be modified in a non-physiological condition, therefore degrading the quality of results of this type of automatic annotation.

To take into account the complexity of the transcriptional profile of the cell populations, other tools are based on the correlation between the query dataset and a reference database. Some of these methods use bulk RNA seq data as a database, like SingleR (Aran et al. 2019). SingleR uses reference datasets of pure cell types sequenced by bulk RNA seq and performs correlation analysis between the bulk references and the query dataset.

The most recently developed methods are based on supervised classification and machine learning. The identities of the cell populations are learned on reference objects to be reapplied on new object that need annotation. Different algorithms can be used: SingleCellNet is based on a random forest algorithm (Tan and Cahan 2019), OnClass uses a k-nearest-neighbor algorithm (S. Wang et al. 2019), as well as scClassify (Lin et al. 2020). However, the accuracy of these methods can suffer from a noisy reference, or technical differences between the reference and the query data. Finally, we can use deep learning and transfer learning methods to learn annotation algorithms from the reference. For example, SuperCT is based on an artificial neural networks algorithm and uses the Mouse Cell Atlas as a reference (Xie et al. 2019), and scArches uses fine tuning and architecture surgery (Lotfollahi et al. 2022).

To summarize, if the choice of the automatic annotation tool is crucial, the choice of the reference is even more important: to obtain a quality annotation, the reference needs to be extensive, contain healthy samples but also samples from individuals affected by different conditions or diseases, that can be sequenced using different technologies. The reference needs to be carefully annotated, with precision.

g) Methods for differentially expressed gene analysis

Once the different populations of the object have been annotated, one of the first analysis that is interesting to perform is the comparison of groups of cells: either by comparing two cell populations, or by comparing two experimental conditions within one cell populations. Thus, we can detect the differentially expressed genes (DEG) between two conditions. Different methods have been developed in that purpose (T. Wang et al. 2019). At first, the tools developed for bulk RNA sequencing data analysis have been used to analyze single cell RNA sequencing data, like DESeq (Anders and Huber 2010) or edgeR (Robinson, McCarthy, and Smyth 2010). However, due to the low capture efficiency of single cell RNA

sequencing methods, a phenomenon known as “drop-out”, we can observe transcripts highly expressed in some cells and absent in other, and the data present a high sparsity. So, methods developed of bulk RNA sequencing data might not work very efficiently. Therefore, new methods have been developed to consider the particularities of the data. The MAST (Justice et al. 2019) method has been specifically developed to take into account these drop-out events, as well as DEsingle (Miao et al. 2018) that uses a zero-inflated negative binomial regression model to estimate the proportion of drop-out events. Other methods use distance metrics between the matrices corresponding to the two compared conditions to perform the differential gene expression analysis, like SigEMD (T. Wang and Nabavi 2018) or EMDomics (Nabavi et al. 2016).

h) Trajectory analysis

When studying a biological process and how it is influenced by a perturbation (like a disease or injury), it can be interesting to look at cell differentiation or regeneration. To that purpose, some tools have been developed to perform trajectory analysis: it is the study of the connection that different cell populations can have together. When cells are differentiating into another population, the expression of some genes increases and others decreases. Trajectory analysis uses this information to link together populations that could be deriving from one another. Furthermore, with previous knowledge about the cell population of origin, the cells can be ordered in a pseudo time scale to see the order of events from the cell population of origin to the terminal cell population. Different tools allow to do this kind of analysis, like Monocle3 (J. Cao et al. 2019) or Slingshot (Cannoodt, Saelens, and Saeys 2016).

i) RNA velocity analysis

Trajectory analysis can predict relation between cells at a given time (the moment of the sample collection). However, cells can change state or differentiate in response to a stimulus, and it can be very informative to investigate these dynamics. RNA velocity is an analysis method that allows to make prediction on the next state of a cell (La Manno et al. 2018). This tool uses the results of the sequencing of the different mRNA molecules for the different genes for each cell. For each mRNA molecule, it looks for the presence or absence of introns in the mRNA molecule to determine if the mRNA has already been spliced or not. It then postulates that it is possible to use the balance between the quantity of spliced and

unspliced mRNA molecules for a given gene in a given cell to determine if the expression of this gene is increasing or decreasing. Indeed, if unspliced mRNA molecules of a given gene accumulate in a cell, it means that the process of RNA splicing did not have time to occur yet, and all of these newly produced mRNA are the sign of an increasing expression of the gene. On the contrary, if a cell contains a lot of mature mRNAs for a gene, but only a few unspliced mRNA, it means that the cell used to express the gene highly, but it is decreasing. This increasing or decreasing gene expression is calculated for all genes and all individual cells. With this information, it is possible to predict how the transcriptome of the cells will evolve in the near future, and therefore if the cells are transitioning to a different state.

The RNA velocity analysis is performed in two steps: first the identification of the spliced and unspliced mRNAs, and second the prediction of the next state of the cells. There are different tools that have been developed to perform RNA velocity analysis, and they can do one or both steps of the analysis. The first package to allow the RNA velocity analysis was velocyto (La Manno et al. 2018); others have been developed since, like scVelo (Bergen et al. 2020), or cellDancer that uses a deep neural network to predict individual cells kinetics (S. Li et al. 2024).

j) Gene regulatory network analysis

In the different cells of an organism, genes participate in complex networks of regulation. Transcription factors are proteins able to bind to specific DNA sequences and control the transcription of one or several genes, by increasing or decreasing their expression. Genes encoding for transcription factors can be detected in single cell RNA sequencing data and used to reconstruct the network of gene regulation through analysis of changes after a perturbation. That is the purpose of gene regulatory network (GRN) analysis. SCENIC is a method developed with that objective and consists of three steps. First, a gene correlation analysis is performed to identify the genes whose expression is correlated with the transcription factors expression. Then the second step allows the identification of the direct binding targets. This is done through regulatory motifs analysis: the enriched DNA motives around the transcription start site of the gene of interest are identified and associated to the corresponding transcription factors. Finally, a score is computed for each regulon (group of genes or operons that are regulated by a single regulatory protein or transcription factor) to

quantify their activity. These scores can be compared between the different experimental conditions (Aibar et al. 2017).

k) Intercellular interactions

Intercellular interactions can be predicted using single cell RNA sequencing data, and several tools have been developed to that purpose (Dimitrov et al. 2022).

These methods rely on the use of curated databases of ligands and receptors. The results of the analysis can vary greatly with the composition of the databases. So, when choosing a particular tool, it is important to investigate the process of creation of these databases. Some tools can consider the interactions between a ligand and a receptor formed by a complex. Most can differentiate the secreted versus membrane bound ligands and can predict both autocrine and paracrine interactions. Some of the developed packages also allow the user to input a custom database of ligand-receptor interactions.

Another difference between the packages used for intercellular communication prediction is the scoring methods. The tools use the average expression of the ligand in a given cluster and the proportion of cells from this cluster that express it, and the same parameters for the expression of a receptor in another cluster, to predict if these two clusters are likely to interact through this ligand and receptor. Cellchat (S. Jin, Plikus, and Nie 2023) and CellPhoneDB (Troulé et al. 2023) are the two most widely used tools.

CellChat uses a database of ligand-receptor interactions manually curated from literature that comprises multi-subunits receptors, soluble and membrane-bound ligands, agonists, antagonists and co-receptors. For two given cell populations and one given ligand-receptor pair, an interaction score is calculated with the average expression value of the ligand by one cell population and the average expression value of the receptor by the other cell population, as well as the expression of the possible cofactors. Significance of the interactions is calculated based on a statistical test that randomly permutes the cell populations labels.

CellPhoneDB uses as well different peer-reviewed databases that they then manually curated, like Reactome (Jassal et al. 2020), UniProt (UniProt Consortium 2023) and HMRbase (Rashid et al. 2009), that includes autocrine, paracrine and juxtracrine interactions, with

receptors that can be heterodimer or homodimer complexes. For a given pair of cell populations interacting and ligand-receptor, CellphoneDB returns the mean expression of the interacting partners, and calculates the significance of the interaction by performing random permutations between the population labels.

5. Main objectives

Radio-induced pulmonary fibrosis is a progressive, irreversible disease that can occur as a side effect of thoracic radiotherapy treatment for cancer. It results in fatal respiratory failure for most patients in the few months to years after the radiotherapy. In order to find potential treatments to either cure or slow the progression of RIPF, we need a better understanding of the cellular and molecular events leading to fibrosis. Although some of these mechanisms have been described, the complexity of the implication of the different lung populations and molecular pathways remains to be understood.

My PhD supervisors are the first to propose the study of RILI and RIPF using the single cell RNA sequencing technique. This together with the expertise of my research group in radiobiology as well as our collaborations in the domain of myeloid cells reaction to radiation injury (de Leve et al. 2017), and in the domain of spatial transcriptomics (Defard et al. 2024, **ANNEXE VII**) allowed me to conduct my project in a very favorable environment. Furthermore, the Curie Institute BioinfoHub provided expertise and a place of exchange to share experience and knowledge in bioinformatic procedures.

This support allowed me to develop my project that aims to gain a better understanding in the cellular and molecular mechanisms involved in RILI and RIPF, and more especially to identify the regenerative events involved in lung healing after a radiation injury and decipher the intercellular communication network changes induced by irradiation. In that objective, I used a model or full thoracic mouse irradiation with fibrogenic or non-fibrogenic irradiation dose. We also had access to unique samples of irradiated lung from patients that underwent radiotherapy prior to surgical resection of one lung lobe. These samples were analyzed with the help of the single cell RNA sequencing technology, a valuable tool allowing a better comprehension of the cellular diversity and variability of transcriptome of the samples analyzed.

II

CHAPTER II
Material and methods

CHAPTER II - Material and methods

1. Human samples

Human samples were obtained from patients suffering from Pancoast tumor. This kind of tumor is defined by its location at the apex of the lung and the invasion of neighboring tissues. Due to the high complexity of the tumor, the Institut Mutualiste Montsouris (as well as other hospitals) has devised a treatment protocol in which surgery is preceded by local radiotherapy. The patients receive neoadjuvant chemotherapy and radiotherapy (40-45Gy delivered in daily 2Gy fractions, considered sufficient to trigger radio-induced pulmonary fibrosis). The last session of radiotherapy is received six to eight weeks prior to the surgery (therefore, after the acute inflammation phase and before the fibrotic phase). The surgery consists of a lobectomy of the pulmonary lobe containing the tumor. After the surgery, with the help of the dosimetric CT-scans of the patients, areas in the lobe that received little, or no irradiation are identified, and control samples are taken. Irradiated samples of non-tumor tissue are taken from areas closest to the tumor (but avoiding the tumor itself) that received the highest dose of radiation. Samples of 2 cm3 from each of these regions are immediately placed in cold 1x phosphate buffered saline (DPBS) (Gibco, 14190-094) and transported on ice directly to the research lab for single cell dissociation procedures.

These samples were obtained in collaboration with Institut du Thorax (a joint structure between Institut Curie and Institut Mutualiste Montsouris), as well as Cochin Hospital. Informed consent was obtained from each patient before the surgery.

2. Mice irradiation

To study the sequence of events that occur during the period preceding the installation of full radio-induced pulmonary fibrosis, we used a C57BL/6J female mouse model of lung radiation toxicity. The processes of irradiation, time-resolved fluence measurement, chemical dosimetry, depth-dose distribution, anesthesia of the mouse, mouse immobilization and irradiation of mouse thorax were performed by our physicist Sophie Heinrich according to previous experiments conducted in the lab (Vincent Favaudon et al. 2014; Fouillade et al. 2020). The irradiations were conducted on mice aged 10 to 12 weeks old. To summarize, the

mice were anesthetized with a nose cone using 2.5% isoflurane in air, without adjunction of oxygen. Then the mice were installed for irradiation: they were immobilized in a dorsal position and in a vertical position at a distance of 50cm from the electron source.

Two groups of mice were irradiated using two different irradiators. The first group of mice is composed of five control non-irradiated mice, mice irradiated at 10Gy and sacrificed at one, two, three, four or five months post irradiation (one mouse per time point), and mice irradiated at 17Gy and sacrificed at one, two, three, four or five months post irradiation (two mice per time point), for a total of twenty mice. These mice were irradiated with the Kinetron irradiator, and the dose of irradiation was calculated with a dosimetric film at the exit of the beam, where it enters the lungs. The results of the single cell RNA seq analysis of these mice have been published (Curras-Alonso et al. 2023) and are presented in the **chapter III 2)**.

The second group of mice is composed of two control non-irradiated mice, mice irradiated at 10Gy and sacrificed at one, two, three, four or five months post irradiation (two mice per time point), and mice irradiated at 13Gy and sacrificed at 24 hours, one, two, three, four or five months post irradiation (three mice at 24h, two mice per time point for the months one to four, and one mouse at five months), for a total of twenty-four mice. In addition, three mice were irradiated at 10Gy and sacrificed nineteen months post irradiation, at the same time as four age-matching non-irradiated controls. These mice were irradiated with the new Collimation irradiator, and the dosimetry was performed at entry of the machine the using a gafchromic film. With this method, the mice irradiated at 13Gy present a similar evolution (development of fibrosis five months post irradiation) than the mice irradiated at 17Gy with the Kinetron. The results of the single cell RNA seq analysis of these mice are presented in the rest of this manuscript.

During the few months following the irradiation, the presence or absence of pulmonary fibrosis in the mice lung as well as its level, were monitored using three-dimensional X-ray on the cone beam computed tomography (CBCT) module of the Small Animal Radiation Research Platform (SARRP, Xstrahl). In order to immobilize the mice for imaging, 1.5%-2% isoflurane was used. The mice were maintained in vertical upright position using a poly methyl methacrylate (PMMA) vertical stand. Then a 3D reconstruction of the images was performed with 1,440 projections using the integrated software Murislice

(XStrahl). Then, ImageJ/FIJI (ImageJ, NIH, Bethesda, MD) was used to reconstruct the slices. With the imaging of the mice lungs, the level of fibrosis of the mice lungs were determined, as published previously by the lab (Fouillade et al. 2020).

3. Single cell RNA sequencing data generation

a) Tissue dissociation of the human and mice samples

The tissue dissociation of the mouse and human lung samples were conducted as described in the thesis of Sandra Curras Alonso (Alonso), a former PhD student of the lab:

> *"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 20G 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 1x DPBS MgCl2+ and CaCl2+ (Gibco, 14040-091). Then 320 µl of 25 U/ml elastase (Worthington, LS002292) were added and the suspension was homogenized and incubated for 30 min (45 min for human samples) at 37°C with orbital shaking. Enzymatic activity was inhibited with 5 ml of PF10 (1x DPBS containing 10% fetal bovine serum (FBS)) and 20 µl of 0.5 M ethylenediaminetetraacetic acid (EDTA) pH 8 (Invitrogen, AM9260G). Cell suspension was filtered through 100 µm nylon cell strainer (Fisher Scientific, 22363549), which was rinsed with 5 ml of PF10. This was followed by 37.5 µl 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 µm 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, 11814389001) and incubated for 90 s at room temperature 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 1x 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 1x DPBS containing 0.02% BSA."*

b) FACS sorting of the CD45 negative cells of human samples

The samples from one human patient were sorted to enrich the CD45 cells (non-immune cells) using FACS sorting. The samples were stained using a human anti-CD45

antibody conjugated to the PE fluorophore (368510) diluted 1/100 in 1 ml of phosphate-buffered saline (PBS) with 1% FBS and 1% BSA. Sorting was performed at 4°C using a FACS ARIA sorter (BD Bioscience). A negative sort was performed, allowing us to collect only the cells not marked by the CD45 antibody. Approximately 200,000 to 1 million cells were sorted depending on the samples, and the cell suspensions were then centrifuged at 4°C and 150g for 6 minutes before performing two washes in PBS with 0.02% BSA. Finally, a cell count was performed to adjust the final cell concentration to 1 million/ml for loading onto the Chromium chip.

c) Droplet based single cell RNA sequencing of the mouse and human samples

To analyze the transcriptome of the human and mouse lung samples, we used the 10x genomics chromium droplet-based chemistry V3 reagent kit. The samples were loaded into the Chromium controller (10x Genomics, PN-120237) to capture 6.000 cells. To summarize the procedure, the individual cells from the single cell suspension of lung are encapsulated in droplets called gel beads-in-emulsions (GEMs) with barcoded gel beads using a microfluidic chip. The barcoded gene beads solution contains the reverse transcription (RT) reagents. After encapsulation, the cells are lysed, the gel beads dissolve and the 3' end of the RNAs binds to the poly-dT sequences that coated the gel beads. Then the RNAs undergo reverse transcription to form cDNA. All cDNA from the same cell shares the same 10X barcode, but each individual oligo also comprises a unique molecular identifier (UMI), allowing the distinction between each RNA captured.

After the RT, the droplets are broken and the barcoded cDNAs are purified with silane magnetic beads. Then the cDNAs are amplified by polymerase chain reaction (PCR), re-washed and analyzed on a Bioanalyzer (Agilent) for quality control.

Finally, the last step is the preparation of the libraries using a fixed proportion of the cDNAs. To optimize the cDNA amplicon size by enzymatic fragmentation and size selection. 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 using a HiSeq 2500 (Illumina) or NovaSeq sequencer (Illumina).

4. Single cell RNA sequencing data analysis

a) Cell ranger

After the obtention of the raw sequencing data, the first step was to process the data with the Cell Ranger pipeline (10X Genomics) version (V) 3.2.0, 6.0.0 or 7.1.0 depending on the sample (see **ANNEXE II** for the human samples and **ANNEXE III** for the mouse samples) and the reference genome mm10-2020-A for the mouse samples or GRCh38-3.0.0 or GRCh38-2020-A for the human samples (see **ANNEXE II**). The Illumina sequencing bcl2 files were demultiplexed and mapped on the mm10 mouse reference genome or the GRCh38 human reference genome. This pipeline creates a count matrix for each sample and performs a first filtering of the data to identify the empty droplets that were sequenced.

b) Seurat

We analyzed the single cell RNA sequencing data in R V4.3.0 using Seurat V5.0.1 (Y. Hao et al. 2024). Seurat is an R package that allows the manipulation of the single cell RNA sequencing data manipulation, and also integrates different analysis tools thanks to the packages SeuratData V0.2.2.9001, SeuratDisk V0.0.0.9020, SeuratObject V5.0.1 and SeuratWrappers V0.3.19. The data was loaded in R using the Read10X function.

c) SoupX

As explained in the introduction, at the encapsulation step of the experimental generation of the data, cell-free mRNAs are encapsulated with the cells. We used SoupX V1.6.2 (Young and Behjati 2020) to informatically correct the data for this “mRNA soup”. This step generates a new corrected count matrix for each sample analyzed that will then be used for the following analysis.

d) Data pre-processing

The Seurat object was created for each sample using the function CreateSeuratObject. The mitochondrial genes were identified using the “mt-” prefix for the mouse samples or the “MT-” prefix for the human samples, and the percentage of mitochondrial transcripts was

calculated for each cell. A first filtering was then performed to eliminate the cells with more than 15 percent of mitochondrial mRNA, expressing more than 6000 different genes or less than 200 transcripts. However, different cell populations can present a very variable profile of number of transcripts expressed, number of different genes expressed, or percentage of mitochondrial transcripts expressed. Therefore, we did another round of quality controls analysis for each cluster, once the different samples are put together. For each cluster, we eliminated the cells presenting both a high percentage of mitochondrial mRNA and a small number of transcripts or different genes expressed. Finally, we eliminated the cells expressing markers characteristics of different populations, in order to eliminate the remaining doublets.

e) Cell cycle scoring

We identified the phase of the cell cycle the cell is currently in (G1, S or G2-M). This was done with the method described by Tirosh et al. (2016) and computed by the AddModuleScore function from the Seurat package. This tool returns a score for the S related genes and a score for the G2-M related genes, as well as a prediction of the phase of the cell cycle for each cell.

f) Merging of the mouse samples

As the data from the mouse samples presented no batch effect between the samples sequenced in different experiments, we choose to just merge the samples in order to retain a maximum of biological variability. First, all the mouse samples were merged together in one object. Then we used the SCTransform function (Hafemeister and Satija 2019) to regress the cell cycle scores and normalize the data. The heterogeneity coming from the difference in the phase of the cell cycle in the different cells can complicate the identification of the different cell populations, as the cells can tend to group together based on their phase of the cell cycle instead of the cell population. Therefore, we corrected the count matrix by regressing the effect of the cell cycle on the different genes. Then the data was normalized using the SCTransform method.

On this merged dataset, we performed a dimension reduction analysis using a principal component analysis (PCA) with a number of principal components of 20. Using this result, we computed the 20 nearest neighbors for the dataset using the FindNeighbors

function, and finally we use the FindClusters function to identify different clusters of cells based on similarities or differences in the transcriptome profile.

g) Cell annotation

We used two different methods to annotate the mouse data and the human data.

For the human data, we performed an automatic annotation of the different cell types using the ScArches V0.3.5 python package (Lotfollahi et al. 2022). ScArches is a transfer learning algorithm that has been trained notably on the Human Lung Cell Atlas (HLCA) database (Sikkema et al. 2023), a dataset of human lung samples that have been carefully annotated with various levels of precision. We used ScArches to annotate our human samples one by one. The different identities predicted were then verified using the expression of known cell population markers described in the introduction.

To annotate the mouse data, we tried several automated annotation methods, however none of them gave accurate enough results, probably due to the absence of an extensive database of healthy and disease samples that have been carefully annotated with all the currently known mouse lung populations. Therefore, we performed a manual annotation of the object containing the mouse samples merged together, using the expression of the known markers described in the introduction. In order to identify the markers characteristic of the different clusters determined by Seurat, we used the FindAllMarkers function with a Wilcoxon rank sum test. We obtained a list of genes that have been calculated to be specific of the different clusters, with the log fold change of the expression in the cluster compared to all the other cells, the percentage of cells of the cluster expressing the gene, and the percentage of cells outside of the cluster expressing this gene, as well as a p-value for the specificity of the marker. With all this information, matched with the known markers for the main populations, we were able to annotate the clusters. Then for each population of interest, we extracted the cells of the population and re-did this process to identify the different sub-populations and cell states.

h) Integration of the human samples

Contrary to the mouse data, we found a marked batch effect between the different human patients. This could be due to the difference in treatment of the different patients

and/or the differences between the patients (age, smocking habit, environment differences, genetic background). Therefore, we chose to integrate the different patients using the method developed by Seurat (Stuart et al. 2019). First, we grouped the different samples by patients. For each patient, we normalized the data using the NormalizeData function and we identified the most variable features for each patient using the FindVariableFeatures function. Then, we selected the features that are repeatedly variable across patients. Finally, for the integration itself, we first identified the anchors for integration with the FindIntegrationAnchors and then the IntegrateData function was used to calculate the integrated matrix that will later be used for further analysis.

i) Differential gene expression analysis and gene set enrichment analysis

To compare different conditions during the analysis (for example compare two cell states, compare irradiated cells to non-irradiated cells), we performed DEG analysis using the function FindMarkers and the MAST method (Justice et al. 2019). The adjusted p-value threshold was set to 0.05. This adjusted p-value was also used to determine significant changes in expression for different genes of interest. Then, the genes with the highest positive fold change (= the upregulated genes) or the genes with the lowest negative fold change (= the downregulates genes) were studied by performing a gene set enrichment analysis (GSEA) (Aravind Subramanian et al. 2005; Mootha et al. 2003). This allows us to identify the enriched pathways from different databases in one experimental condition compared to another.

j) Data visualization

In order to visualize the single cell RNA sequencing objects, we performed a dimension reduction technique specialized in scRNA seq data: the visualization called UMAP method (McInnes, Healy, and Melville 2020). This method produces a 2-dimensional representation of the scRNA seq data that groups the cells with similar transcriptome profile and separates the cells with more different transcriptome.

We visualized the expression of the genes of interest using violin plots and the VlnPlot function. The gene expression can also be projected to a UMAP using the FeaturePlot function. To visualize the expression of several genes in several conditions, we also used dot plots using the DotPlot function.

The heatmaps for the plotting of gene expression were created using the ComplexHeatmap package, that allows clustering of the cells or the genes based on the similarity of expression of a chosen set of genes (Z. Gu 2022).

To visualize the result of DEGs analysis, we generated volcano plots using the EnhancedVolcano package (Blighe [2018] 2024). This allowed us to plot the fold changes and the p-values of the DEG analysis.

Finally, diverse kinds of plots (bar plots, scatter plots, dot plots) were created using the ggplot2 package (Wickham 2016).

k) Scoring of pathways

In order to estimate the prevalence of a pathway in individual cells under certain conditions, we used the scoring method provided by Seurat. The different lists of genes corresponding to the pathways, processes or cell states were found in the GSEA databases or different publications. The function AddModuleScore calculates and assigns a score for each cell according to the global expression of the cell of the different genes corresponding to the pathway of interest. This value can then be plotted using violin plots or feature plots.

The statistical comparison of the score between two conditions is computed using a Wilcoxon rank sum test.

l) Data analysis

The trajectory analysis was performed using the Monocle3 V1.3.4 (J. Cao et al. 2019). The intercellular interactions analysis was performed with the CellChat V1.1.3 package (S. Jin, Plikus, and Nie 2023) or the CellPhoneDB V4.0.0 (Troulé et al. 2023). The gene regulatory network analysis was performed using the SCENIC V1.3.1 package (Aibar et al. 2017).

III

CHAPTER III
Results part I

CHAPTER III – Results part I

The lung is a complex organ composed of several specialized cell populations interacting with each other to fulfill its main function: perform the gas exchanges between the outside air and the blood, to absorb the oxygen that will then be brought to the different tissues and organs of the body, and evacuate the waste carried by the blood, mainly composed of carbon dioxide. These gas exchanges are performed by the functional units of the lungs: the alveoli. These terminal structures of the lung are composed of epithelial, endothelial, mesenchymal and immune cells that all play a crucial role in the lung function. These different populations are organized in a specific way that allows a maximum surface for gas exchanges, protection from outside threats and intercellular interactions. However, upon injury, this organization can be disrupted, and the lung function can be compromised. Different mechanisms of injury repair and regeneration exist for the lung to return to a homeostasis state.

Radiation can cause such an injury to the lungs. RILI starts with an inflammation stage, with an important involvement of the immune cells, particularly macrophages. If the repair of the injury is not successful, it can then develop into pulmonary fibrosis with the accumulation of scar tissues, composed mainly of fibroblasts, myofibroblasts and extracellular matrix, with a destruction of the structure of the alveoli. While the main events of the development of radio-induced pulmonary fibrosis have been described, the detailed cellular and molecular events occurring along the progression of the disease remain unknown. A better understanding of these processes would help to develop strategies or treatments to prevent or slow down the development of radio-induced pulmonary fibrosis.

In that mindset, we have here four main objectives: 1) identify the key cellular and molecular alterations and their temporality affecting three crucial populations of the lung: epithelial cells, endothelial cells and macrophages, 2) identify the differences in the response of the lung when exposed to a fibrogenic or a non-fibrogenic dose of irradiation, 3) investigate which of these events are conserved in the response of human lung to radiotherapy, 4) develop and provide access to single cell atlases and datasets of the mouse and human lung tissues responses to radiation injury.

To study the temporality of the events leading to pulmonary fibrosis and the effect of different irradiation doses, we used a mouse model of full thoracic irradiation, at fibrogenic 13Gy dose or non fibrogenic 10Gy dose. After a 13Gy irradiation, all the mice developed pulmonary fibrosis four to five months post irradiation, while 10Gy irradiated mice did not develop pulmonary fibrosis. Samples were collected each month, one to five months post irradiation (**FIGURE 3.1**). The details about the different samples can be found in **ANNEXE III**.

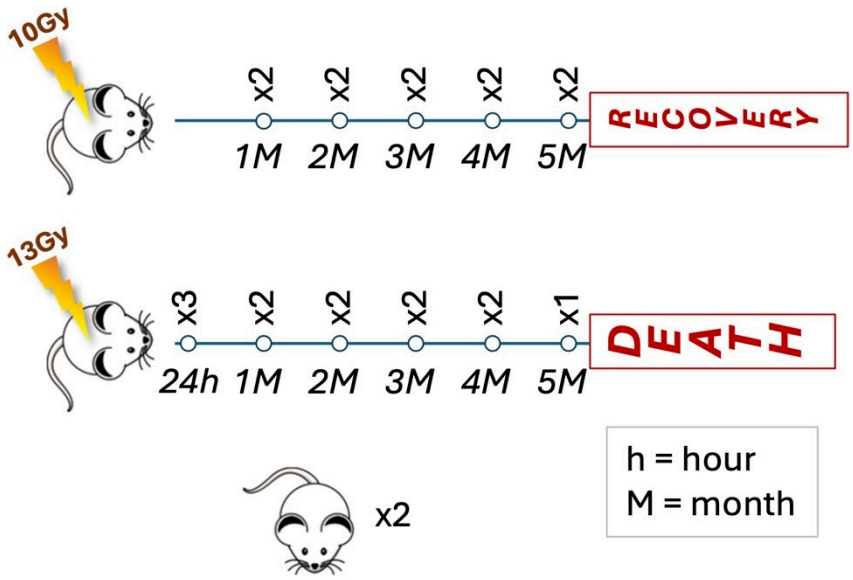


FIGURE 3.1. mouse model used for the analysis of radio-induced pulmonary fibrosis. Collection of 24 time-series mouse samples from full thorax irradiated mice from 24h (24 hours post irradiation) to 5M (5 months post irradiation).

Samples were analyzed using the droplet-based single-cell RNA sequencing method developed by 10X genomics with the V3 protocol.

1. Single cell atlas of the mouse lung response to irradiation

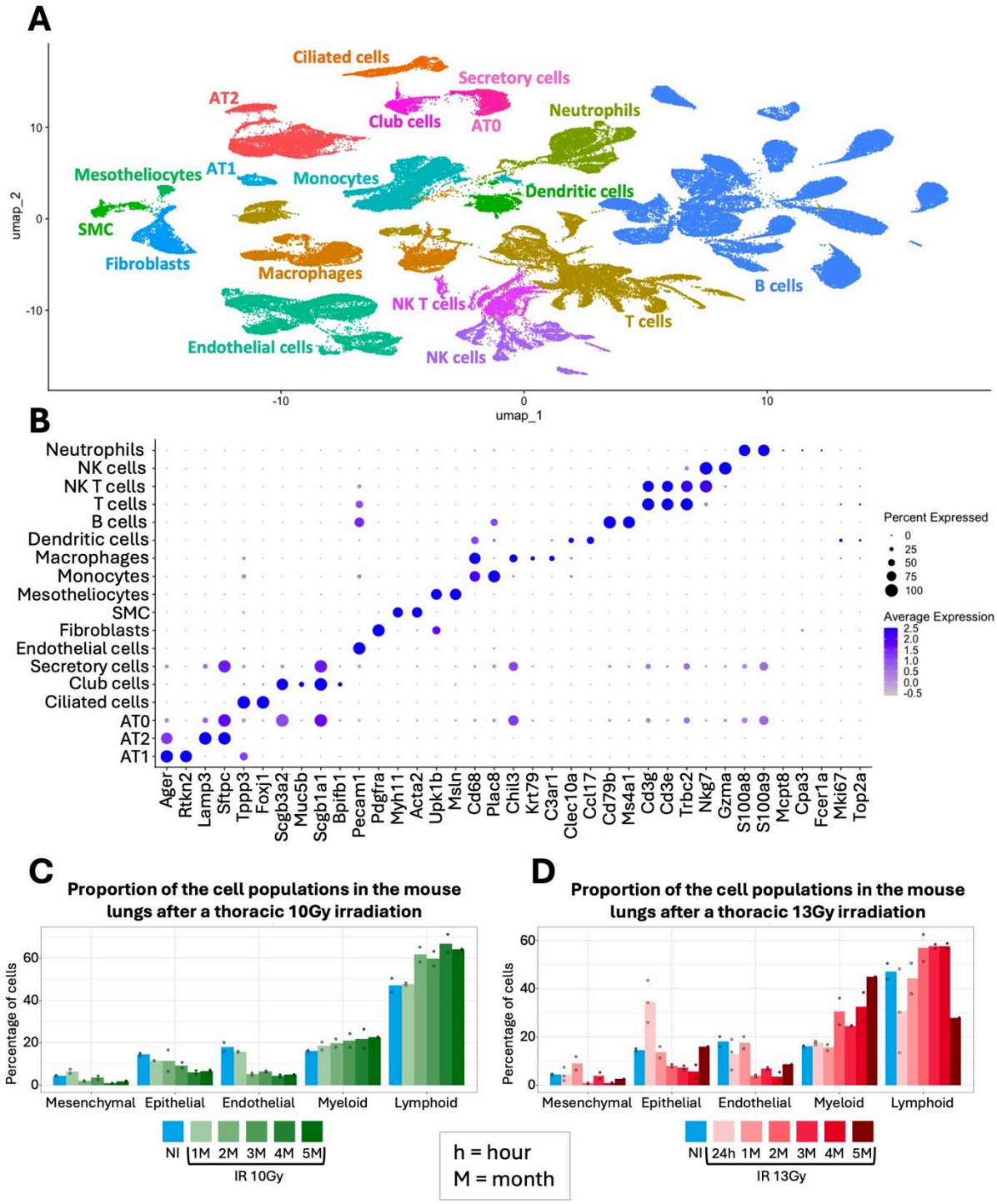


FIGURE 3.2. identification of the different cell populations of the mouse lung.
A: UMAP plot of the single cell RNA sequencing data of the samples from the lungs of 24 mice (123.147 cells), after merging, with the different lung cell populations annotated; B: expression of the canonical markers by the different lung populations; C: Bar plot of the proportion of the different cell populations at the different time points after a 10Gy irradiation (one month-1M to five months-5M post irradiation); D: Bar plot of the proportion of the different cell populations at the different time points after a 13Gy irradiation (one month-1M to five months-5M post irradiation), IR = irradiated

To study the temporality of the events leading to radio toxicity and pulmonary fibrosis, we chose to use a full thoracic irradiation mouse model. As stated above, mice were either irradiated at 10Gy (non fibrogenic dose) or 13Gy (fibrogenic dose), and two non-irradiated mice served as a control. The twenty-four mice were sacrificed at different time points post irradiation (one to five months post irradiation), and the lungs were taken for single cell RNA sequencing analysis. The data obtained were processed with cellranger (Zheng et al. 2017). The different samples were then processed with SoupX (Young and Behjati 2020) for removal of the contaminating mRNA. Then quality controls were performed with filtration of the cells with high mitochondrial content (more than 20% of mitochondrial mRNA, more than 6000 different features expressed, less than 200 different features expressed). All the samples were merged into one Seurat object. As we saw no batch effect between the different experiments, we chose not to integrate the data to preserve a maximum of biological variability. Finally, the different cell populations were annotated using the expression of well-known canonical markers (Travaglini et al. 2020), (**FIGURE 3.2A,B**).

Then, we analyzed the proportion of the different cell populations and their evolution at the different time points after a 10Gy (**FIGURE 3.2C**) or a 13Gy (**FIGURE 3.2D**) irradiation. Contrary to what was expected, the proportion of the mesenchymal population does not present any change during the fibrotic stage (four to five months post 13Gy irradiation). This could be due to the difficulties of dissociation of the extracellular matrix embedded mesenchymal cells. This issue underlines the necessity to be careful when interpreting the changes in population proportions in single cell RNA sequencing experiments.

Furthermore, we observe a drop in the proportion of endothelial cells after irradiation, confirming the observations made for the human samples. Remarkably, 24 hours after a 13Gy irradiation there is an important increase in the proportion of epithelial cells that is consistent within the three replicates.

However, due to dissociation bias (either due to difficulties of dissociation of tissues with high content of extracellular matrix, or destruction of more fragile cell populations), it is difficult to draw strong conclusions from the analysis of cell populations proportions. Therefore, we focused the next analysis on the changes of transcriptome of the different cell populations after irradiation. More precisely, we will focus on several populations that have

been shown to experience major changes during the development of pulmonary fibrosis: the endothelial cells, epithelial cells and macrophages.

Some of the results of these transcriptomic changes happening after irradiation have been published in April 2023 in Nature Communication in the following publication (Curras-Alonso et al. 2023) (see **ANNEXE IV** for the supplementary figures).

In the following publication, we described several populations and how they are affected by irradiation. The samples analyzed are from a previous experiment with mice irradiated using the Kinetron irradiator (see **chapter II 2)** for more precisions).

2. An interactive murine single-cell atlas of the lung responses to radiation injury (*publication*)

nature communications

Article

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

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 physiopathogenesis 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[1,2]. 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[3,4].

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[5]. In parallel, other consortia have gathered efforts to build the Lung Disease Cell Atlas

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

(e.g., the IPF cell atlas[6,7], the COVID cell atlas[8], the COPD cell atlas[9]) 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[10]. 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[11]. 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_{10Gy} mice and 54,131 to the IR_{17Gy} 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[12–14] 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_{10Gy} but it is more progressive and pronounced after IR_{17Gy} (Fig. 2a, Supplementary Fig. 2a). We validated this observation in smFISH experiments where NI, IR_{10Gy}, and IR_{17Gy} 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 $IR5M_{10Gy}$ (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_{17Gy} and to IR_{10Gy} (Fig. 2d). After IR17Gy, the number of upregulated genes increased steadily up to 5M. In comparison, the response to IR_{10Gy} 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*, *Tgfbr2*, *Col6a2*, *Col4a1*, *Col4a2*, *Colgalt1*, *Egfl6*)[15]. To better characterize EMT dynamics in AT2 cells after IR, we selected 31 EMT-associated genes that were significantly upregulated after $IR5M_{17Gy}$ 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_{10Gy}, they only showed strong upregulation at 4 and 5 months after IR_{17Gy} (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 microenvironment 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[16], 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, 1 M after 17 Gy irradiation, 30% of these genes are

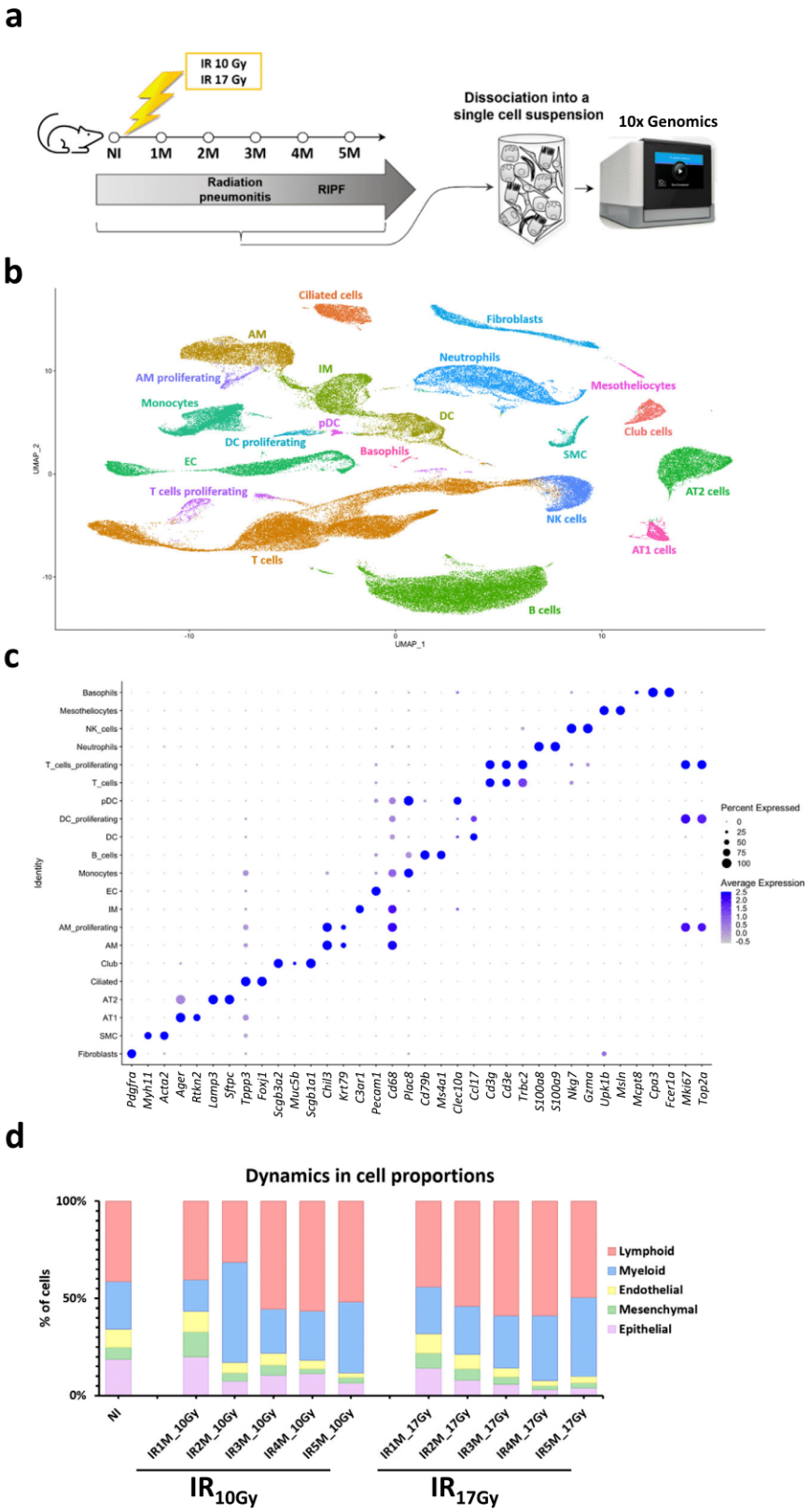


Fig. 1 | Radiation-induced lung injury cell atlas. a Scheme of the experimental set up. Mice are irradiated with a dose of 10 Gy or 17 Gy and monitored by CT-scan to follow the development of pulmonary fibrosis. Mice are sacrificed at 1, 2, 3, 4, and 5 months after IR, and lungs are enzymatically and 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_{10Gy}, one per time point; 10 IR_{17Gy}, 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 5 M after IR_{17Gy} (Supplementary Fig. 2d). This observation lends support to the assumption that a process of AT2 > AT1 transdifferentiation occurs progressively after IR_{17Gy} reaching its strongest expression when fibrosis is already declared (Fig. 2e, Supplementary Fig. 2d). In contrast, when a non-fibrogenic dose is given (IR_{10Gy}), such transdifferentiation appears to be limited both in strength (26.7% of the genes) and time, reaching its maximum at 3 months after IR_{10Gy} (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_{10Gy} ($n = 1$) and IR_{17Gy} ($n = 2$). Error bar refers to the standard deviation of the data. **b** Automatic *Lamp3* mRNA (orange) detection with Big-FISH in NI, $IR5M_{10Gy}$, and $IR5M_{17Gy}$ 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 μm. **c** Quantification and cell volume estimation of the Lamp3+ cells in NI, $IR5M_{10Gy}$, and $IR5M_{17Gy}$ 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 $n = 3$; $IR5M_{10Gy}$ $n = 3$; $IR5M_{17Gy}$ $n = 5$). **d** Dynamics in the significantly upregulated genes in the AT2 cells compared to the NI samples at the different time points after IR_{10Gy} and IR_{17Gy}. **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_{10Gy} and IR_{17Gy}.

(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* + transitional stem cell state that derives from activated AT2 cells and differentiates into AT1 cells[17], 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[18].

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

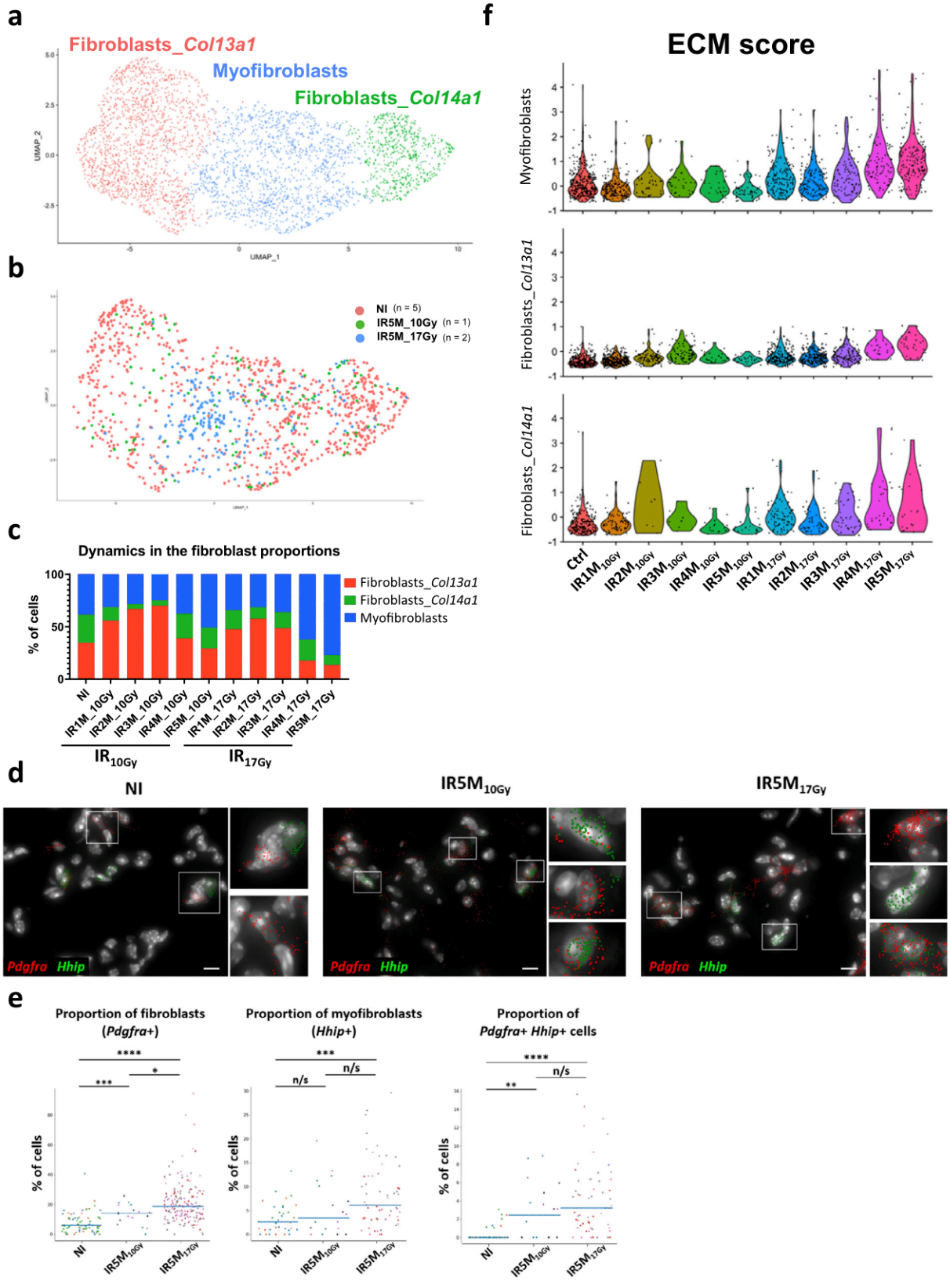
Lung fibroblasts are crucial for maintaining the integrity of the alveolar structure and play key roles in the response to injury through proliferation and remodeling of surrounding tissue. Specific analysis of the fibroblast compartment from all lung samples allowed us to distinguish 3 different sub-populations (Fig. 3a): two different matrix fibroblasts, one *Col13a1*-positive (*Col13a1*+, *Tcf21*+) and one *Col14a1*-positive (*Col14a1*+, *Pi16*+, *Meg3*+)[13], and one subpopulation of myofibroblasts (*Hhip*+, *Cdh11*+, *Pdgfrb*+)[13,19,20] (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% $IR5M_{10Gy}$ and to 4.2% $IR5M_{17Gy}$).

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_{10Gy}, supporting the notion that this is part of the toxic response of fibroblasts specifically associated to lung fibrosis.

Specific macrophage compartments display either proinflammatory or profibrotic profiles after IR17Gy

Analysis of the 11,678 macrophages from NI, IR_{10Gy} and IR_{17Gy} 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 homogenously 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*)[21–23], as well as high levels of proinflammatory cytokines and chemokines (e.g., *Nfkb1*) mainly after IR_{17Gy}, and to a lesser extent after IR_{10Gy} (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_{17Gy} (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*)[21], 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 pro-fibrotic 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_{17Gy} (from 459 µm^3 to 925 µm^3), while it remains constant after

Fig. 3 | Myofibroblasts contribute to the ECM deposition after IR_{17Gy}. **a** UMAP visualization of 3488 cells from the different fibroblast subpopulations annotated by cell type. **b** UMAP visualization of NI ($n = 5$), $IR5M_{10Gy}$ ($n = 1$) and $IR5M_{17Gy}$ ($n = 2$) fibroblasts annotated by time point. **c** Dynamics in the proportion of the fibroblast subpopulations at the different time points after IR_{10Gy} and IR_{17Gy}. **d** Automatic *Pdgfra* (red) and *Hhip* (green) mRNA detection with Big-FISH in NI, $IR5M_{10Gy}$, and $IR5M_{17Gy}$ lung tissue sections. Scale bars, 10 μm. **e** Quantification of the *Pdgfra*+, *Hhip*+ and *Pdgfra*+ *Hhip*+ cells in the NI, $IR5M_{10Gy}$ and $IR5M_{17Gy}$ 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 $n = 3$; $IR5M_{10Gy}$ $n = 3$; $IR5M_{17Gy}$ $n = 5$). **f** Violin plot showing the single cell score calculated based on the ECM expressed genes in the myofibroblasts, fibroblasts *Col13a1*, and fibroblasts *Col14a1*.

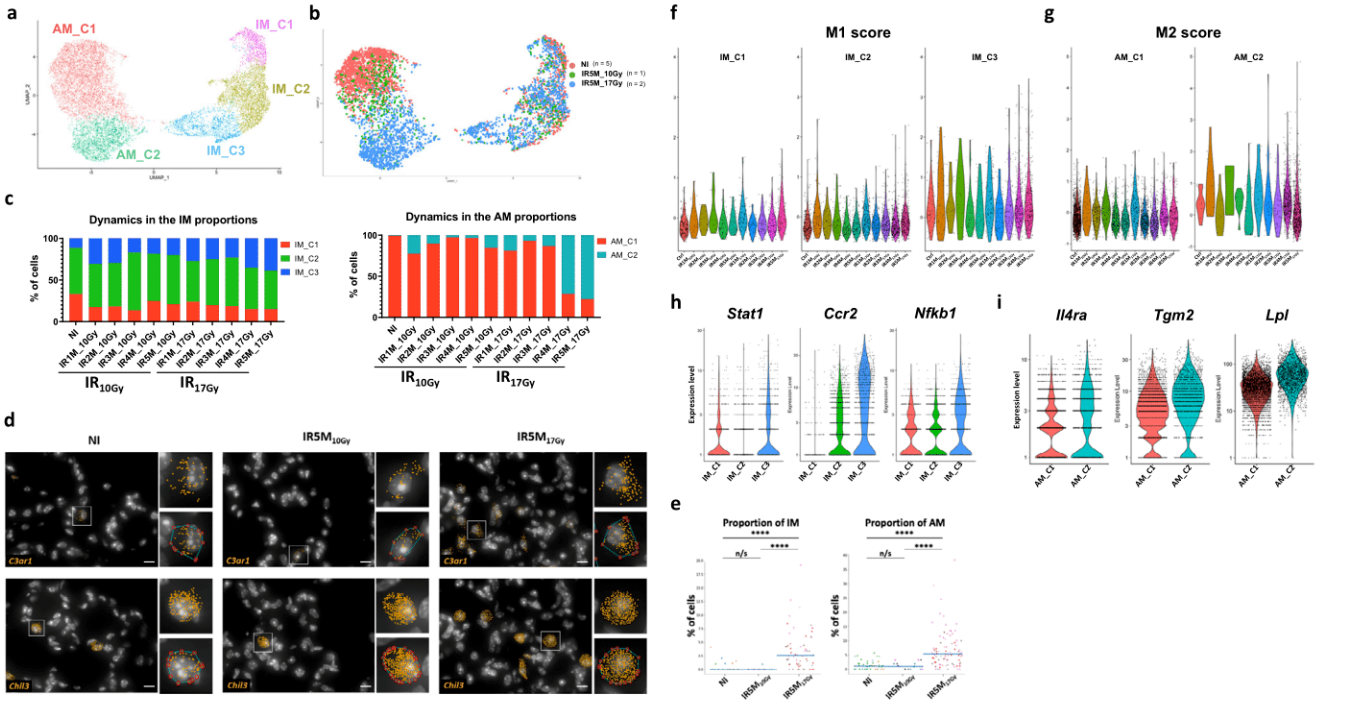


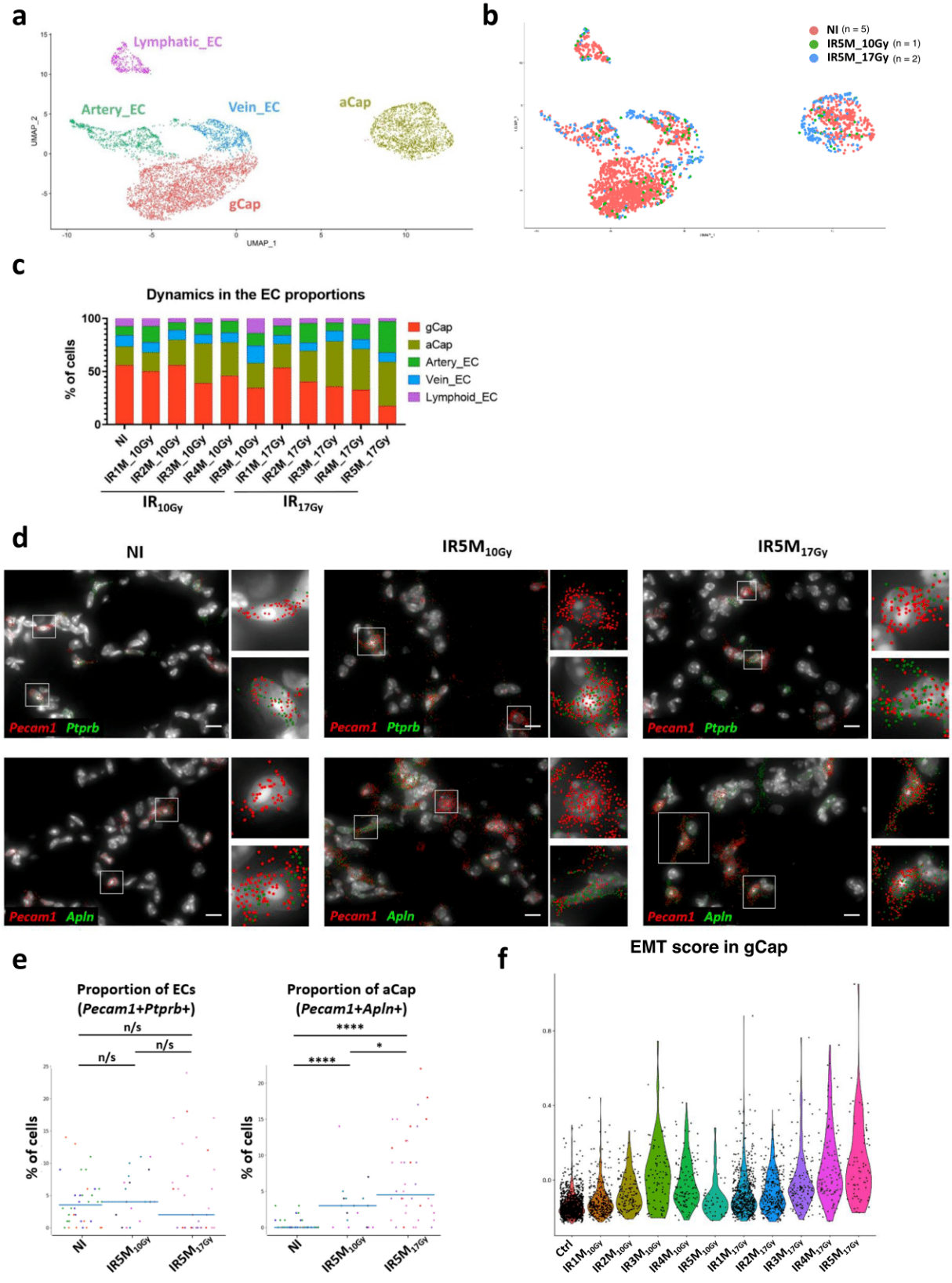
Fig. 4 | Proinflammatory and profibrotic profile of alveolar and interstitial macrophages after fibrogenic doses of IR. **a** UMAP visualization of 11,678 cells from the different IM and AM subpopulations annotated by cell type. **b** UMAP visualization of NI ($n = 5$), $IR5M_{10Gy}$ ($n = 1$) and $IR5M_{17Gy}$ ($n = 2$) 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_{10Gy} and IR_{17Gy}. **d** Automatic *C3ar1* (orange) and *Chil3* (orange) mRNA detection with Big-FISH in NI, $IR5M_{10Gy}$ and $IR5M_{17Gy}$ lung tissue sections. Scale bars, 10 μm. **e** Quantification of the *C3ar1*+ and *Chil3*+ cells in the NI, $IR5M_{10Gy}$, and $IR5M_{17Gy}$ 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 $n = 3$; $IR5M_{10Gy}$ $n = 3$; $IR5M_{17Gy}$ $n = 5$). **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.

$IR5M_{10Gy}$ (Supplementary Fig. 4d). Moreover, the spatial distribution of AMs showed major changes in $IR5M_{17Gy}$. While in the NI control and $IR5M_{10Gy}$ lungs individual AMs are localized within the alveolar compartment, enlarged AMs are gathered in clusters in the fibrotic tissue 5 months after IR_{17Gy} (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_{10Gy}, and IR_{17Gy} 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 *Vwf* 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_{17Gy} (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_{10Gy}.

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*+ *Ptprb*+ 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* + *Apln* +) 5 months after IR, both at 10 and 17 Gy (Fig. 5d, e). Moreover, immunohistochemistry of NI and $IR5M_{17Gy}$ 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_{17Gy}, reaching its maximum 5 M post-IR (Fig. 5f). On the contrary, after IR_{10Gy} 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 ($n = 5$), $IR5M_{10Gy}$ ($n = 1$) and $IR5M_{17Gy}$ ($n = 2$) ECs annotated by time point. **c** Dynamics in the proportion of the EC subpopulations at the different time points after IR_{10Gy} and IR_{17Gy}. **d** Automatic *Pecam1* (red), *Ptprb* (green), and *Apln* (green) mRNA detection with Big-FISH in NI, $IR5M_{10Gy}$, and $IR5M_{17Gy}$ lung tissue sections. Scale bars, 10 μm. **e** Quantification of the *Pecam1* + *Ptprb* + and *Pecam1* + *Apln* + cells in the NI, $IR5M_{10Gy}$ and $IR5M_{17Gy}$ 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 $n = 3$; $IR5M_{10Gy}$ $n = 3$; $IR5M_{17Gy}$ $n = 3$). **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_{10Gy} and IR_{17Gy}.

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[24] 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_{17Gy} compared to NI (Supplementary Fig. 6b) and to IR_{10Gy} (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_{17Gy} compared to IR_{10Gy} (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_{17Gy} and absent in NI and IR_{10Gy} 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/), built using the R package ShinyCell[25], 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, $IR1M_{10Gy}$, $IR2M_{10Gy}$, $IR3M_{10Gy}$, $IR4M_{10Gy}$, $IR5M_{10Gy}$, $IR1M_{17Gy}$, $IR2M_{17Gy}$, $IR3M_{17Gy}$, $IR4M_{17Gy}$, $IR5M_{17Gy}$), 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.

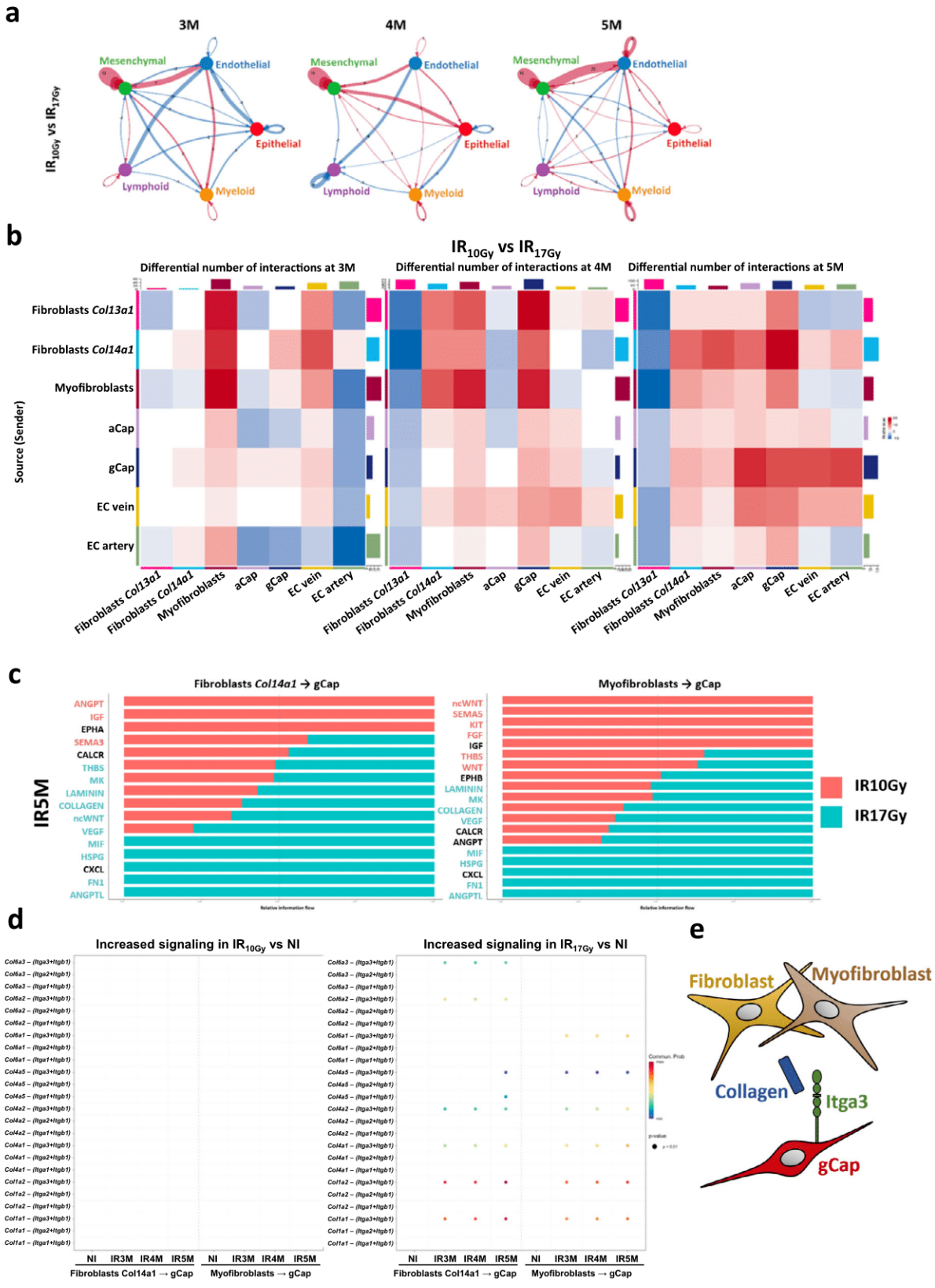
This site also allows a straightforward exploration of the differences in cell proportions using the *Proportion plot* tab, in which the proportions of the main cell compartments, the different cell types (Fig. 7c), and sub cell types along the different conditions (i.e., different time points and doses 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_{10Gy} and IR_{17Gy} mice. Cell type annotations based on published studies identify the 21 main lung cell populations already described using scRNA-seq approaches[26–28]. 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_{10Gy}) versus a fibrogenic dose (IR_{17Gy}) 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[29–32].



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[33]. 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[30]. 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_{10Gy} and IR_{17Gy} 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_{17Gy} compared to the IR_{10Gy}. **b** Heatmap showing the differential number of interactions between IR_{10Gy} and IR_{17Gy} 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_{17Gy} compared to the IR_{10Gy}. 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 $IR5M_{10Gy}$ and $IR5M_{17Gy}$ from the Fibroblasts *Col14a1* and Myofibroblasts (sources) to the gCap (targets). The top signaling pathways colored red are enriched after $IR5M_{10Gy}$, and the ones colored green were enriched after $IR5M_{17Gy}$. **d** Increased signaling ligand-receptor pairs of the Collagen pathway in IR_{10Gy} and IR_{17Gy} 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[34,35].

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[17]. The fact that AT2 cells are known to give rise to AT1 cells[36] and that, after lipopolysaccharide (LPS)-induced lung injury, AT2 cells activate a similar transdifferentiation program[16] 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[37]. 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[38] and a previous study identified hypoxia and TGFβ as EndoMT inducers after radiation injury[39]. 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[40]. 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 RIPF.

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/6 J 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/6 J female mouse model of lung radiation toxicities were used[11]. 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[41,42]. 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 (Kinetron).

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× DPBS $MgCl^{2+}$ and $CaCl^{2+}$ (Gibco, 14040-091). Then 320 µl 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× DPBS containing 10% fetal bovine serum (FBS)) and 20 µl of 0,5 M EDTA pH 8 (Invitrogen, AM9260G). Cell suspension was filtered through 100 µm nylon cell strainer (Fisher Scientific, 22363549), which was rinsed with 5 ml of PF10. This was followed by 37.5 µl 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 µm 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,

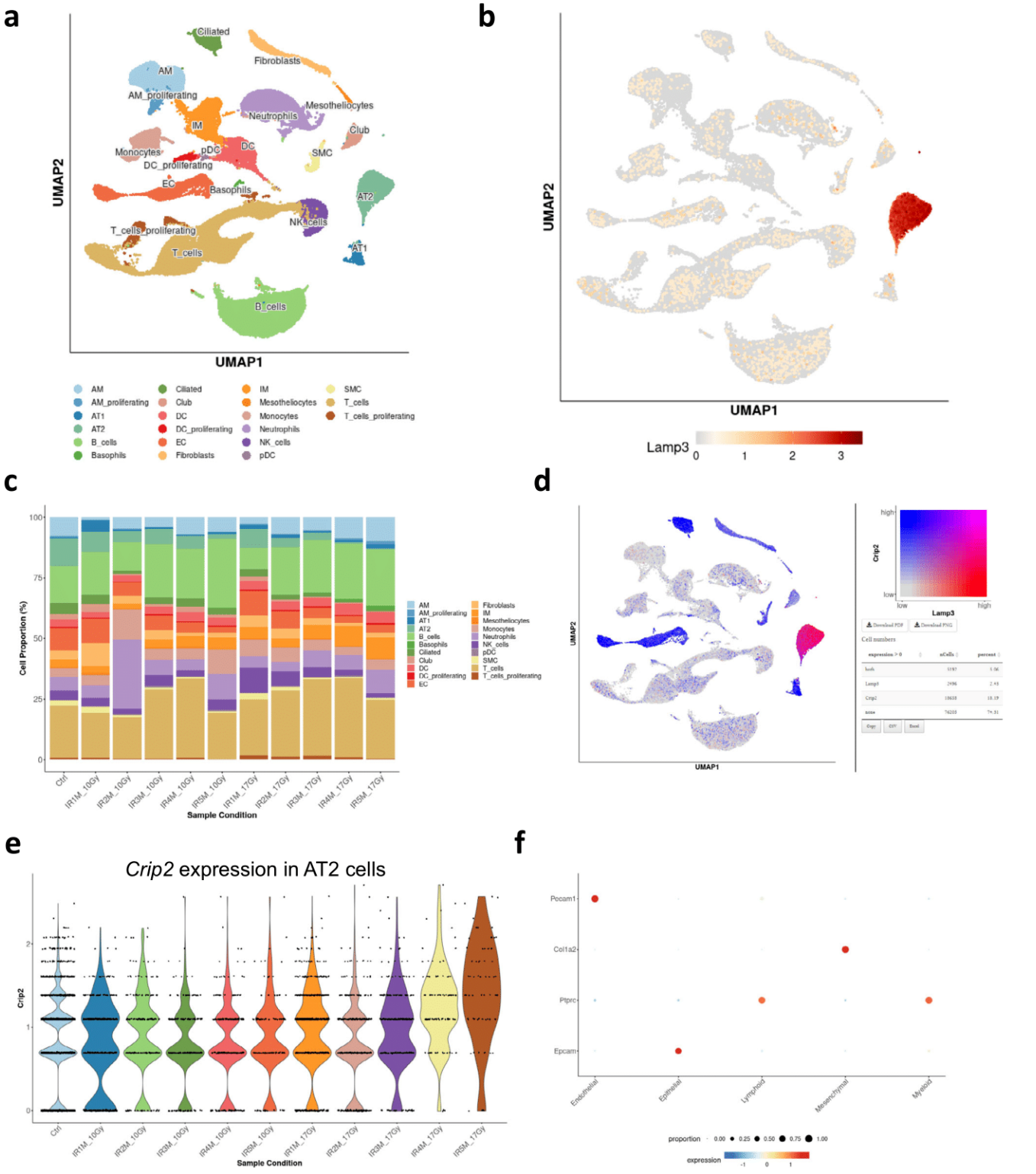


Fig. 7 | Web-based interface for the murine single-cell atlas of the lung response to radiation injury. a UMAP visualization of 102,869 cells from 20 different samples (5 NI; 5 IR_{10Gy}, one per time point; 10 IR_{17Gy}, two per time point) annotated by cell type. **b** UMAP visualization of the expression 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_{10Gy} and IR_{17Gy}. **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_{10Gy}, or IR_{17Gy} 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[43].

In order to remove the contamination from the "soup" of cell-free RNAs, we used SoupX[44]. 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[45]. 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[46]. 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[12–14]. 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[47]. 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[48]. The pseudotime analysis of the AT1 and AT2 cells was done with Monocle 2[49], 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[50]. Lastly, all cell-cell interactions analysis were done using CellChat[24].

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[51]. 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 $\Delta G_{37\,°C}$ 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))[51].

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 μM 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 μM 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 μM. Primary and secondary probe stocks were stored at −20 °C.

smFISH on lung tissue sections

OCT-embedded mouse lung lobes were cut into 16 μm 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 × 20, #1) three times in ethanol (VWR, 20821.310), acetone (Honeywell, 32201), and water. Then, they were sonicated in 1 M KOH (Honeywell, 06005) in H_2O, rinsed with water and dried in an oven at 70 °C for 10 min. For tissue attachment, coverslips were then coated with 2% (3-aminopropyl)triethoxysilane (APTS) (Sigma, A3648-100ML) in H_2O for 2 min and rinsed with water twice. Finally, the coating was activated in the oven at 70 °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 °C, washed twice with cold 1× PBS, and kept in 70% ethanol at 4 °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× NEBuffer 3 (BioLabs, B7003S) and ultra-pure water was incubated in a thermocycler for 3 min at 85 °C, 3 min at 65 °C and 5 min at 25 °C. The hybridization mix was prepared by mixing 2 μl of the smFISH probes and 98 μl of hybridization buffer (100 mg/ml dextran sulfate (Sigma, D8906-10G) and 10% deionized formamide (Invitrogen, AM9342) in 2× SSC (Invitrogen, AM9763)) per sample. Hybridization buffer can be prepared in advance and frozen in aliquots. These aliquots were thaw and heated to 100 °C for 5 min, and let cool to RT. In case of co-staining, 96 μl of hybridization buffer was mixed with 2 μl of primary probes hybridized to FLAP Y-Cy3 and 2 μl 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× SSC) for 3 min at RT, followed by a last incubation in washing buffer II (WBII) (10% deionized formamide in 2× 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 μl droplet of the hybridization mix and incubated at 37 °C o/n.

To finish, tissue sections were washed three times: first in pre-heated WBII at 37 °C for 30 min, second in WBII containing 0.05 μg/μl DAPI (Sigma, D9542-1MG) at 37 °C for 30 min, and third in 1× PBS for 5 min at RT. Samples were mounted in SlowFade Diamonds (Invitrogen #S36963) and stored at 4 °C until image acquisition.

Image acquisition

For each sample, 54 three-dimensional image stacks were captured with an interval of 0.3 μm (total of 16.2 μm) on a widefield microscope (Upright Widefield Apotome Zeiss) equipped with a 63 × 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 × 5 FOV, 16 three-dimensional image stacks were captured with an interval of 0.3 μm (total of 4.8 μm).

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[52] and RNA detection was performed with FISH quant v.2[53]. 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[54] 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_{10Gy}, or IR_{17Gy} 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. Processed data can be explored through an interactive web interface (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 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 (2020).
8. Unterman, A. et al. Single-cell multi-omics reveals dyssynchrony 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 TGFβ 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. Prêle, 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**, giaa151 (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.

Later during my PhD, I performed a deeper analysis of the phenomena described in the publication above, as well as described new ones, using a new group of mice irradiated with the new Collimation irradiator (see **chapter II 2)** for more precisions).

3. Processus of inflammation induced by irradiation

a) Immune cells in the healthy mouse lung

Immune cells are the main actor of the response to outside threat and the maintaining and clearing of inflammation.

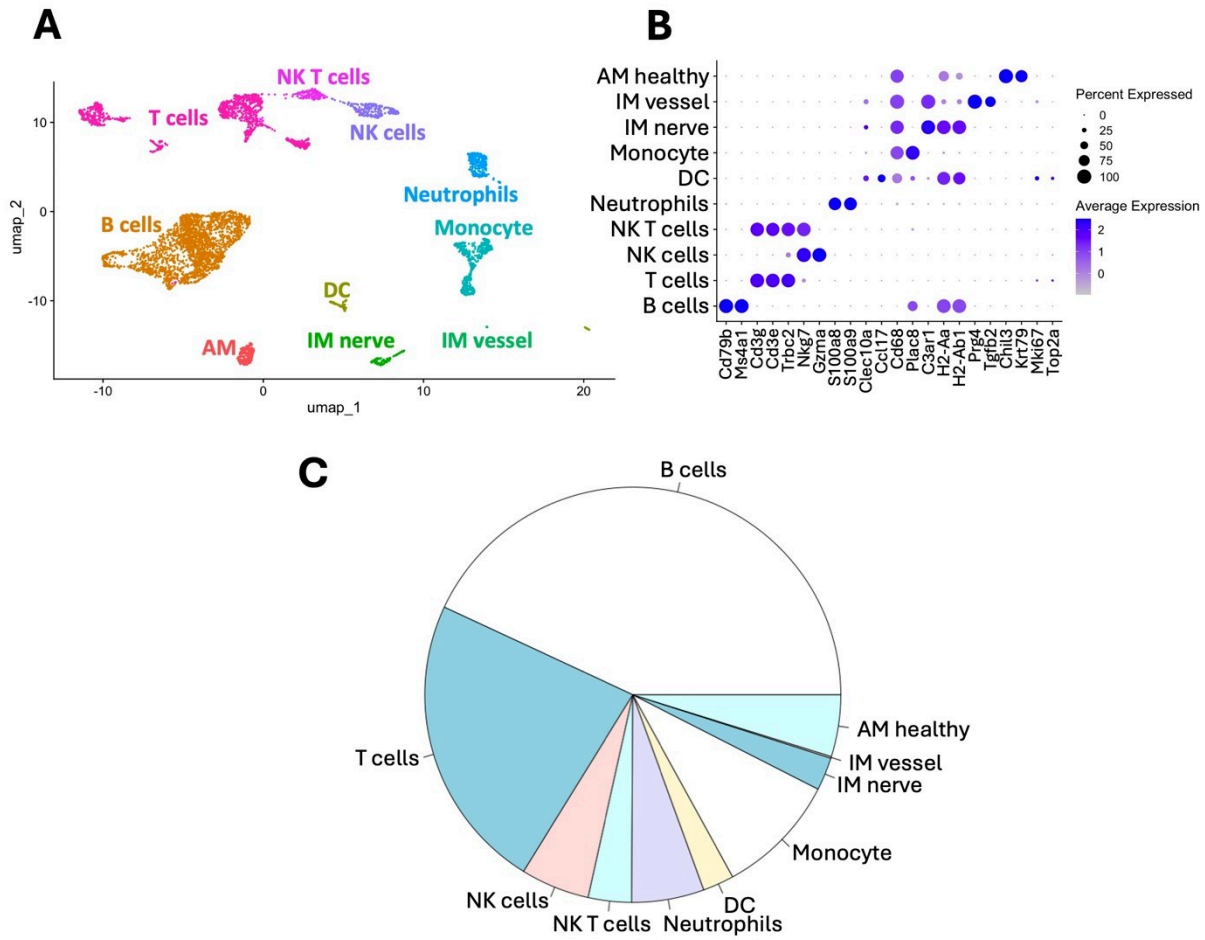


FIGURE 3.3. identification of the different immune cell population in the non-irradiated control mouse lung.
A: UMAP plot of the different immune cell populations (2 samples, 5.243 cell); B: Expression of the markers used for the identification of the immune cell populations; C: Proportion of the different immune cell populations in the non-irradiated mouse lung.

We identified different populations of immune cells from the lymphoid and myeloid lineages in the lungs of non-irradiated mice (**FIGURE 3.3A**): The B cells, T cells, NK cells, NK T

cells, neutrophils, dendritic cells, monocytes, two different populations of interstitial macrophages (the nerve IM and the vessel IM), and alveolar macrophages. These populations have been identified using markers described in the literature (**FIGURE 3.3B**). Contrary to the IM, the AM form one homogenous population.

The main immune population detected is the B cells, accounting for 43 percent of the pool of immune cells. Altogether, lymphoid cells represent 75 percent of the immune cells we detected in the lungs of non-irradiated cells. The other 25 percent are the myeloid cells: mainly monocytes and macrophages, with also neutrophils and dendritic cells. Most of the IM are nerve IM, but a few cells were also identified as vessel IM (**FIGURE 3.3C**).

b) Upregulation of inflammatory pathways in the immune populations after irradiation in mouse lungs

Inflammation is a major process in the initial reaction of the lung to irradiation and the development of pulmonary fibrosis. Therefore, we investigated the changes in transcriptome of the main immune populations of the lung. First, we identified in the T cells an upregulation of oxidative phosphorylation, the MAPK signaling, oxidative stress and TGFβ signaling after both a 10Gy and a 13Gy irradiation. Other upregulated pathways are specific to the 13Gy condition: the p53 signaling pathway and the IL-5 signaling pathway (**FIGURE 3.4A**). Similarly, in B cells, most of the pathways are upregulated after both a 10Gy and a 13Gy irradiation: oxidative phosphorylation, WNT, IL5, TGFβ signaling pathways. There doesn't seem to be upregulation of pathways specific to a 13Gy irradiation (**FIGURE 3.4B**). Dendritic cells exhibit less changes post irradiation, with fewer pathways upregulated, the main one being oxidative phosphorylation (**FIGURE 3.4C**). Finally, in the neutrophils, several signaling pathways are upregulated after both 10Gy and 13Gy irradiation: several interleukins signaling pathways, the TGFβ signaling pathway. Interestingly, the IFNγ signaling pathway and oxidative phosphorylation are specifically upregulated after a 13Gy irradiation (**FIGURE 3.4D**).

As expected, these different immune populations show an upregulation of diverse pathways related to inflammation, like interleukins or TGFβ. Furthermore, all four populations present an upregulation of oxidative phosphorylation, a process that has been shown to be induced by irradiation and that is an indicator of possible mitochondrial damage (Yin et al. 2019).

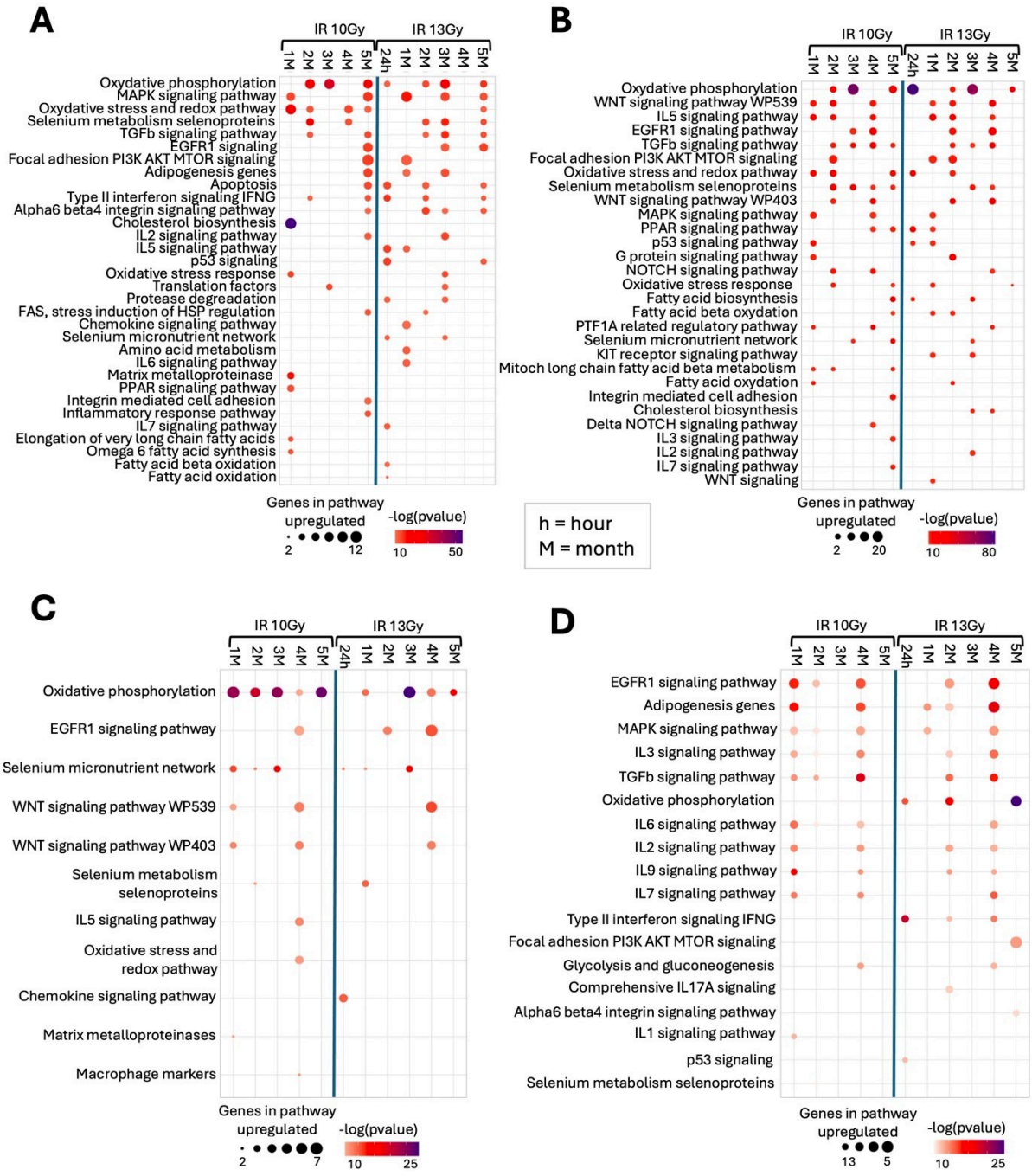


FIGURE 3.4. pathways upregulated in the different immune populations after irradiation: inflammation, oxidative phosphorylation, cytokines
Pathways (from the GSEA WikiPathways database) significantly upregulated (false discovery rate (FDR) adjusted p-value < 0.05) compared to controls at the different time points post irradiation in A: T cells, B: B cells, C: dendritic cells, D: neutrophils.

c) M1 or M2 activated phenotypes of the macrophages after irradiation

Distinct types of macrophages have been shown to play different roles during the development of radio-induced pulmonary fibrosis: M1-activated macrophages promote

inflammation, whereas M2-activated macrophages present a pro-fibrotic activity (Ying, Fang, and Chen 2021). Therefore, we investigated the profile of the macrophages in our model of mouse full thoracic irradiation (**Curras-Alonso et al. 2023**). We found a sub population of alveolar macrophages appearing in the latest months after a fibrogenic dose of irradiation that expresses genes characteristic of the M2 activated macrophages state. Similarly, the interstitial macrophages become M1-activated four to five months post irradiation. So, in our model, interstitial macrophages seem to have a pro-inflammatory activity, whereas alveolar macrophages present a pro-fibrotic activity. These results are presented in the publication displayed in **chapter III 2)**.

d) Identification of different sub populations of macrophages

To better characterize the different populations of macrophages, we investigated how they are affected by irradiation injury and the changes they experience during the healing mechanisms and the processes leading to pulmonary fibrosis.

We can identify the classical macrophage lung cell populations: the alveolar macrophages and the interstitial macrophages (**FIGURE 3.5A**) using well known markers (**FIGURE 3.5C**) (Travaglini et al. 2020). Furthermore, the interstitial macrophage population can be sub-divided in three populations: nerve-associated interstitial macrophages expressing H2-Aa H2-Ab1 Lgals3 and Cd81, vessel-associated interstitial macrophages expressing Prg4 and Tgfb2, and an intermediate population composed of cells expressing markers characteristics of both interstitial macrophages and monocytes (**FIGURE 3.5B, D**). The intermediate monocyte – interstitial macrophages are mainly present 24 hours after a 13Gy irradiation. We can make the hypothesis that they are part of the early healing processes, during which the monocytes are able to differentiate into interstitial macrophages to replenish the pool of this population (Shi et al. 2021). However, due to the lack of samples irradiated at 10Gy at the same time point, it is not possible to determine whether this process is specific for pro-fibrotic condition. The other two populations of interstitial macrophages present a stable proportion after 10Gy and 17Gy irradiation (**FIGURE 3.5F,G**).

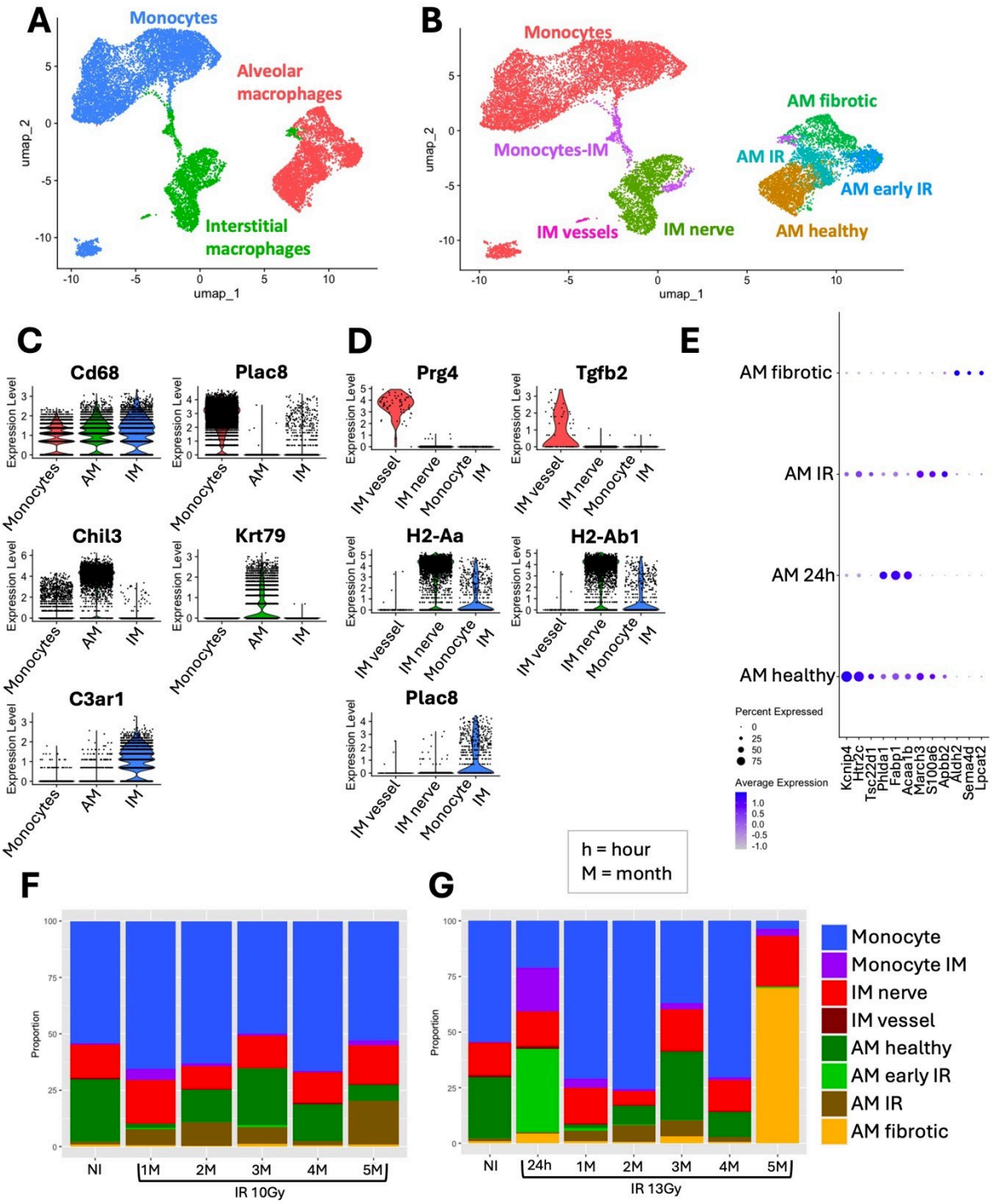


FIGURE 3.5. identification of the different macrophage populations in the mouse lung before and after irradiation.

