



A pulmonologist's guide to perform and analyse cross-species single lung cell transcriptomics

Peter Pennitz^{1,16}, Holger Kirsten^{2,16}, Vincent D. Friedrich^{2,3}, Emanuel Wyler⁴, Cengiz Goekeri^{1,5}, Benedikt Obermayer⁶, Gitta A. Heinz⁷, Mir-Farzin Mashreghi^{7,8}, Maren Büttner^{9,10}, Jakob Trimpert¹¹, Markus Landthaler^{4,12}, Norbert Suttrop¹, Andreas C. Hocke¹, Stefan Hippenstiel¹, Mario Tönnies¹³, Markus Scholz², Wolfgang M. Kuebler^{14,15}, Martin Witzenrath^{1,15}, Katja Hoenzke¹ and Geraldine Nouailles¹

¹Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Infectious Diseases and Respiratory Medicine, Berlin, Germany. ²University of Leipzig, Institute for Medical Informatics, Statistics, and Epidemiology, Leipzig, Germany. ³Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Leipzig, Germany. ⁴Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin Institute for Medical Systems Biology (BIMSB), Berlin, Germany. ⁵Cyprus International University, Faculty of Medicine, Nicosia, Cyprus. ⁶Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Core Unit Bioinformatics, Berlin, Germany. ⁷Deutsches Rheuma-Forschungszentrum Berlin (DRFZ), A Leibniz Institute, Therapeutic Gene Regulation, Berlin, Germany. ⁸Berlin Institute of Health at Charité – Universitätsmedizin Berlin, BIH Center for Regenerative Therapies (BCRT), Berlin, Germany. ⁹University of Bonn, Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, Bonn, Germany. ¹⁰Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Systems Medicine, Bonn, Germany. ¹¹Freie Universität Berlin, Institute of Virology, Berlin, Germany. ¹²Humboldt-Universität zu Berlin, Institute for Biology, IRI Life Sciences, Berlin, Germany. ¹³HELIOS Clinic Emil von Behring, Department of Pneumology and Department of Thoracic Surgery, Chest Hospital Heckeshorn, Berlin, Germany. ¹⁴Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Physiology, Berlin, Germany. ¹⁵German Center for Lung Research (DZL), Berlin, Germany. ¹⁶Both authors contributed equally to this work.

Corresponding author: Geraldine Nouailles (geraldine.nouailles@charite.de)



Shareable abstract (@ERSpublications)

The COVID-19 pandemic led to an increase in publicly available single-cell RNA sequencing data. This review provides an up-to-date framework and readily adoptable tools to measure such data in lungs and compare it with existing data across species. <https://bit.ly/3wHCoHe>

Cite this article as: Pennitz P, Kirsten H, Friedrich VD, *et al.* A pulmonologist's guide to perform and analyse cross-species single lung cell transcriptomics. *Eur Respir Rev* 2022; 31: 220056 [DOI: 10.1183/16000617.0056-2022].

Copyright ©The authors 2022

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org

Received: 28 March 2022
Accepted: 16 May 2022

Abstract

Single-cell ribonucleic acid sequencing is becoming widely employed to study biological processes at a novel resolution depth. The ability to analyse transcriptomes of multiple heterogeneous cell types in parallel is especially valuable for cell-focused lung research where a variety of resident and recruited cells are essential for maintaining organ functionality. We compared the single-cell transcriptomes from publicly available and unpublished datasets of the lungs in six different species: human (*Homo sapiens*), African green monkey (*Chlorocebus sabaeus*), pig (*Sus domesticus*), hamster (*Mesocricetus auratus*), rat (*Rattus norvegicus*) and mouse (*Mus musculus*) by employing RNA velocity and intercellular communication based on ligand–receptor co-expression, among other techniques. Specifically, we demonstrated a workflow for interspecies data integration, applied a single unified gene nomenclature, performed cell-specific clustering and identified marker genes for each species. Overall, integrative approaches combining newly sequenced as well as publicly available datasets could help identify species-specific transcriptomic signatures in both healthy and diseased lung tissue and select appropriate models for future respiratory research.

Introduction

Single-cell RNA sequencing (scRNA-seq) is becoming widely employed to study biological processes in great depth [1–3]. Its name describes its key advantage: the resolution of transcriptomes at single-cell level. This resolution is highly valuable when sample sources are heterogeneous in their cellular compositions, and during the rapid temporal developments occurring during responses to acute infection, such as the



increasing cellular intervariability. Thus, scRNA-seq is an essential tool to identify interactive networks, cell fate, disease signatures and novel cell types.

Initially, scRNA-seq was primarily utilised in studies involving oocytes, zygotes and other early developmental cells [4, 5]. Later, global initiatives such as the Human Cell Atlas and CZ Biohub provided the murine [6] and human [7, 8] scRNA-seq atlases. In cell biology focused research, scRNA-seq enabled the discovery of formerly unknown pulmonary cells such as ionocytes [9], aerocytes [10] and group 2 innate lymphoid cells subpopulations [11]. Furthermore, the identification of disease states in individual pulmonary fibrosis derived cells led to the human pulmonary fibrosis atlas, serving as a groundwork for personalised therapies [12]. scRNA-seq was effectively adopted in cancer research, aiming to identify individualised treatment-targets and improve patient care [13–15]. Recently, the severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) pandemic further boosted applications of scRNA-seq in lungs due to the need to understand cellular pathomechanisms underlying coronavirus disease 2019 (COVID-19). Here, scRNA-seq became the method of choice to dissect the antiviral immune response both in patient samples and model organisms including transgenic mice, hamsters, ferrets and nonhuman primates [16–21]. Several possible markers predicting severity of COVID-19 cases [22] and potential drug targets [23] were also identified.

From this perspective, pandemic-driven accumulation of substantial lung-centred data allows for the comparison of pulmonary immunity across species and the definition of the advantages and shortcomings of animal models. This requires digital platforms and tools to integrate and compare data derived from human and nonhuman subjects to further support translational research towards new forms of medical diagnosis and (personalised) treatment.

Our main aim is to provide guidance and demonstrate technical feasibility of currently utilised powerful methods for pulmonologists interested in the single-cell transcriptome of lung tissue across different species. For this, we demonstrate current procedures and tools to create, compare and qualitatively assess scRNA-seq data in six different species: human (*Homo sapiens*), African green monkey (*Chlorocebus sabaeus*), pig (*Sus domesticus*), hamster (*Mesocricetus auratus*), rat (*Rattus norvegicus*) and mouse (*Mus musculus*). We show similarities regarding cell differentiation and identify a prominent role of proliferating alveolar macrophages in cell–cell communication.

Basic principles of pulmonary single-cell isolation, barcoding and sequencing

Lungs harbour a variety of resident and recruited cells, which participate in immune responses and are essential for maintaining organ functionality. Ideally, analyses of lung cells should be performed in full complexity to improve the understanding of various pulmonary pathologies. However, subsequent cross-species cell isolation protocols are rarely identical. Therefore, their potential impact on cellular representation and fitness should be taken into consideration. However, different protocols do not hinder cross-species comparisons, as demonstrated here for selected species and protocol examples (protocols corresponding to datasets depicted in table 1).

