The Human Lung Cell Atlas

A high-resolution reference map of the human lung in health and disease

Herbert B. Schiller[1], Daniel T. Montoro[2], Lukas M. Simon[3], Emma L. Rawlins[4], Kerstin B. Meyer[5], Maximilian Strunz[1], Felipe Vieira Braga[5], Wim Timens[6.7], Gerard H. Koppelman[7.8], GR Scott Budinger[9], Janette K. Burgess[6,7], Maarten van den Berge[7,10], Fabian J. Theis[3,11], Aviv Regev[12,13], Naftali Kaminski[14], Jayaraj Rajagopal[2], Sarah A. Teichmann[5], Alexander V. Misharin[9,*], Martijn C. Nawijn[6,7,*]

1 Helmholtz Zentrum München, Institute of Lung Biology and Disease, Group Systems Medicine of Chronic Lung Disease, Member of the German Center for Lung Research (DZL), Munich, Germany.

2 Harvard Stem Cell Institute, Cambridge, MA, USA, and Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA, USA.

3 Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Germany

4 Wellcome Trust/CRUK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QN, UK

5 Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK.

6 University of Groningen, University Medical Center Groningen, Department of Pathology and Medical Biology, Groningen, the Netherlands

7 GRIAC research institute at the University Medical Center Groningen, Groningen, the Netherlands

8 University of Groningen, University Medical Center Groningen, Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, Groningen, the Netherlands

9 Division of Pulmonary and Critical Care Medicine, Northwestern University, Chicago, IL, USA

10 University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, the Netherlands

11 Department of Mathematics, Technische Universität München, Munich, Germany

12 Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA

13 Howard Hughes Medical Institute and Koch Institute for Integrative Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA.

14 Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, CT, USA

* These authors share senior authorship

Abstract

Lung disease accounts for every sixth death globally. Profiling the molecular state of all lung cell types in

health and disease is currently revolutionizing the identification of disease mechanisms and will aid the design of novel diagnostic and personalized therapeutic regimens. Recent progress in high-throughput techniques for single-cell genomic and transcriptomic analyses has opened up new possibilities to study individual cells within a tissue, classify these into cell types, and characterize variations in their molecular profiles as a function of genetics, environment, cell-cell interactions, developmental processes, ageing or disease. Integration of these cell state definitions with spatial information allows the in-depth molecular description of cellular neighborhoods and tissue microenvironments including the tissue resident structural and immune cells, the tissue matrix as well as the microbiome. The Human Cell Atlas consortium aims to characterize all cells in the healthy human body, and has prioritized lung tissue as one of the flagship projects. Here, we present the rationale, the approach and the expected impact of a Human Lung Cell Atlas.

Keywords
Human Cell Atlas, single-cell RNA sequencing, spatial transcriptomics, systems biology

Rationale for a Human Lung Cell Atlas

Lung disease is a leading cause of mortality in the US and worldwide, with more than 7 million deaths attributed to lung disease annually (1). Although the lung has been reported to harbor at least 40 discrete cell types (2), the recent identification of the ionocyte as a novel epithelial cell type in human airway wall (See Boxed item 1) shows that our knowledge of the cells in human lung is incomplete (3, 4). Therefore, further studies to systematically and comprehensively characterize all cell types in human lung tissue, and the changes in cellular composition and function associated with initiation, progression and resolution of human lung disease are urgently needed.

The cell is the fundamental unit of all living organisms (5). The cumulative function of all cells and their interactions with each other and with non-cellular tissue components determines the physiology of any tissue or organ. Disease is believed to be a consequence of cell intrinsic changes, changes in response to environmental insults, altered cell-cell communication, disbalance of cell type proportions and/or perturbed architecture of tissues. Recent progress in single-cell analyses and accompanying computational methodologies, provides an extraordinary opportunity to characterize individual cells and their spatial organization on the basis of their RNA or protein expression profile, allowing *de novo* classification of cell types and a comprehensive description of their dynamic phenotype. Combining these with rapidly evolving spatially resolving methods for the analysis of DNA, RNA or protein profiles allows the interpretation or annotation of the cellular heterogeneity in a tissue context.

The Human Cell Atlas (HCA) Consortium is an international, collaborative effort aiming to define all human cell types in terms of their distinctive patterns of gene expression, physiological states, developmental trajectories, and spatial relationships in tissue (6). As part of this overarching effort, we intend to build a comprehensive atlas of human lung in a staged and integrated approach. Molecular profiles of single cells will be mapped computationally on a common coordinate framework of the entire organ in relation to spatial landmarks of both micro- and macro-anatomy, starting from the histological structure of cell neighborhoods, building up layers of increasingly larger units of spatial organization, up to the full organ.

The Lung Cell Atlas will reveal unprecedented insights about the identities, activities and lineage relationships of all cells in healthy human lung, enabling the modeling of lung homeostatic circuitry (6). The Lung Cell Atlas will then serve as a reference point for the analysis of diseased lung tissue at single-cell resolution and will allow identification of the shifts in cellular repertoire, the changes in cellular states and phenotypes, and the altered cell-cell interactions that disrupt normal lung homeostasis and constitute disease. Comparing lung disease data with the lung development reference within the Human Lung Cell Atlas, may reveal whether cellular changes in certain chronic lung diseases represent a failed recapitulation of organogenesis or the acquisition of entirely new pathologic programs governing cell fate and behavior. Thus, the generation of a high resolution, comprehensive catalogue of the changes in lung cellular composition and function in health and disease is expected to lead to development of novel cell and disease specific biomarkers and advancements in therapeutic strategies for lung disease.

Cellular and Anatomical Complexity of the Human Lung

The primary function of lung tissue - gas exchange between the body and its environment - is dependent on a specialized anatomy involving numerous distinct epithelial and endothelial cell populations, supported by specific tissue-resident leukocyte populations, and various mesenchymal cell types providing structural support. The unique anatomy of the lung is established during development by branching morphogenesis resulting in a tree of airways and alveoli, mirrored with trees of blood and lymphatic vessels. Air flows from the nasal cavity into the conducting airways, involving trachea and the main bronchi, which iteratively bifurcate into the branching bronchial tree. The human trachea and airways are composed of a pseudostratified epithelium on the luminal side, with a basal lamina lined with mesenchyme, cartilage, smooth muscle, blood vessels, immune cells and other rare cell types (7). The airways end in terminal bronchioles, which then lead into composite respiratory units including the respiratory bronchioles, the alveolar duct and the alveolar sac, where gas exchange occurs (7). The airways serve as a conduit for gases from the atmosphere to the alveoli, but also have a role in filtering the inhaled air for solid particles and removing these in a process called mucociliary clearance, thereby protecting the distal saccular structures of the alveoli. Gas exchange happens in the alveolar unit, featuring ultra-thin type-1 pneumocytes wrapped by a layer of capillaries, supported by interstitial cells such as matrix producing fibroblasts and the pulmonary surfactant layer (8). Type-2 pneumocytes produce and secrete surfactant, maintain the fluid balance of the alveolar unit and serve as local facultative progenitor cells for type-1 cells (9–11). Resident immune cells, particularly alveolar macrophages and peribronchial and perivascular “interstitial” macrophages play important roles in orchestration of the immune response and in the maintenance of lung homeostasis (12, 13). Alveolar macrophages and alveolar epithelial cells express reciprocal sets of ligands and receptors, which prevent an excessive inflammatory response to inhaled particles. Several studies in mice have demonstrated that peribronchial macrophages consist of multiple ontologically and anatomically distinct subsets (14), however, little is known about their human counterparts.

Single-cell molecular studies profiled from a comprehensive sampling of human lungs will empower the mapping of all possible cellular identities (or states) onto a framework of a high dimensional phenotypic cellular (gene and protein expression) landscape. Such a phenotypic landscape will describe cell type identity as stable intermediates within a continuum of possible states, including lineage relationships and potential transitory paths between different states. The power of his approach is elegantly shown in a recent review on the cellular and molecular pathways involved in lung organogenesis, where single-cell datasets from mouse lung development are used to identify and map cellular trajectories during the different stages of organogenesis (15). The exciting recent demonstration of genetic lineage tracing in human tissues employing mtDNA mutations will now also enable the combination of lineage inference with gene expression or chromatin state profiles in human tissues (16). Thus, the use of mtDNA mutations as natural genetic barcodes may reveal the clonal architecture in human lung health and disease. The multiple snapshots of molecular identities present along the high-dimensional gene expression landscape can also result in the molecular description of unknown cellular states, which may have functions distinct from the traditional cell type definitions currently recognized. Similarly, we will likely encounter gradients of cellular phenotypes along spatial trajectories, as these `specialized´ cellular states are driven by their microenvironment, as was recently reported for enterocytes in the mouse intestine (17).