A: UMAP plot of the different lung macrophage and monocyte populations (24 samples, 17.517 cells); B: UMAP plot of the different lung macrophage and monocyte sub populations; C: Expression of the markers used for the identification of the alveolar macrophages, interstitial macrophages and monocytes; D: Expression of the markers used for the identification of the sub populations of interstitial macrophages; E: Expression of markers characteristics of the sub populations of alveolar macrophages determined using the function FindAllMarkers; F: Bar plot of the proportion of the different macrophage and monocyte populations at the different time points after a 10Gy irradiation; G: Bar plot of the proportion of the different macrophage and monocyte populations at the different time points after a 13Gy irradiation.

Finally, the alveolar macrophages can also be divided into four sub-populations. Each sub-population appears to be specific to a timepoint post irradiation and/or an intensity of irradiation (**FIGURE 3.5F,G**): the AM healthy are mainly found in the controls, the AM 24h are only found in the samples 24h post irradiation, the AM IR are found across all irradiated samples but not in the controls, and finally the AM fibrotic are only found in the samples five months post 13Gy irradiation. Using the function FindAllMarkers from Seurat, different markers characteristics of these populations were determined (**FIGURE 3.5E**).

e) Interstitial macrophages regeneration

We had a closer look at the intermediate monocyte and IM population. With a trajectory analysis, we identified a single trajectory line connecting the interstitial macrophages and monocytes, going through this intermediate population (**FIGURE 3.6A**).

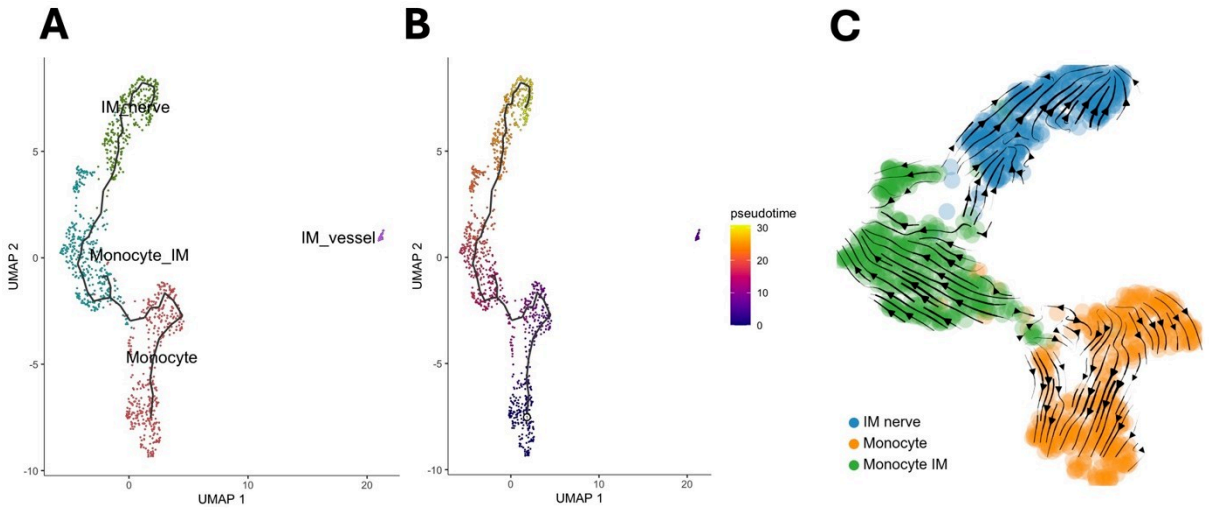


FIGURE 3.6. a potential recruitment of interstitial macrophages from monocytes 24 hours post irradiation
A: trajectory analysis of the IM and monocytes in the samples taken 24 hours after irradiation (3 samples, 1.122 cells); B: pseudo-time analysis of the IM and monocytes in the samples taken 24 hours after irradiation with the AT2 cells as the population of origin; C: RNA velocity analysis of the IM and monocytes in the samples taken 24 hours after irradiation

Furthermore, when placing the origin of the pseudo-time in the monocytes, we find the nerve-associated interstitial macrophages at the end of the pseudo-time (**FIGURE 3.6B**). Finally, the RNA velocity analysis shows cells velocities going from the monocytes to the intermediate population and from this population to the nerve interstitial macrophages (**FIGURE 3.6C**). Therefore, we can hypothesize that after irradiation there is a recruitment of

circulating monocytes that will differentiate into interstitial macrophages, particularly during the early response to irradiation, since the intermediate monocytes-IM population is mainly present at 24h post irradiation (**FIGURE 3.6G**).

f) Increased catabolism and lipid metabolism in alveolar macrophages after irradiation

Then, we investigated the differences between the subsets of alveolar macrophages detected in different conditions after irradiation. We computed the different markers specific to each sub cluster of alveolar macrophages and ran a GSEA analysis to identify the biological processes (from the gene ontology database GO-BP) enriched in the different subsets after irradiation.

First, in the population that appears one day after irradiation, there is an upregulation of the cellular response to stress and of the processes of regulation of programmed cell death (**FIGURE 3.7A**). One month after a 10Gy or 13Gy irradiation, the proportion of alveolar macrophages in the lung is divided by three compared to the control samples (**FIGURE 3.7B,C**). This observation strongly suggests that alveolar macrophages undergo strong acute damage after irradiation, which triggers the removal of a significant proportion of the alveolar macrophage pool. In both in the mice that underwent a 10Gy and 13Gy irradiation, the proportion of alveolar macrophages is restored to a control level at two months post irradiation. However, at five months post 13Gy irradiation the proportion of alveolar macrophages increase by 7-fold while in the case of a 10Gy irradiation it remains close to the control.

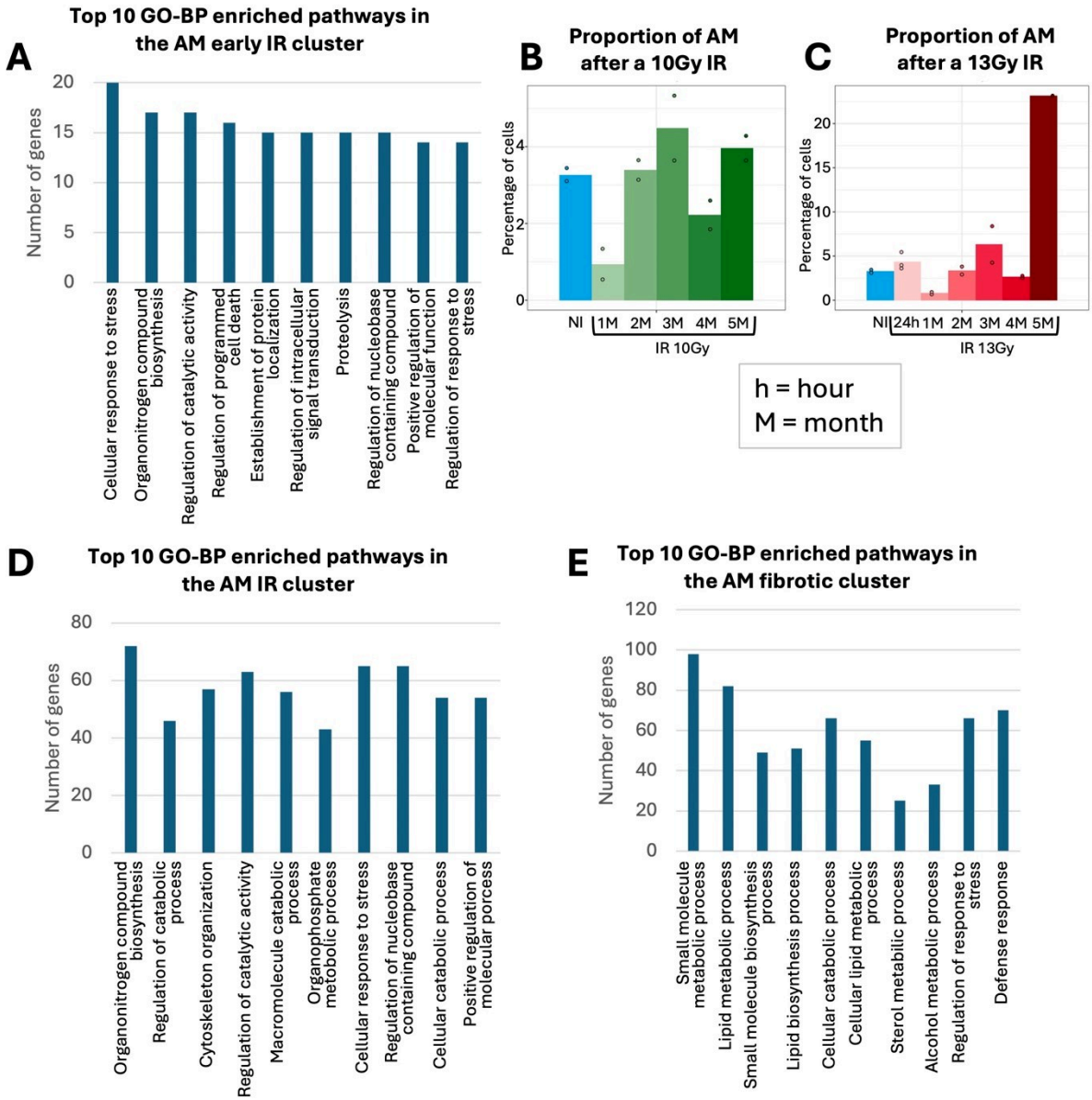


FIGURE 3.7. increase of cell death and lipid metabolism in the different populations of alveolar macrophages after irradiation.
A: Top 10 GO-BP enriched significantly upregulated (FDR adjusted p-value < 0.05) pathways in the AM early IR cluster; B: proportion of AM in the lungs after a 10Gy irradiation; C: proportion of AM in the lungs after a 13Gy irradiation; D: Top 10 GO-BP enriched significantly upregulated (FDR adjusted p-value < 0.05) pathways in the AM IR cluster; E: Top 10 GO-BP enriched significantly upregulated (FDR adjusted p-value < 0.05) pathways in the AM fibrotic cluster. Pathways for A, D and E are ranked by the number of genes in the pathway upregulated.

We also identified a population of alveolar macrophages specific of irradiated samples that have not reached the state of fibrosis (five months post 13Gy irradiation). There is in these samples a similar increase in cellular response to stress, but with the addition of increased catabolic and catalytic activity (**FIGURE 3.7D**). It has been shown that alveolar

macrophages have high catabolic potential, especially concerning lipid catabolism. In homeostasis, one of the functions of alveolar macrophages is the clearing of pulmonary surfactant. When exposed to an increased quantity of lipids in pathological conditions, alveolar macrophages can become activated and increase their catabolic activity (Wculek et al. 2022). Models of bleomycin and silica injury in mice have shown accumulation of some phospholipids in broncho-alveolar lavage fluids (BALF) (Milad and Morissette 2021). This could explain the increased catabolism of alveolar macrophages after irradiation. This tendency is even increased at the stage of fibrosis: the alveolar macrophages population specific of the four to five months post 13Gy irradiation timepoints still show these increased catabolic processes as well as lipid metabolisms processes (**FIGURE 3.7E**). This increased in lipid metabolism is characteristic of the foamy macrophages, a state of macrophages that have been described in several diseases, like tuberculosis (Russell et al. 2009), spinal cord injury (Kong et al. 2020) or atherosclerosis (Yu et al. 2013).

In order to gain a better understanding of the mechanisms driving the differences between the different populations of alveolar macrophages (**FIGURE 3.8A**) after the different conditions of irradiation, we performed a gene regulatory network analysis (**FIGURE 3.8B**). This allowed us to identify the main regulators responsible for the difference in transcriptome of the different AM populations. The main transcription factor expressed in the AM healthy population is Nfia, known to be involved in myeloid cell differentiation (L. Chen et al. 2020). In the AM IR, there is an upregulation of the activity of the transcription factor Nfkb1, a key regulator of inflammatory response (S. Batra, Balamayooran, and Sahoo 2011), which is consistent with the observations already made in macrophages and other populations in reaction to irradiation. Interestingly, the AM 24h presents a strong activity of transcription factors not found in other populations, like Xbp1, involved in inflammation by promoting the expression of genes like Il6 and Tnf (S.-M. Park, Kang, and So 2021).

Finally, many transcription factors present a high activity in the AM fibrotic, specific of five months post 13Gy irradiation. Stat1 has already been shown to be upregulated in the AM after radiation exposure of the lungs (J. Chen et al. 2016). Similarly, Fosl2 expressing macrophages has been shown to contribute to the development of fibrosis, notably by secreting collagen (Ucero et al. 2019). These results show the pro-fibrotic and pro-inflammatory profiles of the AM during the latest stages of fibrosis.

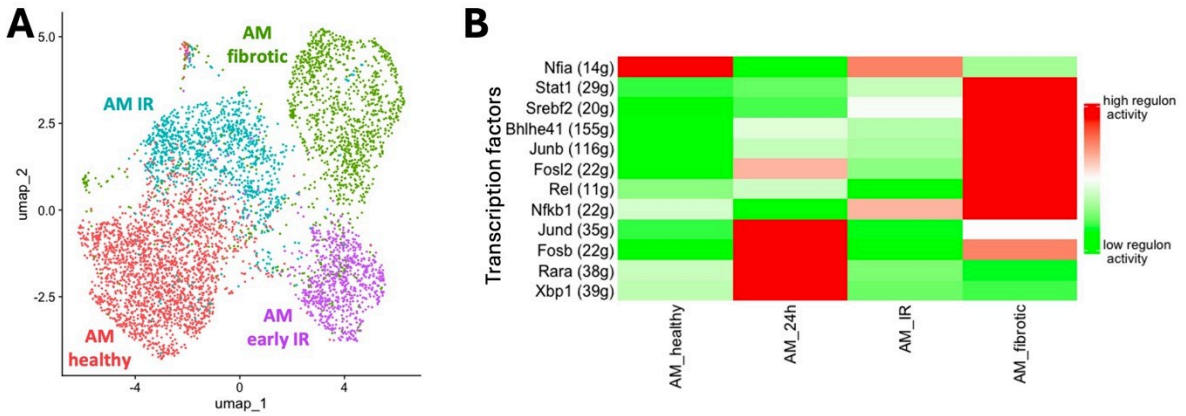


FIGURE 3.8. gene regulatory network analysis of the alveolar macrophages
A: DimPlot of the different populations of alveolar macrophage (AM) identified according to the irradiated conditions where they appear (24 samples, 5.237 cells); B: gene regulatory network analysis of the different populations of AM. In parenthesis, the number of genes regulated by the transcription factor.

g) Late changes in phenotype in the macrophages after a RILI

Finally, radio-induced lung injury does not lead to radio induced pulmonary fibrosis for all patients, however it has been shown that patients treated with radiotherapy can suffer from side effects like reduced pulmonary function years after the treatment (Miller et al. 2003). These late side effects have not been studied extensively and can impact the quality of life of former radiotherapy patients. Therefore, in order to gain a better understanding of the late cellular and molecular events occurring after a non fibrogenic dose of irradiation, we analyzed with single cell RNA sequencing lungs from mice nineteen months post 10Gy irradiation, as well as age matching controls.

The macrophages are important players of the wound healing process and the maintaining or clearing of inflammation. In the different samples collected after a 10Gy irradiation, as well as in the young and old control mice lungs, we were able to identify the monocytes, alveolar macrophages and interstitial macrophages (**FIGURE 3.9A**). However, the samples nineteen months post 10Gy irradiation show a difference in the proportion of the different populations compared to the others irradiated samples or the controls. Indeed, there is in these samples a low percentage of monocytes (five times less than the average of the other samples), and an important increase in the proportion of alveolar macrophages (two to three times more than the other samples) (**FIGURE 3.9B**).

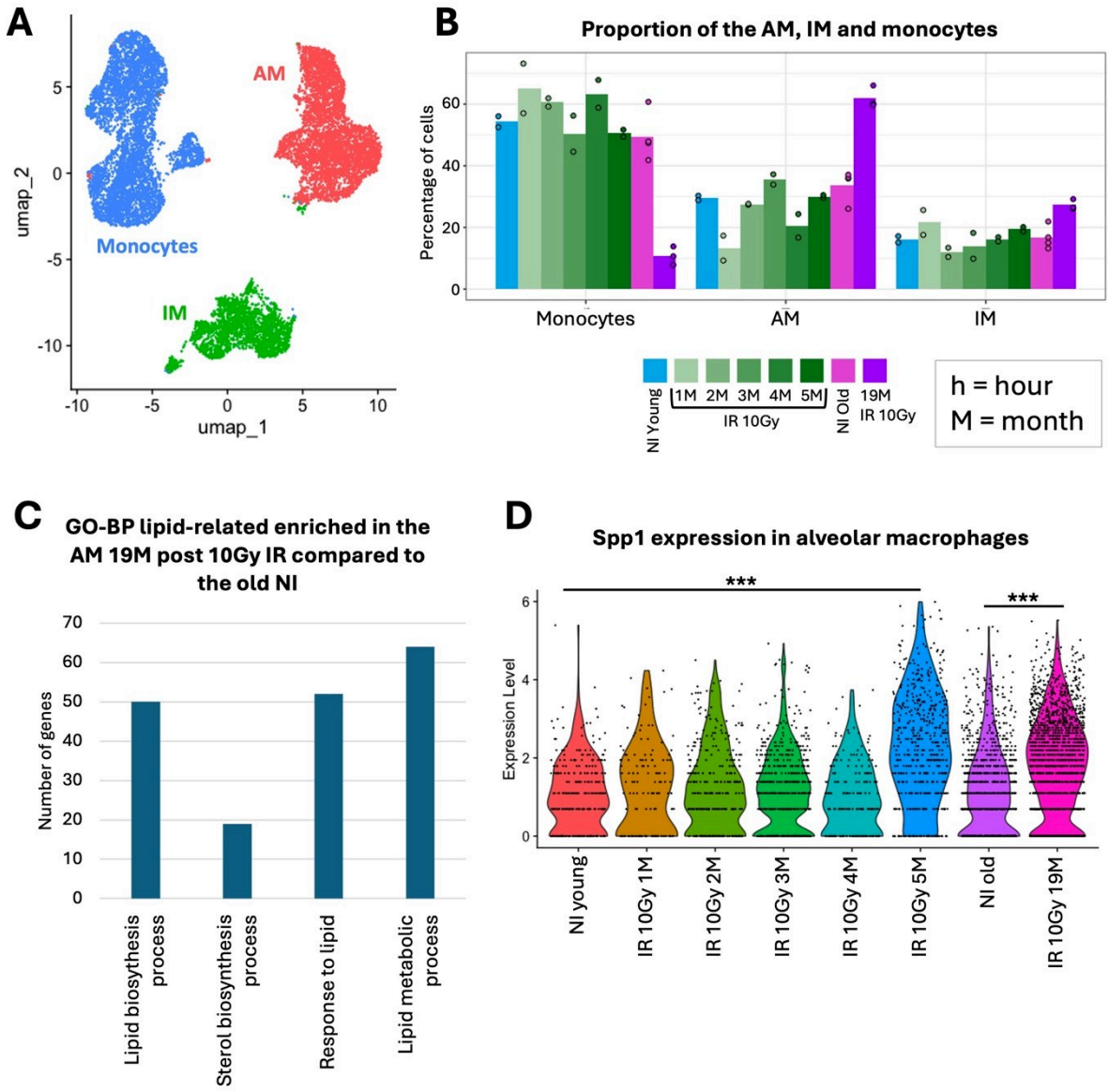


FIGURE 3.9. alveolar macrophages nineteen months post irradiation present a foamy phenotype.
A. UMAP plot of the lung alveolar macrophages, interstitial macrophages and monocytes in the samples after a 10Gy irradiation (19 samples, 13.369 cells); B: Bar plot of the proportion of the different lung alveolar macrophages, interstitial macrophages and monocytes at the different time points after a 10Gy irradiation; C: number of genes related to biological lipid processes significantly upregulated in the alveolar macrophages nineteen months post irradiation compared to the old controls (23 months old mice); D: expression of Spp1 in the alveolar macrophages in the different samples post 10Gy irradiation

Furthermore, these alveolar macrophages nineteen months post irradiation present a similar upregulation of lipidic processes (compared to the old control mice) as the samples after a 13Gy irradiation (**FIGURE 3.9C**), therefore they might also be "foamy macrophages" and contribute to the maintaining of inflammation. They also show a significant upregulation

of Spp1 compared to old controls, one of the described markers of m2-like profibrotic alveolar macrophages (Willemsen and de Winther 2020) (**FIGURE 3.9D**).

4. Regeneration of the epithelial cells: early bipotent progenitor and late processus of transdifferentiation

The lung epithelium is fundamental for the lung to perform its function. It constitutes the surface of gas exchanges between the outside air and the blood. Furthermore, the epithelium contains cells that secrete surfactant, allowing the lungs not to collapse, evacuate the debris and participate in the primary immune response. However, the AT1 cells, the cells that are in charge of the gas exchanges, are fragile and susceptible to destruction or apoptosis upon infection or aggression (Kuwano 2007). Therefore, it is interesting to study their fate after irradiation injury and their potential regeneration.

a) Identification of the epithelial population in healthy mouse lungs

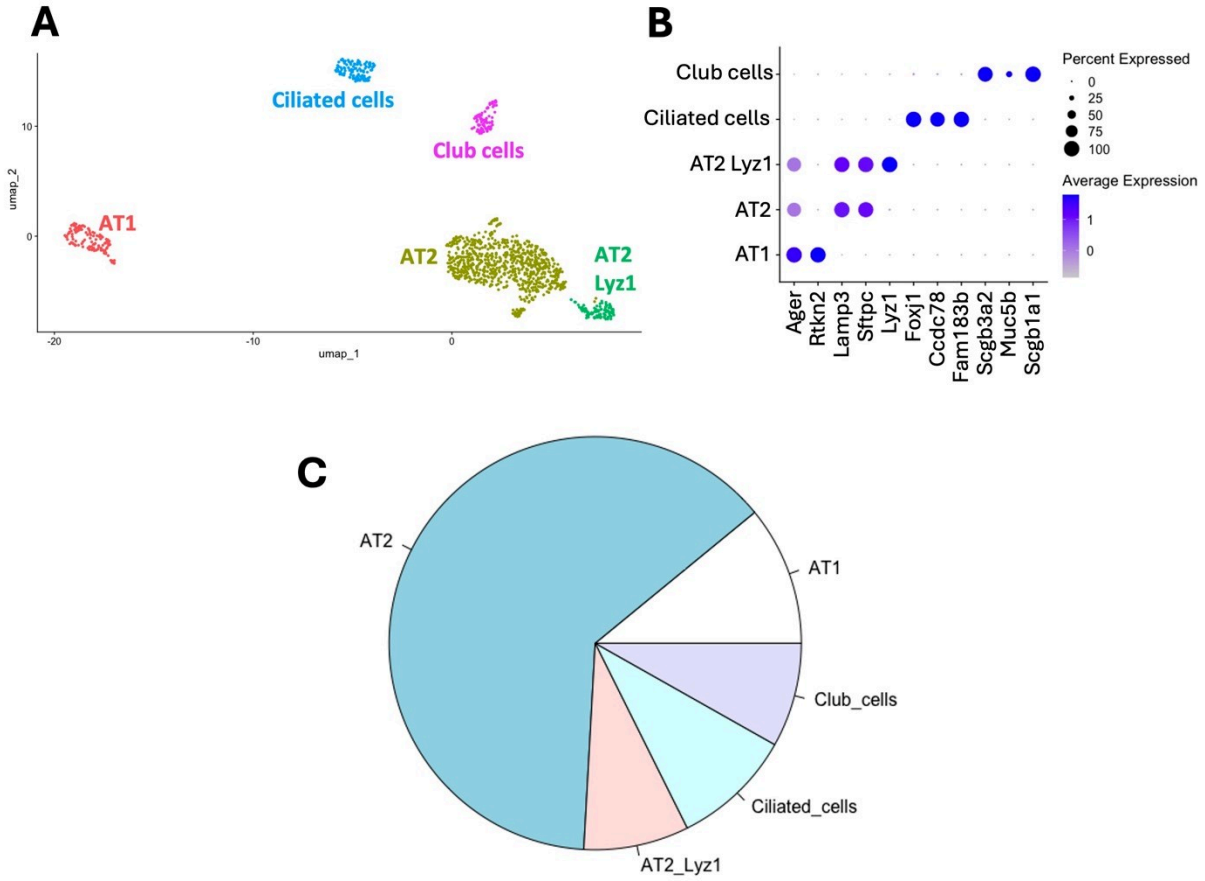


FIGURE 3.10. identification of the different epithelial cell population in the non-irradiated control mouse lung.
A: UMAP plot of the different epithelial cell populations (2 samples, 1.169 cells); B: Expression of the markers used for the identification of the epithelial cell populations; C: Proportion of the different epithelial cell populations in the non-irradiated mouse lung.

Using single cell RNA sequencing, we were able to identify in the non-irradiated mouse lungs the main epithelial cells populations: AT1 cells, AT2 cells, AT2 Lyz1 positive cells, ciliated cells and club cells (**FIGURE 3.10A**), using markers described in the literature (**FIGURE 3.10B**).

The most abundant population we detected are the AT2 cells, accounting for 71 percent of the epithelial cells (AT2 cells + AT2 Lyz1 positive cells). The other populations appear to be equally represented in the lungs of non-irradiated control mice.

b) The late AT2 response to irradiation: AT2 to AT1 transdifferentiation

We first identified a Krt8+ AT2 cells population expressing genes characteristics of AT2 to AT1 differentiation, so that presents the capacities to regenerate the AT1 population (Riemondy et al. 2019; Curras-Alonso et al. 2023). This AT2 to AT1 transdifferentiation signature was found to be mainly upregulated four to five months post irradiation in a small Krt8+ AT2 cells population. These results are presented in the publication displayed in **chapter III 2)**.

c) The AT0 cells, a population involved in the early regeneration of the lung epithelium

In order to investigate more deeply this regeneration potential of the mouse epithelium after an irradiation injury, we used the new mouse samples that were irradiated with the Collimation irradiator, with samples sequenced 24 hours post 13Gy full thorax irradiation. In addition to the classical AT1 and AT2 populations, we were able to identify a Lyz1+ AT2 population that has been shown to be mouse specific (Hurskainen et al. 2021), and the Sftpb+ Scgb1a1+ Scgb3a2+ AT0 population (**FIGURE 3.11A,B**). The AT0 population has been recently described and has been shown to derive from AT2 cells after irradiation and act as a bipotent progenitor that can differentiate into AT1 cells or back to AT2 cells (Kadur Lakshminarasimha Murthy et al. 2022).

The analysis of the cell cycle score of the different epithelial cells populations shows that the AT0 cells present the highest S phase score and G-2M phases score (**FIGURE 3.11C**), which is another argument for the involvement of the AT0 cells in the regeneration of the lung epithelium after an irradiation injury. Furthermore, this AT0 population appears as an

early response, as they are specific of the 24 hours post irradiation timepoint, when they represent more than half of the epithelial cells (**FIGURE 3.11D**).

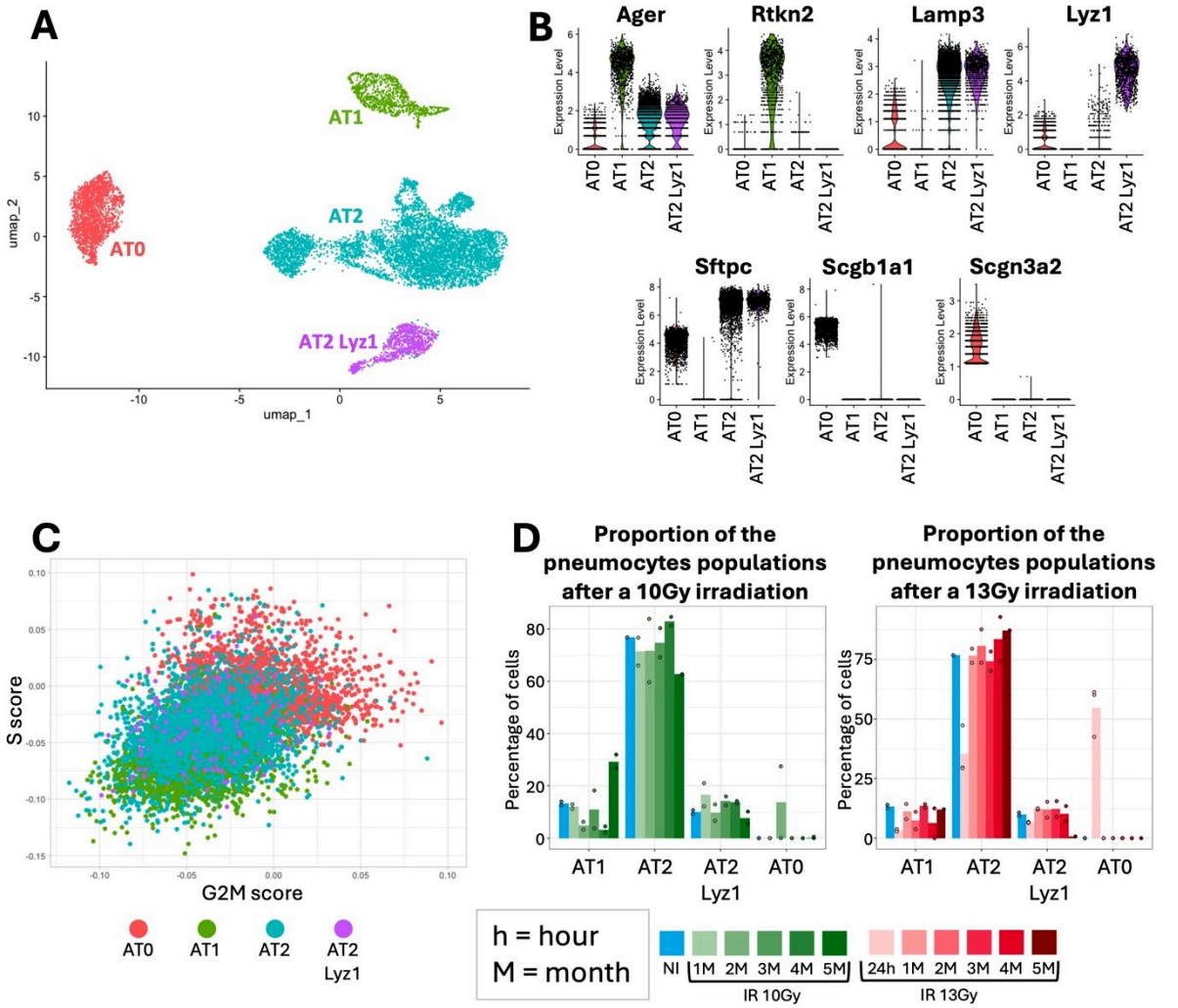


FIGURE 3.11. identification of the AT0 cells, an epithelial cell progenitor population involved in the early lung response to injury.
A: UMAP plot of the different lung pneumocyte populations 24h post irradiation (3 samples, 4.165 cells); B: expression of the markers used to identify the pneumocyte populations 24 hours post irradiation; C: Cell cycle phase scores: S score and G2-M score for the cells of the different pneumocyte populations 24 hours post irradiation; D: Bar plot of the proportion of the different pneumocyte populations at the different time points after a 10Gy or 13Gy irradiation.

With a trajectory analysis of the samples 24h after a 13Gy irradiation, we can predict a connection between the AT2 cells and the AT0 cells, and then another one between the AT0 cells and the AT1 cells (**FIGURE 3.12A**). Furthermore, if we place the origin of the node of origin in the AT2 cells, the pseudo time analysis predicts an order of the populations of 1) AT2 cells, 2) AT0 cells and 3) AT1 cells (**FIGURE 3.12B**).

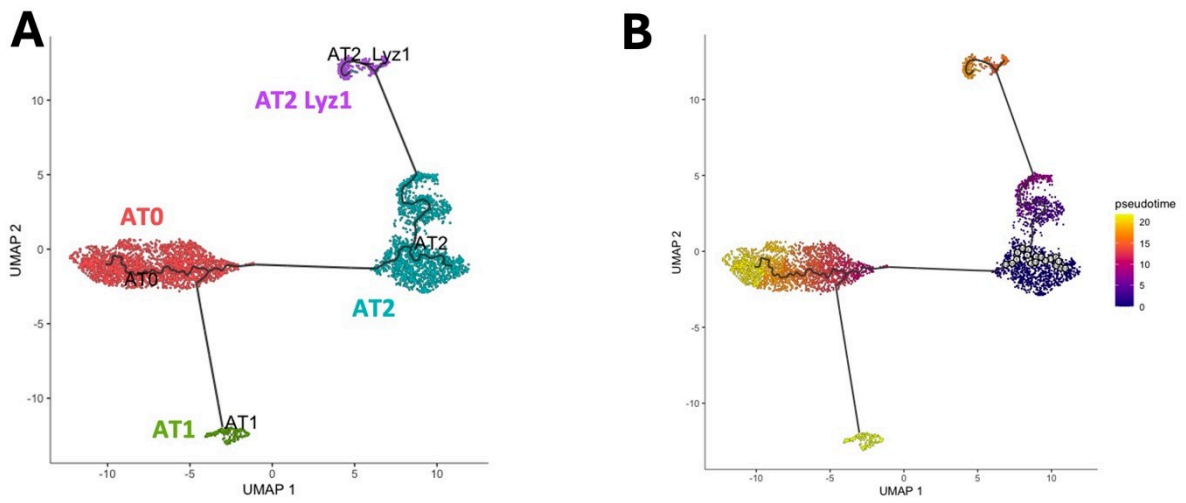


FIGURE 3.12. prediction of the differentiation of AT2 to AT0, then to AT1 early after irradiation injury.
A: trajectory analysis of the AT2, AT1 and AT0 cells in the samples taken 24 hours after irradiation (3 samples, 4165 cells); B: pseudo-time analysis of the AT2, AT1 and AT0 cells in the samples taken 24 hours after irradiation with the AT2 cells as the population of origin.

To study the potential mechanisms involved in these transitions, we studied the gene regulatory network (GRN) of the different pneumocytes populations. First, the AT1 cells do not show a change in the activity of the different regulons after a 10Gy irradiation compared to the non-irradiated control (**FIGURE 3.13A**). However, after a 13Gy irradiation, there is an increase in some regulons starting at one month post irradiation, notably Xbp1, a transcription factor shown to be involved in the induction of cytokines production (Ribeiro and Lubamba 2017). Other regulons became upregulated during the late stage of fibrosis after a 13Gy irradiation. The AT0 cells express a different set of transcription factors (**FIGURE 3.13B**) not found in the other pneumocytes populations, like Cebpb and Atf3, two transcription factors promoting epithelial cell proliferation (Hsueh, Kuo, and Chen 2013). Finally, the AT2 cells display similar regulons activity after a 10Gy or a 13Gy irradiation (**FIGURE 3.13C**).

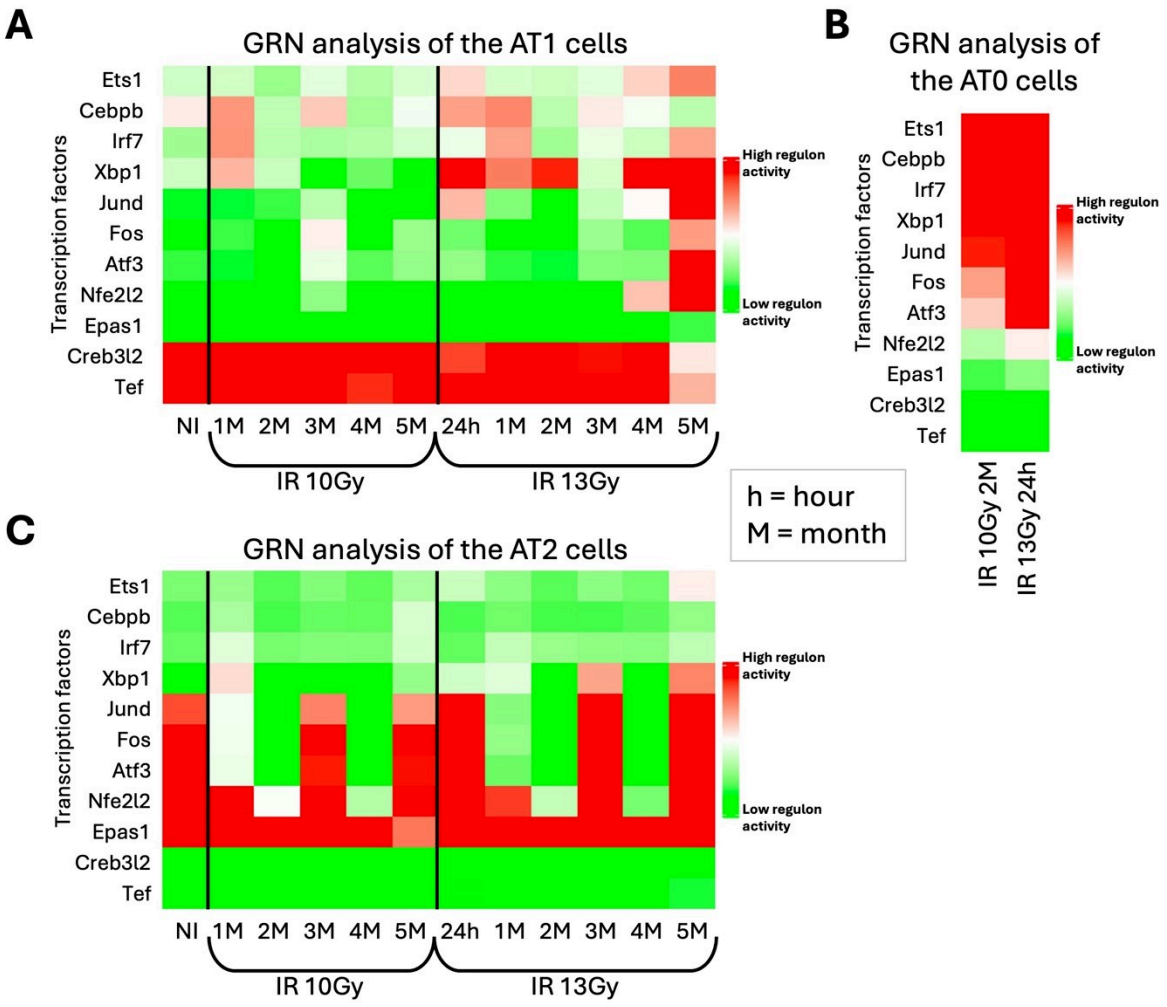


FIGURE 3.13. gene regulatory network analysis of the AT1, AT2 and AT0 cells after irradiation.
Gene regulatory network analysis with of the different pneumocytes populations. Red means a high activity of the transcription factor, and green a low activity. A: analysis of the AT1 cells; B: analysis of the AT0 cells; C: analysis of the AT2 cells.

d) Epithelial to mesenchymal transition induced by alveolar macrophages

We identified a TGFβ1-TGFβR2 interaction, specifically received by the AT2 cells in the early time points (24 hours and one month) or late time point (five months) post 13Gy irradiation, and not after a 10Gy irradiation. This TGFβ1 signaling appears to be sent by the AM at 24h post irradiation as well as by radiation-specific AM during the fibrotic state (**FIGURE 3.14A**). This intercellular interaction has been shown to induce endothelial to mesenchymal transition (EMT) in various contexts, including fibrosis (Nagaraja and Nagarajan 2018). Furthermore, the result of a clustering based on the expression of EMT markers in AT2 cells (**FIGURE 3.14B**) shows a reduced expression of EMT suppressors genes or genes

downregulated by EMT in the samples after a 13Gy irradiation, and particularly the samples five months post 13Gy irradiation. Indeed, in the top left panel of the heatmap in **FIGURE 3.14B**, we see a lower expression of Id2 (Wen et al. 2018), Cadm1 (Sawada et al. 2020), Ppib (Sun et al. 2021), Gja1 (James et al. 2018), Sdc1 (Kumar-Singh et al. 2021) and Cdh1 (Aban et al. 2021) compared to the others panels more enriched in non-irradiated cells and 10Gy irradiated cells.

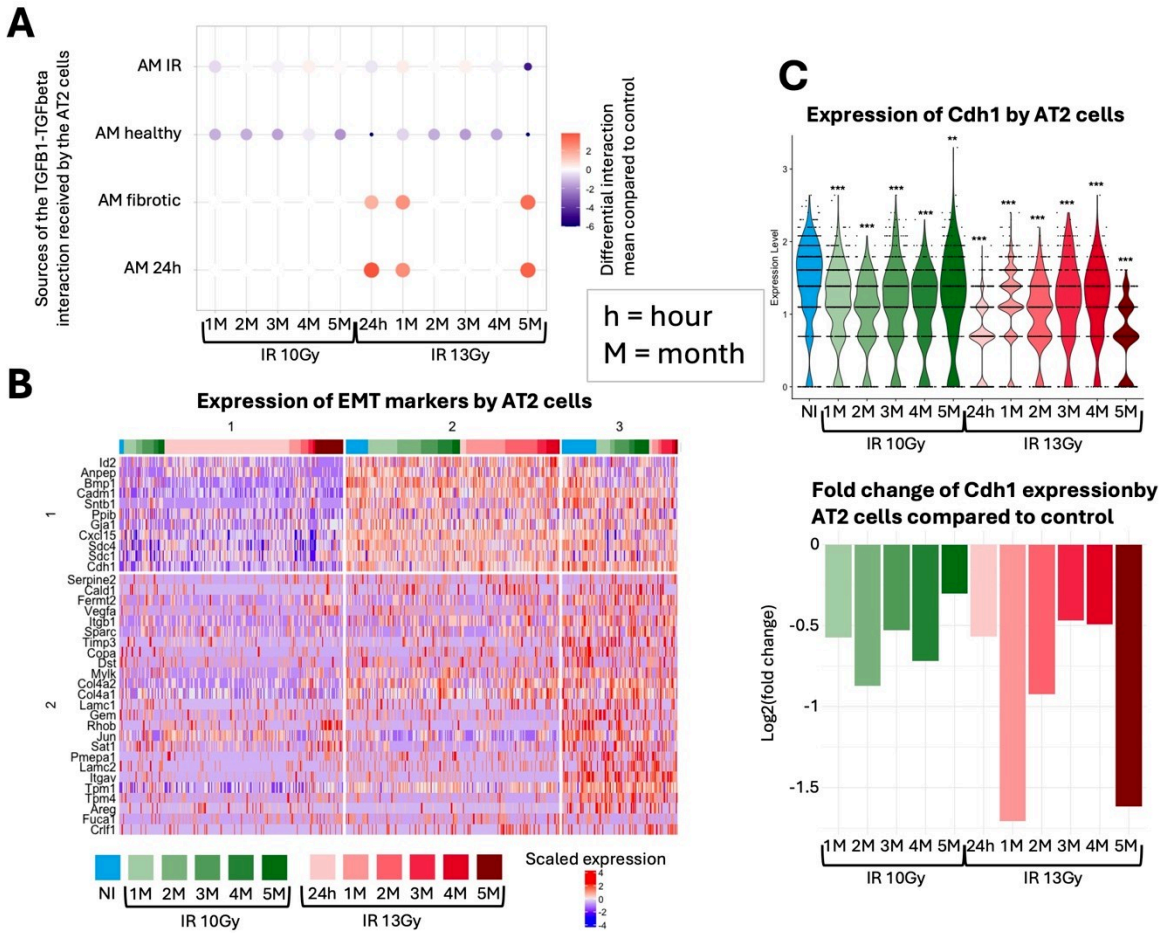


FIGURE 3.14. increase of the EMT-inducing TGFβ1-TGFβR2 interaction between the AM and AT2 cells after irradiation.
A: TGFβ1-TGFβR2 interaction received by the AT2 cells and sent by the different alveolar macrophage sub populations; B: heatmap of expression of EMT markers (from the Hallmark database) by the AT2 cells. Only markers expressed in at least 20% of AT2 cells were kept; C: expression of Cdh1 by the AT2 cells (upper panel), and fold change of the comparison with the non-irradiated control, all the conditions being significantly downregulated compared to non-irradiated (lower panel).

For example, the expression of E-cadherin (Cdh1) in AT2 cells is significantly decreased after irradiation (**FIGURE 3.14C**). The decrease in expression of E-cadherin by epithelial cells

is an early sign of EMT. This decrease is more important after a 13Gy irradiation than after a 10Gy irradiation, especially at one month and five months post irradiation. Therefore, we can make the hypothesis that a radiation induced subpopulation of AM contributes to the observed EMT-like process in AT2 at late stages of RIPF. AM could also be involved in the induction of an EMT-like state in AT2 at earlier stages post-irradiation. However, it is possible that these transitions serve different purposes (regeneration versus survival).

5. Regeneration of the endothelial cells: pro-angiogenic signaling

a) Identification of the different endothelial population in non-irradiated samples

Endothelial cells can be subdivided in different populations that we identified in the non-irradiated mouse lungs.

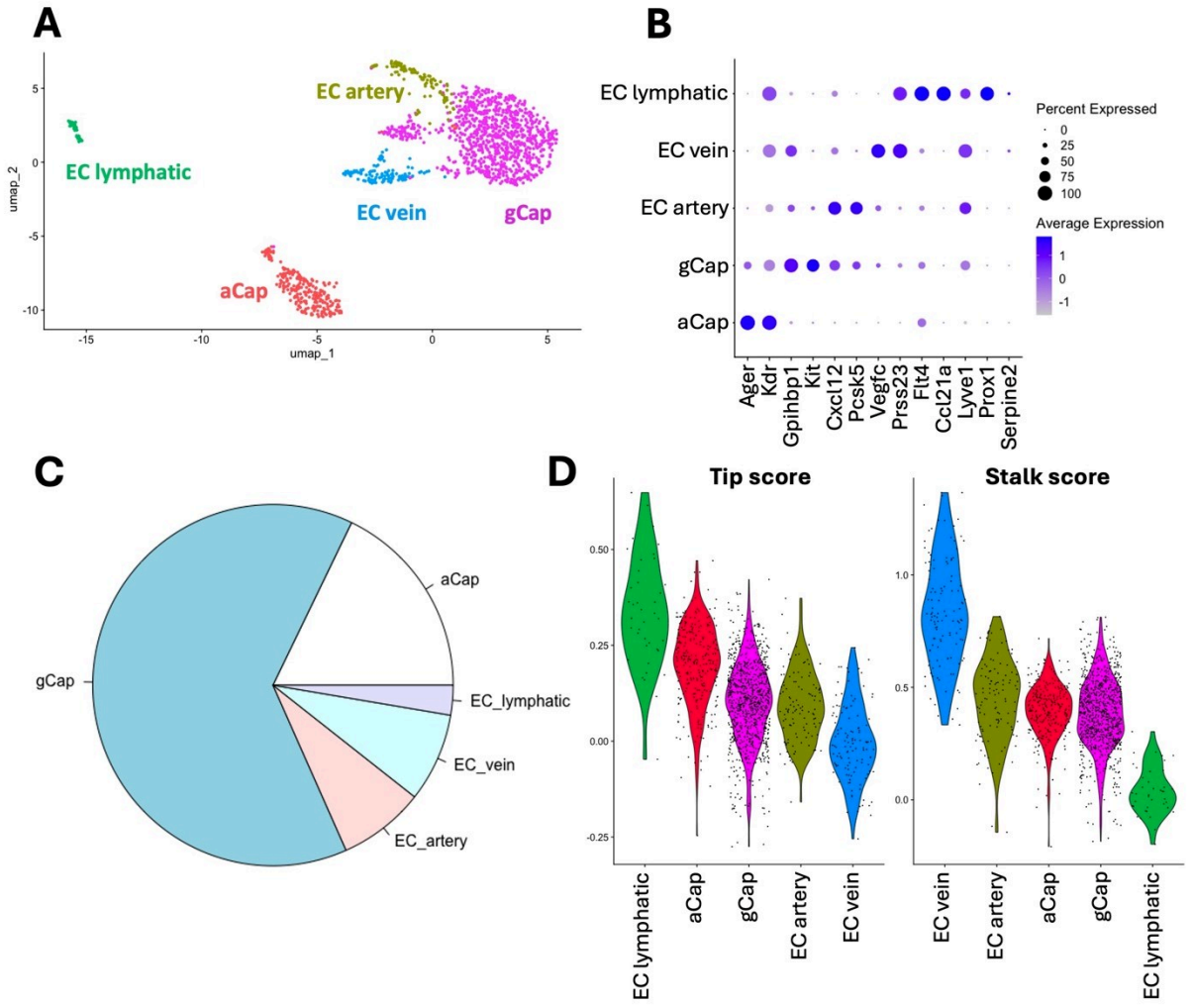


FIGURE 3.15. identification of the endothelial cells populations in the non-irradiated mouse lungs.
A: UMAP visualization of the different mouse endothelial cell subpopulations annotated by sub cell type (2 samples, 1.499 cells); B: Expression of the markers used for the identification of the endothelial cell populations; C: Proportion of the different epithelial cell populations in the non-irradiated mouse lung; D: tip and stalk score in the non-irradiated mouse endothelial cells.

Endothelial vessels are composed of the endothelial vein cells, the endothelial artery cells and the endothelial lymphatic cells. Furthermore, we identified two populations of capillary cells: the aerocytes or aCap, in charge of the gas exchanges, and the general capillary cells or gCap, that act as progenitors (**FIGURE 3.15A**). These different populations were identified using markers described in the literature (**FIGURE 3.15B**). In the non-irradiated lung samples, the main EC subtypes identified are the gCap, accounting for 64% of EC and the aCap accounting for 18% (**FIGURE 3.15C**).

Two endothelial cell states have been described to be involved in vessel growth and repair, physiologically and after injury: tip and stalk cells. As described in the **introduction**, tip cells present motility properties and take the lead of the sprouting of the new vessel, while stalk cells proliferate behind and build the walls of the new vessel. The tip or stalk status can be evaluated using a score based on markers found un the literature (**FIGURE 3.15D**).

b) Vascular damage and repair after a radiation induced lung injury

Vascular damage is a major consequence of lung irradiation. Irradiation causes death of endothelial cells, disorganization of the vascular network, as it has been shown in both human patients and mouse models (Venkatesulu et al. 2018). The replacement of these damaged endothelial cells and vessels is mostly triggered by pro-angiogenic signaling (Boussat et al. 2000).

Angiogenesis is the physiological process through which new blood vessels form from pre-existing vessels. This process is crucial for growth, development, and wound healing. It involves the proliferation, migration, and differentiation of endothelial cells, which line the interior surface of blood vessels. Pro-angiogenic signaling is essential for re-establishing an adequate blood supply to the damaged tissue, ensuring the delivery of oxygen and nutrients necessary for tissue repair and regeneration. Endothelial cells play a pivotal role in angiogenesis, responding to pro-angiogenic signals such as VEGF by proliferating, migrating, repairing damaged endothelial vessels and forming new capillary networks.

Therefore, we studied here the pro-angiogenic molecular and cellular mechanisms triggered by radiation injury in the lungs. We identified the different populations of endothelial cells in the different conditions and time points post irradiation (**FIGURE 3.16A**),

using markers described in the literature (**FIGURE 3.16B**). After either a 10Gy or a 13Gy irradiation there is an increase in the proportion of aCap, mainly at three months post 10Gy or 13Gy irradiation (**FIGURE 3.16C**). Finally, the decrease in the proportion of gCaps is similar after a 10Gy irradiation and a 13Gy irradiation (**FIGURE 3.16D**).

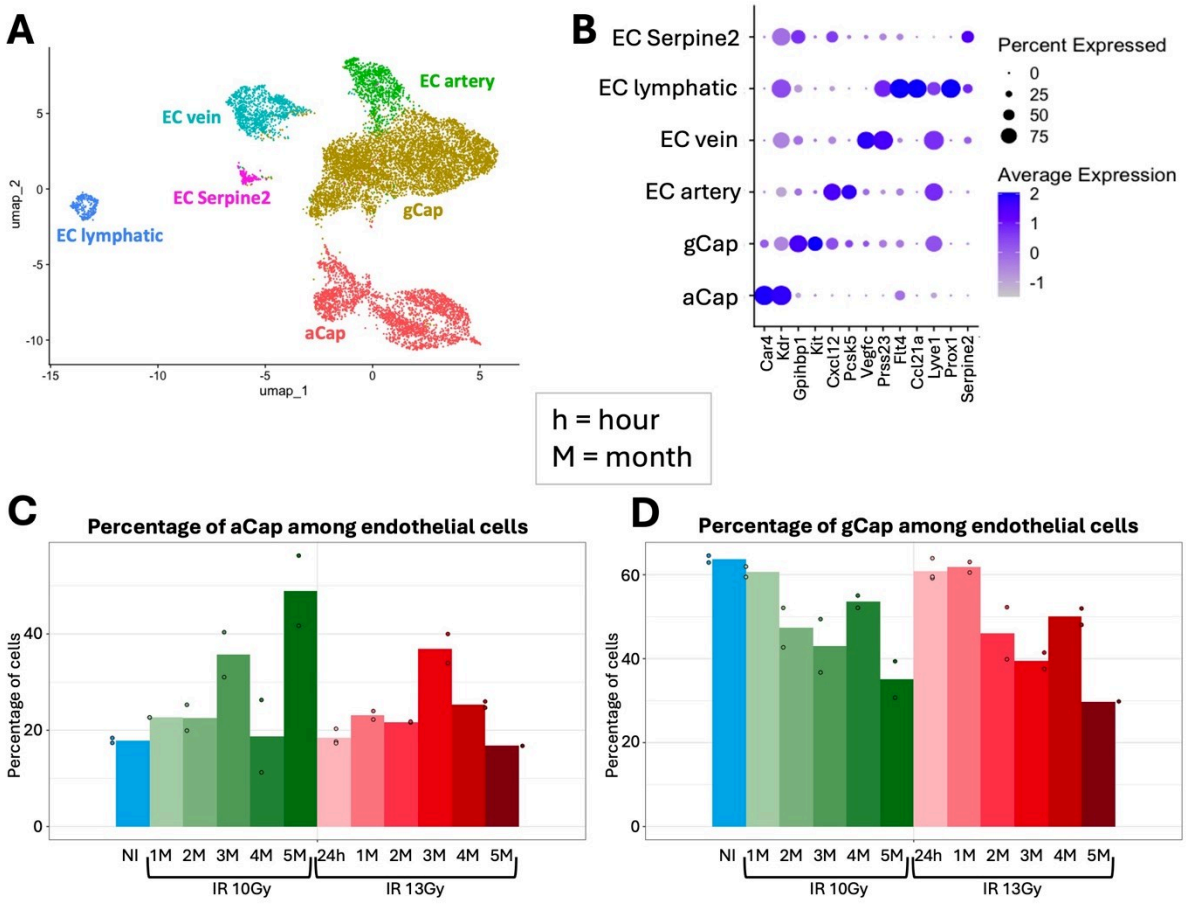


FIGURE 3.16. decrease of the proportion of gCap cells in mouse lungs after irradiation.
A: UMAP visualization of the different mouse endothelial cell subpopulations annotated by sub cell type (24 samples, 11.091 cells); B: DotPlot of the markers used for the identification of the different endothelial cell populations; C: proportion of aCaps among the endothelial cells in the non-irradiated samples and one to five months post 10Gy or 13Gy irradiation; D: proportion of gCaps among the endothelial cells in the non-irradiated samples and one to five months post irradiation a 10Gy or 13Gy irradiation.

The tip and stalk identities were then assessed using a score of the expression of lists of markers of tip and stalk cells published (W. Chen et al. 2019). The score levels are presented in **FIGURE 3.17A**. The tip score is significantly upregulated for all conditions, with a higher fold change five months post 13Gy irradiation. On the contrary, the stalk score is downregulated compared to irradiation in all the time points after a 10Gy irradiation, and one to three

months post 13Gy irradiation. In the end, at the time point when we see the development of fibrosis (five months post 13Gy irradiation), the stalk score is increased and upregulated compared to the control. These results suggest that irradiated gCap cells respond to the increase in pro-angiogenic signaling by acquiring “tip”-like characteristics.

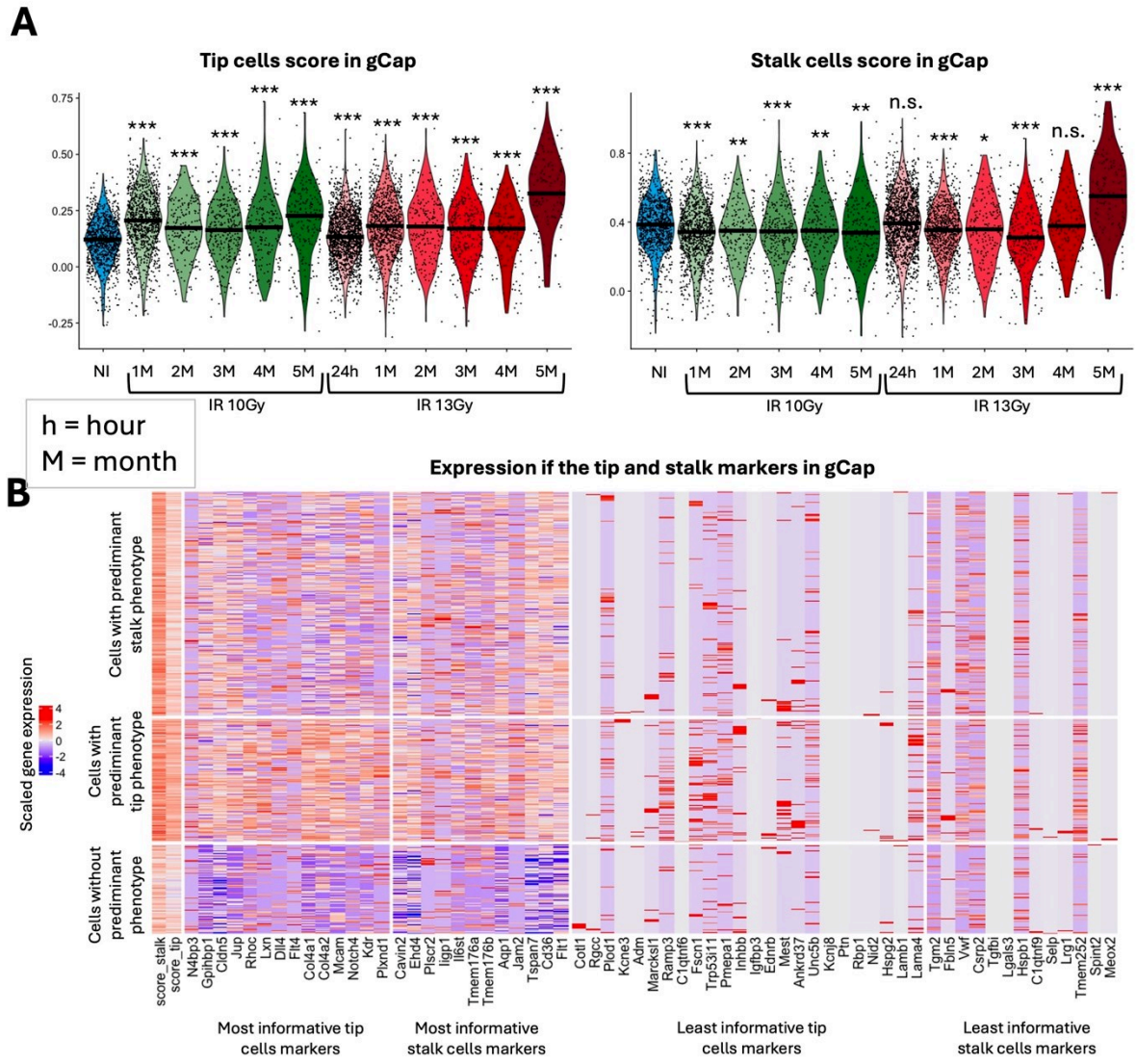


FIGURE 3.17. increase of the tip and stalk identity of the gCap cells in mouse irradiated lungs.
A: Violin plots of the tip and stalk scores in the gCaps, black lines represent the median; stars represent the p-value of the wilcox test compared to the non-irradiated (n.s.: non-significant, *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001); B: heatmap of the expression of the different tip and stalk markers used for the score calculation of the panel A of this figure.

However, as tip and stalk cells are dynamic cell states, it can be difficult to quantify exactly the quantity of tip or stalk cells at each point after irradiation. Indeed, the tip or stalk

identity is defined by the upregulation of an ensemble of genes, rather than just one marker (**FIGURE 3.17B**).

c) Populations sending VEGFA signaling

The main interaction that has been shown to trigger angiogenesis-related processes is the VEGFA-VEGFR2 (also known as KDR) interaction. Therefore, we used CellphoneDB to predict which cell populations are sending this signaling to the gCap, the endothelial cell populations that have been described as the progenitor for the capillary cells (Gillich et al. 2020). Overall, the main sources of the VEFGA signaling received by the gCaps through the VEGFR2 receptor are predicted to be the artery endothelial cells, the AT2 cells, the fibroblasts Col13a1+ as well as an autocrine pathway (**FIGURE 3.18**).

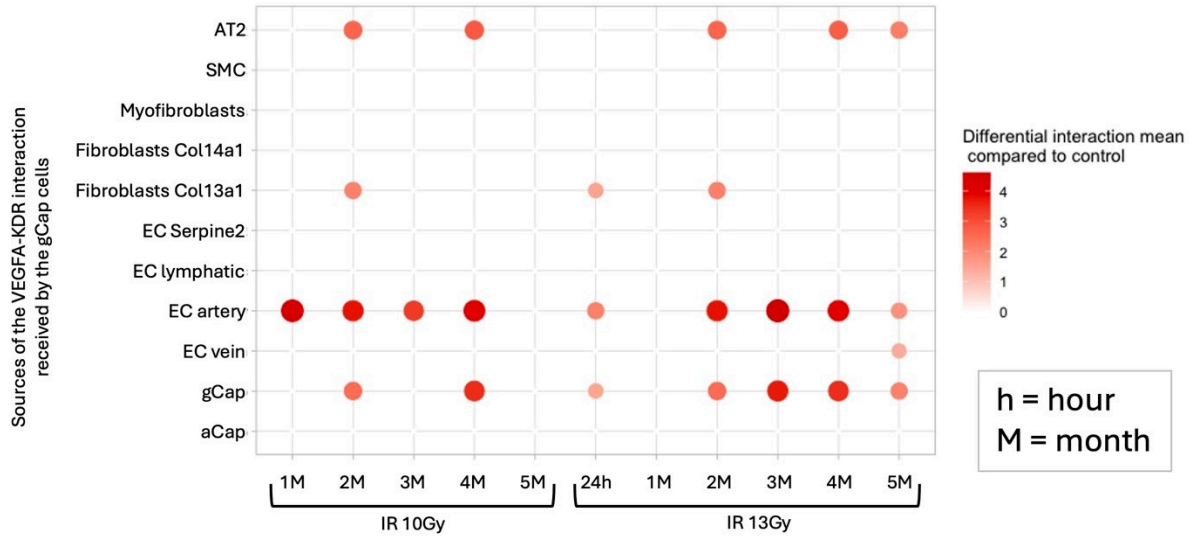


FIGURE 3.18. AT2 cells, fibroblasts and endothelial cells are the main sources of the VEGFA signaling received by the gCap cells
VEGFA-KDR interaction received by the gCap and sent by different cell populations.

Furthermore, although this signaling is present both after a 10Gy irradiation and a 13Gy irradiation, it disappears at five months post 10Gy irradiation. We can make the hypothesis that the vascular repair has been completed a few months after a 10Gy irradiation, while the signaling persists five months after a fibrogenic dose, suggesting that pro-angiogenic activities are associated with the fibrotic state. Whether these activities directly participate in the fibrotic process or are just a consequence of tissue remodeling remains to be demonstrated.

d) Late pro-angiogenic signaling of endothelial cells after RILI healing

In order to understand the late consequences of a healed RILI, we analyzed the different endothelial cells populations nineteen months post a 10Gy irradiation and via clustering of the endothelial cells we identified a population of gCaps that is only present in this specific condition (**FIGURE 3.19A,B**), and not in any other 10Gy irradiated samples or in the old non irradiated lungs.

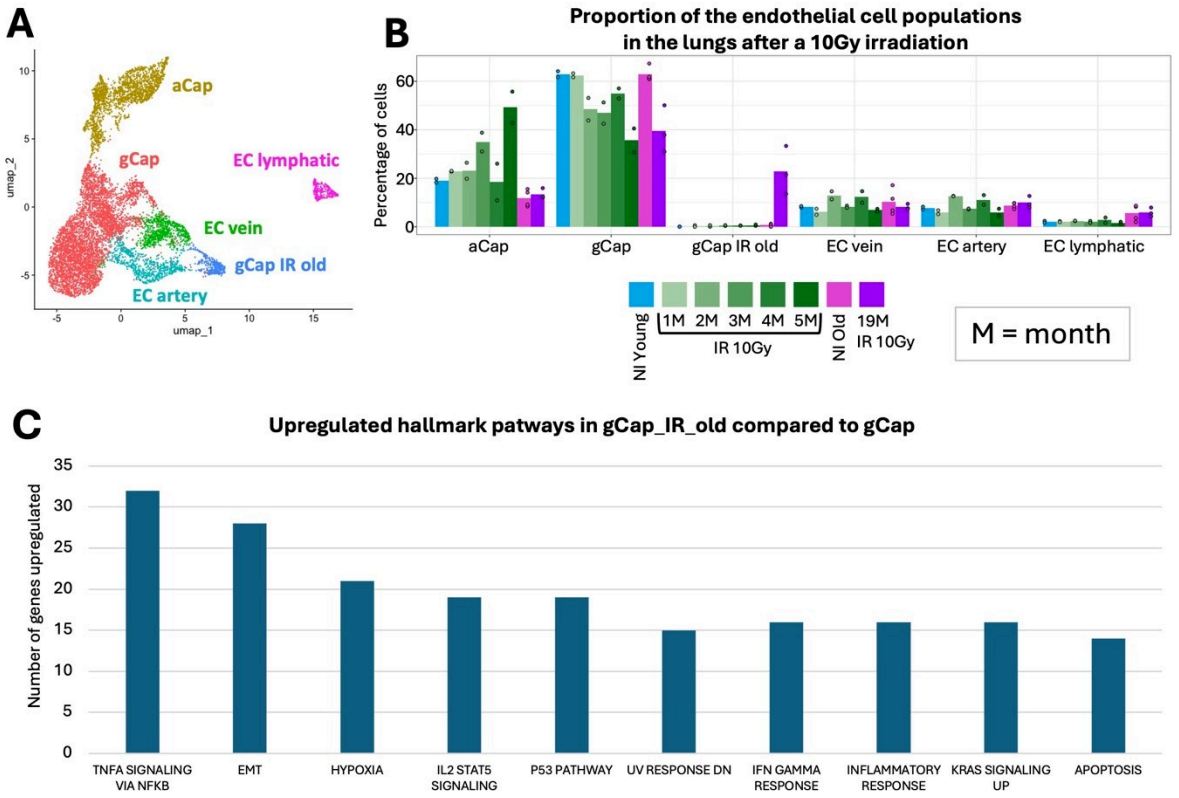


FIGURE 3.19. identification of a gCap population specific of the mice nineteen months post 10Gy irradiation.
A. UMAP plot of the different lung endothelial cell populations (19 samples, 8.867 cells); B: Bar plot of the proportion of the different endothelial cell populations at the different time points after a 10Gy irradiation; C: significantly upregulated (FDR adjusted p-value < 0.05) Hallmark pathways in the gCap IR old population compared to the classical gCap population.

Then, we compared this gCap IR old cells with the classical gCap population and identified an upregulation of different inflammation and stress related pathways: TNFα signaling, IL2 STAT5 signaling and IFNγ signaling. Furthermore, there is also an upregulation of genes involved in hypoxia response (**FIGURE 3.19C**), suggesting abnormalities in the capillary network, or a possible perturbation of the AT1-aCap function, maybe through tissue

disorganization and loss of contact between the partners. Therefore, we investigated the evolution of the pro-angiogenic signaling in the endothelial cells, and more specifically gCaps, in the lungs from old mice, nineteen months post irradiation or non-irradiated.

During the early time points post 10Gy irradiation (one to five months) we see an increase, regarding the non-irradiated control, in the proportion of gCap tip cells. However, this increase is two-fold lower than in the mice irradiated at 13Gy. Regarding the proportion of gCap among the endothelial cells, the 10Gy irradiation triggers a decrease one to five months post irradiation. However, the proportion of gCap does not recover nineteen months post irradiation (**FIGURE 3.20B**), while the aCap proportion returns to a level comparable to the one found in non-irradiated young and old mice (**FIGURE 3.20A**). Furthermore, nineteen months post irradiation we still see an increased tip score after irradiation compared to the old control (**FIGURE 3.20C**).

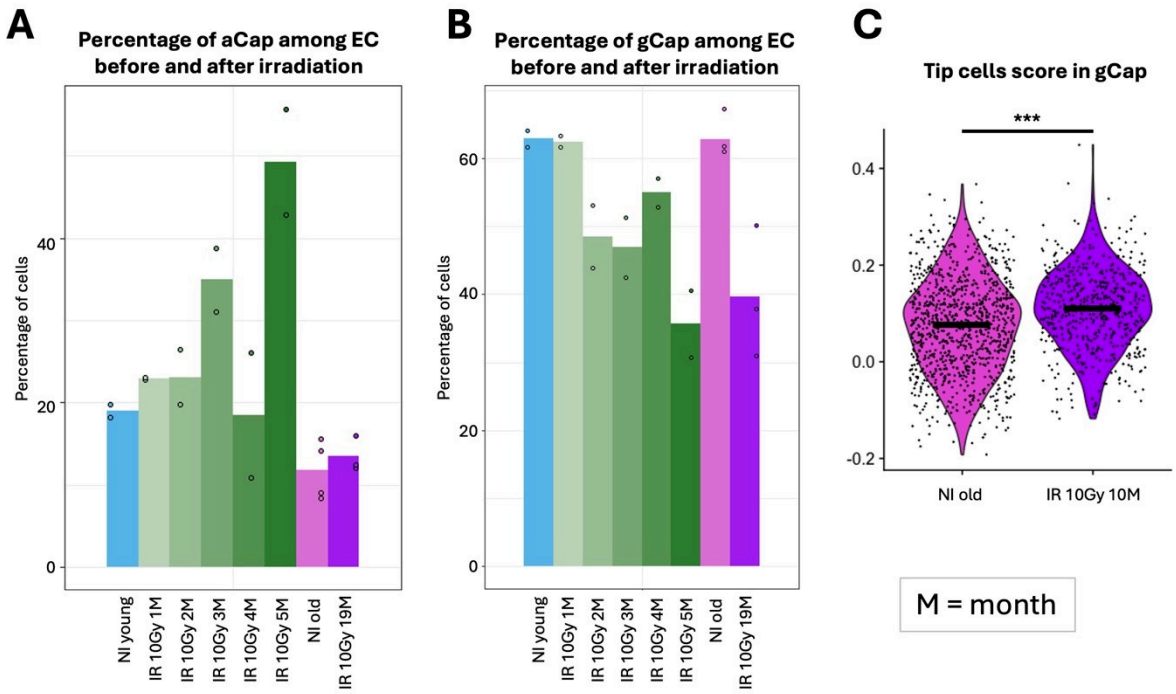


FIGURE 3.20. analysis of the dynamics of the tip, stalk and senescent endothelial cells in the old mice irradiated or non-irradiated.
A: proportion of mouse aCap among the endothelial cells in the non-irradiated samples and one to nineteen months post irradiation, with each dot representing replicate; B: proportion of mouse gCap among the endothelial cells in the non-irradiated samples and one to nineteen months post irradiation, with each dot representing replicate; C: tip score in the old gCap, stars represent the p-value of the wilcox test compared to the non-irradiated (***: p-value < 0.001)

Even if mice irradiated with a 10Gy dose recover after an early phase of pneumonitis and present a normal lifespan, we can still see an impact of the radiation on the endothelial cells nineteen months after the treatment. This could have long term implication on the health and quality of life of the patients that underwent radiotherapy, and it could be interesting to study the impact of radiotherapy decades after the treatment.

6. Processes of RIPF induced senescence

Senescence has been described in the progression of RIPF, mainly in AT2 cells, macrophages and fibroblasts (Y. He et al. 2019; Su et al. 2021). Senescence can be an advantage in the fight against cancer, as senescence induces cell cycle arrest thus can prevent cancer cells to proliferate (Jha et al. 2024). However, senescent cells secrete pro-inflammatory cytokines and other molecules, called senescence-associated secretory phenotype (SASP) (W. Huang et al. 2022). Therefore, if senescent cells accumulate, they can lead to a pro-inflammatory environment either in the tumor itself, promoting tumor growth, or in the surrounding tissues in response to irradiation. We investigate here the presence and particularities of senescence in the different lung populations after irradiation.

a) Senescence of the AT1 and AT2 cells after irradiation

As stated previously, the last stage of RIPF is characterized by the irreversible destruction of the alveolar structure (Wynn 2011). Therefore, we wondered if there could be an impairment of the epithelial cells repair processes. It has already been published that AT2 cells express the senescence marker p21 after high dose of irradiation (Citrin et al. 2013), and during fibrosis (Chung et al. 2021) but never using single cell RNA sequencing.

In our model, after irradiation, there is a significant increase of the expression of p21 in the AT1 only five months post 13Gy irradiation (**FIGURE 3.21A**). Similarly, this population presents an increase of a score representing a global expression of SASP associated genes only after a 13Gy irradiation, with a maximum fold change compared to control five months post 13Gy irradiation (**FIGURE 3.21B**). This is the time point during which the lungs are fibrotic, with a disorganization of the lung architecture and loss of the alveoli structure (Knudsen, Ruppert, and Ochs 2017). We can make the hypothesis that the late AT1 senescence phenotype might be the consequence of the loss of the epithelial-endothelial contact with the aCap.

On the contrary, in the AT2 cells the increase starts as soon as 24 hours post 13Gy irradiation or one month post 10Gy irradiation (**FIGURE 3.21A**), with also a SASP score significantly upregulated compared to irradiation for all time points except four months post 10Gy irradiation and two months post 13Gy irradiation (**FIGURE 3.21B**). This consistent

increase after irradiation is the sign of a global response of this population, however it is not restricted to fibrogenic condition, therefore it is difficult to evaluate the consequences of this process.

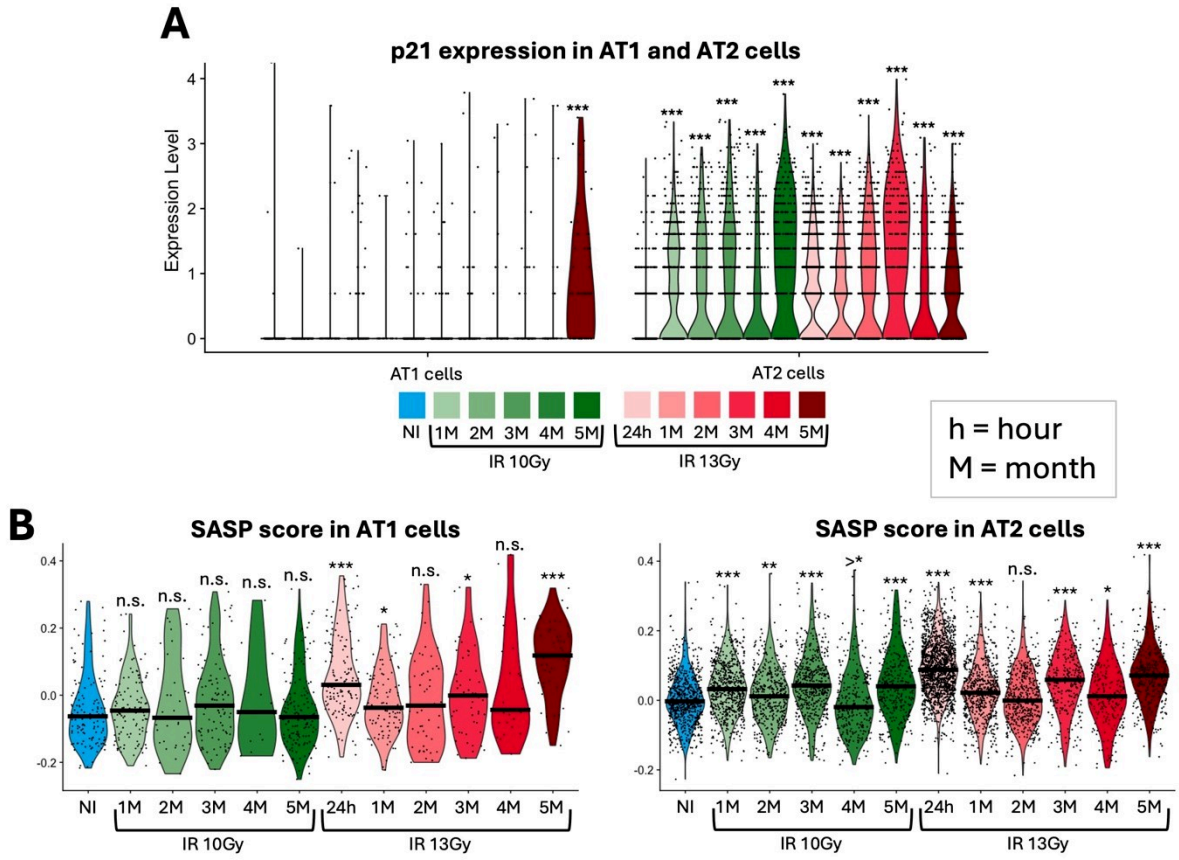


FIGURE 3.21. expression of senescence marker by the epithelial cells after irradiation injury.
A: expression of p21 by the AT1, AT2 and AT2 Lyz1 cells at the different timepoints poste 10Gy or 13Gy irradiation (24 samples). *** indicates a significant upregulation of expression compared to the control with an adjusted p-value of the MAST test inferior to 0.001; B: score based on the reactome senescence associated secretory phenotype (SASP) dataset in the AT1 or in the AT2 cells, stars represent the p-value of the wilcox test compared to the non-irradiated (n.s.: non-significant, *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001, >*: p-value < 0.05 downregulated compared to control)

b) Senescence of the myeloid cells

Senescence has been detected in RIPF in macrophages and these processes have been shown to be detrimental during the development of RIPF (Su et al. 2021).

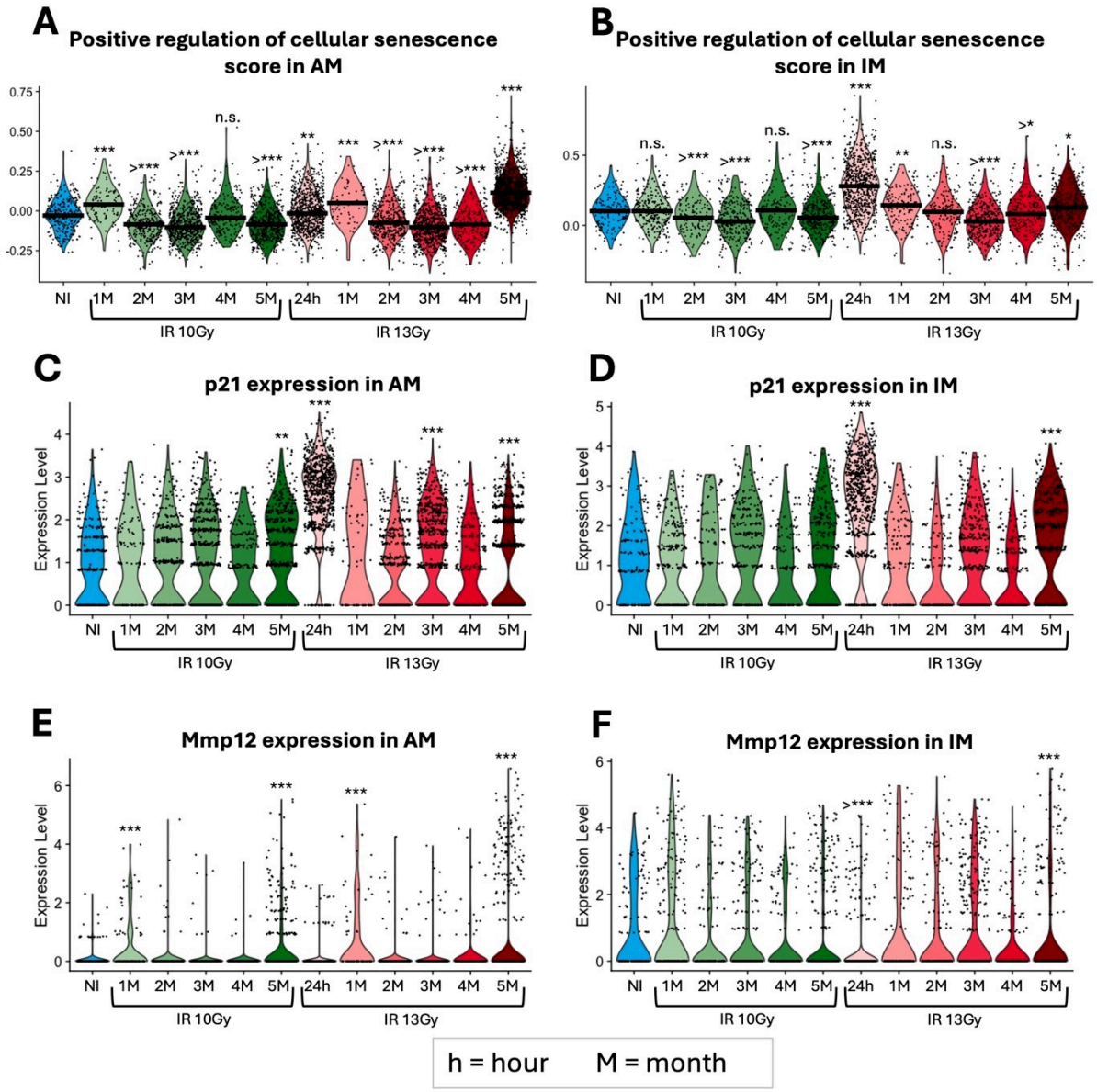


FIGURE 3.22. expression of senescence markers by the macrophages after irradiation injury.
A: SASP score based on the GO-PB positive regulation of cellular senescence dataset in the AM; B: SASP score based on the positive regulation of cellular senescence dataset in the IM, stars represent the p-value of the wilcox test compared to the non-irradiated (n.s.: non-significant, *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001, >: downregulated compared to control); C: p21 expression by the AM; D: p21 expression by the IM; E: Mmp12 expression by the AM; F: Mmp12 expression by the IM, stars represent the adjusted p-value of the MAST test compared to the non-irradiated (n.s.: non-significant, *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001, >: downregulated compared to control).

Therefore, we investigated the presence or not of senescence in our model, and the temporality of it. First, a score of positive regulation of cellular senescence showed some mixed results in the AM. The score was significantly increased one month post 10Gy irradiation and 24h, one- and five-months post 13Gy irradiation, and downregulated or non-significant for the rest of the time points (**FIGURE 3.22A**). Similarly, in the IM compartment, there was a significant upregulation of the senescence score 24h, one- and five-months post 13Gy irradiation (**FIGURE 3.22B**). Therefore, there seemed to be two waves of radiation induced senescence in the macrophages: one during the first weeks after irradiation, and one at the latest timepoint of fibrosis. The expression of other senescence markers follows a similar temporality: p21 (**FIGURE 3.22C,D**) and Mmp12 (**FIGURE 3.22E,F**), a pro-fibrotic SASP matrix metalloproteinase (Su et al. 2021; Freitas-Rodríguez, Folgueras, and López-Otín 2017).

c) Senescence of the endothelial cells

Endothelial cells are a vulnerable population that sustains damages induced by irradiation as demonstrated above. Therefore, we investigated the presence of radiation-induced senescence. First, we identified that p16, a classical marker of senescence (Rayess, Wang, and Srivatsan 2012) is expressed in the endothelial cells only after irradiation, and specifically in the aCap (**FIGURE 3.23A**). Indeed, there is a tendency of increase of expression of p16 during the latest time points after irradiation, however this increase is not significant (**FIGURE 3.23B**). We then computed a SASP score accounting for the global expression of senescence genes by the aCap. We found an initial significant increase of the score one-two month post 10Gy irradiation and 24 hours-one month post 13Gy irradiation, with then a return to levels comparable to controls and a final significant increase of the score five months post 10Gy or 13Gy irradiation, with a higher fold change compared to controls after a 13Gy irradiation (**FIGURE 3.23C**).

Previously, we made the hypothesis that a loss of contact between aCaps and AT1 cells due to alveolar structure destruction might be one of the reasons for the senescence of AT1 cells. Considering these results, it might also participate in the senescence of the aCaps after irradiation. However, this explanation cannot account for the senescence observed in the aCap at the earlier timepoints or after a 10Gy irradiation, therefore there are probably other processes involved in the induction of this phenotype.

