



Breathing fresh air into respiratory research with single-cell RNA sequencing

Michael J. Alexander, G.R. Scott Budinger and Paul A. Reyfman 

Affiliation: Northwestern University, Feinberg School of Medicine, Dept of Medicine, Division of Pulmonary and Critical Care Medicine, Chicago, IL, USA.

Correspondence: Paul A. Reyfman, Northwestern University, Feinberg School of Medicine, Division of Pulmonary and Critical Care Medicine, 240 E Huron, McGaw 2-422, Chicago, IL 60611, USA. E-mail: paul.reyfman@northwestern.edu

 @ERSpublications

Single-cell RNA sequencing is being used more and more in respiratory research. This technology can be leveraged in different ways in studies of lung development, structure and function, and lung diseases from pulmonary fibrosis to asthma to lung cancer. <https://bit.ly/36TdvJT>

Cite this article as: Alexander MJ, Budinger GRS, Reyfman PA. Breathing fresh air into respiratory research with single-cell RNA sequencing. *Eur Respir Rev* 2020; 26: 200060 [<https://doi.org/10.1183/16000617.0060-2020>].

ABSTRACT The complex cellular heterogeneity of the lung poses a unique challenge to researchers in the field. While the use of bulk RNA sequencing has become a ubiquitous technology in systems biology, the technique necessarily averages out individual contributions to the overall transcriptional landscape of a tissue. Single-cell RNA sequencing (scRNA-seq) provides a robust, unbiased survey of the transcriptome comparable to bulk RNA sequencing while preserving information on cellular heterogeneity. In just a few years since this technology was developed, scRNA-seq has already been adopted widely in respiratory research and has contributed to impressive advancements such as the discoveries of the pulmonary ionocyte and of a profibrotic macrophage population in pulmonary fibrosis. In this review, we discuss general technical considerations when considering the use of scRNA-seq and examine how leading investigators have applied the technology to gain novel insights into respiratory biology, from development to disease. In addition, we discuss the evolution of single-cell technologies with a focus on spatial and multi-omics approaches that promise to drive continued innovation in respiratory research.

Introduction

Every cell in the body shares a similar genome, but the epigenome, transcriptome, proteome and metabolome of each cell varies dramatically between tissues and cells. These “omes” beyond the genome dynamically change in response to environmental challenges, disease states and ageing. While technological advances increasingly allow measurement of epigenome, proteome and metabolome in small tissue samples that can be collected as part of clinical care, none are as robust, reproducible or low cost as next-generation sequencing (NGS) technologies to measure the transcriptome [1–3]. NGS technologies first allowed direct measurement of gene expression in composite tissues *via* sequencing of messenger RNA (RNA-seq) in 2008 [4–6]. Applying these technologies to ever-smaller samples allowed profiling of gene expression in a single cell within a year [7]. Since then, commercialisation and standardisation have made these technologies available in most advanced laboratories, supporting an explosion of publications using single-cell RNA-seq (scRNA-seq). Reductions in cost and advances in computational approaches

Provenance: Commissioned article, peer reviewed.

Received: 6 March 2020 | Accepted after revision: 21 May 2020

Copyright ©ERS 2020. This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0.

have allowed the number of cells profiled in these studies to increase exponentially over time reaching >1 million per study [8, 9]. Boosted by these enabling technologies, scRNA-seq is being used in large-scale efforts to provide a high-resolution map of every cell in the human body, offering unparalleled opportunities to explore cellular interactions and trajectories over the course of disease.

The community of respiratory researchers, long hampered by the cellular complexity of the lung, have been leaders in applying scRNA-seq to the study of human disease. These studies have supported a broad array of findings, including insights into respiratory system development, the identification of novel cell types in the human lung and profiles of heterogeneity in respiratory system cell populations in health and disease [10–15]. The ability to address fundamental biological questions is continuously expanding as technologies to collect and process respiratory specimens for scRNA-seq are refined, costs for reagents and sequencing fall and computational platforms become more robust. Rapid advances in spatial transcriptomics, epigenomics, proteomics and metabolomics provide the opportunity for an integrated multi-omic approach to investigating lung disease. Nevertheless, techniques to leverage data generated from scRNA-seq technologies for respiratory research are evolving, and the limitations of these technologies for profiling respiratory samples are incompletely understood. In this review, we aim to provide an overview of scRNA-seq technologies focused on its applications and limitations when applied to studies of the respiratory system. We begin with some illustrative examples from our own group and others that address disease focused questions that can be specifically answered using scRNA-seq.

Case study 1: alveolar macrophage heterogeneity in pulmonary fibrosis

The understanding of alveolar macrophages as a homogenous, nonreplicating cell population continuously replenished from a reservoir of peripheral monocytes changed dramatically when a series of lineage-tracing studies in mice showed that alveolar macrophages are a long-lived, self-renewing population that populates the lung immediately after birth and persists without input from circulating monocytes over prolonged periods of time [16–20]. In murine models of bleomycin- and asbestos-induced fibrosis, we found that monocyte-derived alveolar macrophages recruited in response to lung injury were necessary for fibrosis, while tissue-resident alveolar macrophages were dispensable [21, 22]. We used genetic lineage tracing systems to flow cytometry sort tissue-resident and monocyte-derived alveolar macrophages for bulk RNA-seq, which showed that monocyte-derived alveolar macrophages exhibit a profibrotic transcriptomic signature distinct from their tissue-resident counterparts. These findings predicted the presence of at least two transcriptionally distinct populations of alveolar macrophages in the lungs of patients with pulmonary fibrosis, a question that could only be addressed using scRNA-seq [14]. Applying this technology to the human lung, we identified two populations of alveolar macrophages in the lungs of patients with pulmonary fibrosis, one of which resembled macrophages from normal lungs and one of which differentially expressed profibrotic genes homologous to those we observed in mice. We were able to definitively show this in a remarkably small group of patients (eight patients with lung fibrosis and eight controls), suggesting that cellular heterogeneity rather than true biological variability might have masked signals in previous studies using bulk RNA-seq. Similar results were found by two independent laboratories, highlighting the reproducibility of scRNA-seq even when applied to human disease [23, 24].

Case study 2: discovery of lung ionocytes

Two groups used scRNA-seq to sequence a very large number of cells from samples of normal human airway, thereby describing the transcriptional landscape of the airway at unprecedented resolution. They identified a transcriptionally unique cell type that expressed large quantities of *CFTR* (cystic fibrosis transmembrane conductance regulator), the causal gene of cystic fibrosis [11, 12]. In addition, this cell was characterised by increased expression of several other ion transporters and the transcription factor *FOXI1*, a transcription factor with homology to *foxl1*, the canonical transcription factor of *Xenopus* larval skin ionocytes. The studies describing the discovery of the human airway ionocyte also reported extensive validation to demonstrate the presence of airway ionocytes in the human and murine lung using single-molecule fluorescence *in situ* hybridisation. They further validated their findings by performing functional studies in mice to show that this population of pulmonary ionocytes contributes disproportionately to CFTR function in the airway. These data challenged the existing paradigm that ciliated cells expressing FOXJ1 are the major source of CFTR protein in the lung with potential implications for gene-therapy approaches to cystic fibrosis.

Case study 3: characterising the immune landscape of lung cancer

Immunotherapy has quickly become a frontline therapy in patients with nonsmall cell lung cancers (NSCLC) lacking driver mutations. A minority of patients receiving immune checkpoint blockade will exhibit a durable response to disease, an outcome that was previously unheard of. Unfortunately for the majority of patients, progression of their lung cancer is inevitable. The tumour immune microenvironment

has been found to have a large influence on response to immunotherapy and by understanding the unique compositions of those environments researchers hope to better predict responsiveness and reveal novel targets [25]. Using scRNA-seq, a group of researchers characterised the tumour T-cell landscape of 14 patients with NSCLC compared to surrounding lung and peripheral blood [26]. They found significant heterogeneity in CD8⁺ and regulatory T-cell (Treg)-infiltrating T-cells with increasing proportions of exhausted CD8⁺ subtypes and activated Tregs portending a worse prognosis.