All lung cells are contained within a regionally tailored and highly complex extracellular matrix that forms the basis of various distinct microenvironments, which is further defined by (the signals from) other cells in close proximity and potentially local variations in the microbiome. The effect of the microenvironment on the molecular phenotype of a cell has not been explored into great detail, although diseased human lung extracellular matrix, for instance, has been shown to drive a specific cellular transcriptome (18). Whether the identity of cells within the lung is determined by their specific location and function within the tissue and local microenvironment or by their developmental origin, or by a combination of both, is one of the key questions for the Human Lung Cell Atlas. Interestingly, it was recently reported that distinct developmental lineages can produce the same anatomical cell type by converging to a homogenous transcriptomic state in the final cell divisions of development, indicating that distinct developmental lineages can assume very similar molecular phenotypes as a function of their local niche (19). Location-dependent variation in the (molecular) state of cells in human lung has not been studied in detail yet, although in the mouse model, club cells have been shown to have a transcriptomic phenotype that is dependent on their position along the respiratory tract (3). The relative frequency of epithelial airway cell types varies along the respiratory tract (20). Submucosal glands are present from trachea to the small airways in human lung, but whether their composition is constant along the respiratory tract remains to be established. Similarly, it is unknown whether location within the lung affects the molecular phenotype of the cells that make up the alveolar unit. A unique feature of the lung is that it is a constantly moving organ, with dynamic changes in pressure and shape leading to presumably similar cells being exposed to very distinct mechanical cues – apical versus basal, central versus peripheral - and the cellular transcriptomic adaptations to these variables are completely unknown.

A complete atlas of the healthy human lung must sample the entire coordinate framework of the organ, both macro- and micro-anatomically, and capture the natural variation of cellular states within the healthy population (See Boxed item 2). A comprehensive map of molecular cell states and the possible transitions between them will provide mechanistic insight into developmental processes and regenerative and repair capacity in the lung and their underlying regulatory mechanisms. The Human Lung Cell Atlas will map this landscape of molecular cell states onto a framework of spatial landmarks and enable the direct comparison of samples between multiple healthy and, eventually, diseased individuals.

A Cellular Reference Map of Human Lung Development

The process of building the lung structure, and maturing the cells it contains, occurs over many weeks of embryonic development and beyond in humans and is thought to require numerous different cell-cell and cell-matrix interactions (15, 21). Many of these signaling and physical interactions occur between multiple progenitor cell types which are specific to a developmental time window and do not exist in the adult organ. The epithelium is the best-characterized part of the developing lung and at least three, apparently embryonic-specific, progenitor cell types have been detected. For example, the budding tip of the developing human lung epithelium was shown to be analogous to that of the mouse and contain a major epithelial progenitor population required during branching morphogenesis (22, 23). However, there is no evidence that such cells exist in the mature lung. Characterization of the development-specific, or possibly developmentally-enriched, progenitor cell states, and their

(developmental) trajectories in the transcriptomic landscape, within the Human Cell Atlas Project will first and foremost provide insight into the basic mechanisms of lung development. In addition, it will provide specific and critical data allowing long-standing hypotheses regarding the recapitulation of developmental mechanisms in repair and disease states to be tested. While the initial focus of the Human Lung Cell Atlas is on adult lung, it will integrate data from the Pediatric Cell Atlas (24) and initiatives focused on lung development, such as LungMAP or the recent NIH-funded DEVMAP.

Due to the progressive nature of branching morphogenesis, the developing lung epithelium contains a continuum of cells along the proximal-distal axis in progressive states of maturity from multipotent progenitor to fully mature mucocilliary, or alveolar, epithelium. This is particularly significant given the low rates of cell turn-over in the adult organ (25) and means that analysis of embryonic lung development will allow states of differentiation, particularly intermediate forms, to be analyzed transcriptionally in a way that will be very difficult in the adult lung. In addition, the developing human lung mesenchymal and leukocyte cell lineages are almost completely uncharacterized. For example, it is unknown when artery-venous fate arises in the lung endothelium, or when alveolar macrophages first become resident in the human lungs. A Developmental Lung Cell Atlas will provide baseline data describing the molecular phenotype of the cells in developing lung, facilitating their future functional characterization. It is important to note that the human lung continues to grow after birth well into puberty and beyond. While the cellular changes during this period are completely unknown, they are considered to be of critical importance to development of lung disease later in life (26). The presence of bronchial associated lymphoid tissue (BALT) is rarely observed during fetal lung development and infrequently in newborns and infants in the absence of infection (27), has been reported to increase with age during the first few years of life (28, 29), and absent in healthy adults after the age of 20 (30) except in the context of specific inflammatory conditions (31). This illustrates the intimate interactions between a developing immune system, environmental challenges and a maturing lung during early life, further emphasizing the need for inclusion of neonatal and pediatric studies in the Human Lung Cell Atlas.

The Human Lung Cell Atlas of healthy ageing and disease

The ageing process manifests itself in age-related changes in body composition, the endocrine system and vasculature, which are initially subclinical (32). The lung grows progressively until early adulthood, and lung function peaks between 18 and 25 years of age. After this peak, lung function declines, associated with structural remodeling of the lung, progressive loss of alveolar surface area and enlargement of alveolar size (33, 34). In addition, mucociliary clearance is reduced with age in both upper and lower airways (35, 36). At the cellular level, hallmarks of ageing include genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing mitochondrial dysfunction, stem cell exhaustion, cellular senescence and altered cell-cell communication (37), which is reflected in (bulk) gene expression patterns from lung (38). Cellular senescence contributes to alterations in extracellular matrix and is associated with physiological remodeling of the ageing lung (39), and mitochondrial dysfunction has been observed in age-related lung disorders, including COPD and IPF (40). Integrating deep tissue proteomics and single cell transcriptomics, a first draft of an aging mouse lung atlas reported increased transcriptional noise, indicating loss of epigenetic control (41). Furthermore, age associated changes in composition of the

extracellular matrix, ciliated cell frequency and a pro-inflammatory microenvironment were evident in both proteome and transcriptome (41).

To systematically and comprehensively analyze the effect of age on human lung cells, tissue sources used for the Human Lung Cell Atlas should therefore be collected over the entire adult age range from 20 – 80 years old. To distinguish natural variation from sub-clinical disease and to incorporate a history of relevant environmental and infectious exposures we will need to obtain detailed clinical data, and potentially need to use biological characteristics of aging such telomere length (42) and the epigenetic clock (43) to calibrate the findings and understand natural variation and boundaries of health and disease. The promise of single cell molecular profiling is to deliver unprecedented insight into the cellular origins of human disease. The human genome project and decades of genomic research have enabled investigations into the association of gene polymorphisms with disease. Using single cell transcriptomics we can now move from genetic and genomic associations to more mechanistic questions about disease and cell type associated gene networks and reveal the cellular context of disease associated polymorphisms (44).