Classical droplet-based scRNA-seq technologies such as Drop-seq, In-Drop, Chromium (10× Genomics, Pleasanton, CA, USA) and Nadia (Dolomite Bio, Royston, UK) are widely used due to their high throughput and reasonable pricing. Although recent approaches in bulk RNA and scRNA-seq applications of cryopreserved cells showed promising results for lung epithelial cells, using viable cells from fresh tissue remains the recommended option by the manufacturers (10× Genomics, Dolomite Bio) of commercially available droplet-based scRNA-seq workflows [26, 27]. In plate-based scRNA-seq technologies such as SMART-seq2, single cells are conventionally sorted by flow cytometry into wells in order to overcome the viability shortcoming, as dead cells can be removed during sorting. Another advantage of the SMART-seq2 is the deeper transcriptomic information that can be drawn, much like other full-length RNA-seq methods [28]. The disadvantages include the low number of cells that can be analysed (96 or 384 cells per plate, per run) and significantly higher costs per cell, making it very challenging to utilise for whole organs, especially those that are heterogeneous in structure such as the lungs. The BD Rhapsody system enables high-throughput analyses of the whole transcriptome and adopts a microwell-based approach, resulting in fewer doublet cells compared to the droplet-based 10× Genomics system, albeit with the drawback of allowing the analysis of just one sample per assay [29, 30], an issue that could be solved by additional antibody tagging (cell-hashing).

Ideally, the single-cell suspension, regardless of the method chosen, represents all pulmonary cells in their respective proportion and all original gene signatures. Evidently, cellular dissociation of solid tissues preferably selects migratory cells, such as leukocytes [31]. To dissociate pulmonary mesenchymal,

TABLE 1 Overview of experimental protocols applied to generate viable pulmonary single-cell suspensions; cell isolation procedures corresponding to the interspecies datasets to dissociate lung tissue into a single-cell suspension. Enzymatic digestion and mechanical processing steps applied to dissociate lung tissue into a single-cell suspension vary between species.

	Human lung tissue (Charité)	Murine lung tissue (Charité)	Hamster lung tissue (NOUAILLES <i>et al.</i> [18])	Human lung tissue (TRAVAGLINI <i>et al.</i> [24])	Monkey lung tissue (SPERANZA <i>et al.</i> [21])	Rat lung tissue (RAREDON <i>et al.</i> [25])	Pig lung tissue (RAREDON <i>et al.</i> [25])
Treatment		Mock infection			Mock infection		Heparinisation
1st enzymatic digestion		Intra-alveolar vasculature				Intra-alveolar vasculature	Intra-alveolar
Enzymes		Dispase (50 U·mL ⁻¹)				Collagenase/dispase (1 mg·mL ⁻¹) elastase (3 U·mL ⁻¹) DNase (20 U·mL ⁻¹)	Collagenase/dispase (1 mg·mL ⁻¹) elastase (3 U·mL ⁻¹) DNase (20 U·mL ⁻¹)
Temperature		22°C				37°C	37°C
Duration		5 min					
1st mechanical dissociation	Mincing with scissors	Clapping with tweezers	Clapping with tweezers	Dissection and mincing		Dissociation in rocker	Dissociation in rocker
2nd enzymatic digestion	Distal alveolar tissue	Full lung tissue	Lobus caudalis	Distal alveolar tissue	Lung sections	Full lung tissue	Distal lung regions
Enzymes	Collagenase I (500 U·mL ⁻¹) DNase I (1 U·mL ⁻¹) dispase (5 U·mL ⁻¹)	Collagenase CLS II (750 U·mL ⁻¹) DNase I (1 mg·mL ⁻¹)	Collagenase CLS II (750 U·mL ⁻¹) DNase I (1 mg·mL ⁻¹) Dispase (50 U·mL ⁻¹)	Liberase DL (400 µg·mL ⁻¹) elastase (100 µg·mL ⁻¹)	Liberase (0.1 mg·mL ⁻¹) DNase I (0.02 mg·mL ⁻¹)	Collagenase/dispase (1 mg·mL ⁻¹) elastase (3 U·mL ⁻¹) DNase (20 U·mL ⁻¹)	Collagenase/dispase (1 mg·mL ⁻¹) elastase (3 U·mL ⁻¹) DNase (20 U·mL ⁻¹)
Temperature	37°C	37°C	37°C	37°C	37°C	37°C	37°C
Duration	60 min	30 min	30 min	30 min	NR	20 min	40 min
2nd mechanical dissociation		Shear force <i>via</i> syringe	Shear force <i>via</i> pipette	GentleMACS dissociator		Pressed through sieve [#]	Pressed through sieve [#]
2nd incubation				30 min nutator at 37°C			
3rd dissociation				GentleMACS dissociator +DNase I (100 µg·mL ⁻¹)			
1st cell filtration	70 µm cell strainer	70 µm cell strainer	70 µm cell strainer	100 µm cell strainer	100 µm cell strainer	70 µm cell strainer	70 µm cell strainer
Red blood cell lysis	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2nd cell filtration	40 µm cell strainer	70 µm cell strainer	40 µm cell strainer	70 µm cell strainer		40 µm cell strainer	40 µm cell strainer
Isolation buffer	RPMI, 10% FCS, 1% glutamine	1× PBS, 0.5% BSA	1× PBS, 0.5% BSA	1× PBS, 5% FBS	1× PBS, 0.1% BSA	1× PBS, 0.01% BSA	1× PBS, 0.01% BSA
Centrifugation force	360×g (10 min, 4°C)	300×g (6 min, 4°C)	350×g (6 min, 4°C)	300×g (5 min, 4°C)	NR	300×g (3 and 5 min)	300×g (3 and 5 min)
Single-cell processing	10× Genomics	10× Genomics	10× Genomics	10× Genomics/ SMART-seq2	10× Genomics	10× Genomics	10× Genomics

DL: dispase low; FCS: fetal calf serum; BSA: bovine serum albumin; FBS: fetal bovine serum; NR: not reported. #: the sieve was additionally rinsed with DMEM containing 10% FCS, 1% penicillin/streptomycin, 1% amphotericin and 0.1% gentamicin.

endothelial and epithelial cells, use of enzymatic digestion and enforcement of mechanical stress is necessary [32]. However, these steps increase risks of inducing transcriptomic stress responses that overwrite the cells' original transcriptomic signature [33–35]. A preventive measure is to add reagents during isolation steps to stop *de novo* transcription, such as dithio-bis(succinimidyl propionate) [36], CellCover reagent [37] or Actinomycin D [38]. Alternatively, cold-active proteases can be used at low temperatures to hinder ongoing transcription [33]. Another option is single-nucleus RNA sequencing (snRNA-seq), whereby nuclei, rather than viable cells, are isolated from tissues. Advantages of snRNA-seq include the well-established applicability to frozen tissue, minimal isolation-induced cellular stress and improved representation of tissue cells, albeit at the cost of diminished leukocyte frequencies [39, 40]. Particularly, the enhanced representation of pulmonary epithelial and endothelial cells, without need to undergo prior single-cell dissociation steps, makes this technique highly relevant to lung research focusing on tissue responses, *e.g.* lung barrier integrity in acute lung injury models. Furthermore, snRNA-seq enables the generation of large-scale reference maps from archived tissues as demonstrated by the GTEx Consortium [41]. In autopsy samples from lung, heart, liver and kidney tissues, snRNA-seq outperformed scRNA-seq to systematically access the cellular ecosystem of the tissues [42]. A recent publication comparing scRNA-seq and snRNA-seq data of fresh and cryopreserved murine lung tissue could demonstrate comparable detection rate of genes as well as lower expression of mitochondrial and ribosomal genes, but increased ambient RNA contamination which could not be corrected with the soupX package [40].

