Faculty of Biology

Ludwig-Maximilians-University Munich

Characterization of Emilin-2 in Lung Injury and Regeneration using Single-Cell Transcriptomics

---

Submitted to the Faculty of Biology

Master of Science in Biology

Submitted by

Mert Akgündüz

Munich, February 2020

Title of the Thesis:	Characterization of Emilin2 in lung injury and regeneration using single-cell transcriptomics
Institute:	Comprehensive Pneumology Center (CPC) Institute of Lung Biology and Disease (ILBD) Helmholtz Zentrum München Max-Lebsche-Platz 31 81377 München Germany
Submitted by:	Mert Akgündüz
Student Number:	11815797
External supervisor 1:	Dr. Herbert Schiller
External supervisor 2: (Habilitated supervisor)	Prof. Dr. Silke Meiners
Internal supervisor:	Prof. Dr. Angelika Böttger

Abstract

The lung functions as the body's gas exchange organ. The fibrotic diseases of the lung interstitium, such as idiopathic pulmonary fibrosis (IPF), can be characterized by alveolar epithelial injury and aberrant extracellular matrix deposition by fibroblasts and a faulty in repair mechanisms after chronic tissue injury, which results in loss of normal lung architecture and impairment in gas exchange. Previous study of the time-resolved proteome of the bleomycin lung injury mouse model by Schiller et al. in 2015 highlighted the kinetics of certain ECM proteins, including Emilin2, which was shown to be highly upregulated two weeks after injury and potentially driving a progressive scarring phenotype. In addition, previous studies in the Schiller lab suggest a delay in the resolution of fibrosis in the late stages of the bleomycin model. Here, we employed the bleomycin-induced lung fibrosis model to explore the impact of depletion of Emilin2 after injury by utilizing the single cell RNAseq platform Drop-seq. Investigation of diseased tissue milieu in lung fibrosis is complex due to great cellular heterogeneity. Single cell transcriptomics enables the unbiased single cell profiling and thereby allows to elucidate the contribution of individual cell types to cell population dynamics and transcriptional changes during fibrotic disease states. Analyses presented in this thesis revealed modules of differentially expressed genes in the lungs of Emilin2 knockout mice and the dynamic changes in immune cell populations in the late stages of the model.

Abstract

Die Lunge ist für den Gasaustausch des Körpers verantwortlich. Die fibrotischen Erkrankungen des interstitiellen Raums der Lunge, wie die idiopathische pulmonäre Fibrose (IPF), sind gekennzeichnet durch Verletzung des alveolären Epithels und die vermehrte Ablagerung von extrazellulärer Matrix durch Fibroblasten, sowie durch gestörte Reparationsmechanismen nach chronischer Gewebsverletzung, die in Verlust normaler Gewebsarchitetktur und Beeinträchtigung des Gasaustausches resultieren. Vorherige Studien zum Proteom über den Zeitverlauf des Bleomyzin-induzierten Lungenverletzungsmodell in der Maus durch Schiller et al. in 2015 zeichneten die Kinetiken gewisser extrazellulärer Proteine in diesem Modell heraus. Emilin2 wurde hierbei als überaus hochreguliertes Protein zwei Wochen nach Gabe von Bleomyzin entdeckt und als potenzieller Progressor für den zunehmend vernarbenden Phänotyp diskutiert. Darüber hinaus lassen weitere Studien des Schiller Labors vermuten, dass es eine zeitliche Verzögerung der Auflösung der Fibrose in vor allem den späten Zeitpunkten des Modells gibt. Für die Arbeit hier, nutzten wir das Bleomyzin-induzierte Fibrosemodell, um das Ausmaß einer Depletion von Emilin2 in Mäusen nach der Lungenverletzung mittels der Einzelzell-RNA-Sequenzierungsmethode Dropseq zu studieren. Die Untersuchung von erkranktem Gewebe ist während der Fibrose durch hohe zelluläre Heterogenität schwierig. Transkriptomische Studien auf Einzelzellebene erlauben unvoreingenommene Analysen indem sie es erlauben, die Beteiligung einzelner Zelltypen zu studieren, sowie deren Dynamiken und transkriptomischen Änderungen während der fibrotischen Krankheitszustände. Die Analysen in dieser Thesis brachten differenziell exprimierte Gene ans Licht, die sich zwischen wildtyp und Emilin2 knockout Tieren besonders in den dynamischen Änderungen der Immunzellpopulationen zeigten und zudem auffällig zu den späteren Zeitpunkten des Mausmodells auftraten.

Contents

1 Introduction

The mammalian lung is a specialized organ for gas exchange forming an air-liquid interface between the atmosphere and the bloodstream (Koval and Sidhaye, 2017). In addition to its main functions of air conduction and gas exchange, different tissue niches within the different compartments of the lung have a role in connection to the cardiovascular system, and immune surveillance (Basil and Morrisey, 2019). Within the evolutionary context, the mammalian lung is one of the most recently evolved organs; from the primordial water breathing, evaginated gas exchangers to air breathing lungs in terrestrial life (Carvalho, 2011). Larger mammals and in particular those with upright posture require the larger organ size and a unique gas distribution which eventually contributes to the evolution of structural and cellular complexity, especially in the human lung. Such heterogeneity in cellular architecture and complexity in cell-cell interactions makes it challenging to define the response of various cell types to injuries. To achieve this complexity, the mammalian lung undergoes intensive developmental processes which start with the specification in the anterior foregut endoderm and continues with the postnatal maturation of the alveoli (Zepp and Morrisey, 2019). The human lung is inevitably exposed to air which brings dust, noxious gases, microorganisms, allergens and pollutants. Aetiology of the most interstitial lung disease remains largely unknown, however, constant external insults, aging and genetic background increase susceptibility to IPF which is a leading cause of death worldwide, affecting roughly five million people (Burgstaller et al., 2017). Despite being relatively quiescent organ, the human lung has a varying capacity of regenerative response depending on the severity of injury. Upon chronic lung diseases, such as in pulmonary fibrosis, normal cellular architecture of the lung is not restored due to the faulty in regenerative and repair mechanisms. Instead, some populations, such as fibroblasts, undergo aberrant cellular expansion while others are lost. Newly discovered provisional ECM proteins Emilin2 and collagen-XXVIII was demonstrated as highly regulated in the context of fibrosis in bleomycin mice model. Their upregulation already starts at day 3 and shows a peak at day 14 (Schiller et al., 2015). Emilin2 as a microfibrillar protein is believed to contribute to the fibrotic phenotype which is also consistent with the decrease in the body weight in the late stage of fibrosis in bleomycin treated *Emilin2* knockout mice (Strunz, 2019). Murine animal models have been widely used to study lung diseases and its repair, however, recapitulating human pathophysiology in mice is challenging due to the differences in anatomy, cellular landscape, and environmental exposures. Also, the more complex diseases, such as IPF, are more challenging to model completely in mice. However, most of the knowledge about repair after chronic injury comes from modelling chronic injury in mice. In this regard, bleomycin mouse model is quite established and provides simplicity of use and genetic manipulation. In addition, bleomycin model can provide means of novel investigation when coupled with rapid advancements in technologies, such as single cell RNA-seq. Single cell analyses allow to circumvent difficulty in analyses due to the diverse cellular heterogeneity in the lungs and trachea, and to better elucidate effects of lung disease on respiratory function at the level of individual cells (Basil and Morrisey, 2019).

1.1 Anatomical overview of the respiratory system

“Inevitable Fatum” means “Inevitable Fate” and has been Dryander’s motto to remind people the fragility of life. German physician, anatomist, mathematician and astronomer Johann Dryander used this motto on one of the earliest anatomical work illustrating the trachea, lungs and bronchial system of the lung in the 16th century (Figure 1) (Koehler, 2019).

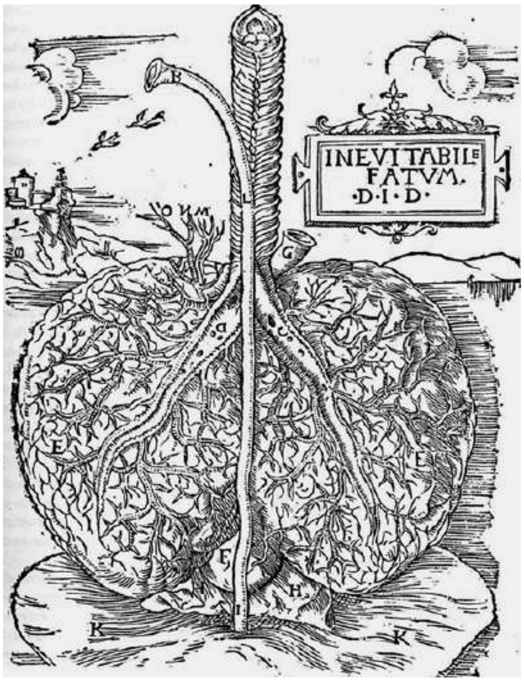

Figure 1. Woodcut representation of the human lung from the posterior view. A = Trachea with cartilage, B = esophagus, C, D = right and left main bronchi, E = bronchial system, F = lower heart impression, G =

aorta, H = pericardium, I = lower esophagus, K = diaphragm, L = esophagus, M, N, O = blood vessels. Illustration is adapted from Koehler, 2019.

The human lung constitutes one of the barrier organs with a surface area of approximately 70 m^2 (Basil and Morrisey, 2019). Anatomically, the respiratory system in mammals can be thought of as divided into the upper respiratory tract and the lower respiratory tract. The upper respiratory tract encompasses the nostrils, nasal cavity, mouth, pharynx, and larynx, while the lower respiratory tract includes the trachea, terminal and respiratory airways and the peripheral alveoli (Zepp and Morrisey, 2019). The upper respiratory tract primarily functions to humidify and filter the inhaled air first in the nasopharynx and then the air passes through conducting airways to the alveoli for gas exchange. Conducting airways, in the lower respiratory tract, begin at the trachea and bifurcate into the main bronchi which further branches into smaller bronchi and ultimately the bronchioles (Koval and Sidhaye, 2017). The healthy lung is soft and light. The human lung is a paired organ which is located at either side of the heart in thoracic cavity with three right and two left lobes resting over concave-shaped diaphragm. The lungs are composed of branched airways and blood vessels that unite at the very distal end to form the alveolus where gas exchange takes place. Lung surfaces are surrounded by a membrane called the visceral pleura, which refers to the mesothelium in mouse, while the thoracic cavity is lined by the parietal pleura. Space between these membranes form the pleural cavity. Pleural membranes are lubricated by secreted pleural fluid which allows lungs to glide over the thoracic wall during breathing. Extending from inferior end of the larynx, trachea, bronchi, and some bronchioles are covered by C-shaped cartilaginous rings providing stability, which is also supported by trachealis muscle at the posterior side. The tracheobronchial tree is formed by numerous divisions of most proximal trachea into narrower airways (Aung et al., 2019). The branched structure of the airways is paralleled by the pulmonary vasculature (Zepp and Morrisey, 2019). The first division of the trachea gives rise to the right and left main stem bronchi which enters the lung through the hilum; and each main bronchus divides into the lobar (secondary) bronchi supplying lobes of the lung; and again, each lobar bronchus divides into the segmental (tertiary) bronchi supplying the bronchopulmonary segments. These segments are separated by connective tissue septa. Each segment has its own arterial supply. Segmental bronchi divide into terminal bronchioles through conducting bronchioles. Bronchi are supported by hyaline cartilage, whereas bronchioles are covered smooth muscle. From terminal bronchioles, microscopic branches or respiratory bronchioles separate and form several alveolar ducts which give rise to alveolar sacs surrounded by clusters of alveoli. Air travels all the way from trachea to alveoli, where the gas exchange with the vasculature occurs (Aung et al., 2019). In mouse transition point between the conducting airways and the alveolar unit is called as Bronchioalveolar Duct Junction (BADJ), whereas, in large mammals including humans, the intermediate area is populated with the respiratory airways. On the other hand, murine animal models exhibit architectural differences compared to the human lung. A mouse lung possesses 6000 times less volume than the human counterpart and smaller surface area for gas exchange with 82 cm^2. The mouse lung has less branches in the airways and thinner alveolar units. (Basil and Morrisey, 2019).

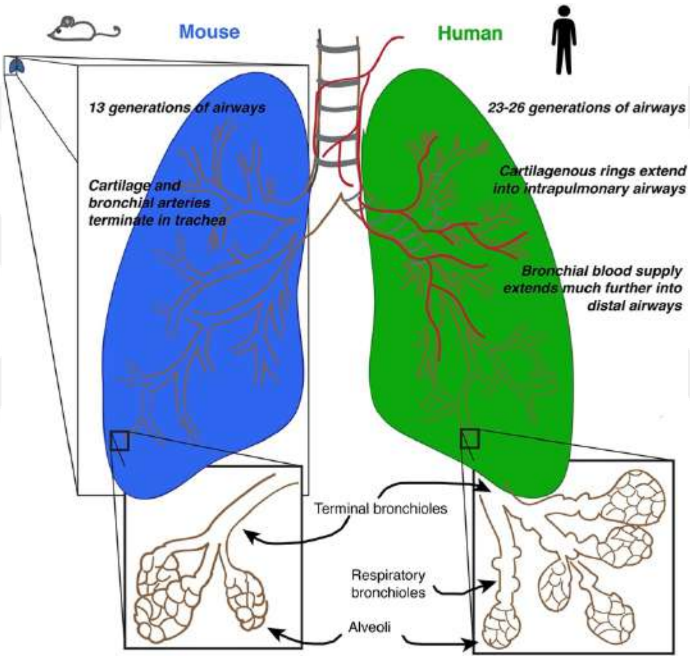

Figure 2. Anatomical differences between the human lung and the murine counterpart. The human lung architecture is larger and more branched with bronchial blood supply extending much further into distal airways. Illustration is adapted from Basil and Morrisey, 2019.

1.2 Cellular composition of the lung and its response to injury

Air-exposed surfaces are covered with epithelium in entire respiratory system of mammals. The human lung possesses more than 40 different cell types (Koval and Sidhaye, 2017). Spatially, the cellular composition differs along the proximal-distal axis (Basil and Morrisey, 2019). In humans as well as in mice, the proximal airways and trachea are lined by a pseudostratified epithelium with varying cell types, including multi-ciliated cells, secretory cells, goblet cells, basal stem/progenitor cells (BSCs), and further subtypes. Goblet and secretory cells can provide a first line of defence by trapping microparticles and microorganisms. Mucous secreting cells form so-called the mucociliary elevator with ciliated cells to clear the irritants out of the airspaces. The mucous layer in the airway surface liquid (ASL) contains mucins which traps the inhaled microparticles and the second layer, periciliary liquid layer (PCL), maintains a distance between these mucins and underlying epithelium to ease mucous clearance (Tarran et al., 2006). The airway surface liquid is maintained by the ion channels in order to optimize mucous clearance process. In the terminal bronchioles, pulmonary surfactant containing surfactant protein A (SP-A), SP-B and club cell specific protein 10 (CC-10) is mainly produced by club cells due to the lack of the submucosal glands and goblet cells. Pulmonary surfactant produced in the alveoli lacks CC-10 (produced by club cells) and instead contain SP-C which results in different biophysical characteristics (Koval and Sidhaye, 2017). Both the human and mouse lung present huge heterogeneity and single-cell studies discover new populations. A smaller number of CFTR-rich pulmonary ionocytes are showed to be present in the both human and mouse lung as a newly defined novel cell population (Plasschaert et al., 2018). Within the context of regeneration, in humans, basal cells; Trp63-positive multi-potent progenitors, can extend into the terminal bronchioles, whereas in mouse, they only reside in the trachea and the main stem bronchi. They can repopulate the pseudostratified epithelium in homeostasis and after injury (Basil and Morrisey, 2019). Secretory cells, or club cells, can also serve as progenitor population with a limited capacity than basal cells. Especially in mice bronchioles where the basal population is missing, secretory cells can differentiate into multi-ciliated lineage but not the other lineages including neuroendocrine, or tuft, cells. Tata et al., 2013 has demonstrated, using lineage tracing, fully committed secretory cells can dedifferentiate into basal stem cells as a response to their ablation in ex vivo culture. Another study from Guha et al., 2017 has identified a rare Uroplakin3a (Upk3a)-expressing club cell subtypes in murine airway localized next to neuroendocrine bodies. These progenitors have better propensity to generate club cells and ciliated cells in an injured airway. They can also contribute to alveolar type I and type II cells in bleomycin injury model (Guha et al., 2017). Pulmonary neuroendocrine cells (PNECs), on the other hand, are relatively rare, and first specialized cell types of airway epithelium. They are not only unique to mammals; they are evolutionarily conserved and found in many species. They have microvilli at their apical surface. PNECs exist both as single scattered cells in the respiratory epithelium and as clustered PNECs, or neuroepithelial bodies (NEBs), located mostly in the intrapulmonary airways, airway bifurcations or bronchioalveolar duct junctions (Song et al., 2012). NEBs can secrete various amines and peptides which are stored in their secretory granules and dense-core vesicles. Their functions include chemo-, mechano-, oxygen sensing in the airway, regulating pulmonary blood flow, controlling bronchial tonus, immune cell recruitment, and maintaining stem cell niche. It was shown that a fraction of proliferating PNECs can generate club cells and ciliated cells in response to naphthalene-induced lung damage. Notch signalling involves in PNEC fate decisions during injury. PNECs are also shown to be associated with human small cell lung cancer (SCLC) (Yao et al., 2018). Also, tuft cells or brush cells, expressing doublecortin-like kinase 1, are detected in upper and distal lung epithelium of mice after severe influenza injury (Rane et al., 2019). Proximal and distal airways possess mesenchyme population that can direct many significant events including respiratory lineage specification, branching morphogenesis, epithelial differentiation, lineage distinction, vascular development, and alveolar maturation (McCulley et al., 2015). ECM surrounds large airways and serve as a structural scaffold. In addition, it involves in cellular repair after epithelial injury. As a final unit in the distal lung parenchyma, alveolar sacs form delicate gas exchange sites. They share similar function and architectural homology from species to species. Alveolar air-liquid interface must be thin enough to enable free diffusion of oxygen and carbon dioxide (Koval and Sidhaye, 2017). Alveolar niche is composed of primary epithelial cell types of squamous alveolar type 1 (AT1) and cuboidal AT2 cells. AT1s mediate gas exchange by providing epithelial interface with vascular capillary. AT2s secrete surface-tension lowering surfactant complex to provide critical level of hydration and to prevent from alveolar collapse upon expiration. Clonal analysis and lineage tracing in developing mice showed that AT1 and AT2s arise independently during development from a bipotent progenitor. Mature AT1 and AT2s appear about 1 day before birth and alveolar development completes several weeks after birth with a switch in AT2s to make them precursors for both AT2s and AT1s (Desai et al., 2014). These precursors exhibit progenitor cell behaviour and they are different than Sox2-derived Krt5+ epithelial cells which migrate and regenerate damaged distal lung regions after severe injury. Their expression profile is enriched for key developmental genes including Fgfr2, Nkx2-1, Id2, Etv4, Etv5, and Foxa1. Also, both mouse and human progenitors were shown to be responsive to Wnt and Fgf signalling by clonal alveolar organoid assays. In alveolar niche, they reside near Wnt-expressing fibroblasts. Juxtracrine Wnts maintain progenitors in alveolar niche, however, upon severe injury, autocrine Wnt signalling can be induced and also ancillary stem cells transiently recruited to expand progenitor pool (Nabhan et al., 2018). Some cytokines and growth factors secreted by immune cells including IL-1β and TNFα were also shown to promote AT2 cell growth (Basil and Morrisey, 2019). In addition to its role of encasing respiratory system, mesothelium has important functions during lung development as generating certain mesenchymal cell populations, secreting paracrine growth factors including fibroblast growth factors (FGF) and WNTs, and inducing mesenchymal cell proliferation and differentiation, and contributing to airway patterning (Zepp and Morrisey, 2019).

