From the:
Comprehensive Pneumology Center and the Institute for Lung Biology and Disease at the
Helmholtz Zentrum München and the Ludwig-Maximilians-University

Dissertation
zum Erwerb des Doctor of Philosophy (Ph.D.)
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

Dissecting lung fibrogenesis and regeneration through the lens of single cell transcriptomics

vorgelegt von:

Maximilian Strunz

.....

aus:

Memmingen, Germany

.....

Jahr:

2020

_____

First supervisor: *Prof. Dr. med. Jürgen Behr*

Second supervisor: *Prof. Dr. rer. nat. Silke Meiners*

Dean: **Prof. Dr. med. dent. Reinhard Hickel**

Datum der Abgabe:

13.11.2019

_____

Datum der Verteidigung:

06.04.2020

_____

Omnis cellula e cellula.
Where a cell arises,
there a cell must have previously existed.

Rudolf Ludwig Carl Virchow - 1858

Contents

Statutory declaration/Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation mit dem Titel

Dissecting lung fibrogenesis and regeneration through the lens of single cell transcriptomics

von mir selbständig und ohne unerlaubte Hilfe angefertigt wurde.

Außerdem erkläre ich hiermit, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist, und dass ich mich nicht anderweitig einer Doktorprüfung ohne Erfolg unterzogen habe.

München, 11. November 2019

Maximilian Strunz

Confirmation of congruency

I hereby declare that the electronic version of the submitted thesis, entitled

Dissecting lung fibrogenesis and regeneration through the lens of single cell transcriptomics

is congruent with the printed version both in content and format.

München, November 11, 2019

Maximilian Strunz

Abbreviations

The following list includes the key abbreviations and acronyms used throughout the thesis.

Abbreviation	Meaning
ADI	alveolar differentiation intermediate
ALI	acute lung injury
ARDS	acute respiratory distress syndrome
AT1	type I alveolar epithelial cell
AT2	type II alveolar epithelial cell
ATP	adenosine triphosphate
CD	cluster of differentiation
cDNA	complementary deoxyribonucleic acid
DPLD	diffuse parenchymal lung disease
ECM	extracellular matrix
EEA	extrinsic allergic alveolitis
EMT	epithelial-to-mesenchymal transition
FDR	false discovery rate
ILD	interstitial lung disease
IPF	idiopathic pulmonary fibrosis
IPAF	interstitial pneumonia with autoimmune features
MHC	major histocompatibility complex
NEB	neuroepithelial body
PAGA	partition-based graph abstraction
PNEC	pulmonary neuroendocrine cell
PC	principal component
PCLS	precision-cut lung slices
PCR	polymerase chain reaction
RNA	ribonucleic acid
RNA-seq	RNA sequencing
scRNA-seq	single cell RNA sequencing
TGFb	transforming growth factor beta
UMI	unique molecular identifier

Summary

Single cell RNA sequencing studies have revealed a plethora of novel findings and leveraged the way in which we study disease development and progression. In the recent past, and with increasingly smarter protocols for single cell transcriptomics, such investigations gave insights into complex cellular composition of intricate biological systems, including alterations in cross-cellular interactions, and led to the discovery of rare but functionally important cell types.

The goal for this present work was to comprehensively study the course of lung injury and subsequent regeneration by using the bleomycin-mediated lung injury mouse model. This model comprises three phases: acute lung injury with the initiation of inflammation, transient fibrogenesis, and its resolution over a 4-8-week time course. Using single cell RNA sequencing and novel bioinformatic tools, the longitudinal sequence of lung injury and regeneration has been charted for the first time. With daily sampling, this cell type- and time-resolved strategy shed new light on cellular and molecular processes involved in lung injury and regeneration, and led to the discovery and identification of a transiently emerging cell state. During alveolar regeneration, epithelial progenitor cells transdifferentiated into squamous-shaped intermediate cell states, marked by high expression of Keratin-8, which were found to be exclusively present during the fibrogenic phase and to exert cellular interactions with both the mesenchyme and macrophages. By computational modeling of cell fate trajectories and by employing cell state predictions using RNA velocity, two cellular origins for the novel cell state were discovered. Both type-II alveolar epithelial cells, as well as airway-derived club cells were identified to transcriptionally converge into the Keratin-8 expressing cellular intermediate state.

In the bleomycin-injured lung, one of the most induced proteins is Emilin-2, an extracellular matrix component with key regulatory functions to provide a provisional scaffold for tissue regeneration. With the help of a constitutive knock-out mouse, the injured and regenerating lung was investigated at different time points of the model. The functional characterization of Emilin-2 was approached using different tools/methods, including tissue histology, flow cytometry, or single cell RNA sequencing. First observations and data presented in this work hint at the functional involvement of Emilin-2 in lung regeneration by impacting the immune response after injury.

1 Introduction

The Italian writer Giovanni Papini knew: Breathing is the greatest pleasure in life. Although Papini made his statement at times when tuberculosis and lung infections were epidemic in Europe, his words still remain meaningful in our days. In a world with increasing environmental pollution and toxic exposures, respiratory diseases and cancers are on the rise and impose a major burden for both the patient and society. Many among those patients suffer from a significant shortage of breath and would undoubtedly consent with Papini's words. In the World Health Statistics of 2016, published by the World Health Organization (WHO), four respiratory diseases are ranked among the top ten global killers: Chronic obstructive pulmonary disease (COPD), lung cancer, as well as tuberculosis and lower respiratory infections [1]. However, although not as potent in increasing prevalence, other forms of lung disease are also on the rise, which makes respiratory research indispensable.

Conducting research on human lung material, however, is restricted due to ethical reasoning. Hence, research is dependent on models that mimic disease to allow for easier estimation of dynamical processes during disease formation and progression. Here, by using the bleomycin-induced lung injury mouse model, injury and subsequent regeneration can be studied at controllable condition. Unfortunately, the interpretation of data is not solely complicated as the bleomycin-injured lung regenerates over a time frame of up to 56 days, which makes a comprehensive study over time difficult. By studying defined time points, the outcome is usually averaged over these single snap shots and fails to reflect complex and dynamic phases during lung regeneration. Combining single cell RNA transcriptomics with rational sampling, such processes can be studied at improved temporal and cellular resolution. Currently, many efforts are made to chart cellular atlases for both the human and the mouse lung, including a human cell reference map, or an aging lung cell atlas. The individual cell types that are present in the lung will be introduced in the following sections, including the bleomycin mouse model, respiratory disease, and single cell RNA sequencing.

1.1 Gross anatomy of the normal lung

The lungs are paired but asymmetrically shaped organs in the chest cavity that facilitate uptake of fresh oxygen into the bloodstream and the release of carbon dioxide produced by the body. Morphologically, the lungs can be divided into structural units describing the

parenchymal tissue for active diffusion of gases (oxygen and carbon dioxide) and airways, including trachea and bronchi. Anatomically, the lung is compartmentalized into segmental lobes and the airway tree, which is branching into different generations of bronchial tissue with varying structure and function (Fig.1a). The tracheobronchial airways summarize bronchioles that furcate from bronchi and further subdivisions into bronchioles and terminal bronchioles (all defined non-respiratory). In contrast, distal to the tracheobronchial tree, the respiratory bronchioles guiding into the alveolar ducts and ultimately into the alveolar sacs/alveoli comprise the pulmonary acini or parenchymal lobes - the area for active gas exchange (Fig.1b) [2, 3].
On cellular levels, both airway and alveolar regions reveal distinct cell populations with varying cell frequencies and cellular functions. Different cell types play different pivotal roles in maintaining tissue homeostasis, defense actions against inhaled exposures and in regeneration processes.
Airway epithelium can be classified by its appearance and cellular composition as stratified, pseudostratified, or simple epithelium (Fig.1c). Simple epithelium describes the concatenation of epithelial cells along the basement membrane, whereas stratified epithelium is characterized by multiple lines of cells [4]. By appearance, the pseudo-stratified epithelium resembles much a simple epithelium as it only comprises a single cell layer, however, the cells' nuclei are not solely positioned towards the basal cell end and hence give the impression of a stratified epithelium [4]. Stratified epithelium has been described in human airways but is more prominent during disease [5]. Under normal conditions, human airways are lined with pseudostratified epithelia that change to simple epithelia towards distal direction. In contrast to humans, mouse airway epithelia are of simple epithelium character and lack the pseudostratified cell appearance [6, 7]. Distal airway cell shapes are described as cuboidal, columnar, or squamous [7].

1.1.1 Airway epithelial cells

Non-secretory airway epithelial cells

Ciliated cells. The upper, conducting airways of humans are mainly lined by non-secretory ciliated cells but contain to a lesser number secretory epithelial cells, including club cells, goblet cells, as well as brush cells, and neuroendocrine cells [8]. In the mouse, by contrast, non-secretory cells are limited to the trachea and secretory cells are the most frequent cell type of the airway [8, 9]. Non-secretory ciliated cells are lined strategically with high abundance along the upper airway and fulfill a barrier function [10]. Using their cilia and microvilli that consist of both structural and motor proteins responsible for an orchestrated directional beating at the apical cell end [8], pathogens and other inhaled particles are

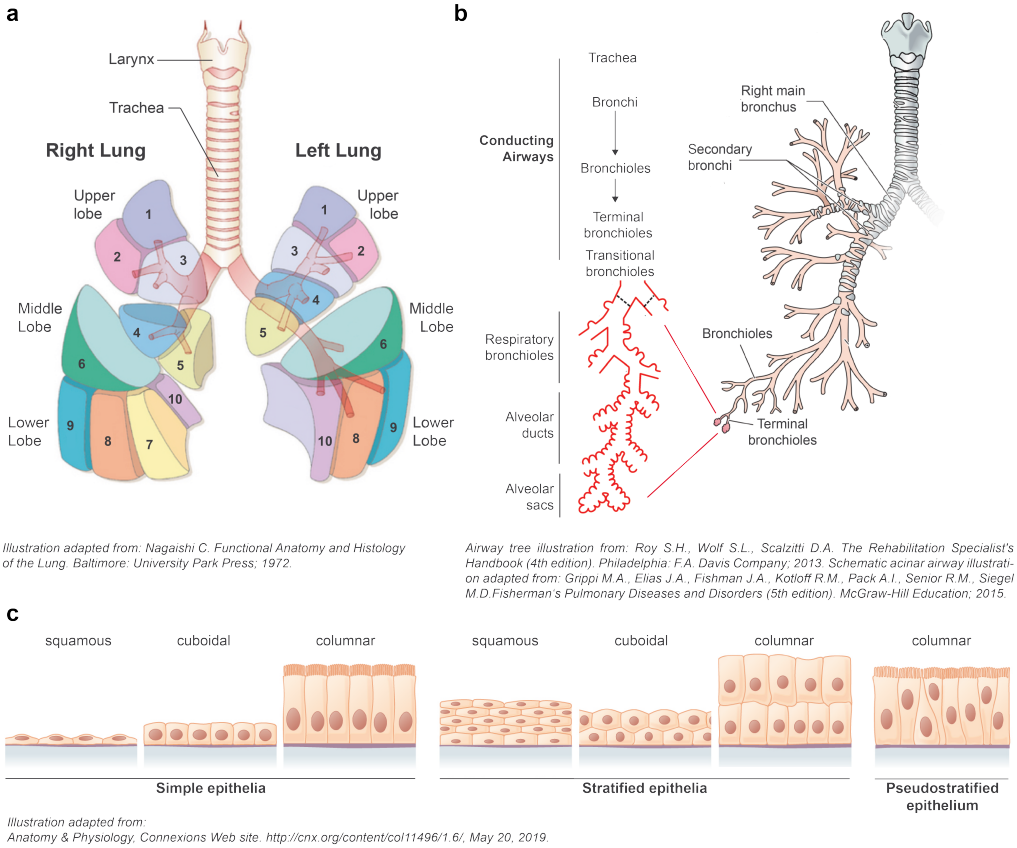

Illustration adapted from: Nagaishi C. Functional Anatomy and Histology of the Lung. Baltimore: University Park Press; 1972.

Airway tree illustration from: Roy S.H., Wolf S.L., Scalzitti D.A. The Rehabilitation Specialist's Handbook (4th edition). Philadelphia: F.A. Davis Company; 2013. Schematic acinar airway illustration adapted from: Grippi M.A., Elias J.A., Fishman J.A., Kotloff R.M., Pack A.I., Senior R.M., Siegel M.D.Fisherman's Pulmonary Diseases and Disorders (5th edition). McGraw-Hill Education; 2015.

Illustration adapted from:
Anatomy & Physiology, Connexions Web site. http://cnx.org/content/col11496/1.6/, May 20, 2019.

Figure 1: Gross human lung architecture, the complex branching of airways, and morphologic specifics of respiratory epithelial cells. **a**. Anatomical representation of the human lung with ten segmental lobes for both the right and left lung. **b**. Bronchopulmonary airway tree branching into continuously narrowing smaller airways with each generation; larger conducting airways span from trachea to terminal bronchioles, ultimately changing via the transitional bronchioles into the acinar airways, including the smallest respiratory bronchioles, alveolar ducts, and alveolar sacs. **c**. overview of the different epithelia types found in the lung; simple and stratified epithelia can be described by distinct cellular shapes; pseudostratified epithelium features a columnar appearance with all cells attached to the basement membrane. *Figure credits are indicated under the respective panel.*

removed from the lung [10]. This mucociliary clearance is only possible as ciliated cells and secretory airway cells synergize. If this balance is impaired, by e.g. ciliary dysfunction, cell disruption, or abnormalities in fluid secretion [8], airways can obstruct and react with immune reactions and hence contribute to disease manifestation and the development of chronic pulmonary diseases, such as asthma, chronic obstructive pulmonary disease (COPD), pulmonary or cystic fibrosis, or primary ciliary dyskinesia [8, 11–13]. The disease aspects relevant for this work are described with more detail in section 1.2. Ciliated cells have been reported to stem from basal and club cell progenitors [8, 14, 15]. Differentiation of

ciliated cells occurs via Notch signaling [16–18] by activation of the Notch receptors with its specific ligands Jagged1 and Jagged2 [16–18]. When the Notch signals are missing, ciliated cell differentiation is induced by implied basal and club progenitors [8, 14, 15]. By contrast, when Notch signaling is active, first secretory cells are differentiated, and with higher Notch activity the goblet cell population [8]. Characteristic for the ciliated cell is the assembly of ciliary proteins [19–22], as already mentioned, but also the activation of transcription by the regulatory factor X (RFX), the MYB proto-oncogene (MYB), and the forkhead box J1 (FOXJ1) transcription factors [20, 21]. Such examples of transcriptional network activities already highlight the complex inter-cellular cross-talks and cellular plasticity, which are important mechanisms needed to react upon abnormalities of foreign, or external stimuli.

Basal cells. In contrast to ciliated cells, basal cells are less malleable in shape as the basal cell shape is dependent on the degree of constriction of the airway. This cell type accounts for approximately 30% of airway epithelial cells [6, 15], which are quite undifferentiated at normal condition, marked by the expression of cytokeratin-5 (KRT5), -14 (KRT14), and the transcription factor Trp-63 (p63) [23, 24]. A main function of this cell population is the anchorage of columnar cells into pseudo-stratified epithelia [25] and the maintenance of the epithelial barrier as basal cells are in positioned in direct contact with the basement membrane [9, 10]. Basal cells can withstand injury or other insults with high persistence thanks to hemidesmosomes, via which they are tightly rooted on the basement membrane [9, 10]. Given their high resilience, it becomes apparent that basal cells exert a pivotal role during mediating inflammation and regeneration. It is therefore consequential that basal cells can function as stem or progenitor cells, at least in the upper airway, with the ability for self-renewal and trans-differentiation into other specialized cell lineages, enclosing secretory and ciliated cells [6, 9, 10, 26]. As for murine airways, basal cells are restricted to the trachea and uppermost airways - likely due to much smaller body size [15]. They are located between secretory, ciliated and neuroendocrine cells [23]. Differences in anatomical architecture as well as cellular composition between the mouse and human lung has been well described by *Rock J.R.*, *Randell S.H.*, and *Hogan B.L.M.* in 2010 [15].

Chemosensory epithelial cells

Brush cells. Although not highly innervated, the lung requires some chemosensory mechanisms to fulfill its proper function. As we breathe, our respiratory system constantly filters and warms up the inhaled air and guarantees appropriate conditions in terms of oxygen supply. In the conducting airway epithelium both brush cells and pulmonary neuroepithelial bodies (NEBs), see below, exert such chemosensory functions. The brush or tuft cell is named by its cell shape or the possession of large and long microvilli, respectively [27], and senses bacterial products, such as quorum sensing molecules [28]. Despite being very rare in both

the mouse and human airways [29], studies have suggested a particularly important role for these solitary cells, which is chemoreception of the airway lining fluid [28,30–32]. Brush cells sense with their G-protein coupled T1R and T2R receptors [31], ultimately leading to the activation of the vagal nerve to initiate defensive respiratory reflexes, such as expectoration [33].

Pulmonary neuroendocrine cells. Notwithstanding that these cells are solitary, pulmonary neuroendocrine cells (PNECs) tend to form cell clusters of 2-25 cells, termed neuroepithelial bodies (NEBs) [34,35]. PNECs are enriched with vesicles packed with neurotransmitters and other functional products [36] and exhibit only small cell surface areas to the airway lumen with small microvilli. Both PNEC and NEB functions vary throughout development and life [34, 35, 37]. Postnatally, NEBs are capable of sensing oxygen levels [36], with PNECs being the oxygen sensors detecting hypoxic conditions [36]. While there is proof for mechanistic actions of these cells in the developmental and postnatal lung, it remains to date unknown whether PNECs or NEBs are able to sense oxygen or carbon dioxide levels in the adult lung [34–36].

Secretory airway epithelial cells

Club cells. Within the group of secretory epithelial cells there are serous and mucous cells. Club cells, formerly termed Clara cells, are non-ciliated secretory cells filled with membrane-bound granules. The cell's main function is the secretion of among others peroxidases, lactoferrin, albumin, lysozyme, and proline-rich proteins [38–40]. Club cells are studied in great detail and with much effort as recent findings have suggested various important functions for club cells in lung development, homeostasis, and especially upon injury and subsequent tissue repair [15, 26, 39]. Conventionally, club cells are characterized by the expression of SCGB1A1 but there is growing evidence for highly heterogeneous cell variants within the club cell family [41]. First subsets were identified thanks to the differentially expressed secretoglobin family proteins SCGB3A1 and SCGB3A2 [42].

Goblet cells. Similar in function to the club cells, goblet cells constitute the mucous cell type of the airway. As the main producer of mucus, goblet cells are specialized to secrete proteins and other factors, but mostly mucins (over 20 mucin genes are known), in order to guarantee proper tissue barrier protection and tissue homeostasis [43]. Under normal physiologic condition, goblet cells produce only a balanced amount of mucus, with MUC5AC and MUC5B being the most abundant [43–45], however, mucus can be rapidly secreted upon allergic or inflammatory stimulus [9, 46]. In response to persistent stimuli, a higher number of goblet cells can be present, leading to excessive and highly viscid mucus pro-

duction and consequently to impaired mucociliary clearance and goblet cell metaplasia, which represents a clinical characteristic for some pulmonary diseases [9, 43, 44, 46]. As mentioned, basal cells can give rise to goblet cells but this is regarded to be valid for upper airways only [6, 47–49]. In the lower airway ducts, club cells are thought to be the leading progenitor or stem cell population [46, 50]. Generally, the epithelial cell compartments of both airway and alveolar tissue underlie complexly orchestrated molecular and physiologic processes.

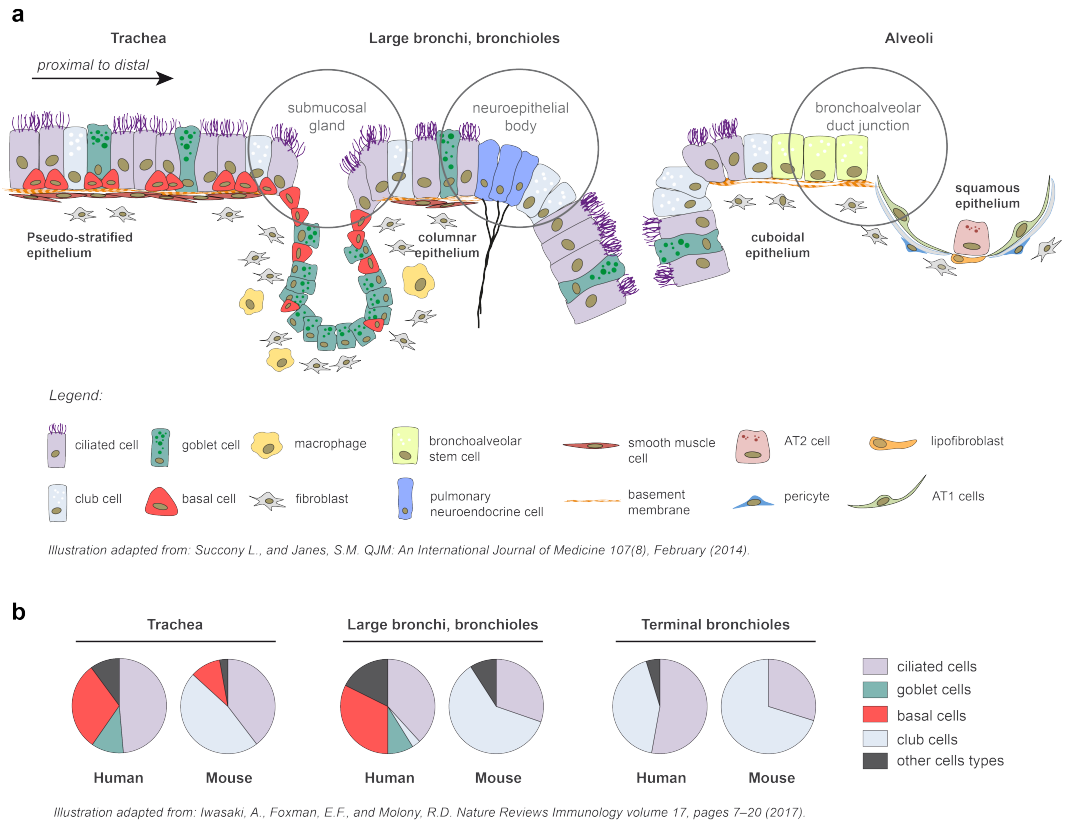

Figure 2: Overview of the airway architecture and cellular composition. **a** Schematic representation of (mouse) airway epithelial cell populations spanning from tracheal cell composition over large bronchi and bronchioles through the delicate epithelia of the alveolar compartment. Cell niches are highlighted in circles. **b** Cellular differences in airway cell composition between mouse and human with respect to different anatomical location. Pie chart representation with approximate proportions of the main airway cell types. Basal cells are rare in the mouse upper airway, goblet cells almost not existing. The human airway is void of club cells in tracheal and large bronchi regions. *Figure credits are indicated under the respective panel.*

1.1.2 Alveolar epithelial cells

The major cell populations in the alveolar region are alveolar type I and type II epithelial cells, endothelial cells, interstitial cells, and immune cells, including alveolar macrophages and dendritic cells.

Type I and II alveolar pneumocytes. Alveolar type I (AT1) cells are thinly shaped cells that span the alveolus and facilitate gas exchange thanks to their direct contact with endothelial cells from the pulmonary capillaries [51]. To enable optimized gas exchange, AT1 cells exhibit a large cell surface area and are the most abundant type of alveolar epithelial cells. Alveolar type II (AT2) cells, by contrast, are rather small and cuboidal in shape with the main function of producing and secreting surfactant for the alveolar space [52, 53]. This surfactant is composed of proteins and lipids to prevent alveolar collapse during breathing and to protect the alveolar architecture [54, 55]. Alveolar epithelial cells were shown to be relatively fragile and, depending on the insult or injury, different mechanisms are exerted to repair destructed alveoli [56]. Using lineage tracing studies in mice, it could have been demonstrated that AT2 cells are able to repopulate lost AT1 cells, which has been substantially studied after bleomycin-mediated lung injury [57–59]. Interestingly, it seems that AT2 cells cannot only give rise to AT1 cells but also generate more AT2 cells after injury by clonal expansion, suggesting a potential role for AT2 pneumocytes as stem or progenitor cells within the alveolar compartment [57]. The formation of newly emerging AT1 cells was proposed to occur via initial flattening of the cell, followed by folding into the alveolar space [60].

1.1.3 Pulmonary/alveolar interstitium

Pulmonary endothelial cells

Under regular conditions, the endothelium represents a tight barrier between the circulating blood and the vascular wall. In larger veins or arteries, smooth muscle cells enclose the vessels, in the fine capillary structures of the alveolus, however, pericytes are the supporting cell type [61]. Endothelial cells are squamous epithelial cells located at the inner vessel wall [62] and are among the most abundant cell types in the lung. Pulmonary endothelial cells have various important functions, including the production and metabolism of nitric oxide, which is a pivotal regulator of endothelial cell response [63]. Moreover, pulmonary endothelial cells closely control the transit of substances from the blood to both the alveolar region and the interstitium [62, 64, 65]. Upon injury or other pathologic state, this barrier gets disabled, resulting in expression of adhesion molecules and the initiation of complex inflammatory reactions by pulmonary endothelial cells [64, 66–68]. Such impairments

accelerate the permeability of the endothelial monolayer and can result in tissue edema and other complications.

Pulmonary fibroblasts and myofibroblasts

Pulmonary fibroblasts are mesenchymal interstitial cells that manifest with a high degree of heterogeneity. The fibroblast's role in lung homeostasis is critical: as they directly interact with epithelial and endothelial cells [69], they can instantaneously respond to extracellular matrix (ECM) alterations and lung architectural changes. Upon activation stimulus, fibroblasts initiate an eclectic cascade with to date not fully understood dimensions. As important mediators after tissue injury and vital players during regeneration, pulmonary fibroblasts are among the most studied cell type in respiratory research. Expressing various receptors for pro-fibrotic molecules, including transforming growth factor beta (TGFb) [70], or platelet-derived growth factor (PDGF) [71, 72], pulmonary fibroblasts can migrate and differentiate into active myofibroblasts. These cells are highly contractile and express alpha smooth muscle actin (alpha-SMA), which is up-regulated upon stress stimuli [73]. In pulmonary fibrotic disease, myofibroblasts accumulate in fibrogenic patches where they assemble excessive amounts of ECM proteins. Several origins for this plasticity have been described: (i) myofibroblast differentiation from pulmonary fibroblasts (as outlined above), (ii) circulating fibrocytes migrating to the site of injury and differentiating into myofibroblasts [74], or (iii) epithelial to mesenchymal transition (EMT), with alveolar epithelial cell transdifferentiation after injury into myofibroblasts [75]. Despite these diverse hypotheses, the findings are conflicting [59, 76, 77]. To highlight is an *in vivo* study, in which mice that were treated with bleomycin (cf. section 1.4) did not show clear evidence of EMT based on both AT2 (alveolar) and club cell (airway) transdifferentiation [59]. Another investigation proposes pericytes as the origin for myofibroblasts [78] due to similar gene expression between these cell types.

Pulmonary macrophages

Pulmonary macrophages fulfill imperative functions in the non-specific immunological defense by clearance of foreign matter or cellular debris and antigen presentation to the host's immune system. Pulmonary macrophages may be classified interstitial macrophages (located between alveolar and vascular epithelium), and alveolar macrophages but other macrophage subpopulations with largely unknown functions are known, including intravascular and pleural macrophages, or airway macrophages [79–82]. Depending on induction with inflammatory cytokines, pulmonary macrophages can be further classified into M1 (classically activated or pro-inflammatory), or M2 (alternatively activated or anti-inflammatory) phenotypic macrophages [83–87], with the latter being the most incumbent phenotype under

both regular or regenerative state [88–90].

Interstitial macrophages. As the designation suggests, these cells are inhabiting the interstitium where they can interact with lymphocytes and actively participate in antigen presentation [91–94]. Due to this interaction, interstitial macrophages exert greater involvement during pro-inflammatory reactions than alveolar macrophages, e.g. by the secretion of interleukin-1 (IL-1) and -6 (IL-6), highlighting their immunoregulatory role [92, 95, 96]. Interstitial macrophages are capable of expressing M2-like factors and acting inhibitory on dendritic cells, suggesting that the interstitial macrophages are involved in adaptive immunity [92, 93, 97].

Alveolar macrophages. These important lung cells are located in the alveoli and are the primary cells to rid trapped particles or pathogens via the mucociliary clearance [98–100]. Alveolar macrophages contribute to antigen presentation and express major histocompatibility complex (MHC)-class II antigens but less so in the presence of professional antigen presenting cells, including pulmonary dendritic cells and interstitial macrophages [92–95, 101, 323]. Their prime immunologic feature is the avoidance of allergenic overload in pulmonary lymphoid cells [103]. High expression of M2 phenotypic scavenger receptors, such as CD68, or CD204, plus low expression of M1 phenotypic markers, like Ly6C [101, 323], typically characterize alveolar macrophages. They further stand out due to slow turnover rates and longevity [322]. Generally, one can differentiate between resident and non-resident alveolar macrophages. Resident alveolar macrophages can augment the number of alveolar macrophages in the lung by cell division or via the egress of interstitial macrophages or myeloid-derived monocytes, which give rise to the implied non-resident macrophage populations [109, 110]. After lung injury, the resident population exhibits the stronger resilience and persistence, suggesting that they contribute to wound healing and repair mechanisms [105–108]. Upon massive loss of all alveolar macrophages, including the tissue resident ones, cells are replenished by mainly monocytes or interstitial macrophages [109].

1.1.4 Basement membrane and the extracellular matrix

Generally, the extracellular matrix (ECM) possesses supportive tissue functions but is influenced by the factors and proteins that are secreted by their encompassing cells. In large airways and the proximal respiratory tract, collapse of the airway is impeded by the tracheal cartilages, composed of collagens and hydrophilic proteoglycans [111]. Among the most abundant ECM proteins are collagen and elastin, both of which allow for structural stability without losing elasticity [112–114]. In the distal alveolar space, the ECM composition is utterly critical and must manifest a structure allowing for optimized gas exchange. As compared to the airway ECM, the alveolar region frequently exhibits fused basement membranes

between alveolar epithelium and endothelium [115]. The ECM and the basement membrane represent a scaffold for surrounding cells, which are actively influenced by the ECM composition with regards to cell migration, proliferation and differentiation [116, 117]. Changes in ECM organization and architecture, such as excessive ECM deposition, decrease gas diffusion efficacy [118] and accelerate the pathologic condition and disease progression.

1.2 Respiratory disease

Lung or respiratory disease affects millions of people worldwide. Patients suffering from respiratory disease or who even developed chronic forms of it, are seriously retrenched in life. While some types of disease affect the airways, for instance, others are rather morbid in the parenchymal or vascular region of the tissue. Although being multifaceted with regards to disease onset and progression, many disorders share disease mechanisms. In this manner, some respiratory diseases elicit the decrease of the lung's elasticity (as observed in pneumonia or pulmonary fibrosis), others, by contrast, obstruct the airways (as in e.g. asthma or chronic obstructive pulmonary disease). Both forms result in reduced air flow and diminished oxygen supply to the body. Despite established knowledge about disease-triggering risk factors, including smoking, environmental pollution, or yet genetic factors, most molecular and etiological principles of respiratory disease remain largely unknown. In virtue of the immense diversity of respiratory disease, the focus will be set on only diseases that impact the herein presented work.

1.2.1 Diffuse parenchymal lung disease

Various chronic lung diseases, such as interstitial lung disease (ILD) can be classified based on distinct features, yet many share certain characteristics, including inflammation and concomitant fibrosis [119]. Interstitial lung disease, or diffuse parenchymal lung disease (DPLD), is marked by progressive tissue scarring and summarizes up to 200 different subforms of the disease [120, 121]. The lung is an immunocompetent and metabolically active organ subserving gas exchange, and filtration for toxic substances and blood components. ILD patients present with evidence of inflamed alveolar tissue, which can be intensified by ongoing infections of the lung [122]. In the case of persisting inflammation and gradual scarring/fibrotization of tissue in the alveolar region, interstitial lung disease can result in lung fibrosis. In consequence, patients exhibit impaired gas exchange capacity, phases of exacerbations, and sometimes heart failure owed to chronic lung disease. Among the class of ILDs, idiopathic interstitial pneumonias (IIPs), including the frequently occurring idiopathic pulmonary fibrosis (IPF), manifest very progressive disorders with unknown causes. IPF is

an end-stage lung disease with a median survival of 2-3 years after diagnosis [123], which cannot be cured but mitigated with drugs that decelerate disease progression, two of which are clinically approved and in practical use: pirfenidone and nintedanib [124, 125]. However, these drugs have strong side effects and are without prospect for regression so that the only curative therapy for patients with poor prognosis is ultimately lung transplantation. Although being considered rare diseases, DPLDs are among the top 40 global diseases with an increase of years of life lost in 2013 of 86% compared to 1990 (Fig.3) [126]. Years of life lost measures premature mortality by identifying years of potential life lost due to causes of premature death [333].

1.2.2 Acute respiratory distress syndrome

A major risk factor and potentially life-threatening event can be the development of an acute respiratory distress syndrome (ARDS), a disorder often linked to variety of lung injury insults leading to fulminant morbidity and acute lung injury (ALI). Due to the massive structural damage in the alveolar tissue, including the extracellular compartment, the lung fails to supply the body with sufficient amounts of oxygen, which makes breathing difficult and usually requires mechanical ventilation to assist the patients. Although the pathologic hallmarks of ALI are epithelial and endothelial dysfunction, including pulmonary edema [127], ALI/ARDS are not classified as interstitial lung disease. Some patients who survive can fully recover, but others may exhibit severe end-stage lung abnormalities as they are frequently observed in ILD patients.

Although the term respiratory disease conflates a wide range of distinct diseases and their subtypes into one hypernym, there are demands for distinguished scientific approaches to better resolve and understand the underlying patho-mechanisms. As access to human material is usually restricted or ethically reprehensible, research is conducted mainly by the use of various *in vitro* and *in vivo* models. Animal models have been proven to be imperative tools for the identification of key processes during disease. There are several models to mimic human ILD features in the mouse but thanks to its pulmonary cytotoxicity, bleomycin-induced lung injury is the most frequently used model to study acute lung injury and subsequent tissue repair and remodeling (described in 1.4).

1.3 Bleomycin-induced pulmonary cytotoxicity

The glycopeptide bleomycin is an antibiotic compound that was isolated and obtained from *Streptomyces verticillus* in 1966 [128]. Thanks to its anti-tumoral properties, bleomycin has been used as a potent drug against various cancers, including Hodgkin lymphomas and

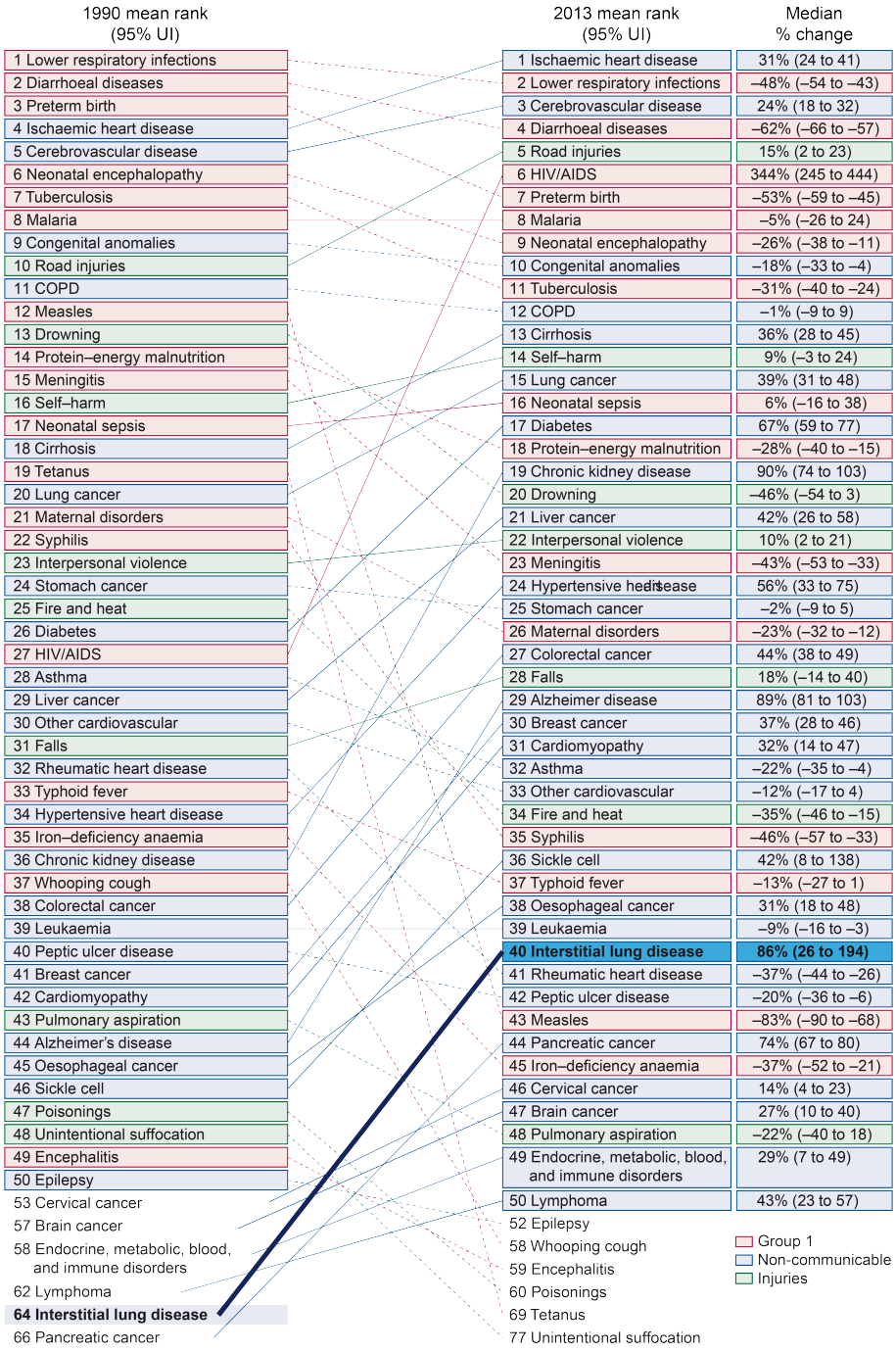

Figure 3: Top 50 causes of global years of life lost in 1990 and 2013 - data and table adapted from the Global Burden of Disease Study 2013; Lancet 2015; 385: 117-71.

germ cell tumors [129, 130]. However, the use of bleomycin is confined by the development of ILD in up to ten percent of treated patients [131–133].
The endonuclease bleomycin exerts anti-neoplastic effects by inhibiting DNA replication and repressing DNA polymerase activity, and by inducing single- or double-strand DNA breaks via a complex of bleomycin, oxygen, and iron [131, 134, 135]. More specifically, bleomycin-induced chelated iron reduces oxygen, which results in subtraction of hydrogen from the deoxyribose carbons, thereby releasing a DNA base and introducing the DNA strand break [134]. The high toxicity in the lung (and also in the skin) is due to a lack of the bleomycin hydrolase that enzymatically deactivates bleomycin. Clinical signs or histologic hallmarks of the so-called bleomycin lung include among others alveolitis, interstitial edema, intra-alveolar and interstitial fibrosis, metaplasia, and dysplasia of the alveolar cells [136]. The distribution of fibrotic lesions are found in both subpleural and periseptal areas [136]. With such features of interstitial lung disease, and the lack of notion via which mechanisms fibrogenesis occurs, there is warranty to make use of bleomycin for disease models that mimic IPF and ILD signs. Studying main pathogenic roads can be achieved using different *in vivo* models, unfortunately, no animal model will be able to fully reflect the human disease character. Nevertheless, bleomycin administration in mice has been proven successfully thanks to reproducible effects recapitulating lung injury and lung fibrogenesis [119].

1.4 Bleomycin mouse model of pulmonary fibrosis

The bleomycin mouse model, which represents the most commonly applied experimental disease model in C57Bl6/J mice [137, 138], has helped tremendously in finding new aspects about the cellular and molecular mechanisms associated with IPF and other ILDs, as well as in developing novel therapeutic approaches. As outlined in the paragraph above (1.3), bleomycin administration to murine lungs induces DNA strand breaks, leading to alveolar epithelial cell death, massive tissue injury with the initiation of complex inflammatory responses and gradual manifestation of fibrogenesis by fibroblast activation, collagen deposition, followed by tissue repair and almost complete regeneration of destructed tissue [138]. Single dose administration of bleomycin leads to transient pulmonary fibrosis [139] and reproduces several features of human ILD, defining this model to be very suitable for the study of both the development of fibrosis as well as its resolution over time (Fig4a). Repetitive administration, by contrast, does not allow for the quick resolution of inflammation and thus of the pulmonary fibrosis (Fig4a) [76, 140, 141]. In a nutshell, the bleomycin mouse model can be applied in different manners, depending on the research interest, as to whether systemic processes over time with regard to successful tissue restoration is of interest (acute model), or the development of non-regenerative permanent fibrosis with increased similarities to human ILD (chronic model).

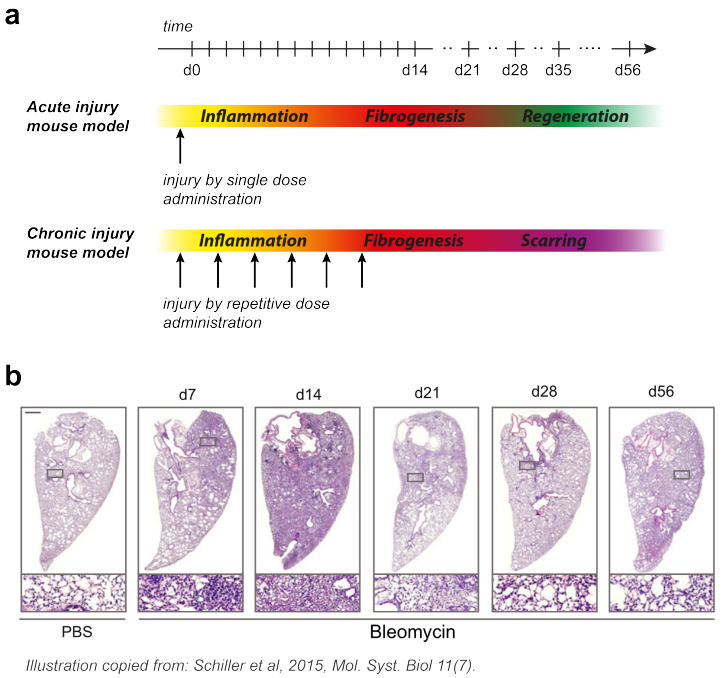

Illustration copied from: Schiller et al, 2015, Mol. Syst. Biol 11(7).

Figure 4: The bleomycin mouse model to induce acute lung injury and pulmonary fibrosis. **a** The acute injury model is initiated by single dose administration of bleomycin, resulting in transient fibrogenesis and subsequent tissue regeneration. Repetitive administration of bleomycin, including during the inflammatory and fibrogenic phase, induces excessive tissue scarring without regeneration. Typically assessed time points are day 3, 7, 14, 21, 28, and 56, but to better capture tissue changes, shorter intervals are also considered. **b** Histological assessment over time post injury. Compared to the PBS instilled control lung, aberrant extracellular matrix deposition and concomitant immune cell influx can be observed towards day 14, the peak of the inflammatory phase. At day 56, the treated lungs almost resemble those that have been treated with PBS only. **c** Schematic representation of the intratracheal bleomycin administration. Using a microsprayer, bleomycin is instilled via an intratracheal catheter as a highly dispersed aerosol. **d** Principle of the oropharyngeal aspiration route. Bleomycin is pipetted at the tip of the oropharynx after blockage of the nostrils using tweezers and pulling firmly at the tongue. This way the mice are forced to deeply aspirate the solution into the lung. *Figure credits are indicated under the respective panel.*

Administration of bleomycin to mice can be achieved by intratracheal instillation, intratracheal microspray aerosolization [144], or oropharyngeal aspiration [142, 143, 147]. In addition, intranasal application is possible, however, scarcely described in the literature. Although intratracheal bleomycin delivery is favorably used [146], the advantages of aerosolized instillation via the trachea (Fig4c) [144] outweigh the use of regular intratracheal injection [142, 145]. A recent study reported on oropharyngeal aspiration routes (Fig4d) that outperformed intratracheal delivery methods due to the ease of handling, the diminished animal burden, and the downsizing of experimental animal death [142].
For the acute bleomycin model, and in order to capture the whole process from lung injury to repair, there are certain time points that are commonly and classically used by the research community [139, 148–150]. At day three after administration of bleomycin, acute damages in the lung epithelium are found and first inflammatory and immune reactions are induced. Day seven denotes an intermediate time point characterized by early signs of fibrogenic actions and a severe immunologic response. At day 14, the phase of active fibrogenesis, several fibrotic foci are formed in the parenchymal tissue, marked by exuberant expression and deposition of ECM proteins. The climax of tissue fibrosis is at around day 21, when secreted and assembled ECM proteins constitute a provisional matrix for subsequent tissue repair. Gradually, from day 28 to day 56, the fibrotic tissue is remodeled and the lung parenchyma almost completely regenerated and fully functional again (Fig4a and b). Although the investigation interval spans the entire acute phase of lung injury and following tissue repair, many important processes might be missed by neglecting the days in between the observed time points. In fact, by studying the acute bleomycin-induced lung injury over time with higher temporal resolution, for instance daily, novel aspects and previously imprecise interpretations of injury or disease mechanisms can be better investigated and scrutinized - which will be a key point in this thesis.

1.5 Single cell RNA sequencing

After the completion of the Human Genome Project in 2003 [151, 152] and the release of the Human Protein Atlas in 2005 [153, 154], scientists are now gathering to create a cell-based reference of the human body - the Human Cell Atlas [155]. This highly collaborative initiative will likely accomplish what has been sought for a long time, namely the mapping of both the geno- and phenotype. Only recent advancements in novel single cell-based RNA-sequencing methods allow for such ambitious endeavors. While the power to study transcriptomic alterations has long been understood, it was previously impossible to deconvolute cellular heterogeneity in both healthy or diseased state [155]. Although same cell types within an organ can fulfill equal functions, the transcriptomic state of each individual cell may mirror solely the activity of only few active genes in that cell [156–158],

thereby defining heterogeneous cell type composition. Stochastic gene expression events that guide biological processes and define cell fate decisions [159, 160] are missed or neglected when performing bulk sequencing studies as these methods fail to resolve the cell-to-cell variability and the distinct transcriptomic signals of individual cells. However, the notion of such differences is required to understand disease development as many diseases can be associated to a dysregulation in protein transcription [161].

Single cell transcriptome sequencing has been initiated by the expansion of single cell-derived cDNA using either in vitro transcription (IVT, linear transcript amplification) or the polymerase chain reaction (PCR, exponential transcript amplification) [162, 163]. With the help of incorporating fluorescent dyes, these methods soon gave rise to DNA-microarrays [164, 165], which measured the increasing fluorescent signal after each round of cDNA amplification [166, 167]. Despite the fact that DNA microarrays and genome-wide association studies (GWAS) levered the discovery of unprecedented transcriptional diversity, only known cDNA sequences could be analyzed because of previously designed oligonucleotide probes. Starting from 2005, various high-throughput sequencing methods were published [168, 169], however, only after the next-generation sequencing platforms were introduced by Illumina, the global market leader for DNA sequencing, the interest for sequencing single cells on a widespread scale aroused. RNA-sequencing allows for highly and massively parallel sequencing of mRNA-derived cDNA fragments [170–173], thereby enabling unbiased profiling of qualitative and quantitative gene expression.

Extraction of single cells is an important but not trivial step for single cell RNA sequencing. Depending on the research interest and sample availability, different isolation procedures are commonly used, including fluorescent-activated cell sorting (FACS), or laser capture microdissection (LCM). While FACS can be the method of choice for precisely selecting cells with antibodies against known cell surface markers it may be too dependent on cell input number and large working volume. And while only distinct cells may be of interest and can be collected even from very solid tissues by LCM, microfluidics-assisted methods can convince with relatively low sample input, ease of handling, and cost efficiency. However, these droplet-based methods require a priori cell extraction from tissues with enzymatic digestion. Up to recently, more than 52 different single cell RNA sequencing methods have been described (Fig.5) [174], demonstrating their broad application range. Common to most methods is the library preparation step, which usually consists of (i) RNA extraction, (ii) mRNA enrichment, e.g. selection of only polyadenylated mRNA species [175], (iii) mRNA reverse transcription into cDNA, and (iiii) cDNA fragmentation. Due to high ribosomal RNA (rRNA) contents in a cell [175] and low reverse transcription rates (10-20% of transcripts) attributed to the Poisson sampling [176], the cell lysis step continues to be the vital factor deciding upon the sequencing results. To boost sequencing output and simultaneously keep the expenses at a minimum, both the 5' and 3' sequencing has been introduced [177, 178].

However, with the novel emergence of single cell-based methods, an improved accuracy in gene detection as well as effective removal of the PCR bias was needed. The solution is the introduction of so-called unique molecular identifiers (UMIs) and cell barcodes during the reverse transcription step [177–179]. These small, only few basepair-long sequences allow for direct molecular transcript counting as contrasted to read-based estimations, such as RPKM/FPKM (read/fragment per kilobase per million mapped reads) [176, 180]. Despite the vast choice of methods, droplet-based technologies transformed both basic and clinical research. With commercially available platforms, including the Fluidigm C1 system providing automated single cell analysis but at limited cell number, or the 10x Genomics Chromium system guaranteeing high throughput and highly efficient cell capture, there is no halt in studying and interrogating a sample's cellular diversity. One of the widely used droplet-based single cell RNA sequencing technologies is Drop-Seq [179], which is primarily due to its cost-efficiency, ease of handling, and high-throughput nature. This technology has been applied for the work presented in this thesis and is explained in more detail under 2.2.2.
The application of single cell RNA sequencing is a vivid field of research and currently much exploited as, for the first time, global gene expression differences can be studies within and across cell types [181–183]. Moreover, previously undescribed cell types or cell populations could have been identified, laying a foundation for potentially major breakthroughs in medical research, or giving rise to further research hypotheses. These include findings in the liver [184], the pancreas [180, 185, 186], the immune system [187, 188], the lung [189, 190, 197], the retina [179, 191], and the brain [192–196].

