UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Spatial Transcriptomics in COPD: Insights into Smoking-Induced Pathophysiological Changes

by Hao Chen

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Philip Hansbro, Dr. Alen Faiz, Dr. Caitlin Margaret Gillis, A/Prof. Lana McClements.

University of Technology Sydney
Faculty of Science

January 2025

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, ***Hao Chen***, declare that this thesis is submitted in fulfilment of the requirements for the award of ***Doctor of Philosophy***, in the ***School of Life Sciences, Faculty of Science*** at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research was supported by an Australian Government Research Training Program (RTP) Scholarship doi.org/10.82133/C42F-K220.

Signature: Production Note:
Signature removed prior to publication.

Date: 09/01/2025

Acknowledgement

This PhD thesis has been a profound and unforgettable journey for me. I am immensely thankful to all my supervisors, with special appreciation reserved for my primary supervisor, Prof. Philip Hansbro, who also serves as the director of *UTS Centenary Centre for Inflammation*, for entrusting me with this opportunity as a PhD candidate. I extend my gratitude to the *Faculty of Science* for the scholarships that enabled my postgraduate studies at *UTS*. Additionally, I am thankful for the support received from the travel funds provided by UTS, grant and speech invitations provided by *10x Genomics*, travel fund and speech invitation provided by the *Thoracic Society of Australia and New Zealand*, and presentation award provided by *New South Wales Airways Meeting*, which have been instrumental in guiding me towards success as a fledgling researcher.

I wish to extend my deepest appreciation to my co-supervisor, Dr. Alen Faiz, for his invaluable guidance, advice, unwavering support, patience, and granting me the opportunity to work in his lab, *Respiratory Bioinformatics and Molecular Biology*. His extensive expertise and wealth of experience have been a constant source of inspiration and motivation in my academic pursuits. His involvement in refining my bioinformatic experimental approach and offering constructive feedback on my ideas has greatly enhanced my skills as a critical thinker in the realm of scientific inquiry.

I am indebted to my co-supervisor, A/Prof. Lana McClements, who served as my supervisor prior to my PhD and continues to collaborate with me. Her ongoing support and partnership have been instrumental throughout my academic journey. I hope that our collaborations will endure for years to come.

I want to express my gratitude to my co-supervisor, Dr. Caitlin Margaret Gillis, for her dedicated hands-on supervision in the laboratory for approximately a year. She consistently went above and beyond, arriving early to assist with monitoring my animal experiments and staying late to support other aspects of our research. I am truly thankful for her continuous support. As she embarks on her transition from academia to the industry, I offer my sincerest wishes for continued success in her future pursuits. Hopefully, our paths will cross again someday.

I also want to express my appreciation to both past and present technical staff members, including Carol Devine, Simon Gao, and Bob Lu, for their dedicated efforts in managing the

animal smoking facility at *UTS Centenary Centre for Inflammation*. A special acknowledgment goes to Solomon Odgers, whose extensive contributions as a technical officer and now as a research assistant have been invaluable. Additionally, I extend my gratitude to our other research assistants, Dr. Duc Nguyen, Dr. Christina Nalkurthi, Nisha Panth, Linda Tong, and Caroline Wilson, for their consistent support.

I extend my heartfelt gratitude to all my colleagues at *UTS Centenary Centre for Inflammation* under the leadership of my primary supervisor, Prof. Phil Hansbro, and at *Respiratory Bioinformatics and Molecular Biology* led by my co-supervisor, Dr. Alen Faiz. A special thanks goes to my friends and fellow colleagues, Rashad Mohammad Mahbub, Dr. Tayyaba Sadaf, Dr. Sobia Idrees, Dr. Saima Firdous Rehman, Dr. Jos van Nijnatten, Dr. Ridhima Wadhwa, Kyle Kovacs, and Fia Sabrina Boedijono, with whom I have greatly enjoyed working, collaborating, and playing. Their support and camaraderie have enriched my experience immensely.

I express my gratitude to *UTS eResearch* for providing me access to their high-performance computing facility, which has been extensively utilised in all chapters of this thesis. Additionally, I extend my appreciation to the Centenary Institute's animal facility for their technical support in conducting my animal experiments, notably highlighted in Chapters 2 and 3. I also wish to thank Dr. Izabela Galvo and Dr. Vrushali Chimankar for generously providing lung tissue sections from their previous animal experiments, which were instrumental in my spatial transcriptomics experiment in Chapters 2 and 3. Furthermore, I am thankful to *10x Genomics* for their invaluable technical guidance, which greatly influenced the experimental design of my spatial transcriptomics study. Additionally, I am grateful to *10x Genomics* for extending an invitation for me to deliver a speech, marking my inaugural talk for biotechnology industry. Lastly, I would like to express my appreciation to the *American Thoracic Society* for granting me the opportunity to deliver a speech in front of thousands of people in my last year of PhD candidature, making it my largest-ever talk in my PhD candidature.

Apart from expressing gratitude to persons, I also want to acknowledge the significance of objects like my boxing gloves, through the avenue of sport, and my piano, through the realm of music. From childhood to the present, they have served as my primary sources of entertainment, offering me a valuable outlet to alleviate stress during my PhD journey.

Last but certainly not least, I must acknowledge the unwavering support of my parents as well as my partner. Their belief in me has been a constant source of strength and motivation throughout this journey. I am deeply grateful for their steadfast support in all my endeavours and for imparting to me the values of hard work, persistence, and resilience. Their dedication, patience, and tireless efforts serve as an inspiring example for me, shaping me into the person I am today. I owe much of my success to their guidance and love.

Format of Thesis Statement

I hereby state that this submission is a conventional thesis.

List of Publications, Conferences, and Awards

All publications, conferences, and awards are presented in chronological order, with only those happened during the thesis author's PhD candidature included.

Publications:

- Richards C[#], **Chen H[#]**, O'Rourke M, Owen G, Volkerling A, Ghosh A, Gorrie C, Gallego-Ortega D, Bottomley A, Padula MP, McGrath K, Cole L, Hansbro PM[*], McClements L. Matrix directs trophoblast differentiation in a bioprinted organoid model of early placental development. *Nat. Commun.* (major revision).
- **Chen H[#]**, Galvao I[#], Donovan C, Kim RY, Idrees S, Gillis CM, Gallego-Ortega D, Schulte-Schrepping J, Schaar AC, Horvat JC, McClements L, Martinez-Nunez RT, Sauler M, Chotirmall SJ, Segal LN, Faner R, Dharmage SC, Belz GT, Kapellos TS, Schultze JL, Faiz A[#], Hansbro PM[#*]. Spatially Resolved Transcriptomic Mapping Reveals Immune Changes in Lung Parenchyma in Cigarette Smoke-Induced COPD. *Nat. Commun.* (under review).
- Idrees S, **Chen H**, Panth N, Paudel KR, Hansbro PM[*]. Exploring Viral-Host Protein Interactions as Antiviral Therapies: A Computational Perspective. *Microorganisms*. 2024 Mar 21;12(3):630.
- Pouwels SD[#], **Chen H[#]**, Rathnayake SNH, Lan A, Mahbub RM, Yeung ACY, Brandsma CA, Lobo TJ, Hansbro PM, Guryev V, Heijink IH, van den Berge M[#], Faiz A[#*]. Epigenetic adaptation of human airway epithelium upon cigarette smoking. *Nat. Commun.* (under review).
- Idrees S[#], Johansen MD[#], Kovacs KJ[#], Boedijono FS[#], **Chen H**, Galvao I, Donovan C, Kim RY, Sikkema L, Strobl DC, Belz GT, Segal LN, Faner R, van den Berge M, Chotirmall SJ, Nawijn M, Gote-Schniering J, Lehmann M, Kapellos TS, Lucken MD, Gallego-Ortega D, Faiz A[#], Hansbro PM[#*]. Single-Cell Mouse Lung Disease Atlas. *Nat. Methods.* (under review).

Conferences:

- INVITED PRESENTATION | *Thoracic Society of Australia and New Zealand (TSANZ)* Education Hub 2024 – Melbourne, Victoria, AU.

- ORAL & CHAIR | *New South Wales Airways Meeting (NAMe)* 2024 – Sydney, New South Wales, AU.
- INVITED PRESENTATION | *University of Sydney Single-cell and Spatial Biology* Monthly Seminar 2024 – Sydney, New South Wales, AU.
- ORAL | *American Thoracic Society (ATS)* International Conference 2024 – San Diego, California, USA.
- INVITED PRESENTATION | *10x Genomics* User Group Meeting 2024 – Sydney, New South Wales, AU.
- POSTER | *TSANZ and Australian and New Zealand Society of Respiratory Science (TSANZSRS)* Annual Scientific Meeting 2024 – Gold Coast, Queensland, Australia.
- ORAL | *Australian Society of Medical Research (ASMR)* National Scientific Conference 2023 – Melbourne, Victoria, AU.
- ORAL | *NAMe* 2023 – Sydney, New South Wales, AU.
- POSTER | *ASMR* New South Wales Annual Scientific Meeting 2023 – Sydney, New South Wales, AU.

Honours & Awards:

- Best Basic Science PhD Student Award (AUD $500) | *NAMe*, Sydney | 2024.
- Travel fund (AUD $250) | *TSANZ* Education Hub, Melbourne | 2024.
- Poster Finalist | *TSANZSRS* Annual Scientific Meeting, Gold Coast | 2024.
- Single Cell and Spatial Technology Access Program Fund (equivalent to ~AUD $16,000) | *10x Genomics* | 2023.
- Vice Chancellor's Conference Funds | *UTS* | 2021, 2022, 2024.
- President's Scholarship | *UTS* | 2021.
- International Research Scholarship | *UTS* | 2021.

Table of Contents

List of Abbreviations

*Many genes appear throughout the thesis; these are presented as standard gene symbols rather than in their full forms.

aCap, alveolar capillary endothelial cell.

aEC, arterial endothelial cell.

ALI, air-liquid interface.

AlvFibro, alveolar fibroblast.

AlvM, alveolar macrophage.

AntiInflam mono, anti-inflammatory monocyte.

aSMC, airway smooth muscle cell.

AT1, pulmonary alveolar cell type 1.

AT2, pulmonary alveolar cell type 2.

BAL, bronchoalveolar lavage.

Cd4 Cd8 T, Cd4-Cd8 double-positive T cell.

COPD, chronic obstructive pulmonary disease.

DC, dendritic cell.

DGE, differential gene expression.

eQTM, cis-expression quantitative trait methylation.

FEV1, forced expiratory volume in 1 second.

FVC, forced vital capacity.

gCap, general capillary endothelial cell.

GOLD, Global Initiative for Chronic Obstructive Lung Disease.

ILC2, group 2 innate lymphoid cell.

IntM, interstitial macrophage.

KO, knockout.

lymphEC, lymphatic endothelial cell.

LymphFoll, lymphoid follicle.

MastBaso, mast cell and basophil.

Meso, mesothelial cell.

Neut, neutrophil.

NGS, next-generation sequencing.

NK, natural killer cell.

PCR, polymerase chain reaction.

Peri, pericyte.

PeriBronc fibro, peribronchial fibroblast.

ProInflam mono, pro-inflammatory monocyte.

QC, quality control.

ROI, region of interest.

SA-β-gal, senescence-associated β-galactosidase.

scRNA-seq, single-cell RNA sequencing.

ST, spatial transcriptomics.

TF, transcription factor.

TSS, transcription start site.

UMAP, Uniform manifold approximation and projection.

vEC, venular endothelial cell.

vSMC, vascular smooth muscle cell.

List of Tables and Figures

Abstract

The spatial architecture of the lung plays a vital role in maintaining homeostasis, ensuring effective gas exchange, and mounting immune responses to external respiratory challenges such as infections and pollutants. Chronic obstructive pulmonary disease (COPD) is a complex, progressive and heterogeneous lung disease characterised by progressive inflammation in the lungs, driving airway remodelling and alveolar destruction (also known as emphysema). These pathological changes ultimately lead to severely impaired lung function and breathing difficulties. COPD is the third leading cause of illness and death worldwide, with cigarette smoking being the predominant risk factor, though other exposures, such as air pollution, also contribute. Current therapies suppress symptoms but fail to halt disease progression or reverse the disease. This therapeutic limitation stems from an incomplete understanding of the disease's pathogenesis in humans and its representation in mouse models. Single-cell transcriptomics has significantly expanded our molecular understanding of COPD, while the spatial context of the disease mechanisms remains largely unexplored.

Chapters 2 and 3 of this thesis investigate the cellular and molecular dynamics across different lung regions, offering new insights into the pathogenesis of the development and progression of COPD. By focusing on the spatial specificity of disease mechanisms, these two chapters inform future strategies for personalised COPD management. Using experimental COPD animal models induced by chronic cigarette smoke exposure, we investigated the molecular mechanisms and their spatial resolution in experimental COPD that has all the hallmark features of the human disease in a controlled environment. The whole lung was also profiled. This platform reduced the inherent heterogeneity seen in COPD patients. Spatial transcriptomics (ST) and single-cell RNA sequencing (scRNA-seq) were employed to define cell-type-specific gene expression patterns within the lung microenvironment, capturing key phenotypic features that mirror those observed in human COPD lungs, such as alveolar macrophages spreading from pulmonary vasculature to distal parenchyma and lymphoid follicle of increasing size along with continuous smoking. These chapters show two regions—distal parenchyma and lymphoid follicles—as areas of previously unexplored pathogenesis, providing novel insights into the development and progression of COPD.

Chapter 4 of this thesis investigated the epigenetic effects of cigarette smoking on human airways. A distinct "smoking signature" gene set was identified, effectively distinguishing current smokers from non-smokers across cohorts. ST and scRNA-seq mapping revealed that smoke-induced damage predominantly affects the airway surface epithelial layer. Since the advent of cooking with fire, smoke exposure has been a consistent aspect of human history, potentially leading to adaptations in the human airway to counteract the harmful effects of smoke. Comparative analysis between the lungs of human and non-human primate species revealed a group of genes we termed the “human lung evolution signature” specifically expressed in the airway surface epithelial layer of human smokers, suggesting an evolutionary adaptation to smoke exposure. We hypothesised that this adaptation may have resulted from prolonged smoke exposure throughout human history, potentially leading to epigenetic modifications. The transcription factors (TFs) *NRF2* and *AhR* were identified as master regulators within the smoking signature. Further epigenetic analysis uncovered their binding sites at *NQO1* and *ALDH3A1*, both of which are upregulated in human airways in response to smoke exposure. Functional knockout studies of these genes demonstrated their protective role in aiding human airways to adapt to smoke.

In chapter 5, we evaluated the role of cellular senescence in COPD using our animal model of experimental COPD. Our findings show that senescent cells accumulate in COPD lungs, contributing to disease pathology. Administering senolytics successfully eliminated these cells, reduced immune cell infiltration, and improved structural damage associated with alveolar destruction. However, while this intervention alleviated certain aspects of lung damage, it did not fully restore impaired lung function. This is likely due to either the relatively small number of senescent cells or the clearance of senescent structural cells that still provide the structural support. Nevertheless, senolytics could serve as a first line therapy for selected patients or complementary therapy alongside other treatments aimed at managing COPD symptoms, because preclinical evidence highlights their ability to alleviate specific disease features, such as reduced immune cell infiltration and emphysema.

In conclusion, this thesis provides a comprehensive framework for understanding COPD pathogenesis, leveraging spatial and molecular insights to uncover novel therapeutic avenues. By bridging basic research with translational potential, this work represents a significant step toward precision medicine in COPD.

Chapter 1: General Introduction

1.1 Overview of COPD and Smoking

Chronic obstructive pulmonary disease (COPD) is a complex, progressive and heterogeneous lung disease characterised by progressive inflammation in the lungs, driving airway remodelling and alveolar destruction (also known as emphysema)[1]. These pathological changes ultimately lead to severely impaired lung function and breathing difficulties[1]. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) provides the most widely adopted guidelines for diagnosis, prognosis, and management of COPD[1]. Following the GOLD definition, approximately 391 million people aged 30-79 years worldwide were estimated to have COPD in 2019[2], representing approximately 10% of the global population. It is now the third leading cause of morbidity and death worldwide[1]. The major current risk factor for COPD is cigarette smoking but its aetiology is multifactorial. Chronic exposure to air pollution, cooking smoke, toxic pollutants, lifetime asthma and childhood events such as infections, passive smoke exposure and prematurity are also significant contributors[1].

The link between smoking and COPD was established decades ago, with early expert consensus in 1984 noting that 80-90% of COPD cases in the United States could be attributed to smoking[3]. In the 2000s, it was suggested that potentially more than 15% of smokers develop COPD[4], and a large-scale systematic review further identified smoking and ageing as the two biggest risk factors[2]. Although nonsmokers can also develop COPD, smokers exhibit a higher prevalence of respiratory symptoms, a faster annual decline in lung function, and a higher COPD mortality rate compared to non-smoker COPD[5].

1.2 Management of COPD

In clinical practice, a diagnosis of COPD is often considered in patients presenting with respiratory symptoms such as shortness of breath, chronic cough, or sputum production, as well as a history of recurrent lower respiratory tract infections and/or exposure to risk factors for the disease, notably cigarette smoking. Confirmation of COPD diagnosis requires spirometry, with a post-bronchodilator FEV1/FVC ratio of less than 0.7. The FEV1 value is often

used to determine the severity of airflow obstruction, categorised by GOLD grades 1 to 4 (mild, moderate, severe, and very severe; Table 1).

Table 1: GOLD diagnostic criteria for different stages of COPD.

Grade	**Stage**	**FEV1 (% predicted)**	**Symptoms**
GOLD 1	Mild	≥80%	Chronic cough and sputum production may be present, but many are asymptomatic.
GOLD 2	Moderate	50–79%	Increased dyspnoea, cough, and sputum production. Symptoms may affect daily activities.
GOLD 3	Severe	30–49%	Further increased shortness of breath, fatigue, and reduced exercise tolerance.
GOLD 4	Very severe	<30%	Chronic respiratory failure symptoms (e.g. hypoxemia, hypercapnia), frequent exacerbations, and a significant impact on quality of life.

COPD, chronic obstructive pulmonary disease; FEV1, forced expiratory volume in 1 second; GOLD, Global Initiative for COPD.

Pharmacological interventions for COPD primarily involve inhaled therapies, typically consisting of either a bronchodilator monotherapy or a dual combination of two long-acting bronchodilators to improve airflow. In addition, inhaled corticosteroid is often added to the dual bronchodilator combination to control inflammation. Surgical or endoscopic lung volume reduction may be considered in selected severe patients with symptomatic emphysema and lung hyperinflation, and lung transplantation is an option for patients with very severe COPD. However, there are currently no pharmacological therapies that stop the progression or reverse the disease. For most patients, non-pharmacological and non-invasive approaches are

currently central to COPD management, primarily including smoking cessation for current smokers, oxygen therapy for COPD exacerbations and pulmonary rehabilitation.

Collectively, current mainstream management plans for COPD primarily focus on symptom control rather than addressing the underlying pathogenesis of the disease, let alone incorporating precision medicine approaches. Furthermore, approximately 40% of current smokers continue smoking after a COPD diagnosis, which accelerates disease progression for the majority of COPD cases[6]. Therefore, basic science research is essential to develop new treatments that target disease pathogenesis and advance precision medicine.

1.3 Pathogenesis of COPD and Smoking

A puff of cigarette smoke contains millions of liquid droplets the size of approximately 0.45µm, each carrying a complex mix of toxic chemicals from tobacco and additives for flavour[7]. In the upper respiratory tract (nose, mouth, larynx) and central conducting airways, larger inhaled particles leave the flowing stream of gas and collide with the airway walls. In the lower respiratory tract, gas flow slows due to their much larger cross-sectional area, and smaller particles reach the deeper airways. Overall, toxic gases and particulates spread throughout the lungs, with particles decreasing in size from upper to lower airways (Fig. 1.1). Smoking damages tissues at all levels, both directly through oxidative stress and indirectly by inducing inflammation.

Elucidating the immunology that underpins pathogenesis is central to understanding the molecular and cellular pathogenesis of COPD. COPD typically involves activation of innate immune responses, with neutrophils and macrophages infiltrating the lungs[8]. Among immune cells, macrophages are the most representative population in the respiratory tract due to their crucial roles in airway surveillance, clearance of cellular debris, immune regulation, and resolution of inflammation[9]. Pulmonary macrophages can be broadly classified into two major populations residing in distinct anatomical compartments: alveolar macrophages, which line the alveolar surface, and interstitial macrophages, which occupy the space between the alveolar epithelium and vascular endothelium. Alveolar macrophages are generally more active in pathogen phagocytosis and can originate from recruited monocytes, whereas interstitial macrophages exhibit lower phagocytic activity and are thought to play a more

regulatory role[9]. In the lungs of COPD patients, numbers of both types of macrophages are markedly increased, correlating with disease severity and regions of tissue destruction[10-12].

Chapter 2 contributes new insights into the spatial characteristics of innate immunity in smoking-induced experimental COPD. In severe COPD, lymphocytes can form tertiary lymphoid structures known as "lymphoid follicles", suggesting that adaptive immune responses may be a pathological feature. Chapter 3 investigates the role of lymphoid follicles in experimental COPD, documenting their origin, development and progression for the first time.

In COPD, structural changes are primarily observed in the airways and lung parenchyma. Airway remodelling results in thickened walls involving the epithelium, lamina propria, smooth muscle, and adventitia of the upper and lower airways. Micro-computed tomography (micro-CT) has shown a 40% reduction in the number of small airways in mild-to-moderate COPD[13] and an 80% reduction in severe-to-very-severe COPD[14]. In the alveoli-rich lung parenchyma, alveolar destruction and collapse (emphysema) occurs. Historically, an imbalance in protease and anti-protease activity, often due to neutrophilic infiltration or reduced anti-protease activity (i.e. a-anti-trypsin deficiency), was thought to drive emphysema[15]. However, since the 2000's, additional emphysema-driving factors have been proposed, including increased apoptosis[16] and necroptosis[17], maintenance failure in the lungs[18], as well as oxidative stress, autoimmunity, malnutrition, or a combination of these factors[19-21]. Chapter 2 provides new insights into the translocation of alveolar macrophages (AlvMs) in emphysema, and Chapter 5 investigates cellular senescence as a contributing factor to emphysema and also demonstrates that eliminating senescent cells can improve emphysema.

Pulmonary circulation can be altered in COPD. A reduction in capillary surface area is often associated with emphysema, increased hypoxic pulmonary vasoconstriction, or both in patients with severe disease[22]. In those with milder airflow limitation, pulmonary vascular inflammation and endothelial dysfunction have also been observed[23]. Chapter 2 illustrates the changes of pulmonary capillaries in COPD.

Cigarette smoking remains the leading risk factor for COPD, but the transition from asymptomatic smoking to COPD is not fully understood. Using a comprehensive approach, Chapter 3 examines the spatiotemporal changes in the lungs over time to profile the development of COPD from smoking and the progression of smoking-induced COPD.

In summary, clinical COPD is a highly heterogeneous disease, encompassing a wide range of features that may or may not manifest in individual cases. To address this complexity, Chapters 2, 3, and 5 focus on studying COPD using experimental animal models. The use of experimental COPD animal models provides a controlled environment to investigate the effects of key exposures, such as cigarette smoke[24], enabling *in vivo* exploration of disease mechanisms while minimising the variability inherent in clinical COPD.

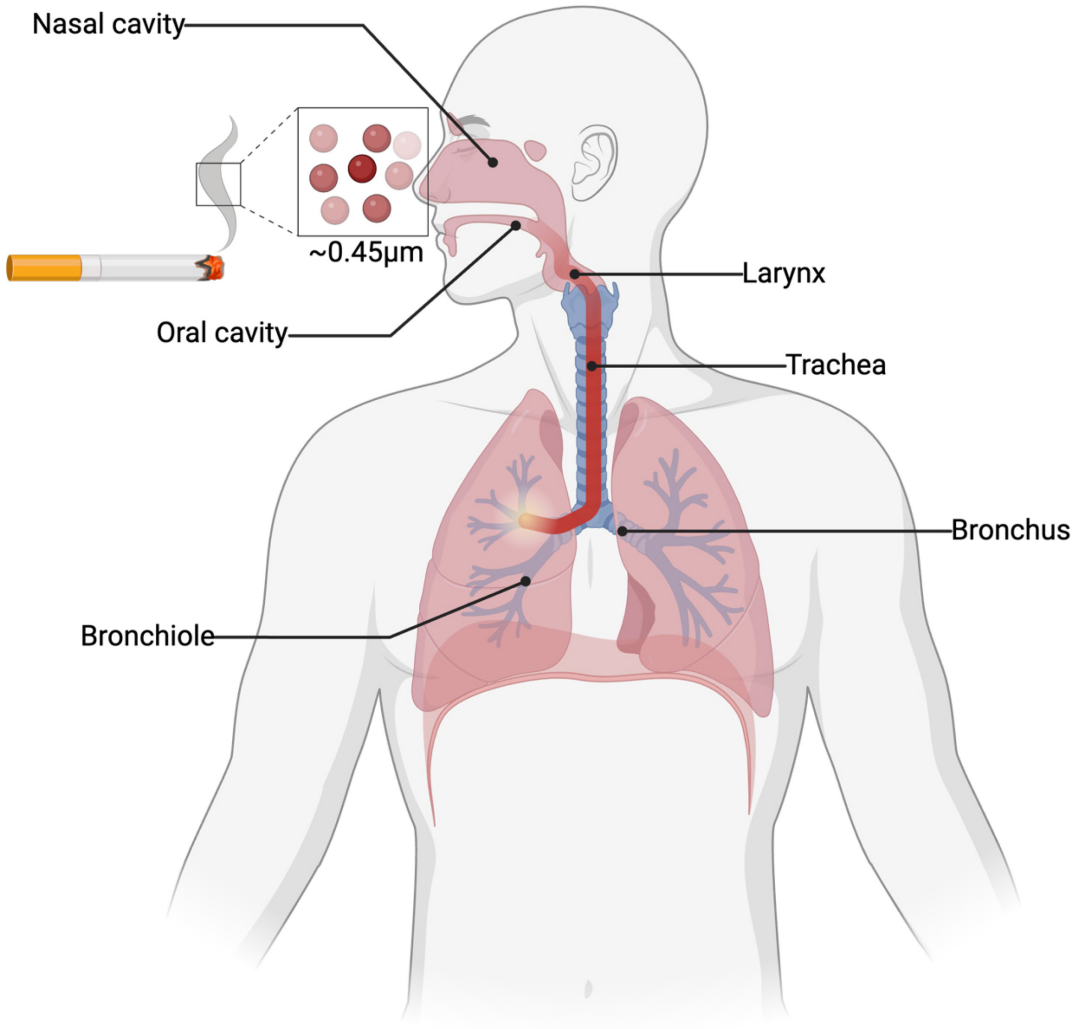


Fig. 1.1: Schematic illustration of the spread of harmful particles from cigarette smoke through the airways.

1.4 Single-Cell Transcriptomics

The transcriptome of individual cells can vary significantly, even within the same tissue. Single-cell RNA sequencing (scRNA-seq) provides an unbiased approach to studying the dynamic cellular transcriptome at single-cell resolution. scRNA-seq technology was recognised as the "Method of the Year 2013" by *Nature Methods*[25], highlighting its already proven transformative potential in uncovering new insights into human physiology and disease by characterising unique gene expression profiles in specific cell types. Briefly, scRNA-seq

employs next-generation sequencing (NGS) to capture nucleic acid sequence information from individual cells. Unlike traditional bulk RNA sequencing, which measures RNA expressions from a mixture of all cells in a sample, scRNA-seq offers cellular resolution by capturing transcripts from each single cell. This enables a deeper understanding of the cellular composition within tissues and the discovery of rare or previously uncharacterised cell types.

Since its' introduction in 2009, the use of scRNA-seq has grown exponentially, with thousands of publications annually[26] and the development of over 1,000 bioinformatic pipelines to process the data[27]. This rapid advancement underscores the sophistication and versatility of scRNA-seq. In respiratory medicine, scRNA-seq has proven particularly impactful. In 2018, two independent studies using scRNA-seq on human and mouse airway tissues identified ionocytes as a rare airway cell type[28,29]. These pulmonary ionocytes were found to be the primary source of the cystic fibrosis transmembrane conductance regulator gene, and whose expression was muted in cystic fibrosis. Using scRNA-seq, our lab uncovered why COPD patients tend to experience more severe symptoms following SARS-CoV-2 infection, revealing the emergence of IFN-responsive club and goblet cells in the airway epithelium during infection[30].

1.5 Spatial Transcriptomics

Despite the success of scRNA-seq, a significant practical challenge remains in isolating viable cells from whole tissues without inducing stress, cell death, or cell aggregation. Immunology research has benefited greatly from scRNA-seq, as many immune cells are relatively easy to isolate from blood, lymphoid organs, peripheral tissues, and even tumours, where they are often not anchored. In contrast, certain cell types, such as neurons, are less suited to scRNA-seq, requiring specialised dissociation protocols to isolate them. Most importantly, scRNA-seq does not provide spatial resolution needed to elucidate cellular crosstalk and how disease networks operate. This has driven interest in performing transcriptomic experiments on intact tissues, preserving spatial information, known as spatial transcriptomics (ST).

Although gene profiling in tissues has existed for decades, spatial transcriptomics was only named "Method of the Year 2020" by *Nature Methods*[31]. The use of spatial transcriptomics is growing rapidly, fuelled by factors such as reduced costs of next-generation sequencing (NGS),

initiatives such as the Human Cell Atlas, computational advancement, and imaging improvements.

Two main platforms for whole-transcriptome spatial transcriptomics are Visium or Visium HD by *10x Genomics* and GeoMx by *nanoString Technologies*. Visium and Visium HD enable whole-transcriptome profiling across the entire tissue within the capture region, whereas GeoMx allows whole-transcriptome profiling only within user-defined regions. Chapters 2 and 3 utilise Visium and GeoMx platforms, respectively.

1.6 Spatial Transcriptomics in Respiratory Medicine

As a technology introduced after 2020, ST remains a very new field in respiratory medicine. In 2023, a spatial atlas of healthy human lungs, covering both the upper and lower respiratory tracts, was established as a baseline[32]. This identified a new immune niche in human lung physiology, where airway submucosal gland epithelial cells recruit B cells and IgA^+ plasma cells, supporting their longevity and local antibody secretion through *CCL28*, *APRIL*, and *IL6* expression.

In a disease context, researchers spatially characterised three key disease-specific niches in idiopathic pulmonary fibrosis, revealing interactions within these niches[33]. These include a fibrotic niche with myofibroblasts and aberrant basaloid cells around the airways, adjacent to an airway macrophage niche in the airway lumen containing *SPP1*$^+$ macrophages, and an immune niche with lymphoid cell foci in fibrotic tissue, surrounded by remodelled endothelial vessels.

In COPD, the GeoMx ST platform was used to profile differences in lymphoid follicles in the lungs of COPD patients with and without emphysema[34]. They found that off-target B cell activation within lymphoid follicles was associated with the pathogenesis of autoimmune-driven emphysema. Building on this work, Chapter 3 profiles the spatiotemporal changes of B cells in relation to lymphoid follicles during the development and progression of experimental COPD. To date, little is known about the spatial aspects of COPD pathogenesis, with Chapters 2 and 3 pioneering this field.

Chapter 2: Spatially Resolved Transcriptomic Mapping Reveals Immune Changes in Lung Parenchyma in Cigarette Smoke-Induced COPD

2.1 Declaration

Chapter 2 is a thesis-adapted version of a manuscript currently under review in *Nature Communications*, in which the author of this thesis is the first author. Additionally, Chapter 2 includes a large single-cell atlas of mouse lungs, titled the "Mouse Lung Disease Cell Atlas," associated with another manuscript in preparation, where the author of this thesis is the second author.

2.2 Abstract

COPD is a prevalent heterogeneous lung disease often driven by prolonged exposure to inhaled noxious particles, notably cigarette smoke. Animal models particularly those induced by cigarette smoke exposure enable the unravelling of disease pathogenesis. Conventional omics technologies lack spatial context, limiting insight into spatial dynamics in diseased lungs. Here, for the first time we employed ST and scRNA-seq to define the spatially resolved gene expression profile of lung tissues from mice with cigarette smoke-induced experimental COPD and controls. Through multi-modal data integration, we identified spatial and cell type changes in pathogenic processes. Initially, we identified deeper penetration of COPD-associated signals into lung structures compared to smoking-related signals in both mouse and human lungs. We then found that AlvMs near vasculature migrated towards the distal parenchyma in experimental COPD lungs. This finding was consistently mirrored by the dispersal pattern of AlvMs from the vasculature to the parenchyma in the lung of a human current smoker. We defined the changes (e.g. increased *Lcn2* and decreased *Klf2*) and the role of the microvasculature in tissue injury and repair processes in the distal parenchyma in COPD. Our study defines how the intricate spatial communication between the distal parenchyma and the vasculature interact to induce inflammation in COPD. Specifically, we spatially determined the mechanisms of the induction of migration of AlvMs from the vasculature to

the distal parenchyma, where they release inflammatory signals that cause damage to alveolar structures in the distal parenchyma.

2.3 Introduction

COPD is a heterogeneous lung condition characterised by inflammation that commences in airways and progresses to lung parenchyma, leading to alveolar destruction[1]. It is now the third leading cause of morbidity and death worldwide[1]. The major current risk factor for COPD is cigarette smoking but its aetiology is multifactorial. Chronic exposure to toxic pollutants, lifetime asthma and childhood events such as infections, passive smoke exposure and prematurity are significant contributors.

The fine spatial architecture of the lung is critical for maintaining homeostasis and its function, to enable effective gas exchange and mount inflammatory and immune responses to protect from respiratory challenges such as infections or inhaled pollutants. It is now well-established that organ-intrinsic niches can dictate effective differentiation of immune cells such as macrophages[35], yet how their spatial organisation determine responses to respiratory challenges remains largely unexplored. Our molecular understanding of COPD has recently been expanded using single-cell transcriptomics[36-41], however, the spatial context of disease mechanisms is largely unexplored. Identifying the cellular and molecular dynamics in different regions of the lungs will provide crucial insights and improve the understanding of COPD disease development and progression in the lungs and inform the development of personalised management with spatial specificity and precision.

The use of experimental animal models enables us to assess the impact of driving exposures such as cigarette smoke[24], offering a controlled setting to study disease mechanisms *in vivo* while mitigating the inherent heterogeneity of COPD. In this study, we employed Visium spot-based ST and analysed scRNA-seq data to comprehensively examine the cell-type-specific gene expressions within the tissue microenvironment in the lung. We analysed mouse lungs following 8- and 12-week exposures to cigarette smoke that drive the development and progression of experimental COPD, respectively, capturing key phenotypic features of disease that closely resemble those observed in human lungs.

We performed Visium spot-based ST on n=6 lung tissue cryo-sections from cigarette smoke-exposed mice and two from normal air-exposed control mice, generating a mouse lung ST dataset of 8,981 spots with an overall median of 15,558 genes detected per spot. We mapped the spatial locations of subpopulations of epithelial, endothelial, stromal, and immune cells. We also investigated intercellular communication between neighbouring cells, uncovering new insights into the recruitment of AlvMs from circulation to the lung periphery in COPD. Key changes in the distal parenchyma, such as increased *Lcn2* and decreased *Klf2*, were identified. In addition, we found that pulmonary endothelial cells, particularly capillary endothelial cells in the distal parenchyma, play a crucial role in the development (when disease features emerge) and progression (increased severity of disease features) of COPD. Our study provides a spatially resolved gene expression profile of the lungs in cigarette smoke-induced experimental COPD. This introduces a new paradigm of spatial heterogeneity of cellular and molecular dynamics in the lung. We have made our spatial gene expression data publicly available as an interactive data-exploratory platform in the Mouse Lung Disease Cell Atlas (manuscript under review in preparation) providing a valuable reference map for future studies and the potential development of novel therapeutics.

2.4 Methods

Ethics

The Animal Welfare Committee of Sydney Local Health District approved all experimental procedures (protocol no.: 2018-030). This study was conducted in accordance with the Australian responsible use of animal code.

Cigarette smoke-induced experimental COPD

The approach for inducing experimental COPD via nose-only cigarette smoke inhalation was performed as previously described[17,42-53]. Briefly, 8-week-old wild-type female C57BL/6J mice (n=6) were simultaneously exposed to cigarette smoke using a custom-designed and purpose-built nose-only, directed flow inhalation and smoke-exposure system (CH Technologies, USA). The mice inhaled the smoke of 12 3R4F reference cigarettes (University of Kentucky, USA) twice per day, five days per week, for 8-12 weeks, with each smoke exposure lasting 75 minutes. Age-matched control mice (n=2) were exposed to normal air. The procedure of

cigarette smoke exposure was approved by the Animal Welfare Committee of Sydney Local Health District (protocol no.: 2018-030). After 8-12 weeks of exposure, mice were euthanised, and the upper left lobes of the lungs were harvested for analysis.

Lung tissue processing

Fresh murine lung cryo-sections were stained with haematoxylin and eosin. Region selection of cryo-sections was not required, as the capture area of ST platform (6.5 x 6.5 mm) was able to include the entire lung sections. Detailed descriptions are in the online methods supplement.

Spatial gene expression assay

Lung cryo-sections (8µm) were placed on pre-chilled optimisation slides (Visium, *10x Genomics*, PN-1000193). The optimisation process determined that the optimal digestion and RNA release from the tissue slide occurred after 18 minutes of permeabilisation. Spatial gene expression slides (Visium, *10x Genomics*, PN-1000187) were used for ST according to the manufacturer's instructions. Brightfield histological images were captured with a 10x objective on an ECLIPSE TiE (*Nikon*) microscope, and the raw images were stitched together using NIS-Elements software (*Nikon*). NGS libraries were prepared as per the Visium user guide. Libraries were loaded at 300pM and sequenced on a NovaSeq 6000 System (*Illumina)* as recommended by *10x Genomics*. Each lung tissue covered ~1,000 of the 5,000 barcoded spots available of the capture area.