Ready-to-use protocols for both human and murine lung tissues are made publicly available, specialising on the isolation of cell nuclei from healthy or diseased (*e.g.* tumorous) tissues [40, 43, 44].

Once a single-cell suspension has been obtained, the individual cell's RNA can be barcoded, reverse-transcribed and cDNA libraries suitable for next-generation sequencing can be created and sequenced. Each cell's individual barcode enables subsequent reassignment of identified RNA sequences to its cell of origin. For basic analysis allowing for identification of individual cell populations and their specific gene expression, usually 50 000 raw reads per cell are recommended when counting unique molecular identifiers (UMIs), *i.e.* the number of sequences before PCR-amplification. Enhanced resolution and correct identification of rare transcripts can be achieved through deeper coverage in RNA-seq [45].

Data processing and analysis

Bioinformatic tools and workflows for processing transcriptomic single-cell data are steadily improving in their scope and usability [46, 47]. The first step in the analysis of each scRNA-seq dataset involves the creation of a table reporting the counts of all detected genes for each measured cell. In the 10× Genomics Chromium system, the vendor's Cell Ranger software is often used. For human and mouse data, the necessary annotation files are publicly available, *e.g.* by the Encyclopedia of DNA Elements (ENCODE) project [48, 49]; however, it might be necessary for these to be generated from sequencing data for other species. For this, gene annotation of related species can be used to identify genes that also match in the *de novo* sequenced data as described previously [50]. In the following steps, software tools written in R and Python dominate the field. In R, Seurat [51] and specific packages from Bioconductor [52] are pre-eminent, while in Python, SCANPY [53] is used frequently. Initial filtering of low-quality data is crucial for high-quality results. Such filters ensure a minimum number of unique measured genes and gene counts per cell and remove cells with high percentages of mitochondrial genes, which is an indication of defective cells [46]. Identified doublets (droplets that have erroneously captured more than one cell) are also often removed.

Exemplary workflow for multispecies data pre-processing and integration

In the following, we demonstrate integration and qualitative assessment of scRNA-seq data across six species from eight datasets: human (*Homo sapiens*), African green monkey (*Chlorocebus sabaeus*), pig (*Sus domesticus*), hamster (*Mesocricetus auratus*), rat (*Rattus norvegicus*) and mouse (*Mus musculus*). We used quality-filtered scRNA-seq data of lungs from different sources (figure 1) using up-to-date analysis workflows. Six datasets are publicly available: resected lung tissue from one patient processed by 1) 10× Genomics (human, TRAVAGLINI *et al.* [24], patient 2) and 2) SMART-seq2 workflow (human, TRAVAGLINI *et al.* [24], patient 2); 3) three hamster lungs (naïve controls from [18]); 4) two rat lungs (data from [25]); 5) two pig lungs (data from [25]); and 6) two monkey lungs (mock-infected controls from [21]). Two further datasets were generated in-house and made available with this article: 7) resected patient lung tissue of four patients (human Charité) and 8) two mouse lungs pooled prior to barcoding (https://github.com/GenStatLeipzig/pulmonologists_interspecies_scRNA). All tissue dissociation procedures for the aforementioned datasets are summarised in table 1.

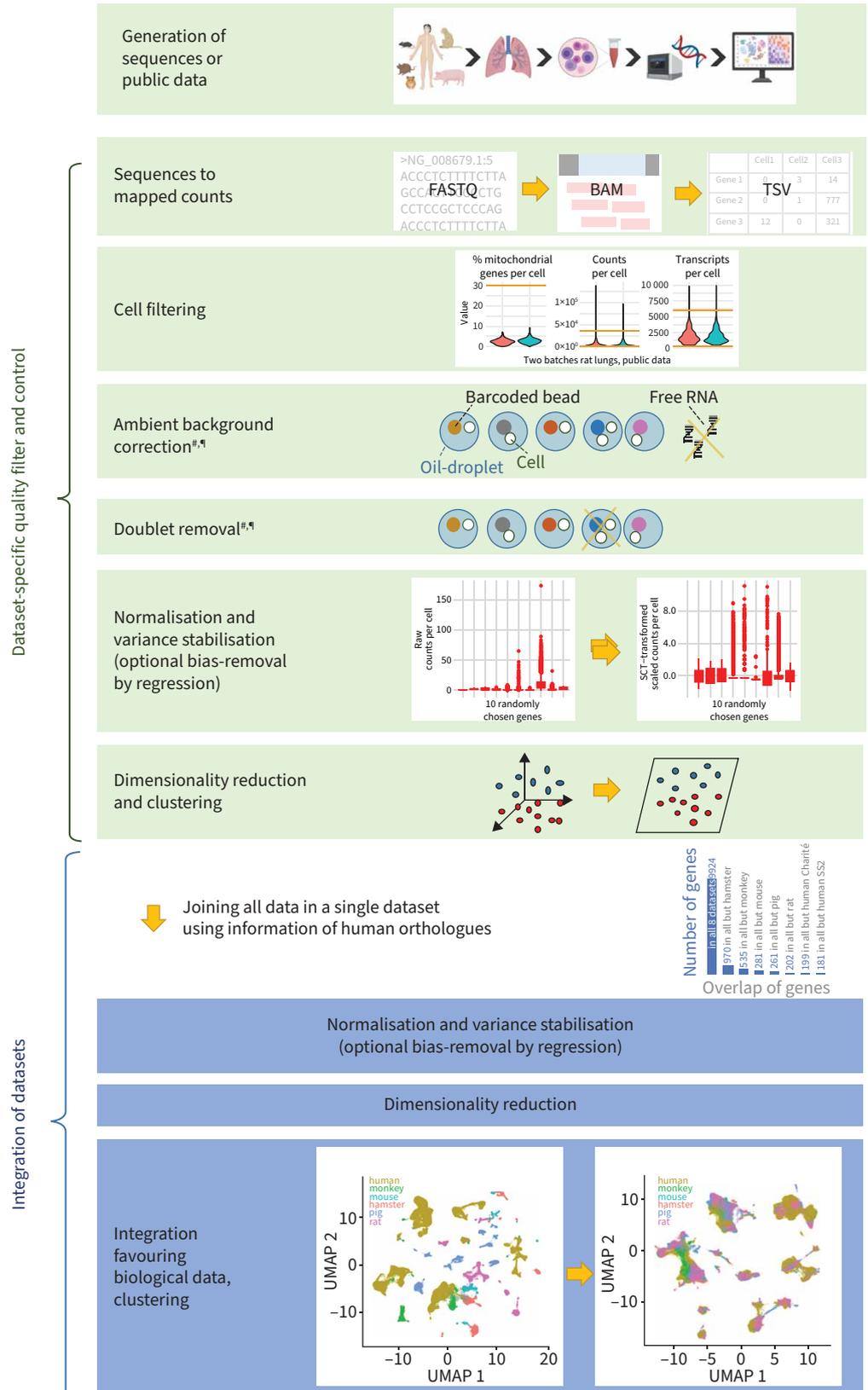


FIGURE 1 Integration and bioinformatic pre-processing of sequenced lung cells. Schematic depicts single-cell RNA sequencing workflow, created with Biorender.com. FASTQ: text-based format for storing sequence and corresponding quality data; BAM: binary alignment map file format; TSV: tab-separated value file format. #: may require a first round of processing for preliminary clustering; ¶: not for Smart Seq2 data.

