ANNUAL REVIEWS

Annual Review of Physiology

Lung Cell Atlases in Health and Disease

Taylor S. Adams,[1] Arnaud Marlier,[2] and Naftali Kaminski[1]

[1]Section of Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, Yale University, New Haven, Connecticut, USA; email: Naftali.kaminski@yale.edu

[2]Department of Neurosurgery, Yale School of Medicine, Yale University, New Haven, Connecticut, USA

Annu. Rev. Physiol. 2023. 85:47–69

First published as a Review in Advance on November 9, 2022

The *Annual Review of Physiology* is online at physiol.annualreviews.org

https://doi.org/10.1146/annurev-physiol-032922-082826

Keywords

single-cell RNA sequencing, pulmonary fibrosis, emphysema, asthma, acute respiratory distress syndrome

Abstract

The human lung cellular portfolio, traditionally characterized by cellular morphology and individual markers, is highly diverse, with over 40 cell types and a complex branching structure highly adapted for agile airflow and gas exchange. While constant during adulthood, lung cellular content changes in response to exposure, injury, and infection. Some changes are temporary, but others are persistent, leading to structural changes and progressive lung disease. The recent advance of single-cell profiling technologies allows an unprecedented level of detail and scale to cellular measurements, leading to the rise of comprehensive cell atlas styles of reporting. In this review, we chronical the rise of cell atlases and explore their contributions to human lung biology in health and disease.

1. INTRODUCTION

The primary function of the lung is the exchange of atmospheric oxygen with the carbon dioxide generated throughout the body's tissues. Air is delivered through the nasal passages, into the airways, in multiple generations of bifurcations of the bronchial tree and then into respiratory units composed of the respiratory bronchioles, the alveolar duct, and the alveolar sac, where gas exchange occurs (1). The airways, which are multilayered tubes, have evolved to regulate air flow, temperature, and humidity and to provide protection from particles and infectious agents, whereas the alveolar unit evolved to allow for unimpeded gas exchange, with single layers of cells and thin interstitium. The normal structure of the lung is achieved through highly distinct and coordinated stages during embryonic and early neonatal life and remains relatively stable through most of adult life. Morphological studies have identified cell populations characteristic of each anatomical region in the lung, and the number of discrete cell population types in the lung have been estimated at >40 (2). To maintain air flow and gas exchange the lungs have the capacity to rapidly shift their cellular repertoire in response to injury. Most acute and chronic lung diseases are characterized by extensive changes in the lung cellular repertoire, with the difference being that some conditions resolve spontaneously or after treatment without substantial sequelae, whereas others lead to permanent and frequently progressive changes in this repertoire (3).

Historically, the characterization of the cellular repertoire of the lung evolved with the availability of novel technologies, such as electron microscopy, immunofluorescence, lineage tracing, and monoclonal antibodies. Cells were defined based on their morphology, location, and the expression of a few molecular markers, most derived from careful lineage studies in mice. The recent emergence of high-throughput single-cell profiling technologies such as single-cell RNA sequencing (scRNA-seq) created a unique opportunity to improve the resolution of previous characterizations of discrete cell populations into a high-throughput signature, function, and systems-based cellular state and phenotype characterization to create comprehensive catalogs of all the cells in an organ system and, eventually, the whole organism (4). The opportunities were met with enthusiasm in nearly every organ system. Publications describing cell atlases have proliferated (**Figure 1*a*,*b***), providing an unprecedented view of the complexity of cellular repertoires of distinct organs. Cell atlases have revealed unexpected outcomes, including new cell types and undiscovered cell states that challenge prior assumptions about cell types, and have provided an outlet for biologists to computationally compare cell behavioral patterns across tissues and experiments in ways previously impossible. Ambitious consortium-lead projects like The Human Cell Atlas or Tabula Muris are seeking to succeed the Human Genome Project by describing all cells found in the healthy human or mouse, respectively (4, 5).

In respiratory research, cell atlas reports have transformed our perception of the lung cellular portfolio and identified previously unrecognized cellular populations in health and disease. The possibility of creating a collection of catalogs of all cells in the lung, as well as their phenotypes, markers, interactions, and specific locations, through different stages of life and in disease, is now more feasible than ever. Such a collection—comprehensive and exhaustive but approachable and accessible—would be the lung cell atlas, a highly useful blueprint for lung health and disease. In this review, we reflect on the technological advances behind the advent of cell atlas reporting and highlight examples where discoveries from cell atlases have disrupted our understanding of lung cellular biology and pathophysiology.

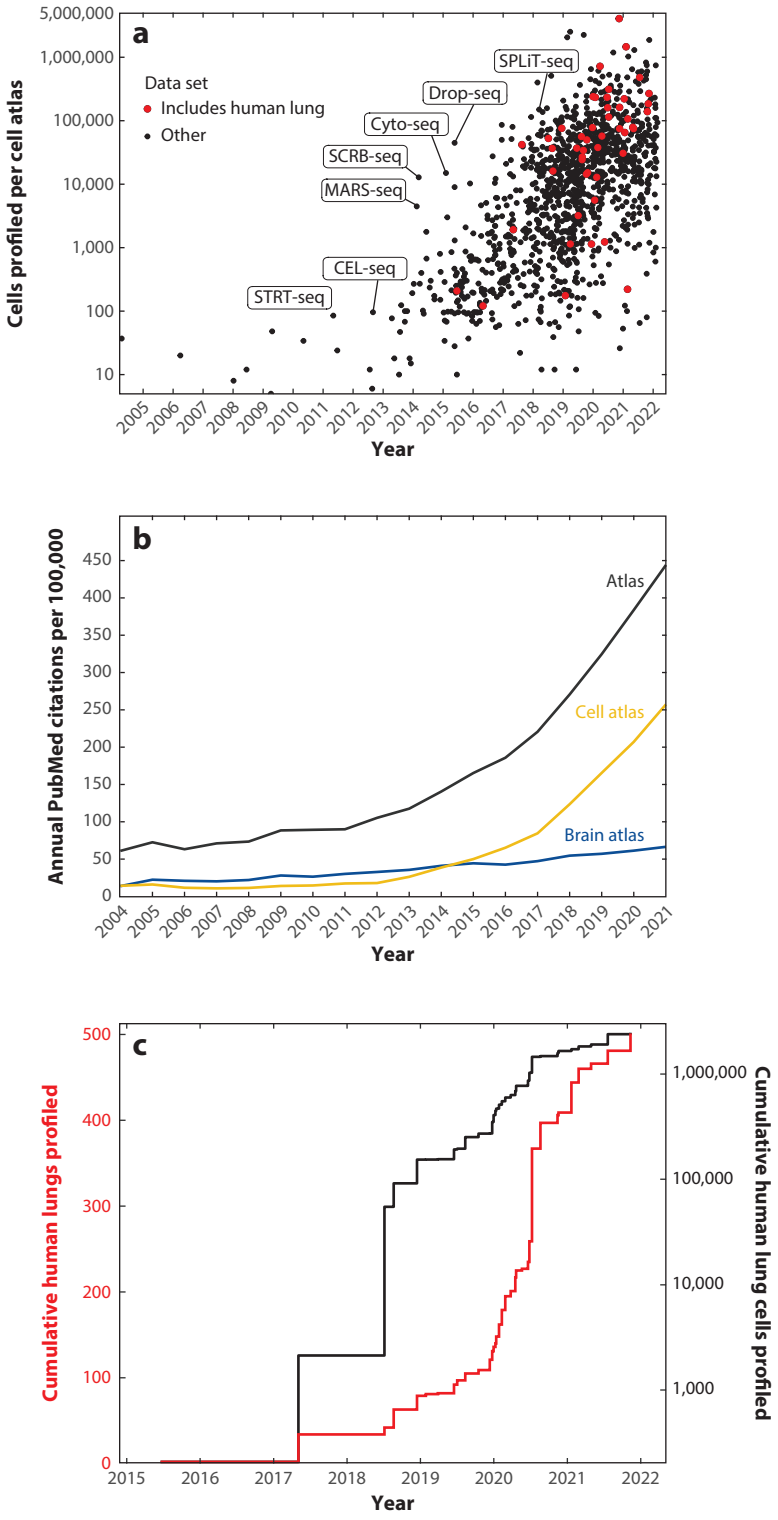

Figure 1

Trends in cell atlas reporting. (*a*) Scatterplot representing the number of cells profiled per cell atlas experiment over time (51). Cell atlas reports that introduced key influential single-cell RNA sequencing methods are labeled. (*b*) Line chart showing the annual proportion of PubMed citations for the terms "atlas," "cell atlas," and "brain atlas" (52). (*c*) Step plot visualizing the cumulative number of human lungs and human lung cells that have been profiled across cell atlases (51).

2. A HISTORICAL TIME LINE OF CELL ATLASES

2.1. From World Mapping to Tissue Mapping

The term atlas is broadly defined as a bound collection of maps, tables, charts, or plates. This definition is derived from the sixteenth-century cartographer Gerardus Mercator, whose atlas compiled various maps of the world at different granularities (6). Mercator considered the harmony of geography, topology, and geology to be divine; his atlas provided an accessible visual guide of the vastness of known cosmography. During the eighteenth century, use of the term atlas would be generalized beyond charts of the physical world to include visual descriptions of more abstract statistical relationships. *The Commercial and Political Atlas* by Scottish economist William Playfair and colleagues nearly exclusively comprises multivariate time-series graphics, the lone exemption being the first documented example of a bar graph (7).

In modern bioscience, the term atlas can be used to describe any comprehensive description of a complex structure or system. Metabolic atlases model all molecular reactions in an organism (8); the Human Protein Atlas provides spatial mapping of all proteins across human tissues (9). Cell atlases provide comprehensive perspectives on how an ensemble of cells collaborates to perform advanced tasks such as the function of an organ, its process of development, or its behavior during disease. Where Mercator's atlas reconciled geographic and political boundaries to explain Earth, and Playfair's atlas reconciled time and financial data to explain the world economy, today's cell atlases reconcile the molecular and spatial properties of cells to provide a global understanding of tissue.