Pre-processing of spatial gene expression data

SpaceRanger software (v1.2, *10x Genomics*) was used to pre-process the sequencing data. The option '–reorient-images' was enabled to allow for automatic image alignment, and the mouse reference genome (mm10) was used for data alignment. Filtered feature-barcode expression matrices from SpaceRanger were used as input for the ST analysis in Seurat (v4.2)[54-57]. Data were filtered to include only spots on the tissues. Individual count matrices were normalised with sctransform[58], and the scaled matrices were calculated for comparative analyses using Seurat's default settings.

scRNA-seq data for experimental COPD

We employed a scRNA-seq dataset of murine lungs derived from the same experiment to ensure a maximally transcriptionally similar reference dataset for cell deconvolution and decomposition of ST. This scRNA-seq dataset was normalised with sctransform using Seurat, and a shared nearest neighbour (SNN) graph was built with the first 30 principal components (PCs) using Seurat's FindNeighbours. Cells were clustered with a Louvain algorithm with FindClusters and annotated at resolution of 0.8.

Cell-type deconvolution of spatial transcriptomic data

Cell-type composition of the sctransform normalised spot-based ST data was decoded to single-cell resolution using SpaTalk (v1.0)[59] through the RCTD deconvolution method (v2.2)[60] as implemented in dec_celltype. We applied a non-negative linear model (NNLM) to obtain the optimal proportion of cell types for each ST spot using the matched scRNA-seq dataset with 27 cell types as a reference. Cell types with varied proportions were then used to project cells from the scRNA-seq dataset onto ST spots.

As recommended by the SpaTalk pipeline developers[59], and others[61-63], we assumed a maximum number of 30 cells in each capture spot to determine the optimal cellular combination for reconstruction of the single-cell ST data matrix. The intra-spot positions of cells are simulated, and the cell-type proportion per spot wes calculated according to the estimated intra-spot cell count.

Spatial map of cell dependencies

We used MISTy's implementation in mistyR (v1.8.0)[64] to infer the importance of the abundance of each cell type in explaining the abundance of the other cell types. We estimated cell-type abundances for all slides using a multi-view model that considered three different spatial contexts: (1) an intrinsic view that measured the relationships between the deconvolution estimations within a spot, (2) a juxta view that summed the observed deconvolution estimations of immediate neighbours (effective radius = 5 spots), and (3) a para view that weighted the deconvolution estimations of more distant neighbours of each cell type (effective radius = 15 spots). The aggregated estimated standardised importances (mean) of each view of all slides were interpreted as cell-type dependencies in different spatial contexts, such as co-localisation or mutual exclusion.

Functional characterisation of spatial transcriptomic data

We estimated all 11 validated signalling pathway activities in each spot of normalised ST data in the PROGENy (v1.20) package[65,66] using the top 1,000 responsive genes ranked by p-value in each pathway gene set. The spatial domains and spatially variable genes were identified by SpaGCN (v1.2.5)[67] using log-normalised ST spots. The hyperparameters pf SpaGCN were set to a=1 and b=49 in calculate_adj_matrix, and n_clusters=5 was used to calculate the optimal resolution for model training. Additionally, we constructed meta genes with optimal performance as markers for individual spatial domains of murine lungs. To locate the smoking- and COPD-associated biological processes in the lungs, we obtained (1) a microarray dataset comparing the lungs of control mice to those exposed to cigarette smoke for a shorter duration (4 weeks, data available upon request), which represents the molecular changes induced by cigarette smoke inhalation, and (2) a human COPD-related gene set from a microarray study comparing smoker's lungs with and without COPD[68]. Hypergeometric tests using these two gene sets were conducted on each ST spot. The module scores of these gene sets were calculated using Seurat's AddModuleScore function.

Definition of molecular niches in spatial gene expression

To generate an integrated atlas of all lung tissue samples, we merged sctransform normalised matrices and performed dimensionality reduction with principal component analysis (PCA) using the top 3,000 most variable genes that were shared between all samples. PCA correction was conducted using Harmony (v0.1.1)[69] to remove sample variability. A SNN graph was built with the first 30 PCs using Seurat's FindNeighbours, and the spots were clustered with a Louvain algorithm with FindClusters at resolution of 0.7 and 0.2. We termed the resulting clusters molecular niches.

Definition of cell-type niches in lung structure

We identified niches in murine lungs, representing clusters of spots that were similar in cell composition and were potentially shared "structural building blocks". Briefly, we transformed estimated cell-type proportions of each ST spot into isometric log ratios and clustered spots into groups. To create a SNN graph, we used scran's (v1.26.2)[70] buildSNNGraph function with k=50 for the numbers of neighbours and performed Louvain clustering. The optimal clustering resolution was determined by selecting the value that maximised the mean silhouette score

of each resulting cluster. To define cell-type niches in ST, we assigned the over-represented cell types in each spatial domain by comparing the distribution of cell-type composition within a cell-type niche and the rest using Wilcoxon tests (FDR < 0.05).

Cell-cell communications

To estimate pairwise ligand-receptor (LR) interactions between the cells in ST, we used the find_lr_path function in SpaTalk[59] to infer spatially resolved LRT signalling networks between each pair of single neighbouring cells in each lung tissue. We focused on manually curated mouse LR pairs provided by CellTalkDB[71]. True cell-cell communications in each sample (the ratio of LR co-expression in receiver cells compared to other cell pairs, P <0.01) were included in the downstream analyses. We extracted the unique cell-cell communications that were specifically present in all COPD lung samples and absent in either of the control samples, and the cell-cell communications that are shared between all samples. In addition, we also obtained the common cell-cell communications that were shared between all mouse lungs exposed to smoke for 8 or 12 weeks. We then compared the combined scores of intracellular and intercellular signalling networks in all shared cell-cell communications using unpaired t-tests. The significance threshold was defined as BH-adjusted P <0.05 for shared cell-cell communications between the corresponding groups of samples in each comparison. LR analysis of the experimental COPD lung scRNA-seq dataset was performed using CellChat (v2.1.2)[72] with default settings. Detailed descriptions are in the online methods supplement.

2.5 Results

Integrative spatial map of cell-type dynamics in the lungs in experimental COPD

COPD is characterised by chronic, non-resolving inflammation, and ineffective wound repair of the small airways and alveolar tissues[1]. It involves complex intercellular interactions that likely dictate mechanistic pathways, from aberrant intracellular and molecular signalling to macroscopic alveolar destruction. We hypothesised that defining disease-related gene expression spatially in the lungs would provide novel insights into the development and progression of COPD. To assess this, we applied an integrative transcriptomic strategy combining spot-based ST (Visium, *10x Genomics*) and scRNA-seq (Chromium, *10x Genomics*) from the same mouse lung tissues to spatially map gene expression patterns and cellular

interactions during homeostasis and in cigarette smoke-induced experimental COPD at unprecedented resolution. We interrogated our established mouse model of experimental COPD that entails chronic cigarette smoke inhalation for 8 or 12 weeks (Fig. 2.1a). After 8 weeks, smoke-exposed mice develop the hallmark features of COPD with lung inflammation, airway fibrosis, emphysema, and impaired lung function, all characteristic features of COPD in human lungs[17,42-53]. We have shown that the damage associated with smoke exposure drives the development of COPD (8-week exposure to cigarette smoke) and progresses and is non-resolving (12-week exposure), resembling severe progressive COPD in humans[42].

We profiled spatial gene expression in lung tissues from 8 mice, including normal air-exposed controls (n=2), and mice exposed to cigarette smoke for 8 (n=3) and 12 weeks (n=3) (Fig. 2.1a). For each lung specimen, we obtained 8μm cryo-sections from the upper left lobe for spot-based spatial gene expression profiling. After quality control, the dataset consisted of a total of 8,981 spots (average 1,122 spots per sample) with an overall median of 15,558 genes detected per spot. This dataset represents the first spatial profiling of mouse lungs with experimental COPD. We devised an integrative data analysis approach spanning two modalities containing ST and scRNA-seq to define cell-specific information and interactions in space, deciphering cellular, molecular and functional characteristics of tissue structure in control lungs and in lungs with cigarette smoke-induced experimental COPD (Fig. 2.1b).

The ST dataset comprised barcoded spots, with each spot containing multiple cells. To deconvolute the cells in each spot most reliably and accurately, we employed our scRNA-seq dataset generated from the same mouse lungs as a referencing map. Our reference dataset was built by clustering all cells based on the integrated scRNA-seq data from all samples after batch correction to determine all cell populations present in the murine lungs irrespective of disease status. The clusters were annotated with curated marker genes from existing knowledge, and a total of 27 cell types were identified (Fig. 2.1c). In the scRNA-seq dataset, AlvMs and neutrophils were predominantly enriched cells in cigarette smoke-induced COPD mouse lungs (Fig. 2.1d), which is consistent with expected immune cell infiltration in the lungs after chronic exposure to inhaled irritants especially cigarette smoke[17,42-53,73,74]. Next, we deconvoluted each ST spot based on the catalogue of cell types in the reference scRNA-seq data, generating a comprehensive cell-type spatial map of mouse lungs (Fig. 2.1c and e).

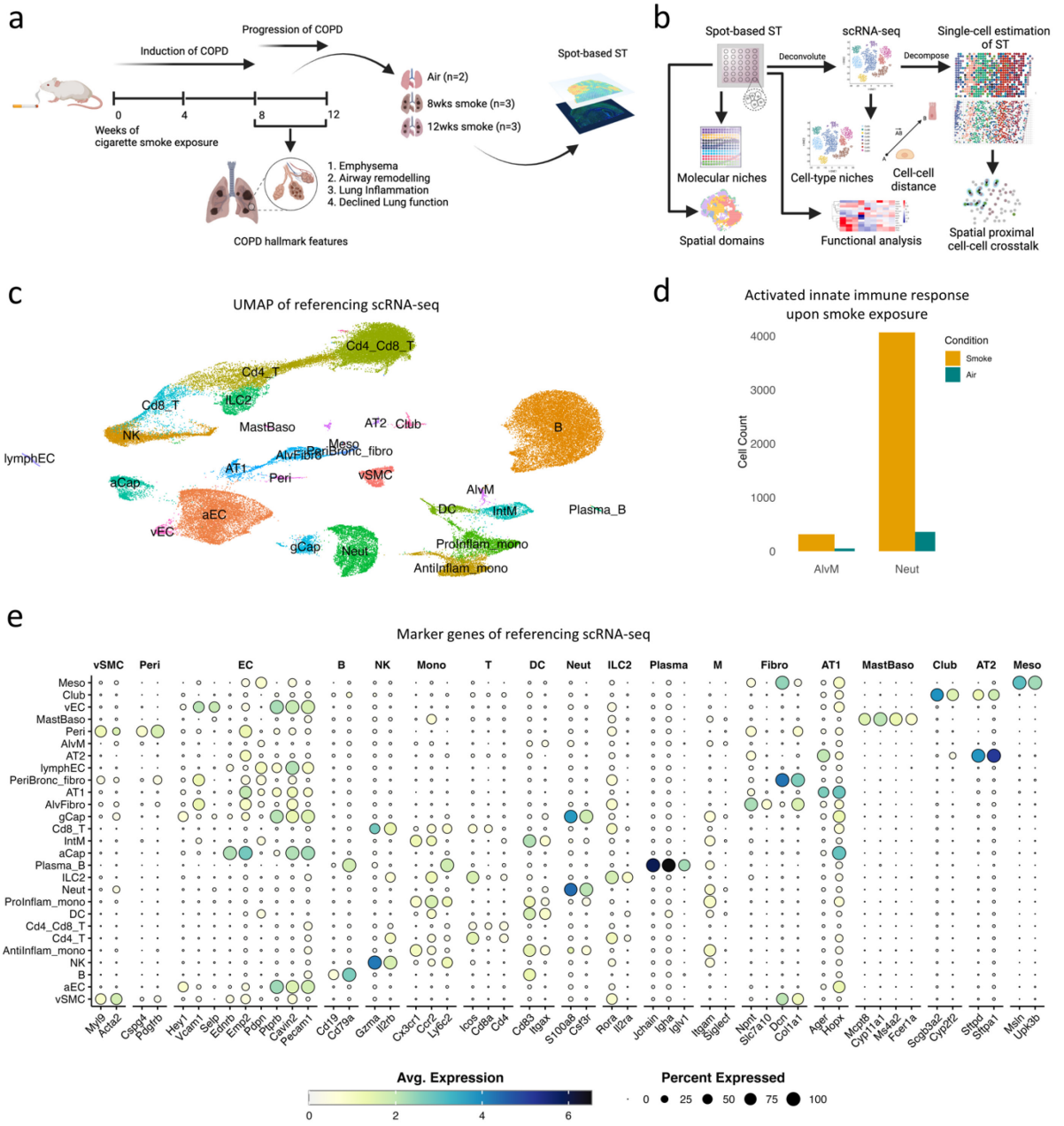


Fig. 2.1: Spatial transcriptomic analysis of murine lungs with and without cigarette smoke-induced experimental COPD.

a, Schematic of experimental design. COPD, chronic obstructive pulmonary disease; ST, spatial transcriptomics. **b**, Schematic of ST data analysis pipeline. **c**, Uniform manifold approximation and projection (UMAP) of reference scRNA-seq data from the same experiment (cells n=69,592). aCap, alveolar capillary endothelial cells; aEC, arterial endothelial cell; AlvFibro, alveolar fibroblast; AlvM, alveolar macrophage; AntiInflam_mono, anti-inflammatory monocyte; aSMC, airway smooth muscle cell; AT1, pulmonary alveolar cell type 1; AT2, pulmonary alveolar cell type 2; B, B cell; Cd4_Cd8_T, Cd4-Cd8 double-positive T cell; DC,

dendritic cell; gCap, general capillary endothelial cell; ILC2, group 2 innate lymphoid cell; IntM, interstitial macrophage; lymphEC, lymphatic endothelial cell; MastBaso, mast cell and basophil; Meso, mesothelial cell; Neut, neutrophil; NK, natural killer cell; Peri, pericyte; PeriBronc_fibro, peribronchial fibroblast; ProInflam_mono, pro-inflammatory monocyte; vEC, venular endothelial cell; vSMC, vascular smooth muscle cell. **d**, Average number of alveolar macrophages and neutrophils per mouse lung in COPD and control groups. **e**, Marker genes for individual cell types in scRNA-seq data.

Spatial organisation of biological process in the lungs in experimental COPD

The localisation of structural cells were spatially restricted to lung anatomical structures, including club cells (markers: *Scgb3a2*, *Cyp2f2*; location: airway lining), alveolar cell type-2 (AT2; markers: *Sftpd*, *Sftpa1*; location: parenchyma), peribronchial fibroblasts (markers: *Dcn*, *Col1a1*; location: airway satellite) and mesothelial cells (markers: *Msln*, *Upk3b*; location: lung edge) (Fig. 2.2a). Interstitial macrophages and dendritic cells were mainly located in the parenchymal region, and AlvMs were restricted to distal parenchyma, particularly in COPD (Fig. 2.2b, left). These immune cells are the major proportion of infiltrated immune cells in human COPD lungs[75,76]. We observed higher proportions of inflammatory immune cells, including AlvMs, interstitial macrophages, neutrophils and dendritic cells, in the parenchyma of COPD compared to control lungs (Fig. 2.2b, right), which is consistent with the recruitment of those inflammatory cells into the lungs of human COPD. Thus, we confirmed the localisation of structural cell types in normal lung structures and aggregated inflammatory immune cells mainly in the parenchyma associated with COPD disease features (e.g. emphysema).

The spatial organisation of cell types mirrored spatially resolved functions in the lungs. To investigate the localisation of common biological processes, we evaluated the activities of 11 well-established signalling pathways using PROGENy[65,66] in each ST spot. Visualisation of the pathway activity in each lung tissue enabled us to link information on cellular abundance and tissue structure to molecular pathways and function in the spatial context of the lung. Notably, we detected over-represented PI3K signalling activity mainly in the regions of airways, indicative of heightened cell metabolism, proliferation and/or *mTOR* activity in the airways (Fig. 2.2c top left), likely associated with fundamental airway functions such as mucus

production, and increased susceptibility to infection and exacerbations[44]. Conversely, signalling activities for *TNF*, *TGFß*, and *NFκB* were lower in the airways but elevated in the parenchyma (Fig. 2.2c), suggesting that the parenchyma has higher levels of pro-inflammatory signals than the airways. These tissue-wide localisations of signalling pathway activities across the mouse lungs were relatively conserved across experimental COPD and control lungs.

To interrogate the relationship between spatially resolved gene expression and COPD-relevant pathology, we applied two gene sets representing the activities of (1) smoking-associated and (2) COPD-associated signals[68]. Upon mapping human smoking-associated signals onto our mouse lung ST dataset, we observed that the molecular changes induced by smoking largely accumulated proximal to the airways (Fig. 2.2d, top left). Conversely, mapping of the human COPD-associated signal revealed its over-representation in the periphery of COPD mouse lungs, exhibiting a diffused pattern (Fig. 2.2d, top right). This implies that the pathogenic signals occur beyond the small airways, and that smoking signals could signify early disease. To validate and translate these findings, these two signals were mapped onto a publicly available healthy human lung Visium ST dataset[32]. Consistent results emerged, with the smoking-associated signal predominantly restricted to the airway lining rather than deeper lung structures (Fig. 2.2d, bottom left); while the COPD-associated signal penetrated through the airway lining to reach deeper lung tissue (Fig. 2.2d, bottom right). Collectively, our findings demonstrate a deeper penetration of pathogenic signals into the lungs resulting from cigarette smoke-induced COPD pathogenesis compared to smoking alone, likely reflecting progressive changes in pathogenic signalling from the airways to the parenchyma during the progression from cigarette smoking to COPD (Fig. 2.2e).

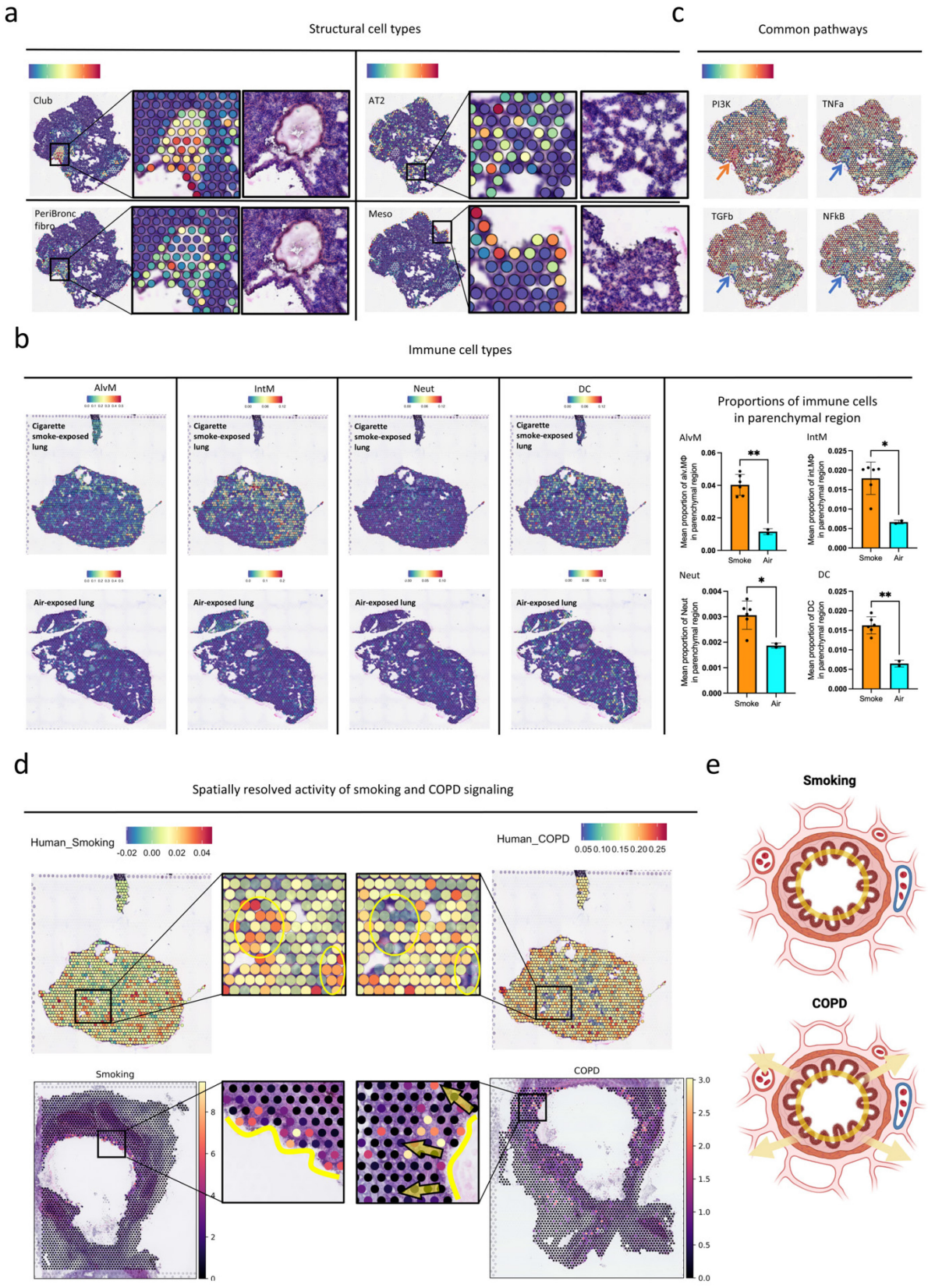


Fig. 2.2: Spatial profiling of cell types and biological processes in murine lungs with and without cigarette smoke-induced experimental and human COPD.

Characterisation of spatial transcriptomic data using cell-type deconvolution for **a**, structural cells and (**b, left**), myeloid cells, and **c**, common pathway activity for mouse lung samples; blue arrow indicates regions with lower signalling score of a pathway; red arrow indicates regions with higher signalling score of a pathway. **b, right**, Mean proportions of myeloid cells in parenchyma of COPD and control groups (adj. p-value using unpaired t-test is shown); **d**, Spatial mapping of human smoking (left) and human COPD (right) gene sets on mouse lungs (top) and human lungs (bottom); **e,** Schematics of the distinct patterns of spatial localisation of smoking and COPD effects in the lungs. In each panel box or circle indicates the focused region of the tissue.

Structural variations of experimental COPD lungs

To directly relate cell localisation to lung structures, we used unsupervised clustering of ST spots based on the similarity of cell-type compositions. This generated 5 clusters, that we referred to as cell-type niches (Fig. 2.3a). These cell-type niches provide a framework of cellular co-localisation with structural building blocks of the mouse lungs, facilitating downstream comparisons of cell-cell distance within lung tissues between COPD and control groups. Among these niches, four (niches 2-5) exhibited enrichment in various immune cells, and one (niche 1) was predominantly composed of structural cells (Fig. 2.3b). Three immune cell-enriched niches (niches 2, 4, 5) were also enriched with parenchyma-located cells (i.e. AT1, AT2 cells, and AlvFibro), indicating co-localisation in the parenchyma with AlvMs, pro-inflammatory monocytes, dendritic cells, and natural killer cells. Likewise, the composition of immune cell-enriched niche 3 suggests that anti-inflammatory monocytes, *Cd4*- and *Cd8*-high T cells, and type-2 innate lymphoid cells co-localised with general capillaries.

We next investigated whether the abundance of cell types within a spot could be predicted by cell types in neighbouring spots, reflecting the distance between target cell types and predictor (neighbouring) cell types at a wider-tissue level rather than in focused areas. This spatial analysis generated an overview of broad tissue landscape based on cell types, which enabled us to compare the lung architecture at cellular resolution between COPD and control lungs. To cover neighbourhood patterns in different ranges, we evaluated three different neighbourhood area sizes using MISTy[64]: (1) cell-type abundances within each spot of 55µm

diameter as intra-spot neighbourhood, (2) in an immediate neighbourhood within a diameter of 5 spots that is 275µm across, and (3) in an extended neighbourhood that expanded to a diameter of 15 spots that is 825µm across (Fig. 2.3c). Our analysis revealed that the overall spatial contexts are different between COPD and control lungs (Fig. 2.3d and Supplementary Table 1). For example, the abundances of structural cells were the most predictive cell types in control lungs, whereas immune cells were more predictive in smoke-exposed lungs, indicating that altered lung architecture is caused by immune cell infiltration. In particular, we found that the distribution of AlvMs was best predicted by nearby vascular smooth muscle cells (vSMC) (intra-spot and immediate neighbourhood) in control lungs, reflecting the close proximity of AlvMs to the vasculature in the lungs during homeostasis (Fig. 2.3f, left). In contrast, we observed reduced proximity of AlvMs to vasculature in COPD compared to normal lungs (Fig. 2.3e and 2.3f, right). This reflects an infiltration pattern of AlvMs during the pathogenesis of COPD. We validated this co-localisation of AlvMs and pulmonary vasculature in a publicly available spot-based ST dataset of an asymptomatic current smoker's lung[32], where we observed that alveolar macrophages dispersed from the blood vessels (Fig. 2.3g). Our findings show a transport route of AlvMs from the vasculature to the distal parenchyma in COPD.

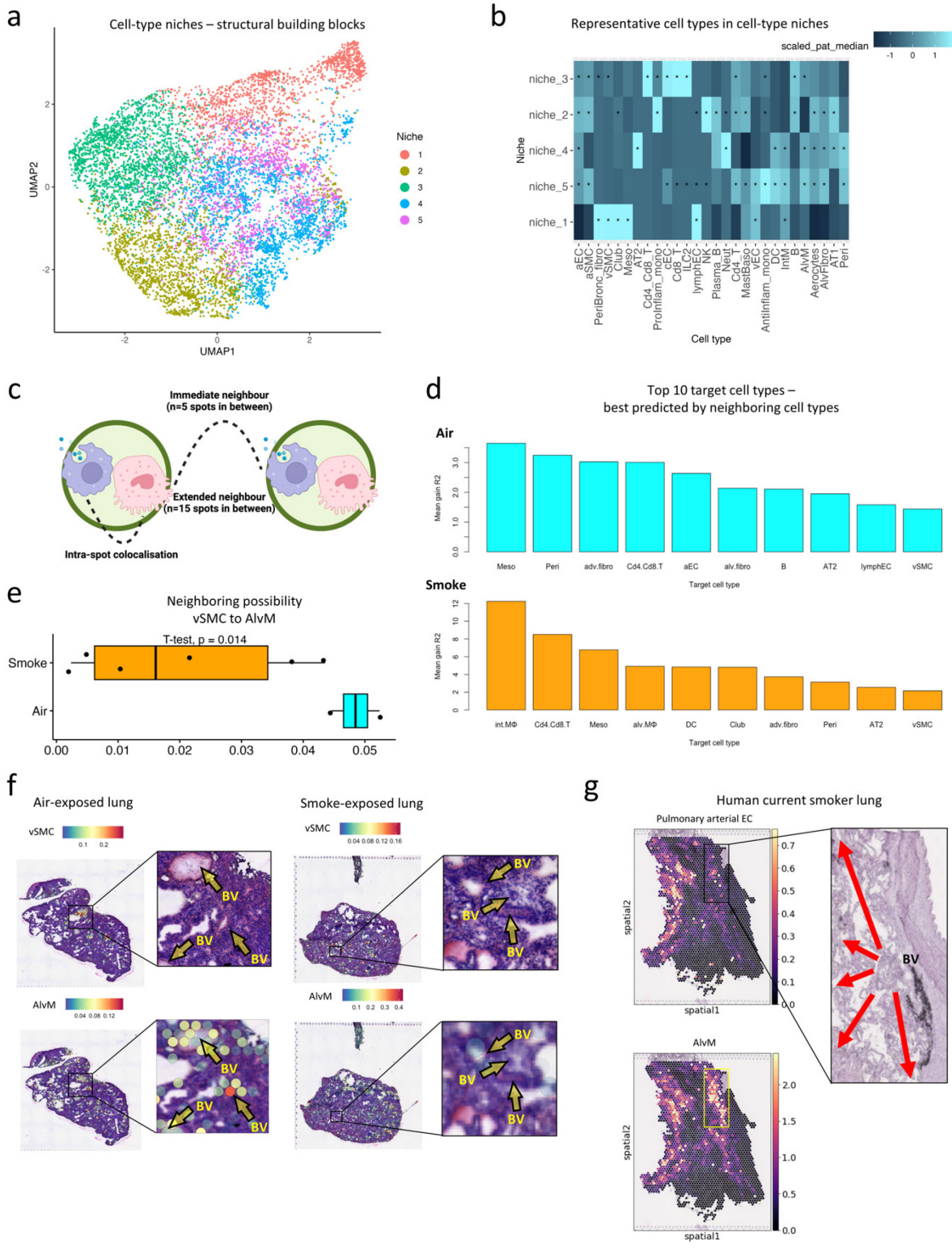


Fig. 2.3: Cell-type compositions in relation to tissue architecture of murine lungs with and without cigarette smoke-induced experimental COPD and human current smokers.

a, UMAP of spatial transcriptomic spots based on cell-type compositions forming cell-type niches in murine lungs. **b**, Scaled median cell-type compositions within individual cell-type

niches. Asterisks indicate more abundant composition of a cell type in a niche compared with other niches (one-sided Wilcoxon rank sum test, adjusted P <0.05). **c**, Schematic of the definition of cell-cell localisation relationship. **d**, Top 10 cell types whose localisations are associated with localisations of other cell types in broader lung tissues of control (top) and COPD (bottom) groups. **e**, Importance of vascular smooth muscle cells (vSMC) abundance (target) in the immediate neighbourhood for prediction of alveolar macrophages (predictor) abundance in COPD and control groups (adj. p-value using unpaired t-test is shown). The center line corresponds to the median, the left and right hinges delineate the first and third quartiles, respectively. COPD group: n=6, control group: n=2. **f**, Broad and focused view of abundances of alveolar macrophages and vSMC in a lung tissue in COPD group (left) and control (right) groups. **g**, Broad and focused view of abundances of alveolar macrophages and pulmonary arterial endothelial cells in a current smoker's lung. Box indicates the region of high abundance of a cell type; red arrows indicate the position of blood vessel; yellow arrows indicate the spread pattern of alveolar macrophages. BV, blood vessel.

Molecular variations of experimental COPD lungs in space

To characterise molecular variations in tissue structures across samples, we employed an unsupervised clustering of gene expressions in each spot, generating a set of molecular niches (Fig. 2.4a). Based on the marker genes in each niche (Fig. 2.4b), we identified molecular niches associated with inflammatory processes (niches 2-4 lung periphery), B and plasma cells (niche 6 equally dispersed in lung tissue), mesothelial cells (niche 7 on the lung edge), club cells (niche 8 in the airway), and vascular smooth muscle cells (niche 10 in the blood vessels) (Fig. 2.4c).

Notably, we observed apparent variations in the expression of marker genes associated with inflammation-featured niches (niches 2-4) between COPD and control groups (Fig. 2.4b). Comparable differentially expressed genes (DEGs) were identified within the inflammation-featured niches, with increased expression of *Lcn2*, *Chil1* and *Ccl6*, as well as decreased expression of *Klf2* in COPD lungs (Fig. 2.4c). A recent study found a strong association of augmented *LCN2* levels in the lungs of smokers with COPD compared to smokers without COPD, probably due to the inflammatory role of *LCN2*[77]. Those authors reanalysed a publicly

available dataset of airway brushings from smokers with and without COPD[78] and observed that the elevated expression of *LCN2* in COPD airway brushings was attenuated by inhaled steroid treatment. Our analysis of molecular features across the lung tissue revealed the elevated expression of *Lcn2* in the lung periphery instead of the airways in smoke-induced experimental COPD, suggesting that the periphery is a site to validate therapeutic targets or biomarkers. The upregulated genes that we identified (Fig. 2.4c) underpin inflammatory processes, and downregulated *Klf2* expression marked impaired barrier function of the pulmonary microvasculature[79], both in the periphery of experimental COPD lungs. These spatially resolved disease-specific DEGs that can only be identified at this level of granularity remain largely unexplored by traditional targeted or untargeted approaches. Thus, here we provide a new molecular map of COPD lungs where inflammatory processes predominantly occur in the lung periphery that may be therapeutic targets.

The molecular niches displayed a segmentation of lung regions in a layer-like gene expression pattern from the periphery (parenchyma-enriched) to core (airway-enriched). To decipher the spatial context of lung tissues, we elucidated the spatial domains in each segment of lung tissue identified by a machine learning approach[67]. The positions of lung domains were in close agreement with our observations of a layer-like pattern of spatial gene expression in the lungs (Fig. 2.4d). We named each domain based on its position in the lungs and its specific marker genes. Interestingly, one domain (the “injury domain”) was uniquely present in COPD lungs, and its location resembled one of the inflammatory niches (molecular niche 4) (Fig. 2.4e and Fig. 2.4c middle right). Apart from the structural cells, AlvMs were the most abundant cell type (~10%) in the injury domain (Fig. 2.4f). A long non-coding RNA, *Gm42418*, was the top gene in the meta gene set of the injury domain and also an upregulated gene of molecular niche 4 in COPD lungs. It is common practice to remove *Gm42418* gene or *Gm42418*-high cell clusters from scRNA-seq datasets[80-82], mainly due to its association with ribosomal RNA contamination. *Gm42418* has been found to be enriched in *NLRP3* inflammasomes which trigger inflammatory cell death and the release of cytokines[83]. Additionally, the chemokine *Ccl6* and the inflammasome activator *Malat1* were highly expressed in the injury domain compared to other regions, suggesting an inflammasome-enriched activity in the injury domain. For the first time, our analysis detected a clear spatial cluster marked by *Gm42418*

associated with lung injury in COPD mouse lungs, providing new biological insights into this transcript in space.

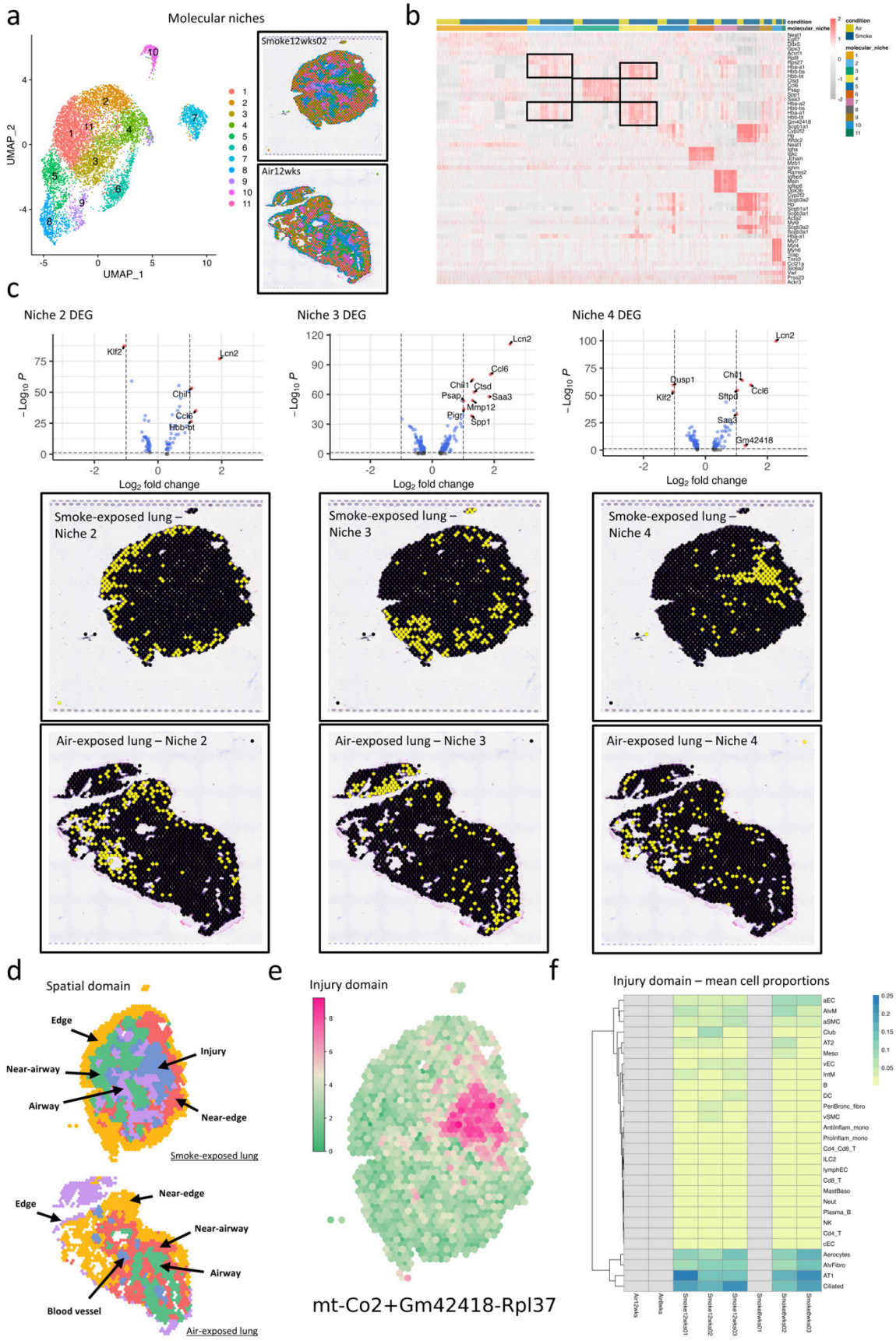


Fig. 2.4: Molecular variations associated with inflammation in murine lungs with and without cigarette smoke-induced experimental COPD.

a, UMAP of spatial transcriptomic spots based on gene expression forming molecular niches in murine lungs (left), and spatial mapping of molecular niches in a COPD and control mouse lung sample (right). **b**, Heatmap of marker genes for individual molecular niches. **c**, Volcano plots of differentially expressed genes (DEGs) of molecular niches 2 (top left), 3 (top middle) and 4 (top right) that are inflammatory genes-enriched in COPD compared to control lung samples, and spatial mapping of molecular niches 2-4 in a COPD (middle) and control (bottom) mouse lung sample. **d**, Spatial domains of mouse lungs identified by a machine learning algorithm in a COPD (top) and control (bottom) mouse lung sample. **e**, Spatial map of the expression of prominent meta genes (represent the expression pattern of a spatial domain) of the injury domain in a COPD mouse lung sample. **f**, Heatmap of cell-type composition in the injury domain in mouse lung samples. Grey indicates the absence or unclear annotation of a domain.