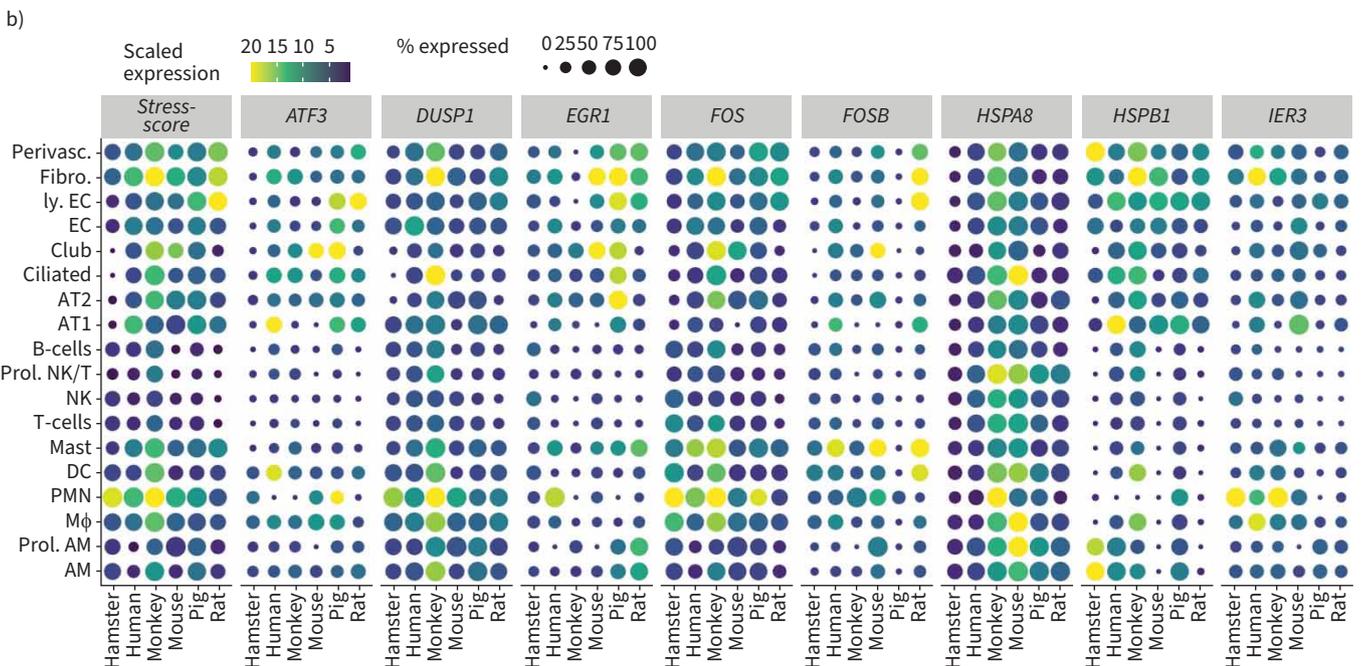
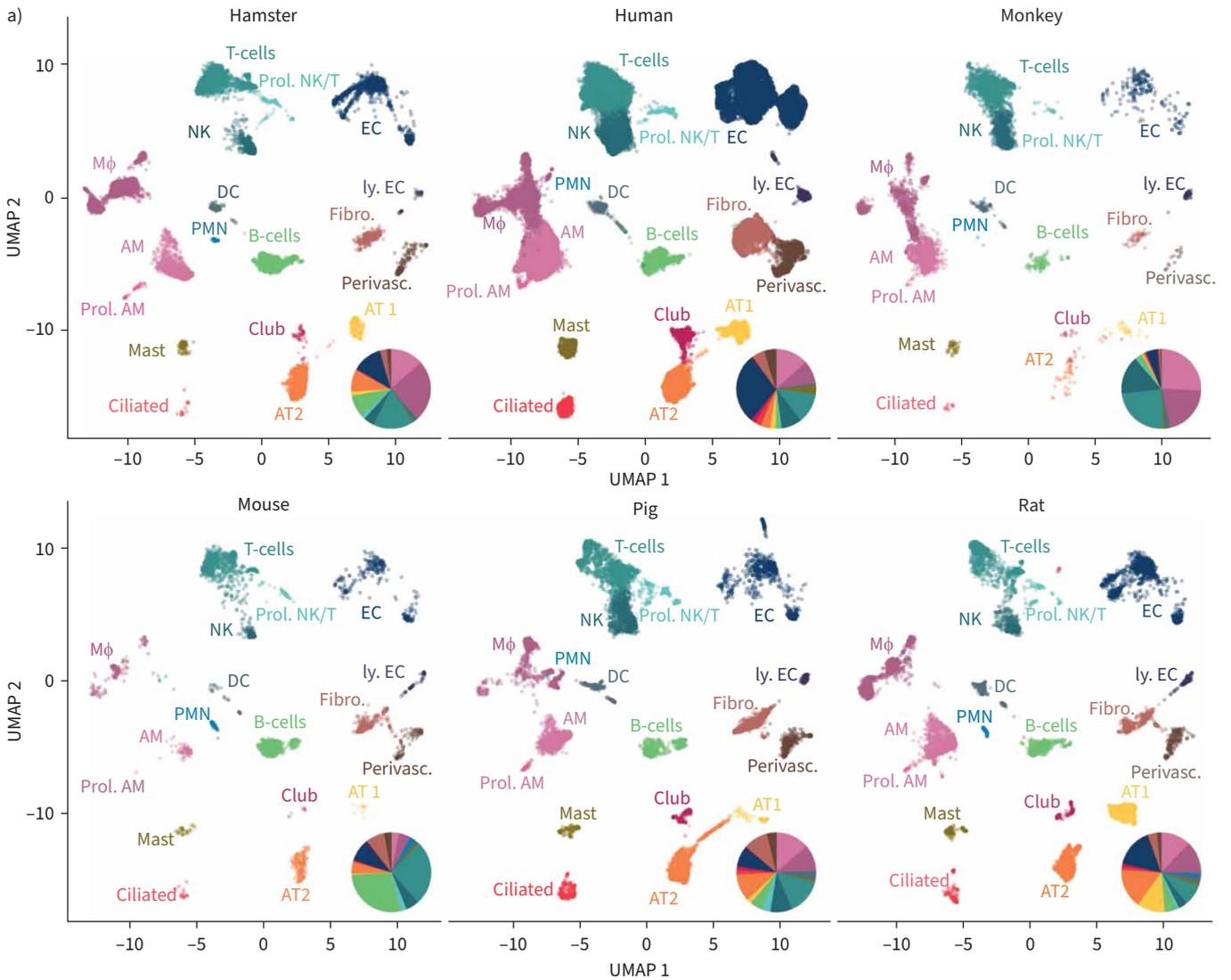