2.2. Brain Atlases: Pioneers of Tissue Atlases

The philosophy behind cell atlases can largely be linked to brain atlas projects spearheaded by neurologists in the late 1990s and early 2000s. Utilizing magnetic resonance imaging (MRI) or functional MRI two-dimensional (2D) scans of serial sections, brain atlases offered three-dimensional (3D) reconstructions of the brain for applications in education and neurosurgery and for assessing functional relationships between neurons (10). Instrumental to the success of early brain atlas projects was the ability of others to interactively explore them. Programs for viewing brain atlases were made available through CD-ROM (11) and later through web-based tools (12), impressive feats given the scale of data and limitations in computer hardware at the time.

The first transcriptome-wide brain atlas was generated by researchers at the Allen Institute using an automated, industrialized process of in situ hybridization for more than 20,000 genes (13). Despite the value and excitement surrounding such projects, several limitations prohibited widespread adoption of the technique. First, the costs remain prohibitively expensive. Second, there are persistent limitations in the reliability of cell segmentation algorithms used to estimate boundaries between cells in 3D space based on 2D images (14). Despite their high-resolution scans, the Allen Institute's 2006 brain atlas article describes their resolution as "cellular, but not single cell" (13, p. 169). As detailed and groundbreaking as the early brain atlases were, the single-cell technologies required to generate cell atlases would still be many years away.

2.3. The Transcriptomic Road Toward Cell Atlases

The ambition to capture the transcriptome of single cells largely paralleled the evolution of transcriptional profiling, moving from serial analysis of gene expression (SAGE) and microarrays to massive parallel sequencing. The development of SAGE in 1995 revolutionized whole-transcriptome analysis by allowing assessments of messenger RNA (mRNA) abundance without prior knowledge of their sequences (15). SAGE's demand for high amounts of input mRNA (16, 17) and poor scalability (18) made applications of single-cell analysis impractical. In contrast,

DNA microarrays that emerged simultaneously (19, 20) offered substantially improved scalability and sensitivity, with the limitation of only targeting known sequences (17, 18). While DNA microarrays allowed analysis of small numbers of cells identified by microscope and isolated by micromanipulation, such as Reed-Sternberg cells from patients with Hodgkin's disease (21), single-neuron microdissections from patients with schizophrenia (22) and Alzheimer's disease (23), or embryonic stem cells (24, 25) in the lung, both technologies were adapted early (26, 27) but were mostly applied to bulk analyses of tissue. The rapid development and superior performance of next-generation sequencing platforms (28) led to their implementation on single cells. The first report of scRNA-seq profiled a single mouse blastomere cell (29). Despite relying on methods optimized for single-cell microarray analysis (24), this RNA-seq experiment returned many more genes than what had previously been possible.

Though most early attention surrounding RNA-seq was focused on its unbiased assessment of transcripts, another key advantage over microarrays would soon emerge and become instrumental to all single-cell experiments: the ability to append synthetic sequences to complementary DNA (cDNA) strands to serve as a barcode of read metadata. Indeed, the earliest and most commonly used application of cDNA barcoding was sample multiplexing. By adding known barcode sequences to each end of a cDNA fragment from each sample, Bar-seq would support pooling up to 96 samples for downstream sequencing in parallel. Reads are later computationally demultiplexed by assigning them to samples based on their respective barcode identity (30). The single-cell tagged reverse transcription (STRT) protocol was the first single-cell method to implement this technique. After isolating individual cells into wells of a 96-well plate, a cell-specific barcode is added during the reverse-transcription step at the beginning of the experiment. By pooling all tagged cDNA before amplification steps, STRT reduced amplification biases in downstream reactions while supporting an unprecedented degree of sample scalability (31). The need to isolate individual cells at the start of each experiment was a limitation to the scalability of such experiments. The C1 platform from Fluidigm offered a solution to this problem by isolating individual cells from suspension into chambers using a sophisticated valve-based microfluidic system via subtle changes in pressure actuation (32). The very first lung cell atlas consisted of 198 embryonic murine lung epithelial cells profiled by using this platform (33).

Improvements in library construction methods and barcoding (34) and the invention of unique molecular identifiers (UMIs), randomly generated barcodes that are appended to transcripts before amplification, were critical to improving the fidelity of sequencing because they prevent nonlinear distortions caused by amplification and decouple measurements of transcript abundance from technical metrics like sequencing depth (35, 36). With this breakthrough, the stage was set for increasing the numbers of cells profiled. Cyto-seq (37) and Microwell-seq (38) achieved low-cost scalability by using an agarose-based microwell array and magnetic beads covered by DNA oligos featuring randomly assigned UMI barcodes. Bead-specific cell barcodes generated via split-pooled combinatorial indexing, Drop-seq (39), and inDrop (40) use microfluidics systems to rapidly encapsulate dissociated cells within nanoliter-sized droplets alongside uniquely barcoded polymer or hydrogel beads. Cell lysis and subsequent mRNA hybridization to barcodes occur within the boundaries of each droplet, after which barcoded material is pooled and collectively processed. The droplet-based methods inherit Cyto-seq's appeal to probability for matching cells with barcodes but avoid partitioning cells into wells by partitioning them into an infinitely scalable number of oil-based droplets that comprise an emulsion. Split-pool ligation-based transcriptome sequencing (SPLiT-seq) avoids the need to isolate single cells by generating combinatorial barcodes for transcripts within the membranes of cells themselves (41). This is achieved by applying a suspension of cells to a split-pool combinatorial barcoding approach similar to the one used for generating the barcoded beads used in Cyto-seq or Drop-seq (37, 39). The number of cells

that can be profiled simultaneously is only limited by the number of split-pooling rounds and the length of the barcode the user is willing to sequence.

Techniques pioneered for scRNA-seq have been adapted to assess different cell compartments or types of molecular information; single-cell multi-omics assays measure more than one. Single nuclear RNA-seq sacrifices sensitivity in exchange for less-biased representations of dissociated cells, a necessary approach for capturing information from neurons or frozen biobanked tissue, which are virtually impossible to assess via scRNA-seq. Methods like CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) (42) and REAP-seq (RNA expression and protein sequencing) (43) allow scRNA-seq assays to simultaneous measure protein levels by labeling cells with protein targeting antibodies conjugated to DNA barcodes that can be captured alongside each cell's mRNA. The assay for transposase-accessible chromatin with sequencing (ATAC-seq) measures cell epigenetics by assessing regions of chromosome accessibility and has been adopted for single-cell implementations. sci-CAR (single-cell combinational indexing) (44) and SNARE-seq (single-nucleus chromatin accessibility and mRNA expression sequencing) (45) perform ATAC-seq and RNA-seq on nuclei or cells in parallel. Single-cell copy number variant assays detect chromosomal duplication events in cancer cells. Single-cell whole genome sequencing can assess single-nucleotide variations that arise from genetic and somatic mutations across cells. Mitochondrial sequencing is used to lineage trace relationships between cells based on somatic mutations and to detect heteroplasmy events: the presence of multiple mitochondrial DNA (mtDNA) variants in the same cell. mtscATAC-seq (mitochondrial single-cell ATAC) (46) simultaneously profiles chromatin accessibility and mtDNA in the same cells by modifying a proprietary droplet-based platform. ASAP-seq (ATAC with select antigen profiling by sequencing) (47) combines mtscATAC-seq with REAP-seq, thus supporting protein, epigenetic, and mtDNA measurements for each cell. DOGMA-seq (47) combines mtscATAC-seq with scRNA-seq and CITE-seq's protein capture strategy. *Nature Methods* chose scRNA-seq as method of the year in 2013 (48), single-cell multimodal omics in 2019 (49), and spatial transcriptomics in 2020 (50). The comprehensive perspectives of cell identity and tissue composition that these technologies afford form the foundation of today's cell atlases.

2.4. Emergence of the Cell Atlas in the Twenty-First Century

Increasing access and cost scalability of single-cell assays over the past decade have triggered an exponential rise in the number of cell atlas reports as well as the typical scale of the reports themselves. In 2013, 11 scRNA-seq papers were published, with a median number of 68 cells profiled, and in 2017, 92 scRNA-seq reports profiled a median of 1,764 cells. The 168 reports in 2021 describe a median number of 27,353 cells (51) (**Figure 1*a***). Concurrent with the rise in cell atlases reporting has been the rise in cell atlas engagement across the scientific community. In 2015, the number of PubMed citations for cell atlas articles would surpass brain atlas citations for the first time (52) (**Figure 1*b***), and the relative contribution of cell atlas reports in the biomedical literature has dramatically increased every year since. Between 2015 and 2021, the proportion of annual PubMed citations corresponding to cell atlases underwent a fivefold increase, and cell atlas representation among any cited article including the word "atlas" nearly doubled from 30.3% to 57.9% (52) (**Figure 1*b***). Within a decade, cell atlases went from being a niche concept to the primary form of atlases discussed across biomedicine. Unsurprisingly, lung biology has participated in and benefited from the rising trend of cell atlases (**Figure 1*a*,*c***). Currently, more than 2.6 million cells have been profiled from over 500 human lungs (51) (**Figure 1*c***). A new preprint from the Human Cell Atlas consortium describes the integration of more than 2.2 million cells from 444 individuals into a common reference data set (53). Likewise, CellCards is a new

project from the LungMAP consortium (54) for providing a unified framework of lung cell identity (55). Such efforts are critically necessary for our ability to understand the cellular makeup of the lung.