Why perform scRNA-seq?

A key limitation of bulk RNA-seq applied to whole-tissue or to purified cell populations is that this approach necessarily averages the gene expression signals of all the cells in a sample (figure 1b). This is a particular concern for complex tissues like the lung, which includes 40 or more cell types [27]. Because many of these populations are relatively rare or difficult to dissociate intact from the underlying tissue, their contributions to disease signatures from bulk RNA-seq analysis are lost. Furthermore, scRNA-seq analyses have raised doubt about the reliability of methods of “deconvolution”, or computational estimation of the composition of individual cell populations in bulk RNA-seq data. While this limitation can be mitigated by purification of cell types or subpopulations using flow cytometry, often no such method of purification is available. Analysis using scRNA-seq estimates the transcriptomes of individual cells in a sample, enabling attribution of gene expression signals to cell types or subpopulations and allowing recognition of rare cells or pathogenic populations of cells that appear only during disease (figure 1). Compared with flow cytometry, which measures levels of surface or intracellular proteins with single-cell resolution, scRNA-seq offers a less biased portrayal of gene expression because it measures RNA corresponding to all genes rather than measuring only a prespecified panel of protein markers. One frequently discussed limitation of scRNA-seq data when compared with bulk RNA-seq data relates to the relatively shallow depth of sequencing within each cell. As a result, single-cell data is susceptible to the problem of “dropout” where a zero value for a gene may or may not reflect a lack of expression in an individual cell. However, recent work suggests that this sensitivity problem in scRNA-seq data can be overcome by sampling and sequencing sufficiently large numbers of cells [28].

Single-cell RNA-sequencing workflows

Several workflows for scRNA-seq have been developed, and most scRNA-seq experiments are now performed using commercial platforms (table 1 and figure 2). A critical step of any scRNA-seq experiment is generation of a single-cell suspension from a sample of interest (figure 2b). For solid tissues, this is usually done using a combination of mechanical dissociation and enzymatic digestion. Existing studies have generally used fresh tissue, but protocols have been developed for cryopreserved samples or for cells isolated using laser-capture microdissection from formalin-fixed, paraffin-embedded samples [41–43]. One

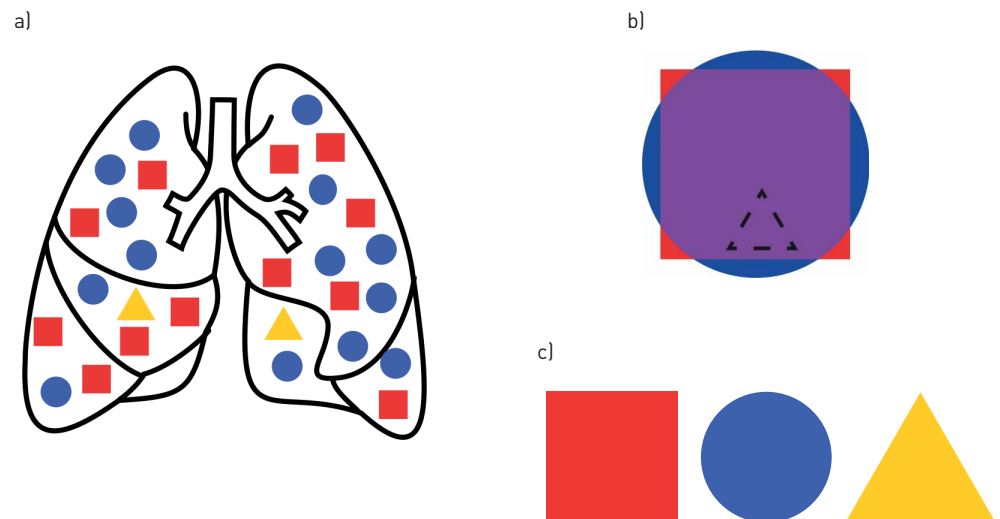


FIGURE 1 Single-cell RNA-sequencing is able to resolve distinct individual cellular transcriptomes compared to bulk RNA-sequencing. a) Whole lung tissue with distinct cellular components represented by blue circles, red squares and yellow triangles; b) transcriptional output of bulk RNA-sequencing experiments results in an averaging of the individual cellular signals. While some marker genes unique to cell types may indicate the presence of that cell type (protruding corners of red square and arcs of blue circle), the overall signal will be a mixture of the cell types (purple) and may obscure the transcriptomic signature of rare cell types (dashed triangle); c) in a single-cell experiment, every cell type (red square, blue circle, yellow triangle) is represented.

TABLE 1 Aspects of single-cell RNA-sequencing experiments

Experimental design [29-31]	
Organism	Culture, animal, human
Replicates/number of cells	General considerations including organism heterogeneity and ability to resolve rare cell populations of interest
Tissue disaggregation	Mechanical dissection, enzymatic disaggregation
Enrichment	FACS, magnetic bead separation, centrifugation
Cell capture	
Microfluidics	
Nanowell [32, 33]	
Droplet [34]	Advantages in throughput and cost
FACS	Relatively lower throughput; protein markers needed
Plate-based	
Micropipette/laser capture microscopy [7, 35]	
Library preparation [36]	
	NGS Illumina <i>versus</i> other cDNA-compatible system; availability depends on protocol
3'	Does not capture splice variants; lower sequencing depth required
5'	Requires sequencing to higher read depth
Full-length transcript	Costly; splice variant, isoform and allelic variation analysis
Quality	
Cell integrity	
FACS	Sorting out dead cells, cell fragments
Imaging	Plate-based systems; detection of empty wells, doublets
RNA quality	
RIN [37]	RNA quality score
Barcoding	
Cell label	One sequence per bead/well
UMI [38]	Different for every oligo on bead; differentiates PCR clone from transcript read
Multiplexing	
RNA spike-in	
Animal doping	Multiplexing with cells of different species, detect doublets
Multiple donor [39, 40]	Using SNPs from unrelated donors or oligo-tagged antibodies to detect doublets, correct for batch effect in multiplexed experiments

FACS: fluorescence-activated cell sorting; RIN: RNA integrity number; UMI: unique molecular identifier; NGS: next-generation sequencing; cDNA: complementary DNA; SNP: single nucleotide polymorphism.

study demonstrated that intact lung tissue could be cold-preserved at 4°C for up to 72 h prior to processing for scRNA-seq without substantial evidence of degradation in the resulting data quality [44]. Following dissociation, individual cells are collected and then lysed in either wells on a plate, oil droplets or nanowells on a chip (figure 2c). This enables capture of messenger RNA (mRNA) molecules, generation of complementary DNA (cDNA) by reverse transcription and barcoding of cDNA for each individual cell in isolation from all other cells. A barcode incorporated into the cDNA during the process of preparing a cDNA “library” for each cell consists of a unique nucleotide sequence for each cell, and often another sequence called a unique molecular identifier (UMI) for each mRNA molecule captured from that cell. Afterwards, the cDNA from all cells can be pooled and sequenced using NGS technology.