FIGURE 2 Cross-species comparison of sequenced lung cells. **a)** Uniform manifold approximation and projection (UMAP) plot of identified cell populations across species including all datasets, with pie charts of lung cell frequencies in indicated species estimated from single-cell ribonucleic acid sequencing (scRNA-seq) data; **b)** dot plots indicating the stress response of lung cells across species for all unique molecular identifier based datasets using eight stress-related genes: *HSPA8*, *FOS*, *DUSP1*, *IER3*, *EGR1*, *FOSB*, *HSPB1* and *ATF3*. AM: alveolar macrophages; prol. AM: proliferating alveolar macrophages; M ϕ : interstitial macrophages/monocytes; PMN: polymorphonuclear leukocytes; DC: dendritic cells; mast: mast cells; NK: natural killer cells; prol. NK/T: proliferating NK- and T-cells; AT1: alveolar epithelial cells type 1; AT2: alveolar epithelial cells type 2; ciliated: ciliated cells; club: club cells; EC: endothelial cells; ly. EC: lymphatic endothelial cells; fibro: fibroblasts; perivasc: perivascular cells. Samples: human Charité, cells from four fresh lung explants; human TRAVAGLINI *et al.* [24], data of patient 2; mouse, cells of two whole lungs pooled prior analysis; hamster, cells from lobus caudalis of three hamsters, data from [18]; monkey cells from two lungs, data from [21]; rat and pig, cells from two lungs each, data from [25]. All experiments involving animals were approved by institutional and governmental authorities (Charité Universitätsmedizin Berlin or Freie Universität Berlin and LaGeSo Landesamt für Gesundheit und Soziales Berlin, Germany). Human Charité dataset biomaterial was obtained following approval of the Charité ethics commission, application ID: EA2/079/13, while written informed consent was obtained from all patients.

As described in the first landmark paper demonstrating cross-species comparison within pulmonary cell types [25], a single unified gene nomenclature is required for subsequent steps. For this, we identified the 9924 human genes with orthologues present in all species from the Ensembl database human genes build GRCh38.p13 (<http://www.ensembl.org>; supplementary figure S1). In the instance of multiple genes relating to a single orthologue, a consensus gene was selected based on the highest cell-specific expression level. In the following, we refer to the orthologue genes by using the human gene name.

Here, the datasets were similarly transformed by defining common cut-offs for high quality cells, performing regression-based adjustment of the percentage of mitochondrial transcripts (if available), excluding doublets *via* R-package DoubletFinder [54] and removing ambient RNA *via* R-package DecontX [55] (figure 1). For the SMART-seq2 dataset, the latter two steps are not required, and a higher cut-off for the minimal number of counts applies as counts here do not reflect UMIs. For normalisation, data was split according to each individual batch and processed with the SCTransform workflow of Seurat making the data more comparable across experiments and species. The SCTransform() function was introduced in 2019 as an improved means of normalisation of scRNA-seq data, based on regularised negative binomial regression [56] and described for integration across different technologies. Applying SCTransform replaces three functions of the previous Seurat workflow: NormalizeData(), ScaleData() and FindVariableFeatures(), we used the version implemented in Seurat_4.0.5. The combining of datasets to create an inter-species data file was performed using Seurat's merge() function, thereby restricting all datasets to genes present in each species. Next, cell-specific clusters with similar expression patterns should be identified. Here, dimension reduction was performed by principal components analysis (PCA), a dimensionality reduction technique to emphasise strong patterns in a dataset considering variation and similarity. R-package Harmony was employed to project the PCA data into a shared embedding [57] to successfully complete the data integration process. The algorithm directs cells to group by cell type rather than according to dataset-specific conditions. As one of many integration tools for scRNA-seq data, Harmony is particularly convincing in terms of usability and was also ranked among top-scoring method [58, 59]. While other top-scoring integration tools are Python-based (scANVI, Scanorama), a second R-based option would be FastMNN or the Seurat IntegrateData() approach applying dimensional reduction to find anchors *via* reciprocal PCA [59]. However, the method of choice may differ depending on the dataset to be analysed, as exemplified by the systematic comparison in [59]. A comparison between Harmony and Seurat's 4.0.5 function IntegrateData() applying reciprocal PCA is shown in supplementary figure S5, quantification of the quality of integration can be done with measures like integration Local Inverse Simpson's Index (LISI) or cell-type LISI [57].

To visualise the resulting cell types, while taking into consideration the multidimensional space in a clustered two-dimensional plot, uniform manifold approximation and projection (UMAP) was used. A key advantage of UMAP is its ability to map cells according to global differences compared to the previous method of choice (t-distributed stochastic neighbour embedding), which only reflects local differences. Furthermore, it is fast and can be scalable to large numbers of cells. Finally, the number of desired clusters was set by tuning the resolution parameter in Louvain clustering (a graph-based method for community detection that aims to partition a network into clusters of strongly correlated cells) while cells belonging to different clusters were weakly correlated [60]. Advantages of Louvain clustering are speed and scalability as well as the lack of necessity to pre-define the number of clusters [61]. In our analysis, we set the resolution parameter to 0.3, resulting in 22 clusters shared among species. These clusters representing accumulation of cells with similar transcriptomic profiles were annotated to a certain cell type, based on