3. CELL ATLASES OF THE HUMAN LUNG

The molecular detail and scalability of cell sampling afforded by single-cell technologies have fundamentally changed the way we can view cells. For the first time, scientists can accurately resolve the genomic identity of cells without relying on assumptions about what cells to look for or expectations of how specific cell types are supposed to be. Single-cell experiments thus have the tremendous potential to disrupt the status quo of cell-type ontology. Not only can previously unrecognized cell types be discovered, but the dogmas of known cell types can be revisited under an unbiased lens. This new perspective also elevates old problems about cell identity that were obscured by traditional binary marker-based cell identification schemes. Notably, what does it mean when markers used to classify a given cell type become expressed by another? Or more broadly, what level of phenotypic variation is sufficient to distinguish fluctuations in cell state from changes in cell type? Cellular biology's problem of defining a cell type is often compared to the taxonomic problem of formalizing the definition of a species (3). In both cases, cognitive biases toward pattern recognition and an overreliance on a limited amount of information can lead to spurious interpretations. One noteworthy zoological example is the goose barnacle myth, where it was believed that the barnacle goose (*Branta leucopsis*) emerged from goose barnacles (*Cirripedia*) (56) until ornithologists formalized the concept of bird migration. A limited understanding of goose localization patterns and the morphological similarities between goose necks and barnacle stalks was all it took for this myth to persist for 500 years. Though geese and barnacles can now easily be distinguished by differences in their DNA, cells from the same organism cannot. All cells have access to the same genetic playbook; any perception of reliability ascribed to genetic markers for classifying cell types is derived from a combination of prior experience and faith that cells will remain committed to adhering to a transcriptional homeostatic state. In the evolution of the lung cell atlases, the most notable observation was that profiling of normal (disease-free tissue) and abnormal (diseased tissue) occurred in parallel. Thus, instead of a rigid process in which first the normal cellular repertoire of the lung is defined and then a catalog of aberrations from the normal is laid, the study of normal, often as control for the abnormal, was interspersed with the abnormal. This is of course a result of practicalities; lung tissue is usually obtained through the course of diagnostic or therapeutic procedures, and research focused on disease is more readily funded. However, the wide availability of known cellular markers, and the surprising robustness of the results, allowed the creation of atlases of both healthy adults and most advanced lung diseases (**Figure 2**). As more data are added, we expect that the extent of the phenotypic diversity of healthy human lung cells and the pathological cell states that exist beyond the bounds of known healthy cell states will be better defined. In the following section, we highlight novel, unexpected, and pathological lung cell phenotypes revealed by lung cell atlases, as well as the cell-type incongruencies that arise between disease and control cell populations. Instead of proceeding by disease, we proceed by general cell families, interweaving disease data in each relevant section.

3.1. Stromal Cells

Lung tissue contains a diverse array of stromal tissues comprised of smooth muscle and pericyte and fibroblast cells; however, the actual number of distinct stromal cell types, their characterizing features, and their ontogenic relationships are still unclear. Single-cell atlases both provided better definition of distinct stromal cell groupings in health and disease and allowed better clarity

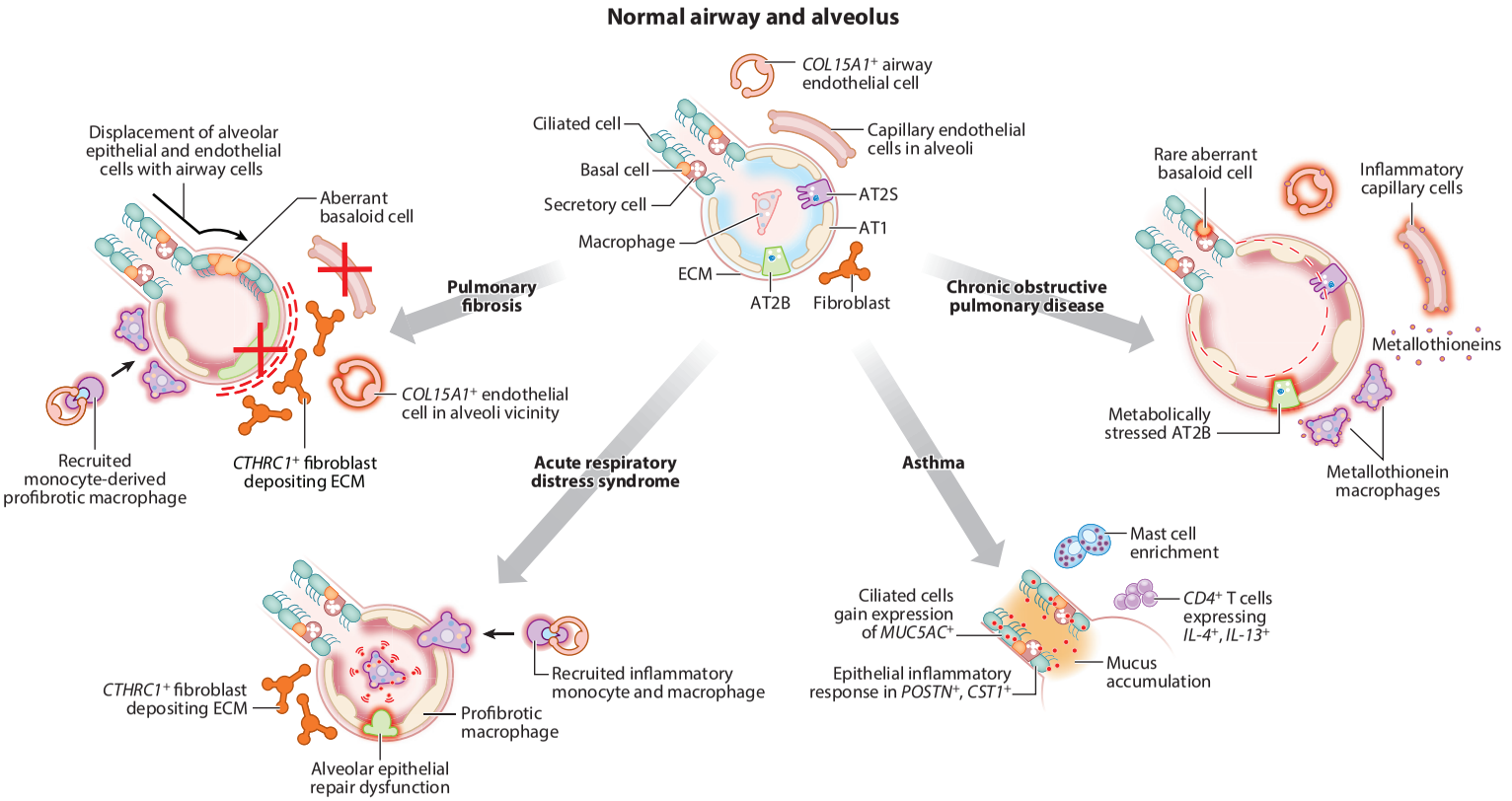

Figure 2

Overview of cells comprising lung tissue in health and disease. This illustrates alterations in cell composition and gene expression from healthy lung (*top*) to (*left to right*) pulmonary fibrosis (61, 62, 92, 101, 102), acute respiratory distress syndrome (67, 96, 97, 108–112), asthma (78, 84) and chronic obstructive pulmonary disease (75, 95). Abbreviations: AT1, alveolar epithelial type 1; AT2, alveolar epithelial type 2; ECM, extracellular matrix; IL, interleukin.

of what is unknown. Below are some of the insights obtained from single-cell profiling of lung stromal cells.

3.1.1. Fibroblasts in development and disease. The number of distinct fibroblast types, their features, and their ontogenic relationships are still unclear, as many commonly used molecular hallmarks of fibroblast subtypes like collagen or α-smooth muscle actin are more prominent in vitro and may only transiently occur under normal homeostasis or pathological conditions. One particularly confusing example includes cells referred to as myofibroblasts. During lung development, cells called myofibroblasts proliferate in the distal lung and function as scaffolds during the development of the alveolar walls. In the adult lung, the term myofibroblast is typically used to describe a pathological phenotype of fibroblast that accumulates in the distal lung during pulmonary fibrosis. In both cases, myofibroblasts are primarily described based on a very limited number of nonspecific marker genes: *ACTA2*, *TAGLN*, and *PDGFRA*. Indeed, much of what is known about fibroblasts, myofibroblasts, and other fibroblast subpopulations emerges from studies of *PDGFRA*-expressing cells during lung development and response to injury (57, 58).

When single-cell analysis was applied to human lungs, a more complex picture emerged; the myofibroblast cells described in both murine and human lung development are more closely related to smooth muscle cells (58, 59) based on reported expression of classical smooth-muscle markers *MYH11* and *ACTG2*. In contrast, the pathological myofibroblasts of pulmonary fibrosis more closely resemble adult resident alveolar fibroblasts based on global gene expression patterns and coexpression of resident alveolar fibroblast (also referred to as lipofibroblasts or *PLIN2*$^{+}$ fibroblasts) markers *ITGA8*, *MYLK*, and *TNC* (60–62). An unbiased assessment of phenotypic concordance between developmental and adult pathological myofibroblast should be conducted because the limited number of nonspecific markers used to describe these cells create a risk of spurious inference. Several independent single-cell analyses of fibrotic lungs have highlighted the poor sensitivity and specificity of *ACTA2* as a classical marker for pathological lung myofibroblasts (63–65). Taken alongside *ACTA2*'s poor correlation with collagen expression levels among pathological fibroblasts in idiopathic pulmonary fibrosis (IPF) (61), expression of the gene *CTHRC1* may serve as a more reliable marker for the pathological fibroblasts closely associated with fibrotic lesions (60, 66). A similar pathological fibroblast phenotype was recently reported in end-stage coronavirus 2019 (COVID-19) patients (67). Similar to the cells described in lesions of chronic pulmonary fibrosis patients, these cells express high levels of *CTHRC1* alongside other pathological hallmarks such as *COL1A1*, *COL3A1*, and *POSTN* (67). An integrated analysis of pathological fibroblasts that identifies the same gene signature is found in pathological fibroblasts of ulcerative colitis and both pancreatic and lung cancers (68); a recent study finds these cells similar to those in synovial tissues of patients with rheumatoid arthritis (69). Such distinctions are critically important, as traditionally, many drug screens aiming to target fibrosis focus on putative interventions to suppress the expression of *ACTA2* in fibroblasts in vitro. Given the poor specificity of *ACTA2* to human pathological lung cell populations, it would be prudent to consider replacing this single marker-based outcome with a more comprehensive disease-related signature gleaned from single-cell data.