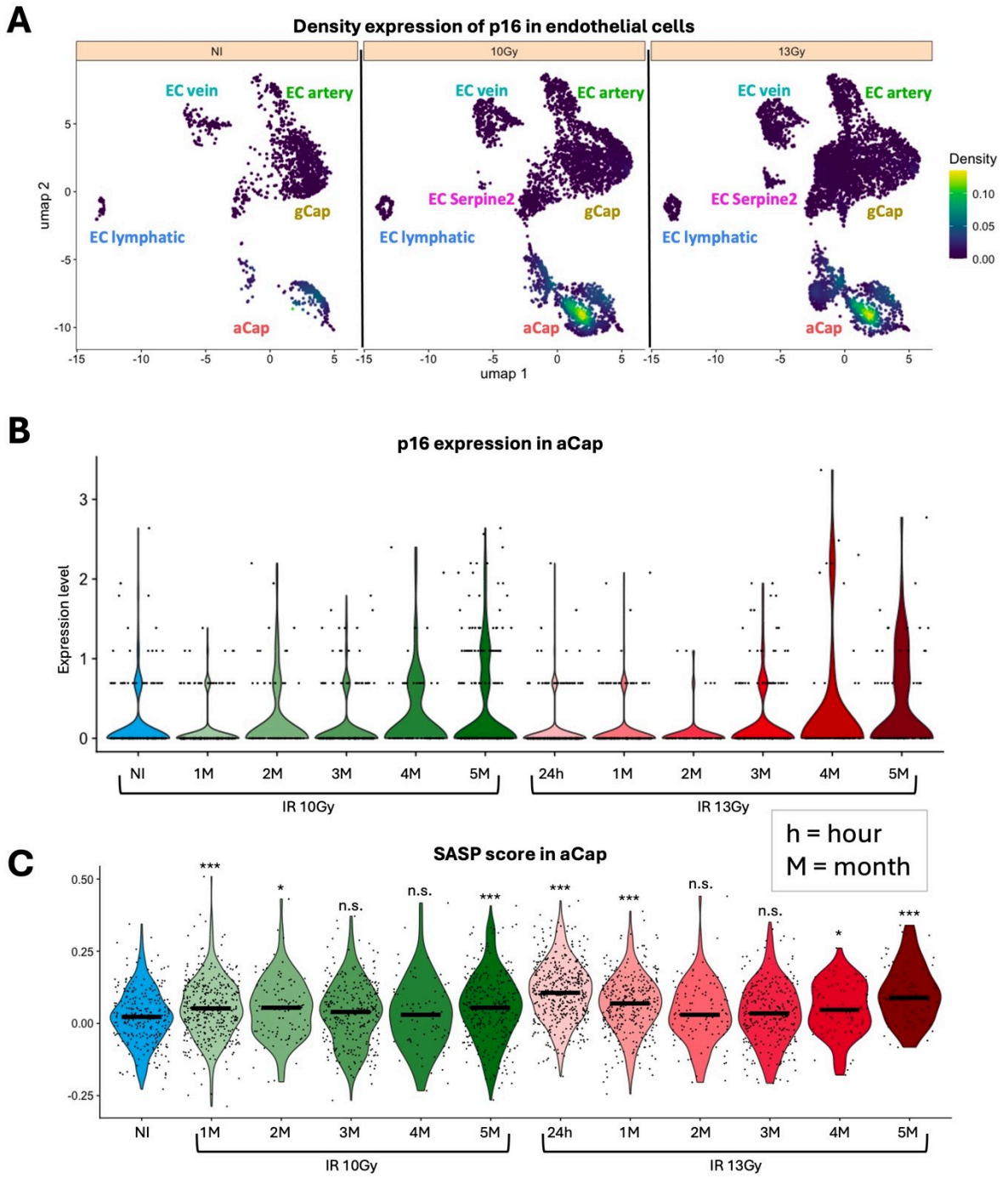


FIGURE 3.23. radiation-induced endothelial cell senescence.
A: density plot of the expression of p16 in endothelial cells in non-irradiated samples, 10Gy irradiated samples or 13Gy irradiated samples; B: violin plots of p16 expression in the non-irradiated post a 10Gy or 13Gy irradiation in aCaps; C: SASP score based on the reactome senescence associated secretory phenotype dataset in the aCap, stars represent the p-value of the wilcox test compared to the non-irradiated (n.s.: non-significant, *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001).

We then investigated whether this SASP phenotype persisted in the long term after a 10Gy irradiation. The expression of p16 in the samples nineteen months post 10Gy irradiation was significantly upregulated compared to both the age-matching control mice and the young control mice (**FIGURE 3.24A**).

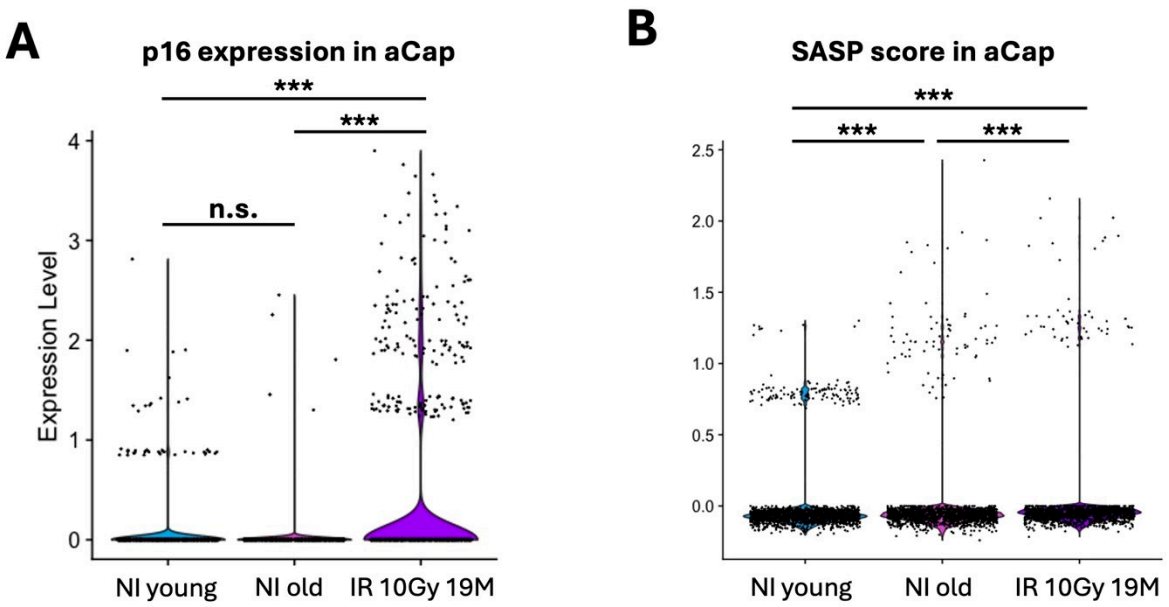


FIGURE 3.24. long term senescence of the aerocytes after irradiation.
A: violin plots of p16 expression in the non-irradiated and one to nineteen months post irradiation mouse aCaps, stars represent the p-value of the MAST test compared to the non-irradiated (n.s.: non-significant, *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001).; B: SASP score based on the reactome senescence associated secretory phenotype dataset in the aCap, stars represent the p-value of the wilcox test compared to the non-irradiated (n.s.: non-significant, *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001).

Indeed, the expression of p16 is significantly upregulated in the samples nineteen months post 10Gy irradiation compared to both the young and old non-irradiated controls (**FIGURE 3.24A**). Furthermore, the SASP score is upregulated in the irradiated sample compared to both controls. Interestingly, there is also a significant upregulation of the score in the old control compared to the young control (**FIGURE 3.24B**). Senescence has been shown to be a hallmark of aging (McHugh and Gil 2017), thus we can make the hypothesis that, in the long term, irradiation accelerates the aging processes.

As presented above, endothelial cells repair processes are activated after a radiation-induced lung injury, in response to the damage done to the micro-vasculature. We must verify that similar processes can be observed in humans, to determine to which extent the mouse model is clinically significant.

Therefore, during the next chapter, I will present the analysis of single cell RNA sequencing data from irradiated human lungs, starting with the endothelial repair processes triggered by pro-angiogenic signaling.

IV

CHAPTER IV
Results part II

CHAPTER IV – Results part II

To achieve these objectives, we chose to use the single cell RNA sequencing technology. We obtained samples from lobectomies from patients suffering from Pancoast tumor. These patients underwent neoadjuvant radiotherapy six to eight weeks prior to surgery. From each patient, one sample was obtained from an irradiated region of the lung and one control from a non-irradiated region of the lung (**FIGURE 4.1**). The details about the different samples can be found in **ANNEXE II**.

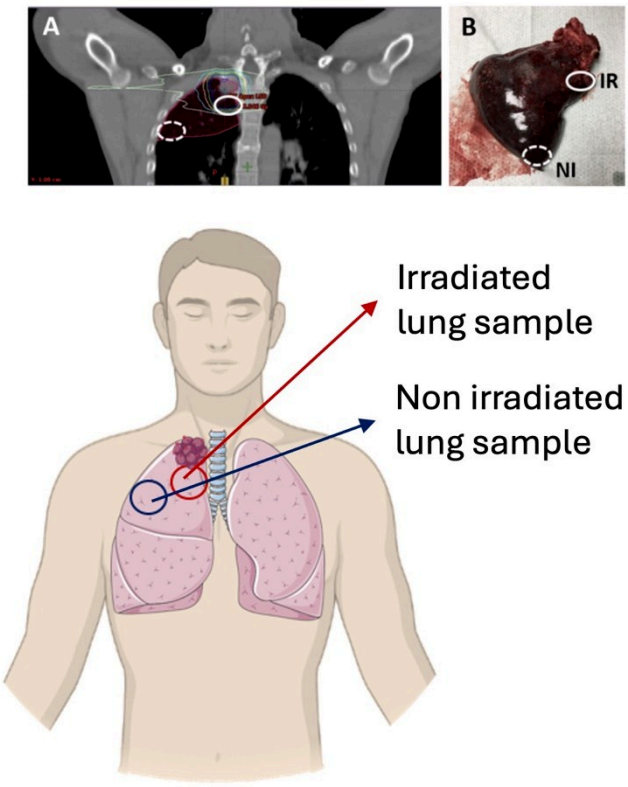


FIGURE 4.1. models used for the analysis of radio-induced pulmonary fibrosis. Collection of the non-tumoral human lung samples from Pancoast patients.

1. Single cell analysis of the human lung response to irradiation

As explained above, we obtained irradiated and non-irradiated samples from 6 patients who underwent neoadjuvant radiotherapy six to eight weeks prior to surgery. The control and irradiated samples from one of the patients were sorted to enrich the CD45 negative cells (the non-immune cells). Indeed, in non-sorted samples, we can see an elevated level of inflammation in the lungs of the patients as revealed by an overwhelming proportion of immune cells. The samples from the other five patients were sequenced entirely. This data set constitutes a unique resource for the study and the understanding of the cellular and molecular mechanisms of radio-induced injury in human lung tissue.

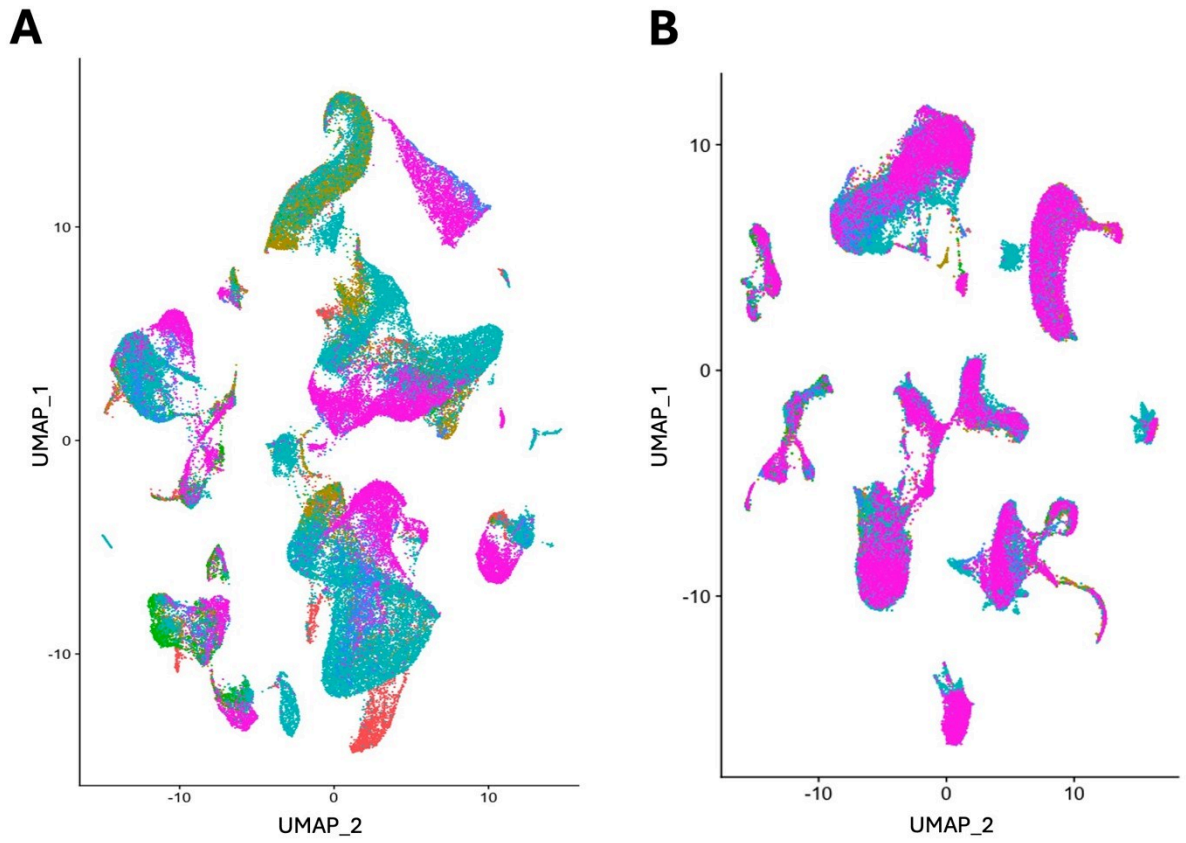


FIGURE 4.2. integration of the data from the six patients.
A: UMAP plot of the single cell RNA sequencing data of the samples from the lungs of the six patients, after merging; B: UMAP plot of the single cell RNA sequencing data of the samples from the lungs of the six patients, after integrating.

After sequencing, the data was processed with 10X Genomics Cell Ranger. One Seurat object for each patient was created and processed with SoupX (Young and Behjati 2020) for removal of the contaminating mRNA. Then quality controls were performed with filtration of

the cells with high mitochondrial content (more than 20% of mitochondrial mRNA, more than 6000 different features expressed, less than 200 different features expressed). The cell populations for each patient object were then annotated using the transfer learning tool scArches (Lotfollahi et al. 2022), and the HLCA (Sikkema et al. 2023) as a reference. Finally, the different objects were merged together to form a Seurat object containing the data from all six patients. When we merged the data from the different patients, we noticed an important batch effect (**FIGURE 4.2A**). Therefore, we chose to integrate the different patients together using the Seurat package for integration (Y. Hao et al. 2024) to correct for the batch effect (**FIGURE 4.2B**).

We were able to identify in the samples from human lungs the different lung cell populations: epithelial cells, endothelial cells, mesenchymal cells, myeloid cells and lymphoid cells (**FIGURE 4.3A**). These different cell populations can also be identified by the expression of canonical markers (**FIGURE 4.3B**), (Travaglini et al. 2020; Zilionis et al. 2019). Thus, this dataset allows us to analyze the human lung response to irradiation at the single cell level.

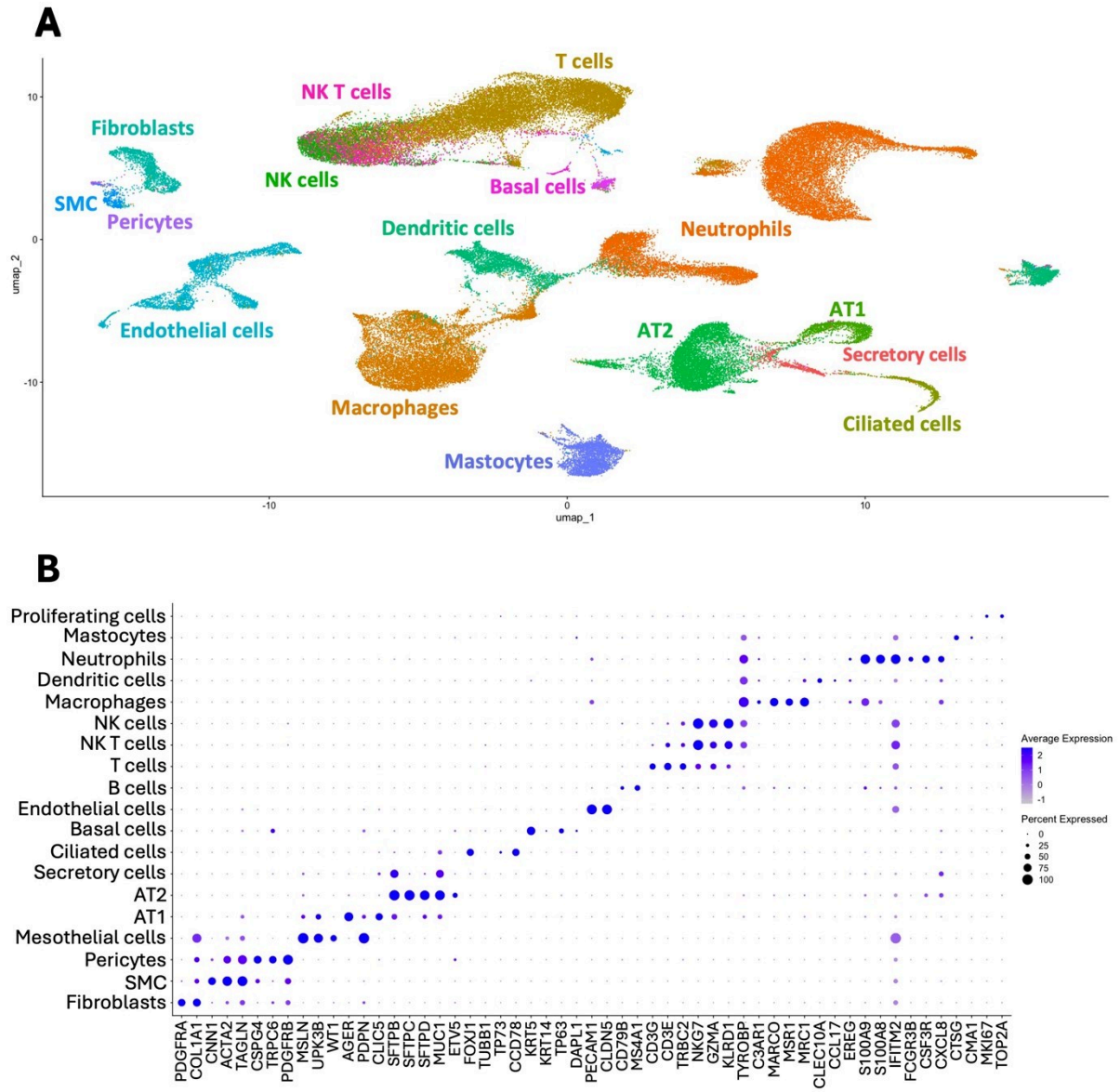


FIGURE 4.3. identification of the different cell populations of the human lung.
A: UMAP with the different lung cell populations annotated (12 samples, 77.048 cells); B: expression of the canonical markers by the different lung populations.

First, we analyzed the proportion of the different main cell populations before and after irradiation in the fully sequenced samples (**FIGURE 4.4**). We can notice the high variability between the different patients, especially in the proportion of immune cells. However, we can still make some observations. There is a decrease in the proportion of epithelial and endothelial cells after irradiation. This is coherent with the already shown vulnerability of these populations to irradiation (Trott, Herrmann, and Kasper 2004; Fuks et al. 1994).

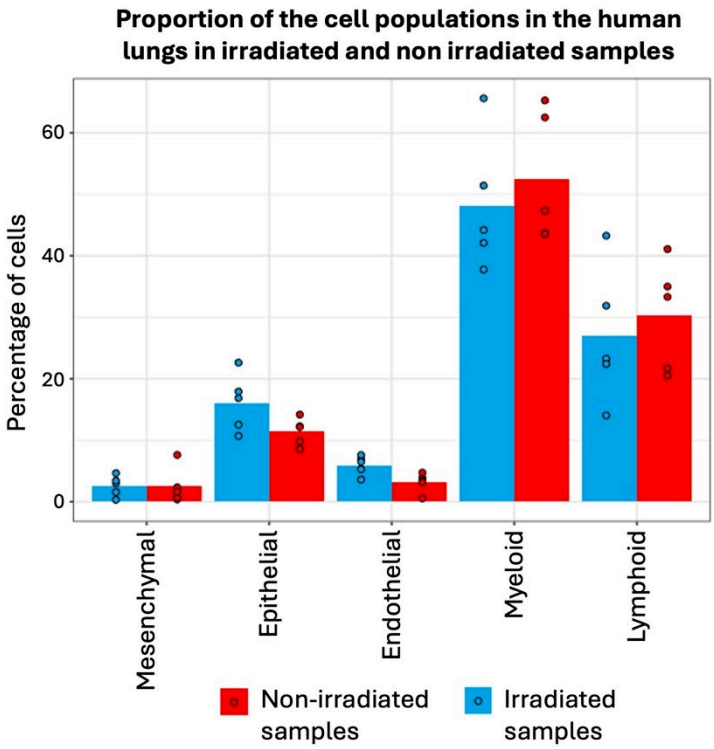


FIGURE 4.4. increase of the proportion of immune cells and decrease of the proportions of epithelial and endothelial cells after irradiation.
Bar plot of the mean percentage of the different cell populations before and after irradiation, each dot representing a different sample

2. Regeneration of the endothelial cells: processus of angiogenesis

a) Vascular damage and repair after a radiation induced lung injury

Vascular damage is a major consequence of lung irradiation. Irradiation causes death of endothelial cells, disorganization of the vascular network, as it has been shown in both human patients and mouse models (Venkatesulu et al. 2018). Pro-angiogenic signaling has been shown to promote vascular repair and capillary growth (Pecoraro et al. 2021).

Pro-angiogenic signaling promotes the sprouting of vessels and can occur in different situations, such as during embryonic development or in response to an injury. This process is crucial for growth, development, and wound healing. It involves the proliferation, migration, and differentiation of endothelial cells, which line the interior surface of blood vessels. Angiogenesis is essential for re-establishing an adequate blood supply to the damaged tissue, ensuring the delivery of oxygen and nutrients necessary for tissue repair and regeneration. Endothelial cells play a pivotal role in angiogenesis, responding to pro-angiogenic signals such as VEGF by proliferating, migrating, and forming new capillary networks.

Therefore, we studied here the molecular and cellular mechanisms triggered by radiation injury in the human lungs.

b) Non-tumor human lung irradiation during radiotherapy triggers a pro-angiogenic signaling (*publication*)

Non-tumor lung irradiation during radiotherapy triggers a pro-angiogenic signaling

Juliette SOULIER[1,2,3], Sandra CURRAS-ALONSO[2,4], Maxime DUBAIL[1,3], Hugo LAPORTE[1,3], Ayan MALLICK[1,3], Chloé LAFOUASSE[5], Marine LEFEVRE[5], Mylène BOHEC[6], Pierre VERRELLE[3,7,8], Nicolas GIRARD[5], Agathe SEGUI-GIVELET[5], Arturo LONDONO VALLEJO[1,3], Charles FOUILLADE[1,3]

[1]Curie Institute UMR3347/U1021, [2]Sorbonne University, [3]Paris Saclay University, [4]Curie Institute UMR3244, [5]Institut du Thorax Curie Montsouris - Institut Mutualiste Montsouris, [6]Curie Institute ICGex platform, [7]Curie Institute UMR9187/U1196, [8]Curie Institute dept of radiation oncology

Abstract

Radiotherapy is one of the main therapeutic options for the treatment of different types of lung cancer. However, radiotherapy can cause severe side effects to the normal tissue, from mild pneumonitis to irreversible and deadly pulmonary fibrosis. The mechanisms leading to this condition remain unclear although the lung microvasculature has been shown to be highly radiosensitive.

We obtained irradiated and non-irradiated non-tumor lung tissues samples from six Pancoast patients that underwent radiotherapy prior to surgery. Samples were analyzed through single cell RNA sequencing and paired irradiated and non-irradiated samples were compared.

We detected in irradiated samples a significant increase of VEGFA/VEFR2 signaling that affected most lung cell compartments. We also detected the emergence of Tip cells, an endothelial cell population responsible for endothelial cell migration at the leading edge of angiogenesis. In addition, we identified a sub-population of alveolar macrophages expressing FLT1+, a receptor for VEGFA, in irradiated lung samples. FLT1+ AM have been predicted to interact with tip cells through the pro-angiogenic interaction IL1B-IL1R. The analysis of single cell RNA sequencing data from a model of mouse radiation-induced pulmonary fibrosis also revealed the presence of tip cells during the inflammatory phase, at a time that may correspond to the one examined in the human samples.

We believe that this study has significantly advanced our understanding of how non-tumoral lung tissue responds to radiation, particularly in terms of microvasculature repair and regeneration. By unraveling these mechanisms, we aim to deepen our comprehension of the processes behind radiation-induced lung injury, which could ultimately lead to the development of new treatments to combat this side effect of radiotherapy.

Abbreviations

aCap = aerocyte
AM = alveolar macrophage
AT1 = type I epithelial cells
AT2 = type II epithelial cells
DC = dendritic cell
EC = endothelial cells
gCap = general capillary
HLCA = human lung cell atlas
IM = interstitial macrophage
NK = natural killer cell
IR = irradiated
NI = non irradiated
RIPF = radio induced pulmonary fibrosis
RILI = radio induced lung injury
scRNA seq = single cell RNA sequencing
SMC = smooth muscle cell

Introduction

Lung cancer is the leading cause of cancer-related death worldwide[1]. Most of the patients treated for this disease undergo radiation therapy. However, the lung is a sensitive organ to radiation[2], therefore the treatment is often limited by dose of irradiation that the lung can sustain. Radiation of healthy lungs induces damage, radio-induced lung injury (RILI), involving DNA damages[2] and oxidative stress[3], leading to inflammation and processes of wound healing in the lung tissues[4]. In some patients, this early toxicity can evolve into a chronic condition called radio-induced pulmonary fibrosis (RIPF)[5].

RIPF is the result of failed tissue regeneration after RILI and the replacement of normal tissue by scar due to excessive deposition of extracellular matrix, proliferation of fibroblasts [6], disruption of the alveolar structure and vascular damages[7]. These changes prevent gas exchange and lung function, leading to respiratory failure and death[5]. There is for the moment no efficient treatment to cure or even stabilize RIPF[8].

Even if some of the molecular and cellular events that occur in RILI and that lead to RILF have been described[9], the detail of the mechanisms and processes leading to fibrosis is not fully understood. Furthermore, most of what is known about RIPF comes from a model of total thoracic irradiation in the mouse[10], mostly because access to fresh human irradiated lung samples remain exceedingly rare.

Here we describe for the first time the effect of radiation injury on human lungs using a single cell RNAseq approach. We collected samples from patients suffering from Pancoast tumor and who underwent radiotherapy previously to a total lobectomy[11].

Material and methods

Human samples availability

Freshly resected lung human samples were obtained from six patients undergoing upper lobectomy of a Pancoast tumor who had previously received a neoadjuvant radiotherapy (40-45 Gy delivered by daily 2 Gy/fraction, considered sufficient to trigger pulmonary fibrosis). Accessibility to the dosimetric computerized tomography (CT)-scans of the patients allowed us to determine a region in the lobe far away from the tumor that did not receive any radiation (NI) and a region next to the tumor that received the highest dose of radiation (IR). A sample of 2 cm3 from each of these regions was resected and immediately placed in cold 1x phosphate buffered saline and transported on ice directly to the research lab for single cell dissociation procedure (**fig. 1a**). Accessibility to human samples was achieved in collaboration with Institut du Thorax and Cochin Hospital. Informed consent was obtained from each patient before the surgery.

Lung tissue dissociation

The human lung tissues dissociation was conducted using the protocol previously described [12], with an incubation of the suspension with elastase for 45 min.

Droplet based single cell RNA-seq (10X GENOMICS)

Single cell suspensions were analyzed with the droplet based single cell RNA-seq method proposed by 10X GENOMICS using the protocol previously described[12].

scRNA seq data analysis

Raw sequencing data were processed using the CellRanger pipeline (version 3.1.0, 6.0.0 or 7.1.0). Count matrices were analyzed using the Seurat package V5.0.1[13]. For each sample, SoupX[14] was used to remove contamination by ambient RNA. The objects from individual patients were annotated using ScArches[15] and the Human Lung Cell Atlas as a reference[16]. Then the Seurat objects of the different patients were integrated using the Seurat method (**fig. 1b**).

Results

The response of the human lung to radiation is characterized by the activation of pro-angiogenic pathways in many cell compartments.

First, we compared the transcriptomic profiles of the different lung populations in the IR and NI areas. We found an upregulation of numerous genes in the IR area of the lung, especially in certain populations (**Fig 1c**) including immune cells (500 to 3.300 overexpressed genes), endothelial cells (EC) (1.500 genes) and type II epithelial (AT2) cells (2.800 genes). We next analyzed the signaling pathways that were stimulated in the irradiated area. Strikingly, the strongest pathway upregulated across most cell populations was the VEGFA-VEGFR2 signaling pathway, known to be part of the vascular repair/regeneration pathway as well as the classic angiogenesis process[17].

To further dissect this prevalent pro-angiogenic signaling, we looked in different cell compartments at the level of expression of major ligands (VEGFA and VEGFB) and their receptors (FLT1 and KDR). Under normal (non-irradiated) conditions, epithelial cells (mostly AT1), endothelial cells (EC), interstitial macrophages (IM) and mastocytes appeared to be the main sources of VEGFA, while alveolar macrophages (AM), smooth muscle cells (SMC) and dendritic cells (DC) appeared to be the main sources of VEGFB (**fig. 2a**). Upon irradiation, the percentages of cells expressing the VEGFA gene increased in AT1, IM, fibroblasts and SMC, while for VEGFB, the most important increases were seen in AM, B cells, fibroblasts and AT1 cells (**fig. 2b**). The expression of receptors KDR and FLT1 was largely prevalent in the different types of ECs (**fig. 2c**), and this expression increased upon irradiation, especially in aCap (aerocytes) and gCap (general capillary cells) for KDR and in aCap and arterial ECs for FLT1 (**fig. 2d**). A low percentage of dendritic cells, neutrophils and AM was also as expressing FLT1 and this percentage increased upon irradiation (**fig. 2c-d**). This observation is interesting, since a FLT1+ AM population has been shown to play an important role in some processes during angiogenesis[18]. Together, these results point to a stimulation of the pro-angiogenic signaling that target both lung capillary EC subsets, aCap and gCap. As the gCap have been described as the progenitor population of the endothelial capillary cells compartment[19], we will focus the following analysis on this cell type.

Lung gCap ECs display gene expression patterns compatible with a Tip phenotype.

To better apprehend the consequences of an increase in the pro-angiogenic signaling pathway in response to radiation, we further focused our analyses on the different subsets of EC (**Fig. 3a**) (**supplementary fig. 3a**). Since the expressions of either KDR or FLT1 have been shown to be characteristic of two cell states involved in angiogenesis (tip and stalk cells, respectively), we defined two different scores, “tip” and “stalk”, using multiple markers described to be associated with either of these states[20]. We found a significant increase of the tip score in gCaps in irradiated lung areas, when compared to non-irradiated areas, that was concomitant with a significant decrease of the stalk score. These results suggest that irradiated gCap cells respond to the increase in pro-angiogenic signaling by acquiring “tip”-like characteristics. We then performed a DEG analysis by comparing gCap cells KDR positive

versus KDR negative. We found an over expression of genes associated with cell motility (**fig. 3c**). Furthermore, there is in the IR area a correlation of expression of several motility related genes with the expression of KDR. For instance, CAV1 promotes EC polarization and movement[21] and ROBO4 is involved in filopodia formation in EC[22] (**supplementary fig. 3a**). Together, these results support the idea that in response to irradiation, lung gCap cells gain "tip"-like properties that may play a key role in the vascular regeneration process.

Alveolar macrophages (AM) display a pro-angiogenic signature in response to radiation.

When looking at the expression of the angiogenic factors in the different cell populations, we identified a FLT1-positive population amongst resident macrophages. FLT1-positive macrophages have been shown to be crucial for an efficient angiogenesis in the lungs[18]. On the other hand, it has been showed that the expression of FLT1 by circulating monocytes is sufficient to attract them to the sites of VEGFA expression and stimulate their migration to the tissue[18, 23]. Analysis of our data show that there is an increase of FLT1 expression specifically in AM in response to radiation (**fig. 4a-b**). Next, we examined the level of expression of ITGAM, a characteristic marker of recruited AM [24]. This marker is expressed by most of the FLT1-positive AM macrophages (**supplementary fig. 4a**), suggesting that they may have been recruited from the circulation. Since such macrophages are thought to interact with ECs and support angiogenesis, we perform an intercellular interaction analysis to identify potential signals received by the gCap tip cells. This analysis revealed a potential IL-1b-mediated interaction between the FLT1-positive AM and gCaps. Indeed, we observe a significant increase of IL-1b expression after irradiation amongst the FLT1-positive AM, but not in the FLT1 negative AM (**fig. 4c**).

Sc RNA-seq analyses in a mouse model of RIFP reveal similar pro-angiogenic responses triggered by radiation.

We have previously published a mouse lung cell atlas of responses to radiation [12]. Given our observations in human lungs described above, we sought to verify whether similar phenomena were occurring in the irradiated mouse lung. In this model, samples were analyzed every month after either a single pro-fibrotic dose (17Gy) or a non-fibrogenic single dose (10Gy). We extracted the information corresponding to ECs from the whole data and annotated the different sub-populations using specific markers[19] (**fig. 5a**). We identified in the mouse the same EC populations identified in the human lungs, with the addition of a Serpine2 positive population that is only observed during the fibrotic phase. We found an increase in the proportion of gCap cells KDR positive starting two months after irradiation after both a 10Gy and a 17Gy irradiation (**fig. 5b**). Furthermore, after a 10Gy irradiation, this percentage starts to decrease again at four months post irradiation, while it increases until five months post a 17Gy irradiation. Similarly, the tip score in the gCap increases as soon as one month post irradiation and continue until five months after the 17Gy irradiation (**fig. 5c**). Contrary to what is observed in human patients after radiotherapy, we observe a slight increase of the stalk score after a 10Gy irradiation, starting at three months post irradiation. Finally, a 10Gy irradiation triggers an initial decrease of stalk score one month post irradiation, then the score increases at three

months post irradiation and decrease again to reach levels similar to controls at five months post irradiation.

We then investigated the motility phenotype described for the human KDR positive gCap. The comparison of gCaps KDR-positive versus KDR-negative cells from 10Gy irradiated samples showed the upregulation of 31 motility-related genes, while the cells from 17Gy irradiated samples showed only 26 motility-related genes upregulated.

Discussion

Radiotherapy triggers microvasculature damage to the healthy tissues[25], and a functioning vasculature is crucial for the proper function of the lungs. Therefore, we studied the mechanisms of vascular repairs and angiogenesis in the human after radiotherapy. The main trigger of pro-angiogenic signaling is the VEGFA-KDR interaction[26] and the VEGFA-FLT1 and VEGFB-FLT1 interactions[23]. A high percentage of EC express FLT1, a receptor that seems to have a complex role in angiogenesis. Indeed, it has a ten times higher affinity for VEGFA than KDR, but its kinase activity is ten times lower. Therefore, it has been thought to be a negative regulator of angiogenesis by trapping the VEGFA ligand[23]. However, studies have shown that expression of FLT1 by the EC is necessary for an efficient angiogenesis. In Flt1 deficient mice the sprouting of new vessels is compromised[27]. Therefore, the expression of FLT1 by the EC is not incompatible with processes of angiogenesis. However, KDR remains the main receptor able to trigger an angiogenesis response in EC[28].

The gCap, that are thought to function as a capillary progenitor in both homeostasis and disease, seems to be among the main angiogenic signaling populations, along with the aCap. An increased proportion of gCap expresses KDR after irradiation. KDR also a marker characteristic for tip cells. The tip cells are specialized in processes of sprouting and ramification. Therefore, this process could be increased in IR area in order to repair the damages to the vasculature. Tip cells develop filipodia and are able to migrate, following the extracellular gradient of VEGFA. Tip cells occupy a leading position in the new vessel formation and are followed by the stalk cells that divide to form the walls of the new vessel[29]. Tip cells receive the VEGF signaling, which triggers the expression of several genes including DLL4. DLL4 acts then as a ligand for the NOTCH receptor expressed by the stalk cells. In response in stalk cells, NOTCH activation affects the expression of the VEGF receptors, forming a VEGF-VEGFR-DLL4-NOTCH-VEGFR feedback loop[20, 30].

Angiogenesis appears to be increased by radiotherapy, and more EC appears to be able to receive it. This signaling could trigger a change in the balance of the cell states involved in angiogenesis that we observe, with an increase of tip cells. Furthermore, these tip cells present a phenotype of cell motility that could allow them to explore the environment through filopodia formation and to increase the ramification of the new vessels to restore the vascular network. Finally, we identified in the IR human lung area a population of alveolar macrophages that have been described to be crucial for angiogenesis. Therefore, it could play an important role in the wound healing and repair processes occurring after a RILI.

Furthermore, a subset of macrophages presents as FLT1 positive and thus are also able to receive the angiogenic signaling. These macrophages also express IL-1, a behavior that has been shown when stimulated by LPS, that is enough to induce angiogenesis in Matrigel plugs[31]. Furthermore, the inactivation of IL-1b in the supernatant of these macrophages is enough to impair this angiogenesis[31].

Overall, we are confident that this work contributed to advance the knowledge about the response of the non-tumoral lung to irradiation, especially concerning the mechanisms of microvasculature repair and regeneration. The deciphering of theses mechanisms participates

into the effort to gain a better understanding of the events involved in radiation induced lung injury and could contribute to the discovery of new therapeutic options to fight this side effect of radiotherapy.

Bibliography (max 40)

1 GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Lond Engl* 2018; 392: 1789–1858.

2 Giuranno L, Ient J, De Ruysscher D, *et al.* Radiation-Induced Lung Injury (RILI). *Front Oncol* [Internet] 2019 [cited 2024 Feb 27]; 9. Available from: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2019.00877.

3 Wei J, Wang B, Wang H, *et al.* Radiation-Induced Normal Tissue Damage: Oxidative Stress and Epigenetic Mechanisms. *Oxid Med Cell Longev* Hindawi; 2019; 2019: e3010342.

4 Zhao W, Robbins MEC. Inflammation and Chronic Oxidative Stress in Radiation-Induced Late Normal Tissue Injury: Therapeutic Implications. *Curr Med Chem* 2009; 16: 130–143.

5 Ding N-H, Jian Li J, Sun L-Q. Molecular Mechanisms and Treatment of Radiation-Induced Lung Fibrosis. *Curr Drug Targets* 2013; 14: 1347–1356.

6 Jarzebska N, Karetnikova ES, Markov AG, *et al.* Scarred Lung. An Update on Radiation-Induced Pulmonary Fibrosis. *Front Med* [Internet] Frontiers; 2021 [cited 2024 Apr 12]; 7. Available from: https://www.frontiersin.org/articles/10.3389/fmed.2020.585756.

7 Venkatesulu BP, Mahadevan LS, Aliru ML, *et al.* Radiation-Induced Endothelial Vascular Injury. *JACC Basic Transl Sci* American College of Cardiology Foundation; 2018; 3: 563–572.

8 Arroyo-Hernández M, Maldonado F, Lozano-Ruiz F, *et al.* Radiation-induced lung injury: current evidence. *BMC Pulm Med* 2021; 21: 9.

9 Chen Z, Wu Z, Ning W. Advances in Molecular Mechanisms and Treatment of Radiation-Induced Pulmonary Fibrosis. *Transl Oncol* 2019; 12: 162–169.

10 Wirsdörfer F, Jendrossek V. Modeling DNA damage-induced pneumopathy in mice: insight from danger signaling cascades. *Radiat Oncol* 2017; 12: 142.

11 PANCOAST HK. IMPORTANCE OF CAREFUL ROENTGEN-RAY INVESTIGATIONS OF APICAL CHEST TUMORS. *J Am Med Assoc* 1924; 83: 1407–1411.

12 Curras-Alonso S, Soulier J, Defard T, *et al.* An interactive murine single-cell atlas of the lung responses to radiation injury. *Nat Commun* Nature Publishing Group; 2023; 14: 2445.

13 Hao Y, Stuart T, Kowalski MH, *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* Nature Publishing Group; 2024; 42: 293–304.

14 Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience* 2020; 9: giaa151.

15 Lotfollahi M, Naghipourfar M, Luecken MD, *et al.* Mapping single-cell data to reference atlases by transfer learning. *Nat Biotechnol* Nature Publishing Group; 2022; 40: 121–130.

16 Sikkema L, Ramírez-Suástegui C, Strobl DC, *et al.* An integrated cell atlas of the lung in health and disease. *Nat Med* Nature Publishing Group; 2023; 29: 1563–1577.

17 Abhinand CS, Raju R, Soumya SJ, *et al.* VEGF-A/VEGFR2 signaling network in endothelial cells relevant to angiogenesis. *J Cell Commun Signal* 2016; 10: 347–354.

18 Murakami M, Zheng Y, Hirashima M, *et al.* VEGFR1 Tyrosine Kinase Signaling Promotes Lymphangiogenesis as Well as Angiogenesis Indirectly via Macrophage Recruitment. *Arterioscler Thromb Vasc Biol* American Heart Association; 2008; 28: 658–664.

19 Gillich A, Zhang F, Farmer CG, *et al.* Capillary cell-type specialization in the alveolus. *Nature* Nature Publishing Group; 2020; 586: 785–789.

20 Chen W, Xia P, Wang H, *et al.* The endothelial tip-stalk cell selection and shuffling during angiogenesis. *J Cell Commun Signal* 2019; 13: 291–301.

21 Beardsley A, Fang K, Mertz H, *et al.* Loss of Caveolin-1 Polarity Impedes Endothelial Cell Polarization and Directional Movement *. *J Biol Chem* Elsevier; 2005; 280: 3541–3547.

22 Sheldon H, Andre M, Legg JA, *et al.* Active involvement of Robo1 and Robo4 in filopodia formation and endothelial cell motility mediated via WASP and other actin nucleation-promoting factors. *FASEB J* 2009; 23: 513–522.

23 Shibuya M. Vascular endothelial growth factor receptor-1 (VEGFR-1/Flt-1): a dual regulator for angiogenesis. *Angiogenesis* 2006; 9: 225–230.

24 Wohnhaas CT, Baßler K, Watson CK, *et al.* Monocyte-derived alveolar macrophages are key drivers of smoke-induced lung inflammation and tissue remodeling. *Front Immunol* [Internet] Frontiers; 2024 [cited 2024 Apr 12]; 15. Available from: https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2024.1325090/full.

25 Kim JH, Jenrow KA, Brown SL. Mechanisms of radiation-induced normal tissue toxicity and implications for future clinical trials. *Radiat Oncol J* 2014; 32: 103–115.

26 Eldridge L, Wagner EM. Angiogenesis in the lung. *J Physiol* 2019; 597: 1023–1032.

27 Kearney JB, Kappas NC, Ellerstrom C, *et al.* The VEGF receptor flt-1 (VEGFR-1) is a positive modulator of vascular sprout formation and branching morphogenesis. *Blood* 2004; 103: 4527–4535.

28 Shibuya M. Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signaling in Angiogenesis. *Genes Cancer* 2011; 2: 1097–1105.

29 Jakobsson L, Franco CA, Bentley K, *et al.* Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting. *Nat Cell Biol* Nature Publishing Group; 2010; 12: 943–953.

30 Pasut A, Becker LM, Cuypers A, *et al.* Endothelial cell plasticity at the single-cell level. *Angiogenesis* 2021; 24: 311–326.

31 Carmi Y, Voronov E, Dotan S, *et al.* The Role of Macrophage-Derived IL-1 in Induction and Maintenance of Angiogenesis1. *J Immunol* 2009; 183: 4705–4714.

Figure 1. Upregulation of angiogenesis in the irradiated human lung

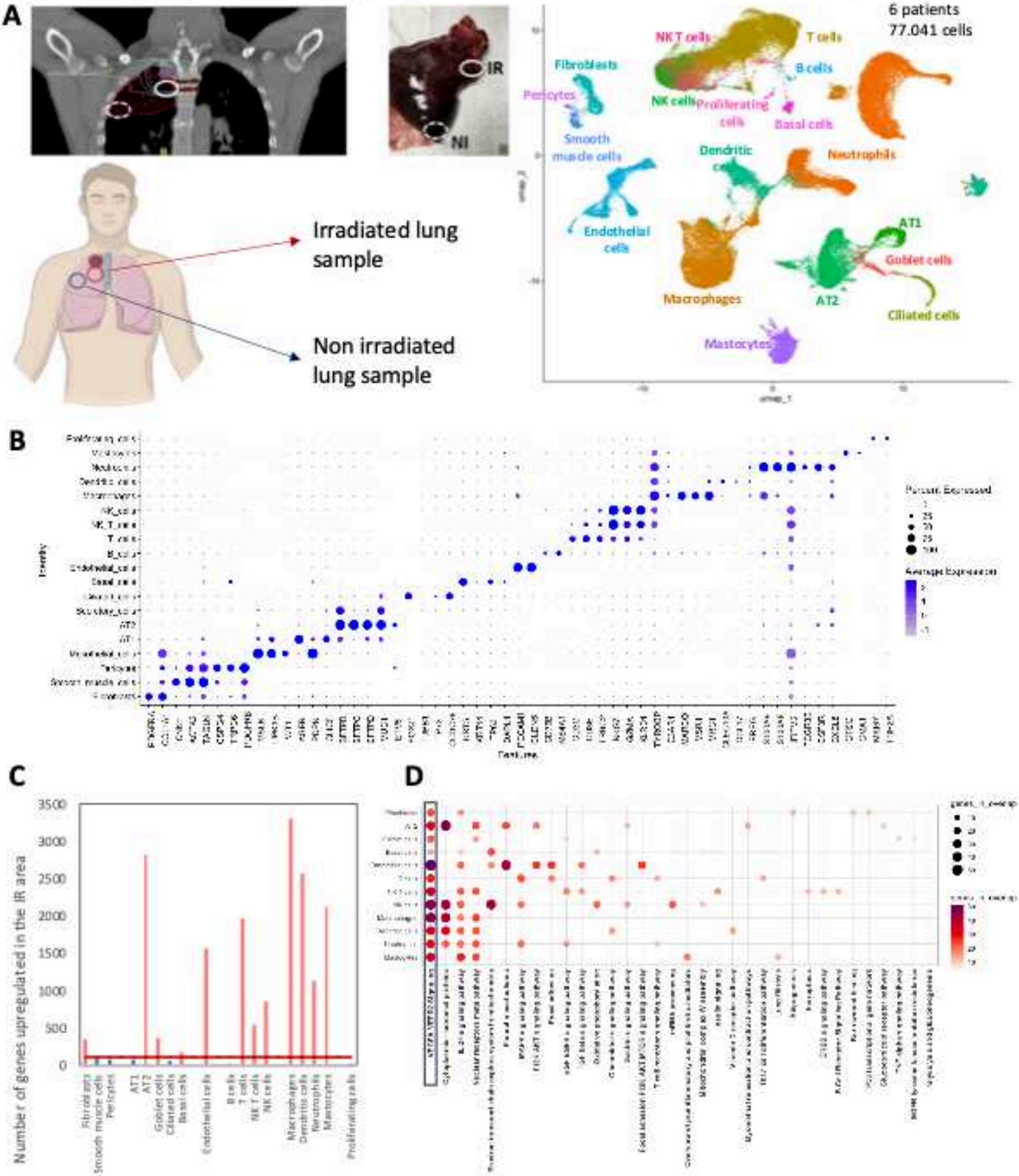


Figure 1. Angiogenesis signaling is upregulated in the irradiated human lung
A: dosimetric CT-scan of a patient who received radiation therapy at Institut Curie and surgery at IMM following a Pancoast tumor, surgical sample (6-8 weeks post radiotherapy), schematic representation of the location of the irradiated and non-irradiated non -tumoral samples from the Pancoast patients and UMAP visualization of 77.048 cells from six patients with 33.328 cells from the NI samples and 43.720 cells from the IR samples , annotated by cell type; B: DotPlot of the expression of characteristic markers of the different lung populations; C: number of genes significantly upregulated in the irradiated area of the lung compared to the irradiated in the different lung populations; D: results of the gene set enrichment analysis of the upregulated genes in the IR area compared to the NI area, main pathways from the WikiPathways database.

Figure 2. An increase after irradiation in angiogenic ligands and receptors in different cell types

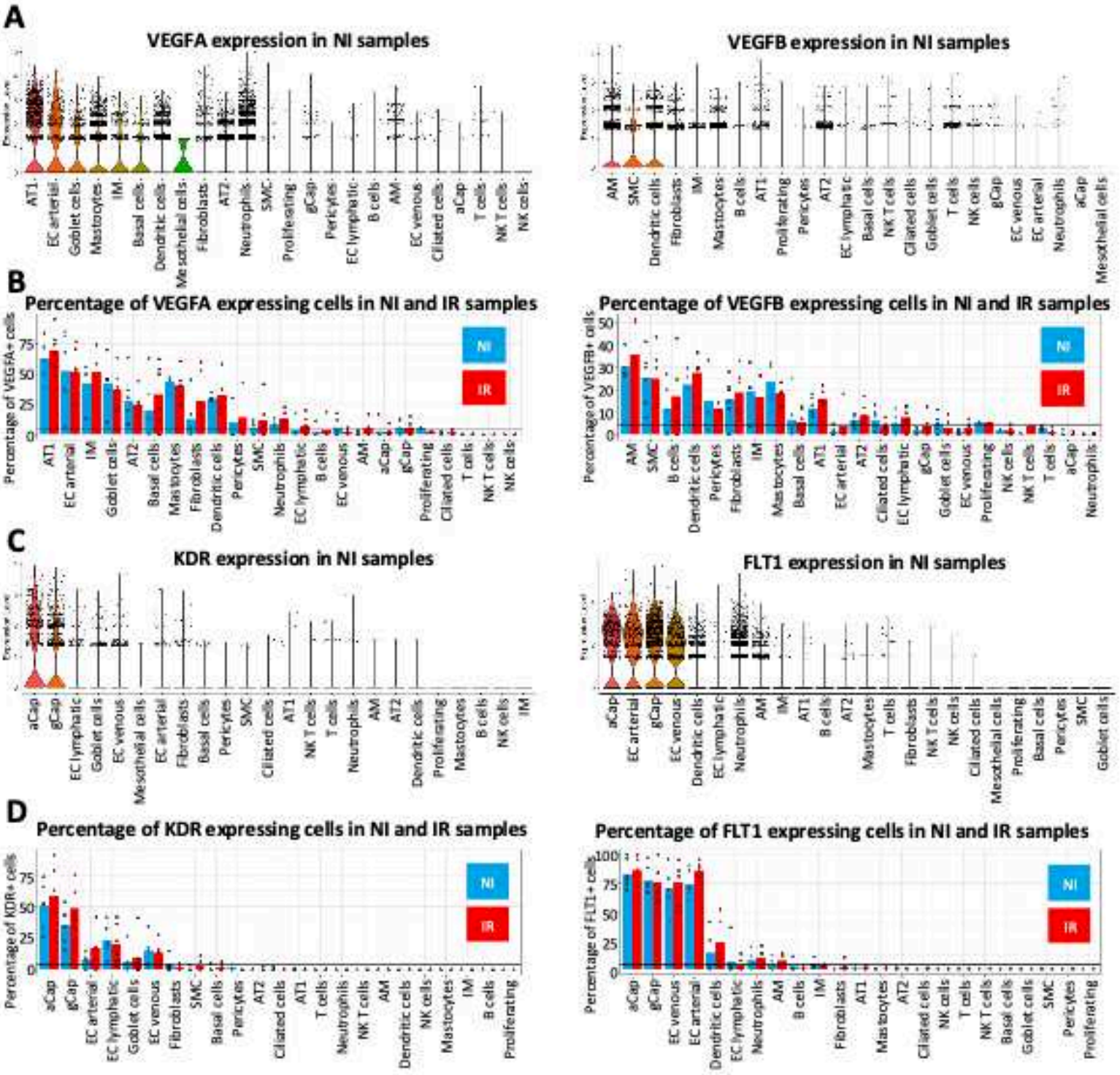


Figure 2. Angiogenic ligands and receptors show an increased expression after irradiation in different cell types
A: violin plots of VEGFA (left panel) and VEGFB (right panel) expression in the non-irradiated area in the different lung populations, sorted by intensity of expression; B: percentage of cells expressing VEGFA (left panel) or VEGF (right panel) in the irradated or non-irradiated area, with each dot representing a patient; C: violin plots of KDR (left panel) and FLT1 (right panel) expression in the non-irradiated area in the different lung populations, sorteb by intensity of expression; D: percentage of cells expressing KDR (left panel) or FLT1 (right panel) in the irradated or non-irradiated area, with each dot representing a patient; E: schematic representation of the increase in VEGF-VEGFR signaling in the irradiated area compared to the non-irradiated area.

Figure 3. Irradiation induces pro-angiogenesis processes in gCaps

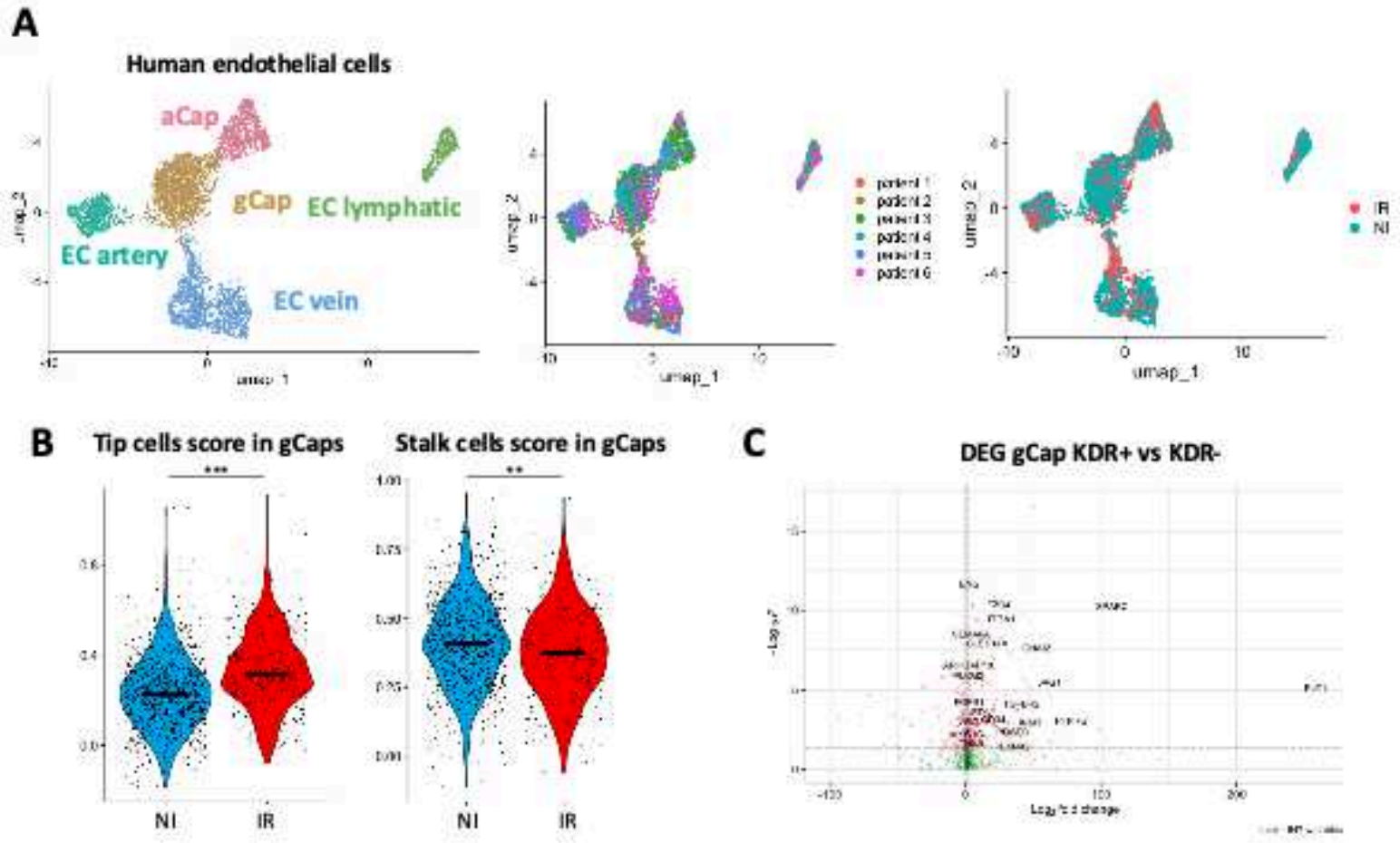


Figure 3. Irradiation induces pro-angiogenesis processes in gCaps
A: UMAP visualization of 3.648 cells from the different endothelial cell subpopulations annotated by sub cell type, by the patient of originand by the area (irrandiated or non-irradiated) of origin; B: violin plot of the stalk and tip cells scores calculated based on the list of markers for these two cell states, the black line represents the median; C: volcano plot representation of the DEG analysis of the gCaps KDR positive versus the gCaps KDR negative (genes overexpressed in the gCaps KDR positive have a positive fold change). The genes named are the genes related to cell motility (from the GOPB cell motility human gene list).

Figure 3. Supplemetary

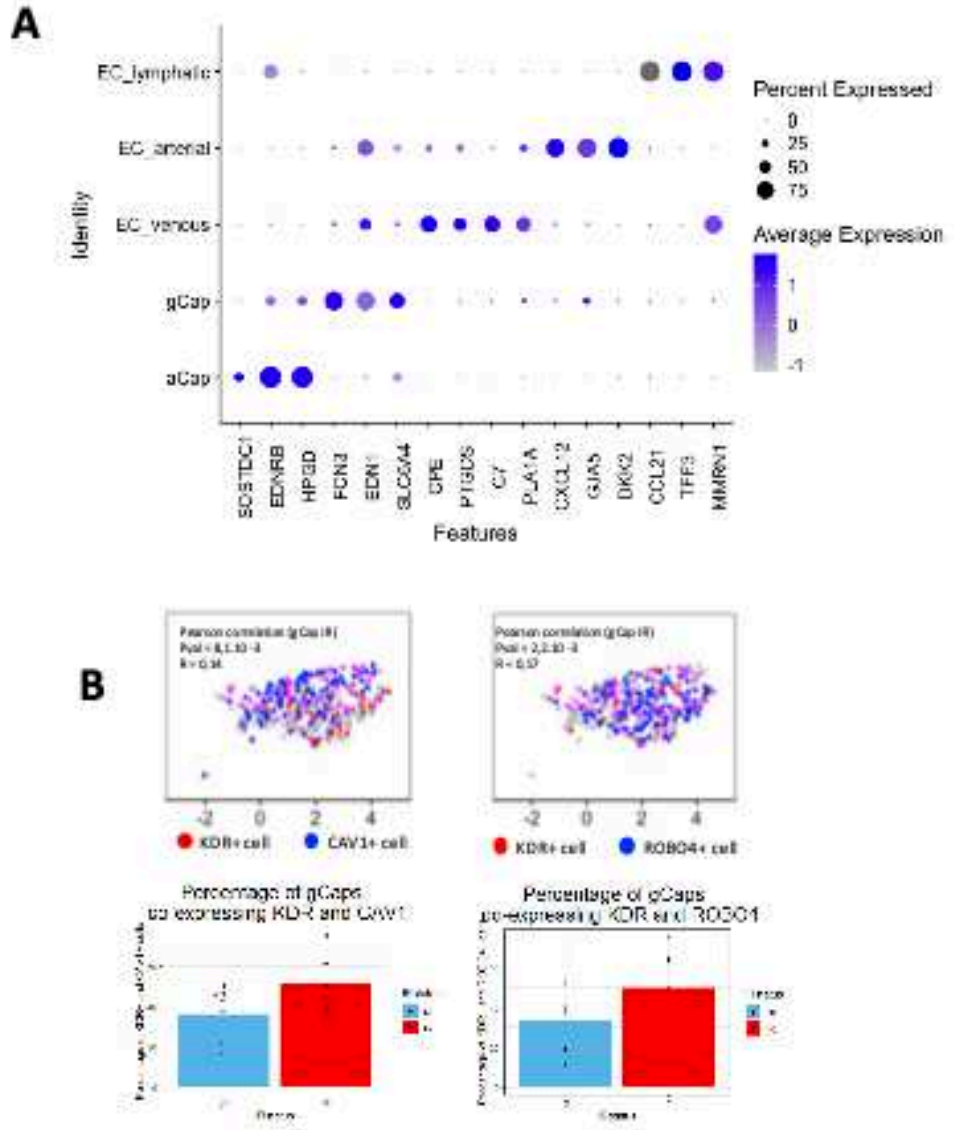


Figure 3. Supplemetary
A: DotPlot of the expression of characteristic markers of the different endothelial cell sub populations in human; B: FeaturePlot representing the coexpression of KDR and CAV1 (left panel) or ROBO4 (right panel) in the gCap irradiated cells. Pink cells are coexpressing the genes and percentage of cells co-expressing KDR and CAV1 (left panel) or ROBO4 (right panel) in the non-irradiated or irradiated areas.

Figure 4. Radiation induces FLT1+ macrophages secreting IL1 signaling

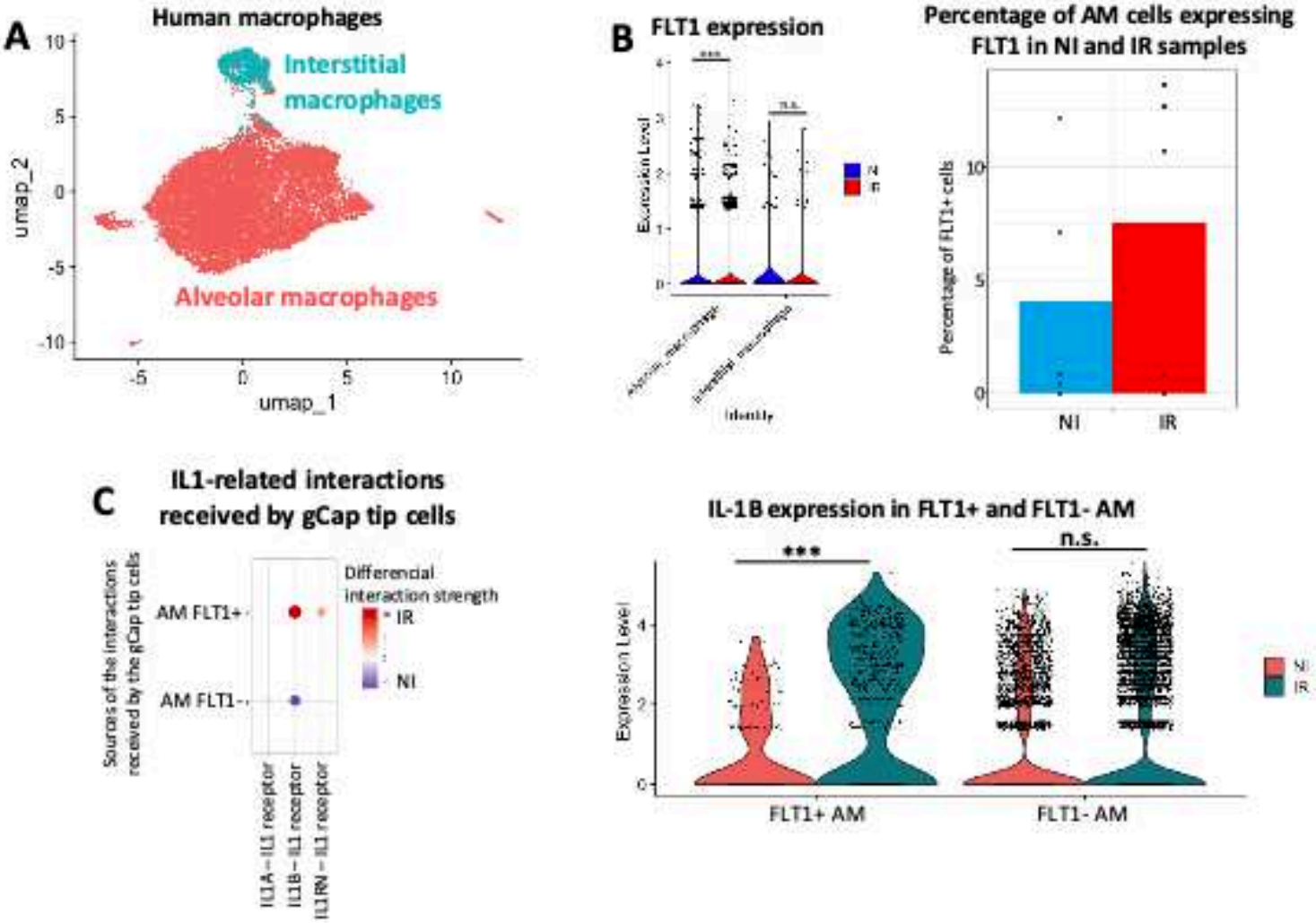


Figure 4. Alveolar macrophages receive and send VEGF signaling after irradiation
A: UMAP visualization of 14.168 cells from the different macrophage subpopulations annotated by sub cell type (4.159 cells from the NI area and 7.009 from the IR area); B: violin plots of FLT1 expression in the non-irradiated and irradiated area in the interstitial and alveolar macrophages and percentage of FLT1 positive alveolar macrophages in non-irradiated and irradiated area, each dot representing a patient; C: DotPlot of the IL-1 related intercellular interactions received by the gCap tip cells and sent by the alveolar macrophages FLT1 positive or negative and violin plots of IL-1 alpha expression in the non-irradiated and irradiated area in the FLT1 positive or FLT1 negative alveolar macrophages.

Figure 4. Supplemetary

A

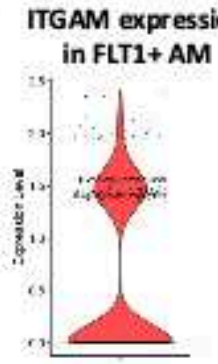


Figure 4. Supplemetary
A: violin plot of ITGAM expression in the FLT1 positive alveolar macrophages.

Figure 5. Dynamic of angiogenesis processes in mouse irradiated lungs

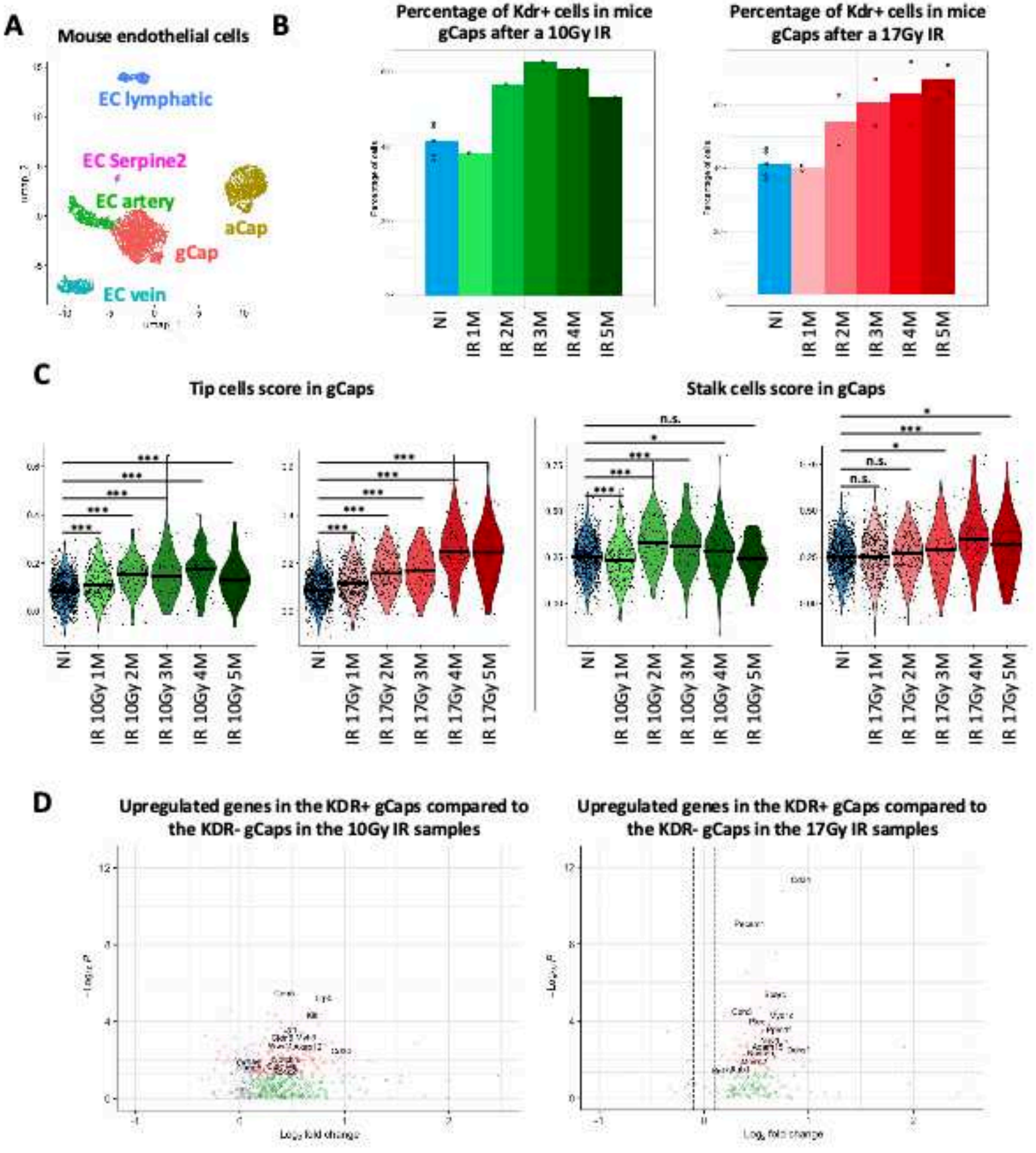


Figure 5. Angiogenesis processes evolve during the months post irradiation in mouse lungs
A: UMAP visualization of 4.958 cells from the different mouse endothelial cell subpopulations annotated by sub cell type; B: proportion KDR+ cells among the gCaps, in the control mice or after a 10Gy or 17Gy irradiation; C: violin plot of the stalk and tip cells scores calculated based on the list of markers for these two cell states, the black line representing the median; D: volcano plot representation of the DEG analysis of the gCaps KDR positive versus the gCaps KDR negative (genes overexpressed in the gCaps KDR positive have a positive fold change) in the 10Gy irradiated gCaps (left panel) or the 17Gy irradiated gCaps (right panel). The genes named are the genes related to cell motility (from the GOPB cell motility mouse gene list).

Figure 5. Supplemetary

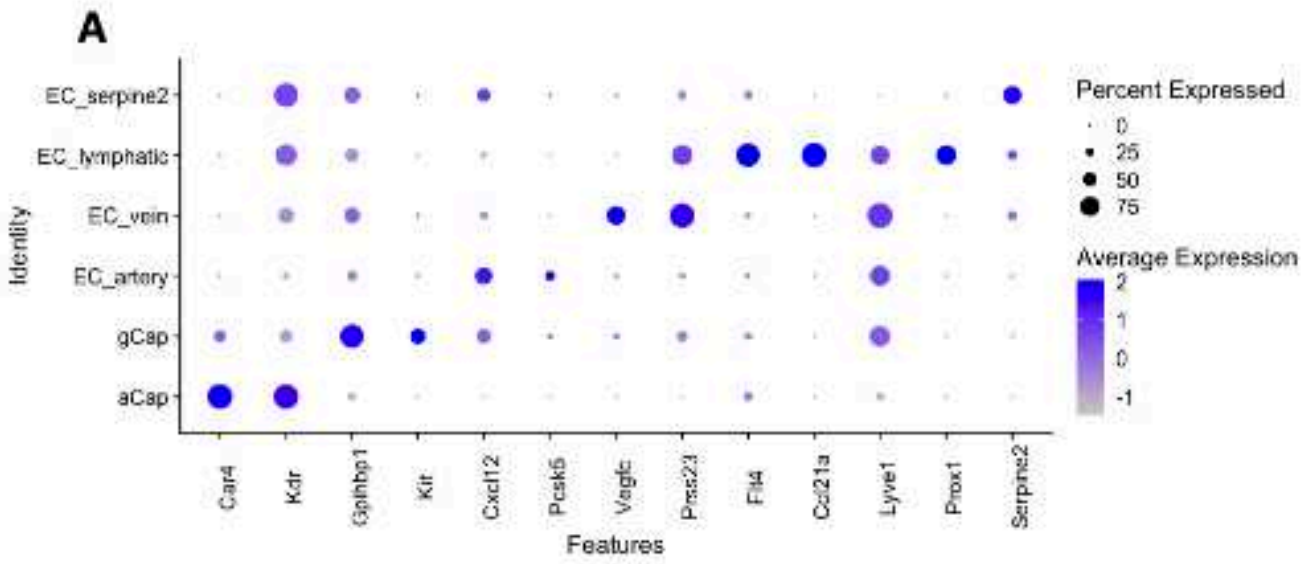


Figure 5. Supplemetary
A: DotPlot of the expression of characteristic markers of the different endothelial cell sub populations in mice.

3. Processes of inflammation induced by radiotherapy in non-tumoral lung tissue

a) Identification of the different immune cell populations in the human lung

The use of single cell RNA sequencing allowed us to identify the different immune populations in both the irradiated and non-irradiated human lung: the B cells, T cells, NK cells and NK-T cells, neutrophils, dendritic cells, alveolar macrophages and interstitial macrophages. All of these populations were present in both the irradiated and non-irradiated samples (**FIGURE 4.5A,B**). The different cell types were identified using markers from literature (Travaglini et al. 2020) (**FIGURE 4.6A**). We can observe some differences in the proportion of the different immune populations in the irradiated and non-irradiated samples: an increase in the proportion of NK-T cells, dendritic cells and alveolar macrophages, and a decrease in the proportion of NK cells and neutrophils (**FIGURE 4.6B,C**).

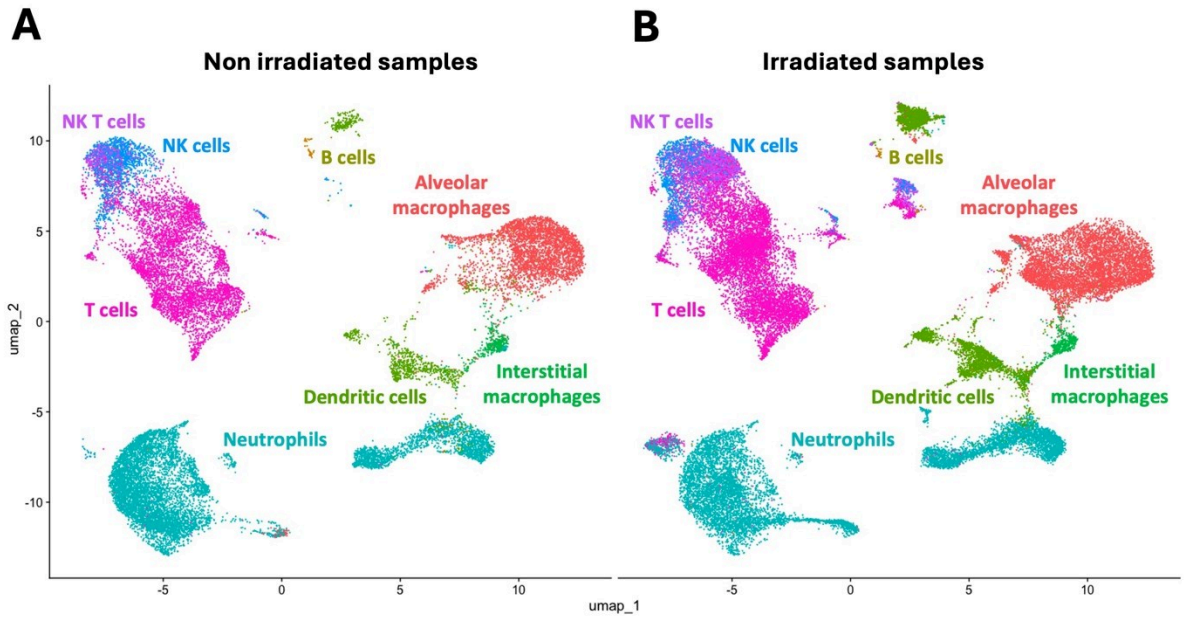


FIGURE 4.5. immune cells in the non-irradiated and irradiated human lung.
A: UMAP visualization of the cells from the different human immune cell subpopulations annotated by sub cell type in the non-irradiated human samples (6 samples, 21.527 cells); B: UMAP visualization of the cells from the different human immune cell subpopulations annotated by sub cell type in the irradiated human samples (6 samples, 34.467 cells).

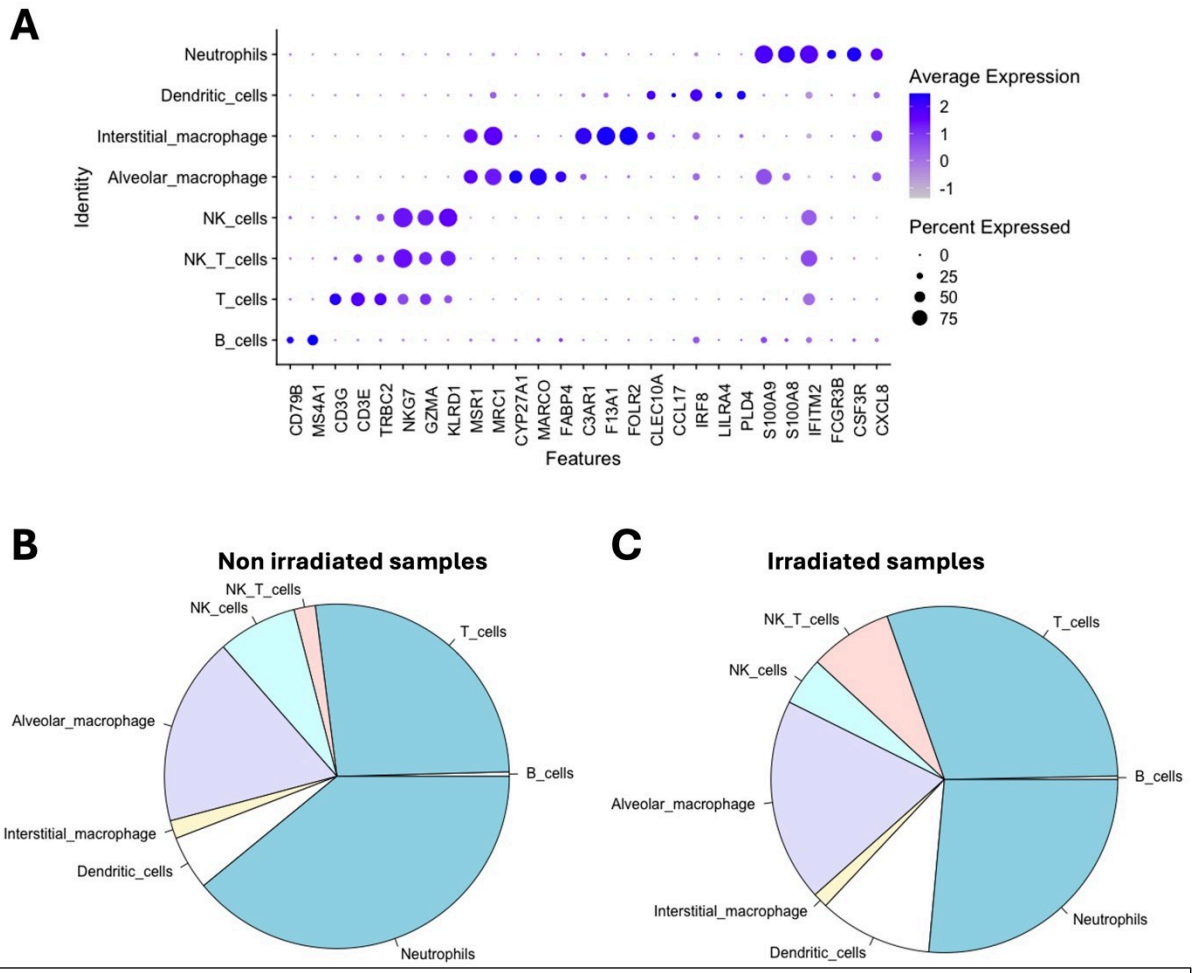


FIGURE 4.6. markers and proportion of the immune cell in the non-irradiated and irradiated human lung.
A: Expression of the markers used for the identification of the immune cell populations; B: Proportion of the different immune cell populations in the non-irradiated human lung; C: Proportion of the different immune cell populations in the irradiated human lung.

b) Upregulation of different pathway in the irradiated human lung

We identified various pathways upregulated in the immune cells from irradiated human samples compared to the non-irradiated samples. Similarly to the mouse samples, several inflammation related pathways are upregulated: the IL-24 signaling pathway in most immune cell populations and other interleukin signaling pathways in diverse immune populations, the MAPK and chemokine signaling pathway (**FIGURE 4.7**).

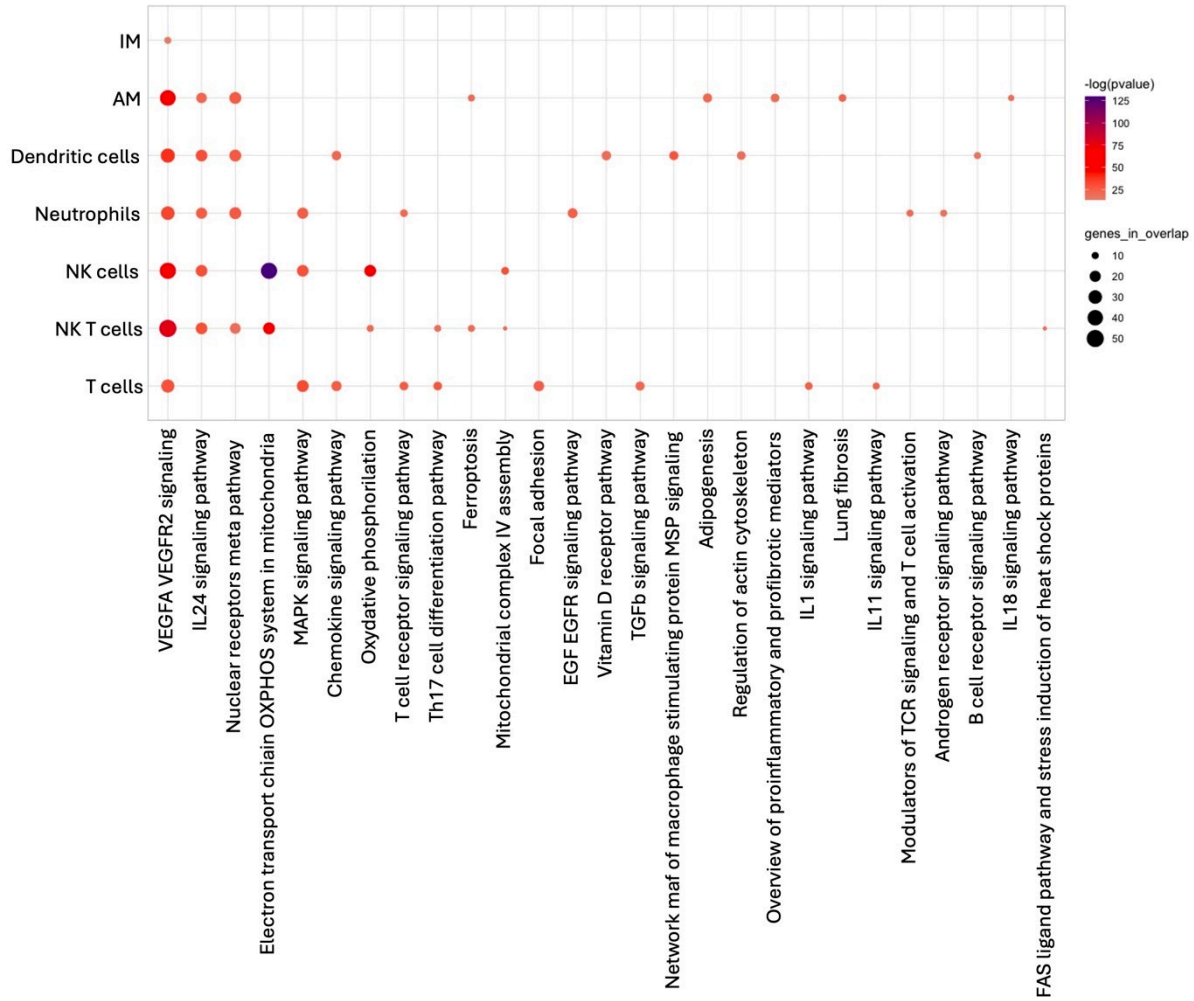


FIGURE 4.7. overview of overexpressed pathways in human immune cells after irradiation.
GSEA WikiPathways significantly upregulated (FDR adjusted p-value < 0.05) in irradiated samples compared to non-irradiated samples in the different immune cell populations.

Furthermore, as in the mouse samples, we also found upregulated several pathways related to oxidative phosphorylation (OXPHOS) and mitochondrial functioning (**FIGURE 4.7**). Several studies have demonstrated the presence of mitochondrial dysfunction and changes in mitochondrial gene expression in several models of radiation induced fibrosis (Yin et al. 2019; Livingston et al. 2020). In the case of liver radiation induced fibrosis, it has been shown that mitochondrial dysfunction contributes to the development of fibrosis (Melin et al. 2022), therefore it could be interesting to investigate this process in lungs.

c) Macrophages

As shown previously with the mouse data analysis, macrophages play an important role in the lung's reaction to irradiation. Therefore, we investigated the possible changes in the human macrophages in the irradiated samples compared to the non-irradiated samples.

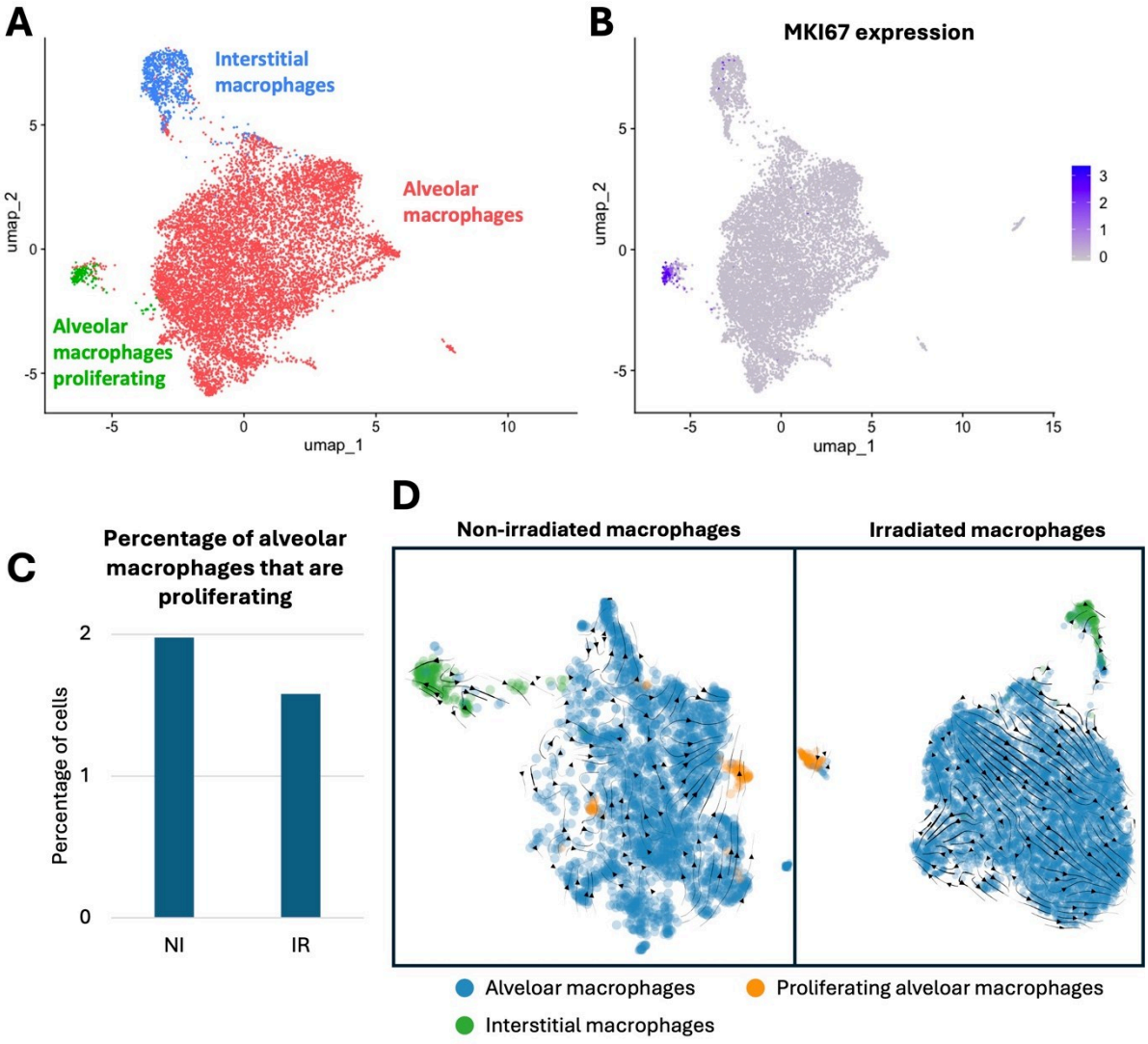


FIGURE 4.8. proliferation of the alveolar macrophages in the irradiated and non-irradiated human lungs.
A: UMAP visualization of human macrophages (12 patients, 11.168 cells); B FeaturePlot of the expression of MKI67 by the macrophages; C: proportion of proliferating alveolar macrophages among the total alveolar macrophages; D: RNA velocity analysis with scvelo of the irradiated and non-irradiated macrophages.