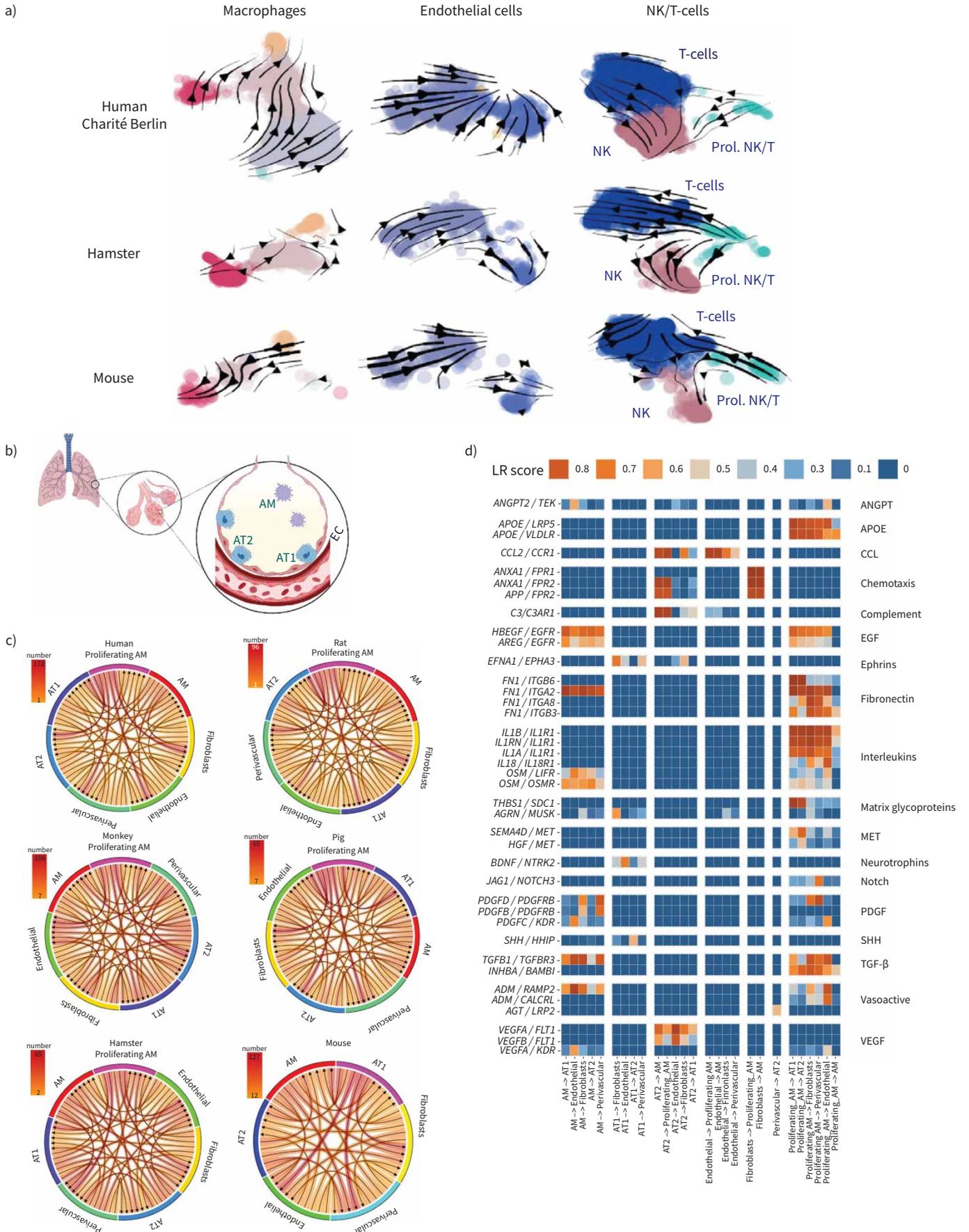


FIGURE 3 RNA velocity in pulmonary cells and alveolar niche communication. **a)** Uniform manifold approximation and projection (UMAP) plots depicting human, hamster and mouse lung cells overlaid with stream arrows derived from the RNA velocity analysis. The colouration indicates chosen cell clusters. **b–d)** Intercellular communication estimated through ligand–receptor co-expression in the indicated cell types: **b)** scheme of the healthy alveolus, created with Biorender.com; **c)** count of relevant receptor–ligand interaction (LRscore ≥ 0.5) (indicated by arrows) in all species for all unique molecular identifier based datasets. In humans and several animals, proliferating alveolar macrophages appear to show increased intercellular communication. **d)** Human ligand–receptor pairs found to be conserved (detected in at least one additional nonhuman species) based on the classification of RAREDON *et al.* [25]. Ligand–receptor pairs found only in humans are shown in supplementary figure S4. NK: natural killer cells; prol. NK/T: proliferating NK and T-cells; AM: alveolar macrophages; AT1: alveolar epithelial cells type 1; AT2: alveolar epithelial cells type 2; EC: endothelial cells; ANGPT: angiopoietin; APOE: apolipoprotein E; CCL: C-C motif chemokine ligand; EGF: epidermal growth factor; PDGF: platelet-derived growth factor; SHH: Sonic Hedgehog; TGF: transforming growth factor; VEGF: vascular endothelial growth factor.

known marker genes and cell-specific annotations transferred from the original dataset. Cell frequencies of both haematopoietic and structural lung cells differ between compared species. While certain differences between species have been described before [62], cell isolation protocols, especially using different enzymatic cocktails, can have a major impact on retrieved cell types. Interspecies research aiming to resolve cell quantitative in addition to qualitative questions should ideally incorporate spatial transcriptomics analyses of tissue sections (figure 2a, supplementary figure S2).

Cross-species analysis of human and nonhuman scRNA-seq data

To demonstrate the quality and comparability of data across species, we investigated the expression of eight reputed stress-related genes [34] to monitor putative cell dissociation-associated stress responses. A stress score was developed by averaging cell-wise expression across these genes (figure 2b). Across all analysed species, the applied cell isolation and analysis strategy led to a suitable stress gene signature, with somewhat higher levels for in human lung-derived cells.

As raw sequencing data files were required, we computed RNA velocity (figure 3a, supplementary figure S3) for the in-house generated datasets only [63]. Here, cell development dynamics and temporal expression information were recovered by comparing spliced and unspliced levels of mRNA with an estimated steady-state. The arrows in the two-dimensional UMAP latent space depict an inference of most probable cell state transitions and cell fates based on the transcriptome. The arrow size indicates the coherence and speed of cell development processes in the corresponding region of the UMAP latent space. For velocity estimation, the Python package scVelo was applied with default stochastic model for transcriptional dynamics [64]. By graphical inspection, the most coherent cell movement across the analysed species can be observed in endothelial and T-cell clusters. Interestingly, for human clusters of macrophages and alveolar macrophages, the derived velocities indicate a common area in the UMAP latent space to which the corresponding cells seem to develop. As RNA velocity infers the most probable cell fate solely based on the transcriptome, the compatibility with biological foundations should be verified.

Intercellular communication based on ligand–receptor co-expression was calculated based on 3251 recently published pairs available through package SingleCellSignalR [25]. We limited cell-communication analysis to UMI-based datasets available for all species. Through analysed species, proliferating alveolar macrophages seem to dominate ligand receptor interaction within the alveolar spaces (figure 3c and d, supplementary figure S4). All cross-species analysis codes and data, including expression analysis of canonical cell type-specific genes, was made available *via* Github (https://github.com/GenStatLeipzig/pulmonologists_interspecies_scRNA).

A major limitation is the poor genome annotation for species other than *Mus musculus* and *Homo sapiens*, exemplified by that of *Mesocricetus auratus* (MesAur1.0), which lacks orthologues in ~25% of all human genes. Integration of further tools into cross-species comparisons, including spatial transcriptomics [65] and new packages from the constantly growing pool of software applications designed for scRNA-Seq data, will further improve our understanding of similarities and differences between cell states and cellular subtypes across species.

Outlook

The meteoric rise of available infection-related patient- and animal model-derived scRNA-seq datasets in response to the COVID-19 pandemic allows for in-depth cross-species comparisons of systemic and pulmonary transcriptomic responses to acute infection. Moreover, such comparisons bare the potential to systematically identify opportunities and limitations of current animal models. Exploiting this unravelled

treasure will allow respiratory researchers and pulmonologists to identify universal and species-specific transcriptional responses to pathogens including SARS-CoV-2 and others.

Provenance: Submitted article, peer reviewed.

Acknowledgements: The authors would like to thank Ulrike Behrendt (Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Infectious Diseases and Respiratory Medicine, Berlin, Germany) for excellent technical assistance. We acknowledge financial support from the Open Access Publication Fund of Charité – Universitätsmedizin Berlin and the German Research Foundation (DFG).