The approach of the Human Lung Cell Atlas: Toolbox and Resources

Single cell transcriptomics

Single cell RNA sequencing (scRNA-seq) aims to comprehensively capture and sequence the RNA contents of a single cell, allowing a description of transcriptional heterogeneity in cell populations. Stable transcriptional phenotypes can be contrasted with differentiation trajectories of cells by inferring an ordered sequence of transcriptional states in pseudotime. The key to the utility of scRNA-seq in tissue profiling has been the massive increase from 10's to over 100,000 cells per experiment in less than a decade (45, 46), due to technological advances as recently reviewed (47). Comprehensive profiling of a tissue under specific (developmental, diseased) conditions requires a high throughput, capturing all cell types and states - including putative novel ones - at a threshold of sensitivity that is determined by both cell input and the observed variation in transcriptional profiles between discrete cell states. Rare cell types with a transcriptional phenotype that is not sufficiently discrete can be missed in unsupervised clustering, as exemplified by the neuroendocrine cells in the recent cell census of airway wall tissue (48). Although enrichment or purification of these cell types prior to scRNA-Seq analysis can be applied, this does require prior knowledge of their identity. Atlasing studies therefore often combine a high throughput approach for cell type and cell marker identification with in depth profiling of purified subpopulations of special interest. To assess completeness of the single cell sampling *in silico* bulk profiles from scRNA-seq can be compared to true bulk RNA-seq data generated from adjacent regions of tissue. Another attractive alternative is the use of new high-density spatial transcriptomic tools, such as Slide-seq or HDST (49, 50). These spatial tools can identify the microanatomical niches that are systematically undersampled and guide efforts for targeted enrichment.

Currently, sequencing of single cells from freshly dissociated tissue yields the greatest number of genes identified per cell. However, this comes at the cost of introducing potential dissociation biases, such as a stress response induced by enzymatic tissue digestion (51) or a bias in cell type composition as cell types differ in their optimal conditions for release from the tissue (52, 53). An alternative technique is the

isolation of nuclei from snap-frozen tissue, which are then assayed by single nuclear RNAseq protocols (snRNA-seq). In general, snRNA-seq can be performed on cryostat sections of lung tissue, with the advantage of improved tissue processing with less biased isolation, and enabling profiling of adjacent tissue or serial sections using targeted imaging techniques in order to put the snRNA-seq data into spatial context (49, 50, 54–57). While single nucleus RNA-seq detects less genes and unique transcripts, than scRNA-seq it allows the use of archived samples (58–61). On the whole, scRNAseq and snRNAseq data correlate well (62) and allow robust cell type identification as well as the derivation of developmental trajectories for the different cell types present in a tissue.

Beyond single cell transcriptomics

While much of single cell biology has focused on the transcriptome, epigenetic profiling of single cells can provide information on a cells' past history and potential future states. Of the various epigenetic profiling methods available, ATAC (Assay for Transposase-Accessible Chromatin)-seq uses transposase tagging to identify regions of open chromatin and regulatory elements within the genome. Recently, single cell ATAC-Seq has become available with the development of upfront bulk tagmentation, followed by DNA fragmentation at the single cell level (63, 64). scATAC-seq has been shown to provide mechanistic insight into the regulatory mechanisms underlying cell states and their transitions (65). Differentiation trajectories of blood cells detected with scATAC-Seq were similar to those from RNA but with earlier branching points, reflecting that the chromatin changes detected by ATAC accessibility precedes expression at the RNA level (66). Thus, scATACseq and scRNA-Seq will provide complementary insights into lung cell states and the trajectories between them. As the cost of high throughput scATAC protocols continues to reduce (67), we expect that this approach will contribute greatly to our understanding of lung development, cell type specification and abnormality in disease.

Additional single-cell 'omics' approaches, such as genome, methylome, proteome, metabolome, are becoming available to complement transcriptomic measurements (68). The true size and shape of the cellular phenotypic landscape of any organ will only be revealed in a multi-omic description of the cells and the non-cellular components. Several multi-omics single-cell approaches have been explored, including integrated single-cell measurement of chromatin accessibility, DNA methylation and transcriptome (69), integrated genetic, epigenetic and transcriptomic measurement (70), and recently integrated scRNA-seq with genotype data to discover cell type-specific expression quantitative trait loci (eQTLs) (71). Recent progress in single cell profiling technologies has made it possible to measure protein and RNA levels within the same cell simultaneously (72–74), which enhances discovery of subtle cellular phenotypes, such as cell type subpopulations (75).

Imaging and spatial transciptomics

Despite the advance of single cell profiling, the function of individual cells within a tissue ultimately depends on their location within the tissue architecture, and their interactions with the extracellular matrix and neighboring cells. Considering the morphological diversity of the lung, characterization of the exact location and local microenvironment where cells are obtained from is critically important. Techniques for measuring gene expression within lung tissue context range from spatial transcriptomics, microdissection approaches, such as laser-capture microdissection, and *in situ* RNA hybridization to *in situ* RNA sequencing. The main limitations on spatial mapping of cell states defined by high dimensional

gene expression patterns, however, are the number of genes detected on the same tissue sample, and the shift from targeted to unbiased detection of gene expression. Spatial transcriptomic approaches measure gene expression in small voxels of tissue, with recent HDST and Slide-Seq technologies bringing the resolution down to around 1-10 μm (49, 50). A single-molecule fluorescent *in situ* hybridization (smFISH)-based protocol to achieve high throughput is the multiplexed error-robust FISH (or MERFISH) technology (76), although application to the lung tissue might require further optimization (77). Another protocol suitable for multiplexed RNA species detection is the proximity ligation and in situ hybridization (PLISH) technology (78). Interestingly, PLISH is compatible with immunohistochemical applications and the usage of antibodies on archived (FFPE) tissue. Similarly, detection of proteins and their posttranslational modification with antibodies in combination with mRNA transcripts in tissue sections was also achieved using a mass cytometric approach (79). Finally, *in situ* RNA sequencing using padlock probes (77, 80) or FISSEQ (fluorescent *in situ* RNA sequencing) technology (81) is of interest. The recently developed STARmap method combined *in situ* amplification of mRNA transcripts with hydrogel tissue chemistry, allowing detection of around a thousand genes in millimeter scale tissue blocks after clearing of cells and lipids, whilst retaining relative location within the tissue blocks' coordinate system (82). Integration of such methods with high-resolution transcriptomic single-cell profiling of matched tissue will likely be key to establish a true atlas of the cellular landscape of healthy or diseased human lung. A final technological challenge is to integrate proteomic data and transcriptomic data from single cells, which requires development of proteomic methods with single cell sensitivity. Currently, a combination of antibody labeling technologies or multi-omics approaches such as combining CyTOF with spatially resolving FISH methodologies remains the approach of choice.

Sources of lung tissue

Lung tissue for the Human Lung Cell Atlas is typically received from (routine) surgical resections, biopsy extraction in routine clinical care or in a research setting when involving healthy volunteers or lung transplantation programs. In this context, it important to highlight the difficulty of obtaining normal lung parenchyma. The most common sources for lung parenchyma are either uninvolved edges of lung resections (mostly for cancer) or tissue obtained from deceased donors through organ procurement organizations which all can be archived or processed fresh. Common sources of archived lung tissue are large tissue banks and repositories that have been established in many centers, mostly associated with academic, medical, or governmental organizations (83). Biobanks or core tissue facilities take care of patient information, obtain informed written consent for tissue donations and biopsy collections, and process and store tissue in an appropriate manner. They organize and monitor tissue distribution for scientific research and store relevant clinical information with respect to preserving patient anonymity, and provide centralized data storage. Specimens are collected according to site-specific protocols and legal regulations. To minimize tissue collection variability between different centers, common standard principles for tissue processing and collection are being developed. Live tissue from bronchoscopic biopsies, organ procurement organizations or surgical resection needs special attention after collection. These specimens need to be transported immediately after acquisition to the research site for preparation of single cell suspensions and downstream processing. Protocols for acquisition, transport and processing of fresh tissue samples are harmonized between the HCA Lung working group members, and available through online platforms (such as: https://www.protocols.io/groups/hca/publications).

Computational challenges

The number of observations in a typical single cell experiment is generally much larger compared to bulk RNA-seq data. Therefore, the underlying gene expression matrices, often encoded with genes as rows and cells as columns, pose novel and complex computational challenges due to their size. Moreover, so called dropout events, in which an expressed gene is simply not detected due to lack of sensitivity, introduce high levels of sparsity into the data. Thirdly, the increased resolution enables more precise capture of biological heterogeneity which in turn leads to increased complexity in the data. Altogether scalability, sparsity and complexity in single cell data pose major challenges when analyzing single cell expression profiles.