The sensitivity of scRNA-seq which enables its unparalleled resolution also renders the technology susceptible to technical and experimental biases that may obscure true biological signals. Differences in tissue dissociation protocols contribute significantly to batch effect and technical variation. The method of mechanical and enzymatic disaggregation, the processing time and the strengths of reagents can all affect downstream analysis. Aggressive or prolonged digestion protocols can cause cell death or cell fragmentation that release ambient RNA into the media (figure 2b and c). Because this RNA is included in each droplet, genes from these dead cells appear to be “expressed” in all of the sequenced cells (figure 2d). Gentle dissociation protocols lead to overrepresentation of cells that are easily liberated from the tissue [29, 45]. This is a particular problem in the lung where diseases (fibrosis, emphysema and pneumonia) dramatically alter the cellular composition and matrix characteristics of the tissue. Even in normal lung tissue this becomes evident as cell types that are more difficult to liberate intact, such as broad, flat, sail-like alveolar type I (AT1) cells and matrix-embedded fibroblasts, are relatively undersampled compared with cuboidal alveolar type II (AT2) cells and alveolar macrophages in published

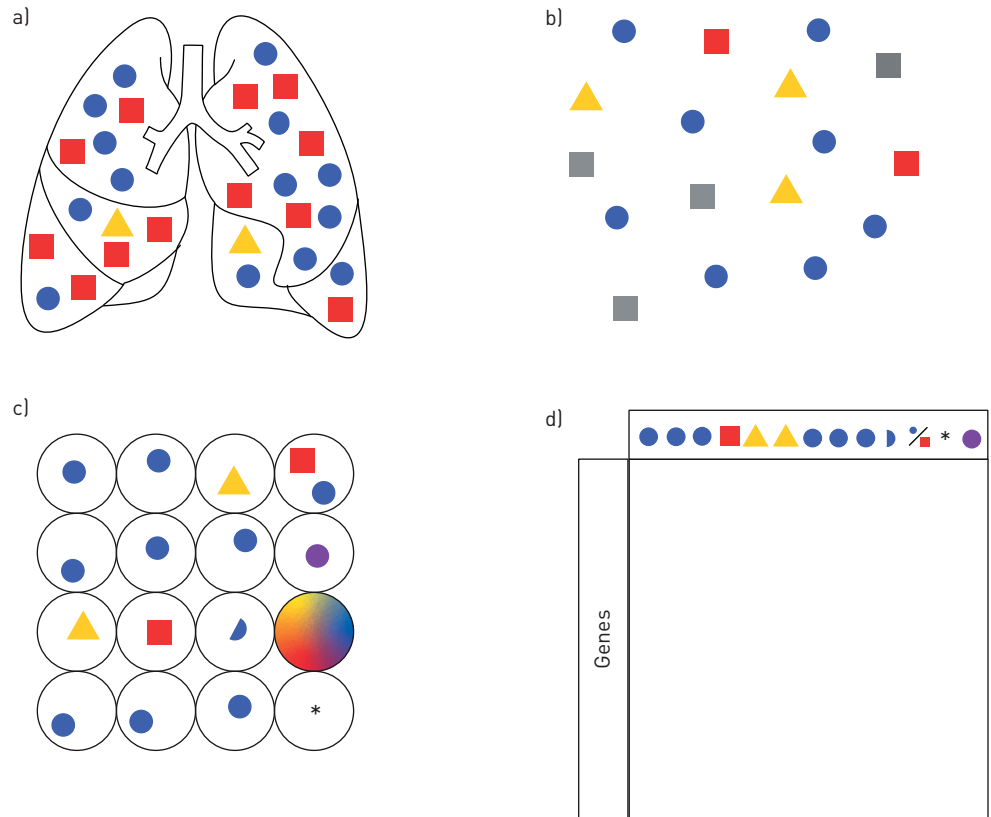


FIGURE 2 Typical workflow of a single-cell RNA-sequencing experiment and possible pitfalls. a) Whole-lung tissue with highly abundant cellular components (blue circle, red square) and rare cell type (yellow triangle). b) A single-cell suspension is created through the mechanical and enzymatic disaggregation of the lung. Cell types that are fragile or difficult to liberate intact from the tissue (grey squares) may be underrepresented in the final dataset. In the lung, these cell types typically include alveolar type 1 epithelial cells and mesenchymal cells including fibroblasts. c) Individual cells are isolated into vessels (plate wells or droplets) with barcoded RNA primers and the cells are lysed to create a unique complementary DNA (cDNA) library for each cell. A number of errors can be introduced at this step, including the inclusion of two or more cells in a vessel, the inclusion of a cell fragment, the creation of a library from an empty vessel (*) or one containing ambient RNA (droplet with all colours), the inclusion of apoptotic cells or the induction of transcriptional changes as a result of processing (purple circle). d) cDNA libraries are sequenced and separated by barcode to get all the individual sequences for the individual cell. Sequences are aligned to a reference genome and successful alignments with known gene sequences are counted. This expression matrix is filtered by several quality criteria and used for downstream analyses.

datasets. Efforts are under way to improve consistency and reliability of scRNA-seq methods and decrease technical variation, including the implementation of automated tissue processing. Data structures designed for storing data from single-cell experiments generally permit the storage of metadata detailing experimental conditions under which the data were obtained. The standardisation and meticulous documentation of implemented protocols will support improved consistency across operators and institutions.

scRNA-seq analysis

The initial processing of raw scRNA-seq data entails sorting reads based on their cell barcodes and aligning the reads to the reference transcriptome of the organism being studied to produce a table of gene counts. In this table, each column corresponds to a cell barcode and each row corresponds to a gene so that each entry contains an integer reflecting the number of reads with that cell barcode mapped to that gene (figure 2d). If libraries have been prepared for sequencing using UMIs, then all reads with the same UMI are treated as amplification duplicates and are collapsed into a single read in the counts table. A threshold is often selected to perform filtering on the counts table which removes cell barcodes with low sequencing depth, or total number of unique reads. These removed barcodes are attributed to capture, during cell isolation, of partial or damaged cells or ambient mRNA. Additional filtering may be performed to remove cell identities based on other metrics reflecting poor quality of input cells, such as having greater than a specific proportion of reads mapping to mitochondrial genes.

Following filtering and other quality-control procedures, machine-learning tools can be used to group cells. Most often this entails clustering using either a similarity- or distance-based metric, and then assigning a cell identity to each resulting cluster based on the most highly expressed genes in that cluster compared with the other clusters [46–48]. Machine-learning algorithms for projecting high-dimensional data in low-dimensional space, such as t-distributed stochastic neighbour embedding (t-SNE) or uniform manifold approximation and projection (UMAP) are usually used for generating visualisations of clustered data [49, 50]. The results of a clustering analysis could provide evidence for a novel cell type or cell state if a distinct cluster was identified that expressed relatively high levels of unexpected genes. Additionally, if an scRNA-seq experiment was designed to compare two different conditions, differential expression analysis could be used to estimate which genes were expressed at a statistically higher level in one condition compared with another between clusters corresponding to the same cell type in samples from the respective conditions. Several computational tools have been developed to assist in combining scRNA-seq data from different experimental conditions or batches or data produced using varying protocols or by distinct groups [46, 51, 52]. These techniques are necessary for generating aggregate datasets and increasing the power to detect small expression differences, but they have to be applied thoughtfully. It is possible to remove expression signatures of true biological differences using these integration techniques, such as when a distinct cell type exists exclusively in one experimental condition.

Auxiliary analyses using scRNA-seq data can answer a number of biologically meaningful questions. If the transition of cells through a biological process such as differentiation is of interest, a pseudotime analysis can be performed where cells are arranged sequentially on a graph according to their similarity, computationally recapitulating a temporal phenomenon occurring in a cross-sectional sampling of cells [53–55]. Another procedure to infer these trajectories, RNA velocity, compares the ratios of spliced, mature mRNA to unspliced, nascent mRNA between cells to identify likely transitional states between cells [56]. If the interest is in uncovering which gene pathways or gene regulatory networks are differentially modulated between cell types or experimental conditions, differentially expressed genes from scRNA-seq clustering analysis can be analysed using a statistical test for enrichment of curated gene sets from a collection such as GO or MSigDB [57, 58]. And if the focus is in uncovering novel cell type interactions, a curated list of ligand–receptor pairs combined with lists of highly expressed genes in different cell clusters from scRNA-seq data can be used to generate hypotheses about signalling interactions between different cell types, such as alveolar macrophage–epithelial cell interactions in the lung [59].