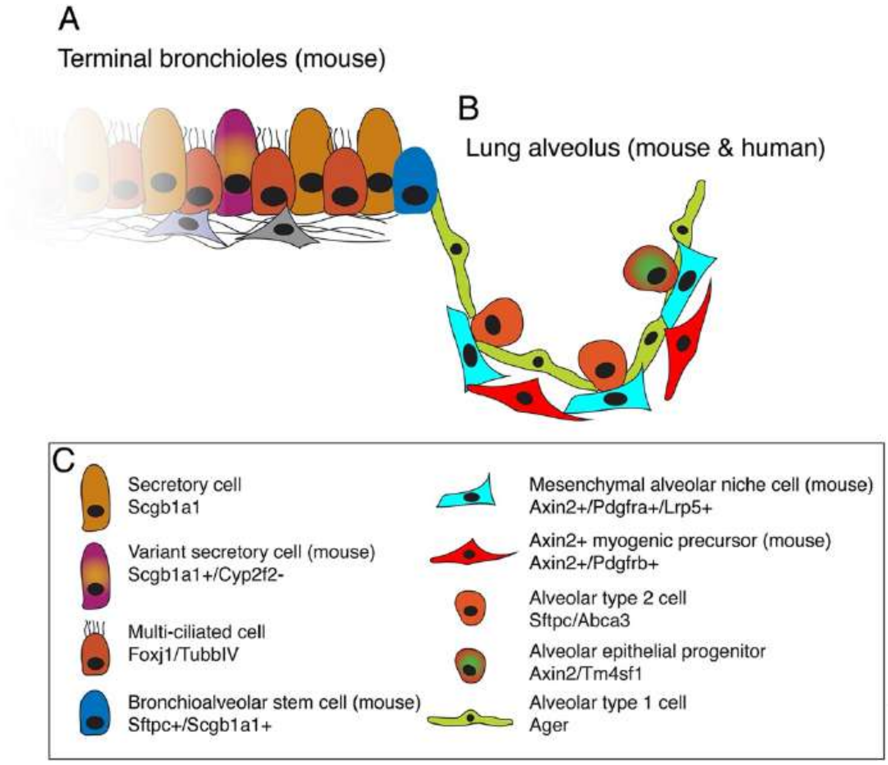

Figure 3. Cell populations that line the intermediate bronchioles in mice and alveoli in both human and mice lung. Illustration is adapted from Basil and Morrisey, 2019.

AT1 cells and the adjacent pulmonary capillary endothelial cells (PCECs) govern the main function of gas exchange in the lung. Alveolar regeneration also includes endothelial cell regeneration. Pulmonary endothelial cells possess substantial heterogeneity throughout the lung (Ding et al., 2011). Primary immune cells at homeostasis are the alveolar macrophages (tissue-resident macrophages) in the lung. They can involve in the stimulation of epithelial proliferation upon injury (Misharin et al., 2017). Pathology of bleomycin-induced lung fibrosis was shown to be worsened when macrophages were depleted during the recovery stage suggesting a dynamic requirement for macrophages in lung maintenance and repair (Gibbons et al., 2011). Alveolar tissue injury or fibrosis can cause the differentiation of monocytes into alveolar macrophages so called the monocyte-derived alveolar macrophages (Misharin et al., 2017). Misharin et al., 2017 has demonstrated monocyte derived alveolar macrophages can persist in the bleomycin treated lung after the resolution of injury. When macrophages arrive at their site of action, they are polarized along a spectrum by microenvironmental cues. Macrophage polarization refers to classically activated macrophages (M1) which expresses distinct set of genes associated with inflammatory conditions, and alternatively activated macrophages (M2) which have upregulated fibrosis related set of genes and have been implicated in wound repair and tissue regeneration. M2-polarized macrophages can originate from either tissue-resident or monocyte-derived macrophages. When induced by cytokines associated with type 2 immunity, such as IL-4 and IL-13, M2 macrophages can activate transcription of *Arg1*, *Fizz1*, *Ym1* and *Mrc1*. (Misharin et al., 2017; Lechner et al., 2017).

1.3 Interstitial lung diseases (ILDs)

All epithelial cells in the respiratory system share barrier function protecting host from environmental insults (Koval and Sidhaye, 2017). Every inhalation exposes lung to air laden with bacteria, viruses, oxidants, pollutants and allergens. In order to minimize inflammatory responses which can easily damage extremely thin and delicate gas exchanging epithelium, lung has several immune mechanisms to facilitate clearance. These mechanisms include mucociliary elevator, immunoglobulins, complement, alveolar macrophages and neutrophils. In addition, pulmonary surfactants contain collectins that promotes pathogen clearance and regulation of adaptive and innate immune-cell functions. Surfactants are composed of phospholipids and contains surfactant proteins of SP-A, SP-B, SP-C and SP-D. AT2 cells can synthesize all four types of surfactant proteins and surfactant lipids through its unique secretory organelle so-called lamellar body. Also, airway surfactants can be produced by club cells and submucosal cells. Airway surfactants lack lipid constituent and do not participate in surface-tension reduction. Interestingly, SP-A, among other surfactant proteins, displays structural similarity to immune protein of the complement cascade, C1q. SP-A and SP-D involves in the clearance of apoptotic cells and direct killing of microorganisms. SP-C has also been shown to bind lipopolysaccharide (LPS) (Wright, 2005). Lung epithelium is impaired in several disease processes including pulmonary fibrosis, interstitial lung disease (ILD), chronic obstructive pulmonary disease (COPD), acute respiratory distress syndrome (ARDS), asthma, and cystic fibrosis (CF) (Koval and Sidhaye, 2017). Interstitial lung diseases (ILDs) comprise a large group of parenchymal lung diseases in which varying degrees of inflammation occurs along with progressive lung scarring and fibrosis. Several known causes and potential risk factors of ILDs include domestic and occupational exposures, tobacco smoking, gastroesophageal reflux, autoimmune connective tissue disease, drug toxicity, and genetic factors, however many ILDs are idiopathic with unknown cause. Idiopathic pulmonary fibrosis (IPF) is a subset of human ILD, accounting for 55% of all, and most common and aggressive form of idiopathic interstitial pneumonia (IIP), which occurs primarily in older adults (Barratt et al., 2018; Glassberg, 2019). IPF presents chronic and progressive lung scarring which is detected by radiographic (usually with high-resolution computed tomography) and/or histopathological (usually with lung biopsy) hallmark of usual interstitial pneumonia (UIP) (Martinez et al., 2017; Sauleda et al., 2018). UIP pattern in IPF presents honeycombing, dilation of the bronchi (traction bronchiectasis), peripheral alveolar septal thickening with alveolar remodelling. Scarring of the lung tissue causes impaired gas exchange leading to symptoms such as dyspnoea, dry cough, and general fatigue in IPF patients (Carrington et al., 2018). IPF is very rarely diagnosed before age 50 and more prevalent in males then females (Glassberg, 2019; Carrington et al., 2018). IPF shows, in the last decades, an elevated incidence and prevalence rate affecting nearly ~3 million people globally (Martinez et al., 2017). IPF can be fatal and is usually associated with end-stage lung disease, inexorable decline in lung function, progressive respiratory failure, and high mortality (Barratt et al., 2018; Martinez et al., 2017). If it is not treated with antifibrotics, epidemiological research has shown median survival for patients is changing between 3 to 5 years from the time of diagnosis with an unpredictable prognosis (Caminati et al., 2015; Glassberg, 2019). The ATS/ERS/JRS/ALAT (American Thoracic Society/European Respiratory Society/Japanese Respiratory Society/Latin American Thoracic Association) guideline for treatment of IPF includes two medications, nintedanib and pirfenidone. Nintedanib function as a tyrosine kinase inhibitor targeting growth factor pathways and pirfenidone is a pleiotropic drug with anti-inflammatory and antifibrotic effects (Sauleda et al., 2018). Disease progression is highly variable and unpredictable among individual patients. The course of IPF can be assessed by pulmonary function tests, including forced vital capacity (FVC) and diffusion capacity for carbon monoxide (DLCO), high resolution computed tomography (CT) scans, and measures of oxygenation. However, these advances do not allow to predict molecular mechanism of the disease (Herazo-Maya et al., 2013). IPF was mainly thought to be an inflammatory disease, however, anti-inflammatory therapies have provided little benefit for IPF patients (Tsujino et al., 2012). Current paradigm suggests alveolar epithelial damage is the leading factor which increases production of mediators for fibroblast migration, proliferation and differentiation into active myofibroblasts (marked by ACTA2 expression or α-smooth muscle actin, α-SMA) which can aberrantly change extracellular matrix architecture (Zepp and Morrisey, 2019). Latter paradigm based on epithelial dysfunction and abnormal wound healing hypothesized by Wanda Haschek and Hanspeter Witschi over 35 years ago (Mulugeta et al., 2015; Uhal and Nguyen, 2013). During injury, as a cardinal feature, healthy cellular landscape is destructed and after, irreversible disruption may occur in the process of recapitulating normal cellular architecture. Instead of building cellularly normal tissue, combination of aberrant tissue and scar formation can occur as a dysplastic response to chronic injury in both mesenchyme and epithelium (Zepp and Morrisey, 2019). Despite its unknown nature, several genetic and environmental factors can be associated with increased risk of IPF. Proper functioning of distal lung relies on health of AT2 cells (Mulugeta et al., 2015). Several mutations facilitating telomere attrition in AT2 cells were shown to contribute in sporadic and familial IPF in patients. These genes include TERT (encoding telomerase reverse transcriptase) and TERC (encoding telomerase RNA component), and some other telomere maintenance genes such as TINF2, DKC1, RTEL1, PARN and NAF1. Also, rare mutations in surfactant system can cause AT2 cell dysfunction. SFTPC (encoding surfactant protein C in AT2 cells) mutation can cause dysfunctional folding in the surfactant protein. Cellular responses to mutant surfactant components can result in cell stress including endoplasmic reticulum (ER) stress, mitochondrial dysfunction, intrinsic apoptosis, and might promote epithelial-mesenchymal transition; and/or it can cause derangement in quality control signatures including unfolded protein response (UPR), ubiquitin proteasome system, and macroautophagy (Mulugeta et al., 2015; Martinez et al., 2017). Variants in Toll-interacting protein (TOLLIP), TLR3, TLR9 and SPPL2C genes are associated with sporadic IPF susceptibility as well as a single nucleotide polymorphism (SNP) of mucin 5B (MUC5B). TOLLIP has a role in Toll-like receptor-mediated innate immune responses and the transforming growth factor β (TGF- β) signalling pathway. TOLLIP can interact with SMAD7 affecting intracellular trafficking and can degrade ubiquitinated TGF- β type 1 receptor (Noth et al., 2013). MUC5B, on the other hand, involves in mucociliary clearance and host defence (Martinez et al., 2017). Similarly, several other variants implicated in sporadic IPF such as OBFC1 encoding oligonucleotide/oligosaccharide-binding fold-containing 1, DSP and DPP9 involving in epithelial integrity. Besides genetic causes, other factors such as several chronic herpesvirus (Epstein-Barr virus (EBV), cytomegalovirus (CMV), Kaposi’s sarcoma herpes virus (KSHV)) infection of the lung can trigger ER stress and UPR activation in AT2 cells (Lawson et al., 2008). Changes in plasma protein concentrations or peripheral blood transcriptome may also be informative about disease presence. Evidence from 2014 study suggest that down-regulation of T-cell activation genes, in particular, CD28, ICOS, LCK, and ITK is associated with shorter transplant-free survival (TFS) in IPF patients (Herazo-Maya et al., 2013). Changes in the lower airway microbiome might be contributing inflammatory response and recurrent alveolar injury due to proliferation of pathologic organisms. Bronchoalveolar lavage (BAL) fluid analysis identified presence of *Staphylococcus* spp. and *Streptococcus* spp. in IPF patients (Huang et al., 2017). Another risk factor contributing to repetitive alveolar injury can be elevated microaspirations of gastric contents due to gastro-oesophageal reflux (Ley and Collard, 2013). Gene expression profiling studies have identified several thousand differentially expressed genes are associated with IPF transcriptome. These genes are involved in ECM formation, degradation, signalling, smooth muscle markers, growth factors, developmental pathways, immunoglobulins, complement, and chemokines. Yang et al., 2014 has shown a link between dynamic changes in IPF transcriptome and epigenetic mechanisms such as upregulation in Wnt-β-catenin signalling. Similarly, dysregulation of miRNAs was shown to have a role in IPF pathogenesis. For instance, increased expression of miR-21 was detected in myofibroblasts. miR-21 can regulate the expression of an inhibitory Smad7 and thereby pro-fibrogenic activity of TGF- β1 in fibroblasts (Liu et al., 2010). In addition, several chromatin changes and histone modifications are associated with IPF. All class I and II histone deacetylases were shown to be upregulated especially in myofibroblasts and in abnormal bronchiolar basal cells. Also, several anti-fibrotic genes were shown to be silenced including CAV1 which is a TGF- β signalling regulator, and PTGS2 which leads to loss of prostaglandin E2, an anti-fibrotic mediator. As another aspect, IPF as a fibroproliferative disease is also characterized by fibroblast resistance to apoptosis. In one study, resistance to apoptosis in fibroblasts was associated with decreased FAS expression (which encodes tumour necrosis factor (TNF) receptor superfamily member 6) in bleomycin treated mice (Martinez et al., 2017; Huang et al., 2013). Increased histone deacetylase expression (as mentioned previously) was shown to be leading to decreased histone acetylation and increased histone 3 lysine 9 trimethylation (H3K9Me3) (huang et al 2013). Also, anti-fibrotic receptor Thy1 membrane glycoprotein (THY1) was also shown to be silenced due to histone modifications (Martinez et al., 2017). Genetic architecture and epigenetic reprogramming cause aberrant activation of epithelial cells. IPF lungs exhibit phenotypically diverse epithelial populations undergoing concurrent epithelial cell proliferation, apoptosis, senescence, and partial epithelial-mesenchymal transition (EMT)-like process. AT2 cell damage and apoptosis as an initial response to injury causes aberrant lung reepithelization usually in areas of active fibrogenesis. Endoplasmic reticulum stress which is also associated with ageing implicated in damaged cells. Reepithelization is an uncertain process. Bronchiolar basal cells, AT2 cells and stem cell-like subtypes might be participating in the process. Accordingly, in some areas, IPF lung exhibit hyperplastic proliferation of epithelium marked by Ki67 antigen and other epithelial cells display EMT. Reestablishing epithelial integrity is critical and requires orchestrated regenerative response. Chronic alveolar damage is associated with dysregulated repair and regeneration in IPF; however, another paradigm suggests impaired cell function and loss of basement membrane integrity alone can disrupt proper reepithelization process and trigger the initiation of the lung fibrotic response. In that regard, tetraspanin CD151 function was investigated since CD151 is essential for firm adhesion of alveolar cells on basal membrane. Mostly laminins, among several ECM molecules, involve in cell-ECM interaction through laminin binding integrins such as α3β1, α6β4, and α6β1. Tsujino et al., 2012 has demonstrated loss of CD151 causes elevated phosphorylated Smad 2 (p-Smad2) in alveolar cells and mesenchymal-like changes (increased α-SMA and decreased E-cadherin) which then leads to myofibroblast propagation and development of IPF. Same study shows bleomycin injury causes moderate decrease in CD151 expression in WT lungs, whereas, it accelerates alveolar epithelium disintegrity with severe fibrosis in CD151 knockout lungs (Tsujino et al., 2012).

1.4 Alveolar injury models and bleomycin-induced pulmonary fibrosis

The incomplete understanding of the disease and the lack of safe and effective treatment for idiopathic pulmonary fibrosis (IPF) requires the development of the experimental models. Several alveolar injury models are developed to study inflammatory and fibrotic changes in the lung interstitium (Izbicki et al., 2002). Pulmonary fibrosis like features can be induced by using different experimental animals and different fibrotic agents in different dosing routes and regimes (Carrington et al., 2018). Animal alveolar injury models include administration of various fibrotic insults such as bleomycin, FITC, irradiation, silica, asbestos, paraquat, transgenes using viral vector, and adoptive transfer of human fibroblasts into immunodeficient mice. These models should be assessed by their own advantages and disadvantages. Bleomycin has its applications in various experimental animals including mice, rats, hamsters, rabbits, guinea pigs, dogs, and primates. Mice and rats are commonly used due to the ease of handling, characterised immune systems and possibility of using transgenic models in mice (Carrington et al., 2018). Although there are considerable differences in the structure and physiology across species, murine models provide significant insights on how different cellular populations can interact in the lung and contribute to the repair process which then can be later translated to humans (Basil and Morrisey, 2019). Bleomycin was recognized as fibrosis inducing agent and bleomycin-induced fibrosis was become the widely used murine model in IPF studies since the early 1970's (Carrington et al., 2018). Bleomycin was originally isolated from *Streptomyces verticillatus*. It was discovered as an anti-neoplastic agent which was effective against squamous cell carcinomas and skin tumors, however, its dose dependent effect on pulmonary fibrosis was discovered as an unwanted effect in some patients (Moore and Hogaboam, 2008). Bleomycin injury is achieved through production of DNA-cleaving superoxide and hydroxide free radicals which causes single and double stranded DNA breaks. This damage preferentially occurs in the lungs cause the lung, compared to the other organs, exhibit very low levels of bleomycin hydrolase which is bleomycin inactivating enzyme (Carrington et al., 2018). The fibrotic response to bleomycin mice is also strain-dependent due to differences in expression of the bleomycin hydrolase. Model provides a number of administration routes and regimes including intraperitoneal (ip), intravenous (iv), subcutaneous (sc), or intratracheal (it) delivery. Method of delivery directly affects the mechanisms of injury (Moore and Hogaboam, 2008). Low-dose intratracheal bleomycin exposure initially causes direct injury to alveolar epithelium and ultimately leads to pulmonary fibrosis (Tsujino et al., 2012). Single intratracheal dose of bleomycin model in mice leads to inflammatory response driven by acute infiltrating neutrophils which lasts 7-10 days and then transforms into a chronic lymphocytic inflammation and a fibrotic response from approximately day 14. Drug tests typically begin once the fibrosis is established which is from day 10 to 14. Terminal investigations are typically performed on day 21 or 28 after initial bleomycin dose. However, single intratracheal dose of bleomycin in mice does not recapitulate all cardinal manifestations of the human disease. It has several limitations including spontaneous resolution of fibrosis beyond 28 days, which does not model the progressive and persistent nature of IPF. Another aspect of human IPF which cannot be modelled is cough as chronic cough is a major symptom of IPF affecting over 80% of patients (Carrington et al., 2018).