Some popular scRNA-seq analysis frameworks such as Scanpy (84) are designed to provide high scalability. Moreover, dimension reduction applications in molecular biology have evolved with the increased size and complexity of single cell data. Linear dimension reduction by principal component analysis is commonly used for the visualization of bulk and initial scRNA-seq data sets. More recent approaches, using nonlinear two dimensional embeddings such as t-distributed stochastic neighbour embedding (t-SNE) and uniform manifold approximation and projection (UMAP), have helped visualize single cell data, and have recently been discussed with respect to quality as well as scalability (85, 86). These nonlinear approaches are particularly suitable for highly complex tissues, such as the lung, where a large number of different cell types exist (41, 53).

The increased number of cells profiled in a single experiment represents a computational challenge but also an opportunity for the application of more sophisticated modeling approaches, including machine learning (46). Deep learning, a subfield of machine learning which has for instance revolutionized image analysis, is particularly powerful when large training data sets exist (87). Some of the recent, larger data sets now open up the possibility to apply deep learning methods to scRNA-seq data. First deep learning algorithms have been developed and tailored towards scRNA-seq data with, for example, the goal of denoising scRNA-seq expression data (88). Machine learning approaches could be very useful for the detection of robust cell type-specific signatures and the integration of datasets generated on different platforms, where sequencing depth and transcript capture bias can vary dramatically depending on the protocol and experimental setup. As more data becomes available and datasets grow in size we anticipate to see more machine learning applications for the analysis of scRNA-seq data.

Unbiased clustering approaches can be applied not only to scRNA-seq or scATAC-seq data, but also to spatial techniques to uncover interactions between the cell types forming lung micro-niches, as was recently demonstrated for epithelial and mesenchymal cells (10, 78, 89). These tools will be particularly helpful for integration of high throughput sequencing and spatial datasets. In combination with neural network based approaches we expect that additional data features such as morphometry - which can for example be used to robustly predict cell cycle stage (90) - will be integrated into spatial analysis to learn about combined cell state and communication.

Another promising computational approach is the statistical deconvolution of bulk RNA-seq data using scRNA-seq data to infer cell type proportions of a complex tissue sample (measured with bulk RNA-Seq) by leveraging information from scRNA-seq data (91). It was recently shown in pancreatic datasets that celltype heterogeneity represents a major covariate in bulk differential expression analysis (92). Lung tissue biopsies can show very high celltype heterogeneity depending on the precise location of surgery (93). Therefore, publicly available lung bulk data from patients and controls is a particularly promising candidate for re-analysis using deconvolution techniques.

Opportunities to build a Human Lung Cell Atlas

The opportunity to build a high resolution single cell and spatial atlas of the human body has generated enormous excitement in the scientific community. The global initiative to build the Human Cell Atlas, launched in 2016, now involves scientists from 62 countries, a number likely to increase. The HCA community is supported by many funding organizations and initiatives, and organized into biological networks, including the Human Lung Cell Atlas. For example, in the specific area of the lung, the NIH has actively supported these efforts through programs including LungMAP (94) in the healthy lung, and disease-focused programs, in particular the Cancer Moonshot Human Tumor Atlas Network (http://bit.ly/NCI-HumanTumorAtlas). Moreover, recognizing these emerging initiatives, and the importance of supporting this ambitious new vision, the NIH Common Fund recently launched the Human BioMolecular Atlas Program (HuBMAP), which includes the lung among its organs of focus (https://commonfund.nih.gov/hubmap). Other relevant efforts are supported by the MRC (http://bit.ly/HCALungMRC) and Wellcome Trust (http://bit.ly/HCALungW) in the UK, and by the Chan Zuckerberg Initiative (http://bit.ly/HCALungCZI1). The Human Lung Cell Atlas will be built by integration of the datasets provided by these and other initiatives. Working within the HCA thus ensures integration and international coordination across multiple funded efforts. Indeed, the HCA design is inherently inclusive and open, allowing parallel initiatives to contribute to the Human Cell Atlas, leveraging expertise, data and technological developments between projects, organ systems, institutes and countries. Consequently, the infrastructure that is under development within the HCA consortium will have a scope and reach beyond any of the organ-specific atlases, making the HCA well-suited for a truly community-wide effort towards establishing a Human Cell Atlas.

Next steps towards establishing a Human Lung Cell Atlas

The first cell censi of upper airways and lung tissue have been published (48, 53, 95, 96), and more will follow with better coverage of regional differences and rare cell types, as evident from the progress meetings at the HCA meetings (http://bit.ly/HCALungTV1, http://bit.ly/HCALungTV2). These studies on single-cell profiling of healthy and diseased lung tissue mainly report scRNA-Seq and ATAC-Seq data (48, 53, 95, 96). With the anticipated increase in studies reporting molecular phenotypes of lung cells, the efforts of the Lung Biological Network within the HCA consortium is rapidly shifting from reporting on the progress of individual groups to coordinating shared and integrated analysis of the cumulative data within the HCA data coordination platform (DCP). The next breakthrough in establishing a Human Lung Cell Atlas will also come from a detailed spatial mapping of cell types onto the tissue architecture, the description of molecular cell states as a function of relative position in the lung along one of the many gradients present in the tissue or as a function of local cellular neighborhood, and the visualization of the cellular profiles in a spatial framework reflecting the structure of the lung. In addition, novel technological advances, especially in spatial mapping and computational approaches (e.g. towards multimodal data integration and combined visualization of profiling and spatial data) will remain an important topic for future meetings of the Human Cell Atlas consortium.

Joining the HCA Lung Biological Network

The HCA consortium is open for anyone in the respiratory community to join (https://www.humancellatlas.org/joinHCA), and open sharing of best practices and proven productive

approaches to tissue acquisition, processing, data generation and analysis as well as technological innovations will help bring the field forward. Progress will be presented at the HCA meetings, but also at the annual meetings of the American Thoracic Society (ATS) and the European Respiratory Society (ERS). Lung has been identified as one of the 12 priority organs and systems within the HCA initiative, with an initial focus on upper and lower airways as well as lung parenchyma (6). To achieve our ambition of establishing the Human Lung Cell Atlas, we call on widespread involvement of the respiratory community to contribute to this effort. This will allow the HCA-Lung Biological Network to achieve the depth and breadth necessary to obtain a meaningful reference for basic, translational and clinical respiratory scientists and which may serve as an example for other, less readily accessible organ systems within the HCA. To facilitate integration of datasets acquired by different research groups at geographically distinct sites, in multiple ethnicities, age groups, and genders, we urgently need standardization in tissue procurement, processing, and storage, as well as in data generation, handling, storage and analysis pipelines. For tissue acquisition, we have developed a series of standardized protocols specifically tested and validated for lung tissue, and continue to expand these, making them available to the community through open access platforms such as protocols.io (See https://www.protocols.io/groups/hca/publications). For processing and integration of data originating from multiple centers and divergent technological platforms, methodologies are being tested using the datasets from the initial lung scRNA-Seq efforts. The computational approaches are also shared with the community, for instance through platforms such as Github (https://github.com/HumanCellAtlas). Methods for integration of genomic, transcriptomic and proteomic datasets from the cellular branch of the Human Lung Cell Atlas with those acquired in the spatial branch will need to be further developed. For data storage and accessibility, the HCA-Lung project will tap into the HCA data coordination platform (DCP) infrastructure that is currently in beta testing. One critical aspect for the Human Lung Cell Atlas to achieve its full impact will be the development of an interactive, tertiary HCA data portals (6), that allow exploration of the data by the respiratory and single-cell community. At this moment, interim data portals allow visualization and limited exploration of the currently available datasets (see for instance http://bit.ly/LungMap-scRNASeq (94); https://lungcellatlas.org (48); https://theislab.github.io/LungAgingAtlas (41), and https://www.nupulmonary.org/resources/ (53)). Incorporation of detailed demographic and clinical metadata from each tissue source in the HCA data coordination portal will be critical to allow the identification of individual and combined contributions of gender, ethnicity, age, respiratory health and disease to the position of each cell type in the transcriptomic and proteomic landscape that defines the Human Lung Cell Atlas.

The Human Lung cell atlas holds the promise of becoming a resource of unsurpassed scale and detail, allowing the respiratory community to acquire novel insights into the biology of the lung, including - but not restricted to - its development, anatomy, and physiology in health, and the changes thereof in lung disease.