Intercellular communications in COPD lungs

Our study leveraged ST to gain insights into cell-cell communications that regulate lung homeostasis and drive COPD pathogenesis. We analysed cell-cell communication in our scRNA-seq data derived from experimental COPD lungs, focusing on the ligand-receptor-target signalling networks (Fig. 2.5a) between each pair of cell types. We found a pivotal role for general capillary endothelial cells (gCap) in mediating intercellular signalling in the lungs. In experimental COPD, AT2 cells receive more communication signals (Fig. 2.5b), which are indicative of impaired structural integrity, such as those mediated by *LAMIN*, *FGF*, and *JAM* (Fig. 2.5c). This signalling pattern correlates with the observed alveolar destruction (data in another manuscript, available upon request) characteristic of COPD. Furthermore, we documented the altered signalling dynamics between AlvMs (sender) and gCap cells (receiver) in experimental COPD (Fig. 2.5d). Notably, there was an upregulation of signals associated with inflammatory infiltration (e.g. *Spp1*, *Cxcl2*, *Ccl6*, *Ccl9*, *Anxa1*), alongside reduced signals linked to angiogenesis (e.g. *Vegfa*) and macrophage migration inhibitory factor (*Mif*).

As a consequence of the spatial constraints of juxtacrine and paracrine signalling, our ST data is particularly useful in understanding the cellular interplay that occur in the immediate proximity between cells. In addition to analysing cell-cell communications between different

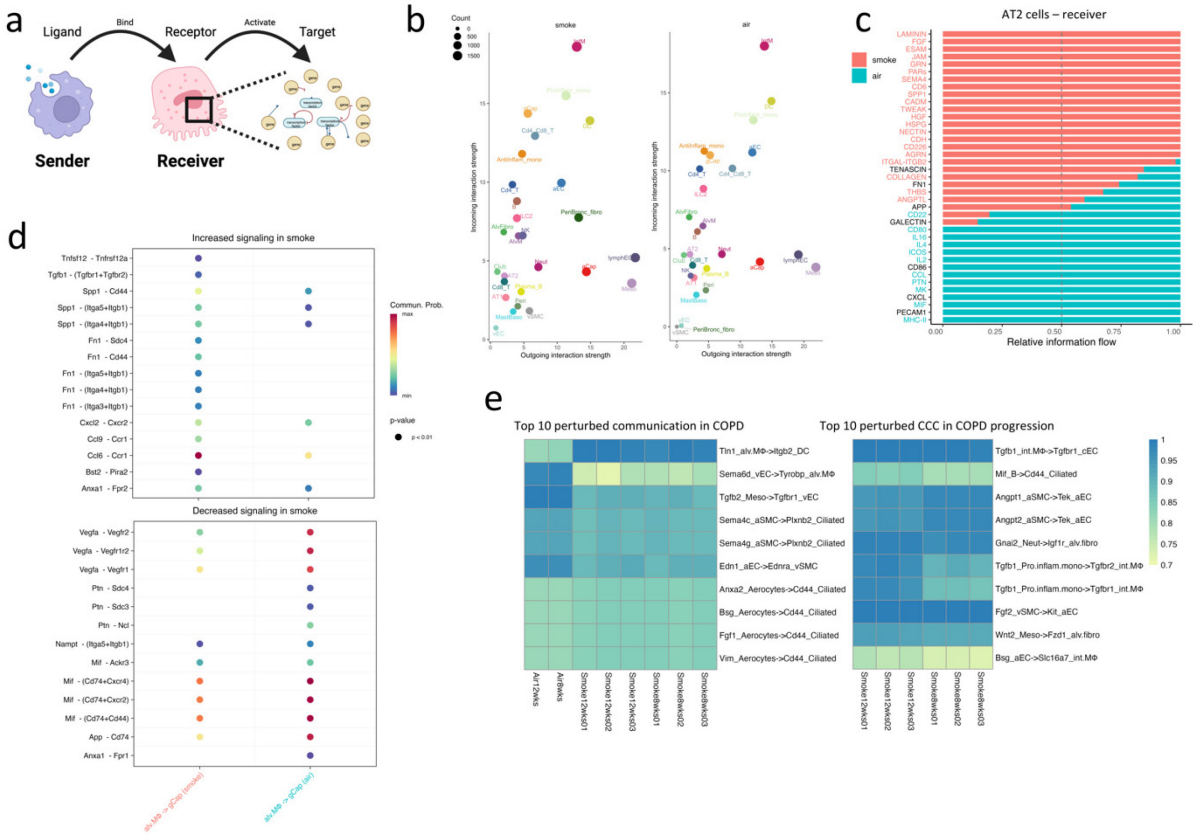
cell types, we also evaluated the communication between individual neighbouring cell pairs, which allowed us to identify spatially distinct patterns of intercellular dynamics in the lungs.

We first investigated the spatially proximal cell-cell communications commonly present in all samples and compared the signalling scores between COPD and control groups (Fig. 2.5e). We focused on the top significant cell-cell communication, *Tln1-Itgb2*, between AlvMs-dendritic cells, that was enhanced in the COPD group. This *Tln1-Itgb2* signalling was spatially restricted to the lung periphery and was more prevalent in the COPD group. The effector function of dendritic cells is crucially linked to their remarkable migratory plasticity, such as the ability to transition from mesenchymal to amoeboid migration mode[84]. Amoeboid migration is well-suited for the rapid movement of leukocytes including dendritic cells in cellular networks[85]. Leukocytes including dendritic cells express various integrins that bind to cell adhesion molecules on endothelial cell surfaces. *Tln1* connects integrins to actin and contributes to the regulation of integrin adhesion complexes[86]. This AlvMs-dendritic cell crosstalk may mediate firm adhesion and spreading of dendritic cells along the endothelium, which leads to dendritic cell crawling and eventually transmigration through the endothelium to sites of inflammation[87].

Next, to evaluate the cellular interplay driving the progression of smoke-induced COPD, we tested the spatially proximal signalling with differential signalling scores generated by (v1.0)[59] between the establishment (8-weeks) and progression (12-weeks) of experimental COPD in the lungs (Fig. 2.5g). We focused on the top significant cell-cell communications, *Tgfb1-Tgfbr1*, between interstitial macrophages-capillary endothelial cells, which was more enhanced in the lung tissues exposed to smoke for 12 compared to 8 weeks (Fig. 2.5g). *Tgfb1* bound to *Tgfbr2* recruits *Tgfbr1* and induces the trans-phosphorylation of the latter, which subsequently phosphorylates *Smad* proteins 2 and 3. These proteins can interact with *Smad4* and regulate the expression of target genes, or with inhibitory Smad proteins like *Smad7*, which functionally inhibit the cascade. This signalling process marks an impairment of the angiogenic capacity of the microvasculature[88]. The elevated score of this cellular interplay in the 12-week smoke-exposed group suggests a greater extent of microvasculature injury during the progression of COPD.

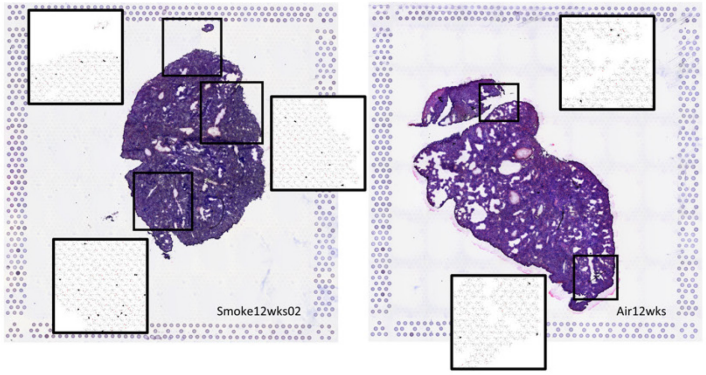
In summary, we developed a new concept of spatial-cellular heterogeneity and intercellular crosstalk between neighbouring cells in the lungs in COPD, where the importance of the lung

periphery in COPD spatial pathogenesis is emphasised. Pulmonary endothelial cells, especially capillary endothelial cells, play a key role in cell-cell communications associated with COPD disease and progression that increases their potential as therapeutic targets.



f

Increased cell-cell communication in smoke compared to air: Tln1 (ligand) from AlvM (sender) to Itgb2 (receptor) in DC (receiver)



g

Increased cell-cell communications in smoke12wks compared to smoke8wks: Tgfb1 (ligand) from IntM (sender) to Tgfbr1 (receptor) in gCap (receiver)

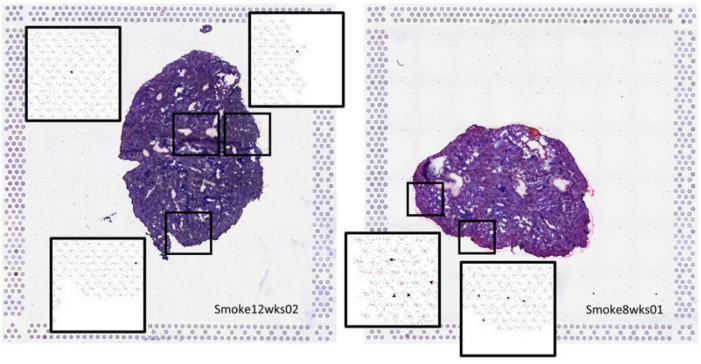


Fig. 2.5: Alterations of cell-cell communications in murine lungs with and without cigarette smoke-induced experimental COPD, and during the progression of experimental COPD disease features.

a, Schematic of the definition of spatially proximal cell-cell communication via a ligand-receptor-target (LRT) signalling network. **b,** The total incoming and outgoing communication probability of each cell type from the scRNA-seq dataset, separated into mouse lungs with and without cigarette smoke-induced experimental COPD. Dot size reflects the number of inferred signals. **c**, All significant signalling pathways were ranked based on their differences of relative information flow within the inferred networks in AT2 cells as the receiver cell type between mouse lungs with and without COPD from the scRNA-seq dataset. The top signalling pathways colored red are more enriched in COPD lungs, those colored black are equally enriched in COPD and control lungs, and those colored green are more enriched in control lungs. **d,** Comparison of the significant ligand-receptor pairs between COPD and control lungs, which contribute to the signalling from alveolar macrophages (AlvM) to general capillary endothelial cells (gCap). Dot color reflects communication probabilities. Empty space means the communication probability is zero. p-values are computed from one-sided permutation test, and only ligand-receptor pairs with p-values <0.01 are shown. **e,** Scores of the top 10 significantly differential spatially proximal cell-cell communications in COPD and control lungs (left), and in mouse lungs exposed to smoke for 8 and 12 weeks (right), from the ST dataset. **f,** Spatial maps of the locations of the top spatially proximal cell-cell communications, including cell-cell communications with significantly different scores in between all COPD and control lung samples, as well as **g**, in between all lungs exposed to cigarette smoke for 8 and 12 weeks. Black-lined box on hematoxylin and eosin-stained tissue indicates the region of focus. Each dot in a cell-cell communication figure represents the simulated position of a single cell.

2.6 Discussion

COPD is a highly heterogeneous respiratory condition often influenced by infections, accompanying comorbidities, and a wide range of environmental risk factors, with cigarette smoking being the foremost among these. The use of experimental COPD animal models in this study allows for the examination of the effects of cigarette smoke, providing a controlled

environment to explore disease mechanisms *in vivo* and reduce the inherent heterogeneity in COPD. Our findings, for the first time, highlight the critical role of the lung periphery in COPD pathogenesis, identifying the lung periphery as the destination of AlvM transport and a site of AlvM accumulation. Indeed, our findings could be combined with screening of patients with early COPD, which already revealed the occurrence of small airway disease defining opportunities to prevent COPD development and progression[89,90]. One promising avenue is the use of targeted molecular imaging, such as ligand- or receptor-specific positron emission tomography tracers, to non-invasively visualise disease-associated cellular or molecular signals in the distal parenchyma, potentially enabling early detection and stratification without the need for invasive lung biopsy.

The interaction between neighbouring cells is crucial for normal function and tissue maintenance in the lungs, which can become aberrant and drive the pathogenesis of diseases. The use of scRNA-seq can identify cellular and molecular changes in diseases, however it lacks spatial context, making it difficult to understand how different cell types interact to regulate tissue functions in space. Little is known about pathogenesis of respiratory diseases in spatial contexts, largely due to tissue homogenisation in many studies. Very recently, the respiratory research community has begun to address this gap by using spatial transcriptomics in human lung specimens, especially for COPD[34,91]. However, the large size of human lungs presents challenges for ST platforms, leaving the spatial profiling unable to cover the entire lung tissue. To address this, we created a detailed map of healthy murine lungs and those affected by cigarette smoke-induced COPD by integrating ST with scRNA-seq data. This enabled us to characterise the spatial arrangement of different cell types and uncover the spatially coordinated changes that occur in experimental COPD validated in human disease. By combining single-cell and spatial data, we gain a more comprehensive understanding of COPD pathogenesis and identify new potential therapeutic targets such as increased *Lcn2* and decreased *Klf2* in the microvasculature in the lung periphery. Promisingly, monoclonal antibodies targeting *Lcn2* proteins have shown efficacy in reducing breast cancer metastasis to the lungs in animal experiment[92]. Indeed, biologic therapies using antibodies have very recently begun entering the market for respiratory diseases. This underscores the value of our spatial data as a rich resource for discovering novel therapeutic targets for COPD.

Our computational analyses enabled an increased resolution of ST by estimating cell-type compositions for each location and also pathway activities. These different layers of biological information enabled us to link the organisation in lung tissue specimens of different histomorphological regions to changes induced by chronic cigarette smoke exposure. Our findings reveal deeper penetration of COPD-associated signals into lung structures compared to smoking-related signals alone. We found that AlvMs near the large vasculature relocated towards the distal parenchyma in cigarette smoke-induced COPD lungs. We also observed inflammation-enriched niches located in the peripheral zone of COPD lungs. *Gm42418* was the best marker for the injury domain associated with AlvMs accumulation and inflammatory cell aggregation in COPD lungs, which was not previously known. Indeed, this was also reflected in our cell-cell communication analysis where we observed increased *Tgfbr-Tgfb* interactions, possibly reflecting direct crosstalk between interstitial macrophages and capillary endothelial cells. For the first time, we used comprehensive spatial gene expression to segment the whole mouse lung tissue into a layer-like pattern and identified the over-represented biological functions in each layer in experimental COPD. Respiratory researchers often rely on easily accessible samples such as airway brushings and sputum to study COPD, which primarily reflect airway pathogenesis, making it challenging to decode the pathogenesis of deep lung tissue. Our identification of the layer-like pattern of gene expression in the lungs provides a novel perspective on the spatial heterogeneity of molecular features and improves the understanding of disease processes in COPD, which facilitates respiratory researchers to obtain insights from deep lung tissues.

Through intercellular communications, we identified that AT2 cells display signals of structural impairment in COPD, mirroring alveolar destruction as a hallmark feature of COPD. The interplay between AlvMs and general capillaries revealed upregulation of inflammatory cytokine signals (including *Anxa1*) and reduced angiogenesis signals, suggesting that general capillaries play a central role in driving COPD pathogenesis. Although targeting the pulmonary microvasculature presents challenges, therapeutics aimed at modulating the disrupted signals between AlvMs and general capillaries may offer benefits. For instance, the mimetic peptide AC2-26 of *Anxa1* has demonstrated positive effects in experimental COPD animal model[93].

By leveraging the spatial information, we characterised the intercellular crosstalk in the lungs, which specifically highlighted the importance of the pulmonary microvasculature associated

with tissue injury and repair processes in peripheral lung region under COPD, centred by impaired angiogenesis of microvasculature. The neighbouring cellular interplay added a spatial heterogeneity on top of cellular and molecular heterogeneity, increasing the understanding of COPD pathogenesis. While some studies have investigated microvasculature impairment in COPD, they predominantly focus on airway regions[94,95], often overlooking deeper lung tissues. In this study, we highlight the significance of microvasculature in the peripheral regions of the lungs. By exploring multiple approaches, we provided a ground to shift the focus of future COPD research toward these critical but frequently neglected peripheral lung regions. However, we acknowledge that these findings are based on mouse lung data, and secondary validation, particularly of the spatially identified pathways, is essential in human lung tissues to confirm their translational relevance. Future studies integrating spatial transcriptomics or proteomics in human COPD samples will be valuable for validating these pathways and further elucidating their role in disease progression.

Our ST dataset serves as a valuable reference for future investigations that combine single-cell genomics with spatial gene expression data in mouse lungs. Our findings also aid in comprehending gene expression patterns and gene regulatory networks in the lungs and represent a useful resource for future research aimed at understanding the spatial-cellular-molecular aspects of pulmonary disease and homeostasis in and beyond COPD.

2.7 Additional Information

Acknowledgements

This work was funded by a fellowship and grants from *National Health and Medical Research Council (NHMRC) of Australia* (GNT1175134) and *UTS* to Prof. Philip Hansbro, the primary supervisor of the thesis author. Dr. Alen Faiz, the co-supervisor of the thesis author, received funds from *Sanofi*. Dr. Caitlin Gillis, the co-supervisor of the thesis author, was supported by an *NHMRC* Early-Career Researcher Fellowship. The author of this thesis is supported by *UTS* President's Scholarship granted by *UTS* and International Research Scholarship granted by Australian government. We acknowledge the support from all animal technicians at the Centre for Inflammation for their contributions to the in vivo experiment: Bob Lu, Caitlin Hubbard, Jacky Lam, Kathleen Bilson, and Solomon Odgers. We also acknowledge the support

from *UTS* eResearch High Performance Compute Facilities for high performance computing services.

Competing interests

All authors report no conflict of interest.

Data availability

Raw data generated by *10x Genomics* Visium spatial transcriptomics platform, mouse lung histological data and physiological data are available upon request and are available (GSE270040; restricted access before published). scRNA-seq data and smoke-effect gene set are only available upon request but will be published imminently. Results of all analyses are in supplementary files and are available in GitHub (https://github.com/HAOLUNG/Visium_COPD_mouse_lungs).

Code availability

All code used for analysis is in supplementary files and are publicly available in GitHub (https://github.com/HAOLUNG/Visium_COPD_mouse_lungs).

Supplementary information

Supplementary Table 1

Cell-to-cell dependencies within each spot and in areas of immediate and extended neighbourhood in each lung tissue sample.

Chapter 3: Spatiotemporal Trajectory of B cells in the Lymphoid Follicles of COPD

3.1 Declaration

Chapter 3 is a thesis-adapted version of a manuscript currently in preparation, in which the author of this thesis is the first author. Chapter 3 also includes a large single-cell atlas of mouse lungs, titled the "Mouse Lung Disease Cell Atlas," associated with another manuscript under review in *Nature Methods*, where the author of this thesis is the second author.

3.2 Abstract

Lymphoid follicles, primarily composed of B cells, are a known pathogenic feature of COPD, but their role remains under-explored. Most studies lack the spatial and longitudinal data necessary to understand their contribution to COPD pathogenesis. Using the *nanostring* GeoMx spatial transcriptomics (ST) platform, this study examines the spatiotemporal formation and progression of lymphoid follicles in a mice exposed to cigarette smoke for different durations (n=8 per timepoint across 2, 4, 6, 8 and 12 weeks) to replicate the onset and progression of human smoking-induced COPD. Early smoke exposure (2 weeks) was sufficient to cause immune cell infiltration, airway inflammation, with emphysema observed at the onset of COPD (8 weeks). B cell levels initially declined in the parenchyma with early smoke exposure. Soon before COPD onset, recruitment of B cells from the circulation through the arteries increased substantially. Instead of distributing evenly, B cells accumulated within these clusters, forming lymphoid follicles that expanded over time due to increased B cell content. Within lymphoid follicles, autoimmune signals (e.g., *Enpp2*) progressively increased, while inflammatory signals (e.g., *Ccl7*) declined over time. We traced follicular B cells to their origins in bone marrow, transitioning through the circulation before homing to lymphoid follicles in the lungs. This study provides novel insights into the spatiotemporal trajectory of B cells in COPD. For the first time, we defined the translocation and accumulation of B cells to form and expand lymphoid follicles along a trajectory of smoke exposure over time, which mirrors the trajectory of the development and progression of human COPD.

3.3 Introduction

COPD is a heterogeneous disease characterised by dysregulated inflammatory and immune responses in the lungs to noxious particles and gases, particularly cigarette smoke. To defend against harmful inhaled exposures in the lower respiratory tract, the airways and lungs rely on innate immune mechanisms, including the epithelial barrier, mucociliary clearance, humoral factors, and innate immune cells such as macrophages, dendritic cells, monocytes, neutrophils, natural killer cells, and mast cells. Dendritic cells bridge innate and adaptive immune responses[96], ultimately leading to the development of lymphoid follicles through activated adaptive immunity in COPD.

Lymphoid follicles are small ectopic tissue masses containing aggregations of inflammatory cells, predominantly B cells, alongside some T cells and dendritic cells. In COPD, B cell-rich lymphoid follicles have been observed in parenchymal regions in close proximity to small airways[89] and within the distal parenchyma[97], particularly in patients with severe and very severe disease. The cellular arrangement of lymphoid follicles facilitates T cell and B cell priming and clonal expansion[98]. The oligoclonal nature of B cells within the follicles suggests antigen-specific induction[97]. However, the specific antigens involved remain unclear, microbial antigens, cigarette smoke-derived antigens, extracellular matrix breakdown products, and autoantigens have been proposed[98]. Therefore, the pathogenic role of follicular B cell responses in COPD is controversial. They may provide a protective benefit by combating microbial colonisation and infection. Conversely, they may be detrimental if the responses are directed against lung tissue antigens, implying an autoimmune component in the pathogenesis of COPD.

Lymphoid follicles, as localised tissue microenvironments, are emendable for the application of ST to study the pathogenesis of COPD at a local niche level. A recent study using ST has identified off-target B cell activation within lymphoid follicles, further supporting a role for autoimmunity in COPD[34]. To advance our current understanding, it is critical to determine the origin of the B cells forming the lymphoid follicles and how these structures change over time. To address this, we exposed mice to cigarette smoke for 2, 4, 6, 8, and 12 weeks, alongside control mice exposed to normal air for 2 and 12 weeks (n=8 per time point), providing a

comprehensive time-course model. ST was performed on the mouse lungs, and scRNA-seq was applied to the systemic samples including bone marrow and blood to profile the extrapulmonary changes. Follicular B cells were traced to their origins in the bone marrow, transitioning through the circulation before localising to lymphoid follicles in the lungs. For the first time, we mapped the spatiotemporal trajectory of B cells forming and expanding lymphoid follicles, mirroring the changes of lymphoid follicles in human COPD over time.

3.4 Methods

Ethics

The Animal Welfare Committee of Sydney Local Health District approved all experimental procedures (protocol no.: 2018-030). This study was conducted in accordance with the Australian responsible use of animal code.

Time-course of cigarette smoke-induced experimental COPD

The induction of experimental COPD via nose-only cigarette smoke inhalation in the mouse model was performed as previously described[17,42-53]. The establishment of this model, similar to that in Chapter 2, included a time-course of smoke exposure. Eight-week-old wild-type female C57BL/6J mice were exposed to cigarette smoke for 2, 4, 6, 8, and 12 weeks, alongside time-matched control mice exposed to normal air (n=8 per time point). Female C57BL/6J mice typically exhibit more stable body weights, lower rates of aggression, and better tolerance to repeated smoke exposure compared to males, which reduces variability and stress-related confounders in long-term inhalation studies.

Airway and lung inflammation, airway remodelling, emphysema and Lung function measures

The approaches to measure baseline structural and cellular changes in respiratory samples from animal model were described previously[17,42-53]. Airway inflammation was determined by differential cell counts in bronchoalveolar lavage (BAL) fluid. Formalin-fixed lung samples were sectioned at 6µm thickness, stained with hematoxylin and eosin, and parenchymal inflammation was assessed by counting inflammatory cells in 10 randomised fields (x100 magnification). Airway remodelling was evaluated by measuring airway epithelium thickening.

Emphysema was quantified using the mean linear intercept method, a standard approach for assessing alveolar diameter in mice. Lung function parameters were evaluated using the flexiVent system with forced oscillation and forced manoeuvre techniques.

Spatial gene expression assay

ST was performed using the *nanoString* GeoMx Digital Spatial Profiler. Formalin-fixed lung sections (6µm) were prepared according to the *nanostring* Manual Slide Preparation guide. Sections were mounted on charged slides (Epredia Colormark Plus, CM-4951WPLUS-001), baked (60°C, 1 hour), and stored at 4°C in vacuum-sealed containers with desiccant for up to 2 weeks. All subsequent steps were conducted under RNase-free conditions with DEPC-treated water. Slides were deparaffinised in 3 sequential 5-minute xylenes washes, followed by ethanol washes (100%, 95%) and PBS rinses. Target retrieval was performed using 1x target retrieval reagent (Invitrogen 00-4956-58 EDTA, pH 9.0, 95°C, 10 minutes). Proteinase K (0.1µg/mL; *Invitrogen*, 25530-049) treatment (37°C, 10 minutes) was followed by post-fixation in 10% neutral-buffered formalin and washes in Tris-Glycine NBF stop buffer. Lung sections were hybridised with the GeoMx Mouse Whole Transcriptome Atlas (v1.0, *nanoString*) probes diluted in Buffer R (37°C, 16–20 hours).

Post-hybridisation stringent washes were performed with formamide/SSC buffer at 37°C, followed by SSC buffer rinses. Tissue morphology was visualised using fluorescently labelled antibodies for Sytox13 - FITC (525nm) for nucleic acid stain, pan-cytokeratin - Cy3 (568nm) for epithelial cell stain, and CD45-Texas Red (615nm) for immune cell staining. Imaging on the GeoMx Digital Spatial Profiler determined the regions of interest (ROIs) representing airways, arteries, proximal parenchyma (ParencNear), distal parenchyma (ParencFar), and lymphoid follicles, where distal and proximal parenchyma were defined based on the proximity to airways. The selection of ROIs was blinded to information about the animal experiment, with a separate individual maintaining records in annotation files to ensure the association between lung tissues and their respective mice was preserved. ROIs were selectively illuminated with ultraviolet light, releasing DNA barcodes hybridised to mRNA targets. The raw data are provided upon request, and the codes used in all analyses are publicly available (https://github.com/HAOLUNG/GeoMx_COPD_mouse_lungs).

NGS library preparation and sequencing for spatial transcriptomics

The barcodes from each ROI were deposited in a collection plate, and polymerase chain reaction (PCR) was used to uniquely index each ROI's barcodes with specific Illumina i5/i7 dual indexing primers. Amplified, indexed libraries were then pooled, purified with SPRI beads (*Beckman Coulter*), and their quality was assessed with capillary electrophoresis using the Fragment Analyser (*Agilent*). Sequencing was performed on *Illumina* NovaSeq6000. Base calling was done by *Illumina* RTA3, and output was demultiplexed and converted to fastq format with bcl2fastq (v1.9.0). The fastq files were then processed through GeoMx NGS Pipeline (v2.0.21) to generate digital count conversion (DCC) files for subsequent analysis.

Analysis of spatial transcriptomics datasets

Raw data files, including DCC files, annotation files, and *nanoString* panel definition files, were imported to construct the GeoMxSet object through GeomxTools (v3.5.0). QC was performed comprehensively to evaluate data reliability and segment performance. Key metrics, including minimum segment reads, percent trimmed, percent stitched, percent aligned, and negative probe geometric means, were assessed, and segments that failed to meet predefined thresholds were excluded. Normalisation of the raw data was carried out using quantile normalisation, scaling the counts to the 75^{th} percentile within each segment, as well as background correction with negative control probes. Following normalisation, UMAP was applied to visualise global transcriptomic patterns and identify clustering of segments according to the ROIs. High-coefficient-of-variation genes were further investigated using hierarchical clustering and heatmaps to delineate their expression across ROIs and exposure condition with time-points. To identify marker genes for each ROI, differentially expressed genes were (DEGs) identified using a linear mixed-effects model, which incorporated both fixed and random effects to account for inter-slide variability while assessing gene expression differences between experimental groups.

Cell-type deconvolution of spatial transcriptomic data

We used *nanoString*'s cell-type deconvolution algorithm, SpatialDecon[99], to estimate cell-type proportions in our spatial transcriptomic data. SpatialDecon's log-normal regression method was applied, using the deconvolution matrix of our Mouse Lung Disease Cell Atlas (manuscript under review) as the reference.

Estimated proportions of cells in ROIs

The relative abundance of B cells in each ROI was quantified and analysed using the spatially deconvoluted data. The cell type proportion matrix was processed to extract the composition of cell types across all ROIs. Deconvolution results were normalised by column sums to calculate cell type proportions for each ROI. We then stratified the processed data by ROI, time-point, and cell type. For statistical analysis, the estimated proportions of cell types were then compared across each time-point separating by exposure condition using one-way ANOVA with Tukey's multiple comparisons test.

Time-series differentially expressed genes

Time-series DEG analysis was performed to investigate transcriptomic changes in each ROI with air- or smoke-exposure across time-points through DESeq2 (v1.42.1). Raw expression counts were firstly subsetted for each ROI. Metadata were adjusted to include experimental conditions (air- vs. smoke-exposure), time-points (e.g., 2, 12 weeks), and their combinations, ensuring proper stratification for downstream analyses. DEG analyses were conducted separately for each ROI. For lymphoid follicles that were only present under smoke conditions, a time-course design was applied to evaluate gene expression trends over the time series. For the other ROIs, a more complex design was used to model the interaction between exposure condition and timepoint. Likelihood ratio tests were performed to compare the full models to reduced models, assessing the significance of time-course-related changes or interactions. Genes with FDR-adjusted $p<0.05$ were deemed differentially expressed. Clustering analysis was conducted on significant genes in lymphoid follicles to classify them based on their temporal expression trends, identifying two groups of genes that showed gradual upregulation or downregulation over time.

scRNA-seq data of experimental COPD lungs

We utilised a scRNA-seq dataset of murine lungs derived from a time-course of smoke exposure. This was from 8-week-old wild-type female C57BL/6J mice exposed to cigarette smoke for 2, 4, 6, 8, and 12 weeks, along with control mice exposed to normal air for 12 weeks (n=8 per time-point). This scRNA-seq dataset was normalised with sctransform using Seurat, and a shared nearest neighbour (SNN) graph was built with the first 30 principal components (PCs) using Seurat's FindNeighbours. Cells were clustered with a Louvain algorithm with FindClusters and annotated at a resolution of 0.8. In addition to lung samples as local

respiratory data, we incorporated two published scRNA-seq datasets of systemic samples, bone marrow and blood, from mice exposed to cigarette smoke or normal air for 12 weeks, generated using the same approach[100].

Clusters of B cells and plasma B cells were extracted from each scRNA-seq dataset. These extracted cells were then integrated, and cells were annotated based on canonical markers for B cell subtypes and plasma B cells.

Integration of B cells extracted from scRNA-seq data of experimental COPD bone marrow, blood, and lungs

scRNA-seq datasets of bone marrow and blood from mice exposed to either cigarette smoke or normal air for 12 weeks were obtained from our previous study[100]. B cells and plasma B cells were isolated from the scRNA-seq datasets of murine bone marrow, blood, and lungs, and subsequently integrated using the Harmony algorithm[69].

Trajectory analysis of B cell scRNA-seq data

Integrated scRNA-seq of B cells was analysed to investigate the differentiation trajectory of B cells using Monocle3[101]. Clustering was performed with a resolution of 0.001. A trajectory graph was constructed without partitioning the dataset, and cells were ordered along pseudotime, with the root defined by S-phase B cells. To further explore differentiation pathways, a branch associated with follicular B cells was identified by subsetting cells from clusters of interest. Pseudotime progression and associated gene expression patterns were analysed within this branch. Significant genes driving the trajectory were identified using graph-based statistical testing, and genes with an FDR<0.05 were considered. Among these, key TFs were identified by cross-referencing with a curated list of mouse TFs from AnimalTFDB (v4.0)[102].

3.5 Results

Baseline time-course pathogenic changes in cigarette smoke-induced COPD

COPD is characterised by chronic, unresolved inflammation and impaired wound repair in small airways and alveolar tissues[1]. It involves intricate intercellular interactions that drive the

disease, spanning from abnormal intracellular and molecular signalling to macroscopic alveolar destruction. Cigarette smoking remains the primary risk factor for COPD, but the transition from asymptomatic smoking to COPD, as well as the progression to more severe stages, is not fully understood. This study provides a spatiotemporal trajectory of changes from asymptomatic smoking to COPD and its progression due to continuous smoking, offering valuable insights for therapeutic target discovery in respiratory medicine.

Here, the temporal changes induced by smoking were captured by exposing mice to cigarette smoke for 2, 4, 6, 8, and 12 weeks, and spatial information was obtained using *nanoString* GeoMx ST to analyse the whole transcriptome of key lung anatomical sites, including airways, arteries, lymphoid follicles, and distal and proximal parenchyma (Fig. 3.1a). Our animal model, previously reported to replicate key features of human COPD[17,42-53], demonstrated hallmark pathological changes. Airway epithelial thickening, collagen deposition in the airways, and emphysema were evident in the lungs of mice after experimental COPD is established (Fig. 3.1b). We observed significant immune cell infiltration in the lungs at all timepoints of smoke exposure, along with compromised lung function from the earliest stage, while emphysema became apparent only after 8 weeks of exposure, confirming the onset of experimental COPD at this timepoint (Fig. 3.1c). The trajectory of intrapulmonary changes aligns with our previous study, demonstrating that smoke-induced lung damage drives the onset of COPD after 8 weeks of exposure, followed by progressive and non-resolving pathology at 12 weeks, replicating severe, progressive COPD in humans[42].

Spatial gene expression analysis revealed clear separation between the selected anatomical sites, with no segregation based on time-points (Fig. 3.1d, e). The exception was the proximal and distal parenchyma, which shared similar transcriptomes due to both being parenchymal tissues but differing in proximity to the airways. Notably, *Cxcl13* and *Tmtc1* were identified as markers to distinguish proximal and distal parenchyma, respectively, a finding corroborated by reproducible visualisation from 10x Visium ST (Chapter 2, Supplementary Fig. 1). To characterise the cellular composition within each anatomical site, we applied *nanoString*'s SpatialDecon[99] cell-type deconvolution algorithm, using the Mouse Lung Disease Cell Atlas as the reference scRNA-seq dataset. We accurately mapped cell types to their corresponding regions, aligning with the defined anatomical site characteristics (Fig. 3.1f). Similarly, marker genes for airways (*Scgb3a2*), arteries (*Acta2*), lymphoid follicles (*Cxcl13*), and parenchyma

(*Ager*) corresponded appropriately to the correct cell types and functions within their respective sites (Fig. 3.1g). For example, B cells were the predominant component of lymphoid follicles, supported by secondary populations of T cells and dendritic cells. The relatively high presence of AT1 and AT2 cells confirmed the localisation of lymphoid follicles within the parenchymal region, mirroring their predominant presence in the parenchyma of human COPD lungs[34]. Altogether, the spatial gene expressions and cell compositions in key lung anatomical regions of our animal model closely resemble those observed in humans, providing a robust baseline for studying human COPD pathogenesis.

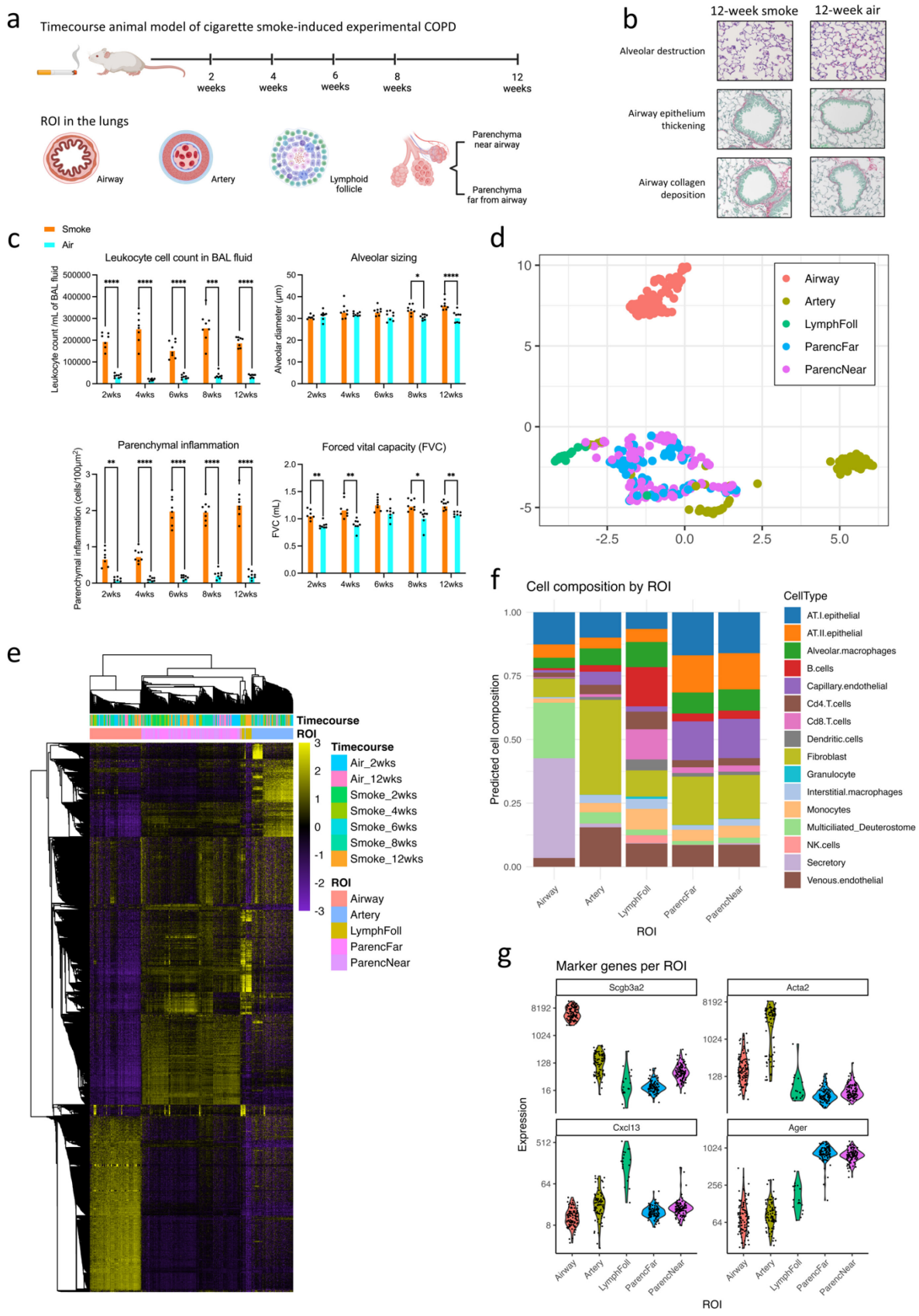


Fig. 3.1: Baseline pathogenic changes of the development and progression of cigarette smoke-induced experimental COPD.

a, Schematic of experimental design. COPD, chronic obstructive pulmonary disease; ROI, region of interest. **b**, Representative histological images showing alveolar size, airway epithelial thickness, and airway collagen in mouse lungs exposed to 12 weeks of cigarette smoke or normal air. **c**, Quantitative analysis of leukocyte counts in BAL fluid, inflammatory cells in the parenchyma, alveolar diameters, and lung function (FVC) in smoke- and air-exposed mice across time-points, 2, 4, 6, 8, and 12 weeks (n=8 per timepoint). Statistical analysis was performed using two-way ANOVA with Sidak's multiple comparisons test. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. BAL, bronchoalveolar lavage; FVC, forced vital capacity. **d**, UMAP representation of selected ROIs of key lung anatomical sites, including airways, arteries, lymphoid follicles (LymphFoll), proximal (ParencNear) and distal parenchyma (ParencFar). **e**, Heatmap of all detected transcriptomes in ST. **f**, Cell composition of selected ROIs of key lung anatomical sites, with Mouse Lung Disease Cell Atlas as the reference scRNA-seq data for cell deconvolution in ST data. ATI, pulmonary alveolar cell type 1; ATII, pulmonary alveolar cell type; NK, natural killer. **g**, Representative marker genes of airways, arteries, parenchyma, and lymphoid follicles (LymphFoll).