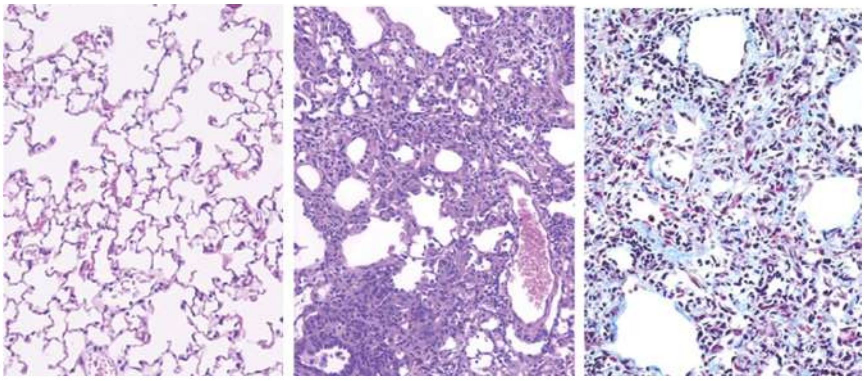

Figure 4. Characteristic pathology after bleomycin treatment. All sections are from paraffin-embedded mice lung tissue. First two images from the left shows hematoxylin and eosin (H&E) staining of untreated lung and bleomycin treated lung tissue (CBA/J mouse) on day 14 after intratracheal instillation. Image in the middle shows main characteristics of fibrosis including insterstitial thickening, inflammation, alveolar collapse, and cystic air speaces. Image on the right shows trichrome staining of a lung section from C57Bl/6 mouse on day 21 after bleomycin treatment. The blue staining indicates collagen deposition. Illustration is adapted from Moore and Hogaboam, 2008.

Bleomycin-induced fibrosis causes inflammation and elevated levels of cytokines, chemokines, and growth factors in the lung tissue. The development of fibrotic lesions is paralleled by recruitment of inflammatory cells such as monocytes, lymphocytes, and fibrocytes, and the release of chemokines such as CCL2 and CCL12 from the injured lung. Transforming growth factor (TGF)- β1, as the profibrotic cytokine, is also involved in fibrotic process. Additional mechanisms contributing to the fibrotic profile includes altered epithelial-mesenchymal interactions, circulating mesenchymal precursors and epithelial- mesenchymal transition, and their regulation by inflammatory mediators. The proinflammatory cytokine, TNF-α has roles both in promoting and repressing disease development upon bleomycin treatment. Upon its administration, lung insterstitium is injured and damage to alveolar epithelium causes leakage of fluid and plasma proteins into alveolar space, alveolar consolidation and the formation of hyaline membranes. Pathological response upon injury is well characterized and involves focal necrosis of AT1 cells and the metaplasia in AT2s. Subpleural fibrosis develops with the accumulation of infiltrating inflammatory cells and excess collagen deposition (Moore and Hogaboam, 2008). Time-course of bleomycin-induced lung fibrosis in C57Bl/6J mice was highlighted by a study from Izbicki et al. Upon intratracheal instillation of bleomycin, analysis of bronchoalveolar lavage (BAL) fluid revealed increased total cell count on days 14 and 21, and increased percentage of neutrophils on days 3 and 6 and also an increase in lymphocytes on days 6, 14 and 21. Accordingly, increase in wall area fraction on day 14 and increased hydroxyproline levels on day 14 and 21 suggest that the most suitable time point for assessing lung fibrosis is 14 days after instillation of bleomycin (Izbicki et al., 2002).

1.5 Overview of the lung matrix composition and the role of EMILINs

Extracellular matrix (ECM) is a structural support network and a complex assembly of hundreds of dynamic macromolecules with their regulatory factors. In addition to its mechanical support, ECM can play an important role in regulating key cellular events such as gene expression patterns, differentiation pathways, proliferation and cell death through selective interactions of the cell with components of the ECM. Thus, exploring specific ECM constituents can help to better understand their fine-tuning function in certain regulatory pathways in health and disease (Balestrini and Niklason, 2015). Disrupted ECM and thereby its altered function can cause and/or exacerbate several chronic and progressive lung diseases in particular such as refractory asthma, chronic obstructive pulmonary disease (COPD) and idiopathic pulmonary fibrosis (IPF). Pulmonary ECM is organized into a three-dimensional (3D) scaffold which primarily consists of basement membrane and interstitial matrices. Basement membranes are thin sheets of glycoproteins that forms a branched tissue architecture covering the basal side of the epithelium, endothelium, muscle, fat, and peripheral nerve cells. Interstitial matrices form a fibril-like meshwork that function in maintaining 3D cohesiveness and interconnects structural cell types within tissues (Burgstaller et al., 2017). It provides structural support to the lung so that the gas exchange is facilitated, and airway collapse can be prevented (Balestrini and Niklason, 2015). Functional lung matrix should be able to provide a barrier to separate blood from air, and have a hierarchial branching geometry providing high surface area for gas exchange and should contain a perfusable microvasculature that is resistant to thrombosis, and should be mechanically robust to allow for ventilation and physiological mechanical stresses (Calle et al., 2014). ECM components comprise biomechanical and biochemical properties of the lung during its development, homeostasis and disease. Lung matrix is mainly composed of collagens and elastin fibres that are interwoven with glycosaminoglycans (GAGs), fibronectin fibrils, proteoglycans (PGs), and water sequestered by PGs and GAGs (Balestrini and Niklason, 2015). Collagens are severely altered in many lung diseases. The fibrillar collagen subtypes that constitute in the lung are types I, II, III, V, and XI in which type I and III forms interstitial ECM of the alveoli and a relaxed meshwork. Collagens provide great tensile strength but low elasticity. In contrast to collagen content, elastic fibres provide low tensile strength and high elasticity which provides necessary function during respiration since the viscoelasticity in the lung parenchyma is driven by fibre-fibre contacts in the collagen-elastin ECM to some extent. Elastic fibres are structured with an inner core of elastic fibres made by protein elastin, whereas, the outer periphery of the elastic fibres contains 10-15 nm-sized microfibrils. Microfibrils are mainly structured by the large glycoproteins fibrillin-1, -2 and -3 and the other proteins such as microfibril-associated glycoprotein, fibulins, elastin microfibril interface-located protein (EMILINs) and members of the elastin-crosslinking lysyl oxidase (LOX) family are associated with microfibrils or with elastin itself. Fibulins and EMILINs as elastic fibre components, function by interacting with heteromeric transmembrane integrin receptors. EMILINs, in particular, bind integrins via their gC1q-1 domain. On the other hand, proteoglycans exhibit some significant functional roles in fibrogenesis. Decorin and biglycan, in the human lung, were shown to bind to the profibrotic cytokine TGF- β1, however only decorin not biglycan can inhibit the fibrogenic effect of TGF- β1. These proteoglycans also found prominently in matrisome of the healthy adult mouse lung (Burgstaller et al., 2017).

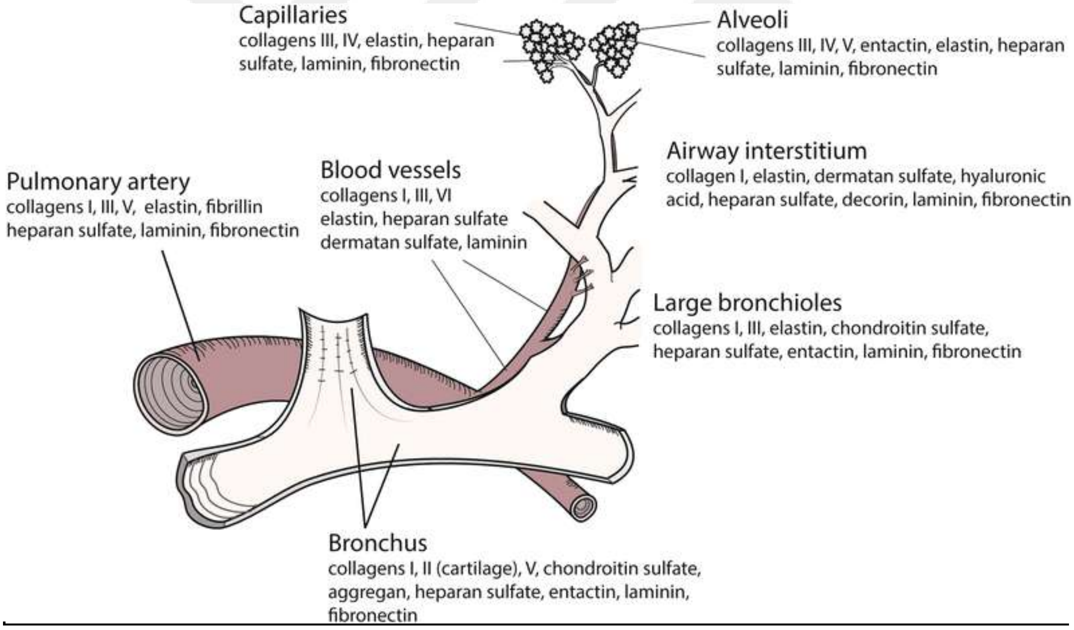

Figure 5. Regional ECM composition in the human lung. Illustration is adapted from Balestrini and Niklason, 2015.

Matrix composition of the lung varies regionally and temporally (Figure 5). The lung matrix function through the combination of coupled interactions between matrix components. Thus, overall ECM function should be considered as joint interaction of each component rather than individual contributions. The depletion or damage of one component can affect the function of the neighbouring matrix proteins, which indicate that *Emilin2* knockout model should be assessed by many aspects of cellular function during injury. Collagen and elastin mostly contribute to the mechanical nature of the lung, whereas, other components of the matrix including fibronectin and laminin are mostly involved in cell adhesion and survival processes. Fibronectin, as a cell-adhesive glycoprotein, involves in adherence of various pulmonary cell types to the ECM, and has an impact on cellular morphology, motility, and differentiation. Intriguingly, fibronectin has a significant role in growth factor storage during remodelling of the matrix (Balestrini and Niklason, 2015). Substrate stiffness in the lung matrix can affect cell shape and influence numerous cell functions including formation of focal adhesions, cell proliferation, apoptosis, growth factor or surfactant production, and stem cell differentiation. Increased matrix stiffness can play a role in the progression of fibrotic diseases of the lung (Eisenberg et al., 2011). As mentioned earlier in the lung anatomy, there are species-specific differences and spatial differences in matrix result in a dramatically different matrix compositions between species (Balestrini and Niklason, 2015).

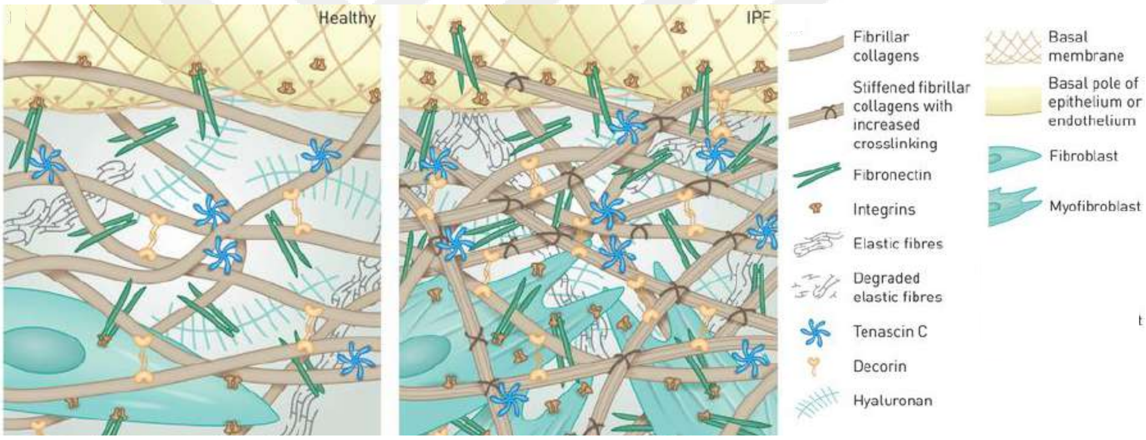

Figure 6. Comparison of pathological changes within the interstitial extracellular matrix (ECM) between healthy (on the left) and IPF lung (on the right). In IPF, myofibroblasts are present to deposit high levels of ECM and also, ECM stiffness increases by enzymatic covalent crosslinking of collagen and elastin. Illustration is modified and adapted from Burgstaller et al., 2017.

IPF is characterized by exaggerated deposition of ECM molecules (Tsujino et al 2012). One of the triggering factors for IPF is the oxidation of ECM proteins which leads to the alteration in the ECM affecting cell-ECM interactions. IPF exhibits an ongoing destruction of elastic parenchymal lung tissue due to repetitive injury and a faulty in repair mechanisms. Eventually, healthy tissue is replaced by malfunctioning stiff scar tissue which then forms large and condensed cell-ECM aggregates known as fibroblastic foci within the pulmonary interstitium. Activated fibroblasts and/or highly contractile myofibroblasts also contribute to the fibroblastic foci formation (Burgstaller et al., 2017). In order to understand the contribution of particular extracellular niche components that selectively involve in progressive scarring phenotype as seen in IPF, Schiller et al., 2015 has described time-resolved proteome of the bleomycin lung injury model.

1.6 EMILIN/Multimerin family proteins

1.6.1 Structural organization

EMILIN/Multimerin family of secreted extracellular matrix glycoproteins comprises four members (EMILIN1, EMILIN2, Multimerin1, and Multimerin2) which are characterized by unique arrangement of structural domains. At N-terminus, microfibril interface domain (EMI) domain and just in downstream, a large α-helical central domain with a high possibility to form coiled-coil structure is located. Central domain is followed by a unique varying region which possibly attains specific functions to each member. At C-terminal, a region homologous to globular C1q domain is located. N-terminal EMI domain is shared with a sequence of about 80 cysteine-rich amino acids which also define the EMI domain endowed (EDEN) superfamily in mammals. EDEN superfamily of proteins composed of three families including EMILIN/Multimerins based on their major protein domains. First family proteins are encoded by *Emu1* and *Emu2* genes and share only EMI domain. However, second family of proteins are formed by EMILIN-truncated (EMILIN-T) gene which gives a structure similar to third family except the lack of C-terminal gC1q domain. EMILIN/Multimerins as a third family harbour N-terminal EMI domain, an extended central region of 700 amino acids which may facilitate coiled-coil structures, and a C-terminal region homologous to globular domain of C1q that directs formation of trimers. EMI domain exhibits high sequence homology about 60 to 70% between family members. EMILIN/Multimerins have seven cysteine residues in their EMI domain except MMRN1 which lacks second cysteine residue. EMI has unique and conserved consensus motif of WRCCPG(Y/F)xGxxC toward C-terminus of protein. Large central region next to EMI domain possess structural homology with a high probability for coiled-coil formation. Downstream of large central coiled-coil part to C-terminal, EMILIN1 possess a leucine zipper motif and a short collagenic region of 17 GXY triplets forming a trypsin resistant triple helix. This region between central part and C-terminal gC1q is unique among EMILIN/Multimerin members conferring their unique structural and/or functional properties (Colombatti et al., 2011).

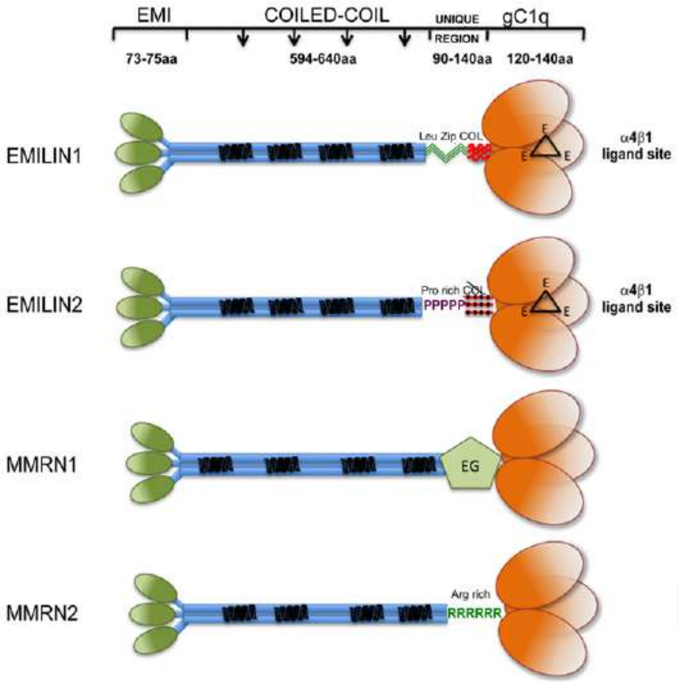

Figure 7. Graphical view of EMILIN/Multimerin family members. Illustration is adapted from (Colombatti et al., 2011).

Similarly, EMILIN2 has its unique region composed by 53 amino acid long proline-rich motif which assures more flexibility and possible interactions with other ligands. In contrast, Multimerin members MMRN1 and MMRN2 present unique regions with EGF-like domain and arginine rich sequence, respectively. C-terminal gC1q domain shares significant homology with C1q/TNF superfamily members. NMR based homology model of human 52-kDA homotrimer of the EMILIN1 gC1q domain exhibits nine β strands and a bigger central cavity lumen in contrast to other gC1q structures. Quaternary structure shows three identical subunits and each protomer conserves jelly roll topology of two β sheets made of antiparallel strands. Intriguingly, NMR locates glutamic acid residue E933 at the apex of flexible unstructured loop which hosts an interaction site via binding of three E933 carboxylate groups to integrin metal ions. Sequence alignments suggest human EMILIN2 gC1g domain share common features with EMILIN1 homolog and a corresponding loop interacting with integrin should also occur, however, there is no data available describing possible interaction (Colombatti et al., 2011; Schiavinato et al., 2016).

1.6.2 Genomic organization

EMILIN genes are located on chromosome 18p11.3 consisting of eight exons and seven introns. Genes are similar and well conserved, however, EMILIN1 is a bit more compact.