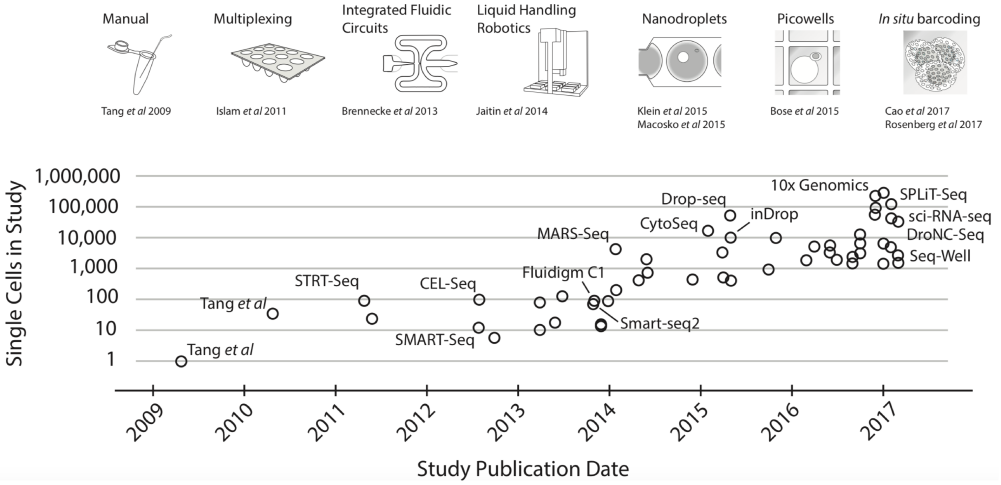

Illustration copied from: Svensson, Vento-Tormo, and Teichmann, 2018, Nature Protocols 13:599-604.

single cell RNA seq method	Reference	single cell RNA seq method	Reference
Tang protocol	*Tang et al. 2009*	Div-Seq	*Habib et al., 2016*
STRT-seq	*Islam et al. 2011*	LCM-seq	*Nichterwitz et al., 2016*
Smart-seq	*Ramskold et al. 2012*	scTrio-seq	*Hou et al., 2016*
CEL-seq	*Hashimshony et al. 2012*	scM&T-seq	*Angermueller et al., 2016*
Digital RNA sequencing	*Shiroguchi et al., 2012*	MASC-seq	*Vickovic et al., 2016*
Quartz-Seq	*Sasagawa et al., 2013*	Spatial transcriptomics	*Stahl et al, 2016*
DP-seq	*Barghava et al., 2013*	SureCell	*Illumina 2016*
Smart-seq2	*Picelli et al., 2013*	MATQ-seq	*Sheng et al., 2017*
Smart-seq/C1	*Pollen et al., 2014*	10x Genomics	*Zheng et al., 2017*
TIVA	*Lovatt et al., 2014*	Seq-Well	*Gierahn et al., 2017*
STRT-UMI	*Islam et al. 2014*	SPLiT-seq	*Rosenberg et al., 2017*
SCRB-seq	*Soumillon et al., 2014*	sci-RNA-seq	*Cao et al., 2017*
CEL-seq UMI	*Grün et al., 2014*	DroNc-seq	*Habib et al., 2017*
MARS-seq	*Jaitin et al., 2014*	sNucDrop-Seq	*Hu et al., 2017*
Drop-seq	*Macosko et al., 2015*	STRT-2i	*Hochgerner et al., 2017*
InDrops	*Klein et al., 2015*	snoRNA-seq	*Faridani et al., 2017*
Cyto-seq	*Fan et al., 2015*	Quartz-Seq2	*Sasagawa et al., 2017*
SC3-seq	*Nakamura et al., 2015*	GEO-seq	*Chen et al., 2017*
BAT-seq	*Velten et al., 2015*	REAP-seq	*Peterson et al., 2017*
SUPeR-seq	*Fan et al., 2015*	CITE-seq	*Stoeckius et al., 2017*
DR-seq	*Dey et al., 2015*	mcSCRB-seq	*Bagnoli et al., 2017*
G&T-seq	*Macaulay et al., 2015*	Niche-seq	*Medaglia et al., 2017*
Hi-SCL	*Rotem et al., 2015*		
SORT-seq	*Muraro et al., 2016*		
CEL-seq2	*Hashimshony et al. 2016*		
Patch-seq	*Cadwell et al., 2016*		
Patch-seq	*Fuzik et al., 2016*		
Smart-seq2-nuclei	*Krishnaswami et al., 2016*		
FRISCR	*Thomsen et al., 2016*		

Illustration adapted from: Ziegenhain et al, 2018, Brief Funct Genomics 17(4):220–232.

Figure 5: Overview of scRNA sequencing protocols: over the past decade, more than 50 sequencing protocols emerged; thanks to fluidics-based applications, thousands of single cells can be sequenced in a very short time. Top panel shows the most commonly used protocols with original publication date. Lower panel lists a variety of different protocols until 2017, including the latest innovations. Later protocols are usually adaptations from these. *Figure credits are indicated under the respective panel.*

2 Material and methods

2.1 Animal handling and bleomycin administration

All mice used in the herein presented investigations were purchased from Charles River, Germany, and maintained under pathogen-free conditions at the animal husbandry of the Helmholtz Zentrum München, Munich, Germany. Emilin-2 knockout animals were provided by Dr. Paolo Bonaldo (University of Padova, Italy). Mice were given food and water *ad libitum* at all times. All experimental procedures were conducted in accordance with the regulations of the Bavarian government and the guidelines of the European Union and are registered and approved under the registry 55.2-1-54-2532-130-2014 and ROB-55.2-2532.Vet_02-16-208 [198]. Control mice were treated with 80 μl of saline solution (phosphate-buffered saline, PBS), bleomycin-treated animals were administered with either 2 U/kg (oropharyngeal instillation) or 3 U/kg (intratracheal instillation) bodyweight bleomycin. Mice were instilled with either PBS or bleomycin after intraperitoneal injection of medetomidin 500 μg/kg body weight, midazolam 5 mg/kg body weight, and fentanyl 50 μg/kg body weight in sterile sodium chloride solution at a maximal volume of 100 μl via a 27 gauge cannula. Intratracheal instillation was performed using the Micro-Sprayer Aerosolizer, Model IA-1C (Penn-Century, Wyndmoor, PA). Oropharyngeal administration was achieved by bleomycin application to the oropharynx by pulling out the tongue and closing the nostrils to induce deep breathing through the oral cavity. Directly after treatment, mice were antagonized by subcutaneous injection of atipamezol 2.5 mg/kg body weight, flumazenil 500 μg/kg body weight, and naloxon at 1.2 mg/kg body weight, dissolved in sterile sodium chloride solution at a maximal volume of 100 μl via a 27 gauge cannula. Lung injury was induced by single-dose administration of bleomycin (Sigma Aldrich), dissolved in sterile PBS [198]. Treated animals were continuously observed during the experimental phase and sacrificed in case of abnormal behavior or noticeable signs of phenotypic changes caused by the administration [198]. Experimental end points were chosen based on the commonly investigated time points (days 3, 7, 14, 21, 28, and 56), or even daily to better track down changes at increased resolution of time after lung injury (days 1-14, 21, 28, 35, 42, and 56) [198].

2.2 Single cell analysis of the mouse lung

2.2.1 Generation of single cell suspensions from whole mouse lungs

Mice were sacrificed by cervical dislocation of the spinal column from the brain and dissected as described in the submitted manuscript [198]. After opening the thorax, the lung was perfused with sterile PBS through the right ventricle of the heart. For Drop-Seq, the right lung was tied off at the main bronchus and cut apart for tissue dissociation. When histology samples were collected, the left lung was fixed by slow inflation of the tissue with 4% paraformaldehyde and subsequently prepared for formalin-fixed paraffin embedded (FFPE) tissue blocks. As the Drop-Seq method requires a single cell suspension, the right lobes were minced and transferred for mild enzymatic digestion for 20-30 min at 37°C in an enzyme mix containing dispase (50 caseinolytic U/ml), collagenase (2 mg/ml), elastase (1 mg/ml), and DNase (30 μg/ml) (described previously in [201]). Inhibition of enzyme activity was achieved by adding 10 ml of PBS supplemented with 10% fetal calf serum (FCS). As described previously [198], single cells were harvested by straining the digested tissue suspension through a 40 micron mesh and by centrifugation at 300 x g for 5 min at 4°C. Cell pellets were taken up in PBS supplemented with 10% FCS, singularized by thorough mixing through pipetting, and ultimately counted and assessed for overall cell viability, which is required to be at 85% or higher. For the Drop-Seq run, single cells were aliquoted in PBS supplemented with 0.04% of bovine serum albumin (BSA) at a final concentration of 100 cells/μl [198, 201].

2.2.2 Single cell RNA-sequencing using Drop-Seq

Production of microfluidic devices for Drop-Seq

Drop-Seq is a microfluidics-based method to encapsulate single cells into nanoliter-sized droplets and uses passive-flow polydimethylsiloxane (PDMS) devices [179]. Using a silicon wafer and a photomask that contains the device design, master molds for the production of microfluidic devices can be generated using photoresistant substances. A commonly used photoresist in soft photolithography is SU-8 thanks to its highly applicative properties, including the mold's longevity [199]. A huge benefit for the production of PDMS replicas is the unnecessity of working at cleanroom conditions as all fabrication steps downstream of the master mold production can be executed in a normal laboratory environment. After the master molds are finished, PDMS is mixed with a cross-linking agent in a 1:10 ratio, casted over the mold, and baked in an oven to allow for polymerization and cross-linking. After the cure and cross-linking of the PDMS, the polymer replica is peeled off the mold and both inlets and outlets for the tubing connections are punched into the replica. Next, both the replica as well as a glass slide are oxygen plasma treated to ultimately activate the surfaces

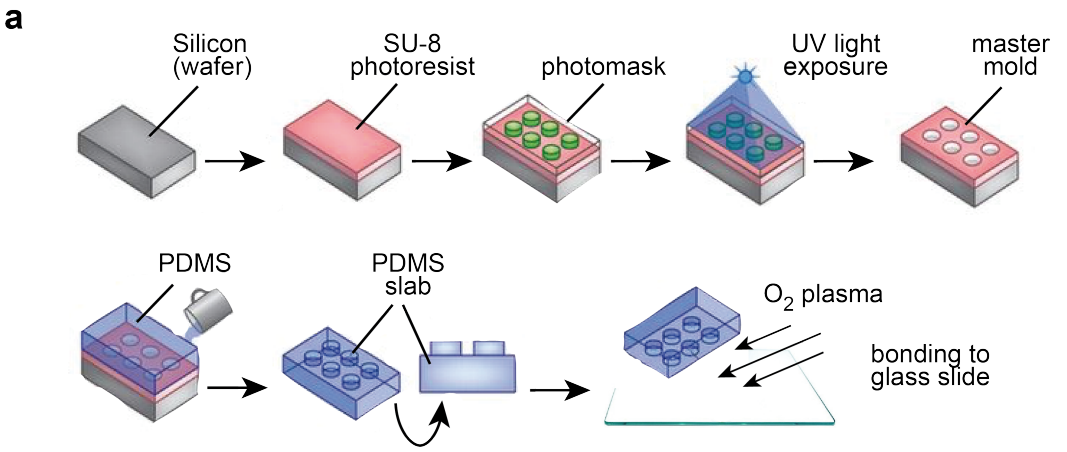

Illustration adapted from: Bhatia and Ingber, 2014, Nature Biotechnology 32:760-772.

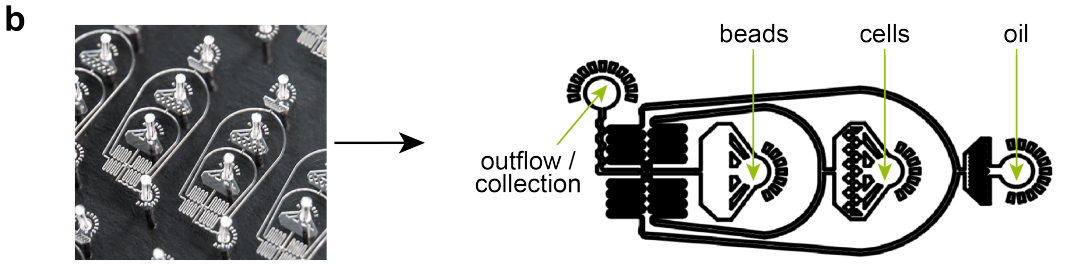

Illustration copied from: Macosko et al, 2015, Cell 161(5): 1202-1214.

Figure 6: Overview of the PDMS chip production. **a** Fabrication process of a PDMS device for single cell RNA sequencing. After the master mold production using a photoresist, PDMS slabs are generated, plasma treated, and bonded to glass slides. **b** Design of microfluidic devices as used in the Drop-Seq protocol. *Figure credits are indicated under the respective panel.*

for bonding and to generate closed channels [200]. For the Drop-Seq application, and to ameliorate the flows, all channels are treated with a hydrophobic repellent.

Here, to produce the microfluidics devices needed for the Drop-Seq platform, master molds were fabricated after the Drop-Seq device design (CAD file available as a download from: http://mccarrolllab.org/dropseq/) using standard soft lithography [198]. For the experiments here, either commercially available chips were used (Nanoshift, USA), or in collaboration with Dr. Igor Kukhtevic, in house produced chips (Helmholtz Zentrum München, Germany). Briefly, using the SU-8 photoresist (MicroChem, USA) on a 3-inch silicon wafer, 125 micron-thick uniform layers were produced [198]. The master molds were filled with a 1:10 mixture of PDMS and Sylgard 184 (PDMS Kit, Dow Corning, USA) and baked at 60°C for 4 hours to finalize the cross-linking. The resulting PDMS replica was de-molded by peeling and ports for the inlet and outlet channels (both 1 mm in diameter) were punched into the PDMS slab. Using a 2x3 inch microscopic slide, both the replica and the slide were oxygen plasma

treated for device assembly. Ultimately, all channels were treated with the repellent Aquapel (Pittsburgh Glass Works, USA) to render all channels hydrophobic [198].

Drop-Seq technology and generation of sequencing libraries

The Drop-Seq method makes use of microfluidics to encapsulate single cells into nanoliter-sized droplets [179]. With the help of syringe pumps, cells are pumped into a microfluidic device, in which they are co-flown with specialized microbeads and oil to form droplets. The microbeads are covered with oligonucleotides that include a cell-specific barcode (12 bp) and a unique molecular identifier (UMI, 8 bp) for transcript identification and counting [179]. When a single cell and a microbead are encapsulated into a droplet, the cell is instantly lysed and its cytosolic content released (Fig.7). The freely accessible mRNA transcripts can be captured by the microbeads as the oligonucleotides possess a 30 bp-long poly(dT) sequence [179]. Beads that are saturated with mRNA transcripts are referred to as 'single-cell transcriptomes attached to microparticles', or shortly STAMP. All droplets from a sample are collected and broken to release the STAMPs. After washing steps, STAMPS are reverse-transcribed by template switching, thereby generating stable cDNA strands from the mRNA transcripts, including the barcode and UMI. Subsequently, cDNA is first amplified using PCR and later tagmented, which describes the random cut of the transcripts and the addition of sequencing adapters. The result are single cell libraries, which are ready for sequencing.

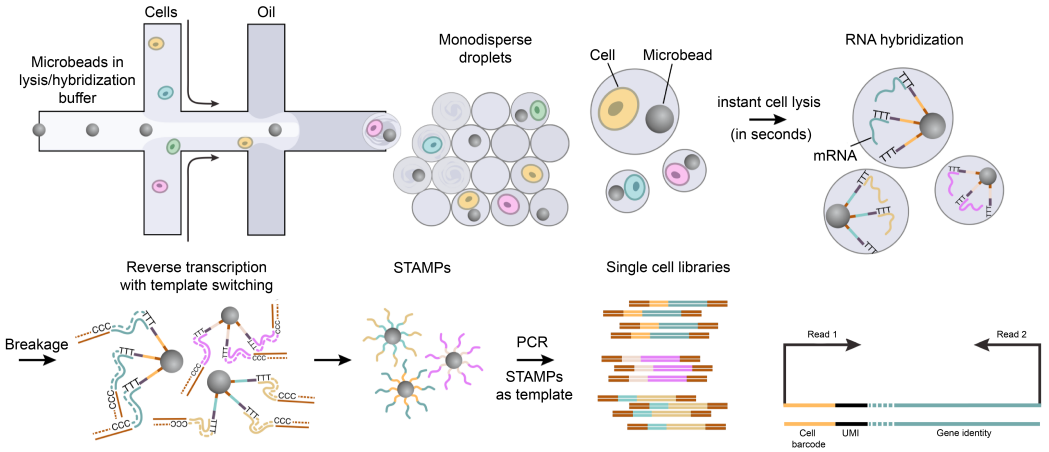

Illustration copied and adapted from: Macosko et al, 2015, Cell

Figure 7: The Drop-Seq principle. Cells are co-encapsulated with specialized microbeads to capture polyadenylated mRNA transcripts after cell lysis. Thus captured transcribed are reverse transcribed and PCR amplified for single cell library preparation. Libraries are sequenced in a paired-end run (read 1 and 2) on Illumina flow cells.*Figure credits are indicated under the respective panel.*

All Drop-Seq experiments were conducted as previously published [179,201]. Isolated single cells (100/µl) were compartmentalized with barcoded microbeads (120/µl, ChemGenes, USA) at flow rates of 4000µl/hr. After 10-20 min, monodisperse droplets were broken by harsh shaking and released STAMPS were collected by the addition of perfluorooctanol (Sigma-Aldrich, Germany). Collected transcripts were reverse-transcribed using template switch oligos (Eurofins, Germany) and Maxima RT (Thermo Fisher, Germany). In the case of unused primers on the microbeads, excessive single strands were removed by exonuclease I (New England Biolabs, USA). For cDNA amplification, 2000 beads per reaction tube were aliquoted and amplified in a 12-14 cycle-PCR using PCR primers at 100 µM concentration, and the 2x KAPA HiFi Hotstart Ready-mix (KAPA Biosystems, USA). Cycling conditions and primer sequences were used as described by *Macosko et al, 2015*. Clean-up of PCR products was performed according to the manufacturer's protocols of the clean-up beads (CleanNA, USA). Sample quality was assessed on the BioAnalyzer using DNA High Sensitivity chips (both Agilent Technologies, USA). Tagmentation and adapter ligation, complimentary to immobilized primers on the sequencing flow cell, were performed using the Nextera XT DNA Sample Prep Kit (Illumina, USA) and a customized P5-primer (Integrated DNA Technologies, Germany) (according to the Drop-Seq laboratory protocol, version 2.1; http://mccarrolllab.com/dropseq/).

RNA-sequencing on Illumina platforms

All libraries that were produced for this work were sequenced on the HiSeq 4000 machine from Illumina, USA. Illumina is the current global market leader and skims the biggest part of the next-generation sequencing market [202]. The HiSeq 4000 sequencer is fast and produces higher output as compared to other sequencers thanks to a patterned flow cell technology and kinetic exclusion amplification [203]. Patterned flow cells have got nanowells at fixed locations on the sequencing lane to increase uniform cluster spacing, higher cluster densities, and improved unambiguous cluster identification [203]. During cluster generation, the exclusion amplification only allows for one single DNA template to bind and form a monoclonal cluster within a single nanowell of the flow cell (Fig.8c). The principle is that after template binding to the nanowell-anchored seeding primers, instantaneous and rapid amplification takes place, thus preventing other templates from binding in that particular nanowell and forming a polyclonal cluster [203, 204]. Once, a template is hybridized to a nanowell, it is immediately extended and denatured (Fig.8a), followed by bridge formation and bridge amplification [205]. In a final step, reverse strands are cleaved from the resulting monoclonal clusters, which are now ready for next-generation sequencing-by-synthesis (Fig.8b). Sequencing-by-synthesis makes use of cyclic reversible termination [206] and describes the addition of four fluorescently-labeled DNA bases/nucleotides at the same time. In more detail, with each sequencing cycle, all fluorescently-tagged nucleotides compete

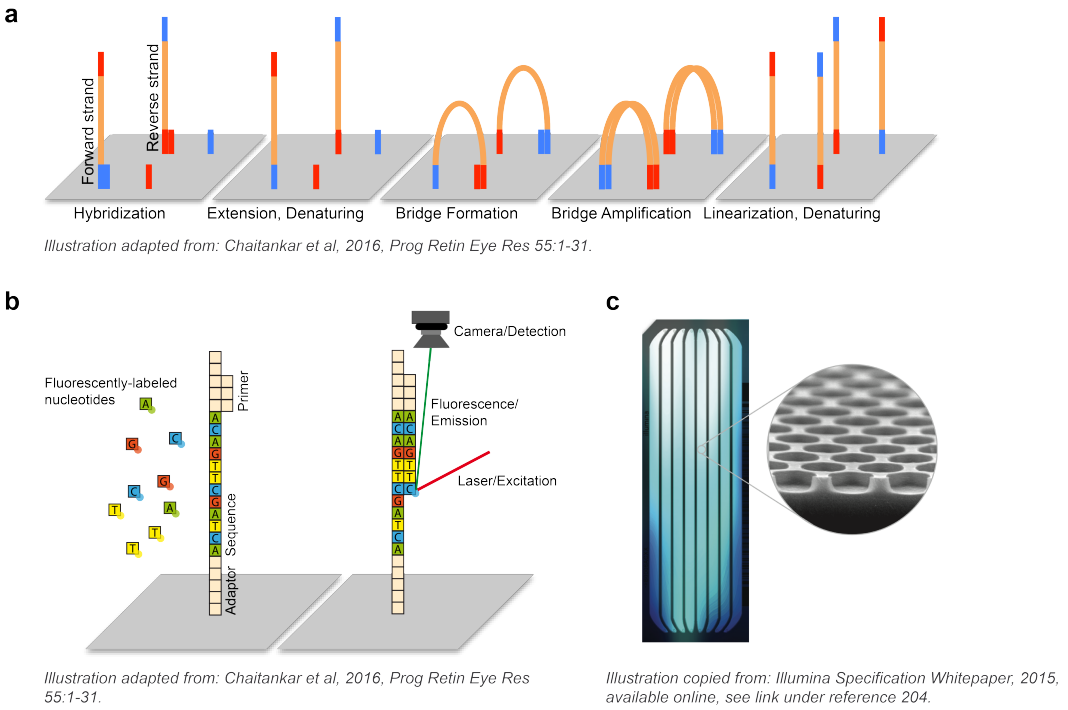

Illustration adapted from: Chaitankar et al, 2016, Prog Retin Eye Res 55:1-31.

Illustration adapted from: Chaitankar et al, 2016, Prog Retin Eye Res 55:1-31.

Illustration copied from: Illumina Specification Whitepaper, 2015, available online, see link under reference 204.

Figure 8: RNA sequencing on Illumina flow cells. **a** Schematic overview of the cluster generation process before sequencing. Sample or template strands are hybridized to a nanowell on the flow cell. After extension and denaturation, strands form bridges for subsequent bridge amplification resulting in monoclonal cluster formation. **b** Principle of the sequencing-by-synthesis reaction. In each sequencing cycle, fluorescently-labeled nucleotides compete for the incorporation to a newly growing strand, which after laser excitation are detected by a camera. **c** Illumina HiSeq 4000-compatible patterned flow cell with eight sequencing lanes containing billions of nano wells. *Figure credits are indicated under the respective panel.*

for the addition but only one labeled nucleotide is subjoined to the growing strand, which terminates the polymerization. After the addition of each nucleotide, i.e. each sequencing cycle, the sequencer detects the fluorescent dye by light excitation and thus identifies the newly incorporated nucleotide (Fig.8b). Hence, both emission wavelengths, as well as signal intensity determine the base call [203, 204]. After detection, the fluorescent label is enzymatically cleaved, giving way for the next nucleotide to be incorporated. A clear benefit is granted by the natural competition of all four bases, which reduces nucleotide incorporation bias. Additionally, as base calls are derived directly and immediately after each sequencing cycle, less detection errors are observed [207, 208]. Finally, when the last sequencing cycle has been reached, base calling from the gathered fluorescence data is performed [209].

For the experiments presented in the thesis, samples were pooled across the sequencing lane using multiplex indices. As the i5 index is customized in Drop-Seq, samples were only indexed using the i7 indices (Illumina Nextera XT DNA Kit, v2). Final libraries were first purified using twice 0.6x clean-up beads (CleanNA) and then sequenced in a 100

bp paired-end run on the HiSeq 4000 sequencer (Illumina, USA). Denatured sequencing sample was loaded at 0.2 nM concentration and mixed with 5% of PhiX spike-in to increase sample complexity. The priming of Read1 was achieved with 0.5 μM Read1CustSeqB primer (see online Drop-Seq protocols) because of the Drop-Seq-specific sequence at the 3-prime end.

2.3 Bioinformatic single cell data analysis

2.3.1 Processing of single cell RNA-sequencing reads

After sequencing, Drop-Seq pools were demultiplexed with the standard bcl2fastq2 Conversion Software v2.20 (Illumina standard pipeline for the HiSeq 4000 sequencer). During each sequencing cycle, the real-time analysis (RTA) software produces a base call (BCL) file, which is converted into a FASTQ file - the input file for downstream analysis of the data [210]. Based on the index sequence, the software is able to demultiplex the pooled sample back into single samples. All samples underwent adapter trimming and were mapped to the mouse genome (mm10) as previously described and published [179]. Preprocessing analysis was performed using the published Drop-Seq pipeline, version 2.0 [179]. The mouse genome reference is available as a download (GEO accession code: GSE63269). In brief, alignments of reads were determined by STAR, version 2.5.2a [211] using default parameters. For barcode filtering, barcodes with less than 200 detected genes were excluded. As a high proportion (> 10%) of transcript counts derived from mitochondrial-encoded genes, which may indicate low cell quality, such cells were removed from the downstream analysis. Similar, cells with an incomparably high number of UMI counts (> 5000) were excluded as these may represent cell doublets [198]. The Drop-Seq pipeline eventually produces a digital gene expression (DGE) matrix, which was further processed using the R-based *Seurat* package, version 2.3. The *Seurat* package was created to specifically analyze single-cell RNA-sequencing data [212]. Using default settings, the DGE matrices were normalized with the *Seurat* function NormalizeData() [198]. Single cell gene expression data was normalized by total expression of all cells, then multiplied with a scaling factor (here: 10,000) and log-transformed. To mitigate the effects of unwanted sources of cell-to-cell variation, the number of UMI counts was regressed from the data using the *Seurat* function ScaleData() [198].

2.3.2 Bioinformatic tools for downstream data analysis

R packages

The *R* distribution is a free-to-use language and environment for statistical computation and graphical display of computed data (https://www.r-project.org/about.html). Although a plethora of statistical functions and techniques are implemented, R can be readily extended by multiple packages that are available online on various websites, including those from CRAN and Bioconductor, for instance. Single cell-specific R packages used in this thesis are explained in more detail in the following paragraphs.

Seurat. This R toolkit provides many options and functions to analyze and integrate single cell data [212–214]. The package offers algorithms and functions for quality control, data filtration, calculation of highly-variable genes, graph-based clustering, dimensional reduction, and the identification of cluster-specific markers using specific vignettes (source: https://satijalab.org/seurat/vignettes.html). This allows for the identification and interpretation of cell heterogeneity from single cell RNA-seq experiments. Later developments included vignettes for differential gene expression testing, cell cycle regression, and improved visualization methods. This R package is continuously improved with frequently updated versions, available at https://satijalab.org/seurat/. In this work, *Seurat*, version 2.3.0) was used.

Limma. Originally, this package was created for gene expression analysis for data derived from microarray studies or RNA-seq studies [215]. With the surge of novel single cell-based RNA-seq methods, however, *limma* is a valuable tool for data analysis, linear modeling and differential gene expression analysis [215, 216]. Additionally, *limma* provides tools to compare between complicated or multifactor-designed experiments, thus being ideal for the analysis of complex experiments with diverging experimental conditions [215, 216].

SoupX. This R package has been designed to quantify contaminant RNA transcripts from droplet-based single cell RNA-seq data and consequently remove them from the data set [217]. Briefly, as is assumed that all captured mRNA transcripts in collected droplets are endogenous to cells, *SoupX* searches for contaminant background RNA that is likely to be derived from ambient sources. Due to varying experimental conditions and the ubiquitous presence of ambient RNAs, considered as *the soup*, the goal is to remove the background contamination and to reduce batch effects to allow for ameliorated quality control and data interpretation [217]. In this work, *SoupX*, version 3.0.0 was used.

Splines. Analysis of non-linear relationships require modeling of the data using non-linear regression [218]. Although polynomial regression is a common way to address this issue,

spline regression is the favorable method to use [218]. Splines fit smooth curves between fixed polynomial segments, called knots, by computing polynomial regression between the knots [218]. Generalized additive models (GAM, cf. R package *gam*) are applied when splines are modeled with automated knot selection [218]. The R package *splines* was used in version 3.6.1.

Dbscan. This R package contains algorithms for density-based clustering [219]. *Dbscan* is the abbreviation for density-based spatial clustering and application with noise [219]. It is applied for large datasets with outliers or noise to detect clusters independently of shape or cluster size [219]. I.e., compared to other partitioning methods, such as *k* means, that *dbscan* can find any cluster shape and is not restricted to spherical-shaped or convex cluster shape [219] - a useful tool when analyzing oddly shaped clusters derived from single cell RNA-seq data. Here, version 1.1.3 was used.

Slingshot. This package was crafted to predict cell fate trajectories and pseudotimes from single cell RNA-seq data after dimensionality reduction and data clustering [220]. A huge benefit of *Slingshot* is the addition of prior knowledge through supervised graph construction [220]. This renders *Slingshot* a very flexible but easy-to-handle toolkit in order to assess lineage branching events in low-dimensional data [220]. In this work, the version 1.2.0 was used.

Scanpy packages

Scanpy is a scalable Python-based single cell analysis toolkit that has been developed to overcome analysis limitations as they have been observed using R-based frameworks (https://github.com/theislab/Scanpy) [221]. *Scanpy* provides functions for the preprocessing, clustering, visualization, pseudotime and trajectory inference, simulation of gene regulatory networks, or differential expression testing [221]. As single cell data sets become increasingly larger and analysis demands faster and better analysis options, *Scanpy* represents an easy implementation with advanced machine-learning packages that can easily be run on interactive web-based environments, such as the Juypter notebooks (https://jupyter.org).

Partition-based graph abstraction. This algorithm, which has been implemented into *Scanpy*, is a method that provides a graph-like map of the data manifold, which is easy to interpret and that covers different approaches, such as development over time or cell grouping by properties [222]. The resulting *partition-based graph abstraction (PAGA)* maps are created by estimating the connectivity of manifold partitions (https://github.com/theislab/paga) [222]. The clear benefit is accounted for by the higher computational efficiency of the workflow and the preservation of the global topology of the data [222]. Whilst similar properties between

cells are reflected by clustering, or developmental pathways by trajectory inference, *PAGA* combines both methods in a single analysis, which allows to assess cellular transitions between cell types and states at unprecedented resolution [222].

scVelo. This package is the implemented version of the *velocyto* package, which is available for both R and Python. *Velocyto* or *scVelo* is a tool for analyzing gene expression dynamics in single cell RNA-seq data [223]. By using the information from unspliced and spliced transcript reads, RNA velocities are estimated [223]. As the collection of cells using Drop-Seq, for instance, capture only cells at the time of isolation, i.e. they represent a snapshot in time, the time-dependent relationship of unspliced and splices transcripts needs to be modeled [223]. RNA velocity is represented by the first time derivative of the spliced mRNA abundance between spliced and unspliced transcripts and mRNA degradation [223]. RNA velocities are plotted as high-dimensional vectors that reflect the future state of a single cell based on the balance of spliced and unspliced mRNAs [223].

Diffusion pseudotime. This algorithm has been created to infer different developmental processes from complex cellular data displaying asynchronous states; similar to the static snapshot problem described for the *scVelo* package [224]. Pseudotemporal ordering of cells by similarity of gene expression reconstitutes the sequence of cellular dynamics, however, can not be applied to data with branching lineages [224]. To overcome this challenge, the diffusion pseudotime algorithm makes use of random-walk-based distances, computed via Euclidean distances, to estimate the temporal order of cells during differentiation processes [224]. Hence, diffusion pseudotime enables to study cellular transitions and to reconstruct branching and differentiation events [224].

Batch-balanced k *nearest neighbor*. The BBKNN is a fast and easy-to-use tool for batch effect removal and is implemented in *Scanpy* (https://github.com/Teichlab/bbknn) [225]. Frequently, during data integration, batch effects become prominent, be it due to technical or biological reason. BBKNN is a graph-based function that creates a neighbor graph which in turn is used for clustering, pseudotime, and the UMAP visualization [225]. The data is first split into smaller batches and then submitted to the *k* nearest neighbor determination for each individual cell within each small batch group [225]. This way, in contrast to taking the entire data set for *k* nearest neighbor calculation, cell connectivities are only created between analogous cells from different batches, and thereby leaving the data structure, like the PCA space, intact and cell counts unchanged [225].

2.3.3 Data analysis of whole lung tissue

For the whole lung single cell survey spanning six time points, highly variable genes were calculated per sample, selecting the top 7,000 genes with a mean expression between 0.01 and 8 [198]. After excluding homologs of known cell-cycle marker genes [226], a total of 18,893 genes were subjected to independent component analysis [198]. The first 50 independent components were used as input to the FindClusters() function with the resolution parameter set to two and the RunUMAP() function with the n_neighbors parameter set to 10 [198]. For dimensionality-reduced data visualization the Uniform Manifold Approximation and Projection (UMAP) method was used [227].

Multi-omic data integration. To confirm global expression changes observed at the single cell level, previously published bulk RNA-seq data and proteomics data were integrated, both obtained from murine whole lungs examined 14 days after bleomycin-induced injury and controls [139, 198]. Multi-omic data integration was performed as previously described [139, 198, 201]. Briefly, by summing all counts within an individual mouse sample, *in silico* bulk samples were created. Using the voom () function of the *limma* R package, data from whole lung tissue bulk and *in silico* bulk was normalized [198, 228]. After normalization, both data sets were merged with the proteomics data based on gene sets detected in all the data sets and subsequently quantile normalized. The resulting data matrix was used as input for principal component analysis (PCA) [198].

Discovery of cell type identity marker genes. To identify cluster-specific marker genes, the Seurat FindAllMarkers() function was applied, restricted to genes detected in more than 10% of cells and with an average fold change difference of 0.25 or more [198]. Based on these derived marker genes and manual curation, all clusters were assigned to cell type and meta-cell type identities [198]. Cell type frequencies were calculated by dividing the number of cells annotated to a specific cell type identity by the total number of cells for each mouse sample [198]. As background mRNA contamination, so-called ambient RNA, is frequently observed in droplet-based single cell RNAseq data, the function inferNonExpressedGenes() from SoupX [198, 217] was applied to identify a set of 80 ambient RNAs. Ambient RNA contamination is believed to derive from dying cells releasing their cellular content upon cell lysis [198]. During the Drop-Seq protocol, such contamination can be distributed to many droplets and beads, leading to a blurred expression signal [198].

Time course differential expression analysis. For the identification of differential gene expression patterns across time within a given cell type the R packages *splines* and *lmtest* were used [198]. First, Louvain clusters were manually combined into 26 cell types to generate a more coarse-grained cell type annotation for the time course differential expression

analysis [198]. Within each of these groups, gene expression was modeled as a binomial response where the likelihood of detection of each gene within each mouse sample was the dependent variable [198]. Therefore, the sample size of the model conformed the number of mouse samples (n = 28) and not the number of cells [198]. To assess significance, a likelihood-ratio test was performed between the following two models: for the first model, the independent variables contained an offset for the log-transformed average total UMI count and a natural splines fit of the time course variable with two degrees of freedom; the independent variables of the second model contained solely the offset for the log-transformed average total UMI count [198]. The dependent variable of both models was the number of cells with UMI count > 0 out of all cells for a given cell type and mouse sample [198]. To account for potential false positive signal derived from ambient RNA levels, cell type marker genes were calculated for the 26 cell type annotations using the *Seurat* FindAllMarkers() function [198]. Consequently, for all 80 candidate ambient RNAs, all regression p-values were set to 1 in cell types where the gene was not simultaneously a marker gene with an adjusted p-value of less than 0.1 and a positive average log fold change [198].

Cell-cell communication analysis. For the analysis of cell-cell communication networks, a list of annotated receptor-ligand pairs was downloaded [198, 229]. The receptor-ligand information was integrated with the whole lung data's cell type marker genes, followed by the generation of cell-cell communication networks. This is facilitated by the creation of edges between two cell types if both share a receptor-ligand pair as marker genes [198].

Macrophage analysis. It is not entirely understood whether monocyte-derived macrophages contribute to the development of lung fibrosis [198]. In order to see if the data reflects published models of monocyte recruitment, bulk RNA-seq data from FACS-sorted macrophage populations after bleomycin-induced lung fibrosis was integrated [198, 230, 231]. This data set contained bulk RNA-seq gene expression of tissue-resident alveolar macrophages (TR-AMs), monocyte-derived alveolar macrophages (Mo-AMs), interstitial macrophages (IM), and monocytes (Mono) for both day 14 and day 19 after bleomycin injury, including additional measurements for TR-AMs at day 0 [198]. To derive a gene expression signature from the bulk RNA-seq data, the R package *limma* was applied [198, 215]. Using the default settings of the *limma* workflow [198, 230], differentially expressed genes between these four populations were detected. Next, the whole lung single cell RNA-seq data was subsetted to clusters which only expressed known macrophage markers, followed by a selection of a new set of variable genes [198]. Thereafter, PCA and UMAPs were recreated for this subset, using 20 PCs and 20 n_neighbors in *Seurat's* functions [198]. The macrophages from the whole lung data were scored according to their similarity to these bulk-derived signatures using Pearson correlation [198]. For each of the four bulk-derived groups, the log

fold changes of the 500 most differentially expressed genes were correlated with the scaled expression values of each macrophage cell in the present single cell RNA-seq data [198]. For the separation of potential monocyte-derived macrophages from interstitial macrophages, each cell was assigned to the category with the higher correlation coefficient, provided that the difference was greater than 0.05 [198].

2.3.4 Data analysis of lung epithelial cells

In contrast to the whole lung-derived data, the MACS-sorted epithelial cell data was applied to SoupX for ambient RNA removal using the pCut parameter 0.3 within each sample before merging the count matrices together [198]. The merged expression table was then pre-processed as described in 2.3.3 with the alteration that the upper threshold for the number of total UMI counts per cell was set to 3,000 [198].

For computational analysis, both the frequently used *Seurat* packages [212], as well as the *Scanpy* code [221] was applied. Undesirable sources of variation, including mitochondrial reads, cell-cycle effects, and the total number of UMI counts, were regressed from the data by applying the *Seurat* functions CellCycleScoring() and ScaleData() [198]. 17,038 genes showed variable expression in at least two samples and were used for the principal component analysis [198]. The majority of the cells were airway and alveolar epithelial cells, although non-epithelial cells were also captured [198]. Further data filtering was achieved by clustering so that clusters expressing non-epithelial markers could be excluded from the data set [198]. The resulting object was subsequently converted into a -.h5ad file for downstream analysis using the python package *Scanpy* [198]. The aligned -.bam files were used as input for *Velocyto* [223] to derive the counts of unspliced and spliced reads in the loom format [198]. Next, the sample-wise loom files were combined, normalized and log transformed using *scVelo's* (https://github.com/theislab/scvelo) functions normalize_per_cell() and log1p() [198]. After merging the loom information to the exported -.h5ad file using *scVelo's* merge() function, the object was scaled and the neighborhood graph constructed [198]. With the batch-balanced *k* nearest neighbor (BBKNN) [225] tool, the different PCR cycles used in the experiment were accounted for by setting the neighbors_within_batch=15 and n_pcs=40 [198]. Two dimensional visualization and clustering was carried out with the *Scanpy* functions tl.louvain() at resolution=2 and tl.umap() [198]. Neuroendocrine cells (PNEC) formed a distinct cluster in the UMAP, however, they were only assigned to a single cluster at higher resolutions. In order to separate them from basal cells, PNEC were clustered with the dbscan tool [219] using the UMAP coordinates (assigned as neuroendocrine cluster, cf. Fig.21b) [198]. After manual curation of the markers, the remaining 20 cluster were combined, leading to 13 final meta cell types [198].

Cell-cycle analysis. The proliferating cell cluster (cf. Fig.21b) of the high-resolution epithelial cell data set was subjected to cell type deconvolution analysis [198]. Cell cycle phases (S.Score, G2M.Score) were regressed using the *Seurat* ScaleData() function [198]. Next, PCA was performed using all unique marker genes and the *Seurat* RunPCA() function [198]. UMAP-embedding and Louvain clusters were calculated using the first 20 principal components with the *Seurat* RunUMAP() and FindClusters() functions, respectively [198]. Upon manual curation of the marker genes for the generated embedding, four distinct clusters were identified [198]. The frequency of cells undergoing proliferation was calculated by dividing the number of cells in cluster 14 by the number of total cells for each mouse sample [198].

PAGA analysis. In order to assess how Louvain clusters are connected in a global space, partition-based graph abstraction (PAGA) was applied [198, 222]. The *Scanpy*-integrated tl.paga() function was used to calculate the connectivities and the Louvain clusters as partitions [198]. Weighted edges are described by a statistical test to measure the connectivity between the single partitions [198]. Connections with a weight less than 0.3 were removed [198].

Velocity analyses. The future cellular states of individual cells can be inferred by making use of the spliced and unspliced information of the cells [198]. Here, scVelo was applied using the previously normalized and log transformed data as a starting point to calculate first and second order moments for each cell across its nearest neighbors (scvelo.pp.moments(n_pcs = 40, n_neighbors = 15)) [198]. The velocities were estimated and the velocity graph constructed using the scvelo.tl.velocity() function with the mode set to stochastic and the scvelo.tl.velocity_graph() function, respectively [198]. Velocities were visualized on top of the previously calculated UMAP coordinates with the scvelo.tl.velocity_embedding() function [198]. Terminal state likelihood of a subset of cells was assessed with the scvelo.tl.terminal_states() function at default parameters [198].

Trajectory differential expression analysis. To identify genes showing significantly altered expression across the differentiation trajectory towards the Krt8+ cell state, the following approach was used: The high-resolution data set of the epithelial cells was restricted to cells from Louvain clusters 2 (Krt8+ cells), 10 (MHC-II+ cells) and 11 (activated AT2 cells), cf. Fig.21b and 22b [198]. The dbscan() function [219] was used to identify outlier cells, which were subsequently removed from further analysis [198]. The R package *slingshot* [220] was used to infer the pseudotemporal ordering across the trajectory of the first two diffusion components of all remaining cells [198]. Next, the analysis was restricted to genes with more than 10 total UMI counts in more than 5 mouse samples [198]. For each gene, the following generalized additive model was fitted using the R package *gam* [198, 232, 233].

Gene expression was used as the explanatory variable and defined as a binary outcome, representing the detection of the gene (UMI count > 0) [198]. Log transformed total UMI counts were included as a covariate in the model to account for differences in library size [198]. A smooth *Loess* fit of the pseudotemporal coordinate was included as the second independent variable [198]. Using the *Loess* non-parametric local regression approach, smooth lines can be fitted via the pseudotemporal ordering [198]. As the interpretation of p-values across pseudotime is difficult, all genes with a marginal p-value of less than 0.00001 were defined to be significant [198]. Gene expression patterns along the pseudotemporal trajectories were visualized using local polynomial regression fitting as implemented in the R-based Loess() function at default parameters [198].

2.3.5 Cellular pathway analysis

In order to forecast pathway activities and cellular functions with respect to the observed gene expression changes as presented in the findings of this work, the Ingenuity Pathway Analysis platform (IPA, Qiagen, Redwood City, www.qiagen.com/ingenuity) was used as previously described [139, 198]. IPA analysis deploys a retinue of algorithms and other tools to infer and score up-stream regulatory networks by analyzing gene expression data from large-scale causal networks, derived from Ingenuity Knowledge Base [198].

2.4 Whole lung RNA transcriptome analysis

RNA was isolated from whole lung tissue that was immediately processed after the mouse was sacrificed. Shortly, mice were killed and the lungs excised. The lung lobes were minced into smaller pieces using scalpel blades and mixed on a glass slide to avoid bias in the selection of a single lobe. A part of these minced lung pieces were transferred to reaction tubes and covered in RNA stabilization buffer, following the instructions of the RNeasy Kit (Qiagen, Germany). Cells were lysed and homogenized and subsequently frozen for later RNA isolation using QiaShredders (Qiagen, Germany) and the RNeasy Kit (Qiagen, Germany). RNA integrity and quality was assessed using the RNA Nano Kit and the BioAnalyzer (both from Agilent, Germany). Samples were eventually given to the sequencing facility at the Helmholtz Zentrum München, where samples were prepared for Sequencing using the TruSeq Kit (Illumina, Germany). Analysis of bulk RNA sequencing data was analyzed with the freely available R package DESeq2 [366].

2.5 Magnetic-activated cell sorting

First described in 1990, the MACS technology provides an easy and fast way to isolate cells from complex tissues [303, 304]. Magnetic activated cell sorting (MACS) reliably enriches cell populations based on cell coupling to microbeads via specific antibody binding. During the procedure, antibody-coated microbeads are incubated with a cell suspension and later magnetically retained by column-based purification. Once eluted, the enriched cell population can be used for further analysis. Principally, both positive and negative cell selection is possible: positive selection means the direct separation of the cells of interest by bead-binding, whereas the negative selection describes the 'flow-through' pool of cells, of which undesired cell populations were removed by prior positive selection.
For the presented work, cells from single cell suspensions were strained using a 40 μm mesh size and red blood cells (RBC) were eliminated by lysis (RBC lysis buffer, ThermoFisher) [198]. For positive epithelial cell selection, cells were stained with CD326-AlexaFluor647 antibody (Biolegend, 118212) for 30 min at 4°C in the dark, and after washing, incubated with microbeads specific against AlexaFluor647 (Miltenyi Biotec, 130-091-395) for 15 min at 4°C [198]. MACS LS columns (Miltenyi Biotec, 130-042-401) were prepared as outlined by the manual [198]. Cells were applied to the columns, and positively-labeled epithelial cells were retained in the column [198]. The flow-through was collected separately for later mesenchymal cell enrichment (negative MACS selection) and kept on ice [198]. Epithelial cells were eluted from the LS columns and prepared for Drop-Seq [198]. Mesenchymal cells from the flow-through were further enriched by negative depletion of CD31+ (Invitrogen, 17-0311-82), CD45+ (Biolegend, 103112), Lyve1+ (Invitrogen, 50-0443-82), Ter119+ (Biolegend, 116218), and CD326+ cells (Biolegend, 118212) [198]. After antibody staining, 100 μl per 10 million cells of MACS dead cell removal beads (Miltenyi Biotec, 130-090-101) were added and incubated according to the product's accompanying protocols [198]. Depletion of undesired cell types was achieved by the use of micro-beads specific for APC (Miltenyi Biotec, 130-090-855), which ensured magnetic retention of these cells [198]. Likewise to epithelial cells, negatively-selected mesenchymal cells were applied to the Drop-Seq workflow [198].

2.6 Fluorescent-activated cell sorting and flow cytometry

Single cell suspensions derived from whole lung tissue was prepared for analysis using flow cytometry. After depletion of red blood cells by red blood cell lysis buffer (Invitrogen, ThermoFisher), cell suspensions were stained with anti-mouse CD45-PE-Vio770 (Miltenyi Biotec, 130-110-661), CD326-BV421 (Biolegend, 118225), Krt8/TROMA-I (DSHB-Developmental Studies Hybridoma Bank at the University of Iowa), and av-beta-6-specific monoclonal

antibody 6.3G9 (Itgb6-3G9; kindly provided by Prof. Dr. Dean Sheppard, available through Biogen Idec, USA). Cells were stained for surface markers in the dark at 4°C for 20 min, followed by cell fixation and permeabilization (Fix Perm, Life Technologies, GAS004) for intracellular staining of Krt8. From the CD326 (EpCam)-positive population, epithelial cells were selected by the exclusion of the CD45+/CD326+ fraction. Within the epithelial cell gate, Krt8+, Itgb6+, or Krt8+/Itgb6+ cells were identified and quantified by their geometric mean fluorescence signal intensity. For exclusion of non-specific antibody binding and autofluorescence signal, fluorescence-minus-one (FMO) controls were included in the measurement. Antibody staining per 1,000,000 cells was performed in the following dilutions: CD326 (1:500), CD45 (1:20), Krt8 (1:35), Itgb6 (1:1000). Data was acquired using the BD LSRII flow cytometer (Becton Dickinson, Germany) and analyzed using the FlowJo software (TreeStart Inc., Ashland, OR, USA) [198]. All gatings were defined by thresholding the mean fluorescence intensity signals based on isotype-labeled and unstained controls [198].

2.7 Precision-cut lung slices

Precision cut lung slices (PCLS) were obtained by following previously published instructions [306]. Shortly, directly after euthanasia, mouse lungs were inflated by intratracheal instillation of warm 2% (w/v) low-gelling temperature melting point agarose (Sigma Aldrich, A9414) dissolved in sterile DMEM/Ham's F12 cultivation medium (Gibco, 12634010), supplemented with 100 U/ml penicillin, 100 μg/ml streptomycin, and 2.5 μg/ml amphotericin B (Sigma Aldrich, A2942) [198]. Upon full inflation, lungs were harvested and immediately chilled in ice-cold cultivation media to enhance the gelling of the agarose [198]. The actual cutting of single 300 μm-thick slices was facilitated by a vibratome (Hyrax V55; Zeiss, Jena, Germany) [198]. Final PCLS were methanol-fixed for 20 min and prepared for microscopy as previously described [198, 307]. Briefly, primary antibodies diluted in 1% bovine serum albumin (BSA, Sigma Aldrich, 84503) in PBS (1:100) were incubated for 16 hours at 4°C and 3x washed with PBS for 5 min [198]. Secondary antibodies were likewise diluted at a concentration of 1:200, incubated for 4 hours at room temperature, and finally 3x washed with PBS for 5 min [198]. Primary antibodies: rat anti-Krt8/TROMA-I (1:200; DSHB-Developmental Studies Hybridoma Bank at the University of Iowa), rabbit anti-pro-SPC (1:200; Millipore, AB3786), goat anti-Pdpn (1:200; RD Systems, AF3244). Nuclear counterstaining was achieved with DAPI (4',6-diamidino-2-phenylindole, Sigma-Aldrich, 1:2,000) [198]. Confocal high-resolution 3D imaging of the PCLS was accomplished by placing the PCLS into a glass-bottomed 35 mm CellView cell culture dish (Greiner BioOne, 627870) as a wet chamber. Acquisition of PCLS z-stack images succeeded with an inverted microscope stand of an LSM 710 (Zeiss) confocal module [198]. The microscope was used in multitrack mode with these listed objectives: Plan-Apochromat W 40x/1.0 M27 and Plan-Apochromat W 63x/1.3 M27 [198].

The microscopy software was ZEN2009 (Zeiss). The collected confocal z-stack images were surface rendered using the Imaris 9.3 software (Bitplane) [198]. For 3D cell shape analysis, the Imaris statistical analysis tool (Measurement Pro) was used [198]. Basis for the analysis was the morphometric parameter sphericity, in which a value of 1 corresponds to a perfect sphere [198].

2.8 Human lung material

Human lung tissue and human lung explant material were received from the CPC-M bioArchive at the Comprehensive Pneumology Center (CPC), Munich [198]. Lung tissue from patients with ILD diagnosis (n=5) was derived from lung explant material, thereby reflecting non-resolving end-stage disease [198]. ARDS-lung tissue sections (n=2) were provided by the Institute of Pathology at the University Hospital of Ludwig Maximilians University, Munich [198]. Healthy control tissue (n=7) originated from tumor resections from the Asklepios Lung Specialty Clinics in Gauting, near Munich, Germany [198]. All participants gave written informed consent [198]. This study and the handling of human tissue, which required special training, were performed in agreement with the locally authorized ethics committee of the Ludwig Maximilians University, Munich, Germany (vote number: 333-10) [198].

2.9 Immunofluorescence microscopy

After euthanasia, mouse lungs were immediately inflated with 4% paraformaldehyde. For formalin-fixed, paraffin-embedded (FFPE) lung tissue, sections were cut at 3.5 μm, followed by deparaffinization, rehydration, and antigen retrieval by pressure-cooking (30 sec at 125°C and 10 sec at 90°C) in citrate buffer (10 mM, pH 6.0) [198]. Sections were blocked for 1 hour at room temperature with 5% bovine serum albumin, and incubated in primary antibodies overnight at 4°C, followed by secondary antibody (1:250) incubation for 2 hours at room temperature [198]. These primary (1) and secondary (2) antibodies were used: (1) rat anti-Krt8 (170 μg/ml; University of Iowa Hybridoma Bank, 1:200), rabbit anti-pro-SPC (1:200; Millipore, AB3786), goat anti-Pdpn (1:200; RD Systems, AF3244), rabbit anti-SPC (1:150; Sigma-Aldrich, HPA010928), mouse anti-alphaSMA (1:1,000, Sigma-Aldrich, A5228), rabbit anti-Areg (1:50; LSBIO, LS-B13911), rabbit anti-Hbegf (1:200; Bioss Antibodies, bs-3576R), rabbit anti-Ki67 (1:200; Abcam, ab16667), mouse anti-CC10 (1:200; Santa Cruz, sc-365992), rabbit anti-Cst3 (1:100; Abcam, ab109508); (2) donkey anti-rabbit AlexaFluor568 (Invitrogen, A10042), donkey anti-rat AlexaFluor488 (Invitrogen, A21208), donkey anti-goat

AlexaFluor647 (Invitrogen, A21447), goat anti-mouse AlexaFluor647 (Invitrogen, A21236) [198]. All images were taken using the LSM 710 microscope (Zeiss) [198].