Spatiotemporal changes of B cell levels in arteries and parenchyma leading to lymphoid follicle expansion during cigarette smoke exposure

Lymphoid follicles, small B cell-rich tissue aggregates, are predominantly observed in the lung parenchyma of patients with severe-to-very-severe COPD[89,97], and also in those with mild disease[34]. The presence of lymphoid follicles appears to be specific to smoking, as several studies report little to no lymphoid follicles in individuals without a history of smoking, regardless of COPD diagnosis[34,97,103]. Despite this strong association with smoking, it remains unclear when lymphoid follicles begin to develop during smoke exposure. Our mouse model faithfully replicated the morphology of lymphoid follicles observed in human lungs (Fig. 3.2a), providing a valuable platform to study their formation and progression during smoking, which is challenging to explore in humans.

Lymphoid follicles of observable size were first detected at 4 weeks of cigarette smoke exposure, defined as the early stage of smoke exposure and representing the induction phase of COPD features[42]. Although lymphoid follicle formation is traditionally considered a late

event in COPD progression, the use of animal models of COPD suggests that early adaptive immune responses are sufficient to drive lymphoid follicle formation during early COPD. Notably, B cell proportions within lymphoid follicles gradually increased with prolonged smoke exposure (4 vs. 12 weeks, $p<0.05$; Fig. 3.2b), facilitating the expansion of tissue masses and a local adaptive immune niche.

We also analysed B cell levels across other lung regions to address the broader spatial distribution of B cells. In arteries, B cell levels began to rise at 4 weeks of smoke exposure (Fig. 3.2c), coinciding with the first appearance of observable lymphoid follicles. This surge in B cell levels at 4 weeks suggests a dramatic recruitment of B cells from circulation into the lungs at this time-point, likely driven by the presence of lymphoid follicles. Before the 4-week time-point, upon the onset of smoke exposure, B cell levels in the parenchyma declined and did not return to baseline during continued exposure (Fig. 3.2d). This indicates that alveoli are not the primary sites where recruited B cells reside at any time-point during cigarette smoking; rather, the parenchyma loses its B cells.

Collectively, this data shows that B cells trafficked from the parenchyma and were re-directed into smaller tissue aggregates in the beginning of smoke exposure. There they act as precursors for the formation of lymphoid follicles after 4 weeks of exposure, where additional B cells are recruited from circulation to aggregate in the follicles (Fig. 3.2e). For the first time, our study provides a comprehensive map of the central role of B cells in the formation and expansion of lymphoid follicles across multiple lung regions and over time during cigarette smoke exposure and the development and progression of experimental COPD.

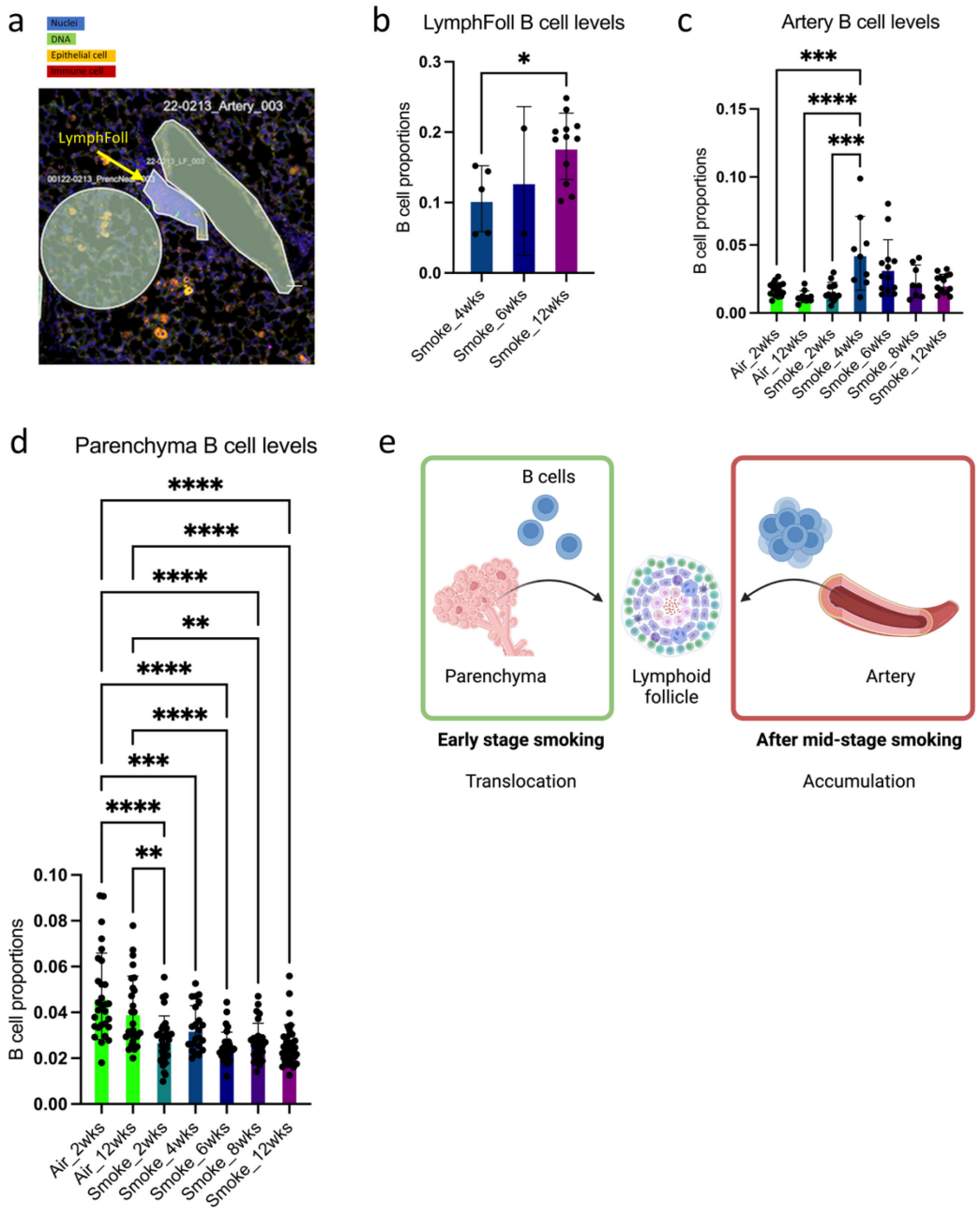


Fig. 3.2: Spatiotemporal changes of B cell levels associated with lymphoid follicles during cigarette smoking.

a, Representative histological image of a lymphoid follicle. DNA is stained with Sytox13 - FITC (525nm), epithelial cells with PanCK - Cy3 (568nm), and immune cells with CD45 - Texas Red (615nm). **b-e**, Quantitative analysis of the proportions of B cells in lymphoid follicles (**b**), arteries (**c**) and parenchyma (**d**) across timepoints from smoke- and normal air-exposed mice. Statistical analysis was performed using one-way ANOVA with Tukey's multiple comparisons

test. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. **e**, Schematic representation of B cell trafficking within the lungs in response to cigarette smoke, based on multi-site changes in B cell levels over time.

Local immune niche of lymphoid follicles primes autoimmunity

The oligoclonality of B cells within lymphoid follicles over time during cigarette smoking shows the presence of local immune responses. To investigate this further, we first compared the morphology of lymphoid follicles from their initial formation to the end-stage of smoke exposure and experimental COPD. Over time, lymphoid follicles became denser with increased follicular cellular contents, enhanced local macrophage trafficking, and reduced attachment to extracellular matrix as a structural base (Fig. 3.3a). To provide molecular evidence supporting these morphological observations, we performed a time-series of DEG analysis to track gene changes in lymphoid follicles with smoke exposure. The gene changes corroborated the morphological changes, revealing a progressive increase in the expression of genes associated with autoimmunity in lymphoid follicles over time during smoking (Fig. 3.3b). *Enpp2*, the most DEG, produces lysophosphatidic acid, which facilitates lymphocyte trafficking to lymphoid follicles, a hallmark of many autoimmune diseases[104]. However, the autoimmune component of lymphoid follicles remains complex due to the concurrent gradual increase in *Mfge8* expression. *Mfge8* is an inhibitor of autoimmunity, facilitating the clearance of apoptotic lymphocytes by macrophages[105,106]. This may explain the observed increase in macrophage trafficking to lymphoid follicles at later time-points of smoke exposure. The rise in *Mfge8* likely represents a compensatory mechanism to counteract the pro-autoimmune signals driven by *Enpp2*. Interestingly, genes associated with type I interferon signalling (e.g., *Ifit1*, *Ifit3*) were downregulated in lymphoid follicles over time (Fig. 3.3c), suggesting a shift from T-cell-driven or antiviral-like immune responses to B-cell-mediated immunity. In addition, the detachment of lymphoid follicles from lung structures such as vascular and alveolar walls over time was reflected in the decreased expression of *Col1a1*, which encodes collagen. This detachment may result from alveolar destruction in later time-points of smoke exposure, causing lymphoid follicles to become standalone tissue masses.

Further investigation identified two distinct gene clusters within the lymphoid follicles: one with a gradual increase in expression (*LymphFoll_up*) and the other with a gradual decrease (*LymphFoll_down*). These gene signatures were mapped onto scRNA-seq data from lungs of mice exposed to cigarette smoke over the same time-course. The majority of upregulated genes in lymphoid follicles during smoke exposure were expressed by B cells, followed by dendritic cells, monocytes, and macrophages (Fig. 3.3d), indicating progressive local autoimmune priming. Conversely, the majority of downregulated genes were expressed by stromal cells, supporting the observed detachment of lymphoid follicles from their structural base over time.

Thus, for the first time, we characterised the lymphoid follicle as a local adaptive immune microenvironment in COPD, revealing its role in priming autoimmunity during continuous smoking signified by gradual increase of *Enpp2*.

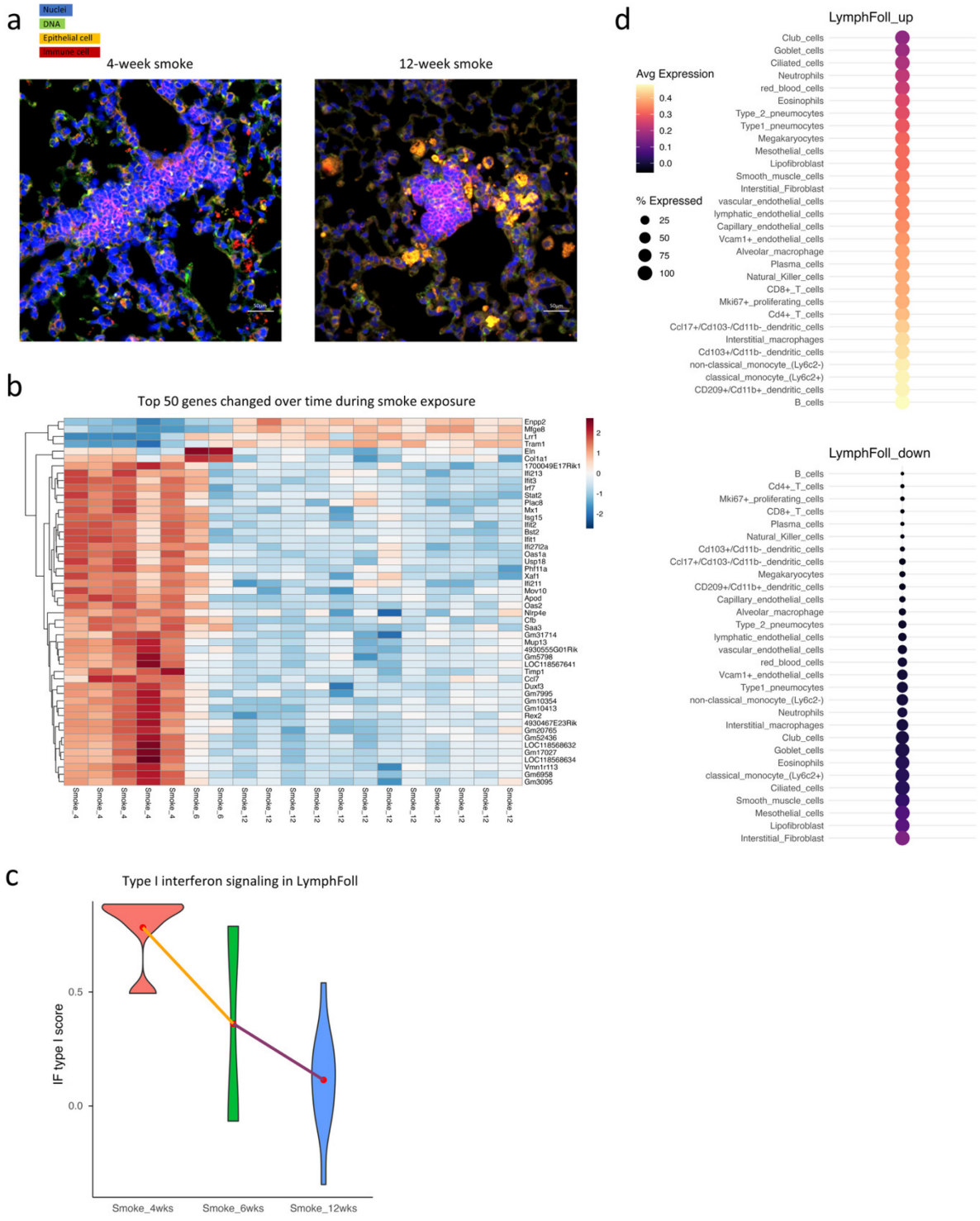


Fig. 3.3: Local immune microenvironment of lymphoid follicles induced by cigarette smoke exposure.

a, Representative histological image of a lymphoid follicle. DNA is stained with Sytox13-FITC (525nm), epithelial cells with PanCK-Cy3 (568nm), and immune cells with CD45-Texas Red

(615nm). **b**, Expression profiles of the top 50 genes altered across time-points of cigarette smoke exposure in lymphoid follicles. Differentially expressed genes (DEGs) were identified through a time-series analysis of each ROI at all time-points. Likelihood ratio tests were used to compare full models against reduced models, with DEGs determined by a false discovery rate (FDR)-adjusted $p<0.05$. **c**, Mapping of the type I interferon gene set onto the lymphoid follicles, across time-points of cigarette smoke exposure. **d**, Contribution of cell types to the expression of genes showing increases (LymphFoll_up) or decreases (LymphFoll_down) in lymphoid follicles over time. LymphFoll_up and LymphFoll_down clusters were derived from the time-series DEG analysis in lymphoid follicles.

Systemic-to-lung developmental trajectory of follicular B cells in experimental COPD

The development and migration of B cells follow a tightly regulated process, beginning in the bone marrow, passing through circulation, and ultimately reaching target organs. Using an animal model across different experimental batches, we analysed B cells extracted from scRNA-seq datasets of bone marrow[100], blood[100] and lungs of mice exposed to cigarette smoke across the time-course. This approach enabled us to create a multi-organ integration of B cells affected by cigarette smoking (Fig. 3.4a).

Tnfrsf17, a key marker of lymphoid follicles was identified in our ST data and is known as a marker of immunoglobulin-secreting plasma B cells, was mapped onto the multi-organ B cell dataset. We observed dominant *Tnfrsf17* expression in a sub-cluster within the plasma B cell sub-population (Fig. 3.4b), which was predominantly present in the lungs (Fig. 3.4c), identifying a follicular B cell component within lymphoid follicles. Indeed, the hallmark of lymphoid follicles is the intra-follicle plasma cells producing high titres of antibodies. We then extracted the top 10 marker genes of follicular B cells and mapped them onto the entire plasma B cell population in the lungs. We observed a general gradual upregulation of the follicular B cell signature with prolonged smoke exposure (Fig. 3.4d). Interestingly, the follicular B signature in lung plasma B cells declined at 8 weeks of smoke exposure in the scRNA-seq dataset, mirroring the absence of lymphoid follicles in the GeoMx ST dataset at the same time-point. This alignment between datasets, despite being generated from separate experiments, indicates that temporal changes of lymphoid follicles are not linear. This

exception at 8 weeks of smoke exposure coincided with the onset of the emergence of chronic features of COPD such as emphysema. Here, structural changes in the lungs, such as alveolar destruction, may have disrupted the aggregation of follicular B cells, as alveoli provide a structural base for lymphoid follicles.

Integrating B cells from bone marrow, blood, and lungs allowed us to construct a systemic-to-lung trajectory of B cell development. Pre-B cells in the bone marrow were identified as the root of this trajectory, which progressed through maturation stages in the blood and culminated in follicular B cells in the lungs, marking the terminal stage of a developmental branch (Fig. 3.4e). Next, we investigated the TF genes driving their development from immature B cells, predominantly present in the blood, to follicular B cells, that dominated B cell populations in the lungs. Notable TFs included *Foxp1*, *Hmgb1*, *Xbp1*, and *Mxd3* (Fig. 3.4f). *Foxp1* plays a critical role in the early stages of B cell development[107], but is downregulated as B cells differentiate into plasma B cells[108]. Using primary human B cells differentiating into plasma cells, van Keimpema et al.,[109] demonstrated that *Foxp1* repression is essential for further differentiation beyond the plasma B cell stage. Additionally, trajectory analysis confirmed known driver TFs of B cell maturation, such as *Xbp1*, which is essential for differentiation into plasma B cells[110]. *Hmgb1* and *Mxd3* emerged as newly identified regulators of the development of follicular B cells in association with lymphoid follicles. Both *Hmgb1* and *Mxd3* reached their highest expression at the terminal point of the developmental trajectory, predominantly located in the lungs (Fig. 3.4g). *Hmgb1* is released into the extracellular milieu by plasma B cells during maturation[111], where it exerts pro-inflammatory effects, promoting autoimmune disease and chronic inflammation. *Mxd3* confers anti-apoptotic effects on B cells[112], reflecting resistance to apoptosis of follicular B cells within lymphoid follicles at the terminal stages of B cell development.

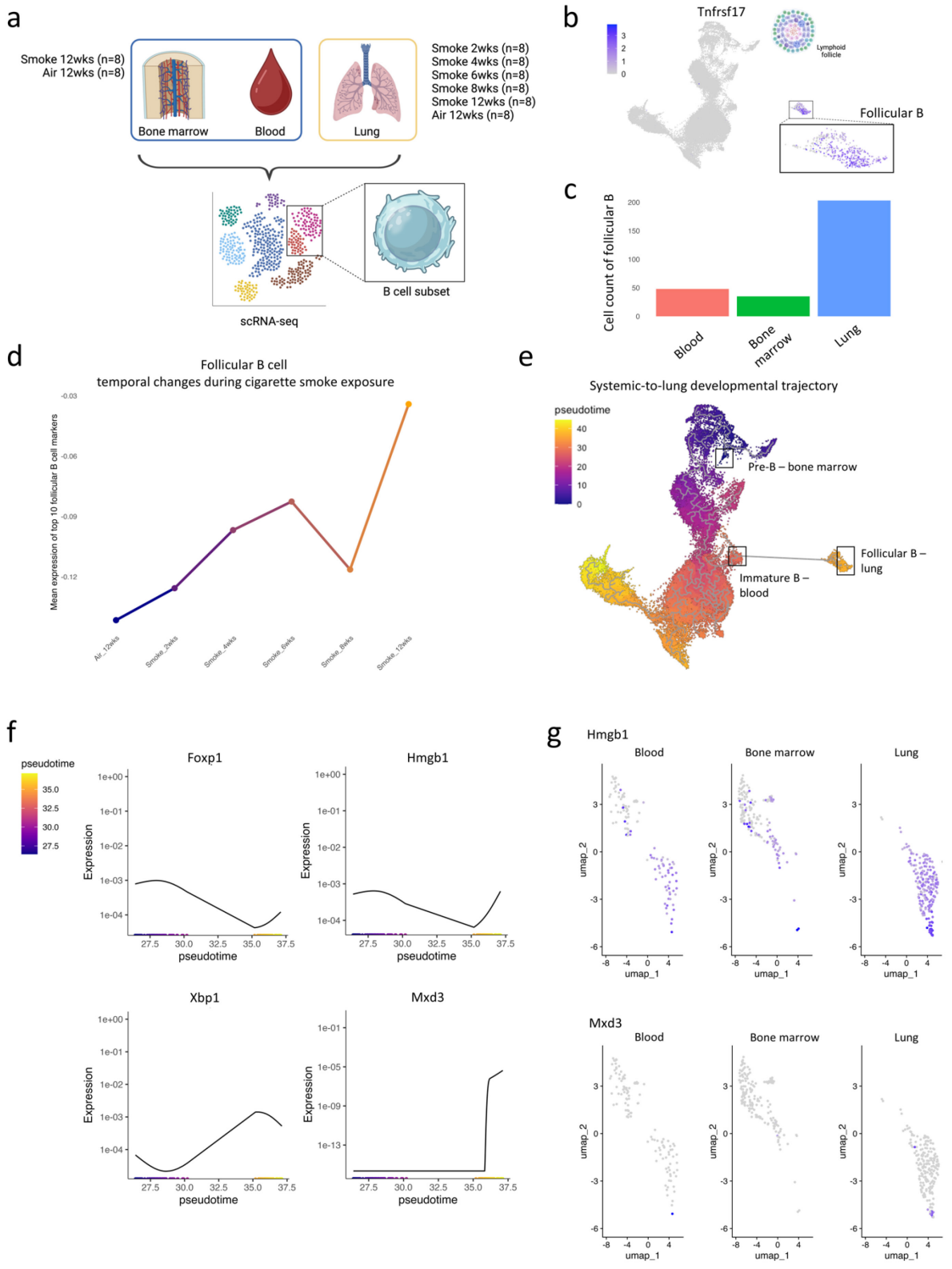


Fig. 3.4: Systemic-to-lung fate of follicular B cells during cigarette smoke exposure and the development and progression of experimental COPD.

a, Schematic illustration of the integrated B cell dataset extracted from scRNA-seq of bone marrow, blood, and lungs of mice exposed to cigarette smoke for 2-12 weeks. **b**, Expression of *Tnfrsf17*, a marker gene of lymphoid follicles identified in the GeoMx ST dataset, mapped in the multi-organ integrated B cell dataset. Zoomed-in regions (top) highlight the plasma B cell population with predominant *Tnfrsf17* expression, and (bottom) indicates the organ origin of the plasma B cell population. **c**, Number of follicular B cells identified across murine blood, bone marrow, and lungs. **d**, Mapping of the top 10 marker genes of follicular B cells, identified from the multi-organ integrated B cell dataset, onto the plasma B cell population in the lungs, along the time-course of cigarette smoke exposure. **e**, Developmental trajectory of B cells spanning bone marrow, blood, and lungs in mice exposed to cigarette smoke over the time-course. Highlighted populations include pre-B cells in the bone marrow, immature B cells in the blood, and follicular B cells in the lungs. **f**, Representative transcription factor genes with altered expression along the developmental trajectory from immature B cells in the blood to follicular B cells in the lungs. **g**, Mapping of representative transcription factor genes onto the plasma B cell population across blood, bone marrow, and lungs.

3.6 Discussion

COPD is a highly heterogeneous respiratory disease affected by infections, comorbidities, and a variety of environmental risk factors, with cigarette smoking being the most significant risk factor. The interplay between smoking-related factors such as the duration and intensity of smoke exposure, airflow dynamics, and the different cross-sectional area across lung compartments, suggests the presence of a spatiotemporal spectrum during the transition from smoking to clinical COPD and its advance to more severe disease. Although screening for these spatiotemporal changes in the lungs of smokers is critical, it remains challenging in humans due to the difficulty of obtaining whole-lung biopsies from pre-clinical smokers. Our study addresses this gap by providing, for the first time, a comprehensive spatiotemporal profile of gene expression changes within key lung anatomical structures using a mouse model of cigarette smoke-induced experimental COPD.

Initially, we observed increased immune cell influx into the lungs and a decline in lung function from the beginning of smoke exposure in the animal model, which parallels observations in

human smokers without chronic respiratory symptoms[113]. Emphysema emerged at 8 weeks in the animal model, marking the onset of chronic COPD features. This timing is consistent with previous findings from experimental COPD mouse models conducted in different studies[17,42-53]. Together, these baseline observations confirm that the animal model accurately replicates the effects of human smoking and COPD, providing a robust platform to investigate spatiotemporal gene changes in the lungs during smoke exposure.

Immunopathology is a hallmark feature of lung damage in COPD. Innate immune responses have been the primary focus for decades, whereas adaptive immunity has been overlooked but has recently garnered increased attention. The presence of lymphoid follicles in the lungs is a marker of adaptive immune responses in smoking-induced COPD. In our study, lymphoid follicles were first observed in the lungs of mice exposed to cigarette smoke for 4 weeks, prior to the onset of emphysema, highlighting the role of smoking in their formation. Lymphoid follicles were captured using ST for gene profiling. Early in smoke exposure, B cells trafficked from the parenchyma into smaller aggregates, which served as precursors of lymphoid follicles formed after 4 weeks of exposure. Subsequently, B cells were recruited from circulation and accumulated within these lymphoid follicles. Over time, lymphoid follicles became denser, with increased follicular cellular content, enhanced local macrophage trafficking, and reduced attachment to the extracellular matrix.

The role of lymphoid follicles in promoting autoimmunity over time became evident through the gradual upregulation of *Enpp2*, which encodes lysophosphatidic acid. This factor facilitates lymphocyte trafficking to lymphoid follicles, a hallmark of autoimmune diseases[104]. However, the autoimmune nature of lymphoid follicles is complex, as evidenced by the concurrent gradual increase in *Mfge8* expression. *Mfge8* plays a role in inhibiting autoimmunity by aiding macrophages to clear apoptotic lymphocytes[105,106]. This explains the observed increase in macrophage trafficking to lymphoid follicles at later stages of smoke exposure. The rise in *Mfge8* likely reflects a compensatory mechanism to counterbalance pro-autoimmune signals driven by *Enpp2*. Interestingly, type I interferon signalling genes (e.g., *Ifit1*, *Ifit3*) were downregulated in lymphoid follicles over time, suggesting a shift from T-cell-driven or antiviral-like immune responses to B-cell-dominated immunity. Furthermore, the progressive detachment of lymphoid follicles from lung structures such as vascular and alveolar walls was marked by reduced expression of *Col1a1* over time, a collagen-encoding gene. This

detachment likely results from emphysema at later stages of smoke exposure, leading to lymphoid follicles becoming standalone tissue masses in COPD.

To extend our understanding beyond pulmonary changes of B cells, we also incorporated extra-pulmonary B cells from bone marrow and blood in mice exposed to cigarette smoke, providing insights into the systemic origins of follicular B cell in experimental COPD. Here, we constructed a systemic-to-lung developmental trajectory of B cells, starting from pre-B cells in the bone marrow, progressing through maturation in the blood, and culminating in follicular B cells in the lungs. The gradual upregulation of follicular B cell signatures during smoke exposure underscored their progressive role in lymphoid follicle expansion, while a transient decline at 8 weeks coincided with emphysema onset, likely reflecting structural disruptions in the lungs. Key TFs (e.g. *Foxp1*, *Xbp1*) driving B cell maturation were identified, aligning with established roles in plasma B cell differentiation[108,110]. Additionally, *Hmgb1* and *Mxd3* emerged as novel regulators of follicular B cell development, with their highest expression in the lungs occurring during the progression to more severe experimental COPD. *Hmgb1* promotes pro-inflammatory responses, linking follicular B cells to autoimmune and chronic inflammatory processes[111]. *Mxd3* enhances resistance to apoptosis[112], a characteristic of follicular B cells within lymphoid follicles. Collectively, the identification of the driver genes involved in follicular B cell development shed light on the systemic origins and lung-specific dynamics of lymphoid follicles in smoking-induced COPD, presenting potential therapeutic targets to inhibit lymphoid follicle formation.

However, this study is limited by the absence of human lung samples for direct comparison. Future studies could aim to validate these findings in human lungs exposed to cigarette smoke over varying durations to better understand the temporal dynamics of disease progression. Obtaining such samples, however, remains a major challenge due to the invasive nature of distal lung biopsies required for spatial and molecular analyses.

In conclusion, our study offers the first comprehensive spatiotemporal profiling of the lungs during the development of smoking-induced COPD and its progression to severe disease. By integrating pulmonary and systemic data, we reveal the critical role of B cells and lymphoid follicles in COPD immunopathogenesis, emphasising their increasing autoimmune priming over time. Notably, we identified local autoimmune priming genes within lymphoid follicles (e.g., *Enpp2*) and systemic driver genes of follicular B cell development (e.g., *Hmgb1*, *Mxd3*),

which present promising therapeutic targets for intra-pulmonary lymphoid follicle formation in pre-clinical smokers and COPD patients.

3.7 Additional Information

Acknowledgements

Hao Chen was supported by *UTS* President's Scholarship granted by *UTS* and International Research Scholarship granted by Australian government. This work was funded by a fellowship and grants from *National Health and Medical Research Council (NHMRC) of Australia* (GNT1175134) and *UTS* to Prof. Philip Hansbro. Dr. Alen Faiz, received funds from *Sanofi*. We acknowledge the support from all animal technicians at the Centre for Inflammation for their contributions to in vivo experiments: Bob Lu, Caitlin Hubbard, Jacky Lam, Kathleen Bilson, and Solomon Odgers.

Competing interests

All authors report no conflict of interest.

Data availability

Raw data generated by *nanoString* GeoMx spatial transcriptomics platform, mouse lung histological data and physiological data are available upon request and at (GSE270040; restricted access before publishing). scRNA-seq data and smoke-effect gene set are only available upon request but will be published imminently. Results of all analyses are available in GitHub (https://github.com/HAOLUNG/GeoMx_COPD_mouse_lungs).

Code availability

All code used for analysis is publicly available in GitHub (https://github.com/HAOLUNG/GeoMx_COPD_mouse_lungs).

Chapter 4: Epigenetic Adaptation of Human Airway Epithelium upon Cigarette Smoking

4.1 Declaration

Chapter 4 is a thesis-adapted version of a manuscript currently under review in *Nature Communications*, in which the author of this thesis is the co-first author.

4.2 Abstract

Cigarette smoke first interacts with the lung through airway epithelium, driving the development of most smoke-associated chronic lung diseases. Here, we profiled a group of genes ("smoking signature") that differentiates between current smokers and non-smokers well across multiple cohorts. By mapping the smoking signature onto single-cell and spatial transcriptomics of human respiratory tract, we showed the affected site by smoking is predominately smokers' airway surface epithelial layer. We then compared the gene differences of lungs of human and non-human primates, and observed a cluster of human-specific genes ("human lung evolution") which is specifically expressed in human smokers' airway surface epithelial cells, suggesting an evolutionary adaptation to smoke exposure in human airway. To elucidate this adaptation to smoke, we examined two transcription factors, *NRF2* and *AhR*, where they were identified as master regulators of gene expressions in the smoking signature. Transcription factors regulate gene expression by binding to specific DNA sequences near the target genes. To find the root of the regulation of *NRF2* and *AhR*, we performed epigenetics study where we identified their binding sites of *NQO1* and *ALDH3A1*, both of which are also increased in human airway upon smoke exposure in protein level. Accompanying knockout studies of *ALDH3A1* and *NQO1* in epithelial cells revealed their roles in a protective mechanism that aids human airways in adapting to smoke.

4.3 Introduction

Cigarette smoking remains one of the leading causes of preventable death worldwide. It is a well-known risk factor of developing various serious diseases, including lung cancer,

cardiovascular disease, and COPD. Smoking, whether acute or chronic, profoundly impacts gene expressions throughout the airways[114-116]. Airway epithelium provides the first line of physical defence to protect submucosal layers from inhaled toxicants, such as cigarette smoke, to minimise cellular damage[117,118].

Airway epithelium acts as a chemical barrier by releasing antioxidants, antiproteases, antimicrobial peptides, and bioactive cytokines, thereby initiating innate immune responses and safeguarding the airways against foreign toxicants and pathogens[119]. Airway lining comprises a complex structure featuring various specialised cell types such as ciliated cells, secretory cells, and basal cells within a ciliated pseudostratified columnar and differentiated epithelial layer, as well as rare epithelial cells such as tuft cells, pulmonary ionocytes, and pulmonary neuroendocrine cells[28,120]. Among these cell types, basal cells, located in the inner layer of the airway epithelium, serve as key progenitors that differentiate into secretory and ciliated cells in the airway surface epithelial layer[121-123].

Our previous findings have demonstrated that cigarette smoke impairs airway epithelial barrier function[124]. With the evolution of humans and the advent of cooking with fire, exposure to smoke has been a consistent aspect of human history, potentially driving adaptations in the human airway to counteract the detrimental effects of smoke exposure. Over generations, prolonged exposure to smoke may have induced epigenetic modifications, such as changes in DNA methylation patterns, that influenced the expression of genes critical for adaptive responses to environmental toxins[125]. These modifications could represent a form of evolutionary plasticity, where the human genome has responded to repeated smoke exposure through regulatory changes that enhance protection against lung damage. Here, we hypothesise that epigenetic changes induce increased expression of genes associated with oxidative stress response and detoxification, thereby offering adaptive protection against lung damage in human smokers. Specifically, these DNA methylation sites likely reside within or near the binding sites of TFs sensitive to cigarette smoke, enabling direct regulation of key genes that mediate protein-level changes in response to smoke exposure.

In this study, we first identified a "smoking signature" gene set distinguishing current smokers from never smokers across multiple cohorts. ST showed that this signature is specifically enriched in the airway of human smokers, not in other lung structures. We mapped smoking signature onto scRNA-seq datasets of human respiratory tract, revealing predominant effects

on smokers' airway surface epithelial layer which is in direct contact with inhaled smoke. Comparison between human and non-human primate lung transcriptomes showed unique gene expressions in airway surface epithelium of human smokers, indicating that evolutionary adaptation to smoke in human airway overlaps with smoking signature. We examined two TFs in smoking signature, *NRF2* and *AhR*, to unravel the adaptation to smoke in epigenetic and proteomic levels, where we revealed the role of *ALDH3A1* and *NQO1* as guardians to maintaining epithelial integrity upon smoke exposure.

4.4 Methods

Study cohorts

The Study to Obtain Normal Values of Inflammatory Variables From Healthy Subjects (NORM; ClinicalTrials accession no.: NCT00848406) includes asymptomatic smokers and never-smokers, defined by the absence of respiratory symptoms, normal lung function of FEV_1/FVC >70% and FEV_1 >80% predicted, and no bronchial hyperresponsiveness to methacholine. Bronchial biopsies were obtained for comprehensive analysis of whole-transcriptome expression and methylation patterns using RNA sequencing and 450K methylation arrays, respectively.

Methylation analysis

Processing and quality control of the Infinium Human Methylation 450K BeadChip has been previously described[126]. Differential methylation analysis was conducted between a subset of NORM cohort containing current (n=33) and never smokers (n=36) (Supplementary table 2), using the R package limma[127], adjusting for age, gender and principle component analysis based on technical probes[128]. An Benjamini–Hochberg correction for multiple testing was applied and an adjusted p-value of <0.05 was considered statistically significant.

Gene expression analysis

RNA extraction, sample preparation and high-throughput sequencing have previously been described in detail[129]. Bronchial biopsies were taken from segmental divisions of the main bronchi, 3rd-6th generation. Biopsies frozen in Tissue-Tek (VWR, Radnor, PA) at −80 °C were thawed at room temperature and cut from the blocks when they were semi-solid. Total RNA

was extracted using AllPrep DNA/RNA Mini kits (Qiagen, Venlo, the Netherlands). Samples were lysed in 600 µL RLT-plus buffer using an IKA Ultra Turrax T10 Homogenizer, and RNA was purified according to the manufacturer's instructions. RNA samples were dissolved in 30 µL RNAse free water. Concentrations and quality of RNA were checked using a Nanodrop-1000 and run on a Labchip GX (PerkinElmer, Waltham, MA). RNA samples were further processed using TruSeq Stranded Total RNA Sample Preparation Kits (Illumina, San Diego, CA), using an automated procedure in a Caliper Sciclone NGS Workstation (PerkinElmer, Waltham, MA). All cytoplasmic and mitochondria rRNA was removed (RiboZero Gold kits). The obtained cDNA fragment libraries were loaded in pools of multiple samples unto an Illumina HiSeq2500 sequencer using default parameters for paired-end sequencing (2×100 bp).

Gene expression quantification

The trimmed fastQ files were aligned to build b37 human reference genome using HISAT (v0.1.5)[130], allowing for two mismatches. Before gene quantification, SAMtools (v1.2)[131] was used to sort the aligned reads. Gene-level quantification was performed by HTSeq (v0.6.1p1)[132] using Ensembl v75 as gene annotation database.

Quality control

QC metrics were calculated for the raw sequencing data, using the FastQC tool (v0.11.3). Alignments of 77 subjects were obtained. QC metrics were calculated for the aligned reads using Picard-tools (v1.130), CollectRnaSeqMetrics, MarkDuplicates, CollectInsertSize-Metrics and SAMtools flagstat. In addition, we checked for concordance between sex-linked (XIST and Y-chromosomal genes) gene expression and reported sex. All samples were concordant. This resulted in high-quality RNA-seq data from 77 subjects.