EMILIN1 and EMILIN2 has gene size of about 8kb and 40kb due to several larger intron size in latter suggesting random loss and/or uptake of intervening sequences occurred during divergent evolution of these two genes. EMILIN-2 in particular mostly expressed in mesenchymal cells and by cells of central nervous system (Colombatti et al., 2011; Doliana et al., 2001).

1.6.3 mRNA expression of EMILIN/Multimerin member genes

Mouse development studies have revealed mRNA expression of EMILIN/Multimerin member genes are mainly restricted to cardio-vascular system and mesenchymal cells with several exceptions in expression territories. At pre-implantation stage, EMILIN-2 mRNA expression was detected in both morulae and blastocysts similar with EMILIN-1. At early post-implantation stage (E6.5-E8.5), EMILIN2 was not detected in extraembryonic tissues while EMILIN-1 was strongly expressed. Extraembryonic visceral endoderm and ectoplacental cone were shown to be slightly expressing EMILIN-2. Also, EMILIN-2 was present in decidual cells mostly at the hilar side. Intriguingly, expression analyses in later stages of whole mount embryos from E6.5 to E16.5 stage revealed that EMILIN-2 was present in somites, neural tube, and mesenchyme of branchial arches, limb buds, intervertebral disks and perineural tube except endothelial cells. In addition, mesenchymal cells of most organs including lung, liver, intestine, and bladder expressed EMILIN-2 and it was still present at the latest stage (E16.5) investigated. During postnatal development, EMILIN/Multimerin genes were expressed by a variety of tissues and its deposition decreases in adulthood. In adult, EMILINs were mostly present in cardiovascular system and EMILIN-2 was particularly abundant in spleen and uterus; and weak in kidney and gut. RNA blot hybridization in various fetal and adult human tissues has revealed that mRNA expression for EMILIN-2 was much more restricted than EMILIN-1. It was more abundant in fetal heart and adult lung which is a striking evidence in the scope of this study. It was expressed in intermediate levels in peripheral leukocytes, placenta, and spinal cord and a low expression was observed in fetal brain, spleen, thymus, and lung; and in adult heart, aorta, testis, bone marrow, small intestine, thymus, lymph node and appendix. Intriguingly, adult lung had more abundant expression of EMILIN-2 as compared to fetal lung and all other tissues which has striking importance in the scope of this study. EMILIN-2 expression was shown to be induced by epidermal growth factor (EGF) and insulin-like growth factor (IGF) (Braghetta et al., 2004; Colombatti et al., 2011; Doliana et al., 2001).

1.6.4 Function

Elastic fibres in ECM give elasticity to many tissues most prominently to lung, dermis, and large blood. EMILIN-2 is found to be at the interface between main constituents of elastic fibres. Amorphous core of elastin, giving elastic properties, and the surrounding microfibrils located at the periphery and composed mainly of microfibrillin-1 and/or -2. In contrast to highly organized fibrillin containing microfibrils, amorphous elastic core is poorly organized structure. Several contributing factors for elastic fibre organization was identified as microfibril-associated proteins 1-4, latent transforming growth factor β-binding proteins 1-4, fibulins 1-2, microfibril associated glycoprotein-2 and EMILIN. EMILIN has a unique multidomain organization in contrast to other elastin-associated proteins. Given that its unique properties, EMILINs may have a role in elastogenesis and association with other ECM components. EMILINs show similarity to C1q/tumour necrosis factor superfamily of proteins due to its C-terminal gC1q domain which is also present in type VIII and type X collagens of ECM (Colombatti et al., 2011; Doliana et al., 2000).

ECM component EMILIN-2 can act as a modulator of cell survival. EMILIN-2 was shown to possess growth-suppressive effect on varying cell lines. When transfected with full-length EMILIN-2 cDNA, different cell lines have shown similar results of reduced growth and striking decrease in cell viability in a dose and time dependent manner. EMILIN-2 can induce extrinsic apoptosis during development or following tissue damage. It may also be a mechanism to maintain cell number homeostasis. EMILIN-2 can directly activate DR4 and DR5 of tumour necrosis factor (TNFR) superfamily death receptors. Mechanism was shown to be similar to TRAIL induced extrinsic apoptosis. After receptor engagement, particularly DR4, receptors were clustered and localized in lipid rafts to initiate caspase cascade. EMILIN-2 binding lead to formation of DISC assembly and activation of initiator caspases; caspase-8 and capsase-10. When challenged with EMILIN-2, HT1080 cells displayed significant level of DNA fragmentation with TUNEL assay (Mongiat et al., 2007). Proapoptotic effect was distinctive of EMILIN-2 among the other EMILIN/Multimerin family of proteins. Also, these effects were only relevant for cancer cells since viability was not affected in normal dermal fibroblasts. In a following study, ∆4 deletion mutant (90-amino acid residue-long coiled-coil fragment toward N-terminus) was shown to encompass antitumorigenic features of entire molecule. Also, this molecule mimicked three-dimensional homotrimeric structure of putative antineoplastic TRAIL although they did not share any sequence homology (Colombatti et al., 2011).

Conversely, another study indicated that EMILIN-2 can restrict tumour growth by lowering viability in an apoptosis independent mechanism. In fact, EMILIN-2 is mostly methylated in breast, lung and colorectal tumours. When challenged with EMILIN-2, breast tumour MDA-MB-231 showed less migration and significant reduction in viability, particularly in S-phase cells. EMILIN-2 treatment blocked Wnt signalling and its target genes; thereby leading to a negative regulatory effect on cell cycle progression by down modulating cyclin-D1, c-Myc and Axin2 genes. EMILIN-2 EMI domain shares 42% sequence homology with cysteine-rich domains (CRD) of the Frizzled receptors. EMILIN-2 interaction with Wnt1 showed a dose-dependent decrease in β-catenin activation and impaired Wnt activation in E293 cells. In addition, EMILIN-2 treatment reduced Wnt1-induced evasion of breast tumour cells in Matrigel. Overall, EMILIN-2, as a constituent of a tumour microenvironment, has shown tumour suppressor effect and therapeutic potential (Colombatti 2014). EMILIN-2 also stimulates angiogenesis. EMILIN-2 and ∆4 deletion mutant led to EC proliferation and motility which most likely resulted in increased tumour vascularization. Taken together, EMILIN-2 encompass antitumoral drug potential with lack of cytotoxicity to normal cells as an advantage. Efficacy might be enhanced when administered with an antiangiogenic (Colombatti et al., 2011).

EMILIN-2 was represented as an important regulator of IL-8 expression via the epidermal growth factor (EGFR/EGF) pathway. In this regard, HUVEC cells were challenged with ectopic expression of EMILIN-2 and showed upregulated IL-8 expression. Solid-phase analyses showed an interaction between EMILIN-2 and EGFR suggesting a possible crosstalk with the JAK2/STAT3 pathway leading to IL-8 expression. Also, human dermal fibroblasts (NHDF) were increased their IL-8 production due to over-expression of EMILIN-2. Over-expression of EMILIN-2 by fibroblasts induced EC motility and sprouting. In order to investigate EMILIN-2 function on vascularization in vivo, *Emilin2* knockout mouse model was established and also this model was kindly gifted and used in the single cell RNA-seq experiments of this study. *Emilin2*$^{-/-}$ was generated by replacing exon 1 of the *Emilin2* gene with *neo*^{R} cassette. MEF cultures from *Emilin2*$^{-/-}$ animal exhibited lower EGFR and STAT3 phosphorylation. To show impairing effects of lacking *Emilin2*, murine bEnd.3 brain ECs were treated with media from *Emilin2*$^{-/-}$ MEFs. They displayed lower proliferation rate and a similar migration in contrast to wild-type treatment. Also, *Emilin2*$^{-/-}$ MEF culture media led ECs to have a few and short sprouts. In addition, aortic rings from *Emilin2*$^{-/-}$ mice had lower number of microvessels. In

Emilin2$^{-/-}$ P5 pups, impaired vascularization and reduced levels of MIP-2, IL-8 murine ortholog, was observed in developing retina. Skin melanoma growth was analyzed in both wild-type and *Emilin2*$^{-/-}$ mice. EGFR dependent increase in vascularization and efficient growth rate were observed in wild-type mice with skin melanoma, whereas, these effects were significantly abolished in *Emilin2*$^{-/-}$ mice indicating that lack of EMILIN-2 expression and thereby MIP-2 resulted in impaired growth and vascularization. However, proangiogenic effects of EMILIN-2 can entail better chemotherapy efficacy by improving drug delivery in tumour. Accordingly, *Emilin2*$^{-/-}$ mice displayed less sensitivity to chemotherapy agent, melphalan, in contrast to wild-type animals (Paulitti et al., 2018).

2 Methods

2.1 Animal model

Emilin2 knockout (Emilin2$^{-/-}$ or Emilin2 KO) mice were used in the single cell RNA-seq experiments of this study and this model was kindly provided by Dr. Paolo Bonaldo (University of Padova, Italy). Emilin2$^{-/-}$ was generated by replacing exon 1 of the Emilin2 gene with neo^{R} cassette in embryonic stem cells. Positive clones for the knockout were generated by homologous recombination and injected into C57BL/6NCrl blastocysts as described by Paulitti et al., 2018. All wild-type mice used in this study were purchased from Charles River, Germany. All animals were maintained under pathogen-free conditions at the animal husbandry of the Helmholtz Zentrum München, Munich, Germany. All animal experiments 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. Animal handling, bleomycin/PBS instillation, and lung dissection procedures was performed by Maximilian Strunz (PhD Student, Schiller Lab).

2.2 Experimental design and bleomycin-induced pulmonary fibrosis model

Emilin2$^{-/-}$ and wild-type mice were both given a single dose of bleomycin hydrochloride (Sigma Aldrich) in a weight-adjusted dosage. 2U/kg bleomycin (dissolved in sterile phosphate-buffered saline, PBS) was delivered through transoral instillation. Transoral instillation method is noninvasive requiring no surgery, and the bleomycin is administered in the lung with natural inhalation by the mouse. It produces a uniform distribution of the drug due to gravity. For the endotracheal delivery of the liquid, tongue of the mice was extended to one side and bleomycin was delivered into the back of the oral cavity during inspiration (Liu et al., 2010, Strunz et al., 2019). Control animals were given equal volume of sterile PBS instead of bleomycin solution. Treated mice were monitored daily; body weights and other parameters including phenotypic changes and abnormal behaviour were evaluated and recorded until the animal is euthanized for analysis (Strunz et al). Both Emilin2$^{-/-}$ and wild-type animals were divided into two groups as PBS and bleomycin treated. Experimental end points were chosen as days 10, 14, and 28.

2.3 Lung dissection and cell isolation for single cell RNA analyses

Single-cell suspensions were prepared from the adult mouse lung to process through Drop-seq. Emilin2$^{-/-}$ and wild-type mice were euthanized at the time points of interest by cervical dislocation. Toe pinch reflex was checked to verify euthanasia. Mouse was laid down in supine position and arms and legs were pinned down. The abdomen and chest area of the animal was dampened with 80% ethanol spray. The skin of the lower abdomen was pulled up with forceps and a transverse incision was made to create a small opening into the abdominal cavity. Then, a midline incision was made in a cranial direction up to the level of the diaphragm by starting from the small incision in the abdominal cavity. Midline incision was extended gently up to the neck and then lateral incisions were made at both directions. Ribcage was exposed. Sternum and ribcage were cut carefully with appropriate scissors. Intestines were gently moved to down. Aorta was transected and blood was absorbed with paper towels. To remove blood from the lung vasculature, heart apex was hold with forceps and 10 ml syringe loaded with sterile PBS was inserted to the right ventricle of the heart to perfuse the lungs. After perfusion, trachea was exposed in the neck area and separated from surrounding tissues. Trachea was lifted up to thread a suture underneath the trachea. Then, trachea was punctured with a scissor to make an opening to insert a blunt-end needle. After inserting needle, suture was tied on the needle tip/trachea to hold it in place. 3 ml syringe was filled with 2 ml enzymatic digestion mix containing dispase (50 caseinolytic U/ml, from Corning, Cat. No. 354235), collagenase (2 mg/ml, from Merck, Cat. No. C1-28), elastase (1mg/ml, from Serva, Cat. No. 20931), and DNase (30µg/ml, form Qiagen, Cat. No. 79254). Then, syringe is attached to the needle which was inserted in the trachea and the lung was inflated with digestion mix. After filling the lung with digestion mix, lung was promptly removed from the body and placed on a glass petri dish to mince (tissue pieces at approx. 1mm2) with another 2ml digestion mix. Minced lung with digestion mix was transferred in 50 mL Falcon tube and was kept on ice until the incubation step. For mild enzymatic digestion, the lung suspension was incubated for 20-30min at 37℃ (Liu et al., 2010, Strunz et al., 2019). After this step, lung dissection process is completed, and the preparation of single cell suspension begins. The procedures of dissection and further single cell preparation steps should be undertaken as rapidly as possible to maximize recovery of the intact mRNA and viable cells for further analysis. After incubation, enzyme activity was inhibited by adding 10 ml of PBS supplemented with 10% fetal calf serum (FCS). Then, cell suspension was passed through a 40 µm strainer into a 15 ml Falcon tube and the filter was rinsed with ice cold PBS with 10% FCS. Cells were centrifuged at 300 x g for 5 min at 4℃. Supernatant was removed. RBC lysis buffer (Thermo Fisher 00-4333-57) was added at room temperature to initiate red blood cell lysis. Lysis was monitored and usually after 1- or 2-mins, cells were washed with ice cold 10 ml of PBS with 10% FCS and pelleted at 300 x g for 5 min at 4℃. After removing supernatant, cells were resuspended thoroughly in ice cold PBS with 10% FCS and counted with viability of 85% or higher (Strunz et al., 2019). Cells were aliquoted with a final concentration of 100 cells per microliter in ice cold PBS supplemented with 0.04% of bovine serum albumin (BSA) and loaded for Drop-seq. After this step, in brief, single-cell suspension was processed through Drop-seq to generate single-cell complementary DNA (cDNA) libraries attached to microbeads. Microbeads were counted, and amplified by PCR, and the 3’ end of the cDNA was prepared for sequencing. Single-cell libraries were constructed with the use of the Illumina Nextera XT DNA Sample Preparation kit.

2.4 Drop-seq run and generation of single-cell cDNA libraries

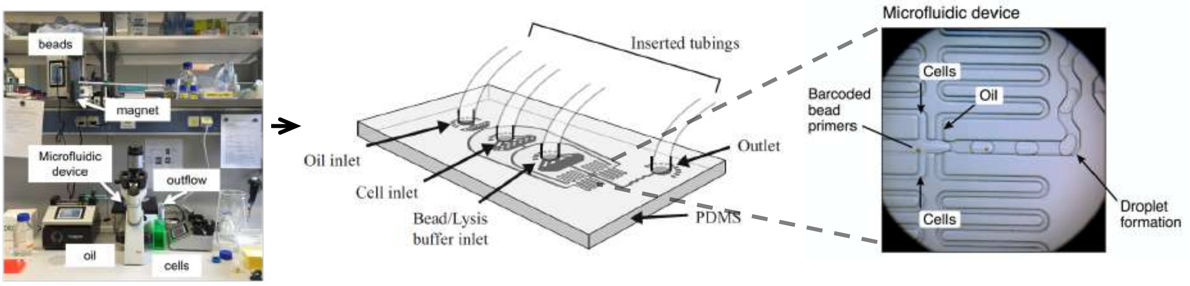

Figure 8. Illustration of the PDMS microfluidic chip with attached tubings. Illustration is adapted from Bageritz and Raddi, 2019 and Macosko et al., 2015.

Drop-seq workflow requires multiple initial steps before the downstream analysis of the data. These steps include single-cell dissociation from both whole lung of Emilin2$^{-/-}$ and wild-type mice, single-cell isolation, library construction, and sequencing (Luecken and Theis, 2019). Drop-seq profiles thousands of individual cells by encapsulating them into nanoliter-sized droplets. Each droplet ideally contains a single cell and a uniquely barcoded mRNA capture microbead. All beads contain a common sequence called PCR handle for further PCR amplification, and each bead has more than 10^8 individual primers that share the same cell barcode which identifies each unique single cell. Cell barcodes are generated with a total of 12 split-pool cycles of equally sized oligonucleotide synthesis reaction which yields 4^{12} (16,777,216) possible sequences to differentiate individual cells. Just after cell barcode, each primer has a unique molecular identifiers (UMIs) to label individual mRNA transcripts from a single cell. UMIs are generated after split-and-pool synthesis cycles of cell barcode by subjecting beads together to eight rounds of degenerate synthesis with all four DNA bases so that each primer receives one of 4^8 (65,536) possible UMI sequences to label individual mRNA. Lastly, primer sequence ends with a 30-bp oligo dT to capture individual mRNA. After droplet formation, each cell is lysed within a droplet by the action of lysis buffer. Cell membranes break down and the released mRNA binds to the primers on its companion microparticle via the 30-bp oligo dT sequence at the end of all primers. After capturing on beads, intracellular mRNAs are reverse-transcribed to cDNA molecules and amplified to initiate library construction. Concisely, cellular cDNA libraries have a unique label (cell barcode) from the respective cell and a label (UMI) for each intracellular mRNA read. After reverse transcription of mRNAs, each bead is called single-cell transcriptomes attached to microparticles (STAMPs). cDNAs of STAMPs are further amplified by PCR to increase its probability of being measured. Despite being amplified, UMIs makes it possible to distinguish between amplified copies of the same RNA and separate mRNA molecules from the same gene (Macosko et al., 2015). After library construction and amplification, libraries are pooled together and sent for sequencing. In the following sections, Drop-seq run is explained in detail. All buffers were prepared by using nuclease-free water (Ambion, Cat No. AM9932). Centrifugation steps were carried out at 4℃. A detailed protocol provided online on the group home page of Steve McCarroll by Macosko and colleagues (Drop-seq-Protocolv3.1-Dec-2015) and the methods provided by Bageritz and Raddi, 2019. These two sources collectively served the basis for the following sections of methods parts regarding the Drop-seq run.

2.4.1 *Preparation of Barcoded Beads*

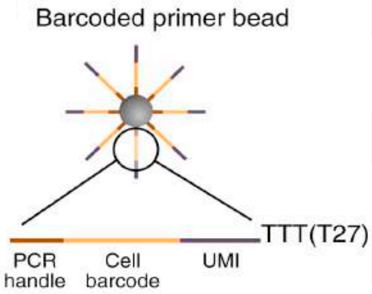

Figure 9. Sequence of primers on the microparticle. The primers on all beads contain a common sequence (“PCR handle”) to enable PCR amplification after STAMP formation. Each microparticle contains more than 10^8 individual primers that share the same “cell barcode”. Figure legend and the illustration is adapted from Macosko et al., 2015.