3.1.2. A distinct pathological stromal cell. The possibility of discovering completely novel cell populations emerges with scRNA-seq, as has been the case in lymphangioleiomyomatosis (LAM), a rare disease caused by *TSC1* and *TSC2* mutations. scRNA-seq assessment of both lungs and uterus of LAM patients led to the identification of a disease-restricted mesenchymal cell that presented a common molecular and morphological phenotype in both organs (70) with a potential mechanistic role in the disease. This finding highlights the potential power of single-cell

profiling, but also a common limitation. Because of dissociation methods, most scRNA-seq studies now do not contain a large enough number of mesenchymal cells in humans, which allow only identification of the significant effects but not detailed and granular analyses. Methods that are independent of dissociation biases, larger numbers of tissues, and potentially cross-tissue analyses are required to expand on the observations mentioned above to create a truly detailed atlas of lung mesenchymal cells in health and disease.

3.2. Endothelial Cells

The primary function of the lung is to perform gas exchange on behalf of the respiratory system. Fittingly, the lung is a highly vascularized organ and the only mammalian organ with two circulatory systems: The ordinary systemic vasculature that supplies oxygen to tissue is normally restricted to major bronchial airways and lung pleura, while pulmonary circulation extends throughout the distal lung alveoli, exchanging carbon dioxide for oxygen. The diverse physiological and anatomical properties of lung endothelial cells make them compelling subjects of cell atlases. Vascular endothelial (VE) cells found in the lung can broadly be classified under three groups. Arterial VE cells transport blood away from the heart and can be molecularly identified by gap junction proteins GJA4 and GJA5 and venule VE cells transport blood back to the heart and express genes that facilitate immune cell recruitment such as *ACKR1* and *SELP*. VE capillary cells are microvessels that connect arterioles in venules in circulation and perform "last mile" transportation to tissues. VE capillary cells in the lung can collectively be identified by expression of *CA4* and *PRX*. scRNA-seq allowed better classification of known cell populations and distinction between previously indistinguishable cell populations (71).

3.2.1. New distinctions of vascular endothelial capillary cells. scRNA-seq revealed that the lung is populated by two distinct phenotypes of capillary cells: a generic capillary cell and a specialized capillary cell called an aerocyte. Aerocyte cells are dedicated to alveolar gas exchange and possess a phenotype that is uniquely found in lung alveoli (71–74). Molecularly, these cells can be distinguished from other capillary cells by expression of endothelin receptor *EDNRB*, transforming growth factor beta (TGF-β) inhibitor *SOSTDC1*, and the transcription factors *TBX2* and *FOXP2* (71, 74). Curiously, aerocytes uniquely express the prostaglandin metabolizing gene *HPGD* and completely lack expression of key coagulation and thrombosis mediators *VWF*, *SELP*, and *THBD* (71), suggesting that aerocytes constitutively repress inflammation and have a relatively limited ability to induce clotting after injury when compared to other lung endothelial cells. The macrolevel distribution of general and aerocyte capillary cells in lung parenchyma appears mosaic-like (71, 74), but on the microscale, aerocyte capillary cells each interface directly with an alveolar epithelial cell type 1 (AT1) epithelial cell (74), presumably where molecular oxygen and carbon dioxide are rapidly exchanged. In a large scRNA-seq study comparing chronic obstructive pulmonary disease (COPD) lungs, the authors noted an increased expression of inflammatory markers in capillary VEs, with the largest number of differentially expressed genes in general capillary cells. Connectomic analysis highlighted the expression of *CXCL12* by general capillary cells in COPD lungs, highlighting their potential role as regulators of lung inflammation in this disorder (75).

3.2.2. New distinctions of venular vascular endothelial cells. Lung cell atlases have also revealed that venous endothelial cells in the lung can also be separated into two distinct groups based on molecular phenotype. Pulmonary venous VE cells uniquely express *CPE*, *DKK3*, and prostaglandin synthases *PTGS1* and *PTGIS* and are found in the pulmonary-perfused lung parenchyma. In contrast, venule VE cells that are localized to systemically perfused bronchi and

visceral pleura of the lung have a distinct molecular phenotype based on expression of the connective tissue collagen *COL15A1*, as well as other distinct genes such as *VWA1*, *PLVAP*, and *ZNF385D* (71). A large scRNA-seq study comparing IPF and control lungs identified a significant increase in fibrotic lungs of the proportion of venule VE cells that have the phenotype associated with the systemically perfused, peribronchial venules. Cells with this phenotype were localized around lesions in affected regions of the distal lung parenchyma (61). It remains unclear whether these vessels are in fact systemically perfused, or perhaps the vasculature of affected tissues assumes this phenotype due to bronchiolization of the distal lung or because of mechanical stiffening in a fibrotic environment.

A recent multi-omics lung cell atlas study validates the paradigm of pulmonary and systemic venule VE cells being both spatially and molecularly distinct from one another. Interestingly, the authors used similar spatial-omics localization patterns to distinguish pulmonary and systemic arterial cells as well, but they did not report any notable differences in molecular phenotype between the two populations of arterial cells (76).

Together, these observations highlight the discovery potential of single-cell atlases but also their limitations. As the lung environment is distinct in its mechanical pressure and compliance gradients, it is possible that some of the characteristics are driven by the lung environment or blood supply, whereas others may be cell autonomous. A detailed analysis of VE cells in distinct lung compartments and regions, as well as in multiple pathological conditions, is needed.

3.2.3. Pulmonary lymphatic endothelium. The lymphatic system plays a complementary role to the blood circulatory system by draining interstitial fluid back to blood vascular veins and as channels for transporting immune cells between lymph nodes and peripheral tissues. A multicohort assessment of pulmonary endothelial cell atlases found that lymphatic endothelial cells share many core gene expression patterns with blood VE cells, including classical endothelial markers *PECAM1* (CD31), *CDH5* (VE-Cadherin), and *CLDN5* but are readily distinguished by expression of canonical features *PROX1*, *LYVE1*, *PDPN*, and *FLT4* and the chemotactic cytokine *CCL21* (71). scRNA-seq analysis of lungs from patients with LAM observed lymphatic endothelial-specific upregulation of vascular remodeling genes *UNC5B*, *ESAM*, and *ENG*, as well as immunomodulating genes *CD200* and *NECTIN2* (70), consistent with the disease's titular aberrant lymphangiogenesis that arises secondarily to the disease's characteristic stromal metastases. A detailed description of lymphatic endothelial cells in lung health and disease, as well as in distinct lung compartments, is not yet available.

3.3. Lung Epithelial Cells

Although airway and alveolar epithelial cells are probably among the best characterized cells in the lung, single-cell atlases revealed numerous exciting novel findings, including novel cells, cellular phenotypes, and their role, leading to a new appreciation of the phenotypic plasticity and diversity of lung epithelial cells in humans and mice.

3.3.1. An overview of airway epithelial cells. Human airways comprise a diverse array of epithelial cells with specialized functions, and basal airway stem cells give rise to ciliated cells alongside many hyperspecialized types of secretory cells of different rarities. These disparate cells operate in close harmony with each other to form a protective barrier from the environment and constantly clear debris away from the lower lung. Diseases such as asthma, ciliary dyskinesia, and cystic fibrosis are characterized by dysfunctional behavior of specific cells residing in this tissue. In addition to their diversity, airway epithelial cells make attractive candidates for cell atlases because

they are relatively accessible. Upper airway and nasal brushings are only marginally invasive. Moreover, the airway liquid interface (ALI) in vitro model allows primary basal cells to differentiate into their diverse array of cell progeny in a similar pseudostratified pattern as the in vivo airway wall. ALI is a powerful tool for studying multicellular airway epithelial function and development. Several single-cell atlas studies characterized potential airway cell subpopulations, including the distinct phenotypes of airway basal cells (77). Below, we focus on novel cells and disease states (53, 73, 78, 79).

3.3.2. Ionocytes. Perhaps the most widely influential contribution to lung biology made by single-cell atlases was the discovery of airway epithelial ionocyte cells. These rare cells are named for their role in regulating fluids at the epithelial surface via ion transport, are found in bronchial gland ducts of airways in both humans and mice, and can be identified by the transcription factor *FOXI1/Foxi1* (80, 81). These cells express the highest levels of CFTR of any cell in the lung, and mutations in this gene cause cystic fibrosis. Genetic knockout of *Foxi1* results in increased mucous viscosity and ciliary beat frequency similar to *Cftr* knockout models of cystic fibrosis in mice (81), further underscoring the ionocyte's putative relevance to cystic fibrosis. *FOXI1* knockout in a human ALI model resulted in increased ionic transmembrane potential similar to *CFTR* knockout models of cystic fibrosis (82), supporting the role of these cells in maintaining the ion transport potential of airway epithelium. A recent cell atlas of end-stage cystic fibrosis and control airway epithelium observed the increased expression of mutant *CFTR* among ionocytes from cystic fibrosis lungs (83). However, the relative rarity of ionocytes and the lack of cell-type specificity of *CFTR* expression indicated that the aggregate CFTR levels found in the lung primarily come from the abundance of basal and secretory cells (83), an observation also made in human control ALI (82).

3.3.3. Airway epithelial dysfunction in asthma. The number of mucin-secreting goblet cells was much higher in asthmatic airways than controls, and MUC5AC—a gel-forming mucin protein normally specific to secretory cells—was found expressed among a population of ciliated cells among asthma patients. Though mucous-ciliated cells are a known hallmark of remodeled asthmatic airways, global transcriptomic relationships between these cells and other airway epithelial cells indicated that this population represents ciliated cells that have acquired specific mucosal features, rather than a distinct transdifferentiation event between secretory and ciliated cell types (78). Many similar findings are described, including a shared inflammatory response signal across secretory and ciliated cells of *POSTN1* and *CST1* (**Figure 2**). The gain of MUC5AC expression was also observed among interleukin 13 (IL-13)-treated ciliated cells (84). No MUC5AC protein was detected, and other genes associated with functional gel mucosa formation such as *FCGBP*, *ITLN1*, and *SCIN* were absent from mucosal ciliated cells. One hypothesis for this phenomenon involves ciliated cells attempting to generate mucin protein without the necessary machinery, leading to endoplasmic reticulum stress and subsequent cell death. While this is an interesting explanation for the decline in ciliated cells and reduced mucosal clearance in asthma, it remains unknown whether this finding is relevant to asthmatic ciliated cells in vivo.