Acknowledgements

The authors are grateful to Dana Pe'er and Jennifer Rood for discussions on the paper outline and feedback on the manuscript. HBS acknowledges funding by the Helmholtz Association and the German Center for Lung Research (DZL). LMS acknowledges funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 753039. ELR is supported by the UK Medical Research Council (G0900424).GRSB is supported by NIH grants ES013995, HL071643, AG049665, The Veterans Administration Grant BX000201 and Department of Defense grant PR141319. AVM is supported by NIH grants HL135124 and AI135964 and Department of Defense grant PR141319. NK is funded by NIH grants RO1HL141852, RO1HL127349, UHHL3123886, UO1HL122626, UG3TR002445 and Department of Defence grant PR151124. MCN acknowledges funding from the Netherlands Lung Foundation (Grant no. 5.1.14.020 and 4.1.18.226).

References

1. Gibson GJ, Loddenkemper R, Lundbäck B, Sibille Y. Respiratory health and disease in Europe: the new European Lung White Book. *Eur Respir J* 2013;42:559–563.
2. Franks TJ, Colby TV, Travis WD, Tuder RM, Reynolds HY, Brody AR, Cardoso WV, Crystal RG, Drake CJ, Engelhardt J, Frid M, Herzog E, Mason R, Phan SH, Randell SH, Rose MC, Stevens T, Serge J, Sunday ME, Voynow JA, Weinstein BM, Whitsett J, Williams MC. Resident cellular components of the human lung: current knowledge and goals for research on cell phenotyping and function. *Proc Am Thorac Soc* 2008;5:763–766.
3. Montoro DT, Haber AL, Biton M, Vinarsky V, Lin B, Birket SE, Yuan F, Chen S, Leung HM, Villoria J, Rogel N, Burgin G, Tsankov AM, Waghray A, Slyper M, Waldman J, Nguyen L, Dionne D, Rozenblatt-Rosen O, Tata PR, Mou H, Shivaraju M, Bihler H, Mense M, Tearney GJ, Rowe SM, Engelhardt JF, Regev A, Rajagopal J. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* 2018;560:319–324.
4. Plasschaert LW, Žilionis R, Choo-Wing R, Savova V, Knehr J, Roma G, Klein AM, Jaffe AB. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* 2018;560:377–381.
5. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M, Clevers H, Deplancke B, Dunham I, Eberwine J, Eils R, Enard W, Farmer A, Fugger L, Göttgens B, Hacohen N, Haniffa M, Hemberg M, Kim S, Klenerman P, Kriegstein A, Lein E, Linnarsson S, Lundberg E, Lundeberg J, *et al.* The Human Cell Atlas. *Elife* 2017;6.:
6. Regev A, Teichmann S, Rozenblatt-Rosen O, Stubbington M, Ardlie K, Amit I, Arlotta P, Bader G, Benoist C, Biton M, Bodenmiller B, Bruneau B, Campbell P, Carmichael M, Carninci P, Castelo-Soccio L, Clatworthy M, Clevers H, Conrad C, Eils R, Freeman J, Fugger L, Goettgens B, Graham D, Greka A, Hacohen N, Haniffa M, Helbig I, Heuckeroth R, *et al.* The Human Cell Atlas White Paper. *arXiv*

[q-bioTO] 2018;at <http://arxiv.org/abs/1810.05192>.

7. Hogan BLM, Barkauskas CE, Chapman HA, Epstein JA, Jain R, Hsia CCW, Niklason L, Calle E, Le A, Randell SH, Rock J, Snitow M, Krummel M, Stripp BR, Vu T, White ES, Whitsett JA, Morrisey EE. Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell* 2014;15:123–138.
8. Weibel ER. On the tricks alveolar epithelial cells play to make a good lung. *Am J Respir Crit Care Med* 2015;191:504–513.
9. Whitsett JA, Wert SE, Weaver TE. Diseases of pulmonary surfactant homeostasis. *Annu Rev Pathol* 2015;10:371–393.
10. Nabhan AN, Brownfield DG, Harbury PB, Krasnow MA, Desai TJ. Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* 2018;359:1118–1123.
11. Zacharias WJ, Frank DB, Zepp JA, Morley MP, Alkhaleel FA, Kong J, Zhou S, Cantu E, Morrisey EE. Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature* 2018;555:251–255.
12. Lim HY, Lim SY, Tan CK, Thiam CH, Goh CC, Carbajo D, Chew SHS, See P, Chakarov S, Wang XN, Lim LH, Johnson LA, Lum J, Fong CY, Bongso A, Biswas A, Goh C, Evrard M, Yeo KP, Basu R, Wang JK, Tan Y, Jain R, Tikoo S, Choong C, Weninger W, Poidinger M, Stanley RE, Collin M, *et al.* Hyaluronan Receptor LYVE-1-Expressing Macrophages Maintain Arterial Tone through Hyaluronan-Mediated Regulation of Smooth Muscle Cell Collagen. *Immunity* 2018;49:326–341.e7.
13. Gibbings SL, Thomas SM, Atif SM, McCubbrey AL, Desch AN, Danhorn T, Leach SM, Bratton DL, Henson PM, Janssen WJ, Jakubzick CV. Three Unique Interstitial Macrophages in the Murine Lung at Steady State. *Am J Respir Cell Mol Biol* 2017;57:66–76.
14. Tan SYS, Krasnow MA. Developmental origin of lung macrophage diversity. *Development* 2016;143:1318–1327.

15. Whitsett JA, Kalin TV, Xu Y, Kalinichenko VV. Building and Regenerating the Lung Cell by Cell. *Physiol Rev* 2019;99:513–554.
16. Ludwig LS, Lareau CA, Ulirsch JC, Christian E, Muus C, Li LH, Pelka K, Ge W, Oren Y, Brack A, Law T, Rodman C, Chen JH, Boland GM, Hacohen N, Rozenblatt-Rosen O, Aryee MJ, Buenrostro JD, Regev A, Sankaran VG. Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. *Cell* 2019;176:1325–1339.e22.
17. Moor AE, Harnik Y, Ben-Moshe S, Massasa EE, Rozenberg M, Eilam R, Bahar Halpern K, Itzkovitz S. Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis. *Cell* 2018;175:1156–1167.e15.
18. Parker MW, Rossi D, Peterson M, Smith K, Sikström K, White ES, Connett JE, Henke CA, Larsson O, Bitterman PB. Fibrotic extracellular matrix activates a profibrotic positive feedback loop. *J Clin Invest* 2014;124:1622–1635.
19. Packer JS, Zhu Q, Huynh C, Sivaramakrishnan P. A lineage-resolved molecular atlas of C. elegans embryogenesis at single cell resolution. *BioRxiv* 2019;at <https://www.biorxiv.org/content/10.1101/565549v2.abstract>.
20. Tata PR, Rajagopal J. Plasticity in the lung: making and breaking cell identity. *Development* 2017;144:755–766.
21. Nikolić MZ, Sun D, Rawlins EL. Human lung development: recent progress and new challenges. *Development* 2018;145.:
22. Nikolić MZ, Johnson J-A, Sun D, Caritg O, Laresgoiti U, Brady J, Allen G, Giangreco A, Rawlins EL. Development of a genetically modifiable epithelial in-vitro culture system from human embryonic lung epithelial stem cells: towards human lung regeneration in end-stage respiratory failure. *Lancet* 2017;389:S74.
23. Miller AJ, Hill DR, Nagy MS, Aoki Y, Dye BR, Chin AM, Huang S, Zhu F, White ES, Lama V, Spence JR.

In Vitro Induction and In Vivo Engraftment of Lung Bud Tip Progenitor Cells Derived from Human Pluripotent Stem Cells. *Stem Cell Reports* 2018;10:101–119.