Upon arrival, DNA-barcoded beads (120/µl, ChemGenes, USA) were washed twice with 30 ml absolute ethanol (PanReac AppliChem, Cat. No. A3678, 1000, stored at room temperature), and twice with 30 ml TE-TW (Tris-EDTA – Tween buffer: 10mM Tris pH 8.0 – Invitrogen, Cat. No. 00679473, 1mM EDTA – Invitrogen, Cat. No. 15575-038-100ml, 0.01% Tween-20 – Sigma Aldrich, Cat. No. 93773-250G, filtered by using 0.22 µm pore size filter – Merck, Cat. No. SLGP033RS, stored at room temperature). Beads were passed through a 100 µm strainer and counted using C-Chip disposable hemocytometer (NanoEntek, Cat. No. DHC-B01). For counting, 20 µl of beads were loaded with P200 to achieve even distribution of beads on the chip. After counting beads were aliquoted with a final concentration of 120.000 beads per aliquot (for a microfluidics device that produces droplets of 125 µm) and stored at 4℃. For pelleting barcoded beads, DNA low-binding tubes (Sarstedt, Cat. No. 72.706.700) were used. Before the Drop-seq run, following arrangements were prepared.

2.4.2 Prerun Setup

Three syringe pumps were arranged close to the inverted microscope with cell and oil syringe pump in a horizontal position and the bead/lysis buffer syringe pump in an upright position with the syringe pointing down. Magnetic stirrer was positioned close to the bead/lysis buffer syringe pump. Three tubings (Smiths, Fine Bore Polythene Tubing, 0.38mm ID, 1.09mm OD, Cat. No. 800/100/120) were cut of the appropriate lengths for the oil, bead/lysis buffer, cell inlet, and outlet, respectively, by using a scalpel and making a slight angle at the ends of tubings. Microfluidics chip was placed on the microscope stage and a clean device was chosen for the run without any fabrication defects. Droplet generation oil (Bio-Rad, Cat. No. 1864006) was loaded into 20 ml syringe (. Oil can be filtered with 0.22 µm pore size filter before using. Air bubbles were removed by bringing syringe in an upright position and slowly pushing the plunger of the syringe until oil comes out of the free-end tubing. Then, 26 G needle (B. Braun, 26G, Cat. No. 466 5457) was attached to syringe and oil tubing was affixed to needle carefully. Oil syringe was placed in the pump. Then, the right syringe type and flow rate for oil (15.000 µl/h for oil) was set in the pump. Tubing was primed with oil and free-end tubing of oil was inserted carefully in the oil inlet of the selected device. Outlet tubing was inserted into outlet channel of the same device and its free-end was placed into a 50 ml waste Falcon tube. 120.000 bead aliquot was spinned down in a tabletop centrifuge and the TE-TW buffer was removed and beads were resuspended in 950µl of lysis buffer (6% Ficoll PM-400 - Sigma Aldrich, Cat. No. F5415-50ml, 0.2% Sarkosyl - Sigma Aldrich, Cat. No. L7414-10ml, 20 mM EDTA, 200 mM Tris pH 7.5 – Sigma Aldrich, Cat. No. T2944-100ml, in water, filtered by using 0.22 µm pore size filter, stored at room temperature) without DTT (PanReac AppliChem, Cat. No. A2948,0025, aliquots are stored at -20℃, repeated freezing and thawing cycles was avoided). Cell lysis buffer is only complete when 50µl, 50mM DTT was added in the lysis buffer/bead aliquot. A clean magnetic mixing disc (V&P Scientific, Cat. No. VP 772DP-N42-5-2) was placed into a 3ml syringe (B.Braun, NORM-JECT, Luer-Lock Solo, Cat. No. NJ-4606701) before loading the beads. Beads were loaded and air bubbles was removed from the syringe. After affixing 26G needle and bead tubing, syringe was placed in the pump with the right syringe and flow rate (4.000 µl/h for beads) settings. Magnetic stirrer was turned onto mix beads and positioned in a good angle so that beads are always homogenously distributed in the syringe. Tubing for beads was primed to push out air bubbles and free-end of tubing was inserted in bead/lysis buffer inlet of the selected device. For cells, single cell suspension of 100 cells/µl was loaded into a 3ml syringe. 26 G needle was attached and tubing for cells was affixed. Cell syringe was placed in the pump with the right settings for syringe type and flow rate (4.000 µl/h for cells). Tubing for cells is primed to remove air bubbles and its free-end was placed in the cell inlet of the selected device.

2.4.3 Cell Encapsulation and STAMP Generation

Drop-seq run begins by first starting the cell, then the bead/lysis buffer and lastly the oil syringe pump (wait for few seconds between each step). With this order, cell lysis outside the droplets is avoided. Flow was waited to be stabilized for about 30s. Flow was constantly checked for the production of uniform droplets which should make a blurry elongated triangle at the droplet generation junction. In order to evaluate the droplet quality, some droplets can be placed on a glass petri dish and their appearance can be checked under microscope. Once the stable droplets were observed, collection was started by placing outgoing tubing into a fresh 50ml falcon tube for the selected sample. Collection tube was always kept on ice. Collection time was around 20 min. Collection time can be changed according to desired cell number to be collected and accordingly, the volumes of oil, bead/lysis buffer and cells should be adjusted. In order to reload any solution again, pumps can be stopped in the order of bead/lysis buffer, cell and lastly the oil syringe. After waiting for few seconds, all three ingoing can be pulled out of the device at the same time. When flow is stopped, outgoing tubing can also be pulled out. In order to start again, oil, cell, and bead/lysis buffer tubings should be connected. Formation of droplets was regularly checked during entire collection time. When the sample collection is finished, flow was stopped in the order of bead/lysis buffer, cell and the oil. Procedure is preceded by droplet breakage. Sample was always kept on ice. First, most of the oil was removed from the bottom of the collection tube using a P1000. Then, 30ml of 6X SSC (diluted from 20X SSC: saline sodium citrate buffer, Sigma Aldrich, Cat. No. S6639-1L, stored at room temperature) was added on top of beads. Breakage was achieved by adding 1 ml of perfluorooctonal (PFO, Alf Aesar, Cat. No. 10188144, stored at room temperature, used only under the fume hood) and sample was shaken firmly by ten times. Amount of PFO should be adjusted if the collection time is longer than 15min. Sample was centrifuged for 1min at 1000 x g with the cap being loosened. After centrifuge, a relatively thin, white line should be observed as an indicative of successful breakage. Upper aqueous phase was removed by using a 20 ml pipette until few millilitres remains above the interface. Beads lying on top of the interface were kicked up by addition of 30 ml 6X SSC. Oil was waited to sink for few seconds and then the supernatant with beads was transferred into a fresh 50 ml falcon tube. Transfer of oil should be avoided at this step. Beads were centrifuged for 1min at 1000 x g, and supernatant was removed until about 1 ml liquid remains at the bottom. Beads were resuspended in the remaining 1ml of SSC and transferred into a fresh 1.5 ml tube. Remaining beads on the bottom surface of the falcon tube was transferred by using about 500 µl 6X SSC and transferred into same 1.5 ml tube. Beads were washed twice with 1ml of 6X SSC, and oil was removed if any observed. After this step, reverse transcription of hybridized mRNA transcripts begins. Firstly, beads were washed with about 300 µl 5X RT buffer. Then, 190 µl RT mix (1X RT buffer, 4% Ficoll PM-400, 1mM dNTPs – Promega, Cat. No. U1515, stored at -20℃, 200 U RNase Inhibitor – NxGen, Cat. No. F83923-1, 2.5 µM Template Switch Oligo – TSO: 5'-AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG – 3', in water, stored at -20℃, aliquoted as 190 µl each) was mixed with 10 µl of Maxima H Minus RTase (Thermo Scientific, Cat. No. EP0752) just prior to start with reverse transcription. 200 µl of RT mix was used for a maximum of 90.000 recovered beads. Beads were spinned down in a tabletop centrifuge and supernatant was removed. Beads were resuspended in RT mix including enzyme. Sample was incubated for 30 min at room temperature with gentle rotation and then transferred to 42 ℃ and incubated for another 90 min with the same rotation. After incubation, reaction was stopped by washing sample once with 1 ml of TE-SDS (Tris-EDTA – SDS buffer: 10 mM Tris pH 8.0, 1 mM EDTA, 0.5% SDS – Sigma Aldrich, stored at room temperature, aliquoted as 10ml each) and twice with TE-TW. After this step, reverse transcription is completed, and sample can be stored at 4℃. Procedure is preceded with exonuclease treatment to remove unused primers on the beads. Sample was washed with 1 ml of 10 mM Tris pH 8.0. Beads were pelleted by a tabletop centrifuge and supernatant was removed. Exonuclease I digest was carried out by adding 200 µl of the exonuclease mix (1X Exo I buffer -New England Biolabs, Cat. No. B0293S, and 200 U Exonuclease I -New England Biolabs, Cat. No. M0293L, in water, reagents were kept separately at -20℃) to the pelleted beads. 200 µl of the exonuclease mix was freshly prepared on ice and used for up to 90.000 beads. Sample was incubated for 45 min at 37℃ with gentle rotation. Reaction was stopped by washing sample with 1 ml TE-SDS once and 1 ml TE-TW twice. After this step, sample can be stored at 4℃. Procedure is preceded by library preparation.

2.4.4 Library Preparation

Beads were resuspended (with P1000) in 1 ml TE-TW solution for counting. 20 µl of beads were loaded with P200 to achieve even distribution of beads on the chip. Beads were counted by using a C-Chip hemocytometer and bead concentration was calculated. Beads were aliquoted first for the desired number of STAMPs taking into account the estimation that ~5% of beads yield to STAMPs. Before aliquoting for PCR, beads were washed with 1 ml of water. For the pre-amplification of cDNA, each PCR reaction tube contained 200 STAMPs and 50 µl of PCR mix (0.8 µM Smart PCR primer: 5’ – AAGCAGTGGTATCAACGCAGAGT – 3’, 100 µM, 1x KAPA HiFi Hotstart Ready-mix, KAPA Biosystems, Cat. No. KM2602, in water). PCR mix was prepared freshly on ice. Tubes were mixed well and transferred to thermocycler. Cycle conditions were 3 min 95℃, 2 cycles of 20 s 98℃, 45 s 65℃, 3 min 72℃, followed by 7 cycles of 20 s 98℃, 20 s 67℃, 3 min 72℃, then 5 min at 72 ℃. After PCR, amplified cDNA was transferred and pooled into a new 1.5 ml tube without carrying any beads. PCR product was purified twice by using clean-up beads (CleanNA, Cat. No. CPCR-0050, stored at 4℃) with a bead-to-DNA ratio of 0.6. Before mixing with cDNA, clean-up beads were mixed well with vortex and only used at room temperature. After mixing, sample was incubated for 5 min at room temperature. Then, placed in a magnetic rack (Life Technologies, DynaMag-2, Cat. No. 12321D) and waited until the supernatant is clear. Supernatant is removed and clean-up beads were washed with freshly prepared 80% ethanol twice. Clean-up beads were left for air-dry for 15 min and then eluted in 50 µl of water. Clean-up procedure was repeated for a second time and cDNA was eluted in 12 µl of water. Transcript integrity, purity and amount were assessed by loading 1 µl of the sample on DNA High Sensitivity Chip (Agilent, Cat. No. 5067-4626) on the 2100 Bioanalyzer (Agilent). In this step, cDNA should have the average sizes from 1.300-2.000bp for human and mouse. After this step, procedure is preceded with cDNA tagmentation and amplification using the Nextera XT DNA Library Preparation Kit (Illumina, Cat. No. 15032354). 1 ng of pre-amplified cDNA was combined with water to a total volume of 5µl. Sample was treated with 5µl of Nextera TD Buffer and 5µl of Amplicon Tagment Enzyme. Sample was mixed and incubated at 55℃ for about 5 min. After incubation, sample was immediately treated with 5µl of Neutralization Buffer (NT Buffer, room temperature) and incubated for 5 min at room temperature. Amplification was carried out by using 15 µl Nextera PCR mix,1 µl 10 µM New-P5-SMART PCR hybrid Oligo, 1 µl 10 µM Nextera N70X oligo, and 8 µl of water per reaction. Tagmented and amplified library was purified with the same procedure as mentioned above and quality of the final library was assessed by running 1 µl of the sample on a Bioanalyzer High Sensitivity Chip. At this step, high-quality libraries should show a smooth profile, with a size average from 500 to 680 bp. After this step, samples were pooled (multiplexed) with equimolar pooling and stored at -20 ℃.

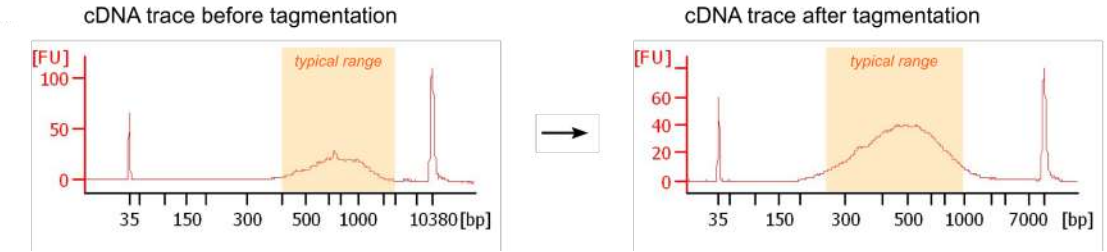

Figure 10. Illustration of the typical bioanalyzer results.

2.4.5 *Sequencing strategy*

After equimolar pooling, libraries were quantitated by an Agilent Bioanalyzer using the High Sensitivity DNA analysis kit and then libraries were sent for sequencing with the depth of 3.3B to 4.1B reads (NovaSeq 6000; Illumina).

2.5 Computational analysis of Drop-seq single cell data

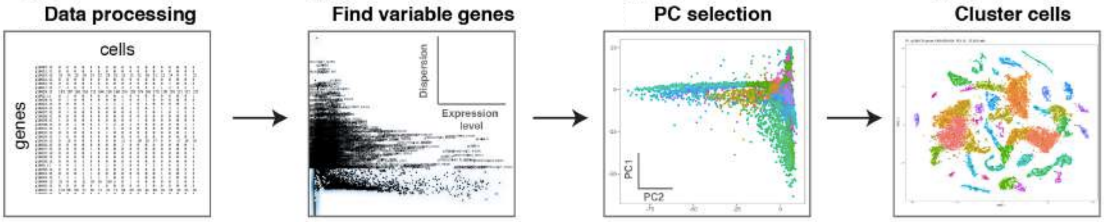

Figure 11. Schematic of scRNA-seq analysis workflow. Raw sequencing data is first processed and aligned to produce count matrices. Then the count matrices are further analysed with pre-processing and downstream analysis. Scanpy was used for the identification of genes with relatively high average expression and variability, and subsequent principal component analysis. Clusters were visualized on Uniform Approximation and Projection (UMAP) plots. Illustration is adapted from Siebert et al., 2019.

Sequencing produces raw data of mRNA reads, which needs to undergo quality control, and grouping based on their assigned barcodes (demultiplexing), and alignment in read processing pipelines. Count matrices are generated from reads by further demultiplexing based on UMI codes. Resulting matrices have the dimension number of cells and number of transcripts (Luecken and Theis, 2019). Sequencing reads were mapped to a reference transcriptome and single-cell expression data was analysed with Python-based toolkit Scanpy (Wolf et al., 2018). Before starting the analysis with Scanpy, SoupX (provided as a R package) was applied to correct the count data. SoupX aims to model and remove ambient RNA expression profiles from single-cell transcriptomic data. Cell free mRNAs or the ambient mRNAs can significantly confound the biological interpretation of the scRNA-seq data so that the SoupX method is necessary to increase the accuracy and cellular specificity of the transcriptional reads. Ambient gene expression can originate from the mRNA contamination of lysed cells prior to library construction. SoupX provided a corrected table of counts which was used for the next steps of analysis (Young and Behjati, 2020). Steps of a Drop-seq scRNA-seq analysis starts with the pre-processing which covers quality control, normalization, data correction, feature selection, and dimensionality reduction. Then the pre-processing is followed by the downstream analysis such as marker gene identification, cluster annotation and differential expression tests.

3 Results

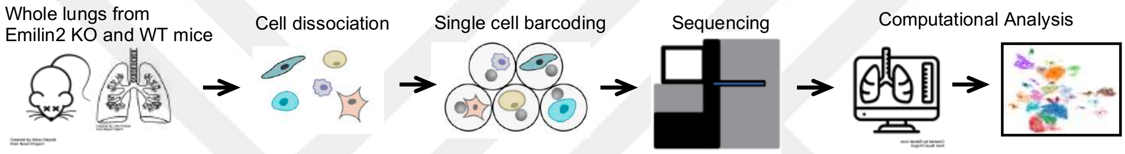

Figure 12. Experimental design – whole lung single-cell suspensions of Emilin2 knockout and wild-type mice were analyzed using the Dropseq workflow. Illustration is modified from Macosko et al., 2015.

Table 1. Number of mice (*n*) used for each condition (WT: wild-type, KO: knockout, Bleo: Bleomycin treatment, PBS: control). Cell counts are estimated numbers and does not represent actual counts in the analysis.