3.3.4. Alveolar epithelial cells. Alveolar epithelial cells continuously cover 99% of the surface area of the lung (~100 m^2) and comprise mainly two cell types, AT1 and alveolar epithelial cell type 2 (AT2) (85). Because of functional constraints—gas exchange requires an extremely thin layer—the alveolus cannot contain a multitude of cells; thus, these two cell types are tasked with the complex functions of gas exchange, maintenance of alveolar stability, and fluid balance and local repair. AT1 cells that occupy 96% of the lung surface are extremely thin cells overlying

capillaries and characterized by the expression of *AGER*, *AQP5*, and *PDPN*. AT2 cells make and secrete pulmonary surfactant to keep the alveolus open during breathing, are the progenitor cells for AT1 cells, absorb fluid from the alveolus, and regulate lung innate immunity (86, 87). Characteristically, they are large cuboidal cells located in the alveolar corners and contain lamellar bodies. Their most common markers include surfactant proteins, especially SFTPC. Single-cell connectomic analysis of mammalian lungs identified a role for AT1 in the homeostatic regulation of the alveolar niche (88), which has also been observed in lung development (89). A later analysis of human lungs suggested that alveolar epithelial cells may exhibit a diversity of cell states and phenotypes, potentially including two subtypes of AT2 cells. One subtype, AT2B, that expresses high levels of surfactant genes *WIF1*, *HHIP*, and *CA2*$^{+}$, potentially represents an AT2 subpopulation with a role in alveolar maintenance, and another subtype called AT2S that expresses WNT signaling genes is potentially representative of a group with more stem cell properties (73). The COPD cell atlas study mentioned above identified the same subclasses of AT2 cells. AT2B had the highest number of differentially expressed genes when compared to controls (**Figure 2**) and the greatest enrichment for the expression of genes with polymorphisms associated with COPD-related traits among epithelial cells. Among the genes reduced in AT2B in COPD lungs was *NUPR1*, a cellular stress response gene and positive regulator of antioxidants (75). In IPF, scRNA-seq revealed a substantial decrease in AT1 and AT2 in the distal lung parenchyma and an increase in airway epithelial cells (61), highlighting the loss of gas exchange units caused by the aberrant lung remodeling in the disease.

3.3.5. Terminal airway secretory cells as alveolar progenitors. Two recent studies describe how a unique subset of terminal airway secretory cells serve as progenitors to alveolar epithelial cells. Terminal respiratory bronchial stem cells (90) or respiratory airway secretory (91) cells are characterized by coexpression of *SFTPB* and *SCGB3A2* and are only found in the distal airways at the junction of bronchioles and alveoli. Histological exploration of these cells across tissues of various lung diseases revealed an enrichment of these cells in severely affected regions of IPF lungs (90), supporting the idea that abnormal bronchiolization occurs in place of ordinary alveolar epithelial repair (**Figure 2**). In lungs of patients with COPD, a subpopulation of AT2 cells found featuring the expression of *SCGB3A2* were enriched (91), suggesting an abnormal, incomplete differentiation process from secretory to alveolar epithelium (**Figure 2**). This phenomenon of partial AT2 status in lungs was further observed in exposure models of cigarette smoke in ferrets (91), indicating that smoking-induced injuries to lung epithelium are an explanation for this phenomenon in COPD.

3.3.6. Aberrant basaloid cells in advanced lung disease. Two of the largest lung cell atlas papers simultaneously reported the existence of a distinct cell state exclusive to the affected tissue regions of patients with advanced lung disease (61, 62). Aberrant basaloid cells express a conflicting signature of features normally restricted to either airway or alveolar epithelial cells. Similar to airway basal cells, they express *TP63* and *KRT17*, but unlike airway basal cells, they have minimal downstream expression of the *TP63*-induced hallmark genes *KRT5*, *KRT15*, or *MIR205HG*. Similar to alveolar epithelial cells, they express high levels of *SOX9*, *NAPSA*, *ITGB6*, and collagen type IV genes, but lack canonical AT1 or AT2 markers such as *AGER* or *SFTPC*. Notably, these cells demonstrated a wide variety of pathogenic features, including epithelial-mesenchymal transition features; *CDH2*, *VIM*, *FN1*, and *COL1A1*; markers of injury such as *MDK*, *GDF15*, and *PTGS2*; senescence features such as *CDKN1A*, *CDKN2A*, *CDKN2B*, *CCDN2*, *MDM2*, and *HMGA2*; the TGF-β-activating integrin genes *ITGB6*, *ITGB8*, and *ITGAV*; and *MMP7*, an IPF biomarker. These cells have been detected in the lungs of several different advanced lung diseases

including IPF, non-specific interstitial pneumonia (NSIP), hypersensitivity pneumonitis, and systemic scleroderma, and rarely in COPD (61, 62, 92–94). In IPF, these cells spatially localize to the surface of fibroblastic foci (61, 62) (**Figure 2**). In COPD, analysis of clonally expanded $p63^+$ epithelial cells revealed similar basaloid cells that show elevated *CLDN4*, *SOX9*, *FN1*, *PTGS2*, *MMP7*, *CDKN2B*, and *CCND2*. Immunohistochemistry of end-stage COPD distal lung localized these cells to the surface of metaplastic lesions (95). Similar cell phenotypes have also been described lining the cysts of transplanted lungs from two patients with severe COVID-19 (96) as well as in COVID-19 autopsy samples (97).

In an effort to explain the cellular source of aberrant basaloid cells, it has been proposed that these cells most closely resemble a transitional intermediate state between AT2 and AT1 cells observed in murine models (98). Pre-alveolar type-1 transitional cell state (PATS) (99) or $Krt8^+$ transitional cells (100) lack canonical markers of murine AT1 or AT2, but similar to aberrant basaloid cells, they show elevated *Cldn4*, *Cdkn1a*, *Ccdn1*, and *Fn1*. Importantly, these murine intermediate cells belong to a normal process of lung repair and resolution, whereas aberrant basaloid cells have only been identified in advanced parenchymal lung disease. The signature that includes expression of airway basal markers such as *TP63* and *KRT17*—but not *KRT5*—senescence markers (*CDKN1A*, *CDKN2A*), epithelial to mesenchymal transition (*CDH2*, *COL1A1*), as well as others listed above and previously (61) remains unique to aberrant basaloid cells, and their origins in humans are unclear. The unexpected novel finding of the aberrant basaloid cells underscores the discovery power of single-cell profiling technologies. While some features of these cells were observed, their existence as a distinct entity with a unique combination of markers was not expected or hypothesized, and their discovery opens new pathways for the development of novel therapeutics and diagnostics.

3.4. Immune Cells

Although the lung has a wide and rapidly changing repertoire of inflammatory cells, many of the insights into this organ have been on macrophages and monocytes. We expect this repertoire to widen and expand as more samples, conditions, and compartments are studied. Below, we describe some of these insights, mostly in the context of fibrosis and acute respiratory distress syndrome.

3.4.1. Macrophages. There are two phenotypically distinct macrophage populations commonly identified in the healthy human lung. Tissue-resident alveolar macrophages (r-AM) arise from embryonic progenitor cells, can maintain themselves through self-renewal, and play specialized roles as both sentinels and custodians of the distal lung by phagocytosing foreign particles and recycling surfactant proteins. r-AM can be transcriptionally distinguished by a stereotypic gene expression signature of the lipid-interacting genes *FABP4*, *APOC*, and *PPARG* and delineated from other macrophages by expression of *SIGLEC1* (CD163). In contrast, monocyte-derived macrophages have a less stereotypic phenotypic signature and can be observed assuming relatively extremal phenotypes of inflammation or wound repair when needed. While many other findings have been described, we focus here on two populations: profibrotic macrophages found in fibrotic lung diseases and inflammatory myeloid populations found in severe COVID-19 lung disease.

3.4.2. Profibrotic macrophages. scRNA-seq analysis of end-stage pulmonary fibrosis and control lung dissociates performed by Reyfman et al. (101) described a distinct subpopulation of macrophages only in fibrotic lungs with elevated expression of profibrotic genes, a phenomenon that has since been confirmed across many subsequent cell atlases of human pulmonary fibrosis, including human systemic scleroderma, hypersensitivity pneumonitis, and IPF (61, 101, 102).

These cells are defined by elevated expression of extracellular matrix–interacting genes *SPP1*, *MMP9*, *SPARC*, and *FN1*; elevated expression of lysosomal activity genes *LPL*, *LIPA*, *LGMN*, and *CTSK*; and coexpression of genes *MERTK*, *PLA2G7*, *SIGLEC15*, *CHIT1*, and *CHI3L1*. These cells were observed coexisting in the same alveolar spaces of fibrotic lungs among typical *FABP4*$^+$ r-AM (101), though seen in higher frequency in the lower, more affected lobes of IPF patients (102). scRNA-seq analysis of lung immune cells from West Highland terriers with canine IPF describes a similar macrophage phenotype based on overexpression of *SPP1*, *FN1*, and *CXCL8* (103). scRNA-seq and bulk RNA-seq of sorted cells of murine lung fibrosis models have likewise revealed a similar phenotype among macrophages based on coexpression of *Chil3*, *Il18*, *Ctsk*, *Sparc*, and *Mertk* (104, 105). Murine lung macrophages with this phenotype arise from newly recruited monocyte-derived macrophages. Monocyte depletion preceding injury attenuates fibrosis, whereas r-AM depletion has no effect, indicating an active role of these cells during wound repair. Following recruitment and activation, these cells persist in the lung indefinitely, gradually losing their profibrotic phenotype until becoming functionally indistinguishable from r-AM (105, 106). Acute injuries from both bleomycin and influenza are shown to irreversibly disrupt the relative makeup of embryonically derived r-AM (105). Recent cell atlases have identified a similar profibrotic macrophage population in bronchial alveolar lavage fluid from severe COVID-19 patients (107, 108) and in tissue dissociates of end-stage COVID-19 lung explants (96). scRNA-seq interrogation of a COVID-19 infection model in ferret lungs describes the arrival of *CHIT1*$^+$/*SPP1*$^+$ macrophages alongside inflammatory, *IL1B*-expressing macrophages two days after infection. These persist after inflammatory macrophage proportions decline (109), which is consistent with the latent wound repair process associated with this cell phenotype. Impressively, human monocyte exposure to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was sufficient to induce a profibrotic macrophage phenotype in vitro but not type A influenza or ligands of viral RNA sensors (108). Collectively, these findings indicate that SARS-CoV-2 has a distinct capacity to induce a fibrotic phenotype among myeloid cells, and this profibrotic phenotype is a common hallmark of many different types of severe lung injury.