24. Taylor DM, Aronow BJ, Tan K, Bernt K, Salomonis N, Greene CS, Frolova A, Henrickson SE, Wells A, Pei L, Jaiswal JK, Whitsett J, Hamilton KE, MacParland SA, Kelsen J, Heuckeroth RO, Potter SS, Vella LA, Terry NA, Ghanem LR, Kennedy BC, Helbig I, Sullivan KE, Castelo-Soccio L, Kreigstein A, Herse F, Nawijn MC, Koppelman GH, Haendel M, *et al.* The Pediatric Cell Atlas: Defining the Growth Phase of Human Development at Single-Cell Resolution. *Dev Cell* 2019;doi:10.1016/j.devcel.2019.03.001.
25. Teixeira VH, Nadarajan P, Graham TA, Pipinikas CP, Brown JM, Falzon M, Nye E, Poulsom R, Lawrence D, Wright NA, McDonald S, Giangreco A, Simons BD, Janes SM. Stochastic homeostasis in human airway epithelium is achieved by neutral competition of basal cell progenitors. *Elife* 2013;2:e00966.
26. Bush A. Lung Development and Aging. *Ann Am Thorac Soc* 2016;13 Suppl 5:S438–S446.
27. Gould SJ, Isaacson PG. Bronchus-associated lymphoid tissue (BALT) in human fetal and infant lung. *J Pathol* 1993;169:229–234.
28. Tschernig T, Kleemann WJ, Pabst R. Bronchus-associated lymphoid tissue (BALT) in the lungs of children who had died from sudden infant death syndrome and other causes. *Thorax* 1995;50:658–660.
29. Heier I, Malmström K, Sajantila A, Lohi J, Mäkelä M, Jahnsen FL. Characterisation of bronchus-associated lymphoid tissue and antigen-presenting cells in central airway mucosa of children. *Thorax* 2011;66:151–156.
30. Hiller AS, Tschernig T, Kleemann WJ, Pabst R. Bronchus-associated lymphoid tissue (BALT) and larynx-associated lymphoid tissue (LALT) are found at different frequencies in children, adolescents and adults. *Scand J Immunol* 1998;47:159–162.
31. Tschernig T, Pabst R. Bronchus-associated lymphoid tissue (BALT) is not present in the normal adult

lung but in different diseases. *Pathobiology* 2000;68:1–8.

32. Partridge L, Deelen J, Slagboom PE. Facing up to the global challenges of ageing. *Nature* 2018;561:45–56.
33. Ochs M, Nyengaard JR, Jung A, Knudsen L, Voigt M, Wahlers T, Richter J, Gundersen HJG. The number of alveoli in the human lung. *Am J Respir Crit Care Med* 2004;169:120–124.
34. Hecker L. Mechanisms and consequences of oxidative stress in lung disease: therapeutic implications for an aging populace. *Am J Physiol Lung Cell Mol Physiol* 2018;314:L642–L653.
35. Proença de Oliveira-Maul J, Barbosa de Carvalho H, Goto DM, Maia RM, Fló C, Barnabé V, Franco DR, Benabou S, Perracini MR, Jacob-Filho W, Saldiva PHN, Lorenzi-Filho G, Rubin BK, Nakagawa NK. Aging, diabetes, and hypertension are associated with decreased nasal mucociliary clearance. *Chest* 2013;143:1091–1097.
36. Svartengren M, Falk R, Philipson K. Long-term clearance from small airways decreases with age. *Eur Respir J* 2005;26:609–615.
37. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell* 2013;153:1194–1217.
38. de Vries M, Faiz A, Woldhuis RR, Postma DS, de Jong TV, Sin DD, Bossé Y, Nickle DC, Guryev V, Timens W, van den Berge M, Brandsma C-A. Lung tissue gene-expression signature for the ageing lung in COPD. *Thorax* 2017;doi:10.1136/thoraxjnl-2017-210074.
39. Calhoun C, Shivshankar P, Saker M, Sloane LB, Livi CB, Sharp ZD, Orihuela CJ, Adnot S, White ES, Richardson A, Le Saux CJ. Senescent Cells Contribute to the Physiological Remodeling of Aged Lungs. *J Gerontol A Biol Sci Med Sci* 2016;71:153–160.
40. Meiners S, Eickelberg O, Königshoff M. Hallmarks of the ageing lung. *Eur Respir J* 2015;45:807–827.
41. Angelidis I, Simon LM, Fernandez IE, Strunz M, Mayr CH, Greiffo FR, Tsitsiridis G, Ansari M, Graf E, Strom T-M, Nagendran M, Desai T, Eickelberg O, Mann M, Theis FJ, Schiller HB. An atlas of the aging

lung mapped by single cell transcriptomics and deep tissue proteomics. *Nat Commun* 2019;10:963.

42. Everaerts S, Lammertyn EJ, Martens DS, De Sadeleer LJ, Maes K, van Batenburg AA, Goldschmeding R, van Moorsel CHM, Dupont LJ, Wuyts WA, Vos R, Gayan-Ramirez G, Kaminski N, Hogg JC, Janssens W, Verleden GM, Nawrot TS, Verleden SE, McDonough JE, Vanaudenaerde BM. The aging lung: tissue telomere shortening in health and disease. *Respir Res* 2018;19:95.
43. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* 2018;19:371–384.
44. Hemberg M. Summing up the parts of the hypothalamus. *Nat Neurosci* 2017;20:378–379.
45. Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. *Nat Protoc* 2018;13:599–604.
46. Angerer P, Simon L, Tritschler S, Alexander Wolf F, Fischer D, Theis FJ. Single cells make big data: New challenges and opportunities in transcriptomics. *Current Opinion in Systems Biology* 2017;4:85–91.
47. Chen X, Teichmann SA, Meyer KB. From Tissues to Cell Types and Back: Single-Cell Gene Expression Analysis of Tissue Architecture. *Annu Rev Biomed Data Sci* 2018;1:29–51.
48. Vieira Braga FA, Kar G, Berg M, Carpaij OA, Polanski K, Simon LM, Brouwer S, Gomes T, Hesse L, Jiang J, Fasouli ES, Efremova M, Vento-Tormo R, Affleck K, Palit S, Strzelecka P, Firth HV, Mahbubani KTA, Cvejic A, Meyer KB, Saeb-Parsy K, Luinge M, Brandsma C-A, Timens W, Angelidis I, Strunz M, Koppelman GH, van Oosterhout AJ, Schiller HB, *et al.* A cellular census of healthy lung and asthmatic airway wall identifies novel cell states in health and disease. *bioRxiv* 2019;doi:10.1101/527408.
49. Sanja Vickovic, Gökcen Eraslan, Johanna Klughammer, Linnea Stenbeck, Fredrik Salmén, Tarmo Äijö, Richard Bonneau, Ludvig Bergenstråhle, Joshua Gould, Mostafa Ronaghi, Jonas Frisén, Joakim Lundeberg*, Aviv Regev and Patrik L Ståhl. High-density spatial transcriptomics arrays for in situ

tissue profiling. *bioRxiv* 2019;at <https://www.biorxiv.org/content/10.1101/563338v1.abstract>.

50. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, Welch J, Chen LM, Chen F, Macosko EZ. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 2019;363:1463–1467.
51. van den Brink SC, Sage F, Vértesy Á, Spanjaard B, Peterson-Maduro J, Baron CS, Robin C, van Oudenaarden A. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat Methods* 2017;14:935–936.
52. Cohen M, Giladi A, Gorki A-D, Solodkin DG, Zada M, Hladik A, Miklosi A, Salame T-M, Halpern KB, David E, Itzkovitz S, Harkany T, Knapp S, Amit I. Lung Single-Cell Signaling Interaction Map Reveals Basophil Role in Macrophage Imprinting. *Cell* 2018;175:1031–1044.e18.
53. Reyfman PA, Walter JM, Joshi N, Anekalla KR, McQuattie-Pimentel AC, Chiu S, Fernandez R, Akbarpour M, Chen C-I, Ren Z, Verma R, Abdala-Valencia H, Nam K, Chi M, Han S, Gonzalez-Gonzalez FJ, Soberanes S, Watanabe S, Williams KJN, Flozak AS, Nicholson TT, Morgan VK, Winter DR, Hinchcliff M, Hrusch CL, Guzy RD, Bonham CA, Sperling AI, Bag R, *et al.* Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2018;doi:10.1164/rccm.201712-2410OC.
54. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, Mollbrink A, Linnarsson S, Codeluppi S, Borg Å, Pontén F, Costea PI, Sahlén P, Mulder J, Bergmann O, Lundeberg J, Frisén J. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016;353:78–82.
55. Pinskiy V, Jones J, Tolpygo AS, Franciotti N, Weber K, Mitra PP. High-Throughput Method of Whole-Brain Sectioning, Using the Tape-Transfer Technique. *PLoS One* 2015;10:e0102363.
56. Carow B, Hauling T, Qian X, Kramnik I, Nilsson M. Spatial and temporal localization of immune transcripts defines hallmarks and diversity in the tuberculosis granuloma. *bioRxiv* 2019;at

<https://www.biorxiv.org/content/10.1101/509547v1.abstract>.