As in mice, we were able to identify the two macrophage populations: the interstitial macrophages and the alveolar macrophages (**FIGURE 4.8A**). Furthermore, a distinct

population of alveolar macrophages express a proliferation marker: MKI67 (**FIGURE 4.8B**). Therefore, we labelled this population “proliferating alveolar macrophages”. The proportion of alveolar macrophages that are proliferating in irradiated samples does not seem to be different than in non-irradiated samples (**FIGURE 4.8C**). However, an RNA velocity analysis showed more connections between the non-proliferating and proliferating alveolar macrophages in the non-irradiated samples than in the irradiated samples (**FIGURE 4.8D**). With this result, we can make the hypothesis that the replenishment of the alveolar macrophage population might be impaired after irradiation.

4. Regeneration of the epithelial cells after an irradiation injury

a) Identification of the different epithelial cell populations in the human lung

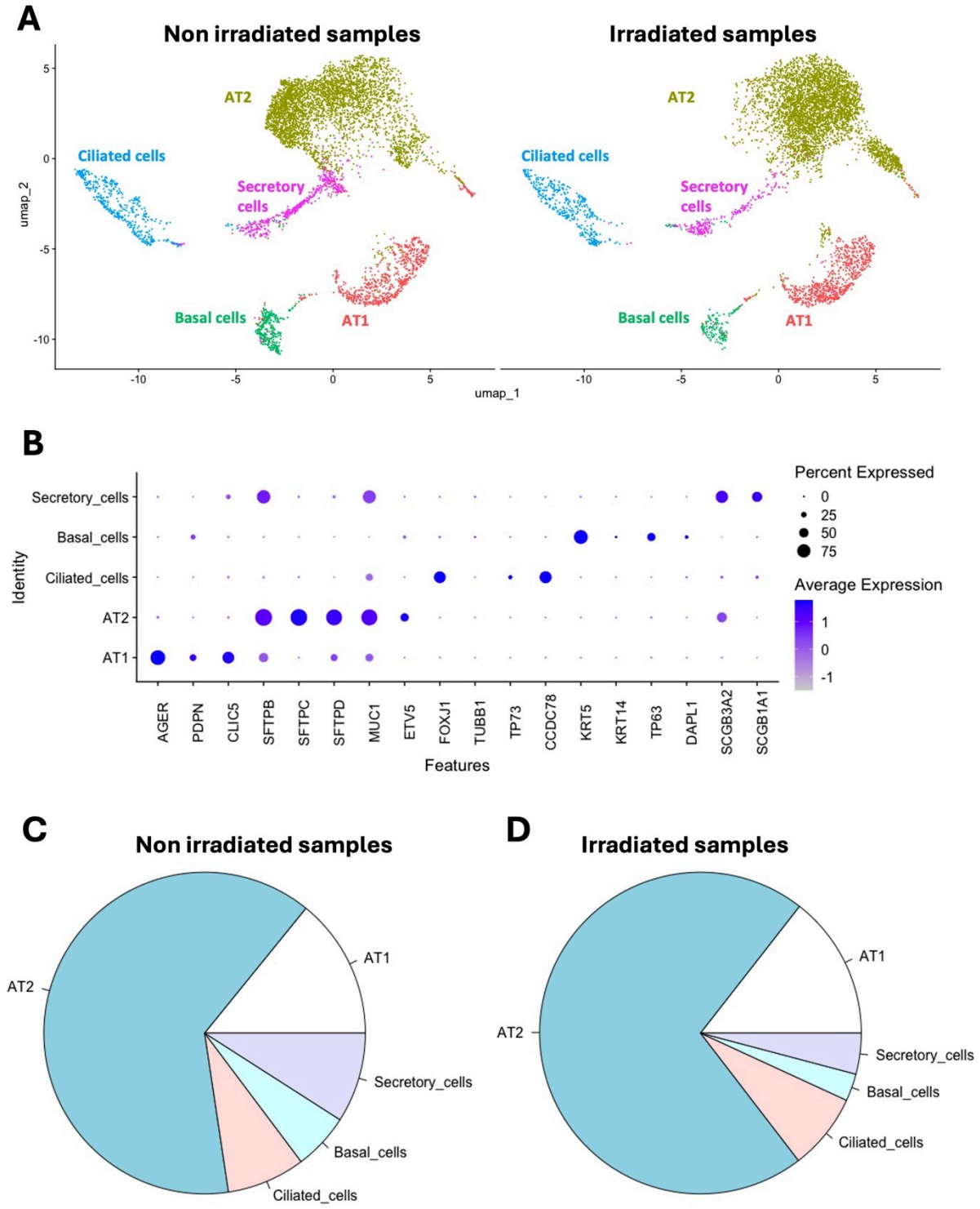


FIGURE 4.9. epithelial cells in the non-irradiated and irradiated human lung.
A: UMAP visualization of the cells from the different human epithelial cell subpopulations annotated by sub cell type (12 samples, 11.348 cells); B: Expression of the markers used for the identification of the epithelial populations; C: Proportion of the different epithelial populations in the non-irradiated human lung; D: Proportion of the different epithelial cell populations in the irradiated human lung.

As for the other cell compartments, we were able to identify the different epithelial cells population in the human lungs using single cell RNA sequencing: the ciliated cells, basal cells, secretory cells and the two pneumocytes populations, the AT1 and AT2 cells (**FIGURE 4.9A**). These populations were identified using markers described in the literature (Travaglini et al. 2020; Kadur Lakshminarasimha Murthy et al. 2022) (**FIGURE 4.9B**). We didn't detect any important changes in proportion of the epithelial cells in the irradiated samples compared to the non-irradiated samples. The AT2 cells are the majority of the population, followed by the AT1 cells and the other epithelial cell types (**FIGURE 4.9C,D**).

b) A shift in the pneumocytes transcriptome

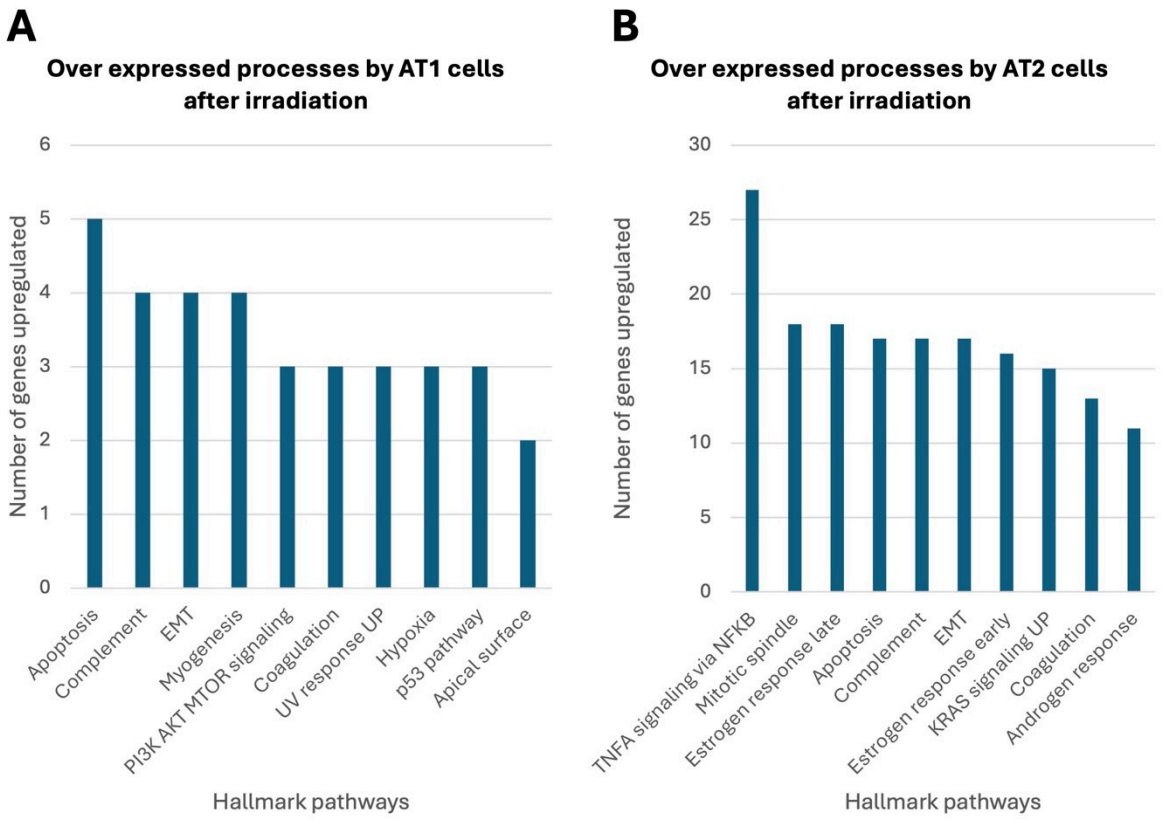


FIGURE 4.10. enrichment of hypoxia and EMT related pathways in the different AT1 and AT2 cells.
A: Top 10 HALLMARK pathways significantly upregulated (FDR adjusted p-value < 0.05) (ranked by the number of genes in the pathway upregulated) in the AT1 from the irradiated samples compared to the non-irradiated samples; B: Top 10 HALLMARK pathways upregulated (ranked by the number of genes in the pathway upregulated) in the AT2 from the irradiated samples compared to the non-irradiated samples

We then analyzed the differences between the irradiated and non-irradiated conditions for the two pneumocytes populations. In the AT1 cells, we see different pathways

upregulated indicating of damages to the population and a remodeling of the environment: apoptosis, epithelial to mesenchymal transition, response to hypoxia that could indicate a loss of contact with the aCap and p53, a marker of senescence (Zhou et al. 2022) (**FIGURE 4.10A**). Similar pathways are upregulated in the irradiated samples in the AT2 cells: the TNFα signaling pathway, apoptosis and EMT (**FIGURE 4.10B**).

CHAPTER V
Discussion

CHAPTER V – Discussion

The side effects of radiotherapy on healthy tissues are complex and many of the mechanisms leading to RILI and RIPF remain to be understood. Through the use of single cell RNA sequencing analysis, we gained new insights in the processes occurring in the different lung cell populations leading to fibrosis after an irradiation of healthy lung tissues. In this work, the focus was put mainly on epithelial cells, fibroblasts, macrophages and endothelial cells (**FIGURE 5.1**) as these populations has been shown to be particularly affected by irradiation and play an important part in the development or radiation induced pulmonary fibrosis (Ding, Jian Li, and Sun 2013; I. Y. Adamson and Bowden 1983).

These populations were studied using an extensive dataset of 34 single cell RNA seq samples from irradiated and control mice of 123.147 cells, with the addition of 37.292 cells from 6 irradiated and control samples from twenty-three-month-old mice. The comparison of the effect of non-fibrogenic (10Gy) and fibrogenic (13Gy) doses allowed us to compare the effect of irradiation on the lungs to the events leading to radiation-induced pulmonary fibrosis.

Furthermore, through the analysis of single cell RNA seq of non-irradiated and irradiated non tumoral lungs samples data from 6 patients and 77.048 cells, we were able to gain unique insight in the mechanisms of human lung response to irradiation.

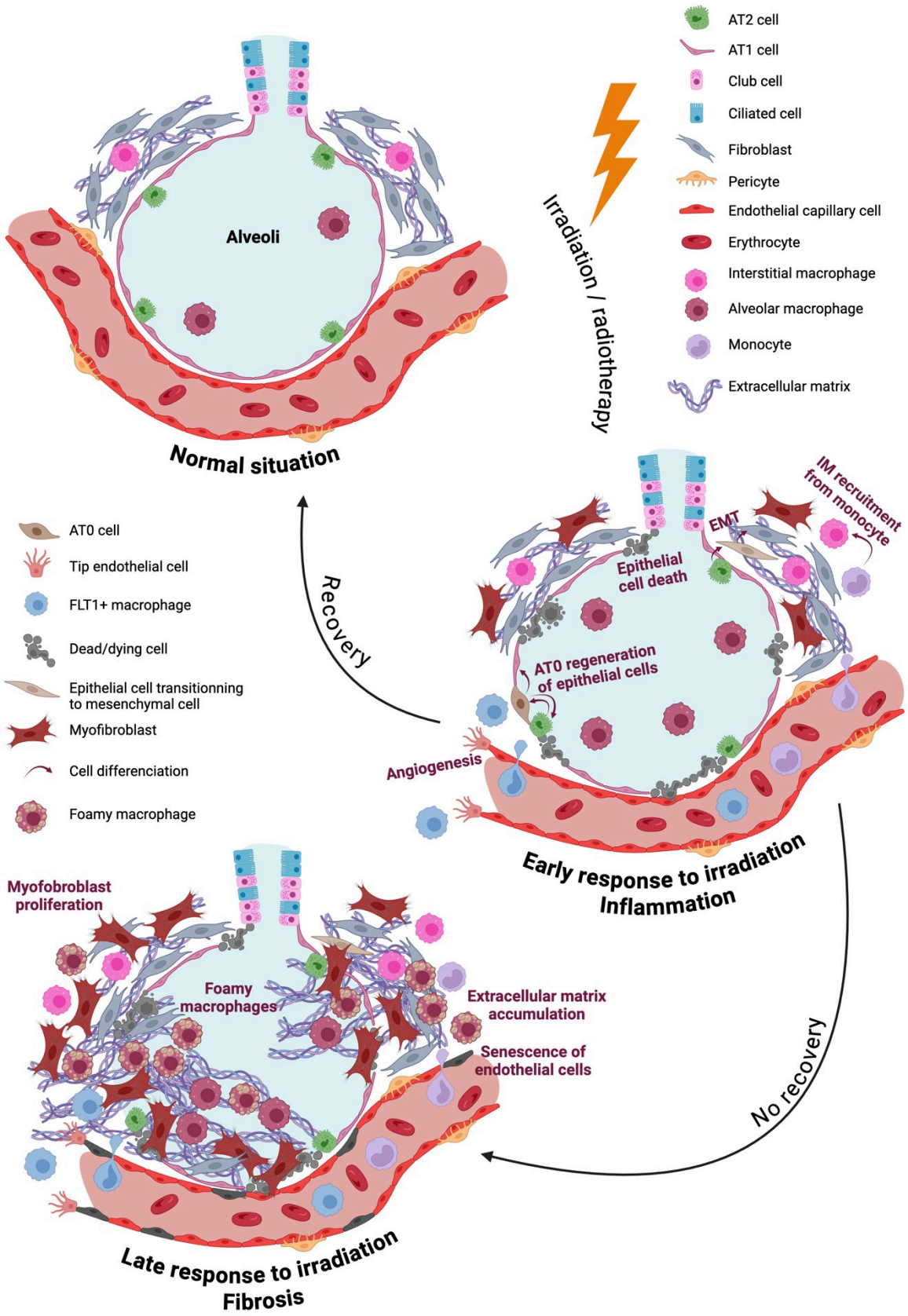


FIGURE 5.1. recapitulative figure of the thesis results

1. Impairment of the epithelial cell functions

Different processes are affecting the epithelial cells in reaction to irradiation. First, in the first hours after irradiation, a new population of epithelial cells appears: the AT0 cells. This population is characterized by the co-expression of the genes Sftpb, Scgb1a1 and Scgb3a2 and is thought to be derived from the AT2 cells through a "dedifferentiating" process (Kadur Lakshminarasimha Murthy et al. 2022). This AT0 population then acts as a progenitor population that can proliferate and differentiate to new AT1 and AT2 cells to restore the damage alveoli epithelium. During the weeks and months following irradiation, another process participates in the restoration of the population of AT1 cells: the transdifferentiation of AT2 cells to AT1 cells (Riemondy et al. 2019). It is the direct transformation of AT2 cells to AT1 cells, and we observe an increase of this process in the AT2 cells populations during the months leading to fibrosis after a 13Gy irradiation, but not a 10Gy irradiation. However, as after a 13Gy irradiation there is not a recovery of the lung and alveoli structure, the transdifferentiation might not allow an efficient reconstruction of the alveoli epithelium.

Finally, we observe processes of epithelial to mesenchymal transition (EMT) in the AT2 cells mainly after 13Gy irradiation and we identified two alveolar macrophages sub populations as the potential populations triggering this process through the TGFβ1-TGFβR2 interaction. EMT has been shown to be central in the pathogenesis of fibrosis development, and part of the myofibroblast population is demonstrated to be of epithelial origin (Willis, duBois, and Borok 2006). Finally, activated macrophages seem to promote these processes (H.-R. Park, Jo, and Jung 2019).

The epithelial population has been shown to be a major driver in the development of lung fibrosis in the context of idiopathic pulmonary fibrosis (Selman and Pardo 2020; 2006), through diverse processes. Notably, an important AT2 cell death has been observed, in reaction to endoplasmic reticulum stress or mitochondrial dysfunction. Furthermore, in the context of IPF, AT2 cells present impaired renewal and progenitor capacity (Parimon et al. 2020) and have been shown to express senescence markers, especially p21 and p16 (Zhou et al. 2022). However, in our data the AT2 cells only show a slight upregulation of the expression of p21, but not p16. The observations made in the other studies were mainly done in the context of idiopathic pulmonary fibrosis (Yao et al. 2021), and one study using X-ray mouse

thoracic irradiation (Citrin et al. 2013). The epithelial cell senescence induced by electron irradiation could present with a different phenotype than the senescence observed in IPF patients or in X-ray irradiated mice.

All these mechanisms participate in the loss of structure of the alveoli and a loss of function of the lungs in the terminal stages of fibrosis (Z. Chen, Wu, and Ning 2019).

2. Crucial role of the different subtypes of lung macrophages

The macrophages also play an important role in events leading to RIPF. Early after irradiation, the damaged interstitial macrophage compartment is replenished through differentiation of circulating macrophages, mainly into nerve interstitial macrophages. The transcriptome profile of the alveolar macrophages is as well affected by irradiation: different sub populations appear at different time points post irradiation and develop a “foamy” phenotype, characterized by the expression of genes involved in the lipidic metabolism and implicated in various diseases (Kong et al. 2020; Russell et al. 2009; Yu et al. 2013). We found this phenotype after both a fibrogenic and non fibrogenic dose of irradiation, but it is increased during the timepoints where we observe the development of fibrosis. Indeed, it has been shown that these foamy macrophages present a possible pro-fibrotic role in the context of atherosclerosis (Thomas et al. 2015). Furthermore, in the context of tuberculosis, it has been shown that the accumulation of lipid droplet and the increase of cholesterol content in alveolar macrophages can lead to apoptosis and necrosis of this population, releasing toxins in the tissues and participating in a chronic pro-inflammatory feedback loop (Russell et al. 2009; S. N. et al. 2019). However, other evidence shows a possible protective role of foamy macrophages in smocking-induced emphysema (Hirama et al. 2007), underlining the already demonstrated ambivalent role of macrophages in the development of pulmonary fibrosis (Yang et al. 2023).

Recent studies have also highlighted the role of senescent macrophages in RIPF. Irradiation has been shown to induce cellular senescence in macrophages, leading to an increased expression of senescence-associated secretory phenotype genes (Su et al. 2021). These factors contribute to a pro-fibrotic environment by promoting inflammation and ECM remodelling, thereby facilitating the progression of fibrosis. The presence of senescent

macrophages in the lung tissue post-radiation suggests that they may serve as a therapeutic target for preventing or treating RIPF.

In this work, we showed the more important change in transcriptome of the alveolar macrophages compared to the interstitial macrophages. However, a study showed that depletion of alveolar macrophages didn't allow any improvement in the development of radiation induced pulmonary fibrosis, whereas the depletion of interstitial macrophages in mice blocked the development of fibrosis (Meziani et al. 2018). More needs to be done to understand these apparently contradictory results.

3. Damages and changes of the endothelial cells in response to irradiation

Finally, endothelial cells present major alterations after irradiation and undergo extensive regeneration mechanisms, mainly through pro angiogenic signaling. These processes start shortly after irradiation through changes in the proportion of two endothelial capillaries cell states: the tip cells and stalk cells. These sub-populations are activated by the reception of VEGF signaling. Furthermore, the VEGFA secretion in the sites of vascular damages triggers the recruitment of a particular population of circulating FLT1 expressing alveolar macrophages that also triggers angiogenesis through a IL1B-IL1 receptor interaction. However, four to five months after irradiation the endothelial capillary cells express markers of senescence that could indicate replicative exhaustion and failure to regenerate a proper microvasculature and participating into the improper healing of the lung after a RILI, leading to RIPF. Overall, these findings support the demonstrated role of the vasculature in the development of radiation induced pulmonary fibrosis, with a loss of capillary and a permeabilization of the micro vasculature (Engelbrecht, Kooistra, and Knipe 2022; Caporarello and Ligresti 2023).

Furthermore, it has been shown that endothelial cells can undergo endothelial-to-mesenchymal transition, similarly as epithelial cells, where they lose their endothelial identity and acquire characteristics typical of mesenchymal cells. This transition has been shown to contribute significantly to the accumulation of fibroblasts and extracellular matrix in the lungs (W. Zhao et al. 2023). A study with a bleomycin-induced fibrosis model indicated that

approximately 16% of fibroblasts in fibrotic lung tissues originated from endothelial cells through endothelial-to-mesenchymal transition (Hashimoto et al. 2010). These processes need to be further investigated and could be potential therapeutic targets, as the maintaining of the endothelial barrier and network is crucial for lung function.

4. Spatial context of the changes induced by irradiation

The lung is a complex tissue composed of different cell populations interacting with each other. Therefore, it is important to put back the finding of this study in the context of the tissue. We already described several intercellular interactions between different cell populations of the lung and their possible consequences. The interactions between alveolar macrophages and epithelial cells have been shown to trigger epithelial to mesenchymal transition (Nagaraja and Nagarajan 2018). We also identified an interaction sent by the FLT1 positive alveolar macrophages and received by the gCap tip cells that participates in the induction of pro-angiogenic signaling.

Another couple of populations that are of great relevance for the function of the lungs are the aerocytes and AT1 cells. Indeed, in normal conditions, these two populations are in close contact with each other to perform the function of the lungs: the gas exchanges between the blood and the outside air (L. Song et al. 2024). After irradiation, we observe a disorganization and a collapse of the structure of the alveoli (Knudsen, Ruppert, and Ochs 2017). Therefore, we can make the hypothesis that there might be a loss of contact between the two populations. At the late stages after irradiation, during fibrosis, we observe evidence of senescence in both aerocytes and AT1 cells. This senescence could be a consequence of the disruption of the lung tissue, the inflammatory context, or possibly replicative senescence. However, the loss of contact between the aerocytes and the AT1 cells could also partly be a consequence of the senescence of the populations: it has been shown that the presence of senescent cells can disrupt tight junctions and disorganize monolayers of cells (Krouwer et al. 2012). Like many other processes occurring during the development of radiation-induced pulmonary fibrosis, it is a difficult task to determine if a particular event is a cause or a consequence. In order to identify potential treatments, it is important to elucidate the sequential activation of the distinct pathways and how they are spatially connected to each other.

CHAPTER VI
Perspectives

CHAPTER VI – Perspectives

All the results presented in this study are based on the analysis of single cell RNA sequencing data. While this type of analysis is a great way to obtain a better understanding of the complexity of a tissue and its response to a stimulus using a low number of samples, the results obtained need to be validated using other methods: *in vivo* or *in vitro* experiments, immunostaining, co-culture experiments, spatial transcriptomics (Curras-Alonso et al. 2021), FACS. Therefore, we plan to perform these experiments in order to verify the previous conclusions made during this thesis.

1. Study of the spatial context of the changes induced by irradiation

First, we will use the spatial transcriptomics technique Xenium from 10X Genomics to analyze the spatial changes induced by irradiation to the human lungs from Pancoast patients we already analyzed with single cell RNA sequencing.

Furthermore, we are in collaboration with a research group in Institut Pasteur to develop an in-house sequential single molecular FISH method called autoFISH (**FIGURE 5.2**) that we are planning to use to map the human and mouse lungs and the structural changes triggered by radiotherapy or irradiation.

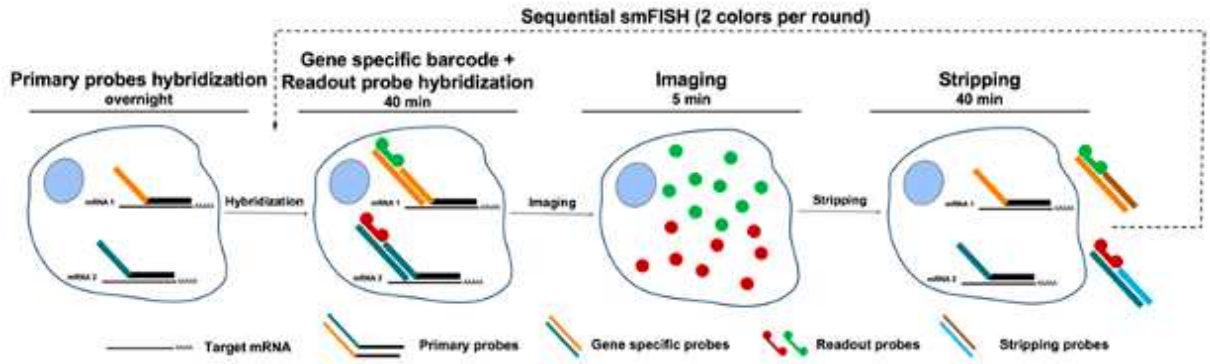


FIGURE 6.1. principle of the autoFISH method

The first step of autoFISH is the overnight hybridization of the primary probes for all the genes of interest. The probes are composed of a segment specific to the mRNA of the

gene and binding to it, and of a gene-specific segment. This first step is followed by several rounds of hybridization, imaging and stripping allowing the analysis of two targets per round. During each round, the secondary gene specific probe partly hybridizes with the gene-specific segment and partly with a readout fluorescent probe. The tissue is then images, and stripping probes are used to remove the secondary probes before the next round. This innovative method will allow us to image the presence of mRNAs of the different cell population markers, as well as cell states markers. Furthermore, we will be able to validate the predicted intercellular communications by analyzing the position of the partners of the communication as well as the expression of the ligands and receptors. The computational analysis of the results of the autoFISH is challenging, especially the cell segmentation. Indeed, the cells of the lungs present various shapes that can be very difficult to delimitate. Therefore, as part of our collaboration with Institut Pasteur, a new computational method called ComSeg was developed for a more efficient segmentation (Defard et al. 2024, **ANNEXE VII**).

2. Lineage tracing study of the endothelial cells

Endothelial cells appear to have a decisive role in the regeneration of the lung and the development of pulmonary fibrosis. In order to precise the fate of the progenitor gCap cells and study the origin of the senescent aCap cells, we are developing mouse models of aCap (Apln Cre) and gCap (Aplnr Cre). This will allow us to perform lineage tracing analysis and obtain an unprecedented understanding of the role of the different populations of endothelial cells in the development of RIPF.

3. Senescence and the use of senolytiques

As we saw in this study, regenerative processes and its dysfunction play a major role in the failure to heal properly after a RILI, leading to pulmonary fibrosis. The single cell RNA sequencing data from both the epithelial cells and the endothelial cells indicated the possible presence of senescence at the stage of pulmonary fibrosis. These two populations play a major role in the proper functioning of the lungs, and senescence could have a great impact on their functioning. Senescence is characterized by a cell cycle arrest and metabolic changes including the secretion in the cell environment of pro-inflammatory cytokines and other molecules, called senescence-associated secretory phenotype (SASP) (W. Huang et al. 2022).

Senescence can be a defense mechanism against tumor cells, to prevent their proliferation and accumulation. Senescent cells have also been shown to play a role in physiological condition, and elimination of p16 high cells reduces the lifespan of mice (Grosse et al. 2020). However, in some pathological cases, senescence can be detrimental. Secretion of SASP by senescent cells can induce the neighboring cells to turn senescent as well. In this manner, senescence can spread across the tissue and impair the normal functioning of the organ.

Senescence can be cause by different processes. Telomeres are protective caps at the ends of chromosomes that shorten with each cell division. When telomeres become critically short, it triggers a DNA damage response and senescence. Various forms of DNA damage, such as double-strand breaks, can activate the DNA damage response pathways and induce senescence as a protective mechanism against propagating damaged cells. Excessive production of ROS or deficiencies in antioxidant defences can cause oxidative damage to cellular components, including DNA, leading to senescence. Changes in chromatin structure and epigenetic modifications, such as DNA methylation and histone modifications, can contribute to the induction of senescence through the p16-RB pathway. Impaired mitochondrial function and increased mitochondrial oxidative stress can trigger senescence through various mechanisms, including DNA damage and metabolic dysregulation. Chronic inflammation and the associated secretion of pro-inflammatory cytokines can promote cellular senescence, particularly in the context of aging and age-related diseases.

These causes often interact and reinforce each other, creating a complex network of signaling pathways that ultimately lead to a global senescence of the tissue. This seems to be the case of RIPF, with diverse mechanisms leading to senescence and failure to heal the RILI. Therefore, one of the possibilities to prevent RIPF is the use of senolytics, a class of drugs that selectively eliminate senescent cells (Kirkland and Tchkonia 2020). This is currently under investigation in the context of IPF and various studies present positive results. A first human open label pilot study involving fourteen patients suffering from IPF treated with a combination of two senolytic drugs (dasatinib and quercetin) showed an alleviation of the respiratory symptoms (Justice et al. 2019). Other studies have shown the potential of senolytics in the context of IPF (Lee et al. 2024; de Godoy, Macedo, and Gambero 2024; Lehmann et al. 2017). A few studies using a mouse model of ionizing thoracic radiation also show some alleviation of pulmonary symptoms (Meng et al. 2021; Nguyen et al. 2018; Zhou

et al. 2022) and even a reversion of the fibrosis in one study (Pan et al. 2017). However, the effect of senolytics on the development of RIPF is still poorly understood and was never studied using scRNAseq. Therefore, we plan to test the administration of senolytics to mice at various time points before, after or at the moment of irradiation, in order to determine the most efficient moment for the administration of the drugs, as well as perform scRNAseq experiment on irradiated mice treated with senolytics in order to obtain a better understanding of the mechanisms of fibrosis alleviation after administration of senolytics.

4. The use of FLASH radiotherapy to spare the healthy tissues

Another possibility for the reduction of radiotherapy side effects is the use of innovative radiotherapy modality. The FLASH radiotherapy method is one of them, and was developed by Vincent Favaudon in my research group (V. Favaudon, Fouillade, and Vozenin 2015; Fouillade et al. 2020). FLASH radiotherapy delivers radiation at ultra-high dose rates, typically greater than 40 Gy/s, to treat tumors, with the entire radiation dose delivered in less than 200 milliseconds. FLASH radiotherapy treatment shows reduced radiation-induced toxicity in healthy tissues with the same tumoral control as conventional radiotherapy. Preclinical studies have shown that FLASH radiotherapy can control tumors while minimizing normal tissue toxicity compared to CONV radiotherapy. This effect has been shown in different organisms (mouse, cat, dog, pig, zebrafish) and different organs (lung, skin, brain, gut) (Limoli and Vozenin 2023) and the first results on human patients are very promising (Bourhis et al. 2019). Currently, the mechanisms of radioprotection of FLASH are under investigation in my research group, using a similar approach as the one presented in this manuscript with single cell RNA sequencing data analysis of mice irradiated with the FLASH radiotherapy method, at several time points post irradiation, and with the study of irradiated lung slices from human and mice (Dubail et al. 2023).

5. Tutorials for single cell RNA sequencing data analysis

This project allowed us to gain a better understanding of the complex mechanisms of lung response to radiation injury and development of radiation induced pulmonary fibrosis, thanks to the use of single cell RNA sequencing data analysis. This method is very effective to decipher the events occurring in complex tissues composed of numerous cell types and

states. However, it can be difficult to start the analysis of this very complex data. Therefore, I wrote tutorials for the analysis of single cell RNA sequencing data for beginners that can be found in **ANNEXE V**. They should allow more people to exploit the full potential of single cell RNA sequencing.

6. Conclusion

Overall, we are confident that this work has significantly advanced our understanding of radiation-induced lung injuries, laying a strong foundation for further investigation into many unresolved questions. The application of single-cell RNA sequencing has provided a powerful tool for elucidating the diverse normal tissue cellular responses and intercellular interactions initiated by radiation-induced lung injury. Furthermore, our findings highlight potential new therapeutic targets and pave the way for innovative drug use to combat this deadly radiotherapy side effect.

Bibliography

BIBLIOGRAPHY

Aban, C. E., A. Lombardi, G. Neiman, M. C. Biani, A. La Greca, A. Waisman, L. N. Moro, G. Sevlever, S. Miriuka, and C. Luzzani. 2021. 'Downregulation of E-Cadherin in Pluripotent Stem Cells Triggers Partial EMT'. *Scientific Reports* 11 (1): 2048. https://doi.org/10.1038/s41598-021-81735-1.

Abhinand, Chandran S., Rajesh Raju, Sasikumar J. Soumya, Prabha S. Arya, and Perumana R. Sudhakaran. 2016. 'VEGF-A/VEGFR2 Signaling Network in Endothelial Cells Relevant to Angiogenesis'. *Journal of Cell Communication and Signaling* 10 (4): 347–54. https://doi.org/10.1007/s12079-016-0352-8.

Adams, Taylor S., Jonas C. Schupp, Sergio Poli, Ehab A. Ayaub, Nir Neumark, Farida Ahangari, Sarah G. Chu, et al. 2020. 'Single-Cell RNA-Seq Reveals Ectopic and Aberrant Lung-Resident Cell Populations in Idiopathic Pulmonary Fibrosis'. *Science Advances* 6 (28): eaba1983. https://doi.org/10.1126/sciadv.aba1983.

Adamson, I. Y., and D. H. Bowden. 1983. 'Endothelial Injury and Repair in Radiation-Induced Pulmonary Fibrosis.' *The American Journal of Pathology* 112 (2): 224–30.

Adamson, I. Y. R. 1985. 'Cellular Kinetics of the Lung'. In *Toxicology of Inhaled Materials: General Principles of Inhalation Toxicology*, edited by Hanspeter Witschi and Joseph D. Brain, 289–317. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-69521-6_11.

Ahmad, Ijaz, and Mansour A. Balkhyour. 2020. 'Occupational Exposure and Respiratory Health of Workers at Small Scale Industries'. *Saudi Journal of Biological Sciences* 27 (3): 985–90. https://doi.org/10.1016/j.sjbs.2020.01.019.

Aibar, Sara, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert Hulselmans, Florian Rambow, et al. 2017. 'SCENIC: Single-Cell Regulatory Network Inference and Clustering'. *Nature Methods* 14 (11): 1083–86. https://doi.org/10.1038/nmeth.4463.

Albertine, Kurt H. 2016. '1 - Anatomy of the Lungs'. In *Murray and Nadel's Textbook of Respiratory Medicine (Sixth Edition)*, edited by V. Courtney Broaddus, Robert J. Mason, Joel D. Ernst, Talmadge E. King, Stephen C. Lazarus, John F. Murray, Jay A. Nadel, Arthur S. Slutsky, and Michael B. Gotway, 3-21.e5. Philadelphia: W.B. Saunders. https://doi.org/10.1016/B978-1-4557-3383-5.00001-4.

AlJanahi, Aisha A., Mark Danielsen, and Cynthia E. Dunbar. 2018. 'An Introduction to the Analysis of Single-Cell RNA-Sequencing Data'. *Molecular Therapy Methods & Clinical Development* 10 (September):189–96. https://doi.org/10.1016/j.omtm.2018.07.003.

Almeida, Celine, Devipriya Nagarajan, Jian Tian, Sofia Walder Leal, Kenneth Wheeler, Michael Munley, William Blackstock, and Weiling Zhao. 2013. 'The Role of Alveolar Epithelium in Radiation-Induced Lung Injury'. *PLOS ONE* 8 (1): e53628. https://doi.org/10.1371/journal.pone.0053628.

Alonso, Sandra Currás. 2021. 'Lung Responses to Radiation Injury at the Single Cell Level'. Phdthesis, Université Paris sciences et lettres. https://pastel.hal.science/tel-03917500.

Alysandratos, Konstantinos-Dionysios, Michael J. Herriges, and Darrell N. Kotton. 2021. 'Epithelial Stem and Progenitor Cells in Lung Repair and Regeneration'. *Annual Review of Physiology* 83 (1): 529–50. https://doi.org/10.1146/annurev-physiol-041520-092904.

Anbazhagan, Kolandaswamy, Isabelle Duroux-Richard, Christian Jorgensen, and Florence Apparailly. 2014. 'Transcriptomic Network Support Distinct Roles of Classical and Non-Classical Monocytes in Human'. *International Reviews of Immunology* 33 (6): 470–89. https://doi.org/10.3109/08830185.2014.902453.

Anders, Simon, and Wolfgang Huber. 2010. 'Differential Expression Analysis for Sequence Count Data'. *Genome Biology* 11 (10): R106. https://doi.org/10.1186/gb-2010-11-10-r106.

Andrade Carvalho, Heloisa de, and Rosangela Correa Villar. 2018. 'Radiotherapy and Immune Response: The Systemic Effects of a Local Treatment'. *Clinics* 73 (January). https://doi.org/10.6061/clinics/2018/e557s.

Ang, Yvonne L. E., Hon-Lyn Tan, and Ross A. Soo. 2015. 'Best Practice in the Treatment of Advanced Squamous Cell Lung Cancer'. *Therapeutic Advances in Respiratory Disease* 9 (5): 224–35. https://doi.org/10.1177/1753465815581147.

Antoniou, Katerina M., George A. Margaritopoulos, Sara Tomassetti, Francesco Bonella, Ulrich Costabel, and Venerino Poletti. 2014. 'Interstitial Lung Disease'. *European Respiratory Review: An Official Journal of the European Respiratory Society* 23 (131): 40–54. https://doi.org/10.1183/09059180.00009113.

Aran, Dvir, Agnieszka P. Looney, Leqian Liu, Esther Wu, Valerie Fong, Austin Hsu, Suzanna Chak, et al. 2019. 'Reference-Based Analysis of Lung Single-Cell Sequencing Reveals a Transitional Profibrotic Macrophage'. *Nature Immunology* 20 (2): 163–72. https://doi.org/10.1038/s41590-018-0276-y.

Argelaguet, Ricard, Anna S. E. Cuomo, Oliver Stegle, and John C. Marioni. 2021. 'Computational Principles and Challenges in Single-Cell Data Integration'. *Nature Biotechnology* 39 (10): 1202–15. https://doi.org/10.1038/s41587-021-00895-7.

Arroyo-Hernández, Marisol, Federico Maldonado, Francisco Lozano-Ruiz, Wendy Muñoz-Montaño, Mónica Nuñez-Baez, and Oscar Arrieta. 2021. 'Radiation-Induced Lung Injury: Current Evidence'. *BMC Pulmonary Medicine* 21 (1): 9. https://doi.org/10.1186/s12890-020-01376-4.

Artal Cortés, Ángel, Lourdes Calera Urquizu, and Jorge Hernando Cubero. 2015. 'Adjuvant Chemotherapy in Non-Small Cell Lung Cancer: State-of-the-Art'. *Translational Lung Cancer Research* 4 (2): 191–97. https://doi.org/10.3978/j.issn.2218-6751.2014.06.01.

Asti, C., V. Ruggieri, S. Porzio, R. Chiusaroli, G. Melillo, and G. F. Caselli. 2000. 'Lipopolysaccharide-Induced Lung Injury in Mice. I. Concomitant Evaluation of Inflammatory Cells and Haemorrhagic Lung Damage'. *Pulmonary Pharmacology & Therapeutics* 13 (2): 61–69. https://doi.org/10.1006/pupt.2000.0231.

Bae, Geon Ho, Ye Seon Kim, Ji Ye Park, Mingyu Lee, Sung Kyun Lee, Ji Cheol Kim, Jang Gyu Kim, et al. 2022. 'Unique Characteristics of Lung-Resident Neutrophils Are Maintained by

PGE2/PKA/Tgm2-Mediated Signaling'. *Blood* 140 (8): 889–99. https://doi.org/10.1182/blood.2021014283.

Bagnato, Gianluca, and Sergio Harari. 2015. 'Cellular Interactions in the Pathogenesis of Interstitial Lung Diseases'. *European Respiratory Review* 24 (135): 102–14. https://doi.org/10.1183/09059180.00003214.

Barnes, Peter J. 2016. 'Inflammatory Mechanisms in Patients with Chronic Obstructive Pulmonary Disease'. *Journal of Allergy and Clinical Immunology* 138 (1): 16–27. https://doi.org/10.1016/j.jaci.2016.05.011.

Barnes, Peter J. 2017. 'Cellular and Molecular Mechanisms of Asthma and COPD'. *Clinical Science (London, England* 131 (13): 1541–58. https://doi.org/10.1042/cs20160487.

Batra, Hitesh, and Veena B. Antony. 2015. 'Pleural Mesothelial Cells in Pleural and Lung Diseases'. *Journal of Thoracic Disease* 7 (6). https://doi.org/10.3978/j.issn.2072-1439.2015.02.19.

Batra, Sanjay, Gayathriy Balamayooran, and Malaya K. Sahoo. 2011. 'Nuclear Factor-κB: A Key Regulator in Health and Disease of Lungs'. *Archivum Immunologiae et Therapiae Experimentalis* 59 (5): 335. https://doi.org/10.1007/s00005-011-0136-z.

Becht, Etienne, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel W. H. Kwok, Lai Guan Ng, Florent Ginhoux, and Evan W. Newell. 2019. 'Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP'. *Nature Biotechnology* 37 (1): 38–44. https://doi.org/10.1038/nbt.4314.

Bergen, Volker, Marius Lange, Stefan Peidli, F. Alexander Wolf, and Fabian J. Theis. 2020. 'Generalizing RNA Velocity to Transient Cell States through Dynamical Modeling'. *Nature Biotechnology* 38 (12): 1408–14. https://doi.org/10.1038/s41587-020-0591-3.

Bernhardt, Erica B., and Shadia I. Jalal. 2016. 'Small Cell Lung Cancer'. In *Lung Cancer: Treatment and Research*, edited by Karen L. Reckamp, 301–22. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-40389-2_14.

Blighe, Kevin. (2018) 2024. ‘Kevinblighe/EnhancedVolcano’. R. https://github.com/kevinblighe/EnhancedVolcano.

Boopathi, Ettickan, Robert B. Den, and Chellappagounder Thangavel. 2023. ‘Innate Immune System in the Context of Radiation Therapy for Cancer’. *Cancers* 15 (15): 3972. https://doi.org/10.3390/cancers15153972.

Bourdin, A., P.-R. Burgel, P. Chanez, G. Garcia, T. Perez, and N. Roche. 2009. ‘Recent Advances in COPD: Pathophysiology, Respiratory Physiology and Clinical Aspects, Including Comorbidities’. *European Respiratory Review* 18 (114): 198–212. https://doi.org/10.1183/09059180.00005509.

Bourhis, Jean, Wendy Jeanneret Sozzi, Patrik Gonçalves Jorge, Olivier Gaide, Claude Bailat, Fréderic Duclos, David Patin, et al. 2019. ‘Treatment of a First Patient with FLASH-Radiotherapy’. *Radiotherapy and Oncology*, FLASH radiotherapy International Workshop, 139 (October):18–22. https://doi.org/10.1016/j.radonc.2019.06.019.

Boussat, Sandrine, Saadia Eddahibi, André Coste, Virginie Fataccioli, Mallaury Gouge, Bruno Housset, Serge Adnot, and Bernard Maitre. 2000. ‘Expression and Regulation of Vascular Endothelial Growth Factor in Human Pulmonary Epithelial Cells’. *American Journal of Physiology-Lung Cellular and Molecular Physiology* 279 (2): L371–78. https://doi.org/10.1152/ajplung.2000.279.2.L371.

Büttner, Maren, Zhichao Miao, F. Alexander Wolf, Sarah A. Teichmann, and Fabian J. Theis. 2019. ‘A Test Metric for Assessing Single-Cell RNA-Seq Batch Correction’. *Nature Methods* 16 (1): 43–49. https://doi.org/10.1038/s41592-018-0254-1.

Cannoodt, Robrecht, Wouter Saelens, and Yvan Saeys. 2016. ‘Computational Methods for Trajectory Inference from Single-Cell Transcriptomics’. *European Journal of Immunology* 46 (11): 2496–2506. https://doi.org/10.1002/eji.201646347.

Cao, Junyue, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M. Ibrahim, Andrew J. Hill, Fan Zhang, et al. 2019. ‘The Single-Cell Transcriptional Landscape of Mammalian Organogenesis’. *Nature* 566 (7745): 496–502. https://doi.org/10.1038/s41586-019-0969-x.

Cao, Yinghao, Xiaoyue Wang, and Gongxin Peng. 2020. 'SCSA: A Cell Type Annotation Tool for Single-Cell RNA-Seq Data'. *Frontiers in Genetics* 11. https://www.frontiersin.org/articles/10.3389/fgene.2020.00490.

Caporarello, Nunzia, and Giovanni Ligresti. 2023. 'Vascular Contribution to Lung Repair and Fibrosis'. *American Journal of Respiratory Cell and Molecular Biology* 69 (2): 135–46. https://doi.org/10.1165/rcmb.2022-0431TR.

Castranova, V., J. Rabovsky, J. H. Tucker, and P. R. Miles. 1988. 'The Alveolar Type II Epithelial Cell: A Multifunctional Pneumocyte'. *Toxicology and Applied Pharmacology* 93 (3): 472–83. https://doi.org/10.1016/0041-008X(88)90051-8.

Chargari, Cyrus, Eric Deutsch, Pierre Blanchard, Sebastien Gouy, Hélène Martelli, Florent Guérin, Isabelle Dumas, et al. 2019. 'Brachytherapy: An Overview for Clinicians'. *CA: A Cancer Journal for Clinicians* 69 (5): 386–401. https://doi.org/10.3322/caac.21578.

Chen, Felicia, and Alan Fine. 2016. 'Stem Cells in Lung Injury and Repair'. *The American Journal of Pathology* 186 (10): 2544–50. https://doi.org/10.1016/j.ajpath.2016.05.023.

Chen, Jing, Xiaoli Tian, Zijie Mei, Yacheng Wang, Ye Yao, Shimin Zhang, Xin Li, Hui Wang, Junhong Zhang, and Conghua Xie. 2016. 'The Effect of the TLR9 Ligand CpG-Oligodeoxynucleotide on the Protective Immune Response to Radiation-Induced Lung Fibrosis in Mice'. *Molecular Immunology* 80 (December):33–40. https://doi.org/10.1016/j.molimm.2016.11.001.

Chen, Li, Liqun Hu, Xiang Zhu, Yan Wang, Qing Li, Jian Ma, and Hongqi Li. 2020. 'MALAT1 Overexpression Attenuates AS by Inhibiting Ox-LDL-Stimulated Dendritic Cell Maturation via miR-155-5p/NFIA Axis'. *Cell Cycle* 19 (19): 2472–85. https://doi.org/10.1080/15384101.2020.1807094.

Chen, Wenqi, Peng Xia, Heping Wang, Jihao Tu, Xinyue Liang, Xiaoling Zhang, and Lisha Li. 2019. 'The Endothelial Tip-Stalk Cell Selection and Shuffling during Angiogenesis'. *Journal of Cell Communication and Signaling* 13 (3): 291–301. https://doi.org/10.1007/s12079-019-00511-z.

Chen, Zhongjie, Zhiqiang Wu, and Wen Ning. 2019. 'Advances in Molecular Mechanisms and Treatment of Radiation-Induced Pulmonary Fibrosis'. *Translational Oncology* 12 (1): 162–69. https://doi.org/10.1016/j.tranon.2018.09.009.

Chetta, Alfredo, Emilio Marangio, and Dario Olivieri. 2004. 'Pulmonary Function Testing in Interstitial Lung Diseases'. *Respiration* 71 (3): 209–13. https://doi.org/10.1159/000077416.

Choi, Noah, Michael Baumann, M Flentjie, Pirkko Kellokumpu-Lehtinen, Suresh Senan, Nikolaos Zamboglou, and Paris Kosmidis. 2001. 'Predictive Factors in Radiotherapy for Non-Small Cell Lung Cancer: Present Status'. *Lung Cancer* 31 (1): 43–56. https://doi.org/10.1016/S0169-5002(00)00156-2.

Choi, Yo Won, Reginald F. Munden, Jeremy J. Erasmus, Kyung Joo Park, Woo Kyung Chung, Seok Chol Jeon, and Choong-Ki Park. 2004. 'Effects of Radiation Therapy on the Lung: Radiologic Appearances and Differential Diagnosis'. *RadioGraphics* 24 (4): 985–97. https://doi.org/10.1148/rg.244035160.

Chung, Eun Joo, Seokjoo Kwon, Jessica L. Reedy, Ayla O. White, Joon Seon Song, Ilseon Hwang, Joon Yong Chung, Kris Ylaya, Stephen M. Hewitt, and Deborah E. Citrin. 2021. 'IGF-1 Receptor Signaling Regulates Type II Pneumocyte Senescence and Resulting Macrophage Polarization in Lung Fibrosis'. *International Journal of Radiation Oncology*Biology*Physics* 110 (2): 526–38. https://doi.org/10.1016/j.ijrobp.2020.12.035.

Citrin, Deborah E., Uma Shankavaram, Jason A. Horton, William Shield III, Shuping Zhao, Hiroaki Asano, Ayla White, Anastasia Sowers, Angela Thetford, and Eun Joo Chung. 2013. 'Role of Type II Pneumocyte Senescence in Radiation-Induced Lung Fibrosis'. *JNCI: Journal of the National Cancer Institute* 105 (19): 1474–84. https://doi.org/10.1093/jnci/djt212.

Clark, Shane B., and Saud Alsubait. 2024. 'Non–Small Cell Lung Cancer'. In *StatPearls*. Treasure Island (FL): StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK562307/.

Cook, Peter C., and Andrew S. MacDonald. 2016. 'Dendritic Cells in Lung Immunopathology'. *Seminars in Immunopathology* 38 (4): 449–60. https://doi.org/10.1007/s00281-016-0571-3.

Cosio, Manuel G., Joaquim Majo, and Monica G. Cosio. 2002. 'Inflammation of the Airways and Lung Parenchyma in COPD: Role of T Cells'. *Chest* 121 (5, Supplement): 160S-165S. https://doi.org/10.1378/chest.121.5_suppl.160S.

Curras-Alonso, Sandra, Juliette Soulier, Thomas Defard, Christian Weber, Sophie Heinrich, Hugo Laporte, Sophie Leboucher, et al. 2023. 'An Interactive Murine Single-Cell Atlas of the Lung Responses to Radiation Injury'. *Nature Communications* 14 (1): 2445. https://doi.org/10.1038/s41467-023-38134-z.

Curras-Alonso, Sandra, Juliette Soulier, Thomas Walter, Florian Mueller, Arturo Londoño-Vallejo, and Charles Fouillade. 2021. 'Spatial Transcriptomics for Respiratory Research and Medicine'. *European Respiratory Journal* 58 (1). https://doi.org/10.1183/13993003.04314-2020.

Dar, Tahir B., Regina M. Henson, and Stephen L. Shiao. 2019. 'Targeting Innate Immunity to Enhance the Efficacy of Radiation Therapy'. *Frontiers in Immunology* 9 (January). https://doi.org/10.3389/fimmu.2018.03077.

Defard, Thomas, Hugo Laporte, Mallick Ayan, Juliette Soulier, Sandra Curras-Alonso, Christian Weber, Florian Massip, et al. 2024. 'A Point Cloud Segmentation Framework for Image-Based Spatial Transcriptomics'. *Communications Biology* 7 (1): 1–13. https://doi.org/10.1038/s42003-024-06480-3.

Della Latta, Veronica, A. Cecchettini, S. Del Ry, and M. A. Morales. 2015. 'Bleomycin in the Setting of Lung Fibrosis Induction: From Biological Mechanisms to Counteractions'. *Pharmacological Research* 97 (July):122–30. https://doi.org/10.1016/j.phrs.2015.04.012.

Demedts, M., A. U. Wells, J. M. Antó, U. Costabel, R. Hubbard, P. Cullinan, H. Slabbynck, et al. 2001. 'Interstitial Lung Diseases: An Epidemiological Overview'. *European Respiratory Journal* 18 (32 suppl): 2S-16S. https://doi.org/10.1183/09031936.01.18s320002.

Devine, Megan Stuebner, and Christine Kim Garcia. 2012. 'Genetic Interstitial Lung Disease'. *Clinics in Chest Medicine* 33 (1): 95–110. https://doi.org/10.1016/j.ccm.2011.11.001.

Dimitrov, Daniel, Dénes Türei, Martin Garrido-Rodriguez, Paul L. Burmedi, James S. Nagai, Charlotte Boys, Ricardo O. Ramirez Flores, et al. 2022. 'Comparison of Methods and Resources for Cell-Cell Communication Inference from Single-Cell RNA-Seq Data'. *Nature Communications* 13 (1): 3224. https://doi.org/10.1038/s41467-022-30755-0.

Ding, Nian-Hua, Jian Jian Li, and Lun-Quan Sun. 2013. 'Molecular Mechanisms and Treatment of Radiation-Induced Lung Fibrosis'. *Current Drug Targets* 14 (11): 1347–56.

Dolivo, David M., Charlotte R. Reed, Kristine A. Gargiulo, Adrian E. Rodrigues, Robert D. Galiano, Thomas A. Mustoe, and Seok Jong Hong. 2023. 'Anti-Fibrotic Effects of Statin Drugs: A Review of Evidence and Mechanisms'. *Biochemical Pharmacology* 214 (August):115644. https://doi.org/10.1016/j.bcp.2023.115644.

Dubail, Maxime, Sophie Heinrich, Lucie Portier, Jessica Bastian, Lucia Giuliano, Lilia Aggar, Nathalie Berthault, et al. 2023. 'Lung Organotypic Slices Enable Rapid Quantification of Acute Radiotherapy Induced Toxicity'. *Cells* 12 (20): 2435. https://doi.org/10.3390/cells12202435.

Duncan, Steven R. 2023. 'A Randomized, Double-Blind, Placebo-Controlled Phase II Clinical Trial of GKT137831 in Patients With Idiopathic Pulmonary Fibrosis'. Clinical trial registration NCT03865927. clinicaltrials.gov. https://clinicaltrials.gov/study/NCT03865927.

El Agha, Elie, Alena Moiseenko, Vahid Kheirollahi, Stijn De Langhe, Slaven Crnkovic, Grazyna Kwapiszewska, Marten Szibor, et al. 2017. 'Two-Way Conversion between Lipogenic and Myogenic Fibroblastic Phenotypes Marks the Progression and Resolution of Lung Fibrosis'. *Cell Stem Cell* 20 (2): 261-273.e3. https://doi.org/10.1016/j.stem.2016.10.004.

Engelbrecht, Eric, Tristan Kooistra, and Rachel S. Knipe. 2022. 'The Vasculature in Pulmonary Fibrosis'. *Current Tissue Microenvironment Reports* 3 (4): 83–97. https://doi.org/10.1007/s43152-022-00040-9.

Erle, David J., and Dean Sheppard. 2014. 'The Cell Biology of Asthma'. *Journal of Cell Biology* 205 (5): 621–31. https://doi.org/10.1083/jcb.201401050.

Ettinger, David S., Wallace Akerley, Gerold Bepler, Matthew G. Blum, Andrew Chang, Richard T. Cheney, Lucian R. Chirieac, et al. 2010. 'Non–Small Cell Lung Cancer'. *Journal of the National Comprehensive Cancer Network* 8 (7): 740–801. https://doi.org/10.6004/jnccn.2010.0056.

Favaudon, V., C. Fouillade, and M. -C. Vozenin. 2015. 'Radiothérapie « Flash » à Très Haut Débit de Dose : Un Moyen d'augmenter l'indice Thérapeutique Par Minimisation Des Dommages Aux Tissus Sains ?' *Cancer/Radiothérapie*, 26e Congrès national de la Société française de radiothérapie oncologique SFRO), 19 (6): 526–31. https://doi.org/10.1016/j.canrad.2015.04.006.

Favaudon, Vincent, Laura Caplier, Virginie Monceau, Frédéric Pouzoulet, Mano Sayarath, Charles Fouillade, Marie-France Poupon, et al. 2014. 'Ultrahigh Dose-Rate FLASH Irradiation Increases the Differential Response between Normal and Tumor Tissue in Mice'. *Science Translational Medicine* 6 (245): 245ra93. https://doi.org/10.1126/scitranslmed.3008973.

Fernandes, Darren J, John F McConville, Alastair G Stewart, Vladimir Kalinichenko, and Julian Solway. 2004. 'Can We Differentiate Between Airway and Vascular Smooth Muscle?' *Clinical and Experimental Pharmacology and Physiology* 31 (11): 805–10. https://doi.org/10.1111/j.1440-1681.2004.04084.x.

Fouillade, Charles, Sandra Curras-Alonso, Lorena Giuranno, Eddy Quelennec, Sophie Heinrich, Sarah Bonnet-Boissinot, Arnaud Beddok, et al. 2020. 'FLASH Irradiation Spares Lung Progenitor Cells and Limits the Incidence of Radio-Induced Senescence'. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 26 (6): 1497–1506. https://doi.org/10.1158/1078-0432.CCR-19-1440.

Freitas-Rodríguez, Sandra, Alicia R. Folgueras, and Carlos López-Otín. 2017. 'The Role of Matrix Metalloproteinases in Aging: Tissue Remodeling and Beyond'. *Biochimica et*

Biophysica Acta (BBA) - Molecular Cell Research, Matrix Metalloproteinases, 1864 (11, Part A): 2015–25. https://doi.org/10.1016/j.bbamcr.2017.05.007.

Fujino, Naoya, Hiroshi Kubo, Takaya Suzuki, Chiharu Ota, Ahmed E. Hegab, Mei He, Satoshi Suzuki, et al. 2011. 'Isolation of Alveolar Epithelial Type II Progenitor Cells from Adult Human Lungs'. *Laboratory Investigation* 91 (3): 363–78. https://doi.org/10.1038/labinvest.2010.187.

Fuks, Zvi, Roger S. Persaud, Alan Alfieri, Maureen McLoughlin, Desiree Ehleiter, Jeffery L. Schwartz, Andrew P. Seddon, Carlos Cordon-Cardo, and Adriana Haimovitz-Friedman. 1994. 'Basic Fibroblast Growth Factor Protects Endothelial Cells against Radiation-Induced Programmed Cell Death in Vitro and in Vivo1'. *Cancer Research* 54 (10): 2582–90.

Garrison, Annelise T., Rebecca E. Bignold, Xinhui Wu, and Jill R. Johnson. 2023. 'Pericytes: The Lung-Forgotten Cell Type'. *Frontiers in Physiology* 14 (March):1150028. https://doi.org/10.3389/fphys.2023.1150028.

GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. 2018. 'Global, Regional, and National Incidence, Prevalence, and Years Lived with Disability for 354 Diseases and Injuries for 195 Countries and Territories, 1990-2017: A Systematic Analysis for the Global Burden of Disease Study 2017'. *Lancet (London, England)* 392 (10159): 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7.

Gibbings, Sophie L., Stacey M. Thomas, Shaikh M. Atif, Alexandra L. McCubbrey, A. Nicole Desch, Thomas Danhorn, Sonia M. Leach, et al. 2017. 'Three Unique Interstitial Macrophages in the Murine Lung at Steady State'. *American Journal of Respiratory Cell and Molecular Biology* 57 (1): 66–76. https://doi.org/10.1165/rcmb.2016-0361OC.

Gillich, Astrid, Fan Zhang, Colleen G. Farmer, Kyle J. Travaglini, Serena Y. Tan, Mingxia Gu, Bin Zhou, Jeffrey A. Feinstein, Mark A. Krasnow, and Ross J. Metzger. 2020. 'Capillary Cell-Type Specialization in the Alveolus'. *Nature* 586 (7831): 785–89. https://doi.org/10.1038/s41586-020-2822-7.

Giroux Leprieur, Etienne, Coraline Dumenil, Catherine Julie, Violaine Giraud, Jennifer Dumoulin, Sylvie Labrune, and Thierry Chinet. 2017. 'Immunotherapy Revolutionises Non-

Small-Cell Lung Cancer Therapy: Results, Perspectives and New Challenges'. *European Journal of Cancer* 78 (June):16–23. https://doi.org/10.1016/j.ejca.2016.12.041.

Giuranno, Lorena, Jonathan Ient, Dirk De Ruysscher, and Marc A. Vooijs. 2019. 'Radiation-Induced Lung Injury (RILI)'. *Frontiers in Oncology* 9. https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2019.00877.

Godoy, Maria Carolina Ximenes de, Juliana Alves Macedo, and Alessandra Gambero. 2024. 'Researching New Drug Combinations with Senolytic Activity Using Senescent Human Lung Fibroblasts MRC-5 Cell Line'. *Pharmaceuticals* 17 (1): 70. https://doi.org/10.3390/ph17010070.

Grosse, Laurent, Nicole Wagner, Alexander Emelyanov, Clement Molina, Sandra Lacas-Gervais, Kay-Dietrich Wagner, and Dmitry V. Bulavin. 2020. 'Defined p16High Senescent Cell Types Are Indispensable for Mouse Healthspan'. *Cell Metabolism* 32 (1): 87-99.e6. https://doi.org/10.1016/j.cmet.2020.05.002.

Gu, Q., and L. -Y. Lee. 2006. 'NEUROPHYSIOLOGY | Neural Control of Airway Smooth Muscle'. In *Encyclopedia of Respiratory Medicine*, edited by Geoffrey J. Laurent and Steven D. Shapiro, 138–45. Oxford: Academic Press. https://doi.org/10.1016/B0-12-370879-6/00253-2.

Gu, Zuguang. 2022. 'Complex Heatmap Visualization'. *iMeta* 1 (3): e43. https://doi.org/10.1002/imt2.43.

Gul, Aman, Fangyong Yang, Cong Xie, Wenjing Du, Nabijan Mohammadtursun, Bin Wang, Jingjing Le, and Jingcheng Dong. 2023. 'Pulmonary Fibrosis Model of Mice Induced by Different Administration Methods of Bleomycin'. *BMC Pulmonary Medicine* 23 (1): 91. https://doi.org/10.1186/s12890-023-02349-z.

Guler, Sabina A., and Tamera J. Corte. 2021. 'Interstitial Lung Disease in 2020: A History of Progress'. *Clinics in Chest Medicine* 42 (2): 229–39. https://doi.org/10.1016/j.ccm.2021.03.001.

Guo, Peng, Rui Li, Tie Hua Piao, Chun Lan Wang, Xiao Lu Wu, and Hong Yan Cai. 2022. 'Pathological Mechanism and Targeted Drugs of COPD'. *International Journal of Chronic Obstructive Pulmonary Disease* 17 (July):1565–75. https://doi.org/10.2147/COPD.S366126.

Habiel, David M., and Cory M. Hogaboam. 2017. 'Heterogeneity of Fibroblasts and Myofibroblasts in Pulmonary Fibrosis'. *Europe PMC* 5:101–10. https://doi.org/10.1007/s40139-017-0134-x.

Hafemeister, Christoph, and Rahul Satija. 2019. 'Normalization and Variance Stabilization of Single-Cell RNA-Seq Data Using Regularized Negative Binomial Regression'. *Genome Biology* 20 (1): 296. https://doi.org/10.1186/s13059-019-1874-1.

Hahn, Jonas, Christine Schauer, Christine Czegley, Lasse Kling, Lenka Petru, Benjamin Schmid, Daniela Weidner, et al. 2019. 'Aggregated Neutrophil Extracellular Traps Resolve Inflammation by Proteolysis of Cytokines and Chemokines and Protection from Antiproteases'. *The FASEB Journal* 33 (1): 1401–14. https://doi.org/10.1096/fj.201800752R.

Hanania, Alexander N., Walker Mainwaring, Yohannes T. Ghebre, Nicola A. Hanania, and Michelle Ludwig. 2019. 'Radiation-Induced Lung Injury'. *Chest* 156 (1): 150–62. https://doi.org/10.1016/j.chest.2019.03.033.

Hao, Pengyi, Kun You, Haozhe Feng, Xinnan Xu, Fan Zhang, Fuli Wu, Peng Zhang, and Wei Chen. 2020. 'Lung Adenocarcinoma Diagnosis in One Stage'. *Neurocomputing* 392 (June):245–52. https://doi.org/10.1016/j.neucom.2018.11.110.

Hao, Yuhan, Tim Stuart, Madeline H. Kowalski, Saket Choudhary, Paul Hoffman, Austin Hartman, Avi Srivastava, et al. 2024. 'Dictionary Learning for Integrative, Multimodal and Scalable Single-Cell Analysis'. *Nature Biotechnology* 42 (2): 293–304. https://doi.org/10.1038/s41587-023-01767-y.

Hare, E.S. 1838. 'Tumor Involving Certain Nerves'. *Lond Med Gaz*, no. 1, 16–18.

Hashimoto, Naozumi, Sem H. Phan, Kazuyoshi Imaizumi, Masaki Matsuo, Harunori Nakashima, Tsutomu Kawabe, Kaoru Shimokata, and Yoshinori Hasegawa. 2010. 'Endothelial–Mesenchymal Transition in Bleomycin-Induced Pulmonary Fibrosis'. *American Journal of*

Respiratory Cell and Molecular Biology 43 (2): 161–72. https://doi.org/10.1165/rcmb.2009-0031OC.

He, Yonghan, Dinesh Thummuri, Guangrong Zheng, Paul Okunieff, Deborah E. Citrin, Zeljko Vujaskovic, and Daohong Zhou. 2019. 'Cellular Senescence and Radiation-Induced Pulmonary Fibrosis'. *Translational Research*, Pathologic fibrosis: disease mechanisms and novel therapeutics, 209 (July):14–21. https://doi.org/10.1016/j.trsl.2019.03.006.

He, Zhen-feng, Nan-shan Zhong, and Wei-Jie Guan. 2022. 'Impact of Chronic Respiratory Diseases on the Outcomes of COVID-19'. *Archivos de Bronconeumología* 58 (1): 5–7. https://doi.org/10.1016/j.arbres.2021.05.009.

Hellström, Mats, Li-Kun Phng, and Holger Gerhardt. 2007. 'VEGF and Notch Signaling'. *Cell Adhesion & Migration* 1 (3): 133–36.

Hervier, Baptiste, Jules Russick, Isabelle Cremer, and Vincent Vieillard. 2019. 'NK Cells in the Human Lungs'. *Frontiers in Immunology* 10. https://www.frontiersin.org/articles/10.3389/fimmu.2019.01263.

Hill, R P. 2005. 'Radiation Effects on the Respiratory System'. *British Journal of Radiology* 78 (Supplement_27): 75–81. https://doi.org/10.1259/bjr/34124307.

Hinton, Geoffrey E, and Sam Roweis. 2002. 'Stochastic Neighbor Embedding'. In *Advances in Neural Information Processing Systems*. Vol. 15. MIT Press. https://proceedings.neurips.cc/paper_files/paper/2002/hash/6150ccc6069bea6b5716254057a194ef-Abstract.html.

Hirama, Noriyuki, Yoko Shibata, Kazuhisa Otake, Jun-ichi Machiya, Toshihiro Wada, Sumito Inoue, Shuichi Abe, Noriaki Takabatake, Makoto Sata, and Isao Kubota. 2007. 'Increased Surfactant Protein-D and Foamy Macrophages in Smoking-Induced Mouse Emphysema'. *Respirology* 12 (2): 191–201. https://doi.org/10.1111/j.1440-1843.2006.01009.x.

Hodge, Greg, and Sandra Hodge. 2019. 'Therapeutic Targeting Steroid Resistant Pro-Inflammatory NK and NKT-Like Cells in Chronic Inflammatory Lung Disease'. *International Journal of Molecular Sciences* 20 (6): 1511. https://doi.org/10.3390/ijms20061511.

Hou, Fei, Kun Xiao, Li Tang, and Lixin Xie. 2021. 'Diversity of Macrophages in Lung Homeostasis and Diseases'. *Frontiers in Immunology* 12. https://www.frontiersin.org/articles/10.3389/fimmu.2021.753940.

Hoy, Haley, Thuy Lynch, and Monica Beck. 2019. 'Surgical Treatment of Lung Cancer'. *Critical Care Nursing Clinics of North America*, Cardiothoracic Surgical Critical Care, 31 (3): 303–13. https://doi.org/10.1016/j.cnc.2019.05.002.

Hsueh, Yi-Jen, Pei-Ching Kuo, and Jan-Kan Chen. 2013. 'Transcriptional Regulators of the ΔNp63: Their Role in Limbal Epithelial Cell Proliferation'. *Journal of Cellular Physiology* 228 (3): 536–46. https://doi.org/10.1002/jcp.24160.

Hu, Xiaoyu, Yuan Tian, Shihui Qu, Yongguo Cao, Shumin Li, Wenlong Zhang, Zecai Zhang, Naisheng Zhang, and Yunhe Fu. 2017. 'Protective Effect of TM6 on LPS-Induced Acute Lung Injury in Mice'. *Scientific Reports* 7 (1): 572. https://doi.org/10.1038/s41598-017-00551-8.

Huang, Weijun, LaTonya J. Hickson, Alfonso Eirin, James L. Kirkland, and Lilach O. Lerman. 2022. 'Cellular Senescence: The Good, the Bad and the Unknown'. *Nature Reviews Nephrology* 18 (10): 611–27. https://doi.org/10.1038/s41581-022-00601-z.

Huang, Yijuan, Weiqiang Zhang, Fangrong Yu, and Fu Gao. 2017. 'The Cellular and Molecular Mechanism of Radiation-Induced Lung Injury'. *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research* 23 (July):3446–50. https://doi.org/10.12659/MSM.902353.

Hung, CF. 2020. 'Origin of Myofibroblasts in Lung Fibrosis'. *Current Tissue Microenvironment Reports* 1 (4): 155–62. https://doi.org/10.1007/s43152-020-00022-9.

Hurskainen, Maria, Ivana Mižíková, David P. Cook, Noora Andersson, Chanèle Cyr-Depauw, Flore Lesage, Emmi Helle, et al. 2021. 'Single Cell Transcriptomic Analysis of Murine

Lung Development on Hyperoxia-Induced Damage'. *Nature Communications* 12 (1): 1565. https://doi.org/10.1038/s41467-021-21865-2.

Hutchinson, Barry D., Girish S. Shroff, Mylene T. Truong, and Jane P. Ko. 2019. 'Spectrum of Lung Adenocarcinoma'. *Seminars in Ultrasound, CT and MRI*, Interstitial lung disease, 40 (3): 255–64. https://doi.org/10.1053/j.sult.2018.11.009.

Jakobsson, Lars, Claudio A. Franco, Katie Bentley, Russell T. Collins, Bas Ponsioen, Irene M. Aspalter, Ian Rosewell, et al. 2010. 'Endothelial Cells Dynamically Compete for the Tip Cell Position during Angiogenic Sprouting'. *Nature Cell Biology* 12 (10): 943–53. https://doi.org/10.1038/ncb2103.

James, Carissa C., Michael J. Zeitz, Patrick J. Calhoun, Samy Lamouille, and James W. Smyth. 2018. 'Altered Translation Initiation of Gja1 Limits Gap Junction Formation during Epithelial–Mesenchymal Transition'. *Molecular Biology of the Cell* 29 (7): 797–808. https://doi.org/10.1091/mbc.E17-06-0406.

Jarzebska, Natalia, Ekaterina S. Karetnikova, Alexander G. Markov, Michael Kasper, Roman N. Rodionov, and Peter M. Spieth. 2021. 'Scarred Lung. An Update on Radiation-Induced Pulmonary Fibrosis'. *Frontiers in Medicine* 7 (January). https://doi.org/10.3389/fmed.2020.585756.

Jassal, Bijay, Lisa Matthews, Guilherme Viteri, Chuqiao Gong, Pascual Lorente, Antonio Fabregat, Konstantinos Sidiropoulos, et al. 2020. 'The Reactome Pathway Knowledgebase'. *Nucleic Acids Research* 48 (D1): D498–503. https://doi.org/10.1093/nar/gkz1031.

Jha, Saurav Kumar, Gabriele De Rubis, Shankar Raj Devkota, Yali Zhang, Radhika Adhikari, Laxmi Akhileshwar Jha, Kunal Bhattacharya, et al. 2024. 'Cellular Senescence in Lung Cancer: Molecular Mechanisms and Therapeutic Interventions'. *Ageing Research Reviews* 97 (June):102315. https://doi.org/10.1016/j.arr.2024.102315.

Jin, Hee, Youngjo Yoo, Younghwa Kim, Yeijin Kim, Jaeho Cho, and Yun-Sil Lee. 2020. 'Radiation-Induced Lung Fibrosis: Preclinical Animal Models and Therapeutic Strategies'. *Cancers* 12 (6): 1561. https://doi.org/10.3390/cancers12061561.

Jin, Suoqin, Maksim V. Plikus, and Qing Nie. 2023. 'CellChat for Systematic Analysis of Cell-Cell Communication from Single-Cell and Spatially Resolved Transcriptomics'. Preprint. Bioinformatics. https://doi.org/10.1101/2023.11.05.565674.

Justice, Jamie N., Anoop M. Nambiar, Tamar Tchkonia, Nathan K. LeBrasseur, Rodolfo Pascual, Shahrukh K. Hashmi, Larissa Prata, et al. 2019. 'Senolytics in Idiopathic Pulmonary Fibrosis: Results from a First-in-Human, Open-Label, Pilot Study'. *eBioMedicine* 40 (February):554–63. https://doi.org/10.1016/j.ebiom.2018.12.052.

Kadur Lakshminarasimha Murthy, Preetish, Vishwaraj Sontake, Aleksandra Tata, Yoshihiko Kobayashi, Lauren Macadlo, Kenichi Okuda, Ansley S. Conchola, et al. 2022. 'Human Distal Lung Maps and Lineage Hierarchies Reveal a Bipotent Progenitor'. *Nature* 604 (7904): 111–19. https://doi.org/10.1038/s41586-022-04541-3.

Kavaliauskaite, Gabija, and Jesper Grud Skat Madsen. 2023. 'Automatic Quality Control of Single-Cell and Single-Nucleus RNA-Seq Using valiDrops'. *NAR Genomics and Bioinformatics* 5 (4): lqad101. https://doi.org/10.1093/nargab/lqad101.

Kim, Eric S. 2016. 'Chemotherapy Resistance in Lung Cancer'. In *Lung Cancer and Personalized Medicine: Current Knowledge and Therapies*, edited by Aamir Ahmad and Shirish Gadgeel, 189–209. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-24223-1_10.

Kirkland, J. L., and T. Tchkonia. 2020. 'Senolytic Drugs: From Discovery to Translation'. *Journal of Internal Medicine* 288 (5): 518–36. https://doi.org/10.1111/joim.13141.

Knudsen, Lars, Clemens Ruppert, and Matthias Ochs. 2017. 'Tissue Remodelling in Pulmonary Fibrosis'. *Cell and Tissue Research* 367 (3): 607–26. https://doi.org/10.1007/s00441-016-2543-2.

Komlósi, Zsolt I., Willem van de Veen, Nóra Kovács, Gergő Szűcs, Milena Sokolowska, Liam O'Mahony, Mübeccel Akdis, and Cezmi A. Akdis. 2022. 'Cellular and Molecular Mechanisms of Allergic Asthma'. *Molecular Aspects of Medicine*, Molecular Aspects of Asthma, 85 (June):100995. https://doi.org/10.1016/j.mam.2021.100995.

Kong, Fan-Qi, Shu-Jie Zhao, Peng Sun, Hao Liu, Jian Jie, Tao Xu, An-Di Xu, et al. 2020. 'Macrophage MSR1 Promotes the Formation of Foamy Macrophage and Neuronal Apoptosis after Spinal Cord Injury'. *Journal of Neuroinflammation* 17 (1): 62. https://doi.org/10.1186/s12974-020-01735-2.

Korpela, Elina, and Stanley K. Liu. 2014. 'Endothelial Perturbations and Therapeutic Strategies in Normal Tissue Radiation Damage'. *Radiation Oncology* 9 (1): 266. https://doi.org/10.1186/s13014-014-0266-7.

Korsunsky, Ilya, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-ru Loh, and Soumya Raychaudhuri. 2019. 'Fast, Sensitive and Accurate Integration of Single-Cell Data with Harmony'. *Nature Methods* 16 (12): 1289–96. https://doi.org/10.1038/s41592-019-0619-0.

Kotton, Darrell N., and Alan Fine. 2008. 'Lung Stem Cells'. *Cell and Tissue Research* 331 (1): 145–56. https://doi.org/10.1007/s00441-007-0479-2.

Krouwer, Vincent J. D., Liesbeth H. P. Hekking, Miriam Langelaar-Makkinje, Elsa Regan-Klapisz, and Jan Andries Post. 2012. 'Endothelial Cell Senescence Is Associated with Disrupted Cell-Cell Junctions and Increased Monolayer Permeability'. *Vascular Cell* 4 (1): 12. https://doi.org/10.1186/2045-824X-4-12.

Kumar-Singh, Ashish, Malgorzata Maria Parniewska, Nikolina Giotopoulou, Joman Javadi, Wenwen Sun, Tünde Szatmári, Katalin Dobra, Anders Hjerpe, and Jonas Fuxe. 2021. 'Nuclear Syndecan-1 Regulates Epithelial-Mesenchymal Plasticity in Tumor Cells'. *Biology* 10 (6): 521. https://doi.org/10.3390/biology10060521.

Kunkel, Eric J., and Eugene C. Butcher. 2003. 'Plasma-Cell Homing'. *Nature Reviews Immunology* 3 (10): 822–29. https://doi.org/10.1038/nri1203.

Kuwano, Kazuyoshi. 2007. 'Epithelial Cell Apoptosis and Lung Remodeling'. *Cellular & Molecular Immunology* 4 (6): 419–29.

Kwong, King F., Martin J. Edelman, Mohan Suntharalingam, Lindsay B. Cooper, Ziv Gamliel, Whitney Burrows, Petr Hausner, L. Austin Doyle, and Mark J. Krasna. 2005. 'High-

Dose Radiotherapy in Trimodality Treatment of Pancoast Tumors Results in High Pathologic Complete Response Rates and Excellent Long-Term Survival'. *The Journal of Thoracic and Cardiovascular Surgery* 129 (6): 1250–57. https://doi.org/10.1016/j.jtcvs.2004.12.050.

La Manno, Gioele, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, et al. 2018. 'RNA Velocity of Single Cells'. *Nature* 560 (7719): 494–98. https://doi.org/10.1038/s41586-018-0414-6.

Lambrecht, B. N., J-, B. Prins, and H. C. Hoogsteden. 2001. 'Lung Dendritic Cells and Host Immunity to Infection'. *European Respiratory Journal* 18 (4): 692–704. https://doi.org/10.1183/09031936.01.18040692.

Lee, Jin Young, Nabora S. Reyes, Supriya Ravishankar, Minqi Zhou, Maria Krasilnikov, Christian Ringler, Grace Pohan, et al. 2024. 'An in Vivo Screening Platform Identifies Senolytic Compounds That Target *P16^{INK4a+}* Fibroblasts in Lung Fibrosis'. *The Journal of Clinical Investigation* 134 (9). https://doi.org/10.1172/JCI173371.

Lehmann, Mareike, Martina Korfei, Kathrin Mutze, Stephan Klee, Wioletta Skronska-Wasek, Hani N. Alsafadi, Chiharu Ota, et al. 2017. 'Senolytic Drugs Target Alveolar Epithelial Cell Function and Attenuate Experimental Lung Fibrosis Ex Vivo'. *European Respiratory Journal* 50 (2). https://doi.org/10.1183/13993003.02367-2016.

Leve, Simone de, Florian Wirsdörfer, Federica Cappuccini, Alexandra Schütze, Alina V. Meyer, Katharina Röck, Linda F. Thompson, Jens W. Fischer, Martin Stuschke, and Verena Jendrossek. 2017. 'Loss of CD73 Prevents Accumulation of Alternatively Activated Macrophages and the Formation of Prefibrotic Macrophage Clusters in Irradiated Lungs'. *The FASEB Journal* 31 (7): 2869–80. https://doi.org/10.1096/fj.201601228R.

Li, Fengqi, Federica Piattini, Lea Pohlmeier, Qian Feng, Hubert Rehrauer, and Manfred Kopf. 2022. 'Monocyte-Derived Alveolar Macrophages Autonomously Determine Severe Outcome of Respiratory Viral Infection'. *Science Immunology* 7 (73): eabj5761. https://doi.org/10.1126/sciimmunol.abj5761.

Li, Shengyu, Pengzhi Zhang, Weiqing Chen, Lingqun Ye, Kristopher W. Brannan, Nhat-Tu Le, Jun-ichi Abe, John P. Cooke, and Guangyu Wang. 2024. 'A Relay Velocity Model Infers

Cell-Dependent RNA Velocity'. *Nature Biotechnology* 42 (1): 99–108. https://doi.org/10.1038/s41587-023-01728-5.

Limoli, Charles L., and Marie-Catherine Vozenin. 2023. 'Reinventing Radiobiology in the Light of FLASH Radiotherapy'. *Annual Review of Cancer Biology* 7 (Volume 7, 2023): 1–21. https://doi.org/10.1146/annurev-cancerbio-061421-022217.

Lin, Yingxin, Yue Cao, Hani Jieun Kim, Agus Salim, Terence P Speed, David M Lin, Pengyi Yang, and Jean Yee Hwa Yang. 2020. 'scClassify: Sample Size Estimation and Multiscale Classification of Cells Using Single and Multiple Reference'. *Molecular Systems Biology* 16 (6): e9389. https://doi.org/10.15252/msb.20199389.

Liu, Xue, Simon C. Rowan, Jiurong Liang, Changfu Yao, Guanling Huang, Nan Deng, Ting Xie, et al. 2021. 'Categorization of Lung Mesenchymal Cells in Development and Fibrosis'. *iScience* 24 (6): 102551. https://doi.org/10.1016/j.isci.2021.102551.

Livingston, Katie, Rachel A. Schlaak, Lindsay L. Puckett, and Carmen Bergom. 2020. 'The Role of Mitochondrial Dysfunction in Radiation-Induced Heart Disease: From Bench to Bedside'. *Frontiers in Cardiovascular Medicine* 7 (February). https://doi.org/10.3389/fcvm.2020.00020.

Lorusso, Bruno, Angela Falco, Denise Madeddu, Caterina Frati, Stefano Cavalli, Gallia Graiani, Andrea Gervasi, et al. 2015. 'Isolation and Characterization of Human Lung Lymphatic Endothelial Cells'. *BioMed Research International* 2015:747864. https://doi.org/10.1155/2015/747864.

Lotfollahi, Mohammad, Mohsen Naghipourfar, Malte D. Luecken, Matin Khajavi, Maren Büttner, Marco Wagenstetter, Žiga Avsec, et al. 2022. 'Mapping Single-Cell Data to Reference Atlases by Transfer Learning'. *Nature Biotechnology* 40 (1): 121–30. https://doi.org/10.1038/s41587-021-01001-7.

Luckheeram, Rishi Vishal, Rui Zhou, Asha Devi Verma, and Bing Xia. 2012. 'CD4+T Cells: Differentiation and Functions'. *Clinical and Developmental Immunology* 2012:925135. https://doi.org/10.1155/2012/925135.

Luecken, Malte D., M. Büttner, K. Chaichoompu, A. Danese, M. Interlandi, M. F. Mueller, D. C. Strobl, et al. 2022. 'Benchmarking Atlas-Level Data Integration in Single-Cell Genomics'. *Nature Methods* 19 (1): 41–50. https://doi.org/10.1038/s41592-021-01336-8.

Lun, Aaron T. L., Samantha Riesenfeld, Tallulah Andrews, The Phuong Dao, Tomas Gomes, John C. Marioni, and participants in the 1st Human Cell Atlas Jamboree. 2019. 'EmptyDrops: Distinguishing Cells from Empty Droplets in Droplet-Based Single-Cell RNA Sequencing Data'. *Genome Biology* 20 (1): 63. https://doi.org/10.1186/s13059-019-1662-y.

Lütge, Almut, Joanna Zyprych-Walczak, Urszula Brykczynska Kunzmann, Helena L. Crowell, Daniela Calini, Dheeraj Malhotra, Charlotte Soneson, and Mark D. Robinson. 2021. 'CellMixS: Quantifying and Visualizing Batch Effects in Single-Cell RNA-Seq Data'. *Life Science Alliance* 4 (6). https://doi.org/10.26508/lsa.202001004.

Ma, Jonathan, Bruce K. Rubin, and Judith A. Voynow. 2018. 'Mucins, Mucus, and Goblet Cells'. *Chest* 154 (1): 169–76. https://doi.org/10.1016/j.chest.2017.11.008.

Maaten, L.J.P. van der, and G.E. Hinton. 2008. 'Visualizing High-Dimensional Data Using t-SNE'. *Journal of Machine Learning Research* 9 (nov): 2579–2605.

McGinnis, Christopher S., Lyndsay M. Murrow, and Zev J. Gartner. 2019. 'DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors'. *Cell Systems* 8 (4): 329-337.e4. https://doi.org/10.1016/j.cels.2019.03.003.

McHugh, Domhnall, and Jesús Gil. 2017. 'Senescence and Aging: Causes, Consequences, and Therapeutic Avenues'. *Journal of Cell Biology* 217 (1): 65–77. https://doi.org/10.1083/jcb.201708092.

McInnes, Leland, John Healy, and James Melville. 2020. 'UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction'. arXiv. http://arxiv.org/abs/1802.03426.

Mehta, Vivek. 2005. 'Radiation Pneumonitis and Pulmonary Fibrosis in Non–Small-Cell Lung Cancer: Pulmonary Function, Prediction, and Prevention'. *International Journal of*

*Radiation Oncology*Biology*Physics* 63 (1): 5–24. https://doi.org/10.1016/j.ijrobp.2005.03.047.

Melin, Nicolas, Tural Yarahmadov, Daniel Sanchez-Taltavull, Fabienne E. Birrer, Tess M. Brodie, Benoît Petit, Andrea Felser, et al. 2022. 'A New Mouse Model of Radiation-Induced Liver Disease Reveals Mitochondrial Dysfunction as an Underlying Fibrotic Stimulus'. *JHEP Reports* 4 (7): 100508. https://doi.org/10.1016/j.jhepr.2022.100508.