3.4.3. Inflammatory myeloid cells in COVID acute respiratory distress syndrome. scRNA-seq analysis of bronchoalveolar lavage (BAL) from moderate, severe, and critical COVID-19 patients shows a common inflammatory signature of *IL1B*, *IL6*, *TNF*, *CCL2*, *CCL3*, *CCL4*, *CCL7*, *CXCL9*, *CXCL10*, and *CXCL11* that correlates with disease severity. This signature was evident among both classical monocytes (*CD14*, *FCN1*, *S100A8* expressing) and monocyte-derived macrophages (*SPP1*, *CCL2*, *CCL3*, *CXCL10* expressing) (107), with the T cell recruiting of *CXCL16* highest among inflammatory macrophages in moderate COVID-19 patients, perhaps representative of a transition from innate to adaptive immunity response. A similar scRNA-seq study of COVID-19 BAL from differentially affected patients reported a similar disease severity–associated phenotype among classical monocytes (*CD14*, *FCN1*, *S100A8/9*). Both inflammatory monocytes and macrophages increase with disease severity (110). A multi-tissue COVID-19 cell atlas comprising nasopharyngeal brushings, bronchial brushings, and BALs of COVID-19 patients with various disease severity also identified this inflammatory myeloid signature (111). A reference atlas comparing immune cells from COVID-19 patients with separate inflammatory or interstitial lung diseases determined that these *FCN1*$^+$ and *CXCL10*$^+$/*CCL2*$^+$ inflammatory myeloid cells found abundantly in the lungs of patients with severe COVID-19 are remarkably similar to inflammatory myeloid cells found in synovial tissue of patients with rheumatoid arthritis or in the intestinal walls of patients with inflammatory bowel diseases (112). Such cross-tissue, cross-disease comparisons may provide important insights for drug repurposing. Notably, the presence of these inflammatory macrophages and monocytes was not observed among

COVID-19 cell atlases that used samples collected following either transplant or autopsy, suggesting that the inflammatory mechanisms driving acute respiratory distress syndrome (ARDS) severity and fibrosis may differ (113).

4. IMPACT AND CAVEATS OF LUNG CELL ATLASES

The impact of lung cell atlases on the understanding of the cellular repertoires and potential mechanisms of advanced lung disease has been transformative (**Figure 2**). As foreshadowed by the early discovery of ionocytes, scRNA-seq studies of human lungs are loaded with novel findings. No field has been more influenced than pulmonary fibrosis, as perception of the disease has been completely transformed by multiple novel insights. The extent of the loss of specialized alveolar epithelial and capillary endothelial cells, reflecting the loss of gas exchange units, and the proximalization of the distal lung were not appreciated before. The presence of previously undescribed cell populations such as systemic venular VE cells, aberrant basaloid cells, and the profibrotic macrophages was unknown before scRNA-seq was applied to human lungs. And the need to redefine the molecular signature of the hallmark cell of pulmonary fibrosis, the myofibroblast, has never been more acute than now. Impressively, similar findings were found in all pulmonary fibrosis cell atlases (61, 62, 101, 102, 105, 108), as can be gleaned in the IPF Cell Atlas data sharing and dissemination portal (**http://www.IPFCellAtlas.com**; 114). In COPD, the multitude of insights, including the potential specific involvement of AT2B and the role of endothelial inflammation and metallothionein-expressing macrophages, are all important; some are completely novel, and some are confirmatory but with an unprecedented depth and detail (75). In the airway, the discovery of ionocytes, the *MUC5AC*-expressing ciliated cells, and the depth of immune cell population asthma and cystic fibrosis is shifting our understanding of the airways and their complex response stimuli in common and rare disease (78, 81, 84, 115). Thus, even at this early phase, it is hard not to be excited about the progress and potential future impact of single-cell lung atlases but also to recognize the challenges they pose. Perhaps the biggest challenge relates to cell types.

4.1. Challenges with Cell Classification

Borrowing from how taxonomists describe the species problem (116), the problem of defining cell types can be thought of as the result of two conflicting goals of cellular biologists: to categorize and label cell types as naturally distinct entities and learn the ways cells change or give rise to other cells. It is important to consider that, when applying ontological constructs to classify cells and their relationships to one another, distinctions between one ontological model and another will not always translate. Many traditional cell classifiers have direct biological meaning, such as the cell's anatomical niche, physiological function, or developmental trajectory; heuristic approaches such as marker genes are commonly used as surrogates for all of these attributes. However, such marker genes may not necessarily represent the functional essence of a cell type or subpopulation but represent a feature that can appear in different cell types with different functions as parts of distinct developmental programs or response to injury. They are also dependent on technical characteristics of the cell profiling technology. Indeed, using such markers in cell atlas data often reveals that they only reflect large groupings of cells and miss granular subtypes, or worse, as in the case of *ACTA2*, they may not even reflect or associate with the defining features of the cell subpopulation. Such examples serve as the modern equivalent of the curved neck morphology behind the barnacle goose myth. Simultaneous assessments of single chromatin accessibility and protein markers may help to ameliorate this issue, but the most important answer is conceptual. To fully capitalize on the power of lung cell atlases we must avoid the habit of defining cells based

on small sets of markers often derived in vitro or in animal systems and instead prioritize cellular signatures that are reproducible across multiple individuals.

This brings us to a second important point, which is the need to increase both the diversity and number of individuals studied, as well as the number of cells profiled. The number of lung cells profiled in a recent report of the comprehensive lung cell exceeds 2.2 million cells obtained from more than 400 individuals (53). This is an impressive number, until one considers that the average number of cells in one human lung is estimated at more than 20×10^{10} cells (117). It is unclear that ~5,000 cells represent an adequate sampling of the lung, especially when we move beyond the large effect sizes of disease and try to create a real compendium of all the cells and their states in the human lung. Similarly, the diversity of human attributes from race, ethnicity, and sex; activity, altitude, and climate; environmental, occupational, and recreational exposures; and non-respiratory medical conditions and medications is likely to influence the repertoire of cellular states in healthy adults. Thus, approaches that do not minimize variance by integration but instead embrace and capture the range of individual variance in health and disease are required.

4.2. The Importance of Integrated Biological and Computational Expertise

In this review we intentionally avoid detailed technical discussions of computational methods. However, the role of computational approaches cannot be understated. Single-cell experiments generate vast amounts of data that require substantial computational manipulation before interpretation is possible. The tools and approaches to data quality control, normalization, correction, feature selection, and dimensionality reduction as well as cell annotation and gene expression analyses and result visualization are usually beyond the training of most experimental biologists (118). Even for data scientists, single-cell profiling poses unique challenges, and the term single-cell data science was coined because of the unique challenges posed by scRNA-seq that include data sparsity, uncertainty, varying resolution, and complexity (119). Add to that the rapid pace of analytical tool development, the nontrivial computational power required to run these analyses, and the vast amount of biological information associated with the results and it becomes clear that data scientists, and potentially data science teams with expertise in computer science, biostatistics, and bioinformatics, must play a central role in cell atlas teams. However, it is also important to resist putting blind faith into computational methods when approaching the data. Qualitative assessments of the results by biological and medical domain experts, as well as experimental validations and iterative interactions in multidisciplinary teams, are required, as artifacts of computational tools may lead to spurious interpretations or overlooking unexpected but important results. Thus, the road toward generating each cell atlas is profoundly interdisciplinary and often benefits from an iterative and collective approach to investigation rather than a jurisdictional baton pass between biologists, clinicians, and computational researchers.

5. FUTURE ISSUES

The scale of data behind every cell atlas far exceeds what can be described in a scientific manuscript, requiring authors to make tactical decisions about what cell types or features to highlight while leaving an inordinate amount of information unmentioned. This information is usually deposited in publicly available databases but remains invisible to most of the scientific community, as substantial computational expertise and time commitment are required to mine it. Thus, it is critically important that data sharing and dissemination portals are made available to allow researchers with limited computational expertise to have exploratory access to cell atlas results. The recent success of the online IPF Cell Atlas (114; **http://www.IPFCellAtlas.com**) highlights this unmet need;

since its inception, 9,500 individual users looked for 15,595 genes in 314,000 sessions, indicating the substantial demand for user-friendly, easily accessible, single-cell data portals.

In this review, we summarize the recent impressive insights generated by the human lung cell atlases. Single-cell profiling technologies are already transforming biomedical research with novel insights in all domains of lung health and disease. The incorporation of additional types of information, including proteomics, metabolomics, extracellular matrix composition, biomechanical metrics, nitric oxide signaling patterns, and spatial technologies, will be necessary to achieve a comprehensive, systems-level understanding of the lung as a multicellular organ. Similarly, to create a real compendium of all cells and their states in the human lung, systematic expansion of the diversity of samples analyzed is required. This must include information on race, ethnicity, and sex; activity, altitude, and climate; environmental, occupational, and recreational exposures; and medical conditions and medications that are likely to influence the repertoire of cellular states across health and disease. Finally, cell atlas studies of complex in vitro models like cell organoids or ex vivo models of precision cut lung slices can provide detailed perspectives of intercellular relationships and tissue mechanics that cannot directly be modeled in humans. With the addition of biological domains, enhanced diversity of samples, and use of such models, we can now start envisioning a real comprehensive atlas of the human lung in health and disease that will serve as a blueprint to develop approaches to cure advanced lung disease.

DISCLOSURE STATEMENT

N.K. has served as a consultant to Boehringer Ingelheim, Third Rock, Pliant, Samumed, NuMedii, Theravance, LifeMax, Three Lake Partners, OptiKira, AstraZeneca, CohBar, Veracyte, Augmanity, CSL Behring, Galapagos, Gilead, Chiesi, Arrowhead, Spinnova, and Thyron over the last three years. N.K. reports equity in Pliant and Thyron, grants from Veracyte, Boehringer Ingelheim, and Bristol Myers Squibb, and nonfinancial support from miRagen and AstraZeneca. N.K. has intellectual property on novel biomarkers and therapeutics in idiopathic pulmonary fibrosis licensed to Biotech. T.S.A. and A.M. are not aware of any additional affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors are grateful to the many colleagues who provided input and the patient communities that participate in research. This work was partially supported by the US National Institutes of Health (NIH) grants R01HL141852, R01HL127349, U01HL145567, R21HL161723, and U01HL122626 and generous donations from Three Lakes Partners and the Collins Foundation.

