Mapping the immune landscape in the alveolar space reveals alterations of multiple macrophage states in chronic obstructive pulmonary disease

Dissertation

zur Erlangung des Doktogrades (Dr. rer. nat.)

der Mathematisch-Naturwissenschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

in der Fachrichtung Molekulare Biomedizin

vorgelegt von

Kevin Simon Baßler

Bonn

Juni, 2020

Angefertigt mit Genehmigung und nach den Richtlinien der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Joachim L. Schultze

2. Gutachter: Priv.-Doz. Dr. Marc Beyer

Tag der Promotion: 22.01.2021

Erscheinungsjahr: 2021

Acknowledgement

This thesis has benefited greatly from the support of many people, and I would like to take this opportunity to sincerely thank some of them.

First and foremost, I am deeply grateful to my supervisor Prof. Dr. Joachim L. Schultze for offering me such an interesting topic for my thesis and for his constant support, guidance and expertise. His constant openness for discussions gave me the confidence to believe in my results and to push the project forward. I would also like to take this opportunity to thank Pr. Doz. Dr. Marc Beyer for the interesting scientific discussions throughout my PhD and for being the second examiner in the committee of my final examination.

I would also like to acknowledge my national and international cooperation partners for their contribution and support in the analysis of COPD data: the team of FASTGenomics, especially Dr. Benedikt Reiz, for the development of the machine learning-based cell-type annotation approach; Erika Dudkin and Prof. Dr. Jan Hasenauer for the development of the innovative DE analysis approach; Dr. Malte Lücken and Prof. Dr. Fabian J. Theis for the discussions and efforts in the analysis of monocyte-to-macrophage trajectories; Allon Wagner and Prof. Dr. Nir Yosef for helping me to model the metabolic states of the alveolar macrophages; Prof. Dr. Christoph Thiele for performing lipidomics; Prof. Dr. Andreas Schlitzer and Prof. Dr. Florent Ginhoux for providing murine monocyte data. Finally yet importantly, I also thank Prof. Dr. Dirk Skowasch and team for recruiting patients and providing BALF and blood samples.

The underlying data of the present study would not exist without the help of many people from the Schultze group, some of whom I would like to highlight and extend my heartfelt gratitude to. First of all, my beloved lung group comprising Dr. Wataru Fujii, Dr. Theodore Kapellos and Arik Horne – we have gone through ups and downs together! A special thank also goes to Collins Osei-Sarpong, who taught us how to use the Seq-Well technology. Further thanks are due to Nico Reusch, Stefanie Warnat-Herresthal, Jonas Schulte-Schrepping, Emily Hinkley, Dr. Lorenzo Bonaguro, Dr. Matthias Becker, Dr. Kristian Händler, and Dr. Anna Aschenbrenner for all their support and refreshing work-related as well as private discussions – many thanks for this unforgettable time!

The path of a PhD student is often steep and filled with hurdles. Fortunately, I had many companions on my journey, who gave me support and stability. For this, I would like to thank my parents, Simone und Hubert, for showing their interest and support, which encouraged me throughout this time of intense work. The same applies to my close friends Patrick Günther and Dr. Thomas Ulas.

However, my greatest thanks go to my wife Julia, who was the driving force behind my work with her empathy, care and interest – I love you!

Finally, but first in my heart, I would like to thank my grandparents and Tadeusz Jelenkowski, who have always supported me in my dream of becoming a scientist – I will never stop missing you!

Content

Table index

Figure index

Abbreviations

Remark: The list of abbreviations does not include gene names, chemicals, and acronyms of computational tools or wet lab technologies.

acc. to	according to
AM	alveolar macrophage
APC	antigen-presenting cell
approx.	approximately
ATM	adipose tissue macrophage
AUC	area under the curve
AUPR	area under the precision recall curve
BALF	bronchoalveolar lavage fluid
BCR	B cell receptor
CAT	COPD assessment test
CCA	canonical correlation analysis
cDC	conventional DC
cMoP	common monocyte progenitor
COPD	chronic obstructive pulmonary disease
COVID-19	coronavirus disease 2019
CPFE	combined pulmonary fibrosis and emphysema
ctrl.	control
CyTOF	cytometry by time of flight
DC	dendritic cell
DE	differentially expressed
degrad.	degradation
DNA	deoxyribonucleic acid
e.g.	for example
ENCODE	encyclopedia of DNA elements
FACS	fluorescence-activated cell sorting
FEV1	forced expiratory volume in one second