scRNA-seq collaborations and data repositories

Collaborative efforts to generate and publish data using scRNA-seq on the cellular compositions of normal tissues have coalesced around several “atlases”. These efforts are motivated by the goal of making raw scRNA-seq data available as quickly and as widely as possible with the recognition that the value of a single dataset can be increased through careful reanalysis in combination with other datasets. Consistent with such a fundamentally collaborative approach, many researchers using scRNA-seq have adopted the practice of making publications available before peer review on preprint servers such as ArXiv (<https://arxiv.org>), bioRxiv (www.biorxiv.org/) and medRxiv (www.medrxiv.org/). The idea that the free sharing and synergistic integration of scRNA-seq datasets can accelerate the pace of discovery is predicated on the quality of those datasets. Collaboration is only possible through corroboration. Computational genomics is not immune to the reproducibility challenges in other branches of science, but in data science any result should be easily reproducible from the raw data and code used to generate the final analysis. We believe that the full disclosure of the experimental procedures, raw data, metadata and code used in a study is critical to the integrity of research. In the United States, the National Institutes of Health (NIH) has published requirements for genomic data sharing for all NIH-funded research and maintains several large databases to meet this need [60, 61]. In Europe, the European Molecular Biology Laboratory through the European Bioinformatics Institute operates similar repositories committed to the free sharing of genomic data [62]. Our practice is to upload our raw sequencing data to repositories maintained by the NIH including the Sequence Read Archive or the Database of Genotypes and Phenotypes (dbGaP, for human data) and to post our code to public repositories. Many journals now mandate that data and code generated in support of an article be uploaded to a repository and freely accessible to other researchers [63, 64].

Challenges and future directions of scRNA-seq

The computational identification of a cell population or cell state using any of the commonly employed clustering algorithms always requires independent validation with a complementary technology (table 2). Ideally for tissue analyses, this validation should add spatial information that defines the “neighbourhood” in which a putative new cell population or cell state resides. To this end, the single-cell research community is sharply focused on developing multiplexed, high-resolution spatial transcriptomic and

TABLE 2 Emerging single-cell technologies and multi-omics

Single nucleus [65, 66]	Analysis of frozen/archival tissue
Spatial	
smFISH [67]	Single-molecule fluorescent <i>in situ</i> hybridisation
ISH+amplification [68]	PLISH
ISH+amplification+multiplexing [69–71]	MERFISH, SCRINSHOT, STARmap (3D)
Oligo microarray [72]	
LCM capture [73, 74]	Single cells on array, captured, sequenced
Multi-omics	
Surface markers	
Epitope capture [75, 76]	Simultaneous sequencing of mRNA and oligo-labelled antibodies; CITE-seq, REAP-seq
TCR antigen specificity [77]	Oligo-tagged MHC multimer libraries
Epigenome	
scRBBS [78]	Reduced-representation bisulfite sequencing
scWGBS [79]	Whole-genome bisulfite sequencing
Parallel transcriptome and methylome [80–82]	
CHIP-seq [83]	Chromatin immunoprecipitation sequencing
ATAC-seq [84]	Assay for transposase-accessible chromatin
Metabolome/proteome	
Mass spectrometer/MALDI-TOF	Matrix-assisted laser desorption/ionisation
Mass cytometry [85, 86]	Cytometry TOF; panel of antibody-labelled ions

ISH: *in-situ* hybridisation; LCM: laser capture microscopy; TCR: T-cell receptor; TOF: time of flight; CITE-seq: cellular indexing of transcriptomes and epitopes by sequencing; REAP-seq: RNA expression and protein sequencing; MHC: major histocompatibility complex.

proteomic methods to both validate and extend data from scRNA-seq. To date, most studies in the lung have used variants of single-molecule RNA fluorescence *in situ* hybridisation to validate their findings. However, these techniques are limited by the small number of probes that can be imaged simultaneously. To address this concern, multiplexed techniques that allow sequential imaging of multiple probes on a tissue section are being optimised [68–70]. Tissue clearing and permeabilisation protocols combined with light sheet microscopy offer the opportunity for three-dimensional imaging. All of these spatial transcriptomic techniques will necessitate advances in terms of processing and storage given the size and high complexity of these datasets [71–74, 87].

While the technology to measure the transcriptome at single-cell resolution is more robust and less costly when compared with those that measure the epigenome, proteome or metabolome, these other technologies are advancing rapidly. Examples of integrated single-cell transcriptomics and single-cell ATAC-seq (assay for transposase-accessible chromatin, which measures regions of open chromatin) are already present [80, 81]. While not yet available at single-cell resolution, reduced representation bisulfite sequencing to query the DNA methylome and “cut and run” technology to measure chromatin modifications can be done with very small numbers of cells [88–90]. Multiparameter flow cytometry and mass cytometry (CyTOF) can currently measure several proteins simultaneously to complement scRNA-seq with proteomic analysis [91]. Even more exciting is a method for incorporating oligonucleotide-tagged antibodies into a droplet-based scRNA-seq workflow called CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) [75]. This technology permits simultaneous measurement of gene expression and surface protein levels in the same single cell.

The liberation of cells from tissues for scRNA-seq differs depending on the tissue digestion protocols used. Therefore, scRNA-seq cannot reliably recapitulate the cellular composition of tissues [14, 92]. Single-nucleus RNA-seq (snRNA-seq) offers a solution to this problem. Protocols using this technology typically take advantage of the relative resistance of the nuclear membrane compared to the plasma membrane to fracture during a freeze–thaw cycle. The intact nucleus is then used in typical droplet-based sequencing protocols. While this procedure necessarily increases contamination by ambient RNA, overrepresents long noncoding and small nuclear RNA and reduces the number of measured genes, it has shown promise when applied to the brain [65, 93, 94]. Most importantly, snRNA-seq can be performed on frozen tissue, offering the promise to generate transcriptomic information at single-cell resolution from archival tissues [65, 93–97].

Costs of scRNA-seq remain high because of costs of reagents in library preparation workflows and costs of sequencing [8, 36]. Sequencing costs are likely to continue to decrease, but current technologies may reach a limit such that further lowering of cost will require the development of novel NGS platforms [98]. Innovations in experimental protocols have produced great decreases in library preparation cost and are likely to continue to do so. Techniques for adding barcodes to samples prior to single-cell isolation, or for using single nucleotide variants to computationally determine sample genotypes will allow pooling of multiple samples for library preparation [39, 40, 99, 100]. In addition, these techniques are helpful for mitigating the effect of doublets, a single shared cell ID barcode in a scRNA-seq dataset that in actuality corresponds to two distinct cells.

Preparation of libraries for scRNA-seq using droplet-based methods results in the capture of ambient RNA that is present in the input single-cell suspension [14]. Thus, genes, particularly those expressed at high levels in cells that are highly prevalent in the input sample, will be counted for cells in which they may not actually be expressed. Computational tools have been developed to address this issue by estimating contamination and adjusting counts accordingly [101, 102]. However, it is challenging to validate these tools, and the effect of ambient RNA contamination may vary between different tissues, experimental conditions and single-cell library preparation protocols.

Applications of scRNA-seq to respiratory research

Cellular composition of the normal and diseased lung

According to an often-cited estimate, the lung is composed of ~40 different cell types [27]. However, this estimate is based largely on studies that used microscopy, which may not be able to distinguish between cell populations similar in appearance but having distinct transcriptional or proteomic states and correspondingly distinct functions. It is likely that scRNA-seq will expand this number. Indeed, a recent preprint describes the detection of 58 cell populations from scRNA-seq of human lung [103]. Furthermore, technologies like snRNA-seq offer promise to more precisely quantify the changes in cellular composition in the lung that develop in response to environmental challenge, disease and ageing. While still in the future, three-dimensional imaging techniques with automated quantification may enable precisely localising and quantifying cell populations within the three-dimensional volume of the lung, obviating the need for statistical estimates based on stereology [104–107].