LITERATURE CITED

1. Whitsett JA, Kalin TV, Xu Y, Kalinichenko VV. 2019. Building and regenerating the lung cell by cell. *Physiol. Rev.* 99:513–54
2. Franks TJ, Colby TV, Travis WD, Tuder RM, Reynolds HY, et al. 2008. Resident cellular components of the human lung. *Proc. Am. Thorac. Soc.* 5:763–66
3. Clevers H, Rafelski S, Elowitz M, Klein A, Shendure J, et al. 2017. What is your conceptual definition of "cell type" in the context of a mature organism? *Cell Syst.* 4:255–59
4. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, et al. 2017. The Human Cell Atlas. *eLife* 6:e27041
5. Tabula Muris Consortium. 2018. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* 562:367–72
6. Mercator G, Mercator R. 1595. *Atlas sive cosmographicæ meditationes de fabrica mvndi et fabricati figvra.* Duisburg, Ger.: Dvisbvrgi Clivorvm

7. Playfair W. 1786. *The Commercial and Political Atlas: Representing, by Means of Stained Copper-Plate Charts, the Exports, Imports, and General Trade of England; the National Debt, and Other Public Accounts; with Observations and Remarks*. London: J. Debrett
8. Robinson JL, Kocabaş P, Wang H, Cholley P-E, Cook D, et al. 2020. An atlas of human metabolism. *Sci. Signal.* 13:eaaz1482
9. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, et al. 2015. Proteomics. Tissue-based map of the human proteome. *Science* 347:1260419
10. Nowinski WL. 2017. Human brain atlasing: past, present and future. *Neuroradiol. J.* 30:504–19
11. Lanzenberger R, Beisteiner R, Prayer D. 2002. CD-ROM review: discussion of the brain atlas for functional imaging (BAFI). *Eur. J. Radiol.* 41:255–56
12. Mikula S, Trotts I, Stone JM, Jones EG. 2007. Internet-enabled high-resolution brain mapping and virtual microscopy. *Neuroimage* 35:9–15
13. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, et al. 2007. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445:168–76
14. Dries R, Chen J, Del Rossi N, Khan MM, Sistig A, Yuan GC. 2021. Advances in spatial transcriptomic data analysis. *Genome Res.* 31:1706–18
15. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. 1995. Serial analysis of gene expression. *Science* 270:484–87
16. Tremain N, Korkko J, Ibberson D, Kopen GC, DiGirolamo C, Phinney DG. 2001. MicroSAGE analysis of 2,353 expressed genes in a single cell-derived colony of undifferentiated human mesenchymal stem cells reveals mRNAs of multiple cell lineages. *Stem Cells* 19:408–18
17. Ye SQ, Lavoie T, Usher DC, Zhang LQ. 2002. Microarray, SAGE and their applications to cardiovascular diseases. *Cell Res.* 12:105–15
18. van Ruissen F, Ruijter JM, Schaaf GJ, Asgharnegad L, Zwijnenburg DA, et al. 2005. Evaluation of the similarity of gene expression data estimated with SAGE and Affymetrix GeneChips. *BMC Genom.* 6:91
19. Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–70
20. Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *PNAS* 91:5022–26
21. Cossman J. 2001. Gene expression analysis of single neoplastic cells and the pathogenesis of Hodgkin's lymphoma. *J. Histochem. Cytochem.* 49:799–800
22. Hemby SE, Ginsberg SD, Brunk B, Arnold SE, Trojanowski JQ, Eberwine JH. 2002. Gene expression profile for schizophrenia: discrete neuron transcription patterns in the entorhinal cortex. *Arch. Gen. Psychiatry* 59:631–40
23. Ginsberg SD, Che S, Counts SE, Mufson EJ. 2006. Single cell gene expression profiling in Alzheimer's disease. *NeuroRX* 3:302–18
24. Kurimoto K, Yabuta Y, Ohinata Y, Ono Y, Uno KD, et al. 2006. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res.* 34:e42
25. Chiang MK, Melton DA. 2003. Single-cell transcript analysis of pancreas development. *Dev. Cell* 4:383–93
26. Ning W, Li CJ, Kaminski N, Feghali-Bostwick CA, Alber SM, et al. 2004. Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease. *PNAS* 101:14895–900
27. Zuo F, Kaminski N, Eugui E, Allard J, Yakhini Z, et al. 2002. Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. *PNAS* 99:6292–97
28. Holt RA, Jones SJ. 2008. The new paradigm of flow cell sequencing. *Genome Res.* 18:839–46
29. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, et al. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6:377–82
30. Smith AM, Heisler LE, St. Onge RP, Farias-Hesson E, Wallace IM, et al. 2010. Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic Acids Res.* 38:e142
31. Islam S, Kjällquist U, Moliner A, Zajac P, Fan JB, et al. 2011. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* 21:1160–67

32. Tan SJ, Phan H, Gerry BM, Kuhn A, Hong LZ, et al. 2013. A microfluidic device for preparing next generation DNA sequencing libraries and for automating other laboratory protocols that require one or more column chromatography steps. *PLOS ONE* 8:e64084
33. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, et al. 2014. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509:371–75
34. Hashimshony T, Wagner F, Sher N, Yanai I. 2012. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep*. 2:666–73
35. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, et al. 2014. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* 11:163–66
36. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, et al. 2013. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498:236–40
37. Fan HC, Fu GK, Fodor SP. 2015. Combinatorial labeling of single cells for gene expression cytometry. *Science* 347:1258367
38. Han X, Wang R, Zhou Y, Fei L, Sun H, et al. 2018. Mapping the mouse cell atlas by microwell-Seq. *Cell* 172:1091–107.e17
39. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, et al. 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–14
40. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, et al. 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161:1187–201
41. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, et al. 2018. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* 360:176–82
42. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, et al. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* 14:865–68
43. Peterson VM, Zhang KX, Kumar N, Wong J, Li L, et al. 2017. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* 35:936–39
44. Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, et al. 2018. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* 361:1380–85
45. Chen S, Lake BB, Zhang K. 2019. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* 37:1452–57
46. Lareau CA, Ludwig LS, Muus C, Gohil SH, Zhao T, et al. 2021. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat. Biotechnol.* 39:451–61
47. Mimitou EP, Lareau CA, Chen KY, Zorzetto-Fernandes AL, Hao Y, et al. 2021. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat. Biotechnol.* 39:1246–58
48. Nature Methods. 2014. Method of the Year 2013. Methods to sequence the DNA and RNA of single cells are poised to transform many areas of biology and medicine. *Nat. Methods* 11:1
49. Nature Methods. 2020. Method of the Year 2019: single-cell multimodal omics. *Nat. Methods* 17:1
50. Nature Methods. 2021. Method of the Year 2020: spatially resolved transcriptomics. *Nat. Methods* 18:1
51. Svensson V, da Veiga Beltrame E, Pachter L. 2020. A curated database reveals trends in single-cell transcriptomics. *Database* 2020:baaa073
52. Sperr E. 2016. *PubMed by year*. **https://esperr.github.io/pubmed-by-year/**
53. Sikkema L, Strobl D, Zappia L, Madissoon E, Markov N, et al. 2022. An integrated cell atlas of the human lung in health and disease. bioRxiv 483747. **https://doi.org/10.1101/2022.03.10.483747**
54. Ardini-Poleske ME, Clark RF, Ansong C, Carson JP, Corley RA, et al. 2017. LungMAP: the molecular atlas of lung development program. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 313:L733–40
55. Sun X, Perl AK, Li R, Bell SM, Sajti E, et al. 2022. A census of the lung: CellCards from LungMAP. *Dev. Cell* 57:112–45.e2
56. Nissan E. 2015. A British cultural history of biological taxa. *La Ric. Folk.* 70:305–12
57. Green J, Endale M, Auer H, Perl A-KT. 2016. Diversity of interstitial lung fibroblasts is regulated by platelet-derived growth factor receptor α kinase activity. *Am. J. Respir. Cell Mol. Biol.* 54:532–45
58. Li R, Bernau K, Sandbo N, Gu J, Preissl S, Sun X. 2018. *Pdgfra* marks a cellular lineage with distinct contributions to myofibroblasts in lung maturation and injury response. *eLife* 7:e36865