		Time points			
		d10	**d14**	**d28**	**Total number of mice per condition**
Condition	**WT PBS**	5	3	3	11
	KO PBS	3	2	3	8
	WT Bleo	9	3	4	16
	KO Bleo	7	5	5	17
	Total number of mice per time point	24	13	15	
	Total cell count per time point	31600	24200	36942	
	Total cell count	**92742**			

~43,000 single-cell transcriptomes from bleomycin-induced murine lungs were generated by using Drop-seq. Gene expression changes were uncovered collectively for three time points of day 10, day 14, and day 28 after single dose transoral administration of bleomycin. This data revealed the dynamics of gene expression that occur after bleomycin injury in *Emilin2*$^{-/-}$ and wild-type mice. Fifty-two droplet-based single-cell RNA-seq (Drop-seq) libraries were prepared from dissociated whole murine adult lungs. In the pre-processing of scRNA-seq data, cell quality control (QC) metrics were assessed to make sure that all cellular barcode data correspond to viable cells. QC measurements ensures sufficient data quality for the downstream analysis. QC was performed based on three covariates: the number of counts per barcode (count depth), the number of genes per barcode, and the fraction of counts from mitochondrial genes per barcode. Outlier barcodes such as barcodes with a low count depth, few detected genes, and a high fraction of mitochondrial counts were filtered out by thresholding since they can be indicative of cells whose cytoplasmic mRNA has leaked out through a broken membrane during gathering of single cell suspension, and thus, only mRNA located in the mitochondria can still be conserved (Luecken and Theis, 2019). Also, the high mitochondrial reads can indicate that a cell was stressed or dying (Siebert et al., 2019). Reason for a high-count depth threshold in UMI counts is that the cells with unexpectedly high counts may represent doublets (or multiplets) which can be caused by capturing multiple cells in one single droplet (Luecken and Theis, 2019). Cut-offs were performed for minimum of 100 detected genes and 150 to 6,000 unique molecular identifiers (UMIs), resulting in a dataset with a detected median of 366 genes and 512 UMIs per cell. Total cell and gene number detected were 49,158 and 36,609 before applying filtering cut-offs, respectively. 304 cells that have more than 6000 counts, 1335 cells that have less than 150 counts, and 44 cells that have less than 100 genes expressed were filtered out. For doublet detection in single-cell RNA-seq data, Phyton code for identifying doublets called Scrublet (Single-Cell Remover of Doublets; Wolock et al., 2019) was used and identified only 4 doublets in the dataset. QC metrics were left relatively permissive. Genes that are expressed in at least three cells were kept and thereby, 9452 genes that are detected in less than 3 cells were filtered out. After applying 2% mitochondrial reads cut-off, count matrix had only 43,373 cells and 27,157 genes. After this point, the data were normalized and scaled by using a so-called size factor which is proportional to the count depth per cell. Normalization is necessary since the library preparation steps including reverse transcription, further amplification and sequencing can cause variability in count depths for identical cells. Normalization overcomes this issue and makes count data comparable between cells by scaling count data to obtain correct relative gene expression abundances between cells (Luecken and Theis, 2019). After normalization, data matrices are log-transformed. Although it is normalized, data can still contain unwanted variability due to technical and biological covariates. To remove effects of the cell cycle on the transcriptome, cell cycle scores were computed and linear regression against cell cycle score was performed. Additionally, mitochondrial gene expression was also regressed out to better interpret scRNA-seq data. After this step, the analysis is continued with downstream analysis tools. High number of genes from single cells causes dataset to be complex with high dimensionality. In this dataset, many genes had zero counts for some cells and not every gene was assumed to be informative of variability. Feature selection was applied to reduce high dimensions, and thereby to reduce the noise in the data, and to keep only informative genes. 9592 highly variable genes (HVGs) variable in at least three samples were selected for the downstream analysis. Then the principal components (PCs) was calculated on these variable genes. Principal component analysis (PCA) reduce the data to its essential components. PCA generates reduced dimensions by maximizing the captured residual variance in each further dimension (Luecken and Theis, 2019). Top 30 PCs were selected to be used in further dimensionality reduction in visualization by the Uniform Approximation and Projection method (UMAP; McInnes et al., 2018). To better describe biological insights, cells were grouped into clusters based on similarity of their gene expression profiles. Louvain algorithm was used with the resolution parameter at one, which determines the scale of the cluster partition. Cluster identity was annotated by matching gene signatures of each cluster with the published gene expression patterns from the literature. Some marker genes were also deciphered by using online webtool as a reference dataset for the bleomycin-induced injury model, which is accessible at http://146.107.176.18:3838/Bleo_webtool_v2/. List of top ranked genes (marker genes) for clusters was generated by applying differential expression (DE) testing between the cells in one cluster and all other cells in the dataset. Wilcoxon rank-sum test was used to rank genes by their difference in expression levels. In the clustering analysis, cells were nicely separated according to cell lineage, and populations were observed in the expected lineage. Cell clustering was biologically meaningful. A clear separation of clusters based on time points in certain cell types such as macrophages and capillary endothelial cells was evident in the UMAP. This separation might be caused by different states of the same cell type since they belong to different stages of the injury model. Further analysis was carried out by separating day 28 cells from the other samples. Same cut-offs were applied and resulting dataset had 22,102 cells and 23,178 genes. Several marker genes for subtypes of macrophages were checked on this dataset. On the other hand, the identification of genes that are differentially expressed between Emilin2$^{-/-}$ and wild-type mice at only single time point of day 28 provides a basis for understanding the highly regulated gene sets and defining functional roles of Emilin2 after injury. To show differentially expressed genes between two treatment conditions; bleomycin (*n*=6) and PBS control (*n*=6), and two genotype conditions; Emilin2$^{-/-}$ (*n*=6) and wild-type (WT) (*n*=6), *in silico* bulk sample were created by summing all counts within an individual mouse sample for day 28 after bleomycin treatment. First, the reading depth is corrected by dividing counts by the sum. Then, genes with zero counts were filtered out which resulted in a list with 6418 genes. After normalization and filtering, counts were log transformed, and z scored and *t*-test statistics was applied for differential expression testing. Differential gene expression testing between treatment conditions identified 76 significantly regulated genes (*t*-test statistics, false discovery rate (FDR) <5%) between two different treatment conditions of bleomycin (*n*=6) and PBS control (*n*=6). Resulting data matrix with significant genes was used as input for the principal component analysis (PCA), volcano plots, and hierarchical clustering in Figure 4. With similar approach, differential expression testing was applied between two genotype conditions and no significance was observed. Resulting data matrix similarly used as an input for the volcano plot, pathway enrichment analysis, and hierarchical clustering şn Figure 5 by selecting top regulated genes with lower *p* values. Using gene annotation enrichment analysis on the Gene Ontology (GO) terms (Benjamini-Hochberg test was applied with a false discovery rate (FDR) < 10%), several pathways were found upregulated including laminin complex, regulation of neutrophil apoptosis, and neutrophil aggregation in the wild-type bleomycin treated mice when compared to *Emilin2* knockout mice at day 28 after bleomycin treatment.

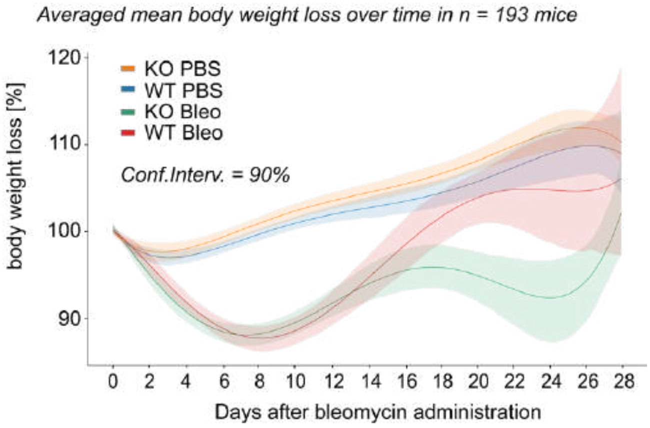

Figure 13. Changes in the body weight after bleomycin injury until day 28. From initial time points to day 10, bleomycin treated animals lose their weight due to inflammation and fibrosis. After day 10, both bleomycin treated mice regained weight, however, weight loss of *Emilin2* KO mice is continued in the later stage. Control mice (PBS) gradually gained weight. Illustration is adapted from Strunz et al., 2019.

Prolonged weight loss in the later stages of bleomycin model suggests a functional role for Emilin2 in persistent fibrosis.

Table 2. Summary statistics shows median of counts and genes, percentage of mitochondrial reads and number of cells for each sample before applying filtering cut-offs.

	identifier	n_counts	n_genes	percent.mito	n_cells
0	muc6642	567,5	372,5	0,020961	378
1	muc6643	495	344,5	0,017212	544
2	muc6644	576	348	0,091159	589
3	muc6645	488	331	0,025926	225
4	muc6646	473	327	0,027272	286
5	muc6647	522	347	0,17885	387
6	muc6648	461	298	0,170886	209
7	muc6649	508	374	0,042685	219
8	muc6650	0	0	0,141957	350
9	muc6651	547	353	0,145438	325
10	muc6652	592,5	409	0,108292	478
11	muc6653	377	290	0,045403	198
12	muc6654	441	316	0,08167	435
13	muc5215	480	354	0,043125	450
14	muc5216	497,5	361,5	0,034079	298
15	muc5217	495	360	0,030295	412
16	muc5218	489	335	0,074079	1003
17	muc5219	524	373	0,041062	1751
18	muc5220	455	309	0,039676	521
19	muc5221	444	322,5	0,038865	820
20	muc5222	420	334	0,060764	741
21	muc5223	368	275	0,027532	81
22	muc5224	493	347	0,014649	483
23	muc5225	0	0	0,081768	691
24	muc4668	617,5	399	0,023104	466
25	muc4669	537	371	0,036442	1263
26	muc4672	529	374	0,063495	489
27	muc4673	490	339	0,071517	779
28	muc4674	524	365	0,05749	633
29	muc12927	433	296	0,075816	2596
30	muc12928	528	330	0,132875	1680
31	muc12929	312	198	0,07192	561
32	muc12930	534	353	0,076086	1560
33	muc12931	328	252	0,050206	2280
34	muc12932	226	176	0,05296	3233
35	muc12933	438	291	0,121941	1200
36	muc12934	447	331	0,040709	1440
37	muc12935	431	301	0,033342	960
38	muc12939	255	205	0,03788	2156
39	muc12940	257	195	0,028457	1661
40	muc13932	858	526,5	0,024115	1680
41	muc13933	947,5	588,5	0,059642	1560
42	muc13934	1183	734	0,048353	1200

43	muc13935	1192	753,5	0,01519	1200
44	muc13936	1291,5	768	0,019238	1200
45	muc13937	263	160	0,116437	549
46	muc13938	1288,5	743,5	0,048527	1200
47	muc13939	663	423	0,057343	1200
48	muc13940	312	171	0,164549	938
49	muc13941	862	542,5	0,06758	1200
50	muc13942	1349,5	810,5	0,052661	1200
51	muc13943	1905	1026,5	0,016916	1200
				Sum n_cells	49158

Table 3. Summary statistics shows median of counts and genes, percentage of mitochondrial reads and number of cells for each sample after applying filtering cut-offs.

	identifier	**name**	**n_counts**	**n_genes**	**percent.mito**	**n_cells**
0	muc6642	WT_PBS_d10	563	369	0,019874	371
1	muc6643	WT_Bleo_d10	493	344	0,017229	543
2	muc6644	KO_Bleo_d10	619	385	0,05491	507
3	muc6645	KO_PBS_d10	488	331	0,025926	225
4	muc6646	WT_PBS_d10	478	331	0,022082	282
5	muc6647	KO_Bleo_d10	516	387	0,100806	195
6	muc6648	WT_Bleo_d10	395,5	283,5	0,105195	120
7	muc6649	KO_PBS_d10	508	374	0,042685	219
8	muc6650	KO_Bleo_d10	868,5	536	0,107927	98
9	muc6651	KO_Bleo_d10	535	366	0,102662	235
10	muc6652	WT_Bleo_d10	576	403	0,075449	380
11	muc6653	WT_Bleo_d10	377	290	0,045403	198
12	muc6654	KO_PBS_d10	432	322	0,066025	397
13	muc5215	KO_Bleo_d10	481	355	0,04277	449
14	muc5216	KO_Bleo_d10	497,5	361,5	0,031421	290
15	muc5217	KO_Bleo_d10	495	360	0,030295	412
16	muc5218	WT_Bleo_d10	517,5	366,5	0,044564	896
17	muc5219	WT_Bleo_d10	524	373	0,040089	1733
18	muc5220	WT_Bleo_d10	460	310,5	0,036854	516
19	muc5221	WT_PBS_d10	443	323	0,037268	810
20	muc5222	WT_PBS_d10	427	346	0,048269	695
21	muc5223	WT_Bleo_d10	368	275	0,027532	81
22	muc5224	WT_Bleo_d10	493	347	0,014649	483
23	muc5225	WT_PBS_d10	786	482,5	0,072852	280
24	muc4668	KO_Bleo_d14	608	393	0,023536	457
25	muc4669	KO_Bleo_d14	534,5	370	0,035353	1248
26	muc4672	KO_PBS_d14	563	384	0,049991	465
27	muc4673	KO_Bleo_d14	527	369	0,039375	689
28	muc4674	WT_Bleo_d14	571	378	0,03932	591
29	muc12927	WT_Bleo_d28	450	316	0,049975	2205
30	muc12928	WT_Bleo_d28	566	378	0,020483	1205

31	muc12929	WT_Bleo_d28	322	208	0,039629	483
32	muc12930	KO_Bleo_d14	549	374	0,039464	1294
33	muc12931	WT_PBS_d14	330	253	0,047887	2249
34	muc12932	KO_PBS_d14	239	187	0,04259	2805
35	muc12933	WT_PBS_d14	443,5	312	0,056076	804
36	muc12934	KO_Bleo_d14	459	337	0,032472	1395
37	muc12935	WT_Bleo_d14	433	302	0,03209	953
38	muc12939	WT_Bleo_d14	270	216	0,034039	1981
39	muc12940	WT_PBS_d14	281	210	0,027032	1481
40	muc13932	WT_PBS_d28	870	534	0,017884	1621
41	muc13933	WT_PBS_d28	1082	655	0,009938	1269
42	muc13934	WT_PBS_d28	1235	768	0,01737	1051
43	muc13935	KO_PBS_d28	1190	753,5	0,012986	1174
44	muc13936	KO_PBS_d28	1282	761,5	0,019296	1188
45	muc13937	KO_PBS_d28	268	178	0,058501	367
46	muc13938	WT_Bleo_d28	1320,5	764	0,02848	1082
47	muc13939	WT_Bleo_d28	693	450	0,026867	1063
48	muc13940	WT_Bleo_d28	306	204	0,043302	595
49	muc13941	KO_Bleo_d28	899	564	0,044694	1057
50	muc13942	KO_Bleo_d28	1423	863	0,017262	1037
51	muc13943	KO_Bleo_d28	1867	1010	0,017286	1149
					Sum n_cells	43373

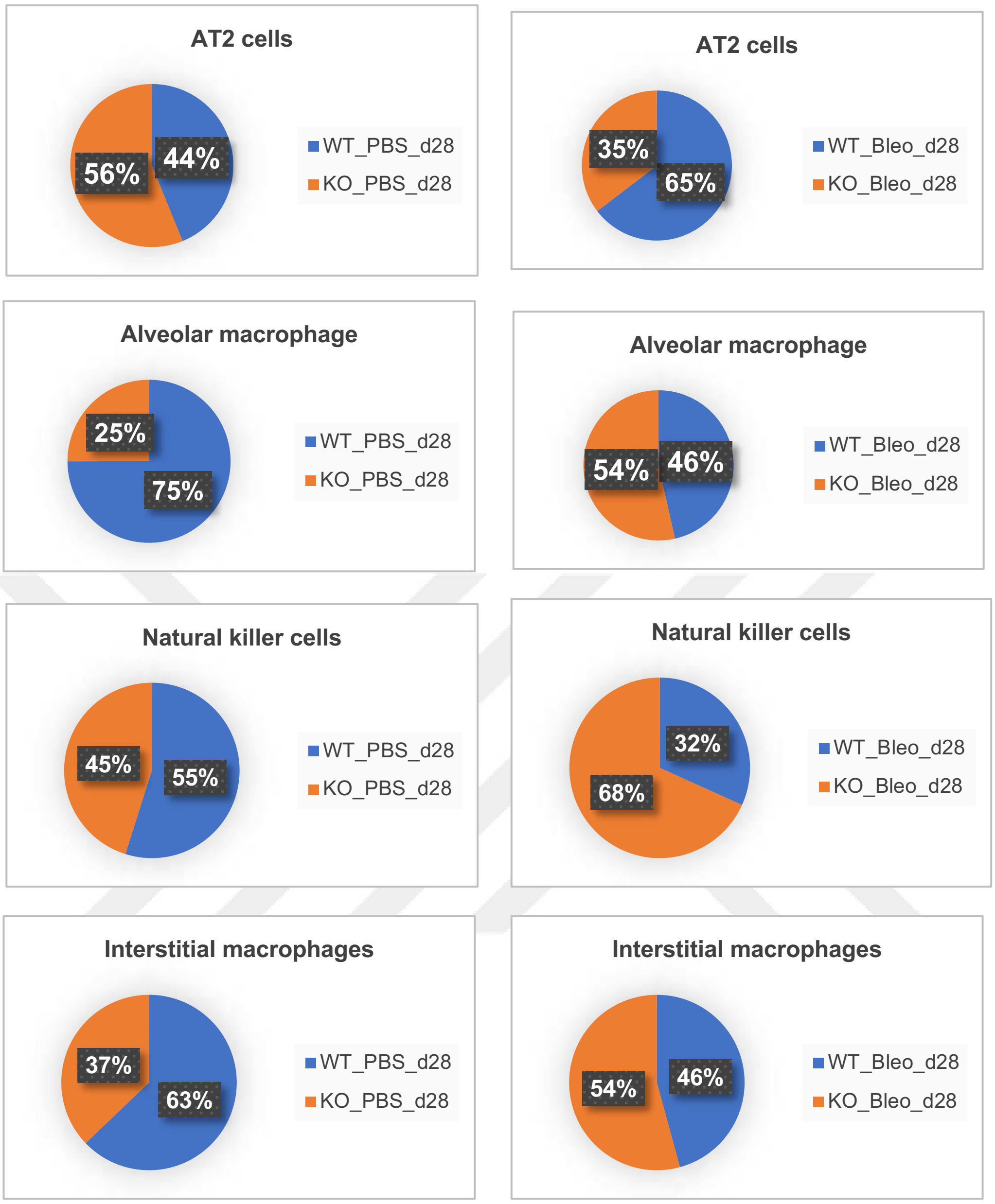

Figure 14. Pie charts show relative frequency comparisons. Depletion of Emilin2 causes increase in number of inflammatory cells and a decrease in AT2 cells in the late stage of bleomycin induced fibrosis model.