Meng, Jingshu, Yan Li, Chao Wan, Yajie Sun, Xiaomeng Dai, Jing Huang, Yan Hu, et al. 2021. 'Targeting Senescence-like Fibroblasts Radiosensitizes Non–Small Cell Lung Cancer and Reduces Radiation-Induced Pulmonary Fibrosis'. *JCI Insight* 6 (23): e146334. https://doi.org/10.1172/jci.insight.146334.

Merad, Miriam, Priyanka Sathe, Julie Helft, Jennifer Miller, and Arthur Mortha. 2013. 'The Dendritic Cell Lineage: Ontogeny and Function of Dendritic Cells and Their Subsets in the Steady State and the Inflamed Setting'. *Annual Review of Immunology* 31:10.1146/annurev-immunol-020711–950. https://doi.org/10.1146/annurev-immunol-020711-074950.

Merkt, Wolfgang, Marta Bueno, Ana L. Mora, and David Lagares. 2020. 'Senotherapeutics: Targeting Senescence in Idiopathic Pulmonary Fibrosis'. *Seminars in Cell & Developmental Biology*, Autophagy in host-microbe interactions, 101 (May):104–10. https://doi.org/10.1016/j.semcdb.2019.12.008.

Meziani, Lydia, Michele Mondini, Benoît Petit, Alexandre Boissonnas, Vincent Thomas de Montpreville, Olaf Mercier, Marie-Catherine Vozenin, and Eric Deutsch. 2018. 'CSF1R Inhibition Prevents Radiation Pulmonary Fibrosis by Depletion of Interstitial Macrophages'. *European Respiratory Journal* 51 (3). https://doi.org/10.1183/13993003.02120-2017.

Miao, Zhun, Ke Deng, Xiaowo Wang, and Xuegong Zhang. 2018. 'DEsingle for Detecting Three Types of Differential Expression in Single-Cell RNA-Seq Data'. *Bioinformatics* 34 (18): 3223–24. https://doi.org/10.1093/bioinformatics/bty332.

Milad, Nadia, and Mathieu C. Morissette. 2021. 'Revisiting the Role of Pulmonary Surfactant in Chronic Inflammatory Lung Diseases and Environmental Exposure'. *European Respiratory Review* 30 (162). https://doi.org/10.1183/16000617.0077-2021.

Miller, Keith L, Su-Min Zhou, Robert C Barrier, Timothy Shafman, Rodney J Folz, Robert W Clough, and Lawrence B Marks. 2003. 'Long-Term Changes in Pulmonary Function Tests after Definitive Radiotherapy for Lung Cancer'. *International Journal of Radiation Oncology*Biology*Physics* 56 (3): 611–15. https://doi.org/10.1016/S0360-3016(03)00182-2.

Miura, Yoko, Maggie Lam, Jane E. Bourke, and Satoshi Kanazawa. 2022. 'Bimodal Fibrosis in a Novel Mouse Model of Bleomycin-Induced Usual Interstitial Pneumonia'. *Life Science Alliance* 5 (1). https://doi.org/10.26508/lsa.202101059.

Mootha, Vamsi K., Cecilia M. Lindgren, Karl-Fredrik Eriksson, Aravind Subramanian, Smita Sihag, Joseph Lehar, Pere Puigserver, et al. 2003. 'PGC-1α-Responsive Genes Involved in Oxidative Phosphorylation Are Coordinately Downregulated in Human Diabetes'. *Nature Genetics* 34 (3): 267–73. https://doi.org/10.1038/ng1180.

Myers, David J., and Jason M. Wallen. 2024. 'Lung Adenocarcinoma'. In *StatPearls*. Treasure Island (FL): StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK519578/.

Nabavi, Sheida, Daniel Schmolze, Mayinuer Maitituoheti, Sadhika Malladi, and Andrew H. Beck. 2016. 'EMDomics: A Robust and Powerful Method for the Identification of Genes Differentially Expressed between Heterogeneous Classes'. *Bioinformatics* 32 (4): 533–41. https://doi.org/10.1093/bioinformatics/btv634.

Nagaraja, Sunilgowda S., and Devipriya Nagarajan. 2018. 'Radiation-Induced Pulmonary Epithelial-Mesenchymal Transition: A Review on Targeting Molecular Pathways and Mediators'. *Current Drug Targets* 19 (10): 1191–1204. https://doi.org/10.2174/1389450119666180207092234.

Nam, Jae-Kyung, A-Ram Kim, Seo-Hyun Choi, Ji-Hee Kim, Su Chul Han, Seungwoo Park, Yong Jin Lee, et al. 2021. 'Pharmacologic Inhibition of HIF-1α Attenuates Radiation-Induced Pulmonary Fibrosis in a Preclinical Image Guided Radiation Therapy'. *International Journal of Radiation Oncology*Biology*Physics* 109 (2): 553–66. https://doi.org/10.1016/j.ijrobp.2020.09.006.

Nguyen, Hoang Quy, Nhu Hanh To, Patricia Zadigue, Stéphane Kerbrat, Alexandre De La Taille, Sabine Le Gouvello, and Yazid Belkacemi. 2018. 'Ionizing Radiation-Induced Cellular

Senescence Promotes Tissue Fibrosis after Radiotherapy. A Review'. *Critical Reviews in Oncology/Hematology* 129 (September):13–26. https://doi.org/10.1016/j.critrevonc.2018.06.012.

Padem, Nurcicek, and Carol Saltoun. 2019. 'Classification of Asthma'. *Allergy and Asthma Proceedings* 40 (6): 385–88. https://doi.org/10.2500/aap.2019.40.4253.

Pan, Jin, Deguan Li, Yanfeng Xu, Junling Zhang, Yueying Wang, Mengyi Chen, Shuai Lin, et al. 2017. 'Inhibition of Bcl-2/Xl With ABT-263 Selectively Kills Senescent Type II Pneumocytes and Reverses Persistent Pulmonary Fibrosis Induced by Ionizing Radiation in Mice'. *International Journal of Radiation Oncology*Biology*Physics* 99 (2): 353–61. https://doi.org/10.1016/j.ijrobp.2017.02.216.

Panagopoulos, Nikolaos, Vasilios Leivaditis, Efstratios Koletsis, Christos Prokakis, Panagiotis Alexopoulos, Nikolaos Baltayiannis, Antonios Hatzimichalis, et al. 2014. 'Pancoast Tumors: Characteristics and Preoperative Assessment'. *Journal of Thoracic Disease* 6 (Suppl 1): S108–15. https://doi.org/10.3978/j.issn.2072-1439.2013.12.29.

PANCOAST, HENRY K. 1924. 'IMPORTANCE OF CAREFUL ROENTGEN-RAY INVESTIGATIONS OF APICAL CHEST TUMORS'. *Journal of the American Medical Association* 83 (18): 1407–11. https://doi.org/10.1001/jama.1924.02660180025007.

Parimon, Tanyalak, Changfu Yao, Barry R. Stripp, Paul W. Noble, and Peter Chen. 2020. 'Alveolar Epithelial Type II Cells as Drivers of Lung Fibrosis in Idiopathic Pulmonary Fibrosis'. *International Journal of Molecular Sciences* 21 (7): 2269. https://doi.org/10.3390/ijms21072269.

Park, Hae-Ran, Sung-Kee Jo, and Uhee Jung. 2019. 'Ionizing Radiation Promotes Epithelial–to–Mesenchymal Transition in Lung Epithelial Cells by TGF-β-Producing M2 Macrophages'. *In Vivo* 33 (6): 1773–84. https://doi.org/10.21873/invivo.11668.

Park, Sung-Min, Tae-Il Kang, and Jae-Seon So. 2021. 'Roles of XBP1s in Transcriptional Regulation of Target Genes'. *Biomedicines* 9 (7): 791. https://doi.org/10.3390/biomedicines9070791.

Pasquini, Giovanni, Jesus Eduardo Rojo Arias, Patrick Schäfer, and Volker Busskamp. 2021. 'Automated Methods for Cell Type Annotation on scRNA-Seq Data'. *Computational and Structural Biotechnology Journal* 19:961–69. https://doi.org/10.1016/j.csbj.2021.01.015.

Pasut, Alessandra, Lisa M. Becker, Anne Cuypers, and Peter Carmeliet. 2021. 'Endothelial Cell Plasticity at the Single-Cell Level'. *Angiogenesis* 24 (2): 311–26. https://doi.org/10.1007/s10456-021-09797-3.

Pecoraro, Anthony R., Brian D. Hosfield, Hongge Li, William Christopher Shelley, and Troy A. Markel. 2021. 'Angiogenesis: A Cellular Response to Traumatic Injury'. *Shock* 55 (3): 301. https://doi.org/10.1097/SHK.0000000000001643.

Petersen, Ivy A, Michael G Haddock, John H Donohue, David M Nagorney, Joseph P Grill, Daniel J Sargent, and Leonard L Gunderson. 2002. 'Use of Intraoperative Electron Beam Radiotherapy in the Management of Retroperitoneal Soft Tissue Sarcomas'. *International Journal of Radiation Oncology*Biology*Physics* 52 (2): 469–75. https://doi.org/10.1016/S0360-3016(01)02595-0.

Pfeiffer, F., C. David, M. Burghammer, C. Riekel, and T. Salditt. 2002. 'Two-Dimensional X-Ray Waveguides and Point Sources'. *Science* 297 (5579): 230–34. https://doi.org/10.1126/science.1071994.

Phan, Thị Hằng Giang, Panagiotis Paliogiannis, Gheyath K. Nasrallah, Roberta Giordo, Ali Hussein Eid, Alessandro Giuseppe Fois, Angelo Zinellu, Arduino Aleksander Mangoni, and Gianfranco Pintus. 2021. 'Emerging Cellular and Molecular Determinants of Idiopathic Pulmonary Fibrosis'. *Cellular and Molecular Life Sciences* 78 (5): 2031–57. https://doi.org/10.1007/s00018-020-03693-7.

Picelli, Simone, Omid R. Faridani, Åsa K. Björklund, Gösta Winberg, Sven Sagasser, and Rickard Sandberg. 2014. 'Full-Length RNA-Seq from Single Cells Using Smart-Seq2'. *Nature Protocols* 9 (1): 171–81. https://doi.org/10.1038/nprot.2014.006.

Polverino, Francesca, Leen J. M. Seys, Ken R. Bracke, and Caroline A. Owen. 2016. 'B Cells in Chronic Obstructive Pulmonary Disease: Moving to Center Stage'. *American Journal*

of Physiology-Lung Cellular and Molecular Physiology 311 (4): L687–95. https://doi.org/10.1152/ajplung.00304.2016.

'Prevalence and Attributable Health Burden of Chronic Respiratory Diseases, 1990–2017: A Systematic Analysis for the Global Burden of Disease Study 2017'. 2020. *The Lancet. Respiratory Medicine* 8 (6): 585–96. https://doi.org/10.1016/S2213-2600(20)30105-3.

Raghu, G., F. Nyberg, and G. Morgan. 2004. 'The Epidemiology of Interstitial Lung Disease and Its Association with Lung Cancer'. *British Journal of Cancer* 91 (2): S3–10. https://doi.org/10.1038/sj.bjc.6602061.

Rannou, Emilie, Agnès François, Aurore Toullec, Olivier Guipaud, Valérie Buard, Georges Tarlet, Elodie Mintet, et al. 2015. 'In Vivo Evidence for an Endothelium-Dependent Mechanism in Radiation-Induced Normal Tissue Injury'. *Scientific Reports* 5 (1): 15738. https://doi.org/10.1038/srep15738.

Rashid, Mamoon, Deepak Singla, Arun Sharma, Manish Kumar, and Gajendra PS Raghava. 2009. 'Hmrbase: A Database of Hormones and Their Receptors'. *BMC Genomics* 10 (1): 307. https://doi.org/10.1186/1471-2164-10-307.

Rayess, Hani, Marilene B. Wang, and Eri S. Srivatsan. 2012. 'Cellular Senescence and Tumor Suppressor Gene P16'. *International Journal of Cancer* 130 (8): 1715–25. https://doi.org/10.1002/ijc.27316.

Regev, Aviv, Sarah A Teichmann, Eric S Lander, Ido Amit, Christophe Benoist, Ewan Birney, Bernd Bodenmiller, et al. 2017. 'The Human Cell Atlas'. Edited by Thomas R Gingeras. *eLife* 6 (December):e27041. https://doi.org/10.7554/eLife.27041.

Ribeiro, Carla M. P., and Bob A. Lubamba. 2017. 'Role of IRE1α/XBP-1 in Cystic Fibrosis Airway Inflammation'. *International Journal of Molecular Sciences* 18 (1): 118. https://doi.org/10.3390/ijms18010118.

Riemondy, Kent A., Nicole L. Jansing, Peng Jiang, Elizabeth F. Redente, Austin E. Gillen, Rui Fu, Alyssa J. Miller, et al. 2019. 'Single-Cell RNA Sequencing Identifies TGF-β as a Key

Regenerative Cue Following LPS-Induced Lung Injury'. *JCI Insight* 4 (8): e123637. https://doi.org/10.1172/jci.insight.123637.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2010. 'edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data'. *Bioinformatics* 26 (1): 139–40. https://doi.org/10.1093/bioinformatics/btp616.

Rodero, Mathieu P, Lucie Poupel, Pierre-Louis Loyher, Pauline Hamon, Fabrice Licata, Charlotte Pessel, David A Hume, Christophe Combadière, and Alexandre Boissonnas. 2015. 'Immune Surveillance of the Lung by Migrating Tissue Monocytes'. *eLife* 4 (July):e07847. https://doi.org/10.7554/eLife.07847.

Russell, David G., Pere-Joan Cardona, Mi-Jeong Kim, Sophie Allain, and Frédéric Altare. 2009. 'Foamy Macrophages and the Progression of the Human Tuberculosis Granuloma'. *Nature Immunology* 10 (9): 943–48. https://doi.org/10.1038/ni.1781.

S. N., Sunil Gowda, Raghavi Raviraj, Devipriya Nagarajan, and Weiling Zhao. 2019. 'Radiation-Induced Lung Injury: Impact on Macrophage Dysregulation and Lipid Alteration – a Review'. *Immunopharmacology and Immunotoxicology* 41 (3): 370–79. https://doi.org/10.1080/08923973.2018.1533025.

Sawada, Yu, Emi Mashima, Natsuko Saito-Sasaki, and Motonobu Nakamura. 2020. 'The Role of Cell Adhesion Molecule 1 (CADM1) in Cutaneous Malignancies'. *International Journal of Molecular Sciences* 21 (24): 9732. https://doi.org/10.3390/ijms21249732.

Sayan, Mutlay, and Brooke T. Mossman. 2016. 'The NLRP3 Inflammasome in Pathogenic Particle and Fibre-Associated Lung Inflammation and Diseases'. *Particle and Fibre Toxicology* 13 (1): 51. https://doi.org/10.1186/s12989-016-0162-4.

Schlegel, Wolfgang. 2006. 'New Technologies in 3D Conformal Radiation Therapy: Introduction and Overview'. In *New Technologies in Radiation Oncology*, edited by Wolfgang Schlegel, Thomas Bortfeld, and Anca-Ligia Grosu, 1–6. Medical Radiology. Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-29999-8_1.

Schwartz, Christian, Joerg U. Eberle, and David Voehringer. 2016. 'Basophils in Inflammation'. *European Journal of Pharmacology*, Pharmacological modulation of Mast cells and Basophils, 778 (May):90–95. https://doi.org/10.1016/j.ejphar.2015.04.049.

Schyns, Joey, Fabrice Bureau, and Thomas Marichal. 2018. 'Lung Interstitial Macrophages: Past, Present, and Future'. *Journal of Immunology Research* 2018 (April):5160794. https://doi.org/10.1155/2018/5160794.

Seguin, Laetitia, Manon Durandy, and Chloe C. Feral. 2022. 'Lung Adenocarcinoma Tumor Origin: A Guide for Personalized Medicine'. *Cancers* 14 (7): 1759. https://doi.org/10.3390/cancers14071759.

Selman, Moisés, and A Pardo. 2020. 'The Leading Role of Epithelial Cells in the Pathogenesis of Idiopathic Pulmonary Fibrosis'. *Cellular Signalling* 66 (February):109482. https://doi.org/10.1016/j.cellsig.2019.109482.

Selman, Moisés, and Annie Pardo. 2006. 'Role of Epithelial Cells in Idiopathic Pulmonary Fibrosis'. *Proceedings of the American Thoracic Society* 3 (4): 364–72. https://doi.org/10.1513/pats.200601-003TK.

Shao, Xin, Jie Liao, Xiaoyan Lu, Rui Xue, Ni Ai, and Xiaohui Fan. 2020. 'scCATCH: Automatic Annotation on Cell Types of Clusters from Single-Cell RNA Sequencing Data'. *iScience* 23 (3): 100882. https://doi.org/10.1016/j.isci.2020.100882.

Shi, Ting, Laura Denney, Huazhang An, Ling-Pei Ho, and Yuejuan Zheng. 2021. 'Alveolar and Lung Interstitial Macrophages: Definitions, Functions, and Roles in Lung Fibrosis'. *Journal of Leukocyte Biology* 110 (1): 107–14. https://doi.org/10.1002/JLB.3RU0720-418R.

Sikkema, Lisa, Ciro Ramírez-Suástegui, Daniel C. Strobl, Tessa E. Gillett, Luke Zappia, Elo Madissoon, Nikolay S. Markov, et al. 2023. 'An Integrated Cell Atlas of the Lung in Health and Disease'. *Nature Medicine* 29 (6): 1563–77. https://doi.org/10.1038/s41591-023-02327-2.

Silverman, Edwin K. 2020. 'Genetics of COPD'. *Annual Review of Physiology* 82 (February):413–31. https://doi.org/10.1146/annurev-physiol-021317-121224.

Song, Chang W., Eli Glatstein, Lawrence B. Marks, Bahman Emami, Jimm Grimm, Paul W. Sperduto, Mi-Sook Kim, Susanta Hui, Kathryn E. Dusenbery, and L. Chinsoo Cho. 2021. 'Biological Principles of Stereotactic Body Radiation Therapy (SBRT) and Stereotactic Radiation Surgery (SRS): Indirect Cell Death'. *International Journal of Radiation Oncology*Biology*Physics*, A Red Journal Special Issue: HyTEC, 110 (1): 21–34. https://doi.org/10.1016/j.ijrobp.2019.02.047.

Song, Licheng, Kuan Li, Huaiyong Chen, and Lixin Xie. 2024. 'Cell Cross-Talk in Alveolar Microenvironment: From Lung Injury to Fibrosis'. *American Journal of Respiratory Cell and Molecular Biology* 71 (1): 30–42. https://doi.org/10.1165/rcmb.2023-0426TR.

Stuart, Tim, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M. Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. 2019. 'Comprehensive Integration of Single-Cell Data'. *Cell* 177 (7): 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

Su, Lulu, Yinping Dong, Yueying Wang, Yuquan Wang, Bowen Guan, Yanhua Lu, Jing Wu, et al. 2021. 'Potential Role of Senescent Macrophages in Radiation-Induced Pulmonary Fibrosis'. *Cell Death & Disease* 12 (6): 1–12. https://doi.org/10.1038/s41419-021-03811-8.

Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, et al. 2005. 'Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles'. *Proceedings of the National Academy of Sciences* 102 (43): 15545–50. https://doi.org/10.1073/pnas.0506580102.

Subramanian, Ayshwarya, Mikhail Alperovich, Yiming Yang, and Bo Li. 2022. 'Biology-Inspired Data-Driven Quality Control for Scientific Discovery in Single-Cell Transcriptomics'. *Genome Biology* 23 (1): 267. https://doi.org/10.1186/s13059-022-02820-w.

Sun, Xun, Kexin Li, Rongrong Zha, Shengzhi Liu, Yao Fan, Di Wu, Misato Hase, et al. 2021. 'Preventing Tumor Progression to the Bone by Induced Tumor-Suppressing MSCs'. *Theranostics* 11 (11): 5143–59. https://doi.org/10.7150/thno.58779.

Sung, Yeoun Eun, Uiju Cho, and Kyo Young Lee. 2020. 'Peripheral Type Squamous Cell Carcinoma of the Lung: Clinicopathologic Characteristics in Comparison to the Central Type'. *Journal of Pathology and Translational Medicine* 54 (4): 290–99. https://doi.org/10.4132/jptm.2020.05.04.

Tan, Yuqi, and Patrick Cahan. 2019. 'SingleCellNet: A Computational Tool to Classify Single Cell RNA-Seq Data Across Platforms and Across Species'. *Cell Systems* 9 (2): 207-213.e2. https://doi.org/10.1016/j.cels.2019.06.004.

Taylor, A, and M E B Powell. 2004. 'Intensity-Modulated Radiotherapy—What Is It?' *Cancer Imaging* 4 (2): 68–73. https://doi.org/10.1102/1470-7330.2004.0003.

Thandra, Krishna Chaitanya, Adam Barsouk, Kalyan Saginala, John Sukumar Aluru, and Alexander Barsouk. 2021. 'Epidemiology of Lung Cancer'. *Contemporary Oncology (Poznan, Poland)* 25 (1): 45–52. https://doi.org/10.5114/wo.2021.103829.

Thomas, Anita C., Wouter J. Eijgelaar, Mat J. A. P. Daemen, and Andrew C. Newby. 2015. 'Foam Cell Formation In Vivo Converts Macrophages to a Pro-Fibrotic Phenotype'. *PLOS ONE* 10 (7): e0128163. https://doi.org/10.1371/journal.pone.0128163.

Tilley, Ann E., Matthew S. Walters, Renat Shaykhiev, and Ronald G. Crystal. 2015. 'Cilia Dysfunction in Lung Disease'. *Annual Review of Physiology* 77 (1): 379–406. https://doi.org/10.1146/annurev-physiol-021014-071931.

Tirosh, Itay, Benjamin Izar, Sanjay M. Prakadan, Marc H. Wadsworth, Daniel Treacy, John J. Trombetta, Asaf Rotem, et al. 2016. 'Dissecting the Multicellular Ecosystem of Metastatic Melanoma by Single-Cell RNA-Seq'. *Science* 352 (6282): 189–96. https://doi.org/10.1126/science.aad0501.

Travaglini, Kyle J., Ahmad N. Nabhan, Lolita Penland, Rahul Sinha, Astrid Gillich, Rene V. Sit, Stephen Chang, et al. 2020. 'A Molecular Cell Atlas of the Human Lung from Single-Cell RNA Sequencing'. *Nature* 587 (7835): 619–25. https://doi.org/10.1038/s41586-020-2922-4.

Travis, Elizabeth L. 1980. 'The Sequence of Histological Changes in Mouse Lungs after Single Doses of X-Rays'. *International Journal of Radiation Oncology*Biology*Physics* 6 (3): 345–47. https://doi.org/10.1016/0360-3016(80)90145-5.

Travis, William D. 2011. 'Pathology of Lung Cancer'. *Clinics in Chest Medicine* 32 (4): 669–92. https://doi.org/10.1016/j.ccm.2011.08.005.

Trimm, Emily, and Kristy Red-Horse. 2023. 'Vascular Endothelial Cell Development and Diversity'. *Nature Reviews Cardiology* 20 (3): 197–210. https://doi.org/10.1038/s41569-022-00770-1.

Trott, Klaus Rüdiger, Thomas Herrmann, and Michael Kasper. 2004. 'Target Cells in Radiation Pneumopathy'. *International Journal of Radiation Oncology*Biology*Physics* 58 (2): 463–69. https://doi.org/10.1016/j.ijrobp.2003.09.045.

Troulé, Kevin, Robert Petryszak, Martin Prete, James Cranley, Alicia Harasty, Zewen Kelvin Tuong, Sarah A. Teichmann, Luz Garcia-Alonso, and Roser Vento-Tormo. 2023. 'CellPhoneDB v5: Inferring Cell-Cell Communication from Single-Cell Multiomics Data'. arXiv. https://doi.org/10.48550/arXiv.2311.04567.

Tsikis, S. T., S. C. Fligor, T. I. Hirsch, A. Pan, L. J. Yu, H. Kishikawa, M. M. Joiner, P. D. Mitchell, and M. Puder. 2022. 'Lipopolysaccharide-Induced Murine Lung Injury Results in Long-Term Pulmonary Changes and Downregulation of Angiogenic Pathways'. *Scientific Reports* 12 (June):10245. https://doi.org/10.1038/s41598-022-14618-8.

Tsukui, Tatsuya, Kai-Hui Sun, Joseph B. Wetter, John R. Wilson-Kanamori, Lisa A. Hazelwood, Neil C. Henderson, Taylor S. Adams, et al. 2020. 'Collagen-Producing Lung Cell Atlas Identifies Multiple Subsets with Distinct Localization and Relevance to Fibrosis'. *Nature Communications* 11 (1): 1920. https://doi.org/10.1038/s41467-020-15647-5.

Ucero, Alvaro C., Latifa Bakiri, Ben Roediger, Masakatsu Suzuki, Maria Jimenez, Pratyusha Mandal, Paola Braghetta, et al. 2019. 'Fra-2–Expressing Macrophages Promote Lung Fibrosis'. *The Journal of Clinical Investigation* 129 (8): 3293–3309. https://doi.org/10.1172/JCI125366.

UniProt Consortium. 2023. ‘UniProt: The Universal Protein Knowledgebase in 2023’. *Nucleic Acids Research* 51 (D1): D523–31. https://doi.org/10.1093/nar/gkac1052.

Ural, Basak B., Stephen T. Yeung, Payal Damani-Yokota, Joseph C. Devlin, Maren de Vries, Paola Vera-Licona, Tasleem Samji, et al. 2020. ‘Identification of a Nerve-Associated, Lung-Resident Interstitial Macrophage Subset with Distinct Localization and Immunoregulatory Properties’. *Science Immunology* 5 (45): eaax8756. https://doi.org/10.1126/sciimmunol.aax8756.

Venkatesulu, Bhanu Prasad, Lakshmi Shree Mahadevan, Maureen L. Aliru, Xi Yang, Monica Himaani Bodd, Pankaj K. Singh, Syed Wamique Yusuf, Jun-ichi Abe, and Sunil Krishnan. 2018. ‘Radiation-Induced Endothelial Vascular Injury’. *JACC: Basic to Translational Science* 3 (4): 563–72. https://doi.org/10.1016/j.jacbts.2018.01.014.

Wang, Fa, Christopher Ting, Kent A. Riemondy, Michael Douglas, Kendall Foster, Nisha Patel, Norihito Kaku, et al. 2023. ‘Regulation of Epithelial Transitional States in Murine and Human Pulmonary Fibrosis’. *The Journal of Clinical Investigation* 133 (22). https://doi.org/10.1172/JCI165612.

Wang, Renying, Peijing Zhang, Jingjing Wang, Lifeng Ma, Weigao E, Shengbao Suo, Mengmeng Jiang, et al. 2023. ‘Construction of a Cross-Species Cell Landscape at Single-Cell Level’. *Nucleic Acids Research* 51 (2): 501–16. https://doi.org/10.1093/nar/gkac633.

Wang, Sheng, Angela Oliveira Pisco, Aaron McGeever, Maria Brbic, Marinka Zitnik, Spyros Darmanis, Jure Leskovec, Jim Karkanias, and Russ B. Altman. 2019. ‘Unifying Single-Cell Annotations Based on the Cell Ontology’. https://doi.org/10.1101/810234.

Wang, Tianyu, Boyang Li, Craig E. Nelson, and Sheida Nabavi. 2019. ‘Comparative Analysis of Differential Gene Expression Analysis Tools for Single-Cell RNA Sequencing Data’. *BMC Bioinformatics* 20 (1): 40. https://doi.org/10.1186/s12859-019-2599-6.

Wang, Tianyu, and Sheida Nabavi. 2018. ‘SigEMD: A Powerful Method for Differential Gene Expression Analysis in Single-Cell RNA Sequencing Data’. *Methods*, Data mining methods for analyzing biological data in terms of phenotypes, 145 (August):25–32. https://doi.org/10.1016/j.ymeth.2018.04.017.

Wculek, Stefanie K., Gillian Dunphy, Ignacio Heras-Murillo, Annalaura Mastrangelo, and David Sancho. 2022. 'Metabolism of Tissue Macrophages in Homeostasis and Pathology'. *Cellular & Molecular Immunology* 19 (3): 384–408. https://doi.org/10.1038/s41423-021-00791-9.

Wen, Xiao-Fen, Min Chen, Yang Wu, Min-Na Chen, Aleksandra Glogowska, Thomas Klonisch, and Guo-Jun Zhang. 2018. 'Inhibitor of DNA Binding 2 Inhibits Epithelial-Mesenchymal Transition via Up-Regulation of Notch3 in Breast Cancer'. *Translational Oncology* 11 (5): 1259–70. https://doi.org/10.1016/j.tranon.2018.07.015.

White, Eric S. 2015. 'Lung Extracellular Matrix and Fibroblast Function'. *Annals of the American Thoracic Society* 12 (Suppl 1): S30–33. https://doi.org/10.1513/AnnalsATS.201406-240MG.

Whitsett, Jeffrey A., Tanya V. Kalin, Yan Xu, and Vladimir V. Kalinichenko. 2019. 'Building and Regenerating the Lung Cell by Cell'. *Physiological Reviews* 99 (1): 513–54. https://doi.org/10.1152/physrev.00001.2018.

Wickham, Hadley. 2016. *Ggplot2*. https://link.springer.com/book/10.1007/978-3-319-24277-4.

Wijerathne, Harshani, Jordan C. Langston, Qingliang Yang, Shuang Sun, Curtis Miyamoto, Laurie E. Kilpatrick, and Mohammad F. Kiani. 2021. 'Mechanisms of Radiation-Induced Endothelium Damage: Emerging Models and Technologies'. *Radiotherapy and Oncology : Journal of the European Society for Therapeutic Radiology and Oncology* 158 (May):21–32. https://doi.org/10.1016/j.radonc.2021.02.007.

Wijsenbeek, Marlies, Atsushi Suzuki, and Toby M. Maher. 2022. 'Interstitial Lung Diseases'. *The Lancet* 400 (10354): 769–86. https://doi.org/10.1016/S0140-6736(22)01052-2.

Willemsen, Lisa, and Menno PJ de Winther. 2020. 'Macrophage Subsets in Atherosclerosis as Defined by Single-Cell Technologies'. *The Journal of Pathology* 250 (5): 705–14. https://doi.org/10.1002/path.5392.

Willis, Brigham C., Roland M. duBois, and Zea Borok. 2006. 'Epithelial Origin of Myofibroblasts during Fibrosis in the Lung'. *Proceedings of the American Thoracic Society* 3 (4): 377–82. https://doi.org/10.1513/pats.200601-004TK.

Wirsdörfer, Florian, and Verena Jendrossek. 2016. 'The Role of Lymphocytes in Radiotherapy-Induced Adverse Late Effects in the Lung'. *Frontiers in Immunology* 7 (December). https://doi.org/10.3389/fimmu.2016.00591.

Wynn, Thomas A. 2011. 'Integrating Mechanisms of Pulmonary Fibrosis'. *The Journal of Experimental Medicine* 208 (7): 1339–50. https://doi.org/10.1084/jem.20110551.

Xiang, Ruizhi, Wencan Wang, Lei Yang, Shiyuan Wang, Chaohan Xu, and Xiaowen Chen. 2021. 'A Comparison for Dimensionality Reduction Methods of Single-Cell RNA-Seq Data'. *Frontiers in Genetics* 12 (March). https://doi.org/10.3389/fgene.2021.646936.

Xie, Peng, Mingxuan Gao, Chunming Wang, Jianfei Zhang, Pawan Noel, Chaoyong Yang, Daniel Von Hoff, Haiyong Han, Michael Q Zhang, and Wei Lin. 2019. 'SuperCT: A Supervised-Learning Framework for Enhanced Characterization of Single-Cell Transcriptomic Profiles'. *Nucleic Acids Research* 47 (8): e48. https://doi.org/10.1093/nar/gkz116.

Xiong, Yumei, and Rémy Bosselut. 2012. 'CD4-CD8 Differentiation in the Thymus: Connecting Circuits and Building Memories'. *Current Opinion in Immunology* 24 (2): 139–45. https://doi.org/10.1016/j.coi.2012.02.002.

Yan, Yujie, Jiamei Fu, Roman O. Kowalchuk, Christopher M. Wright, Ran Zhang, Xuefei Li, and Yaping Xu. 2022. 'Exploration of Radiation-Induced Lung Injury, from Mechanism to Treatment: A Narrative Review'. *Translational Lung Cancer Research* 11 (2): 307–22. https://doi.org/10.21037/tlcr-22-108.

Yang, Ganghao, Yang Yang, Yiping Liu, and Xiaoshu Liu. 2023. 'Regulation of Alveolar Macrophage Death in Pulmonary Fibrosis: A Review'. *Apoptosis* 28 (11–12): 1505–19. https://doi.org/10.1007/s10495-023-01888-4.

Yao, Changfu, Xiangrong Guan, Gianni Carraro, Tanyalak Parimon, Xue Liu, Guanling Huang, Apoorva Mulay, et al. 2021. 'Senescence of Alveolar Type 2 Cells Drives Progressive

Pulmonary Fibrosis'. *American Journal of Respiratory and Critical Care Medicine*, March. https://doi.org/10.1164/rccm.202004-1274OC.

Yin, Zhongyuan, Guanghai Yang, Sisi Deng, and Qiong Wang. 2019. 'Oxidative Stress Levels and Dynamic Changes in Mitochondrial Gene Expression in a Radiation-Induced Lung Injury Model'. *Journal of Radiation Research* 60 (2): 204–14. https://doi.org/10.1093/jrr/rry105.

Ying, Hang-Jie, Min Fang, and Ming Chen. 2021. 'Progress in the Mechanism of Radiation-Induced Lung Injury'. *Chinese Medical Journal* 134 (02): 161–63. https://doi.org/10.1097/CM9.0000000000001032.

Young, Matthew D., and Sam Behjati. 2020. 'SoupX Removes Ambient RNA Contamination from Droplet-Based Single-Cell RNA Sequencing Data'. *GigaScience* 9 (12): giaa151. https://doi.org/10.1093/gigascience/giaa151.

Yu, Xiao-Hua, Yu-Chang Fu, Da-Wei Zhang, Kai Yin, and Chao-Ke Tang. 2013. 'Foam Cells in Atherosclerosis'. *Clinica Chimica Acta* 424 (September):245–52. https://doi.org/10.1016/j.cca.2013.06.006.

Zaynagetdinov, Rinat, Taylor P. Sherrill, Peggy L. Kendall, Brahm H. Segal, Kevin P. Weller, Robert M. Tighe, and Timothy S. Blackwell. 2013. 'Identification of Myeloid Cell Subsets in Murine Lungs Using Flow Cytometry'. *American Journal of Respiratory Cell and Molecular Biology* 49 (2): 180–89. https://doi.org/10.1165/rcmb.2012-0366MA.

Zhang, Allen W., Ciara O'Flanagan, Elizabeth A. Chavez, Jamie L. P. Lim, Nicholas Ceglia, Andrew McPherson, Matt Wiens, et al. 2019. 'Probabilistic Cell-Type Assignment of Single-Cell RNA-Seq for Tumor Microenvironment Profiling'. *Nature Methods* 16 (10): 1007–15. https://doi.org/10.1038/s41592-019-0529-1.

Zhang, Nu, and Michael J. Bevan. 2011. 'CD8+ T Cells: Foot Soldiers of the Immune System'. *Immunity* 35 (2): 161–68. https://doi.org/10.1016/j.immuni.2011.07.010.

Zhang, Shijie, Xiang Tong, Sitong Liu, Jizhen Huang, Li Zhang, Tianli Zhang, Dongguang Wang, and Hong Fan. 2023. 'AAV9-Tspyl2 Gene Therapy Retards Bleomycin-Induced

Pulmonary Fibrosis by Modulating Downstream TGF-β Signaling in Mice'. *Cell Death & Disease* 14 (6): 1–14. https://doi.org/10.1038/s41419-023-05889-8.

Zhao, Quan, Jianghui Wang, Ilya V. Levichkin, Stan Stasinopoulos, Michael T. Ryan, and Nicholas J. Hoogenraad. 2002. 'A Mitochondrial Specific Stress Response in Mammalian Cells'. *The EMBO Journal* 21 (17): 4411–19. https://doi.org/10.1093/emboj/cdf445.

Zhao, Weiming, Lan Wang, Yaxuan Wang, Hongmei Yuan, Mengxia Zhao, Hui Lian, Shuaichen Ma, Kai Xu, Zhongzheng Li, and Guoying Yu. 2023. 'Injured Endothelial Cell: A Risk Factor for Pulmonary Fibrosis'. *International Journal of Molecular Sciences* 24 (10): 8749. https://doi.org/10.3390/ijms24108749.

Zheng, Grace X. Y., Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, et al. 2017. 'Massively Parallel Digital Transcriptional Profiling of Single Cells'. *Nature Communications* 8 (1): 14049. https://doi.org/10.1038/ncomms14049.

Zhou, Shenghui, Jiaojiao Zhu, Ping-Kun Zhou, and Yongqing Gu. 2022. 'Alveolar Type 2 Epithelial Cell Senescence and Radiation-Induced Pulmonary Fibrosis'. *Frontiers in Cell and Developmental Biology* 10 (November). https://doi.org/10.3389/fcell.2022.999600.

Zilionis, Rapolas, Camilla Engblom, Christina Pfirschke, Virginia Savova, David Zemmour, Hatice D. Saatcioglu, Indira Krishnan, et al. 2019. 'Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species'. *Immunity* 50 (5): 1317-1334.e10. https://doi.org/10.1016/j.immuni.2019.03.009.

Annexes

ANNEXE I – Markers used for the identification of the lung populations.

TABLE 1 – markers used for the identification of the lung epithelial cells populations.

Cell type	Markers human	Markers mouse	Reference
AT1	AGER, PDPN, CLIC5	Ager, Clib5, Pdpn	(Travaglini et al. 2020)
AT2	SFTPB, SFTPC, SFTPD	Sftpc, Muc1, Etv5	
Club cells	CYP2F2, SCGB3A2, CCKAR	Scgb3a2, Cyp2f2, Cckar	
Goblet cells	MUC5B, MUC5AC, SPDEF	Tff2, Muc5b, Spdef	
Ciliated cells	FOXJ1, TUBB1, TP73	Foxj1, Ccdc78, Fam183b	
Basal cells	KRT5, KRT14, TP63, DAPL1	Krt5, Trp63, Dapl1	
Aberrant basaloid cells	TP63, KRT17, LAMB3, LAMC2		(Adams et al. 2020)
AT0	SFTPB, SFTPC, SCGB3A2	Sftpb, Sftpc, Scgb3a2	(Kadur Lakshminarasimha Murthy et al. 2022)
Secretory cells	SFTPB, SCGB3A2, SCGB1A1	Sftpb, Scgb3a2, Scgb1a1	

TABLE 2 – markers used for the identification of the lung mesenchymal cells populations.

Cell type	Markers human	Markers mouse	Reference
Fibroblasts	COL1A1, PDGFRA	Pdgfra	(Travaglini et al. 2020)
Myofibroblasts	COL1A1, PDGFRA, ELN, ACTA2	Wif1, Fgf18, Aspn	
Adventitial fibroblasts		Serpinf1, Pi16, Entpd2	
Alveolar fibroblasts		Fgfr4, Slc7a10, Slc38a5	
Pericytes		Trpc6, Higd1b, Vtn	
Pericytes	CSPG4, TRPC6, PDGFRB		
Vascular smooth muscle cells	CNN1, ACTA2, TAGLN, RGS5		
Airway smooth muscle cells	CNN1, ACTA2, TAGLN, DES, LGR6		
Smooth muscle cells		Acta2, Cnn1, Tagln	
Peribronchial fibroblasts		Fgf8, Hhip	(Tsukui et al. 2020)
Lipofibroblasts	COL1A1, PDGFRA, PLIN2, APOE		(Travaglini et al. 2020), (Liu et al. 2021)

TABLE 3 – markers used for the identification of the lung endothelial cells populations.

Cell type	Markers human	Markers mouse	Reference
aCap	SOSTDC1, EDNRB, HPGD	Car4, Ednrb, Fibin	(Gillich et al. 2020)
gCap	FCN3, EDN1, SLC6A4	Gpihbp1, Pvalp, Cd93, Ptprb	
Vein EC	CPE, PTGDS, C7, PLA1A	Nr2f2, Vwf	
Artery EC	CXCL12, GJA5, DKK2	Gja5, Bmx, Vwf	
Lymphatic EC	CCL21, TFF3, MMRN1	Pdpn, Prox1	
Tip cells	ADM, ANKRDd37, C1QTNF6, CLDN5, COL4A1, COL4A2, COTL1, DLL4, EDNRB, FSCN1, GPIHBP1, HSPG2, IGFBP3, INHBB, JUP, KCNE3, KCNJ8, LAMA4, LAMB1, LXN, MARCKSL1, MCAM, MEST, N4 BP3, NID2, NOTCH4, PLOD1, PLXND1, PMEPA1, PTN, RAMP3, RBP1, RGCC, RHOC, TRP53ILL, UNC5B, KDR, FLT4	Adm, Ankrd37, C1qtnf6, Cldn5, Col4a1, Col4a2, Cotl1, Dll4, Ednrb, Fscn1, Gpihbp1, Hspg2, Igfbp3, Inhbb, Jup, Kcne3, Kcnj8, Lama4, Lamb1, Lxn, Marcksl1, Mcam, Mest, N4 bp3, Nid2, Notch4, Plod1, Plxnd1, Pmepa1, Ptn, Ramp3, Rbp1, Rgcc, Rhoc, Trp53ill, Unc5B, Kdr, Flt4	(W. Chen et al. 2019)
Stalk cells	ACKR1, AQP1, C1QTNF9, CD36, CSRP2, EHD4, FBLN5, HSPB1, LIGP1, IL6ST, JAM2, LGALS3, LRG1, MEOX2, PLSCR2, SDPR, SELP, SPINT2, TGFBI, TGM2, TMEM176A, TMEM176B, TMEM252, TSPAN7, FLT1, VWF	Ackr1, Aqp1, C1qtnf9, Cd36, Csrp2, Ehd4, Fbln5, Hspb1, Ligp1, Il6st, Jam2, Lgals3, Lrg1, Meox2, Plscr2, Sdpr, Selp, Spint2, Tgfbi, Tgm2, Tmem176a, Tmem176b, Tmem252, Tspan7, Flt1, Vwf	

TABLE 4 – markers used for the identification of the lung myeloid cells populations.

Cell type	Markers human	Markers mouse	Reference
Conventional DC	CD1C, CD141	Cd45, Cd8, Cd4, Cd11b	(Cook and MacDonald 2016; Merad et al. 2013)
Plasmacytoid DC	LILRB4, IRF8, LILRA4	Siglech, Cd300c, Klk1	(Travaglini et al. 2020)
Monocyte derived DC	MHCII, CLEC9A, LAMP3, CD1C, PLD4	MHCII, Cd24a, Clec9a, Sirpa, Itgam	
Classical monocytes	CD14, S100A8	Cd14, F13a1, Ly6c2	
Non classical monocytes	S100A8, CD16	Emr4, Itgax, Treml4	
Macrophage	MARCO, MSR1, MRC1		
Neutrophils	S100A8, S100A9, IFITM2, FCGR3B	Retnlg, S100a8, S100a9	
AM	CYP27A1, MARCO, FABP4	Marco, Msr1, Mcr1	(Sikkema et al. 2023)
IM	F13A1, FOLR2	MHCII, C1qa, Trem2	
Resident AM		Siglec-F, Cd11c	(Shi et al. 2021)
Recruited AM		Cd11b	
Vessel associated IM		Lyve1, Prg4, Tgfb2	(Ural et al. 2020; Gibbings et al. 2017)
Nerve associated IM		MHCII	

TABLE 5 – markers used for the identification of the lung lymphoid cells populations.

Cell type	Markers human	Markers mouse	Reference
B cells	CD79A, CD24, MS4A1, CD19	Cd79a, Cd19, Ms4a1	(Travaglini et al. 2020)
Plasma cells	CD79A, CD27, SLAMF7	Cd79a, Cd27, Slamf7	
T cells	CD3E	CD3e, Lck	
CD8+ T cells	CD8, GZMH	Cd8e	
CD4+ T cells	CD4	Cd4	
NK cells	KLRD1, NKG7, TYROBP	Klrb1a, Tyrobp, Gzma	
Basophils	MS4A2, CPA3, TPSAB1	Cpa3, Ms4a2, Mcpt8	

ANNEXE II – Metadata of the human samples sequenced.

TABLE 6 – metadata of the human samples sequenced.

SAMPLE NAME	SEXE	AGE	IR STATUS	SORTED	CONDITION	NUMBER OF CELLS	CELLRANGER VERSION	REFERENCE GENOME
PATIENT1_NI	M	64	NI	NO	Pancoast_NI	4833	3.1.0	GRCh38-3.0.0
PATIENT1_IR	M	64	IR	NO	Pancoast_IR	3675	3.1.0	GRCh38-3.0.0
PATIENT2_NI	M	61	NI	NO	Pancoast_NI	6184	3.1.0	GRCh38-3.0.0
PATIENT2_IR	M	61	IR	NO	Pancoast_IR	5881	3.1.0	GRCh38-3.0.0
PATIENT3_NI	M	53	NI	CD45neg	Pancoast_NI	1984	3.1.0	GRCh38-3.0.0
PATIENT3_IR	M	53	IR	CD45neg	Pancoast_IR	1479	3.1.0	GRCh38-3.0.0
PATIENT4_NI	M	42	NI	NO	Pancoast_NI	5894	6.0.0	GRCh38-2020-A
PATIENT4_IR	M	42	IR	NO	Pancoast_IR	25175	6.0.0	GRCh38-2020-A
PATIENT5_NI	M	42	NI	NO	Pancoast_NI	3376	7.1.0	GRCh38-2020-A
PATIENT5_IR	M	42	IR	NO	Pancoast_IR	1876	7.1.0	GRCh38-2020-A
PATIENT6_NI	F	75	NI	NO	Pancoast_NI	11057	7.1.0	GRCh38-2020-A
PATIENT6_IR	F	75	IR	NO	Pancoast_IR	5634	7.1.0	GRCh38-2020-A

ANNEXE III – Metadata of the mouse samples sequenced.

TABLE 7 – metadata of the mouse samples sequenced.

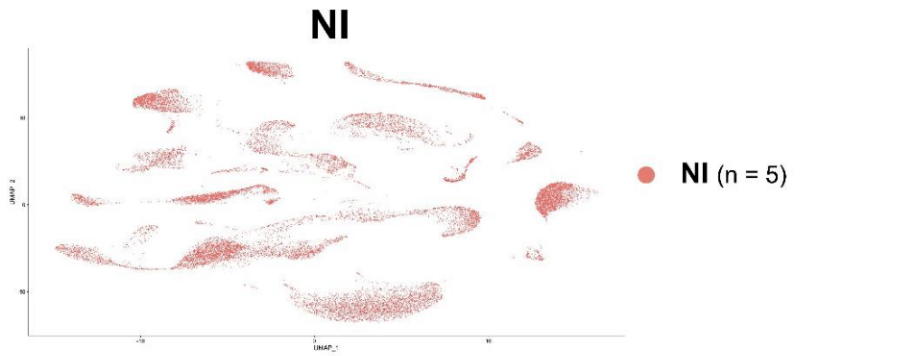
SAMPLE NAME	EXPERIMENT	IR DOSE	TIME POINT	CONDITION	NUMBER OF CELLS	CELLRANGER VERSION
NI_YOUNG_1	M51	NI	NI	NI	4241	7.1.0
NI_YOUNG_2	M51	NI	NI	NI	4062	7.1.0
IR_CONV_10GY_1M_1	M48	10Gy	1M	IR_CONV_10Gy_1M	5673	7.1.0
IR_CONV_10GY_1M_2	M48	10Gy	1M	IR_CONV_10Gy_1M	4622	7.1.0
IR_CONV_10GY_2M_1	M46	10Gy	2M	IR_CONV_10Gy_2M	3947	7.1.0
IR_CONV_10GY_2M_2	M46	10Gy	2M	IR_CONV_10Gy_2M	5003	7.1.0
IR_CONV_10GY_3M_1	M47	10Gy	3M	IR_CONV_10Gy_3M	8105	7.1.0
IR_CONV_10GY_3M_2	M47	10Gy	3M	IR_CONV_10Gy_3M	3901	7.1.0
IR_CONV_10GY_4M_1	M45	10Gy	4M	IR_CONV_10Gy_4M	4508	7.1.0
IR_CONV_10GY_4M_2	M45	10Gy	4M	IR_CONV_10Gy_4M	4383	7.1.0
IR_CONV_10GY_5M_1	M51	10Gy	5M	IR_CONV_10Gy_5M	4284	7.1.0
IR_CONV_10GY_5M_2	M51	10Gy	5M	IR_CONV_10Gy_5M	7802	7.1.0
IR_CONV_13GY_24H_1	M42	13Gy	24H	IR_CONV_13Gy_24H	6599	6.0.0
IR_CONV_13GY_24H_2	M42	13Gy	24H	IR_CONV_13Gy_24H	7215	6.0.0
IR_CONV_13GY_24H_3	M42	13Gy	24H	IR_CONV_13Gy_24H	4991	6.0.0
IR_CONV_13GY_1M_1	M48	13Gy	1M	IR_CONV_13Gy_1M	4527	7.1.0
IR_CONV_13GY_1M_2	M48	13Gy	1M	IR_CONV_13Gy_1M	3587	7.1.0
IR_CONV_13GY_2M_1	M46	13Gy	2M	IR_CONV_13Gy_2M	5984	7.1.0
IR_CONV_13GY_2M_2	M46	13Gy	2M	IR_CONV_13Gy_2M	6285	7.1.0
IR_CONV_13GY_3M_1	M47	13Gy	3M	IR_CONV_13Gy_3M	5846	7.1.0
IR_CONV_13GY_3M_2	M47	13Gy	3M	IR_CONV_13Gy_3M	4424	7.1.0
IR_CONV_13GY_4M_1	M44	13Gy	4M	IR_CONV_13Gy_4M	4622	7.1.0
IR_CONV_13GY_4M_2	M44	13Gy	4M	IR_CONV_13Gy_4M	3805	7.1.0
IR_CONV_13GY_5M_1	M41	13Gy	5M	IR_CONV_13Gy_5M	4731	6.0.0
NI_OLD_1	M37	NI	19M	NI_19M	5578	6.0.0
NI_OLD_2	M37	NI	19M	NI_19M	5754	6.0.0
NI_OLD_3	M43	NI	26M	NI_26M	2550	7.1.0
NI_OLD_4	M43	NI	26M	NI_26M	3274	7.1.0
IR_CONV_10GY_19M_1	M37	10Gy	19M	IR_CONV_10Gy_19M	5073	6.0.0
IR_CONV_10GY_19M_2	M37	10Gy	19M	IR_CONV_10Gy_19M	6579	6.0.0
IR_CONV_10GY_19M_3	M37	10Gy	19M	IR_CONV_10Gy_19M	8484	6.0.0

ANNEXE IV – An interactive murine single-cell atlas of the lung responses to irradiation injury (*publication*) – supplementary figures.

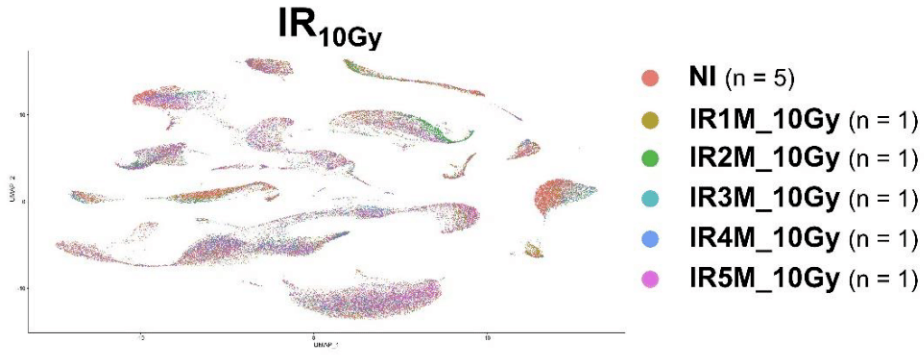
An interactive murine single-cell atlas of the lung responses to radiation injury

SUPPLEMENTARY FIGURES

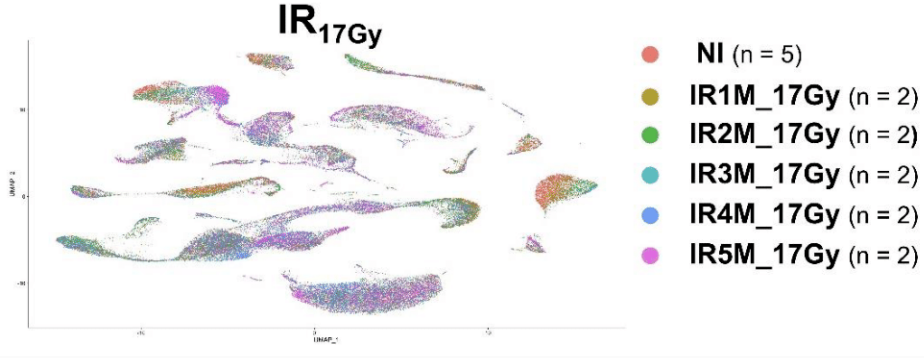
a

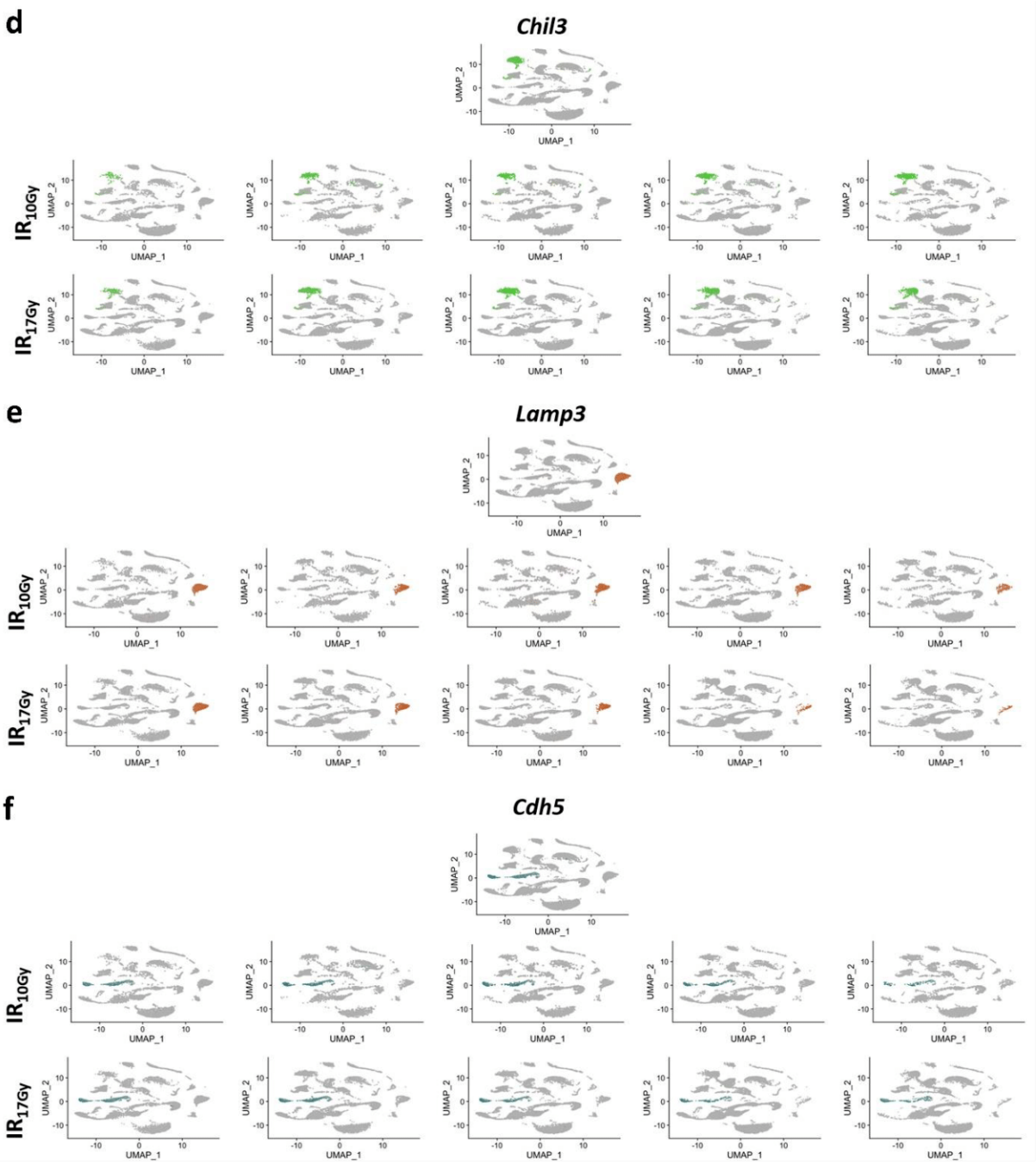


b

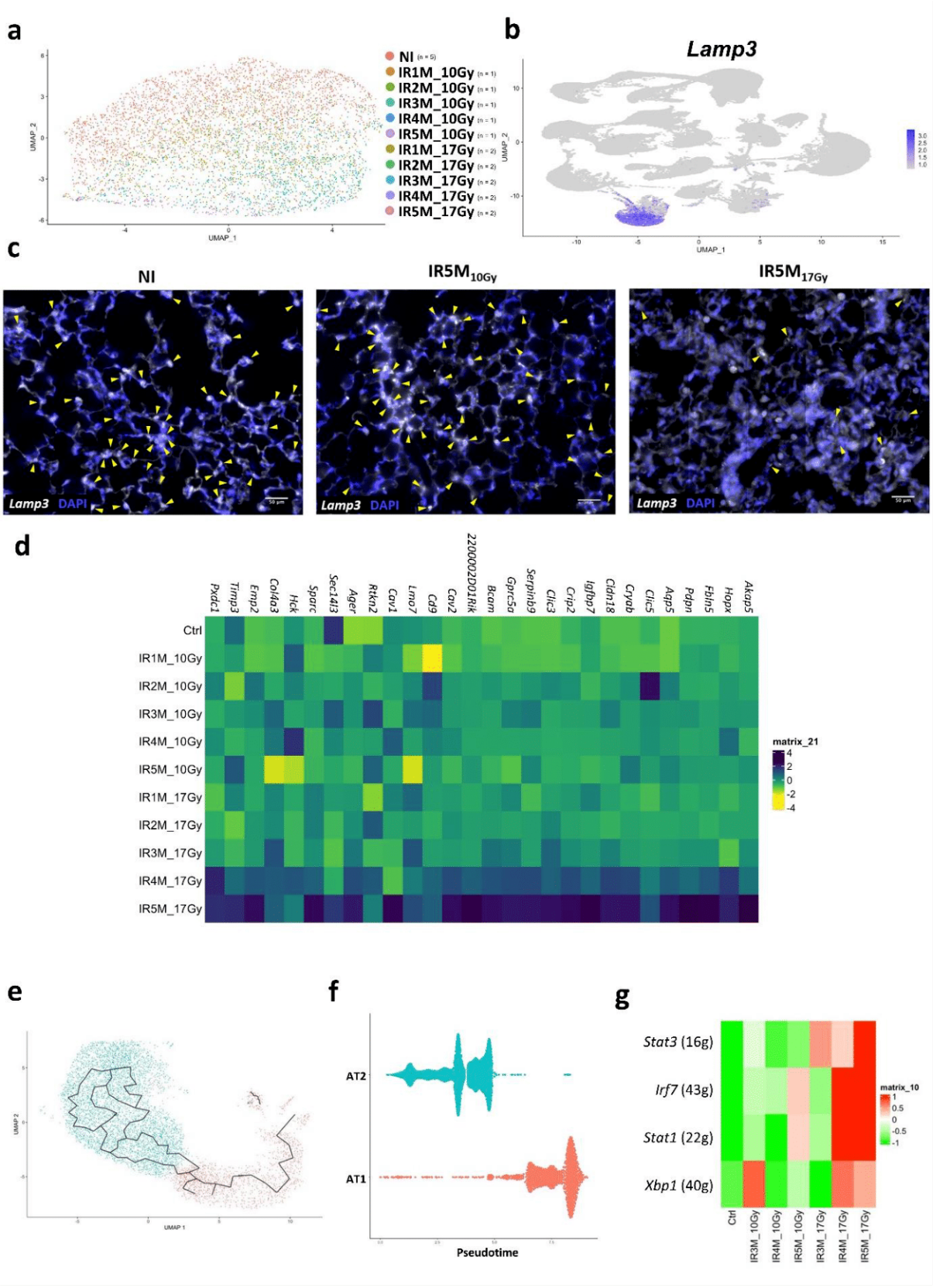


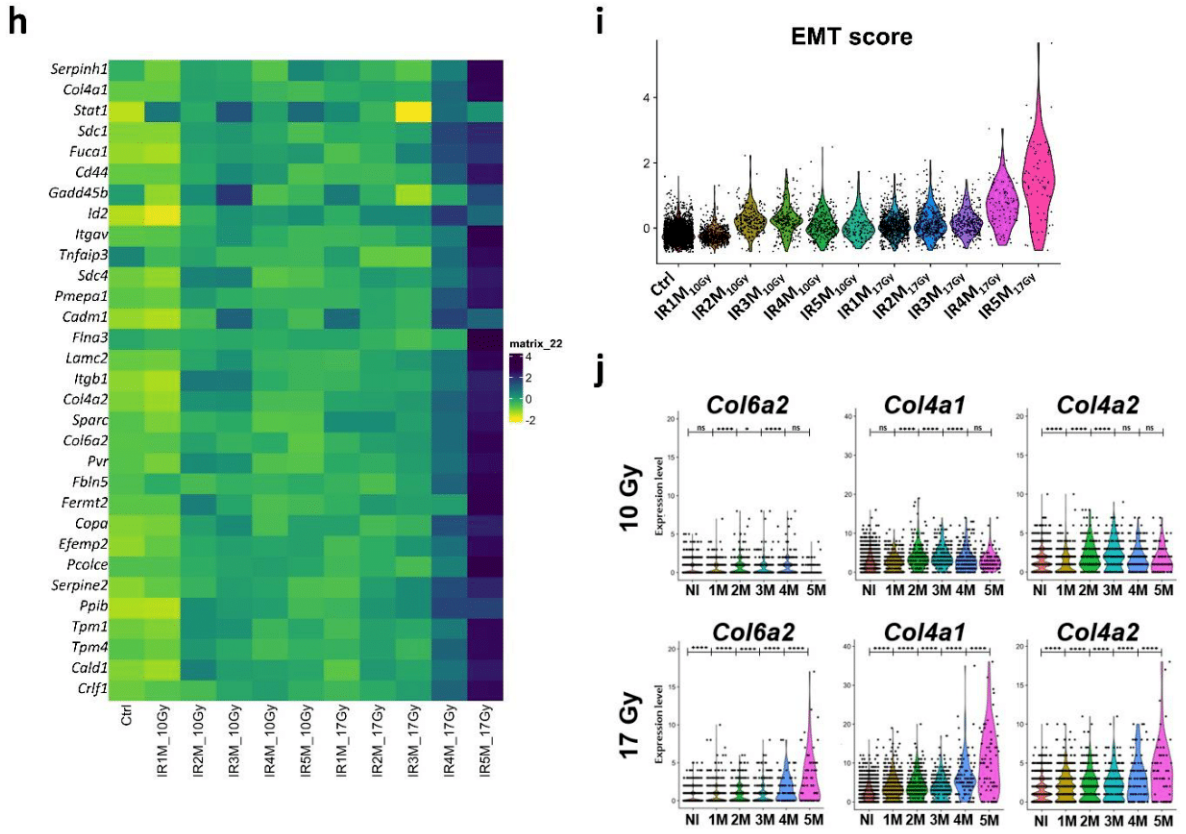
c



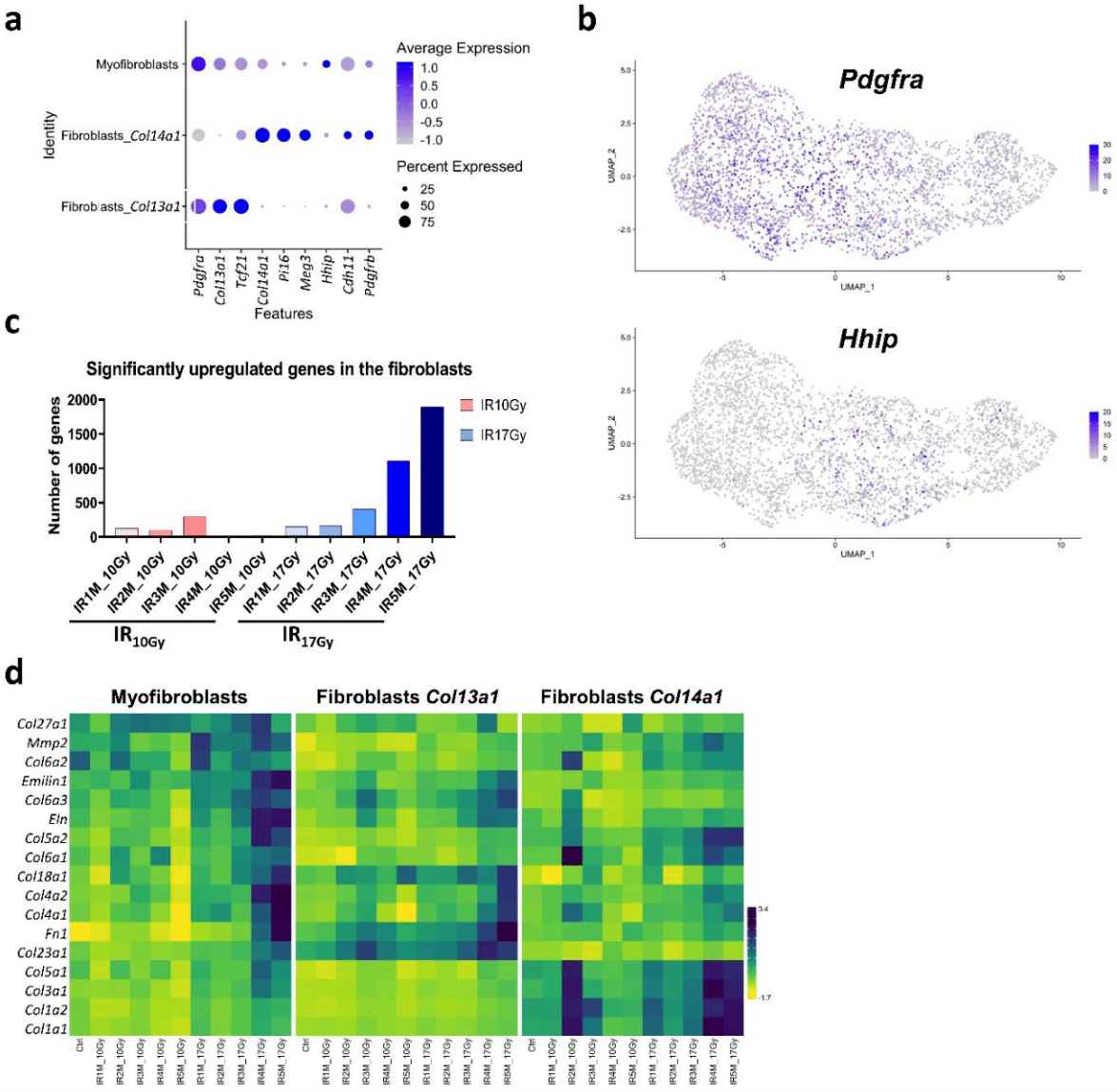


Supplementary Figure 1. Single-cell data visualization of the NI and IR lungs and cell type specific markers identification. UMAP visualization from **a** 5 NI samples alone (22,378 cells), **b** 5 NI samples and 5 IR_{10Gy} samples from 1 to 5M (one sample per time point) (26,360 cells) and **c** 5 NI samples and 10 IR_{17Gy} samples from 1 to 5M (two samples per time point) (54,131 cells). UMAP visualization of *Chil3* **d**, *Lamp3* **e** and *Cdh5* **f** expression in the NI samples and the different time points from 1 to 5 months after IR_{10Gy} (n =1) and IR_{17Gy} (n=2).

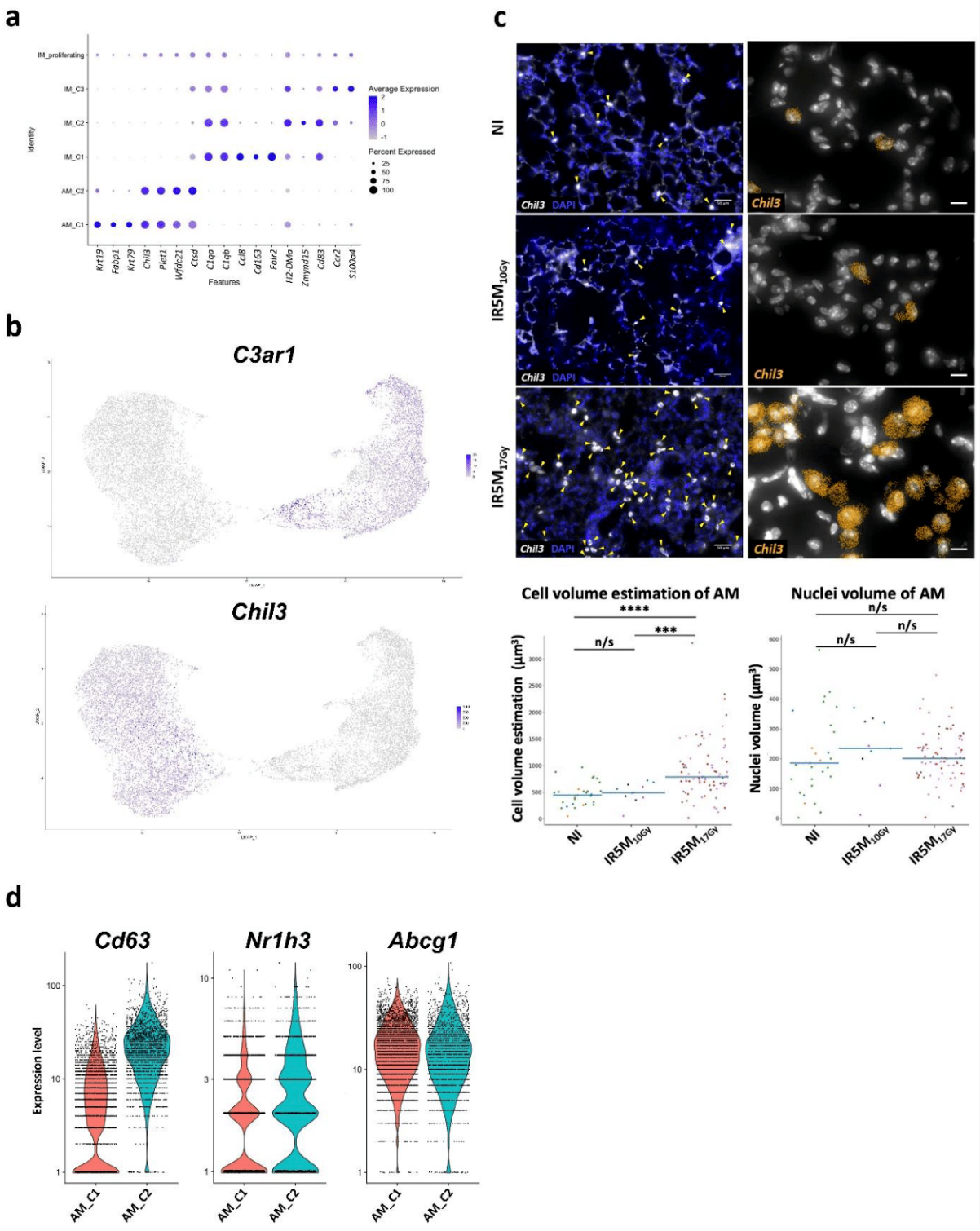

a
NI (n = 5)
IR1M_10Gy (n = 1)
IR2M_10Gy (n = 1)
IR3M_10Gy (n = 1)
IR4M_10Gy (n = 1)
IR5M_10Gy (n = 1)
IR1M_17Gy (n = 2)
IR2M_17Gy (n = 2)
IR3M_17Gy (n = 2)
IR4M_17Gy (n = 2)
IR5M_17Gy (n = 2)
UMAP_1
UMAP_2
b
Lamp3
c
NI
IR5M10Gy
IR5M17Gy
Lamp3 DAPI
d
Pxdc1
Timp3
Emp2
Col4a3
Htk
Sparc
Sec14l3
Ager
Rtkn2
Cav1
Lmo7
Cd9
Cav2
2200002D01Rik
Bcam
Gprc5a
Serpinb9
Clic3
Crip2
Igfbp7
Cldn18
Cryab
Clic5
Aqp5
Pdpn
Fbln5
Hopx
Akap5
Ctrl
IR1M_10Gy
IR2M_10Gy
IR3M_10Gy
IR4M_10Gy
IR5M_10Gy
IR1M_17Gy
IR2M_17Gy
IR3M_17Gy
IR4M_17Gy
IR5M_17Gy
matrix_21
e
f
AT2
AT1
Pseudotime
g
Stat3 (16g)
Irf7 (43g)
Stat1 (22g)
Xbp1 (40g)
matrix_10
Ctrl
IR3M_10Gy
IR4M_10Gy
IR5M_10Gy
IR3M_17Gy
IR4M_17Gy
IR5M_17Gy



Supplementary Figure 2. Cellular and molecular characterization of AT2 cells. a UMAP visualization of the 5,550 NI (n = 5), IR_{10Gy} (n = 1) and IR_{17Gy} (n = 2) AT2 cells annotated by time point. **b** UMAP visualization of the expression of *Lamp3*. **c** *Lamp3* (white) staining in lung tissue sections from NI (n = 3), $IR5M_{10Gy}$ (n = 3) and $IR5M_{17Gy}$ (n = 5) mice. Sections were counterstained with DAPI (blue). Images are shown as a maximum intensity projection (16 z-stacks, 5 µm). Images were acquired using the tiles tool (5x5) on an apotome microscope with a 63X objective. Yellow arrows point at AT2 cells. Scale bars, 50 µm. **d** Heatmap of the expression of transdifferentiation related genes in the AT2 cells across the different samples. **e** Monocle trajectory analysis of the AT2 and AT1 clusters. **f** Monocle pseudotime analysis of the AT2 and AT1 cells using the marker genes of the transdifferentiating AT2 cluster 3 to order the cells. **g** Gene Regulatory network analysis of the AT2 cells in the NI samples and 3, 4 and 5 months after IR_{10Gy} and IR_{17Gy}. **h** Heatmap of the expression of EMT genes in the AT2 cells across the different samples. **i** Violin plot showing the single cell score calculated based on the EMT expressed genes in the AT2 cells. **j** Violin plots of EMT genes expression in the AT2 cells in the NI samples and at the different time points after IR_{10Gy} and IR_{17Gy}. (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).

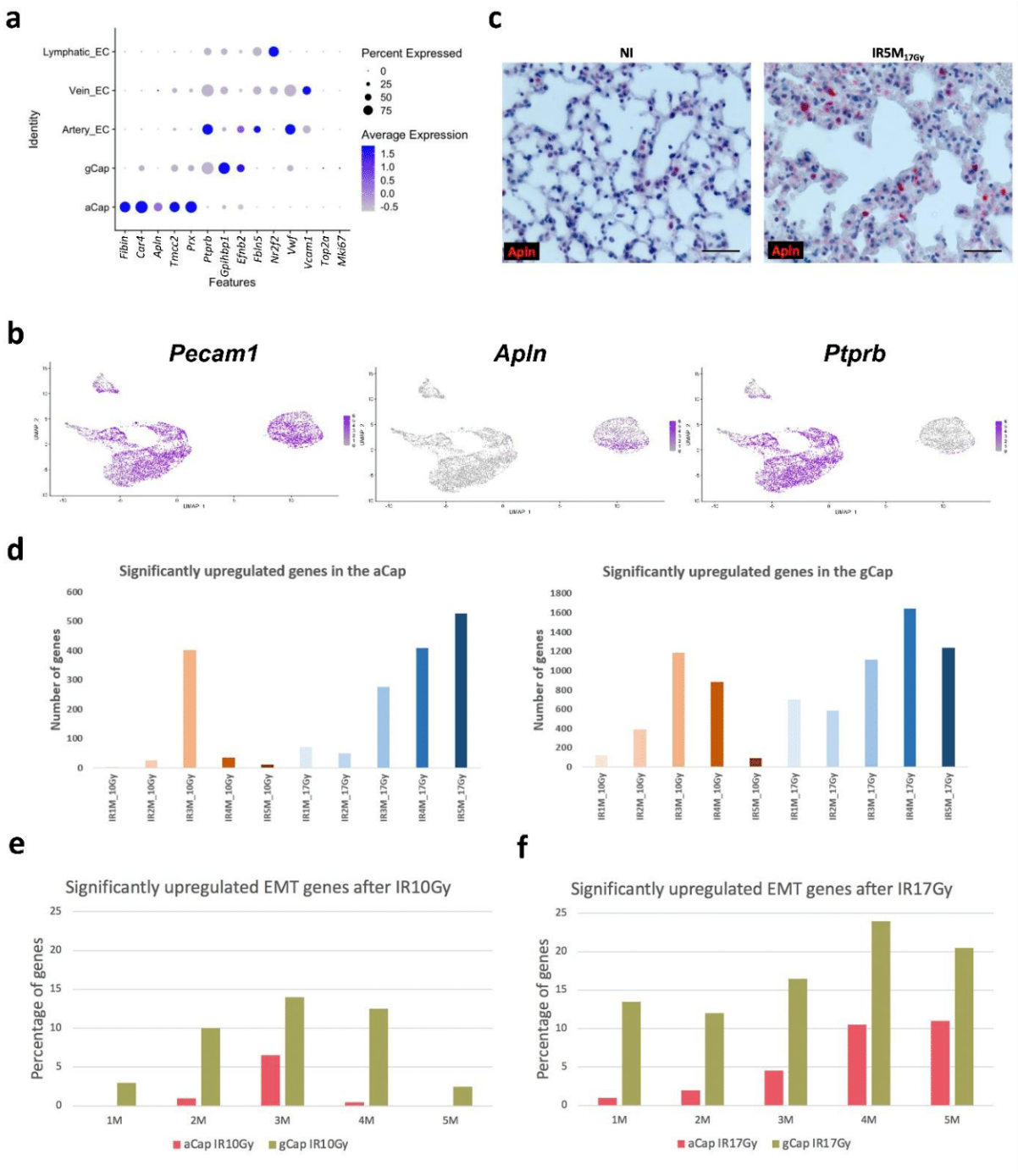


Supplementary Figure 3. Molecular characterization of the mesenchymal cells. a DotPlot of the expression of the marker genes used to identify the fibroblasts subpopulations: *Col13a1* and *Tcf21* for *Col13a1*+ fibroblasts; *Col14a1*, *Pi16* and *Meg3* for *Col14a1*+ fibroblasts; *Hhip*, *Cdh11* and *Pdgfrb* for myofibroblasts. **b** UMAP visualization of the expression of *Pdgfra* and *Hhip*. **c** Dynamics in the significantly upregulated genes in the fibroblast subpopulations compared to the NI samples at the different time points after IR10Gy and IR17Gy. **d** Heatmap of the expression of ECM related genes in the fibroblast subpopulations across the different samples.

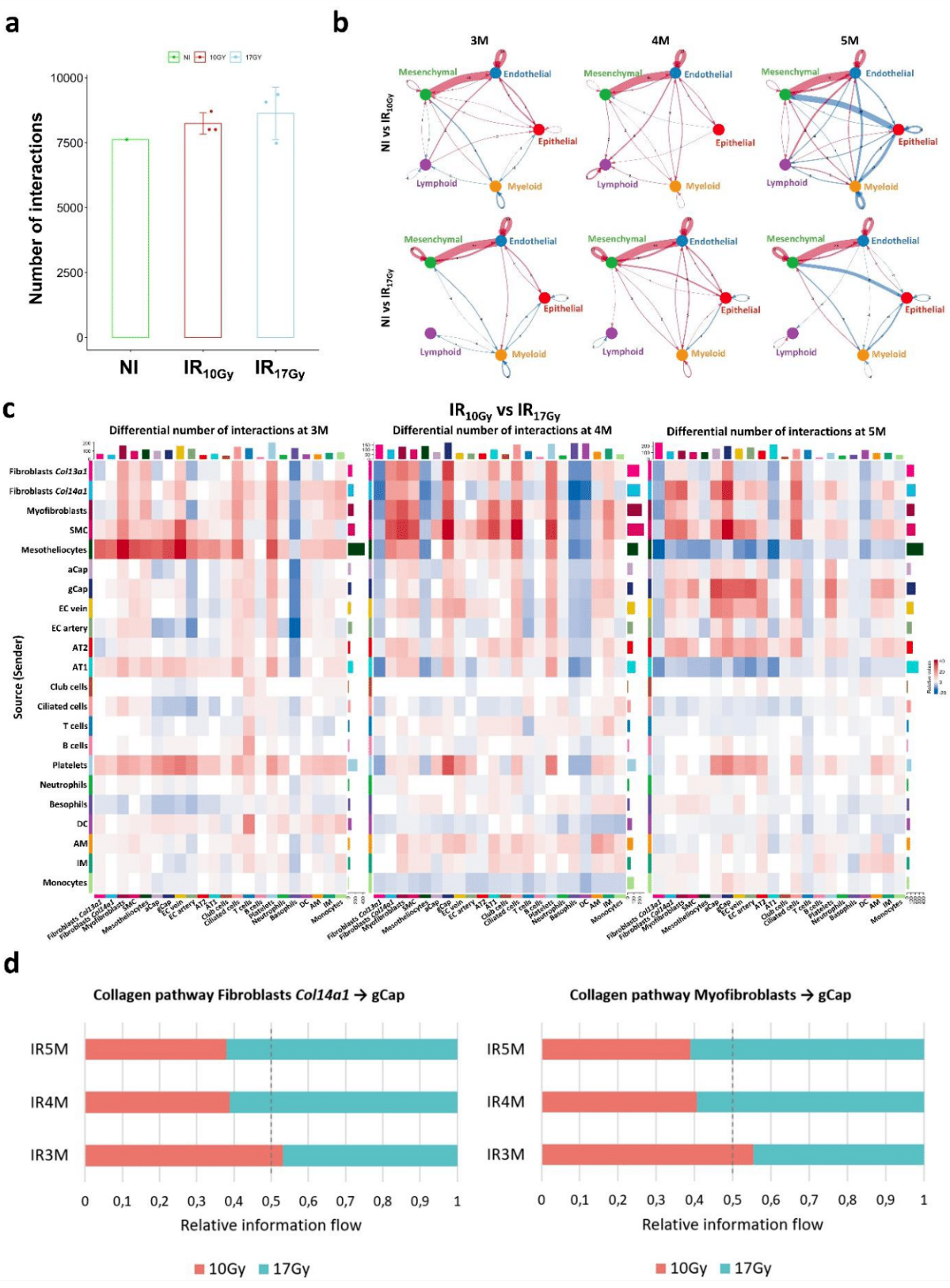


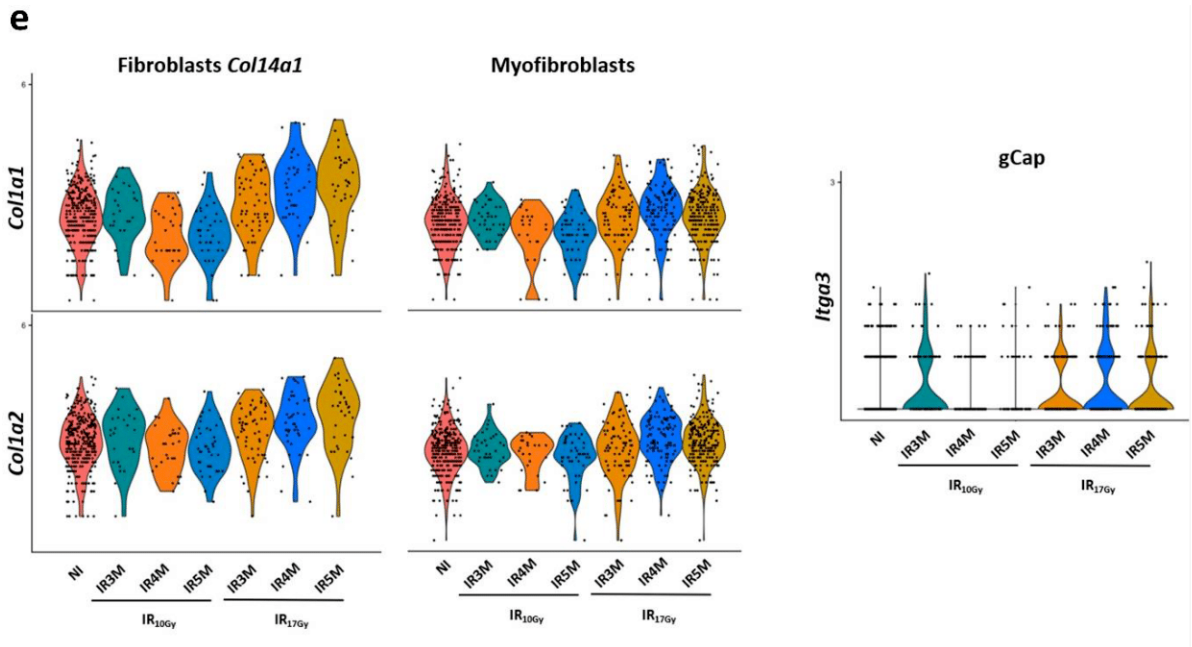
Supplementary Figure 4. Cellular and molecular characterization of the alveolar and interstitial macrophages. a DotPlot of the expression of the marker genes used to identify the different IM and AM subpopulations. **b** UMAP visualization of the expression of *C3ar1* and *Chil3*. **c** *Chil3* staining in lung tissue sections from NI, IR5M$_{10Gy}$ and IR5M$_{17Gy}$ mice. Left column: sections were counterstained with DAPI (blue). Images are shown as a maximum intensity projection (16 z-stacks, 5 µm). Images were acquired using the tiles tool (5x5) on an apotome microscope with a 63X objective. Yellow arrows point at AM. Scale bars, 50 µm. Right column: automatic *Chil3* mRNA detection with Big-FISH. Scale bars, 10 µm. Cell and nuclei volume estimation of the *Chil3*+ cells in NI, IR5M$_{10Gy}$ and IR5M$_{17Gy}$ 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$_{10Gy}$ n = 3; IR5M$_{17Gy}$ n = 5). **d** Violin plots of foam genes expression in the different AM subpopulations.