Conflict of interest: H. Kirsten reports support for the present manuscript from German Federal Ministry of Education and Research (BMBF) grants e:Med CAPSyS (01ZX1304A) and e:Med SYMPATH (01ZX1906B). E. Wyler has received payment or honoraria for lectures, presentations, speakers' bureaus, manuscript writing or educational events from Podcast Gegenblende by DGB, outside the submitted work. M. Landthaler reports support for the present manuscript from Berlin Institute of Health. M. Scholz has received grants or contracts from Pfizer Inc. for a project not related to this research. W.M. Kuebler reports support for the present manuscript from German Research Foundation (KU 1218/9-1, KU 1218/11-1, CRC TR84 A02, CRC TR84 C09, CRC 1449 B01), Germany Ministry for Research and Education (SYMPATH, PROVID consortia), German Center for Cardiovascular Research (Partner site project Berlin) and Berlin Institute of Health (Focus Area Vascular Biology). M. Witzentrath has received grants or contracts from Deutsche Forschungsgemeinschaft, Bundesministerium für Bildung und Forschung, Deutsche Gesellschaft für Pneumologie, European Respiratory Society, Marie Curie Foundation, Else Kröner Fresenius Stiftung, Capnetz Stiftung, International Max Planck Research School, Quark Pharma, Takeda Pharma, Noxxon, Pantherna, Silence Therapeutics, Vaxxilon, Actelion, Bayer Health Care, Biotest and Boehringer Ingelheim, outside the submitted work. M. Witzentrath has received personal fees for consulting from Noxxon, Pantherna, Silence Therapeutics, Vaxxilon, Aptarion, GlaxoSmithKline, Sinoxa and Biotest, and for lectures, presentations, speakers' bureaus, manuscript writing or educational events from AstraZeneca, Berlin Chemie, Chiesi, Novartis, Teva, Actelion, Boehringer Ingelheim, GlaxoSmithKline, Biotest and Bayer Health Care, outside the submitted work. M. Witzentrath has the following patents planned, issued or pending: EPO 12181535.1: IL-27 for modulation of immune response in acute lung injury (issued: 2012); WO/2010/094491: Means for inhibiting the expression of Ang-2 (issued: 2010); and DE 102020116249.9: Camostat/Niclosamide cotreatment in SARS-CoV-2 infected human lung cells (issued: 2020/21). G. Nouailles receives funding from Biotest AG for a project not related to this work. All other authors have nothing to disclose.

Support statement: G. Nouailles and M. Witzentrath are supported by the German Federal Ministry of Education and Research (BMBF) and the Agence nationale de la recherche (ANR) in the framework of MAPVAP (16GW0247). M. Witzentrath is supported by the German Research Foundation (DFG) grants SFB-TR84 C06, C09, and SFB-1449 B02; the BMBF in the framework of e:Med CAPSyS (01ZX1604B), PROVID (01KI20160A), e:Med SYMPATH (01ZX1906A) and NUM-NAPKON (01KX2021); the BIH (CM-COVID). J. Trimpert is supported by DFG grant number SFB-TR84 Z01b. M-F. Mashreghi receives funding from the State of Berlin and the "European Regional Development Fund" with grant ERDF 2014-2020, EFRE 1.8/11. M-F. Mashreghi is part of the Leibniz Association (Leibniz Collaborative Excellence, TargArt). W.M. Kuebler receives funding from the DFG (SFB-TR84 A02 and C09, SFB 1449 B01, KU1218/9-1 and KU1218/11-1), the BMBF in the framework of e:Med SYMPATH (01ZX1906A), PROVID (01KI20160A), and the German Center for Cardiovascular Research (DZHK) Partner site Berlin. N. Suttorp is supported by DFG (grant number SFB-TR84 B01, C09 and Z02) and BMBF (e:Med CAPSyS grants 01ZX1304B and 01ZX1604B). M. Scholz and H. Kirsten are supported by the BMBF in the framework of e:Med CAPSyS (01ZX1304A), e:Med SYMPATH (01ZX1906B) and V.D. Friedrich within the project "Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig" (01IS18026B). K. Hoenzke was supported by the SFB-TR84 (DFG), "NUM-COVID 19 Organo-Strat" (BMBF), Charité 3R and the "Einstein Center 3R" (Einstein Foundation Berlin). A.C. Hocke and S. Hippenstiel were supported by Berlin University Alliance GC2 Global Health (Corona Virus Pre-Exploration Project), the SFB-TR84 (DFG), RAPID and "NUM-COVID 19 Organo-Strat" (BMBF) as well as the Berlin Institute of Health (BIH), Charité 3R and the "Einstein Center 3R" (Einstein Foundation Berlin). P. Pennitz and C. Goekeri are supported by Charité 3R. Funding information for this article has been deposited with the Crossref Funder Registry.

References

- 1 Cheung P, Khatri P, Utz PJ, *et al.* Single-cell technologies – studying rheumatic diseases one cell at a time. *Nat Rev Rheumatol* 2019; 15: 340–354.
- 2 Paik DT, Cho S, Tian L, *et al.* Single-cell RNA sequencing in cardiovascular development, disease and medicine. *Nat Rev Cardiol* 2020; 17: 457–473.