2.10 Microscopic image quantification

The fluorescence intensity of Krt8 expression in selected regions of immunofluorescence microscopy images was measured (excluding airway signal) using FIJI (ImageJ) [305]. For quantification of Krt8 expression in human FFPE sections, and likewise, for the Hbegf and Areg quantification in the mouse sections, the mean overall fluorescence intensities were measured. For quantification of cell proliferation, cells were stained with Ki67 and Krt8 and counted manually for Ki67 positive cells.

2.11 Mouse lung function

After anesthesia, mice were tracheotomized and a cannula inserted into the trachea. In more detail, the skin at the neck was opened by a marginal incision, followed by the separation of the submaxillary gland. Once the trachea was exposed, it was cut open in close location to the larynx, ensued by the insertion of a cannula and its fixation in the trachea by the aid of a suture. Thereafter, mice were immediately connected to a computer-controlled piston ventilator (flexiVent, SCIREQ Inc, Montreal, Canada), which executed a predefined forced oscillation and deep inflation program [198].

2.12 Histology and collagen quantification

For histologic assessment of the degree of interstitial fibrosis and collagen content, mouse lung tissue was fixed in 4% paraformaldehyde prior to embedding in paraffin. Sections with 3.5 micron thickness were mounted on microscopy slides, deparaffinized and stained using a commercially available Masson's Trichrome staining kit (Sigma Aldrich, Germany). Briefly, following the manufacturer's instructions, sections were washed in deionized water and incubated in preheated Bouin's solution for 15 min at 56°C. Slides were cooled in deionized water and counterstained with Weigert's iron hematoxylin solution and Biebrich scarlet-acid Fuchsin. After treatment with phosphotungstic/hosphomolybdic acid, slides were put into Anilin blue solution and 1% acetic acid. Finally, slides were dehydrated and conserved using Entellan mounting medium (Merck Millipore, Germany). Collagen content was quantified by segmentation image analysis (cf. Fig.38b) to define large vessel space (excluded for the analysis) and differentiate between cellular staining and whitespace ares (mainly the alveolar space). Additionally, Anilin-blue positive staining was quantified using the calculation as

shown in Fig.38b. Likewise, the formula for the quantification of tissue density is noted in Fig.38b, describing the degree of how condensed the tissue appeared in response to bleomycin administration.

2.13 Flow cytometry for BAL and tissue analysis

Bronchoalveolar lavage. During the procedure of lung function measurement, lungs of sedated mice were filled with 2 ml of sterile PBS, supplemented with 0.1 mM EDTA and cOmplete protease inhibitors (Sigma Aldrich, Germany) through the previously prepared cannula in the trachea. Collected BAL fluid was centrifuged at 300x*g* for 10 min and cells were taken up in FACS buffer (PBS with 10% fetal bovine serum) and stored on ice until measurement.

Whole lung single cell extraction. Single cell suspensions from whole lung tissue were generated as described earlier (cf. 2.2.1).

Flow cytometry. Cells were counted and per sample, 2x 100.000 cells were stained with 7-aminoactinomycin D (7-AAD; ThermoFisher) for dead-live cell discrimination. Subsequently, cells were blocked using 5% PBS/normal mouse serum, 5% PBS/normal rat serum and 1% PBS/mouse Fc blocking reagent (eBioscienes, USA). Following antibodies were used for the staining of the cells: CD45 (clone 30-F11, rat, 1:800, eBiosciences), Ly6G (clone 1A8, rat, 1:200, BD Biosciences), CD11c (clone HL3, armenian hamster, 1:100, BD Biosciences), CD11b (clone M1/70, rat, 1:200, eBiosciences), MHC-II (clone M5/114.15.2, rat, eBiosciences), CD24 (clone M1/69, rat, 1:1000, Biolegend), CD64 (clone X54-5/7.1, mouse, 1:400, Biolegend), Ly6C (clone AL-21, rat, 1:200, BD Biosciences), and SiglecF (clone E50-2440, rat, 1:1500, BD Biosciences). Cells were incubated for 20-30 min at room temperature and fixed using BD CytoFix Fixation Buffer (BD Biosciences, USA). Data acquisition was performed on the BD LSRII cytometer (Becton Dickinson, Germany). Data was analyzed using the FlowJo software (TreeStart Inc., Ashland, OR, USA) and freely-accessible R script codes.

3 Results

3.1 Single cell transcriptomics reveals a novel alveolar epithelial progenitor

The work that is presented is this chapter (3.1) has been drafted into a manuscript for submission to a scientific journal. Hence, this work has not yet been peer-reviewed by the time of submission of this doctoral thesis. Nevertheless, a preprint version of this manuscript has been uploaded onto the *bioRxiv* server, which is a free online archive and distribution service for unpublished preprints in the life sciences and operated by the Cold Spring Harbor Laboratory, NY, USA. The file is accessible through the *bioRxiv* website (https://www.biorxiv.org) or directly via the digital object identifier: https://doi.org/10.1101/705244. For the following chapter, and as has been used in parts for the material and methods section (2), the citation for this preprint will be 'Strunz et al., 2019': [198].
Being the first author of this manuscript, I was involved in all parts of the study, including the second project, which has not yet been published. I performed all animal experiments and practical wet lab work, executed programming code, and drafted all figure panels. Some of the panels I could not create on my own and was therefore thankful to receive computational support. Credits need to be given to Dr. Simon Lukas for data integration methods, trajectory modeling and gene fit modeling, as well as for providing code for the cellular deconvolution analysis: Fig.11a; Fig.22 b; Fig.28 c, d; Fig.29; Fig.32; Fig.43 f; Likewise I want to acknowledge the contributions and the continuous support of Meshal Ansari in solving issues that emerged during the single cell data analysis: Fig.13 d, e; Fig.14 b; Fig.26 b.

3.1.1 Aims and hypothesis

The lung is the organ, which is constantly exposed to our environment by the air that we inhale. It is an organ without high cellular turnover but there are cellular niches that are home to stem cells and stem cell-like progenitor cells that can react to local injury. The alveolar niche has been described to react to injury by induction of alveolar type-II cells that transdifferentiate to alveolar type-I cells. But also local airway-derived cell types are involved during regeneration processes. However, the molecular events guiding this transdifferentiation have not yet been defined. Such findings are based on classical *in*

vivo lineage tracing studies in mice, which rely on the tracing of a single gene and may be therefore misleading considering the very plastic remodeling. While data suggests a remarkable cellular plasticity during alveolar regeneration, it remains difficult to prove that plasticity, as stringent lineage tracing with high temporal labeling of cell types is almost not feasible. For the first time, with the advent of single cell transcriptomics, hundreds to thousands of genes from a single cell can be 'traced' giving rise to unprecedented possibilities in studying repair mechanisms.
As many pathways and factors play a role during regeneration, the overall aim centralized around the resolution of the many molecular events that guide and support but also counteract to a proper tissue repair response. Combining single cell RNA sequencing with a mouse model that uses bleomycin to induce alveolar damage, it was aimed to denote the longitudinal transcriptional changes and mechanisms via which cells regenerate the destructed alveolar architecture. More tangibly, to cast light on the molecular events during AT2-to-AT1 conversion, the hypothesis stated that AT1 cell regeneration after injury must be dependent on intermediate progenitor cells or cellular states that are tightly regulated over time.

3.1.2 Introduction

Despite being extensively studied, the complex mechanisms after lung injury remain elusive. In the mouse, many efforts have been made to better understand regenerative responses after lung injury and to ascertain cell populations that guide or drive tissue regeneration, such as tissue-resident stem and progenitor cells [198]. Using genetic reporter mice with the aid of the *Cre* recombination technology, several cell lineages were described to promote tissue regeneration with respect to both the location and the severity of the injured lung [38, 198, 234, 235]. Epithelial cells from both the airway, as well as the alveolar compartment contribute to regeneration thanks to a highly plastic source of stem and progenitor cells that get activated by complex cell-cell communications and niche signals after injury [198, 236]. However, this intricate orchestration remains largely unknown, including how tightly controlled is the spatiotemporal crosstalk between epithelial, immune, and mesenchymal cells after injury [198, 237, 238].

As introduced in Chapter 1.1.2, AT1 and AT2 cells constitute the main cell populations in the alveolus. There is evidence that AT2 cells can serve as a progenitor population for AT1 reconstitution after alveolar injury or during homeostatic turnover [57]. While this AT2 transdifferentiation potential seems valid for minor injuries, and is maybe restricted to local damage, studies revealed that upon massive loss of alveolar epithelial cells after severe injury, airway-derived stem cells may be recruited for tissue regeneration [198, 239–243]. On the molecular level, recent findings suggest a cell cycle activation of AT2 cells for alveolar

regeneration by the inflammation-induced NFkB signaling [244]. Yet, other results spotted Yap/Taz activity in AT2 cells to be essential for correct alveolar regeneration [198, 245]. As part of the Hippo pathway, these proteins negatively regulate NFkB. For the AT2-into-AT1 transdifferentiation, AT2 cells require to undergo a cellular reorganization, which hence seems to be guided by signals from multiple pathways. Additionally to the NFkB and Hippo signaling pathways, TGFb signaling has been identified as a major regenerative driver by mediating an arrest of the cell cycle in AT2 cells [198, 246]. Nevertheless, a comprehensive knowledge of the molecular details is missing.

Combining single cell RNA sequencing using the Drop-Seq platform [179] with novel computational methods (see 2 for more details), new biological aspects can be addressed at unprecedented resolution. For instance, using RNA velocity [223], future cellular states can be predicted on the basis of a spliced-to-unspliced ratio of the transcripts within a single cell [198]. Moreover, cell fate trajectories can be modeled by deploying the diffusion pseudotime algorithm [247, 248]. Interrogation of the clonal history of cells can be achieved by studying allele frequencies of somatic mitochondrial mutations [198, 248]. Such analyses can be exceptionally powerful using high-resolution longitudinal data, which is generated by timely well-coordinated sampling, e.g. hourly or daily sampling [198, 249, 250]. Additional benefits from such data is the possibility to investigate dynamic cellular communications between cell types by the construction of cell-cell communication networks, approximated by screening receptor-ligand databases [198, 249, 251].

The following sections will summarize the results of the herein conducted investigations. Using the bleomycin-mediated lung injury mouse model, 28 lung cell types were identified from single cells, which were derived from the dissociation of whole lungs [198]. Single cell sequencing data was used to study the dynamics of gene expression, as well as cell-cell communication routes during a 4-week regeneration course [198]. The analysis revealed the discovery of an intermediate cell state with progenitor features during alveolar regeneration, concomitant with the increasing accumulation of myofibroblasts and M2-phenotypic macrophages [198]. In order to gain enhanced insights into this novel cellular state, a second experiment was conducted, focusing on only epithelial cells from the murine lung. In a skydive approach [252], sorted epithelial cells were sampled at very high temporal resolution and were analyzed with respect to cell state transitions after injury, including potential routes towards AT1 cell fate transitions [198].

3.1.3 A single cell portrait of lung regeneration

Bleomycin-induced lung injury was studied at six time points after injury - day 3, 7, 10, 14, 21, and day 28, n=4 for each time point. Uninjured control mice (n=7) were treated

with sterile PBS. Per examined time point, at least four replicate mice with each approx. 1,000 cells were sequenced, which finally, after filtering and quality control, resulted in a data set containing 29,297 cells [198]. Single cell transcriptomes were subjected to data dimensionality reduction using the Uniform Manifold Approximation and Projection (UMAP) method [253], and were displayed with a color code for each sampling time point (Fig.9a). Unsupervised clustering of the data revealed 46 cell clusters. By manual annotation of these clusters with known canonical marker genes from published single cell RNA sequencing studies [201, 254], 28 lung cell populations could be tagged and separated into major lung cell lineages (Fig.9b) [198]. Thanks to the sampling at different time points, the UMAP embedding of the single cell transcriptomes even unveiled cellular gene expression states that changed over time; such gradual movements suggest cell differentiation processes during regeneration (Fig.9a) [198].

For the validation of the degree of reproducibility with respect to technical quality and the assignment of cell types across all mice [198], the UMAP visualizations were color-coded according to treatment groups (PBS and bleomycin), or assessed by overlaying the individual replicates (Fig.10a, b). Analysis of the fraction of cells for each cell type across all the replicate mice confirmed good technical reproducibility (Fig.10c), and the summary statistics for the alignment of sequencing reads were comparable across all mice (Fig.10d).

For a comparison of the single cell-derived data with prior published bulk RNA sequencing and proteomics data from the day 14 time point after bleomycin injury [139], the single cell sequencing data was used to generate an *in silico* bulk (scRNA-seq) data set (see chapter 2.3.3) [198]. Bulk RNA-seq data and the *in silico* bulk (scRNA-seq) data were merged with the proteome data and quantile normalized [198]. Principal component analysis shows the clustering by data modality (RNA-seq, *in silico* scRNA-seq, proteome) in the first two principal components (PC), and how the third principal component separates the bleomycin-treated mice from the control animals (Fig.11a). Checking for the highest gene loadings of PC3 demonstrated clear shared features across all data modalities and hence, the global bleomycin-induced effects (Fig.11d) [198]. Studying of the loadings of PC3 over time showed a clear peak of those shared features at day 10 with a gradual decline over the regeneration course (Fig.11b) [198]. The cell type specific expression is hereby mainly restricted to alveolar epithelial cells, macrophages and fibroblasts (Fig.11c) [198]. In order to better chart the molecular dynamics of each of the cell types, differential expression time course analysis was executed (see chapter 2.3.3) [198]. In total, 6,660 genes revealed to be significantly changed after bleomycin-mediated lung injury in at least one cell type (FDR < 0.1) [198]. To allow for an easy exploration of this data, and to track changes of gene expression in the lung after injury, as well as during the regenerative phase, an interactive web tool has been

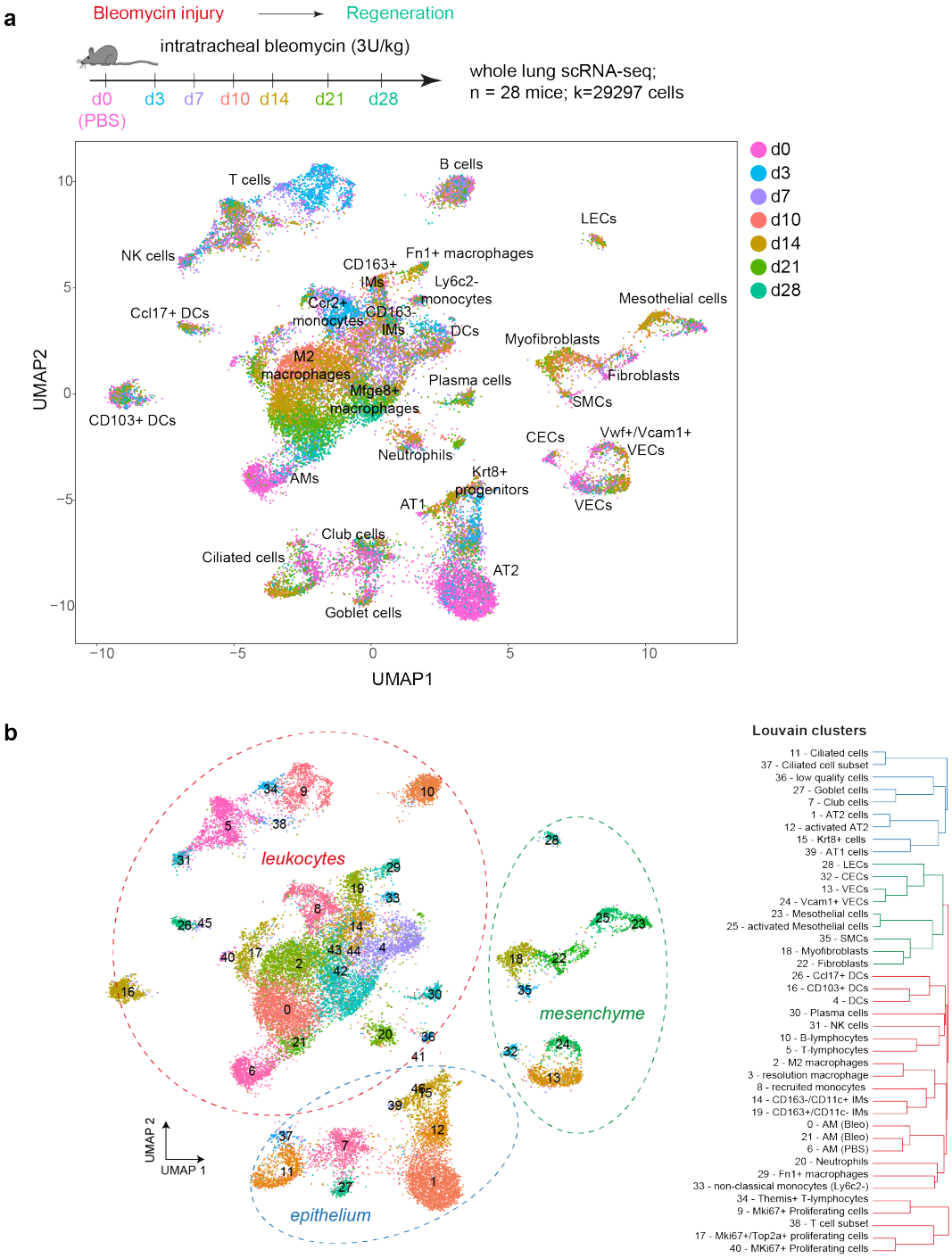

Figure 9: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Longitudinal single cell RNA-seq analysis of lung regeneration reveals different cellular dynamics over time. **a** Single cell suspensions from whole mouse lungs were analyzed using scRNA-seq at the indicated time points after bleomycin-mediated lung injury. The color code in the UMAP embedding shows shifts of the indicated cell types in gene expression space during the regeneration time course. **b** UMAP embedding colored by Louvain clusters demonstrates separation of cells into major lineages; unsupervised hierarchical clustering of the Louvain clusters recapitulates known hierarchical cell type topology.

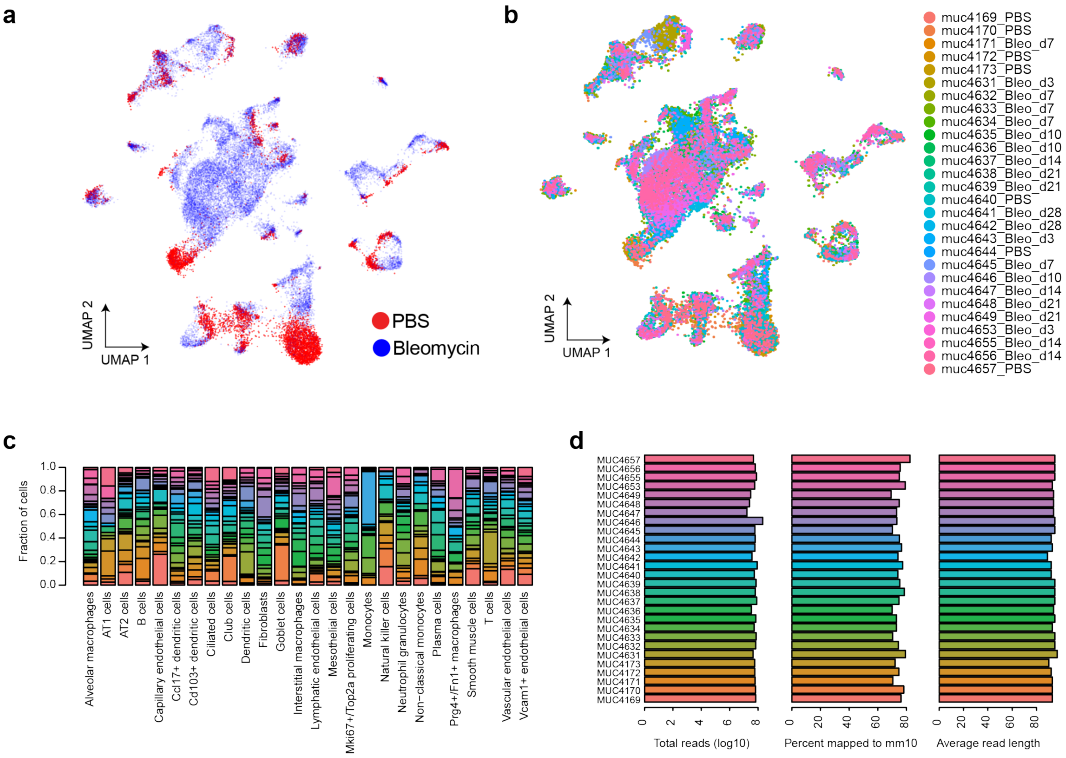

Figure 10: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Good technical agreement of whole lung single cell transcriptomes of 28 individual mice. UMAP embeddings show good overlap between treatment conditions (**a**) and individual mouse replicates (**b**). **c** Bar plot shows high overlap of mouse samples across cell types. **d** Alignment summary statistics are comparable across mouse samples.

created, accessible at http://146.107.176.18:3838/Bleo_webtool_v2/ [198].

Exploration of relative cell frequencies over time helps to understand the dynamic processes during lung regeneration and to address potential cellular movements. For instance, the relative cell frequency of proliferative T cells showed a massive increase after injury at day 3 (Fig.12a) [198]. Likewise, monocyte recruitment after injury is noticeable (Fig.12b), the accumulation of Arg1-positive M2-phenotypic macrophages culminating at day 10 and onwards after injury (Fig.12c), or the transient appearance of myofibroblasts (Fig.12d) [198]. Another example to show is a resolution macrophage cell state, marked by the expression of Mfge8, which is gradually rising in number over time (Fig.12e).

There is evidence that myofibroblasts can convert back to fibroblast phenotypes, as found in the regular tissue steady state, during the course of resolution of fibrosis [198, 255]. Fig.12d demonstrates the transient emergence of myofibroblasts, which are known fibrogenic cells marked by the expression of matrix genes, such as type-1 collagen (Fig.12f), alpha-smooth muscle actin (Fig.12g), and tenascin-C (Fig.12h); shown here are relative expression levels of the respective genes [198]. Comparing the gene expression levels between myofibroblasts (right side in the volcano plot) and fibroblasts (left side in the volcano plot) reveals distinct expression of known markers for each of the two cell populations (Fig.12i), and extracellular matrix components that are shared between the mesenchymal lineages (Fig.12j).

Since it is known that bleomycin administration engenders a massive influx and activation of monocytes and macrophages, immune cell populations were subclustered in the data, which revealed several distinct phenotypes (Fig.13c, shown are commonly used markers for distinct macrophage subtypes) at different time points (Fig.13a, b) [198]. In order to score individual cells, the data was compared to published bulk RNA-seq signatures, which derived from lineage tracing experiments of monocyte-derived macrophages using the bleomycin lung injury model [198, 231]. This comparison showed a distinct assignment of the cells to either interstitial or alveolar macrophages (Fig.13d), the latter of which is marked by the expression of Cebpb and Ear2 (Fig.13f, g), and that monocyte-derived macrophages precede the alveolar macrophage phenotype, as well as the M2 phenotypic and resolution macrophages (Fig.13e), marked by the expression of Arg1 and Mfge8 (Fig.13h, i), respectively [198].

For an improved understanding of cellular interactions, receptor-ligand pairs across all cell types were used to implement a cell-cell communication network (see chapter 2.3.3) (Fig.14a) [198]. By identifying receptor-ligand pairs with at least one gene being significantly expressed over time in the corresponding cell type, expression dynamics could be integrated [198]. Based on the proportion of receptor-ligand pairs, the highest possible cellular communication

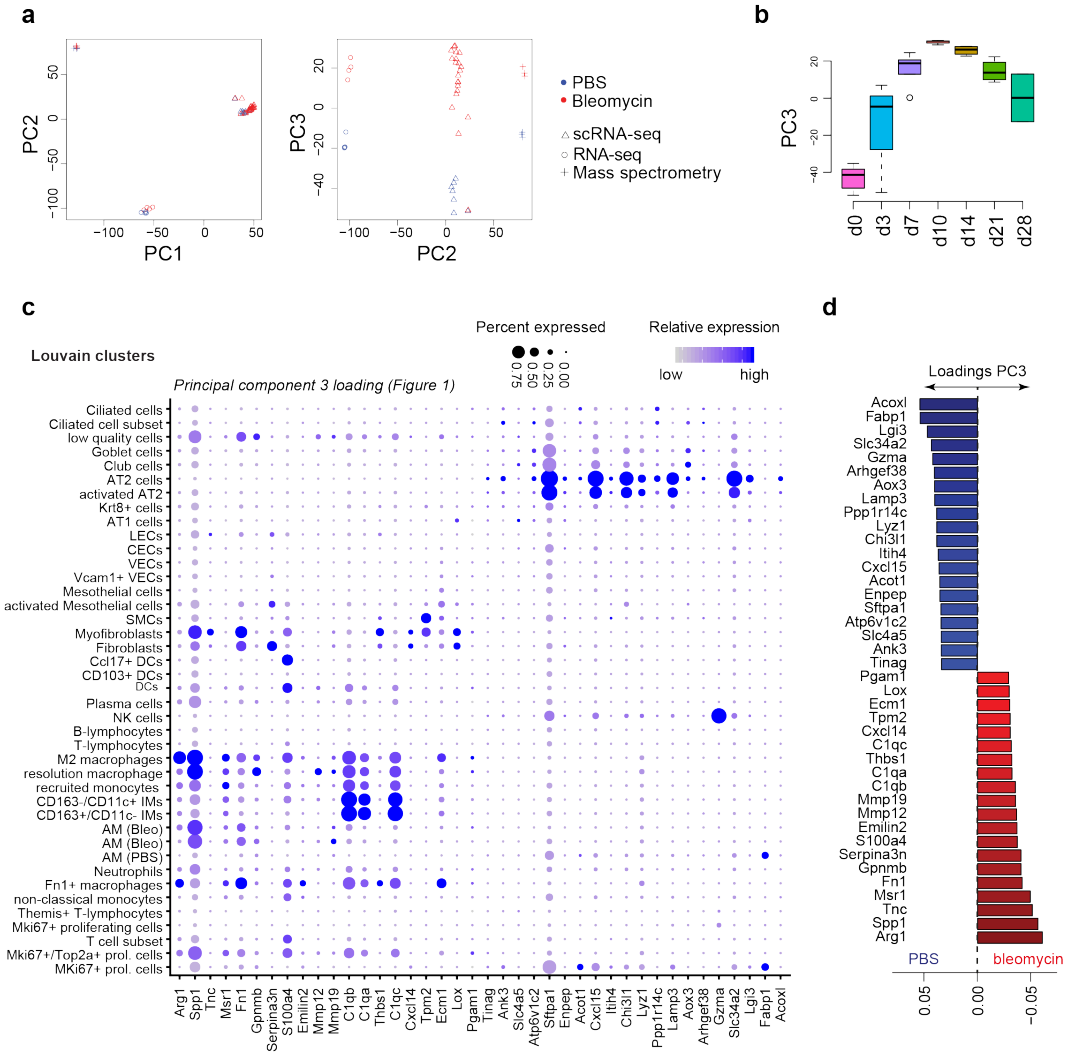

Figure 11: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
a Normalized bulk (RNA-seq) and *in silico* bulk (scRNA-seq) data were merged with proteome data (mass spectrometry) and quantile normalized. Bulk and protein data contain samples from day 14 after bleomycin-induced injury and controls. The first two principal components show clustering by data modality. The third principal component separates bleomycin samples from controls across all three data modalities. Blue and red colors indicate control and bleomycin samples, respectively. **b** The box plot shows the time-resolved loading of PC3 peaking at day 10. The boxes represent the interquartile range, the horizontal line in the box is the median, and the whiskers represent 1.5 times the interquartile range. **c** The dot plot shows average expression of genes with top PC3 loadings across cell type identities. **d** The barplot depicts genes with the highest loadings for PC3.

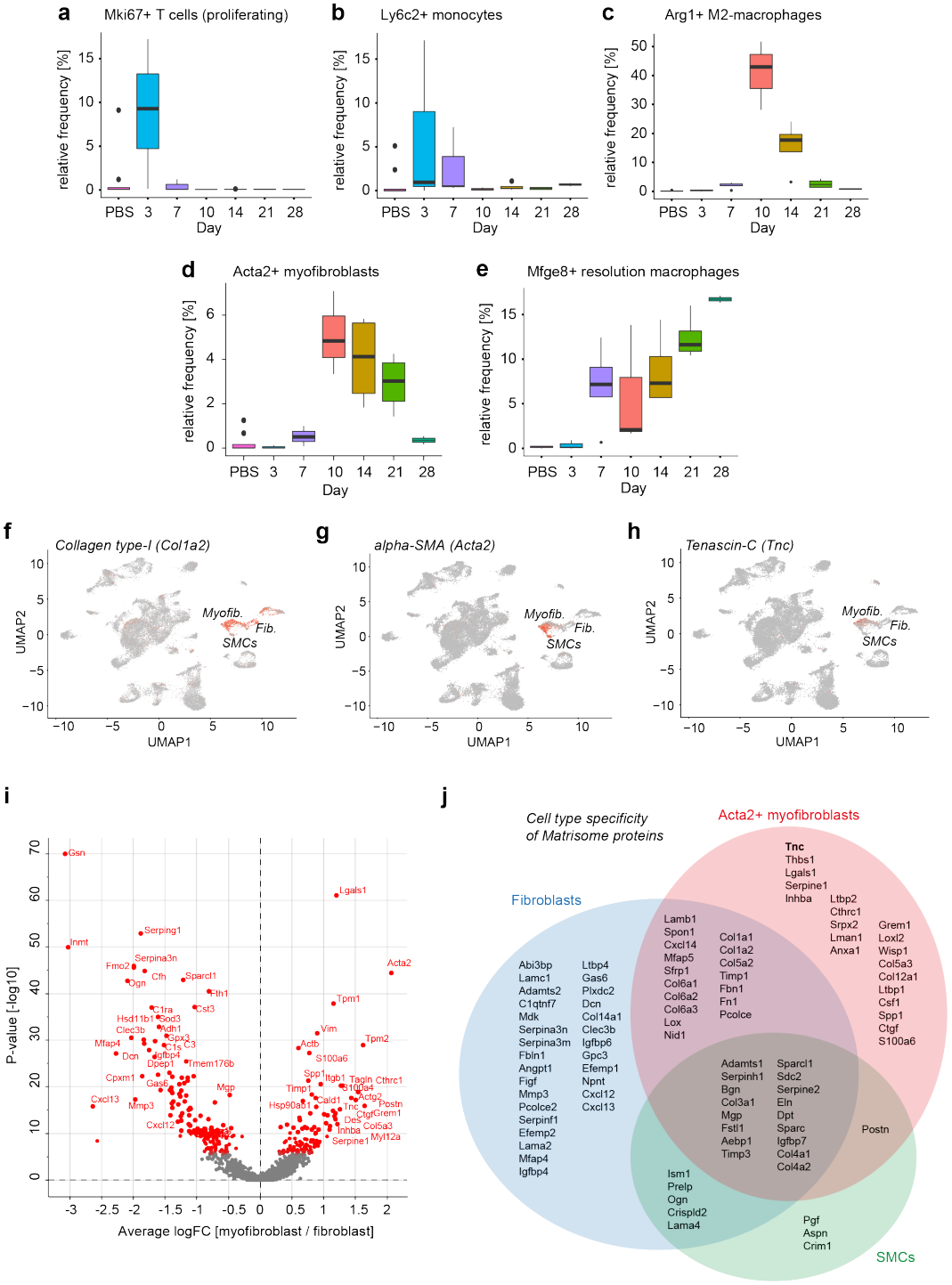

Figure 12: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
a-e Relative frequency of the indicated cell types relative to all other cells was calculated for individual mice at the indicated time points after injury (n=4) and for PBS treated control mice (n=7). The boxes represent the interquartile range, the horizontal line in the box is the median, and the whiskers represent 1.5 times the interquartile range. **f-h** Relative expression levels of Col1a2 (**f**), Acta2 (**g**), and Tnc (**h**) are shown on the UMAP embedding. **i** The volcano plot shows differential gene expression between myofibroblasts (right side) and fibroblasts (left side). **j** Single cell analysis was used to derive the myofibroblast specific ECM components in comparison to fibroblasts and smooth muscle cells.

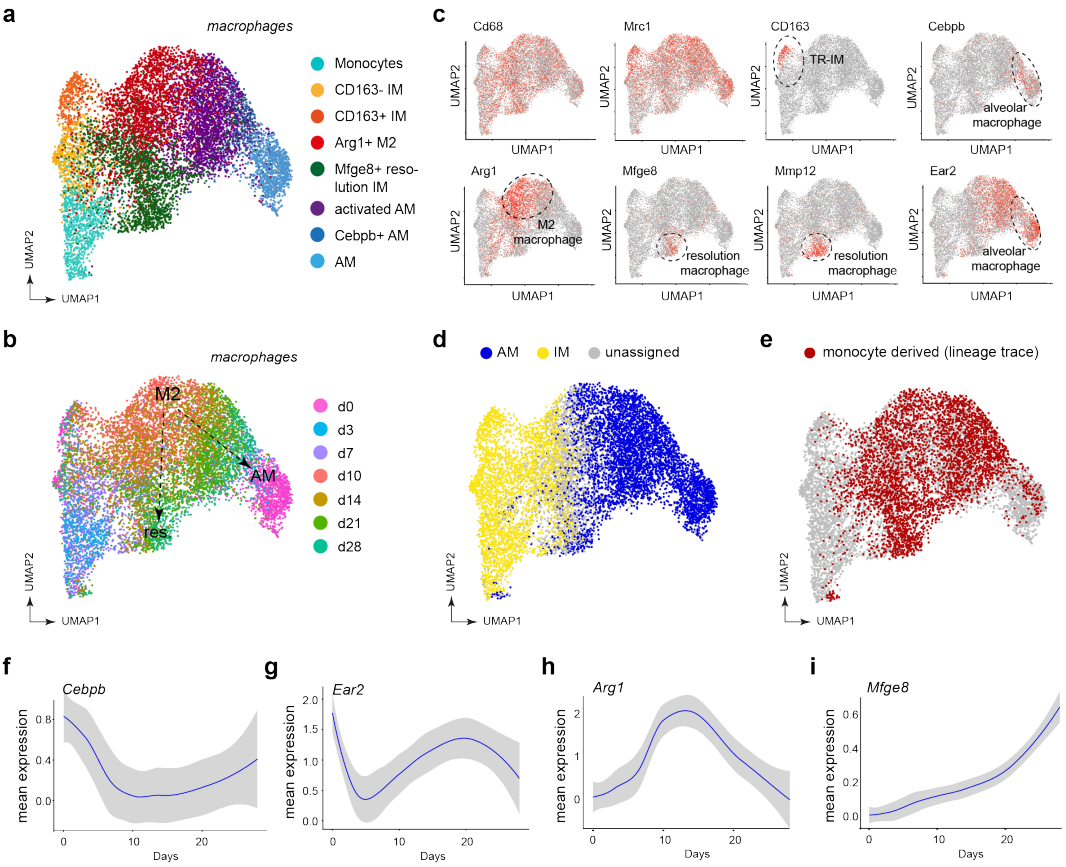

Figure 13: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Dynamics of macrophage states in lung tissue regeneration. **a, b** UMAP embedding of 10,379 cells that express known macrophage markers, colored by cluster identity (in **a**) and time points (in **b**). Following cells along the time course after reaching the peak of inflammation at day 10 and 14, two potential trajectories from M2-phenotypic macrophages can be discerned. **c** Several macrophage populations can be identified. These clusters uniformly express the macrophage markers Cd68 and Mrc1, while also showing distinct expression of other specific genes. **d** Previously published gene signatures from bulk RNA experiments were used to reveal potential origins of macrophage cells. In that data set, FACS-sorting allowed to differentiate between tissue-resident alveolar (AM), interstitial (IM) and monocyte-derived macrophage populations [231]. Similarity score of each cell is calculated as correlation to differentially expressed genes and corresponding log fold changes in the three sorted populations. Cells are assigned to either AM or IM category, if the difference in scores for either category is higher than 0.05. Alveolar macrophages in this herein presented data set indeed show the highest score on the tissue-resident AM. **e** Potentially monocyte-derived cells based on scoring (at threshold of 0.1). There is a separation in the potentially monocyte-derived cells, which concurs with the real-time trajectories in (**b**). Relative cell type frequency per time point of alveolar macrophages (**f, g**), M2 macrophages (**h**), and Mfge8+ macrophages (**i**) across all samples and the smoothed expression per time point of Cebpb (**f**), Ear2 (**g**), Arg1(**h**), and Mfge8 (**i**) in the macrophage subset with the confidence interval of 0.95.

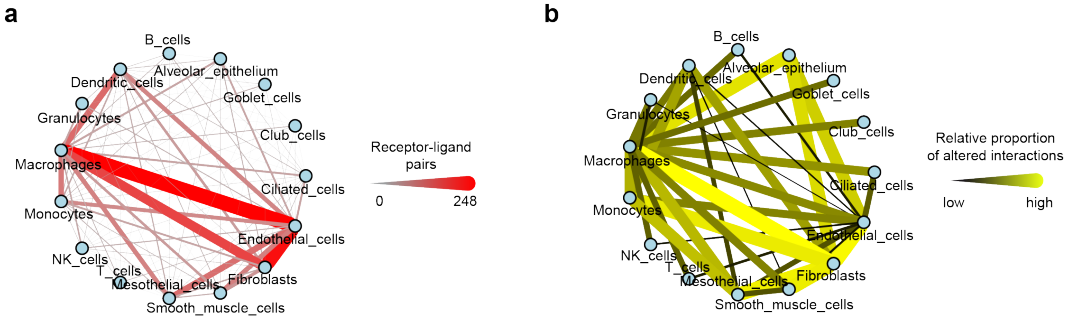

Figure 14: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Longitudinal single cell RNA-seq analysis of lung regeneration reveals cell communication dynamics. **a** The network shows 15 meta-cell type identities and their putative communication structure. Edge weight and color illustrate the number of receptor-ligand pairs between cell types. **b** The edges represent the relative proportion of receptor-ligand pairs between cell types with altered expression after injury.

after injury was observed between fibroblasts and macrophages and their altered interaction with the epithelium (Fig.14b) [198].

3.1.4 Alveolar regeneration involves a novel cell state

During the annotation of the cell clusters with canonical markers, in particular one alveolar epithelial cell cluster showed a distinct but unknown gene expression pattern with a notable expression of Keratin-8 (Krt8) at day 10 and later after injury. Partitioning of the alveolar epithelium by subclustering of the corresponding single cell transcriptomes demonstrated a split into 4 distinct clusters (Fig.15a), which in large parts represented different time points (Fig.15b). When observed at more detail, the connectivity between AT1 and AT2 cells is mainly derived from intermediate time points [198], indicating that cells are undergoing cellular transdifferentiation. To better trace these cellular dynamics, hierarchical clustering was employed, resulting in two AT2 cell clusters marked by Sftpc expression, with one of which representing an activated cell state defined by the expression of the injury-induced genes Lcn2 and Il33 (Fig.15c, d, e) [198]. While AT1 cells were identified by the expression of known markers, such as Pdpn, Ager, or Hopx, one cluster remained unassigned as no canonical genes for AT1 or AT2 cells could be detected (Fig.15c, f). However, clustering disclosed a marked expression of Krt8 among other genes. The gene expression plot in the UMAP embedding (Fig.15g) confirms that this newly discovered cell state is only found at stages during fibrogenesis and distinguishable by Krt8 expression. In order to address whether there is a possible transition from AT2 to AT1 cells, the single cell analysis method scVelo was used (see chapter 2.3.2 for details) [198]. This method makes use of splicing events of each individual cell and thus infers RNA velocities [223] to predict a cell's future state [198]. Overlaying the vectors resulting from the scVelo analysis with the UMAP

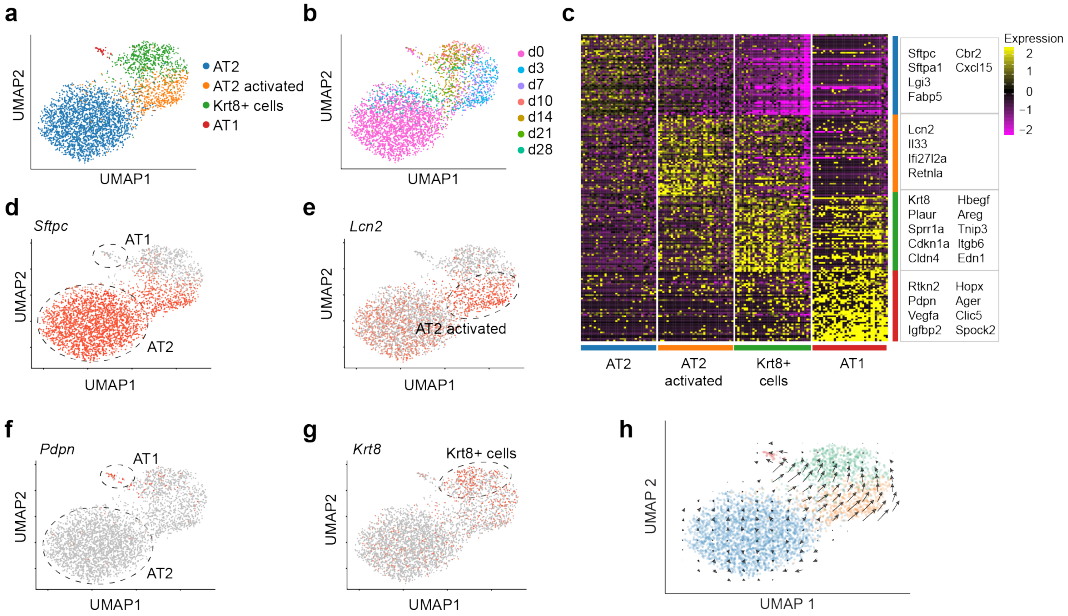

Figure 15: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Discovery of a novel cellular state involved in alveolar tissue regeneration. **a** UMAP embedding of alveolar epithelial cells shows four distinct gene expression states. **b** The color code shows the time points of sampling on the UMAP embedding. **c** Heatmap shows the 50 most differentially expressed genes for the four alveolar cell states. The box shows gene names of selected examples. **d-g** UMAP embedded visualizations of single cells colored by gene expression for the four distinct gene expression states: AT2-Sftpc, AT2 activated-Lcn2, AT1-Pdpn, alveolar Krt8+-Krt8. **h**RNA velocity information is plotted onto the UMAP embedding. Each arrow represents the local direction of transcriptome dynamics, estimated by comparing spliced vs. unspliced transcripts. Arrows are pointing towards the alveolar Krt8+ cell state after bleomycin-mediated injury.

embedding suggests the establishment of the Krt8-positive cell state to be derived from the AT2 via the activated AT2 cell state, potentially even reaching to AT1 cells (Fig.15g, h) [198].

To confirm the presence of these cells in the lung, and to assess whether this expression on RNA level is also present on the protein level, immunofluorescence staining of control (PBS) and bleomycin-treated lung tissue sections were executed. In the lung, Krt8 expression is found to be located in airway epithelial cells, which served as a control for the success of the staining (Fig.16a, highlighted time points). Compared to control sections, indeed, the staining confirmed the appearance of Krt8+ cells in the parenchyma of bleomycin-treated lungs. Furthermore, the characteristic of transient *de novo* Krt8 expression during the time course became evident (Fig.16b, all time points stained). Quantification of the mean fluorescence intensity of Krt8 signal, excluding Krt8-positive airway cells, additionally corroborated the transient expression in the lung parenchyma (Fig.16c). Along with these findings, there were rare Krt8+ cells found in the alveolar space of uninjured or PBS-treated lung sections (Fig.16d, highlighted by yellow arrowheads), raising the possibility that the observed cell state also represents a naturally occurring homeostatic cell state [198].

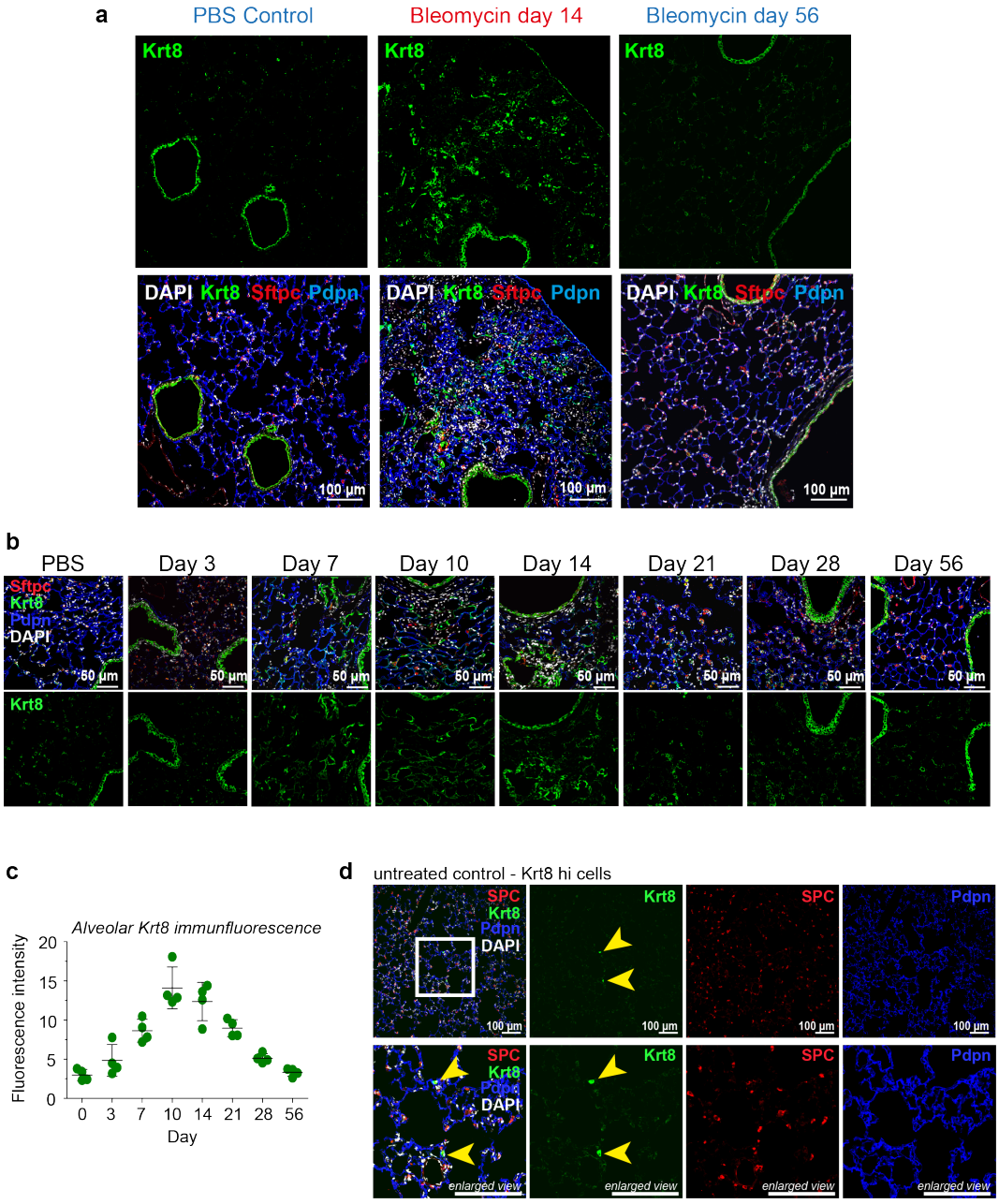

Figure 16: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
a Fluorescent immunostainings and confocal imaging of lung sections from the indicated conditions. Nuclei (DAPI) are colored in white, Krt8 appears in green, Sftpc (AT2 cells) in red, and Pdpn (AT1 cells) in blue. The scale bar indicates 100 microns. **b** Immunostaining of Krt8 (green) at the indicated time points after bleomycin injury. FFPE tissue sections were co-stained with the AT2 marker Sftpc (red), and the AT1 marker Pdpn (blue). Nuclei were labeled using DAPI (white). Scale bar = 50 microns. **c** Krt8 expression quantified by the mean fluorescence intensity of selected regions in the alveolar space, excluding Krt8+ airways. Alveolar Krt8 expression is highest at day 10 and day 14 (n=4 per time point, mean with SD). **d** Rare Krt8+ cells in normal alveolar homeostasis. Fluorescent immunostainings and confocal imaging of lung sections from untreated control lungs. Nuclei (DAPI) are colored in white, Krt8 appears in green, Sftpc (AT2 cells) in red, and Pdpn (AT1 cells) in blue. The scale bar indicates 100 microns.

The transient protein expression character of Krt8 was also validated by mass spectrometry-based whole lung tissue proteomics (Fig.17a) [198]. Next to Krt8 protein quantification based on immunofluorescence staining and mass spectrometry, flow cytometry-based quantification was added (Fig.17b). EpCam-positive epithelial cells were further subdivided into Krt8+ and integrin-beta6 (Itgb6)+ cells (Fig.17b-d), both of which were found at increased number after bleomycin-mediated lung injury. Surface expression of Itgb6 was found to be higher on alveolar Krt8+ cells (Fig.17d).

The single cell data exhibited striking expression of the low-affinity ligands Areg and Hbegf, which bind to the epidermal growth factor receptor [198]. The increased protein expression of both Areg and Hbegf in bleomycin-injured lungs was verified by immunofluorescence staining and its quantification by measuring the mean fluorescence signal intensities (Fig.18a, b) [198].

With the help of precision-cut lung slices (PCLS), whole alveolar cell morphologic analysis was performed. 300 micron-thick PCLS were immediately fixed after cutting and immunostained against Krt8 (new cell state), Sftpc (AT2 cells), and Pdpn (AT1 cells). Confocal microscopy allowed for the generation of high-resolution images of stained alveoli and the quantification of cell sphericity [198]. Maximum projection of confocal *z*-stack images showed low-level Krt8 expression in cuboidal AT2 cells (Sftpc+) of uninjured lung sections but high Krt8 expression levels in sections received from mice at day 14 post bleomycin injury (Fig.19a) [198]. Notably, the expression of Krt8 was detected in cuboidal AT2 cells (Krt8/Sftpc double positive cells) or in Sftpc-negative cells that were found in larger number and with elongated and squamous morphology [198]. Analysis of the cell shape clearly identified single Krt8+ cells with significantly reduced sphericity factor (Fig.19b) [198]. In addition, after lung injury, Krt8+ AT2 cells featured a significantly flatter shape (Fig.19c), indicating that the squamous flat Krt8+ cells may cover the alveolus similarly to AT1 cells [198].

These findings arose the question whether the formation of the Krt8 cell state was restricted to the bleomycin-induced injury model. In fact, efforts from collaborating groups could demonstrate the emergence of Krt8 high cells after exposure of neonatal mice to either hypoxic or hyperoxic condition (cf. preprint - [198]). This challenge has been reported to increase the susceptibility for lung injury and fibrosis after influenza type A infection in adult age [198,259–261]. In another approach, the exposure of adult mice to hyperoxia, which has been reported to raze out AT1 cells [198,262] showed Krt8+ cells in FFPE lung sections and were identifiable already at day three after 60 hours of the hyperoxic challenge (cf. preprint - [198]).