ALI gene expression profiling during smoke exposure

Two publicly available microarray datasets from airway epithelial cells cultured at ALI from healthy controls were treated with whole gaseous phase smoke were analysed (GSE30660[133], n = 3; and GSE82137[134], n = 4). ALIs in the GSE30660 dataset were exposed for 30 minutes on four separate days to whole cigarette smoke (n = 4) compared to air exposure. Those from the GSE82137 dataset were treated with a 48-minute exposure on day one with whole cigarette smoke and then rested for 24 hours, compared to air exposure. Microarray analysis was conducted using R software v3.02 with limma package[127], and normalised using robust multi-

array average. A paired linear analysis was conducted using limma comparing treatment and control.

Transcription factor enrichment analysis

The enrichment analysis for *NRF2* and *AhR* was conducted using GSEA on the following publicly available databases GSE38332[135] and GSE109576[136], respectively. GSE38332 investigated the effects of *NRF2* inhibition on A549 expression using siRNA; control A549 cells (n=3) compared to *NRF2* siRNA A549 cells (n=3), while GSE109576 investigated the effects of AhR inhibition on A549 expression with 4 hours of treatment with CH-223191 (a selective inhibitor of AHR); control A549 cells (n=3) compared to A549 cells + CH-223191 (n=3).

Transcription factor ChIP-Seq analysis

ChIP-Seq analysis was conducted on two publicly available datasets during the activation of *NRF2* and *AhR*, respectively. We analysed ChIP-Seq for *AhR* in the human breast cancer epithelial cell line MCF-7 in the presence and absence of treatment with TCDD, an activator of *AhR*, for 24 hours (GSE90550), as previously described in detail[137]. We analysed ChIP-Seq for NRF2 in the airway epithelial cell line BEAS-2B in the presence and absence of treatment with the dietary isothiocyanate, sulforaphane (SFN), an activator of *NRF2* (GSE75812), as previously described in detail[138].

Proteomic analysis

Sputum proteomics analysis was conducted on a publicly available dataset (PXD001977) of current (n=54) and never smokers (n=54)[139]. Briefly, counts were corrected with internal controls and then analysed using the R package limma[127].

Comparing gene expression between human and primates

Gene expression profiles of human and non-human primate lung tissue were compared. Publicly available RNA-seq data from healthy lung tissue of Korean non-smoker females (n = 6) including technical replicates was downloaded from GEO, accession GSE3775. RNA-seq data from healthy lung tissues of several non-human primates was taken from the Nonhuman Primate Reference Transcriptome Resource (NHPRTR)[140], including Baboon, Chimpanzee (technical replicates n=2), Rhesus macaque, Cynomolgus macaque (both Indochinese and Mauritian), Squirrel monkey, Sooty mangabey, Pig-tailed macaque, Marmoset and Mouse

Lemur. The quality of all raw RNA-seq data was checked using FastQC. Reads were mapped against the primary human genome assembly (GRCh38) and quantified with STAR aligner (v2.4.2a)[141] using release 83 of the human gene annotation from Ensembl (http://www.ensembl.org). Differentially expressed genes were selected from a list of 232 smoking-related genes requiring an average expression level of at least 1 CPM (n=176 genes). The normalised CPM expression values were z-transformed before clustering and plotting was done using pheatmap package.

Spatial transcriptomics analysis

To investigate the specificity of the genes associated with smoking, we applied the gene signature derived from all significantly upregulated genes in smokers compared to non-smokers to publicly available spot-based ST dataset of healthy human lungs (ArrayExpress: E-MTAB-11640)[32] obtained through 10x Visium platform. Our analysis focused specifically on samples extracted from the parenchyma-enriched region of the lungs, comprising data from asymptomatic current smokers (n=1) and never smokers (n=1), as these were the only relevant samples available in the dataset relevant to our investigation. Airway club cell marker genes (*SCGB1A1* and *SCGB3A1*) and visualization of hematoxylin and eosin (H&E)-stained lung tissue sections were employed to identify and locate the small airways within the lung parenchyma region.

scRNA-seq analysis

To investigate which cell types express genes that change upon smoke exposure the following cell types (Club 1, Basal 1, Goblet 1, Basal2, Ciliated 1, Club 2, Goblet 2, Club 3, Ciliated 2) were analysed from scRNA-seq data from current (n=6) and never smokers (n=6) (GSE131391)[142]. In a second analysis these cell types (Basal, Club, Goblet, and Ciliated) were merged. A signature was created of the genes that were higher in current compared to never smokers from the RNA-seq data (FC>1, FDR<0.05) and genes that were more highly expressed in the lungs of humans compared to other primates. UMAPs and violin plots were created based on smoking status.

CRISPR knockout creation

To generate CRISPR-Cas9 KO's the plasmid (Px458) was used as previously described [143]. Guide RNA (gRNA) were designed using the online web tool Benchling (v2018). A single gRNA was created for each of the four different genes (ALDH3A1, CYP1A1, CYP1B1, NQO1), targeting a common exon contained in all known splice variants determined by Ensembl. The CRISPR-Cas9 construct was transfected into A549 cells seeded at 100,000 cells/well in growth medium (RPMI 1640 medium, Lonza, Basel, Switzerland) supplemented with 10% Fetal Calf Serum (FCS, Sigma-Aldrich, St. Louis, USA) and 1% Penicillin/Streptomycin (P/S, Gibco, California, USA) in 12 well plates (Sigma-Aldrich). The transfection of the px458 CRISPR-Cas9 construct was performed with Lipofectamine 3000 Transfection Reagent (ThermoFisher Scientific) according to the manufacturer's protocol.

After 24 hours of incubation, transfected cells were harvested and sorted based on GFP using a single cell sorter (SH800S Cell Sorter; Sony Biotechnology, Weybridge, UK) into a 96 well plate. To validate the knock-out of selected genes, DNA was extracted using QIAamp DNA Investigator Kits (Ref. No. 56504; QIAGEN) according to the manufacturer's protocol, regions surrounding the cut site amplified by PCR and the samples sequenced at BaseClear (Leiden, Netherlands), and subsequently assessed with TIDE web tool and Benchling.

Cigarette smoke extract

Two 3R4F research cigarettes without filters (Tobacco Research & Development Center, Lexington, KY) were bubbled through 25 mL of RPMI 1640 medium (Lonza) supplemented with 1% P/S (Gibco), using a high flow peristaltic pump (Watson Marlow 603S, Rotterdam, the Netherlands). The obtained solution was used as 100% cigarette smoke extract. The extract was prepared freshly for each experiment and used within 30 minutes.

Electric cell-substrate impedance sensing

Electrical resistance was recorded in real-time using the ECIS Z Theta system (Applied Biophysics) with associated software. 8W10E+ PET electrode chamber arrays were equilibrated with serum-free growth medium for 2 hours. A549 cells were seeded in duplicates at 25,000 cells/well. After serum-starvation for 24 hours the cells were exposed to medium or 20% cigarette smoke extract for 24 hours. Resistance was measured from 0-72 hours at a low-frequency resistance of 40 Hz, to assess cell-cell contacts[144].

Annexin-V/PI staining

A549 cells were treated with 0%, 50% and 100% cigarette smoke extract for 4 hours. Afterwards, cells were incubated with fresh serum-free medium for 20 hours. Cells were collected and washed twice using Cell Staining Buffer (BioLegend; Cat. No. 420201) and centrifuged (1,200 x*g*, 5 minutes). Cells were resuspended and stained with 40:42 Annexin-V Binding Buffer (BioLegend; Cat. No. 422201), 1:42 Annexin-V FITC (ImmunoTools; Cat. No. 31490013) and 1:42 Propidium Iodide [1.0 mg/mL], and incubated (15 minutes, RT, protected from light). Fluorescence was assessed using a FACSCalibur (Becton-Dickinson Medical Systems, Heidelberg, Germany).

4.5 Results

Generation and evaluation of the smoking signature

Cigarette smoke first interacts with the lung through the airway epithelium, contributing to the development of many chronic lung diseases. Most existing studies used the airway epithelium in the upper respiratory tract (trachea)[145] or *in vitro* airway epithelial cell culture[146], overlooking the direct effect of smoke exposure on the airways within the lungs and the impact across multiple layers of omics. To address this, we employed a comprehensive multi-omics approach on respiratory samples exposed to smoke, comparing the transcriptomic differences between human and non-human primate lungs (Fig. 4.1a). Our study highlights an evolutionary adaptation in the human airway to withstand smoke exposure.

To precisely define asymptomatic smokers, we first recruited individuals without respiratory symptoms (NORM; ClinicalTrials accession no.: NCT00848406; Supplementary Table 2). We then applied bulk RNA-sequencing on bronchial biopsies of asymptomatic current smokers (n=37) and never smokers (n=40) from the NORM cohort. We identified the 194 significantly upregulated genes in response to cigarette smoking, referred to as the "smoking signature" (Fig. 4.1b, Supplementary Table 3). This signature includes genes related to detoxification, mucus production, and cancer (log2 fold change >1, adjusted $p<0.05$, Supplementary Table 4), consistent with known pathogenic effects of cigarette smoking. The robustness of the smoking signature was confirmed by its upregulation in two independent datasets of *in vitro* smoke-exposed airway epithelial cells cultured in air-liquid interface (ALI)[133,134] (Fig. 4.1c).

To verify the specificity of the smoking signature to the airway, we mapped it onto a publicly available *10x Visium* spatial transcriptomics dataset of asymptomatic human lungs[32]. Smoking signature is enriched in the airway of current smokers, even in a lung region predominantly consisting of alveoli, showing higher expression compared to never smokers (Fig. 4.1d). For greater cellular resolution, we further examined scRNA-seq data from bronchial brushes of current smokers (n=6) and never smokers (n=6)[142]. Smoking signature is primarily expressed by goblet/club cells and ciliated cells from the bronchial brush samples (Fig. 4.1e), with minimal changes observed in basal cells, indicating that the direct response to smoke occurs mainly in the airway surface epithelial layer, which is directly exposed to inhaled smoke. We further validated this observation using the Human Lung Cell Atlas[147], confirming the predominant expression of the smoking signature in goblet, club, and ciliated cells, while basal cells showed no prominent alterations (Fig. 4.1f). These findings underscore the critical role of the airway surface epithelial layer in responding to cigarette smoke (Fig. 4.1g).

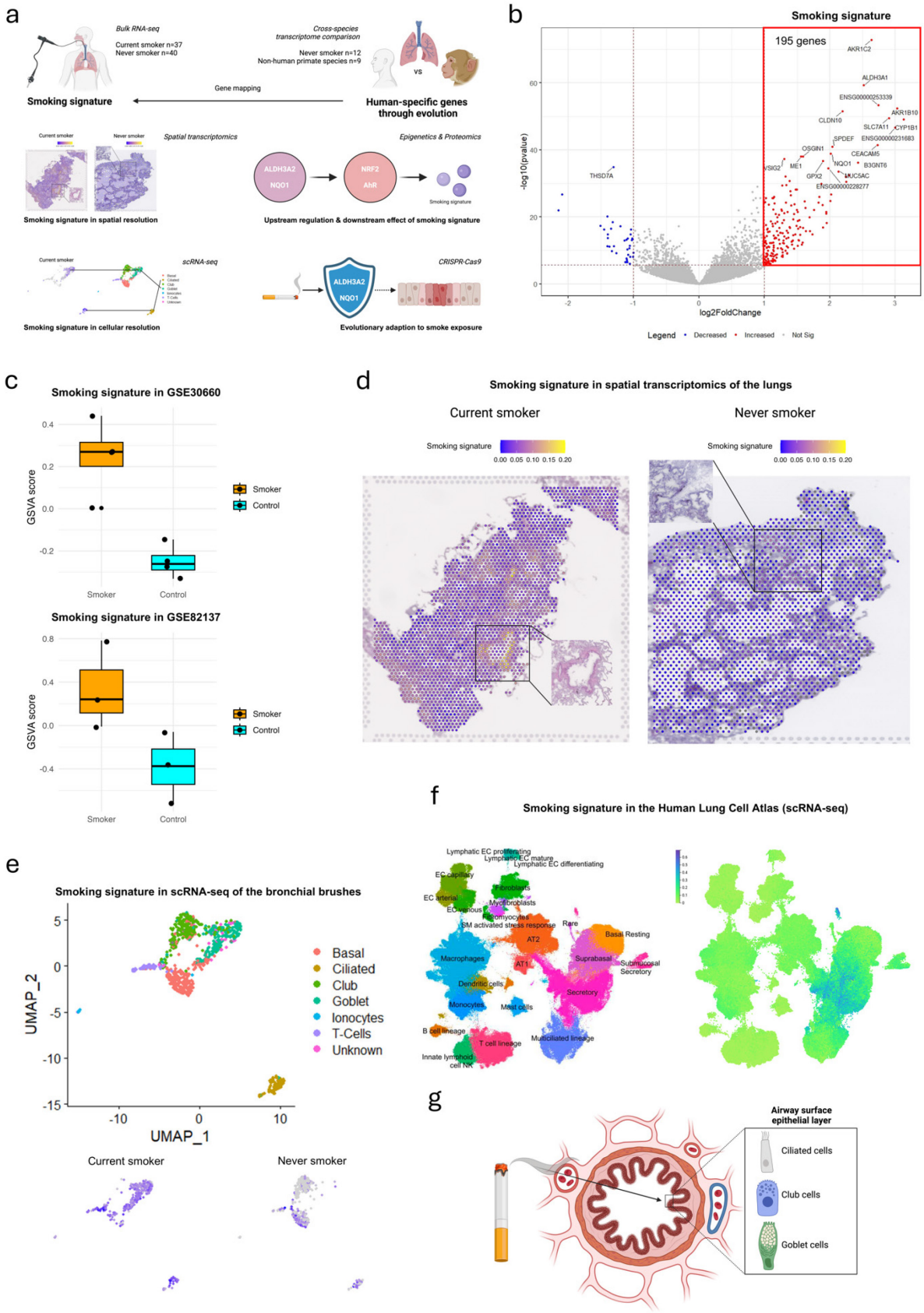


Fig. 4.1: Generation and evaluation of the smoking signature at the transcriptome level.

a, Schematic of experimental design of the study. **b**, Volcano plot of significantly differentially expressed genes in bronchial biopsies from current smokers (n=37) compared to never smokers (n=40) within the NORM cohort (ClinicalTrials accession no.: NCT00848406). A total of 194 significantly upregulated genes, highlighted within the boxed area, constitute the "smoking signature". Significance was determined by a log2 fold change >1 and a Benjamini-Hochberg (BH) adjusted p-value <0.05. **c,** Boxplot of gene set variation analysis (GSVA) signalling scores of the smoking signature in two independent datasets of *in vitro* smoke-exposed airway epithelial cells cultured in air-liquid interface (ALI)[133,134]. The centre line corresponds to the median, the top and bottom hinges delineate the first and third quartiles, respectively. **d,** Mapping score of the smoking signature in 10x Visium spatial transcriptomics dataset[32] of lung regions from a current smoker and a never smoker, predominantly consisting of alveoli with minimal airway presence. The boxed area highlights the airway region. **e,** Uniform manifold approximation and projection (UMAP) of a single-cell RNA sequencing (scRNA-seq) dataset of bronchial brushes of current smokers (n=6) and never smokers (n=6)[142] (top), and mapping score of the smoking signature based on the smoking status (bottom). **f,** UMAP of the Human Lung Cell Atlas (left) and the mapping scores of the smoking signature (right). **g,** Illustration of the main effect of cigarette smoke on the airway surface epithelial layer.

Evolutionary adaptation to smoke in human airway

Over generations, prolonged exposure to smoke may have induced epigenetic modifications, such as changes in DNA methylation patterns, that influenced the expression of genes critical for adaptive responses to environmental toxins[125]. These modifications could represent a form of evolutionary plasticity, where the human genome has adapted to repeated smoke exposure through regulatory changes that enhance protection against lung damage. Since the use of fire for cooking is a relatively recent development in human evolution, we sought to determine whether this protective mechanism is unique to humans or conserved across primate species. To address this, we compared the lung transcriptomes of humans to those of other primates, including chimpanzee, baboon, cynomolgus macaque, marmoset, pig-tailed macaque, rhesus macaque, sooty mangabey, and squirrel monkey, all of which were collected from Nonhuman Primate Reference Transcriptome Resource[140] (Fig. 4.2a). Cross-species transcriptomic

comparison identified four distinct clusters of genes with lineage-specific upregulation (Fig. 4.2b). Of which, a cluster (named “human lung evolution”) showed significantly higher expression in human lungs compared to non-human primates. Interestingly, we found a high overlap of the human lung evolutionary genes with the smoking signature (Fig. 4.2c) as well as a high expression of human lung evolutionary genes in the bronchial biopsies of human smokers (Fig. 4.2d), evidenced that evolution plays a role in human’s response to smoke exposure. Mapping of human lung evolution genes onto the Human Lung Cell Atlas[147] revealed predominant enrichment in goblet, club, and ciliated cells (Fig. 4.2e), all of which exactly overlap with the cell types affected by smoking signature, suggesting that the key evolutionary differences between human and non-human primate lungs lie in the same airway surface epithelial layer that directly responds to smoke exposure.

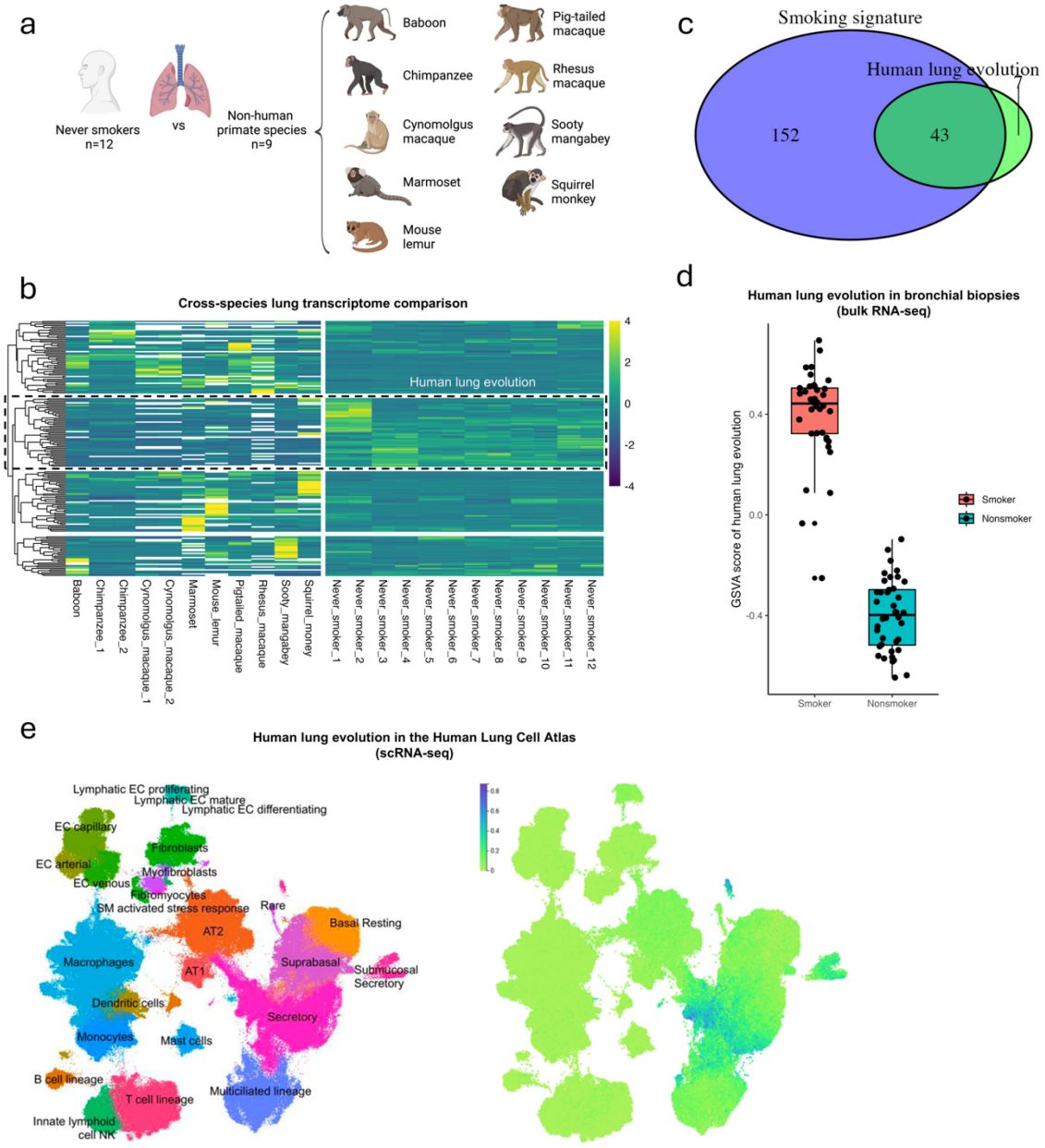


Fig. 4.2: Evolutionary adaptation to cigarette smoke exposure in human airway.

a, Illustration of cross-species lung transcriptome analysis comparing human never smokers (n=12) and non-human primate species (n=9 species). **b**, Heatmap of a total of four clusters containing significantly differentially expressed genes between the lungs of human never smokers and non-human primates. The boxed cluster, with higher expression in humans, represents the "human lung evolution" signature. **c,** Venn diagram showing the overlap between the "human lung evolution" signature and the smoking signature. **d,** Boxplot of GSVA signalling scores of the "human lung evolution" signature in the bulk RNA-seq of bronchial biopsies collected from current smokers and never smokers from the NORM cohort. The

centre line corresponds to the median, the top and bottom hinges delineate the first and third quartiles, respectively. **e,** UMAP of the Human Lung Cell Atlas (left) and the mapping scores of the "human lung evolution" signature (right).

NRF2 and *AhR* as master regulators of the smoking signature

We then investigated whether these human-specific genes are regulated by the same epigenetic mechanisms controlling gene expressions in the smoking signature. Among genes in the smoking signature, there are two key TFs, including *NRF2* and *AhR*. Suppression of *NRF2*[135] and *AhR*[136] in A549 human alveolar basal epithelial cells led to decrease in the smoking signature upon smoke exposure, suggesting their role as master regulator of the smoking signature (Fig. 4.3a, b). We then investigated the proximity of *NRF2* and *AhR* binding to other genes in the smoking signature using ChIP-Seq analysis prior- and post-activation with D,L-sulforaphane16 and 2,3,7,8-tetrachlorodibenzo-p-dioxin[137], respectively, in *NRF2*- and *AhR*-inhibited A549 epithelial cell lines. Initially, we examined genes in smoking signature that are downregulated during *AhR* suppression (Supplementary Table 5). In *AhR*-inhibited cells, 60.0% of the suppressed genes in smoking signature exhibit an *AhR* binding peak within a 100KB region flanking their transcription start site (TSS), indicating direct binding of the majority of *AhR*-altered genes by the protein. Next, in *NRF2*-inhibited cells, only 21.9% suppressed genes in smoking signature (Supplementary Table 5) show an *NRF2* binding peak within a 100KB region flanking their TSS, suggesting that the majority of *NRF2*-regulated genes are indirectly altered by it.

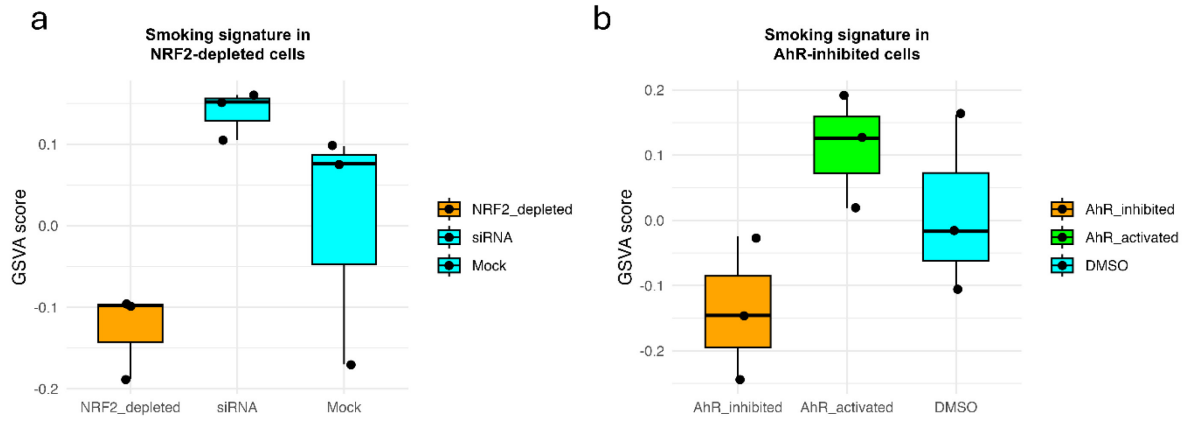


Fig. 4.3: *NRF2* and *AhR* as master regulators of the smoking signature.

Boxplots of GSVA signalling scores of the smoking signature in *NRF2*-depleted cells (**a**) and in *AhR*-inhibited cells (**b**). The centre line corresponds to the median, the top and bottom hinges delineate the first and third quartiles, respectively.

ALDH3A1 and *NQO1* as upstream epigenetic regulators of the smoking signature

Initially, DNA methylation profiling of bronchial biopsies from asymptomatic current (n=33) and never smokers (n=36) revealed 14,186 differentially methylated CpG sites, well-differentiated between smokers and non-smokers (Fig. 4.4a-c). Notably, CpG methylation was over-represented in the 5' UTR and gene body regions (Fig. 4.4d), suggesting direct regulation of gene expression by smoking-induced DNA methylation given their position in the promotor region. Key CpG sites include cg00370022, upstream of the detoxification gene *CYP1A1*, and cg23576855 in *AhR* repressor (*AhRR*), both involve in the *AhR* signalling (Fig. 4.4e, Supplementary Table 6).

To bridge how the changes of DNA methylation regulate gene expression of the smoking signature, we performed cis-expression quantitative trait methylation (eQTM) analysis (1 MB flanking the gene) using bronchial biopsies from asymptomatic current (n=29) and never smokers (n=35). Among 1,882 eQTMs identified, 143 (7.6%) are located within the binding sites of TFs, including NRF2, AhR, and AhRR. Notably, the *CYP1B1* gene shows the strongest eQTM, located upstream of the TSS of *CYP1B1* and within the TF binding sites of both *AhR* and *AhRR* (Fig. 14f). Similarly, eQTMs for the ALDH3A1 gene are found within AhR binding sites

and positioned 25,000 bp downstream of its TSS (Fig. 14g). Additionally, eQTM for *AhRR* gene is situated within the TF binding sites of *AhR* and *AhRR*, positioned 50,000 bp downstream to the TSS of *ALDH3A1* gene (Fig. 4.4h). Moreover, eQTMs for *NQO1* gene are located 2,500 bp downstream to the TSS of *NRF2* gene (Fig. 4.4i).

We further explored protein-level changes by analysing the sputum proteome of current and never smokers using a publicly available dataset (PXD001977)[139]. Among the 499 proteins identified across >50% of the samples, 24 showed significantly differential abundance between current and never smokers (log2 fold change >1.5, adjusted p-value <0.05, Supplementary Table 8), with 11 upregulated and 13 downregulated in smokers (Fig. 4.4j). String network analysis of these 24 proteins shows significant enrichment of interactions, indicating biological connectivity and functional relevance (Supplementary Fig. 2). The proteome shows enrichment for "*RAGE receptor binding*", "*Toll-like receptor 4 binding*", upregulation of detoxification pathway "*Metabolism of xenobiotics by cytochrome P450*", and cancer-related pathway "*Chemical carcinogenesis*" (adjusted p<0.05, Supplementary Table 9), paralleling transcriptional findings. Notably, *ALDH3A1* and *NQO1*, altered at both transcriptional and epigenetic levels, are among the proteins differentially expressed in smokers.

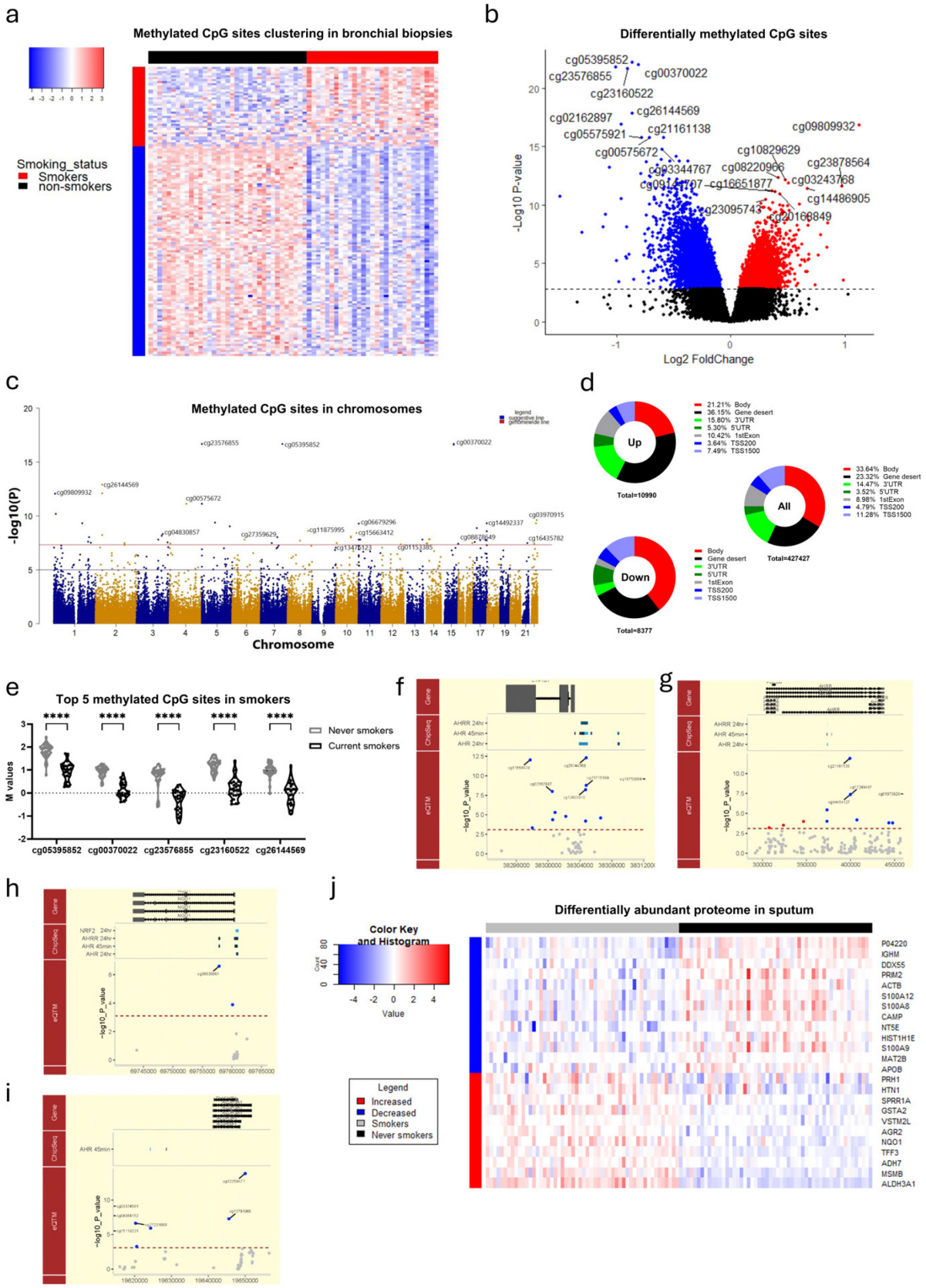


Fig. 4.4: *ALDH3A1* and *NQO1* as upstream epigenetic regulators of the smoking signature.

Heatmap (**a**) and volcano plot (**b**) of the significantly differentially methylated CpG sites in the bronchial biopsies collected from current smokers (n=33) and never smokers (n=36) from the NORM cohort. BH-adjusted p-value <0.05. **c**, Manhattan plot of the differentially methylated CpG sites between current and never smokers and their positions on chromosomes. **d,** Positional methylated CpG sites upon smoke exposure. **e**, Violin plot of the top 5 significantly methylated CpG sites. Data shown as individual data points, **** p<0.0001. Expression quantitative trait methylation (eQTM) analysis in relation to transcription factor binding regions for CYP1B1 (**f**), ALDH3A1 (**g**), AhRR (**h**), and NQO1 (**i**). Genes located in the identified genomic region are shown in the top panel, the black arrows indicate the direction of transcription. The middle panel shows the binding of transcription factors by ChIP-Seq analysis. Bottom panel demonstrates the eQTM, the –log10 p-value (y-axis) against genomic location (x-axis). eQTMs with a false discovery rate (FDR) <0.05 and with a negative association with gene expression are represented as blue points, while positive associations are represented as red points. All other eQTMs that were not found to be significant are depicted as grey points. **j**, Heatmap of 24 sputum proteins significantly altered between smokers and never-smokers (log2 fold change >1.5 or <1.5, BH-adjusted p-value <0.05).

ALDH3A1 and *NQO1* as guardians of epithelial survival and barrier function upon smoke exposure

Previous studies have shown upregulation of genes related to aldehyde and ketone metabolism, including *ALDH3A1* and *NQO1* in whole lungs[148] and in airway epithelial cells upon smoking[149,150]. To investigate their role in cellular functions such as survival and epithelial barrier integrity upon smoke exposure, CRISPR/Cas9 knockout (KO) cell lines were generated using A549 human alveolar basal epithelial cells. Cell death was assessed by measuring necrosis and apoptosis in CRISPR/Cas9 KO A549 cells after 4 hours of cigarette smoke extract exposure, followed by a 24-hour recovery. Wild-type A549 cells exposed to 100% cigarette smoke extract showed decreased viability and increased necrosis (Fig. 4.5a). *ALDH3A1*-KO cells exhibited even higher necrosis levels compared to wild-type cells, while no significant differences were observed in other KO cells.

Next, the barrier function in response to smoke exposure was measured in real-time by monitoring electrical resistance and capacitance during and after the establishment of epithelial monolayers using ECIS. Although baseline barrier integrity was similar, wild-type cells exposed to 20% cigarette smoke extract experienced reduced monolayer integrity, which was further compromised in *ALDH3A1*- and *NQO1*-KO cells (Fig. 4.5b, c). In contrast, no significant differences were observed in *CYP1A1*- and *CYP1B1*-KO cell lines. These findings suggest that *ALDH3A1* plays a protective role, as its knockout leads to increased necrosis and epithelial barrier disruption upon cigarette smoke exposure.

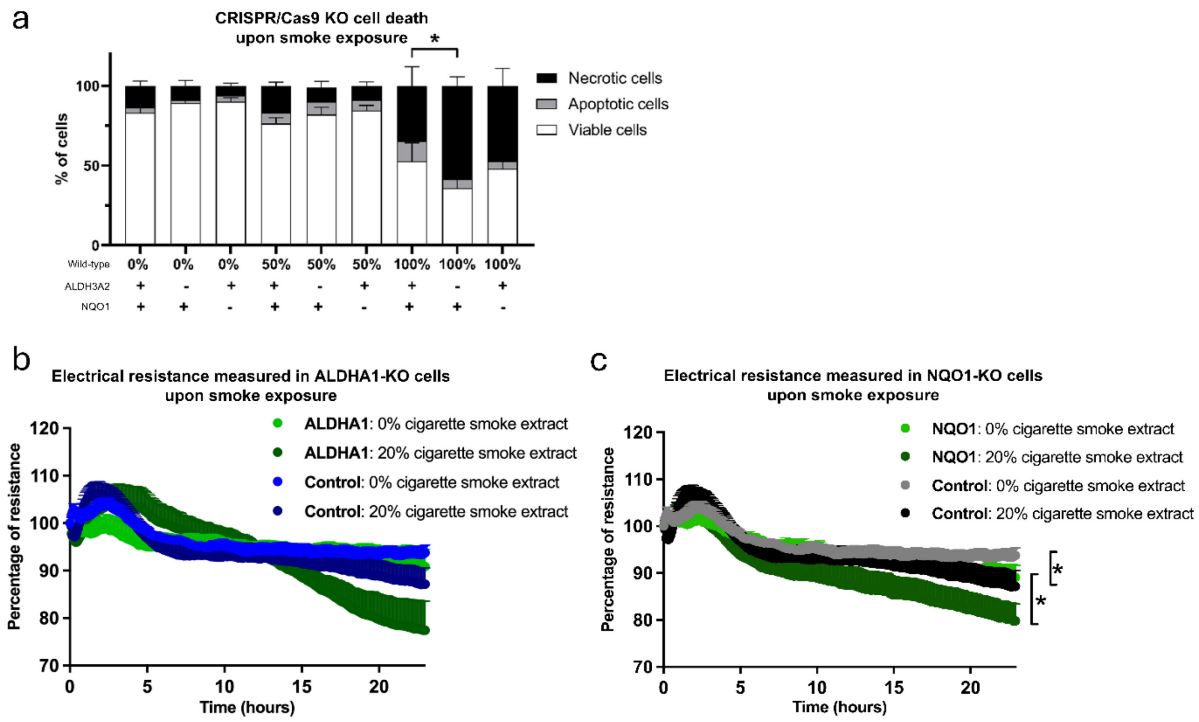


Fig. 4.5: *ALDH3A1* and *NQO1* as guardians of epithelial survival and barrier function upon smoke exposure.

a, A549 epithelial cell death upon 4 hours of exposure to 0-100% cigarette smoke extract in ALDH3A1 and NQO1 gene knockout condition through CRISPR-Cas9. Epithelial cell barrier function change in response to 0-20% cigarette smoke extract exposure for 24 hours, which was measured in real-time monitoring of electrical resistance and capacitance during and upon the establishment of epithelial monolayers using electric cell-substrate impedance sensing (ECIS) in *ALDH3A1*- (**b**) and *NQO1*-knockout (**c**) A549 cells. * p-value <0.05.

4.6 Discussion

This study represents the first comprehensive investigation employing a broad multi-omics approach to elucidate the biological responses of airway epithelial cells to chronic smoke exposure. We delineated the transcriptional and epigenetic profile of the airway epithelium upon cigarette smoke exposure. Furthermore, we demonstrated a robust association between epigenetic regulation and the transcriptional response of oxidative stress response and detoxification genes to this exposure. Cigarette smoke induced epigenetic marks to regulate gene expression, being positioned within smoking active TF binding sites. The upregulated genes, particularly *ALDH3A1*, showed protective effects against cigarette smoke exposure in lung epithelial cells, as loss of function of these genes exacerbated the loss of epithelial integrity and increased epithelial cell death upon cigarette smoke exposure. these genes were predominantly expressed by ciliated and goblet cells, indicating that the oxidative stress and detoxification response are localised to the surface layer of the airway wall.