Two prominent atlas projects using scRNA-seq, focused on mouse and human data respectively, are the *Tabula Muris* project and the Human Cell Atlas (HCA) project [108–111]. Each of these atlases aims to produce data about normal structure and function of organs and tissues. The *Tabula Muris* is a single publicly available dataset including scRNA-seq from almost 100 000 cells spanning 20 murine organs and tissues, and it includes data from a substantial number of respiratory system cells, which can be included as a basis of comparison in studies of disease models (<https://tabula-muris.ds.czbiohub.org>). The HCA is an ongoing collaborative project aimed at producing data from many different organs and tissues from healthy humans (www.humancellatlas.org/). Furthermore, the HCA contains 38 different seed networks spanning a number of different organs and tissues including a dedicated HCA lung network focused on the respiratory system [112]. Major strengths of the HCA lung network include a welcoming stance that allows any investigator to participate and incorporation of multiple assay modalities including single-cell epigenetic and spatial transcriptomic tools together with scRNA-seq. Recently, during the coronavirus disease (COVID-19) global pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), members of the HCA lung network have demonstrated the power and speed of this consortium in being able to respond to novel biological challenges. Rapidly, members of the HCA lung network produced several studies reporting analyses of datasets combined from different groups detailing anatomical expression of the SARS-CoV-2 receptor *ACE2* and identifying clinical covariates associated with different expression levels in various tissues of viral entry mediators [113–115].

Development

The first published study using scRNA-seq in the respiratory field by TREUTLEIN *et al.* [10] attempted to elucidate the process of lung epithelial cell development in a murine embryonic model. The authors performed clustering on gene expression profiles of lung epithelial cells from embryonic and adult mice to support their finding of a common progenitor cell population of AT1 and AT2 epithelial cells during development. A more recent study reported on scRNA-seq of ~2 million cells from 61 mouse embryos between embryonic days 9.5 and 13.5 [9]. This study was labelled by its authors as an “atlas” in acknowledgement of it being descriptive and of the data it produced being potentially useful for incorporation into future studies. Cells belonging to lung epithelium, a lung epithelial developmental trajectory and other developmental trajectories related to components of the respiratory system are all identified in this study.

Ageing

Ageing is a critical risk factor for the development of many lung diseases including cancer, lung fibrosis, COPD and lung infections [116]. Studying ageing in the lung using scRNA-seq is attractive because it offers the promise of comparing distributions of transcriptional states within cell types at different extremes of age. A study that combined scRNA-seq with proteomics of lungs from young and old mice found that ageing was associated with increased transcriptional noise, an increased ratio of ciliated cells to club cells and increased cholesterol biosynthesis [117]. Additionally, an integrated analysis of the scRNA-seq and proteomic data from that study allowed the investigators to identify the most likely cellular sources of differentially regulated proteins with age. In another study, investigators analysed scRNA-seq data derived from lung as well as kidney and spleen tissue from young and old mice. The findings from this study suggested that the effect of ageing on transcriptional noise varied between different cell types and was greater in lung stromal cell types, for example, than AT2 cells. Investigating the biology of ageing in the lung using scRNA-seq of human tissue is particularly challenging because obtaining normal lung tissue is difficult, particularly from multiple individuals of different ages over the lifespan. As more and more scRNA-seq datasets of human lung tissue are published, such as within the HCA, a meta-analysis of human lung scRNA-seq data over the lifespan will become feasible, that can include enough samples from different groups to control for batch effects.

Lung cancer

Tissue samples from patients with lung cancer are relatively accessible because treatment of lung cancer frequently includes surgical resection. Accordingly, several studies have thus focused on scRNA-seq of lung tissue obtained from patients with NSCLC [26, 118–121]. The majority of these studies highlight the heterogeneity of tumour cells and, as in the study highlighted in case study 3, tumour-associated immune cells [122–124]. In addition, the results of two studies suggested that the presence of certain subsets of tumour-associated immune cell markers was associated with prognosis in patients with NSCLC [26, 121].

Interstitial lung disease and pulmonary fibrosis

While stromal, epithelial and immune cells have all been implicated in contributing to the pathogenesis of pulmonary fibrosis, the first published study leveraging scRNA-seq to investigate pulmonary fibrosis pathobiology focused on epithelial cells [13, 125]. This study of human lung tissue demonstrated, and work from our group and others later confirmed, that pulmonary fibrosis was characterised by the emergence of transcriptionally distinct epithelial cell populations [14, 24, 126]. Our group was among the first to apply scRNA-seq to the analysis of patients with pulmonary fibrosis to address the question of macrophage heterogeneity during fibrosis [14]. Our findings were subsequently confirmed and extended in two independent datasets, highlighting the reproducibility of these approaches [23, 24]. Importantly, all of the groups responsible for these studies have already made their data available or committed to doing so, and some have made these data available in a format that can be explored by the community [15, 23, 24, 44, 126–130]. While most investigators have used scRNA-seq at a single time point during the course of disease, time-series single-cell data offer particular promise. For example, Schiller and colleagues performed scRNA-seq in murine lungs harvested over the course of pulmonary fibrosis in mice induced by bleomycin [131]. The investigators identified and validated a novel Krt8-positive transitional epithelial cell progenitor population that receives input from AT2 and club cells and is responsible for replacing AT1 cells during alveolar repair. This finding was supported by an enrichment analysis indicating higher expression of genes related to proliferation in this novel Krt8-positive population and by an RNA velocity analysis indicating convergence of the transcriptional states of AT2 and club cells on the Krt8-positive state.

Lung injury and repair

Mouse models of lung injury and repair have been used to understand acute lung injury in humans and also to gain insights into related processes involved in lung development and fibrosis. In one study, investigators used scRNA-seq of lung macrophages to confirm a previous observation that following lung injury induced by lipopolysaccharide (LPS), a population of recruited alveolar macrophages expresses increased levels of inflammatory genes compared with resident alveolar macrophages [132, 133]. In another study, investigators performed scRNA-seq on AT2 cells after LPS-induced lung injury and found that a relative decrease in *Tgfb2* expression was associated with AT2 cells occupying a transdifferentiating state rather than a cell cycle arrest state during repair [134].

Asthma

One study has reported on scRNA-seq analysis of samples from patients with asthma. The investigators used a distance-based trajectory analysis, termed “pseudotime”, to estimate developmental trajectories among epithelial cells from healthy controls and patients with asthma [15]. The investigators found that

there were altered epithelial cell differentiation pathways in asthma and increased numbers of goblet and mucous ciliated cells compared with healthy controls. Additionally, there were increased numbers of type 2 helper T (Th2) effector cells in asthma and increased cell-to-cell signalling involving Th2 cells. This finding is further supported by a study using scRNA-seq to analyse Th-cells in a murine house dust mite (HDM) model of allergic airway disease, which identified a distinct Th2 cell gene expression signature in the airways of mice exposed to HDM [135].

Other respiratory system diseases

We anticipate that, for any diseases affecting the respiratory system for which studies using scRNA-seq have not yet been published, such studies are probably forthcoming. For example, for cystic fibrosis, COPD and pulmonary hypertension, analyses of scRNA-seq data have not yet been published, but there have been multiple studies utilising bulk RNA-seq approaches [136–141]. As techniques for tissue processing, library preparation and data analysis improve, and as cost continues to decrease, more studies containing larger numbers of cells can be expected. The discovery of the pulmonary ionocyte, a cell type that expresses high levels of CFTR, provides an exceptional motivation for using scRNA-seq to gain insight into the function of this cell type in cystic fibrosis [11, 12].

Biomarker development

Bulk RNA-seq has been used for biomarker development for a variety of respiratory system diseases, including NSCLC and idiopathic pulmonary fibrosis [142–147]. Developing biomarkers using scRNA-seq to assist in diagnosis, prognosis and selection of appropriate therapy in a variety of lung diseases may be feasible, but several questions will have to be clarified. It is not known whether tissue proximal to the disease is required for biomarker development using scRNA-seq, or whether samples obtained by less invasive means, such as blood or nasal epithelial samples, can be used. Furthermore, it is not known what approaches to scRNA-seq analysis can support biomarker development. The extent to which scRNA-seq can be part of a successful platform for developing clinically relevant biomarkers will need to be determined through future research, although the existing studies hinting at the prognostic relevance of certain tumour-associated immune cell populations are promising [26, 121].