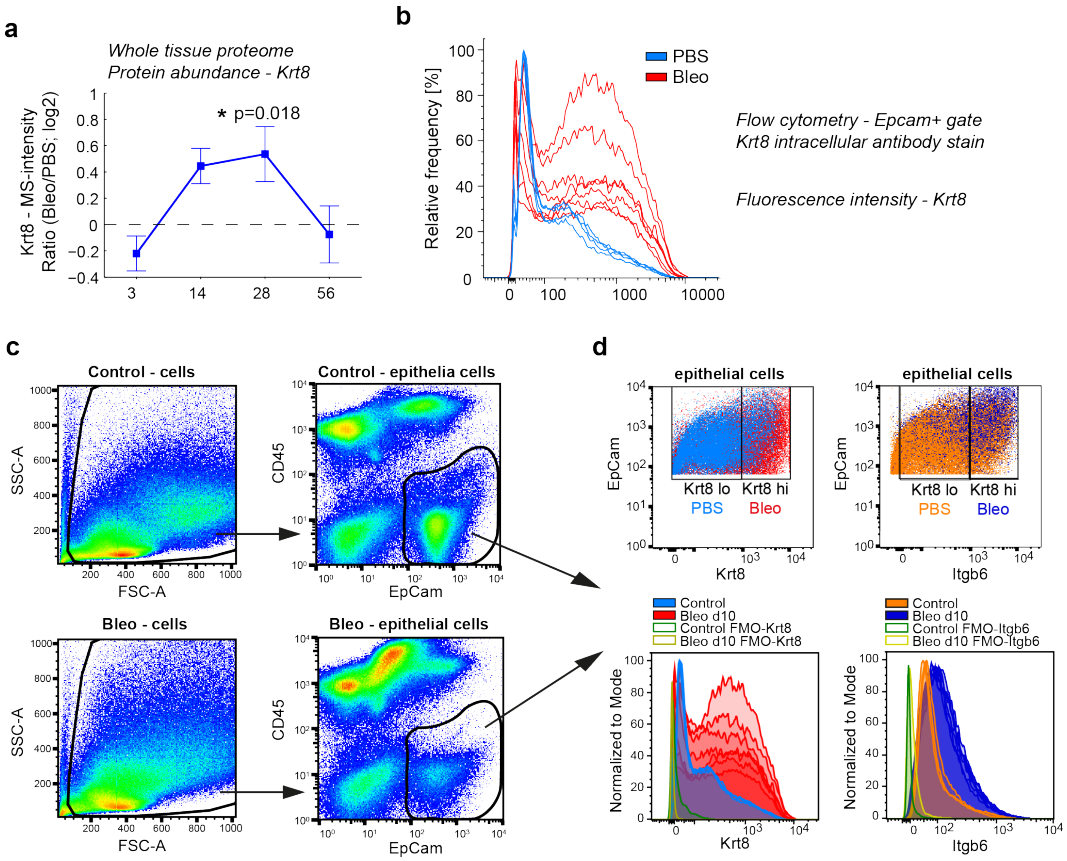

Figure 17: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Protein validation of the alveolar Krt8+ cell signature. **a** Protein abundance of Krt8 in total lung homogenates was assessed by mass spectrometry [139]. The line plot shows the log2 ratio of Krt8 MS-intensity of mice at day 14 after bleomycin injury (n=4) and PBS control mice (n=4). Error bars show the standard error of the mean. **b** The histogram shows Krt8 fluorescence intensity quantified by flow cytometry using a CD45 negative and Epcam positive gate to select epithelial cells. PBS control mice (n = 5, blue color) and mice at day 10 after bleomycin (n=7, red color) are shown. **c** Gating strategy for the analysis of CD45-/Epcam+ epithelial cells. **d** The scatter plots and histograms show increased expression of Krt8 and Itgb6 at day 10 after bleomycin in Epcam+ epithelial cells. Highest expression of Itgb6 was observed on Krt8 high cells. Fluorescence-minus-one (FMO) controls were used for both the Krt8 and Itgb6 quantification.

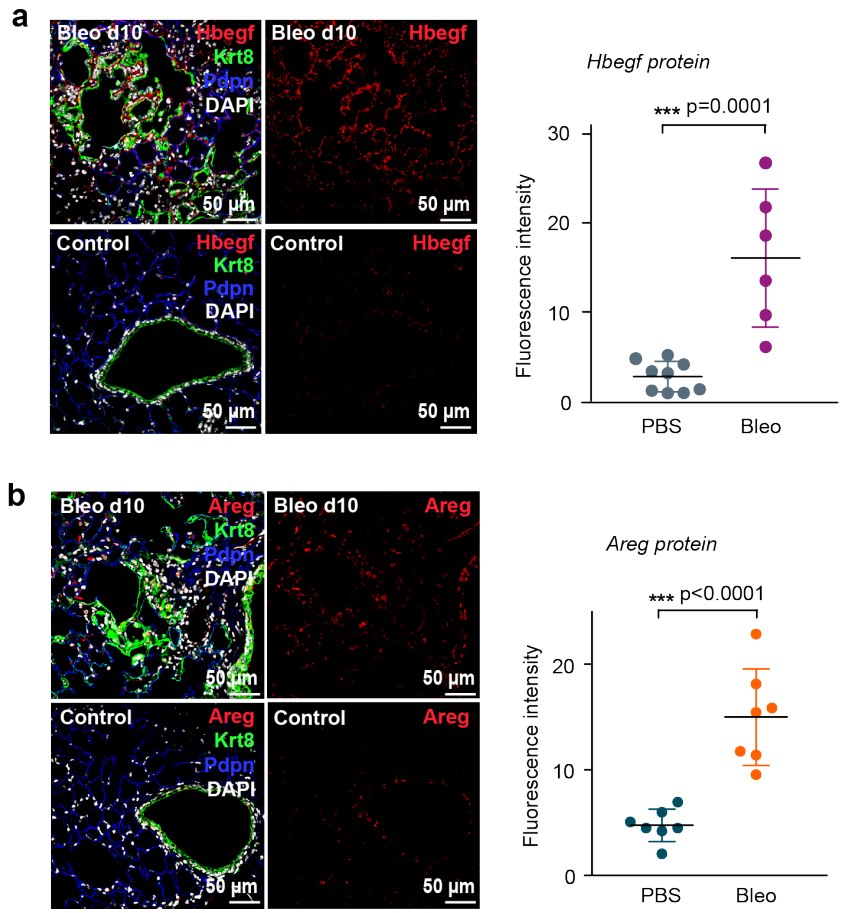

Figure 18: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
a Increased Hbegf (red) expression in bleomycin treated lung tissue, showing partial overlap with Krt8 (green) signal. Quantification of the mean fluorescence signal intensities confirmed increased Hbegf expression (unpaired t-test *** p = 0.0001). Sections were co-stained with Pdpn (blue); scale bar = 50 microns. **b** Immunostainings of Areg (red) and Krt8 (green) expression in the lung, co-stained with Pdpn (blue) and quantified by mean fluorescence intensity. Unpaired t-test *** p < 0.0001. Scale bar = 50 microns.

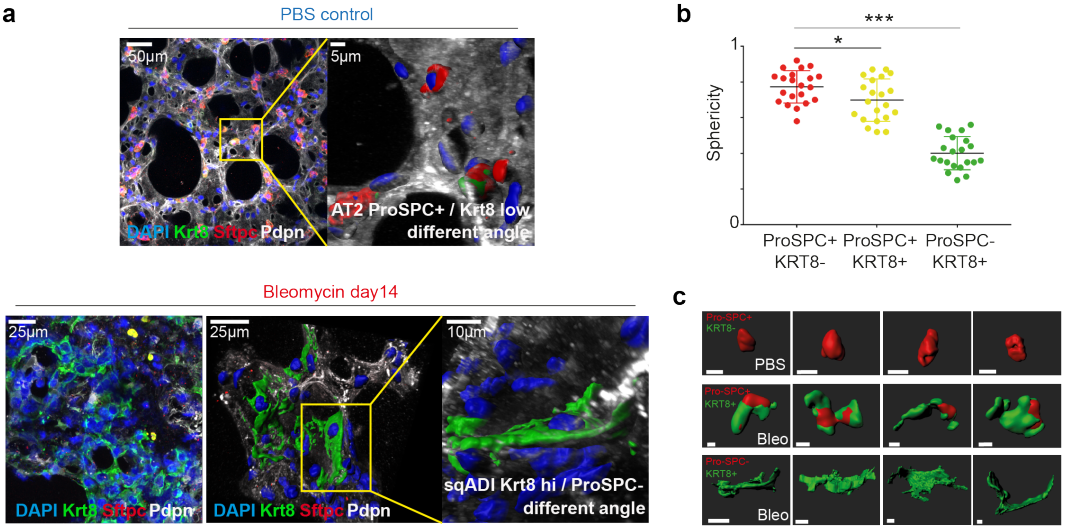

Figure 19: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
PCLS imaging reveals elongated cell shape for Krt8+ cells. **a** Maximum projections of confocal z-stacks taken from immunostained 300 micron-thick precision cut lung slices (PCLS) are shown for a representative PBS control mouse and a mouse at day 14 after bleomycin injury. Nuclei (DAPI) are colored blue, Krt8 appears in green, Sftpc (AT2 cells) in red, and Pdpn (AT1 cells) in white. **b** Alveolar cell sphericity analysis of 21 cells per condition revealed elongated cell shapes for Krt8+ cells in IF-stained precision cut lung slices (in **a**). Sphericity of 1 indicates round, cuboidal cells; 0 indicates flat cells. One-way ANOVA with Dunnett's post testing: * $p = 0.0376$, *** $p < 0.0001$. **c** The images indicate the cell morphologies found in one healthy ROI within one PBS PCLS, and in two fibrotic ROI within one bleomycin PCLS. Note that Krt8+ cells form larger networks and clusters.

Similar to the approach used in mice, a human cohort was included for immunostaining of the Krt8+ alveolar cell state. FFPE sections were stained for KRT8, SFTPC, and alpha-SMA. Lung tissue from ALI patients (n=2; ARDS/Influenza A and ARDS/Sarcoidosis) and ILD patients (EEA, n=1; IPF, n=3; and IPAF, n=1) were received from lung explant material from transplantations, whereas non-injured control lung tissue (n=7) was derived from tumor-free material after excision of lung tumor mass [198]. While control sections were almost void of KRT8+ cells in the alveolar space, massive expression of KRT8 was detectable in parenchymal tissue sections of both ALI and ILD patients (Fig.20a-c) [198].

In summary, the herein collected data proves the transient appearance of squamous Krt8-positive alveolar cells in different lung injury mouse models, as well as in human lung disease. The single cell RNA-seq data allows for the modeling of cellular dynamics over the time course of the bleomycin lung injury model. Hence, it was hypothesized whether, over time, the Krt8+ alveolar cells can function as progenitor cells for reconstitution of AT1 cells and their maturation after lung injury [198].

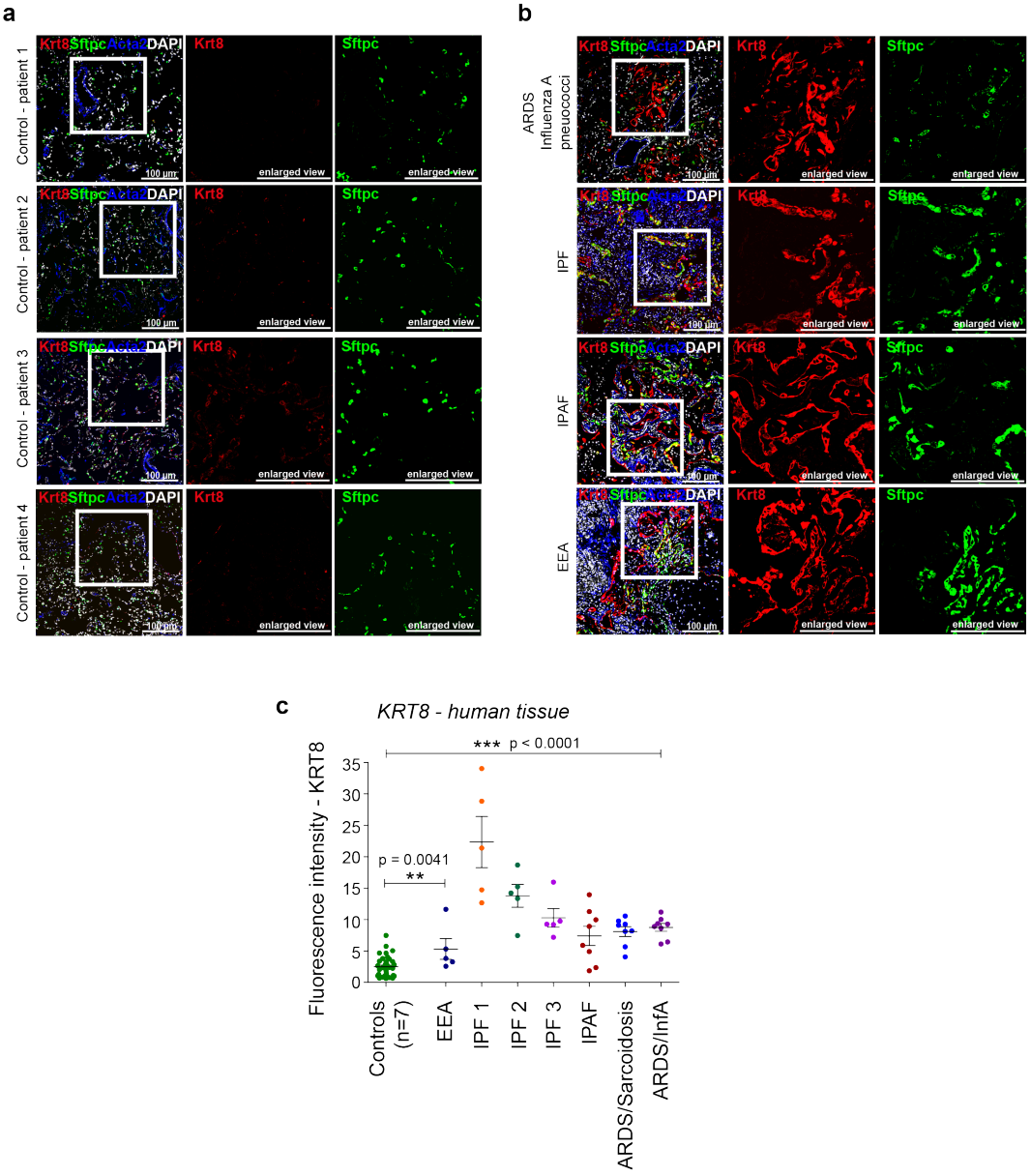

Figure 20: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Krt8+ alveolar cells appear in human acute lung injury and fibrosis. **a** FFPE sections from non-fibrotic control parenchyma derived from non-involved areas in tumor resections were stained against Krt8 (red), Sftpc (green), and Acta2 (blue). Scale bar = 100 microns. **b** Human lung tissue sections were stained as in (a) revealing pronounced Krt8 expression at the site of acutely injured lesions (ARDS diagnosis) and fibrotic regions of ILD patient lungs (IPAF, IPF and EAA diagnosis). Scale bar = 100 microns. **c** Fluorescence intensity of Krt8 stainings was quantified from 5-8 representative areas of control tissue (n=7), ILD tissue (n=5), and ARDS (n=2). One-way ANOVA statistical analysis: *** $p < 0.0001$, ** $p = 0.0041$.

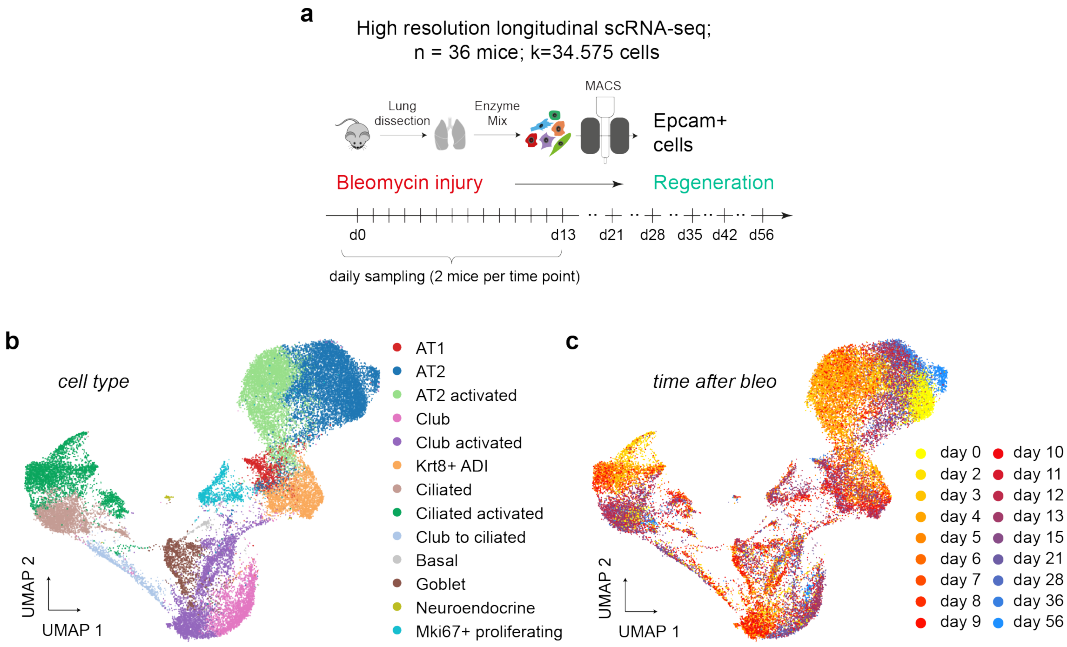

Figure 21: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
a A high-resolution longitudinal data set was generated by subjecting sorted cells from the epithelial compartment to scRNAseq from the 18 indicated time points. UMAP embedding displays cells colored by cell type identity (**b**) and time point (**c**).

3.1.5 A sky dive into epithelial cell fate transitions

After the identification of the novel Krt8-positive cell state, its progenitor potential was tested in a refined experiment. For better time- and cell type-resolved resolution with the goal to comprehensively model the emergence of Krt8+ cells from alveolar epithelial precursor cells, mice were again treated with bleomycin or PBS [198]. Single cells were extracted from the lung but prior to the scRNA-seq experiment subjected to epithelial cell selection via the EpCam surface protein by magnetic-activated cell sorting (MACS). Mice were sacrificed daily (n=2/day) until day 13, and weekly until day 56, resulting in a final scRNA-seq data set of the epithelium at 18 time points with n=36 mice and k=34,575 cells (Fig.21a) [198]. Sampling until day 56, which constitutes the fully-regenerated baseline in the bleomycin injury model, ensured to track the dynamic changes over the entire injury and regeneration course [198].

This high resolution data set of the epithelial compartment led to the classification of the same cell type identities as were found in the whole lung screen [198]. However, thanks to the improved temporal and cellular resolution, neuroendocrine cells (PNEC) and basal stem cells were additionally collected (Fig.21b, c) [198]. As some cell types revealed a remarkable transcriptional heterogeneity during the time course of regeneration, certain cell clusters

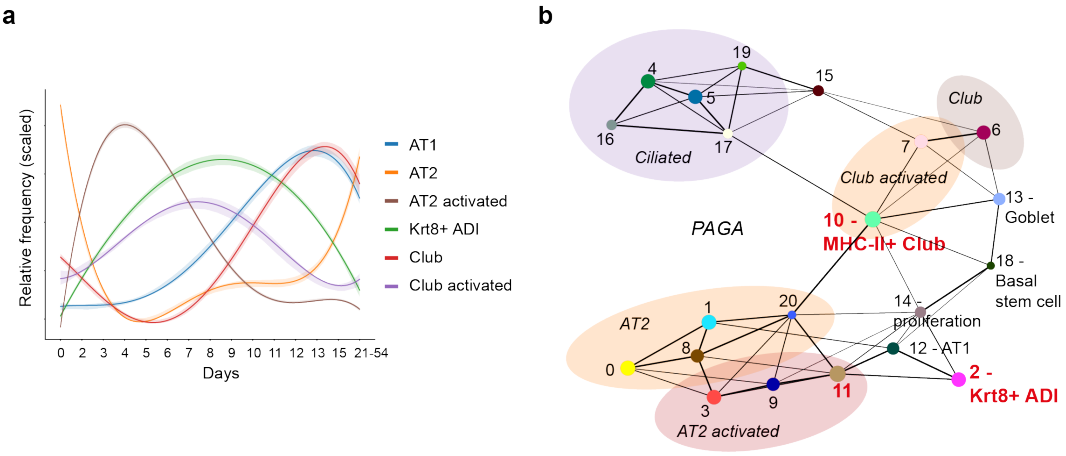

Figure 22: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
a The scaled lines represent smoothed relative frequencies of alveolar cell types and club cells over the time course. Confidence interval derived from the smoothing fit is shown. **b** PAGA graph visualizes potential cell-type transitions and the topology of the data manifold. Nodes represent Louvain clusters and thicker edges indicate stronger connectedness between clusters.

were categorized into *activated* cell states (Fig.21b, c) [198]. To understand which lineages give rise to the Krt8+ cells, relative cell frequency over time was assessed. Activated AT2 cell states preceded the appearance of Krt8+ cells just as activated club cells did, however, with the difference that the relative frequency of activated AT2 cell states was observed to be increased at earlier time points (Fig.22a) [198]. To analyze global connectivity and infer on potential cell trajectories, partition-based graph abstraction (PAGA) was used [198]. PAGA is among the best performing pseudotime methods [247] and establishes a graph-like map for the interpretation of the data manifold (Fig.22b) [198, 222]. Studying the PAGA map with more detail, several potential transdifferentiation routes emerge between the connectivity nodes of the cell clusters. To highlight are the strong connectivity of the nodes of cell clusters 10 (airway club cell subset), 2 (Krt8+ cells), and 11 (activated AT2 cells) [198].

Application of differential gene expression analysis to the club cell subsets only (clusters 6, 7, and 10 in Fig.23a), allowed for the split of club cells (cluster 6) versus activated club cells (cluster 7) [198]. Cluster 10, by contrast, remained a distinct cluster and was most prominently different by the high expression of MHC-II complex genes, including H2-Ab1 and the cysteine proteinase inhibitor Cystatin-C (gene name: Cst3) [198]. The latter has been demonstrated to be co-expressed with MHC-II by dendritic cells (Fig.23a-d) [198, 263]. The contradistinction between activated club cells and club cells is best appreciated by enriching gene categories with significant enrichment (FDR<5%) in both cell populations (Fig.23e) [198]. Expression of ribosomal genes and ATP synthesis were increased in activated club cells, while the Wnt signaling pathway was found downregulated [198]. The

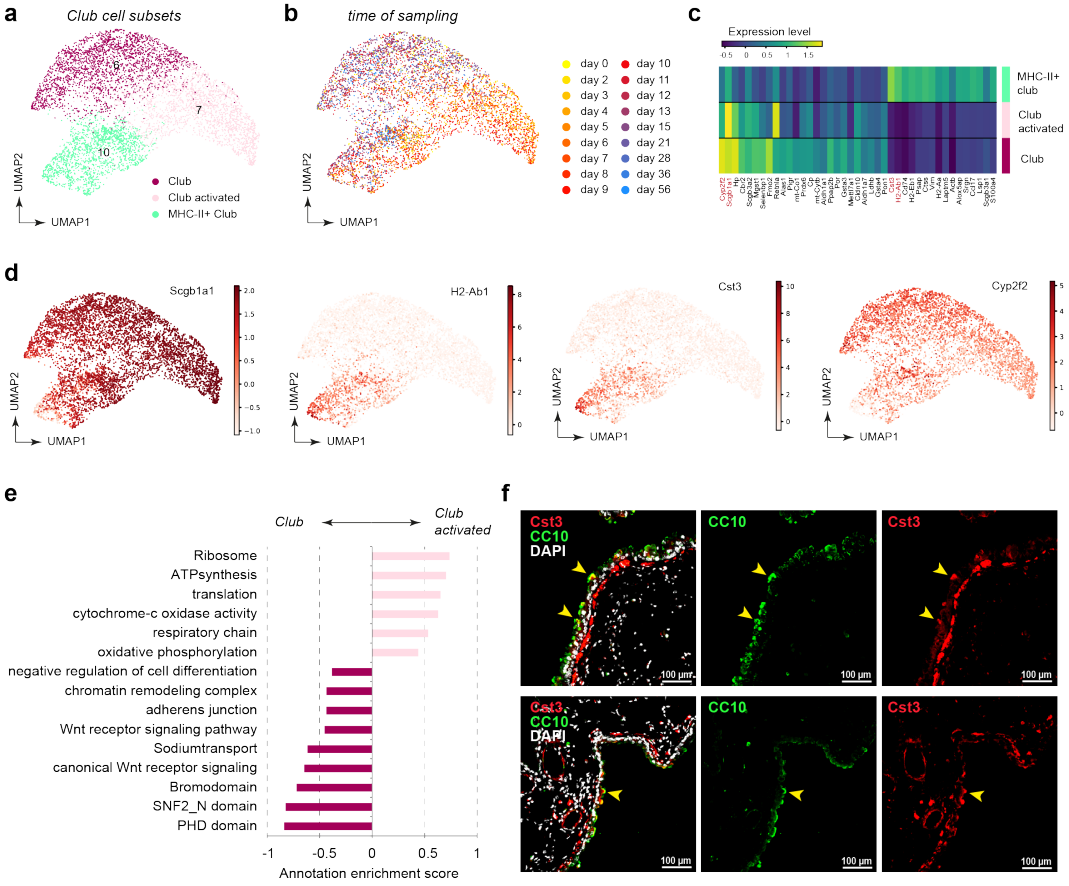

Figure 23: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Discovery of a MHC-II positive club cell subset. Plots visualize the UMAP embedding of Club cells colored by Louvain clustering (**a**) and by time point (**b**). **c** The heatmap shows the average expression levels of marker genes across the three Club cell clusters. **d** UMAP embedding shows distinct expression patterns for selected marker genes. **e** The bar graph shows the annotation enrichment score68 for selected examples of gene categories with significant enrichment (FDR < 5%) in either activated Club (positive scores) or Club cells (negative scores). **f** Immunofluorescence staining of mouse airways shows CC10+ club cells (green) and Cst3+ cells (red), DAPI (white). Note the partial overlap of Cst3+/CC10+ airway cells (highlighted by yellow arrowheads). Scale bar = 100 microns.

MHC-II+ club cell subset also features higher amounts of ribosomal genes, as well as increased cytokine and intermediate filament expression, suggesting that this subset may play a distinct role in mediating response to injury in the airway epithelium [198]. Staining against Cst3 and Scgb1a1 (CC10) confirmed the presence of Cst3+ subset cells in the airway epithelium as only few Scgb1a1+ cells were found to co-express Cst3 (Fig.23f, highlighted with yellow arrowheads).

In order to predict cellular functions and cell pathway activity based on the expressed genes of each cell type, Ingenuity Pathway Analysis (IPA, Qiagen Redwood City, USA; www.qiagen.com/ingenuity) was performed. IPA-embedded algorithms infer and score

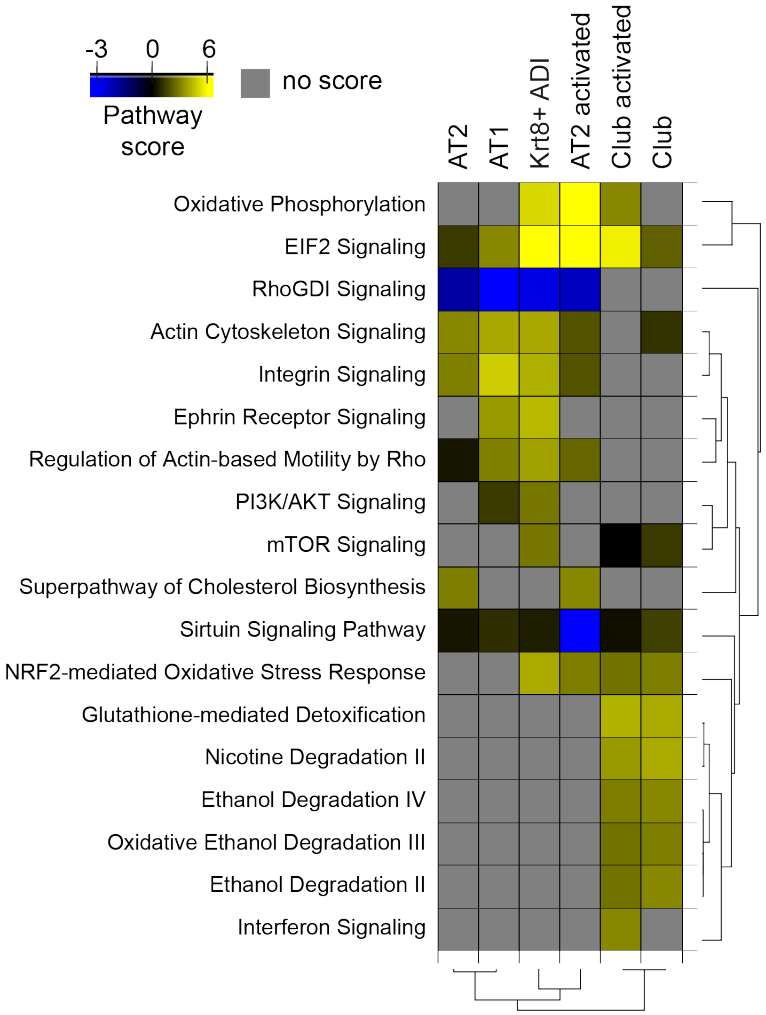

Figure 24: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Ingenuity pathway analysis was used to score the activity of pathways within the signatures of the indicated cell states. The pathway activation z-scores were grouped by hierarchical clustering using their Pearson correlation.

pathway activities and assign respective *z*-scores, which were hierarchically clustered using the Pearson correlation.

The analysis revealed that Krt8+ cells, activated AT2 cells, as well as activated club cells possess higher oxidative phosphorylation (Fig.24), indicating that these cells are metabolically active and undergoing cellular differentiation [198, 264]. Strikingly, EIF2 signaling is very active in the activated cell states, as well is the up-regulation of ribosomal genes and N-Myc target genes, suggesting that cells re-enter the cell cycle (Fig.25) [198]. Comparison of the Krt8+ cells with AT1, AT2, activated AT2, club, and activated club cells in the Ingenuity upstream regulator analysis (IPA, Qiagen) revealed target genes significantly enriched for the TNF, TGFb, and NRF2 pathway, which indicates an active hypoxic response (Fig.25).

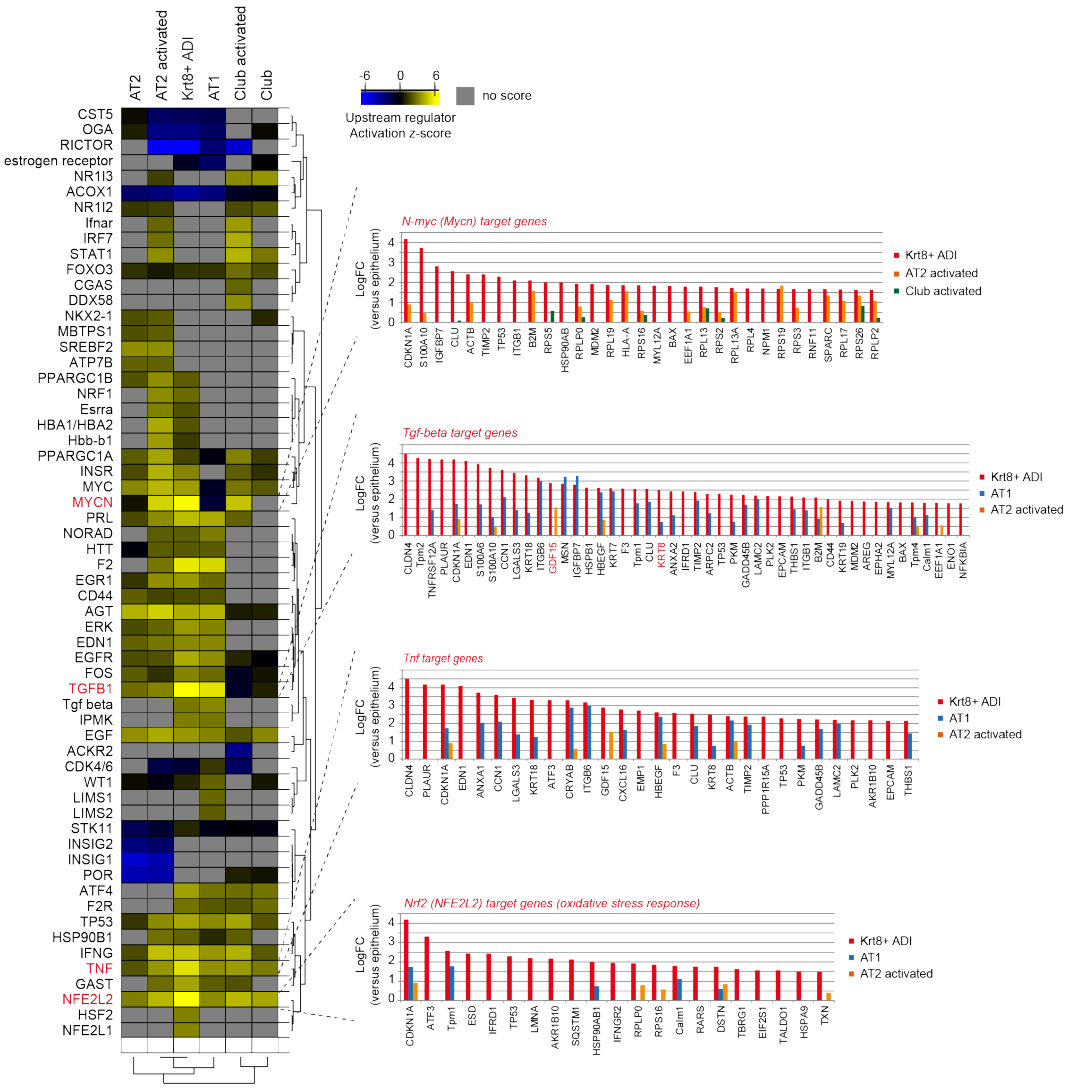

Figure 25: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Ingenuity upstream regulator analysis was used to score the activity of upstream regulators within the signatures of the indicated cell states. The activation z-scores were grouped by hierarchical clustering using their Pearson correlation. Bar graphs show target genes sorted by highest expression in Krt8+ cells relative to all other cells.

Moreover, in both the UMAP embedded data (Fig.21b) and the PAGA representation (Fig.22b), a proliferative cluster of Mki67+ and Top2a+ cells (denoted cluster 14 in the PAGA plot) became prominent, interestingly connecting activated AT2 cells and the MHC-II+ club cell subset [198]. Studying the relative cell frequency of cluster 14 showed a marked Mki67 expression around day 15 after bleomycin-mediated lung injury (Fig.26a). Cell cycle regression analysis within that cluster resulted in the determination of the cell type identities that all shared the proliferative gene signature (Fig.26b); after lung injury, Krt8+ cells, AT2 cells, club cells, and the MHC-II+ club cells all proliferated (Fig.26c) [198]. Staining of bleomycin-treated lung sections and control tissue against Ki67 confirmed the proliferative peak of cells around day 15 (Fig.26e), with a sudden inhibition of Ki67 expression at day 28 after injury [198]. Quantification by counting all Ki67+ cells within randomly chosen regions of interest (n=4-7 per time point) manifested a curve quite similar to the relative frequency curve (compare Fig.26d with Fig.26a) [198]. To assess the degree of proliferation among the Krt8+ cell population, the stained images were further used to specifically screen for Krt8/Ki67 double-positive cells at day 10 after lung injury (Fig.27). In fact, double-positive cells were detected, highlighted with yellow arrowheads, corroborating that Krt8+ cells are actively proliferating during the fibrogenic phase.

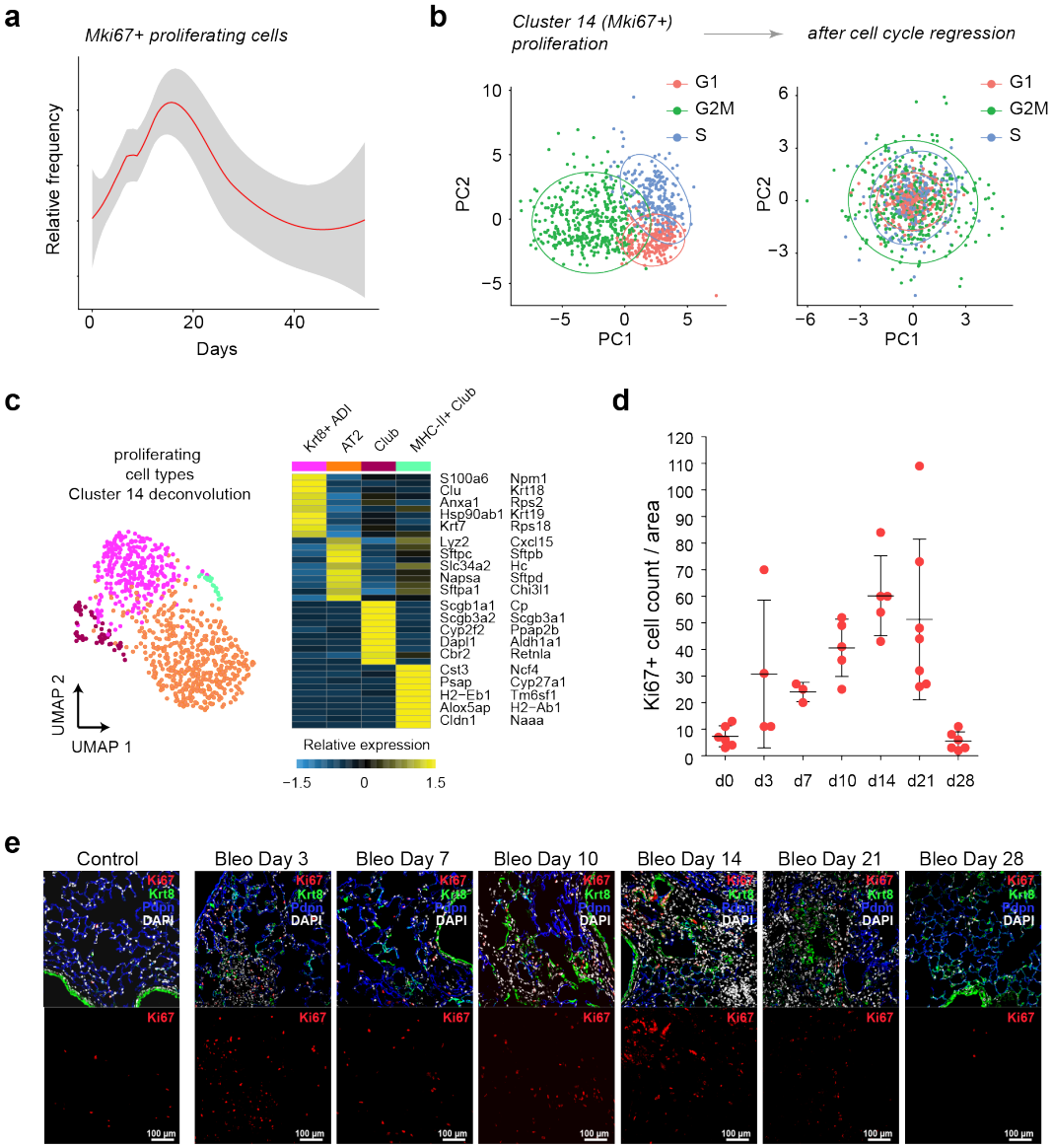

Figure 26: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Cell cycle analysis reveals proliferation of Krt8+ alveolar progenitor cells. **a** Relative frequency of Mki67+ proliferating cells plotted over time. **b** The scatter plots show cells from proliferating cell cluster 14 before and after cell cycle regression, colored by inferred cell cycle phase. Regression removes cell cycle effects from principal component data manifold. Re-analysis of cell cycle corrected expression deconvolves cell type identities of proliferating cells. **c** UMAP of cell cycle corrected cluster 14 cells visualizes four distinct clusters, which contain Krt8+ progenitors, AT2, club, and MHC-II+ club cells. The heatmap shows the average expression levels of selected marker genes. **d** Quantification of Ki67+ cells (cf. panel e) from immunostained FFPE sections demonstrates transiently increasing cellular proliferation of cells after bleomycin treatment. Randomly selected areas were analyzed for Dapi+/Ki67+ cells, n(areas)=3. **e** Immunostaining of FFPE lung tissue over time after bleomycin-induced injury. Scale bar indicates 100 microns.

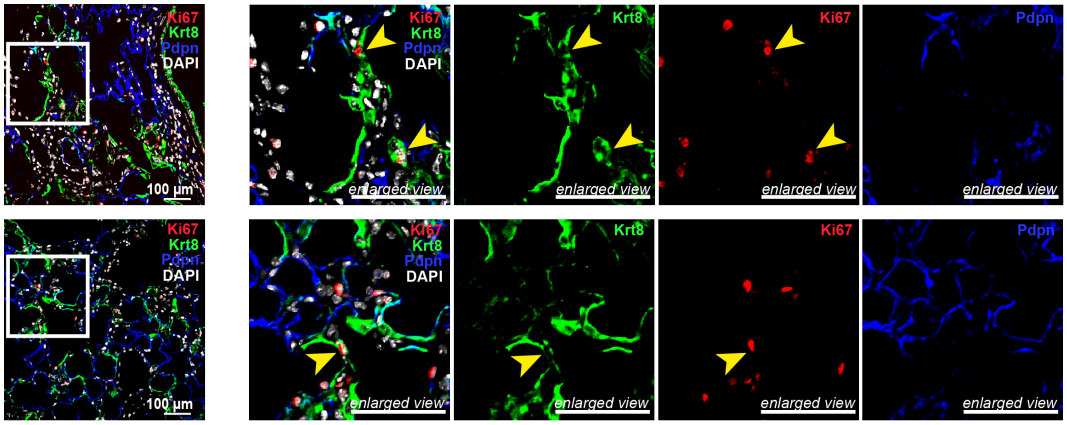

Figure 27: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Immunostaining of day 10 post bleomycin-injured lungs with enlarged views on proliferative Krt8+ alveolar progenitor cells (Ki67+/Krt8+), highlighted with yellow arrowheads. Scale bar indicates 100 microns.

3.1.6 Transcriptional convergence of AT2 and club cells towards Krt8+ cells

To comprehend transcriptional transdifferentiation dynamics towards the newly identified Krt8+ cell state, RNA velocities [223] were calculated for the skydive epithelial data set and overlaid onto the UMAP data representation (Fig.28a) [198]. Being consistent with published literature, the RNA velocity vectors pointed from the club cells towards the ciliated and goblet cells [38, 198]. Scrutinizing all the velocity vectors, the Krt8+ cell state seemed to be of dual origin via both the AT2 cells and club cells (Fig.28a, b). By calculating the terminal state likelihood using *scVelo* of only the AT2 cells, Krt8+ cells, and club cells, a differentiation axis of activated AT2 cells (cluster 11) and the MHC-II+ club cells (cluster 10) towards the Krt8+ cell state emerged (Fig.28c). The terminal state likelihood algorithm describes the terminally differentiated cells by analysis of their spliced and unspliced transcript ratio and a random walk bioinformatic analysis (see chapter 2.3.4 for details). Pseudotemporal trajectory analysis identified 3,036 genes with distinct expression along the calculated trajectories (Fig.28d) [198]. Cell identities are represented by their relative expression levels as smoothened lines along the pseudotime. The gene expression patterns along the converging differentiation trajectories based on the inferred likelihood of detection revealed 3,036 altered genes (Fig.29a). Both club cells and AT2 cells show distinct transcriptional changes towards the Krt8+ cell state, categorized into gene clusters A-F (Fig.29a). The most significant regulating transcription factors for each cluster are presented next to the heatmap (Fig.29a). Closely studying the transcription factors, there was a gradual decline in the expression of the homeobox protein Nkx-2.1 noticeable [198], a factor important in lung development and for the lung epithelial identity (Fig.29b) [198, 265]. Similarly, Foxp2 and Cebpa were found to be decreased during the conversion of club and AT2 cells to Krt8+

cells; Krt8+ cell state is represented by the dashed line as defined by the peak expression in Fig.28d. Foxp2 has been reported to be critical in repressing the differentiation of the lung epithelium [198, 266, 267], and Cepba to play important roles in lung development and cell maintenance of club and AT2 cells (Fig.29b) [198, 268–270]. Etv5 expression, a master transcription factor for AT2 cell maintenance [198, 271], was found to be drastically reduced, whereas the prime regulator for AT1 cell differentiation, Gata6, gradually increased during the differentiation towards the Krt8+ cell state [198, 272, 273]. Both the AT2 and club cell trajectories showed an overlap of genes that correspond to the Krt8+ cell state [198]. These genes include the p53 interactor Nupr1, which is induced by stress, as well as the key regulator for epithelial to mesenchymal transition Sox4, and genes involved in chromatin remodeling, such as the histone demethylase Kdm5c (Fig.29b) [198].

Based on computational analysis, the question remained open whether these predictions can be validated *in vivo*. With collaborative support, reporter mice for the AT2 cell lineage (Sftpc-Cre/ER), and later the club/airway cell lineage (Sox2-Cre/ER), were treated with bleomycin to confirm a dual origin for the Krt8+ cell state (cf. preprint - [198]). Ten days after treatment, they found high Krt8 expression in both Sftpc-Cre/ER-positive and -negative cells in stained sections of the lung, suggesting that indeed some Krt8+ cells might be airway-derived.
In a nutshell, the acquired data and analyses demonstrate a dual origin for Krt8+ alveolar epithelial cells of both airway-derived progenitors from the club cell lineage and AT2 cells [198]. Since both AT2 cells and club cells focalize on the same transcriptional state with peak expression of Krt8 during alveolar regeneration [198], and due to the observation that Krt8+ cells are transiently present in the course of lung injury and regeneration, the assumption was raised that Krt8+ cells can act as progenitor epithelial cells for AT1 cell maturation after lung injury, as drafted in a first hypothetical model (Fig.30). [198].

3.1.7 Krt8+ cells are alveolar progenitors that reconstitute the AT1 layer

In order to corroborate the mentioned hypothesis and to model potential cell trajectories, the Krt8+ and AT1 cells were submitted for a separate analysis [198]. A first idea of potential differentiation was given by the RNA velocity-based trajectory between Krt8+ cells and mature AT1 cells (Fig.31a). Making use of the RNA velocity and evaluating the proportion of spliced and unspliced reads, a steady transcriptional induction of Ager, the prime AT1 cell marker, could be observed around day 14 after bleomycin-mediated injury (Fig.31b) [198]. Comparing control (day 0, no bleomycin administration) and late time points (days 36 and 56) during the time course of regeneration with time points that cover the acute injury response

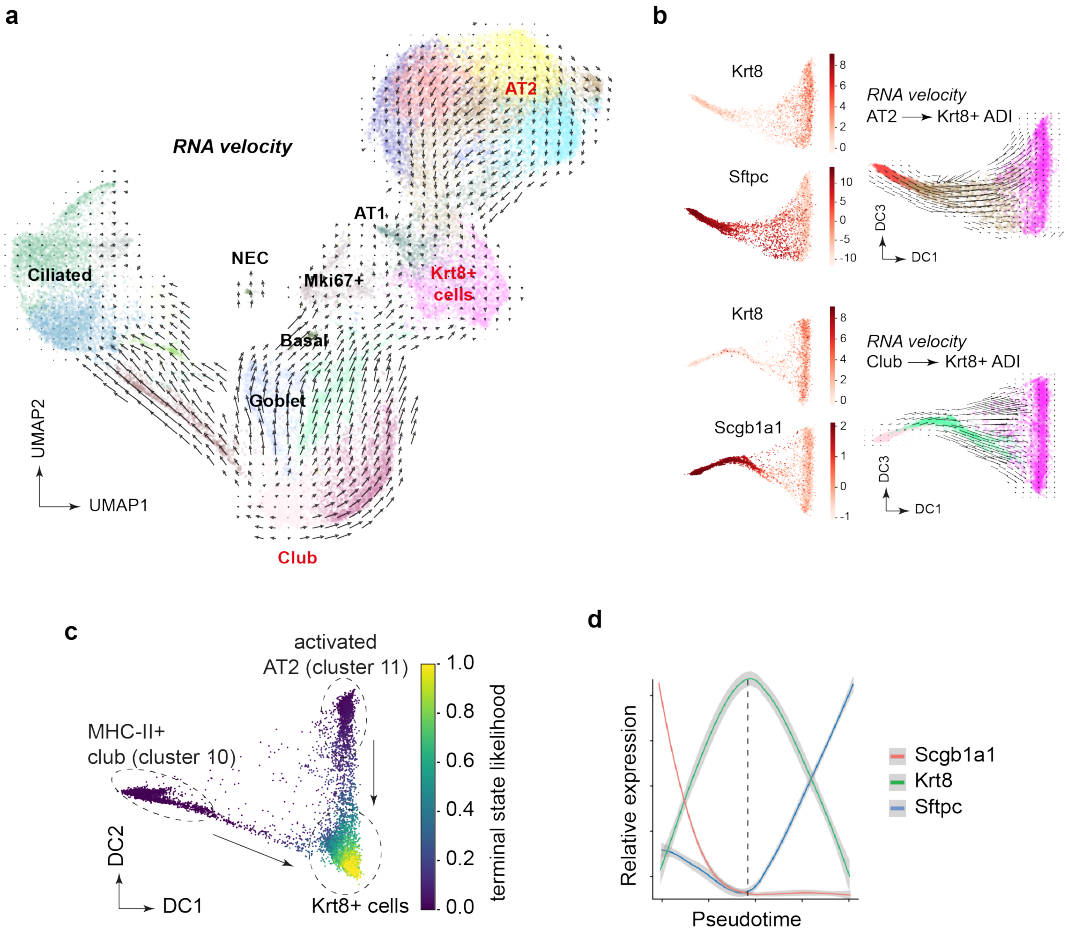

Figure 28: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Transcriptional convergence of club and AT2 cells onto the alveolar Krt8+ cell state. **a** Velocity plot displays the UMAP embedding colored by Louvain clusters with velocity information overlaid (arrows), indicating convergence of AT2 and club cells onto the alveolar Krt8+ cell state. **b** Velocity plots on AT2 and club cell subsets corroborate the transcriptional convergence towards the Krt8+ cell state. **c** Diffusion map of Louvain clusters 2, 10, and 11, colored by inferred terminal state likelihood, reveals two distinct transdifferentiation trajectories from activated AT2 and MHC-II+ club cells towards a Krt8+ cell state. **d** The lines illustrate smoothed expression levels of Scgb1a1, Krt8, and Sftpc across the trajectory, marking cell identities. The dashed vertical line indicates the peak of Krt8 expression. Grey colors represent the confidence interval derived from smoothing fit.