Our findings regarding the higher expression of genes related to oxidative stress and detoxification pathways upon chronic smoke exposure are consistent with previous studies[151,152]. Furthermore, these genes are likely under the direct regulation of the TF candidates *NRF2* and *AhR* of which the binding sites are epigenetically regulated. Both *NRF2* and *AhR* play significant roles in cellular responses to environmental pollutants[153,154]. Mice deficient in *AhR* display more severe pulmonary inflammation upon smoking[155], while mice deficient in *NRF2* are highly susceptible to cigarette smoke-induced emphysema[156], providing evidence that both TFs have essential roles in regulating genes that protect against cigarette smoke. Our enrichment analysis of smoking-induced methylation sites supports the hypothesis that the expression of detoxification and antioxidant genes is tightly regulated, and that their upregulation in response to prolonged smoke exposure is likely a result of hypomethylation of these sites.

Comparative studies showed that the lungs of humans exhibit higher expression of genes associated with the response to smoke compared to other primates, likely due to our historical practice of cooking food over open flames, resulting in direct exposure to associated smoke[157]. *NQO1* is considered to have an antioxidant role during smoke exposure through the reduction

of endogenous quinones, such as p-benzosemiquinone, a prominent constituent of cigarette smoke. In contrast, *ALDH3A1* acts as an oxidant to catalyse the oxidation of reactive aldehydes in cigarette smoke. In our study, knocking out *ALDH3A1* led to reduced epithelial barrier function and cell survival. Supporting these results, a previous study showed that suppression using siRNA targeting *ALDH3A1* or *AhR* decreased the survival of human bronchial epithelial cells during smoke exposure[158]. Our study revealed that *ADLH3A1* and *NQO1* are specifically expressed in differentiated epithelial cells localized to the lumen, indicating that the oxidative stress response in the airways is confined to cells on the airway luminal surface directly exposed to inhaled smoke. In cell culture, our findings and those of others have shown that these genes are upregulated in undifferentiated basal cells in response to cigarette smoke extract, indicating that the lack of basal cell responses to smoke exposure is not due to an inability to respond but likely due to a lack of direct exposure. This has important implications on the validity of most smoking models, which directly expose cells found deeper in the airways, such as basal cells.

Although the observed expression patterns raise the hypothesis that certain airway-enriched genes may have evolved in response to environmental exposures such as cigarette smoke, this remains speculative. Our study did not include phylogenetic or selective pressure analyses to directly assess evolutionary adaptation. Future comparative genomic studies, ideally controlling for smoke exposure across species, would be necessary to confirm this concept. However, such experiments involving controlled smoke exposure in non-human primates present considerable ethical and logistical limitations. Here, we sought to provide preliminary insight by comparing transcriptomes between human non-smokers and non–smoke-exposed primates.

Overall, our study shows that the airways are epigenetically regulated to promote the expression of oxidative stress-response genes, thereby mitigating the effects of chronic smoke exposure. Importantly, this response is specific to airway surface epithelial cells and does not extend to the basal cell layer.

4.7 Additional Information

Acknowledgements

This work was funded by a fellowship and grants from *National Health and Medical Research Council (NHMRC)* of Australia (GNT1175134) and *UTS* to Prof. Philip Hansbro, the primary supervisor of the thesis author. Dr. Alen Faiz, the co-supervisor of the thesis author, received funds from *Sanofi*. The author of this thesis is supported by *UTS* President's Scholarship granted by *UTS* and International Research Scholarship granted by Australian government.

Competing interests

All authors report no conflict of interest.

Data availability

All data are available in the main text. Results of all analyses are available in GitHub (https://github.com/HAOLUNG/smoker_airway_epigenetics).

Chapter 5: Cellular Senescence and its targeting in experimental COPD

5.1 Abstract

Cellular senescence has emerged as a key driver of chronic lung diseases, particularly COPD. Senescence is primarily triggered by the activation of *p53* and $p16^{INK4a}$, which subsequently induce $p21^{CIP1}$-mediated cell cycle arrest. In COPD, senescent cells accumulate in the lungs and contribute to chronic low-grade inflammation through the secretion of inflammatory proteins collectively known as the senescence-associated secretory phenotype. Senolytics, a class of drugs that selectively induce apoptosis in senescent cells while sparing healthy cells, have recently shown potential for therapeutic intervention for other chronic lung conditions such as pulmonary fibrosis. Here, we used a pre-clinical experimental COPD animal model to evaluate the efficacy of a senolytic cocktail of the senolytics quercetin and dasatinib in cigarette smoke-induced COPD. We demonstrated the presence of senescent cells in the lungs in experimental COPD and showed that oral administration of senolytics effectively eliminated these cells in the lungs. Senolytic treatment reduced immune cell infiltration to the lungs and improved pulmonary structural integrity. Our study highlights the potential of quercetin and dasatinib as a promising front-line therapy and as an adjuvant therapy alongside conventional symptom-management treatments for COPD.

5.2 Introduction

COPD is a complex and heterogeneous lung condition characterised by persistent inflammation that originates in the airways and progressively involves the lung parenchyma, ultimately leading to airway remodelling, alveolar destruction and impaired lung function[1]. Cigarette smoking is the primary risk factor for COPD although air pollution and cooking smoke are also important causal factors. In 2019, it was estimated that approximately 391 million individuals aged 30–79 years worldwide were affected by COPD[2], accounting for nearly 10% of the global population. COPD is now recognised as the third leading cause of morbidity and mortality worldwide[1]. In healthy individuals, FEV1 and FVC typically peak around the age of

25 and gradually decline with age[159]. This age-related decline may be attributed to low-grade pulmonary inflammation, peripheral airway narrowing, and alveolar space enlargement, which are features that closely resemble the key pathological hallmarks of COPD.

Cellular senescence is a hallmark of ageing and arises primarily from repeated cellular divisions, leading to progressive telomere shortening and activation of DNA damage response pathways. This process activates the tumour suppressor *p53*. Alternatively, cellular senescence can also be induced by cellular stressors, such as oxidative stress, radiation, or cytotoxic therapies, which activate the cyclin-dependent kinase inhibitor $p16^{INK4a}$. These pathways may converge to activate $p21^{CIP1}$, a downstream effector that causes cell cycle arrest by inhibiting cyclin-dependent kinases 2 and 4. Senolytics, a class of drugs that selectively target senescent cells, induce apoptosis in these cells while having minimal effect on proliferating cells[160]. Notably, the senolytic cocktail of quercetin and dasatinib (Q+D) has substantial beneficial effects in aged mice, reducing senescent cell accumulation, extending lifespan by 36%, and decreasing mortality risk by 65%[161].

Senescent cells, including alveolar epithelial and endothelial cells, have been shown to accumulate in the lungs of patients with COPD[162-164]. Similarly, in pulmonary fibrosis, cellular senescence is observed in alveolar epithelial cells and fibroblasts[165]. To date, the first and only clinical trial using Q+D senolytic therapy was conducted in patients with idiopathic pulmonary fibrosis aged over 50 years. This pilot study demonstrated slight clinical improvements, such as improved 6-minute walk distance, but the findings remained inconclusive due to the small sample size[166]. To build confidence for future clinical trials of senolytics in COPD, we conducted a pre-clinical study to evaluate the efficacy of oral administration of the Q+D senolytic cocktail in an experimental COPD animal model induced by cigarette smoke exposure. We confirmed the accumulation of senescent cells in the lungs in experimental COPD and demonstrated that senolytic treatment effectively eliminated these cells. Although spirometry-like lung function was not restored, senolytic therapy reduced immune cell infiltration in the lungs and improved pulmonary structural integrity. Our findings underscore the potential of Q+D as a promising front-line and adjuvant therapy alongside conventional treatments for managing COPD symptoms.

Senolytics are administered intermittently to effectively target and eliminate accumulated senescent cells. The intermittent dosing strategy offers several advantages: allowing for efficient clearance of senescent cells, while minimising drug exposure to healthy cells thereby reducing the risk of side effects. Intermittent treatment also prevents complete depletion of senescent cells, preserving their occasional beneficial roles such as promotion of wound healing[167], and aligns with the natural accumulation cycle of senescent cells[160], ensuring the therapy is optimally administered when the burden is highest.

5.3 Methods

Ethics

The Animal Welfare Committee of Sydney Local Health District approved all experimental procedures (protocol no.: 2018-030). This study was conducted in accordance with the Australian responsible use of animal code.

Animal model of cigarette smoke-induced experimental COPD

The protocol for inducing experimental COPD via nose-only cigarette smoke inhalation in the mouse model was performed as previously described[17,42-53]. The establishment of this model, similar to those in Chapters 2 and 3. Briefly, 8-week-old wild-type female C57BL/6J mice were exposed to cigarette smoke for 8 weeks. Mice were treated either weekly or bi-weekly with oral administration of the senolytic cocktail Q+D or vehicle control on a weekly basis. Time-matched control groups were exposed to normal air, with or without the same treatments (n=8 per group).

Senolytic administration

The senolytic cocktail consisting of quercetin (50 mg/kg) and dasatinib (5 mg/kg) was prepared in 100% PEG400. Mice received the treatment via oral gavage, with three doses administered per treatment cycle. A treatment cycle was defined as either one week (weekly administration) or two weeks (bi-weekly administration), depending on the designated frequency for each group. Vehicle-treated groups received 100% PEG400 following the same administration schedule.

Airway inflammation, emphysema and lung function measures

Airway inflammation was determined by differential cell counts in bronchoalveolar lavage (BAL) fluid. The approaches to measure baseline structural and cellular changes in respiratory samples from animal model were described previously[17,42-53]. Emphysema was quantified using the mean linear intercept method, a standard approach for assessing alveolar diameter in mice. Lung function parameters were evaluated using the flexiVent system with forced oscillation and forced manoeuvre techniques.

RT-qPCR for Gene Expression Analysis

Total RNA was extracted from mouse lung homogenates. cDNA synthesis was performed using a high-capacity reverse transcription kit with random primers. RT-qPCR was performed to assess the expression of genes that were associated with anti-apoptotic (*Bcl-w*, *Bcl-xl*), senescence-related (*p21*), senescence-associated secretory phenotype (SASP) (*Pai1*, *Mmp12*), and the anti-ageing (*Sirt1*) processes. Gene-specific primers were designed and validated to ensure amplification specificity. RT-qPCR reactions were conducted using a SYBR Green master mix on a real-time thermal cycler, with triplicate technical replicates for each sample. Gene expression levels were normalised to the housekeeping gene *Gapdh*, and relative expression was calculated using the ΔΔCt method.

SA-β-gal staining

Frozen murine lung tissue sections were used to stain for senescent cells with senescence-associated β-galactosidase (SA-β-gal). Once thawed, a 5-minute PBS wash was used to remove OCT from tissues, with two additional PBS washes to ensure thorough cleaning. Tissue sections were fixed with 100µL of freshly prepared 2% paraformaldehyde (PFA, in PBS, 10 minutes, room temperature, protected from light). After fixation, sections were washed twice with 100 µL PBS to remove residual PFA. A working solution of the SA-β-gal SGP probe (ThermoFisher #C10841) was freshly prepared by diluting the probe 1:1000 in pre-warmed senescence buffer (37°C). Each tissue section was stained with 100µL of the working solution and incubated (37°C without CO_2, 1 hour, protected from light). Following incubation, the working solution was removed, and the slides were washed three times with PBS. Sections

were counterstained with DAPI and imaged using a fluorescence microscope with a 488nm/FITC filter set and an exposure setting of 1000.

5.4 Results

Improvement of intrapulmonary immune cell infiltration and alveolar integrity by senolytics

To determine the optimal dosing frequency for targeting senescent cells in the lungs during regular cigarette smoke exposure, we evaluated the administration of Q+D on a weekly and bi-weekly schedule in a pre-clinical model of cigarette smoke-induced COPD (Fig. 5.1a). Smoke-exposed mice had reduced body weight compared to air-exposed mice (Fig. 5.1b), which is due to a reduction a water content. Q+D treatment did not reverse this effect similar to most other treatments. In terms of lung function, significant impairment was observed in the experimental COPD group, whereas Q+D administration did not lead to any measurable improvement (Fig. 5.1c). In fact, a slight reduction in static compliance (Cst) was observed with bi-weekly administration of Q+D, potentially due to the senolytics promoting pulmonary hypertension at this dosing frequency[168].

Q+D had potent beneficial effects in reducing immune cell infiltration in BAL fluid, with a marked decrease in macrophages, neutrophils, and lymphocytes (Fig. 5.1d). Furthermore, while emphysema was clearly observed in the experimental COPD group, senolytic treatment almost completely inhibited this feature (Fig. 5.1e). These findings highlight the therapeutic potential of Q+D senolytics in experimental COPD, as evidenced through reductions in intrapulmonary immune cell infiltration and by improvements in alveolar integrity. This would be most beneficial in patients with emphysema at a milder stage.

These findings highlight the therapeutic potential of Q+D senolytics in experimental COPD, as evidenced by reductions in intrapulmonary immune cell infiltration and improvements in alveolar integrity. Such effects may be most relevant to early or milder stages of COPD, particularly in cases presenting with inflammatory or emphysematous features.

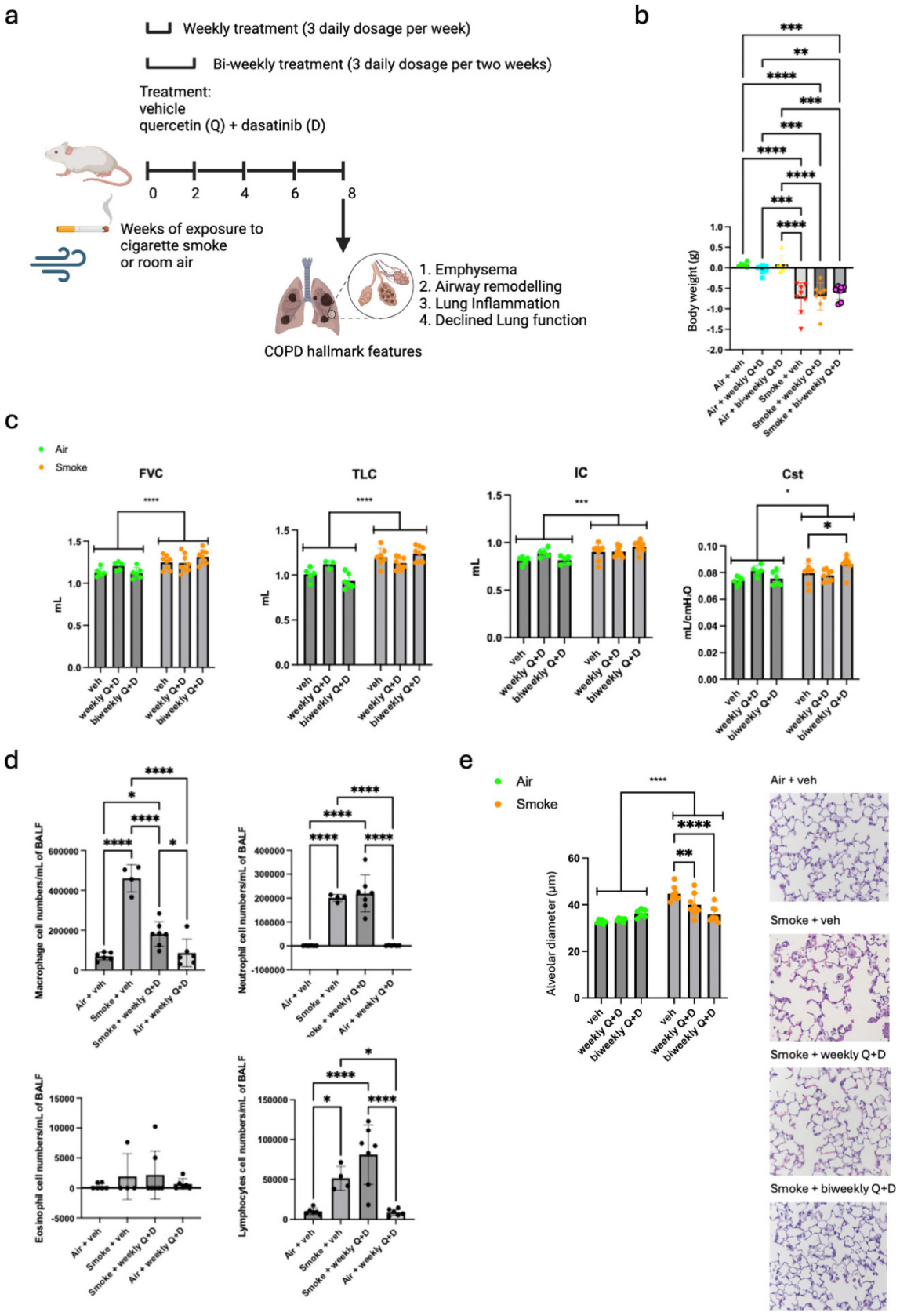


Fig. 5.1: Effects of senolytic administration of Q+D in experimental COPD

a, Schematic representation of the establishment of an experimental COPD mouse model used as a pre-clinical platform to test the effects of Q+D senolytic treatment. **b**, Changes in body weight in mice treated with Q+D or vehicle and/or exposed to ?? weeks of cigarette smoke compared to normal air (n=8 per group). **c**, Inflammatory cell counts in bronchoalveolar lavage (BAL) fluid. **d**, Quantitative analysis of alveolar diameter to evaluate emphysematous changes. **e**, Assessment of lung function parameters, including forced vital capacity (FVC), total lung capacity (TLC), inspiratory capacity (IC), and static lung compliance (Cst). Statistical analysis was performed using two-way ANOVA with Sidak's multiple comparisons test.

Elimination of senescent cells in the lungs by oral treatment with senolytics

SA-β-gal is widely recognised as the standard marker for identifying senescent cells, as its activity is strongly associated with the accumulation of senescence-related lysosomal β-galactosidase. To evaluate the presence of senescent cells, we stained lung tissues with SA-β-gal and quantified the number of positively stained cells. Our results confirmed a substantial accumulation of senescent cells in the lungs of mice with experimental COPD (Fig. 5.2a, b), consistent with findings from clinical studies in humans with COPD[162-164]. This accumulation highlights an association of cellular senescence with COPD, particularly in driving chronic inflammation and tissue damage. Treatment with senolytics via oral administration substantially reduced SA-β-gal-positive cells in the lungs (Fig. 5.2a, b), indicating the effective clearance of senescent cells. The reduction in senescent cells aligns with improvements in immune cell infiltration and pulmonary structural integrity, further supporting the therapeutic potential of senolytic treatment in COPD.

Next, we investigated the effects of senolytic treatment on the expression of senescence-related genes in the lungs. First, we analysed the expression of anti-apoptotic BCL family genes, including *Bcl-xl* and *Bcl-w*, both of which were elevated in experimental COPD (Fig. 5.2c). Weekly senolytic administration effectively reduced the expression of these genes, likely by promoting apoptosis in senescent cells. We then examined the expression of the senescence-related gene *p21* and SASP genes *Pai1* and *Mmp12*, which were significantly upregulated in experimental COPD. Weekly senolytic treatment notably downregulated these

genes, indicating a reduction in senescence-associated signalling pathways (Fig. 5.2d). Additionally, we assessed the expression of the anti-ageing gene *Sirt1*, which was reduced in the lungs in experimental COPD, reflecting accelerated ageing associated with the disease. Bi-weekly administration of Q+D restored *Sirt1* expression, suggesting a potential role in mitigating COPD-related ageing processes. Although senolytics did not significantly improve lung function in COPD, their ability to reduce immune cell infiltration, enhance lung structure and eliminate senescent cells via oral treatment highlights their potential as an adjuvant therapy for COPD.

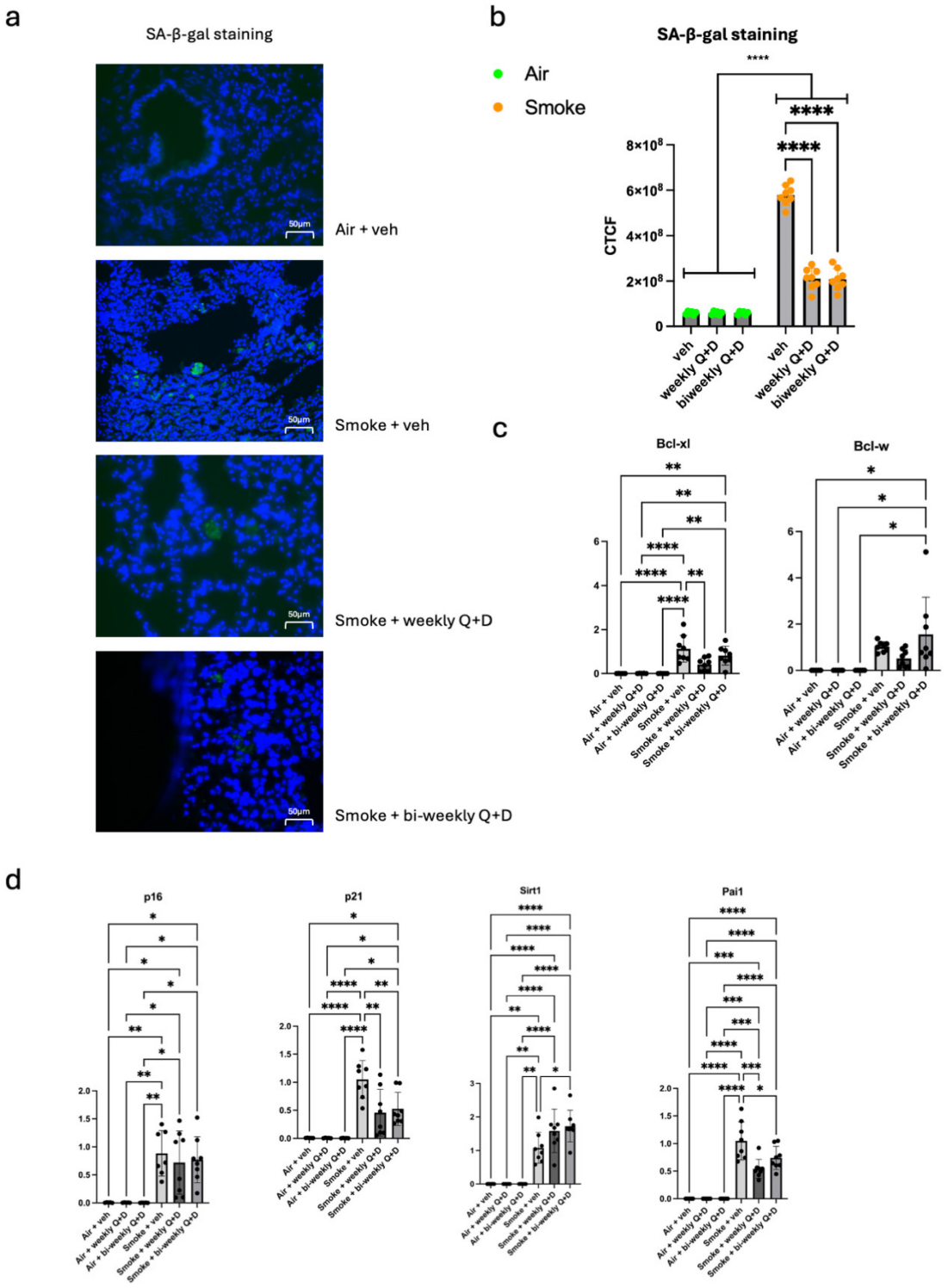


Fig. 5.2: Intrapulmonary effects of oral treatment with Q+D senolytics on senescence.

a, Representative histological images of SA-β-gal staining in lung tissues from experimental COPD and controls. Bright green, senescent cells. **b**, Quantification of SA-β-gal staining presented as corrected total cell fluorescence (CTCF). **c, d**, Expression levels of anti-apoptotic genes (*Bcl-xl* and *Bcl-w*) (**c**), senescence-related gene (*p21*), senescence-associated secretory

phenotype (SASP) genes (*Pai1* and *Mmp12*), and the anti-ageing gene (*Sirt1*) (**d**). Statistical analysis were evaluated using one-way ANOVA with Dunnett's multiple comparisons test.

5.4 Discussion

This observational study provides new insights into the importance of senescence and the therapeutic potential of senolytics in COPD, a condition characterised by chronic inflammation, alveolar destruction and impaired lung function. Although Q+D senolytic treatment did not improve lung function, it demonstrated substantial efficacy in targeting senescent cells, reducing inflammatory cell infiltration, and enhancing lung structure, underscoring its potential as a therapy for COPD.

Cellular senescence is increasingly recognised as a key driver of COPD pathogenesis, contributing to persistent inflammation and tissue remodelling through the SASP. Our findings confirmed the accumulation of senescent cells in the lungs in experimental COPD, consistent with clinical observations in humans[162-164]. The senolytic cocktail Q+D, delivered as an oral treatment, effectively reduced the number of SA-β-gal-positive cells, highlighting their ability to seliminate senescent cells. Clearance was associated with reductions in inflammatory cell infiltration to the lungs as evidenced by decreased macrophages, neutrophils, and lymphocytes in BAL fluid, and improved alveolar integrity and reduced emphysema. These finding are in line with previous studies demonstrating the role of senolytics in mitigating inflammatory cell recruitment and improving lung homeostasis in other situations[165,166].

Although Q+D treatment partially reversed emphysematous changes, it failed to improve lung function parameters. The slight worsening of Cst observed with bi-weekly dosing may reflect potential off-target effects, such as pulmonary hypertension, which warrants further investigation. These findings underscore the complexity of COPD pathophysiology, wherein structural improvements alone may be insufficient to restore physiological function. Senolytic treatment routes and regimes may need to be optimised and/or combined with other therapeutic modalities to maximise clinical benefits. Collectively, senolytics administration

might be most beneficial for patients at a mild stage of COPD (GOLD Stage 1-2) to potentially halt disease progression. However, future studies need to define the dosage and administration frequency clinically.

In conclusion, our study highlights the importance of senescence and the potential of Q+D as a therapy for COPD, with its ability to selectively eliminate senescent cells, mitigate inflammation, and preserve lung structure. Further pre-clinical and clinical studies are needed to refine dosing strategies, explore combination therapies, and assess long-term outcomes. This study represents a step toward understanding the role of senolytics in addressing the cellular and molecular drivers of COPD pathogenesis.

Chapter 6: General Discussion

6.1 Overview

This thesis investigates the complex pathogenesis of COPD through the lens of ST, scRNA-seq, and other advanced omics techniques. The overarching aim was to elucidate the cellular, spatial, and molecular mechanisms driving COPD progression, particularly in response to cigarette smoke exposure. COPD is a highly heterogeneous disease characterised by airway inflammation and chronic bronchitis, systemic immune dysregulation and lung parenchymal destruction and emphysema, with cigarette smoking as the current leading risk factor. Despite the availability of pharmacological and non-pharmacological interventions, effective treatments targeting disease pathogenesis remain limited and are poorly effective.

The integration of ST and scRNA-seq enabled the profiling of spatially resolved gene expression, uncovering novel insights into the interplay of immune cells, structural lung changes, and senescence-driven inflammation. Each chapter 2-5 contributes unique perspectives on these aspects, from spatial immune dynamics (Chapter 2), spatiotemporal evolution of lymphoid follicles (Chapter 3), epigenetic regulation in airway epithelial cells (Chapter 4), to the therapeutic potential of senolytics in addressing senescence-related pathologies in COPD (Chapter 5). Together, these findings pave the way for precision medicine approaches targeting COPD pathogenesis.

6.2 Spatial Immune Dynamics in COPD Pathogenesis

Chapter 2 emphasised the critical role of the lung periphery in COPD pathogenesis. AlvMs were shown to migrate from the vasculature to the distal lung parenchyma during chronic cigarette smoke exposure, forming inflammation-enriched niches near the edge of the lungs. This study expands the conventional airway-centric view of COPD.

Spatial profiling revealed specific transcriptional changes, such as upregulated *Lcn2* and downregulated *Klf2*, implicating microvascular dysfunction as a driver of COPD. The potential

of *Lcn2* as a therapeutic target is supported by preclinical studies demonstrating the efficacy of monoclonal antibodies against *Lcn2* in limiting tumour metastases to the lungs[92]. However, they have not yet been applied to COPD. These findings underscore the translational value of our ST study in identifying actionable therapeutic targets potentially in the form of biologics.

Inter-cellular communication highlighted that impaired angiogenesis signals and upregulated inflammatory pathways, particularly *Anxa1*-mediated cytokine signalling, were elevated in experimental COPD. The peptide AC2-26, mimicking *Anxa1*, demonstrated therapeutic potential in reducing inflammation and tissue damage in COPD animal models[93]. These insights underscore the importance of targeting the pulmonary microvasculature and inflammatory crosstalk to mor holistically address COPD pathogenesis.

6.3 Spatiotemporal Evolution of Lymphoid Follicles

Chapter 3 presented a spatiotemporal trajectory of lymphoid follicle formation in cigarette smoke-induced experimental COPD. Lymphoid follicles were observed after as early as four weeks of cigarette smoke exposure, preceding the onset of emphysema. The occurrence of lymphoid follicles before emphysema onset highlights their roles and potential as early markers of adaptive immune responses in the development of COPD.

This study uncovered a shift in lymphoid follicle composition over time, characterised by increased macrophage trafficking and detachment from lung structures such as vascular and alveolar walls, attributed to emphysema-induced architectural disruption. Key genes, including *Enpp2* (pro-autoimmune) and *Mfge8* (anti-autoimmune), revealed a dynamic balance between autoimmune priming and resolution within lymphoid follicles. This duality underscores the complexity of adaptive immune responses in COPD.

The systemic origins of follicular B cells were traced from pre-B cells in the bone marrow to mature B cells in the lungs, with TFs *Foxp1* and *Xbp1* playing pivotal roles in their development[107,110]. Novel regulators, such as *Hmgb1* and *Mxd3*, linked B cells to chronic inflammation and apoptotic resistance[111,112], suggesting the potential of targeted

pharmacotherapeutic strategies to disrupt lymphoid follicle formation and mitigate autoimmune priming in COPD.

6.4 Epigenetic Adaptation of Airway Epithelial Cells

Chapter 4 highlighted the epigenetic regulation of airway epithelial cells in response to cigarette smoke exposure. We first profiled the "smoking signature" that differentiates between current smokers and non-smokers across multiple cohorts. The affected site by smoking is predominately the smokers' airway surface epithelial layer found using the smoking signature. We then compared the gene differences of the lungs of human and non-human primates and observed that the gene differences are mainly in human smokers' airway surface epithelial cells, suggesting an evolutionary adaptation to smoke exposure in human airway.

The study identified key TFs (*NRF2*, *AhR*) driving oxidative stress responses, with hypomethylation of their binding sites correlating with increased expression of detoxification genes, such as *ALDH3A1* and *NQO1*. These genes, which are predominantly expressed by luminal epithelial cells, have protective effects against smoke-induced epithelial damage[158]. The findings align with prior studies showing essential roles for *NRF2* and *AhR* in mitigating smoke-induced pulmonary inflammation and emphysema[155,156]. The localisation of detoxification responses to the airway surface layer highlights the importance of modelling smoke exposure in differentiated epithelial systems, as basal cells showed limited direct responses. This specificity underscores the need for refined experimental models that mimic the hierarchical organisation of the airway epithelium.

However, this study has certain limitations, primarily the use of the alveolar epithelial cell line A549 instead of a true airway epithelial model. While A549 cells provide valuable mechanistic insights, they do not fully recapitulate the cellular diversity and differentiation hierarchy of the human airway. In future studies, the use of human small airway epithelial cells cultured under ALI conditions will enable a more physiologically relevant model that better mimics the structure and function of the airway epithelium. This approach will allow for more accurate

investigation of smoke-induced epigenetic and transcriptional adaptations in airway epithelial cells.

Furthermore, while the findings raise the hypothesis that certain airway-enriched genes may have undergone adaptive changes in response to environmental exposures such as cigarette smoke, this remains speculative. Our study did not include phylogenetic or selective pressure analyses to directly evaluate evolutionary adaptation. Future comparative genomic studies, ideally controlling for smoke exposure across species, would be required to substantiate this hypothesis. However, conducting such experiments in non-human primates presents considerable ethical and logistical challenges. In this study, we instead provided preliminary insights by comparing transcriptomes between human non-smokers and non–smoke-exposed primates.

6.5 Therapeutic Potential of Senolytics in COPD

Chapter 5 explored the efficacy of the senolytic cocktail Q+D in targeting senescence-driven inflammation in COPD. Senescent cells, identified by SA-β-gal staining, accumulated in experimental COPD lungs, consistent with clinical observations[162-164]. Oral treatment with Q+D effectively eliminated these cells, reduced immune cell infiltration, and partially restored alveolar integrity suppressing the development of emphysema. These findings align with studies demonstrating the anti-inflammatory effects of senolytics in fibrotic lung disease[165,166].

Despite these structural improvements, Q+D did not prevent lung function decline, and exacerbated Cst at bi-weekly dosing, potentially due to off-target effects like pulmonary hypertension[168]. These pre-clinical observations of senolytics in COPD underscore the need for optimised dosing strategies and combination therapies to achieve broader clinical benefits. They also demonstrate the potential of senolytics as therapeutics in COPD.

6.6 Integration and Implications

All the studies in this thesis highlight the interplay between immune dysregulation, structural lung changes, and cellular senescence in COPD pathogenesis. The integration of ST and scRNA-seq provides unprecedented insights into spatially resolved gene expression, revealing critical processes such as macrophage migration and lymphoid follicle evolution in experimental COPD in animal model. These findings were further validated in clinical COPD in humans, exemplified by the observed migration pattern of AlvMs from the vasculature to the distal parenchyma in COPD lungs.

These findings have significant implications for therapeutic development, advocating for a shift toward precision medicine approaches that target peripheral lung regions, immune crosstalk, and senescence-driven inflammation. Furthermore, the identification of actionable targets, such as *Lcn2*, *Enpp2*, and senescent cells, underscores the translational potential of this work.

6.7 Future Directions

Future studies should focus on leveraging advanced spatial profiling tools to refine our understanding of macrophage migration and its interactions with the pulmonary microvasculature. Exploring the molecular drivers of microvascular dysfunction, such as *Lcn2* and *Klf2*, in human COPD cohorts could pave the way for targeted therapies. Functional validation of these targets using genetically modified animal models or CRISPR-based editing in organoid systems will be essential. Additionally, the development of biologics or small molecules targeting *Lcn2* and *Anxa1* signalling pathways offers a promising therapeutic avenue. Investigating the long-term effects of these interventions in pre-clinical experiments and pilot clinical trials would further inform their translational potential.

To better understand the adaptive immune responses in COPD, future research should employ longitudinal sampling of lymphoid follicles in human COPD patients. Single-cell transcriptomics and single-cell ST could provide deeper insights into the dynamic balance between autoimmune priming and resolution within these follicles. Additionally, therapeutic

strategies aimed at modulating follicular B cell activity, such as inhibitors of *Hmgb1* and *Mxd3*, should be explored in pre-clinical studies. Combining these approaches with advanced imaging techniques, such as intravital microscopy, could elucidate the real-time interactions between lymphoid follicles and their surrounding lung microenvironment.

Future directions in the epigenetic area of smoking can aim to unravel the epigenetic landscape of airway epithelial cells in more diverse smoking conditions, including passive and occupational exposure, apart from cigarette smoking. Advanced epigenome-editing techniques, such as dCas9-TET/CRISPR systems, could be employed to experimentally validate the functional role of specific epigenetic modifications in genes like *ALDH3A1* and *NQO1*. Moreover, creating airway organoids or ALI *in vitro* models that mimic human airway architecture would allow for high-throughput drug screening and precise modelling of smoke-induced damage. Comparative studies across diverse human populations and primates could also reveal evolutionary adaptations that contribute to differences in susceptibility to smoke-induced COPD.

Future work about cellular senescence should focus on optimising dosing regimens and combinations of senolytic therapies to minimise off-target effects, such as pulmonary hypertension, while maximising therapeutic efficacy. The use of patient-derived lung organoids and our experimental COPD animal models would provide a more accurate representation of human COPD and allow for personalised medicine approaches. Combining senolytics with anti-inflammatory drugs or agents targeting structural repair mechanisms could offer synergistic benefits. Additionally, large-scale clinical trials are needed to assess the safety, efficacy, and long-term benefits of senolytic therapies in COPD patients, particularly in addressing structural and functional impairments.

6.8 Conclusions

This thesis provides a comprehensive investigation of the complex mechanisms underlying COPD pathogenesis, employing cutting-edge spatial and molecular technologies to address key gaps in our understanding. By integrating ST and scRNA-seq, this work highlights the

critical roles of immune dysregulation, structural lung remodelling, and cellular senescence in COPD.

Chapter 2 advanced our understanding of immune dynamics by demonstrating macrophage migration from the vasculature to the lung periphery, forming inflammation-enriched niches and implicating microvascular dysfunction in disease progression. These findings expand upon the traditional airway-centric focus of COPD pathogenesis and highlight the importance of targeting peripheral lung regions for therapeutic intervention.

Chapter 3 revealed the spatiotemporal evolution of lymphoid follicles in response to cigarette smoke exposure, identifying them as early markers of adaptive immune responses and key players in COPD progression. The discovery of novel regulators such as *Hmgb1* and *Mxd3* provides a foundation for therapeutic strategies aimed at disrupting autoimmune priming.

Chapter 4 explored the epigenetic adaptation of airway epithelial cells to smoke exposure, identifying a "smoking signature" gene set and key transcription factors *NRF2* and *AhR* that mediate oxidative stress responses. The findings suggest an evolutionary adaptation of human airways to prolonged smoke exposure and emphasise the importance of modelling epithelial responses in differentiated systems.

Chapter 5 investigated the therapeutic potential of senolytics in targeting senescence-driven inflammation, demonstrating their ability to clear senescent cells, reduce immune infiltration, and partially restore alveolar integrity. While promising, the findings also highlight the need for optimised dosing and combination therapies to fully address structural and functional impairments in COPD.

Together, these studies pave the way for precision medicine approaches that target the diverse and dynamic mechanisms driving COPD pathogenesis. By identifying actionable therapeutic targets such as *Lcn2*, *Enpp2*, and senescent cells, this thesis bridges the gap between basic research and clinical translation. The integration of advanced omics techniques with functional assays provides a robust framework for future investigations, laying the groundwork for novel therapies that address the unmet needs of COPD patients.