Conclusions

In a relatively short period of time, the impact of research using scRNA-seq for advancing knowledge about respiratory biology and disease has been substantial. The opportunity of data sharing and meta-analysis of data from multiple studies is encouraging the formation of large-scale collaborations [112]. The influence of scRNA-seq on respiratory research will grow as the underlying technologies continue to improve along with experimental and computational approaches for more closely integrating scRNA-seq data with spatial, proteomic and epigenomic data.

Conflict of interest: None declared.

Support statement: Funding support was received from ATS Foundation/Boehringer Ingelheim Pharmaceuticals Inc. Research Fellowship in IPF, National Heart, Lung, and Blood Institute (HL071643, K08HL146943, T32HL076139), National Institute on Aging (AG049665), Parker Foundation (Parker B. Francis Fellowship), and U.S. Department of Veterans Affairs (BX000201). Funding information for this article has been deposited with the Crossref Funder Registry.

References

- 1 Lander ES, Linton LM, Birren B, *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860–921.
- 2 Venter JC, Adams MD, Myers EW, *et al.* The sequence of the human genome. *Science* 2001; 291: 1304–1351.
- 3 van Dijk EL, Auger H, Jaszczyszyn Y, *et al.* Ten years of next-generation sequencing technology. *Trends Genet* 2014; 30: 418–426.
- 4 Lister R, O'Malley RC, Tonti-Filippini J, *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 2008; 133: 523–536.
- 5 Nagalakshmi U, Wang Z, Waern K, *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 2008; 320: 1344–1349.
- 6 Mortazavi A, Williams BA, McCue K, *et al.* Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008; 5: 621–628.
- 7 Tang F, Barbacioru C, Wang Y, *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 2009; 6: 377–382.
- 8 Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. *Nat Protoc* 2018; 13: 599–604.
- 9 Cao J, Spielmann M, Qiu X, *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 2019; 566: 496–502.
- 10 Treutlein B, Brownfield DG, Wu AR, *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 2014; 509: 371–375.

- 11 Montoro DT, Haber AL, Biton M, *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* 2018; 560: 319–324.
- 12 Plasschaert LW, Žilionis R, Choo-Wing R, *et al.* A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* 2018; 560: 377–381.
- 13 Xu Y, Mizuno T, Sridharan A, *et al.* Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* 2016; 1: e90558.
- 14 Reyfman PA, Walter JM, Joshi N, *et al.* Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am J Respir Crit Care Med* 2019; 199: 1517–1536.
- 15 Vieira Braga FA, Kar G, Berg M, *et al.* A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat Med* 2019; 25: 1153–1163.
- 16 Guilliams M, De Kleer I, Henri S, *et al.* Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life *via* GM-CSF. *J Exp Med* 2013; 210: 1977–1992.
- 17 Yona S, Kim KW, Wolf Y, *et al.* Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. *Immunity* 2013; 38: 79–91.
- 18 Scott CL, Henri S, Guilliams M. Mononuclear phagocytes of the intestine, the skin, and the lung. *Immunol Rev* 2014; 262: 9–24.
- 19 Kopf M, Schneider C, Nobs SP. The development and function of lung-resident macrophages and dendritic cells. *Nat Immunol* 2015; 16: 36–44.
- 20 van de Laar L, Saelens W, De Prijck S, *et al.* Yolk sac macrophages, fetal liver, and adult monocytes can colonize an empty niche and develop into functional tissue-resident macrophages. *Immunity* 2016; 44: 755–768.
- 21 Misharin AV, Morales-Nebreda L, Reyfman PA, *et al.* Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J Exp Med* 2017; 214: 2387–2404.
- 22 Joshi N, Watanabe S, Verma R, *et al.* A spatially restricted fibrotic niche in pulmonary fibrosis is sustained by M-CSF/M-CSFR signalling in monocyte-derived alveolar macrophages. *Eur Respir J* 2020; 55: 1900646.
- 23 Valenzi E, Bulik M, Tabib T, *et al.* Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease. *Ann Rheum Dis* 2019; 78: 1379–1387.
- 24 Habermann AC, Gutierrez AJ, Bui LT, *et al.* Single-cell RNA-sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *bioRxiv* 2019; preprint [https://doi.org/10.1101/753806].
- 25 Binnewies M, Roberts EW, Kersten K, *et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med* 2018; 24: 541–550.
- 26 Guo X, Zhang Y, Zheng L, *et al.* Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med* 2018; 24: 978–985.
- 27 Franks TJ, Colby TV, Travis WD, *et al.* 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.
- 28 Cain MP, Hernandez BJ, Chen J. Quantitative single-cell interactomes in normal and virus-infected mouse lungs. *Dis Model Mech* 2020; in press [https://doi.org/10.1242/dmm.044404].
- 29 Nguyen QH, Pervolarakis N, Nee K, *et al.* Experimental considerations for single-cell RNA sequencing approaches. *Front Cell Dev Biol* 2018; 6: 108.
- 30 Chen G, Ning B, Shi T. Single-cell RNA-seq technologies and related computational data analysis. *Front Genet* 2019; 10: 317.
- 31 Lafzi A, Moutinho C, Picelli S, *et al.* Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nat Protoc* 2018; 13: 2742–2757.
- 32 Han X, Wang R, Zhou Y, *et al.* Mapping the mouse cell atlas by microwell-seq. *Cell* 2018; 173: 1307.
- 33 Gierahn TM, Wadsworth MH 2nd, Hughes TK, *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* 2017; 14: 395–398.
- 34 Macosko EZ, Basu A, Satija R, *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 2015; 161: 1202–1214.
- 35 Frumkin D, Wasserstrom A, Itzkovitz S, *et al.* Amplification of multiple genomic loci from single cells isolated by laser micro-dissection of tissues. *BMC Biotechnol* 2008; 8: 17.
- 36 Haque A, Engel J, Teichmann SA, *et al.* A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* 2017; 9: 75.
- 37 Schroeder A, Mueller O, Stocker S, *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* 2006; 7: 3.
- 38 Islam S, Zeisel A, Joost S, *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* 2014; 11: 163–166.
- 39 Kang HM, Subramaniam M, Targ S, *et al.* Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotechnol* 2018; 36: 89–94.
- 40 Stoeckius M, Zheng S, Houck-Loomis B, *et al.* Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol* 2018; 19: 224.
- 41 Guillaumet-Adkins A, Rodríguez-Esteban G, Mereu E, *et al.* Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol* 2017; 18: 45.
- 42 Wohnhaas CT, Leparac GG, Fernandez-Albert F, *et al.* DMSO cryopreservation is the method of choice to preserve cells for droplet-based single-cell RNA sequencing. *Sci Rep* 2019; 9: 10699.
- 43 Foley JW, Zhu C, Jolivet P, *et al.* Gene expression profiling of single cells from archival tissue with laser-capture microdissection and Smart-3SEQ. *Genome Res* 2019; 29: 1816–1825.
- 44 Madissoon E, Wilbrey-Clark A, Miragaia RJ, *et al.* scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biol* 2019; 21: 1.
- 45 van den Brink SC, Sage F, Vértessy A, *et al.* Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat Methods* 2017; 14: 935–936.
- 46 Stuart T, Butler A, Hoffman P, *et al.* Comprehensive integration of single-cell data. *Cell* 2019; 177: 1888–1902.
- 47 Kiselev VY, Kirschner K, Schaub MT, *et al.* SC3: consensus clustering of single-cell RNA-seq data. *Nat Methods* 2017; 14: 483–486.