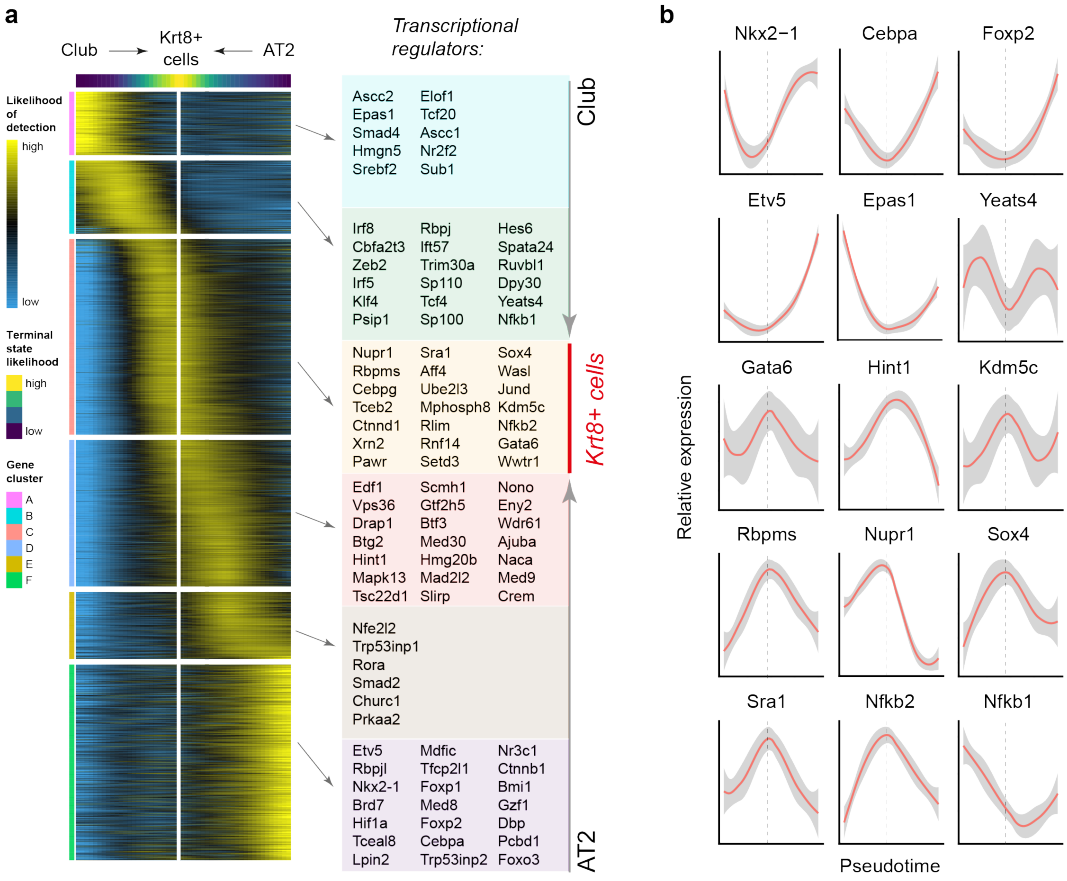

Figure 29: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
a The heatmap shows the gene expression patterns along the differentiation trajectory based on the inferred likelihood of detection for 3,036 altered genes. **b** Line plots show the smoothed relative expression levels of selected transcriptional regulators across the converging trajectories. The dashed vertical line indicates the peak of Krt8 expression. Grey colors represent the confidence interval derived from smoothing fit.

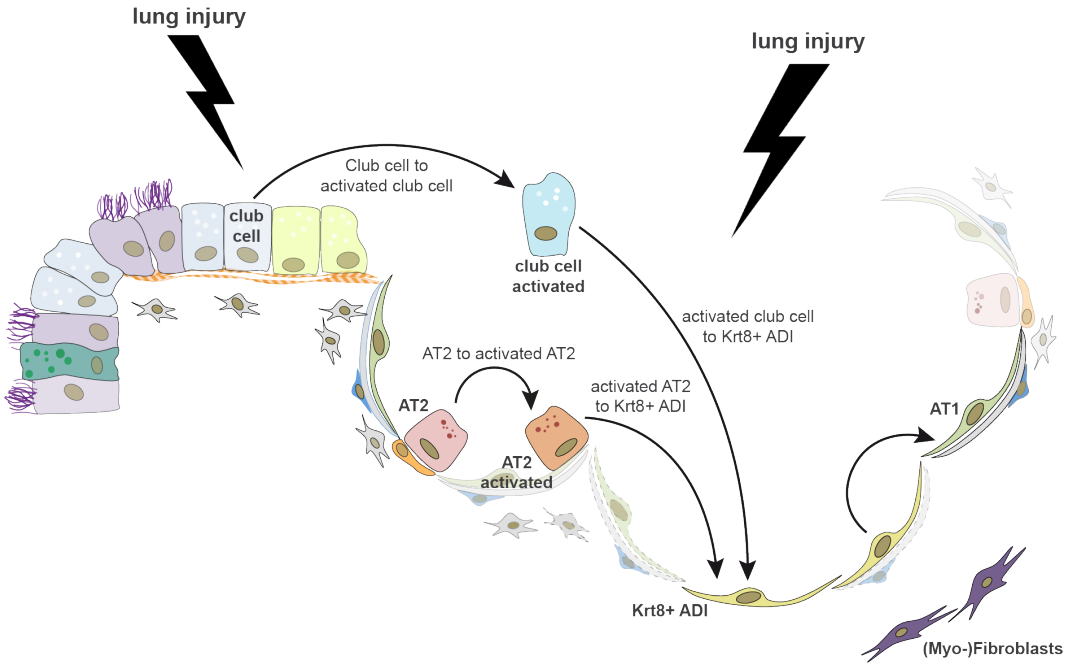

Figure 30: Novel proposed model for alveolar regeneration after bleomycin-mediated lung injury. Lineage tracing experiments together with scRNAseq experiments show that both airway progenitors and AT2 cells give rise to the same Krt8+ cell transcriptional state.

and the fibrogenic phase (days 2-28), a significantly lower ratio of unspliced over spliced Ager reads was detected (Fig.31c; Wilcoxon Rank Sum test, p<1e-46) [198]. The analysis of RNA velocity predicted an increasing expression of Ager along the cell trajectory towards mature AT1 cells (Fig.31d), which was corroborated by plotting the total expression levels of Ager onto the UMAP embedding (Fig.31e) [198].

In order to figure out which genes and transcriptional regulators drive these observed gene expression dynamics, a trajectory was calculated based on the pseudotime, which revealed 1,150 significantly regulated genes to be involved in the Krt8+ progenitor to AT1 cell transition (Fig.32) [198]. The pseudotime trajectory allowed for segregation of temporal changes during the transition from the Krt8+ progenitor cell towards the terminally differentiated AT1 cell [198]. The transdifferentiation trajectory could be distinguished into four phases, separated by distinct expression of transcriptional regulators driving different cellular responses or actions [198]. Determination of key genes was achieved by searching for the keywords cell adhesion, developmental protein, differentiation, ECM, growth factor, and keratin (Fig.32). Accordingly, pathways could have been assigned to each of the four phases, which showed that the initial phase during the transition from Krt8+ progenitor to

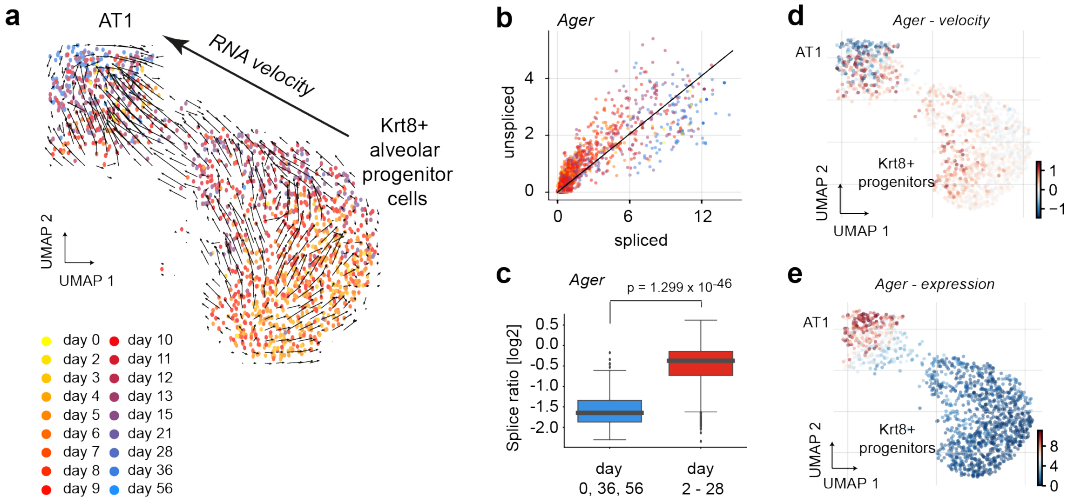

Figure 31: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Pseudotime analysis reveals Krt8+ cells as alveolar progenitors that give rise to AT1 cells. **a** Velocity plot displays the UMAP embedding colored by time point with velocity information overlaid (arrows), indicating terminal differentiation of Krt8+ progenitors into AT1 cells. **b** Velocity phase plot shows the number of spliced and unspliced reads of the AT1 marker Ager for each cell (points) on the X and Y axes, respectively. Cells are colored by time point and the black line represents the linear steady-state fit. Cells above and below the diagonal are predicted to be in inductive or repressive states, respectively. **c** The boxplot shows the log2 ratio of unspliced over spliced Ager reads for days 0, 36 and 56 (blue) and all other time points (red). To avoid division by zero, one was added to both counts. UMAP embedding colored by Ager velocity (**d**) and expression (**e**) displays a gradual increase along the inferred trajectory.

AT1 cell is featured by cell growth and cell cycle exit, as demonstrated, for instance, via MYC targets and oxidative phosphorylation (Fig.32) [198].

Towards the culmination of Krt8 expression along the differentiation trajectory, stress-related signaling, like the induction of p53 and the unfolded protein response pathways, join the transcriptional landscape, which is additionally noticed by the enhanced expression of transcription factors, such as Atf4 and Trp53inp1 (Fig.32, Fig.33) [198]. Moreover, non-canonical NFkB signaling seems to characterize the Krt8+ cell state as the expression of the EGFR-ligand amphiregulin (Areg), the nuclear factor NF-kappa-B p100 subunit (Nfkb2), as well as the Tweak receptor CD266 (Tnfrsf12a) were highly correlated with the Krt8 expression (Fig.33) [198]. Towards a terminal differentiation into an AT1 cell, the trajectory helped to distinguish a critical pre-AT1 phase, in which Krt8 expression was found down-regulated while concomitant, a genetic profile reminiscent of the epithelial-to-mesenchymal transition (EMT) was induced, e.g. by the key regulator Sox4 [198, 274]. An interesting finding in this pre-AT1 cell stage was the expression of TAZ (Wwtr1), which is a downstream regulator of the Hippo signaling pathway with crucial functions in AT2 cell differentiation after alveolar injury [198, 245]. As both the Hippo and Wnt pathways have been ascribed a role in mediating pulmonary disease, the expression of beta-catenin (Ctnnb1) is worthwhile to mention here. Beta-catenin is the key downstream signaling molecule in the canonical Wnt pathway [275] and peaked transiently in the pre-AT1 cell stage [198]. In the context of fibrogenesis, this

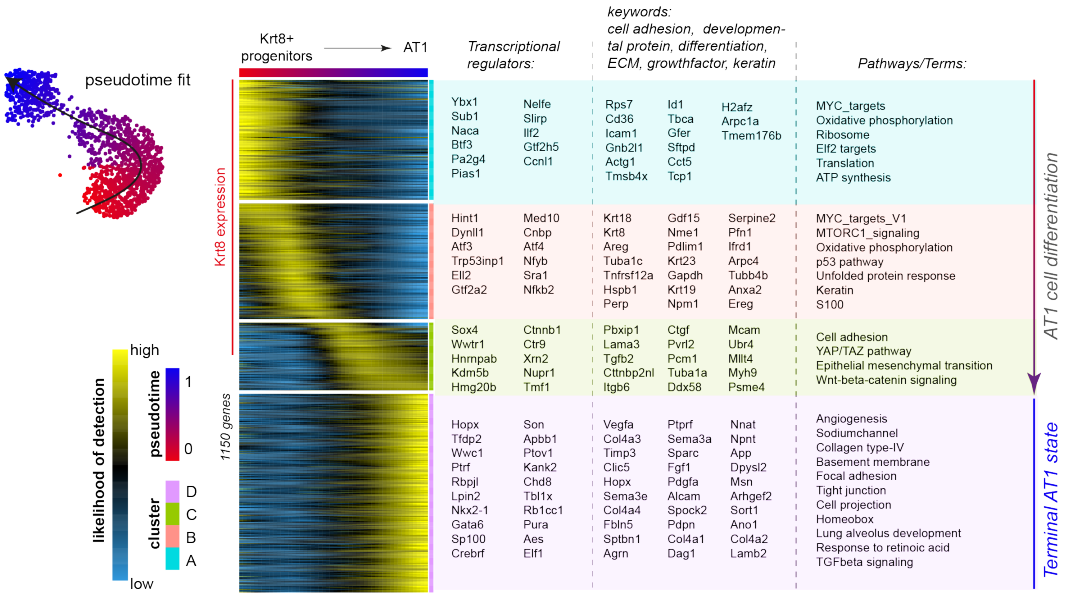

Figure 32: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
The heatmap shows the gene expression patterns across the differentiation trajectory for 1,150 altered genes. The analysis allowed for segregation into four groups based on temporal gene expression changes. Based on transcriptional regulators, cellular responses and pathways were predicted.

cell stage was further marked by the expression of many TGFb target genes, including Sox4, integrin beta-6 (Itgb6), and connective tissue growth factor (Ctgf) (Fig.33), suggesting a pivotal role for TGFb in alveolar regeneration [198, 246]. Considering the squamous cellular morphology of the Krt8+ cells (as was demonstrated in 3.1.4), the expression of the non-muscle myosin heavy chain IIa (Myh9) in the pre-AT1 cell stage confirmed that additional cytoskeletal rearrangements and increased cell contractility constitute final steps in the maturation of the Krt8+ progenitor cells towards AT1 cells [198]. Ultimately, the fully differentiated and terminated AT1 cell state showed expression of commonly valid AT1 cell markers and transcription factors, including Hopx, Gata6 and Wwc1, as well as ECM proteins and growth factors, such as Fgf1, Npnt and Agrn (Fig.33) [198].

Although having surveyed the genetic and transcriptional dynamics of the Krt8+ progenitor cell, the functional role of these cells in the context of alveolar regeneration remained still unclear. Considering the fact, that many ECM-related and EMT-related genes were found to be up-regulated during the Krt8-to-AT1 transdifferentiation, an important role for intercellular communication was raised for the Krt8+ alveolar epithelial progenitor [198]. Using the whole lung single cell RNA-seq data (cf. Fig.9), a cell communication network was established by pairing receptor-ligand interactions between cell types [198].
Here, most pairs were found in fibroblasts, macrophages, and (capillary) endothelial cells (Fig.34a) [198]. AT1 and capillary endothelial cells are in close contact due to their shared

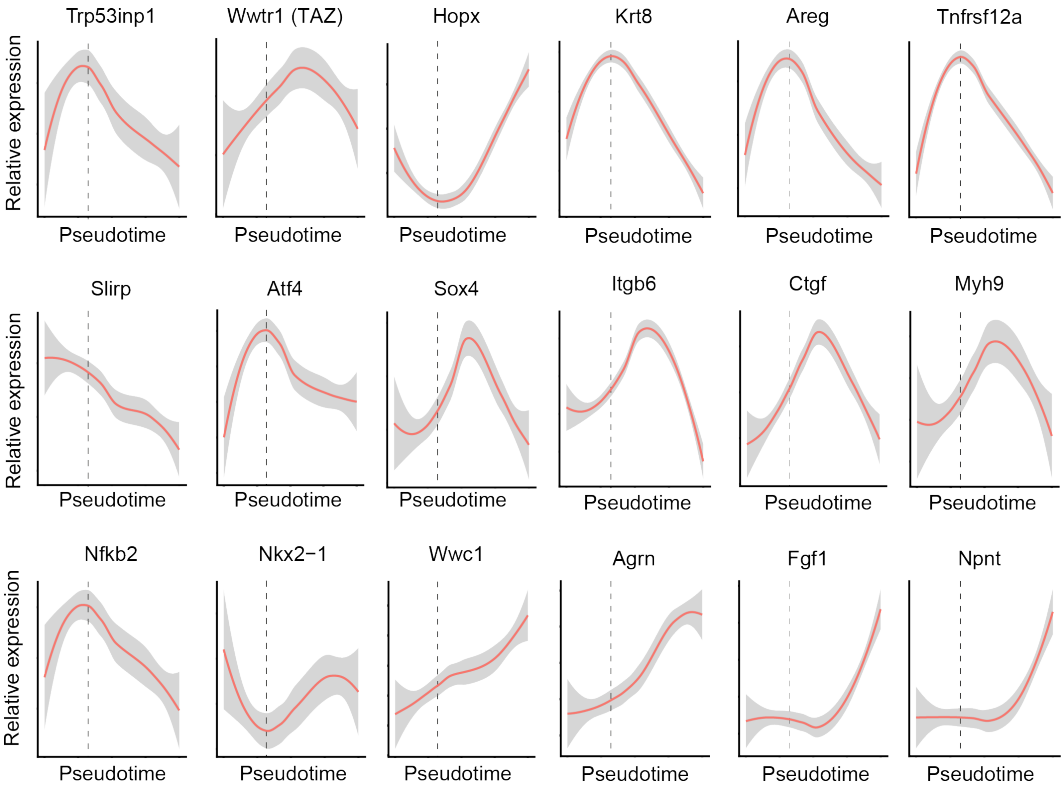

Figure 33: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
The line plots illustrate smoothed expression across the differentiation trajectory for a number of exemplary genes. Grey colors represent the confidence interval derived from the smoothing fit. The dotted line indicates the peak of Krt8 expression.

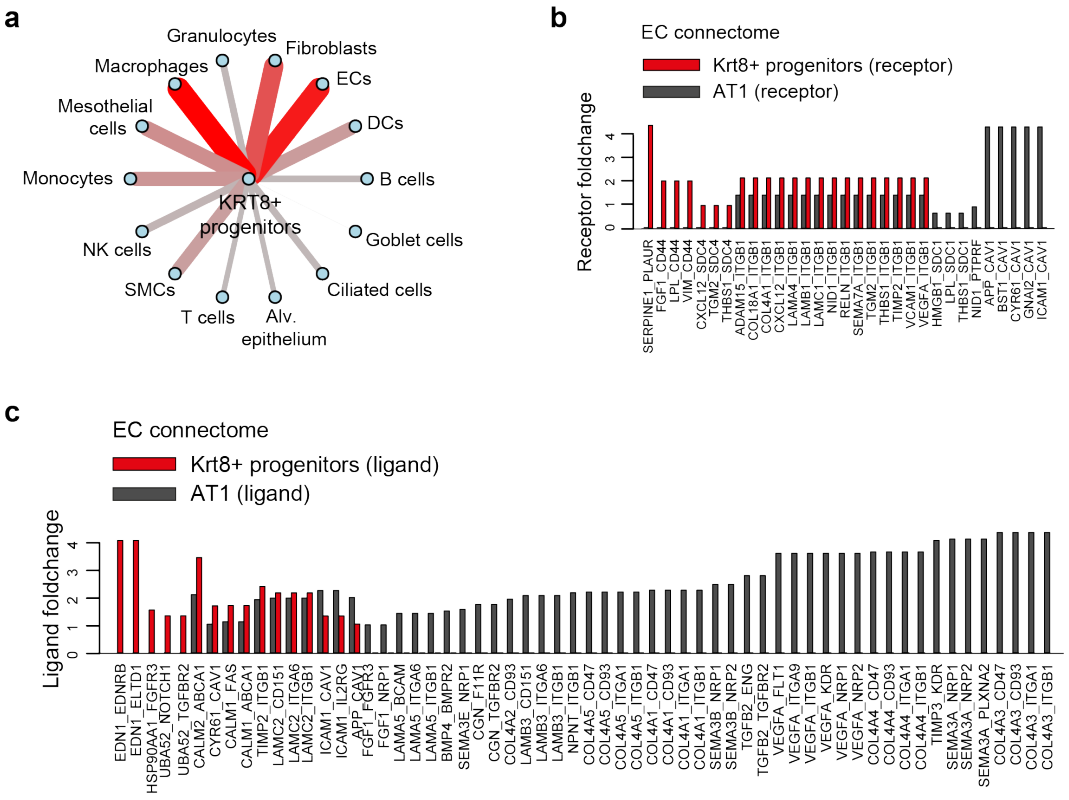

Figure 34: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
a The cell-cell communication network displays the number of receptor-ligand pairs between the molecular markers of the Krt8+ alveolar progenitor cell state and all other meta cell type identities. **b, c** The bar graphs show the average log2 fold change of either ligands (**b**) or receptors (**c**) within the endothelial cell (EC) connectome for Krt8+ alveolar progenitors and AT1 cells.

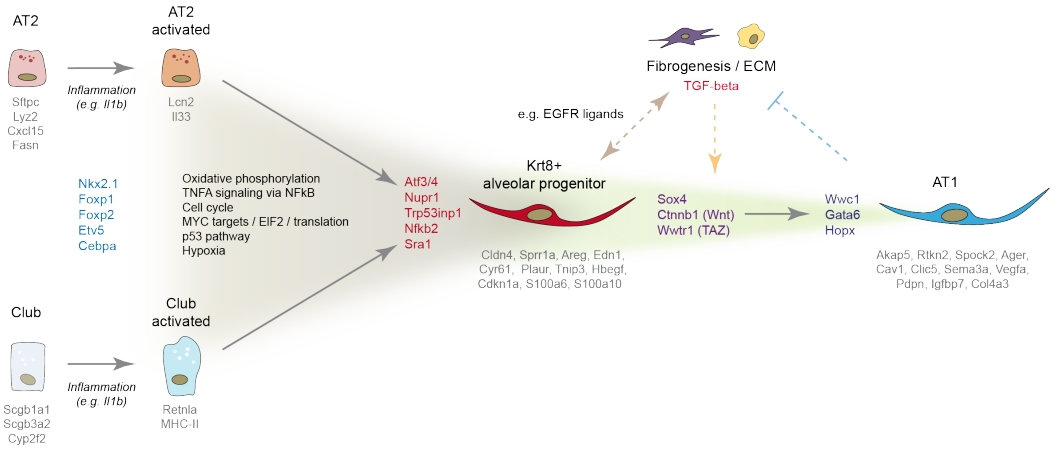

Figure 35: *Figure and figure legend: copied from the manuscript 'Strunz et al., 2019'*: [198]
Summary of observations and revised model of alveolar regeneration. AT2 and club cells enter an activated state characterized by a distinct set of pro-inflammatory genes as a result of inflammatory cytokine driven NFkB activation early after injury. Both club and AT2 cells can converge on the same Krt8+ alveolar progenitor cell state as a consequence of metabolic reprogramming and the local niche in injured alveoli. Activated club and AT2 cells lose cell identity genes such as Cebpa and Etv5 and re-enter the cell cycle, followed by a Myc-driven phase of cell growth and differentiation that is marked by increased oxidative phosphorylation and a drastic change in shape towards a squamous morphology. The Krt8+ alveolar progenitors are characterized by a stress response gene expression program involving p53, hypoxia response, and ER-stress related gene programs. Terminal differentiation into AT1 cells involves a transient peak of the transcriptional regulators Sox4, Ctnnb1 and Wwtr1, and their target genes. This indicates involvement of the Wnt, Yap/Taz, and TGF-beta signaling pathways in committing Krt8+ alveolar progenitors towards AT1 cell fate, which then becomes fixed by a peak in Hopx expression. Krt8+ alveolar progenitors peak in numbers during fibrogenesis and feature a highly distinct connectome of receptor-ligand pairs with endothelial cells, fibroblasts, and macrophages, suggesting that these cells serve to instruct the fibrogenic phase of lung regeneration. Conversely, it is likely that the fibrogenic niche, which as we show is largely driven by fibroblasts and macrophages, plays an instructive role in terminal AT1 differentiation. Finally, mature AT1 cells likely provide signals important for resolution of fibrogenesis.

functional role in facilitating gas exchange between inhaled air and blood [198]. Analysis of the connectome, which are all receptor-ligand interactions, of endothelial cells (EC) with both the Krt8+ progenitor and AT1 cells, revealed signal transduction between the endothelin-receptor (Ednrb) on EC via the ligand endothelin-1 (Edn1), which was exclusively expressed on the Krt8+ cells [198] (Fig.34c). In contrast, receptor-ligand pairs specific to only AT1 cells were also spotted, including for instance Vegfa-Flt1, Vegfa-Nrp1/2, or Sema3e-Nrp1/2 [198]. Similarly, differences observed for the receptors included the urokinase plasminogen activator receptor (Plaur), which is specifically expressed on the Krt8+ progenitor cells and binds to the EC-derived ligand plasminogen activator inhibitor-1 (PAI; Serpine1) (Fig.34b) [198].

Combined, this data suggests an important role for the newly-identified Krt8+ alveolar epithelial progenitor cell in mediating fibrogenesis and alveolar regeneration. There are distinct transcriptional programs following a temporal order, which has been deciphered in the presented analysis. Compliant with the data, a hypothetical model for regeneration

including all transcriptional and genetic regulators and pathways was updated with the molecular data and is summarized in Fig.35.

3.1.8 Discussion

The presented work describes the dynamics of mouse lung regeneration at single cell resolution and the discovery of an alveolar progenitor cell state that gives rise to AT1 cells [198]. This novel cell state was found to highly express Krt8 and only appeared transiently but with highest expression during the fibrogenic phase. After having surveyed all cell populations of the murine lung, spanning the six commonly studied time points of the bleomycin lung injury model, a special focus was set on EpCam+ epithelial cells over 18 time points, thereby increasing the longitudinal resolution [198]. Deploying the potential of pseudotemporal modeling [198, 224, 247, 276, 277], epithelial transdifferentiation was studied based on gene regulation, which enabled temporal gene expression programs and main transcriptional regulators [198]. Besides cell fate and trajectory modeling, receptor-ligand networks helped to unravel potential cell-cell communications and their temporal dynamics [198].

Combining high-throughput single cell sequencing with advanced computational methods allowed for unprecedented insights into the dynamics of lung regeneration and lineage reconstruction, which complements traditional *in vivo* lineage tracing experiments. Transdifferentiation processes are complex and frequent, and involve intermediate gene expression states with distinct genotypic differences between initial and terminal cell state [198]. Traditional views on tissue regeneration are mainly focused on stem cell differentiation with gradual loss of stem cell-specific genes towards terminally matured tissue cells. Many of such stem cell subtypes and even niches have been described in the lung. For the regeneration of the alveolar compartment, different mechanisms have been discovered, often in dependence of the severity of injury [234]. Bronchioalveolar stem cells (BASC) and p63+/Krt5+ distal airway stem cells (DASC) have been proposed as the key drivers for alveolar regeneration [198]. Although a major contribution of Krt5+ cells can be neglected based on the herein presented data, there was clear evidence for airway-assisted alveolar regeneration by Krt8+ cells (as seen in the Sox2-Cre/ER experiments) [198]. Activation of club cells after injury, in particular the MHC-II subset, as well as the activation of AT2 cells led to the emergence of cellular intermediate states, marked by high Krt8 expression. Despite such convincing data, an involvement of BASCs cannot be ruled out [198]. Due to the observation that two distinct lineages (club cell- and AT2-derived) converged in the same Krt8+ transcriptional state implies that not the lineages themselves but the injured alveolar niche determines the (progenitor) cell identities [198].

The pathways that were carved out in this work can be associated with injury-induced stress- and inflammation responses and allow for drawing the conclusion that environmental alterations might evoke the emergence of the Krt8+ alveolar progenitor [198]. Consistent with the presented data, independent work demonstrated a loss in AT2 cell proliferation after injury in AT2-specific IL1-receptor knock-out mice via IL1b and NFkB activation [244], providing a molecular tie between inflammation and epithelial regeneration [198]. Environmental and inflammatory stimuli might promote cell plasticity by activation of alternative cell fate programs [198]. Cultivation of AT2 cells *in vitro* may lead to spontaneous AT1 cell drifts, indicative for the necessity to receive niche-derived signals for proper AT2 cell maintenance [198]. A five-day *in vitro* AT2 to AT1 transdifferentiation experiment revealed highest Krt8 protein levels at day 3, with a peak expression of Pdpn (AT1 cell marker) at day 5 [198, 278].

The Krt8+ alveolar progenitor has been introduced here as an important intermediate cell state during lung regeneration. However, although very low in number, there were cells with high Krt8 expression in parenchymal lung tissue of control mice. The function of these rare cells remain obscure and has only been described as an observation in this present work. Neither club cell-, AT2-, nor AT1 cell marker genes were found in this novel Krt8+ cell state, suggesting that it truly represents a distinct reprogrammed cell type and progenitor role [198]. There is no evidence that could disprove a potential AT2 reconstitution by the Krt8+ alveolar progenitor, however, the data strongly suggests a compelling propensity towards a final AT1 cell differentiation [198].

Terminal AT1 cell differentiation during development has been shown to occur via a non-proliferative two-step process of cell flattening and cell folding [60, 198]. By morphology, the Krt8+ progenitor cell appears in squamous shape, which is congruent with the cell flattening. For proper regeneration, the alveolar architecture needs to be re-established, which, on the molecular level, is guided by morphogenic factors, such as Vegfa and semaphorins which coordinate angiogenesis and alveolar morphogenesis [60, 198]. The connectome analysis presented here demonstrated the specific expression of these morphogens in only the mature AT1 cells after injury. By contrast, the Krt8+ progenitor cells expressed Endothelin-1 (End1) for the stimulation of capillary endothelial cells [198].

Transcriptional regulators, including TAZ (Wwtr1), Sox4, and beta-catenin (Ctnnb1) constitute key switch points in terminal AT1 differentiation [198]. Using small molecule inhibitors *in vivo* and conditional knockout mouse models, Wwtr1 (involved in the YAP/TAZ-Hippo pathway) was recently shown to be of high mechanistic relevance in the AT2 to AT1 transdifferentiation [198, 245, 279]. Additionally, beta-catenin and the canonical Wnt signaling pathways are crucial factors for AT2 in vitro differentiation [198, 278]. The data suggested that cell fate-deciding switches towards the AT1 phenotype are operated by regulators, such as Sox4, but this awaits further experimentation [198]. Generally, the contribution and crosstalk of all the involved pathways of TGFb, Wnt, or Hippo in shaping the transdifferentiation from

Krt8+ alveolar progenitor towards the AT1 cell is fairly obscure and needs more in-depth research [198].
Nevertheless, there is a complex coordination of cell types during fibrogenesis, which includes active or inflammatory (monocyte-derived) macrophage populations and fibroblasts that can transform to myofibroblasts for exacerbated expression of extracellular matrix proteins. Microenvironmental stimuli that drive communications, such as intercellular contacts, or the cell to cell communication remain loosely defined. During tissue regeneration, the matrix takes over a master regulator function to ensure proper healing. This includes the polarization of macrophages into phenotypes promoting fibrogenesis and, at a later stage, promoting scar resolution [198, 256, 257]. By the interaction with fibroblasts, also monocyte-derived macrophages can contribute to a pro-fibrotic processes [198, 231, 237, 258]. Further research is necessary to better define the local niche signals that drive fibrogenesis and a proper resolution towards tissue regeneration.

3.2 Effects of Emilin-2 in experimental pulmonary fibrosis

The work presented is this chapter summarizes the findings up to the completion date of this thesis. This project is currently ongoing to discern the molecular effects of Emilin-2 in the context of lung injury and repair.

3.2.1 Aims and hypothesis

This side project is anchored in previously published findings by Schiller et al. [139] who studied lung remodeling using the bleomycin model with a proteomics approach. The data revealed among others the peak expression of the extracellular matrix protein Emilin-2 at day 14 of the model. In a follow up approach, it was aimed at understanding the effects of Emilin-2 after lung injury using a knockout mouse line. As the peak expression was found to be highest at the inflammatory and fibrogenic phase, the hypothesis was drafted that there is a more severe phenotype during fibrogenesis when Emilin-2 is absent/knocked out. Two central questions initiated this project: (i) which cells express the protein Emilin-2, and (ii) is there a specific phenotype after lung injury in knockout animals?

3.2.2 Introduction

The lung can be injured in various ways. Upon injury, a complex cascade is initiated leading to local inflammation, which causes additional alterations of tissue homeostasis, as well as cellular and extracellular changes on behalf of a timely response to injury.

Local alveolar injury involves the immediate recruitment of neutrophils [280], concomitant to the development of alveolar edema and the deposition of collagen fibers [281–283]. Due to the neutrophil-derived proteases, necrotic cells are detached from the basement membrane [284], which requires macrophage populations at the sites of injury for dead cell removal. Collagen fibers and other ECM constituents, mostly expressed by fibroblasts, are assembled to facilitate cellular migration at the lesion site. Under regular condition, these fibers, or the scar, respectively, is ultimately removed for complete tissue regeneration. If this delicately orchestrated response is impaired, however, the excessive expression of ECM components can lead to exuberant scarring of the tissue and to tissue fibrosis [59, 285–288], which greatly hampers the ability for proper gas exchange in the body.

Although many regulatory processes have been described, the convoluted mechanisms that involve immune cells, their mediators, and mesenchymal populations during inflammation and regeneration are insufficiently understood [289, 290]. Upon injury and during wound healing, the ECM undergoes massive rearrangement that has effects on cell growth and cell behavior, cellular differentiation, and cell viability [291–294]. Moreover, the ECM plays an important role in regulating signals derived from developmental signaling pathways that are active during tissue repair, including the TGFb, Wnt, sonic hedgehog (Shh), or bone morphogenic protein (Bmp) pathways [139, 295].

Using the bleomycin-induced lung injury mouse model and mass spectrometry-driven tissue proteomics, *Schiller et al.* identified key regulator proteins during tissue injury and regeneration [139]. In their work, a novel quantitative detergent solubility profiling (QDSP) method was used to comprehensively chart temporal events in ECM changes after injury [139]. Among others, the highly-regulated proteins Emilin-2 and Collagen-XXVIII emerged as provisional ECM components for tissue repair [139]. Protein expression of Emilin-2 was found highest during the peak inflammatory phase at day 14 [139]. While Emilin-2 expression was found to be restricted to perivascular and peribronchiolar location in healthy lungs, the bleomycin-treated mice also revealed Emilin-2 expression in the alveolar regions, especially in areas where alpha-SMA-positive myofibroblasts were found to be accumulated [139].

Emilin-2 belongs to the Emilin/Multimerin family [296, 297] and has been named after its biochemical properties: elastin microfibril interface located protein [297]. Emilin-2 is an elastic fiber-associated glycoprotein with a C1q domain at the C-terminus and a cysteine-rich domain at the N-terminus, termed *EMI* after the Emilin family [297]. Other protein domains include a short collagenous region and a coiled-coil region [297]. The 40 kb large protein is encoded on chromosome 18p11.3 with the gene spanning eight exons and seven introns [296] and has experimentally found to have highest expression in the fetal heart and the adult lung [297]. Emilin-2 is predominantly expressed in diffuse meshwork

pattern in the extracellular space [296, 297]. The different domains have been assigned to different protein functions, including apoptosis, tumor growth, and angiogenesis: Emilin-2 activates the extrinsic apoptotic pathway by binding to the death receptors 4 and 5 and can thus mimic death receptor ligand activity [296, 298]. Creating tumors in Emilin-2 null mice corroborated the pro-apoptotic function of Emilin-2 as tumors were found at reduced sizes as compared to wildtype mice [298, 299]. Interestingly, Emilin-2 increased the overall vessel density in tumors, additionally giving proof of the pro-angiogenic functions for Emilin-2 [296, 300]. In a different experiment, the promotion of angiogenesis by Emilin-2 was demonstrated by direct binding of the epidermal growth factor receptor (EGFR), which can cause augmentation of IL-8 levels and hence, the proliferation and migration of vascular endothelial cells [300]. Emilin-2 was shown to regulate angiogenesis by IL-8 expression through the EGFR/EGF pathway [300]. Apart from these functions, Emilin-2 has additionally described to associate with elastic fibers and the lung interstitium [297], and to bind to Wnt ligands in the ECM [301]. The data generated in these studies suggest a crucial role for Emilin-2 in regulating extracellular growth factors from the Bmp/TGFb and Wnt signaling pathway [139, 302]. For the study of functional effects *in vivo*, a constitutive knockout of Emilin-2 in C57Bl/6 mice was created [300].

3.2.3 Characteristic features of the injured Emilin-2 knockout lung

Knowing that Emilin-2 is highly up-regulated at day 14 in the bleomycin-injured lung, the aim was to assess the functional role of the protein in the context of lung injury and regeneration. A straightforward approach was the usage of an *in vivo* knockout mouse model, which for the case of Emilin-2, was kindly provided by Dr. Paolo Bonaldo at the University of Padova in Italy. To generate a constitutive knockout of Emilin-2 in C57Bl6/J mice, exon 1 of the Emilin-2 gene was replaced by a neoR cassette (further details are described in the cited paper) [300]. To study the effects of Emilin-2 under the bleomycin toxicity, in total, almost 200 mice were studied over the time points of the model until day 28. During each experimental round, mice were regularly weighed for loss of body weight, which could not exceed a loss of more than 20% of the initial body weight (in accordance with the animal welfare protection requirements). Having four groups in each cohort (wildtype/Emilin-2 knockout with PBS/bleomycin), the weight curves could be tracked over time (Fig.36a). After treatment, independent of the assigned group (PBS or bleomycin), a loss of weight was usually noticeable due to the acute burden of the instillation. While PBS-treated animals gained back to regular weights within the first 3-6 days, bleomycin-treated animals discontinued their food intake due to the malady induced by the treatment. Weight loss measurement can be a rough indicator for the successful administration of bleomycin, for the disease progression, and at times even for the severity of injury. Studying the weight loss curves, there was no difference between the two PBS-treated groups. While the control mice even exceeded their initial weight over

time, the bleomycin-treated animals lost weight rapidly during the first days and failed to uptake enough food during the inflammatory phase until day 14 (red and green curves in Fig.36a). Despite gross variances over time, the knockout bleomycin group only resumed almost initial weights after 3 weeks post injury, while in contrast, their wildtype litter mates made up for the weights of the control mice. Only in the late days of week four post-injury, the bleomycin-treated knockout mice reached to similar weights as their wildtype litter mates. In a different approach, some animals could be subjected to lung function measurement at day 10 and day 14 after bleomycin administration (Fig.36b). Pulmonary compliance measures the lung's distensibility and confirmed the effect of bleomycin in both the wildtype and Emilin-2 knockout mice. As was expected, the lungs were even less compliant at day 14, however, no differences could be determined between the bleomycin-treated groups.

With unobtrusive results of the lung function measurement, tissue was collected and processed for histologic assessment using Masson's Trichrome/Anilin-blue staining for the detection of collagen content in the sections (Fig.37). Comparing day 3, resembling the initial inflammatory time point, with later time points of day 10, 14, 21, and 28, the bleomycin-induced changes in the lung became very prominent. PBS-treated mice generally did not feature obvious relevant morphologic alterations. Although clear spaces around perivascular and peribronchiolar regions were detected, which is presumably attributable to installation and fixation of the tissue, these did not differ between the PBS groups (Fig.37a). Bleomycin lung sections, especially at day 10 and later, exhibited interstitial pneumonia, however, a wide inter-individuality was observed with varying degree of severity. This highly inter-individual effect is likely the reason for non-significant observations during the quantification of collagen content in these sections (Fig.37a and Fig.39). Tissue destruction was mainly characterized by perivascular and perbronchiolar edema, e.g. in WT bleomycin mice numbers 56 or 3-91, by vascular congestion, e.g. in KO bleomycin mouse number 9-75 or WT bleomycin mouse number 3-91, and inflammatory cell infiltration (cf. Fig.38a highest magnification of the bleomycin section). Inflammatory cells and fibroblasts extended into the pulmonary parenchyma where alveolar expansion and the presence of alveolar macrophages was noticed (Fig.38a). The bleomycin-treated animals showed an apparent overall progression of lesions from day 3 (generally more acute inflammatory changes) to day 28 (inflammatory lesions with more accentuated interstitial fibrosis), however again, considerable inter-individual variation was discerned in mice of the same necropsy dates. Generally, the bleomycin treatment had no clearly discernible or consistent differences between the two genetic groups (wildtype and Emilin-2 knockout), which is consistent with the quantification results (Fig.39). In order to quantify the degree of fibrosis and the tissue density, scanned sections were adjusted by increasing the background contrast, followed by segmentation into areas excluded for the analysis (i.e. main vessels), and areas that made up the tissue density (Fig.38b). First, PBS-treated lungs were analyzed to allow for the differentiation between

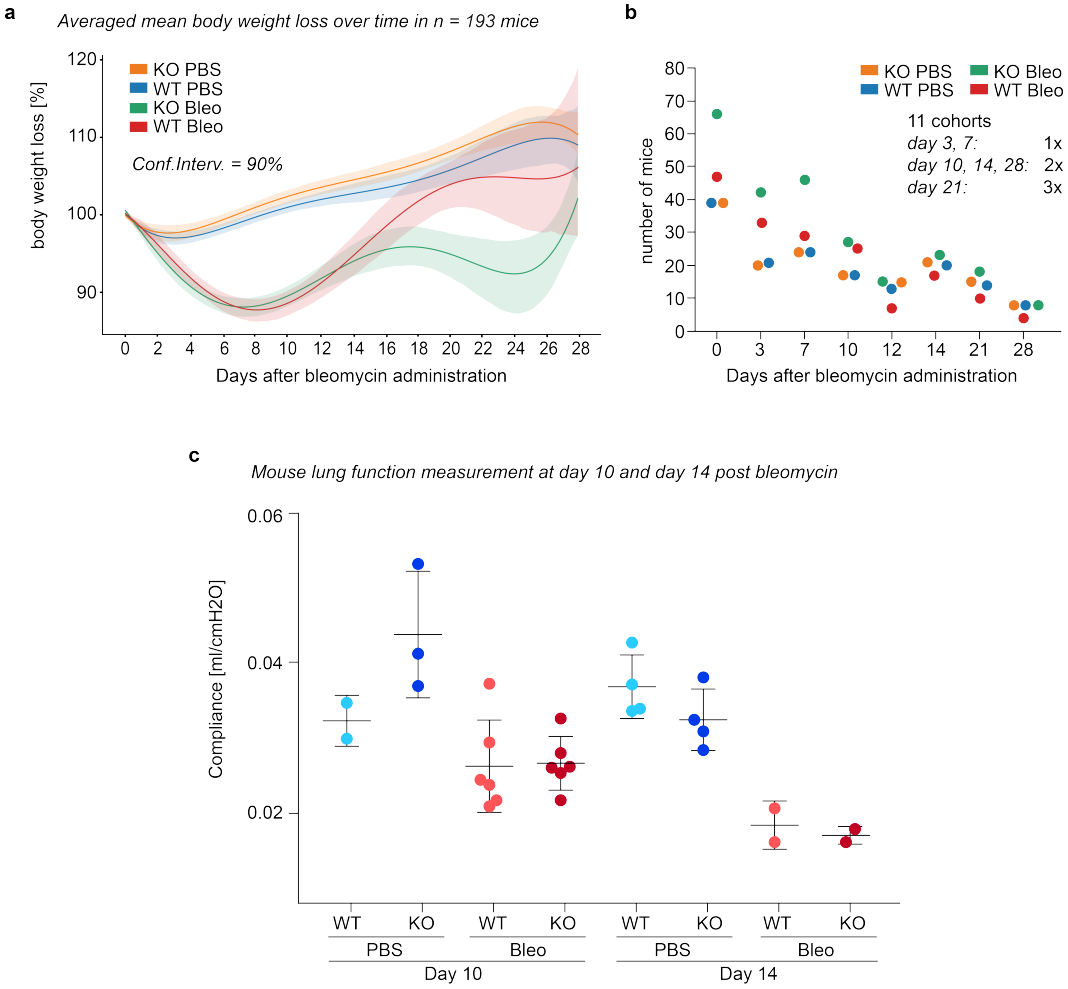

Figure 36: **a** Emilin-2 knockout mice recover more slowly after bleomycin administration. Mice from 11 instillation cohorts (n = 193 mice) were observed with regard to average mean body weight loss in their respective group (PBS wildtype or knockout, Bleomycin wildtype or knockout). Bleomycin-treated animals lose body weight in a similar manner until day 14; thereafter, wildtype mice re-gain body weight faster than knockout animals. The lines represent smoothed body weight loss over the time course. Confidence interval (CI = 90%) derived from the smoothing fit is included. **b** Distribution of numbers of mice used in the different experiments. Data was collected from 193 mice from 11 instillation cohorts. Each one cohort was sacrificed at day 3 and 7, two cohorts at day 10, 14, and 28, and 3 cohorts at day 21. Mice that died during the experiment were removed from the analysis. Cohorts of later time points generally include less mice explaining the wide confidence interval ranges observed in **a**. Statistical tests between the different conditions were not significant at day 14 to day 28. **c** Lung function measurement in the mouse lung reveals no differences between wildtype and Emilin-2 knockout animals. Bleomycin-treated test mice show expectedly less compliance of the lung, which is increasing at day 14.

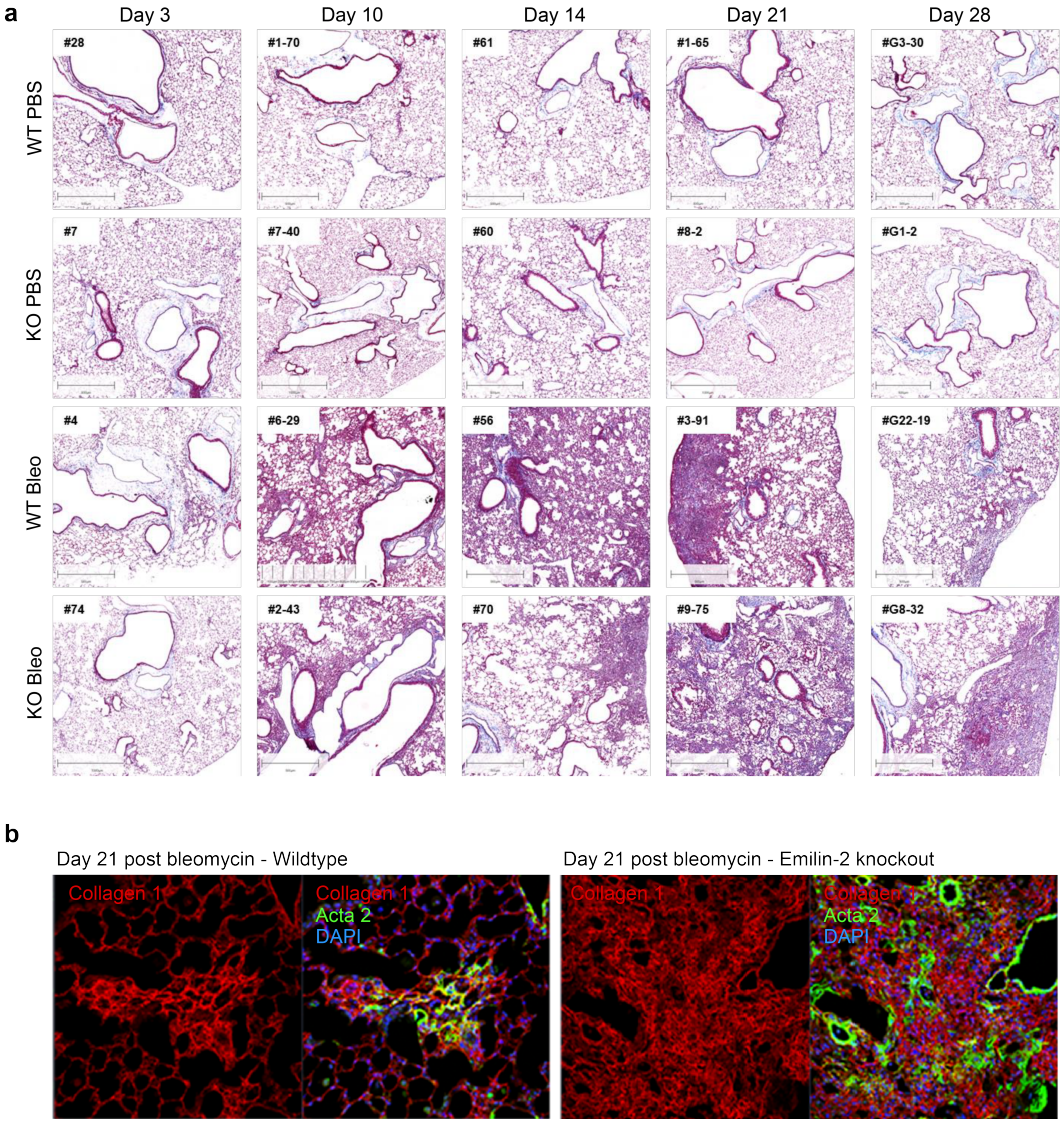

Figure 37: Histologic analysis using Masson's Trichrome staining shows emerging interstitial fibrosis over time. **a** PBS-treated mice have no histologic signs for tissue abnormalities independent of the genetic background. Bleomycin mice show typical features of the bleomycin lung with alveolar wall thickening, excessive ECM production, especially after day 14. At day 28 after bleomycin administration, Emilin-2 knockout animals tend to resolve less rapidly the fibrotic areas with anilin-blue positive Collagen staining, suggestive by histology that there might be a delayed response to injury when Emilin-2 is absent. Shown are representative sections; the scale bars indicate 500 microns. **b** Immunofluorescence staining of wildtype and Emilin-2 knockout bleomycin-treated lung sections reveal increased density of Collagen 1 fibers and increased deposition of alpha-SMA (Acta2) protein in the knockout lungs.

peribronchiolar or perivascular positive Anilin-blue staining versus the interstitial staining. Bleomycin-specific effects, i.e. interstitial fibrosis by mature collagen fiber staining with Anilin-blue (cf. Fig.38a) were then quantified by division of the Anilin-blue positive space by the entire tissue space without whitespace areas (cf. calculation formula in Fig.38b). For the quantification, 3-5 fields of view/regions of interest (ROI), depending on the lobe size, were assigned to the scanned lung sections in an unbiased manner, however, the ROI needed to lay within tissue space only. For the mice analyzed, at day 10, there were only few ROIs detected in the wildtype cohort as compared to the Emilin-2 knockout mice (Fig.39a; each dot represents a ROI with detectable Anilin-blue+ interstitial fibrosis). This trend was generally valid for all investigated time points, suggesting that there may be more mature Collagen 1 content in knockout lungs, in particular around day 21 and day 28. Compared to their wildtype littermates, however, no significant changes were observed (Fig.39c). Immunofluorescence staining of bleomycin-treated lungs after 21 days showed a massive increase in collagen 1-positive signal, together with the expression of Emilin-2 in the knockout mice (Fig.37b). These contradicting results evoked a critical assessment of the Anilin-blue staining of affected lung tissue. A different approach of studying the degree of tissue abnormality is using the tissue density as calculated and described in Fig.38b. All ROIs of all time time points could be integrated into the analysis. The results demonstrated that there were no noticeable differences between the PBS-treated cohorts (Fig.39b). After lung injury, wildtype mice revealed tissue density increases up to day 14, followed by a slow decline towards regeneration (day 28). Prominent changes were measured only between day 10 and day 14. By contrast, the Emilin-2 knockout cohort demonstrated massive increase in tissue density at day 21 with significant changes compared to day 10 and day 14 (Fig.39b). It was surprising to see that Emilin-2 knockout mice revealed a highly significant reduction in tissue density within one week of time (changes between day 21 and day 28; Fig.39b). Juxtaposition of both genotypes after bleomycin administration over time showed significant differences only at day 14 and day 21 (Fig.39d).