57. Ziegenhain C, Vieth B, Parekh S, Hellmann I, Enard W. Quantitative single-cell transcriptomics. *Brief Funct Genomics* 2018;17:220–232.

58. Lacar B, Linker SB, Jaeger BN, Krishnaswami SR, Barron JJ, Kelder MJE, Parylak SL, Paquola ACM, Venepally P, Novotny M, O'Connor C, Fitzpatrick C, Erwin JA, Hsu JY, Husband D, McConnell MJ, Lasken R, Gage FH. Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat Commun* 2016;7:11022.

59. Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, Choudhury SR, Aguet F, Gelfand E, Ardlie K, Weitz DA, Rozenblatt-Rosen O, Zhang F, Regev A. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* 2017;14:955–958.

60. Habib N, Li Y, Heidenreich M, Swiech L, Avraham-Davidi I, Trombetta JJ, Hession C, Zhang F, Regev A. Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* 2016;353:925–928.

61. Krishnaswami SR, Grindberg RV, Novotny M, Venepally P, Lacar B, Bhutani K, Linker SB, Pham S, Erwin JA, Miller JA, Hodge R, McCarthy JK, Kelder M, McCorrison J, Aevermann BD, Fuertes FD, Scheuermann RH, Lee J, Lein ES, Schork N, McConnell MJ, Gage FH, Lasken RS. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc* 2016;11:499–524.

62. Lake BB, Codeluppi S, Yung YC, Gao D, Chun J, Kharchenko PV, Linnarsson S, Zhang K. A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci Rep* 2017;7:6031.

63. Chen X, Miragaia RJ, Natarajan KN, Teichmann SA. A rapid and robust method for single cell chromatin accessibility profiling. *Nat Commun* 2018;9:5345.

64. Chen X, Litzenburger U, Wei Y, Schep AN, LaGory EL, Choudhry H, Giaccia AJ, Greenleaf WJ, Chang H. Joint single-cell DNA accessibility and protein epitope profiling reveals environmental regulation

of epigenomic heterogeneity. 2018;doi:10.1101/310359.

65. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 2015;523:486–490.
66. Buenrostro JD, Ryan Corces M, Lareau CA, Wu B, Schep AN, Aryee MJ, Majeti R, Chang HY, Greenleaf WJ. Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* 2018;173:1535–1548.e16.
67. Mezger A, Klemm S, Mann I, Brower K, Mir A, Bostick M, Farmer A, Fordyce P, Linnarsson S, Greenleaf W. High-throughput chromatin accessibility profiling at single-cell resolution. *Nat Commun* 2018;9:3647.
68. Colomé-Tatché M, Theis FJ. Statistical single cell multi-omics integration. *Current Opinion in Systems Biology* 2018;7:54–59.
69. Clark SJ, Argelaguet R, Kapourani C-A, Stubbs TM, Lee HJ, Alda-Catalinas C, Krueger F, Sanguinetti G, Kelsey G, Marioni JC, Stegle O, Reik W. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat Commun* 2018;9:781.
70. Hou Y, Guo H, Cao C, Li X, Hu B, Zhu P, Wu X, Wen L, Tang F, Huang Y, Peng J. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res* 2016;26:304–319.
71. van der Wijst MGP, Brugge H, de Vries DH, Deelen P, Swertz MA, LifeLines Cohort Study, BIOS Consortium, Franke L. Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat Genet* 2018;50:493–497.
72. Valdes AM, Glass D, Spector TD. Omics technologies and the study of human ageing. *Nat Rev Genet* 2013;14:601–607.
73. Darmanis S, Gallant CJ, Marinescu VD, Niklasson M, Segerman A, Flamourakis G, Fredriksson S,

Assarsson E, Lundberg M, Nelander S, Westermark B, Landegren U. Simultaneous Multiplexed Measurement of RNA and Proteins in Single Cells. *Cell Rep* 2016;14:380–389.

74. Peterson VM, Zhang KX, Kumar N, Wong J, Li L, Wilson DC, Moore R, McClanahan TK, Sadekova S, Klappenbach JA. Multiplexed quantification of proteins and transcripts in single cells. *Nat Biotechnol* 2017;35:936.
75. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R, Smibert P. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 2017;14:865–868.
76. Moffitt JR, Hao J, Wang G, Chen KH, Babcock HP, Zhuang X. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc Natl Acad Sci U S A* 2016;113:11046–11051.
77. Moor AE, Itzkovitz S. Spatial transcriptomics: paving the way for tissue-level systems biology. *Curr Opin Biotechnol* 2017;46:126–133.
78. Nagendran M, Riordan DP, Harbury PB, Desai TJ. Automated cell-type classification in intact tissues by single-cell molecular profiling. *Elife* 2018;7.:
79. Schulz D, Zanotelli VRT, Fischer JR, Schapiro D, Engler S, Lun X-K, Jackson HW, Bodenmiller B. Simultaneous Multiplexed Imaging of mRNA and Proteins with Subcellular Resolution in Breast Cancer Tissue Samples by Mass Cytometry. *Cell Syst* 2018;6:531.
80. Ke R, Mignardi M, Pacureanu A, Svedlund J, Botling J, Wählby C, Nilsson M. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* 2013;10:857–860.
81. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, Turczyk BM, Yang JL, Lee HS, Aach J, Zhang K, Church GM. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc* 2015;10:442–458.
82. Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan C,

Liu J, Nolan GP, Bava F-A, Deisseroth K. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 2018;361.:

83. Haga SB, Beskow LM. Ethical, legal, and social implications of biobanks for genetics research. *Adv Genet* 2008;60:505–544.
84. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 2018;19:15.
85. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, Ginhoux F, Newell EW. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 2018;doi:10.1038/nbt.4314.
86. Kobak D, Berens P. The art of using t-SNE for single-cell transcriptomics. 2018;doi:10.1101/453449.
87. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–444.
88. Eraslan G, Simon LM, Mircea M, Mueller NS, Theis FJ. Single cell RNA-seq denoising using a deep count autoencoder. 2018;doi:10.1101/300681.
89. Zepp JA, Zacharias WJ, Frank DB, Cavanaugh CA, Zhou S, Morley MP, Morrisey EE. Distinct Mesenchymal Lineages and Niches Promote Epithelial Self-Renewal and Myofibrogenesis in the Lung. *Cell* 2017;170:1134–1148.e10.
90. Eulenberg P, Köhler N, Blasi T, Filby A, Carpenter AE, Rees P, Theis FJ, Wolf FA. Reconstructing cell cycle and disease progression using deep learning. *Nat Commun* 2017;8:463.
91. Shen-Orr SS, Gaujoux R. Computational deconvolution: extracting cell type-specific information from heterogeneous samples. *Curr Opin Immunol* 2013;25:571–578.
92. Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A, Ryu JH, Wagner BK, Shen-Orr SS, Klein AM, Melton DA, Yanai I. A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst* 2016;3:346–360.e4.
93. Baharom F, Thomas S, Rankin G, Lepzien R, Pourazar J, Behndig AF, Ahlm C, Blomberg A, Smed-Sörensen A. Dendritic Cells and Monocytes with Distinct Inflammatory Responses Reside in

Lung Mucosa of Healthy Humans. *J Immunol* 2016;196:4498–4509.