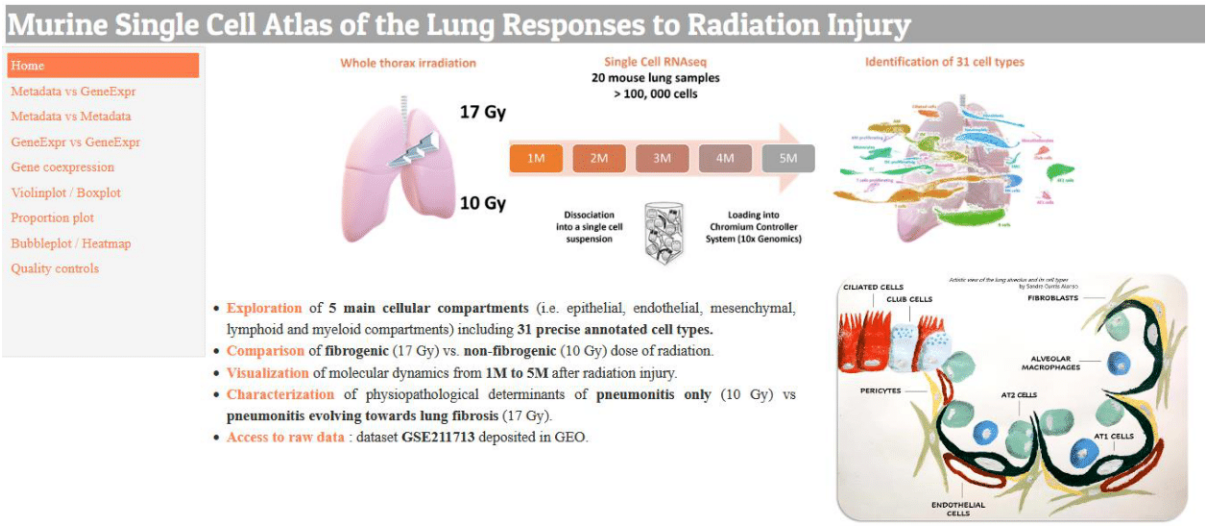


Supplementary Figure 5. Molecular profile of the different endothelial cell sub-compartments. a DotPlot of the expression of the marker genes used to identify the different EC subpopulations. **b** UMAP visualization of the expression of *Pecam1*, *Apln* and *Ptprb*. **c** Immunohistochemistry staining in NI (n = 3) and IR5M$_{17Gy}$ (n = 3) lung tissue sections using an anti-Apln antibody (red). Scale bars, 50 µm. **d** Dynamics in the significantly upregulated genes in the aCap and gCap compared to the NI samples at the different time points after IR$_{10Gy}$ and IR$_{17Gy}$. Percentage of the significantly upregulated EMT genes at the different time points after IR$_{10Gy}$ **e** and IR$_{17Gy}$ **f** in the aCap and gCap.


a
NI
10GY
17GY
Number of interactions
10000
7500
5000
2500
0
NI
IR10Gy
IR17Gy
b
3M
4M
5M
NI vs IR10Gy
NI vs IR17Gy
Mesenchymal
Endothelial
Epithelial
Lymphoid
Myeloid
c
IR10Gy vs IR17Gy
Differential number of interactions at 3M
Differential number of interactions at 4M
Differential number of interactions at 5M
Source (Sender)
Fibroblasts Col13a1
Fibroblasts Col14a1
Myofibroblasts
SMC
Mesotheliocytes
aCap
gCap
EC vein
EC artery
AT2
AT1
Club cells
Ciliated cells
T cells
B cells
Platelets
Neutrophils
Basophils
DC
AM
IM
Monocytes
d
Collagen pathway Fibroblasts Col14a1 → gCap
Collagen pathway Myofibroblasts → gCap
IR5M
IR4M
IR3M
0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1
Relative information flow
10Gy
17Gy



Supplementary Figure 6. Cell Chat interaction analysis of the major cell compartments. **a** Mean number of interactions in the NI samples (n=5 mice lungs were pulled together for the analysis) and the 3M, 4M and 5M IR_{10Gy} (n=3 mice lung) and IR_{17Gy} (n=6 mice lung; 2 samples per time point were pulled together for the analysis) samples. Error bars refer to the standard deviation of the data. **b** Circle plot showing the differential number of interactions between IR_{10Gy} and NI and IR_{17Gy} and NI in the main cellular compartments at 3M, 4M and 5M post-IR: mesenchymal, endothelial, epithelia, myeloid and lymphoid. Red (or blue) colored edges represent increased (or decreased) signaling in the IR compared to the NI. **c** Heatmap showing the differential number of interactions between IR_{10Gy} and IR_{17Gy} in all the different lung subpopulations at 3M, 4M and 5M post-IR. Red (or blue) represents increased (or decreased) signaling in the IR_{17Gy} compared to the IR_{10Gy}. 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). **d** Dynamics of the relative information flow of the Collagen pathway from 3 to 5 months post-IR from the Fibroblasts *Col14a1* and Myofibroblasts to the gCap. **e** Gene expression distribution of signaling genes related to the Collagen pathway in the NI samples and 3M, 4M and 5M after IR_{10Gy} and IR_{17Gy}: *Col1a1* and *Col1a2* ligands in Fibroblasts *Col14a1* and Myofibroblasts; *Itga3* receptor in gCap.



Supplementary Figure 7. Interactive Mouse Radio-induced Pulmonary Fibrosis Atlas webpage. Outline of the homepage of the interactive webpage containing the open-access transcriptomic data for all the scientific community. Bottom left image is an artistic view of the alveolus realized by Sandra Currás-Alonso. The website has been designed by Sophie Heinrich using the *ShinyCell* package.

ANNEXE V – Tutorials to start in single cell RNA seq data analysis.

Single cell RNA sequencing data analysis can be challenging at first, with the variety of tools and methods that exist. Therefore, I wrote tutorials for beginners to more easily start the analysis, with the different steps, why they are done, what are some of the alternatives and links to additional resources to better understand the different scripts. These tutorials were shared through the Bioinfo-Hub of Curie Institute, a grouping of the Institute bioinformaticians that aims to share and mutualize knowledge and experience on bioinformatic and biostatistics.

1) Pre-processing of the data
2) Merging of several samples
3) Integration of several samples to correct batch effect
4) Cell populations identification
5) Visualization and simple analysis
6) Trajectory analysis with Monocle3

Tutorial 1 : 1st step of preprocessing

Juliette SOULIER

4/16/2024

This document is presenting the fist step of preprocessing of 10X chromium single cell RNA sequencing data. 10X chromium provides a fluidic droplet based single cell RNA sequencing method. This tutorial presents the analysis of 10X single cell RNA seq samples sequenced with Illumina and after cellranger alignment, filtering, barcode counting, and UMI counting.

Some part of the code of this tutorial needs to be adjusted to your data before running it (for example to indicate the folder where you stored the data). This will be indicated in **bold**.

Download the data

Before starting the analysis, the fist thing to do is download your data. This can be done in several ways: you may already have your own data, or you can download it from a database, like the gene expression omnibus (GEO) database. You can find here the data from the paper published by Curras-Alonso et al. (2023, Nature Communication) about the lung responses to radiation injury in mice. For this tutorial, we will be using the samples *Ctrl_1*, *IR_17Gy_5M_1* and *IR_17Gy_5M_2*: we will analyse one sample from non irradiated mouse lungs, and two duplicate samples from mice that have been irradiated at 17Gy in the full thorax, the samples were taken five months post irradiation. For each sample, you need to retrieve the *barcodes.tsv.gz*, *count_matrix.mtx.gz* and *genes.tsv.gz* files.

Install the necessary packages

Packages are collection of tools created for specific analysis. To process single cell RNA sequencing data, we will need to use some of them. First the packages need to be installed. This only needs to be done one for a given computer.

Seurat package

Seurat is a popular R tool to perform quality controls and data exploration on single cell RNA sequencing data. It is also usefull to install SeuratWrappers, a useful package enabling the use of different packages with Seurat objects, and SeuratData, another support package for Seurat.

```
remotes::install_github("satijalab/seurat", "seurat5", quiet = TRUE)
remotes::install_github("satijalab/seurat-wrappers", "seurat5", quiet = TRUE)
remotes::install_github("satijalab/seurat-data", "seurat5", quiet = TRUE)
```

SoupX package

SoupX is used to remove contaminating mRNA coming from mRNA floating in the medium and encapsulated with cells during the preparation of the samples.

```
install.packages('SoupX')
```

scran package

scran implements some analysis methods for single cell RNA sequencing analysis, for example assignment of cell cycle phase, detection of highly variable and significantly correlated genes, identification of marker genes...

```
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager"
BiocManager::install("scran")
```

Load the necessary packages

At the beginning of a new session, you need to load the packages you want to use, to have access to the functions stored in these packages.

```
library(Seurat)
library(scran)
```

choosing of the working directory

The working directory is the directory were you data are stored, where the results will be saved. You can choose it at the beggining of each script. **Replace the path by the path of your workind firectory in your computer**.

```
setwd("/Users/labo/Desktop/Tutos_scRNA_seq_R")
```

Load the data

Now you have everything to start the preprocessing of your single cell RNA sequencing data. The steps described in this tutorial needs to be done for each sample you want to analyse, one sample at a time. Fist you need to upload your data in R. To do so you need to put the path of the *filtered_feature_bc_matrix* folder of the sample you want to analyse. This folder contains the three files : *barcodes.tsv.gz*, *features.tsv.gz* and *matrix.mtx.gz* (the files must be named like that). **Replace the path in the Read10X function by the path of the filtered_feature_bc_matrix folder in your computer**.

```
data <- Read10X(data.dir = "data_example/Ctrl_1/filtered_feature_bc_matrix")
```

SoupX filtering

This is an optional steps that will not be run on the data of this tutorial. During the dissociation and the preparation of the sample, before the encapsulation in droplets, some cells die and release their cytoplasmic content in the medium. Therefore when the cells are encapsulated some floating mRNA from dead cells are also encapsulated and are labelled like belonging to this cell. SoupX uses the content of the empty droplets to infer the composition of the "soup" and computationally remove it from all the droplets.

First SoupX requires to create a Seurat object and perform clustering of the data.

```
#library(SoupX)
#srat <- CreateSeuratObject(counts = data)
#srat[["percent.mt"]] <- PercentageFeatureSet(srat, pattern = "^mt-")
#srat <- SCTransform(srat, vars.to.regress = "percent.mt", verbose = T)
#srat <- RunPCA(srat, npcs = 20, verbose = T)
#srat <- RunUMAP(srat, reduction = "pca", dims = 1:20)
#srat <- FindNeighbors(srat, reduction = "pca", dims = 1:20)
#srat <- FindClusters(srat, resolution = 0.5)
```

Then you can run the actual SoupX part. As SoupX decontamination is based on the analysis of the empty droplets, it needs to take as an input both *filtered_feature_bc_matrix* and *raw_feature_bc_matrix* folders. To do so, you need to **put the parent folder containing both of the *filtered_feature_bc_matrix* and *raw_feature_bc_matrix* folders in the argument of the load10X function**.

```
#sc <- load10X("data_example/Ctrl_1")
#sc <- setClusters(sc, srat@active.ident)
```

One parameter that can be adjusted is the contamination fraction. By default it is set to 0.3, which means that we consider that 30% of the sequenced mRNA are contaminating mRNA from the soup.

```
#sc <- setContaminationFraction(sc, 0.3)
#genes_QC_SoupX <- rownames(head(sc$soupProfile[order(sc$soupProfile$est,decreasing=TRUE),],n=20))
#data <- adjustCounts(sc)
```

```
#print(genes_QC_SoupX)
```

genes_QC_SoupX will give you the most highly expressed genes in the calculated soup. The adjusted counts after the removal of the soup are stored in the *out* object and that is what we will be using for the next parts of this analysis.

Creation of the Seurat object

Now that the soup contamination has been filtered out, we can create the final Seurat object. The *min.cells* argument determine the minimum number of cells a gene must be expressed in to be kept in the data. The *min.features* argument allows to exclude cells with few mRNA molecules sequenced to remove the empty droplets from the object.

```
seurat <- CreateSeuratObject(counts = data, min.cells = 3, min.features = 200)
```

Definition of the metadata

Defining the metadata of the sample analyzed is a very important step. First, you need to put a unique identifier (here *orig.ident*) to be able to distinguish every sample from each other when all the samples will be put together in one object.

Then you can put information about the experimental conditions, for example the *dataset*, *project*, *experiment*, *singlecell_version*, *sexe*...

Furthermore, during the analysis of the data it can be useful to regroup the data according to certain criteria, like the time point after treatment, the genotype, the type of treatment... If you want to analyse these elements (example *genotype*, *IRstatus*, *IRdose*, *timepoint*, *sorted*).

Finally, it can be useful to have a name that regroups all these information to be able to group the duplicates. Here we stored this information in the *condition* metadata item.

The names, the number and the content of the metadata information must be adapted to your sample. To add a new metadata item, you can follow this example : seurat$metadata_slot_name <- "information_about_sample"

```
seurat$orig.ident <- "Ctrl_1"
seurat$IRstatus <- "NI"
seurat$condition <- "Ctrl"
```

Calculation of the percentage of mitochondrial genes per cell

The name of all the mitochondrial genes in mouse starts by *mt-*. This line of code calculates for each cell the percentage of transcripts that are from mithochondrial genes. **If this tutorial is done with human samples, *mt-* needs to be replaced by *-MT*.**

```
seurat[["percent.mt"]] <- PercentageFeatureSet(seurat, pattern = "^mt-")
```

Filter out the cells with a high percentage of mitochondrial genes

A high percentage of mitochondrial genes in a cell is an indicator of apoptotic, stressed and low-quality cells. Therefore we filter out the cells with the highest percentage of mitochondrial genes. Here we set the threshold to the mean of percentage of mitochondrial genes + two standard deviation, but **it can also be set to a fixed number (example 20%)**.

```
max.mt <- mean(seurat$percent.mt) + 2*sd(seurat$percent.mt)
seurat <- subset(seurat, subset = percent.mt <= max.mt)
```

Normalization of the data

A lot of technical factors can affect the data, for example the sequencing depth, the number of molecules detected in each cells... To remove the technical variability we perform normalization of the data. Here we choose the SC Transform normalization, but other methods can be used.

```
seurat <- SCTransform(seurat, vars.to.regress = "percent.mt", verbose = T)
```

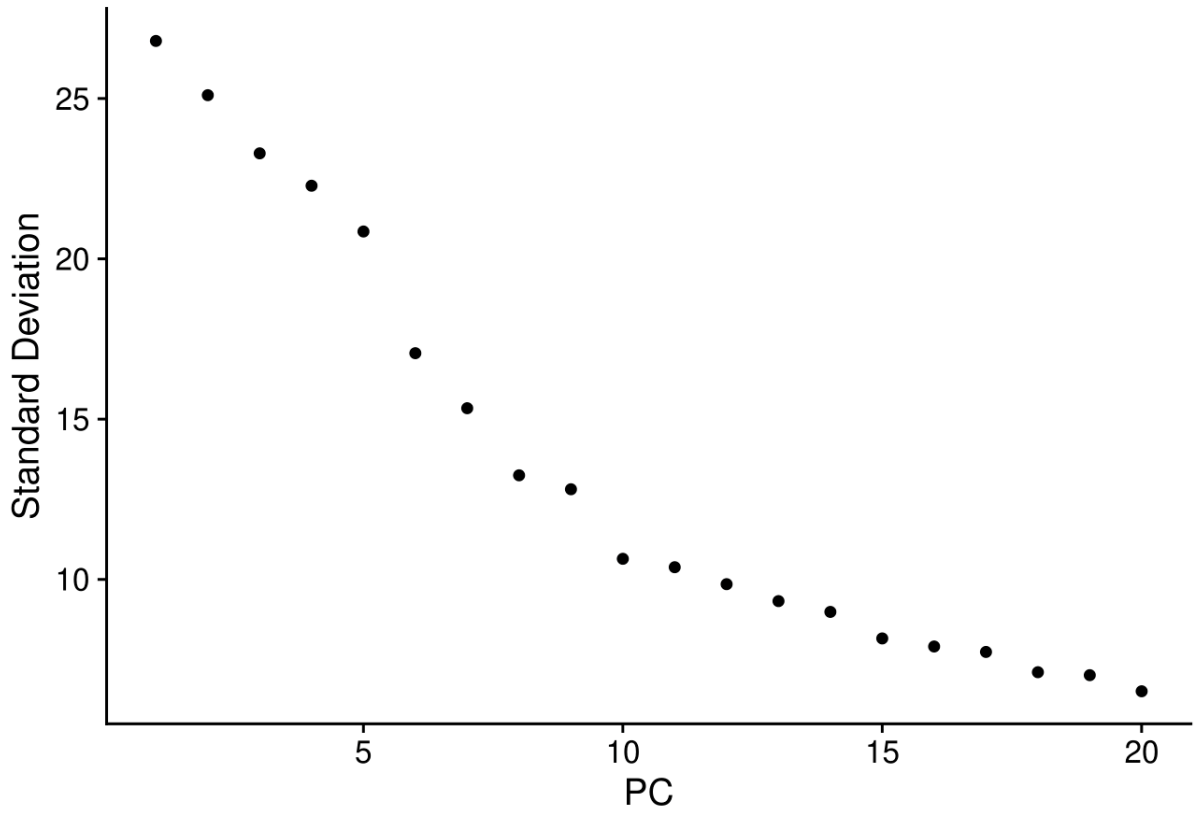
Dimension reduction

A single cell RNA sequencing dataset is an object with a very important number of dimensions. In order to compute the visualisation, the cluster calculation, we need to work with a lower number of dimensions. To do so, the function *RunPCA* is used to run the principal component analysis. The number of principal component (here set to 20) can be adjusted.

```
seurat <- RunPCA(seurat, npcs = 20, verbose = T)
```

An elbow plot can then be used to verify that the number of principal component chosen recapitulates most of the variability of the dataset.

```
ElbowPlot(seurat, ndims = 20, reduction = "pca")
```



Visualization coordiates calculation

As explained above, a single cell RNA sequencing dataset is an object with a very important number of dimensions. We often visualize the data in two dimensions, so we need to calculate the coordinates of each cell in this two-dimension space in order to find the best way to recapitulate the similarities and differences

between the transcriptional profiles of the cells. Various methods can be used to do that. Here we use UMAP, but methods like t-SNE are also often used.

```
seurat <- RunUMAP(seurat, reduction = "pca", dims = 1:20)
```

Clusters calculation

A very important point of the analysis is the calculation of the clusters. With the *FindNeighbors* and *FindClusters*, Seurat classifies the cells in several groups based on their similarities. This will be used later to identify the different cell types.

```
seurat <- FindNeighbors(seurat, reduction = "pca", dims = 1:20)
seurat <- FindClusters(seurat, resolution = 0.5)
```

```
seurat
```

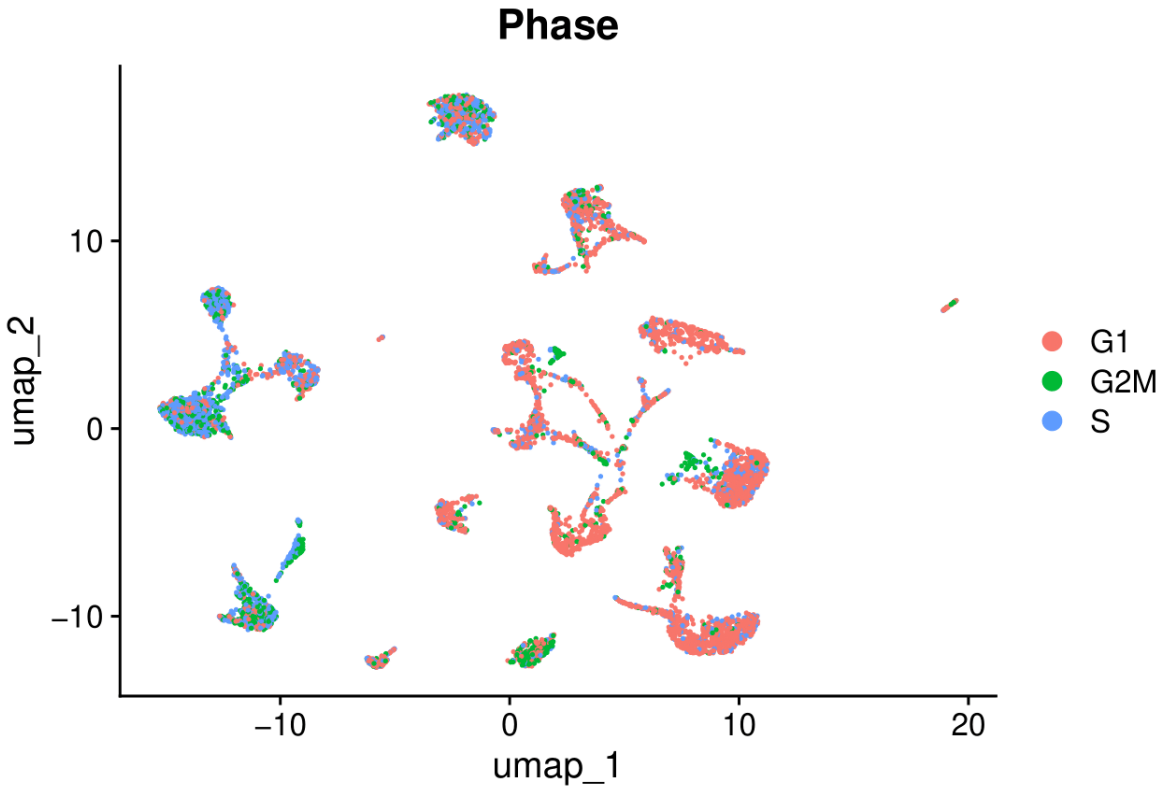
```
## An object of class Seurat
## 37338 features across 6751 samples within 2 assays
## Active assay: SCT (18137 features, 3000 variable features)
##  3 layers present: counts, data, scale.data
##  1 other assay present: RNA
##  2 dimensional reductions calculated: pca, umap
```

Assignation of cell cycle scores

Finally, we use a set of genes provided by Seurat to identify the phase of the cell cycle for each cell.

```
s.genes <- cc.genes$s.genes
g2m.genes <- cc.genes$g2m.genes
seurat <- CellCycleScoring(seurat, s.features = s.genes, g2m.features = g2m.genes)
```

```
DimPlot(seurat, reduction = "umap", label = F, group.by = "Phase")
```



In this plot, each dot represents one cell, and the cells are colored according to their phase in the cell cycle.

Saving of the preprocessed sample

The first step of preprocessing is over, now you can **save your sample in a chosen folder** and process another sample, or go to the next part of the analysis.

```
saveRDS(seurat, "processed_objects/Ctrl_1.rds")
```

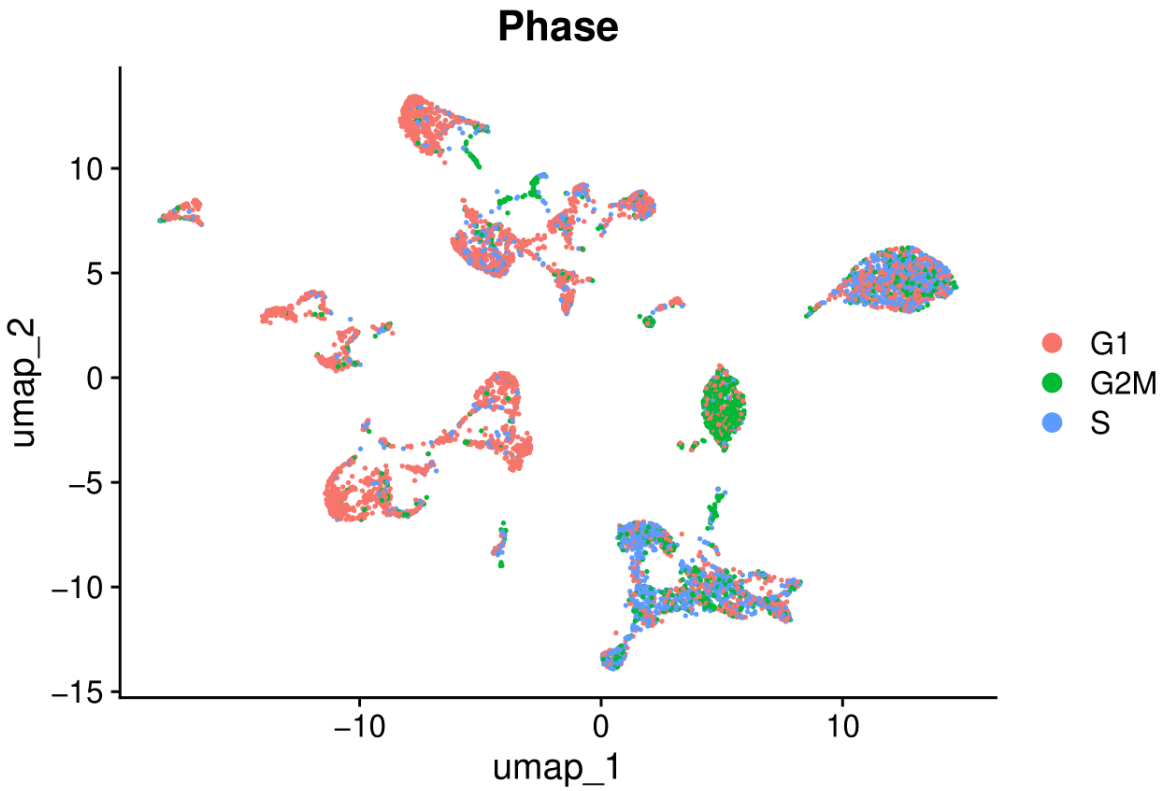
Preprocessing of the other samples

This process needs to be repeated for the other samples that you wish to analyse. Here, we will repeat it for the *IR_17Gy_5M_1* and *IR_17Gy_5M_2* samples.

First, the *IR_17Gy_5M_1* sample:

```
data <- Read10X(data.dir = "data_example/IR_17Gy_5M_1/filtered_feature_bc_matrix")
seurat <- CreateSeuratObject(counts = data, min.cells = 3, min.features = 200)

seurat$orig.ident <- "IR_17Gy_5M_1"
seurat$IRstatus <- "IR_17Gy"
seurat$condition <- "IR_17Gy_5M"
```

```
seurat[["percent.mt"]] <- PercentageFeatureSet(seurat, pattern = "^mt-")
max.mt <- mean(seurat$percent.mt) + 2*sd(seurat$percent.mt)
seurat <- subset(seurat, subset = percent.mt <= max.mt)

seurat <- SCTransform(seurat, vars.to.regress = "percent.mt", verbose = T)

seurat <- RunPCA(seurat, npcs = 20, verbose = T)
seurat <- RunUMAP(seurat, reduction = "pca", dims = 1:20)
seurat <- FindNeighbors(seurat, reduction = "pca", dims = 1:20)
seurat <- FindClusters(seurat, resolution = 0.5)

s.genes <- cc.genes$s.genes
g2m.genes <- cc.genes$g2m.genes
seurat <- CellCycleScoring(seurat, s.features = s.genes, g2m.features = g2m.genes)

DimPlot(seurat, reduction = "umap", label = F, group.by = "Phase")
```

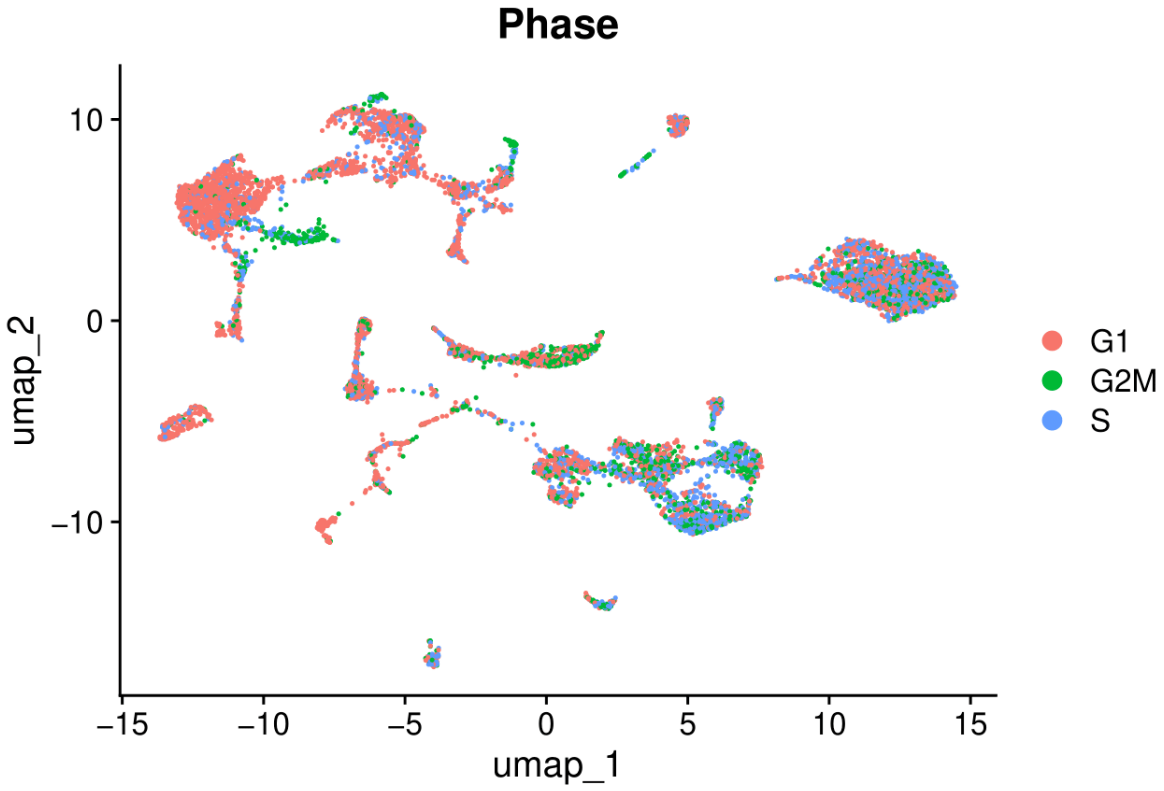



```
saveRDS(seurat, "processed_objects/IR_17Gy_5M_1.rds")
```

Then, the *IR_17Gy_5M_2* sample:

```
data <- Read10X(data.dir = "data_example/IR_17Gy_5M_2/filtered_feature_bc_matrix")
seurat <- CreateSeuratObject(counts = data, min.cells = 3, min.features = 200)
```

```
seurat$orig.ident <- "IR_17Gy_5M_2"
seurat$IRstatus <- "IR_17Gy"
seurat$condition <- "IR_17Gy_5M"

seurat[["percent.mt"]] <- PercentageFeatureSet(seurat, pattern = "^mt-")
max.mt <- mean(seurat$percent.mt) + 2*sd(seurat$percent.mt)
seurat <- subset(seurat, subset = percent.mt <= max.mt)

seurat <- SCTransform(seurat, vars.to.regress = "percent.mt", verbose = T)

seurat <- RunPCA(seurat, npcs = 20, verbose = T)
seurat <- RunUMAP(seurat, reduction = "pca", dims = 1:20)
seurat <- FindNeighbors(seurat, reduction = "pca", dims = 1:20)
seurat <- FindClusters(seurat, resolution = 0.5)

s.genes <- cc.genes$s.genes
g2m.genes <- cc.genes$g2m.genes
seurat <- CellCycleScoring(seurat, s.features = s.genes, g2m.features = g2m.genes)

DimPlot(seurat, reduction = "umap", label = F, group.by = "Phase")
```




```
saveRDS(seurat, "processed_objects/IR_17Gy_5M_2.rds")
```

Tutorial 2 : 2nd step of preprocessing: merging

Juliette SOULIER

4/16/2024

Now that the data from the samples you want to analyse has been pre processed, the next step is to group them together in one object to be able to compare them. There are two ways to do that: merge the data or integrate the data.

- Merge the data consists in only group the different samples in one object without doing any correction. It is usually the one to use if the experimental conditions for all the samples are very similar (same sequencing technique, same protocol, same strain of mice...). It preserves all the biological differences but doesn't correct for batch effect.

- Integrate the data consists in doing corrections for a batch effect due to experimental differences (samples processed in two different laboratories, different version of reagent used...). It allows to analyse together samples from the literature, or processed differently but can erase some biologically relevant processes.

Usually the samples are first merged, and if some batch effect is seen the analysis is redone by integrating the samples.
This tutorial will present how to merge samples, and the tutorial 3 will present how to integrate them.

Load the necessary package

```
library(Seurat)
```

Load the samples

The first step is to load all the preprocessed samples from the tutorial 1. **In the argument of the *readRDS* functions, put the paths to your preprocessed samples. You can add or remove samples if you need to**.

```
setwd("/Users/labo/Desktop/Tutos_scRNA_seq_R")
```

```
sample1 <- readRDS("processed_objects/Ctrl_1.rds")
sample2 <- readRDS("processed_objects/IR_17Gy_5M_1.rds")
sample3 <- readRDS("processed_objects/IR_17Gy_5M_2.rds")
```

Merge the samples into one object

Then you can merge all of these samples into one object. **The first sample goes into the first argument, and then the others in the *y* argument. In the *add.cell.ids* you need to put the unique identifier of each of your samples. And then you can name your *project* as you want.**

```
lung_merged <- merge(sample1, y = c(sample2, sample3),
                     add.cell.ids = c("Ctrl_1", "IR_17Gy_5M_1", "IR_17Gy_5M_2"), project = "RIPF")
lung_merged[["RNA"]] <- JoinLayers(lung_merged[["RNA"]])
```

Calculation of the percentage of mitochondrial genes per cell

The name of all the mitochondrial genes in mouse starts by *mt-*. This line of code calculates for each cell the percentage of transcripts that are from mithochondrial genes. **If this tutorial is done with human samples, *mt-* needs to be replaced by *-MT*.**

```
lung_merged <- PercentageFeatureSet(lung_merged, pattern = "^mt-", col.name = "percent.mt")
```

Regression of the cell cycle score

Cells that cycle can be driven to cluster together because of the expression of cell cycling genes. To avoid this, we use SCTransform to regress the expression of the cell cycle genes.

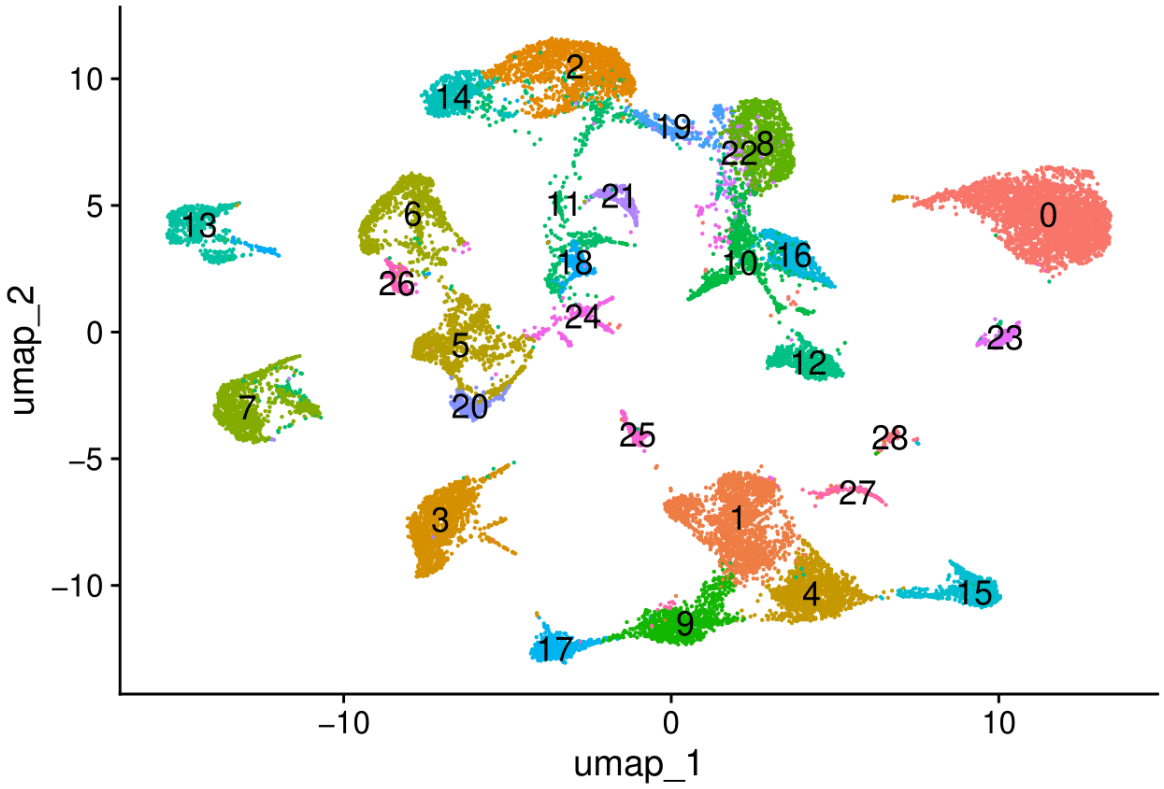
```
lung_merged$CC.Difference <- lung_merged$S.Score - lung_merged$G2M.Score
lung_merged <- SCTransform(lung_merged, vars.to.regress = c("CC.Difference", "percent.mt"), verbose = T
```

Calculating the UMAP coordinates and clusters

Then the PCA, UMAP coordinates for visualization and division into clusters are recalculated taking into account the data from all the samples.

In the *FindClusters* function, the resolution parameter allows to choose for a number more or less important of clusters. *Use a value above (below) 1.0 if you want to obtain a larger (smaller) number of communities.*

```
lung_merged <- RunPCA(lung_merged, npcs = 20, verbose = T)
lung_merged <- RunUMAP(lung_merged, reduction = "pca", dims = 1:20)
lung_merged <- FindNeighbors(lung_merged, reduction = "pca", dims = 1:20)
lung_merged <- FindClusters(lung_merged, resolution = 0.7)
```



Saving of the merged object

The second step of preprocessing is over, now you can **save your object in a chosen folder**. Then you can try the tutorial 3 to integrate your data, or go directly to the tutorial 4 to perform the cell type labeling of the clusters.

```
saveRDS(lung_merged, "processed_objects/Seurat_object_merged.rds")
```

Tutorial 3 : 2nd step of preprocessing: integrating

Juliette SOULIER

4/16/2024

This tutorial will explain you how to integrate a Seurat object. This step needs to be done only if you observe batch effect in your object, because it can erase some biologically relevant processes.

Install the necessary packages

patchwork package

patchwork is used to combine separate plots in the same graphic.

```
install.packages('patchwork')
```

Load the necessary packages

```
library(Seurat)
library(SeuratData)
library(patchwork)
```

Load the samples

The first step is to load the merged object generated with the tutorial 2. **In the argument of the *readRDS* function, put the path to your merged object**. Then in order to integrate the samples you need to split RNA layer (layers can store raw, un-normalized counts (layer='counts'), normalized data (layer='data'), or z-scored/variance-stabilized data (layer='scale.data')). Here, we split the layers by samples individually. **You need to choose to split the layers according to the metadata element you think is responsible for the batch effect. For example, you can split by the samples generated in different experiment. This is done in the *f* argument of the *split* function**.

```
setwd("/Users/labo/Desktop/Tutos_scRNA_seq_R")

lung_merged <- readRDS("processed_objects/Seurat_object_merged.rds")

lung_merged[["RNA"]] <- split(lung_merged[["RNA"]], f = lung_merged$orig.ident)
```

Integration of the samples into one object

First you need to perform a normalization and identify the variable features.

```
DefaultAssay(lung_merged) <- "RNA"

lung_merged <- NormalizeData(lung_merged)
lung_merged <- FindVariableFeatures(lung_merged)
lung_merged <- ScaleData(lung_merged)
lung_merged <- RunPCA(lung_merged)

lung_merged <- FindNeighbors(lung_merged, dims = 1:30, reduction = "pca")
lung_merged <- FindClusters(lung_merged, resolution = 2, cluster.name = "unintegrated_clusters")
```

Then the *IntegrateLayers* performs several steps. First it selects the features that are repeatedly variable across datasets for integration. And then it performs the integration of the data. Using the ** argument, you can choose the method of integration that you want to use. The full Seurat5 integration vignette with the different integration methods can be find here.

```
lung_integrated <- IntegrateLayers(
  object = lung_merged, method = CCAIntegration,
  orig.reduction = "pca", new.reduction = "integrated.cca",
  verbose = FALSE
)
```

Finally, we re-join the layers together

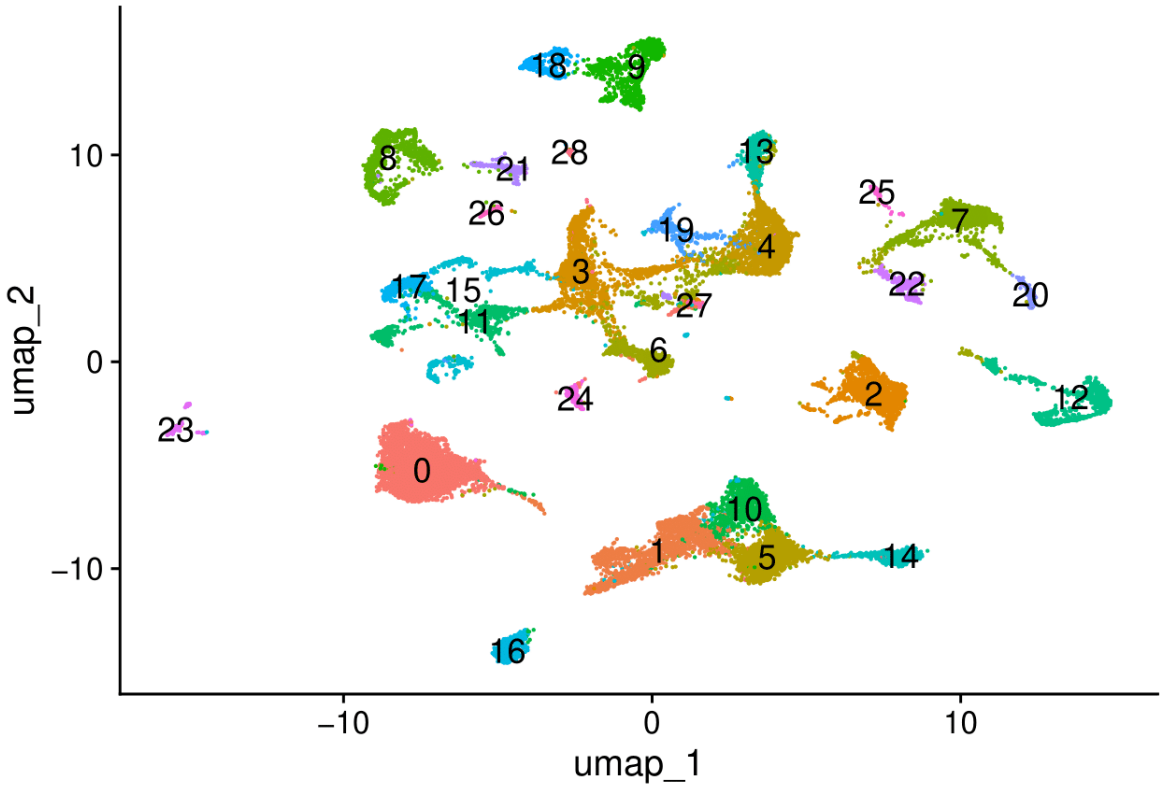
```
lung_integrated[["RNA"]] <- JoinLayers(lung_integrated[["RNA"]])
```

Calculating the UMAP coordinates and clusters

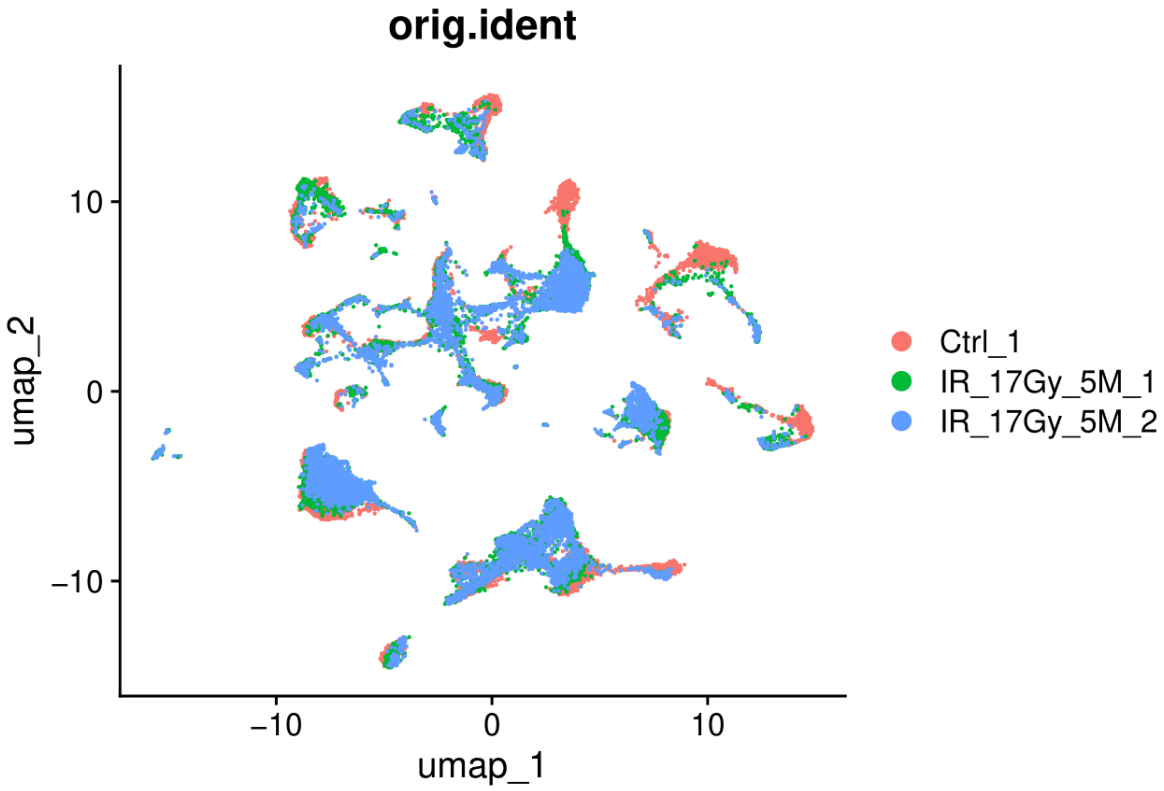
Then the PCA, UMAP coordinates for visualization and division into clusters are recalculated taking into account the data from all the samples.

In the *FindClusters* function, the resolution parameter allows to choose for a number more or less important of clusters. **Use a value above (below) 1.0 if you want to obtain a larger (smaller) number of communities**.

```
lung_integrated <- RunPCA(lung_integrated, verbose = FALSE)
lung_integrated <- RunUMAP(lung_integrated, reduction = "pca", dims = 1:30)
lung_integrated <- FindNeighbors(lung_integrated, reduction = "pca", dims = 1:20)
lung_integrated <- FindClusters(lung_integrated, resolution = 0.5)
```



You can check if the batch effect seems to be corrected by ploting the different samples or conditions in a UMAP.



Saving of the integrated object

The second step of preprocessing is over, now you can **save your object in a chosen folder**. Then you can go to the tutorial 4 to perform the cell type labeling of the clusters.

```
saveRDS(lung_integrated, "processed_objects/Seurat_object_integrated.rds")
```

Tutorial 4 : cell type labeling of the clusters and cleaning of the object

Juliette SOULIER

4/16/2024

The identification of the cellular composition of an object is a crucial and delicate part of the analysis. It will affect the results of all the analysis that will follow. Therefore it is very important to be meticulous with this step. To identify the cell types we use known markers from the literature. There are more and more papers describing and recapitulating cell type identification markers for different organs and organisms. Here we will use the example of the lung in mice samples.

Here the manual method of single cell RNA sequencing data annotation is described. Methods of automatic annotation also exists. However, for the automatic annotation to work properly, you need to find the right algorithm and, more importantly, an extensive single cell RNA sequencing database carefully annotated with all the cell types that you wish to identify. This can be challenging, there is always the need to verify the expression of the known canonical markers to check the proper annotation of the data.

Install the necessary package

ggplot2 package

ggplot2 is a package allowing plot customization.

```
install.packages('ggplot2')
```

Load the necessary packages

```
library(Seurat)
library(ggplot2)
```

Load the object

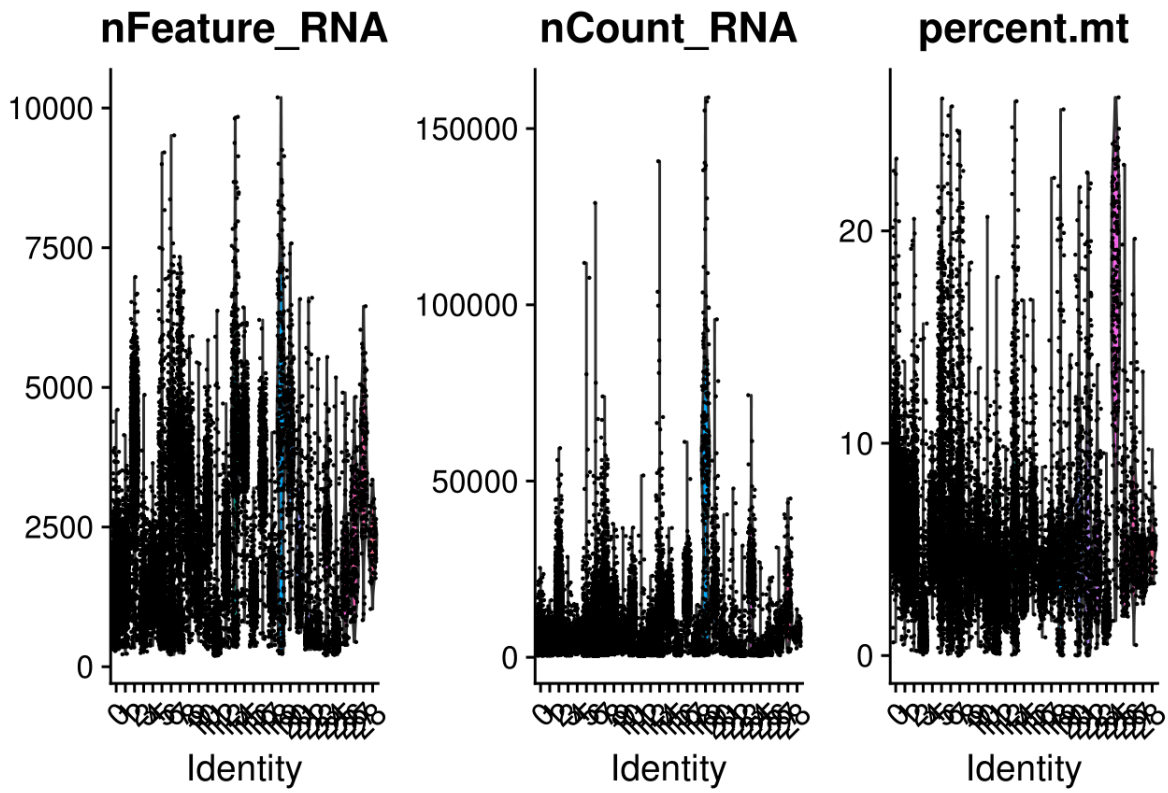
The first step is to load the merged object generated with the tutorial 2 or the integrated object generated with the tutorial 3. For this example we will work with the merged object. **In the argument of the *readRDS* function, put the path to your merged or integrated object**.

```
setwd("/Users/labo/Desktop/Tutos_scRNA_seq_R")
lung_merged <- readRDS("processed_objects/Seurat_object_merged.rds")
```

Quality controls of the clusters

Some clusters can group cells of poor quality, it is important to remove them or they will affect the analysis. A cluster is considered of poor quality when most of its cells present high *percent.mt* and low *nCount_RNA* / *nFeature_RNA*. *nFeature_RNA* is the number of genes with at least one mRNA sequenced in the cell, *nCount_RNA* it the number of mRNA sequenced in the cell and *percent.mt* is the percentage of mitochondrial genes over the total number of genes.

```
VlnPlot(lung_merged, features = c("nFeature_RNA", "nCount_RNA", "percent.mt"))
```



Here the cluster 24 high *percent.mt* and low *nCount_RNA* and *nFeature_RNA*, so we choose to keep all of them.

Removal of the low quality clusters

If during the previous step there were some low quality clusters, we need to remove them from the data and recalculate the clusters and UMAP coordinates (if not, skip this step). Here we will remove the **cluster 24**.

First the function **subset** allows to extract some clusters from the object. Here we remove the **cluster 24** from your object:

```
lung_merged <- subset(lung_merged, idents = c("24"), invert = TRUE)
```

The *invert = TRUE* means that we keep all the clusters except that the ones in the *idents* parameter. If *invert* is set to *FALSE*, the subset function will keep only the clusters specified in the parameter *ident*.

Then you need to redo the UMAP coordinates and cluster calculation, and retest the quality of the newly calculated clusters.

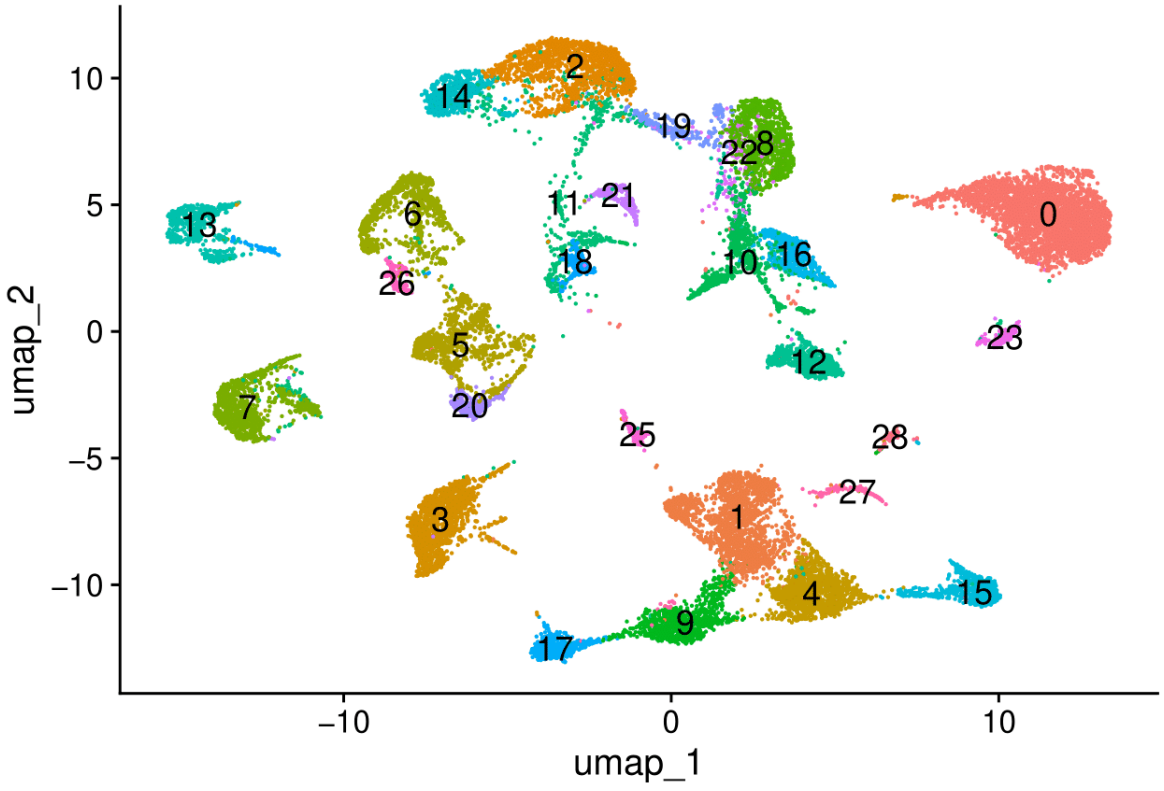
```
lung_merged <- RunPCA(lung_merged, npcs = 20, verbose = T)
lung_merged <- RunUMAP(lung_merged, reduction = "pca", dims = 1:20)
lung_merged <- FindNeighbors(lung_merged, reduction = "pca", dims = 1:20)
lung_merged <- FindClusters(lung_merged, resolution = 0.7)

VlnPlot(lung_merged, features = c("nFeature_RNA", "nCount_RNA", "percent.mt"))
```

If there are still low quality clusters in the new object, this step can be repeated several times.

To divide the cells into more clusters, the *resolution* parameter of the *FindClusters* can be increased (or decreased to divide the cells into less clusters).

Identification by calcul of the markers

Then, we can visualize the object and the clusters.



The Seurat function *FindAllMarkers* allows the identification of the genes that are the most specific of each cluster. This is very helpful to identify the identity of the clusters. This function returns a table that can then be saved to work on the identification with it. **In the argument of the *write.csv* function, put the path of the folder where you want to save the marker table**.

```
DefaultAssay(lung_merged) <- "SCT"
Idents(lung_merged) <- "seurat_clusters"
table_markers <- FindAllMarkers(lung_merged, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25, m
write.csv(table_markers, "results_analysis/Markers_lung_merged_SCT.csv")
```

X	p_val	avg_log2FC	pct.1	pct.2	p_val_adj	cluster	gene
Cd79a	0	6.376864	0.992	0.042	0	0	Cd79a
Ebf1	0	5.911433	0.969	0.043	0	0	Ebf1
Cd79b	0	5.796725	0.943	0.055	0	0	Cd79b
Ly6d	0	4.665536	0.947	0.059	0	0	Ly6d
Ms4a1	0	6.579055	0.864	0.020	0	0	Ms4a1
Bank1	0	5.914312	0.800	0.025	0	0	Bank1
Ighd	0	6.237278	0.792	0.021	0	0	Ighd
Fcmr	0	6.479551	0.766	0.016	0	0	Fcmr
Iglc3	0	3.206417	0.760	0.030	0	0	Iglc3
Mef2c	0	3.791679	0.882	0.160	0	0	Mef2c
H2-DMb2	0	4.104860	0.769	0.056	0	0	H2-DMb2
Iglc2	0	4.057703	0.744	0.034	0	0	Iglc2
Cd37	0	3.205517	0.925	0.258	0	0	Cd37
Ighm	0	1.136586	0.979	0.332	0	0	Ighm
Gm31243	0	6.286738	0.658	0.012	0	0	Gm31243

This is the first lines of the result of the *FindAllMarkers* function. In the lines there are the genes found to be characteristic of the different clusters. Then the different columns give several information about these genes:

- p_val –> p-value not adjusted for multiple test correction
- avg_log2FC –> Average log2 fold change. Positive values indicate that the gene is more highly expressed in the cluster
- pct.1 –> The percentage of cells where the gene is detected in the cluster
- pct.2 –> The percentage of cells where the gene is detected on average in the other clusters
- p_val_adj –> Adjusted p-value, based on bonferroni correction using all genes in the dataset, used to determine significance
- cluster –> Identity of the cluster. Here it is just a number as the clusters haven't been identified yet
- gene –> Ensembl gene ID

Here we see that the genes Cd79b and Ms4a1 are specific of cluster 0, and these genes are markers of B cells type. So the cluster 0 is composed of B cells.
By doing the same analysis for all the clusters we can annotate the identity of all the cells of the object.

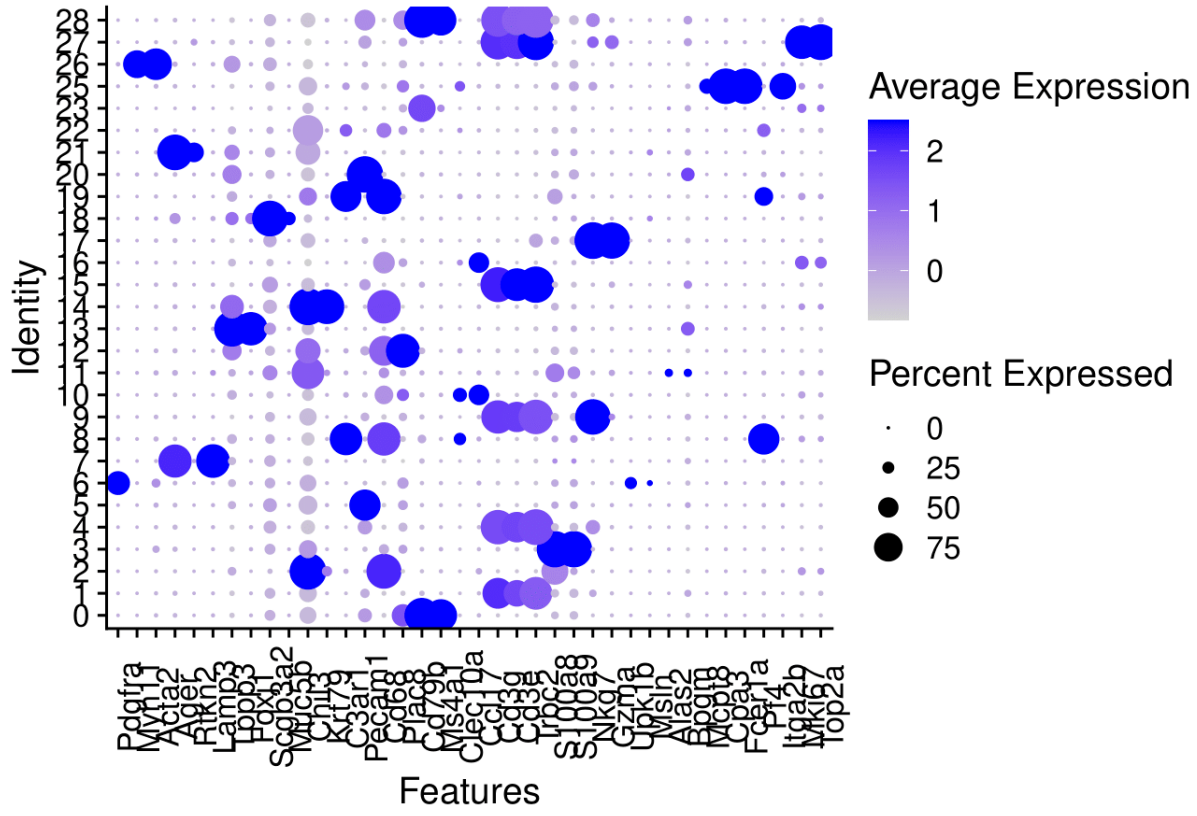
Identification by ploting of the markers

To identify the identity of the different clusters, we can also plot the expression of specific markers for each cell type of the organ, in each Seurat cluster. First we need to choose the markers. Here is a list of markers of the lung cell types :

```
markers <- c( "Pdgfra", #Fibroblasts
              "Myh11", "Acta2", #Smooth muscle cells

              "Ager", "Rtkn2", #AT1
              "Lamp3", #AT2
              "Tppp3", "Foxj1", #ciliated cells
              "Scgb3a2", "Muc5b",  #club cells

              "Chil3", "Krt79", #alveolar macrophages
              "C3ar1", #interstitial macrophages

              "Pecam1", #endothelial cells

              "Cd68", "Plac8", #monocytes
              "Cd79b", "Ms4a1", #B cells
              "Clec10a", "Ccl17", #dendritic cells
              "Cd3g", "Cd3e", "Trbc2", #T cells
              "S100a8", "S100a9", #neutrophils
              "Nkg7", "Gzma", #natural killer cells
              "Upk1b", "Msln", #mesotheliocyte

              "Alas2", "Bpgm", #erythrocytes
              'Mcpt8', 'Cpa3', 'Fcer1a', #basophils
              'Pf4', 'Itga2b', #platelets

              "Mki67", "Top2a" #proliferating
)
```

Then to have the expression of these genes in all of the clusters in one plot, we use the *Dotplot* function:

```
plot <- DotPlot(lung_merged, features = markers)
plot + theme(axis.text.x = element_text(angle = 90))
```



A large dot indicates that a high proportion of the cells of the cluster express a given gene, and a dark blue dot indicates a high mean expression of a given gene by the cells of the cluster.
If we look again at the cluster 0, it expresses the B cells markers Cd79b and Ms4a1.

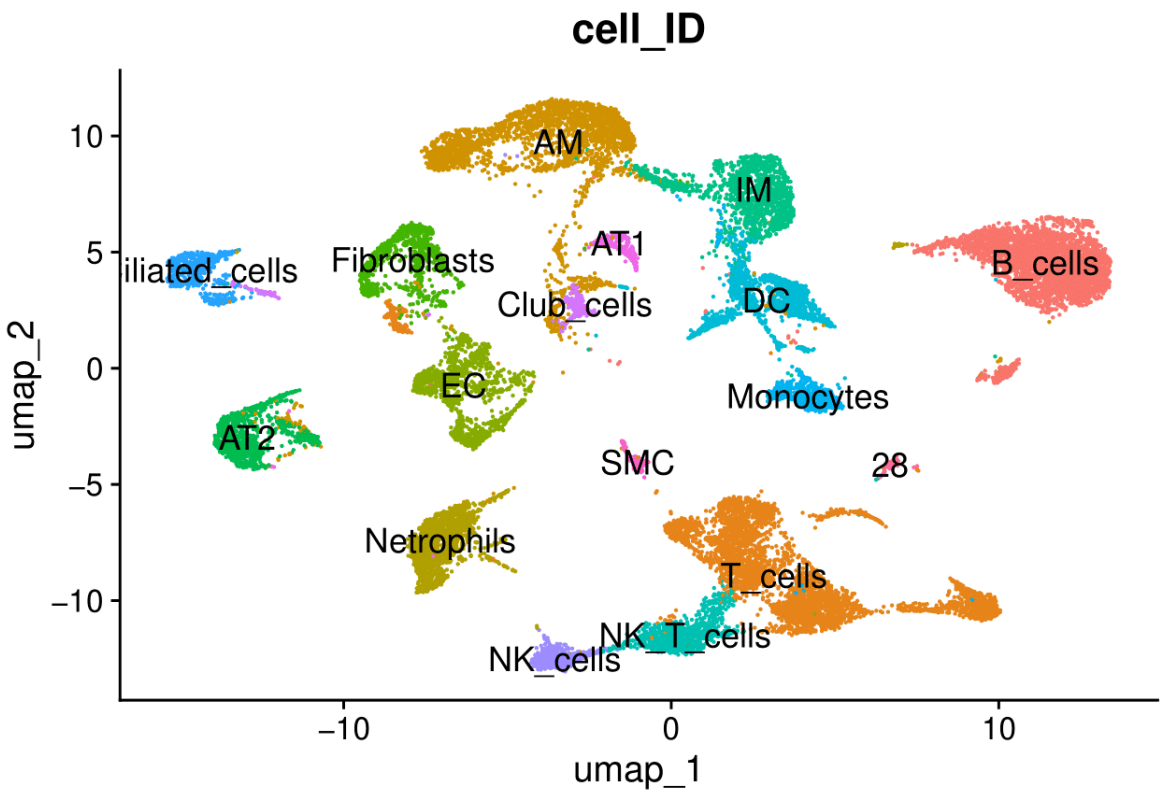
Assignation of the cell types

Now that we identified the cell types corresponding to the different clusters, we need to save this information in the object, by first renaming the *idents* and then create a new metadata slot with the cell identity. **Change the number of clusters according to what you have and assign them according to the cell types that you identified**.

```
lung_merged_named <- RenameIdents(lung_merged,
                                  `0` = "B_cells",
                                  `1` = "T_cells",
                                  `2` = "AM",
                                  `3` = "Netrophils",
                                  `4` = "T_cells",
                                  `5` = "EC",
                                  `6` = "Fibroblasts",
                                  `7` = "AT2",
                                  `8` = "IM",
                                  `9` = "NK_T_cells",

                                  `10` = "DC",
                                  `11` = "AM",
```

```
                            `12` = "Monocytes",
                            `13` = "Ciliated_cells",
                            `14` = "AM",
                            `15` = "T_cells",
                            `16` = "DC",
                            `17` = "NK_cells",
                            `18` = "Club_cells",
                            `19` = "IM",

                            `20` = "EC",
                            `21` = "AT1",
                            `22` = "IM",
                            `23` = "B_cells",
                            `24` = "Basophils",
                            `25` = "SMC",
                            `26` = "T_cells",
                            `27` = "T_cells"
)

lung_merged_named@meta.data$cell_ID <- Idents(lung_merged_named)
```

```
DimPlot(lung_merged_named, label = T, label.size = 4.5, group.by = "cell_ID", raster = F) + NoLegend()
```



Quality controls and cleaning of the cell types

At the beginning of this tutorial, we had a first look at the quality controls indicators. However, different populations can show very different ranges of normal percentage of mitochondrial genes expresses, number of transcripts expressed and number of different genes expressed. Therefore, to obtain a cleaner dataset, we can re-perform these quality controls and cleaning for each of the cell types identified.

We will see here the procedure on the example of the B cells. The first step is to extract population of interest. The subseting of an object can be done on a number of parameters:

- a metadata entry (for example a cell types, or only some samples)
- the *nFeature_RNA*, *nCount_RNA* or *percent.mt* parameters
- the expression of a choosen gene

Here we will extract the **B cells**. Before that, we need to specify the location of the information in the metadata. In our case it is located in the **cell_ID** metadata slot.

```
Idents(lung_merged_named) <- "cell_ID"
object_B_cells <- subset(lung_merged_named, idents = c("B_cells"), invert = FALSE)
```

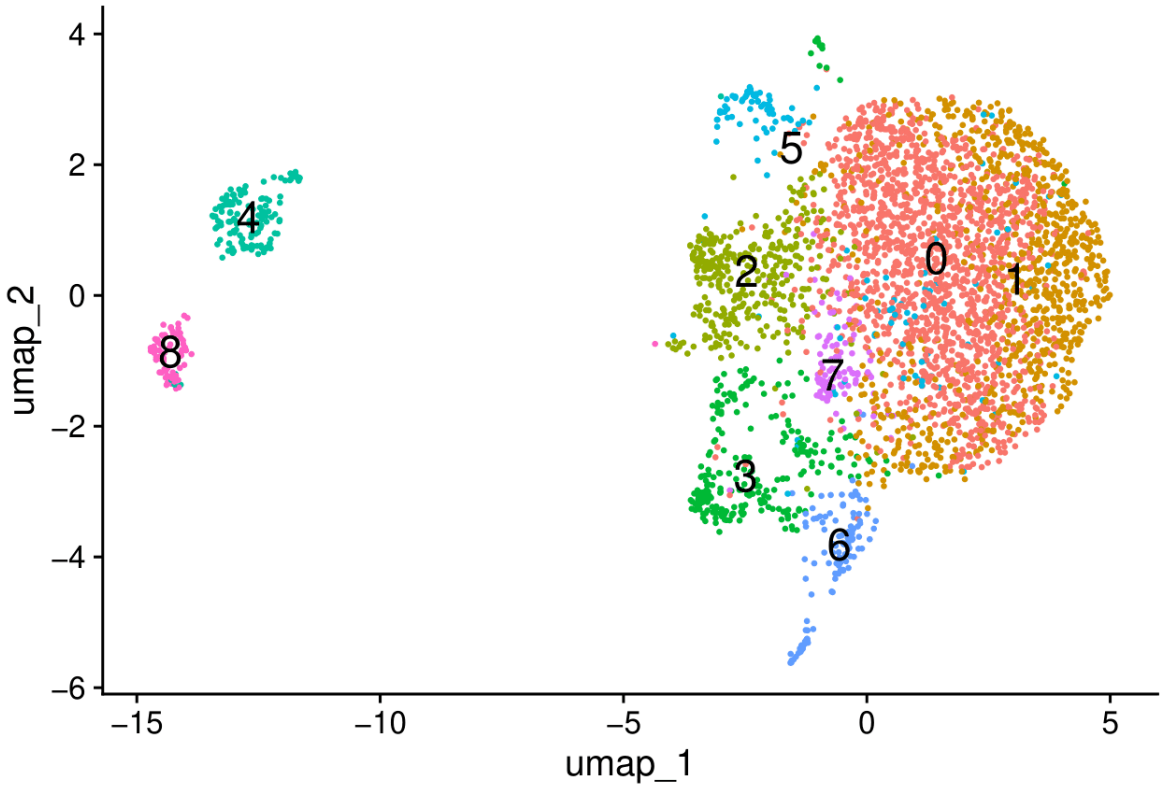
The parameter *invert = FALSE* means that we keep only the clusters specified in the parameter *idents*. If *invert* is set to TRUE, the subset function will keep all the clusters except that the ones in the *idents* parameter.

Then you need to redo the UMAP coordinates and cluster calculation.

```
object_B_cells <- RunPCA(object_B_cells, npcs = 20, verbose = T)
object_B_cells <- RunUMAP(object_B_cells, reduction = "pca", dims = 1:20)
object_B_cells <- FindNeighbors(object_B_cells, reduction = "pca", dims = 1:20)
object_B_cells <- FindClusters(object_B_cells, resolution = 0.5)
```

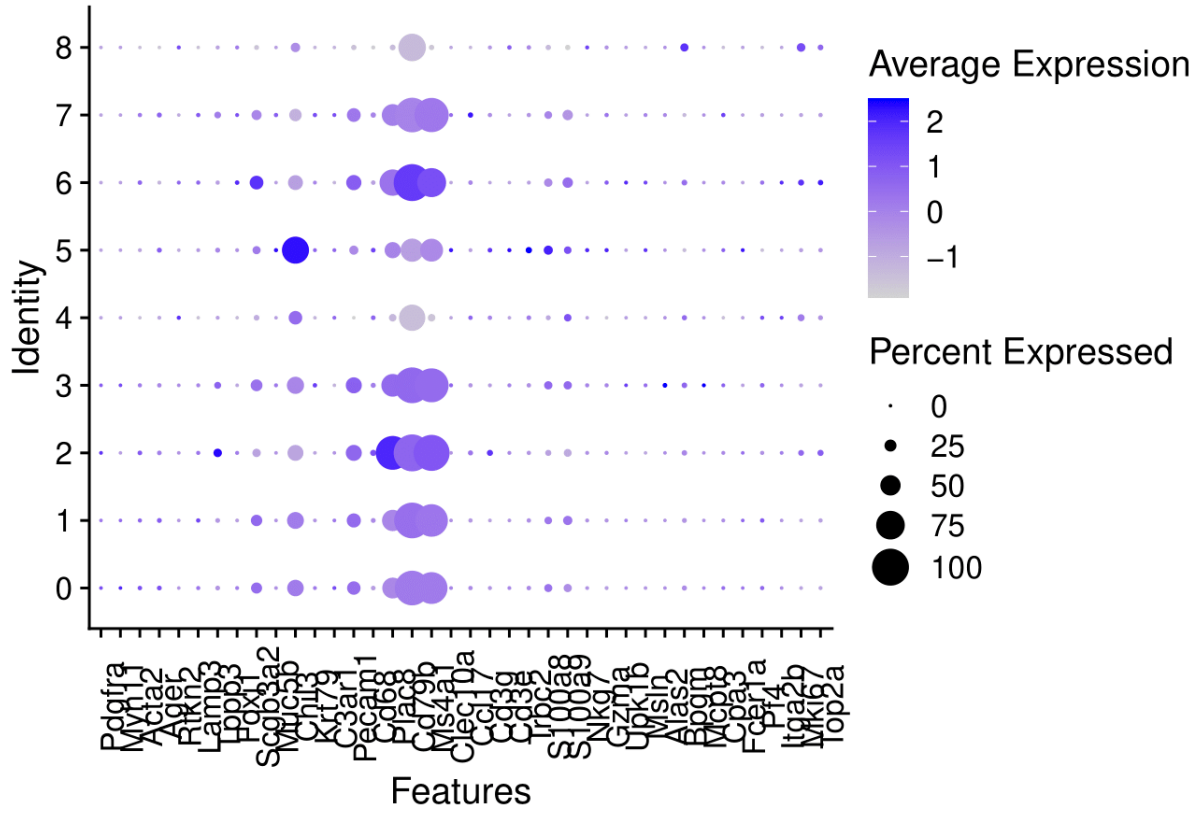
This cell type object is now divided in several clusters.

```
DimPlot(object_B_cells, label = T, label.size = 6, raster = F) + NoLegend()
```



First, we plot the expression of the markers for the different populations. If a cluster express markers non specific of B cells, it is probably a cluster constituted of doublets (two cells encapsulated and sequenced together as one).

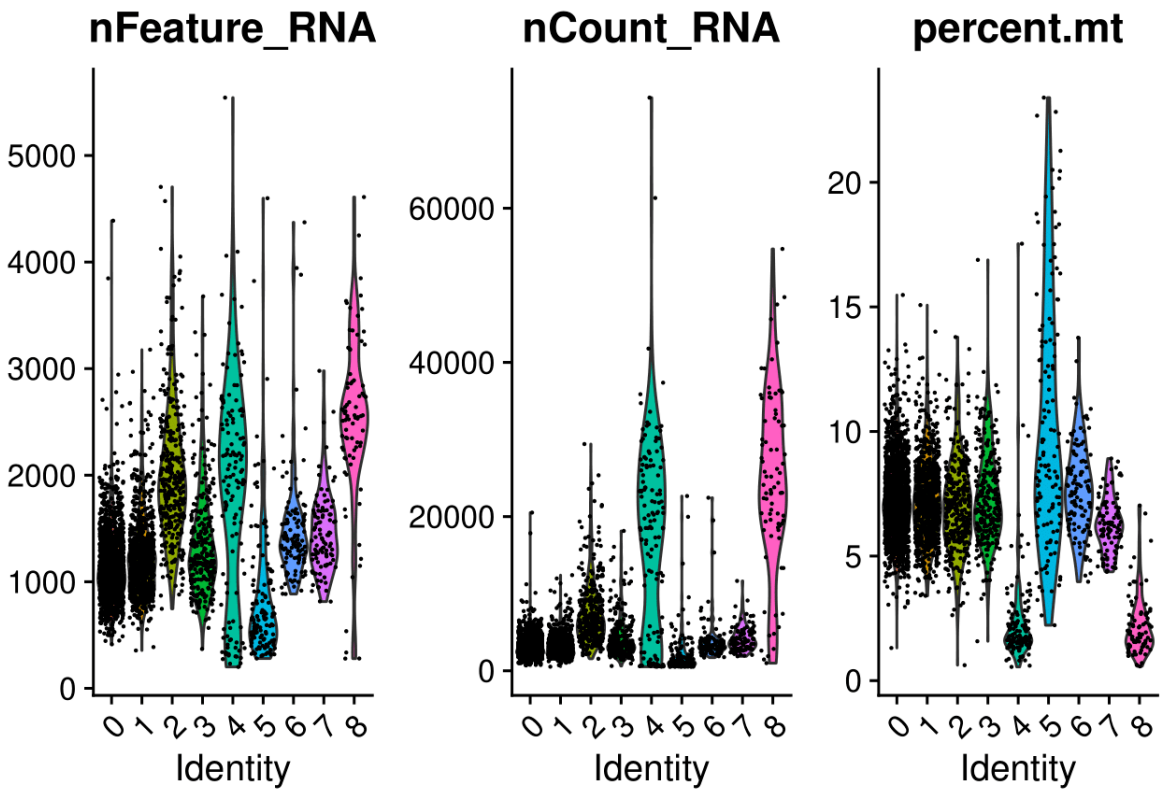
```
plot <- DotPlot(object_B_cells, features = markers)
plot + theme(axis.text.x = element_text(angle = 90))
```



Here we can see that the clusters 4 and 7 don't express any specific markers, they probably contains low quality cells.

We can then plot the *percent.mt*, *nCount_RNA* and *nFeature_RNA* to check the quality of the different clusters.

```
VlnPlot(object_B_cells, features = c("nFeature_RNA", "nCount_RNA", "percent.mt"))
```



With this plot, we can confirm that the clusters 4 and 7 are low quality cells. They present a very high number of transcripts. Furthermore, the clutser 8 shows a twice as high percentage of mitochondrial genes than the other populations, with a low number of genes and transcripts expressed. Therefore, we will remove these cells from the object. To do that, we need the list of cells constituting these clusters.

```
B_cells_to_remove <- subset(object_B_cells, idents = c("4", "7", "8"), invert = FALSE)
cells_to_remove <- colnames(B_cells_to_remove)
```

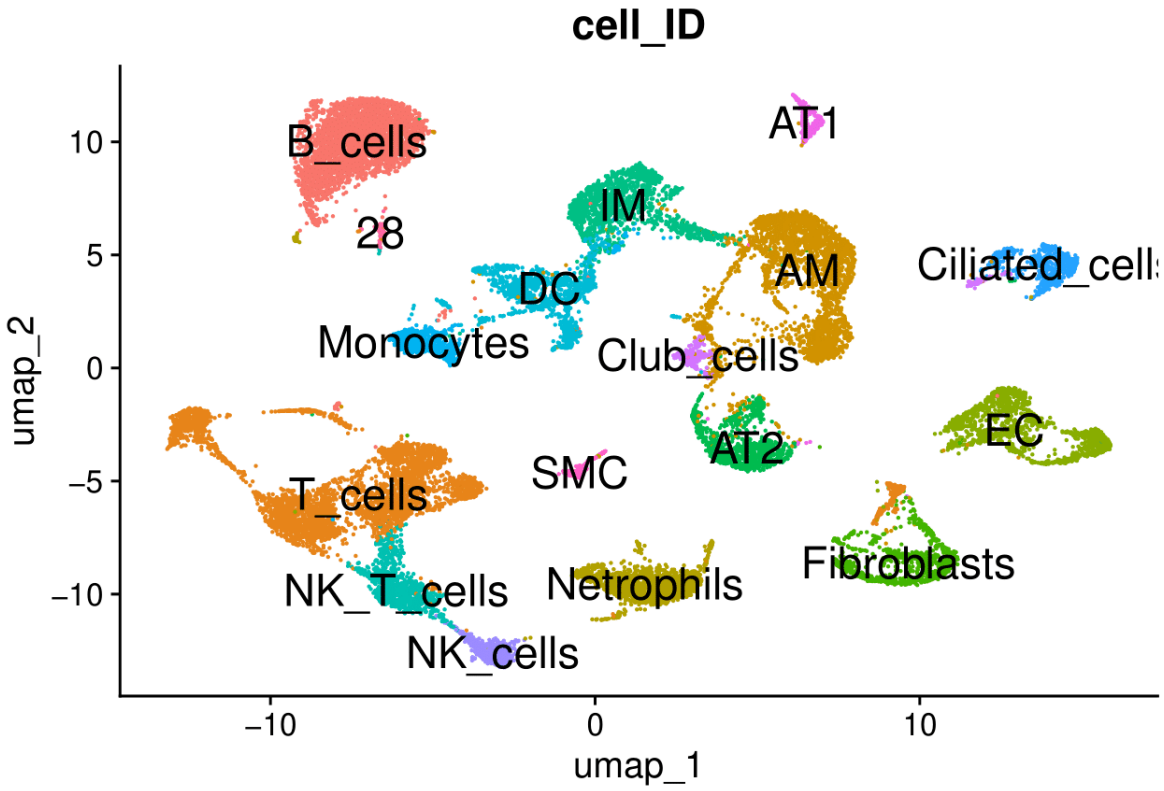
We can then remove this list of cells from the main object. We use the parameter *invert* = *TRUE* to specify that we want to keep all the cells except the list of low quality cells.

```
lung_merged_cleaned <- subset(lung_merged_named, cells = cells_to_remove, invert = TRUE)
```

This processed needs to be done for all the cell population in order to obtain quality results in the following analysis. Then, we redo the UMAP coordinates and cluster calculation.

```
lung_merged_cleaned <- RunPCA(lung_merged_cleaned, npcs = 20, verbose = T)
lung_merged_cleaned <- RunUMAP(lung_merged_cleaned, reduction = "pca", dims = 1:20)
lung_merged_cleaned <- FindNeighbors(lung_merged_cleaned, reduction = "pca", dims = 1:20)
lung_merged_cleaned <- FindClusters(lung_merged_cleaned, resolution = 0.5)
```

```
DimPlot(lung_merged_cleaned, label = T, label.size = 6, raster = F, group.by = "cell_ID") + NoLegend()
```



Identification of the different sub populations

If you are interested in the study of a specific cell population, you may want to annotate this population more precisely, with sup populations or cell states. In order to do so, the population is first extracted with the *subset* function. Then the calculation of the UMAP coordinates and cluster are redone. You can then reanotate this sub object using the same procedure as the main object.

Saving of the named and cleaned object

The cell type labeling is over, now you can **save your object in a chosen folder**.

```
saveRDS(lung_merged_cleaned, "processed_objects/Seurat_object_named.rds")
```

Tutorial 5 : visualization and differentially expressed genes analysis

Juliette SOULIER

4/16/2024

This tutorial explains the basis of the analysis of single cell RNA seq data, with visualization, comparison of different samples/conditions/cell types...

Install the necessary package

EnhancedVolcano package

EnhancedVolcano is a package to make customizable Volcano plots.

```
if (!requireNamespace('BiocManager', quietly = TRUE))
  install.packages('BiocManager')
BiocManager::install('EnhancedVolcano')
```

Load the necessary packages

```
library(Seurat)
library(ggplot2)
library(EnhancedVolcano)
```

Load the object

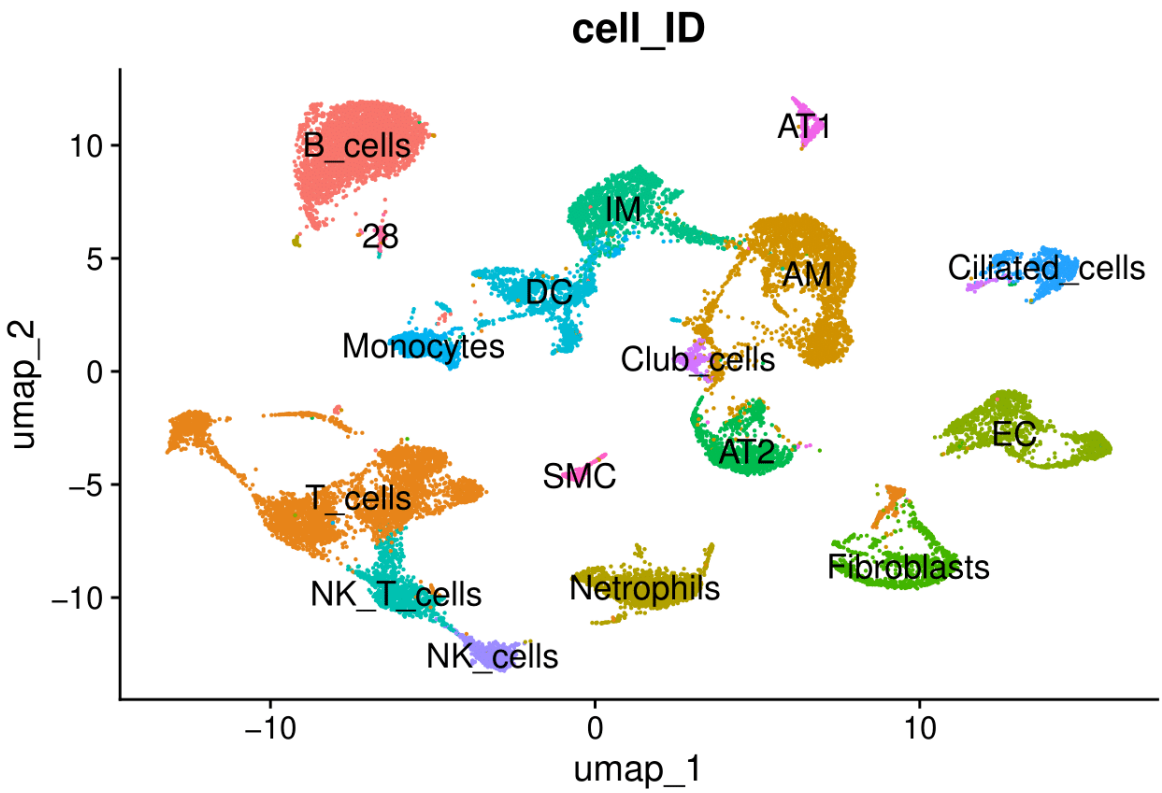
The first step is to load the named object generated with the tutorial 4. **In the argument of the *readRDS* function, put the path to your named object**.