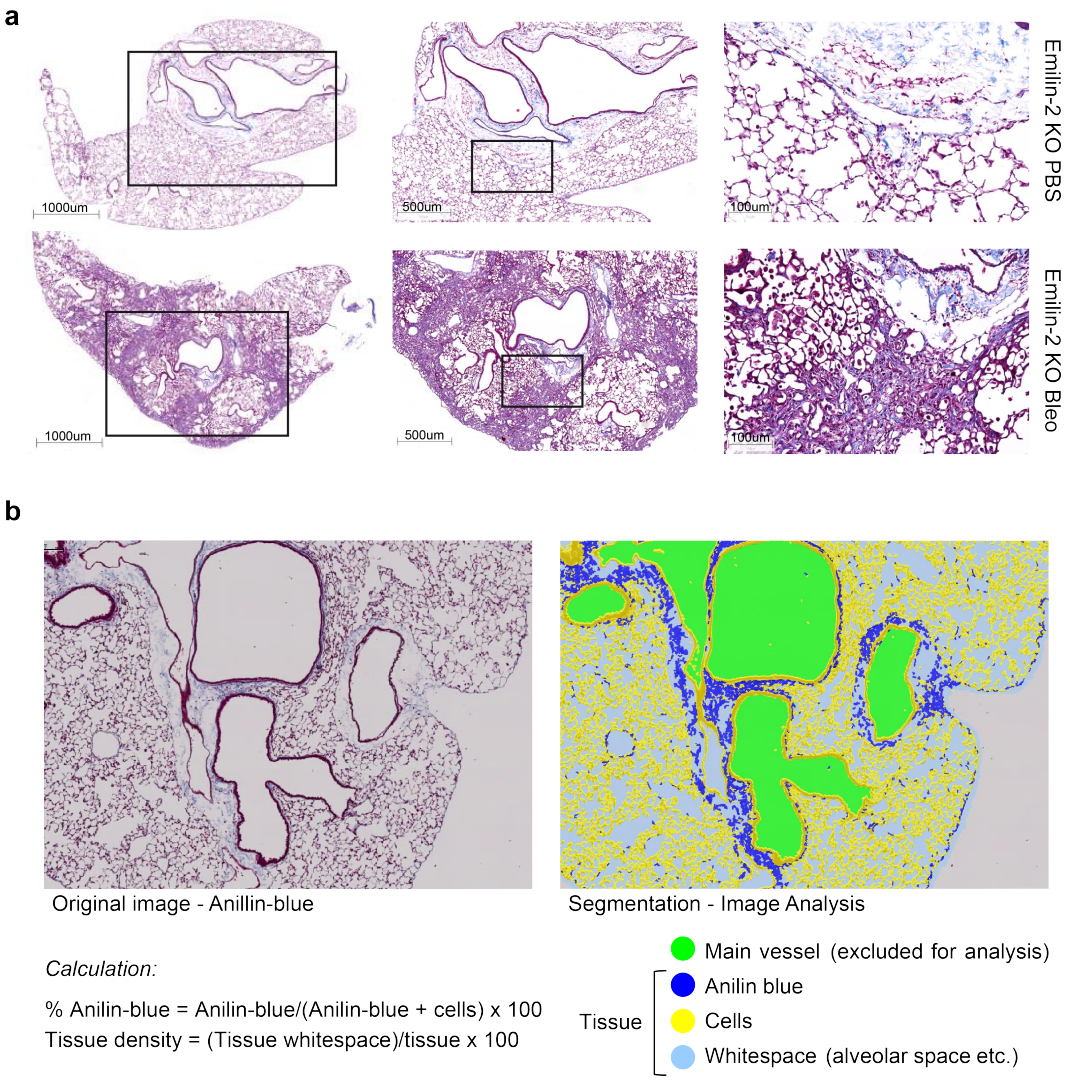

Figure 38: Estimating the degree of fibrosis in Emilin-2 knockout animals after bleomycin-mediated lung injury. **a** Scanned Masson's Trichrome stained sections were analyzed by closely studying histologic features at enlarged view. To differentiate between peribronchiolar or perivascular positive staining versus interstitial staining, Emilin-2 knockout PBS-treated lungs were compared to bleomycin lungs for the detection of true interstitial fibrosis. In bleomycin sections, the infiltration of immune cells into the alveolar space and tissue thickening becomes prominent (cf. enlarged view - Bleo). Scale bars indicate from low to high magnification: 1,000, 500, 100 microns. **b** For quantification of the degree of interstitial fibrosis, at least 3 scenes of a lung's lobe were enriched by contrast in image background and used for segmentation analysis. Lumina of large vessels, shown in green, were excluded from the quantification analysis. Whitespace areas (light blue), cells (yellow), and Anilin-positive regions (blue) were considered tissue. Calculation of % Collagen content (i.e. Anilin-blue positive areas) and tissue density was achieved by using the indicated calculations (lower left).

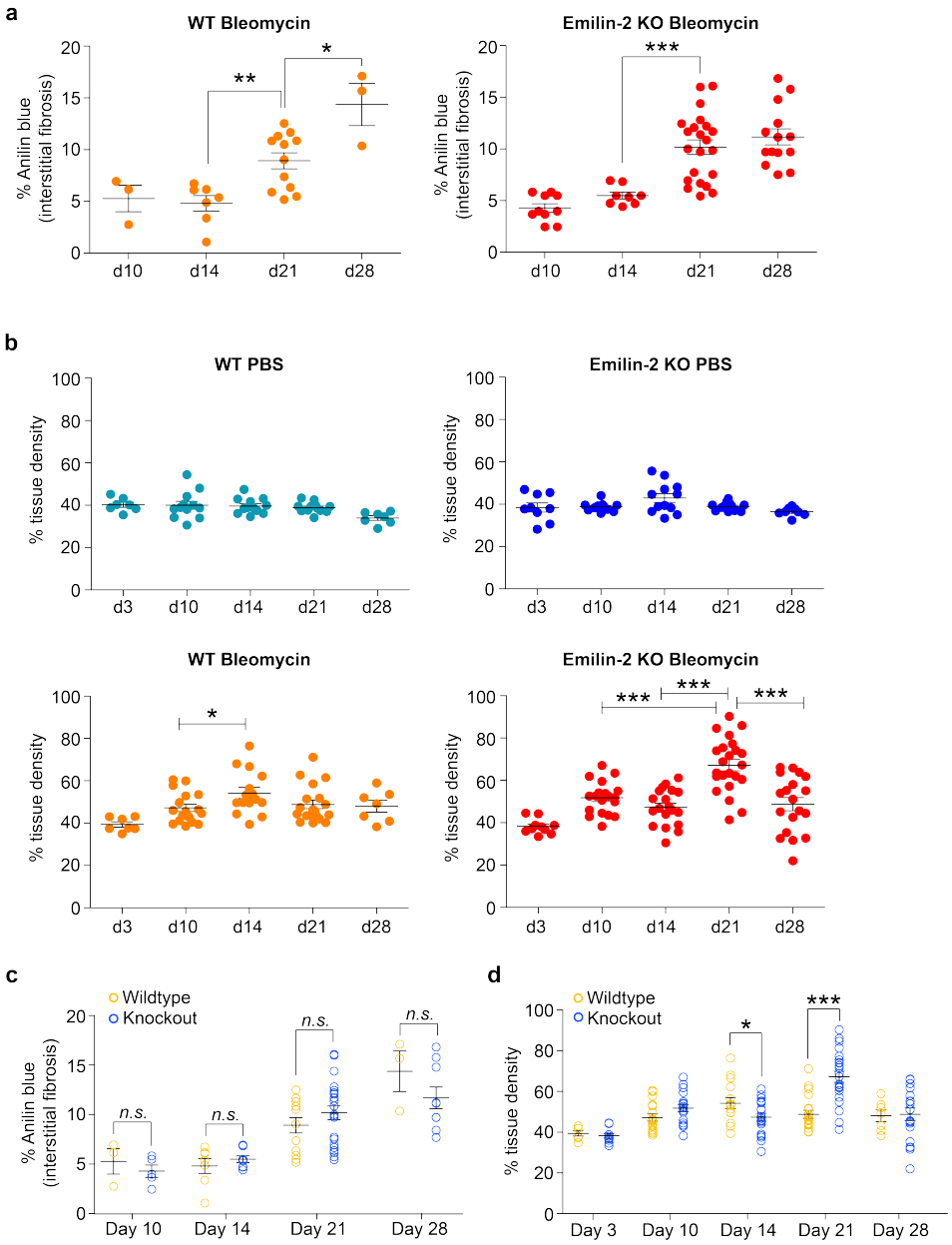

Figure 39: Quantification of the degree of fibrosis in Emilin-2 knockout animals. **a** Measurement of interstitial fibrosis by positive Anilin-blue-staining; stained lung sections were scanned and quantified as described in at least three areas. Emilin-2 knockout mice may develop more fibrotic lesions as compared to their wildtype litter mates by assessing the number of scanned sections possible; Statistics: *t*-test comparison with 95% confidence interval; *p = 0.0109; **p = 0.0028; ***p = 0.0005; data is plotted with SEM. **b** Comparison of tissue density as effect after the bleomycin lung injury shows no difference between genotypes in the PBS-treated mice. In the bleomycin lungs, the tissue density follows time-dependent changes with a peak in density at day 14 in wildtype animals, while in the knockout mice the density changed to massive increases at day 21; note that tissue density at day 28 is again comparable between the genotypes. **c** Data juxtaposition of the Anilin-blue data between wildtype and Emilin-2 knockout over the analyzed time points. **d** Data comparison between the two genotypes with respect to observed tissue density changes over time. Statistics: **a-d** *t*-test comparison with 95% confidence interval; *p < 0.05; ***p < 0.0001; data is plotted with the standard error of the mean.

3.2.4 Ablation of Emilin-2 causes lung immunophenotypic changes

Emilin-2 has been demonstrated to be expressed, among others, in the cochlear basilar membrane [309] and outside of blood vessels [310]. Although there is evidence that Emilin-2 is also expressed in the lung, and especially after lung injury [139], it remained unclear which cell types account most for its expression. Due to the cysteine-rich EMI domain, which are also found in other ECM proteins [311–313], fibroblasts could be a major source for Emilin-2. Surprisingly, by studying the time course single cell data as described in 3.1.3, the gene expression of Emilin-2 in the injured mouse lung was majorly detected within the leukocyte populations (Fig.40a). In fact, Emilin-2 gene expression seemed to be predominantly restricted to cell types of the immune lineage, with the highest average relative expression in fibronectin-positive macrophages (Fig.40b).

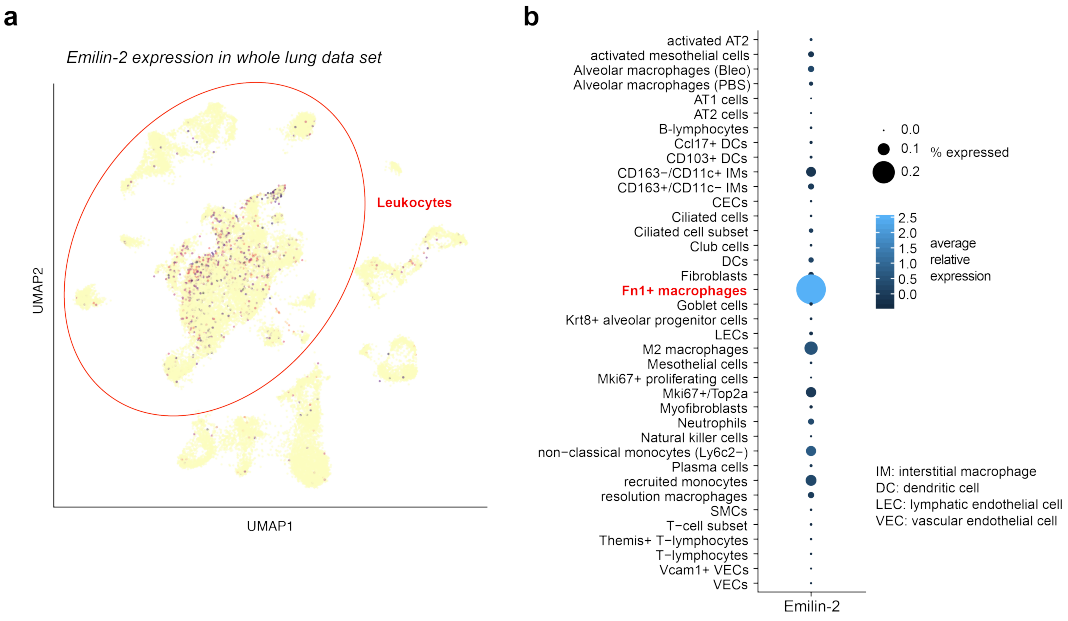

Figure 40: Emilin-2 expression in leukocyte populations. Plots were generated from the web tool as introduced in 3.1.3. **a** UMAP representation of Emilin-2 gene expression in the bleomycin-injured lung. **b** Dot plot representation of Emilin-2 gene expression in all detected cell types; highest expression was observed in fibronectin-expressing macrophages and various other leukocyte populations.

This quick survey suggested a potential contribution of immune cell-derived Emilin-2 during the bleomycin-induced lung injury time course. M2 phenotypic macrophages showed a differentiation over time in the whole lung survey, so the question arose whether Emilin-2 is involved in the processes during the inflammatory phase. Using flow cytometry, cells from bronchoalveolar lavage (BAL) at day 7 and day 10 post bleomcyin-injury were investigated from both Emilin-2 knockout mice and their wildtype litter mates. Following a gating strategy for a comprehensive immune cell panel (Fig.41a) [314], there were noticeable changes

between genotypes in the alveolar macrophages (Fig.41b). Despite being a heterogeneous population, alveolar macrophages from bleomycin-treated lungs have been shown to develop a remarkable pro-fibrotic M2-phenotypic profile [315, 316]. However, in the first line of defense, many alveolar macrophages undergo apoptosis after bleomycin administration [317, 318]. In the collected BAL cells, almost no alveolar macrophages were detected after bleomycin instillation as compared to control instillations (Fig.41b). Although alveolar macrophages participate in experimental lung fibrosis [315,316], and by checking the alveolar macrophage cell cluster in the single cell data from chapter 3.1.3, especially in Fig.13, it could be that remaining classical alveolar macrophages, defined by the expression of the gene Ear2 or the protein CD170/SiglecF, transcriptionally react to the lung injury stimulus by down-regulation of CD170 expression and Ear2 expression (cf. Fig.13g) with concomitant up-regulation of genes specific for an M2-macrophage phenotype. Additionally, it has been described that in the early phase of the bleomycin lung injury model, mice showed a decrease in alveolar macrophage numbers, while at the same time, interstitial macrophages and CD11b+ dendritic cells were found to be augmented [101]. The real time changes in the single cell data (Fig.13a, b) may corroborate this hypothesis. Investigation of the monocytes and interstitial macrophages from the bronchoalveolar lavage revealed only minor changes, with the exception that among the monocytes, the Emilin-2 knockout group showed higher cell numbers after bleomycin treatment at both day 7 and day 10 (Fig.41b). Similar observations were made for the natural killer cells with almost no differences between the bleomycin-treated groups, nor the PBS-treated groups. Among the myeloblast lineage, both neutrophils (in the PMN cluster in the gating strategy) and eosinophils showed an increase in cell number after bleomycin (Fig.41b). Again, the Emilin-2 knockout mice exceeded in numbers with the presence of neutrophils in the BAL, while for eosinophils, the numbers were found to be decreased compared to wildtype litter mates. Considering that there are elevated cell numbers of monocytes in response to injury in the Emilin-2 knockout mice, a higher number of interstitial macrophages would be expected in the lungs of knockout animals after bleomycin administration. The rationale behind this hypothesis was that short-lived macrophages recruited from the blood were shown to actively exacerbate lung inflammation [79, 319, 320] so that the presence of more monocytes in the knockout lungs after injury should give rise to more monocyte-derived interstitial macrophages. Moreover, it has been demonstrated that the presence of active interstitial macrophages may facilitate the influx of neutrophils into the lung [321, 322], which could explain the higher observed amounts of neutrophils in the Emilin-2 knockout animals. However, since the cells collected from bronchioalveolar lavage only represent the alveolar space but not the lung interstitium [323, 324], conclusions from the data need to be drawn carefully. To overcome this impediment, fresh single cell lung suspensions from both the wildtype and knockout cohorts were created. Similar to the BAL flow cytometry analysis, the leukocyte populations from whole lung cell suspensions were

stained and analyzed (Fig.42a). Comparing all cellular populations at four different time points, day 3, 14, 21, and 28, with respect to treatment and genotype group, several cellular changes became apparent (Fig.42a, b): for instance, the afore-mentioned observations in the alveolar macrophage population were similarly found in the whole lung flow cytometry analysis. Quite strikingly, the interstitial macrophages were low in numbers in animals that were treated only for three days, although bleomycin-treated animals already showed a minor increase (Fig42c). However, 14 days after bleomycin treatment, a higher number of interstitial macrophages was detected in wildtype animals. Surprisingly, the massive appearance of interstitial macrophages in the Emilin-2 knockout cohort was most prominent 21 days after the bleomycin-induced injury (Fig.42c). Comparing these cells between the days after treatment with the respective PBS control groups, two major clusters of interstitial macrophages stood out (Fig.42b). The fraction of interstitial macrophages that appeared only after injury partially coincided with identified inflammatory macrophages. Despite a clear differentiation of certain leukocyte populations in the healthy lung (PBS-treated lungs), the herein applied gating strategy is rather challenging for the fibrotic lung. The high cellular plasticity of especially the macrophage populations makes a distinct classification based on surface markers difficult.
Other cell types, e.g. the natural killer cells, behaved also differently in the model between the genotypes (Fig.42b, c). While there was a general increase in cell number over time up to day 21 after bleomycin treatment, only Emilin-2 knockout animals showed a continuous decline over time, which was significantly different at the highest state of fibrosis (day 21) in these mice (Fig.42c).

In a different experiment, whole lung tissue transcriptomics (bulk RNA-sequencing) from 14 day-treated mice were performed and the resulting data matched with cell type marker signatures derived from the single cell data (presented in 3.1.3) to discern enriched cell types in the bulk RNA-seq data. This deconvolution experiment showed an enriched presence of the signatures for natural killer cells and granulocytes in the Emilin-2 knockout (both treatment groups combined) (Fig.43a). With this interesting finding, single cell RNA sequencing data from Emilin-2 knockout mice, 10 and 14 days after bleomycin treatment, were merged with those from the wildtype litter mates (Fig.43b, c). Both genotypes resulted in a data set with good sample overlap, thus allowing for the discrimination of cell types that are rather driven by only one of the genotypes. Interestingly, especially the clusters for granulocytes and natural killer cells (highlighted in red in Fig.43d) seemed to be almost exclusively derived from cells of the Emilin-2 knockout bleomycin-treated lungs (Fig.43b, c). On the gene level, there was a noticeable enrichment of cell type frequency for granulocytes and natural killer cells in only the Emilin-2 knockout lungs (Fig.43e). When comparing to the flow cytometry data, a slight increase in cell number of eosinophil granulocytes can be confirmed. However, the group of granulocytes consist of different cell types, including as

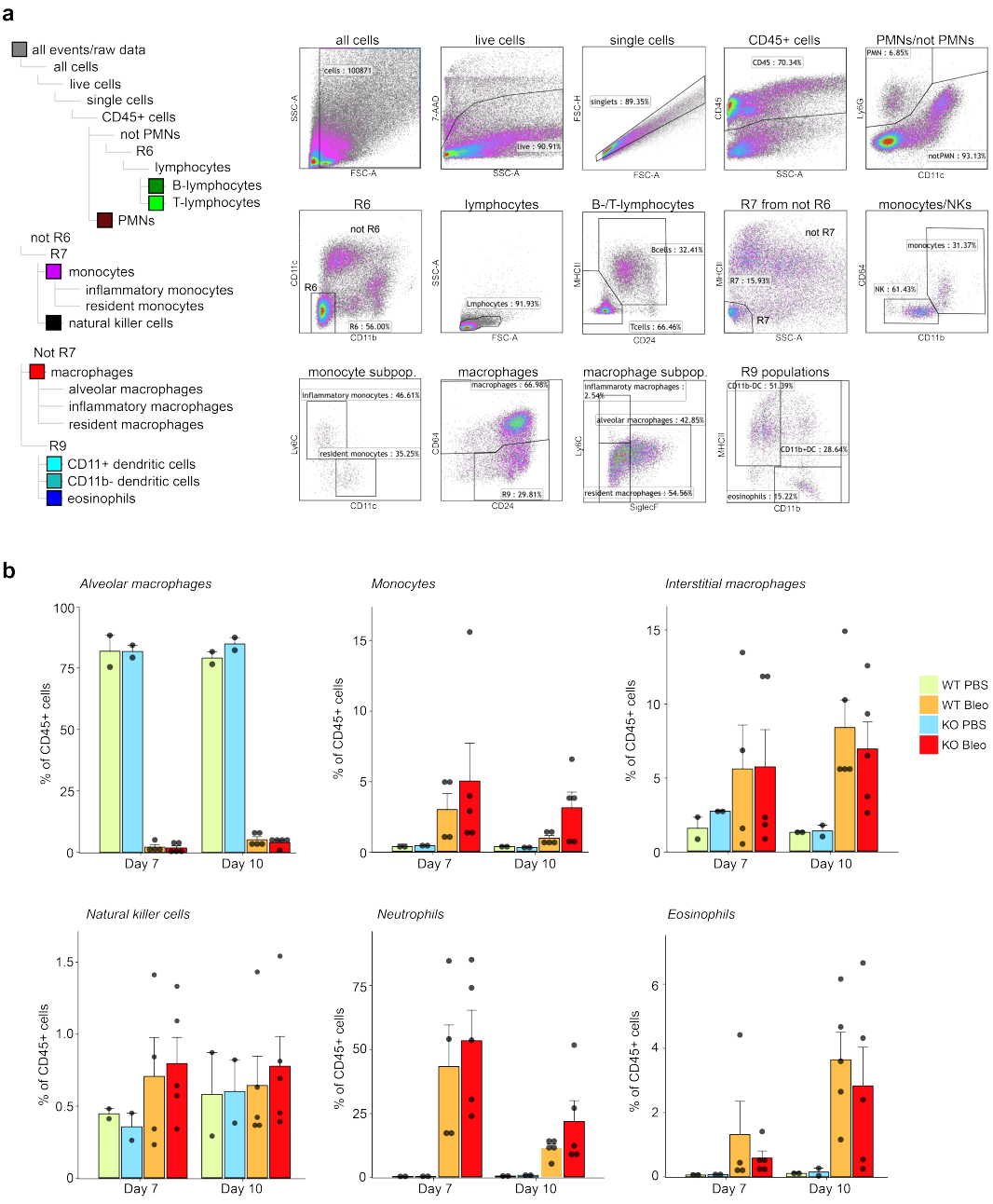

Figure 41: Flow cytometry analysis of bronchoalveolar lavage (BAL) from mouse lungs after bleomycin treatment in both wildtype and Emilin-2 knockout animals. **a** Gating strategy according to Yu et al., 2016. the node tree on the left represents the schematic strategy, while the scatter plots demonstrate the actual gates that were drawn. **b** Bar graph representation of BAL cells at day 7 and day 10 after injury. Shown are the percentages of total 100.000 cells of each group.

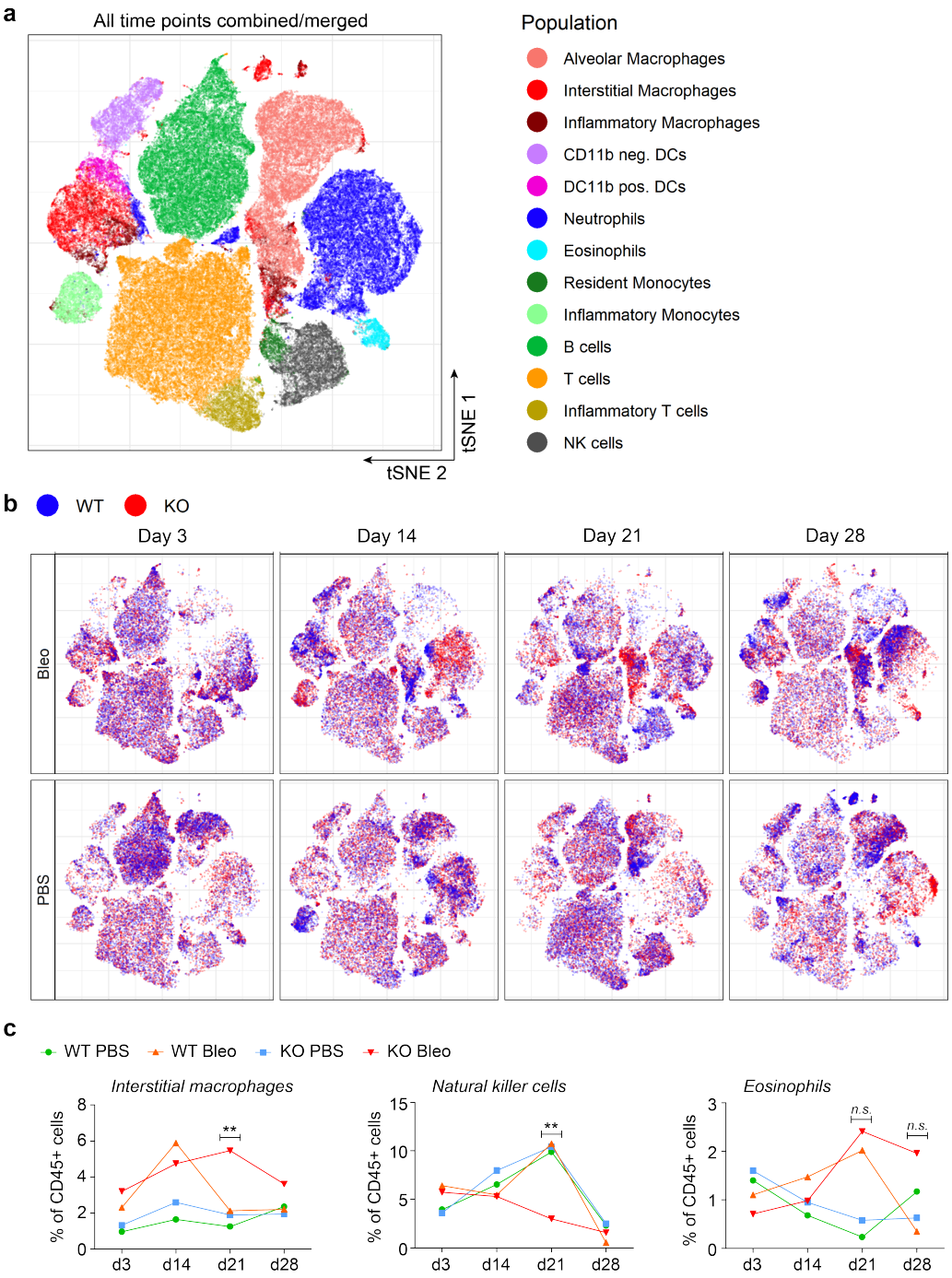

Figure 42: Flow cytometry analysis of whole lunge tissue from mice after bleomycin injury at different time points (day 3, 14, 21, and 28). **a** *t*SNE-map showing all recorded cells and cellular populations. The color code on the right describe the different lung cell types. **b** Comparison of cellular distributions in both bleomycin- and PBS-treated lungs over time in Emilin-2 knockout mice and their wildtype litter mates. Major differences include the interstitial macrophages and natural killer cells (both at day 21). **c** Quantification plots of significantly altered cellular distributions in interstitial macrophages (bleomycin day 21, Wilcoxon test: **p = 0.0095), natural killer cells (bleomycin day 21, Wilcoxon test: **p = 0.001), and high levels of eosinophils at day 28 in the knockout mice after bleomycin (Wilcoxon test: p = 0.1904; n.s.).

well basophil and neutrophil granulocytes, the latter being the most abundant granulocyte cell type. So better distinction between these granulocytes is necessary in future experiments. Studying genes related to type-II immunity in both natural killer cells and granulocytes revealed a small increase in the inflammatory response to bleomycin-induced injury in the knockout mice (Fig.43f). Mxd1 has been described as a transcriptional repressor to Myc, thereby limiting cell proliferation [337] and influencing gene expression [336]. Proteins of the Ifitm-family are induced by type-I and type-II interferons and play an important role in immune response signaling [338]. This is congruent with the hypothesis that Emilin-2 knockout mice may not adequately respond to type-II immune signals. The increased expression of the Il1r2 gene may additionally corroborate these findings due to its inhibiting function of the IL1R-cytokine signaling [339]. The fact that eosinophils have been found to be present at higher levels towards the late time points in the knockout lungs (Fig.42c) suggested an augmented involvement of ILC2s and type-II immunity in these mice. The natural killer cells demonstrated relatively higher expressed genes in the knockout lungs, including Id2, Jak1, and Ptprc (Fig.43f). Elevated levels of Id2 were detected in knockout mice. Id proteins inhibit DNA binding and promote T- and B- lymphocyte differentiation [340]. Moreover, the higher expression of e.g. the genes Jak1 or Rac2 (signaling G-protein) may suggest increased kinase activity in Emilin-2 ablated mice. The most prominent difference between the two genotypes was the expression of the Ptprc gene, which is the CD45 antigen (Fig.43f). Genes that seemed to be relatively higher expressed by the knockout mice included CD45, Sell (lymphocyte adhesion molecule-1), which again highlighted and corroborated an altered immune response upon bleomycin lung injury in the lung in these animals. Together, these findings revealed interesting differences between the genotypes, especially upon bleomycin-induced lung injury. More research is necessary to understand the molecular shifts and the functions of Emilin-2 in this context.

3.2.5 Keratin-8 expression in knockout lungs

The novel finding that Krt8-positive cells constitute an important intermediate cell state during fibrogenesis and the resolution of pulmonary fibrosis, evoked the question whether there are differences between wildtype or Emilin-2 knockout treated animals. Is Emilin-2 a key protein to guide proper alveolar regeneration? The expression patterns of both genes and proteins over time are of similar shape (Fig.44). Overlaying the protein expression profiles of both Krt8 and Emilin-2 over time in the bleomycin lung injury model (data from *Schiller et al., 2015*) resulted in similar progression courses over time, suggesting that there may be a functional correlation between these two proteins. The average gene expression profiles from the single cell data, irrespective of the genotype, were similar to the protein time courses. Expression of Krt8 and Emilin-2 peaked at day 10 and day 14 (Fig.44b; for

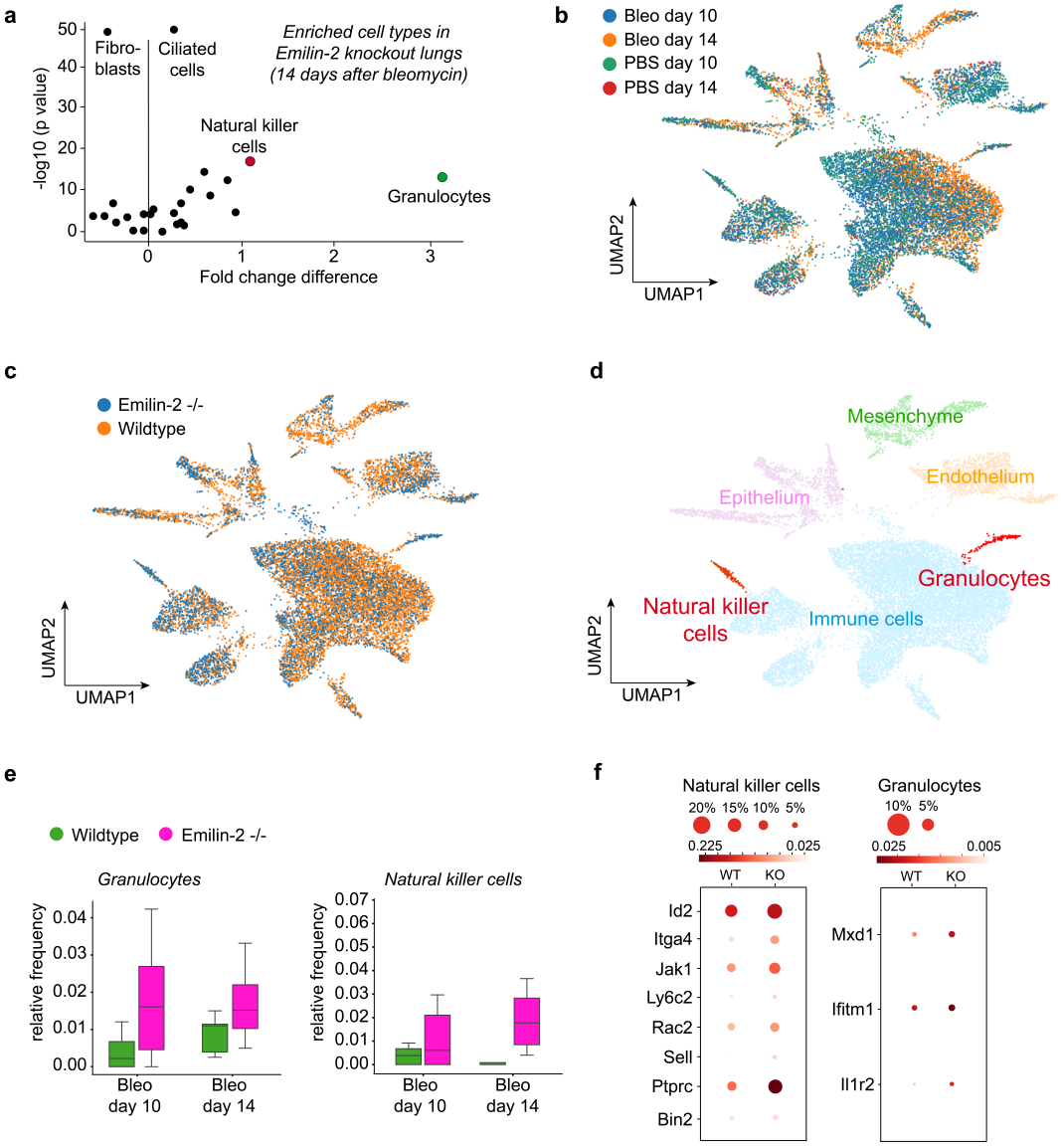

Figure 43: (Single cell) RNA sequencing shows increased natural killer cells and granulocytes in Emilin-2 knockout mice (total n = 34). **a** Deconvolution of RNA bulk transcriptomic data collected from lungs harvested at day 14 post bleomycin injury. Data was merged with cell type signatures derived from the whole lung single cell survey. Natural killer cells and granulocytes were found enriched on bulk RNA levels. Each dot represents an a-priori annotated cell type. **b** UMAP embedding of day 10 and day 14 treated lungs from both wildtype and knockout animals with respective treatment. **c** UMAP embedding of genotype - Emilin-2 knockout and wildtype littermates. **d** Annotation of clusters reveals common cell lineages with marker genes for immune cells, endothelial cells, mesenchymal cells, and epithelial cells; highlighted in red are the natural killer cells and granulocytes, both of which are found to be rather driven by the knockout phenotype, cf. panel c. **e** Relative frequency of granulocytes and natural killer cells in both genotypes; both cell types are found to be increased in cell numbers in the Emilin-2 knockout mice after bleomycin-induced lung injury. **f** Dot plot representation of regulated genes related to type-II immunity by both natural killer cells and granulocytes; data derived from time points day 10 and day 14 only.

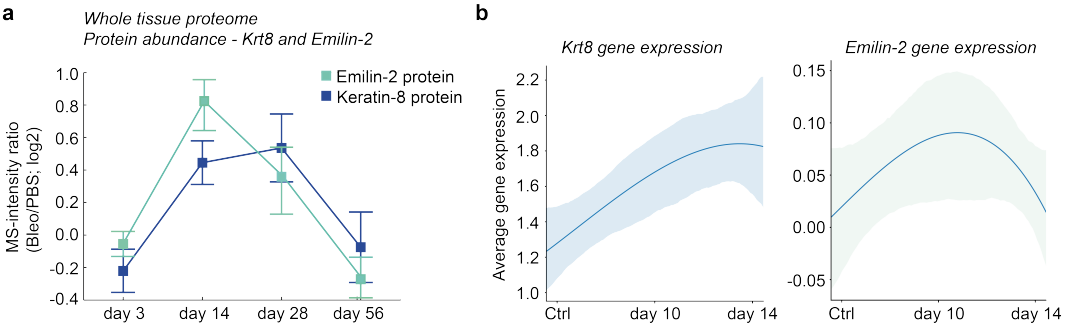

Figure 44: Expression profiles of Krt8 and Emilin-2 on protein and transcript level. **a** Protein abundance of whole lung tissue homogentates measured by mass spectrometry over time of the bleomycin lung injury model. Both proteins show similar time profiles. Plots are adapted from [139]. **b** Average gene expression profiles of Krt8 (left panel) and Emilin-2 (right panel) over day 10 and day 14 of the bleomycin model. Krt8 expression is plotted from alveolar epithelial cells only, Emilin-2 expression from immune cell populations.

Krt8: average gene expression in only epithelial cells; for Emilin-2: average gene expression in only immune cells).

Surveying the lungs of knockout mice after immunofluorescence staining against Krt8 (novel intermediate cell state), Sftpc (AT2 cells), and Pdpn (AT1 cells), revealed an increased presence of cells by positive nuclear staining (Fig.45a). Comparison between the genotypes revealed that after bleomycin injury, the emergence of Krt8-positive cells in knockout lungs is likewise triggered over time as observed in the wildtype litter mates. However, the major histologic difference was the remaining presence of still highly Krt8-positive cells in the lungs of Emilin-2 knockout mice 28 days after bleomycin treatment (Fig.45a, c). Although these lungs featured areas with active regeneration, bigger fibrotic foci with elongated Krt8+ cells were still spotted (Fig.45c, highlighted with yellow arrow heads). Quantification of the fluorescence intensities of the Krt8-positive signal (Fig.45b) demonstrated one the one hand shifted signal intensities in the knockout mice, and on the other hand an augmented number in Krt8+ cells at later time points, indicating that Emilin-2 knockout mice may have a deferred response to bleomycin-induced lung injury. However, such conclusions are rather speculative at this stage of data availability.

In order to study a potential co-expression of Emilin-2 and Krt8 with Pdpn, lung sections were stained accordingly and analyzed (Fig.46). Assessing the staining quality, Emilin-2 signals were observed in the knockout lungs giving need to improve the specificity for the antibody staining. In contrast to PBS-treated animals, where Emilin-2 expression was found to be restricted around vessels (congruent with the findings in [300]), a large Emilin-2 network was detected in day 10 lungs after bleomycin administration (Fig.46a). These networks are mainly found in fibrotic foci of the parenchymal tissue but never co-express Krt8. However, a co-localization of Krt8-positive cells in close proximity to Emilin-2-positive

regions was usually observed, suggesting that Emilin-2 may occupy a critical role during alveolar regeneration.

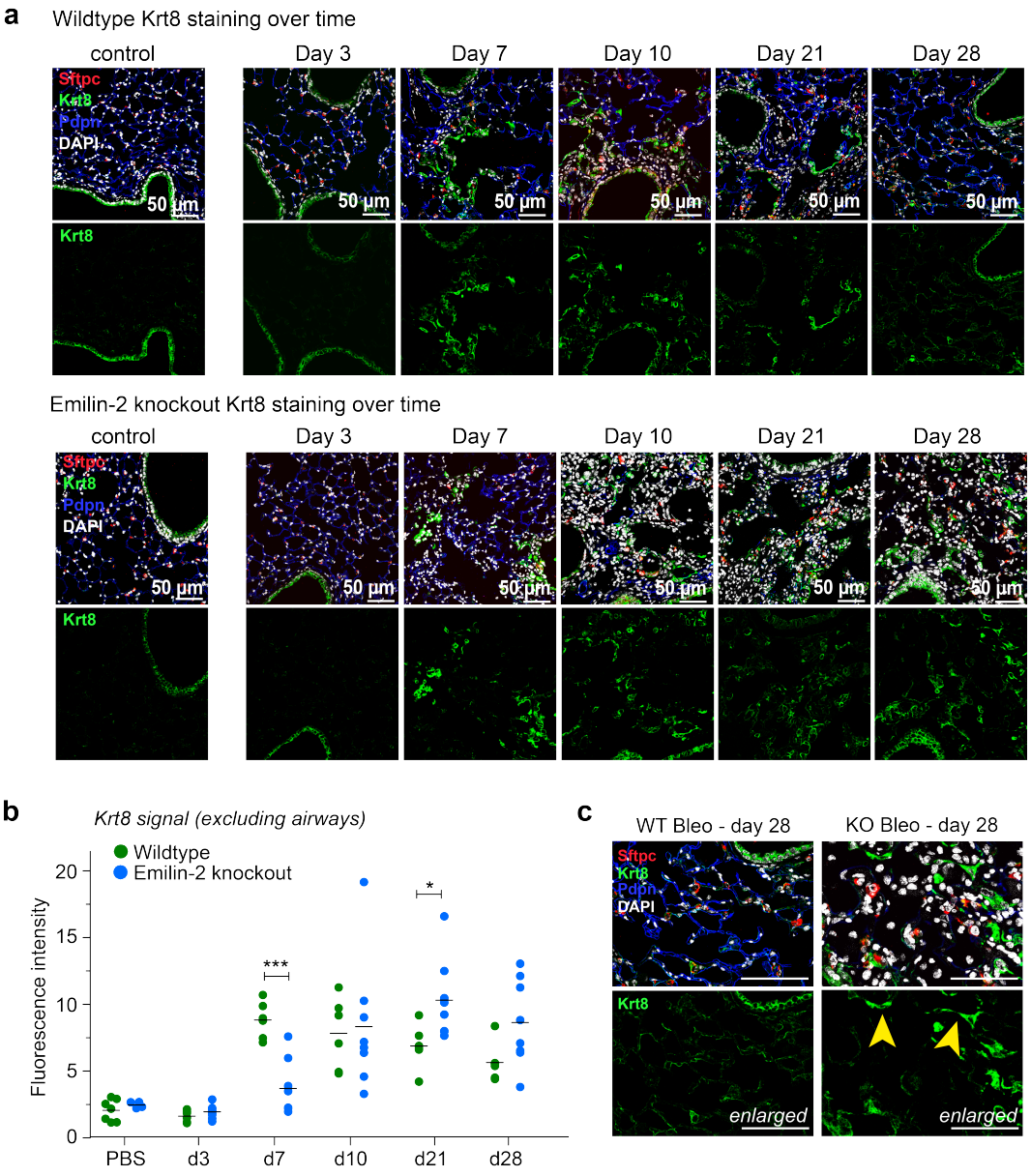

Figure 45: **a** Krt8 immunofluorescence staining in wildtype and Emilin-2 knockout lung sections at different time points of the bleomycin model. Knockout animals have noticeably higher numbers in cells (Dapi staining) at day 10 and onward; Apart from the airway cells, Krt8-positive cells (in green) undergo morphological changes as observed in wildtype mice, including elongation during fibrogenesis, but are retained in the lungs of Emilin-2 knockout animals much longer, cf. day 28. AT2 cells were stained with Sftpc (red), AT1 cells with Pdpn (blue). Scale bars indicate 50 microns. **b** Quantification of the Krt8 mean fluorescence signal by excluding airway regions corroborates the observation of delayed alveolar reconstruction by the remaining presence of Krt8-positive and elongated cells in knockout lungs at day 28 post bleomycin injury; data shown with SD and *p = 0.0266; ***p = 0.0002. **c** Enlarged view of day 28 wildtype and knockout lungs to demonstrate the presence of still elongated Krt8-positive cells in the alveolar regions (highlighted by yellow arrowheads); note the lack of Pdpn staining in the knockout lungs, indicating that there is still incomplete alveolar regeneration.

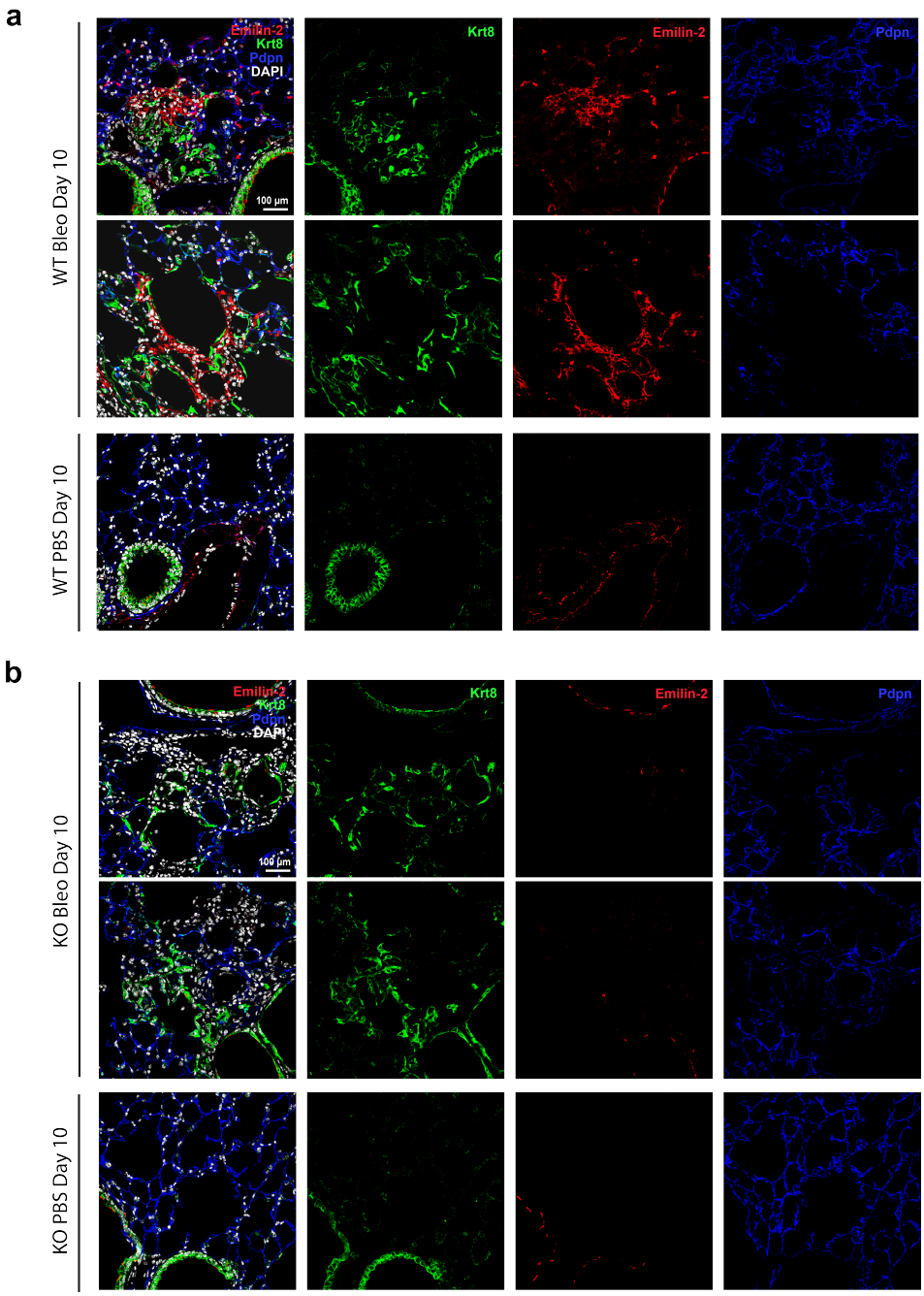

Figure 46: Co-localization of Emilin-2 and Keratin-8 in wildtype and Emilin-2 knockout lungs. **a** Wildtype mice show distinct expression of Emilin-2 in the extracellular space, mostly located within fibrotic foci of the bleomycin lung. Krt8+ cells are in close proximity to these foci, suggesting that there is an important functional role between these cell types. In PBS-treated lungs, Emilin-2 expression is restricted to vascular regions, as previously described in the literature. **b** In the Emilin-2 knockout lungs, there are no differences in Krt8-expression; positively stained cells reveal similar cell shapes and intensity signals as compared to wildtype mice.

3.2.6 Discussion

This chapter introduced Emilin-2, an extracellular protein with a potential role during fibrogenesis after bleomycin-induced pulmonary fibrosis. In both proteomic and transcriptomic analysis, Emilin-2 showed highly significant gene expression [139]. With a peak expression at day 14 after injury, a putative function in mediating fibrogenesis had been suggested [139]. With the help of the Emilin-2 knockout mouse it was aimed to characterize and define differences in the bleomycin-injured lungs of these animals. The Emilin-2 knockout mice are viable and, by direct comparison with their wildtype litter mates, there were no noticeable differences. Mice included in the experiments, did not show any obvious signs in behavior or appearance in response to bleomycin other than the typical features seen after the administration. Despite a measurable difference in loss of body weight between wildtype and knockout after day 21, the knockout animals recovered and re-gained weight slower and less effectively. Lung function analysis showed no compliance differences between the genotypes. However, measurements were only performed at the phases of peak inflammation and fibrogenesis, which only allows to draw conclusions for the acute response but lacks full comprehensibility as later time points are still missing. As lung function in the mouse is an invasive experiment and time-consuming per animal, the experimental setup limited the inclusion of many mice per time point. Moreover, the cytotoxic effects of bleomycin administration elicits a heavy burden for the mouse allowing only small time frames to measure lung function parameters without risking the animal's death.
Using histology, the degree of fibrosis was assessed and quantified. While all mice showed massive loss of functional alveolar tissue after bleomycin administration, followed by slow tissue regeneration, the lungs of Emilin-2 knockout mice expressed higher amounts of collagen 1 and alpha smooth muscle actin at day 21 as compared to wildtype mice. Masson's Trichrome staining of lung sections was not suitable to detect differences in collagen production between control or bleomycin groups. Scrutinizing condensed tissue areas allowed for the identification of interstitial fibrosis by excessive ECM production and matrix deposition. Tissue density assessment proved to be the more reliable method to discern true interstitial fibrosis than quantification of Anilin-blue-positive Collagen 1 signals. Histologic and morphologic observations included a leukocyte accumulation in the alveolar regions and around the interstitial fibrotic regions, which appeared by eye to be intensified in the Emilin-2 knockout lungs. Together with other data, these histologic features of the bleomycin-treated lungs may be used to corroborate key findings.

Addressing the previously acquired single cell RNA sequencing data on Emilin-2 expression throughout the bleomycin time course, led to the identification of leukocyte populations as a major expression source for Emilin-2. In order to trace potential contributions by the immune compartment, first, bronchoalveolar lavage (BAL) samples were analyzed using

a comprehensive flow cytometry panel for lung leukocytes. Later, whole lung tissue samples were used for a likewise analysis. The BAL results permitted to gain first notion of the distribution of immune cells in the alveolar space of affected mice. In order to draw more comprehensive conclusions by not solely studying the alveolar regions with BAL samples, the whole tissue approach allowed for the determination of interstitial changes after bleomycin-induced lung injury. The inclusion of different time points and the representation of data in dimensionality-reduced plots (tSNE embedding), simplified the interpretation of changes within cell types over time. In particular interstitial macrophages showed an interesting abundance curve, considering that these cells are recruited upon injury from the circulation. Further experiments for the improved characterization of these macrophages are indispensable.