94. Ardini-Poleske ME, Clark RF, Ansong C, Carson JP, Corley RA, Deutsch GH, Hagood JS, Kaminski N, Mariani TJ, Potter SS, Pryhuber GS, Warburton D, Whitsett JA, Palmer SM, Ambalavanan N, LungMAP Consortium. LungMAP: The Molecular Atlas of Lung Development Program. *Am J Physiol Lung Cell Mol Physiol* 2017;313:L733–L740.
95. Ordovas-Montanes J, Dwyer DF, Nyquist SK, Buchheit KM, Vukovic M, Deb C, Wadsworth MH 2nd, Hughes TK, Kazer SW, Yoshimoto E, Cahill KN, Bhattacharyya N, Katz HR, Berger B, Laidlaw TM, Boyce JA, Barrett NA, Shalek AK. Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* 2018;560:649–654.
96. Xu Y, Mizuno T, Sridharan A, Du Y, Guo M, Tang J, Wikenheiser-Brokamp KA, Perl A-KT, Funari VA, Gokey JJ, Stripp BR, Whitsett JA. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* 2016;1:e90558.

Boxed item 1 / Discovery of pulmonary ionocytes:

Members of the HCA initiative recently led an investigation to characterize the diversity of epithelial cells that line the airways of the lung(3). In addition to finding an unexpected diversity of club, tuft, and goblet cells in the murine airway, they also identified a new airway epithelial cell type. This novel cell type specifically expresses the transcription factor Foxi1 and its composite gene expression profile resembles those of specialized ion transporting cells in fish gills, frog skin, and mammalian kidney. These diverse cell types are collectively referred to as ionocytes, and thus the new pulmonary cell type has been coined a 'pulmonary ionocyte'. Surprisingly, these cells also expressed the majority of the cystic fibrosis transmembrane conductance regulator (Cftr), whose mutation causes Cystic Fibrosis (CF). The investigators found that deletion of Foxi1 in mice resulted in the loss of mature ionocytes and significantly altered mucus viscosity. They also found that FOXI1^{+} pulmonary ionocytes line the airways of human lungs, and that human pulmonary ionocytes also are high expressors of CFTR. Their role in ion transport, regulation of mucus, and high CFTR expression collectively suggests that pulmonary ionocytes play a critical role in CF biology and disease.

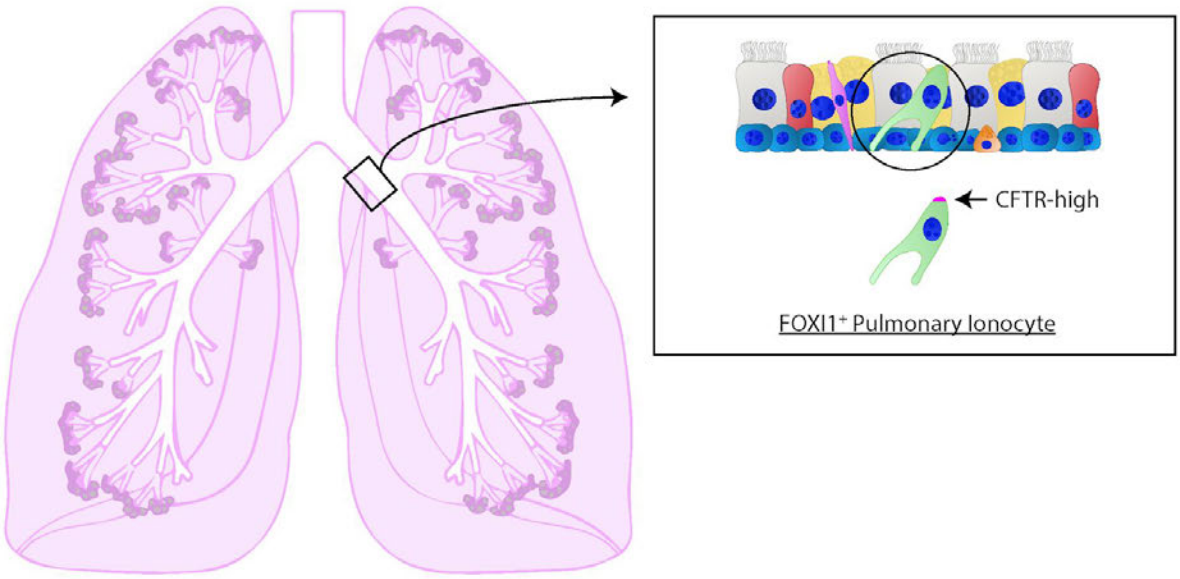

Figure 1. Pulmonary ionocytes line the epithelium of large airways of human lungs, uniquely express the transcription factor FOXI1, and more highly express CFTR than any other cell type in the lung.

Boxed item 2 / From cell type identity to dynamic transitions in cellular phenotypes in space and time

First published pilot datasets by members of the HCA initiative on healthy and diseased human lung (48, 53, 95) have established marker signatures of the main cell types in human lungs. These datasets were derived from both healthy donor and diseased tissue (e.g. Asthma, ILD/IPF). Of note, in many cases cell type specific clusters do not align well in high dimensional gene expression space, even if only healthy donor controls are compared. This is typically discussed as sample specific `batch´ effects, however very likely also represents true biological variation between donors. The human lung is exposed to a great variety of environmental perturbations (e.g. infections, allergens, nanoparticles, smoke) and individual donors will differ in predispositions to react to these environmental insults (e.g. genetics, age). Thus, the observed inter-individual variation of cellular phenotypes will partially represent specific cellular responses to perturbations and subsequent return to normal homeostasis. Data acquired for the Human Lung Cell Atlas will enable to generate predictive models of cell type specific gene regulation. To experimentally test these models we propose that it will be necessary to develop and employ human ex vivo perturbation assays on organoid and tissue culture models.

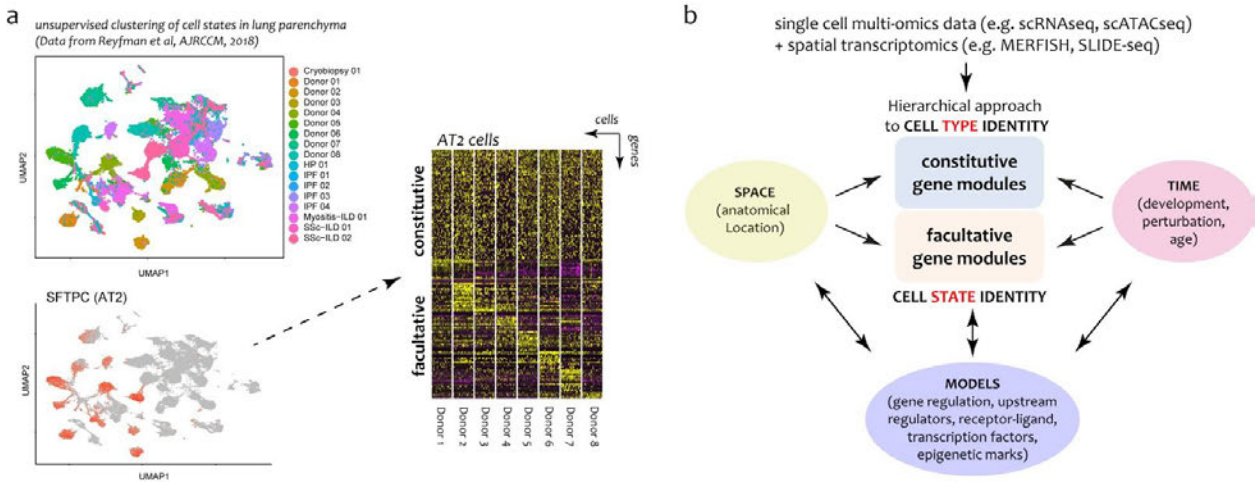

Figure 2 . (a) Surfactant protein C (SFTPC) positive AT2 cell clusters are shown in a dimension reduced UMAP plot (left upper and lower panel). The heatmap in the right panel shows constitutive genes defining AT2 cell type identity across 8 healthy donors, and clusters of genes that have been found to significantly differ between AT2 cells of different individuals. **(b)** Data collected in the Human Lung Cell Atlas project will be used to define cell type identity in a *hierarchical approach*, which defines cell type by similarity metrics of their marker genes across many individuals. Based on these cell type identities and observed variation in facultative gene modules within the different cell types we will generate predictive models of gene regulation that can be experimentally tested. These models will put (disease-) genes in context of their cell type specific `gene regulatory environments´ and predict both the appearance of constitutive and facultative gene modules as a function of time and space as occurring during lung development or regenerative and immune responses.