- 48 Qiu X, Hill A, Packer J, *et al.* Single-cell mRNA quantification and differential analysis with census. *Nat Methods* 2017; 14: 309–315.
- 49 Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008; 9: 2579–2605.
- 50 McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv*: 2018; preprint [https://arxiv.org/abs/1802.03426].
- 51 Johansen N, Quon G. scAlign: a tool for alignment, integration, and rare cell identification from scRNA-seq data. *Genome Biol* 2019; 20: 166.
- 52 Korsunsky I, Millard N, Fan J, *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 2019; 16: 1289–1296.
- 53 Trapnell C, Cacchiarelli D, Grimsby J, *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014; 32: 381–386.
- 54 Bendall SC, Davis KL, El-Ad DA, *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* 2014; 157: 714–725.
- 55 Street K, Risso D, Fletcher RB, *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 2018; 19: 477.
- 56 La Manno G, Soldatov R, Zeisel A, *et al.* RNA velocity of single cells. *Nature* 2018; 560: 494–498.
- 57 Ashburner M, Ball CA, Blake JA, *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; 25: 25–29.
- 58 Subramanian A, Tamayo P, Mootha VK, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; 102: 15545–15550.
- 59 Efremova M, Vento-Tormo M, Teichmann SA, *et al.* CellPhoneDB v2.0: Inferring cell-cell communication from combined expression of multi-subunit receptor-ligand complexes. *Nat Protoc* 2020; 15: 1484–1506.
- 60 National Institutes of Health Office of Science Policy. NIH Genomic Data Sharing Date last accessed: June 11, 2020. Date last updated: June 09, 2020. <https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing/>.
- 61 National Institutes of Health National Human Genome Research Institute. Genomic Data Sharing Policy Date last accessed: June 11, 2020. Date last updated: June 09, 2020. www.genome.gov/about-nhgri/Policies-Guidance/Genomic-Data-Sharing.
- 62 European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI). Data Submission: EMBL-EBI Date last accessed: June 11, 2020. Date last updated: June 08, 2020. www.ebi.ac.uk/submit/.
- 63 Nature Research. Reporting Standards and Availability of Data, Materials, Code and Protocols Date last accessed: June 11, 2020. Date last updated: April 16, 2015. www.nature.com/nature-research/editorial-policies/reporting-standards.
- 64 American Association for the Advancement of Science. Science Journals: Editorial Policies Date last accessed: June 11, 2020. Date last updated: June 09, 2020. www.sciencemag.org/authors/science-journals-editorial-policies.
- 65 Grindberg RV, Yee-Greenbaum JL, McConnell MJ, *et al.* RNA-sequencing from single nuclei. *Proc Natl Acad Sci USA* 2013; 110: 19802–19807.
- 66 Krishnaswami SR, Grindberg RV, Novotny M, *et al.* Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc* 2016; 11: 499–524.
- 67 Raj A, van den Bogaard P, Rifkin SA, *et al.* Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* 2008; 5: 877–879.
- 68 Nagendran M, Riordan DP, Harbury PB, *et al.* Automated cell-type classification in intact tissues by single-cell molecular profiling. *Elife* 2018; 7: e30510.
- 69 Xia C, Babcock HP, Moffitt JR, *et al.* Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Sci Rep* 2019; 9: 7721.
- 70 Sountoulidis A, Lontos A, Nguyen HP, *et al.* SCRINSHOT, a spatial method for single-cell resolution mapping of cell states in tissue sections. *bioRxiv* 2020; preprint [https://doi.org/10.1101/2020.02.07.938571].
- 71 Wang X, Allen WE, Wright MA, *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 2018; 361: eaat5691.
- 72 Ståhl PL, Salmén F, Vickovic S, *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016; 353: 78–82.
- 73 Nichterwitz S, Chen G, Aguila Benitez J, *et al.* Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nat Commun* 2016; 7: 12139.
- 74 Chen J, Suo S, Tam PP, *et al.* Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nat Protoc* 2017; 12: 566–580.
- 75 Stoeckius M, Hafemeister C, Stephenson W, *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 2017; 14: 865–868.
- 76 Peterson VM, Zhang KX, Kumar N, *et al.* Multiplexed quantification of proteins and transcripts in single cells. *Nat Biotechnol* 2017; 35: 936–939.
- 77 Bentzen AK, Marquard AM, Lyngaa R, *et al.* Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat Biotechnol* 2016; 34: 1037–1045.
- 78 Guo H, Zhu P, Wu X, *et al.* Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res* 2013; 23: 2126–2135.
- 79 Smallwood SA, Lee HJ, Angermueller C, *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods* 2014; 11: 817–820.
- 80 Hu Y, Huang K, An Q, *et al.* Simultaneous profiling of transcriptome and DNA methylome from a single cell. *Genome Biol* 2016; 17: 88.
- 81 Angermueller C, Clark SJ, Lee HJ, *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* 2016; 13: 229–232.
- 82 Hou Y, Guo H, Cao C, *et al.* Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res* 2016; 26: 304–319.
- 83 Rotem A, Ram O, Shores N, *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* 2015; 33: 1165–1172.
- 84 Buenrostro JD, Wu B, Litzenburger UM, *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 2015; 523: 486–490.

- 85 Bandura DR, Baranov VI, Ornatsky OI, *et al.* Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal Chem* 2009; 81: 6813–6822.
- 86 Spitzer MH, Nolan GP. Mass cytometry: single cells, many features. *Cell* 2016; 165: 780–791.
- 87 Moncada R, Barkley D, Wagner F, *et al.* Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol* 2020; 38: 333–342.
- 88 Meissner A, Gnirke A, Bell GW, *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* 2005; 33: 5868–5877.
- 89 Singer BD. A practical guide to the measurement and analysis of DNA methylation. *Am J Respir Cell Mol Biol* 2019; 61: 417–428.
- 90 Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* 2017; 6: e21856.
- 91 Yao Y, Liu R, Shin MS, *et al.* CyTOF supports efficient detection of immune cell subsets from small samples. *J Immunol Methods* 2014; 415: 1–5.
- 92 Wu H, Malone AF, Donnelly EL, *et al.* Single-cell transcriptomics of a human kidney allograft biopsy specimen defines a diverse inflammatory response. *J Am Soc Nephrol* 2018; 29: 2069–2080.
- 93 Habib N, Avraham-Davidi I, Basu A, *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* 2017; 14: 955–958.
- 94 Wu H, Kirita Y, Donnelly EL, *et al.* Advantages of single-nucleus over single-cell RNA sequencing of adult kidney: rare cell types and novel cell states revealed in fibrosis. *J Am Soc Nephrol* 2019; 30: 23–32.
- 95 Habib N, Li Y, Heidenreich M, *et al.* Div-Seq: single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* 2016; 353: 925–928.
- 96 Lake BB, Ai R, Kaeser GE, *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* 2016; 352: 1586–1590.
- 97 Lacar B, Linker SB, Jaeger BN, *et al.* Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat Commun* 2016; 7: 11022.
- 98 Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 2016; 17: 333–351.
- 99 Shin D, Lee W, Lee JH, *et al.* Multiplexed single-cell RNA-seq via transient barcoding for simultaneous expression profiling of various drug perturbations. *Sci Adv* 2019; 5: eaav2249.
- 100 McGinnis CS, Patterson DM, Winkler J, *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods* 2019; 16: 619–626.
- 101 Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet based single cell RNA sequencing data. *bioRxiv* preprint [2018 <https://doi.org/10.1101/303727>].
- 102 Yang S, Corbett SE, Koga Y, *et al.* Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol* 2020; 21: 57.
- 103 Travaglini KJ, Nabhan AN, Penland L, *et al.* A molecular cell atlas of the human lung from single cell RNA sequencing. *bioRxiv* 2019; preprint [<https://doi.org/10.1101/742320>].
- 104 Haies DM, Gil J, Weibel ER. Morphometric study of rat lung cells. I. Numerical and dimensional characteristics of parenchymal cell population. *Am Rev Respir Dis* 1981; 123: 533–541.
- 105 Crapo JD, Barry BE, Gehr P, *et al.* Cell number and cell characteristics of the normal human lung. *Am Rev Respir Dis* 1982; 126: 332–337.
- 106 Crapo JD, Young SL, Fram EK, *et al.* Morphometric characteristics of cells in the alveolar region of mammalian lungs. *Am Rev Respir Dis* 1983; 128: S42–S46.
- 107 Hsia CC, Hyde DM, Ochs M, *et al.* An official research policy statement of the American Thoracic Society/ European Respiratory Society: standards for quantitative assessment of lung structure. *Am J Respir Crit Care Med* 2010; 181: 394–418.
- 108 Schaum N, Karkanas J, Neff NF, *et al.* Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*. *Nature* 2018; 562: 367–372.
- 109 Rozenblatt-Rosen O, Stubbington MJT, Regev A, *et al.* The Human Cell Atlas: from vision to reality. *Nature* 2017; 550: 451–453.
- 110 Regev A, Teichmann SA, Lander ES, *et al.* The Human Cell Atlas. *Elife* 2017; 6: e27041.
- 111 Regev A, Teichmann S, Rozenblatt-Rosen O, *et al.* The Human Cell Atlas white paper. *arXiv* 2018; preprint [<https://arxiv.org/abs/1810.05192>].
- 112 Schiller HB, Montoro DT, Simon LM, *et al.* The Human Lung Cell Atlas: a high-resolution reference map of the human lung in health and disease. *Am J Respir Cell Mol Biol* 2019; 61: 31–41.
- 113 Sungnak W, Huang N, Bécavin C, *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat Med* 2020; 26: 681–687.
- 114 Ziegler CGK, Allon SJ, Nyquist SK, *et al.* SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell* 2020; 181: 1016–1035.
- 115 Muus C, Luecken MD, Eraslan G, *et al.* Integrated analyses of single-cell atlases reveal age, gender, and smoking status associations with cell type-specific expression of mediators of SARS-CoV-2 viral entry and highlights inflammatory programs in putative target cells. *bioRxiv* 2020; preprint [<https://doi.org/10.1101/2020.04.19.049254>].
- 116 Thannickal VJ, Murthy M, Balch WE, *et al.* Blue Journal Conference. Aging and susceptibility to lung disease. *Am J Respir Crit Care Med* 2015; 191: 261–269.
- 117 Angelidis I, Simon LM, Fernandez IE, *et al.* An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. *Nat Commun* 2019; 10: 963.
- 118 Min JW, Kim WJ, Han JA, *et al.* Identification of distinct tumor subpopulations in lung adenocarcinoma via single-cell RNA-seq. *PLoS One* 2015; 10: e0135817.
- 119 Ma KY, Schonnesen AA, Brock A, *et al.* Single-cell RNA sequencing of lung adenocarcinoma reveals heterogeneity of immune response-related genes. *JCI Insight* 2019; 4: e121387.
- 120 Damiani C, Maspero D, Di Filippo M, *et al.* Integration of single-cell RNA-seq data into population models to characterize cancer metabolism. *PLoS Comput Biol* 2019; 15: e1006733.