59. Liu X, Rowan SC, Liang J, Yao C, Huang G, et al. 2021. Categorization of lung mesenchymal cells in development and fibrosis. *iScience* 24:102551
60. Tsukui T, Sun KH, Wetter JB, Wilson-Kanamori JR, Hazelwood LA, et al. 2020. Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat. Commun.* 11:1920
61. Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, et al. 2020. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* 6:eaba1983
62. Habermann AC, Gutierrez AJ, Bui LT, Yahn SL, Winters NI, et al. 2020. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci. Adv.* 6:eaba1972
63. Xue L, Simon CR, Jiurong L, Changfu Y, Guanling H, et al. 2021. Categorization of lung mesenchymal cells in development and fibrosis. *iScience* 24:102551
64. Sun KH, Chang YG, Reed NI, Sheppard D. 2016. α-Smooth muscle actin is an inconsistent marker of fibroblasts responsible for force-dependent TGFβactivation or collagen production across multiple models of organ fibrosis. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 310:L824–36
65. Peyser R, MacDonnell S, Gao Y, Cheng L, Kim Y, et al. 2019. Defining the activated fibroblast population in lung fibrosis using single-cell sequencing. *Am. J. Respir. Cell Mol. Biol.* 61:74–85
66. Valenzi E, Bulik M, Tabib T, Morse C, Sembrat J, et al. 2019. Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease. *Ann. Rheum. Dis.* 78:1379–87
67. Melms JC, Biermann J, Huang H, Wang Y, Nair A, et al. 2021. A molecular single-cell lung atlas of lethal COVID-19. *Nature* 595:114–19
68. Buechler MB, Pradhan RN, Krishnamurty AT, Cox C, Calviello AK, et al. 2021. Cross-tissue organization of the fibroblast lineage. *Nature* 593:575–79
69. Korsunsky I, Wei K, Pohin M, Kim EY, Barone F, et al. 2022. Cross-tissue, single-cell stromal atlas identifies shared pathological fibroblast phenotypes in four chronic inflammatory diseases. *Med* 3:481–518.e14
70. Guo M, Yu JJ, Perl AK, Wikenheiser-Brokamp KA, Riccetti M, et al. 2020. Single-cell transcriptomic analysis identifies a unique pulmonary lymphangioleiomyomatosis cell. *Am. J. Respir. Crit. Care Med.* 202:1373–87
71. Schupp JC, Adams TS, Cosme C Jr., Raredon MSB, Yuan Y, et al. 2021. Integrated single-cell atlas of endothelial cells of the human lung. *Circulation* 144:286–302
72. Elmentaite R, Dominguez Conde C, Yang L, Teichmann SA. 2022. Single-cell atlases: shared and tissue-specific cell types across human organs. *Nat. Rev. Genet.* 23:395–410
73. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, et al. 2020. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 587:619–25
74. Gillich A, Zhang F, Farmer CG, Travaglini KJ, Tan SY, et al. 2020. Capillary cell-type specialization in the alveolus. *Nature* 586:785–89
75. Sauler M, McDonough JE, Adams TS, Kothapalli N, Barnthaler T, et al. 2022. Characterization of the COPD alveolar niche using single-cell RNA sequencing. *Nat. Commun.* 13:494
76. Madissoon E, Oliver AJ, Kleshchevnikov V, Wilbrey-Clark A, Polanski K, et al. 2021. A spatial multi-omics atlas of the human lung reveals a novel immune cell survival niche. bioRxiv 470108. **https://doi.org/10.1101/2021.11.26.470108**
77. Carraro G, Mulay A, Yao C, Mizuno T, Konda B, et al. 2020. Single-cell reconstruction of human basal cell diversity in normal and idiopathic pulmonary fibrosis lungs. *Am. J. Respir. Crit. Care Med.* 202:1540–50
78. Vieira Braga FA, Kar G, Berg M, Carpaij OA, Polanski K, et al. 2019. A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.* 25:1153–63
79. Deprez M, Zaragosi L-E, Truchi M, Becavin C, Ruiz García S, et al. 2020. A single-cell atlas of the human healthy airways. *Am. J. Respir. Crit. Care Med.* 202:1636–45
80. Plasschaert LW, Zilionis R, Choo-Wing R, Savova V, Knehr J, et al. 2018. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* 560:377–81

81. Montoro DT, Haber AL, Biton M, Vinarsky V, Lin B, et al. 2018. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* 560:319–24
82. Goldfarbmuren KC, Jackson ND, Sajuthi SP, Dyjack N, Li KS, et al. 2020. Dissecting the cellular specificity of smoking effects and reconstructing lineages in the human airway epithelium. *Nat. Commun.* 11:2485
83. Carraro G, Langerman J, Sabri S, Lorenzana Z, Purkayastha A, et al. 2021. Transcriptional analysis of cystic fibrosis airways at single-cell resolution reveals altered epithelial cell states and composition. *Nat. Med.* 27:806–14
84. Jackson ND, Everman JL, Chioccioli M, Feriani L, Goldfarbmuren KC, et al. 2020. Single-cell and population transcriptomics reveal pan-epithelial remodeling in type 2-high asthma. *Cell Rep.* 32:107872
85. Weibel ER. 2015. On the tricks alveolar epithelial cells play to make a good lung. *Am. J. Respir. Crit. Care Med.* 191:504–13
86. Mason RJ. 2006. Biology of alveolar type II cells. *Respirology* 11:S1215
87. Guillot L, Nathan N, Tabary O, Thouvenin G, Le Rouzic P, et al. 2013. Alveolar epithelial cells: master regulators of lung homeostasis. *Int. J. Biochem. Cell Biol.* 45:2568–73
88. Raredon MSB, Adams TS, Suhail Y, Schupp JC, Poli S, et al. 2019. Single-cell connectomic analysis of adult mammalian lungs. *Sci. Adv.* 5:eaaw3851
89. Zepp JA, Morley MP, Loebel C, Kremp MM, Chaudhry FN, et al. 2021. Genomic, epigenomic, and biophysical cues controlling the emergence of the lung alveolus. *Science* 371:eabc3172
90. Murthy PKL, Sontake V, Tata A, Kobayashi Y, Macadlo L, et al. 2022. Human distal lung maps and lineage hierarchies reveal a bipotent progenitor. *Nature* 604:111–19
91. Basil MC, Cardenas-Diaz FL, Kathiriya JJ, Morley MP, Carl J, et al. 2022. Human distal airways contain a multipotent secretory cell that can regenerate alveoli. *Nature* 604:120–26
92. Valenzi E, Tabib T, Papazoglou A, Sembrat J, Trejo Bittar HE, et al. 2021. Disparate interferon signaling and shared aberrant basaloid cells in single-cell profiling of idiopathic pulmonary fibrosis and systemic sclerosis-associated interstitial lung disease. *Front. Immunol.* 12:595811
93. DePianto DJ, Heiden JAV, Morshead KB, Sun KH, Modrusan Z, et al. 2021. Molecular mapping of interstitial lung disease reveals a phenotypically distinct senescent basal epithelial cell population. *JCI Insight* 6:e143626
94. Wang J, Zhang L, Luo L, He P, Xiong A, et al. 2022. Characterizing cellular heterogeneity in fibrotic hypersensitivity pneumonitis by single-cell transcriptional analysis. *Cell Death Discov.* 8:38
95. Rao W, Wang S, Duleba M, Niroula S, Goller K, et al. 2020. Regenerative metaplastic clones in COPD lung drive inflammation and fibrosis. *Cell* 181:848–64.e18
96. Bharat A, Querrey M, Markov NS, Kim S, Kurihara C, et al. 2021. Lung transplantation for patients with severe COVID-19. *Sci. Transl. Med.* 12:eabe4282
97. Delorey TM, Ziegler CGK, Heimberg G, Normand R, Yang Y, et al. 2021. COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature* 595:107–13
98. Huang KY, Petretto E. 2021. Cross-species integration of single-cell RNA-seq resolved alveolar-epithelial transitional states in idiopathic pulmonary fibrosis. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 321:L491–506
99. Kobayashi Y, Tata A, Konkimalla A, Katsura H, Lee RF, et al. 2020. Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat. Cell Biol.* 22:934–46
100. Strunz M, Simon LM, Ansari M, Kathiriya JJ, Angelidis I, et al. 2020. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat. Commun.* 11:3559
101. Reyfman PA, Walter JM, Joshi N, Anekalla KR, McQuattie-Pimentel AC, et al. 2019. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 199:1517–36
102. Morse C, Tabib T, Sembrat J, Buschur KL, Bittar HT, et al. 2019. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur. Respir. J.* 54:1802441
103. Fastres A, Pirottin D, Fievez L, Tutunaru AC, Bolen G, et al. 2020. Identification of pro-fibrotic macrophage populations by single-cell transcriptomic analysis in West Highland white terriers affected with canine idiopathic pulmonary fibrosis. *Front. Immunol.* 11:611749

104. Aran D, Looney AP, Liu L, Wu E, Fong V, et al. 2019. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* 20:163–72
105. Misharin AV, Morales-Nebreda L, Reyfman PA, Cuda CM, Walter JM, et al. 2017. Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J. Exp. Med.* 214:2387–404
106. McQuattie-Pimentel AC, Ren Z, Joshi N, Watanabe S, Stoeger T, et al. 2021. The lung microenvironment shapes a dysfunctional response of alveolar macrophages in aging. *J. Clin. Investig.* 131:e140299
107. Liao M, Liu Y, Yuan J, Wen Y, Xu G, et al. 2020. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* 26:842–44
108. Wendisch D, Dietrich O, Mari T, von Stillfried S, Ibarra IL, et al. 2021. SARS-CoV-2 infection triggers profibrotic macrophage responses and lung fibrosis. *Cell* 184:6243–61.e27
109. Lee JS, Koh JY, Yi K, Kim YI, Park SJ, et al. 2021. Single-cell transcriptome of bronchoalveolar lavage fluid reveals sequential change of macrophages during SARS-CoV-2 infection in ferrets. *Nat. Commun.* 12:4567
110. Wauters E, Van Mol P, Garg AD, Jansen S, Van Herck Y, et al. 2021. Discriminating mild from critical COVID-19 by innate and adaptive immune single-cell profiling of bronchoalveolar lavages. *Cell Res.* 31:272–90
111. Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, et al. 2020. COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* 38:970–79
112. Zhang F, Mears JR, Shakib L, Beynor JI, Shanaj S, et al. 2021. IFN-γ and TNF-α drive a *CXCL10+ CCL2+* macrophage phenotype expanded in severe COVID-19 lungs and inflammatory diseases with tissue inflammation. *Genome Med.* 13:64
113. Justet A, Zhao AY, Kaminski N. 2022. From COVID to fibrosis: lessons from single-cell analyses of the human lung. *Hum. Genom.* 16:20
114. Neumark N, Cosme CC Jr., Rose K-A, Kaminski N. 2020. The idiopathic pulmonary fibrosis cell atlas. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 319:L887–92
115. Schupp JC, Khanal S, Gomez JL, Sauler M, Adams TS, et al. 2020. Single-cell transcriptional archetypes of airway inflammation in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* 202:1419–29
116. Hey J. 2001. The mind of the species problem. *Trends Ecol. Evol.* 16:326–29
117. Crapo JD, Barry BE, Gehr P, Bachofen M, Weibel ER. 1982. Cell number and cell characteristics of the normal human lung. *Am. Rev. Respir. Dis.* 126:332–37
118. Luecken MD, Theis FJ. 2019. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* 15:e8746
119. Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, et al. 2020. Eleven grand challenges in single-cell data science. *Genome Biol.* 21:31