This thesis represents a significant contribution to the field of respiratory disease research and underscores the potential of spatially resolved and molecularly informed strategies to transform the diagnosis and treatment of COPD.

References

1 Agusti, A. *et al.* Global Initiative for Chronic Obstructive Lung Disease 2023 Report: GOLD Executive Summary. *Eur Respir J* **61** (2023). https://doi.org:10.1183/13993003.00239-2023

2 Adeloye, D. *et al.* Global, regional, and national prevalence of, and risk factors for, chronic obstructive pulmonary disease (COPD) in 2019: a systematic review and modelling analysis. *Lancet Respir Med* **10**, 447-458 (2022). https://doi.org:10.1016/S2213-2600(21)00511-7

3 Health, U. D. o. & Services, H. The health consequences of smoking chronic obstructive lung desease: a report of the Surgeon General. (1984).

4 Rennard, S. I. & Vestbo, J. COPD: the dangerous underestimate of 15%. *Lancet* **367**, 1216-1219 (2006). https://doi.org:10.1016/S0140-6736(06)68516-4

5 Kohansal, R. *et al.* The natural history of chronic airflow obstruction revisited: an analysis of the Framingham offspring cohort. *Am J Respir Crit Care Med* **180**, 3-10 (2009). https://doi.org:10.1164/rccm.200901-0047OC

6 Montes de Oca, M. Smoking Cessation/Vaccinations. *Clin Chest Med* **41**, 495-512 (2020). https://doi.org:10.1016/j.ccm.2020.06.013

7 Kang, M. J. *et al.* Cigarette smoke selectively enhances viral PAMP- and virus-induced pulmonary innate immune and remodeling responses in mice. *J Clin Invest* **118**, 2771-2784 (2008). https://doi.org:10.1172/JCI32709

8 Barnes, P. J. Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. *J Allergy Clin Immunol* **138**, 16-27 (2016). https://doi.org:10.1016/j.jaci.2016.05.011

9 Byrne, A. J., Mathie, S. A., Gregory, L. G. & Lloyd, C. M. Pulmonary macrophages: key players in the innate defence of the airways. *Thorax* **70**, 1189-1196 (2015). https://doi.org:10.1136/thoraxjnl-2015-207020

10 O'Beirne, S. L. *et al.* Alveolar Macrophage Immunometabolism and Lung Function Impairment in Smoking and Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **201**, 735-739 (2020). https://doi.org:10.1164/rccm.201908-1683LE

11 Bassler, K. *et al.* Alveolar macrophages in early stage COPD show functional deviations with properties of impaired immune activation. *Front Immunol* **13**, 917232 (2022). https://doi.org:10.3389/fimmu.2022.917232

12 Hogg, J. C. Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *Lancet* **364**, 709-721 (2004). https://doi.org:10.1016/S0140-6736(04)16900-6

13 Koo, H. K. *et al.* Small airways disease in mild and moderate chronic obstructive pulmonary disease: a cross-sectional study. *Lancet Respir Med* **6**, 591-602 (2018). https://doi.org:10.1016/S2213-2600(18)30196-6

14 McDonough, J. E. *et al.* Small-airway obstruction and emphysema in chronic obstructive pulmonary disease. *N Engl J Med* **365**, 1567-1575 (2011). https://doi.org:10.1056/NEJMoa1106955

15 Barnes, P. J. Chronic obstructive pulmonary disease. *N Engl J Med* **343**, 269-280 (2000). https://doi.org:10.1056/NEJM200007273430407

16 Tuder, R. M., Petrache, I., Elias, J. A., Voelkel, N. F. & Henson, P. M. Apoptosis and emphysema: the missing link. *Am J Respir Cell Mol Biol* **28**, 551-554 (2003). https://doi.org:10.1165/rcmb.F269

17 Lu, Z. *et al.* Necroptosis Signaling Promotes Inflammation, Airway Remodeling, and Emphysema in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **204**, 667-681 (2021). https://doi.org:10.1164/rccm.202009-3442OC
18 Tuder, R. M., Yoshida, T., Fijalkowka, I., Biswal, S. & Petrache, I. Role of lung maintenance program in the heterogeneity of lung destruction in emphysema. *Proc Am Thorac Soc* **3**, 673-679 (2006). https://doi.org:10.1513/pats.200605-124SF
19 Coxson, H. O. *et al.* Early emphysema in patients with anorexia nervosa. *Am J Respir Crit Care Med* **170**, 748-752 (2004). https://doi.org:10.1164/rccm.200405-651OC
20 Sharafkhaneh, A., Hanania, N. A. & Kim, V. Pathogenesis of emphysema: from the bench to the bedside. *Proc Am Thorac Soc* **5**, 475-477 (2008). https://doi.org:10.1513/pats.200708-126ET
21 Taraseviciene-Stewart, L. *et al.* Is alveolar destruction and emphysema in chronic obstructive pulmonary disease an immune disease? *Proc Am Thorac Soc* **3**, 687-690 (2006). https://doi.org:10.1513/pats.200605-105SF
22 Chaouat, A., Naeije, R. & Weitzenblum, E. Pulmonary hypertension in COPD. *Eur Respir J* **32**, 1371-1385 (2008). https://doi.org:10.1183/09031936.00015608
23 Peinado, V. I. *et al.* Inflammatory reaction in pulmonary muscular arteries of patients with mild chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **159**, 1605-1611 (1999). https://doi.org:10.1164/ajrccm.159.5.9807059
24 Jones, B. *et al.* Animal models of COPD: What do they tell us? *Respirology* **22**, 21-32 (2017). https://doi.org:10.1111/resp.12908
25 Method of the year 2013. *Nat Methods* **11**, 1 (2014). https://doi.org:10.1038/nmeth.2801
26 Jovic, D. *et al.* Single-cell RNA sequencing technologies and applications: A brief overview. *Clin Transl Med* **12**, e694 (2022). https://doi.org:10.1002/ctm2.694
27 Zappia, L. & Theis, F. J. Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape. *Genome Biol* **22**, 301 (2021). https://doi.org:10.1186/s13059-021-02519-4
28 Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319-324 (2018). https://doi.org:10.1038/s41586-018-0393-7
29 Plasschaert, L. W. *et al.* A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377-381 (2018). https://doi.org:10.1038/s41586-018-0394-6
30 Johansen, M. D. *et al.* Increased SARS-CoV-2 Infection, Protease, and Inflammatory Responses in Chronic Obstructive Pulmonary Disease Primary Bronchial Epithelial Cells Defined with Single-Cell RNA Sequencing. *Am J Respir Crit Care Med* **206**, 712-729 (2022). https://doi.org:10.1164/rccm.202108-1901OC
31 Marx, V. Method of the Year: spatially resolved transcriptomics. *Nat Methods* **18**, 9-14 (2021). https://doi.org:10.1038/s41592-020-01033-y
32 Madissoon, E. *et al.* A spatially resolved atlas of the human lung characterizes a gland-associated immune niche. *Nat Genet* **55**, 66-77 (2023). https://doi.org:10.1038/s41588-022-01243-4
33 Mayr, C. H. *et al.* Spatial transcriptomic characterization of pathologic niches in IPF. *Sci Adv* **10**, eadl5473 (2024). https://doi.org:10.1126/sciadv.adl5473

34 Rojas-Quintero, J. *et al*. Spatial Transcriptomics Resolve an Emphysema-Specific Lymphoid Follicle B Cell Signature in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **209**, 48-58 (2024). https://doi.org:10.1164/rccm.202303-0507LE

35 Schyns, J. *et al*. Non-classical tissue monocytes and two functionally distinct populations of interstitial macrophages populate the mouse lung. *Nat Commun* **10**, 3964 (2019). https://doi.org:10.1038/s41467-019-11843-0

36 Sauler, M. *et al*. Characterization of the COPD alveolar niche using single-cell RNA sequencing. *Nat Commun* **13**, 494 (2022). https://doi.org:10.1038/s41467-022-28062-9

37 Adams, T. S. *et al*. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* **6**, eaba1983 (2020). https://doi.org:10.1126/sciadv.aba1983

38 Huang, Q. *et al*. Single-cell transcriptomics highlights immunological dysregulations of monocytes in the pathobiology of COPD. *Respir Res* **23**, 367 (2022). https://doi.org:10.1186/s12931-022-02293-2

39 Watanabe, N. *et al*. Anomalous Epithelial Variations and Ectopic Inflammatory Response in Chronic Obstructive Pulmonary Disease. *Am J Respir Cell Mol Biol* **67**, 708-719 (2022). https://doi.org:10.1165/rcmb.2021-0555OC

40 Basil, M. C. *et al*. Human distal airways contain a multipotent secretory cell that can regenerate alveoli. *Nature* **604**, 120-126 (2022). https://doi.org:10.1038/s41586-022-04552-0

41 Li, X. *et al*. Single cell RNA sequencing identifies IGFBP5 and QKI as ciliated epithelial cell genes associated with severe COPD. *Respir Res* **22**, 100 (2021). https://doi.org:10.1186/s12931-021-01675-2

42 Beckett, E. L. *et al*. A new short-term mouse model of chronic obstructive pulmonary disease identifies a role for mast cell tryptase in pathogenesis. *J Allergy Clin Immunol* **131**, 752-762 (2013). https://doi.org:10.1016/j.jaci.2012.11.053

43 Hansbro, P. M. *et al*. Importance of mast cell Prss31/transmembrane tryptase/tryptase-gamma in lung function and experimental chronic obstructive pulmonary disease and colitis. *J Biol Chem* **289**, 18214-18227 (2014). https://doi.org:10.1074/jbc.M114.548594

44 Hsu, A. C. *et al*. Targeting PI3K-p110alpha Suppresses Influenza Virus Infection in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **191**, 1012-1023 (2015). https://doi.org:10.1164/rccm.201501-0188OC

45 Fricker, M. *et al*. Chronic cigarette smoke exposure induces systemic hypoxia that drives intestinal dysfunction. *JCI Insight* **3** (2018). https://doi.org:10.1172/jci.insight.94040

46 Haw, T. J. *et al*. A pathogenic role for tumor necrosis factor-related apoptosis-inducing ligand in chronic obstructive pulmonary disease. *Mucosal Immunol* **9**, 859-872 (2016). https://doi.org:10.1038/mi.2015.111

47 Liu, G. *et al*. Fibulin-1 regulates the pathogenesis of tissue remodeling in respiratory diseases. *JCI Insight* **1** (2016). https://doi.org:10.1172/jci.insight.86380

48 Hsu, A. C. *et al*. MicroRNA-125a and -b inhibit A20 and MAVS to promote inflammation and impair antiviral response in COPD. *JCI Insight* **2**, e90443 (2017). https://doi.org:10.1172/jci.insight.90443

49 Van Eeckhoutte, H. P. *et al*. RIPK1 kinase-dependent inflammation and cell death contribute to the pathogenesis of COPD. *Eur Respir J* **61** (2023). https://doi.org:10.1183/13993003.01506-2022

50 Starkey, M. R. *et al*. IL-22 and its receptors are increased in human and experimental COPD and contribute to pathogenesis. *Eur Respir J* **54** (2019). https://doi.org:10.1183/13993003.00174-2018

51 Liu, G. *et al*. Adverse roles of mast cell chymase-1 in COPD. *Eur Respir J* **60** (2022). https://doi.org:10.1183/13993003.01431-2021

52 Kim, R. Y. *et al*. A microRNA-21-mediated SATB1/S100A9/NF-kappaB axis promotes chronic obstructive pulmonary disease pathogenesis. *Sci Transl Med* **13**, eaav7223 (2021). https://doi.org:10.1126/scitranslmed.aav7223

53 Cooper, G. E. *et al*. Antiviral Responses of Tissue-resident CD49a(+) Lung Natural Killer Cells Are Dysregulated in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **207**, 553-565 (2023). https://doi.org:10.1164/rccm.202205-0848OC

54 Hao, Y. *et al*. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021). https://doi.org:10.1016/j.cell.2021.04.048

55 Stuart, T. *et al*. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019). https://doi.org:10.1016/j.cell.2019.05.031

56 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018). https://doi.org:10.1038/nbt.4096

57 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502 (2015). https://doi.org:10.1038/nbt.3192

58 Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 296 (2019). https://doi.org:10.1186/s13059-019-1874-1

59 Shao, X. *et al*. Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nat Commun* **13**, 4429 (2022). https://doi.org:10.1038/s41467-022-32111-8

60 Cable, D. M. *et al*. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat Biotechnol* **40**, 517-526 (2022). https://doi.org:10.1038/s41587-021-00830-w

61 Moss, A. *et al*. A single cell transcriptomics map of paracrine networks in the intrinsic cardiac nervous system. *iScience* **24**, 102713 (2021). https://doi.org:10.1016/j.isci.2021.102713

62 Zhao, E. *et al*. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol* **39**, 1375-1384 (2021). https://doi.org:10.1038/s41587-021-00935-2

63 Longo, S. K., Guo, M. G., Ji, A. L. & Khavari, P. A. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat Rev Genet* **22**, 627-644 (2021). https://doi.org:10.1038/s41576-021-00370-8

64 Tanevski, J., Flores, R. O. R., Gabor, A., Schapiro, D. & Saez-Rodriguez, J. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. *Genome Biol* **23**, 97 (2022). https://doi.org:10.1186/s13059-022-02663-5

65 Schubert, M. *et al*. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun* **9**, 20 (2018). https://doi.org:10.1038/s41467-017-02391-6

66 Holland, C. H., Szalai, B. & Saez-Rodriguez, J. Transfer of regulatory knowledge from human to mouse for functional genomics analysis. *Biochim Biophys Acta Gene Regul Mech* **1863**, 194431 (2020). https://doi.org:10.1016/j.bbagrm.2019.194431

67 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). https://doi.org:10.1038/s41592-021-01255-8

68 Ning, W. *et al*. Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease. *Proc Natl Acad Sci U S A* **101**, 14895-14900 (2004). https://doi.org:10.1073/pnas.0401168101

69 Korsunsky, I. *et al*. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296 (2019). https://doi.org:10.1038/s41592-019-0619-0

70 Lun, A. T., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122 (2016). https://doi.org:10.12688/f1000research.9501.2

71 Shao, X. *et al*. CellTalkDB: a manually curated database of ligand-receptor interactions in humans and mice. *Brief Bioinform* **22** (2021). https://doi.org:10.1093/bib/bbaa269

72 Jin, S. *et al*. Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021). https://doi.org:10.1038/s41467-021-21246-9

73 Thorley, A. J. & Tetley, T. D. Pulmonary epithelium, cigarette smoke, and chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* **2**, 409-428 (2007).

74 Brusselle, G. G., Joos, G. F. & Bracke, K. R. New insights into the immunology of chronic obstructive pulmonary disease. *Lancet* **378**, 1015-1026 (2011). https://doi.org:10.1016/S0140-6736(11)60988-4

75 Vassallo, R. *et al*. Cigarette smoke promotes dendritic cell accumulation in COPD; a Lung Tissue Research Consortium study. *Respir Res* **11**, 45 (2010). https://doi.org:10.1186/1465-9921-11-45

76 Shapiro, S. D. The macrophage in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **160**, S29-32 (1999). https://doi.org:10.1164/ajrccm.160.supplement_1.9

77 Treekitkarnmongkol, W. *et al*. Augmented Lipocalin-2 Is Associated with Chronic Obstructive Pulmonary Disease and Counteracts Lung Adenocarcinoma Development. *Am J Respir Crit Care Med* **203**, 90-101 (2021). https://doi.org:10.1164/rccm.202004-1079OC

78 Steiling, K. *et al*. A dynamic bronchial airway gene expression signature of chronic obstructive pulmonary disease and lung function impairment. *Am J Respir Crit Care Med* **187**, 933-942 (2013). https://doi.org:10.1164/rccm.201208-1449OC

79 Huang, R. T. *et al*. Experimental Lung Injury Reduces Kruppel-like Factor 2 to Increase Endothelial Permeability via Regulation of RAPGEF3-Rac1 Signaling. *Am J Respir Crit Care Med* **195**, 639-651 (2017). https://doi.org:10.1164/rccm.201604-0668OC

80 Liu, Y. *et al.* Single-Cell Profiling Reveals Divergent, Globally Patterned Immune Responses in Murine Skin Inflammation. *iScience* **23**, 101582 (2020). https://doi.org:10.1016/j.isci.2020.101582
81 Kimmel, J. C., Hwang, A. B., Scaramozza, A., Marshall, W. F. & Brack, A. S. Aging induces aberrant state transition kinetics in murine muscle stem cells. *Development* **147** (2020). https://doi.org:10.1242/dev.183855
82 Hammond, T. R. *et al.* Single-Cell RNA Sequencing of Microglia throughout the Mouse Lifespan and in the Injured Brain Reveals Complex Cell-State Changes. *Immunity* **50**, 253-271 e256 (2019). https://doi.org:10.1016/j.immuni.2018.11.004
83 Zhang, P., Cao, L., Zhou, R., Yang, X. & Wu, M. The lncRNA Neat1 promotes activation of inflammasomes in macrophages. *Nat Commun* **10**, 1495 (2019). https://doi.org:10.1038/s41467-019-09482-6
84 Friedl, P. & Wolf, K. Plasticity of cell migration: a multiscale tuning model. *J Cell Biol* **188**, 11-19 (2010). https://doi.org:10.1083/jcb.200909003
85 Lammermann, T. *et al.* Rapid leukocyte migration by integrin-independent flowing and squeezing. *Nature* **453**, 51-55 (2008). https://doi.org:10.1038/nature06887
86 Sun, Z., Costell, M. & Fassler, R. Integrin activation by talin, kindlin and mechanical forces. *Nat Cell Biol* **21**, 25-31 (2019). https://doi.org:10.1038/s41556-018-0234-9
87 Timmerman, I., Daniel, A. E., Kroon, J. & van Buul, J. D. Leukocytes Crossing the Endothelium: A Matter of Communication. *Int Rev Cell Mol Biol* **322**, 281-329 (2016). https://doi.org:10.1016/bs.ircmb.2015.10.005
88 Maroni, D. & Davis, J. S. TGFB1 disrupts the angiogenic potential of microvascular endothelial cells of the corpus luteum. *J Cell Sci* **124**, 2501-2510 (2011). https://doi.org:10.1242/jcs.084558
89 Hogg, J. C. *et al.* The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N Engl J Med* **350**, 2645-2653 (2004). https://doi.org:10.1056/NEJMoa032158
90 Ritchie, A. I. *et al.* Structural Predictors of Lung Function Decline in Young Smokers with Normal Spirometry. *Am J Respir Crit Care Med* **209**, 1208-1218 (2024). https://doi.org:10.1164/rccm.202307-1203OC
91 Yang, C. X. *et al.* Lung Spatial Profiling Reveals a T Cell Signature in COPD Patients with Fatal SARS-CoV-2 Infection. *Cells* **11** (2022). https://doi.org:10.3390/cells11121864
92 Leng, X. *et al.* Inhibition of lipocalin 2 impairs breast tumorigenesis and metastasis. *Cancer Res* **69**, 8579-8584 (2009). https://doi.org:10.1158/0008-5472.CAN-09-1934
93 Possebon, L. *et al.* Mimetic peptide AC2-26 of annexin A1 as a potential therapeutic agent to treat COPD. *Int Immunopharmacol* **63**, 270-281 (2018). https://doi.org:10.1016/j.intimp.2018.08.011
94 Nana-Sinkam, S. P. *et al.* Prostacyclin prevents pulmonary endothelial cell apoptosis induced by cigarette smoke. *Am J Respir Crit Care Med* **175**, 676-685 (2007). https://doi.org:10.1164/rccm.200605-724OC
95 Minakata, Y. *et al.* Microvascular hyperpermeability in COPD airways. *Thorax* **60**, 882 (2005). https://doi.org:10.1136/thx.2005.045765

96 Lambrecht, B. N. & Hammad, H. The role of dendritic and epithelial cells as master regulators of allergic airway inflammation. *Lancet* **376**, 835-843 (2010). https://doi.org:10.1016/S0140-6736(10)61226-3

97 van der Strate, B. W. *et al.* Cigarette smoke-induced emphysema: A role for the B cell? *Am J Respir Crit Care Med* **173**, 751-758 (2006). https://doi.org:10.1164/rccm.200504-594OC

98 Brusselle, G. G., Demoor, T., Bracke, K. R., Brandsma, C. A. & Timens, W. Lymphoid follicles in (very) severe COPD: beneficial or harmful? *Eur Respir J* **34**, 219-230 (2009). https://doi.org:10.1183/09031936.00150208

99 Danaher, P. *et al.* Advances in mixed cell deconvolution enable quantification of cell types in spatial transcriptomic data. *Nat Commun* **13**, 385 (2022). https://doi.org:10.1038/s41467-022-28020-5

100 Kapellos, T. S. *et al.* Systemic alterations in neutrophils and their precursors in early-stage chronic obstructive pulmonary disease. *Cell Rep* **42**, 112525 (2023). https://doi.org:10.1016/j.celrep.2023.112525

101 Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019). https://doi.org:10.1038/s41586-019-0969-x

102 Shen, W. K. *et al.* AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. *Nucleic Acids Res* **51**, D39-D45 (2023). https://doi.org:10.1093/nar/gkac907

103 Polverino, F. *et al.* B Cell-Activating Factor. An Orchestrator of Lymphoid Follicles in Severe Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **192**, 695-705 (2015). https://doi.org:10.1164/rccm.201501-0107OC

104 Magkrioti, C. *et al.* Autotaxin and chronic inflammatory diseases. *J Autoimmun* **104**, 102327 (2019). https://doi.org:10.1016/j.jaut.2019.102327

105 Hanayama, R. *et al.* Autoimmune disease and impaired uptake of apoptotic cells in MFG-E8-deficient mice. *Science* **304**, 1147-1150 (2004). https://doi.org:10.1126/science.1094359

106 Hanayama, R. *et al.* Identification of a factor that links apoptotic cells to phagocytes. *Nature* **417**, 182-187 (2002). https://doi.org:10.1038/417182a

107 Hu, H. *et al.* Foxp1 is an essential transcriptional regulator of B cell development. *Nat Immunol* **7**, 819-826 (2006). https://doi.org:10.1038/ni1358

108 Sagardoy, A. *et al.* Downregulation of FOXP1 is required during germinal center B-cell function. *Blood* **121**, 4311-4320 (2013). https://doi.org:10.1182/blood-2012-10-462846

109 van Keimpema, M. *et al.* The forkhead transcription factor FOXP1 represses human plasma cell differentiation. *Blood* **126**, 2098-2109 (2015). https://doi.org:10.1182/blood-2015-02-626176

110 Iwakoshi, N. N. *et al.* Plasma cell differentiation and the unfolded protein response intersect at the transcription factor XBP-1. *Nat Immunol* **4**, 321-329 (2003). https://doi.org:10.1038/ni907

111 Vettermann, C. *et al.* Proteome profiling suggests a pro-inflammatory role for plasma cells through release of high-mobility group box 1 protein. *Proteomics* **11**, 1228-1237 (2011). https://doi.org:10.1002/pmic.201000491

112 Barisone, G. A. *et al*. Loss of MXD3 induces apoptosis of Reh human precursor B acute lymphoblastic leukemia cells. *Blood Cells Mol Dis* **54**, 329-335 (2015). https://doi.org:10.1016/j.bcmd.2014.12.002

113 Willemse, B. W., Postma, D. S., Timens, W. & ten Hacken, N. H. The impact of smoking cessation on respiratory symptoms, lung function, airway hyperresponsiveness and inflammation. *Eur Respir J* **23**, 464-476 (2004). https://doi.org:10.1183/09031936.04.00012704

114 Imkamp, K. *et al*. Nasal epithelium as a proxy for bronchial epithelium for smoking-induced gene expression and expression Quantitative Trait Loci. *Journal of Allergy and Clinical Immunology* (2018).

115 Billatos, E. *et al*. The Impact of Acute Exposure to Cigarette Smoke on Airway Gene Expression. *Physiological Genomics* (2018).

116 Conickx, G. *et al*. microRNA profiling in lung tissue and bronchoalveolar lavage of cigarette smoke-exposed mice and in COPD patients: a translational approach. *Scientific reports* **7**, 12871 (2017).

117 Aghapour, M., Raee, P., Moghaddam, S. J., Hiemstra, P. S. & Heijink, I. H. Airway epithelial barrier dysfunction in chronic obstructive pulmonary disease: role of cigarette smoke exposure. *American journal of respiratory cell and molecular biology* **58**, 157-169 (2018).

118 Pouwels, S. *et al*. DAMPs activating innate and adaptive immune responses in COPD. Mucosal immunology. 2014; 7 (2): 215-26. (Epub 2013/10/24. https://doi.org/10.1038/mi. 2013.77 PMID: 24150257).

119 Smith, K. G., Kamdar, A. A. & Stark, J. M. in *Kendig's Disorders of the Respiratory Tract in Children Lung Defenses: Intrinsic,*

Innate, and Adaptive Ch. 8, 120-133.e122 (2019).

120 Hewitt, R. J. & Lloyd, C. M. Regulation of immune responses by the airway epithelial cell landscape. *Nat Rev Immunol* **21**, 347-362 (2021). https://doi.org:10.1038/s41577-020-00477-9

121 Rock, J. R. *et al*. Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc Natl Acad Sci U S A* **106**, 12771-12775 (2009). https://doi.org:10.1073/pnas.0906850106

122 Evans, M. J., Van Winkle, L. S., Fanucchi, M. V. & Plopper, C. G. Cellular and molecular characteristics of basal cells in airway epithelium. *Exp Lung Res* **27**, 401-415 (2001). https://doi.org:10.1080/019021401300317125

123 Rock, J. R. & Hogan, B. L. Epithelial progenitor cells in lung development, maintenance, repair, and disease. *Annu Rev Cell Dev Biol* **27**, 493-512 (2011). https://doi.org:10.1146/annurev-cellbio-100109-104040

124 Heijink, I. H., Brandenburg, S. M., Postma, D. S. & van Oosterhout, A. J. Cigarette smoke impairs airway epithelial barrier function and cell–cell contact recovery. *European Respiratory Journal* **39**, 419-428 (2012).

125 Lee, K. W. & Pausova, Z. Cigarette smoking and DNA methylation. *Front Genet* **4**, 132 (2013). https://doi.org:10.3389/fgene.2013.00132

126 Vermeulen, C. J. *et al*. Differential DNA methylation in bronchial biopsies between persistent asthma and asthma in remission. *European Respiratory Journal* **55** (2020).

127 Ritchie, M. E. *et al*. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015). https://doi.org:10.1093/nar/gkv007
128 Smyth, G. K. in *Bioinformatics and computational biology solutions using R and Bioconductor* 397-420 (Springer, 2005).
129 Faiz, A. *et al*. Cigarette smoke exposure decreases CFLAR expression in the bronchial epithelium, augmenting susceptibility for lung epithelial cell death and DAMP release. *Scientific reports* **8**, 12426 (2018).
130 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360 (2015). https://doi.org:10.1038/nmeth.3317
131 Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009). https://doi.org:10.1093/bioinformatics/btp352
132 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015). https://doi.org:10.1093/bioinformatics/btu638
133 Faiz, A. *et al*. Cigarette smoke exposure decreases CFLAR expression in the bronchial epithelium, augmenting susceptibility for lung epithelial cell death and DAMP release. *Sci Rep* **8**, 12426 (2018). https://doi.org:10.1038/s41598-018-30602-7
134 Moses, E. *et al*. Molecular Impact of Electronic Cigarette Aerosol Exposure in Human Bronchial Epithelium. *Toxicol Sci* **155**, 248-257 (2017). https://doi.org:10.1093/toxsci/kfw198
135 Singh, A. *et al*. Transcription factor NRF2 regulates miR-1 and miR-206 to drive tumorigenesis. *J Clin Invest* **123**, 2921-2934 (2013). https://doi.org:10.1172/JCI66353
136 Prochazkova, J. *et al*. Adaptive changes in global gene expression profile of lung carcinoma A549 cells acutely exposed to distinct types of AhR ligands. *Toxicol Lett* **292**, 162-174 (2018). https://doi.org:10.1016/j.toxlet.2018.04.024
137 Yang, S. Y., Ahmed, S., Satheesh, S. V. & Matthews, J. Genome-wide mapping and analysis of aryl hydrocarbon receptor (AHR)-and aryl hydrocarbon receptor repressor (AHRR)-binding sites in human breast cancer cells. *Archives of toxicology* **92**, 225-240 (2018).
138 Wang, X. *et al*. A polymorphic antioxidant response element links NRF2/sMAF binding to enhanced MAPT expression and reduced risk of Parkinsonian disorders. *Cell reports* **15**, 830-842 (2016).
139 Titz, B. *et al*. Alterations in the sputum proteome and transcriptome in smokers and early-stage COPD subjects. *J Proteomics* **128**, 306-320 (2015). https://doi.org:10.1016/j.jprot.2015.08.009
140 Pipes, L. *et al*. The non-human primate reference transcriptome resource (NHPRTR) for comparative functional genomics. *Nucleic Acids Res* **41**, D906-914 (2013). https://doi.org:10.1093/nar/gks1268
141 Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013). https://doi.org:10.1093/bioinformatics/bts635
142 Duclos, G. E. *et al*. Characterizing smoking-induced transcriptional heterogeneity in the human bronchial epithelium at single-cell resolution. *Sci Adv* **5**, eaaw3413 (2019). https://doi.org:10.1126/sciadv.aaw3413

143 Pouwels, S. D. *et al*. Acute cigarette smoke-induced eQTL affects formyl peptide receptor expression and lung function. *Respirology* **26**, 233-240 (2021).
144 Heijink, I. H. *et al*. Characterisation of cell adhesion in airway epithelial cell types using electric cell–substrate impedance sensing. *European Respiratory Journal* **35**, 894-903 (2010).
145 Goldfarbmuren, K. C. *et al*. Dissecting the cellular specificity of smoking effects and reconstructing lineages in the human airway epithelium. *Nat Commun* **11**, 2485 (2020). https://doi.org:10.1038/s41467-020-16239-z
146 Wohnhaas, C. T. *et al*. Cigarette Smoke Specifically Affects Small Airway Epithelial Cell Populations and Triggers the Expansion of Inflammatory and Squamous Differentiation Associated Basal Cells. *Int J Mol Sci* **22** (2021). https://doi.org:10.3390/ijms22147646
147 Sikkema, L. *et al*. An integrated cell atlas of the lung in health and disease. *Nat Med* **29**, 1563-1577 (2023). https://doi.org:10.1038/s41591-023-02327-2
148 Morissette, M. C. *et al*. Impact of cigarette smoke on the human and mouse lungs: a gene-expression comparison study. *PLoS One* **9**, e92498 (2014). https://doi.org:10.1371/journal.pone.0092498
149 Zuo, W. L. *et al*. Cell-specific expression of lung disease risk-related genes in the human small airway epithelium. *Respir Res* **21**, 200 (2020). https://doi.org:10.1186/s12931-020-01442-9
150 Shahdoust, M., Hajizadeh, E., Mozdarani, H. & Chehrei, A. Finding genes discriminating smokers from non-smokers by applying a growing self-organizing clustering method to large airway epithelium cell microarray data. *Asian Pac J Cancer Prev* **14**, 111-116 (2013). https://doi.org:10.7314/apjcp.2013.14.1.111
151 Imkamp, K. *et al*. Nasal epithelium as a proxy for bronchial epithelium for smoking-induced gene expression and expression quantitative trait loci. **142**, 314-317. e315 (2018).
152 Billatos, E. *et al*. Impact of acute exposure to cigarette smoke on airway gene expression. **50**, 705-713 (2018).
153 Beamer, C. A. & Shepherd, D. M. Role of the aryl hydrocarbon receptor (AhR) in lung inflammation. *Semin Immunopathol* **35**, 693-704 (2013). https://doi.org:10.1007/s00281-013-0391-7
154 Cho, H. Y. & Kleeberger, S. R. Nrf2 protects against airway disorders. *Toxicol Appl Pharmacol* **244**, 43-56 (2010). https://doi.org:10.1016/j.taap.2009.07.024
155 Thatcher, T. H. *et al*. Aryl hydrocarbon receptor-deficient mice develop heightened inflammatory responses to cigarette smoke and endotoxin associated with rapid loss of the nuclear factor-κB component RelB. *The American journal of pathology* **170**, 855-864 (2007).
156 Iizuka, T. *et al*. Nrf2-deficient mice are highly susceptible to cigarette smoke-induced emphysema. **10**, 1113-1125 (2005).
157 Aarts, J. M. *et al*. Fire usage and ancient hominin detoxification genes: protective ancestral variants dominate while additional derived risk variants appear in modern humans. *PLoS One* **11**, e0161102 (2016).
158 Jang, J.-H. *et al*. Aldehyde dehydrogenase 3A1 protects airway epithelial cells from cigarette smoke-induced DNA damage and cytotoxicity. *Free Radical Biology and Medicine* **68**, 80-86 (2014).

159 Bowdish, D. M. E. The Aging Lung: Is Lung Health Good Health for Older Adults? *Chest* **155**, 391-400 (2019). https://doi.org:10.1016/j.chest.2018.09.003
160 Kirkland, J. L., Tchkonia, T., Zhu, Y., Niedernhofer, L. J. & Robbins, P. D. The Clinical Potential of Senolytic Drugs. *J Am Geriatr Soc* **65**, 2297-2301 (2017). https://doi.org:10.1111/jgs.14969
161 Xu, M. *et al*. Senolytics improve physical function and increase lifespan in old age. *Nat Med* **24**, 1246-1256 (2018). https://doi.org:10.1038/s41591-018-0092-9
162 Tsuji, T., Aoshiba, K. & Nagai, A. Alveolar cell senescence in patients with pulmonary emphysema. *Am J Respir Crit Care Med* **174**, 886-893 (2006). https://doi.org:10.1164/rccm.200509-1374OC
163 Rutten, E. P. *et al*. Various Mechanistic Pathways Representing the Aging Process Are Altered in COPD. *Chest* **149**, 53-61 (2016). https://doi.org:10.1378/chest.15-0645
164 Amsellem, V. *et al*. Telomere dysfunction causes sustained inflammation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **184**, 1358-1366 (2011). https://doi.org:10.1164/rccm.201105-0802OC
165 Schafer, M. J. *et al*. Cellular senescence mediates fibrotic pulmonary disease. *Nat Commun* **8**, 14532 (2017). https://doi.org:10.1038/ncomms14532
166 Nambiar, A. *et al*. Senolytics dasatinib and quercetin in idiopathic pulmonary fibrosis: results of a phase I, single-blind, single-center, randomized, placebo-controlled pilot trial on feasibility and tolerability. *EBioMedicine* **90**, 104481 (2023). https://doi.org:10.1016/j.ebiom.2023.104481
167 Munoz-Espin, D. & Serrano, M. Cellular senescence: from physiology to pathology. *Nat Rev Mol Cell Biol* **15**, 482-496 (2014). https://doi.org:10.1038/nrm3823
168 Born, E. *et al*. Eliminating Senescent Cells Can Promote Pulmonary Hypertension Development and Progression. *Circulation* **147**, 650-666 (2023). https://doi.org:10.1161/CIRCULATIONAHA.122.058794

Supplementary Tables and Figures

Supplementary table 1: Cell-to-cell dependencies within each spot and in areas of immediate and extended neighbourhood in each lung tissue sample.

target	intercept	intra	juxta_5	para_15	p.intra	p.juxta_5	p.para_15
alv.M_	-0.0132791	0.68271918	0.28781519	0.32707931	1.33E-15	0.00213843	3.34E-06
Ciliated	-0.0691958	0.84257217	0.05721949	0.45174092	0	0.44688109	1.89E-14
Cd4.T	-0.0008097	0.60952441	0.17036474	0.18813574	1.80E-12	0.1147151	0.01564691
AT1	-0.1581267	1.27698022	0.31309554	0.21844141	0	2.84E-05	6.25E-05
adv.fibro	-0.0014322	0.4377633	0.17567585	0.5271163	1.31E-08	0.06693126	3.41E-11
ILC2	0.00089977	0.54128513	0.09625118	0.14219901	9.86E-11	0.34921041	0.06881509
Meso	-0.0092575	0.84542126	0.16279299	0.3475594	0	0.01026407	7.33E-10
Plasma.cells	0.00232799	-0.0542602	0.07214468	0.06132943	0.17756178	0.06084922	0.11582812
Anti.inflam.mono	0.00015615	0.77109655	0.06517158	0.07217624	0	0.54540753	0.34673212
Cd8.T	-0.0010671	0.80046431	0.13756631	0.09892133	0	0.16121128	0.22199563
aEC	-0.035176	1.11743185	0.19059622	0.34599337	0	0.06399151	3.32E-08
int.M_	-0.0096015	0.56192034	0.23236623	0.61096755	6.66E-16	0.00327525	0
Club	-0.0034346	0.92836306	0.06305824	0.21705987	0	0.12301226	2.28E-10
AT2	-0.0083914	0.9797321	-0.0386608	0.3386403	0	0.66930883	8.54E-08
Pro.inflam.mono	0.00388447	0.50438051	0.0749784	0.05599484	1.28E-07	0.50530524	0.44262091
Neut	0.00115316	0.53592961	0.05302995	-0.0207729	2.46E-08	0.58286115	0.79492457
aSMC	-0.0137508	0.7277115	0.188499	0.35138343	0	0.0521591	1.32E-07
Aerocytes	-0.0531395	0.96282941	0.36714666	0.08887185	0	3.35E-05	0.19790243

cEC	-0.0022189	0.81935628	0.10485778	0.25839858	0	0.26070823	0.00026944
B	-0.007034	0.81436068	0.25686885	0.15503096	0	0.00592573	0.0265596
DC	-0.0067762	0.67521395	0.32078108	0.30713923	3.35E-13	0.00147984	4.34E-06
lymphEC	-0.0003383	0.59371288	0.18407374	0.1567503	1.34E-11	0.05870603	0.04631208
vEC	-0.0056705	1.0931399	-0.0092515	0.12235348	0	0.92735929	0.08293363
NK	0.00250226	0.37691868	-0.155328	0.18763262	8.93E-05	0.16811619	0.00461336
alv.fibro	-0.0406271	0.99470678	0.23223705	0.11634752	0	0.03571614	0.12178538
Cd4.Cd8.T	-0.0029756	0.6069215	0.08551395	0.52459568	0	0.27803073	0
Mast.baso	0.00060019	0.39730226	0.12846989	0.14998239	1.18E-05	0.24152112	0.06350879
vSMC	0.00018767	0.57755311	0.12942684	0.17287605	1.86E-10	0.20423936	0.02165177

Supplementary Table 2: Clinical characteristics of NORM donors.