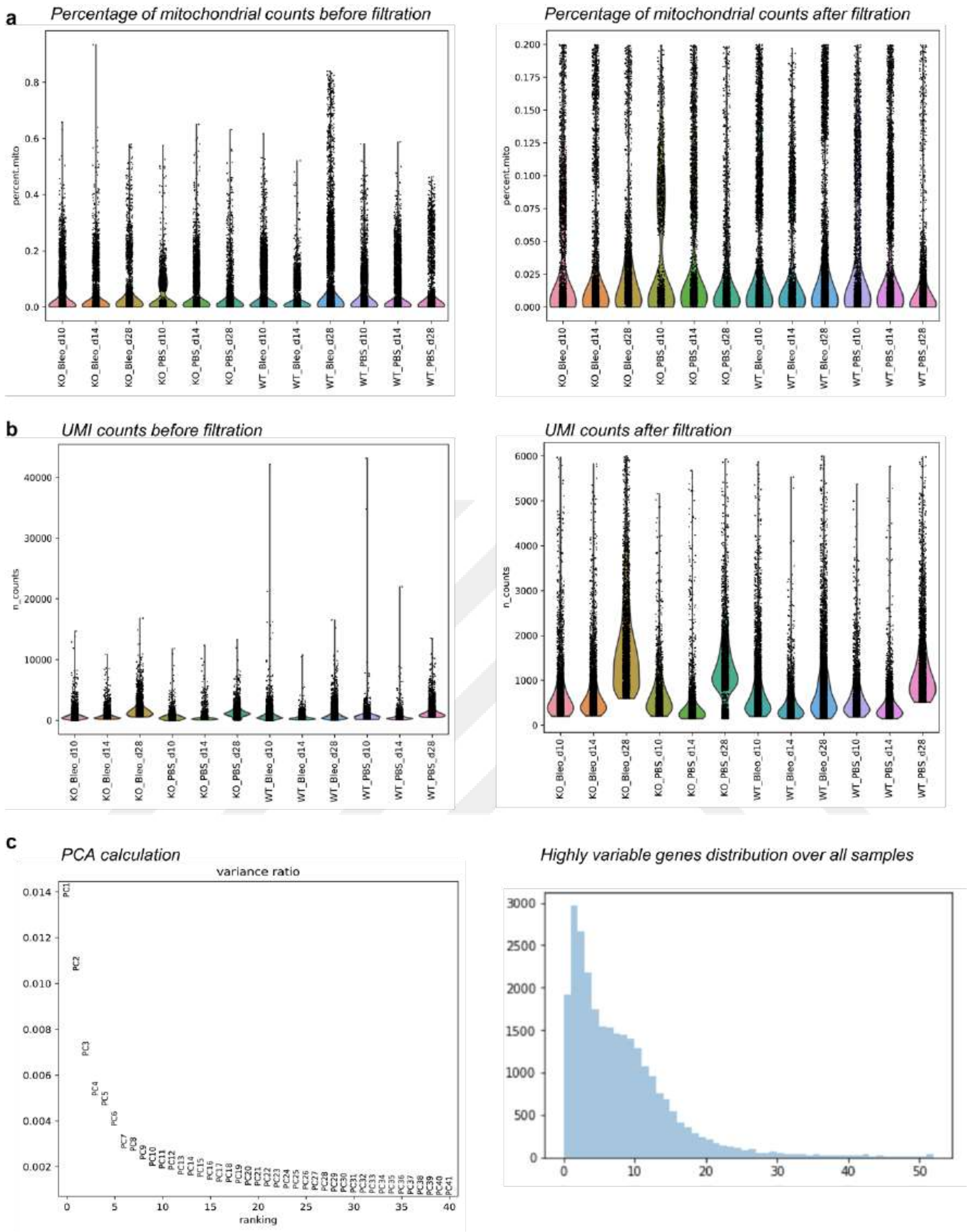

Figure 15. Pre-processing of scRNA-seq data. Violin plots in (a) shows thresholding for percentage of mitochondrial counts to exclude non-informative reads. Violin plots in (b) shows thresholding for UMI counts to exclude doublets from the dataset. (c) Principal components to reduce dimensionality. (d) Feature selection for highly variable genes.

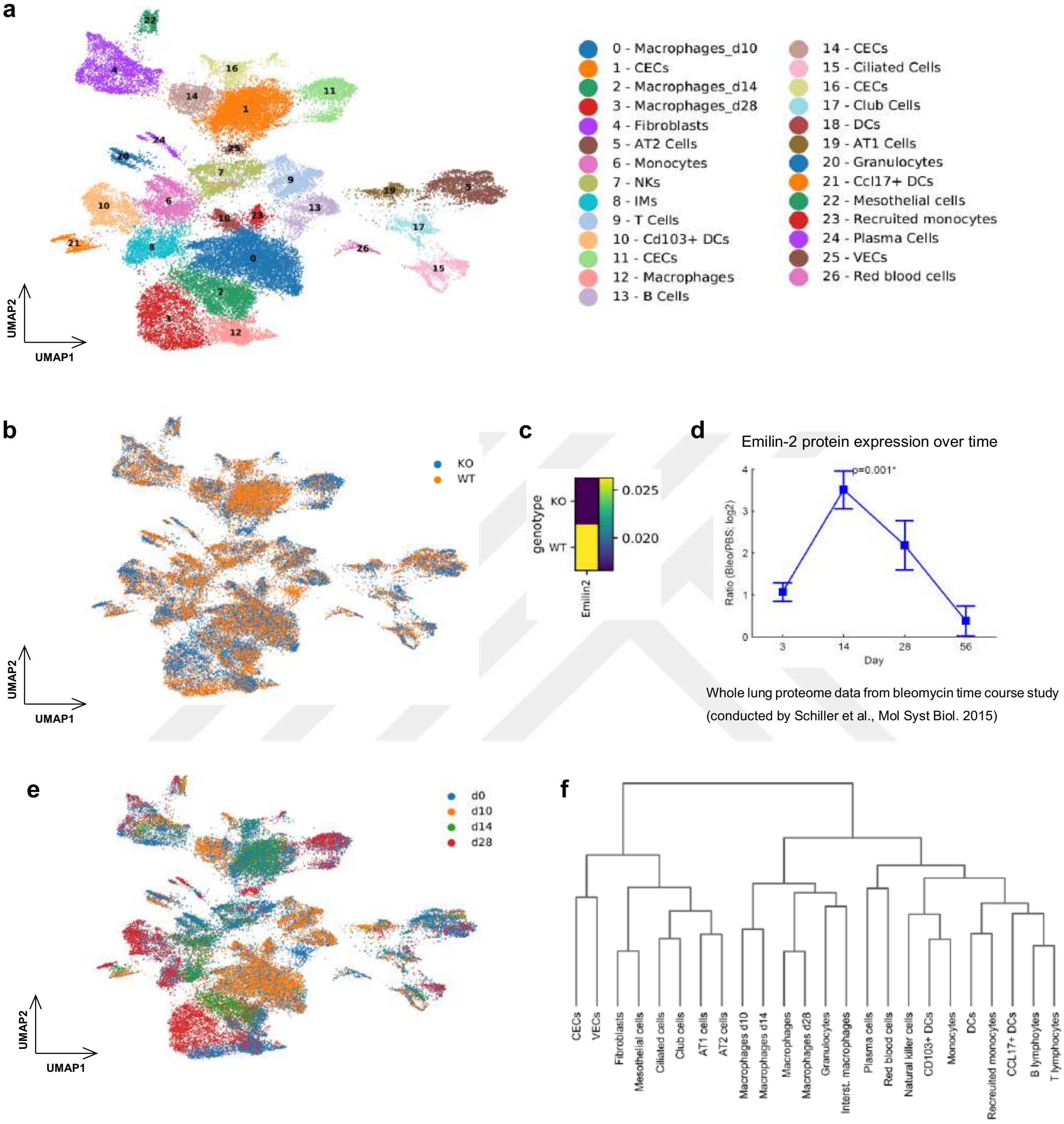

Figure 16. Single-cell RNA sequencing of 43,373 lung cells from bleomycin-induced lung tissue of *Emilin2* knockout and wild-type mice at three time points of day10, day14 and day28 after single dose transoral treatment of bleomycin. **(a)** UMAP representation of clustered cells annotated with cell state. **(b)** The indicated color code in UMAP representation shows the genotype difference. **(c)** The matrixplot shows *Emilin2* expression in knockout and wild-type mice **(d)** Emilin2 abundance changes over time relative to healthy control mice. Illustration is adapted from Schiller et al., 2015 **(e)** UMAP representation of same clusters (as in a and b) with a color code of time points. **(f)** Dendrogram tree of the louvain clusters shows clear separation between epithelial cells and immune derived cells.

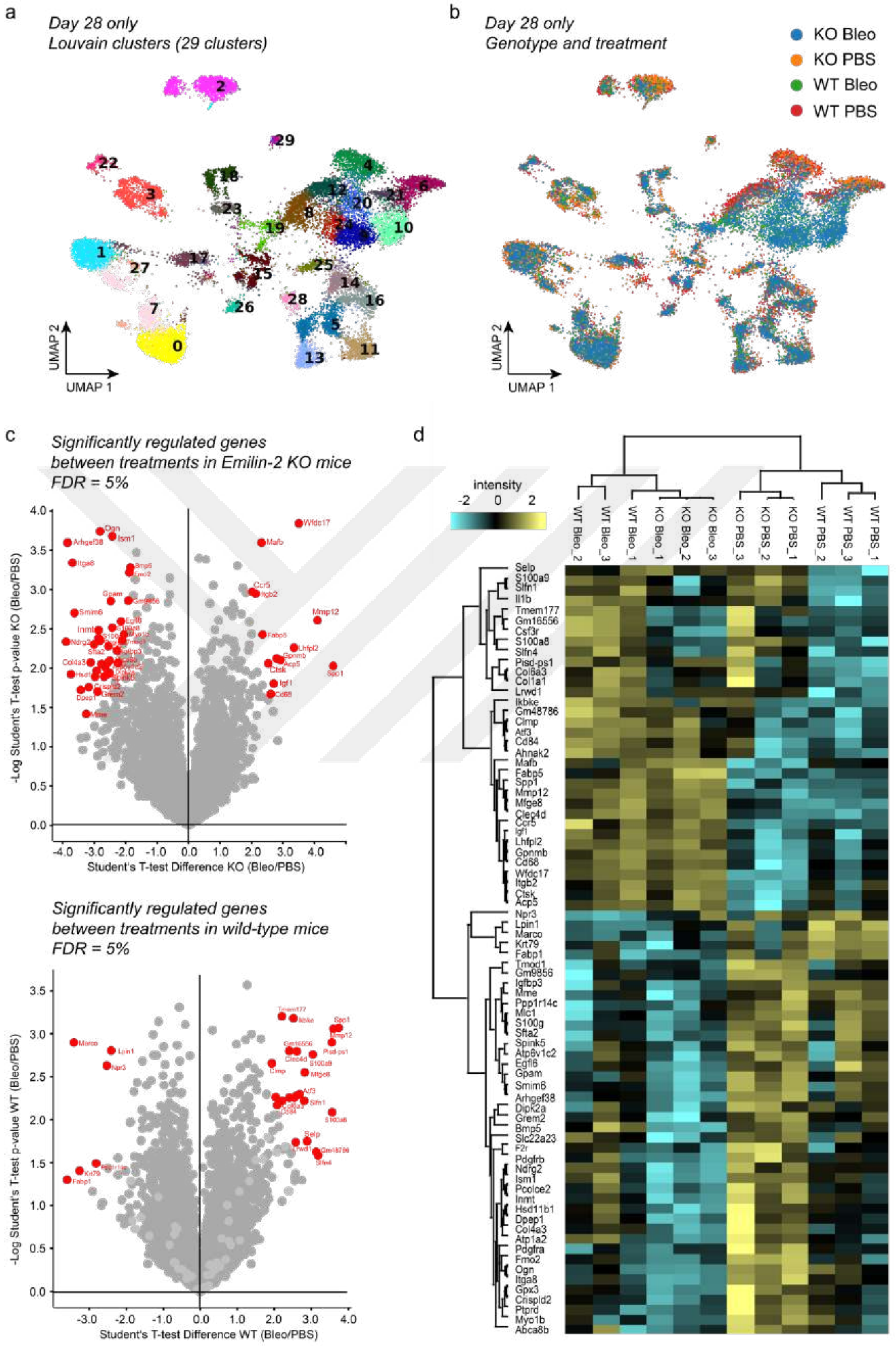

Figure 17. *in silico bulk* (scRNA-seq) data from day 28 after bleomycin-induced injury and controls. Illustrations at (c) and (d) were performed by Perseus software package. **(a)** Louvain clusters in UMAP. **(b)** The indicated color code shows the distribution of treatment conditions (Bleo; bleomycin and PBS; control)

and genotype conditions (KO; knockout and WT; wild-type) in UMAP representation. **(c)** Volcano plot shows differential expression results of the *in silico* bulk RNA-seq data from bleomycin treated (right side) and control mice (left side). Significantly regulated genes with a false discovery rate (FDR) <5% are highlighted in red in the volcano plot showing the indicated fold changes and *p* values derived from two paired *t*-test statistic. Volcano plots were created for two experimental conditions: bleomycin treatment and PBS (control). **(d)** Unsupervised hierarchical clustering of 76 significantly regulated genes between treatment conditions of bleomycin and PBS (control).

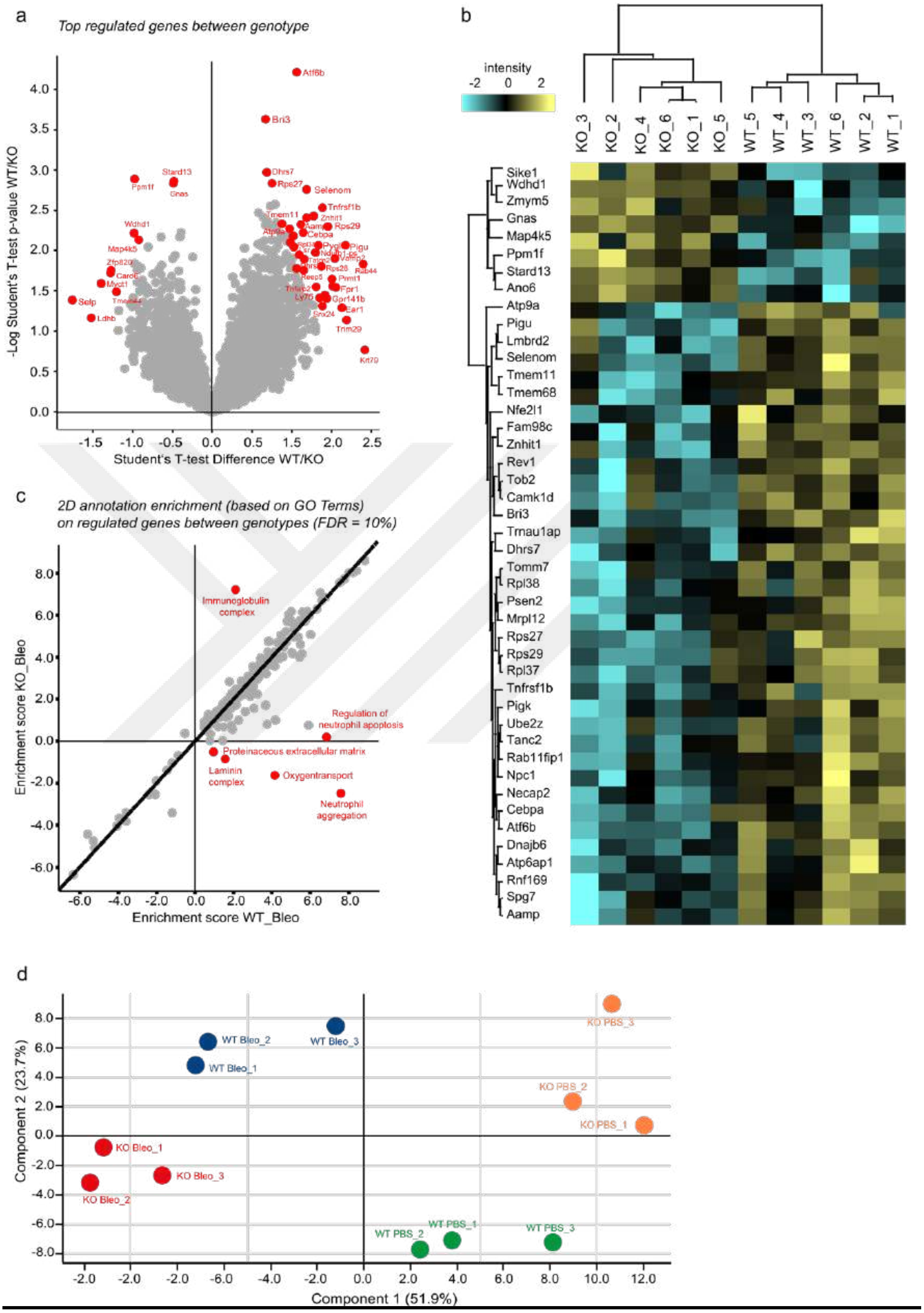

Figure 18. *in silico bulk* (scRNA-seq) data from day 28 after bleomycin-induced injury and controls. All illustrations were performed with the Perseus software package. **(a)** Volcano plot visualize the differential gene expression results of the *in silico* bulk RNA-seq data from *Emilin2* knockout (left side) and wild-type

mice (right side). Top regulated genes with lowest *p* values were highlighted in red in the volcano plot showing the indicated fold changes and *p* values derived from two paired *t*-test statistic (false discovery rate (FDR) <5%). **(b)** Heatmap displays zscores for 44 top regulated genes (rows) with lowest *p* values which were grouped by unsupervised hierarchical clustering. Columns correspond to genotype difference between Emilin2 knockout (KO) and wild-type mice (WT) including both bleomycin treated and control samples. The *p* values are derived from two paired t-test statistic. **(c)** The scatter plot shows the results of a two-dimensional annotation enrichment analysis based on fold changes in the transcriptome of the two genotype conditions in bleomycin-treated mice: wild-type on the x-axis and *Emilin2* knockout on the y axis. Gene Ontology (GO) terms was used for gene annotation and the Benjamini-Hochberg test was applied with a false discovery rate (FDR) < 10%. **(d)** The scatter plot illustrates principal component analysis (PCA) of the *in silico* bulk data (WT Bleo; Bleomycin treated wild-type mice, KO Bleo; Bleomycin treated Emilin2 knockout mice, and PBS is for the control samples). Input matrix for the PCA includes 76 significantly regulated genes (two paired t-test statistic, false discovery rate (FDR) <5%). First principal component separates the bleomycin treated samples from the controls.