- 121 Zilionis R, Engblom C, Pfirschke C, *et al.* Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species. *Immunity* 2019; 50: 1317–1334.
- 122 Patel AP, Tirosh I, Trombetta JJ, *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 2014; 344: 1396–1401.
- 123 Chung W, Eum HH, Lee HO, *et al.* Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun* 2017; 8: 15081.
- 124 Azizi E, Carr AJ, Plitas G, *et al.* Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* 2018; 174: 1293–1308.
- 125 Bagnato G, Harari S. Cellular interactions in the pathogenesis of interstitial lung diseases. *Eur Respir Rev* 2015; 24: 102–114.
- 126 Adams TS, Schupp JC, Poli S, *et al.* Single cell RNA-seq reveals ectopic and aberrant lung resident cell populations in idiopathic pulmonary fibrosis. *bioRxiv* 2019; preprint [https://doi.org/10.1101/759902].
- 127 Morse C, Tabib T, Sembrat J, *et al.* Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur Respir J* 2019; 54: 1802441.
- 128 Division of Pulmonary and Critical Care Medicine Feinberg School of Medicine Northwestern University. UCSC Cell Browser Date last accessed: June 11, 2020. Date last updated: June 03, 2020. www.nupulmonary.org/resources/.
- 129 Lung Aging Atlas. https://theislab.github.io/LungAgingAtlas/.
- 130 IPF Cell Atlas Date last accessed: June 11, 2020. Date last updated: June 08, 2020. http://ipfcellatlas.com/.
- 131 Strunz M, Simon LM, Ansari M, *et al.* Longitudinal single cell transcriptomics reveals Krt8⁺ alveolar epithelial progenitors in lung regeneration. *bioRxiv* 2019; preprint [https://doi.org/10.1101/705244].
- 132 Mould KJ, Barthel L, Mohning MP, *et al.* Cell origin dictates programming of resident *versus* recruited macrophages during acute lung injury. *Am J Respir Cell Mol Biol* 2017; 57: 294–306.
- 133 Mould KJ, Jackson ND, Henson PM, *et al.* Single cell RNA sequencing identifies unique inflammatory airspace macrophage subsets. *JCI Insight* 2019; 4: e126556.
- 134 Riemondy KA, Jansing NL, Jiang P, *et al.* Single cell RNA sequencing identifies TGFβ as a key regenerative cue following LPS-induced lung injury. *JCI Insight* 2019; 5: e123637.
- 135 Tibbitt CA, Stark JM, Martens L, *et al.* Single-cell RNA sequencing of the T helper cell response to house dust mites defines a distinct gene expression signature in airway Th2 cells. *Immunity* 2019; 51: 169–184.
- 136 Kormann MSD, Dewerth A, Eichner F, *et al.* Transcriptomic profile of cystic fibrosis patients identifies type I interferon response and ribosomal stalk proteins as potential modifiers of disease severity. *PLoS One* 2017; 12: e0183526.
- 137 Zoso A, Sofoluwe A, Bacchetta M, *et al.* Transcriptomic profile of cystic fibrosis airway epithelial cells undergoing repair. *Sci Data* 2019; 6: 240.
- 138 Kusko RL, Brothers JF 2nd, Tedrow J, *et al.* Integrated genomics reveals convergent transcriptomic networks underlying chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016; 194: 948–960.
- 139 Morrow JD, Chase RP, Parker MM, *et al.* RNA-sequencing across three matched tissues reveals shared and tissue-specific gene expression and pathway signatures of COPD. *Respir Res* 2019; 20: 65.
- 140 Rhodes CJ, Im H, Cao A, *et al.* RNA sequencing analysis detection of a novel pathway of endothelial dysfunction in pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2015; 192: 356–366.
- 141 Hoffmann J, Wilhelm J, Marsh LM, *et al.* Distinct differences in gene expression patterns in pulmonary arteries of patients with chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis with pulmonary hypertension. *Am J Respir Crit Care Med* 2014; 190: 98–111.
- 142 Kim SY, Diggans J, Pankratz D, *et al.* Classification of usual interstitial pneumonia in patients with interstitial lung disease: assessment of a machine learning approach using high-dimensional transcriptional data. *Lancet Respir Med* 2015; 3: 473–482.
- 143 Pankratz DG, Choi Y, Imtiaz U, *et al.* Usual interstitial pneumonia can be detected in transbronchial biopsies using machine learning. *Ann Am Thorac Soc* 2017; 14: 1646–1654.
- 144 Raghu G, Flaherty KR, Lederer DJ, *et al.* Use of a molecular classifier to identify usual interstitial pneumonia in conventional transbronchial lung biopsy samples: a prospective validation study. *Lancet Respir Med* 2019; 7: 487–496.
- 145 Whitney DH, Elashoff MR, Porta-Smith K, *et al.* Derivation of a bronchial genomic classifier for lung cancer in a prospective study of patients undergoing diagnostic bronchoscopy. *BMC Med Genomics* 2015; 8: 18.
- 146 Silvestri GA, Vachani A, Whitney D, *et al.* A bronchial genomic classifier for the diagnostic evaluation of lung cancer. *N Engl J Med* 2015; 373: 243–251.
- 147 Aegis Study Team. Shared gene expression alterations in nasal and bronchial epithelium for lung cancer detection. *J Natl Cancer Inst* 2017; 109: djw327.