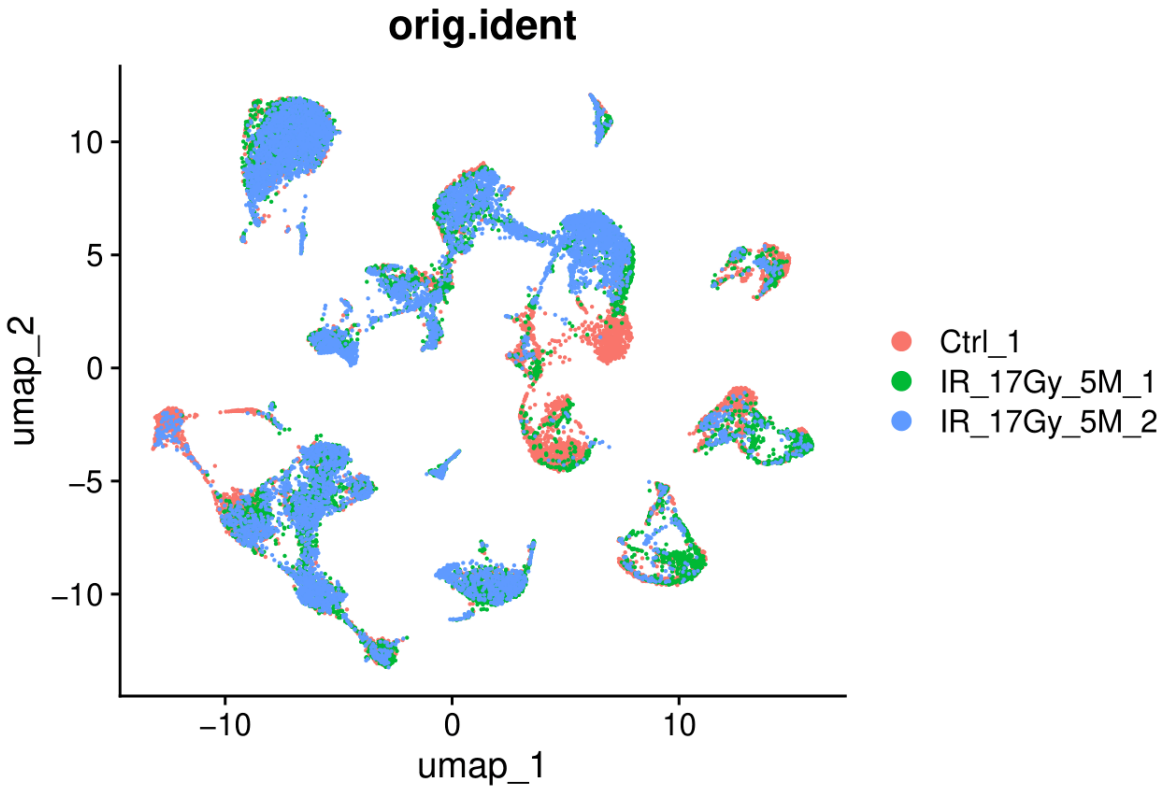
```
setwd("/Users/labo/Desktop/Tutos_scRNA_seq_R")
lung_merged <- readRDS("processed_objects/Seurat_object_named.rds")
```

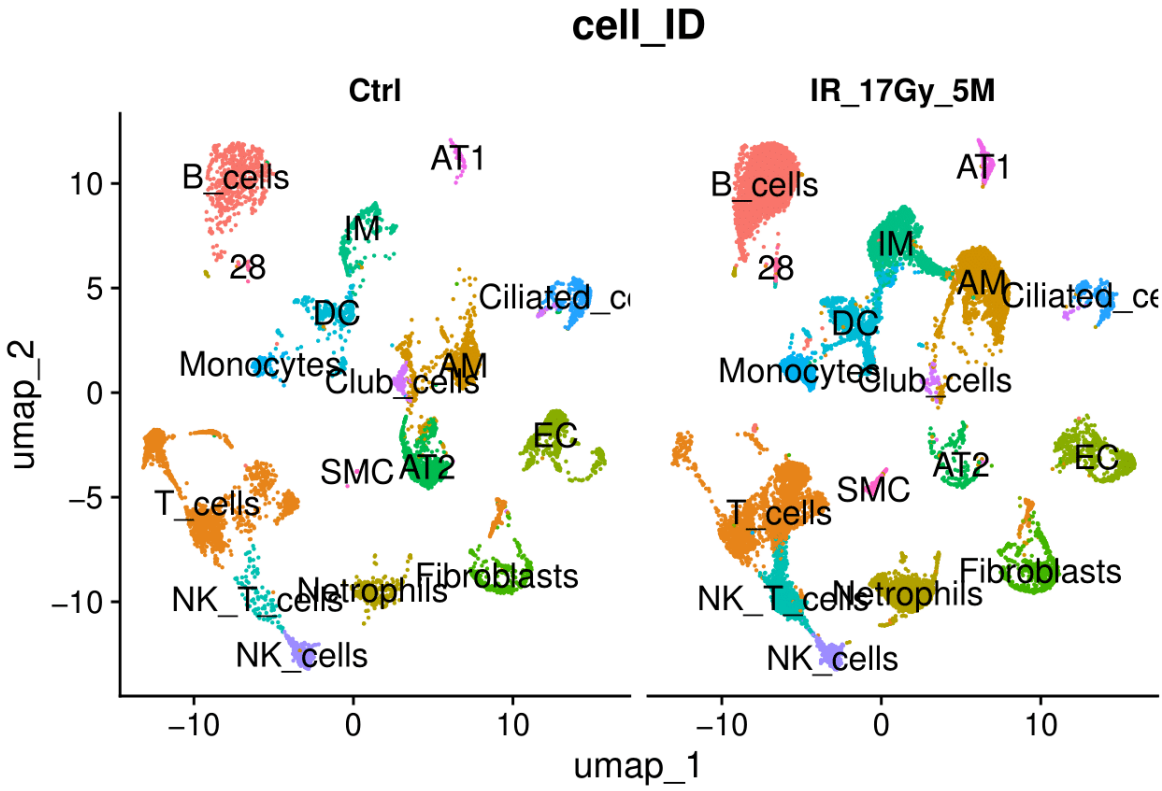
Visualisation of the data and meta data

Single cell data has usually thousands of dimensions, and to visualize it we need to have it plot on a two-dimensoins space. To do so we use the previously calculated reduction of dimensions, and the UMAP algorithm. One dot represent one cells, and the cells are distributed according to their similarities and differences of transcriptome.

You can choose to visualize different meta data parameters (argument *group.by*) or produce separated plots for different modalities (argument *split.by*).



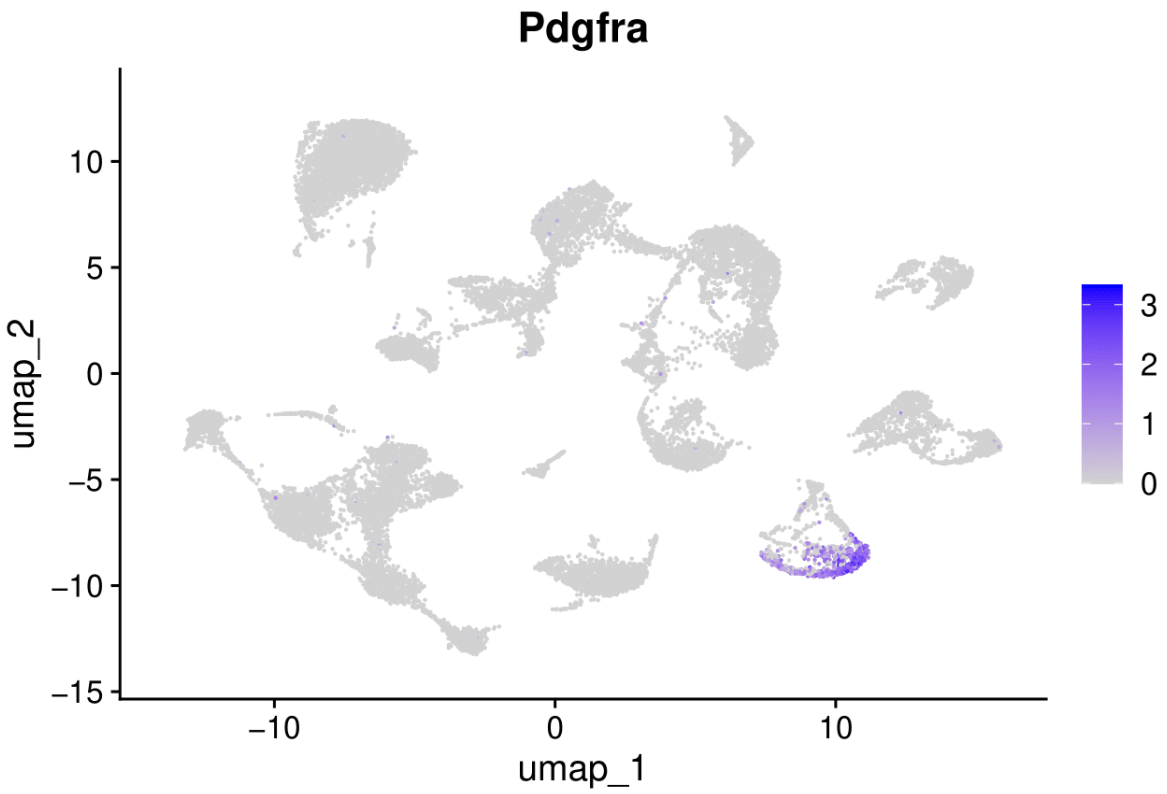

orig.ident
umap_2
umap_1
10
5
0
−5
−10
−10
0
10
Ctrl_1
IR_17Gy_5M_1
IR_17Gy_5M_2

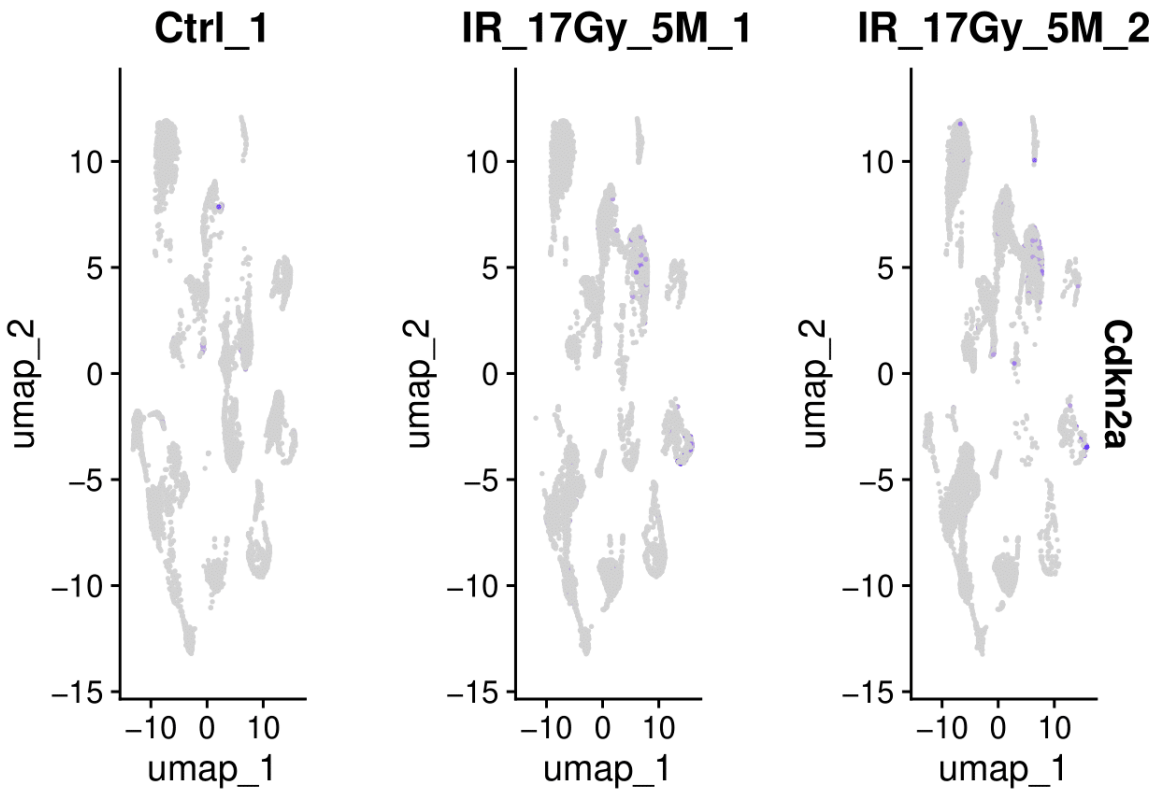


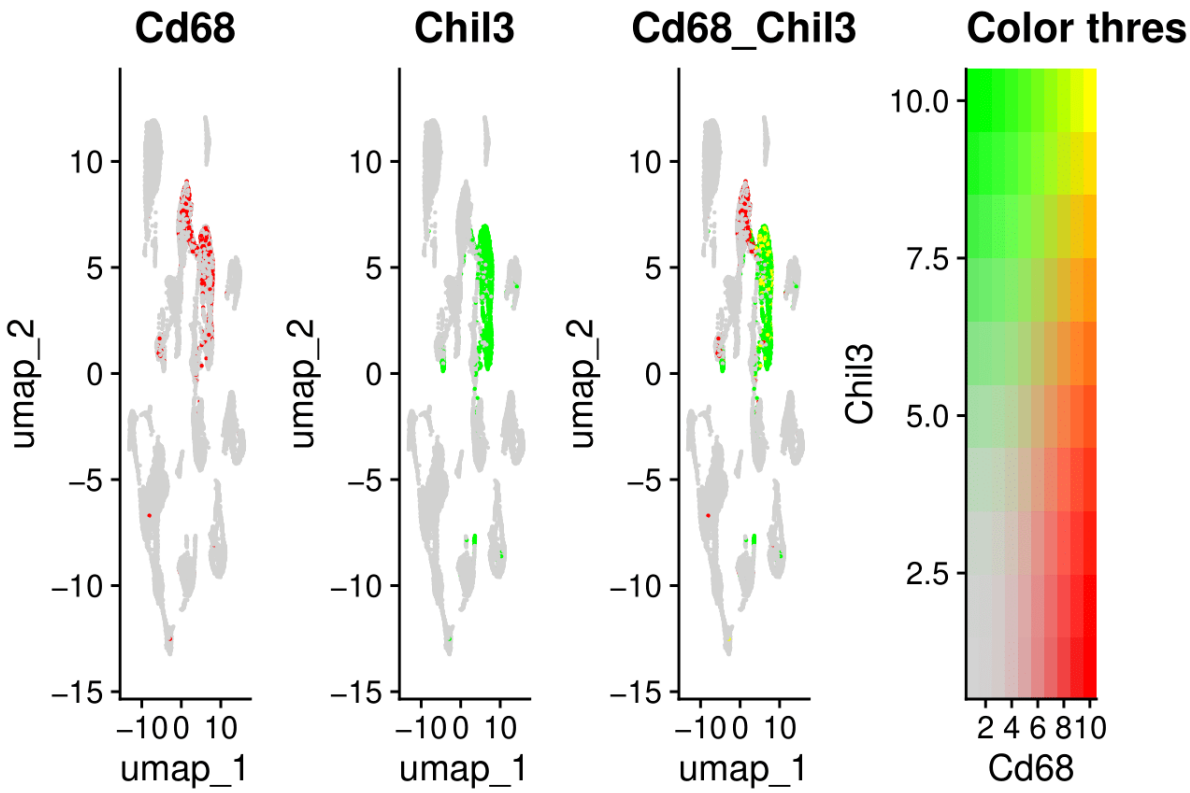
Visualisation of the gene expression

Another useful information to visualize is the gene expression. This can be done in several ways :

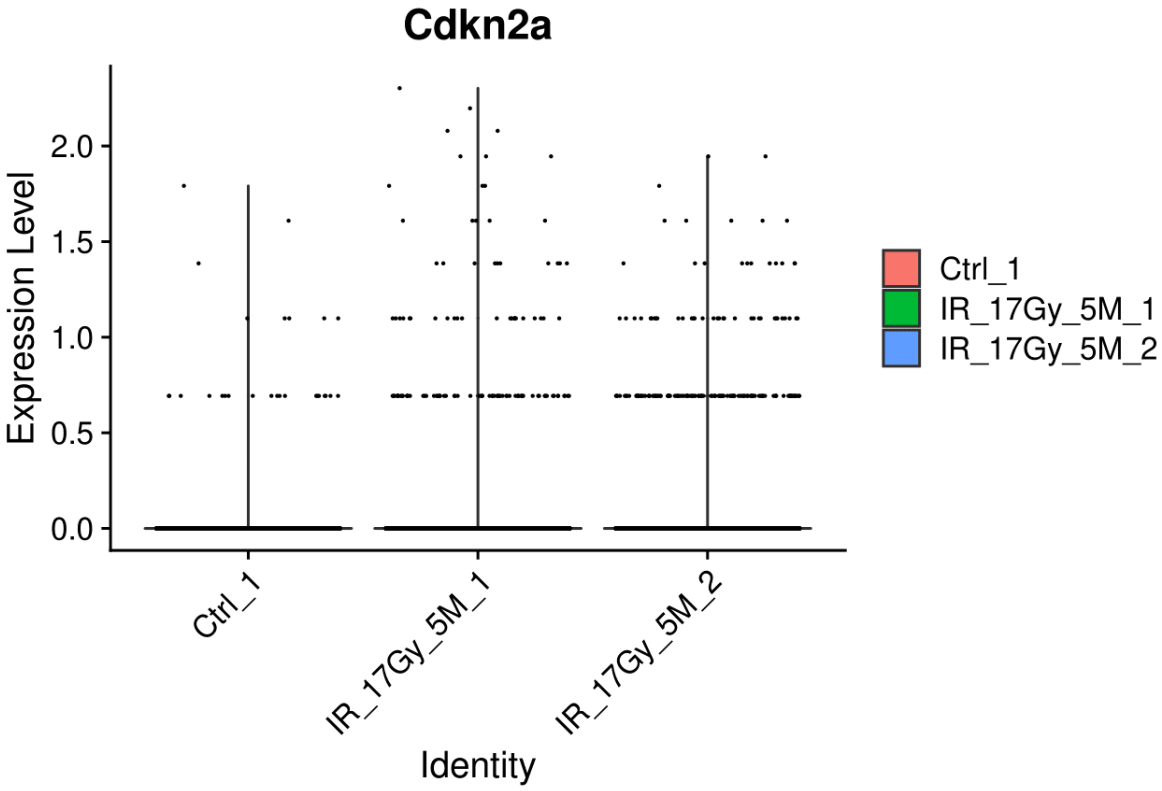
- FeaturePlot: plots the gene expression directly on the UMAP. The plots also be separated according to a modality, and it is also possible to plot the co-expression of two genes.

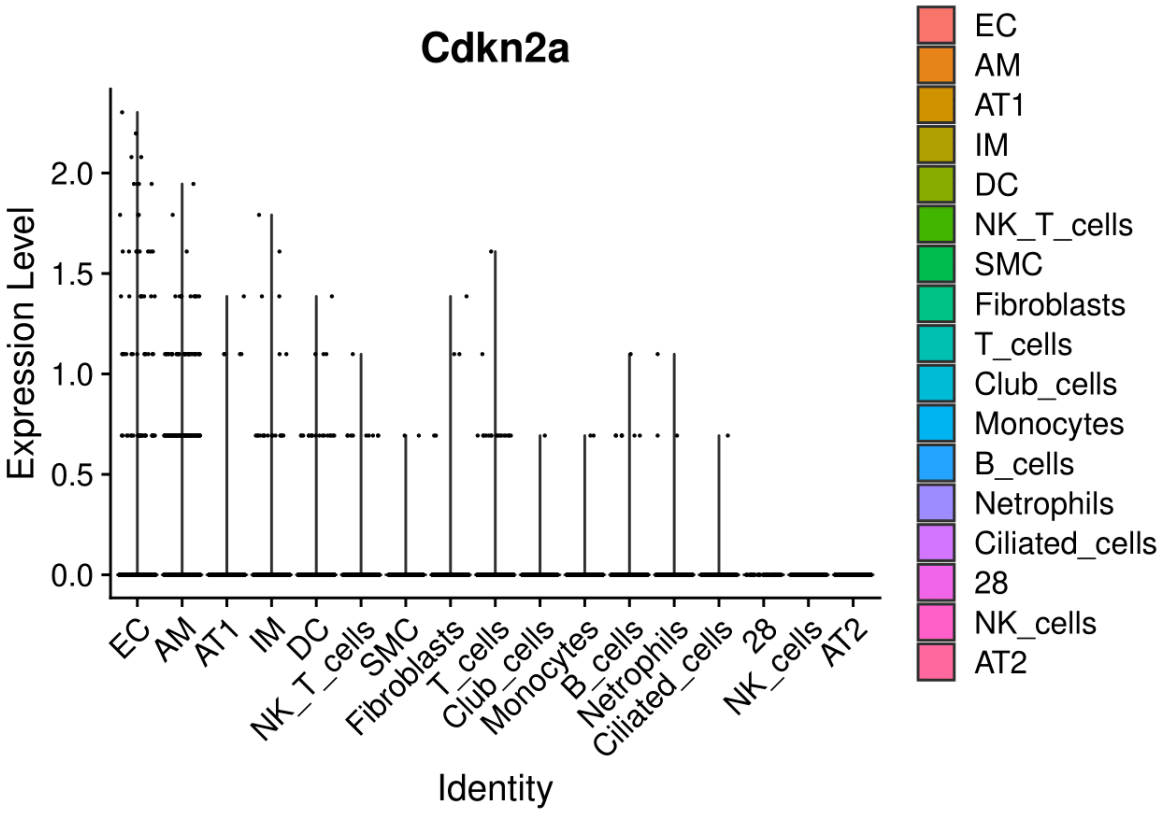

Pdgfra
umap_2
umap_1
10
5
0
−5
−10
−15
−10
0
10
3
2
1
0

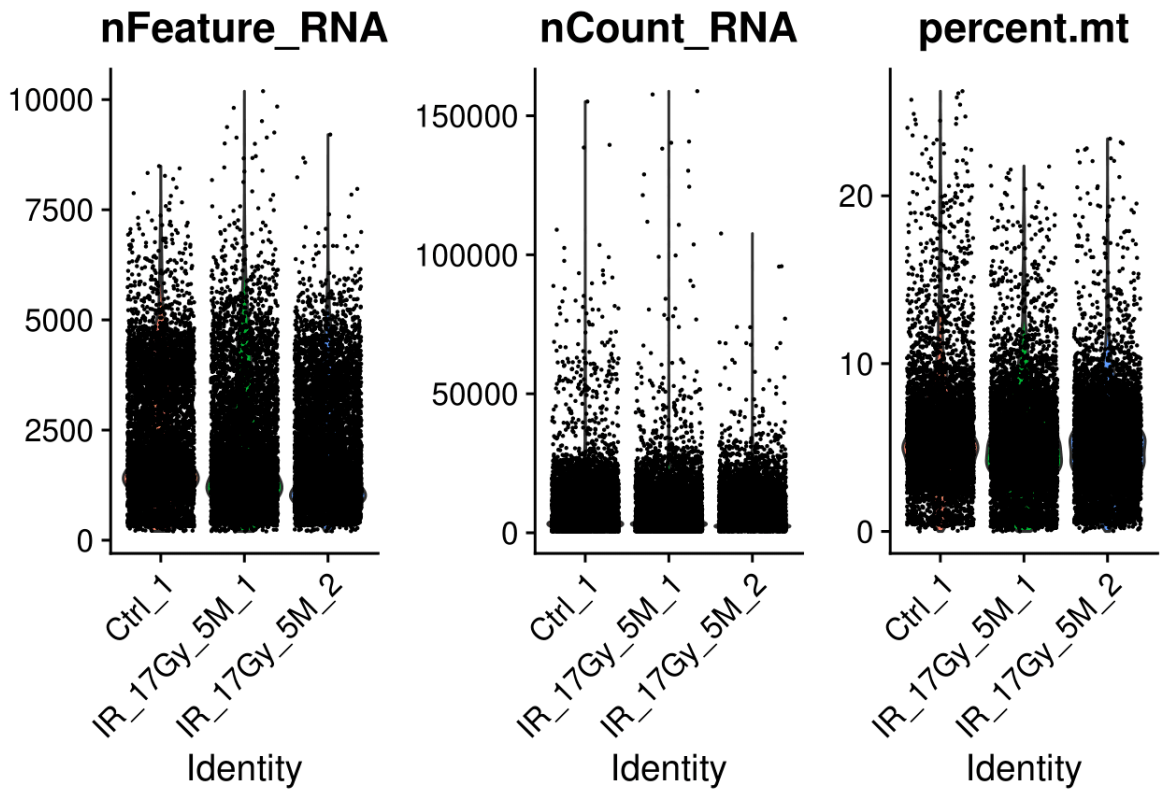

Ctrl_1
IR_17Gy_5M_1
IR_17Gy_5M_2
Cdkn2a
umap_2
umap_1
10
5
0
−5
−10
−15
−10
0
10



- VlnPlot: draws a violin plot of single cell data. It can be gene expression, but also quality controls for example. Like for the UMAP plot, one dot represents one cell, and the violin shape represent a density of cells.


Cdkn2a
Expression Level
2.0
1.5
1.0
0.5
0.0
Ctrl_1
IR_17Gy_5M_1
IR_17Gy_5M_2
Identity
Ctrl_1
IR_17Gy_5M_1
IR_17Gy_5M_2


Cdkn2a
Expression Level
2.0
1.5
1.0
0.5
0.0
EC
AM
AT1
IM
DC
NK_T_cells
SMC
Fibroblasts
T_cells
Club_cells
Monocytes
B_cells
Netrophils
Ciliated_cells
28
NK_cells
AT2
Identity
EC
AM
AT1
IM
DC
NK_T_cells
SMC
Fibroblasts
T_cells
Club_cells
Monocytes
B_cells
Netrophils
Ciliated_cells
28
NK_cells
AT2



nFeature_RNA is the number of genes with at least one mRNA sequenced in the cell, *nCount_RNA* it the number of mRNA sequenced in the cell and *percent.mt* is the percentage of mitochondrial genes over the total number of genes.

- Cell type proportion

The proportion of the cell types in the different samples can be very interesting to study. However this result needs to be analysed with precautions as it can be biased by technical issues, such as the dissociation process.

```
param.1 <- "orig.ident"
param.2 <- "cell_ID"
Idents(lung_merged) <- param.1
cells_per_cluster_and_batch <- as.matrix(table(lung_merged@meta.data[[param.1]], lung_merged@meta.data[
cells_per_cluster <- as.matrix(table(lung_merged@meta.data[[param.1]]))
colnames(cells_per_cluster) <- "Total"
nb.cells <- t(cbind(cells_per_cluster_and_batch, cells_per_cluster))
```

	Ctrl_1	IR_17Gy_5M_1	IR_17Gy_5M_2
B_cells	568	982	1665
T_cells	1507	1088	1263
AM	899	594	1264
Netrophils	234	614	570
EC	634	492	189

	Ctrl_1	IR_17Gy_5M_1	IR_17Gy_5M_2
Fibroblasts	376	525	123
AT2	855	142	16
IM	223	506	731
NK_T_cells	92	298	494
DC	311	487	514
Monocytes	114	169	367
Ciliated_cells	352	161	92
NK_cells	235	144	95
Club_cells	219	42	43
AT1	34	133	91
SMC	7	86	94
28	9	20	37
Total	6669	6483	7648

- Analysis of differentially expressed genes (DEG)

The aim of a DEG analysis is to identify the genes differentially expressed between two conditions. The comparison can be between two cell types, two experimental conditions... First you need to specify the meta data slot where are stored your two conditions. Then the control compartment is put in the *ident.1* argument, and the other one in the *ident.2* argument.

```
Idents(lung_merged) <- "condition"
DEG_NI_IR <- FindMarkers(lung_merged, ident.1 = "IR_17Gy_5M", ident.2 = "Ctrl", logfc.threshold = 0.1,
```

	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
Sftpc	0	-4.282583	0.380	0.989	0
Spp1	0	5.253038	0.728	0.194	0
Igkc	0	1.285935	0.802	0.280	0
Sftpa1	0	-3.623271	0.116	0.574	0
Cbr2	0	-3.611625	0.095	0.507	0
Sftpb	0	-3.920021	0.060	0.358	0
mt-Nd4l	0	-1.385142	0.462	0.747	0
Cyp2f2	0	-2.748017	0.062	0.325	0
Wfdc2	0	-1.895722	0.160	0.422	0
Scgb3a2	0	-2.452463	0.141	0.395	0
Igha	0	5.168471	0.261	0.009	0
Gm10076	0	-1.200278	0.602	0.849	0
Cxcl15	0	-3.586534	0.048	0.287	0
Atp1b1	0	-2.716969	0.086	0.320	0
Selenbp1	0	-2.726413	0.073	0.296	0

This is the fist lines of the result of the FindMarkers function.

- p_val –> p-value not adjusted for multiple test correction
- avg_log2FC –> Average log2 fold change. Positive values indicate that the gene is more highly expressed in the ident.1 compartment
- pct.1 –> The percentage of cells where the gene is detected in the ident.1 compartment
- pct.2 –> The percentage of cells where the gene is detected on average in the ident.2 compartment

- p_val_adj –> Adjusted p-value, based on bonferroni correction using all genes in the dataset, used to determine significance
- row names –> Ensembl gene ID

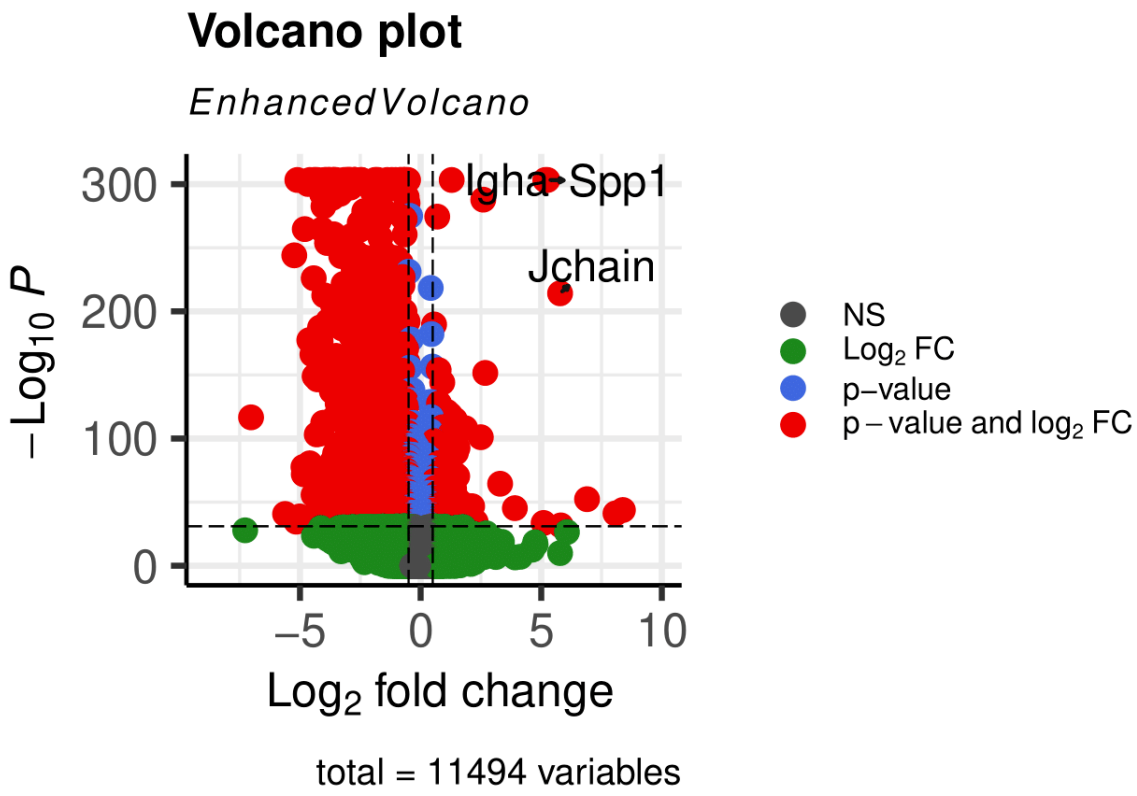
The upregulated and downregulated genes can then be analysed with GSEA.

- Visualization of DEG results: Volcano plot

DEG results can be visualized with a volcano plot. The package EnhancedVolcano is a package to make customizable Volcano plots, the github page proposes many layout options.

```
## Warning: One or more p-values is 0. Converting to 10^-1 * current lowest
## non-zero p-value...
```

```
## Warning: ggrepel: 1705 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



- Analysis of different sub populations

A lot of times, it is much more convenient to work on a subset of the object that contains all the samples and cell types. For example, if you are interested in a specific cell types, it is easier to subset it in order to study it in details and investigate sub populations.

All the visualization and further analysis can be then performed on this population.

Tutorial 6 : trajectory analysis with Monocle3

Juliette SOULIER

4/16/2024

A trajectory analysis is a way to predict a possible transition from a cell state to another or from a cell state to another. Monocle3 is a tool that allows to perform trajectory analysis on single cell RNA sequencing data.

Install the necessary package

Monocle3 package and its dependencies

```
if (!requireNamespace("BiocManager", quietly = TRUE))
install.packages("BiocManager")
BiocManager::install(version = "3.14")
BiocManager::install(c('BiocGenerics', 'DelayedArray', 'DelayedMatrixStats',
                       'limma', 'lme4', 'S4Vectors', 'SingleCellExperiment',
                       'SummarizedExperiment', 'batchelor', 'HDF5Array',
                       'terra', 'ggrastr'))
install.packages("devtools")
devtools::install_github('cole-trapnell-lab/monocle3')
```

Load the necessary packages

```
library(monocle3)
library(Seurat)
library(SeuratWrappers)
library(patchwork)
```

Load the object

The first step is to load the merged object generated with the tutorial 2 or the integrated object generated with the tutorial 3. For this example we will work with the merged object. **In the argument of the *readRDS* function, put the path to your merged or integrated object**.

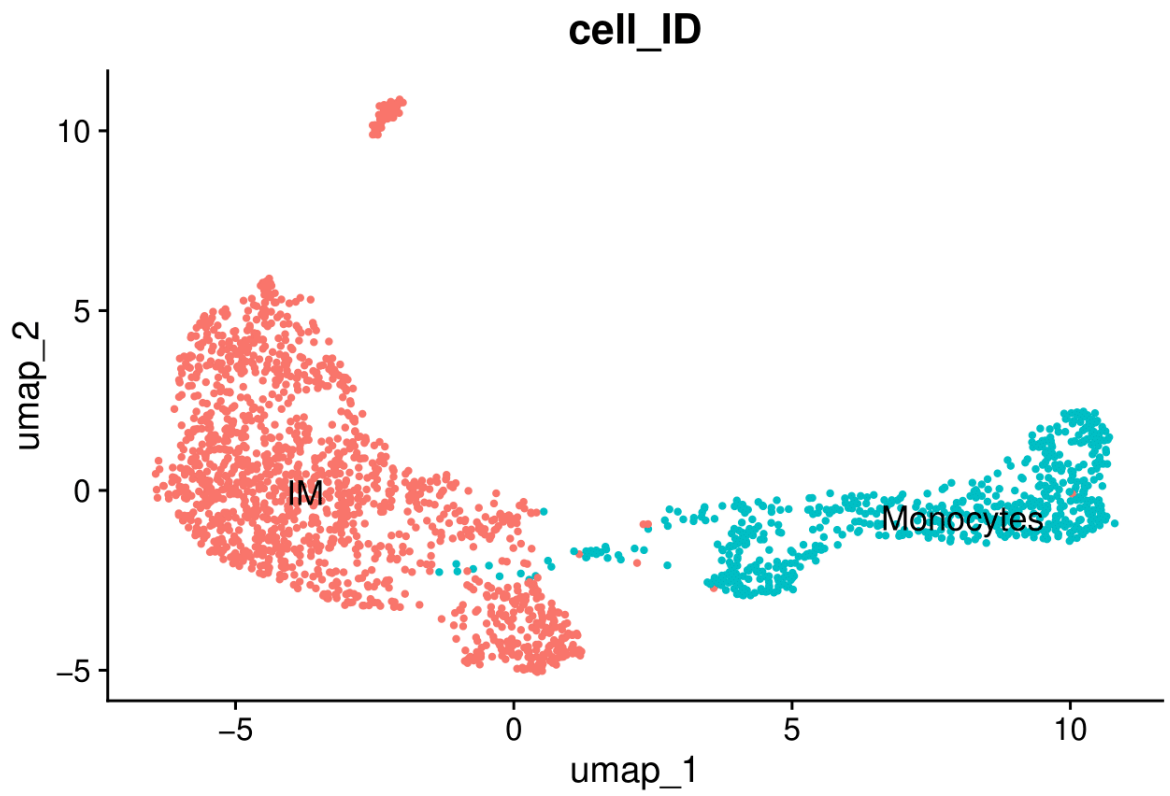
```
setwd("/Users/labo/Desktop/Tutos_scRNA_seq_R")
lung_merged <- readRDS("processed_objects/Seurat_object_named.rds")
```

Subset the object

Trajectory analysis aims to predict a possible transition from a cell population to another. Therefore it is more efficient to perform this analysis of a subset of the cell populations of interest. Here, we take the example if the monocytes and interstitial macrophages.

```
Idents(lung_merged) <- "cell_ID"
object_monocyte_IM <- subset(lung_merged, idents = c("Monocytes", "IM"), invert = FALSE)
```

```
object_monocyte_IM <- RunPCA(object_monocyte_IM, npcs = 20, verbose = T)
object_monocyte_IM <- RunUMAP(object_monocyte_IM, reduction = "pca", dims = 1:20)
object_monocyte_IM <- FindNeighbors(object_monocyte_IM, reduction = "pca", dims = 1:20)
object_monocyte_IM <- FindClusters(object_monocyte_IM, resolution = 0.7)
```



Convert the object

The Seurat object need to be converted to a cell_data_set object in order to be processed by Monocle3.

```
cds <- as.cell_data_set(object_monocyte_IM)
cds <- cluster_cells(cds)
```

Learn the trajectory graph

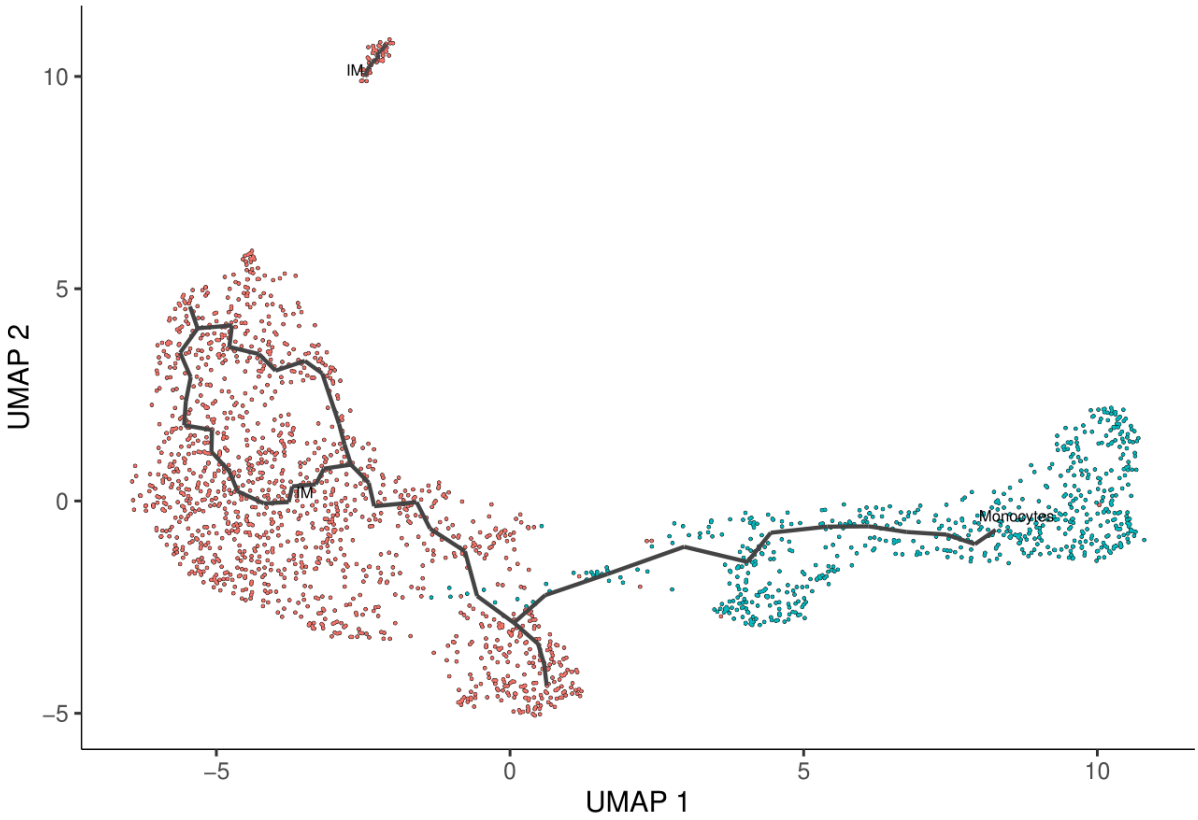
The object is then ready to commute the trajectory analysis.

```
cds <- learn_graph(cds)
```

Plot the trajectory graph

The trajectory can then be plot on the data.

```
plot_cells(cds,
           color_cells_by = "cell_ID",
           label_groups_by_cluster=T,
           label_leaves=FALSE,
           label_branch_points=FALSE)
```



Order and plot the cells in pseudotime

With prior biological knowledge, it is possible to specify the origin point of the trajectory (= the cell type of origin) in order to perform a pseudotime analysis.

```
cds <- order_cells(cds) #open interactive graphical user interface for selecting one or more root nodes
```

This pseudotime can also be plot on the data.