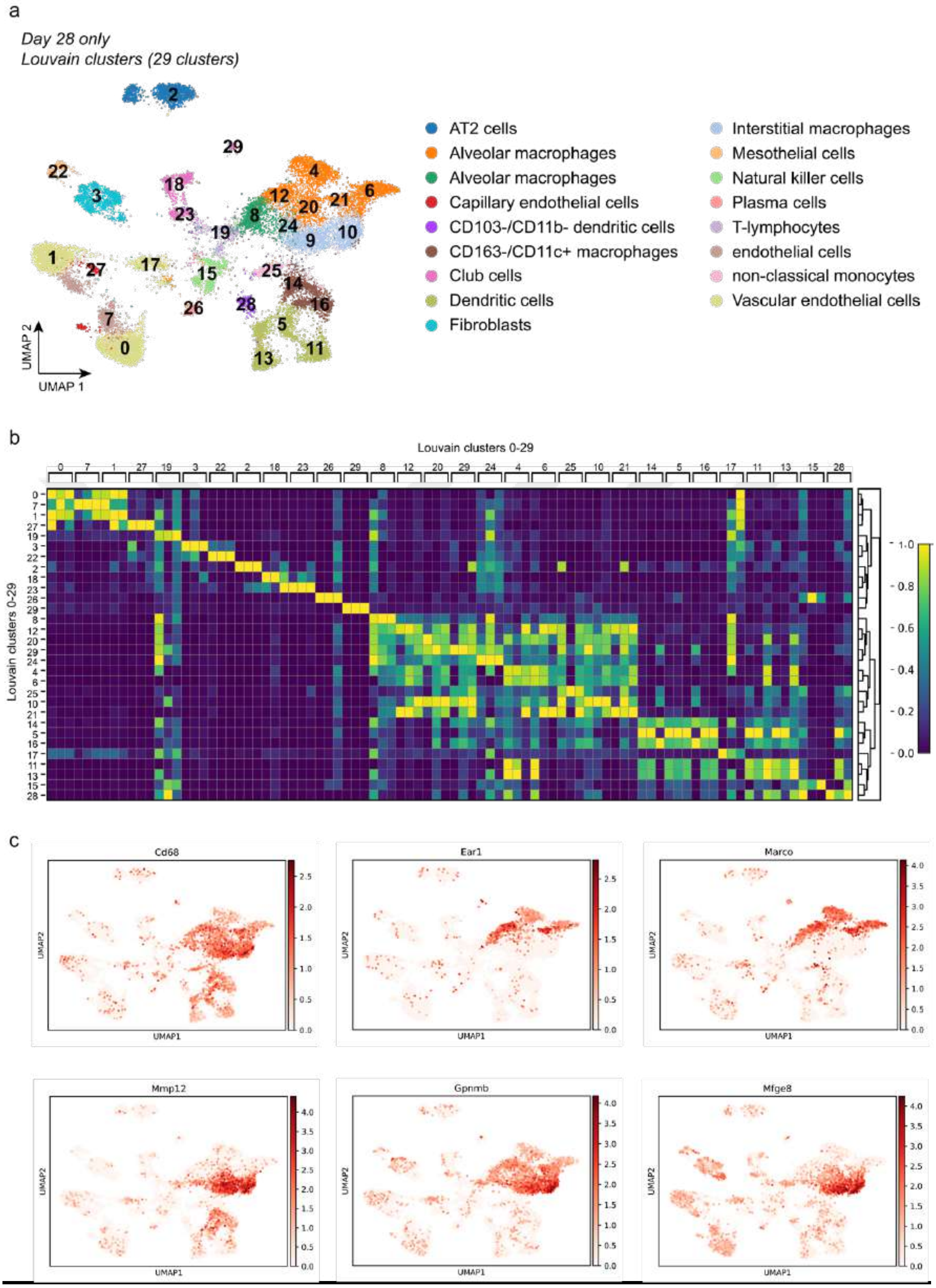

Figure 19. **(a)** 22,102 cells from day 28 were clustered in UMAP representation colored by cluster identity. Clusters were annotated based on marker genes. **(b)** The matrixplot shows the average expression levels for the top 5 marker genes for each Louvain cluster. Rows represent hierarchically clustered Louvain clusters. **(c)** Feature plots of selected marker genes for macrophage subtypes. Red colors indicate higher expression levels.

4 Discussion

Bleomycin-induced lungs have diverse changes in regenerative and repair pathways upon injury and also the changes cellular states that can be captured with sufficient number of cells, whereas, this dataset had relatively small number of sequenced single cells, especially the first two time points of day 10 and day 14. In addition, it was observed by absolute cell counts that some immune subtypes could not be captured during single-cell collection at day 28 (data not shown). Specific cell types and their response to injury can be better profiled by creation of both spatial and temporal map of gene expression. To determine gene expression localizations, single-cell RNA sequencing can be coupled with novel in situ RNA hybridization techniques. In the UMAP representation of clusters, some cells were observed to be interspersed between clusters creating a background noise. In order to get rid of such low-quality cells, quality control (QC) covariates should be considered again since it can lead to misinterpretation. For instance, high fraction of mitochondrial counts may be involved in respiratory process, and cells with low counts and/or genes may correspond to quiescent cell populations, and cells with high counts may be larger in size. Thresholds should be set as permissive as possible to avoid filtering out viable cell populations unintentionally (Luecken and Theis, 2019). Ambient counts can distort the downstream analysis at the cell- and gene-level. Also, the quality control (QC) decisions can be revisited again to judge better analysis performance during cluster annotation at day 28 analysis in particular. QC thresholds were left relatively permissive; however, stringent values might be necessary to get better quality in the downstream analysis as the larger variation of count depths per cell can mask the heterogeneity between cells. In the UMAP representation of all time points (Figure 14), clear separation in endothelial and macrophage clusters were observed based on time points. This can be indicative of biological effect or this can also be batch effect between time points. Batch effect can occur between different time points and/or between each experiment although it is performed in the same laboratory since cells were sequenced in different sequencing lanes and harvested at different time points. Differences can have an effect on the transcriptome and subsequent downstream analyses. For the differential gene expression testing between treatment and genotype conditions, *in silico* bulk analysis was utilized. Although bulk analysis produces robust results, differential expression testing in the single-cell setting holds an advantage over the bulk differential testing. Weight loss curve suggests fibrotic phenotype which persists in the late stage of the bleomycin model. These findings are in line with the observation that augmented or prolonged fibrosis and expression of certain resolution macrophage markers in bleomycin treated knockout cells (Figure 19c). Some gene signatures including *S100a9, Il1b, Csf3r* were indicated as differentially regulated expression patterns between *Emilin2* knockout and wild-type bleomycin treated mice (Figure 17d). In Figure 14, *Emilin2* lungs showed mildly increased lung inflammatory cells compared to wild-type lungs, whereas the highly ranked pathways showed an upregulation of immunoglobulin complex in the knockout lung and neutrophil aggregation in the wild-type lung which suggests an increase in the number of granulocytes (Figure 18c). To validate contribution of immune cells to fibrotic phenotype, further analysis is required to assess fibrosis levels in both wild-type and knockout lungs. *Emilin2* expression levels (data not shown) over time points was not consistent with the protein abundance levels depicted on Figure 16d, which can be resulted from biological differences in mRNA and protein levels or a misleading data due to bad quality settings during pre-processing steps. Data interpretation is highly challenging due to immense heterogeneity in the cellular makeup of the lung. In order gather more knowledge regarding of Emilin2 function in the progress of IPF phenotype, and to better interpret data, more experimental studies are needed. Best source of information for the human disease is the primary tissue itself. Thereby, more data can be collected by utilizing novel methods of spatial transcriptomics, organoids, precision cut lung slices (PCLS), and *ex vivo* lung systems (Basil and Morrisey, 2019).

References

Aung, H.H., Sivakumar, A., Gholami, S.K., Venkateswaran, S.P., Gorain, B., and Shadab (2019). An Overview of the Anatomy and Physiology of the Lung. Nanotechnology-Based Targeted Drug Delivery Systems for Lung Cancer 1–20.

Bageritz, J., and Raddi, G. (2019). Single-Cell RNA Sequencing with Drop-Seq. Methods in Molecular Biology 73–85.

Balestrini, J.L., and Niklason, L.E. (2015). Extracellular matrix as a driver for lung regeneration. Ann. Biomed. Eng. *43*, 568–576.

Barratt, S.L., Creamer, A., Hayton, C., and Chaudhuri, N. (2018). Idiopathic Pulmonary Fibrosis (IPF): An Overview. J. Clin. Med. Res. *7*.

Basil, M.C., and Morrisey, E.E. (2019). Lung regeneration: a tale of mice and men. Semin. Cell Dev. Biol.

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 Biol. *22*, 549–556.

Burgstaller, G., Oehrle, B., Gerckens, M., White, E.S., Schiller, H.B., and Eickelberg, O. (2017). The instructive extracellular matrix of the lung: basic composition and alterations in chronic lung disease. Eur. Respir. J. *50*.

Calle, E.A., Ghaedi, M., Sundaram, S., Sivarapatna, A., Tseng, M.K., and Niklason, L.E. (2014). Strategies for whole lung tissue engineering. IEEE Trans. Biomed. Eng. *61*, 1482–1496.

Caminati, A., Madotto, F., Cesana, G., Conti, S., and Harari, S. (2015). Epidemiological studies in idiopathic pulmonary fibrosis: pitfalls in methodologies and data interpretation. Eur. Respir. Rev. *24*, 436–444.

Carrington, R., Jordan, S., Pitchford, S.C., and Page, C.P. (2018). Use of animal models in IPF research. Pulmonary Pharmacology & Therapeutics *51*, 73–78.

Carvalho, O. (2011). Comparative Physiology of the Respiratory System in the Animal Kingdom. The Open Biology Journal *4*, 35–46.

Colombatti, A., Spessotto, P., Doliana, R., Mongiat, M., Bressan, G.M., and Esposito, G. (2011). The EMILIN/Multimerin family. Front. Immunol. *2*, 93.

Desai, T.J., Brownfield, D.G., and Krasnow, M.A. (2014). Alveolar progenitor and stem cells in lung development, renewal and cancer. Nature *507*, 190–194.

Ding, B.-S., Nolan, D.J., Guo, P., Babazadeh, A.O., Cao, Z., Rosenwaks, Z., Crystal, R.G., Simons, M., Sato, T.N., Worgall, S., et al. (2011). Endothelial-derived angiocrine signals induce and sustain regenerative lung alveolarization. Cell *147*, 539–553.

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*, 164–168.

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*, 12003–12011.

Eisenberg, J.L., Safi, A., Wei, X., Espinosa, H.D., Budinger, G.S., Takawira, D., Hopkinson, S.B., and Jones, J.C. (2011). Substrate stiffness regulates extracellular matrix deposition by alveolar epithelial cells. Res. Rep. Biol. *2011*, 1–12.

Gibbons, M.A., MacKinnon, A.C., Ramachandran, P., Dhaliwal, K., Duffin, R., Phythian-Adams, A.T., van Rooijen, N., Haslett, C., Howie, S.E., Simpson, A.J., et al. (2011). Ly6Chi monocytes direct alternatively activated profibrotic macrophage regulation of lung fibrosis. Am. J. Respir. Crit. Care Med. *184*, 569–581.

Glassberg, M.K. (2019). Overview of idiopathic pulmonary fibrosis, evidence-based guidelines, and recent developments in the treatment landscape. Am. J. Manag. Care *25*, S195–S203.

Guha, A., Deshpande, A., Jain, A., Sebastiani, P., and Cardoso, W.V. (2017). Uroplakin 3a Cells Are a Distinctive Population of Epithelial Progenitors that Contribute to Airway Maintenance and Post-injury Repair. Cell Rep. *19*, 246–254.

Herazo-Maya, J.D., Noth, I., Duncan, S.R., Kim, S., Ma, S.-F., Tseng, G.C., Feingold, E., Juan-Guardela, B.M., Richards, T.J., Lussier, Y., et al. (2013). Peripheral blood mononuclear cell gene expression profiles predict poor outcome in idiopathic pulmonary fibrosis. Sci. Transl. Med. *5*, 205ra136.

Huang, S.K., Scruggs, A.M., Donaghy, J., Horowitz, J.C., Zaslona, Z., Przybranowski, S., White, E.S., and Peters-Golden, M. (2013). Histone modifications are responsible for decreased Fas expression and apoptosis resistance in fibrotic lung fibroblasts. Cell Death Dis. *4*, e621.

Huang, Y., Ma, S.-F., Espindola, M.S., Vij, R., Oldham, J.M., Huffnagle, G.B., Erb-Downward, J.R., Flaherty, K.R., Moore, B.B., White, E.S., et al. (2017). Microbes Are Associated with Host Innate Immune Response in Idiopathic Pulmonary Fibrosis. Am. J. Respir. Crit. Care Med. *196*, 208–219.

Izbicki, G., Segel, M.J., Christensen, T.G., Conner, M.W., and Breuer, R. (2002). Time course of bleomycin-induced lung fibrosis. Int. J. Exp. Pathol. *83*, 111–119.

Koehler, U. (2019). Ein Pionier der Anatomie – Der Marburger Arzt und Ordinarius Dryander (1500 –1560) und die anatomische Darstellung der Lunge. Pneumologie *73*, 374–376.

Koval, M., and Sidhaye, V.K. (2017). Introduction: The Lung Epithelium. Lung Epithelial Biology in the Pathogenesis of Pulmonary Disease xiii – xviii.

Lawson, W.E., Crossno, P.F., Polosukhin, V.V., Roldan, J., Cheng, D.-S., Lane, K.B., Blackwell, T.R., Xu, C., Markin, C., Ware, L.B., et al. (2008). Endoplasmic reticulum stress in alveolar epithelial cells is prominent in IPF: association with altered surfactant protein processing and herpesvirus infection. Am. J. Physiol. Lung Cell. Mol. Physiol. *294*, L1119–L1126.

Lechner, A.J., Driver, I.H., Lee, J., Conroy, C.M., Nagle, A., Locksley, R.M., and Rock, J.R. (2017). Recruited Monocytes and Type 2 Immunity Promote Lung Regeneration following Pneumonectomy. Cell Stem Cell *21*, 120–134.e7.

Ley, B., and Collard, H.R. (2013). Epidemiology of idiopathic pulmonary fibrosis. Clin. Epidemiol. *5*, 483–492.

Liu, G., Friggeri, A., Yang, Y., Milosevic, J., Ding, Q., Thannickal, V.J., Kaminski, N., and Abraham, E. (2010). miR-21 mediates fibrogenic activation of pulmonary fibroblasts and lung fibrosis. J. Exp. Med. *207*, 1589–1597.

Liu, T., De Los Santos, F.G., and Phan, S.H. (2017). The Bleomycin Model of Pulmonary Fibrosis. Fibrosis 27–42.

Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Mol. Syst. Biol. *15*, e8746.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Martinez, F.J., Collard, H.R., Pardo, A., Raghu, G., Richeldi, L., Selman, M., Swigris, J.J., Taniguchi, H., and Wells, A.U. (2017). Idiopathic pulmonary fibrosis. Nat Rev Dis Primers *3*, 17074.

McCulley, D., Wienhold, M., and Sun, X. (2015). The pulmonary mesenchyme directs lung development. Curr. Opin. Genet. Dev. *32*, 98–105.

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software *3*, 861.

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.N., et al. (2017). Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. J. Exp. Med. *214*, 2387–2404.

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. Mol. Cell. Biol. *27*, 7176–7187.

Moore, B.B., and Hogaboam, C.M. (2008). Murine models of pulmonary fibrosis. Am. J. Physiol. Lung Cell. Mol. Physiol. *294*, L152–L160.

Mulugeta, S., Nureki, S.-I., and Beers, M.F. (2015). Lost after translation: insights from pulmonary surfactant for understanding the role of alveolar epithelial dysfunction and cellular quality control in fibrotic lung disease. American Journal of Physiology-Lung Cellular and Molecular Physiology *309*, L507–L525.

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*, 1118–1123.

Noth, I., Zhang, Y., Ma, S.-F., Flores, C., Barber, M., Huang, Y., Broderick, S.M., Wade, M.S., Hysi, P., Scuirba, J., et al. (2013). Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. Lancet Respir Med *1*, 309–317.

Paulitti, A., Andreuzzi, E., Bizzotto, D., Pellicani, R., Tarticchio, G., Marastoni, S., Pastrello, C., Jurisica, I., Ligresti, G., Bucciotti, F., et al. (2018). The ablation of the matricellular protein

EMILIN2 causes defective vascularization due to impaired EGFR-dependent IL-8 production affecting tumor growth. Oncogene *37*, 3399–3414.

Plasschaert, L.W., Žilionis, 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*, 377–381.

Rane, C.K., Jackson, S.R., Pastore, C.F., Zhao, G., Weiner, A.I., Patel, N.N., Herbert, D.R., Cohen, N.A., and Vaughan, A.E. (2019). Development of solitary chemosensory cells in the distal lung after severe influenza injury. Am. J. Physiol. Lung Cell. Mol. Physiol. *316*, L1141–L1149.

Sauleda, J., Núñez, B., Sala, E., and Soriano, J.B. (2018). Idiopathic Pulmonary Fibrosis: Epidemiology, Natural History, Phenotypes. Med Sci (Basel) *6*.

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. J. Invest. Dermatol. *136*, 1150–1160.

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. Mol. Syst. Biol. *11*, 819.

Siebert, S., Farrell, J.A., Cazet, J.F., Abeykoon, Y., Primack, A.S., Schnitzler, C.E., and Juliano, C.E. (2019). Stem cell differentiation trajectories in resolved at single-cell resolution. Science *365*.

Song, H., Yao, E., Lin, C., Gacayan, R., Chen, M.-H., and Chuang, P.-T. (2012). Functional characterization of pulmonary neuroendocrine cells in lung development, injury, and tumorigenesis. Proc. Natl. Acad. Sci. U. S. A. *109*, 17531–17536.

Strunz, M. (2019). Dissecting lung fibrogenesis and regeneration through the lens of single cell transcriptomics (Unpublished doctoral dissertation). Ludwig-Maximilians-Universität München, Germany.

Tarran, R., Button, B., and Boucher, R.C. (2006). Regulation of normal and cystic fibrosis airway surface liquid volume by phasic shear stress. Annu. Rev. Physiol. *68*, 543–561.

Tata, P.R., Mou, H., Pardo-Saganta, A., Zhao, R., Prabhu, M., Law, B.M., Vinarsky, V., Cho, J.L., Breton, S., Sahay, A., et al. (2013). Dedifferentiation of committed epithelial cells into stem cells in vivo. Nature *503*, 218–223.

Tsujino, K., Takeda, Y., Arai, T., Shintani, Y., Inagaki, R., Saiga, H., Iwasaki, T., Tetsumoto, S., Jin, Y., Ihara, S., et al. (2012). Tetraspanin CD151 protects against pulmonary fibrosis by maintaining epithelial integrity. Am. J. Respir. Crit. Care Med. *186*, 170–180.

Uhal, B.D., and Nguyen, H. (2013). The Witschi Hypothesis revisited after 35 years: genetic proof from SP-C BRICHOS domain mutations. Am. J. Physiol. Lung Cell. Mol. Physiol. *305*, L906–L911.

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

Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst *8*, 281–291.e9.

Wright, J.R. (2005). Immunoregulatory functions of surfactant proteins. Nat. Rev. Immunol. *5*, 58–68.

Yang, I.V., Pedersen, B.S., Rabinovich, E., Hennessy, C.E., Davidson, E.J., Murphy, E., Guardela, B.J., Tedrow, J.R., Zhang, Y., Singh, M.K., et al. (2014). Relationship of DNA methylation and gene expression in idiopathic pulmonary fibrosis. Am. J. Respir. Crit. Care Med. *190*, 1263–1272.

Yao, E., Lin, C., Wu, Q., Zhang, K., Song, H., and Chuang, P.-T. (2018). Notch Signaling Controls Transdifferentiation of Pulmonary Neuroendocrine Cells in Response to Lung Injury. Stem Cells *36*, 377–391.

Young, M.D., and Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet based single-cell RNA sequencing data.

Zepp, J.A., and Morrisey, E.E. (2019). Cellular crosstalk in the development and regeneration of the respiratory system. Nat. Rev. Mol. Cell Biol. *20*, 551–566.