Natural killer cells, as well as natural killer T cells play important roles in the context of pulmonary fibrosis. Being cytotoxic, these cells exert functions by secretion of cytokines that change the extracellular milieu [325, 326]. An interesting observation was recently published and describes the bleomycin-induced activation of group-2 innate lymphoid cells (ILC2) by interferon gamma after bleomycin-induced lung injury [327]. Innate lymphoid cells (ILCs) have been categorized into three subsets: ILC1, ILC2, and ILC3 [341]. While the ILC1 and ILC3 populations can exert pro-inflammatory functions, the ILC2 cells are involved in type-II immunity, and are negatively controlled by ILC1 or ILC3 cells, depending on the environmental stimuli [327, 341, 342]. Natural killer cells, which due to their interferon-gamma expression are allocated to the ILC1 group, were shown to inhibit ILC2 proliferation and hence the production of type-II cytokines [327], including interleukin-5, which has been described to cause eosinophilia [327, 328]. With a lack of adequate cell numbers of natural killer cells in the bleomycin-treated Emilin-2 knockout animals, a hypothesis could be that the fibrotic phenotype in these lungs may not be sufficiently controlled by responding to type-II immunity signals. Hence, in these lungs, natural killer cells would not be capable of the attenuation of inflammation and the resulting pulmonary fibrosis. If ablation of the Emilin-2 gene affected the type-II immunity in the lung, which needs better elucidation, then eosinophilia in bleomycin-treated knockout animals, or the observed relative increase of eosinophils (cf. Fig.42c) in knockout mice at 28 days post-bleomycin injury would be consistent with previous findings. Nevertheless, a more comprehensive study of the contribution of Emilin-2 and immunologic responses in the lung is needed. The fluctuation of natural killer cells as observed in the PBS-treated lung has been described before [334, 335], indicating that bleomycin is not as much affecting natural killer cell numbers but further elucidation is necessary.

In agreement with previously reported findings, low amounts of natural killer cells may influence type-II immunologic responses and directly influence eosinophilia. Type-II immunity

has not yet been studied in the context of the injured lungs of Emilin-2 knockout mice. The contribution of highly plastic macrophage populations and the complex cross-talks with other leukocyte and non-leukocyte populations remain to date elusive. Flow cytometry can not tackle these issues easily. Moreover, Emilin-2 gene expression was found highest in fibronectin-positive macrophages, maybe indicating that these macrophages express several extracellular proteins, including Emilin-2, during the fibrogenic phase of the bleomycin-induced lung injury. In order to address future work, recent findings of immunity-related mechanisms in the injured and fibrotic lung might need to be adduced to describe the observed findings regarding the natural killer cells and the increased eosinophils numbers: Upon bleomycin injury, ILC2 cells get activated due to epithelial damage and its release of among others IL33 and IL25 [343–346]. ILC2 cells are tissue-resident and react sensitively on microenvironmental changes [343, 344]. At this stage, macrophages and monocytes contribute to the expression of pro-inflammatory signals [347]. With the early neutrophil influx, full-length IL33 protein e.g. gets actively cleaved by neutrophils into heavily active peptides [348, 349], which strongly activates ILC2 cells. Upon activation, many type-II cytokines are produced, including IL-13, which has been described to drive fibrogenesis [350–352], IL6, and IL5, the latter of which may cause eosinophilia and eosinophilic inflammation [353]. Interestingly, expression of amphiregulin (Areg) has also been noted in these cells [354, 355], highlighting that ILC2s may actively contribute to alveolar regeneration after epithelial damage (cf. findings in chapter 3.1.4) with beneficial functions (which might not yet be detectable at day 10 and 14 in the herein presented data). In line with this, as such inflammatory and pro-fibrogenic signals augment, a negatively regulating feedback mechanism kicks in. ILC1, including natural killer cells, were shown to suppress ILC2 activation [344]. In fact, due to high cellular plasticity among the three ILC subsets, even ILC2 cells may respond in an autocrine manner to the presence of strong pro-inflammatory signals [356]. Interferon-gamma has been shown to suppress ILC2 function, which denotes a key role for the presence of ILC1 cells and natural killer cells [344]. Since both natural killer cells and ILC1 cells are hard to distinguish by phenotype it could well be that the designation of the *natural killer cells* in the flow cytometry data is misleading, as the cells were only defined by a CD11b-/CD64- gate. The application of further distinct natural killer cell markers could benefit future experiments. Interestingly, intranasal administration of either IL25 or IL33 in mice resulted in severe type-II inflammation in the lungs [357, 358]. Both cytokines together with the co-stimulatory IL2 cytokine, which is primarily derived from T lymphocytes, induce a fast cell proliferation and boost the type-II cytokine production by ILC2 cells [344]. Studies have identified signal transducer and activator of transcription (STAT) pathways as main regulatory source for ILC2 cells [359–362].

Activation and survival of ILC2 cells, triggered by e.g. IL9 derived from T lymphocytes and natural killer cells, are enhanced by STAT5, which is activated by Jak1, a gene that showed

higher expression in the scRNA-seq data of the Emilin-2 knockout mice (cf. Fig.43d). This may explain why during the inflammatory and fibrogenic phase, these mice show increased activation of immune cells, particularly of type-II inflammatory cells. Additionally, this would be in line with the observed increase in CD45+ cells (Ptprc gene expression). At day 10 and day 14, the presence of natural killer cells in the lungs of both wildtype and Emilin-2 knockout mice was at equilibrium. Hence, a putative function of Emilin-2 could be to mediate proper microenvironmental stimuli to enhance the increase in cells that negatively regulate type-II immunity at the peak of inflammation so that proper regenerative responses can be initiated. At a first and not yet comprehensive glance, the knockout mice might lack these important signals, which could lead to a perpetuation of ILC2-mediated inflammation and a potentially concomitant eosinophilic inflammation (provided this hypothesis is true). Therefore, the precise molecular role of Emilin-2 remains to date elusive. A possible link could be the pro-TGFb modulatory function of Emilin-2 [302, 363, 364] to activate the TGFb/BMP pathways for intensified suppression of inflammation via IL5, or fibrogenesis via IL13 [365]. For an improved understanding, other time points need to be studied, and additionally, the conduction of more functional experiments regarding a potential immune-modulatory feature of Emilin-2 is necessary.

Additionally, it cannot be excluded that fibroblast populations exert functional roles in the assembly of Emilin-2 in the extracellular space. Mesenchymal cells, in particular fibroblast populations, may also contribute to Emilin-2 expression but further experiments are necessary.

Ultimately, and because the detection of a novel transient intermediate cell state after lung injury has been described, lung tissue from Emilin-2 knockout mice was stained with Krt8 (for the newly described cell state), Sftpc (for AT2 cells) and Pdpn (for AT1 cells). The analysis of micrographs from all examined time points revealed an increased presence of Krt8+ cells 10 days after bleomycin-induced injury, with the noticeable difference that Emilin-2 knockout animals showed higher overall cell numbers, indicative for increased tissue density, and the persistent presence of the Krt8+ cells in the alveolar regions up to the late time points (day 21, day 28). At this stage, and although there is co-localization of Krt8 and Emilin-2 signals in immunofluorescence staining, potentially concerted roles or functions of Krt8 and Emilin-2 after injury and during regeneration could not be carved out. Additional efforts are essential to understand the underlying molecular processes, with a particular focus on cellular communications and the signaling landscape.

4 Conclusions

In conclusion, the presented work delineated epithelial cell plasticity after injury in the lung and, for the first time, described an intermediate cell state, which is involved in non-gradual transdifferentiation [198]. This cell state is marked by high Krt8 expression and was proposed to function as a progenitor cell in the AT2-to-AT1 conversion after lung injury. A precise definition of the molecular functions of these cells is highly relevant, which can be ideally addressed using single cell RNA sequencing [198]. Considering that this novel intermediate cell state is found in different experimental injury models but also in lungs from heavily diseased patients (both in acute lung injury and fibrotic disease), further experimentation is needed to allow for translational conclusions, i.e. whether the herein presented findings are comparable between human disease and mouse experimental pulmonary fibrosis.

For a better notion of the development of chronic lung disease, researchers team up for current endeavors for the Human Lung Cell Atlas [155]. Together with efforts for murine cell atlases, this project will lay basis to understand lung development, lung homeostasis, as well as lung disease formation and lung regeneration [198]. However, such efforts are only feasible by the generation of longitudinal data derived from many individuals and by highly functional data integration methods [198, 214]. By combining experimental models of pulmonary fibrosis with longitudinal sequencing and novel computational approaches [198, 331], as used in this presented work, complex dynamic processes can be identified, even across samples and species, empowering predictions for future translational work [198].

References

[1] Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016. Geneva, World Health Organization; 2018.

[2] Pinkerton, K.E., Green, F.H., Saiki, C., Vallyathan, V., Plopper, C.G., Gopal, V., Hung, D., Bahne, E.B., Lin, S.S., Mche, M.G. and Schenker, M.B., 2000. Distribution of particulate matter and tissue remodeling in the human lung. *Environmental Health Perspectives*, 108(11), pp.1063-1069.

[3] Weibel, E.R., Cournand, A.F. and Richards, D.W., 1963. Morphometry of the human lung (Vol. 1). Berlin: Springer.

[4] Lowe, J.S., and Anderson, P.G., 2015. Epithelial Cells. *Stevens Lowe's Human Histology* Chapter 3, Fourth Edition, ISBN:978-0-7234-3502-0.

[5] Rigden, H.M., Alias, A., Havelock, T., O'Donnell, R., Djukanovic, R., Davies, D.E. and Wilson, S.J., 2016. Squamous metaplasia is increased in the bronchial epithelium of smokers with chronic obstructive pulmonary disease. *PloS one*, 11(5), p.e0156009.

[6] Rock, J.R., Onaitis, M.W., Rawlins, E.L., Lu, Y., Clark, C.P., Xue, Y., Randell, S.H. and Hogan, B.L., 2009. Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proceedings of the National Academy of Sciences*, 106(31), pp.12771-12775.

[7] Pepicelli, C.V., Lewis, P.M. and McMahon, A.P., 1998. Sonic hedgehog regulates branching morphogenesis in the mammalian lung. *Current biology*, 8(19), pp.1083-1086.

[8] Whitsett, J.A., 2018. Airway epithelial differentiation and mucociliary clearance. *Annals of the American Thoracic Society*, 15(Supplement 3), pp.S143-S148.

[9] Whitsett, J.A. and Alenghat, T., 2015. Respiratory epithelial cells orchestrate pulmonary innate immunity. *Nature immunology*, 16(1), p.27.

[10] Shaykhiev, R., 2015. Multitasking basal cells: combining stem cell and innate immune duties. *European Respiratory Journal*, 46: 894-897.

[11] Bustamante-Marin, X. M., Ostrowski, L. E., 2017. Cilia and mucociliary clearance. *Cold Spring Harbor perspectives in biology*, 9(4): a028241.

[12] Button, B., Anderson, W. H., Boucher, R. C., 2016. Mucus hyperconcentration as a unifying aspect of the chronic bronchitic phenotype. *Annals of the American Thoracic Society*, 13(Supplement 2), S156-S162.

[13] Zhou-Suckow, Z., Duerr, J., Hagner, M., Agrawal, R., Mall, M. A., 2017. Airway mucus, inflammation and remodeling: emerging links in the pathogenesis of chronic lung diseases. *Cell and tissue research*, 367(3), 537-550.

[14] Levardon, H., Yonker, L.M., Hurley, B.P. and Mou, H., 2018. Expansion of airway basal cells and generation of polarized epithelium. *Bio-protocol*, 8(11).

[15] Rock, J.R., Randell, S.H. and Hogan, B.L., 2010. Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. *Disease models mechanisms*, 3(9-10), pp.545-556.

[16] Morimoto, M., Liu, Z., Cheng, H. T., Winters, N., Bader, D., Kopan, R., 2010. Canonical Notch signaling in the developing lung is required for determination of arterial smooth muscle cells and selection of Clara versus ciliated cell fate. *J Cell Sci*, 123(2), 213-224.

[17] Morimoto, M., Nishinakamura, R., Saga, Y., Kopan, R., 2012. Different assemblies of Notch

receptors coordinate the distribution of the major bronchial Clara, ciliated and neuroendocrine cells. *Development*, 139(23), 4365-4373.

[18] Siebel, C., Lendahl, U., 2017. Notch signaling in development, tissue homeostasis, and disease. *Physiological reviews*, 97(4), 1235-1294.

[19] Arbi, M., Pefani, D. E., Kyrousi, C., Lalioti, M. E., Kalogeropoulou, A., Papanastasiou, A. D., ... Lygerou, Z., 2016. GemC1 controls multiciliogenesis in the airway epithelium. *EMBO reports*, 17(3), 400-413.

[20] Campbell, E. P., Quigley, I. K., Kintner, C., 2016. Foxn4 promotes gene expression required for the formation of multiple motile cilia. *Development*, 143(24), 4654-4664.

[21] Stubbs, J. L., Vladar, E. K., Axelrod, J. D., Kintner, C., 2012. Multicilin promotes centriole assembly and ciliogenesis during multiciliate cell differentiation. *Nature cell biology*, 14(2), 140.

[22] Zhou, F., Narasimhan, V., Shboul, M., Chong, Y. L., Reversade, B., Roy, S., 2015. Gmnc is a master regulator of the multiciliated cell differentiation program. *Current Biology*, 25(24), 3267-3273.

[23] Evans, M.J., Van Winkle, L.S., Fanucchi, M.V. and Plopper, C.G., 2001. Cellular and molecular characteristics of basal cells in airway epithelium. *Experimental lung research*, 27(5), pp.401-415.

[24] Daniely, Y., Liao, G., Dixon, D., Linnoila, R.I., Lori, A., Randell, S.H., Oren, M. and Jetten, A.M., 2004. Critical role of p63 in the development of a normal esophageal and tracheobronchial epithelium. *American Journal of Physiology-Cell Physiology*, 287(1), pp.C171-C181.

[25] Evans, M.J., Cox, R.A., Shami, S.G., Wilson, B. and Plopper, C.G., 1989. The role of basal cells in attachment of columnar cells to the basal lamina of the trachea. *Am J Respir Cell Mol Biol*, 1(6), pp.463-469.

[26] Rock, J.R. and Hogan, B.L., 2011. Epithelial progenitor cells in lung development, maintenance, repair, and disease. *Annual review of cell and developmental biology*, 27, pp.493-512.

[27] Reid, L., Meyrick, B., Antony, V.B., Chang, L.Y., Crapo, J.D. and Reynolds, H.Y., 2005. The mysterious pulmonary brush cell: a cell in search of a function. *American journal of respiratory and critical care medicine*, 172(1), pp.136-139.

[28] Krasteva, G., Canning, B.J., Papadakis, T. and Kummer, W., 2012. Cholinergic brush cells in the trachea mediate respiratory responses to quorum sensing molecules. *Life sciences*, 91(21-22), pp.992-996.

[29] Saunders, C.J., Reynolds, S.D. and Finger, T.E., 2013. Chemosensory brush cells of the trachea. A stable population in a dynamic epithelium. *American journal of respiratory cell and molecular biology*, 49(2), pp.190-196.

[30] Kinnamon, S.C. and Reynolds, S.D., 2009. Using taste to clear the air (ways). *Science*, 325(5944), pp.1081-1082.

[31] Kinnamon, S.C., 2012. Taste receptor signalling from tongues to lungs. *Acta physiologica*, 204(2), pp.158-168.

[32] Finger, T.E., Böttger, B., Hansen, A., Anderson, K.T., Alimohammadi, H. and Silver, W.L., 2003. Solitary chemoreceptor cells in the nasal cavity serve as sentinels of respiration. *Proceedings of the National Academy of Sciences*, 100(15), pp.8981-8986.

[33] Krasteva, G., Canning, B.J., Hartmann, P., Veres, T.Z., Papadakis, T., Mühlfeld, C., Schliecker, K., Tallini, Y.N., Braun, A., Hackstein, H. and Baal, N., 2011. Cholinergic chemosensory cells in the trachea regulate breathing. *Proceedings of the National Academy of Sciences*, 108(23), pp.9478-9483.

[34] Reynolds, S.D., Giangreco, A., Power, J.H. and Stripp, B.R., 2000. Neuroepithelial bodies of pulmonary airways serve as a reservoir of progenitor cells capable of epithelial regeneration. *The American journal of pathology*, 156(1), pp.269-278.

[35] Verckist, L., Lembrechts, R., Thys, S., Pintelon, I., Timmermans, J.P., Brouns, I. and Adriaensen, D., 2017. Selective gene expression analysis of the neuroepithelial body microenvironment in postnatal lungs with special interest for potential stem cell characteristics. *Respiratory research*, 18(1), p.87.

[36] Cutz, E., Pan, J., Yeger, H., Domnik, N.J. and Fisher, J.T., 2013, January. Recent advances and contraversies on the role of pulmonary neuroepithelial bodies as airway sensors. *Academic Press - In Seminars in cell developmental biology*, 24(1), pp. 40-50.

[37] Sorokin, S.P., Hoyt Jr, R.F. and Pearsall, A.D., 1983. Comparative biology of small granule cells and neuroepithelial bodies in the respiratory system: short review. *American Review of Respiratory Disease*, 128(2P2), pp.S26-S31.

[38] Hogan, B.L., Barkauskas, C.E., Chapman, H.A., Epstein, J.A., Jain, R., Hsia, C.C., Niklason, L., Calle, E., Le, A., Randell, S.H. and Rock, J., 2014. Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell stem cell*, 15(2), pp.123-138.

[39] Rawlins, E.L., Okubo, T., Xue, Y., Brass, D.M., Auten, R.L., Hasegawa, H., Wang, F. and Hogan, B.L., 2009. The role of Scgb1a1+ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell stem cell*, 4(6), pp.525-534.

[40] Stripp, B.R., 2008. Hierarchical organization of lung progenitor cells: is there an adult lung tissue stem cell?. *Proceedings of the American Thoracic Society*, 5(6), pp.695-698.

[41] Zuo, W., Zhang, T., Wu, D.Z.A., Guan, S.P., Liew, A.A., Yamamoto, Y., Wang, X., Lim, S.J., Vincent, M., Lessard, M. and Crum, C.P., 2015. p63+ Krt5+ distal airway stem cells are essential for lung regeneration. *Nature*, 517(7536), p.616.

[42] Reynolds, S.D., Reynolds, P.R., Pryhuber, G.S., Finder, J.D. and Stripp, B.R., 2002. Secretoglobins SCGB3A1 and SCGB3A2 define secretory cell subsets in mouse and human airways. *American journal of respiratory and critical care medicine*, 166(11), pp.1498-1509.

[43] Corfield, A.P., 2015. Mucins: a biologically relevant glycan barrier in mucosal protection. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1850(1), pp.236-252.

[44] Rogers, D.F., 2003. The airway goblet cell. *The international journal of biochemistry cell biology*, 35(1), pp.1-6.

[45] Fahy, J.V. and Dickey, B.F., 2010. Airway mucus function and dysfunction. *New England Journal of Medicine*, 363(23), pp.2233-2247.

[46] Boucherat, O., Boczkowski, J., Jeannotte, L. and Delacourt, C., 2013. Cellular and molecular mechanisms of goblet cell metaplasia in the respiratory airways. *Experimental lung research*, 39(4-5), pp.207-216.

[47] Rock, J.R., Gao, X., Xue, Y., Randell, S.H., Kong, Y.Y. and Hogan, B.L., 2011. Notch-dependent differentiation of adult airway basal stem cells. *Cell stem cell*, 8(6), pp.639-648.

[48] Whitsett, J.A., Haitchi, H.M. and Maeda, Y., 2011. Intersections between pulmonary development and disease. *American journal of respiratory and critical care medicine*, 184(4), pp.401-406.

[49] Danahay, H., Pessotti, A.D., Coote, J., Montgomery, B.E., Xia, D., Wilson, A., Yang, H., Wang, Z., Bevan, L., Thomas, C. and Petit, S., 2015. Notch2 is required for inflammatory cytokine-driven goblet cell metaplasia in the lung. *Cell reports*, 10(2), pp.239-252.

[50] Tompkins, D.H., Besnard, V., Lange, A.W., Wert, S.E., Keiser, A.R., Smith, A.N., Lang, R. and Whitsett, J.A., 2009. Sox2 is required for maintenance and differentiation of bronchiolar Clara, ciliated, and goblet cells. *PloS one*, 4(12), p.e8248.

[51] Evans, M.J., Cabral, L.J., Stephens, R.J. and Freeman, G., 1975. Transformation of alveolar type 2 cells to type 1 cells following exposure to NO2. *Experimental and molecular pathology*, 22(1), pp.142-150.

[52] Sirianni, F.E., Chu, F.S. and Walker, D.C., 2003. Human alveolar wall fibroblasts directly link epithelial type 2 cells to capillary endothelium. *American journal of respiratory and critical care medicine*, 168(12), pp.1532-1537.

[53] Selman, M. and Pardo, A., 2006. Role of epithelial cells in idiopathic pulmonary fibrosis: from innocent targets to serial killers. *Proceedings of the American Thoracic Society*, 3(4), pp.364-372.

[54] Pattle, R.E., 1965. Surface lining of lung alveoli. *Physiological reviews*, 45(1), pp.48-79.

[55] Bachofen, H. and Schürch, S., 2001. Alveolar surface forces and lung architecture. *Comparative Biochemistry and Physiology Part A: Molecular Integrative Physiology*, 129(1), pp.183-193.

[56] Hogan, B.L., Barkauskas, C.E., Chapman, H.A., Epstein, J.A., Jain, R., Hsia, C.C., Niklason, L., Calle, E., Le, A., Randell, S.H. and Rock, J., 2014. Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell stem cell*, 15(2), pp.123-138.

[57] Barkauskas, C.E., Cronce, M.J., Rackley, C.R., Bowie, E.J., Keene, D.R., Stripp, B.R., Randell, S.H., Noble, P.W. and Hogan, B.L., 2013. Type 2 alveolar cells are stem cells in adult lung. *The Journal of clinical investigation*, 123(7), pp.3025-3036.

[58] Desai, T.J., Brownfield, D.G. and Krasnow, M.A., 2014. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature*, 507(7491), p.190.

[59] Rock, J.R., Barkauskas, C.E., Cronce, M.J., Xue, Y., Harris, J.R., Liang, J., Noble, P.W. and Hogan, B.L., 2011. Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. *Proceedings of the National Academy of Sciences*, 108(52), pp.E1475-E1483.

[60] Yang, J., Hernandez, B.J., Alanis, D.M., del Pilar, O.N., Vila-Ellis, L., Akiyama, H., Evans, S.E., Ostrin, E.J. and Chen, J., 2016. The development and plasticity of alveolar type 1 cells. *Development*, 143(1), pp.54-65.

[61] He, L., Vanlandewijck, M., Mäe, M.A., Andrae, J., Ando, K., Del Gaudio, F., Nahar, K., Lebouvier, T., Lavina, B., Gouveia, L. and Sun, Y., 2018. Single-cell RNA sequencing of mouse brain and lung vascular and vessel-associated cell types. *Scientific data*, 5, p.180160.

[62] Gonzales, J.N., and Verin, D.A. (October 24th 2018); Pulmonary Vascular Endothelial Cells, Endothelial Dysfunction - Old Concepts and New Challenges, Helena Lenasi, IntechOpen, DOI: 10.5772/intechopen.76995. Available from: https://www.intechopen.com/books/endothelial-dysfunction-old-concepts-and-new-challenges/pulmonary-vascular-endothelial-cells

[63] Millar, F.R., Summers, C., Griffiths, M.J., Toshner, M.R. and Proudfoot, A.G., 2016. The pulmonary endothelium in acute respiratory distress syndrome: insights and therapeutic opportunities. *Thorax*, 71(5), pp.462-473.

[64] Rounds, S., Lu, Q., Harrington, E.O., Newton, J. and Casserly, B., 2008. Pulmonary endothelial cell signaling and function. *Transactions of the American Clinical and Climatological Association*, 119, p.155.

[65] Kelly, J.J., Moore, T.M., Babal, P., Diwan, A.H., Stevens, T. and Thompson, W.J., 1998. Pulmonary microvascular and macrovascular endothelial cells: differential regulation of Ca2+ and permeability. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 274(5), pp.L810-L819.

[66] Kofler, S., Nickel, T. and Michael, W.E.I.S., 2005. Role of cytokines in cardiovascular diseases: a focus on endothelial responses to inflammation. *Clinical science*, 108(3), pp.205-213.

[67] Krishnaswamy, G., Kelley, J., Yerra, L., Smith, J.K. and Chi, D.S., 1999. Human endothelium as a source of multifunctional cytokines: molecular regulation and possible role in human disease. *Journal of interferon cytokine research*, 19(2), pp.91-104.

[68] Pries, A.R., Kuebler, W.M., 2006. Normal endothelium. *Handbook of Experimental Pharmacology*, (176 Pt 1):1-40

[69] Tomasek, J.J., Gabbiani, G., Hinz, B., Chaponnier, C. and Brown, R.A., 2002. Myofibroblasts and mechano-regulation of connective tissue remodelling. *Nature reviews Molecular cell biology*, 3(5), p.349.

[70] Farkas, L., Gauldie, J., Voelkel, N.F. and Kolb, M., 2011. Pulmonary hypertension and idiopathic pulmonary fibrosis: a tale of angiogenesis, apoptosis, and growth factors. *American journal of respiratory cell and molecular biology*, 45(1), pp.1-15.

[71] Desmoulière, A., Chaponnier, C. and Gabbiani, G., 2005. Tissue repair, contraction, and the

myofibroblast. *Wound repair and regeneration*, 13(1), pp.7-12.

[72] Hinz, B., Pittet, P., Smith-Clerc, J., Chaponnier, C. and Meister, J.J., 2004. Myofibroblast development is characterized by specific cell-cell adherens junctions. *Molecular biology of the cell*, 15(9), pp.4310-4320.

[73] Goffin, J.M., Pittet, P., Csucs, G., Lussi, J.W., Meister, J.J. and Hinz, B., 2006. Focal adhesion size controls tension-dependent recruitment of ?-smooth muscle actin to stress fibers. *J Cell Biol*, 172(2), pp.259-268.

[74] Strieter, R.M., Keeley, E.C., Hughes, M.A., Burdick, M.D. and Mehrad, B., 2009. The role of circulating mesenchymal progenitor cells (fibrocytes) in the pathogenesis of pulmonary fibrosis. *Journal of leukocyte biology*, 86(5), pp.1111-1118.

[75] Kim, K.K., Kugler, M.C., Wolters, P.J., Robillard, L., Galvez, M.G., Brumwell, A.N., Sheppard, D. and Chapman, H.A., 2006. Alveolar epithelial cell mesenchymal transition develops in vivo during pulmonary fibrosis and is regulated by the extracellular matrix. *Proceedings of the National Academy of Sciences*, 103(35), pp.13180-13185.

[76] Degryse, A.L., Tanjore, H., Xu, X.C., Polosukhin, V.V., Jones, B.R., McMahon, F.B., Gleaves, L.A., Blackwell, T.S. and Lawson, W.E., 2010. Repetitive intratracheal bleomycin models several features of idiopathic pulmonary fibrosis. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 299(4), pp.L442-L452.

[77] Kage, H. and Borok, Z., 2012. EMT and interstitial lung disease: a mysterious relationship. *Current opinion in pulmonary medicine*, 18(5), p.517.

[78] Hung, C., Linn, G., Chow, Y.H., Kobayashi, A., Mittelsteadt, K., Altemeier, W.A., Gharib, S.A., Schnapp, L.M. and Duffield, J.S., 2013. Role of lung pericytes and resident fibroblasts in the pathogenesis of pulmonary fibrosis. *American journal of respiratory and critical care medicine*, 188(7), pp.820-830.

[79] Balhara, J. and Gounni, A.S., 2012. The alveolar macrophages in asthma: a double-edged sword. *Mucosal immunology*, 5(6), p.605.

[80] Benoit, M., Desnues, B. and Mege, J.L., 2008. Macrophage polarization in bacterial infections. *The Journal of Immunology*, 181(6), pp.3733-3739.

[81] Haley, P.J., 2003. Species differences in the structure and function of the immune system. *Toxicology*, 188(1), pp.49-71.

[82] Wynn, T.A., Chawla, A. and Pollard, J.W., 2013. Macrophage biology in development, homeostasis and disease. *Nature*, 496(7446), p.445.

[83] Mills, C., 2012. M1 and M2 macrophages: oracles of health and disease. *Critical Review in Immunology*, 32(6).

[84] Martinez, F.O., Sica, A., Mantovani, A. and Locati, M., 2008. Macrophage activation and polarization. *Front Biosci*, 13(1), pp.453-461.

[85] Laskin, D.L., Sunil, V.R., Gardner, C.R. and Laskin, J.D., 2011. Macrophages and tissue injury: agents of defense or destruction?. *Annual review of pharmacology and toxicology*, 51, pp.267-288.

[86] Sica, A. and Mantovani, A., 2012. Macrophage plasticity and polarization: in vivo veritas. *The Journal of clinical investigation*, 122(3), pp.787-795.

[87] Mantovani, A., Biswas, S.K., Galdiero, M.R., Sica, A. and Locati, M., 2013. Macrophage plasticity and polarization in tissue repair and remodelling. *The Journal of pathology*, 229(2), pp.176-185.

[88] Day, J., Friedman, A. and Schlesinger, L.S., 2009. Modeling the immune rheostat of macrophages in the lung in response to infection. *Proceedings of the National Academy of Sciences*, 106(27), pp.11246-11251.

[89] Alber, A., Howie, S.E., Wallace, W.A. and Hirani, N., 2012. The role of macrophages in healing the wounded lung. *International journal of experimental pathology*, 93(4), pp.243-251.

[90] Chen, W.H., Toapanta, F.R., Shirey, K.A., Zhang, L., Giannelou, A., Page, C., Frieman, M.B., Vogel, S.N. and Cross, A.S., 2012. Potential role for

alternatively activated macrophages in the secondary bacterial infection during recovery from influenza. *Immunology letters*, 141(2), pp.227-234.

[91] Weissler, J.C., Lyons, C.R., Lipscomb, M.F. and Toews, G.B., 1986. Human pulmonary macrophages: functional comparison of cells obtained from whole lung and by bronchoalveolar lavage. *American Review of Respiratory Disease*, 133(3), pp.473-477.

[92] Lohmann-Matthes, M.L., Steinmuller, C. and Franke-Ullmann, G., 1994. Pulmonary macrophages. *European Respiratory Journal*, 7(9), pp.1678-1689.

[93] Franke-Ullmann, G., Pförtner, C., Walter, P., Steinmller, C., Lohmann-Matthes, M.L. and Kobzik, L., 1996. Characterization of murine lung interstitial macrophages in comparison with alveolar macrophages in vitro. *The Journal of Immunology*, 157(7), pp.3097-3104.

[94] Steinmuller, C., Franke-Ullmann, G., Lohmann-Matthes, M.L. and Emmendorffer, A., 2000. Local activation of nonspecific defense against a respiratory model infection by application of interferon-?: comparison between rat alveolar and interstitial lung macrophages. *American journal of respiratory cell and molecular biology*, 22(4), pp.481-490.

[95] Bedoret, D., Wallemacq, H., Marichal, T., Desmet, C., Calvo, F.Q., Henry, E., Closset, R., Dewals, B., Thielen, C., Gustin, P. and de Leval, L., 2009. Lung interstitial macrophages alter dendritic cell functions to prevent airway allergy in mice. *The Journal of clinical investigation*, 119(12), pp.3723-3738.

[96] Arias-Diaz, J., Garcia-Verdugo, I., Casals, C., Sanchez-Rico, N., Vara, E. and Balibrea, J.L., 2000. Effect of surfactant protein A (SP-A) on the production of cytokines by human pulmonary macrophages. *Shock*, 14(3), pp.300-306.

[97] Geiser, M., 2010. Update on macrophage clearance of inhaled micro-and nanoparticles. *Journal of aerosol medicine and pulmonary drug delivery*, 23(4), pp.207-217.

[98] Egan, D.F., Scanlan, C.L., Wilkins, R.L., Stoller, J.K., 1999. Egan's fundamentals of respiratory care. Mosby, St. Louis, USA.

[99] Zhang, Y., Luxon, B.A., Casola, A., Garofalo, R.P., Jamaluddin, M. and Brasier, A.R., 2001. Expression of respiratory syncytial virus-induced chemokine gene networks in lower airway epithelial cells revealed by cDNA microarrays. *Journal of virology*, 75(19), pp.9044-9058.

[100] Salem, H. and Katz, S.A., 2005. Inhalation toxicology. CRC press.

[101] Misharin, A.V., Morales-Nebreda, L., Mutlu, G.M., Budinger, G.S. and Perlman, H., 2013. Flow cytometric analysis of macrophages and dendritic cell subsets in the mouse lung. *American journal of respiratory cell and molecular biology*, 49(4), pp.503-510.

[102] Zaynagetdinov, R., Sherrill, T.P., Kendall, P.L., Segal, B.H., Weller, K.P., Tighe, R.M. and Blackwell, T.S., 2013. Identification of myeloid cell subsets in murine lungs using flow cytometry. *American journal of respiratory cell and molecular biology*, 49(2), pp.180-189.

[103] Lay, J.C., Alexis, N.E., Zeman, K.L., Peden, D.B. and Bennett, W.D., 2009. In vivo uptake of inhaled particles by airway phagocytes is enhanced in patients with mild asthma compared with normal volunteers. *Thorax*, 64(4), pp.313-320.

[104] Maus, U.A., Janzen, S., Wall, G., Srivastava, M., Blackwell, T.S., Christman, J.W., Seeger, W., Welte, T. and Lohmeyer, J., 2006. Resident alveolar macrophages are replaced by recruited monocytes in response to endotoxin-induced lung inflammation. *American journal of respiratory cell and molecular biology*, 35(2), pp.227-235.

[105] Knapp, S., Leemans, J.C., Florquin, S., Branger, J., Maris, N.A., Pater, J., van Rooijen, N. and van der Poll, T., 2003. Alveolar macrophages have a protective antiinflammatory role during murine pneumococcal pneumonia. *American journal of respiratory and critical care medicine*, 167(2), pp.171-179.

[106] Herold, S., Mayer, K. and Lohmeyer, J., 2011. Acute lung injury: how macrophages orches-

trate resolution of inflammation and tissue repair. *Frontiers in immunology*, 2, p.65.

[107] Janssen, W.J., Barthel, L., Muldrow, A., Oberley-Deegan, R.E., Kearns, M.T., Jakubzick, C. and Henson, P.M., 2011. Fas determines differential fates of resident and recruited macrophages during resolution of acute lung injury. *American journal of respiratory and critical care medicine*, 184(5), pp.547-560.

[108] Xiong, Z., Leme, A.S., Ray, P., Shapiro, S.D. and Lee, J.S., 2011. CX3CR1+ lung mononuclear phagocytes spatially confined to the interstitium produce TNF-alpha and IL-6 and promote cigarette smoke-induced emphysema. *The Journal of Immunology*, 186(5), pp.3206-3214.

[109] Landsman, L. and Jung, S., 2007. Lung macrophages serve as obligatory intermediate between blood monocytes and alveolar macrophages. *The Journal of Immunology*, 179(6), pp.3488-3494.

[110] Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., Jung, S. and Amit, I., 2014. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell*, 159(6), pp.1312-1326.

[111] Vuga, L.J., Milosevic, J., Pandit, K., Ben-Yehudah, A., Chu, Y., Richards, T., Sciurba, J., Myerburg, M., Zhang, Y., Parwani, A.V. and Gibson, K.F., 2013. Cartilage oligomeric matrix protein in idiopathic pulmonary fibrosis. *PloS one*, 8(12), p.e83120.

[112] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P., 2007. Molecular biology of the cell: Reference edition.

[113] Rozario, T. and DeSimone, D.W., 2010. The extracellular matrix in development and morphogenesis: a dynamic view. *Developmental biology*, 341(1), pp.126-140.

[114] Frantz, C., Stewart, K.M. and Weaver, V.M., 2010. The extracellular matrix at a glance. *J Cell Sci*, 123(24), pp.4195-4200.

[115] Vaccaro, C.A. and Brody, J.S., 1981. Structural features of alveolar wall basement membrane in the adult rat lung. *The Journal of cell biology*, 91(2), pp.427-437.

[116] Dunsmore, S.E., and Rannels, D.E. (1996). Extracellular matrix biology in the lung. *Am.J.Physiol*, 270: L3?27.

[117] Mammoto, T., Jiang, E., Jiang, A. and Mammoto, A., 2013. Extracellular matrix structure and tissue stiffness control postnatal lung development through the lipoprotein receptor?related protein 5/Tie2 signaling system. *American journal of respiratory cell and molecular biology*, 49(6), pp.1009-1018.

[118] Laurent, G.J. and Shapiro, S.D. eds., 2006. Encyclopedia of respiratory medicine (Vol. 3). *Academic Press*.

[119] Liu T., De Los Santos F.G., Phan S.H. (2017) The Bleomycin Model of Pulmonary Fibrosis. In: Rittié L. (eds) Fibrosis. *Methods in Molecular Biology*, vol 1627. Humana Press, New York, NY

[120] Kreuter, M., Herth, F.J., Wacker, M., Leidl, R., Hellmann, A., Pfeifer, M., Behr, J., Witt, S., Kauschka, D., Mall, M. and Günther, A., 2015. Exploring clinical and epidemiological characteristics of interstitial lung diseases: rationale, aims, and design of a nationwide prospective registry?the EXCITING-ILD registry. *BioMed research international*, 2015.

[121] Valeyre, D., Duchemann, B., Nunes, H., Uzunhan, Y. and Annesi-Maesano, I., 2014. Interstitial lung diseases. Respiratory epidemiology, pp.79-87.

[122] Wallace, W.A., Fitch, P.M., Simpson, A.J. and Howie, S.E., 2007. Inflammation?associated remodelling and fibrosis in the lung?a process and an end point. *International journal of experimental pathology*, 88(2), pp.103-110.

[123] Raghu, G., Collard, H.R., Egan, J.J., Martinez, F.J., Behr, J., Brown, K.K., Colby, T.V., Cordier, J.F., Flaherty, K.R., Lasky, J.A. and Lynch, D.A., 2011. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *American journal of respiratory and critical care medicine*, 183(6), pp.788-824.

[124] Richeldi, L., Du Bois, R.M., Raghu, G., Azuma, A., Brown, K.K., Costabel, U., Cottin, V., Flaherty, K.R., Hansell, D.M., Inoue, Y. and Kim, D.S.,

2014. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *New England Journal of Medicine*, 370(22), pp.2071-2082.

[125] King Jr, T.E., Bradford, W.Z., Castro-Bernardini, S., Fagan, E.A., Glaspole, I., Glassberg, M.K., Gorina, E., Hopkins, P.M., Kardatzke, D., Lancaster, L. and Lederer, D.J., 2014. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *New England Journal of Medicine*, 370(22), pp.2083-2092.

[126] GBD, M., 2013. causes of death collaborators. Global, regional, and national age?sex specific all-cause and cause-specific mortality for 240 causes of death, 1990?2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, 385(2013), pp.117-171.

[127] Butt, Y., Kurdowska, A. and Allen, T.C., 2016. Acute lung injury: a clinical and molecular review. *Archives of Pathology and Laboratory Medicine*, 140(4), pp.345-350.

[128] Meadors, M., Floyd, J. and Perry, M.C., 2006, February. Pulmonary toxicity of chemotherapy. *In Seminars in oncology*, 33(1), pp. 98-105.

[129] Pedersen-Bjergaard, J., Daugaard, G., Hansen, S.W., Rorth, M., Philip, P. and Larsen, S.O., 1991. Increased risk of myelodysplasia and leukaemia after etoposide, cisplatin, and bleomycin for germ-cell tumours. *The Lancet*, 338(8763), pp.359-363.

[130] Saxman, S.B., Nichols, C.R. and Einhorn, L.H., 1997. Pulmonary toxicity in patients with advanced-stage germ cell tumors receiving bleomycin with and without granulocyte colony stimulating factor. *Chest*, 111(3), pp.657-660.

[131] Jules-Elysee, K. and White, D.A., 1990. Bleomycin-induced pulmonary toxicity. *Clinics in chest medicine*, 11(1), pp.1-20.

[132] Sleijfer, S., 2001. Bleomycin-induced pneumonitis. *Chest*, 120(2), pp.617-624.

[133] O?sullivan, J.M., Huddart, R.A., Norman, A.R., Nicholls, J., Dearnaley, D.P. and Horwich, A., 2003. Predicting the risk of bleomycin lung toxicity in patients with germ-cell tumours. *Annals of Oncology*, 14(1), pp.91-96.

[134] Sikic, B.I., 1986. Biochemical and cellular determinants of bleomycin cytotoxicity. *Cancer surveys*, 5(1), pp.81-91.

[135] Chandler, D.B., 1990. Possible mechanisms of bleomycin-induced fibrosis. *Clinics in chest medicine*, 11(1), pp.21-30.

[136] Burkhardt, A., Gebbers, J.O. and Höltje, W.J., 1977. Die Bleomycin-Lunge. *DMW-Deutsche Medizinische Wochenschrift*, 102(08), pp.281-289.

[137] Tashiro, J., Rubio, G.A., Limper, A.H., Williams, K., Elliot, S.J., Ninou, I., Aidinis, V., Tzouvelekis, A. and Glassberg, M.K., 2017. Exploring animal models that resemble idiopathic pulmonary fibrosis. *Frontiers in medicine*, 4, p.118.

[138] Mouratis, M.A. and Aidinis, V., 2011. Modeling pulmonary fibrosis with bleomycin. *Current opinion in pulmonary medicine*, 17(5), pp.355-361.

[139] Schiller, H.B., Fernandez, I.E., Burgstaller, G., Schaab, C., Scheltema, R.A., Schwarzmayr, T., Strom, T.M., Eickelberg, O. and Mann, M., 2015. Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair. *Molecular systems biology*, 11(7).

[140] Cao, Z., Lis, R., Ginsberg, M., Chavez, D., Shido, K., Rabbany, S.Y., Fong, G.H., Sakmar, T.P., Rafii, S. and Ding, B.S., 2016. Targeting of the pulmonary capillary vascular niche promotes lung alveolar repair and ameliorates fibrosis. *Nature medicine*, 22(2), p.154.

[141] Degryse, A.L. and Lawson, W.E., 2011. Progress toward improving animal models for idiopathic pulmonary fibrosis. *The American journal of the medical sciences*, 341(6), pp.444-449.

[142] Barbayianni, I., Ninou, I., Tzouvelekis, A. and Aidinis, V., 2018. Bleomycin revisited. *Frontiers in medicine*, 5, p.269.

[143] Egger, C., Cannet, C., Gérard, C., Jarman, E., Jarai, G., Feige, A., Suply, T., Micard, A., Dunbar, A., Tigani, B. and Beckmann, N., 2013. Administration of bleomycin via the oropharyngeal aspiration route leads to sustained lung fibrosis in mice and rats as quantified by UTE-MRI and histology. *PloS one*, 8(5), p.e63432.

[144] Bivas-Benita, M., Zwier, R., Junginger, H.E. and Borchard, G., 2005. Non-invasive pulmonary aerosol delivery in mice by the endotracheal route. *European Journal of Pharmaceutics and Biopharmaceutics*, 61(3), pp.214-218.

[145] Robbe, A., Tassin, A., Carpentier, J., Declèves, A.E., Mekinda Ngono, Z.L., Nonclercq, D. and Legrand, A., 2015. Intratracheal bleomycin aerosolization: the best route of administration for a scalable and homogeneous pulmonary fibrosis rat model?. *BioMed research international*, 2015.

[146] Jenkins, R.G., Moore, B.B., Chambers, R.C., Eickelberg, O., Königshoff, M., Kolb, M., Laurent, G.J., Nanthakumar, C.B., Olman, M.A., Pardo, A. and Selman, M., 2017. An official American Thoracic Society workshop report: use of animal models for the preclinical assessment of potential therapies for pulmonary fibrosis. *American journal of respiratory cell and molecular biology*, 56(5), pp.667-679.

[147] Bale, S., Sunkoju, M., Reddy, S.S., Swamy, V. and Godugu, C., 2016. Oropharyngeal aspiration of bleomycin: An alternative experimental model of pulmonary fibrosis developed in Swiss mice. *Indian journal of pharmacology*, 48(6), p.643.

[148] Ruscitti, F., Ravanetti, F., Essers, J., Ridwan, Y., Belenkov, S., Vos, W., Ferreira, F., KleinJan, A., van Heijningen, P., Van Holsbeke, C. and Cacchioli, A., 2017. Longitudinal assessment of bleomycin-induced lung fibrosis by Micro-CT correlates with histological evaluation in mice. *Multidisciplinary respiratory medicine*, 12(1), p.8.

[149] Izbicki, G., Segel, M.J., Christensen, T.G., Conner, M.W. and Breuer, R., 2002. Time course of bleomycin?induced lung fibrosis. *International journal of experimental pathology*, 83(3), pp.111-119.

[150] Peng, R., Sridhar, S., Tyagi, G., Phillips, J.E., Garrido, R., Harris, P., Burns, L., Renteria, L., Woods, J., Chen, L. and Allard, J., 2013. Bleomycin induces molecular changes directly relevant to idiopathic pulmonary fibrosis: a model for ?active? disease. *PloS one*, 8(4), p.e59348.

[151] Lander, E.S., 2001. Initial sequencing and analysis of the human genome. International Human Genome Sequencing Consortium. *Nature*, 409, pp.860-921.

[152] Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. and Gocayne, J.D., 2001. The sequence of the human genome. *Science*, 291(5507), pp.1304-1351.

[153] Uhlèn, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjöstedt, E., Asplund, A. and Olsson, I., 2015. Tissue-based map of the human proteome. *Science*, 347(6220), p.1260419.

[154] Thul, P.J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Blal, H.A., Alm, T., Asplund, A., Björk, L., Breckels, L.M. and Bäckström, A., 2017. A subcellular map of the human proteome. *Science*, 356(6340), p.eaal3321.

[155] Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M. and Clevers, H., 2017. Science forum: the human cell atlas. *Elife*, 6, p.e27041.

[156] Li, L. and Clevers, H., 2010. Coexistence of quiescent and active adult stem cells in mammals. *Science*, 327(5965), pp.542-545.

[157] Huang, S., 2009. Non-genetic heterogeneity of cells in development: more than just noise. *Development*, 136(23), pp.3853-3862.

[158] Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N. and Schwartz, S., 2014. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature*, 510(7505), p.363.

[159] Eldar, A. and Elowitz, M.B., 2010. Functional roles for noise in genetic circuits. *Nature*, 467(7312), p.167.

[160] Maamar, H., Raj, A. and Dubnau, D., 2007. Noise in gene expression determines cell fate in Bacillus subtilis. *Science*, 317(5837), pp.526-529.

[161] Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. and Luscombe, N.M., 2009. A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics*, 10(4), p.252.

[162] Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M. and Coleman, P., 1992. Analysis of gene expression in single live neurons. *Proceedings of the National Academy of Sciences*, 89(7), pp.3010-3014.

[163] Brady, G., Barbara, M. and Iscove, N.N., 1990. Representative in vitro cDNA amplification from individual hemopoietic cells and colonies. *Methods Mol Cell Biol*, 2(1), pp.17-25.

[164] Klein, C.A., Seidl, S., Petat-Dutter, K., Offner, S., Geigl, J.B., Schmidt-Kittler, O., Wendler, N., Passlick, B., Huber, R.M., Schlimok, G. and Baeuerle, P.A., 2002. Combined transcriptome and genome analysis of single micrometastatic cells. *Nature biotechnology*, 20(4), p.387.

[165] Kurimoto, K., Yabuta, Y., Ohinata, Y., Ono, Y., Uno, K.D., Yamada, R.G., Ueda, H.R. and Saitou, M., 2006. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic acids research*, 34(5), pp.e42-e42.

[166] Becker-Andre, M. and Hahlbrock, K., 1989. Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a P CR aided transcipt titration assay (PATTY). *Nucleic acids research*, 17(22), pp.9437-9446.

[167] Weis, J.H., Tan, S.S., Martin, B.K. and Wittwer, C.T., 1992. Detection of rare mRNAs via quantitative RT-PCR. *Trends in genetics: TIG*, 8(8), p.263.

[168] Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K. and Sidow, A., 2008. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. *Genome research*, 18(7), pp.1051-1063.

[169] Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z. and Dewell, S.B., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), p.376.

[170] Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, 5(7), p.621.

[171] Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M., 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881), pp.1344-1349.

[172] Wilhelm, B.T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C.J., Rogers, J. and Ber, J., 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, 453(7199), p.1239.

[173] Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M. and Gilad, Y., 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research*, 18(9), pp.1509-1517.

[174] Svensson, V., Vento-Tormo, R. and Teichmann, S.A., 2018. Exponential scaling of single-cell RNA-seq in the past decade. *Nature protocols*, 13(4), p.599.

[175] Choy, J.Y., Boon, P.L., Bertin, N. and Fullwood, M.J., 2015. A resource of ribosomal RNA-depleted RNA-Seq data from different normal adult and fetal human tissues. *Scientific data*, 2, p.150063.

[176] Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P. and Linnarsson, S., 2014. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature methods*, 11(2), p.163.

[177] Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.B., Lönnerberg, P. and Linnarsson, S., 2011. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome research*, 21(7), pp.1160-1167.

[178] Hashimshony, T., Wagner, F., Sher, N. and Yanai, I., 2012. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell reports*, 2(3), pp.666-673.

[179] Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M. and Trombetta, J.J., 2015. Highly parallel genome-wide

expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5), pp.1202-1214.

[180] Grün, D., Kester, L. and Van Oudenaarden, A., 2014. Validation of noise models for single-cell transcriptomics. *Nature methods*, 11(6), p.637.

[181] Deng, Q., Ramsköld, D., Reinius, B. and Sandberg, R., 2014. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science*, 343(6167), pp.193-196.

[182] Kolodziejczyk, A.A., Kim, J.K., Tsang, J.C., Ilicic, T., Henriksson, J., Natarajan, K.N., Tuck, A.C., Gao, X., Bühler, M., Liu, P. and Marioni, J.C., 2015. Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell stem cell*, 17(4), pp.471-485.

[183] Martinez-Jimenez, C.P., Eling, N., Chen, H.C., Vallejos, C.A., Kolodziejczyk, A.A., Connor, F., Stojic, L., Rayner, T.F., Stubbington, M.J., Teichmann, S.A. and de la Roche, M., 2017. Aging increases cell-to-cell transcriptional variability upon immune stimulation. *Science*, 355(6332), pp.1433-1436.

[184] Halpern, K.B., Shenhav, R., Matcovitch-Natan, O., Tóth, B., Lemze, D., Golan, M., Massasa, E.E., Baydatch, S., Landen, S., Moor, A.E. and Brandis, A., 2017. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature*, 542(7641), p.352.

[185] Muraro, M.J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning, E.J. and van Oudenaarden, A., 2016. A single-cell transcriptome atlas of the human pancreas. *Cell systems*, 3(4), pp.385-394.

[186] Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., Ryu, J.H., Wagner, B.K., Shen-Orr, S.S., Klein, A.M. and Melton, D.A., 2016. A single-cell transcriptomic map of the human and mouse pancreas reveals inter-and intra-cell population structure. *Cell systems*, 3(4), pp.346-360.

[187] Villani, A.C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S. and Jardine, L., 2017. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*, 356(6335), p.eaah4573.

[188] Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A. and Amit, I., 2014. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science*, 343(6172), pp.776-779.

[189] Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A. and Quake, S.R., 2014. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature*, 509(7500), p.371.

[190] Plasschaert, L.W., Zilionis, R., Choo-Wing, R., Savova, V., Knehr, J., Roma, G., Klein, A.M. and Jaffe, A.B., 2018. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature*, 560(7718), p.377.

[191] Shekhar, K., Lapan, S.W., Whitney, I.E., Tran, N.M., Macosko, E.Z., Kowalczyk, M., Adiconis, X., Levin, J.Z., Nemesh, J., Goldman, M. and McCarroll, S.A., 2016. Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell*, 166(5), pp.1308-1323.

[192] Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T. and Bertagnolli, D., 2016. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature neuroscience*, 19(2), p.335.

[193] Fuzik, J., Zeisel, A., Máté, Z., Calvigioni, D., Yanagawa, Y., Szabó, G., Linnarsson, S. and Harkany, T., 2016. Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. *Nature biotechnology*, 34(2), p.175.