	Current Smoker	
	no	**yes**
n	69	93
Gender n (Male %)	39(0.57)	52(0.56)
Age (SD)	41.32(18.83)	43.71(16.77)
Blood		
Percentage		
Basophils (%)	0.59(0.32)	0.43(0.21)*
Eosinophils (%)	2.55(1.54)	2.92(1.59)

Neutrophils (%)	53.2(7.78)	55.1(9.15)
Lymphocytes (%)	40.11(32.92)	33.78(8.48)
Monocytes (%)	7.66(1.75)	7.83(1.64)
Total numbers (10^9/L)		
Total leucocyte count	5.51(1.55)	7.04(2.04)*
Neutrophils	2.94(1.2)	3.8(1.83)*
Lymphocytes	1.96(0.56)	2.23(0.54)*
Monocytes	0.41(0.15)	0.53(0.15)*
Eosinophils	0.13(0.08)	0.2(0.11)*
Basophils	0.03(0.02)	0.03(0.02)
Sputum percentage		
Basophils (%)	0(0)	0(0)
Eosinophils (%)	0.3(0.91)	0.6(0.8)
Neutrophils (%)	50.72(23.61)	53.83(20.85)
Lymphocytes (%)	0.72(0.84)	0.3(0.4)*
Macrophages (%)	48.26(23.38)	45.28(20.98)
Bronchial cells (%)	3.3(2.93)	5.37(7)
Squamous (%)	53.56(31.63)	37.62(29.12)*
Non-squamous (%)	46.44(31.63)	62.38(29.12)*

Total numbers (10^9)		
Macrophages	0.14(0.09)	0.17(0.18)
Neutrophils	0.33(0.29)	0.33(0.3)
Eosinophils	0.01(0.02)	0.01(0.01)
Lymphocytes	0(0)	0(0)
Basophils	0(0)	0(0)
Bronchial cells	0.02(0.04)	0.03(0.04)
Percentage of FEV1 predicted	108.28(12.39)	103.26(12.57)*

*<0.05 FEV1= forced expiratory volume in one second.

Supplementary Table 3: Smoking signature.

Gene	log2FoldChange	pvalue	padj
AKR1C2	2.63560903	1.54657097627777e-73	3.51752102844616e-69
ALDH3A1	2.51940043	5.562772527432e-60	6.32598491819567e-56
ENSG00000253339	2.74117032	4.21278227672221e-54	3.193850670059e-50
AKR1B10	3.02703918	3.94606060266998e-53	2.24373005867815e-49
CLDN10	2.19116228	3.03211046522605e-52	1.37924640842203e-48
SLC7A11	2.90390122	3.05219725083204e-50	1.15698623788206e-46
CYP1B1	3.12570031	7.14262619808894e-50	2.32074128927621e-46
ENSG00000231683	2.99633026	2.34943163990971e-47	6.67943415226331e-44

CEACAM5	2.72942187	3.50634213073186e-42	8.8609161579295e-39
SPDEF	2.04052969	1.10072210241699e-41	2.5034823497372e-38
NQO1	2.01165069	1.22114094163098e-39	2.52487541604136e-36
OSGIN1	1.58696421	8.89216344494516e-39	1.68536137826527e-35
ME1	1.55651553	9.85262131968311e-39	1.72375399457594e-35
VSIG2	1.30429795	5.60416108680841e-38	9.10435998274075e-35
GPX2	1.89567525	2.21303440319902e-37	3.35555029775723e-34
B3GNT6	2.43376843	7.05303084001356e-37	1.00258833390793e-33
ENSG00000228277	1.97838145	3.20741790635011e-35	4.05275071455705e-32
MUC5AC	2.13270028	2.74996950926538e-34	3.29185823782799e-31
ENSG00000171658	2.29224105	4.21414624861637e-33	4.79232711392653e-30
FUT6	1.32130572	9.7669057253889e-33	1.0578023991345e-29
H19	2.2448666	1.0477392880371e-32	1.08317192577799e-29
TRIM16L	1.38177126	1.66860541345801e-31	1.65003310972561e-28
ENSG00000261296	2.25025437	3.42720276677676e-31	3.24784582198211e-28
ATP13A5	1.87089192	1.37106707289872e-30	1.24734198024034e-27
PIR	1.33216042	3.31443991052578e-30	2.89937005096148e-27
FAM3D	1.27302678	4.46711084099998e-30	3.76296181361865e-27
FER1L6	1.47820651	6.01776675006362e-29	4.7195892056361e-26
AKR1C3	1.49710523	7.81863049039566e-29	5.9275643957853e-26
FUT3	1.39947682	2.1861366900124e-28	1.60391912508523e-25
ENSG00000166104	2.02079341	2.16385484719782e-27	1.4474916071961e-24
RDH10	1.39975068	3.54561556660019e-27	2.30404229847871e-24
KCNE3	1.14482236	9.98846678584737e-27	6.31049134936979e-24
MUCL1	1.83938427	1.18660729018161e-26	7.29410708321366e-24
MUC1	1.12379569	2.18102810366225e-26	1.27833131954767e-23
GALNT6	1.13312098	2.6101280886667e-26	1.48411883121589e-23

GALNT7	1.0216327	2.75468176613945e-26	1.49172576402561e-23
LUCAT1	2.00544718	1.94814946820149e-25	9.84638033439438e-23
LINC00675	1.61273231	2.45346192542069e-25	1.21307691373409e-22
GALNT12	1.01061472	4.14896150024942e-25	2.00774426301432e-22
ENSG00000258689	1.82578658	5.25374482062895e-25	2.48939942084135e-22
GPR110	1.33032948	5.75242406170728e-25	2.67006393590756e-22
CREB3L1	1.15419186	1.01797660838009e-24	4.63057199619937e-22
AGR2	1.45613435	1.84864290409869e-24	8.24422239427857e-22
TALDO1	1.0649813	2.22009752014385e-24	9.71036499964455e-22
ENSG00000224251	1.88795445	2.59953661268389e-24	1.11554454186571e-21
CABYR	1.41111001	3.29101513974907e-24	1.38612682108246e-21
AJAP1	1.62600556	2.28343946960497e-23	9.27402630298131e-21
AKR1C1	1.26108654	5.73842083655646e-23	2.28973058783579e-20
PHEX	1.13519649	5.92276520253391e-23	2.32254089252468e-20
CEACAM6	1.49670036	7.23593759245985e-23	2.78939262038825e-20
AHRR	1.75489739	9.04178842298509e-23	3.42744059820621e-20
TMEM45B	1.26410738	1.4213295828684e-22	5.13122540202523e-20
SYTL5	1.35140108	3.29617583464867e-22	1.17137848723827e-19
S100P	1.50749161	3.48387189232389e-22	1.20609040845163e-19
ENSG00000238755	1.94257436	3.49991061193316e-22	1.20609040845163e-19
GJB1	1.30484998	7.99428753014849e-22	2.67385405273084e-19
ENSG00000264785	1.94420871	1.10377524603402e-21	3.58632345654253e-19
ADH7	1.17254367	1.21600271098282e-21	3.89531910684412e-19
ATP10B	1.44960044	1.53545609015197e-21	4.85033518255782e-19
SRPX2	1.22119879	2.69013920407809e-21	8.38144192569206e-19
DEFB1	1.6864961	1.14542439159706e-20	3.47353764833115e-18
DHRS9	1.19900154	1.32989054089839e-20	3.97987242923591e-18

CYP1B1-AS1	1.43836231	5.73152840499668e-20	1.67125489799032e-17
KLK13	1.53624152	9.0955880160161e-20	2.58587567295338e-17
RAP1GAP	1.2961176	9.56430675487101e-20	2.65281210771691e-17
OCA2	1.51172837	1.5407771117198e-19	4.22210055770543e-17
TMEM178B	1.09663605	2.08448666074691e-19	5.57759583670915e-17
FA2H	1.19958397	2.8007400106143e-19	7.32184262085191e-17
CYP1A1	1.95045422	5.80330327886919e-19	1.46655921971779e-16
TNNT3	1.27035685	6.28116410549904e-19	1.56987688368649e-16
RNF183	1.16751442	9.67651348624541e-19	2.39220242099093e-16
PGD	1.12536207	1.09160602455089e-18	2.66962230348231e-16
CLDN7	1.02083317	2.71398842783134e-18	6.56669710665914e-16
FER1L6-AS1	1.48889587	4.42651837759965e-18	1.04871597895965e-15
SECTM1	1.12655734	6.65832070221579e-18	1.51436846051196e-15
FAM177B	1.38330344	1.04789763608138e-17	2.33660625833675e-15
NKX3-1	1.54521053	1.16970853360652e-17	2.58289814449969e-15
S100A9	1.89610642	1.41672795540839e-17	3.03981703941589e-15
JAKMIP3	1.38952167	1.68642013815463e-17	3.55147589094343e-15
ENTPD8	1.22887688	1.73808782015418e-17	3.62670361298961e-15
UCHL1	1.59105417	2.00356569015018e-17	4.1426452778887e-15
LRRC31	1.51609497	3.45595410040475e-17	7.01805536246479e-15
CYP4F3	1.04607755	3.53758036296391e-17	7.12024139603993e-15
ENSG00000203472	1.03889807	5.62738873358483e-17	1.09392589193721e-14
GAD1	1.33752377	1.3439268957223e-16	2.48506287124455e-14
FAM3B	1.18909118	1.59584298801856e-16	2.88062324757891e-14
MSMB	1.63275125	1.69461211980298e-16	3.03482346872433e-14
PTPRH	1.69786314	1.87441857342424e-16	3.3306075026532e-14
IGSF11	1.08054823	2.46686777238725e-16	4.2504879253921e-14

ENSG00000232756	1.05262167	2.87554342397541e-16	4.88069848021618e-14
HEPACAM2	1.36124317	4.79539446944178e-16	8.01959204507234e-14
ENSG00000232079	1.24447975	5.06808805341621e-16	8.41376603554003e-14
TFF3	1.24178886	6.14536038732544e-16	1.01282664238645e-13
SPTSSB	1.2763832	6.73060190381049e-16	1.10130078921055e-13
C12orf36	1.73083243	1.49433072640918e-15	2.37671734555597e-13
TPRXL	1.46979331	1.88031623818036e-15	2.92917209049138e-13
UPK1B	1.54629453	2.48658929200466e-15	3.73099057704463e-13
KRT23	1.21149505	2.7242607522062e-15	4.04971153909659e-13
KIAA1644	1.4373885	2.88989265684018e-15	4.24103690759605e-13
ANXA3	1.13818663	3.84061028164197e-15	5.49376353746321e-13
RNF175	1.03342634	5.19157458340755e-15	7.33398585869698e-13
TXN	1.30042592	6.05343554828868e-15	8.39508159209011e-13
GALNT5	1.00202875	1.39407582499477e-14	1.78128430133039e-12
ANO7	1.16700403	1.52018848621604e-14	1.92084260724987e-12
MLC1	1.36627611	1.57666364094626e-14	1.98119546130839e-12
CRLF1	1.18119568	1.80776842099456e-14	2.25911455863188e-12
SDCBP2	1.15515938	2.73620573155565e-14	3.27538227150008e-12
ENSG00000249352	1.33583678	3.05499625355815e-14	3.60014688035888e-12
MSLN	1.22259405	6.7571584426281e-14	7.6460105283151e-12
IL1R2	1.52366944	7.70166722669479e-14	8.62890243369193e-12
HS6ST2	1.28841403	9.82121503968444e-14	1.07910007179992e-11
PACSIN1	1.43359238	9.92105294645184e-14	1.08482898179856e-11
CSTA	1.52469983	1.08874194509145e-13	1.17915937138857e-11
ENSG00000253417	1.42523865	1.27113748879507e-13	1.34039323551089e-11
FOLH1	1.01628666	1.32671449603772e-13	1.38416488522394e-11
ENSG00000249599	1.41488167	2.44055822359775e-13	2.46702472166699e-11

ZNF467	1.0114198	3.51609067009128e-13	3.44698130174811e-11
ENSG00000243486	1.58039039	4.41268392686301e-13	4.26056204380272e-11
ABCA12	1.13412816	4.73821227283513e-13	4.54708438537393e-11
ATP6V0A4	1.0158427	5.03279779822718e-13	4.78937042355142e-11
SERPINB11	1.11004772	5.51153099056111e-13	5.20142161200506e-11
SLC25A47P1	1.08194416	6.00624031594373e-13	5.64487312999274e-11
FOXA3	1.28283899	6.68857344416261e-13	6.23462763991944e-11
CPHL1P	1.29597981	7.2614375843993e-13	6.71358278128364e-11
TSPAN8	1.02947568	9.23084110902465e-13	8.36439243759588e-11
BPIFB1	1.23232733	1.15182955372167e-12	1.0233285691346e-10
TFF1	1.42221828	1.95016143954841e-12	1.70594122234958e-10
SCEL	1.57165779	2.06648386636171e-12	1.793897292234e-10
BCAN	1.14399816	3.62970777340429e-12	2.99108962312707e-10
S100A8	1.51100814	4.5633559681357e-12	3.68045986309498e-10
CAPN9	1.00172569	4.64358457737702e-12	3.73193242501283e-10
WDR72	1.33813166	5.54851995558411e-12	4.38178950936823e-10
LINC00589	1.06090544	6.59390350363144e-12	5.06661288130384e-10
ENSG00000245466	1.07358752	1.09923900681316e-11	8.27850727515184e-10
GOLGA7B	1.14205742	1.21842993481987e-11	9.11218072648449e-10
CA12	1.36081701	1.31070349082741e-11	9.64745637390896e-10
MACC1	1.03664924	1.32941458526244e-11	9.75361462168029e-10
TMEM132B	1.0358398	1.90748312964096e-11	1.37290494622006e-09
DUOX2	1.20099189	2.06940798956667e-11	1.47544248635437e-09
ELFN2	1.25157326	2.3425916050777e-11	1.63938164510422e-09
LOXHD1	1.45476307	2.87923708783248e-11	1.98440510077764e-09
CALCA	1.27294835	4.86595624622692e-11	3.26464037947448e-09
ENSG00000250237	1.13790258	6.54951455538522e-11	4.31774374051251e-09

ARHGAP40	1.32301137	1.07375575073961e-10	6.8026464609531e-09
SHH	1.05258241	1.14053825160581e-10	7.14611625193461e-09
ENSG00000251537	1.14316038	1.63478652886964e-10	9.9682533009681e-09
TCN1	1.27977241	3.78442340250129e-10	2.2183743780023e-08
TMPRSS11A	1.34724374	7.12989657031356e-10	3.9075269300051e-08
ENSG00000260641	1.10855323	7.43703894680644e-10	4.06605802418668e-08
A2ML1	1.34585355	8.15011861177145e-10	4.40299994551377e-08
HGD	1.02747999	8.59725075208011e-10	4.60084402600729e-08
GBP6	1.03958953	8.8530709563838e-10	4.71555610847759e-08
PSCA	1.12813926	1.24507014238556e-09	6.49492553174704e-08
ENSG00000260337	1.19042043	1.3086252555648e-09	6.77981157461635e-08
ENSG00000253789	1.04671387	1.76610022489406e-09	8.9262630033312e-08
HMGCS2	1.27750411	2.64395427874022e-09	1.30160381202744e-07
ENSG00000242770	1.16429146	2.88229841895533e-09	1.41282317329138e-07
MTND4P19	1.09276462	3.59604116417559e-09	1.72186021553704e-07
CCAT1	1.10812711	4.48572304216991e-09	2.10770819149207e-07
S100A14	1.03415912	5.30126677514929e-09	2.44567974713987e-07
SCGB2A1	1.02122873	5.43170717799879e-09	2.50078437361143e-07
ENSG00000237311	1.02004256	5.50438321577897e-09	2.52912508807428e-07
GCNT3	1.13898779	5.60767278943062e-09	2.57138931295988e-07
TMEM211	1.02928331	5.6863314898917e-09	2.59698641377705e-07
LYPD3	1.26846567	6.74205338655744e-09	3.03045972774432e-07
IL1RN	1.27760894	8.3873366791185e-09	3.68977921527797e-07
ENSG00000244137	1.03132866	8.44329293916652e-09	3.70008197704053e-07
KLK12	1.00125922	1.37095036353659e-08	5.77423982745854e-07
COMP	1.18171733	1.38472617213811e-08	5.78937721674799e-07
EGF	1.12832216	3.06827970891399e-08	1.18681894046836e-06

OR5BA1P	1.05588951	3.41801858603467e-08	1.31095134436379e-06
CYP4F11	1.00659322	4.29264790156498e-08	1.61108884279198e-06
UGT1A10	1.00376501	4.92337589129303e-08	1.82076847596047e-06
NMUR2	1.15471297	6.79490180278638e-08	2.45306740639006e-06
PPP2R2C	1.12581648	8.82581428663276e-08	3.12468963990317e-06
B3GALT5	1.02010212	8.86134724647179e-08	3.12468963990317e-06
NOS1	1.14940083	1.01551424050056e-07	3.52623753983888e-06
ENSG00000215895	1.06588038	1.17504810702055e-07	4.0309644262557e-06
PADI1	1.1652185	1.26220991462688e-07	4.29113636745499e-06
MSLNL	1.10632734	1.45530994197873e-07	4.89638599413671e-06
HIST1H1B	1.05202915	1.62262606214281e-07	5.38759228574832e-06
ENSG00000234477	1.0683219	2.28559126258451e-07	7.35268566848967e-06
BPIFB2	1.13124414	2.51078965906896e-07	8.00517647002908e-06
SLC9A2	1.03698047	3.55789019542325e-07	1.10096128713886e-05
HIST1H2AJ	1.01197728	3.95758381852285e-07	1.21145742084097e-05
HIST1H2BB	1.04764601	4.6606352865782e-07	1.39659405741679e-05
GRP	1.01675622	5.77301803433918e-07	1.68335284837193e-05
HIST1H3G	1.03504002	5.97589540849211e-07	1.73361945370848e-05
MAL	1.07311825	7.33574608946445e-07	2.06746231795266e-05
ZG16B	1.08906889	9.74796340402686e-07	2.6776289814153e-05
ANKRD22	1.01831231	1.20525785538453e-06	3.23259253099832e-05
SLC9A4	1.05555025	1.21868982774932e-06	3.25709535162521e-05
FAM83C	1.02664813	2.17399174804695e-06	5.39795505650434e-05
KRT6A	1.03824642	2.19778801620334e-06	5.44975402744544e-05
KRT4	1.05377305	2.36144964662666e-06	5.79701096162625e-05

Supplementary Table 4: KEGG pathway analysis of smoking signature.

Pathway	Adjusted p value
Steroid hormone biosynthesis	0.000145
Mucin type O-glycan biosynthesis	4.44e-06
Retinol metabolism	0.00693
Metabolism of xenobiotics by cytochrome P450	0.00036
Metabolic pathways	5.47e-05
Chemical carcinogenesis	0.000645

Log2 fold change >2 or <-2, adjusted p value <0.05.

Supplementary Table 5: Differential ChIP-Seq analysis of smoking signature for AhR in the presence and absence of TCDD treatment in MCF-7 cells and ChIP-Seq analysis of smoking signature for NRF2 in the presence and absence of dietary isothiocyanate, sulforaphane (SFN) treatment in BEAS2B cells.

chr	start	end	length	abs_summit	pileup	-log10(p-value)	fold_enrichment	-log10(q value)	Gene Name	Annotation	Distance to TSS
AhR											
chr2	38304635	38305021	387	38304799	379	205.05	8.25	198.8	CYP1B1	Intergenic	-1505
chr1	120322615	120322865	251	120322748	230	96.58	5.9	91.72	HMGCS2	Intergenic	-11185
chr17	80291687	80292029	343	80291878	182	91.05	7.32	86.28	SECTM1	promoter-TSS (NM_003004)	63

chr3	193160990	193161283	294	193161117	97	87.62	19.6	82.91	ATP13A5	intron (NM_032279, intron 18 of 29)	-64623
chr11	46295518	46295690	173	46295608	133	67.84	7.44	63.45	CREB3L1	Intergenic	-3585
chr16	69760440	69760960	521	69760822	226	57.73	3.65	53.52	NQO1	promoter-TSS (NM_001025433)	-129
chr3	161111217	161111403	187	161111290	62	54.13	17.47	49.98	SPTSSB	Intergenic	-21439
chr1	21965150	21965558	409	21965335	175	48.75	3.9	44.7	RAP1GAP	intron (NM_001145657, intron 3 of 26)	12994
chr8	124934460	124934645	186	124934537	107	48.54	6.34	44.49	FER1L6	intron (NM_001039112, intron 1 of 40)	70326
chr19	42278430	42278793	364	42278681	230	48.37	3.15	44.32	CEACAM6	Intergenic	19214
chr17	18625183	18625529	347	18625355	159	48.29	4.19	44.24	TRIM16L	promoter-TSS (NM_001037330)	-46
chr6	34513580	34513816	237	34513685	81	42	7.42	38.07	SPDEF	intron (NM_001252294, intron 1 of 4)	10412
chr17	19651651	19652003	353	19651772	89	34.72	5.28	30.93	ALDH3A1	promoter-TSS (NM_000691)	-81
chr10	5122101	5122410	310	5122223	75	30.16	5.42	26.47	AKR1C3	intron (NM_001253908, intron 1 of 8)	-14312
chr7	138560754	138560939	186	138560845	57	26.86	6.35	23.24	ATP6V0A4	intron (NM_020910, intron 12 of 19)	-77906
chr18	21688647	21688847	201	21688741	40	26.53	10	22.92	CABYR	intron (NM_153211, intron 6 of 13)	-30173
chr11	761648	761810	163	761731	33	23.51	11.16	19.97	TALDO1	intron (NM_006755, intron 4 of 7)	14297
chr15	63673179	63673340	162	63673261	45	19.72	5.76	16.27	CA12	intron (NM_206925, intron 1 of 9)	815
chr7	16844822	16845020	199	16844928	84	19.3	3.23	15.86	AGR2	promoter-TSS (NM_006408)	-183
chr3	120380248	120380419	172	120380309	61	18.98	4.08	15.55	HGD	intron (NM_000187, intron 4 of 13)	21084
chr10	5051477	5051663	187	5051573	65	15.4	3.27	12.08	AKR1C2	intron (NM_001354, intron 1 of 10)	-5364
chr5	151782741	151782873	133	151782815	24	11.8	6.12	8.61	NMUR2	intron (NM_020167, intron 1 of 3)	2033
chr17	7167998	7168204	207	7168131	34	11.77	4.29	8.58	CLDN7	Intergenic	-1589
chr13	78179347	78179502	156	78179461	29	11.59	4.9	8.4	SCEL	intron (NM_144777, intron 19 of 32)	69616
chr4	139156806	139157113	308	139156992	40	11.47	3.64	8.29	SLC7A11	intron (NM_014331, intron 2 of 11)	6543

chr7	140862025	140862286	262	140862184	40	10.34	3.33	7.21	TMEM178B	intron (NM_001195278, intron 1 of 3)	88124
chr3	122057818	122057955	138	122057900	20	9.4	5.59	6.31	CSTA	intron (NM_005213, intron 2 of 2)	13876
NRF2											
chr16	69760633	69761102	471	69760914	122	196.9733	56.5502	190.0103	NQO1	Promoter (<=1kb)	-100
chr17	19656312	19656659	349	19656494	75	83.62875	22.46204	77.95403	ALDH3A1	Distal Intergenic	-4566
chr9	1.13E+08	1.13E+08	307	1.13E+08	45	47.38888	16.77407	42.30107	TXN	Promoter (<=1kb)	-183
chrX	15511299	15511516	219	15511420	20	19.62139	10.1975	15.17912	PIR	Promoter (<=1kb)	195
chr4	1.39E+08	1.39E+08	386	1.39E+08	12	14.83409	9.14149	10.60889	SLC7A11	Intron (uc021xrw.1/23657, intron 2 of 11)	6409
chr6	84037523	84037674	153	84037648	10	10.53488	7.01857	6.54085	ME1	Intron (uc003pjy.3/4199, intron 5 of 13)	103264
chr10	5062586	5062760	176	5062722	9	10.47211	7.03192	6.48187	AKR1C2	Promoter (2-3kb)	-2361

Supplementary Table 6: Top 50 CpG sites differentially methylated between the bronchial biopsies collected from asymptomatic current smokers and never smokers

Name	Chromosome	Position	Relation to island	UCSC_RefGene_Name	logFC	P	Adjusted p value
cg05395852	chr7	128750384	OpenSea		-0.861	6.32E-23	2.04E-17
cg00370022	chr15	75015192	N_Shelf	CYP1A1	-0.805	9.55E-23	2.04E-17

cg23576855	chr5	373299	N_Shore	AHRR	-1.007	1.50E-22	2.14E-17
cg23160522	chr15	75015787	N_Shelf	CYP1A1	-0.904	2.22E-22	2.38E-17
cg26144569	chr2	38304804	S_Shore	CYP1B1	-0.862	1.50E-18	1.28E-13
cg02162897	chr2	38300537	N_Shore	CYP1B1	-0.958	1.15E-17	8.17E-13
cg09809932	chr1	6515597	Island	ESPN	1.123	1.46E-17	8.93E-13
cg21161138	chr5	399360	OpenSea	AHRR	-0.587	1.66E-16	7.66E-12
cg05575921	chr5	373378	N_Shore	AHRR	-0.776	1.74E-16	7.66E-12
cg00575672	chr4	100644171	OpenSea		-0.710	1.79E-16	7.66E-12
cg03344767	chr1	10475318	OpenSea	PGD	-0.603	1.65E-15	6.43E-11
cg03970915	chr22	47426893	S_Shelf	TBC1D22A	-0.480	6.88E-15	2.45E-10
cg02417427	chr5	79479211	OpenSea	SERINC5	-0.610	1.24E-14	4.08E-10
cg14492337	chr17	79813507	OpenSea	P4HB	-0.446	1.63E-14	4.80E-10
cg09618125	chr22	43120117	S_Shelf		-0.371	1.80E-14	4.80E-10
cg14181111	chr1	168773313	OpenSea		-0.532	1.74E-14	4.80E-10
cg06679296	chr11	1949032	Island	TNNT3	-0.735	2.02E-14	5.08E-10
cg24180759	chr5	167551234	OpenSea	ODZ2	-0.641	3.97E-14	9.43E-10
cg12258471	chr17	19650069	S_Shore	ALDH3A1	-1.059	5.99E-14	1.35E-09
cg11875995	chr8	123587744	OpenSea		-0.581	1.11E-13	2.38E-09
cg17054691	chr17	79813439	OpenSea	P4HB	-0.661	1.26E-13	2.56E-09
cg15663412	chr10	122864146	OpenSea		-0.785	2.01E-13	3.90E-09
cg04830857	chr3	156394724	S_Shore	TIPARP	-0.594	2.67E-13	4.95E-09
cg21715189	chr2	38304802	S_Shore	CYP1B1	-0.709	3.38E-13	6.01E-09
cg00244111	chr3	149086884	OpenSea	TM4SF1	-0.706	4.08E-13	6.98E-09

cg08220966	chr10	88717364	OpenSea	MMRN2	0.413	4.86E-13	7.98E-09
cg15059474	chr7	97990184	OpenSea	BAIAP2L1	-0.542	6.02E-13	9.19E-09
cg08921300	chr1	210534820	OpenSea	HHAT	-0.468	5.91E-13	9.19E-09
cg10829629	chr10	88717494	OpenSea	MMRN2	0.475	7.13E-13	1.05E-08
cg23732725	chr3	128766093	S_Shore		-0.750	1.13E-12	1.60E-08
cg01821149	chr11	1949090	Island	TNNT3	-0.397	1.22E-12	1.60E-08
cg27359629	chr6	74101601	N_Shelf		-0.413	1.23E-12	1.60E-08
cg24153813	chr14	26731696	OpenSea		-0.446	1.27E-12	1.60E-08
cg08924619	chr13	111160399	N_Shelf	COL4A2	-0.471	1.33E-12	1.63E-08
cg03243768	chr11	9115107	S_Shore	FLJ46111	0.503	1.24E-12	1.60E-08
cg19786648	chr17	71824806	OpenSea	C17orf54	-0.437	1.42E-12	1.66E-08
cg26535158	chr17	40175841	S_Shelf	NKIRAS2	-0.339	1.44E-12	1.66E-08
cg14751552	chr8	74209522	S_Shore	RDH10	-0.473	1.80E-12	1.97E-08
cg20623506	chr17	79813410	OpenSea	P4HB	-0.704	1.79E-12	1.97E-08
cg01395499	chr1	204982603	OpenSea	NFASC	-0.347	2.27E-12	2.43E-08
cg23878564	chr17	7287449	Island	TNK1	0.974	2.63E-12	2.74E-08
cg21515113	chr3	185676537	OpenSea		-0.578	2.79E-12	2.84E-08
cg11072637	chr7	65773051	OpenSea	TPST1	-0.492	3.36E-12	3.25E-08
cg04396721	chr1	220958670	N_Shore	MOSC1	-0.430	3.19E-12	3.17E-08
cg06878649	chr16	84336112	OpenSea	WFDC1	-0.543	3.43E-12	3.25E-08
cg25887076	chr10	5576256	OpenSea		-0.659	3.79E-12	3.38E-08
cg14486905	chr2	171677602	N_Shore	GAD1	0.669	3.79E-12	3.38E-08
cg16435782	chr21	39087399	OpenSea	KCNJ6	-0.388	3.77E-12	3.38E-08

cg24911905	chr4	3433773	OpenSea	RGS12	-0.481	4.06E-12	3.48E-08
cg10806811	chr14	25451658	OpenSea	STXBP6	-0.516	4.08E-12	3.48E-08

Log2 fold change >2 or <-2, adjusted p value <0.05.

Supplementary Table 7: Significant sputum proteins differentially abundant between asymptomatic current and never smokers.

Gene	**logFC**	**p-value**	**Adjusted p value**
ALDH3A1	1.650	4.032E-30	2.012E-27
NQO1	1.047	1.832E-10	8.363E-09
MSMB	0.943	6.750E-13	8.421E-11
AGR2	0.871	2.061E-08	5.414E-07
ADH7	0.808	2.037E-11	1.694E-09
SPRR1A	0.738	1.694E-04	1.190E-03
GSTA2	0.672	3.638E-07	6.052E-06
TFF3	0.648	2.601E-12	2.596E-10
PRH1	0.639	1.406E-03	7.232E-03
VSTM2L	0.624	8.044E-08	1.824E-06
HTN1	0.613	1.621E-04	1.172E-03
APOB	-0.593	1.128E-03	5.989E-03
S100A12	-0.618	7.466E-06	8.279E-05

MAT2B	-0.621	6.514E-04	3.736E-03
S100A9	-0.636	1.142E-04	8.771E-04
CAMP	-0.662	1.497E-05	1.494E-04
IGHM	-0.667	2.815E-08	6.689E-07
ACTB	-0.671	7.271E-06	8.247E-05
HIST1H1E	-0.678	9.315E-05	7.497E-04
S100A8	-0.776	1.174E-05	1.247E-04
DDX55	-0.880	1.393E-06	2.098E-05
P04220	-0.950	5.354E-09	1.670E-07
NT5E	-1.007	9.231E-05	7.497E-04
PRIM2	-1.210	2.748E-06	3.516E-05

Log2 fold change >1.5 or <-1.5, adjusted p value <0.05.

Supplementary Table 8: Significant pathways associated with sputum proteins differentially abundant between asymptomatic current and never smokers.

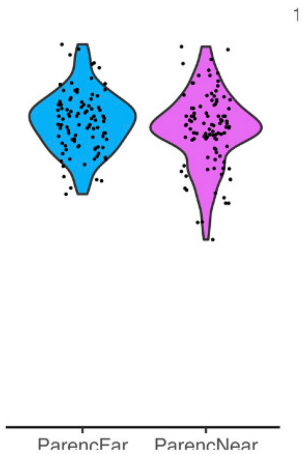
ID	Pathway description	Gene count	False discovery rate
GO.0050786	RAGE receptor binding	3	7.050E-04
GO.0017127	cholesterol transporter activity	3	2.260E-03
GO.0070325	lipoprotein particle receptor binding	3	2.370E-03

GO.0035662	Toll-like receptor 4 binding	2	3.450E-03
GO.0070653	high-density lipoprotein particle receptor binding	2	3.450E-03
GO.0008289	lipid binding	7	5.170E-03
GO.0034190	apolipoprotein receptor binding	2	7.370E-03
GO.0050544	arachidonic acid binding	2	7.370E-03
GO.0060228	phosphatidylcholine-sterol O-acyltransferase activator activity	2	7.370E-03
GO.0005102	receptor binding	8	1.100E-02
GO.0008035	high-density lipoprotein particle binding	2	1.200E-02
GO.0031210	phosphatidylcholine binding	2	3.720E-02
GO.0036094	small molecule binding	11	3.720E-02
GO.0043168	anion binding	11	3.720E-02
GO.0055102	lipase inhibitor activity	2	3.720E-02
GO.0043178	alcohol binding	3	3.990E-02
KEGG: 980	Metabolism of xenobiotics by cytochrome P450	3	0.00792
KEGG: 982	Drug metabolism - cytochrome P450	3	0.00792
KEGG: 5131	Shigellosis	3	0.00792
KEGG: 5204	Chemical carcinogenesis	3	0.00792
KEGG: 4970	Salivary secretion	3	0.0081
KEGG: 5132	Salmonella infection	3	0.0081
KEGG: 4977	Vitamin digestion and absorption	2	0.0186

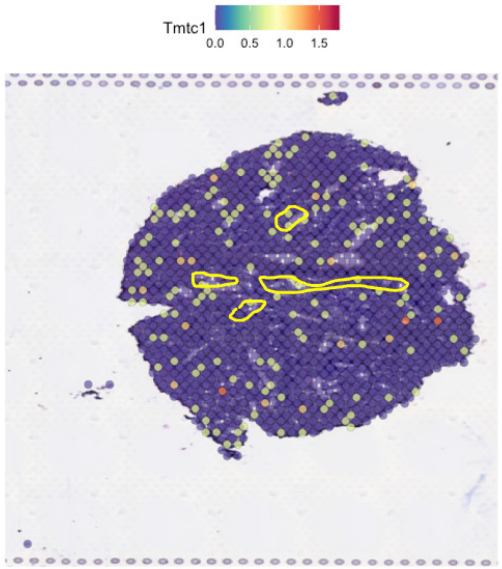
KEGG: 270	Cysteine and methionine metabolism	2	0.0346
KEGG: 350	Tyrosine metabolism	2	0.0346
KEGG: 4975	Fat digestion and absorption	2	0.0346
KEGG: 4015	Rap1 signalling pathway	3	0.0484
KEGG: 4810	Regulation of actin cytoskeleton	3	0.0484
KEGG: 5110	Vibrio cholerae infection	2	0.0484
KEGG: 5130	Pathogenic Escherichia coli infection	2	0.0484
KEGG: 5416	Viral myocarditis	2	0.0484

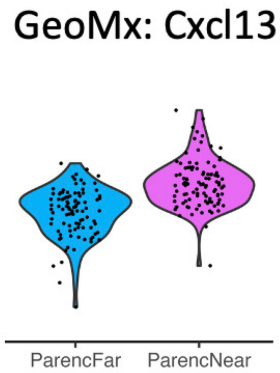
Log2 fold change >1.5 or <-1.5, adjusted p value <0.05.

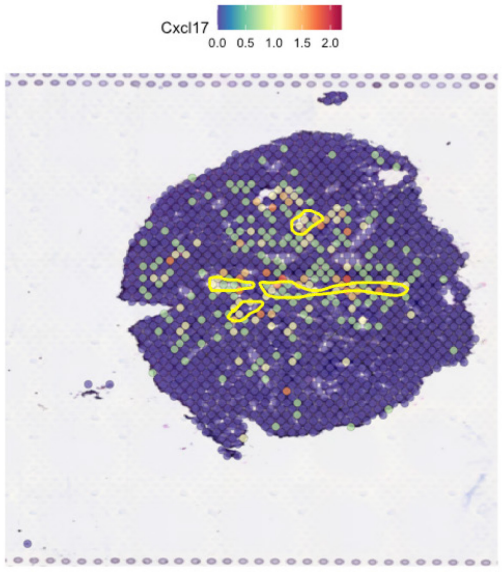
a


GeoMx: Tmtc1
1
ParencFar
ParencNear

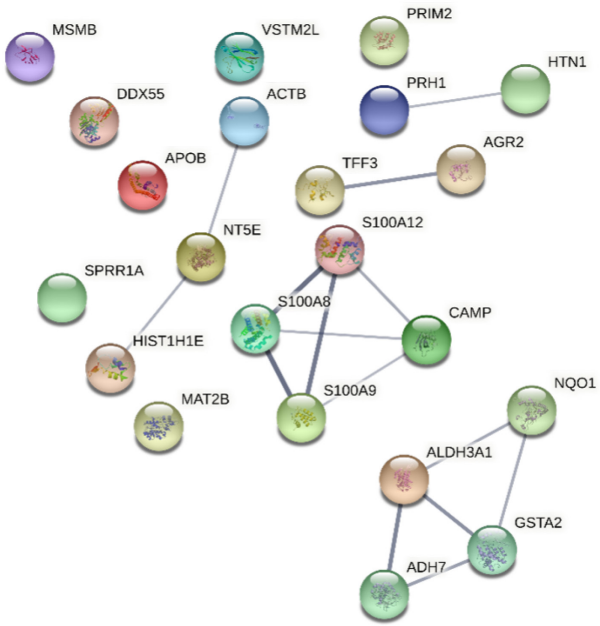
b


Tmtc1
0.0 0.5 1.0 1.5


GeoMx: Cxcl13
ParencFar
ParencNear


Cxcl17
0.0 0.5 1.0 1.5 2.0

Supplementary Fig. 1: Cross-validation of *Tmtc1* as a marker of proximal parenchyma as well as *Cxcl13* as marker of distal parenchyma, between *nanoString* GeoMx and *10x Genomics* Visium ST platforms. **a**, Expressions of *Tmtc13* and *Cxcl13* in proximal (ParencNear) and distal parenchyma (ParencFar) in GeoMx. **b**, Expressions of *Tmtc13* and *Cxcl13* in murine lungs using Visium platform. Areas inside the yellow circles represent the airways confirmed by histological observation.



Supplementary Fig. 2: String analysis of genes significantly altered by smoke exposure in sputum (FDR<0.05).