```
plot_cells(cds,
           color_cells_by = "pseudotime",
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           label_branch_points=FALSE,
           graph_label_size=1.5)
```

Other options for plotting can be found in the Monocle3 website.

ANNEXE VI – Spatial transcriptomics for respiratory research and medicine (*publication*).

EUROPEAN RESPIRATORY JOURNAL
ERJ METHODS
S. CURRAS-ALONSO ET AL.

Spatial transcriptomics for respiratory research and medicine

Sandra Curras-Alonso[1,2], Juliette Soulier[1], Thomas Walter[3,4,5], Florian Mueller[6,7], Arturo Londoño-Vallejo [1,8] and Charles Fouillade [2,8]

[1]Institut Curie, CNRS UMR 3244, PSL Research University, Paris Cedex 05, France. [2]Institut Curie, Inserm U1021-CNRS UMR 3347, University Paris-Saclay, PSL Research University, Centre Universitaire, Orsay Cedex, France. [3]Centre for Computational Biology (CBIO), MINES ParisTech, PSL University, Paris Cedex 06, France. [4]Institut Curie, Paris Cedex, France. [5]INSERM, U900, Paris Cedex, France. [6]Unité Imagerie et Modélisation, Institut Pasteur and CNRS UMR 3691, Paris, France. [7]C3BI, USR 3756 IP CNRS, Paris, France. [8]A. Londoño-Vallejo and C. Fouillade contributed equally to this article as lead authors and supervised the work.

Corresponding author: Charles Fouillade (charles.fouillade@curie.fr)

Shareable abstract (@ERSpublications)
In combination with scRNA-seq, spatial transcriptomics have the potential to lead to an unprecedented view of lung architecture at the single cell level, providing original information on lung physiology and physiopathology https://bit.ly/3uDYir0

Cite this article as: Curras-Alonso S, Soulier J, Walter T, *et al*. Spatial transcriptomics for respiratory research and medicine. *Eur Respir J* 2021; 58: 2004314 [DOI: 10.1183/13993003.04314-2020].

Received: 24 Nov 2020
Accepted: 29 March 2021

Introduction

Recent developments in single-cell technologies, and particularly single cell RNA sequencing (scRNA-seq), have provided invaluable tools to decipher complex biological systems like the lung. In the respiratory system, scRNA-seq analyses have led to the discovery of new cell types, such as ionocytes, as well as to a refined classification of the cells composing the lung [1–4]. Profiling of more than 300 000 cells from patients suffering lung pathologies, such as idiopathic pulmonary fibrosis (IPF), has allowed to identify new sub-populations of aberrant basal and endothelial cells that are specific to IPF [5]. Furthermore, collective efforts such as the Human Cell Atlas, aiming at characterising all cells in the human body at the molecular and spatial levels, have flagged the lung as a priority organ [6]. Ongoing efforts are now directed towards the development of spatial transcriptomic techniques allowing identification of cell localisation and description of prevailing cell–cell interactions in the organ in order to define a physiological cell atlas. Although technologies able to sequence *in situ* the transcriptome of individual cells are rapidly emerging, they still lack the resolution required to depict the extreme cell heterogeneity that characterises the anatomy of the lung. This article will present an overview of the different spatial transcriptomic methodologies that could be applied to the lung, as well as their potential impact for respiratory research and medicine.

One goal, different methodologies

The central goal of spatial transcriptomics is to depict the transcriptome landscape at the single cell level in native histological tissue sections. Two orthogonal techniques to study transcription at the cellular level have been firmly established during the last decade. First, scRNA-seq provides, from a cell suspension, access to an unbiased view of the transcriptome of each sequenced cell. Computational analysis permits the identification of different cell types present along with information on their respective transcriptome. Although scRNA-seq analysis provides key information to understand transcriptional alterations, it lacks sensitivity and, most importantly, loses spatial context due to the experimentally required cell dissociation. Second, conventional (single molecule) fluorescence *in situ* hybridisation (smFISH) detects RNA expression by direct hybridisation of fluorescently labelled probes on tissue sections. This allows the quantitative measurement of RNA expression with high sensitivity and specificity at the single cell level in their native spatial context, but not on the scale of the entire transcriptome [7, 8].

Recent developments further substantially improved these techniques, and innovative combinations of several approaches, such as multiplexing and amplification, provide an even deeper view. Here, we provide a brief overview of some of the recent milestone developments in spatial transcriptomics reached by either

improving approach or by implementing innovative combinations. Benefits and limitations of the different spatial transcriptomic methodologies are summarised in table 1.

scRNA-seq combined with hybridisation-based approaches

Computational analysis of scRNA-seq data allows the identification of distinct cell types present using published canonical markers. The analysis also allows identification of other cell-specific mRNA markers, usually 2–3, for each cell type, that could be used in hybridisation. Next, a set of cell type-specific probes is designed, labelled and hybridised on tissue sections using smFISH protocols (see below), and imaged by either wide-field or confocal microscopy to identify individual mRNAs [9].

The list of classic cell type-specific markers has been considerably enriched by information obtained in scRNA-seq, thus allowing extensive characterisation of cell types in some organs, such as the brain [10]. However, imaging the required number of markers in tissue remains particularly challenging and multiple factors (*i.e.* low expression level, high background due to off-target hybridisation or high tissue autofluorescence) can degrade signal quality. To circumvent these issues, different approaches, which may be combined, aim to: 1) amplify the actual RNA signal [11–15]; 2) increase the specificity of the probes and thus reduce off-target binding and background [16]; and/or 3) remove autofluorescence by tissue clearing [17, 18].

To reliably detect different cell types, several tens to hundreds of mRNA species have to be detected on the same histological sections. Nevertheless, the number of mRNA species that can be detected in conventional smFISH is limited by the number of spectrally resolved fluorophores, usually 3–4. To circumvent this limitation, sequential rounds of hybridisation are performed on microscopes equipped with microfluidic devices that allow, with adequate software, to fully automate imaging and buffer exchange.

TABLE 1 Comparison of the different spatial transcriptomic methodologies

	Hybridisation-based methods	High throughput strategies	*In situ* sequencing methods
Techniques	• seqFISH • osmFISH • HCR# • smHCR# • ClampFISH# • SABER# • PLISH# • SCRINSHOT# • RNAscope# (Bio-Techne)	• seqFISH+ • MERFISH	• *In situ* sequencing for RNA • FISSEQ • ExSeq • STARmap • Slide-seq/Slide-seqV2 • HDST • Visium (10XGenomics)
Imaging requirements	• Widefield (osmFISH, ClampFISH, SABER, SCRINSHOT, RNAscope) • Confocal (HCR, smHCR, PLISH) • Super-resolution (seqFISH)	• Widefield (MERFISH) • Confocal (seqFISH+)	• Widefield (Visium) • Confocal (FISSEQ, STARmap, Slide-seq/ Slide-seqV2, ExSeq)
Benefits	• High sensitivity • Cellular resolution • Low cost	• High throughput (up to 10 000 transcripts) • High sensitivity • Cellular resolution • Untargeted approach	• High throughput • Untargeted approach • *De novo* identification of spatial patterns
Limitations	• Low throughput (2–35 transcripts) • Targeted approach • Time consuming	• Technically challenging • High level of image analysis • Intermediate cost	• Low sensitivity • Low spatial resolution • High cost
Applications in the lung	• PLISH: comparison of AT2 distribution of idiopathic pulmonary fibrosis *versus* control samples, automated cell type identification • SCRINSHOT: spatial mapping of tracheal cell types	NA	• *In situ* sequencing for RNA: spatial mapping of 34 immune markers in granulomas present in *Mycobacterium tuberculosis*-infected lungs

#: with fluorescent signal amplification.

For instance, in the cyclic-ouroboros smFISH method, three sets of probes targeting three different mRNAs are hybridised on a tissue section. After imaging, probes are stripped using high formamide concentration. Next, three new probe sets are hybridised on the same section, imaged and stripped, allowing the number of targets to linearly increase with the number of hybridisation rounds [10]. Computational analysis then uses probabilistic models to infer the cell types in the images when compared to the expression levels measured in previously obtained scRNA-seq.

Multiplexing strategies for hybridisation-based approaches

To increase the number of transcripts that can be detected in RNA FISH, more complex strategies using specific encoding strategies have been developed in the recent years. Here, a given mRNA is not uniquely identified with a single hybridisation but with a unique barcode established over several hybridisation rounds. This permits several species to be targeted simultaneously in each round. The mRNA expression is then computationally inferred, often with error-correction strategies being included in the encoding schemes. Among these approaches are Multiplexed Error-Robust FISH (MERFISH) and seqFISH, which have been improved over several iterations [17, 19, 20]. For instance, with seqFISH+, 10 000 different mRNAs can be imaged in brain sections, allowing precise cell type identification and molecular characterisation in their spatial organisation [21].

In situ *sequencing strategies*

These approaches aim to perform sequencing directly on histological tissue sections. Methods such as FISSEQ (fluorescent *in situ* sequencing) and STARmap (spatially resolved transcript amplicon readout mapping) reverse transcribe the mRNAs within cells and amplify them before sequencing [22–24]. Other approaches, such as Slide-seq/Slide-seqV2 and HDST (high definition spatial transcriptomic) analysis, use coverslips or slides coated with barcoded oligo(dT) to capture mRNAs from tissue sections while recording their position at the same time [25–28]. After next generation sequencing, decoding of the barcodes allows visualisation of expression patterns of mRNAs present on the slide. A recent approach, ExSeq, combines spatial expansion of the biological sample with RNA sequencing, permitting highly multiplexed mapping of RNAs in tissue sections [29]. Compared to the hybridisation-based approach that targets a relatively limited group of markers with high sensitivity, the unbiased approach of sequencing-based methodologies should theoretically depict the whole transcriptome landscape of every cell in the tissue and thus it holds great promise. However, some limitations still need to be overcome, such as the lack of sensitivity and the fact that the spatial resolution does not classically reach the single cell level.

Altogether, spatial transcriptomic methodologies have the potential to lead to an unprecedented understanding of tissue architecture and physiology. However, these technologies remain challenging and researchers need to evaluate what fits best their needs in terms of resolution and throughput (table 1).

Spatial transcriptomics in the lung and future applications

Three spatially resolved transcriptomic studies have been performed in the lung so far. First, PLISH (proximity ligation *in situ* hybridisation) is a FISH-based approach relying on probes that act after a ligation step as a template for signal amplification. This allows to achieve high specificity, high sensitivity and high signal-to-noise ratio [30]. The PLISH protocol works on cryo-preserved as well as formalin-fixed paraffin-embedded human tissues and can be combined with classical immunostaining, making it suitable for clinical studies and diagnosis. As a proof of concept, PLISH was used to compare the spatial distribution of AT2 cells in control and IPF human lung tissue. In addition, the authors demonstrated that images from iterative staining for classical markers of epithelial cells (*e.g.* AT1, AT2, club cells) and macrophages fed into a dedicated analysis pipeline allowed an automated mapping of these cells onto the histological lung section.

Second, SCRINSHOT (single cell resolution *in situ* hybridisation on tissues) is validated in different organs, including the respiratory tract [31]. It uses padlock probes, which have complementary arms to the target mRNA sequence and a common backbone. After hybridisation, padlock probes are ligated using the SplintR ligase to create circular single-stranded DNA molecules that serve as a template for amplification. Staining and analysis of 29 markers by SCRINSHOT allowed mapping of the distinct epithelial cells present in the tracheal epithelium, including the recently discovered ionocytes [1, 2]. In the distal airways and alveoli, 15 markers were used to robustly identify macrophages and epithelial cells, such as AT1 and AT2 cells, as well as club and neuroendocrine cells.

Third, Carow *et al*. [32] used an *in situ* sequencing technique to localise 34 immune transcripts in lung tissue infected by *Mycobacterium tuberculosis*. Briefly, RNAs present in the lung section are reverse

transcribed and padlock DNA probes specific to the targeted transcripts are hybridised. A ligation step allows the circularisation of the probes required for the amplification and then, the amplified products are subjected to sequencing. Here, spatial transcriptomics allowed mapping of immune cells that characterise tuberculosis granulomas.

The spatial transcriptomics field has evolved very rapidly in recent years. With community-driven efforts such as *Tabula Muris* and The Human Cell Atlas that prioritise the respiratory system [6, 33, 34], lung cell classification is being refined and molecular markers for each cell type are now available from lung scRNA-seq datasets. Analysis of these cell type-specific markers with spatial transcriptomics will allow a better characterisation of the physiological interactions between cell types as well as their alterations in respiratory diseases, providing key insights in the understanding of their physiopathology.

In clinical practice, spatial transcriptomics, with an effort to standardise methods and create automated analysis pipelines, may eventually provide a useful complementary tool to analyse molecular markers or spatial patterns specific of lung pathologies on histological sections.

This article has been revised according to the correction published in the February 2023 issue of the *European Respiratory Journal*.

Acknowledgements: The authors would like to thank Arnaud Mailleux, Bruno Crestani, Nicolas Girard and Marine Lefèvre, Sarah Lagha, Simon Lefranc and Agathe Seguin-Givelet for their kind help and fruitful discussions.

Conflict of interest: None declared.

Support statement: We acknowledge financial support from Agence National de la Recherche (ANR-14-CE36-0008-02 program), the Comprehensive Cancer Center "SIRIC" program of Institut Curie (grant INCa-DGOS-4654) as well as the grants ANR-10-EQPX-03 (Equipex), ANR-10-INBS-09-08 (France Génomique Consortium) from the Agence Nationale de la Recherche (Investissements d'Avenir program) for the NGS core facility of Institut Curie. S. Curras-Alonso was supported by the Institut Curie 3-i PhD program under the Marie Skłodowska-Curie grant agreement number 666003. Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Montoro DT, Haber AL, Biton M, *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* 2018; 560: 319–324.
2 Plasschaert LW, Žilionis R, Choo-Wing R, *et al.* A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* 2018; 560: 377–381.
3 Travaglini KJ, Nabhan AN, Penland L, *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 2020; 587: 619–625.
4 Gillich A, Zhang F, Farmer CG, *et al.* Capillary cell-type specialization in the alveolus. *Nature* 2020; 586: 785–789.
5 Adams TS, Schupp JC, Poli S, *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* 2020; 6: eaba1983.
6 Schiller HB, Montoro DT, Simon LM, *et al.* The human lung cell atlas: a high-resolution reference map of the human lung in health and disease. *Am J Respir Cell Mol Biol* 2019; 61: 31–41.
7 Raj A, van den Bogaard P, Rifkin SA, *et al.* Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* 2008; 5: 877–879.
8 Tsanov N, Samacoits A, Chouaib R, *et al.* smiFISH and FISH-quant – a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res* 2016; 44: e165.
9 Halpern KB, Shenhav R, Matcovitch-Natan O, *et al.* Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* 2017; 542: 352–356.
10 Codeluppi S, Borm LE, Zeisel A, *et al.* Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat Methods* 2018; 15: 932–935.
11 Choi HMT, Chang JY, Trinh LA, *et al.* Programmable *in situ* amplification for multiplexed imaging of mRNA expression. *Nat Biotechnol* 2010; 28: 1208–1212.
12 Choi HMT, Beck VA, Pierce NA. Next-generation *in situ* hybridization chain reaction: higher gain, lower cost, greater durability. *ACS Nano* 2014; 8: 4284–4294.
13 Dirks RM, Pierce NA. Triggered amplification by hybridization chain reaction. *Proc Natl Acad Sci USA* 2004; 101: 15275–15278.
14 Kishi JY, Lapan SW, Beliveau BJ, *et al.* SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat Methods* 2019; 16: 533–544.

15 Rouhanifard SH, Mellis IA, Dunagin M, *et al.* ClampFISH detects individual nucleic acid molecules using click chemistry-based amplification. *Nat Biotechnol* 2019; 37: 84–89.
16 Goh JJL, Chou N, Seow WY, *et al.* Highly specific multiplexed RNA imaging in tissues with split-FISH. *Nat Methods* 2020; 17: 689–693.
17 Moffitt JR, Hao J, Wang G, *et al.* High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence *in situ* hybridization. *Proc Natl Acad Sci USA* 2016; 113: 11046–11051.
18 Shah S, Lubeck E, Schwarzkopf M, *et al.* Single-molecule RNA detection at depth by hybridization chain reaction and tissue hydrogel embedding and clearing. *Development* 2016; 143: 2862–2867.
19 Lubeck E, Coskun AF, Zhiyentayev T, *et al.* Single-cell in situ RNA profiling by sequential hybridization. *Nat Methods* 2014; 11: 360–361.
20 Chen KH, Boettiger AN, Moffitt JR, *et al.* Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 2015; 348: aaa6090.
21 Eng C-HL, Lawson M, Zhu Q, *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 2019; 568: 235–239.
22 Ke R, Mignardi M, Pacureanu A, *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* 2013; 10: 857–860.
23 Lee JH, Daugharthy ER, Scheiman J, *et al.* Fluorescent *in situ* sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc* 2015; 10: 442–458.
24 Wang X, Allen WE, Wright MA, *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*; 2018; 361: eaat5691.
25 Rodriques SG, Stickels RR, Goeva A, *et al.* Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 2019; 363: 1463–1467.
26 Ståhl PL, Salmén F, Vickovic S, *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016; 353: 78–82.
27 Vickovic S, Eraslan G, Salmén F, *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat Methods* 2019; 16: 987–990.
28 Stickels RR, Murray E, Kumar P, *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol* 2020; 39: 313–319.
29 Alon S, Goodwin DR, Sinha A, *et al.* Expansion sequencing: spatially precise *in situ* transcriptomics in intact biological systems. *Science* 2021; 371: eaax2656.
30 Nagendran M, Riordan DP, Harbury PB, *et al.* Automated cell-type classification in intact tissues by single-cell molecular profiling. *eLife* 2018; 7: e30510.
31 Sountoulidis A, Liontos A, Nguyen HP, *et al.* SCRINSHOT enables spatial mapping of cell states in tissue sections with single-cell resolution. *PLoS Biol* 2020; 18: e3000675.
32 Carow B, Hauling T, Qian X, *et al.* Spatial and temporal localization of immune transcripts defines hallmarks and diversity in the tuberculosis granuloma. *Nat Commun* 2019; 10: 1823.
33 Regev A, Teichmann SA, Lander ES, *et al.* The human cell atlas. *eLife* 2017; 6: e27041.
34 Schaum N, Karkanias J, Neff NF, *et al.* Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*. *Nature* 2018; 562: 367–372.

ANNEXE VII – A point cloud segmentation framework for image-based spatial transcriptomics (*publication*).

communications biology

Article

https://doi.org/10.1038/s42003-024-06480-3

A point cloud segmentation framework for image-based spatial transcriptomics

Check for updates

Thomas Defard[1,2,3,4,5], Hugo Laporte[6,7], Mallick Ayan[6], Juliette Soulier[6], Sandra Curras-Alonso[6], Christian Weber[4,5], Florian Massip[1,2,3], José-Arturo Londoño-Vallejo[6], Charles Fouillade[6], Florian Mueller[4,5] ✉ & Thomas Walter[1,2,3] ✉

Recent progress in image-based spatial RNA profiling enables to spatially resolve tens to hundreds of distinct RNA species with high spatial resolution. It presents new avenues for comprehending tissue organization. In this context, the ability to assign detected RNA transcripts to individual cells is crucial for downstream analyses, such as in-situ cell type calling. Yet, accurate cell segmentation can be challenging in tissue data, in particular in the absence of a high-quality membrane marker. To address this issue, we introduce ComSeg, a segmentation algorithm that operates directly on single RNA positions and that does not come with implicit or explicit priors on cell shape. ComSeg is applicable in complex tissues with arbitrary cell shapes. Through comprehensive evaluations on simulated and experimental datasets, we show that ComSeg outperforms existing state-of-the-art methods for in-situ single-cell RNA profiling and in-situ cell type calling. ComSeg is available as a documented and open source pip package at https://github.com/fish-quant/ComSeg.

Understanding the spatial organization of tissues at the single-cell level is crucial to study disease and development[1–3]. Molecular profiling of single cells in their spatial context allows us to infer cell states and cell types, cell–cell interactions, and cell-fate decision-making[4], as well as the study of the overall tissue architecture, under normal and diseased conditions, leading to the definition of spatial domains and disease signatures[5].

Spatial transcriptomics denotes a large panel of technologies that allow to measure gene expression and to retrieve the spatial positions of the mRNA molecules. These methods can largely be divided into two groups: sequencing-based[6,7] and imaging-based[8–10]

spatial transcriptomics (SST and IST, respectively). The former measures the entire transcriptome, but has a lower spatial resolution, while the latter relies on measuring a subset of genes at high resolution. Such a subset consists of pre-defined marker genes that are specific for the cell states or types of interest. Since IST captures only a subset of the transcriptome, we will refer to these approaches as RNA profiling. These imaging-based approaches are variants of single-molecule fluorescence in situ hybridization (smFISH) methodologies and can detect RNAs with single-molecule sensitivity with high spatial resolution, several orders of magnitude below the scale of a single cell. A typical dataset consists of 2D or 3D point clouds of the imaged RNA species. One challenge in the analysis of these data is the correct assignment of each RNA to its cell of origin. Indeed, in contrast to single-cell RNA sequencing (scRNA-seq), the information on which RNA molecules belong to the same cell has to be inferred from the image. This assignment is crucial, as it impacts cell type identification and, thus, the major aspect of the analysis we would like to perform.

One frequently used approach to perform this assignment is to segment the cells from additional channels employing markers for cell and nucleus segmentation. RNAs are then assigned to the cells based on their spatial positions with respect to this segmentation. Such stainings typically encompass labeling of the nucleus with DAPI, cellular staining with one or more cell membrane dyes, or labeling all RNAs as a proxy for the cytoplasm[10–12]. However, cell membrane staining is often not an option[13,14]. Besides, staining can be technically complex[15], and may not work equally well for the entire tissue, thus leading to inhomogeneous cell segmentation and cell type calling accuracy. Further, the boundaries of individual cells can be challenging to segment, especially for tissue with complex 3D cellular morphologies.

More recently, several computational approaches have been presented to detect individual cells in the images and establish their RNA profile without relying on a dedicated cell marker. These methods rely only on the RNA positions and in some cases a DAPI stain, to regroup RNAs according

[1]Centre for Computational Biology (CBIO), Mines Paris, PSL University, 75006 Paris, France. [2]Institut Curie, PSL University, 75005 Paris, France. [3]INSERM, U900, 75005 Paris, France. [4]Institut Pasteur, Université Paris Cité, Imaging and Modeling Unit, F-75015 Paris, France. [5]Institut Pasteur, Université Paris Cité, Photonic Bio-Imaging, Centre de Ressources et Recherches Technologiques (UTechS-PBI, C2RT), F-75015 Paris, France. [6]Institut Curie, Inserm U1021-CNRS UMR 3347, University Paris-Saclay, PSL Research University, Centre Universitaire, Orsay, Cedex, France. [7]Institute of Cell Biology (Cancer Research), University Hospital Essen, Essen, Germany. ✉e-mail: fmueller@pasteur.fr; Thomas.Walter@minesparis.psl.eu

to local transcription profiles. Such RNA point clouds can then act as an approximation for cell shape, by establishing a hull that encompasses all RNAs that were deemed to belong together[15–17]. Methods like pciSeq[16] and Baysor[15] leverage statistical models to group RNAs. pciSeq segments cells to match external scRNA-seq datasets and Baysor divides RNAs into spatially homogeneous areas. Recently, deep-learning approaches were proposed. JSTA[17] and BIDCell[18] train a cell membrane segmentation model exploiting nuclei segmentation and external scRNA-seq while SCS trains a membrane segmentation transformer network using solely nuclei segmentation as input. While these approaches hold great promise for the analysis of spatial RNA profiling data, their use is limited by implicit priors on cell shape[15,16] leading to low performance in complex tissues, requirement of additional scRNA-seq data[16,17] or manual annotations which are scarce and error-prone[18,19] (see Table 1).

While there already exist methods for spatial RNA profiling, these methods usually come with requirements that are not always met in practice. For instance, some methods require a high RNA density, which often implies a large panel of marker genes. However, while current commercial spatial RNA profiling approaches provide hundreds of marker genes, they remain extremely costly, and for many questions, fewer marker genes will be sufficient, which can be probed with simpler custom-built solutions[20,21]. Furthermore, most methods[15,16] implicitly assume convex or even round cell shapes. In contrast, the cell shapes in some tissues can be complex and often deviate from such simple shapes. Approaches relying on strong assumptions on cell shape are, therefore, suboptimal for many of these tissue types. Lastly, some methods require parallel scRNA-seq data[16–18], which in practice is not always readily available, so ideally the use of such information should be only optional.

To alleviate these issues, we propose a method named ComSeg. ComSeg uses as input only the coordinates of the RNA molecules and the nucleus positions obtained with staining such as DAPI. ComSeg then groups RNAs with similar expression profiles aided by these landmarks. It does not require scRNA-seq data or cytoplasmic markers, nor does it make implicit use of any prior assumptions regarding cell morphology. Instead, the method describes RNA point clouds as graphs weighted by gene co-expression and relies on graph community detection. Our method is easy to apply by design as it does not require complex machine learning model training. Furthermore, we provide the tool as an open-source Python package (https://github.com/fish-quant/ComSeg) with extensive documentation: https://comseg.readthedocs.io, compatible with the scverse environment[22]. To facilitate the application of ComSeg on large datasets, we ensured its compatibility with SOPA[23], as documented on https://comseg.readthedocs.io.

Development of such analysis approaches requires annotated ground truth to assess their performance. Experimental ground truth can be obtained in some cases by employing membrane markers, from which the cytoplasmic membrane might be segmented with deep neural networks or manual annotation[24]. However, the ground truth quality is affected by the heterogeneity of the staining quality which may be low on complex and dense tissue[15]. Furthermore, experimental data does not permit a more systematic exploration of parameters such as the expression level or cell morphology. Hence, similarly to previous studies[16,17], we address the lack of high-quality ground truth by generating simulated data, allowing us to control the complexity of the input data. We developed *SimTissue* (https://github.com/tdefa/SimTissue), an open-source Python simulation package to reproduce in silico fluorescent-based spatial transcriptomic experiments.

We used the *SimTissue* framework to validate ComSeg on simulated data with increasing complexity in terms of RNA abundance, the number of marker genes, and tissue morphology. ComSeg outperforms other methods in terms of the Jaccard index for RNA-cell association in most of the tested scenarios. We also validated ComSeg on an in-house experimental dataset of mouse lung tissue imaged with smFISH and human embryonic lung tissue[25] imaged with HyBISS[26]. On these experimental data, ComSeg estimates accurate RNA profiles that match established scRNA-seq datasets. Finally, we benchmark ComSeg on two MERFISH[8] datasets with cell membrane staining, the mouse ileum dataset[15] and a human breast cancer dataset from VizGen from the MERSCOPE FFPE Human Immuno-Oncology Data Release. We leverage areas with high-quality membrane staining to obtain segmentation ground truth. Overall, the shape-agnostic approach of ComSeg demonstrates notable efficacy for complex tissue composed of cells with non-convex shapes.

Results

ComSeg overview

Here we present ComSeg, a graph-based method to perform cell segmentation from spatial RNA profiling data. The method operates directly on RNA point clouds and leverages nuclear staining to increase segmentation quality. For this, we define a KNN graph, where each RNA molecule is a node and where edges are weighted according to the co-expression score of the corresponding genes. Instead of relying on external data to compute these co-expression weights, we leverage the input images by estimating the local gene expression vector in a close environment of each RNA molecule (see Material and Methods). We then detect groups of RNA molecules with similar gene expression in their local environment, by the use of a modified version of the Louvain community detection method[27]. Our community detection method can leverage spatial landmarks like DAPI segmentation as prior knowledge. Lastly, we group the detected RNA communities with similar expression profiles using the Leiden algorithm[28] and obtain a transcriptomic domain map at the tissue scale. To obtain single-cell RNA profiles, we exploit the nucleus position concurrently with the transcriptomic domain map. An overview of the method is displayed in Fig. 1a.

State-of-the-art methods for cell segmentation

We benchmarked ComSeg against methods that can be used in an equivalent setting i.e., single-cell spatial RNA profiling approaches requiring no external dataset. As a first baseline, we calculated the Watershed transformation[29] as this method is often used for RNA-nuclei association in tissue[30,31]. This method effectively calculates a Voronoi tessellation with the nuclear regions as markers with an additional distance constraint, and is thus equivalent to assigning RNA transcripts to their nearest nucleus. The method, hence, perfectly works for convex cells of homogeneous size if all nuclei are detected, but may fail otherwise. Second, we benchmarked Baysor[32], a cell marker-free segmentation method optimizing the joint likelihood of transcriptional composition and prior cell morphology. It is particularly suited for cases where only nuclear staining or weak cytoplasmic staining are available. Baysor uses an elliptic function as cell shape prior. We also applied a method leveraging external scRNA-seq data, pciSeq[16], a Bayesian model leveraging prior scRNA-Seq data to estimate a probability of cell assignment for each read. Given the observed RNA spot configuration, the method performs cell assignment to match known transcriptomic profiles from scRNA-seq. Hence, it also simultaneously assigns each cell to a cell type. Leveraging prior knowledge may help pciSeq avoid wrong

Table 1 | Characteristics of existing methods for spatial RNA profiling

	Watershed	pciSeq[16]	JSTA[17]	SSAM[32]	Baysor[15]	BIDCell[18]	SCS[19]	ComSeg
Convex shape prior	Yes	Yes	Yes	No	Yes	Yes	No	No
Require external scRNA-seq dataset	No	Yes	Yes	No	No	Yes	No	No
Require nucleus positions	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Single-cell profiling	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes

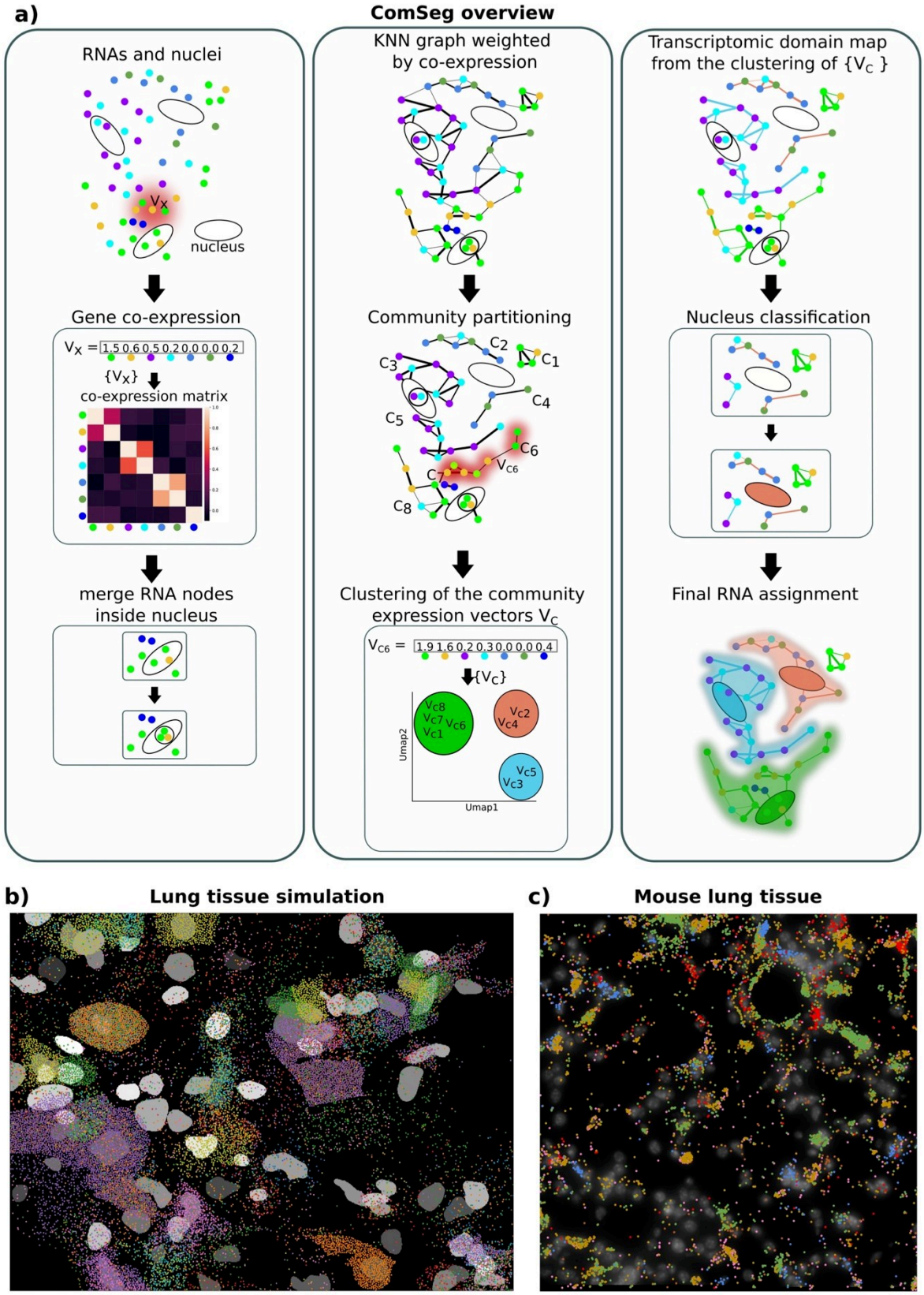


Fig. 1 | Overview of the method and simulated and experimental data. a Overview of the ComSeg method. **b** 3D lung tissue simulation with 34 marker genes represented by different colors. **c** Mouse Lung tissue from our in-house dataset with six marker genes represented by different colors.

RNA-cell assignment. However, pciSeq implicitly uses a spherical cell shape prior that may hinder its application on complex tissue. Finally, we benchmarked our method against SCS[19], a recently published deep-learning-based method primarily designed for Stereo-seq data, but also applicable to IST data. The method trains transformers to predict, for each transcriptomic spot measurement, the direction to the nucleus it belongs to.

All of these methods require hyperparameter tuning. For the Watershed method, we had to set the maximum distance parameter to avoid the

incorrect assignment of RNA to nuclei that were too distant. This maximum distance parameter had to be optimized to account for differences in tissue complexity and cell density. For pciSeq and Baysor, we kept the default settings for all experiments, with the exception of the scaling parameter for Baysor, which was automatically set using prior nucleus segmentation. For the analysis of the mouse ileum with Baysor, we used the compartment-specific gene list provided[15] as additional parameters. SCS was trained over 100 epochs with a bin size of 15 pixels for all the experiments. Lastly, the mean cell size parameter of ComSeg was manually adjusted for each dataset. More details regarding hyperparameter settings of the benchmarked methods can be found in Supplementary Note 1.

Benchmark on simulation of lung tissue

To benchmark the methods on challenging data, we first turned to simulations mimicking lung tissue (see Methods). Here, cells have complex shapes in 3D, and airways add empty space devoid of any transcripts. Moreover, we also simulated some cells without nuclei, which can occur in tissue sections. Further, we sampled real expression profiles from our recent scRNA-seq data[33]. We simulated 34 cell-type marker genes selected with the NS-forest algorithm[34]. This marker list enabled us to classify 19 different cell types present in lung tissues with an accuracy of 0.88 for cell type calling when having a perfect RNA assignment to cells (Methods).

To quantitatively compare the different methods, we implemented different metrics (see Methods). We assessed the quality of the RNA-cell assignment with the mean Jaccard index per cell, which is calculated for the RNA positions.

The ultimate goal of all the methods is to obtain a gene expression vector for each cell. In this context, not all types of RNA-cell association errors are equivalent. If a cell misses some RNAs, it may still be possible to accurately assess its RNA profile, similar to cell type classification in scRNA-seq, where only a fraction of all RNAs are sequenced. However, if some RNAs are wrongly associated with a cell, this can create a mixed expression profile resulting in incorrect cell type classification. We thus reported the mean percentage of wrongly associated RNA per cell (WA) and the mean percentage of missing RNA per cell (MS) separately. Lastly, we assessed cell type calling accuracy by comparing each method's results to the ground truth cell type defined by scRNA-seq data. To classify cell types, we computed the cosine distance between predicted cell expression vectors derived from images and the scRNA-seq cell type cluster centroid (see Methods).

Examples of RNA-cell assignment from the different approaches on lung tissue simulation are shown in Fig. 2a. ComSeg outperforms Baysor, pciSeq, SCS, and the Watershed algorithm with a Jaccard index of 0.57 against 0.30, 0.33, 0.21, and 0.50 for Baysor, pciSeq, SCS, and Watershed, respectively (Fig. 2b, left panel). The lower Jaccard index of SCS could be attributed to its original design for sequencing-based data with access to the complete transcriptome, whereas in our case, we only simulate a limited number of markers. Furthermore, SCS does not utilize the 3D information provided by the simulation and the spatial resolution must be reduced to accommodate SCS to image data. A closer look at the results reveals that the type of error is not the same for the four models. Baysor, SCS and pciSeq have a high percentage (more than 50%) of missing RNA per cell (Fig. 2c, right panel) while Watershed has a very low (15%) mean percentage of missing RNA per cell. This is not surprising, as the Watershed computes a Voronoi tessellation and hence assigns all RNAs except those that are very far from nuclei. Watershed, SCS, and pciSeq have a higher mean percentage of wrongly associated RNA per cell (about 40%), while it is low for Baysor and ComSeg (roughly 20% on average) (Fig. 2c, left panel).

Next, we compared the estimated expression profiles from the simulated images to the known, underlying scRNA-seq ground truth, by performing cell type calling. Here, ComSeg reaches 74% accuracy, compared to 60, 62, 35, and 62% for Watershed, pciSeq, SCS, and Baysor respectively (Fig. 2b, right panel). The higher accuracy of ComSeg for cell-type calling can be explained by the lower misassociation error rate (Fig. 2c, left panel). Of note, the best cell type calling accuracy that could be achieved is 88%, which is thus the value that would be reached if all RNAs were correctly assigned to the cell they belong to. This value is not 100% because the selected marker genes do not perfectly recapitulate the full transcriptomic profiles. On the other hand, we observe that cell-type calling is heavily impacted by wrong RNA assignments.

In conclusion, ComSeg performs better in terms of RNA-cell association and substantially improves downstream tasks like cell type calling on lung tissue simulation compared to the current state-of-the-art. In view of these results, we next turned in Supplementary Note 2 to simulations with a simpler tissue geometry, in order to better understand the limitations of each method.

Application to experimental data without cell membrane staining

We applied ComSeg to two different experimental lung datasets with solely nuclei staining, and each with a specific challenge for the analysis. The first dataset was created in-house. In this experiment, we visualized 6 different marker genes in 3D mouse lung tissue. Our approach has a very high RNA detection efficiency, but many cells display no RNA transcripts as only a subset of cell types is targeted. The second dataset is from a recent study of human embryonic lung mapping 147 genes in 2D[25]. The HybISS approach used for this dataset has a lower capture rate[15] but enables the visualization of many more genes.

For such experimental data, no ground truth is available. In tissue with complex morphology such as lung, having no ground truth makes it particularly challenging to assess the method's quality. An existing validation strategy is based on the gene expression correlation between the overlap region provided by different methods and the non-overlapping regions[15,19]. However, this validation does not compare with respect to a ground truth. Errors in existing methods could thus be propagated to the next generation of methods. Moreover, the method implicitly assumes a homogeneous spatial distribution of RNAs, including for the nuclear region. In the absence of direct access to ground truth for these imaging datasets, we chose to leverage available scRNA-seq datasets obtained from identical organ samples. These scRNA-seq datasets serve as a means to assess the consistency of single-cell spatial RNA profiling. As for simulations, we calculate the cosine distance between cell expression vectors derived from images and the nearest scRNA-seq cluster centroid. Consequently, the cosine distance between single-cell RNA profiles from the image dataset and scRNA-seq clusters serves as a surrogate measure for the quality of single-cell spatial RNA profiling.

Similarly to what we did with simulations, we applied SCS, Baysor, pciSeq, Watershed and, ComSeg on the 3D mouse lung tissue dataset (Fig. 3a) and on the 2D embryonic lung tissue dataset.

On mouse lung tissue, ComSeg identifies more cells than the other tested methods, for which an RNA profile can be assigned (Supplementary Fig. 3). We obtain similar results for human embryonic lung tissue (Supplementary Fig. 3) where ComSeg also detects more cells than the four other methods we tested. We took into account only cells with more than five RNAs. Importantly, the number of cells with matching expression profiles in the scRNA-seq data is higher for ComSeg both for mouse (Fig. 3b) and human lung tissue (Fig. 3c) at the different cosine distance thresholds.

In summary, ComSeg detects more cells than other methods, and our analysis revealed that the RNA profiles measured in the segmented cells also better fit external datasets, thus suggesting better segmentation quality.

Application to experimental data with cell membrane staining

In order to investigate the performance of experimental data, we applied ComSeg to two publicly available MERFISH datasets, containing both DAPI and membrane staining. While the cell membrane staining does not cover all cells, it provides a valuable ground truth in some parts of the image.

The first dataset comprises a mouse ileum tissue section measuring $400 \times 600\ \mu m$[15] with 241 genes. This dataset presents challenges, as mouse ileum is known to contain RNAs with preferential intracellular distributions (Fig. 4a and Supplementary Fig. 4). The authors of this dataset leveraged a pan-cell-type cell surface marker to visualize the cell membrane. However, this surface marker tends to work more effectively in specific locations and

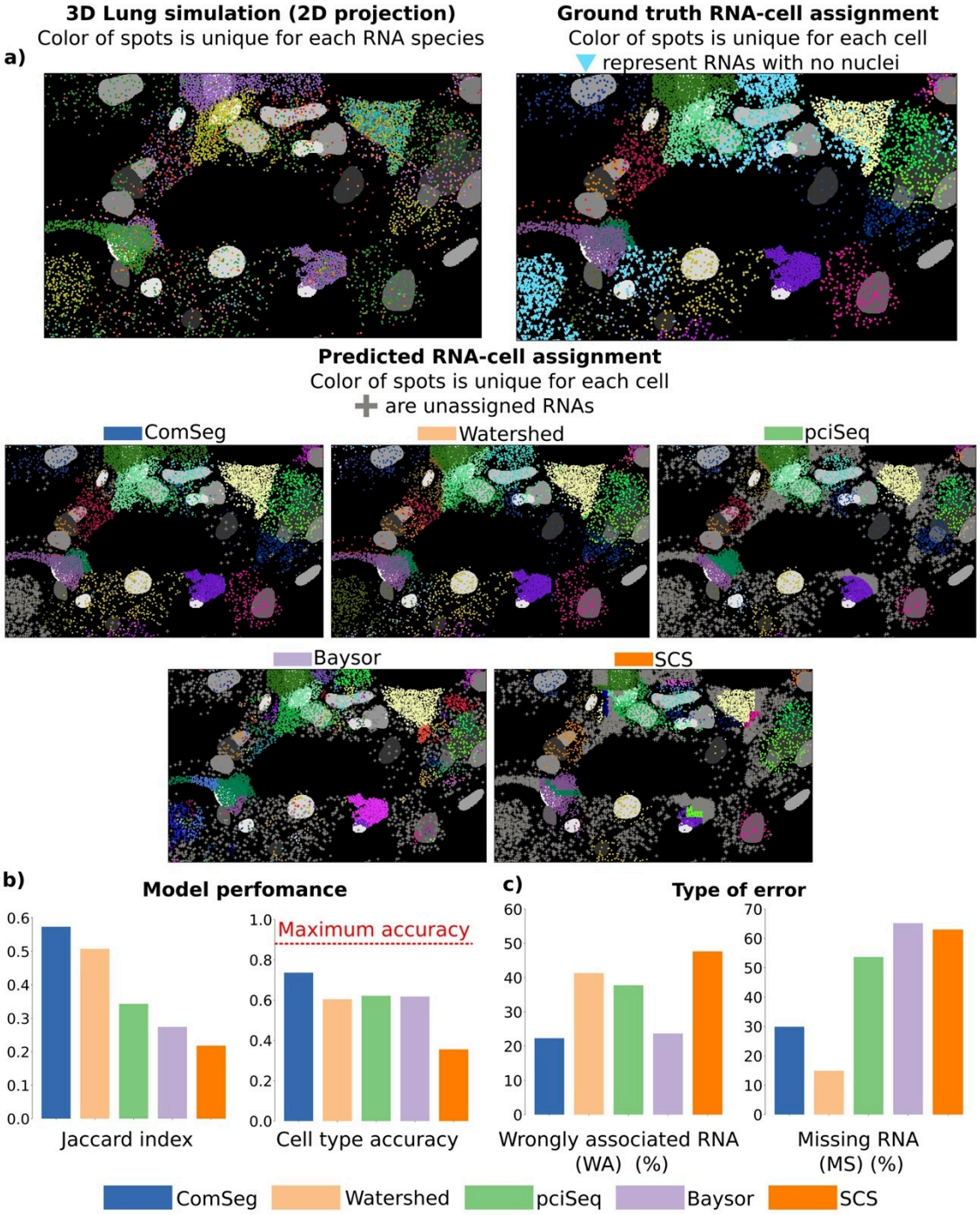


Fig. 2 | Benchmarking on simulated lung tissue. a 3D lung simulation and RNA assignment of the different models. The ground truth takes into account only cells with a nucleus. **b** Performance metrics of the benchmarked models with the mean Jaccard index per cell and cell type calling accuracy, the red line is the maximum accuracy when having a perfect RNA assignment to cells. **c** Error metrics of the benchmarked models with the mean percentage of wrongly associated RNA per cell (*WA*) and mean percentage of missing RNA per cell (*MS*).

cell types, as evidenced in Fig. 4a, where the membrane staining is more pronounced on the tissue periphery.

In addition, we applied the benchmarked methods to another MERFISH dataset featuring human breast cancer tissue. This dataset is part of the publicly released *MERSCOPE FFPE Human Immuno-Oncology Data Release* by the Vizgen company. It encompasses 550 genes spatially resolved over a section of approximately 10 mm × 8 mm. For the purpose of this study, we analyzed a 2D subsection measuring 1900 × 970 µm. Similarly to the previous dataset, it includes nuclei staining and membrane staining.

These two datasets exhibit partial membrane staining, which we used as annotated ground truth to evaluate the performance of the benchmarked method. To ensure a substantial number of annotated cells, we automated the annotation process as follows: first, we independently segmented the nuclei and the cell membrane using Cellpose[35]. Then we defined as ground truth the cells containing exactly one nucleus, thus removing cell segmentation errors likely to correspond to artifacts (e.g., visual patterns in the image that Cellpose erroneously detected as cells) and segmentation errors due to low-quality membrane stain, resulting in multiple nuclei per cell (e.g., by missing the separating membrane). As a result, we obtained a

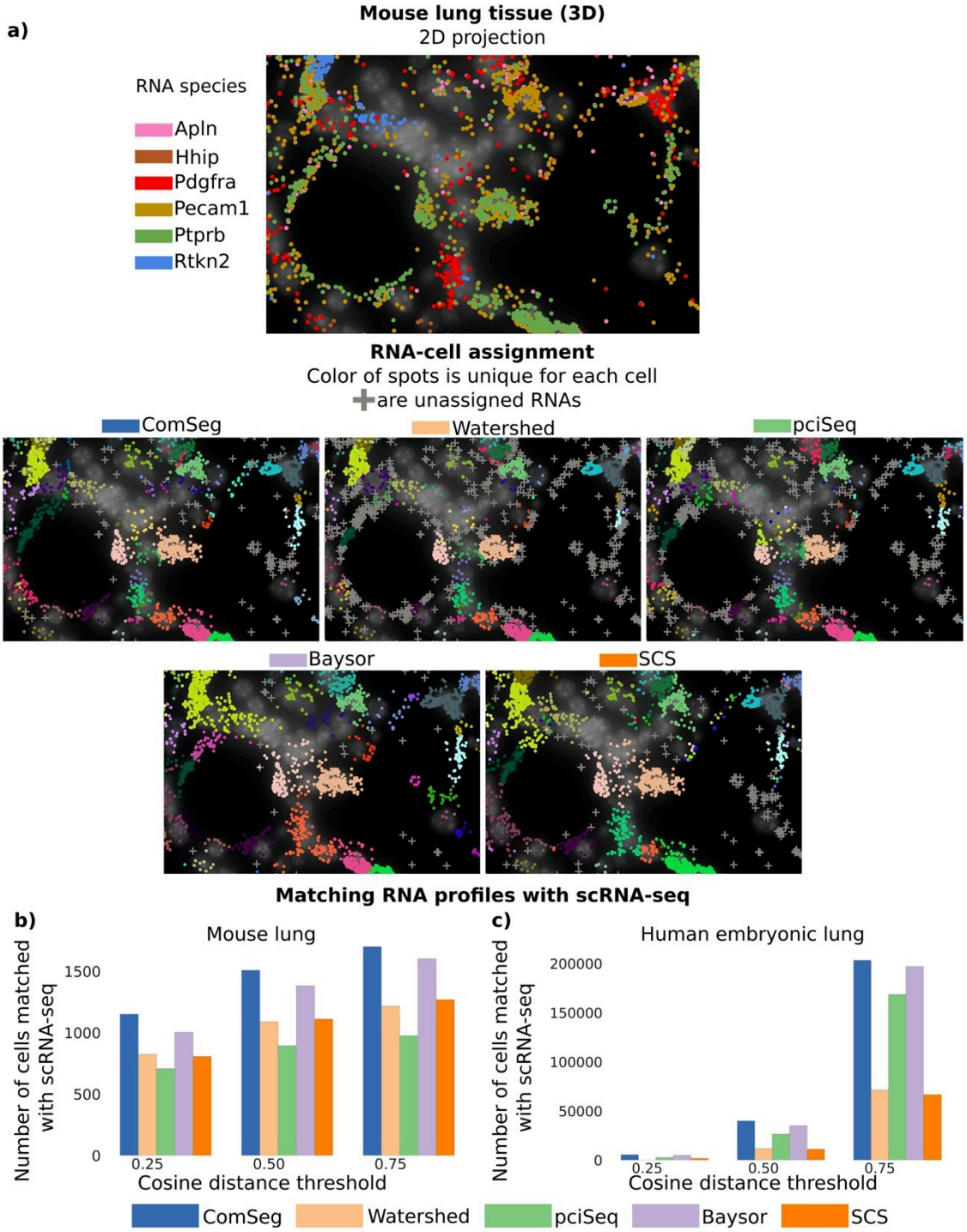


Fig. 3 | Benchmark on experimental data without cell membrane staining. **a** Mouse lung tissue and RNA assignment of the benchmarked models. **b**, **c** Number of single-cell RNA profiles identified in images matched with a scRNA-seq reference dataset at different cosine distance thresholds for mouse lung tissue (**b**) and human embryonic lung tissue (**c**).

high-confidence ground truth segmentation map (see Methods for details). The ground truth segmentation maps used in this study, as well as the corresponding segmentation map from the benchmarked methods, can be found at (https://zenodo.org/records/11237477). An example of the annotations generated using this process is depicted in Fig. 4a.

We used these high-confidence datasets to evaluate the performance of the benchmarked methods. We first associated each annotated cell with its best match from each benchmarked method (Fig. 4b). Subsequently, we calculated the Jaccard index on the RNA set. On the mouse ileum dataset, Watershed and ComSeg performed slightly better than pciSeq and Baysor with a Jaccard index of around 0.4 against 0.36 and 0.31 for Baysor and pciSeq (Fig. 4c). In the breast cancer dataset, Baysor, pciSeq, Watershed, and ComSeg yield comparable Jaccard index, around 0.5 (Fig. 4d). Across both datasets, SCS exhibited a lower Jaccard index than the other methods, potentially due to the fewer resolved RNA species compared to the original study and the need for SCS to lower spatial resolution to be applied to IST data. Besides, the Jaccard index metric can hide disparities in terms of segmentation error as shown in the simulation section.

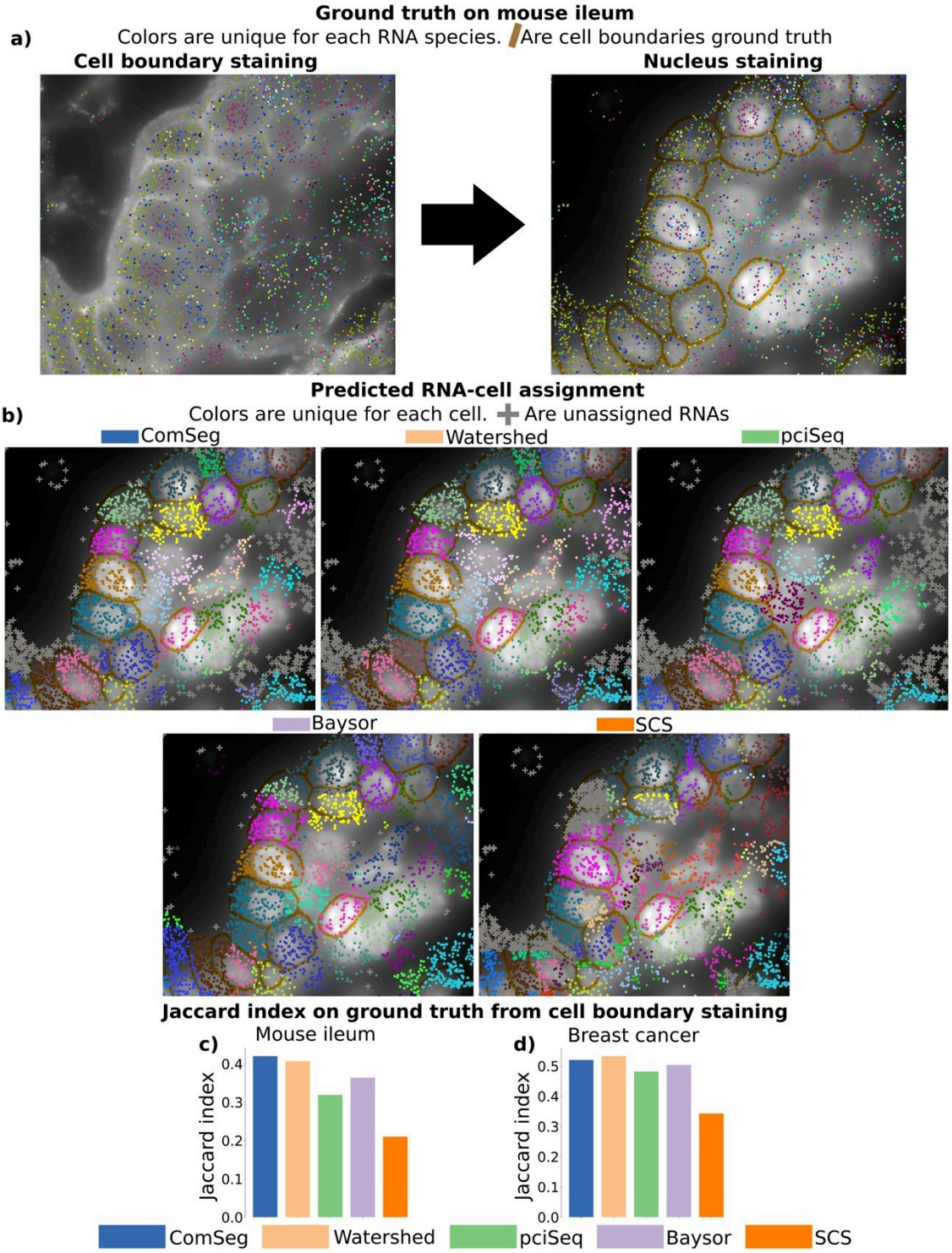


Fig. 4 | Benchmark on experimental data with partial cell membrane staining. **a** Mouse ileum tissue with Membrane staining (right) is employed to establish a segmentation ground truth (left). **b** RNA assignment of the benchmarked models on the mouse ileum dataset. **c**, **d** Jaccard index on the established ground truth from membrane staining for benchmarked models on mouse ileum tissue (**c**) and human breast cancer tissue (**d**).

In summary, this last section shows that ComSeg can handle datasets with a high number of genes and with RNA with preferential intracellular distribution like in the mouse ileum dataset (Supplementary Fig. 4).

Discussion

Imaging-based spatial RNA profiling methods provide RNA point cloud coordinates without information about the cell of origin of each molecule. Still, it is essential to correctly assign RNAs to cells to perform downstream tasks at the single-cell level such as cell type calling or inference of cell interaction. In this study, we present a method called ComSeg, a graph-based method operating directly on the RNA coordinates and nucleus positions. In addition, ComSeg is cell shape agnostic which makes it particularly useful for complex tissues, where cell shape convexity cannot be assumed. As such, ComSeg is a flexible method able to handle various situations, demonstrating good performances even in challenging experimental settings.

In order to compare ComSeg to other methods, we have developed a simulation framework (*SimTissue*) that provides tissue architectures which are generated from real images and are therefore reasonably realistic. Our simulation environment also encompasses simple geometric shapes which are ideally suited to study failure modes and limitations of algorithms thanks to the simplified geometry. Indeed, quality assessment is a critical aspect of segmentation, but actually difficult to perform. So far, there are no manually annotated ground truth datasets for imaging-based spatial RNA profiling. In this study, we use an automated workflow to create ground-truth data for cells with high-quality membrane stains. However, considering solely these cells for assessing spatial RNA profiling performance could introduce bias as membrane staining may work more effectively in specific locations of the tissue or only for certain cell types. Comparison to a consensus is an option that has been adopted by several authors[15,19]. We argue that this can lead to the propagation of systematic errors. For instance, all competing methods rely on shape priors that are not always met in complex tissues. A comparison to the consensus would not be able to reveal such systematic errors. Furthermore, in most cases, the consensus region will contain the nuclear region, and it is possible that the expression profile measured in the nuclear region is different from peripheral regions in the cell.

We found that most methods have difficulties with non-convex shapes and large differences in gene expression density. Our simulations suggest that error rates in cell-type calling are non-negligible: as much as 26% of cell-type assignments are erroneous because of segmentation errors when using previously published methods. ComSeg outperforms competing methods by a large margin on complex tissues. However, even with ComSeg, cell classification errors due to wrong segmentations amount to 14%, thus suggesting that the development of novel segmentation methods remains an important topic for imaging-based spatial RNA profiling data. Besides, as the number of methods and datasets continues to increase, finding optimal parameters for each of them to obtain fair comparison becomes increasingly challenging. We believe that independent benchmark studies for spatial RNA profiling data would be beneficial for the community.

Beyond these benchmark results, ComSeg has several other compelling aspects that make it interesting for the scientific community. First, it does not require membrane stainings. Indeed, existing membrane stainings are highly variable and not very robust. Moreover, they are inhomogeneous and can, therefore, lead to spatial biases in cell-type calling accuracies. In this context, methods like ComSeg, directly operating on RNA point clouds, complement cell membrane staining-based approaches like Cellpose, in cases where cell membrane staining is unfeasible, e.g., in regions of low staining quality or for certain cell types. Second, ComSeg does not require external datasets, such as scRNA-seq, which makes it also applicable in small-scale studies, where such data is not available. Moreover, ComSeg does not rely on annotated data, which is very tedious and sometimes impossible to provide. Its modular structure makes it easy to tailor ComSeg to particularities in the datasets. Lastly, ComSeg demands only basic desktop computational resources and exhibits reasonable runtime and memory requirements (see Supplementary Table 1).

One of the limitations of ComSeg is that it is dependent on the choice of cell type marker genes. The model may fail if the spatially resolved RNA species are not discriminative of cell type or cell state. A potential improvement of ComSeg would be to incorporate in the model several landmark stainings. Another potential enhancement could be to include more spatial information in the clustering of the RNA communities expression profiles. In ComSeg, RNA communities with similar expression profiles are grouped into clusters using the Leiden algorithm which leverages solely expression profiles. However recent clustering methods for spatial transcriptomic were proposed leveraging both spatial information and expression profile[36,37]. For instance, if an RNA community with a given transcriptomic profile is erroneously linked to another profile, this mistake could potentially be corrected by taking neighboring profiles into account. For instance, if the RNA community is surrounded by other communities from the same correct transcriptomic profile, a wrong assignment could be avoided. Finally, ComSeg relies on the position of single RNA molecules which prevents its application on sequencing-based spatial transcriptomic data with low spatial resolution. Indeed, the edge weights are defined with gene co-expression as similarity metrics. Applying ComSeg on sequencing-based data with low spatial resolution would require defining an expression vector similarity metric between binned spots which might be an interesting extension. Finally, ComSeg could, in principle, be used on sequencing-based data if the spatial resolution is close to IST, such as Stereo-seq[14].

Altogether, we believe that this model will be of great interest to the community and has the potential to overcome current shortcomings in cell segmentation and cell type calling from spatial RNA profiling data. To facilitate the use of ComSeg, we have made it available as an open-source and documented Python package: https://github.com/fish-quant/ComSeg. We also make the simulation framework *SimTissue* (https://github.com/tdefa/SimTissue) available to the community, which might help researchers in the future to benchmark their methods.

Material and methods

Description of ComSeg algorithm

We assume that we have for each cell its centroid. In practice, this cell centroid is inferred from the nuclear stain, which are available in virtually all IST datasets. ComSeg associates the detected RNA molecules with their corresponding cell centroid. It leverages a k-nearest neighbor (KNN) graph, where the nodes are the detected RNAs. The method can be decomposed into five steps:

1. Computation of a proximity-weighted expression matrix
2. Construction of KNN graph weighted by co-expression
3. Graph community detection using the cell nucleus.
4. Cell segmentation-free in situ clustering of communities
5. Final RNA assignment

The only hyper-parameters exposed to the user are the mean cell diameter D and the maximum cell radius $R_{\max}$. The numerical values of these hyper-parameters are detailed in Supplementary Note 1.

Proximity-weighted expression matrix

ComSeg leverages gene co-expression information. In principle, co-expression information can originate from parallel scRNA-seq data or from other published resources. However, such external datasets are not always available for the biological system under study. Hence, we estimate co-expression using only the spatial arrangements of detected RNA molecules in the image. For this, we leverage the spatial correlation between the different RNA species molecules as a proxy for gene co-expression.

For each RNA position x, we define a local proximity-weighted expression vector (PE) $V(x)$. For this, we consider a maximum of $K = 40$ neighbors y, positioned at a maximal distance of $\boldsymbol{R_{PE}} = \boldsymbol{D/2}$ from x. A subset of these neighbors are transcripts of the gene g. We thus define a local expression score of gene g in proximity of x:

$$V_g(x) = \Sigma_{y \in KNN(x), gene(y)=g} (\boldsymbol{R_{PE}} - ||x - y||)/R_{PE} \quad (1)$$

We note that each transcript at position y is weighted by a score that linearly decreases between x and the circle around x with radius $\boldsymbol{R_{PE}}$. The underlying idea is that close transcripts should contribute more to the local expression estimation.

The local proximity-weighted expression vector V(x) is then defined as:

$$V(x) = [V_{g1}(x), V_{g2}(x), \ldots, V_{gN}(x)] \quad (2)$$

Of note, the proposed PE are similar to the Neighborhood Composition Vectors proposed in ref. 15 with the difference that it weights RNA positions by a distance score, such that closer RNAs contribute more than RNAs that are far.

By stacking the $V(x)$ for all positions x, we get a proximity-weighted expression matrix $V \in \mathbb{R}^{N_x \times N_g}$ where N_x is the number of detected RNA and N_g the number of marker genes. From this, we can finally compute the

co-expression matrix $W \in \mathbb{R}^{N_g \times N_g}$, where each element $w_{i,j}$ is the Pearson correlation of the columns of V corresponding to the expression of genes i and j:

$$w_{i,j} = Corr(V_i, V_j) \quad (3)$$

The computation of these co-expression values is implemented by the Python class *ComSegDataset* of our package. Of note, the co-expressions could alternatively be computed from external data, e.g., from single-cell RNA sequencing data.

KNN graph weighted by co-expression

Our algorithm operates on a weighted KNN graph, where the RNA molecules (across all genes) are the nodes ($K = 10$ and the max edge distance $\boldsymbol{R_{knn}}$ between molecules is set to $\boldsymbol{D/4}$) and the weights are the estimated co-expression values $w_{i,j}$ defined in (3): edges between RNA molecules from strongly co-expressed genes obtain a large weight, while RNA molecules from genes that are not co-expressed are assigned a small weight. The rationale is that RNAs from co-expressed genes are likely to belong to the same cell, while RNA molecules from genes that are usually not expressed together are more likely to belong to different cells.

It is worth mentioning that we opted to keep the value of K fixed at K = 10. Indeed, modifying this parameter has a negligible impact on performance, as illustrated in Supplementary Fig. 5.

Graph community detection algorithm

In the previous section, we generated a graph strongly connecting RNA nodes that are likely to belong to the same cell. Now, our objective is to partition this graph into sets of RNAs belonging to the same cells. To achieve this, we developed a modified version of the Louvain method[27] for community detection so it can accommodate prior knowledge given by nuclei segmentation or other landmarks.

The original Louvain algorithm is a widely used method to partition a graph into sets of strongly connected nodes. These strongly connected sets of nodes are called communities. The algorithm optimizes a metric of graph structure called the modularity, noted Q[38]. Q is the sum of the differences between intra-community weights and their expected value in a randomly rewired graph. Q can be expressed as a sum over the edges (u,v) of the graph.

$$Q = \frac{1}{2m}\Sigma_{u,v}\left[w_{g_u g_v} - \frac{k_u k_v}{2m}\right]\delta(C_u, C_v) \quad (4)$$

Where g_u and g_v are the gene index of the RNA nodes u and v. $w_{g_u g_v}$ is the corresponding weight from the co-expression matrix W. k_u is the degree of node u defined as $k_u = \Sigma_v w_{g_u g_v}$, m is the sum of the weights in the network $m = \frac{1}{2}\Sigma_{u,v} w_{g_u g_v}$ and the function δ is 1 if the nodes u and v belong to the same community C (i.e., $C_u = C_v$) and 0 otherwise.

This method iterates two elementary phases: in the first step, modularity Q is greedily optimized. For this, we start from an initialization where each node is assigned to its own community. Nodes are then moved in a random order to neighboring communities to greedily maximize modularity. This first step stops when no node move improves the modularity Q. The randomness of this phase has a negligible effect on the final ComSeg output as studied in Supplementary Note 3.

In the second step, a new aggregated network is built where communities found in the first step become nodes. Edge weights between those new nodes are calculated as the sum of the edge weights between the identified communities. We can then re-apply the first step on this aggregated network until there is no modularity gain.

When applying the Louvain method, we only consider positively weighted edges as negatively co-expressed genes are not likely to belong to the same cell. Besides, in order to introduce prior knowledge in the form of landmark segmentation, we modify the method as follows: before running the community detection method, RNA nodes inside the same segmented nucleus or chosen cell landmark are merged together into one node, which we refer to as "cell nodes". When we apply the Louvain method, different cell nodes cannot be merged together. During the first step of local moving of nodes, nodes take the cell label of the community they are assigned to.

Of note, in most cases, nuclear staining is available, and this therefore represents the most frequent use case. However, the landmarks can also originate from other stainings (such as membrane staining).

The input graph strongly connects RNAs that are likely to belong to the same cell. Hence the resulting community partitions of RNAs are supposed to form sets of RNAs belonging to only one cell. In contrast, a cell might contain several communities.

The graph construction and partitioning are implemented in the Python class *ComSegGraph* of our Python package.

Cell segmentation-free in situ clustering

In the previous step, we split our graph into communities of RNAs that are supposed to belong to the same cell. Hence, while a cell can be composed of several RNA communities, we assume that each community does not extend beyond cytoplasmic boundaries. This mimics the mechanism of superpixel segmentation in computer vision[39].

To ease the identification of RNA communities that may belong to the same cell, we first identify communities that have similar RNA profiles.

We associate with each community C of RNA, an expression vector V_C.

$$V_C = \frac{1}{C}\Sigma_{x \in C} V(x) \quad (5)$$

where $V(x)$ are proximity-weighted expression vectors (PE) computed as defined above. Summing the PE of each RNA node in the community helps to capture the local transcriptomic information missing in the global co-expression matrix W. Hence, each community is associated with a community expression vector V_C composed of co-expressed genes at the global scale but also containing the local transcriptomic information of the cell it belongs to.

Then, similarly to what is commonly done in scRNA-seq analysis[40], we cluster our set of community expression vectors V_C using optionally PCA for dimensionality reduction (depending on the number of marker genes) and the modularity-based algorithm Leiden[28]. It defines community clusters $\{L_i\}$ that exhibit similar expression profiles. Each community (and each member x) thus receives a community profile label L_i. However, we do not cluster community expression vectors of less than three RNAs as they might be too small to reliably capture their local transcriptomic neighborhood. We assign to these small communities the majority label of the K nearest neighbors.

This step thus provides a map of RNAs labeled with their community profile label L_i. We refer to this map of labeled RNAs as the transcriptomic domain map. The in situ clustering step is implemented in the class *InSituClustering* of our Python package.

Of note, previous studies have utilized KNN graphs to investigate transcriptomic similarity[41–43]. These studies aimed to identify similarities in gene expression profiles between individual cells or groups of cells (spots). However, our method differs from these approaches in that we seek to uncover similarities between groups of RNA molecules that are not cells but represent subsets of cells.

Final RNA assignment

In the final step, the goal is to associate the RNAs with the cell they belong to. First, we add a centroid node, a node that does not correspond to an RNA molecule. If the landmarks contain nuclear staining, this centroid node is the centroid of the nucleus, all nuclear RNAs are merged into this node, and the centroid node gets the community profile L_i of the nuclear RNAs. In case there is no nuclear RNA, the community profile of the centroid is defined as the majority profile among the K nearest neighbors ($K = 15$, $\mathbf{R} = \mathbf{D}/\mathbf{2}$). In case there is no nuclear staining, the centroid node can be defined based on other landmarks (e.g., the maximum distance function of a cellular landmark). Cell centroid nodes are required to estimate the single-cell RNA profiles.

Once every cell centroid is associated with a community profile L_i, we can finally estimate the single-cell RNA profiles. We associate each cell

centroid of label L_i to their nearest RNAs of the same label. We employ the geodesic distances between cell centroids and RNA nodes. The geodesic distance is the graph's shortest path distance between the centroids and RNA nodes and is computed with the Dijkstra algorithm[44]. We apply the Dijkstra algorithm with Euclidean distance weight on the edges. Besides, when an RNA node's geodesic distance to its nearest centroid is superior to a chosen maximum cell radius R_{max}, the RNA molecule is not associated with any cell. Using the geodesic distance helps to better estimate the cell size of non-convex cells when applying the cell radius R_{max} cut-off. Moreover, geodesic distance accommodates potential lacunar spaces within the tissue when calculating the centroid-RNA distance.

In summary, our method treats RNA positions as nodes in a graph with edges weighted by co-expression. This permits an accurate separation of cells with different expression profiles without the need for explicit cell segmentation. Our method can further use landmarks, such as nuclei, to initiate community detection in spatial RNA graphs. In the absence of clear differences in expression profiles, cells will then be separated based on RNA-centroid distance in the graph.

Our method does not use an explicit cell shape prior other than mean cell diameter D and maximal cell radius R_{max} as the goal is to make this method suitable for tissues harboring arbitrary cell shapes. Furthermore, ComSeg is not based on machine learning, and does therefore not need annotated data or time-consuming learning steps; which simplifies its application and interpretability.

Simulation Python package

In order to validate our method and perform quantitative benchmarks against other approaches, we designed a simulation framework *SimTissue* capable of creating ground-truth data with tunable complexity for tissue morphology, cell type compositions, and marker-gene expression levels.

Our simulation framework can be divided into two steps:

1. Simulation of tissue morphology.
2. Simulation of RNA composition and spatial distribution.

SimTissue is implemented in Python and available at https://github.com/tdefa/SimTissue and documented at https://simtissue.readthedocs.io.

Simulation of tissue morphology

Our framework offers two possible simulation scenarios. In the first scenario, we consider regular geometric patterns. While these are not realistic scenarios, they are well suited to point to potential problems and limitations of the algorithms. They can thus be seen as a purely methodological test scenario. Examples include the checkerboard arrangement or the simulation of clamped L-shapes with random nuclei positions.

In the second scenario, we consider more realistic tissue simulations. For this, *SimTissue* takes as input segmented nuclei from experimental data. These segmentation masks can often be generated easily from DAPI or other nuclei stainings and are widely used for image-based spatial transcriptomic experiments[8,13,33]. Individual cytoplasms are defined by growing cells from segmented nuclei. Each cell grows at a random speed to add irregularity to the cell size. Still, some organs, such as lung tissue, contain lacunar space without cells. Therefore, optional masks can be added to indicate these lacunar spaces where cells cannot grow into.

In experimental data, the tissue section is cut at an arbitrary location and some nuclei are removed from the rest of the cell. It is, hence, possible to simulate cells without nuclei to better mimic experimental fluorescent-based experiments.

Simulation of RNA composition and spatial distribution

Once the nuclei positions and cell shapes are simulated, we have to simulate the RNA composition and distribution within each cell. RNA expression levels can either be set as constant or be directly sampled from experimental measures, e.g. from scRNA-seq data. When sampling profiles from scRNA-seq, we sample for each simulated cell an expression profile of a single cell from scRNA-seq, then multiply the number of RNAs by a factor. Here, we choose a factor of 3 as the fraction of mRNA transcripts captured per cell in scRNA-seq data can reach 30% (depending on reagent chemistry)[45] while the capture rate of smFISH experiment is close to 100%[46]. Finally, RNA molecules are randomly positioned in the available space of the cell with a uniform spatial distribution.

In summary, *SimTissue* allows to simulate experiments of incremental complexity. The full control over the ground truth and the difficulty of the segmentation task aims to understand the limitations of the benchmarked methods.

Simulations in this article

Regular pattern simulations. We simulated 2D square grids (15 µm × 15 µm) and nested L-shape patterns (four squares of 15 µm × 15 µm). The pixel size is 0.150 µm and thus similar to the pixel size obtained with a 60x objective. Nuclei are spheres of 3.75 µm rays. Nuclei are in the center of the cell for square cell shape. For an L-shaped cell, the nucleus is randomly positioned in the center of one of the four squares composing the L-shaped cell.

Lung tissue simulations. We simulated 3D mouse lung tissue leveraging experimental FISH data from ref. 33. The original data were composed of images of 112 µm × 150 µm in XY and 15 µm in Z with a DAPI staining and a Cy3 fluorescent channel. The original pixel size is 0.103 × 0.103 µm and the Z spacing is 0.300 µm. The positions of the nuclei in our simulation are the positions of the segmented nuclei in the original images. The nuclei were segmented with Cellpose[35] on DAPI staining.

Next, we identified the space occupied by cells by thresholding the Cy3 FISH signal (first quintile of the Cy3 distribution). Individual cytoplasms were defined by growing cells with random irregular speed from segmented nuclei inside the allowed space. Random growth speed allows to add irregularity in the cell size. Finally, we removed 20% of nuclei in our simulation to simulate nuclei missed during sample preparation, as explained above.

We used a list of 34 cell-type marker genes with the objective to classify 19 different cell types present in mouse lung tissue. The list of marker genes was selected using the NS-forest algorithm[34] on our external scRNA-seq dataset of mouse lung tissue[33]. Then, we associated each individual cytoplasm with an expression vector sampled from our scRNA-seq dataset so that all the RNA profiles in our simulation are taken from real experimental data. This dataset can be found at https://zenodo.org/records/10172316.

Statistics and reproducibility

We benchmark ComSeg against pciSeq[16], Baysor[15], SCS[19], and Watershed method on both simulation and experimental data. These are among the frequently cited and most widely used methods today if membrane markers are absent. The hyperparameter setting of the benchmarked methods can be found in the Supplementary Note 1. The details of the metrics and datasets that were used are described in the following sections.

Benchmarking on simulations

To assess the RNA profiling quality we compute the mean Jaccard index as follows:

For each cell c, $J_c = \frac{|X_c \cap Y_c|}{|X_c \cup Y_c|}$ where X_c is the ground truth set of RNAs associated with the cell c (ground truth) and Y_c is the set of RNAs predicted as associated with the cell c. The final mean Jaccard index per cell is:

$$J = \frac{1}{\{c\}} \Sigma_c J_c \tag{6}$$

For each cell c, we compute the percentage of wrongly associated RNA WA_c as follows (False Discovery Rate):

$$WA_c = \frac{|Y_c X_c|}{|Y_c|} \tag{7}$$

We compute the percentage of missing RNA MS_c (False Negative Rate) per cell as follows:

$$MS_c = \frac{|X_c Y_c|}{|X_c|} \quad (8)$$

To perform cell type calling from the cell expression vector from Baysor, Watershed, SCS, and Comseg, we first normalize the count matrix from both scRNA-seq and from RNA-cell association using the scTransform normalization[47]. Then we compute the cosine distance between the cell expression vector and the cell type median centroid defined in the reference scRNA-seq data from[33]. Cells are classified into their nearest cell type cluster in terms of cosine distance. For lung tissue simulation, we also employ this cell-type calling method on the ground truth expression vector of each cell to assess the maximum accuracy achievable with the 34 selected markers. Conversely, to previously cited methods, pciSeq performs cell type classification and RNA-nuclei association simultaneously. For this reason, the cell type classification described here was not applied to pciSeq.

Experimental evaluation

We applied the benchmarked methods on two datasets of lung tissue. The first one exhibits 6 different marker genes in 3D mouse lung tissue acquired with a home-built sequential smFISH system. This mouse lung tissue was irradiated with 17 gy 5 months prior to mouse sacrifice as described in ref. 33. The second dataset was acquired with HybISS[26], consisting of a human embryonic lung and 147 genes in 2D[25]. In both cases, we perform a nucleus segmentation with Cellpose[35] and use this segmentation as initialization in our benchmark.

Reference scRNA-seq cluster

As we do not have ground truth for all experimental data, we leverage scRNA-seq to check the consistency of the single-cell spatial RNA profiles obtained from images. We argue that it should be possible to match the spatial profiles obtained from images to the scRNA-seq data, and that the percentage of cells that can be matched is thus a quality metric for the segmentation method.

Reference clusters from scRNA-seq for mouse lung tissue

We re-cluster our single-cell data[33] using only the six mapped genes and cells from the same condition (i.e., 5 months after irradiation with 17 gy). The final clustering is composed of five different clusters.

Reference clusters from scRNA-seq for embryonic lung tissue

In the original study[25], the authors applied pciSeq to perform single-cell spatial RNA profiling. Their final expression matrix provided in the study displays 89 genes over the 147 genes map in the HyBISS data. We also keep the same subset of 89 genes for evaluation. We use the same reference scRNA-seq dataset and clustering annotation as provided by the authors.

Matching in situ single-cell RNA profile and scRNA-seq

As for simulation, we normalize both scRNA-seq data and the count expression matrix from spatially resolved RNA profiling data with scTransform[47]. Many methods exist to match single-cell spatial transcriptomic data and scRNA-seq[48]. We chose to leverage the cosine distance as it is robust with respect to different capture rates among the two modalities.

Each cell expression vector was matched to the closest median centroid cluster from scRNA-seq. The cosine distance acts as a proxy for the matching quality.

Automatic generation of ground truth for high-quality staining area

We segment the nuclei and the cell using Cellpose 2.0[49]. We manually fine-tuned the Cellpose models with the provided human-in-the-loop GUI except for the nuclei human breast cancer dataset where we kept the default cellpose nuclei model. Specifically, in the human breast cancer dataset, we segmented cells based on the third cellbound marker across three different available. For the cells in the mouse ileum dataset, we kept the fine-tuned Cellpose segmentation proposed in the original publication. After segmentation, we identified cells with exactly one nucleus. However, in dense tissue areas, cells might mistakenly include a few pixels from another nearby nucleus due to segmentation inaccuracies. In order to be robust with respect to this kind of variations, we assigned a nucleus to a cytoplasmic region if there was an overlap of at least 20-pixel width between them. The final list of cells containing exactly one nucleus is then employed to generate the ground truth, as these cells are likely to correspond to high-quality staining areas.

To compute the Jaccard index on the generated ground truth mask, we associate each cell from the ground truth with the predicted cell with the most molecules in common. We then apply the formula described in (6).

Ethics statement

For the generation of the in-house mouse lung tissue dataset displayed in Fig. 3: 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.

Data availability

The simulated dataset of lung tissue can be downloaded from Zenodo (https://zenodo.org/records/10172316). The generated ground truth of the two MerFISH datasets and the corresponding segmentation results from our benchmark are available at https://zenodo.org/records/11237477. Our in-house mouse dataset is available at https://zenodo.org/records/11068509.

Code availability

ComSeg is available at https://github.com/fish-quant/ComSeg. The implementation code of SimTissue is available at https://github.com/tdefa/SimTissue. Scripts to reproduce the benchmark are available also on github https://github.com/tdefa/script_Benchmark and also available at this Zenodo repository[50]: https://doi.org/10.5281/zenodo.11505979.

Received: 6 December 2023; Accepted: 20 June 2024;
Published online: 06 July 2024

References

1. Larsson, L., Frisén, J. & Lundeberg, J. Spatially resolved transcriptomics adds a new dimension to genomics. *Nat. Methods* **18**, 15–18 (2021).
2. Xiaowei, Z. Spatially resolved single-cell genomics and transcriptomics by imaging. *Nat. Methods* **18**, 15–18 (2021).
3. Seferbekova, Z., Lomakin, A., Yates, L. R. & Gerstung, M. Spatial biology of cancer evolution. *Nat. Rev. Genet.* **24**, 295–313 (2023).
4. Palla, G., Fischer, D. S., Regev, A. & Theis, F. J. Spatial components of molecular tissue biology. *Nat. Biotechnol.* **40**, 308–318 (2022).
5. Lomakin, A. et al. Spatial genomics maps the structure, nature and evolution of cancer clones. *Nature* **611**, 594–602 (2022).
6. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).
7. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
8. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).

9. Eng, C. H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH + . *Nature* **568**, 235–239 (2019).
10. Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
11. Codeluppi, S. et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* **15**, 932–935 (2018).
12. Moffitt, J. R. et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).
13. Salas, S. M. et al. Optimizing Xenium In Situ data utility by quality assessment and best practice analysis workflows. Preprint at *bioRxiv* https://doi.org/10.1101/2023.02.13.528102 (2023).
14. Wei, X. et al. Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science* **377**, eabp9444 (2022).
15. Petukhov, V. et al. Cell segmentation in imaging-based spatial transcriptomics. *Nat. Biotechnol.* **40**, 345–354 (2022).
16. Qian, X. et al. Probabilistic cell typing enables fine mapping of closely related cell types in situ. *Nat. Methods* **17**, 101–106 (2020).
17. Littman, R. et al. Joint cell segmentation and cell type annotation for spatial transcriptomics. *Mol. Syst. Biol.* **17**, e10108 (2021).
18. Fu, X. et al. BIDCell: Biologically-informed self-supervised learning for segmentation of subcellular spatial transcriptomics data. *Nat. Commun.* **15**, 509 (2024).
19. Chen, H., Li, D. & Bar-Joseph, Z. SCS: cell segmentation for high-resolution spatial transcriptomics. *Nat. Methods* **20**, 1237–1243 (2023).
20. Mateo, L. J., Sinnott-Armstrong, N. & Boettiger, A. N. Tracing DNA paths and RNA profiles in cultured cells and tissues with ORCA. *Nat. Protoc.* **16**, 1647–1713 (2021).
21. Moffitt, J. R. & Zhuang, X. RNA imaging with multiplexed error-robust fluorescence in situ hybridization (MERFISH). *Methods Enzymol.* **572**, 1–49 (2016).
22. Virshup, I. et al. The scverse project provides a computational ecosystem for single-cell omics data analysis. *Nat. Biotechnol.* **41**, 604–606 (2023).
23. Blampey, Q. et al. Sopa: a technology-invariant pipeline for analyses of image-based spatial omics. *Nat. Commun.* **15**, 4981 (2024).
24. Greenwald, N. F. et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* **40**, 555–565 (2022).
25. Sountoulidis, A. et al. A topographic atlas defines developmental origins of cell heterogeneity in the human embryonic lung. *Nat. Cell Biol.* **25**, 351–365 (2023).
26. Gyllborg, D. et al. Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. *Nucleic Acids Res.* **48**, E112 (2020).
27. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
28. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
29. Beucher, S. & Lantuéjoul, C. Use of watersheds in contour detection.http://cmm.ensmp.fr/~beucher/publi/watershed.pdf (1979).
30. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016).
31. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
32. Park, J. et al. Cell segmentation-free inference of cell types from in situ transcriptomics data. *Nat. Commun.* **12**, 3545 (2021).
33. Curras-Alonso, S. et al. An interactive murine single-cell atlas of the lung responses to radiation injury. *Nat. Commun.* **14**, 2445 (2023).
34. Aevermann, B. et al. A machine learning method for the discovery of minimum marker gene combinations for cell type identification from single-cell RNA sequencing. *Genome Res.* **31**, 1767–1780 (2021).
35. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
36. Teng, H., Yuan, Y. & Bar-Joseph, Z. Clustering spatial transcriptomics data. *Bioinformatics* **38**, 997–1004 (2022).
37. Avesani, S. et al. Stardust: improving spatial transcriptomics data analysis through space-aware modularity optimization-based clustering. *Gigascience* **11**, giac075 (2022).
38. Newman, M. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E.* **69**, 026113 (2003).
39. Achanta, R. et al. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 2274–2281 (2012).
40. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
41. Li, J., Chen, S., Pan, X., Yuan, Y. & Shen, H.-B. Cell clustering for spatial transcriptomics data with graph neural networks. *Nat. Comput. Sci.* **2**, 399–408 (2022).
42. Long, Y. et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat. Commun.* **14**, 1155 (2023).
43. Hu, J. et al. SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat. Methods* **18**, 1342–1351 (2021).
44. Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Math.* **1**, 269–271 (1959).
45. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
46. Williams, C. G., Lee, H. J., Asatsuma, T., Vento-Tormo, R. & Haque, A. An introduction to spatial transcriptomics for biomedical research. *Genome Med.* **14**, 68 (2022).
47. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
48. Miller, B. F., Huang, F., Atta, L., Sahoo, A. & Fan, J. Reference-free cell type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data. *Nat. Commun.* **13**, 2339 (2022).
49. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 (2022).
50. Thomas, D. Scripts to reproduce the benchmark of ComSeg. *Zenodo* https://doi.org/10.5281/zenodo.11505979 (2024).

Acknowledgements

This work has received financial support through the Agence Nationale de la Recherche (ANR) for grants (LUSTRA, reference ANR-19-CE14-0015-04 to J.-A.L.-V., C.F., F.Mu., and T.W.) and (TRANSFACT, reference ANR-19-CE12-0007-02, to F.Mu. and T.W.). F.Mu. and C.W. acknowledge funding by Institut Pasteur. 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). S.C.-A. and J.-A.L.-V. also received support from La Ligue Contre Le cancer. S.C.-A. and M.A. were recipients of PhD fellowships from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 666003 and No 847718 respectively. H.L. is the recipient of a PhD fellowship from the International Student program from Paris-Saclay University. J.S. is the recipient of a PhD fellowship from the French Ministry of Education, research and Industry.

Author contributions

T.D., F.Ma, F.Mu, and T.W. designed and developed the method ComSeg and carried out the computational analysis. H.L., M.A., J.S., S.C.-A., C.W.,

J.-A.L.-V., and C.F. performed the experimental studies for the in-house dataset. T.D., F.Ma, F.Mu, and T.W. participated in the writing of the paper. F.Mu and T.W. supervised the work. All the co-authors contributed to the 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/s42003-024-06480-3.

Correspondence and requests for materials should be addressed to Florian Mueller or Thomas Walter.

Peer review information *Communications Biology* thanks Viktor Petukhov, Xun Xu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Ruby Yun-Ju Huang and Manuel Breuer. A peer review file is available.

Supplementary Note 1: Hyper-parameter setting of benchmarked methods

ComSeg

We apply ComSeg with mean cell diameter $D = 10\mu m$ and $R_{max} = 15\mu m$ in checkerboard simulation (Sim 1-2-3-4) and increase to $R_{max} = 40\mu m$ in Sim 5 while keeping the same value for D.

We set $D = 15\mu m$ and $R_{max} = 30\mu m$ in lung tissue simulation. We use $D = 20\mu m$ and $R_{max} = 40\mu m$ on mouse lung tissue and $D = 40\, pixels$ and $R_{max} = 50\, pixels$ on embryonic lung tissue. For the Mouse ilieum dataset we use $D = 10\mu m$ and $R_{max} = 8\mu m$. For the Breast cancer dataset we use $D = 5\mu m$ and $R_{max} = 12\mu m$.

Baysor:

We apply Baysor by incorporating nucleus segmentation mask priors and let Baysor estimate the scale parameter automatically. All other parameters were left as default. In addition, for the mouse ileum dataset, we kept the original parameters set by the authors including a compartment specific gene list as parameter, and we added the nuclei segmentation mask priors.

Baysor does not perform RNA-nucleus assignment but groups RNAs into cells without referring to the given nucleus segmentation index of the prior segmentation mask. On simulated datasets, in order to compare Baysor to other methods for RNA-nucleus assignment, we associated each predicted cell index by Baysor with a nucleus segmentation index from the provided segmentation mask ground truth. To that end, each predicted cell index by Baysor was associated with the nucleus index from the ground truth with the most molecules in common.

pciSeq

We apply pciSeq on geometric simulation using simulated scRNA-seq data containing the simulated RNA profiles. In lung tissue simulations, we use the same scRNA-seq dataset as the one sampled to simulate RNA profiles. All other parameters were left as default. We use the version 0.0.46 of the pciSeq Python PyPI package. We use maximum projection to apply pciSeq to 3D data as pciSeq is only designed for 2D.

For the MERFISH human breast cancer and MERFISH mouse ileum datasets, inspired by the original publication of the MERFISH mouse ileum dataset[1], we use the RNA in the nuclei as reference scRNA-seq data.

Watershed

We apply Watershed by taking as input the inverse distance map from segmented nuclei. In lung simulation, as a mask, we use the inferred cytoplasm from the Cy3 signal. Otherwise, we use the Watershed on the inverse distance map with a maximum distance from the nucleus of 2 µm on mouse lung tissue, of 16 pixels for embryonic lung tissue and of 8 µm on the MERFISH human breast cancer and MERFISH mouse Ileum datasets. The mask helps to prevent the misassignment of RNA to nuclei too far apart and was adapted to the tissues density and complexity

SCS

We use the SCS code released by the author (https://github.com/chenhcs/SCS/) as well as their training parameter display in their example (100 epoch and a learning rate of 0.001). To adapt SCS to image based spatial transcriptome data we use a binning of 15 pixels. Like Baysor, SCS does not perform RNA-nucleus assignment but groups RNAs into cells. Hence we use the same method as for Baysor to assess performance. We use a crop of 3000x3000 for the mouse ileum dataset and crop of 9000x9000 for the human lung and merfish breast cancer dataset. For all datasets we use Cellpose[2] to segment the nuclei instead of the originally proposed watershed.

Supplementary note 2 : Validation on simple simulations of regular patterns

In this section we aim to better understand the strengths and limitations of each benchmarked method. To this end, for validation and benchmarking, we generated five types of simulated datasets of gradually increasing complexity and specific cases.

To study the effect of the different possible expression profiles without any cell shape complexity, we simulate a checkerboard. This is a similar cell size scale to what we can observe in mouse tissue[3,4]. To test the effect of non-convex shapes, we also simulate L-shaped cells (see Methods). Each simulated set contains 10 images of 110 cells except the last one containing 144 cells per image.

1) **Simulation 1 (Sim 1). Variation of expression level:** The objective of this simulation is to benchmark the methods when markers have the same expression level versus the case where one marker is sparsely expressed. We started by simulating only two cell types expressing one marker each, A and B. We fixed the number of transcripts for the first cell type to A=100, while we set the number of transcripts for the second to B=100

(Sim1a) or B=10 (Sim1b) in two runs (Supplementary Figure 1a, left panels). For this geometry, the Watershed performs a perfect assignment, as the nuclei are centered and the cell shapes are convex. Besides, in the simple case where the two markers are equally expressed, Baysor and ComSeg have also a Jaccard index close to 1 whereas pciSeq has a Jaccard index below 0.8. In fact, pciSeq only assigns RNAs in the close neighborhood of the nucleus and with a spherical shape prior. These missed RNAs are penalized by the Jaccard index.

When the expression of the second cell type becomes sparse (B=10), the performance of all methods leveraging RNA spatial distribution, Baysor, pciSeq and ComSeg drops. Still, ComSeg performs better in terms of Jaccard index (over 0.8, Supplementary Figure 1b, center left panel) and cell type accuracy (of 0.99, Supplementary Figure 2) than Baysor and pciSeq. This result can be attributed to the utilization of a cell shape prior by Baysor and pciSeq in RNA assignment. When expression becomes sparse, Baysor and pciSeq may not find RNA point clouds matching their shape prior. Finally, the shape agnostic strategy of ComSeg appears to be more adapted for sparse input.

2) **Simulation 2 (Sim 2). Shared marker genes:** In the subsequent simulation, we aimed to investigate the impact of shared marker genes across diverse cell types. To achieve this, we categorized three distinct cell types: cell type A, with 100 RNAs from gene A; cell type B, with 100 RNAs from gene B; and cell type C with 100 RNAs from gene A and gene B (Supplementary Figure 1a, center right panel). In this case pciSeq and Baysor have slightly better performance in terms of Jaccard index than ComSeg (Supplementary Figure 1b, center right panel). Still, all models get an almost perfect cell type calling (Supplementary Figure 2). Without surprise, Watershed obtains perfect RNA-nuclei assignment due to the simplicity of the tissue geometry.
3) **Simulation 3 (Sim 3). Experimental expression profile:** In this simulation, we sample RNA profiles from experimental data. We simulate 34 marker genes as described above for lung tissue simulation (Supplementary Figure 1a, right panel). On real RNA profiles, ComSeg has a better Jaccard index than Baysor and pciSeq. As the cell shapes are still convex, Watershed gets a Jaccard index of 1 (Supplementary Figure 1b, right panel).
4) **Simulation 4 (Sim 4). Experimental expression profile with missing nuclei:**
In this simulation, a scenario akin to the previous one was replicated, where some nuclei were intentionally omitted to simulate conditions akin to experimental data (Supplementary Figure 1c, left panel). Specifically, 20% of the cells were simulated without nuclei, mirroring situations encountered in tissue experiments. Remarkably, under these conditions, ComSeg continued to exhibit a superior Jaccard index compared to both Baysor and pciSeq (Supplementary Figure 1e, left panel). As expected, the Watershed method cannot cope with missing nuclei, as the method uses nuclei as seeds. As a consequence, accuracy in cell type identification drops as compared to other models (Supplementary Figure 1f, left panel). A visual representation of the benchmarked methods RNA assignments can be found in Supplementary Figure 1c.
5) **Simulation 5 (Sim 5). Experimental expression profile with missing nuclei and L shape:**

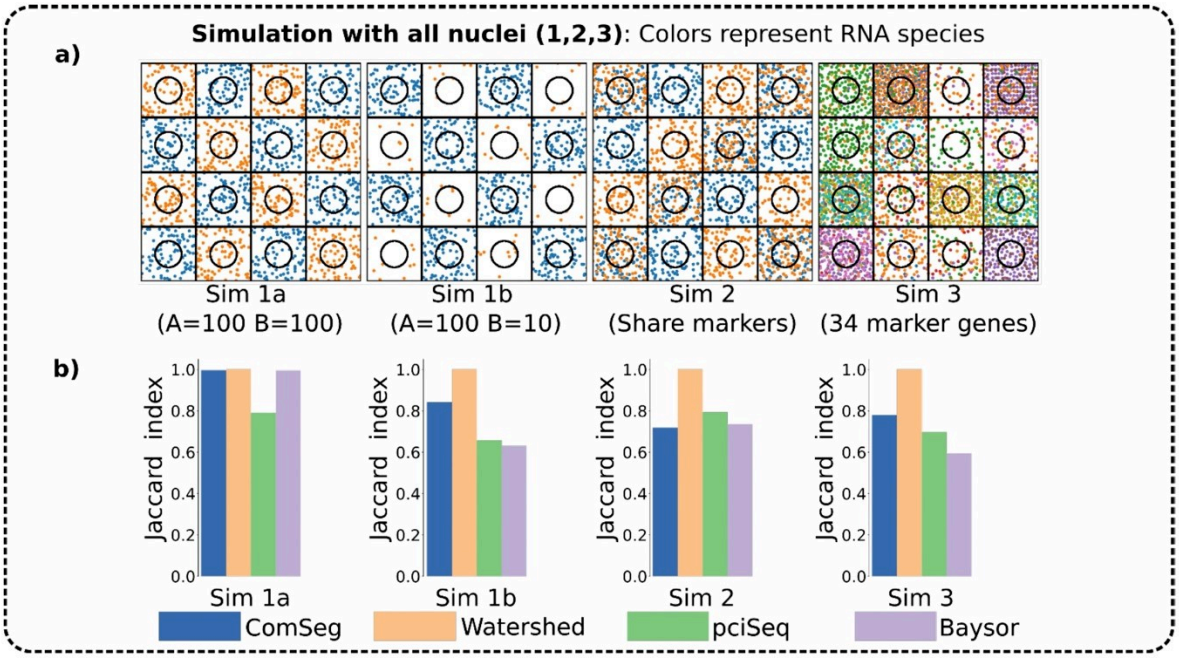
Lastly we also add cells with an L-shape to test non-convex examples (Supplementary Figure 1d, left panel). In this case, ComSeg outperforms other models for the Jaccard index confirming that it is designed to deal with irregular non-convex cell shape. Conversely, Watershed has a very low Jaccard index because it cannot cope with non-convex shapes by construction (Supplementary Figure 1e). Also, Baysor and pciSeq are underperforming in accurately identifying the L-shaped cells owing to their inherently convex cell shape assumptions. As a consequence, ComSeg also exhibits superior cell type calling performances in this more complex case as compared to all other methods (Supplementary Figure 1f). A visual representation of the RNA assignments generated by all models can be found in Supplementary Figure 1d.

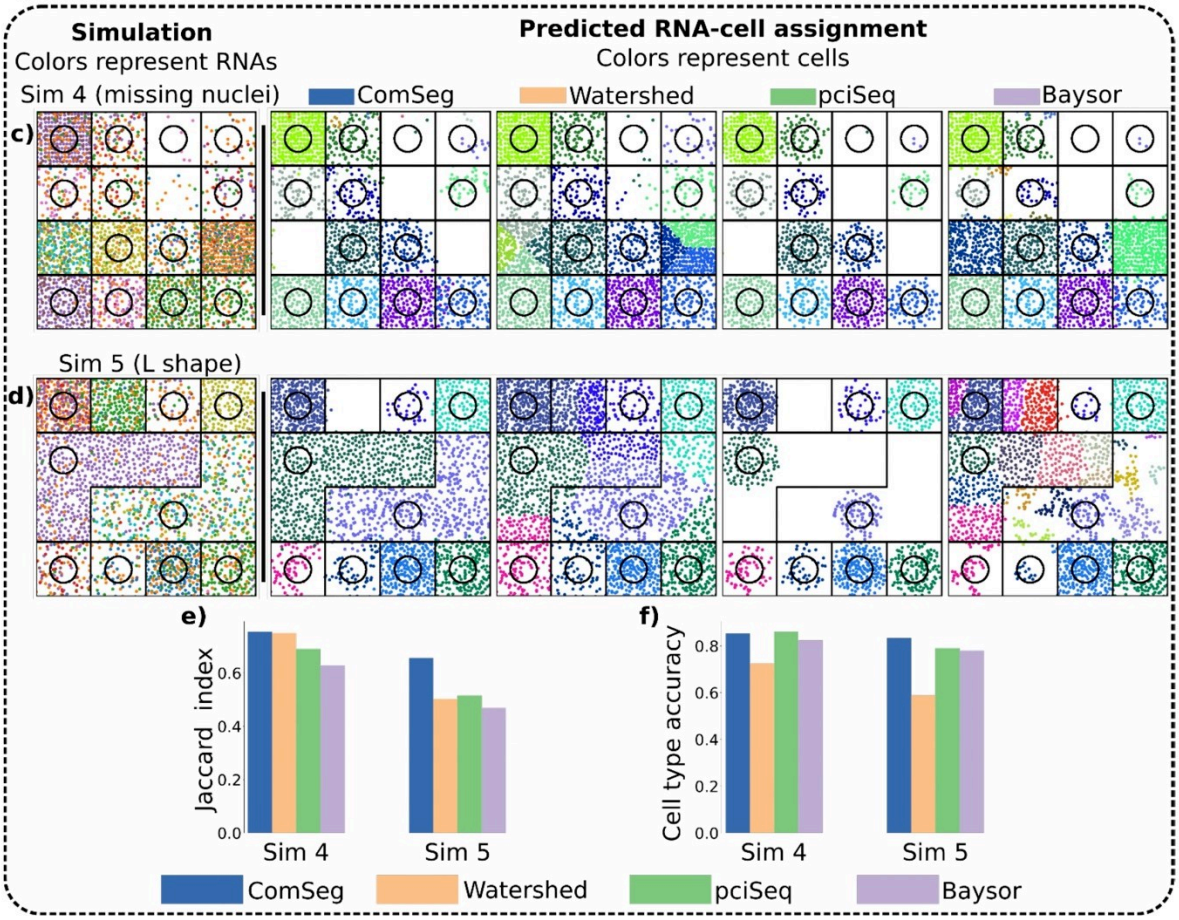
In summary, all methodologies encounter difficulties as cell shapes and RNA profiles become increasingly complex and as marker expression becomes sparse. Notably, Watershed proves to be an optimal choice for cells with convex shapes. PciSeq and Baysor, while capable of estimating valid single-cell spatial RNA profiles in terms of cell type calling, exhibit limitations in capturing a substantial portion of transcripts. Moreover, the disparity in RNA-cell assignment performance between ComSeg and the other benchmarked methods widens notably as cell shapes and expression patterns grow in complexity.

Supplementary Note 3 : Influence of initialization seed on ComSeg

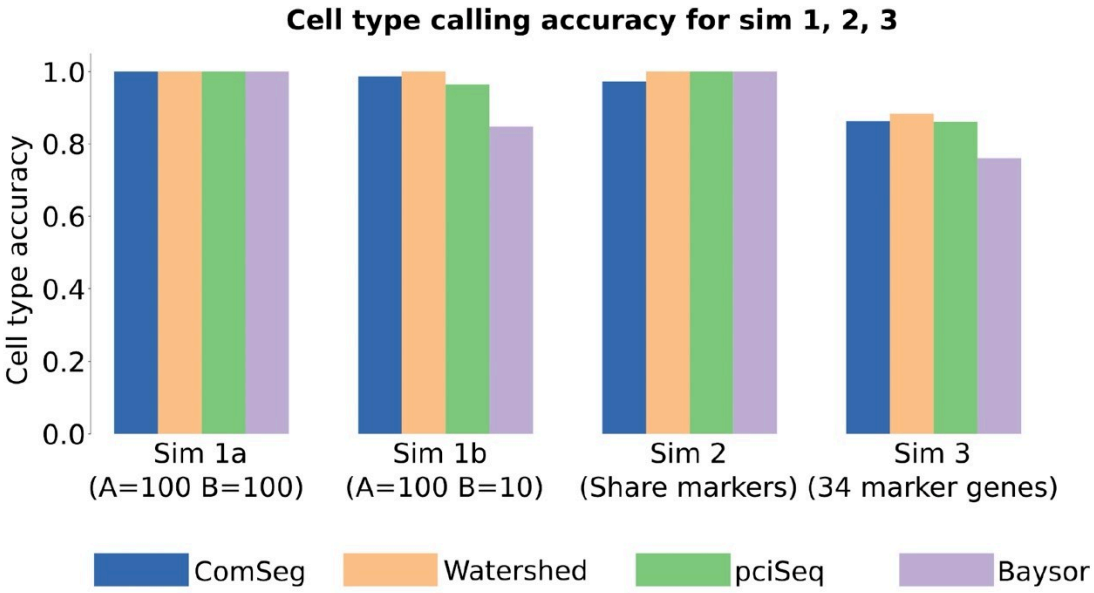
ComSeg employs random initialization within the Louvain method when computing the communities and the Leiden method when clustering the $\{V_C\}$. To evaluate the impact of the random initialization, we executed the algorithm 50 times on the lung simulation and compared the mean Jaccard index per cell for each run with a different random initialization (Supplementary Figure 6). The standard deviation of the mean Jaccard index per cell is less than 0.002, demonstrating the negligible influence of the initialization seed.
Similarly, we compute the Rand index between the RNA assignments over 10 runs on Supplementary Figure 7. This analysis shows that the Rand index between pairs of runs is stable with a value of approx 0.96, indicating that 96% of the RNA are grouped into a similar way across different initializations.

Supplementary figures

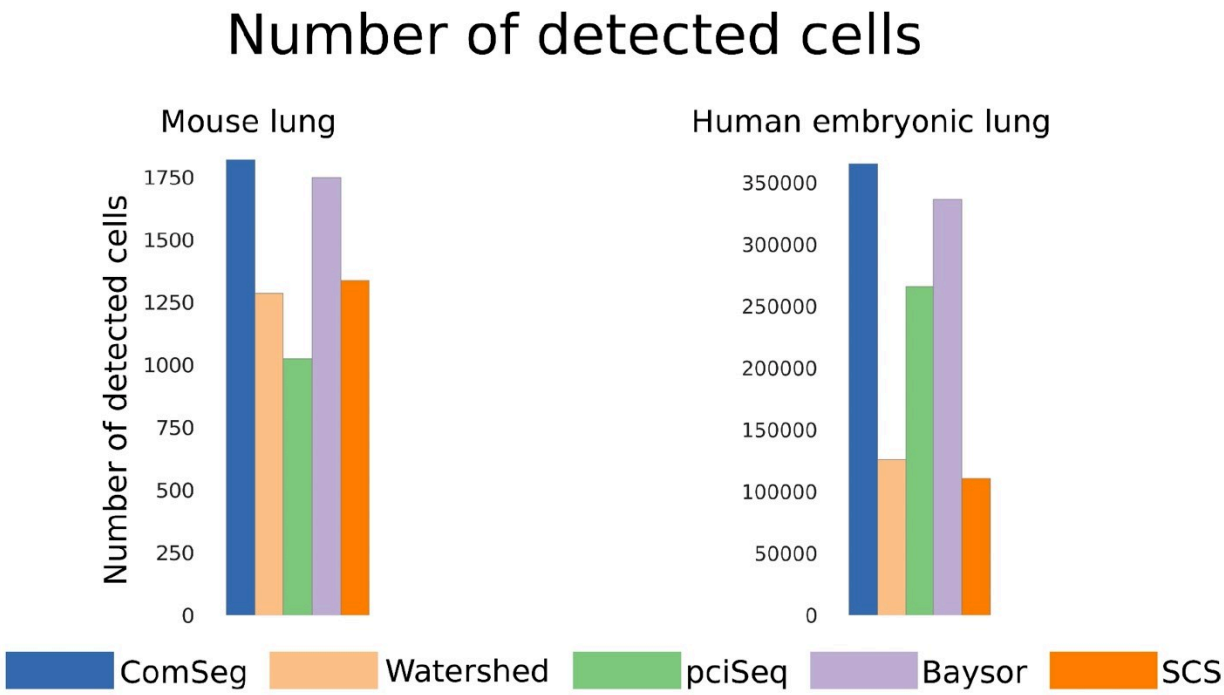

a)
Simulation with all nuclei (1,2,3): Colors represent RNA species
Sim 1a
(A=100 B=100)
Sim 1b
(A=100 B=10)
Sim 2
(Share markers)
Sim 3
(34 marker genes)
b)
Jaccard index
0.0
0.2
0.4
0.6
0.8
1.0
Sim 1a
Sim 1b
Sim 2
Sim 3
ComSeg
Watershed
pciSeq
Baysor


Simulation
Colors represent RNAs
Predicted RNA-cell assignment
Colors represent cells
Sim 4 (missing nuclei)
ComSeg
Watershed
pciSeq
Baysor
c)
Sim 5 (L shape)
d)
e)
Jaccard index
0.0
0.2
0.4
0.6
Sim 4
Sim 5
f)
Cell type accuracy
0.0
0.2
0.4
0.6
0.8
Sim 4
Sim 5
ComSeg
Watershed
pciSeq
Baysor

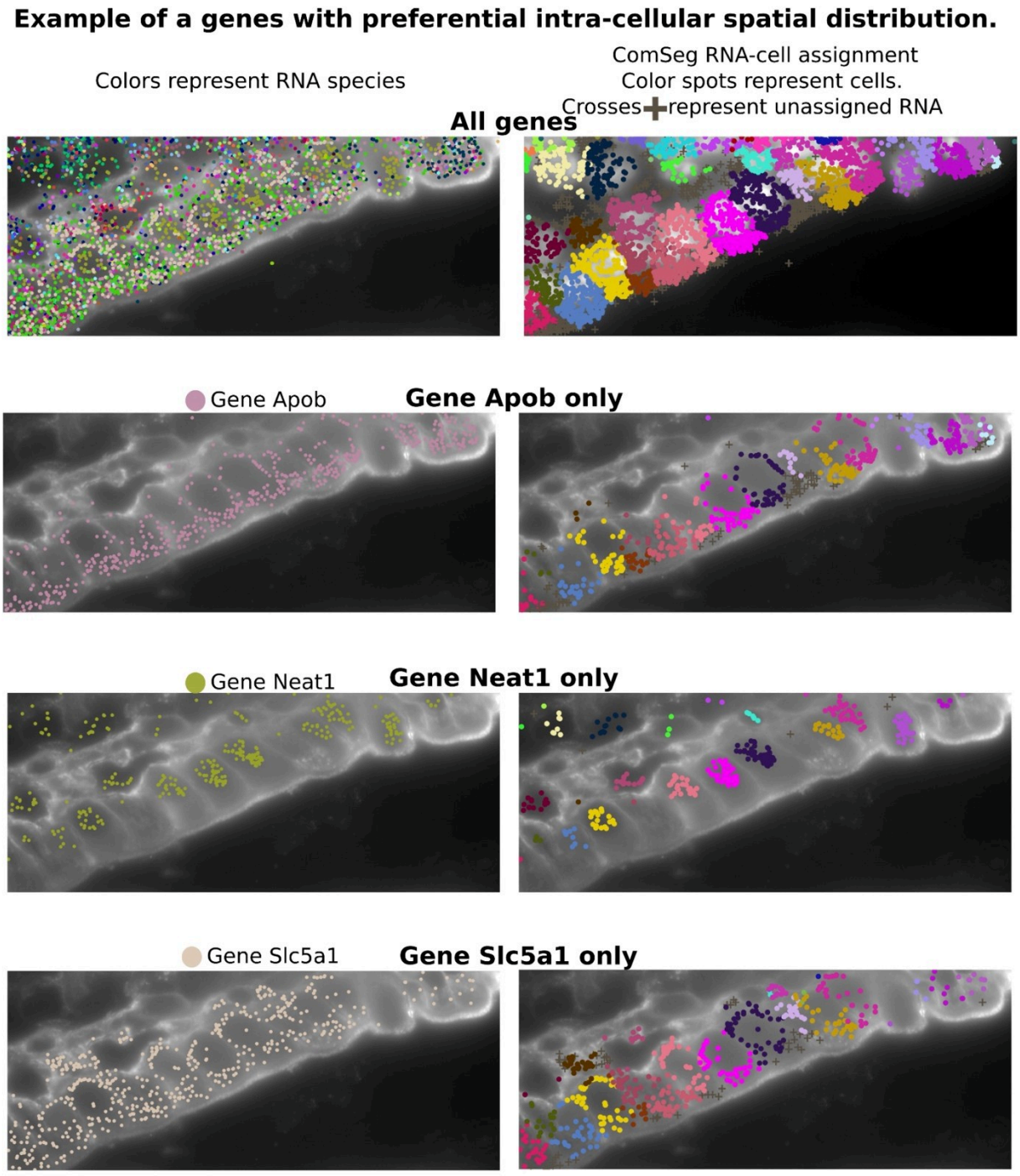
Supplementary Figure 1 : Simulations and results on regular patterns. **a** Examples of simulation where all cells have a nucleus. From left to right: Sim 1 with variation of expression level, Sim 2 with share marker genes, Sim 3 with experimental RNA profiles with 34 markers. **b** Mean Jaccard index per cell for Sim 1-2-3. **c** Example of Sim 4 with experimental RNA profiles with 34 markers and the corresponding RNA-cell assignment for the benchmarked methods. **d** Example of Sim 5 akin to Sim 4 but with L-shaped cells and the corresponding RNA-cell assignment for the benchmarked methods. **f** Mean Jaccard index per cell for sim 4 and 5. **e** Cell type calling accuracy for Sim 4 and Sim 5.



Supplementary Figure 2: Cell type calling accuracy on checkerboard cell for the Sim 1 with different levels of expressions, Sim 2 with share markers and sim 3 with 34 marker genes with RNA profiles sample from scRNA-seq.



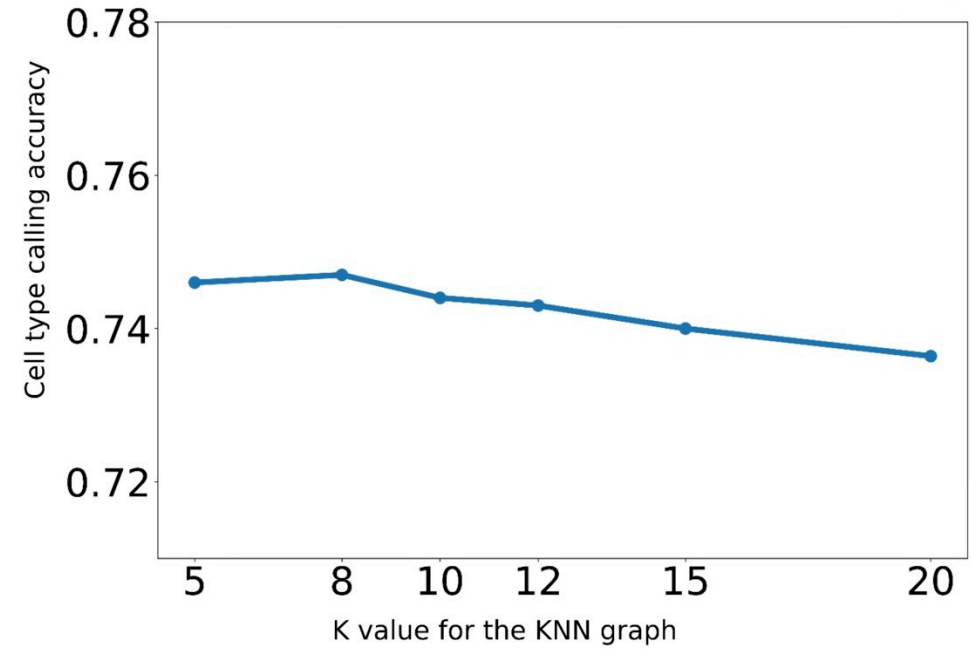
Supplementary Figure 3 : Number of cells associated with an RNA profile (more than 5 RNAs) in the mouse lung tissue dataset (right) and in the Human embryonic lung tissue dataset (Left) for the tested method.



Supplementary Figure 4 : Example of genes with non uniform intracellular spatial distribution in the mouse ileum dataset. All the rows display the membrane staining in gray. The first row of images shows plots with all the genes. The second, third and fourth rows show only results with the gene Apob, Neat1 and Slc5a1 respectively displaying preferential intra-cellular spatial distribution.

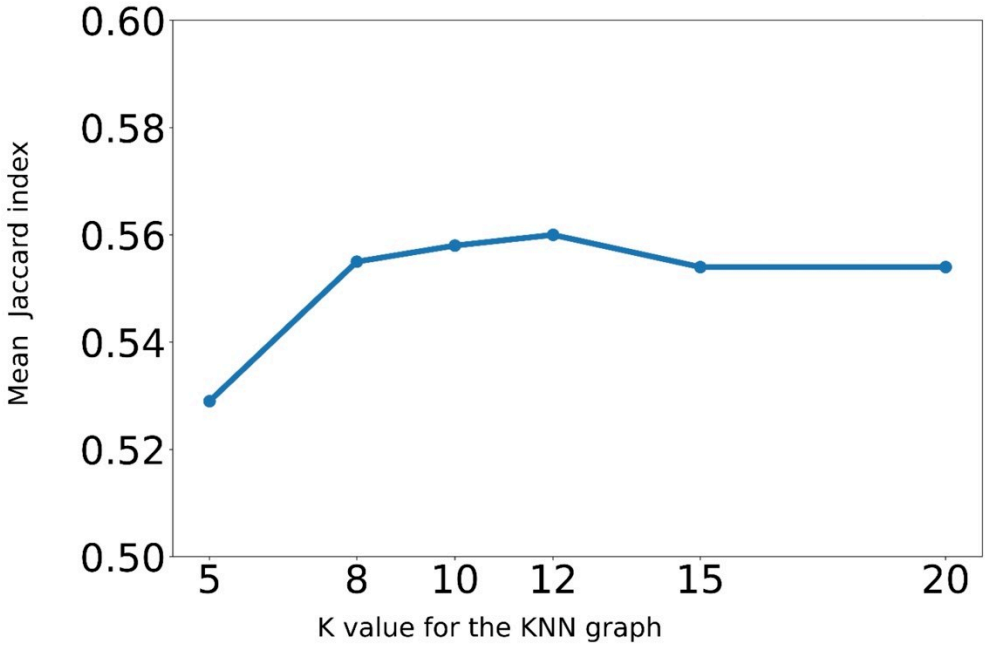
Accuracy for different K values for the KNN graph on lung simulation

a)

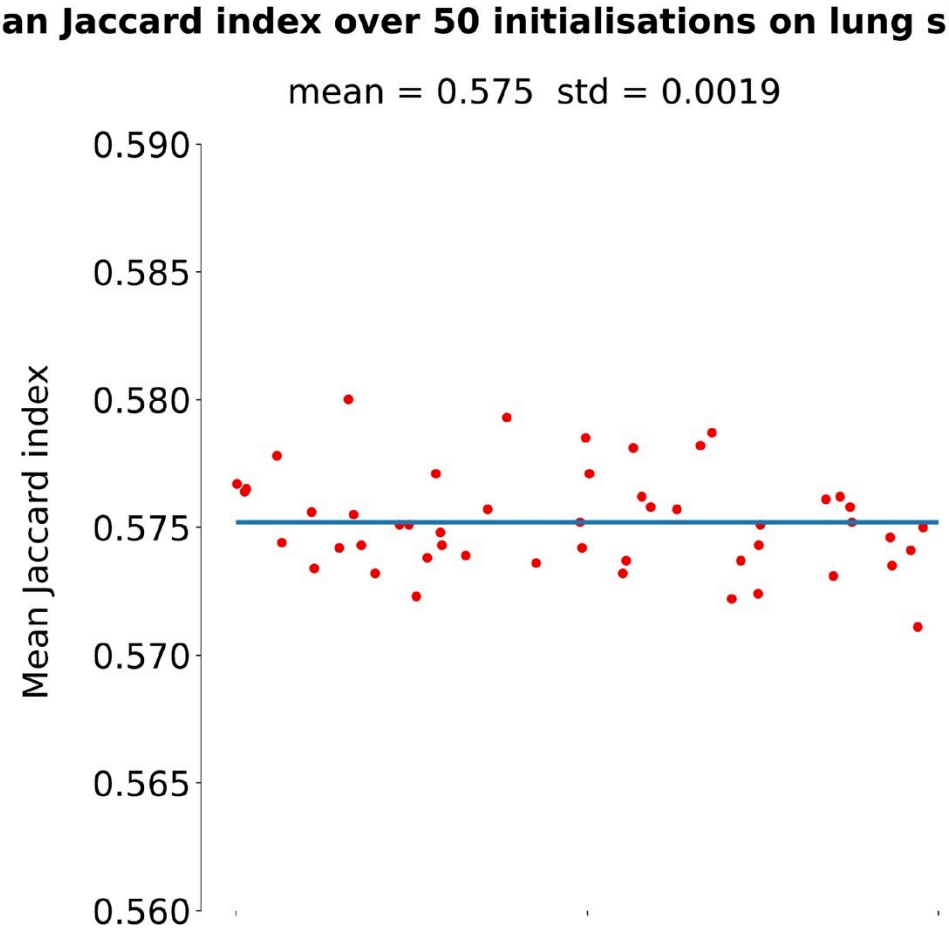


Jaccard index for different K values for the KNN graph on lung simulation

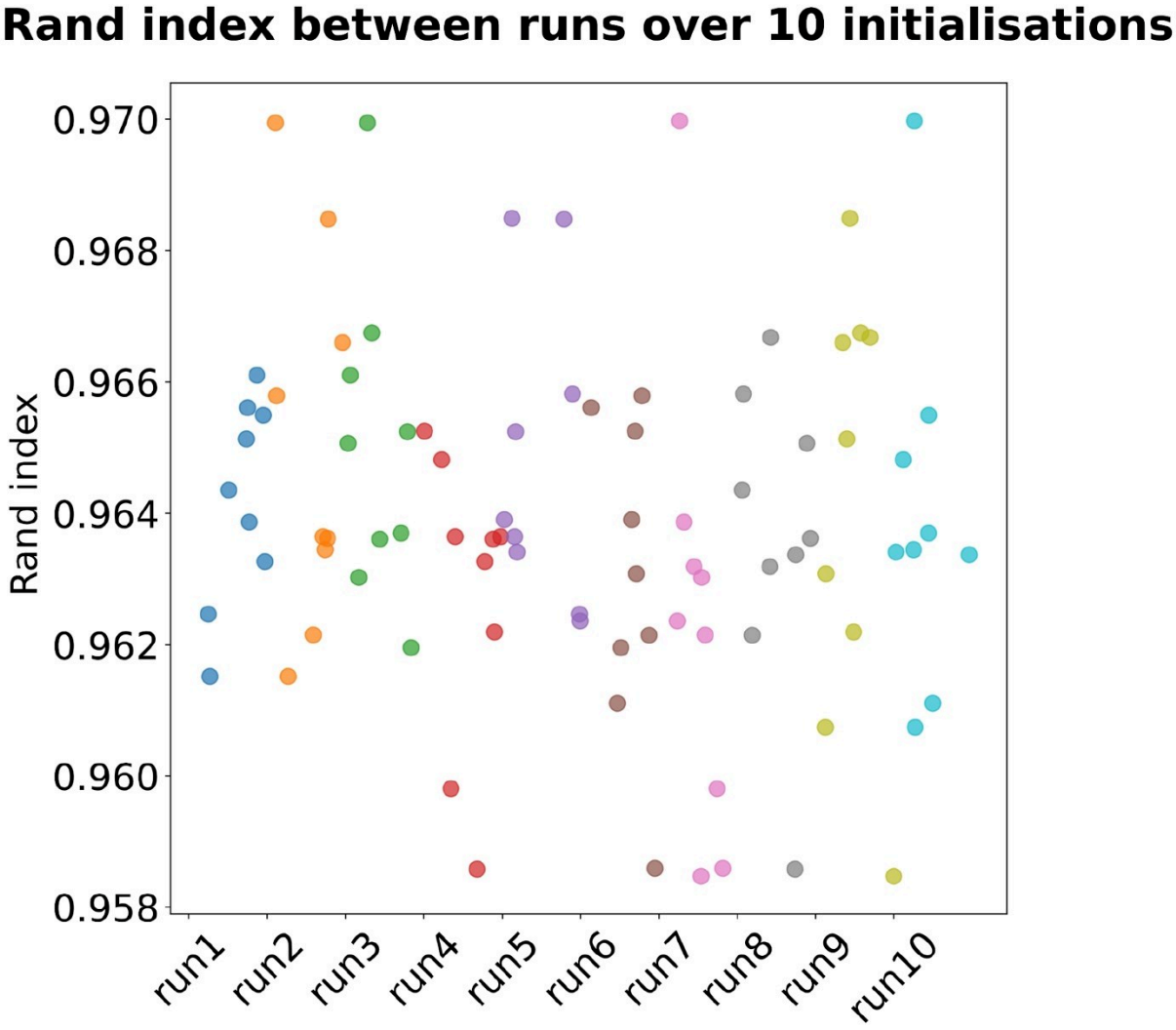
b)



Supplementary Figure 5: Performance of ComSeg with different K values for generating the KNN graph on lung simulation. We measure the cell type calling accuracy (**a**) and the Jaccard index (**b**).



Supplementary Figure 6: The mean Jaccard index per cell obtained for 50 different random initializations of ComSeg on our simulated dataset of lung tissue. The blue line is the mean of all the Jaccard index obtained for 50 different random initializations.



Supplementary Figure 7: Each spot corresponds to the rand index between two different initializations of ComSeg on the simulated dataset. The color of the spots correspond to the run taken as reference. The rand index can be interpreted here as the proportion of RNA in the same cell for the pair compared runs.

Supplementary Tables

Supplementary table 1 : Runtime and memory usage analysis for different methods on the mouse ileum dataset. Experiments were done on a laptop with a 16 cores CPU Intel(R) Core(TM) i9-10885H CPU @ 2.40GHz and with a GPU Quadro T2000 with 4Go of memory. The memory requirement was measured with the *memory_profiler* pip package.

	ComSeg	Watershed	pciSeq	Baysor	SCS*
Memory requirement	1,34Go	0,52Go	2,83Go	8,38Go	6.32go
Time	29 minutes	1 minutes	3 minutes	29 minutes	51 minutes
GPU	No	No	No	No	Yes

*for SCS, we divided SCS into crops of 3000x3000 pixels so the training fits our laptop memory and GPU.

References

1. Petukhov, V. *et al.* Cell segmentation in imaging-based spatial transcriptomics. *Nat. Biotechnol.* 40, 345–354 (2022).

2. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).

3. Turgay, Y. *et al.* The molecular architecture of lamins in somatic cells. *Nature* 543, 261–264 (2017).

4. Curras-Alonso, S. *et al.* An interactive murine single-cell atlas of the lung responses to radiation injury. *Nat. Commun.* **14**, 2445 (2023).

INDEX OF FIGURES

INDEX OF TABLES

RÉSUMÉ POUR LE GRAND PUBLIC

La radiothérapie est l'un des traitements principaux du cancer du poumon. Cependant, les tissus entourant la tumeur reçoivent également une dose d'irradiation. Certains patients ont en réaction une pneumopathie qui peut se compliquer pour les cas les plus graves en fibrose pulmonaire. Cette maladie présente une évolution progressive, irréversible et mortelle en raison des tissus cicatriciels se développant dans les poumons et empêchant la respiration. Le but de mon projet est d'étudier les mécanismes encore mal connus de cette maladie en utilisant une technologie permettant d'analyser la réponse de chaque cellule à la radiothérapie. En utilisant des échantillons provenant de patients ayant subi une radiothérapie et également de modèles de souris, nous avons permis l'avancée des connaissances des processus cellulaires et moléculaires impliqués dans la fibrose pulmonaire radio-induite. L'optique étant de trouver des traitements afin d'améliorer la qualité de vie des patients.

ABSTRACT FOR THE GENERAL PUBLIC

Radiotherapy is one of the main treatments for lung cancer. However, the tissues surrounding the tumor also receive a dose of radiation. Some patients in reaction have a pneumonia, which for the most serious cases can complicate to pulmonary fibrosis. This disease has a progressive, irreversible and fatal evolution due to the scar tissue developing in the lungs and preventing breathing. The aim of my project is to study the still poorly understood mechanisms of this disease by using a technology that allow the analysis of the response of each individual cell to radiotherapy. By using samples from patients who underwent radiotherapy and also from mouse models, we have advanced knowledge of the cellular and molecular processes involved in radiation-induced pulmonary fibrosis. The aim is to find treatments to improve the quality of life of patients.

RÉSUMÉ

La radiothérapie est l'une des principales options thérapeutiques pour le traitement du cancer thoracique, y compris le cancer du poumon, première cause de mortalité par cancer dans le monde. Cependant, cinq à vingt pour cent des patients qui subissent une radiothérapie souffriront de toxicités pulmonaires liées aux radiations, allant de la pneumopathie à l'irréversible fibrose pulmonaire radio-induite (FPRI). La FPRI se caractérise par une destruction progressive et irréversible de la structure des alvéoles, ainsi que par un dépôt de collagène. Ce développement de tissu cicatriciel altère la fonction normale du poumon : les échanges gazeux entre l'air et le sang, et l'évacuation des déchets gazeux, conduisant à une insuffisance respiratoire et finalement à la mort. Certains des principaux événements moléculaires et cellulaires survenant au cours du développement de cette pathologie ont été décrits, mais les mécanismes détaillés et leur temporalité restent largement inconnus. Par conséquent, l'objectif de mon projet est de mieux comprendre les événements cellulaires et moléculaires précis induits par les radiations, du stade initial de la pneumopathie au stade terminal de la fibrose pulmonaire.

L'utilisation du séquençage d'ARN de cellules uniques (scRNA seq) nous a permis de décrire avec des détails inédits les processus induits par l'irradiation au sein des tissus non tumoraux et l'interaction entre les différentes populations cellulaires, ainsi que la dynamique de ces processus. Dans un premier temps, nous avons utilisé un modèle murin pour comparer l'effet d'une dose d'irradiation fibrogène ou non fibrogène et analyser la séquence temporelle du développement des événements après l'irradiation. Nous avons observé l'initiation de divers processus d'inflammation et de réparation de différentes populations (cellules endothéliales, épithéliales et myéloïdes) de manière similaire après une dose d'irradiation fibrogène ou non fibrogène : régénération des pneumocytes par dédifférenciation des pneumocytes de type II ou transdifférenciation, signalisation pro-angiogénique induisant des processus de réparation de la microvasculature, différenciation des monocytes en macrophages interstitiels. Cependant, trois mois après l'irradiation, les souris irradiées avec une dose non fibrogène guérissent, tandis que les souris irradiées avec une dose fibrogène entrent dans une inflammation chronique, présentent un changement dans le phénotype des fibroblastes qui se différencient en myofibroblastes et produisent une matrice extracellulaire, induisent une transition épithéliale vers mésenchymateuse des pneumocytes.

De plus, ce projet présente la première analyse scRNA seq d'échantillons de tissus pulmonaires non tumoraux de patients ayant subi une radiothérapie pour un cancer du poumon. Les résultats ont mis en évidence l'activation importante de la signalisation pro-angiogénique par l'irradiation dans les tissus pulmonaires non tumoraux et nous ont permis de décrire certains des processus de réparation de la microvasculature, à travers différents états cellulaires endothéliaux tels que les cellules « tip » et « stalk », avec une communication intercellulaire importante avec un sous-type spécifique de macrophages alvéolaires recrutés depuis la circulation.

Ce travail a permis de mieux comprendre les mécanismes conduisant à la FPRI, avec un aperçu unique de la réponse pulmonaire du patient à la radiothérapie. De plus, la disponibilité des données temporelles de souris et des données humaines pourrait faciliter d'autres études. Dans l'ensemble, mon projet a participé à l'effort visant à accroître les connaissances sur la physiopathologie de la fibrose pulmonaire radio-induite afin de trouver des options thérapeutiques pour prévenir, arrêter le développement ou traiter l'évolution de cet effet secondaire mortel de la radiothérapie.

Mots clés : Fibrose pulmonaire radio induite, Séquençage d'ARN de cellule unique, Analyses bioinformatiques

ABSTRACT

Radiotherapy is one of the main therapeutic options for thoracic cancer treatment, including lung cancer, the first cause of cancer-related death worldwide. However, five to twenty percent of the patients that undergo radiotherapy will suffer from lung radiation toxicities, from pneumonitis to irreversible radiation induced pulmonary fibrosis (RIPF). RIPF is characterized by a progressive and irreversible destruction of the alveoli structure, along with collagen deposition. This scar tissue development impairs the normal function of the lung: the gas exchanges between the air and the blood, and the evacuation of gaseous waste, leading to pulmonary failure and ultimately death. Some of the main molecular and cellular events occurring during the development of this pathology have been described, however the detailed mechanisms and their temporality remain largely unknown. Therefore, the objective of my project is to gain a better understanding of the precise cellular and molecular events induced by radiation, from the initial stage of pneumonitis to end-stage pulmonary fibrosis.

The use of single cell RNA sequencing (scRNA seq) allowed us to describe with unprecedented details the processes induced by irradiation in the non-tumoral tissue and the interplay between the different cell populations, as well as the dynamics of these processes. First, we used a mouse model to compare the effect of a fibrogenic dose of irradiation or a non fibrogenic dose of irradiation and the temporal sequence of events development after irradiation. We observed the initiation of various processes of inflammation and repair of different populations (endothelial, epithelial and myeloid cells) in a similar way after either a fibrogenic or non-fibrogenic dose of irradiation: regeneration of the pneumocytes through a dedifferentiation of the type II pneumocytes or transdifferentiation, pro-angiogenic signaling inducing processes of microvasculature repair, monocyte differentiation to interstitial macrophages. However, three months post irradiation, the mice irradiated at a non-fibrogenic dose recover, whereases the mice irradiated at a fibrogenic dose enter chronic inflammation, present a shift in the phenotype of fibroblasts that differentiate into myofibroblasts and produced extracellular matrix, induce epithelial to mesenchymal transition of the pneumocytes.

Furthermore, this project presents the first scRNA seq analysis of non-tumoral lung tissue samples from patients who underwent radiotherapy for lung cancer. The results highlighted the important activation of pro-angiogenic signaling by irradiation in non-tumoral lung tissue and allowed us to describe some of the repair processes of the micro-vasculature, through different endothelial cell states such as the tip and stalk cells, with important intercellular communication with a specific subtype of recruited alveolar macrophages.

This work provided a better understanding of the mechanisms leading to RIPF, with unique insight about the patient's lung response to radiotherapy. Furthermore, the availability of both the mouse temporal dataset and the human dataset could facilitate further studies. Overall, my project participated in the effort increase knowledge on the physiopathology or radiation-induced pulmonary fibrosis in order to find therapeutic options to prevent, stop the development or treat the advancement of this deadly side effect of radiotherapy.

Keywords: Radio-induced pulmonary fibrosis, Single cell RNA sequencing, Bioinformatic analysis