[194] Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lönnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggström, J., Kharchenko, O., Kharchenko, P.V. and Linnarsson, S., 2015. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nature neuroscience*, 18(1), p.145.

[195] Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C. and Rolny, C., 2015. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*, 347(6226), pp.1138-1142.

[196] Lake, B.B., Ai, R., Kaeser, G.E., Salathia, N.S., Yung, Y.C., Liu, R., Wildberg, A., Gao, D., Fung, H.L., Chen, S. and Vijayaraghavan, R., 2016. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science*, 352(6293), pp.1586-1590.

[197] Braga, F.A.V., Kar, G., Berg, M., Carpaij, O.A., Polanski, K., Simon, L.M., Brouwer, S., Gomes, T., Hesse, L., Jiang, J. and Fasouli, E.S., 2019. A cellular census of human lungs identifies novel cell states in health and in asthma. *Nature medicine*, p.1.

[198] Strunz, M., Simon, L.M., Ansari, M., Mattner, L.F., Angelidis, I., Mayr, C.H., Kathiriya, J., Yee, M., Ogar, P., Sengupta, A. and Kukhtevich, I., 2019. Longitudinal single cell transcriptomics reveals Krt8+ alveolar epithelial progenitors in lung regeneration. *bioRxiv*, p.705244.

[199] Faustino, V., Catarino, S.O., Lima, R. and Minas, G., 2016. Biomedical microfluidic devices by using low-cost fabrication techniques: A review. *Journal of biomechanics*, 49(11), pp.2280-2292.

[200] Kim, S.H., Cui, Y., Lee, M.J., Nam, S.W., Oh, D., Kang, S.H., Kim, Y.S. and Park, S., 2011. Simple fabrication of hydrophilic nanochannels using the chemical bonding between activated ultrathin PDMS layer and cover glass by oxygen plasma. *Lab on a Chip*, 11(2), pp.348-353.

[201] Angelidis, I., Simon, L.M., Fernandez, I.E., Strunz, M., Mayr, C.H., Greiffo, F.R., Tsitsiridis, G., Ansari, M., Graf, E., Strom, T.M., Nagendran, M., ... , Theis, F.J., and Schiller, H.B., 2019. An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. *Nature communications*, 10(1), p.963.

[202] Zimmerman, E., 2014. 50 Smartest Companies: Illumina. *MIT Technology Review*. Available at: https://www.technologyreview.com/s/524531/why-illumina-is-no-1/ [Accessed July 27, 2019]

[203] Illumina, 2015. Patterned Flow Cell Technology. *Illumina Technology Whitepaper*. Available at: https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/patterned-flow-cell-technology-technical-note-770-2015-010.pdf [Accessed July 26, 2019].

[204] Illumina, 2015. HiSeq 3000/HiSeq 4000 Sequencing Systems. *Illumina Specification Whitepaper*. Available at: https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/hiseq-3000-4000-specification-sheet-770-2014-057.pdf [Accessed July 26, 2019].

[205] Fedurco, M., Romieu, A., Williams, S., Lawrence, I. and Turcatti, G., 2006. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic acids research*, 34(3), pp.e22-e22.

[206] Turcatti, G., Romieu, A., Fedurco, M. and Tairi, A.P., 2008. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic acids research*, 36(4), pp.e25-e25.

[207] Ross, M.G., Russ, C., Costello, M., Hollinger, A., Lennon, N.J., Hegarty, R., Nusbaum, C. and Jaffe, D.B., 2013. Characterizing and measuring bias in sequence data. Genome biology, 14(5), p.R51.

[208] Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. and Law, M., 2012. Comparison of next-generation sequencing systems. *BioMed Research International*, 2012.

[209] Kircher, M., Stenzel, U. and Kelso, J., 2009. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome biology*, 10(8), p.R83.

[210] Illumina, 2019. bcl2fastq2 Conversion Software v2.20. *Illumina Software Guide*. Available at: https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/bcl2fastq/bcl2fastq2-v2-20-software-guide-15051736-03.pdf [Accessed August 1, 2019].

[211] Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M.

and Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15?21.

[212] Satija, R., Farrell, J.A., Gennert, D., Schier, A.F. and Regev, A., 2015. Spatial reconstruction of single-cell gene expression data. *Nature biotechnology*, 33(5), p.495.

[213] Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology*, 36(5), p.411.

[214] Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck III, W.M., Hao, Y., Stoeckius, M., Smibert, P. and Satija, R., 2019. Comprehensive Integration of Single-Cell Data. *Cell*.

[215] Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, 43(7), pp.e47-e47.

[216] Phipson, B., Lee, S., Majewski, I.J., Alexander, W.S. and Smyth, G.K., 2016. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *The annals of applied statistics*, 10(2), p.946.

[217] Young, M.D. and Behjati, S., 2018. SoupX removes ambient RNA contamination from droplet based single cell RNA sequencing data. *bioRxiv*, p.303727.

[218] Bruce, P. and Bruce, A., 2017. Practical statistics for data scientists: 50 essential concepts. O'Reilly Media, Inc..

[219] Ester, M., Kriegel, H.P., Sander, J. and Xu, X., 1996, August. A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd (Vol. 96, No. 34, pp. 226-231).

[220] Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E. and Dudoit, S., 2018. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics*, 19(1), p.477.

[221] Wolf, F.A., Angerer, P. and Theis, F.J., 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome biology*, 19(1), p.15.

[222] Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L. and Theis, F.J., 2019. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome biology*, 20(1), p.59.

[223] La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A. and Fan, J., 2018. RNA velocity of single cells. *Nature*, 560(7719), p.494.

[224] Haghverdi, L., Buettner, M., Wolf, F.A., Buettner, F. and Theis, F.J., 2016. Diffusion pseudotime robustly reconstructs lineage branching. *Nature methods*, 13(10), p.845.

[225] Park, J.E., Polański, K., Meyer, K. and Teichmann, S.A., 2018. Fast batch alignment of single cell transcriptomes unifies multiple mouse cell atlases into an integrated landscape. *bioRxiv*, p.397042.

[226] Kowalczyk, M.S., Tirosh, I., Heckl, D., Rao, T.N., Dixit, A., Haas, B.J., Schneider, R.K., Wagers, A.J., Ebert, B.L. and Regev, A., 2015. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome research*, 25(12), pp.1860-1872.

[227] McInnes, L., Healy, J. and Melville, J., 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint* arXiv:1802.03426.

[228] Law, C.W., Chen, Y., Shi, W. and Smyth, G.K., 2014. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology*, 15(2), p.R29.

[229] Ramilowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V.P., Itoh, M., Kawaji, H., Carninci, P., Rost, B. and Forrest, A.R., 2015. A draft network of ligand?receptor-mediated multicellular signalling in human. *Nature communications*, 6, p.7866.

[230] Law, C.W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G.K. and Ritchie, M.E., 2016. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research*, 5.

[231] Misharin, A.V., Morales-Nebreda, L., Reyfman, P.A., Cuda, C.M., Walter, J.M., McQuattie-Pimentel, A.C., Chen, C.I., Anekalla, K.R., Joshi, N., Williams, K.J. and Abdala-Valencia, H., 2017. Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *Journal of Experimental Medicine*, 214(8), pp.2387-2404.

[232] Chambers, J.M., Hastie, T.J., Statistical Models in S, 1991, eds 1 CRC Press.

[233] Hastie, T., 2008. gam: Generalized additive models. R package version, 1(2).

[234] Logan, C.Y. and Desai, T.J., 2015. Keeping it together: pulmonary alveoli are maintained by a hierarchy of cellular programs. *BioEssays*, 37(9), pp.1028-1037.

[235] Herriges, M. and Morrisey, E.E., 2014. Lung development: orchestrating the generation and regeneration of a complex organ. *Development*, 141(3), pp.502-513.

[236] Tata, P.R. and Rajagopal, J., 2017. Plasticity in the lung: making and breaking cell identity. *Development*, 144(5), pp.755-766.

[237] Lodyga, M., Cambridge, E., Karvonen, H.M., Pakshir, P., Wu, B., Boo, S., Kiebalo, M., Kaarteenaho, R., Glogauer, M., Kapoor, M. and Ask, K., 2019. Cadherin-11-mediated adhesion of macrophages to myofibroblasts establishes a profibrotic niche of active TGF-?. *Sci. Signal.*, 12(564), p.eaao3469.

[238] Gieseck III, R.L., Wilson, M.S. and Wynn, T.A., 2018. Type 2 immunity in tissue repair and fibrosis. *Nature Reviews Immunology*, 18(1), p.62.

[239] Vaughan, A.E., Brumwell, A.N., Xi, Y., Gotts, J.E., Brownfield, D.G., Treutlein, B., Tan, K., Tan, V., Liu, F.C., Looney, M.R. and Matthay, M.A., 2015. Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. *Nature*, 517(7536), p.621.

[240] Zuo, W., Zhang, T., Wu, D.Z.A., Guan, S.P., Liew, A.A., Yamamoto, Y., Wang, X., Lim, S.J., Vincent, M., Lessard, M. and Crum, C.P., 2015. p63+ Krt5+ distal airway stem cells are essential for lung regeneration. *Nature*, 517(7536), p.616.

[241] Zacharias, W.J., Frank, D.B., Zepp, J.A., Morley, M.P., Alkhaleel, F.A., Kong, J., Zhou, S., Cantu, E. and Morrisey, E.E., 2018. Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature*, 555(7695), p.251.

[242] Xi, Y., Kim, T., Brumwell, A.N., Driver, I.H., Wei, Y., Tan, V., Jackson, J.R., Xu, J., Lee, D.K., Gotts, J.E. and Matthay, M.A., 2017. Local lung hypoxia determines epithelial fate decisions during alveolar regeneration. *Nature cell biology*, 19(8), p.904.

[243] Liu, Q., Liu, K., Cui, G., Huang, X., Yao, S., Guo, W., Qin, Z., Li, Y., Yang, R., Pu, W. and Zhang, L., 2019. Lung regeneration by multipotent stem cells residing at the bronchioalveolar-duct junction. *Nature genetics*, 51(4), p.728.

[244] Katsura, H., Kobayashi, Y., Tata, P.R. and Hogan, B.L., 2019. IL-1 and TNF? Contribute to the Inflammatory Niche to Enhance Alveolar Regeneration. *Stem cell reports*, 12(4), pp.657-666.

[245] LaCanna, R., Liccardo, D., Zhang, P., Tragesser, L., Wang, Y., Cao, T., Chapman, H.A., Morrisey, E.E., Shen, H., Koch, W.J. and Kosmider, B., 2019. Yap/Taz regulate alveolar regeneration and resolution of lung inflammation. *The Journal of clinical investigation*, 129(5).

[246] Riemondy, K.A., Jansing, N.L., Jiang, P., Redente, E.F., Gillen, A.E., Fu, R., Miller, A.J., Spence, J.R., Gerber, A.N., Hesselberth, J.R. and Zemans, R.L., 2019. Single-cell RNA sequencing identifies TGF-? as a key regenerative cue following LPS-induced lung injury. *JCI insight*, 4(8).

[247] Saelens, W., Cannoodt, R., Todorov, H. and Saeys, Y., 2019. A comparison of single-cell trajectory inference methods. *Nature biotechnology*, 37(5), p.547.

[248] Ludwig, L.S., Lareau, C.A., Ulirsch, J.C., Christian, E., Muus, C., Li, L.H., Pelka, K., Ge, W., Oren, Y., Brack, A. and Law, T., 2019. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell*, 176(6), pp.1325-1339.

[249] Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J.,

Liu, S., Lin, S., Berube, P. and Lee, L., 2019. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4), pp.928-943.

[250] Fischer, D.S., Fiedler, A.K., Kernfeld, E.M., Genga, R.M., Bastidas-Ponce, A., Bakhti, M., Lickert, H., Hasenauer, J., Maehr, R. and Theis, F.J., 2019. Inferring population dynamics from single-cell RNA-sequencing time series data. *Nature biotechnology*, 37(4), p.461.

[251] Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B., Park, J.E., Stephenson, E., Polański, K., Goncalves, A. and Gardner, L., 2018. Single-cell reconstruction of the early maternal?fetal interface in humans. *Nature*, 563(7731), p.347.

[252] Regev, A., Teichmann, S., Rozenblatt-Rosen, O., Stubbington, M., Ardlie, K., Amit, I., Arlotta, P., Bader, G., Benoist, C., Biton, M. and Bodenmiller, B., 2018. The Human Cell Atlas White Paper. *arXiv preprint* arXiv:1810.05192.

[253] Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W., Ng, L.G., Ginhoux, F. and Newell, E.W., 2019. Dimensionality reduction for visualizing single-cell data using UMAP. *Nature biotechnology*, 37(1), p.38.

[254] Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F. and Huang, D., 2018. Mapping the mouse cell atlas by microwell-seq. *Cell*, 172(5), pp.1091-1107.

[255] El Agha, E., Moiseenko, A., Kheirollahi, V., De Langhe, S., Crnkovic, S., Kwapiszewska, G., Szibor, M., Kosanovic, D., Schwind, F., Schermuly, R.T. and Henneke, I., 2017. Two-way conversion between lipogenic and myogenic fibroblastic phenotypes marks the progression and resolution of lung fibrosis. *Cell stem cell*, 20(2), pp.261-273.

[256] Homma, S., Nagaoka, I., Abe, H., Takahashi, K., Seyama, K., Nukiwa, T. and Kira, S., 1995. Localization of platelet-derived growth factor and insulin-like growth factor I in the fibrotic lung. *American journal of respiratory and critical care medicine*, 152(6), pp.2084-2089.

[257] Rennard, S.I., Hunninghake, G.W., Bitterman, P.B. and Crystal, R.G., 1981. Production of fibronectin by the human alveolar macrophage: mechanism for the recruitment of fibroblasts to sites of tissue injury in interstitial lung diseases. *Proceedings of the National Academy of Sciences*, 78(11), pp.7147-7151.

[258] Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R. and Butte, A.J., 2019. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature immunology*, 20(2), p.163.

[259] Yee, M., Domm, W., Gelein, R., Bentley, K.L.D.M., Kottmann, R.M., Sime, P.J., Lawrence, B.P. and O'Reilly, M.A., 2017. Alternative progenitor lineages regenerate the adult lung depleted of alveolar epithelial type 2 cells. *American journal of respiratory cell and molecular biology*, 56(4), pp.453-464.

[260] Buczynski, B.W., Yee, M., Martin, K.C., Lawrence, B.P. and O'Reilly, M.A., 2013. Neonatal hyperoxia alters the host response to influenza A virus infection in adult mice through multiple pathways. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 305(4), pp.L282-L290.

[261] O'Reilly, M.A., Marr, S.H., Yee, M., McGrath-Morrow, S.A. and Lawrence, B.P., 2008. Neonatal hyperoxia enhances the inflammatory response in adult mice infected with influenza A virus. *American journal of respiratory and critical care medicine*, 177(10), pp.1103-1110.

[262] Nabhan, A.N., Brownfield, D.G., Harbury, P.B., Krasnow, M.A. and Desai, T.J., 2018. Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science*, 359(6380), pp.1118-1123.

[263] El-Sukkari, D., Wilson, N.S., Hakansson, K., Steptoe, R.J., Grubb, A., Shortman, K. and Villadangos, J.A., 2003. The protease inhibitor cystatin C is differentially expressed among dendritic cell populations, but does not control antigen presentation. *The Journal of Immunology*, 171(10), pp.5003-5011.

[264] Tatapudy, S., Aloisio, F., Barber, D. and Nystul, T., 2017. Cell fate decisions: emerging roles for metabolic signals and cell morphology. *EMBO reports*, 18(12), pp.2105-2118.

[265] Yuan, B., Li, C., Kimura, S., Engelhardt, R.T., Smith, B.R. and Minoo, P., 2000. Inhibition of distal lung morphogenesis in Nkx2. 1 (?/?) embryos. *Developmental dynamics: an official publication of the American Association of Anatomists*, 217(2), pp.180-190.

[266] Li, S., Morley, M., Lu, M., Zhou, S., Stewart, K., French, C.A., Tucker, H.O., Fisher, S.E. and Morrisey, E.E., 2016. Foxp transcription factors suppress a non-pulmonary gene expression program to permit proper lung development. *Developmental biology*, 416(2), pp.338-346.

[267] Shu, W., Lu, M.M., Zhang, Y., Tucker, P.W., Zhou, D. and Morrisey, E.E., 2007. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development*, 134(10), pp.1991-2000.

[268] Roos, A.B., Berg, T., Barton, J.L., Didon, L. and Nord, M., 2012. Airway epithelial cell differentiation during lung organogenesis requires C/EBPalpha and C/EBPbeta. *Developmental Dynamics*, 241(5), pp.911-923.

[269] Martis, P.C., Whitsett, J.A., Xu, Y., Perl, A.K.T., Wan, H. and Ikegami, M., 2006. C/EBPalpha is required for lung maturation at birth. *Development*, 133(6), pp.1155-1164.

[270] Basseres, D.S., Levantini, E., Ji, H., Monti, S., Elf, S., Dayaram, T., Fenyus, M., Kocher, O., Golub, T., Wong, K.K. and Halmos, B., 2006. Respiratory failure due to differentiation arrest and expansion of alveolar cells following lung-specific loss of the transcription factor C/EBPalpha in mice. *Molecular and cellular biology*, 26(3), pp.1109-1123.

[271] Zhang, Z., Newton, K., Kummerfeld, S.K., Webster, J., Kirkpatrick, D.S., Phu, L., Eastham-Anderson, J., Liu, J., Lee, W.P., Wu, J. and Li, H., 2017. Transcription factor Etv5 is essential for the maintenance of alveolar type II cells. *Proceedings of the National Academy of Sciences*, 114(15), pp.3903-3908.

[272] Yang, H., Lu, M.M., Zhang, L., Whitsett, J.A. and Morrisey, E.E., 2002. GATA6 regulates differentiation of distal lung epithelium. *Development*, 129(9), pp.2233-2246.

[273] Cheung, W.K., Zhao, M., Liu, Z., Stevens, L.E., Cao, P.D., Fang, J.E., Westbrook, T.F. and Nguyen, D.X., 2013. Control of alveolar differentiation by the lineage transcription factors GATA6 and HOPX inhibits lung adenocarcinoma metastasis. *Cancer cell*, 23(6), pp.725-738.

[274] Tiwari, N., Tiwari, V.K., Waldmeier, L., Balwierz, P.J., Arnold, P., Pachkov, M., Meyer-Schaller, N., Schübeler, D., van Nimwegen, E. and Christofori, G., 2013. Sox4 is a master regulator of epithelial-mesenchymal transition by controlling Ezh2 expression and epigenetic reprogramming. *Cancer cell*, 23(6), pp.768-783.

[275] MacDonald, B.T., Tamai, K. and He, X., 2009. Wnt/Beta-catenin signaling: components, mechanisms, and diseases. *Developmental cell*, 17(1), pp.9-26.

[276] Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S. and Rinn, J.L., 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology*, 32(4), p.381.

[277] Tritschler, S., Büttner, M., Fischer, D.S., Lange, M., Bergen, V., Lickert, H. and Theis, F.J., 2019. Concepts and limitations for learning developmental trajectories from single cell genomics. *Development*, 146(12), p.dev170506.

[278] Mutze, K., Vierkotten, S., Milosevic, J., Eickelberg, O. and Königshoff, M., 2015. Enolase 1 (ENO1) and protein disulfide-isomerase associated 3 (PDIA3) regulate Wnt/beta-catenin-driven trans-differentiation of murine alveolar epithelial cells. *Disease models mechanisms*, 8(8), pp.877-890.

[279] Sun, T., Huang, Z., Zhang, H., Posner, C., Jia, G., Ramalingam, T.R., Xu, M., Brightbill, H., Egen, J.G., Dey, A. and Arron, J.R., 2019. TAZ is required for lung alveolar epithelial cell differentiation after injury. *JCI insight*, 4(14).

[280] Abraham, E., 2003. Neutrophils and acute lung injury. *Critical care medicine*, 31(4), pp.S195-S199.

[281] Kotton, D.N. and Morrisey, E.E., 2014. Lung regeneration: mechanisms, applications and emerging stem cell populations. *Nature medicine*, 20(8), p.822.

[282] Gonzalez-Lopez, A., Astudillo, A., Garcia-Prieto, E., Fernandez-Garcia, M.S., Lopez-Vazquez, A., Batalla-Solis, E., Taboada, F., Fueyo, A. and Albaiceta, G.M., 2011. Inflammation and matrix remodeling during repair of ventilator-induced lung injury. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 301(4), pp.L500-L509.

[283] Gonzalez-Lopez, A. and Albaiceta, G.M., 2012. Repair after acute lung injury: molecular mechanisms and therapeutic opportunities. *Critical Care*, 16(2), p.209.

[284] Korpos, E., Wu, C. and Sorokin, L., 2009. Multiple roles of the extracellular matrix in inflammation. *Current pharmaceutical design*, 15(12), pp.1349-1357.

[285] Noble, P.W., Barkauskas, C.E. and Jiang, D., 2012. Pulmonary fibrosis: patterns and perpetrators. *The Journal of clinical investigation*, 122(8), pp.2756-2762.

[286] Henderson, N.C., Arnold, T.D., Katamura, Y., Giacomini, M.M., Rodriguez, J.D., McCarty, J.H., Pellicoro, A., Raschperger, E., Betsholtz, C., Ruminski, P.G. and Griggs, D.W., 2013. Targeting of av-integrin identifies a core molecular pathway that regulates fibrosis in several organs. *Nature medicine*, 19(12), p.1617.

[287] Thannickal, V.J., Zhou, Y., Gaggar, A. and Duncan, S.R., 2014. Fibrosis: ultimate and proximate causes. *The Journal of clinical investigation*, 124(11), pp.4673-4677.

[288] Armulik, A., Genové, G. and Betsholtz, C., 2011. Pericytes: developmental, physiological, and pathological perspectives, problems, and promises. *Developmental cell*, 21(2), pp.193-215.

[289] Bhatia, M. and Moochhala, S., 2004. Role of inflammatory mediators in the pathophysiology of acute respiratory distress syndrome. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, 202(2), pp.145-156.

[290] Lindsay, C.D., 2011. Novel therapeutic strategies for acute lung injury induced by lung damaging agents: the potential role of growth factors as treatment options. *Human experimental toxicology*, 30(7), pp.701-724.

[291] Bissell, M.J., Kenny, P.A. and Radisky, D.C., 2005, January. Microenvironmental regulators of tissue structure and function also regulate tumor induction and progression: the role of extracellular matrix and its degrading enzymes. *In Cold Spring Harbor symposia on quantitative biology* (Vol. 70, pp. 343-356). Cold Spring Harbor Laboratory Press.

[292] Marastoni, S., Ligresti, G., Lorenzon, E., Colombatti, A. and Mongiat, M., 2008. Extracellular matrix: a matter of life and death. *Connective tissue research*, 49(3-4), pp.203-206.

[293] Hynes, R.O., 2009. The extracellular matrix: not just pretty fibrils. *Science*, 326(5957), pp.1216-1219.

[294] Cukierman, E. and Bassi, D.E., 2010, June. Physico-mechanical aspects of extracellular matrix influences on tumorigenic behaviors. *In Seminars in cancer biology* (Vol. 20, No. 3, pp. 139-145). Academic Press.

[295] Kleinman, H.K., Philp, D. and Hoffman, M.P., 2003. Role of the extracellular matrix in morphogenesis. *Current opinion in biotechnology*, 14(5), pp.526-532.

[296] Colombatti, A., Spessotto, P., Doliana, R., Mongiat, M., Bressan, G.M. and Esposito, G., 2012. The EMILIN/multimerin family. *Frontiers in immunology*, 2, p.93.

[297] Doliana, R., Bot, S., Mungiguerra, G., Canton, A., Cilli, S.P. and Colombatti, A., 2001. Isolation and characterization of EMILIN-2, a new component of the growing EMILINs family and a member of the EMI domain-containing superfamily. *Journal of Biological Chemistry*, 276(15), pp.12003-12011.

[298] Mongiat, M., Ligresti, G., Marastoni, S., Lorenzon, E., Doliana, R. and Colombatti, A., 2007. Regulation of the extrinsic apoptotic pathway by the extracellular matrix glycoprotein EMILIN2. *Molecular and cellular biology*, 27(20), pp.7176-7187.

[299] Mongiat, M., Marastoni, S., Ligresti, G., Lorenzon, E., Schiappacassi, M., Perris, R., Frustaci,

S. and Colombatti, A., 2010. The extracellular matrix glycoprotein elastin microfibril interface located protein 2: a dual role in the tumor microenvironment. *Neoplasia* (New York, NY), 12(4), p.294.

[300] Paulitti, A., Andreuzzi, E., Bizzotto, D., Pellicani, R., Tarticchio, G., Marastoni, S., Pastrello, C., Jurisica, I., Ligresti, G., Bucciotti, F. and Doliana, R., 2018. The ablation of the matricellular protein EMILIN2 causes defective vascularization due to impaired EGFR-dependent IL-8 production affecting tumor growth. *Oncogene*, 37(25), p.3399.

[301] Marastoni, S., Andreuzzi, E., Paulitti, A., Colladel, R., Pellicani, R., Todaro, F., Schiavinato, A., Bonaldo, P., Colombatti, A. and Mongiat, M., 2014. EMILIN2 down?modulates the Wnt signalling pathway and suppresses breast cancer cell growth and migration. *The Journal of pathology*, 232(4), pp.391-404.

[302] Schiavinato, A., Keene, D.R., Wohl, A.P., Corallo, D., Colombatti, A., Wagener, R., Paulsson, M., Bonaldo, P. and Sengle, G., 2016. Targeting of EMILIN-1 and EMILIN-2 to fibrillin microfibrils facilitates their incorporation into the extracellular matrix. *Journal of Investigative Dermatology*, 136(6), pp.1150-1160.

[303] Schmitz, B., Radbruch, A., Kümmel, T., Wickenhauser, C., Korb, H., Hansmann, M.L., Thiele, J. and Fischer, R., 1994. Magnetic activated cell sorting (MACS)?a new immunomagnetic method for megakaryocytic cell isolation: comparison of different separation techniques. *European journal of haematology*, 52(5), pp.267-275.

[304] Miltenyi, S., Müller, W., Weichel, W. and Radbruch, A., 1990. High gradient magnetic cell separation with MACS. *Cytometry: The Journal of the International Society for Analytical Cytology*, 11(2), pp.231-238.

[305] Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. and Tinevez, J.Y., 2012. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7), p.676.

[306] Uhl, F.E., Vierkotten, S., Wagner, D.E., Burgstaller, G., Costa, R., Koch, I., Lindner, M., Meiners, S., Eickelberg, O. and Königshoff, M., 2015. Preclinical validation and imaging of Wnt-induced repair in human 3D lung tissue cultures. *European Respiratory Journal*, 46(4), pp.1150-1166.

[307] Burgstaller, G., Sengupta, A., Vierkotten, S., Preissler, G., Lindner, M., Behr, J., Königshoff, M. and Eickelberg, O., 2018. Distinct niches within the extracellular matrix dictate fibroblast function in (cell free) 3D lung tissue cultures. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 314(5), pp.L708-L723.

[308] Yee, M., Gelein, R., Mariani, T.J., Lawrence, B.P. and O'Reilly, M.A., 2016. The oxygen environment at birth specifies the population of alveolar epithelial stem cells in the adult lung. hStem Cells, 34(5), pp.1396-1406.

[309] Amma, L.L., Goodyear, R., Faris, J.S., Jones, I., Ng, L., Richardson, G. and Forrest, D., 2003. An emilin family extracellular matrix protein identified in the cochlear basilar membrane. *Molecular and Cellular Neuroscience*, 23(3), pp.460-472.

[310] Braghetta, P., Ferrari, A., de Gemmis, P., Zanetti, M., Volpin, D., Bonaldo, P. and Bressan, G.M., 2004. Overlapping, complementary and site-specific expression pattern of genes of the EMILIN/Multimerin family. *Matrix biology*, 22(7), pp.549-556.

[311] Pereira, L., D'Alessio, M., Ramirez, F., R. Lynch, J., Sykes, B., Pangilinan, T. and Bonadio, J., 1993. Genomic organization of the sequence coding for fibrillin, the defective gene product in Marfan syndrome. *Human molecular genetics*, 2(7), pp.961-968.

[312] Sinha, S., Nevett, C., Shuttleworth, C.A. and Kielty, C.M., 1998. Cellular and extracellular biology of the latent transforming growth factor-beta binding proteins. *Matrix biology*, 17(8-9), pp.529-545.

[313] Doliana, R., Bot, S., Bonaldo, P. and Colombatti, A., 2000. EMI, a novel cysteine?rich domain of EMILINs and other extracellular proteins, interacts with the gC1q domains and participates in

multimerization. *FEBS letters*, 484(2), pp.164-168.

[314] Yen-Rei, A.Y., O'Koren, E.G., Hotten, D.F., Kan, M.J., Kopin, D., Nelson, E.R., Que, L. and Gunn, M.D., 2016. A protocol for the comprehensive flow cytometric analysis of immune cells in normal and inflamed murine non-lymphoid tissues. *PloS one*, 11(3), p.e0150606.

[315] Xie, N., Cui, H., Ge, J., Banerjee, S., Guo, S., Dubey, S., Abraham, E., Liu, R.M. and Liu, G., 2017. Metabolic characterization and RNA profiling reveal glycolytic dependence of profibrotic phenotype of alveolar macrophages in lung fibrosis. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 313(5), pp.L834-L844.

[316] Romero, F., Shah, D., Duong, M., Penn, R.B., Fessler, M.B., Madenspacher, J., Stafstrom, W., Kavuru, M., Lu, B., Kallen, C.B. and Walsh, K., 2015. A Pneumocyte-macrophage paracrine lipid Axis drives the lung toward fibrosis. *American journal of respiratory cell and molecular biology*, 53(1), pp.74-86.

[317] Ortiz, L.A., Moroz, K., Liu, J.Y., Hoyle, G.W., Hammond, T., Hamilton, R.F., Holian, A., Banks, W., Brody, A.R. and Friedman, M., 1998. Alveolar macrophage apoptosis and TNF-alpha, but not p53, expression correlate with murine response to bleomycin. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 275(6), pp.L1208-L1218.

[318] Zhao, H. W., Hu, S. Y., Barger, M. W., Ma, J. K. H., Castranova, V., Ma, J. Y. C., 2004. Time-dependent apoptosis of alveolar macrophages from rats exposed to bleomycin: involvement of TNF receptor 2. *Journal of Toxicology and Environmental Health*, 67(17), 1391-1406.

[319] Lee, J.S., 2012. Heterogeneity of lung mononuclear phagocytes in chronic obstructive pulmonary disease. *Journal of innate immunity*, 4(5-6), pp.489-497.

[320] Gwyer Findlay, E. and Hussell, T., 2012. Macrophage-mediated inflammation and disease: a focus on the lung. *Mediators of inflammation*, 2012.

[321] Dhaliwal, K., Scholefield, E., Ferenbach, D., Gibbons, M., Duffin, R., Dorward, D.A., Morris, A.C., Humphries, D., MacKinnon, A., Wilkinson, T.S. and Wallace, W.A., 2012. Monocytes control second-phase neutrophil emigration in established lipopolysaccharide-induced murine lung injury. *American journal of respiratory and critical care medicine*, 186(6), pp.514-524.

[322] Maus, U.A., Waelsch, K., Kuziel, W.A., Delbeck, T., Mack, M., Blackwell, T.S., Christman, J.W., Schlöndorff, D., Seeger, W. and Lohmeyer, J., 2003. Monocytes are potent facilitators of alveolar neutrophil emigration during lung inflammation: role of the CCL2-CCR2 axis. *The Journal of Immunology*, 170(6), pp.3273-3278.

[323] Zaynagetdinov, R., Sherrill, T.P., Kendall, P.L., Segal, B.H., Weller, K.P., Tighe, R.M. and Blackwell, T.S., 2013. Identification of myeloid cell subsets in murine lungs using flow cytometry. *American journal of respiratory cell and molecular biology*, 49(2), pp.180-189.

[324] Soroosh, P., Doherty, T.A., Duan, W., Mehta, A.K., Choi, H., Adams, Y.F., Mikulski, Z., Khorram, N., Rosenthal, P., Broide, D.H. and Croft, M., 2013. Lung-resident tissue macrophages generate Foxp3+ regulatory T cells and promote airway tolerance. *Journal of Experimental Medicine*, 210(4), pp.775-788.

[325] Kimura, T., Ishii, Y., Morishima, Y., Shibuya, A., Shibuya, K., Taniguchi, M., Mochizuki, M., Hegab, A.E., Sakamoto, T., Nomura, A. and Sekizawa, K., 2004. Treatment with alpha-galactosylceramide attenuates the development of bleomycin-induced pulmonary fibrosis. *The Journal of Immunology*, 172(9), pp.5782-5789.

[326] Kim, J.H., Kim, H.Y., Kim, S., Chung, J.H., Park, W.S. and Chung, D.H., 2005. Natural killer T (NKT) cells attenuate bleomycin-induced pulmonary fibrosis by producing interferon-gamma. *The American journal of pathology*, 167(5), pp.1231-1241.

[327] Bi, J., Cui, L., Yu, G., Yang, X., Chen, Y. and Wan, X., 2017. NK cells alleviate lung inflammation by negatively regulating group 2 innate lymphoid cells. *The Journal of Immunology*, 198(8), pp.3336-3344.

[328] Nussbaum, J.C., Van Dyken, S.J., Von Moltke, J., Cheng, L.E., Mohapatra, A., Molofsky, A.B., Thornton, E.E., Krummel, M.F., Chawla, A., Liang, H.E. and Locksley, R.M., 2013. Type 2 innate lymphoid cells control eosinophil homeostasis. *Nature*, 502(7470), p.245.

[329] Xu, Y., Mizuno, T., Sridharan, A., Du, Y., Guo, M., Tang, J., Wikenheiser-Brokamp, K.A., Perl, A.K.T., Funari, V.A., Gokey, J.J. and Stripp, B.R., 2016. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI insight*, 1(20).

[330] Xi, Y., Kim, T., Brumwell, A.N., Driver, I.H., Wei, Y., Tan, V., Jackson, J.R., Xu, J., Lee, D.K., Gotts, J.E. and Matthay, M.A., 2017. Local lung hypoxia determines epithelial fate decisions during alveolar regeneration. *Nature cell biology*, 19(8), p.904.

[331] Lotfollahi, M., Wolf, F.A. and Theis, F.J., 2018. Generative modeling and latent space arithmetics predict single-cell perturbation response across cell types, studies and species. *bioRxiv preprint*, p.478503.

[332] Barkauskas, C.E., Chung, M.I., Fioret, B., Gao, X., Katsura, H. and Hogan, B.L., 2017. Lung organoids: current uses and future promise. *Development*, 144(6), pp.986-997.

[333] Aragon, T.J., Lichtensztajn, D.Y., Katcher, B.S., Reiter, R. and Katz, M.H., 2008. Calculating expected years of life lost for assessing local ethnic disparities in causes of premature death. *BMC public health*, 8(1), p.116.

[334] Monnier, J. and Zabel, B.A., 2014. Anti-asialo GM1 NK cell depleting antibody does not alter the development of bleomycin induced pulmonary fibrosis. *PloS one*, 9(6), p.e99350.

[335] Bugaut, H., Bruchard, M., Berger, H., Derangere, V., Odoul, L., Euvrard, R., Ladoire, S., Chalmin, F., Vegran, F., Rebe, C. and Apetoh, L., 2013. Bleomycin exerts ambivalent antitumor immune effect by triggering both immunogenic cell death and proliferation of regulatory T cells. *PLoS One*, 8(6), p.e65181.

[336] O'Hagan, R.C., Schreiber-Agus, N., Chen, K., David, G., Engelman, J.A., Schwab, R., Alland, L., Thomson, C., Ronning, D.R., Sacchettini, J.C. and Meltzer, P., 2000. Gene-target recognition among members of the myc superfamily and implications for oncogenesis. *Nature genetics*, 24(2), p.113.

[337] Larsson, L.G., Bahram, F., Burkhardt, H. and LuEscher, B., 1997. Analysis of the DNA-binding activities of Myc/Max/Mad network complexes during induced differentiation of U-937 monoblasts and F9 teratocarcinoma cells. *Oncogene*, 15(6), p.737.

[338] Lewin, A.R., Reid, L.E., McMahon, M., Stark, G.R. and Kerr, I.M., 1991. Molecular analysis of a human interferon?inducible gene family. *European journal of biochemistry*, 199(2), pp.417-423.

[339] Lang, D., Knop, J., Wesche, H., Raffetseder, U., Kurrle, R., Boraschi, D. and Martin, M.U., 1998. The type II IL-1 receptor interacts with the IL-1 receptor accessory protein: a novel mechanism of regulation of IL-1 responsiveness. *The Journal of Immunology*, 161(12), pp.6871-6877.

[340] Kee, B.L., 2009. E and ID proteins branch out. *Nature Reviews Immunology*, 9(3), p.175.

[341] Wang, Y.M., Bakhtiar, M. and Alexander, S.I., 2017. ILC2: there's a new cell in town. *Am Soc Nephrol*, 1953-1955.

[342] Spits, H., Bernink, J.H. and Lanier, L., 2016. NK cells and type 1 innate lymphoid cells: partners in host defense. *Nature immunology*, 17(7), p.758.

[343] Gasteiger, G., Fan, X., Dikiy, S., Lee, S.Y. and Rudensky, A.Y., 2015. Tissue residency of innate lymphoid cells in lymphoid and nonlymphoid organs. *Science*, 350(6263), pp.981-985.

[344] Moro, K., Kabata, H., Tanabe, M., Koga, S., Takeno, N., Mochizuki, M., Fukunaga, K., Asano, K., Betsuyaku, T. and Koyasu, S., 2016. Interferon and IL-27 antagonize the function of group 2 innate lymphoid cells and type 2 innate immune responses. *Nature immunology*, 17(1), p.76.

[345] Liew, F.Y., Girard, J.P. and Turnquist, H.R., 2016. Interleukin-33 in health and disease. *Nature Reviews Immunology*, 16(11), p.676.

[346] von Moltke, J., Ji, M., Liang, H.E. and Locksley, R.M., 2016. Tuft-cell-derived IL-25 regulates an

intestinal ILC2?epithelial response circuit. *Nature*, 529(7585), p.221.

[347] Gorski, S.A., Hahn, Y.S. and Braciale, T.J., 2013. Group 2 innate lymphoid cell production of IL-5 is regulated by NKT cells during influenza virus infection. *PLoS pathogens*, 9(9), p.e1003615.

[348] Lefrancais, E., Duval, A., Mirey, E., Roga, S., Espinosa, E., Cayrol, C. and Girard, J.P., 2014. Central domain of IL-33 is cleaved by mast cell proteases for potent activation of group-2 innate lymphoid cells. *Proceedings of the National Academy of Sciences*, 111(43), pp.15502-15507.

[349] Cayrol, C., Duval, A., Schmitt, P., Roga, S., Camus, M., Stella, A., Burlet-Schiltz, O., Gonzalez-de-Peredo, A. and Girard, J.P., 2018. Environmental allergens induce allergic inflammation through proteolytic maturation of IL-33. *Nature immunology*, 19(4), p.375.

[350] Klose, C.S. and Artis, D., 2016. Innate lymphoid cells as regulators of immunity, inflammation and tissue homeostasis. *Nature immunology*, 17(7), p.765.

[351] Liu, T., Jin, H., Ullenbruch, M., Hu, B., Hashimoto, N., Moore, B., McKenzie, A., Lukacs, N.W. and Phan, S.H., 2004. Regulation of found in inflammatory zone 1 expression in bleomycin-induced lung fibrosis: role of IL-4/IL-13 and mediation via STAT-6. *The Journal of Immunology*, 173(5), pp.3425-3431.

[352] Belperio, J.A., Dy, M., Burdick, M.D., Xue, Y.Y., Li, K., Elias, J.A. and Keane, M.P., 2002. Interaction of IL-13 and C10 in the pathogenesis of bleomycin-induced pulmonary fibrosis. *American journal of respiratory cell and molecular biology*, 27(4), pp.419-427.

[353] Duerr, C.U., McCarthy, C.D., Mindt, B.C., Rubio, M., Meli, A.P., Pothlichet, J., Eva, M.M., Gauchat, J.F., Qureshi, S.T., Mazer, B.D. and Mossman, K.L., 2016. Type I interferon restricts type 2 immunopathology through the regulation of group 2 innate lymphoid cells. *Nature immunology*, 17(1), p.65.

[354] Monticelli, L.A., Sonnenberg, G.F., Abt, M.C., Alenghat, T., Ziegler, C.G., Doering, T.A., Angelosanto, J.M., Laidlaw, B.J., Yang, C.Y., Sathaliyawala, T. and Kubota, M., 2011. Innate lymphoid cells promote lung-tissue homeostasis after infection with influenza virus. *Nature immunology*, 12(11), p.1045.

[355] Monticelli, L.A., Osborne, L.C., Noti, M., Tran, S.V., Zaiss, D.M. and Artis, D., 2015. IL-33 promotes an innate immune pathway of intestinal tissue protection dependent on amphiregulin?EGFR interactions. *Proceedings of the National Academy of Sciences*, 112(34), pp.10762-10767.

[356] Mohapatra, A., Van Dyken, S.J., Schneider, C., Nussbaum, J.C., Liang, H.E. and Locksley, R.M., 2016. Group 2 innate lymphoid cells utilize the IRF4-IL-9 module to coordinate epithelial cell maintenance of lung homeostasis. *Mucosal immunology*, 9(1), p.275.

[357] Barlow, J.L., Bellosi, A., Hardman, C.S., Drynan, L.F., Wong, S.H., Cruickshank, J.P. and McKenzie, A.N., 2012. Innate IL-13?producing nuocytes arise during allergic lung inflammation and contribute to airways hyperreactivity. *Journal of Allergy and Clinical Immunology*, 129(1), pp.191-198.

[358] Barlow, J.L., Peel, S., Fox, J., Panova, V., Hardman, C.S., Camelo, A., Bucks, C., Wu, X., Kane, C.M., Neill, D.R. and Flynn, R.J., 2013. IL-33 is more potent than IL-25 in provoking IL-13?producing nuocytes (type 2 innate lymphoid cells) and airway contraction. *Journal of allergy and clinical immunology*, 132(4), pp.933-941.

[359] Roy, D.N. and Goswami, R., 2017. IL-9 signaling Pathway: An update. *In Th9 Cells* (pp. 37-50). Humana Press, New York, NY.

[360] Ziegler, S.F., Roan, F., Bell, B.D., Stoklasek, T.A., Kitajima, M. and Han, H., 2013. The biology of thymic stromal lymphopoietin (TSLP). *In Advances in pharmacology* (Vol. 66, pp. 129-155). Academic Press.

[361] Kabata, H., Moro, K., Fukunaga, K., Suzuki, Y., Miyata, J., Masaki, K., Betsuyaku, T., Koyasu, S. and Asano, K., 2013. Thymic stromal lymphopoietin induces corticosteroid resistance in natural helper cells during airway inflammation. *Nature communications*, 4, p.2675.

[362] Möjsberg, J., Bernink, J., Golebski, K., Karrich, J.J., Peters, C.P., Blom, B., te Velde, A.A., Fokkens, W.J., van Drunen, C.M. and Spits, H., 2012. The transcription factor GATA3 is essential for the function of human type 2 innate lymphoid cells. *Immunity*, 37(4), pp.649-659.

[363] Zacchigna, L., Vecchione, C., Notte, A., Cordenonsi, M., Dupont, S., Maretto, S., Cifelli, G., Ferrari, A., Maffei, A., Fabbro, C. and Braghetta, P., 2006. Emilin1 links TGF-? maturation to blood pressure homeostasis. *Cell*, 124(5), pp.929-942.

[364] Schiavinato, A., Becker, A.K.A., Zanetti, M., Corallo, D., Milanetto, M., Bizzotto, D., Bressan, G., Guljelmovic, M., Paulsson, M., Wagener, R. and Braghetta, P., 2012. EMILIN-3, peculiar member of elastin microfibril interface-located protein (EMILIN) family, has distinct expression pattern, forms oligomeric assemblies, and serves as transforming growth factor beta (TGFb) antagonist. *Journal of Biological Chemistry*, 287(14), pp.11498-11515.

[365] Ogasawara, N., Poposki, J.A., Klingler, A.I., Tan, B.K., Weibman, A.R., Hulse, K.E., Stevens, W.W., Peters, A.T., Grammer, L.C., Schleimer, R.P. and Welch, K.C., 2018. IL-10, TGF-b, and glucocorticoid prevent the production of type 2 cytokines in human group 2 innate lymphoid cells. *Journal of Allergy and Clinical Immunology*, 141(3), pp.1147-1151.

[366] Love, M.I., Huber, W. and Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12), p.550.

List of Figures

Publications

In chronologic order:

- Schiller H.B., Mayr C.H., Leuschner G., **Strunz M.**, Staab-Weijnitz C., Preisendörfer S., Eckes B., Moinzadeh P., Krieg T., Schwartz D.A., Hatz R.A., Behr J., Mann M., Eickelberg O. (2017). "Deep proteome profiling reveals common prevalence of MZB1-positive plasma B cells in human lung and skin fibrosis". *American Journal of Respiratory and Critical Care Medicine* 196.10: 1298-1310.

- Angelidis I., Simon L.M., Fernandez I.E., **Strunz M.**, Mayr C.H., Greiffo F.R., Tsitsiridis G., Ansari M., Graf E., Strom T.M., Nagendran M., Desai T., Eickelberg O., Mann M., Theis F.J., Schiller H.B. (2019). "An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics". *Nature Communications* 10.1: 963.

- **Strunz M.**, Jarrell J.T., Cohen D.S., Rosin E.R., Vanderburg C.R., Huang X. (2019). "Modulation of SPARC/Hevin proteins in Alzheimer's disease brain injury". *Journal of Alzheimer's Disease* 1-16.

- Vieira Braga F.A., Kar G., Berg M., Carpaij O.A., Polanski K., Simon L.M., Brouwer S., Gomes T., Hesse L., Jiang J., Fasouli E.S., Efremova M., Vento-Tormo R., Affleck K., Palit S., Strzelecka P., Firth H.V., Mahbubani K.T.A., Cvejic A., Meyer K.B., Saeb-Parsy K., Luinge M., Brandsma C.A., Timens W., Angelidis I., **Strunz M.**, Koppelman G.H., van Oosterhout A.J., Schiller H.B., Theis F.J., van den Berge M., Nawijn M.C., Teichmann S.A. (2019). "A cellular census of healthy lung and asthmatic airway wall identifies novel cell states in health and disease". *Nature Medicine* 1.

- Schiller H.B., Montoro D.T., Simon L.M., Rawlins E.L., Meyer K.B., **Strunz M.**, Vieira Braga F.A., Timens W., Koppelman G.H., Budinger G.R.S., Burgess J.K., Waghray A., van den Berge M., Theis F.J., Regev A., Kaminski N., Rajagopal J., Teichmann S.A., Misharin A.V., Nawijn M.C. (2019)."The Human Lung Cell Atlas: A high-resolution reference map of the human lung in health and disease". *American Journal of Respiratory Cell and Molecular Biology* 61.1: 31-41.

- **Strunz M.**, Simon L.M, Ansari M., Mattner L.F., Angelidis I., Mayr C.H., Kathiriya J., Yee M., Ogar P.I., Sengupta A., Voss C., Stöger T., Kukhtevich I., Schneider R., Neumann J.H.L., Behr J., Lehmann M., Königshoff M., Burgstaller G., O'Reilly M., Chapman H.A.,

Theis F.J., Schiller H.B. (2019). "Longitudinal single cell transcriptomics reveals Krt8+ alveolar epithelial progenitors in lung regeneration". *submitted, preprint available on bioRxiv: doi: https://doi.org/10.1101/705244*. #

- Koopmans T., Fischer A., Ramesh P., Christ S., **Strunz M.**, Aichler M., Feuchtinger A., Walch A., Ansari M., Theis F.J., Neumann P.A., Schiller H.B., Rinkevich Y. (2019). "Post-surgical adhesions are cell fusions between organ surfaces caused by mesothelial membrane bridges". *Nature Communications - in Revision*.

- Paschke S., **Strunz M.**, Sander P., Scheiber, C. Schneider J., Bechter K., Steinacker J., Georgieff M., Schneider E.M. (2019). "Microparticle-derived miRNA from macrophages and their impact on the diagnostics of affective disorders and schizophrenia". *submitted*.

- Mayr C.H., Ansari M., Simon L.M., Ogar P., **Strunz M.**, Angelidis I., Liontos A., Samakovlis C., Morty R., Ahlbrecht K., Theis F.J., Schiller HB. (2019). "Single cell profiling reveals diversity of mouse and human lung mesenchymal cell types and their distinct activation states in fibrogenesis". *in preparation*.

- Leuschner G., Mayr C.H., Geyer P., **Strunz M.**, Angelidis I., Böhm S., Adler H., Lindner M., Eickelberg O., Simon L.M., Theis F.J., Hilgendorff A., Mann M., Behr J., Schiller H.B. (2019). "Proteomic signatures of subclinical exacerbations and smoking in interstitial lung disease patients". *in preparation*.

this publication resulted from this thesis.

Acknowledgements

First and foremost, I am deeply grateful to my PhD advisor and mentor Dr. Herbert Schiller who provided full support and scientific aid throughout my time at the CPC. I appreciate your opinion and help, especially when new problems appeared and a guiding hand was needed. Thank you for your input and the discussions, which really helped me to better understand lung biology and in particular lung regeneration. I have learned so much throughout the past four years, even beyond science, of which I am sure will help and guide me through my future career. Thank you for all the conferences and retreats that I could go to. Additionally, I want to gratefully acknowledge all the activities we shared outside the lab - from annual Oktoberfest visits, lab get-togethers at restaurants and bars at Christmas time, or small retreats for summer BBQs, beer garden visits, ... your generosity was and is very much appreciated! I will treasure the times of both in and outside the lab just as much and am approaching new life goals with great memories of the past.

A special and big thank you to Prof. Dr. Jürgen Behr for the support and supervision of my PhD work. I benefited from the medical input to my research activities and cherished the participation and presence during my thesis advisory committee meetings. Along these lines, I want to express my gratitude to Prof. Dr. Silke Meiners and Prof. Dr. Reinhard Fässler for their contribution to my thesis advisory committee and fruitful discussions and ideas for the experimental work. Especially to Silke whom I could easily walk by her door to ask for input or help related or unrelated to the PhD project. Thank you all for your efforts and time and for the continuous feedback throughout the past four years.

Beyond, I would like to acknowledge everyone in the lab who contributed to the success of this PhD thesis. Some of you have accompanied me on my way for a long time. I want to thank you for your patience, your understanding, for the strong support and the amazing atmosphere we have generated in the lab. Over the years, the lab size has grown continuously but we all made sure to stay one group. Thanks for all the lunch and coffee breaks, the after work events and your friendship. Because you guys stuck with me the longest and we well connected: thank you Laura, Ilias and Christoph; I will never forget your help and the amazing time when we had long days working with mice in rooms without windows.

I want to thank the CPC Research School and in particular Dr. Doreen Franke, Dr. Darcy Wagner, and Dr. Claudia Staab-Weijnitz for the organization and success of the curriculum. Thank you for including me in some of the event planning and for cherishing my work as the student's representative. We shared fun times together which I will surely miss soon.

Last, but not least, a huge thank you to my family and friends for their support. Without you, this experience would not have been the same. In four years time, there are many people I want to acknowledge but most importantly to name are my grand parents and parents for the continuous solicitude, my brothers for support and their visits in Munich. My full appreciation to Rainer for sympathy, endless encouragement and the unconditional support when I encountered medical and private incidents. To Reno for scientific assistance, life advice, the best times and your support to guide my future career. Thank you all for sharing this important step with me and your continuous support.