- 3 Ramachandran P, Matchett KP, Dobie R, *et al.* Single-cell technologies in hepatology: new insights into liver biology and disease pathogenesis. *Nat Rev Gastroenterol Hepatol* 2020; 17: 457–472.
- 4 Tang F, Barbacioru C, Wang Y, *et al.* mRNA-seq whole-transcriptome analysis of a single cell. *Nat Methods* 2009; 6: 377–382.
- 5 Guo G, Huss M, Tong GQ, *et al.* Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell* 2010; 18: 675–685.
- 6 Tabula Muris Consortium. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* 2018; 562: 367–372.
- 7 Regev A, Teichmann SA, Lander ES, *et al.* The Human Cell Atlas. *Elife* 2017; 6: e27041.
- 8 Rozenblatt-Rosen O, Stubbington MJT, Regev A, *et al.* The Human Cell Atlas: from vision to reality. *Nature* 2017; 550: 451–453.
- 9 Montoro DT, Haber AL, Biton M, *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* 2018; 560: 319–324.
- 10 Schupp JC, Adams TS, Cosme C Jr, *et al.* Integrated single-cell atlas of endothelial cells of the human lung. *Circulation* 2021; 144: 286–302.
- 11 Ghaedi M, Shen ZY, Orangi M, *et al.* Single-cell analysis of ROR α tracer mouse lung reveals ILC progenitors and effector ILC2 subsets. *J Exp Med* 2020; 217: jem.20182293.
- 12 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.
- 13 Wang B, Liu M, Ran Z, *et al.* Analysis of gene signatures of tumor microenvironment yields insight into mechanisms of resistance to immunotherapy. *Front Bioeng Biotechnol* 2020; 8: 348.
- 14 Dimitrov-Markov S, Perales-Patón J, Bockorny B, *et al.* Discovery of new targets to control metastasis in pancreatic cancer by single-cell transcriptomics analysis of circulating tumor cells. *Mol Cancer Ther* 2020; 19: 1751–1760.
- 15 Bicač M, Lückerrath K, Kalidindi T, *et al.* Genetic signature of prostate cancer mouse models resistant to optimized hK2 targeted α -particle therapy. *Proc Natl Acad Sci USA* 2020; 117: 15172–15181.
- 16 Su Y, Chen D, Yuan D, *et al.* Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19. *Cell* 2020; 183: 1479–1495.
- 17 Qin Z, Liu F, Blair R, *et al.* Endothelial cell infection and dysfunction, immune activation in severe COVID-19. *Theranostics* 2021; 11: 8076–8091.
- 18 Nouailles G, Wyler E, Pennitz P, *et al.* Temporal omics analysis in Syrian hamsters unravel cellular effector responses to moderate COVID-19. *Nat Commun* 2021; 12: 4869.
- 19 Ragan IK, Hartson LM, Dutt TS, *et al.* A whole virion vaccine for COVID-19 produced *via* a novel inactivation method and preliminary demonstration of efficacy in an animal challenge model. *Vaccines* 2021; 9: 340.
- 20 Lee JS, Koh JY, Yi K, *et al.* Single-cell transcriptome of bronchoalveolar lavage fluid reveals sequential change of macrophages during SARS-CoV-2 infection in ferrets. *Nat Commun* 2021; 12: 4567.
- 21 Speranza E, Williamson BN, Feldmann F, *et al.* Single-cell RNA sequencing reveals SARS-CoV-2 infection dynamics in lungs of African green monkeys. *Sci Transl Med* 2021; 13: eabe8146.
- 22 Silvin A, Chapuis N, Dunsmore G, *et al.* Elevated calprotectin and abnormal myeloid cell subsets discriminate severe from mild COVID-19. *Cell* 2020; 182: 1401–1408.
- 23 Gordon DE, Jang GM, Bouhaddou M, *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 2020; 583: 459–468.
- 24 Travaglini KJ, Nabhan AN, Penland L, *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 2020; 587: 619–625.
- 25 Raredon MSB, Adams TS, Suhail Y, *et al.* Single-cell connectomic analysis of adult mammalian lungs. *Sci Adv* 2019; 5: eaaw3851.
- 26 Chu SG, Poli De Frias S, Sakairi Y, *et al.* Biobanking and cryopreservation of human lung explants for omic analysis. *Eur Respir J* 2020; 55: 1801635.
- 27 Llamazares-Prada M, Espinet E, Mijošek V, *et al.* Versatile workflow for cell type-resolved transcriptional and epigenetic profiles from cryopreserved human lung. *JCI Insight* 2021; 6: e140443.
- 28 Picelli S, Faridani OR, Björklund AK, *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 2014; 9: 171–181.
- 29 Gao C, Zhang M, Chen L. The comparison of two single-cell sequencing platforms: BD Rhapsody and 10 \times Genomics Chromium. *Curr Genomics* 2020; 21: 602–609.
- 30 Williams JW, Winkels H, Durant CP, *et al.* Single cell RNA sequencing in atherosclerosis research. *Circ Res* 2020; 126: 1112–1126.
- 31 Vieira Braga FA, Miragaia RJ. Tissue handling and dissociation for single-cell RNA-seq. *Methods Mol Biol* 2019; 1979: 9–21.
- 32 Danopoulos S, Bhattacharya S, Mariani TJ, *et al.* Transcriptional characterisation of human lung cells identifies novel mesenchymal lineage markers. *Eur Respir J* 2020; 55: 1900746.

- 33 O'Flanagan CH, Campbell KR, Zhang AW, *et al.* Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. *Genome Biol* 2019; 20: 210.
- 34 Denisenko E, Guo BB, Jones M, *et al.* Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol* 2020; 21: 130.
- 35 van den Brink SC, Sage F, Vértesy A, *et al.* Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat Methods* 2017; 14: 935–936.
- 36 Attar M, Sharma E, Li S, *et al.* A practical solution for preserving single cells for RNA sequencing. *Sci Rep* 2018; 8: 2151.
- 37 Wohnhaas CT, Leparc 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.
- 38 Sunkara V, Heinz GA, Heinrich FF, *et al.* Combining segmental bulk- and single-cell RNA-sequencing to define the chondrocyte gene expression signature in the murine knee joint. *Osteoarthritis Cartilage* 2021; 29: 905–914.
- 39 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.
- 40 Koenitzer JR, Wu H, Atkinson JJ, *et al.* Single-nucleus RNA-sequencing profiling of mouse lung. Reduced dissociation bias and improved rare cell-type detection compared with single-cell RNA sequencing. *Am J Respir Cell Mol Biol* 2020; 63: 739–747.
- 41 Eraslan G, Drokhlyansky E, Anand S, *et al.* Single-nucleus cross-tissue molecular reference maps to decipher disease gene function. *Science* 2022; 376: eabl4290.
- 42 Delorey TM, Ziegler CGK, Heimberg G, *et al.* COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature* 2021; 595: 107–113.
- 43 Joshi N, Misharin M. Single-Nucleus Isolation from Frozen Human Lung Tissue for Single-Nucleus RNA-seq. 2019. Available from: www.protocols.io/view/single-nucleus-isolation-from-frozen-human-lung-ti-kqdg34y81l25/v1 Date last accessed: 30 June 2022. Date last updated: 5 May 2019.
- 44 Slyper M, Porter CBM, Ashenberg O, *et al.* A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat Med* 2020; 26: 792–802.
- 45 Chu SG, Poli De Frias S, Raby BA, *et al.* An RNA-seq primer for pulmonologists. *Eur Respir J* 2020; 55: 1801625.
- 46 Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* 2019; 15: e8746.
- 47 Kharchenko PV. The triumphs and limitations of computational methods for scRNA-seq. *Nat Methods* 2021; 18: 723–732.
- 48 Sloan CA, Chan ET, Davidson JM, *et al.* ENCODE data at the ENCODE portal. *Nucleic Acids Res* 2016; 44: D726–D732.
- 49 Davis CA, Hitz BC, Sloan CA, *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* 2018; 46: D794–D801.
- 50 Andreotti S, Altmüller J, Quedenau C, *et al.* *De novo* whole genome assembly of the Roborovski dwarf hamster (*Phodopus roborovskii*) genome, an animal model for severe/critical COVID-19. *bioRxiv* 2021; preprint [<https://doi.org/10.1101/2021.10.02.462569>].
- 51 Stuart T, Butler A, Hoffman P, *et al.* Comprehensive integration of single-cell data. *Cell* 2019; 177: 1888–1902.
- 52 Amezquita RA, Lun ATL, Becht E, *et al.* Orchestrating single-cell analysis with Bioconductor. *Nat Methods* 2020; 17: 137–145.
- 53 Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 2018; 19: 15.
- 54 McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst* 2019; 8: 329–337.
- 55 Yang S, Corbett SE, Koga Y, *et al.* Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol* 2020; 21: 57.
- 56 Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* 2019; 20: 296.
- 57 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.
- 58 Tran HTN, Ang KS, Chevrier M, *et al.* A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol* 2020; 21: 12.
- 59 Luecken MD, Büttner M, Chaichoompu K, *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nat Methods* 2022; 19: 41–50.
- 60 Blondel VD, Guillaume J-L, Lambiotte R, *et al.* Fast unfolding of communities in large networks. *J Stat Mech* 2008; 2008: P10008.
- 61 Liu Z. Clustering single-cell RNA-seq data with regularized Gaussian graphical model. *Genes* 2021; 12: 311.

- 62 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.
- 63 La Manno G, Soldatov R, Zeisel A, *et al.* RNA velocity of single cells. *Nature* 2018; 560: 494–498.
- 64 Bergen V, Lange M, Peidli S, *et al.* Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* 2020; 38: 1408–1414.
- 65 Curras-Alonso S, Soulier J, Walter T, *et al.* Spatial transcriptomics for respiratory research and medicine. *Eur Respir J* 2021; 58: 2004314.