FMO	fluorescence minus one
FPKM	fragments per kilobase million
freq.	frequent
FVC	forced vital capacity
generat.	generation
GenExPro	gene expression profiler
GEO	gene expression omnibus
GO	gene ontology
GOEA	gene ontology enrichment analysis
GOLD	Global Initiative for Chronic Obstructive Lung Disease
GSEA	gene set enrichment analysis
GSVA	gene set variation analysis
HCA	human cell atlas
HLA	human leukocyte antigen
HPCA	human primary cell atlas
HSC	hematopoietic stem cell
ICS	inhaled corticosteroids
ILC	innate lymphoid cell
IM	interstitial macrophage
intermed.	intermediate
KEGG	kyoto encyclopedia of genes and genomes
kNN	k-nearest neighbor
LABA	long-acting β2 agonists
LAM	lipid-associated macrophage
LAMA	long-acting muscarinic antagonists
Lin	lineage
macro	macrophage
MCFC	multi-color flow cytometry
MDM	monocyte-derived macrophage

MDP	macrophage and DC precursor
MHC	major histocompatibility complex
mito.	mitochondrial
ML	machine learning
MMP	matrix metalloproteinases
mod.	modification
moDC	monocyte-derived DC
mono	monocyte
mRNA	messenger RNA
MSigDB	molecular signatures database
МΦ	macrophage
NET	neutrophil-extracellular trap
neutro	neutrophil
NΦ	neutrophil
PAP	pulmonary alveolar proteinosis
pat.	patient
PBMC	peripheral blood mononuclear cell
PC	principal component
PCA	principal component analysis
PCC	Pearson correlation coefficient
pDC	plasmacytoid DC
PID	pathway interaction database
present.	presentation
proliferat.	proliferating
QC	quality control
regulat.	regulatory
RNA	ribonucleic acid
ROS	reactive oxygen species
RPKM	reads per million kilobase

RPM	reads per million
scRNA-seq	single-cell RNA-seq
sign.	signature
signal.	signaling
SNN	shared nearest neighbor
SOP	standard operating procedure
SR1	SingleR (HPCA)
SR2	SingleR (Blueprint + ENCODE)
TCR	T cell receptor
TF	transcription factor
TPM	transcripts per million kilobase
t-SNE	t-distributed stochastic neighbor embedding
TSO	template switch oligonucleotide
UMAP	uniform multiple approximation and projection
UMI	unique molecular identifier
v	version
visual.	visualization
WHO	world health organization

List of publications

First author publications

Schulte-Schrepping, Jonas, Nico Reusch, Daniela Paclik, **Kevin Baßler**, Stephan Schlickeiser, Bowen Zhang, Benjamin Krämer, et al. 2020.
"*Suppressive Myeloid Cells Are a Hallmark of Severe COVID-19*"
Cell, September. 182 (6): 1419–1440.e23
https://doi.org/10.1016/j.cell.2020.08.001

Kevin Baßler, Wataru Fujii, Theodore S. Kapellos, Arik Horne, Benedikt Reiz, Erika Dudkin, Malte Lücken, et al. 2020.
"*Alterations of Multiple Alveolar Macrophage States in Chronic Obstructive Pulmonary Disease*"
bioRxiv, May. Cold Spring Harbor Laboratory, 2020.05.28.121541
https://doi.org/10.1101/2020.05.28.121541

Patrick Günther, Branko Cirovic, **Kevin Baßler**, Kristian Händler, Matthias Becker, et al. 2019.
"*A rule-based data-informed cellular consensus map of the human mononuclear phagocyte cell space*"
bioRxiv, June. Cold Spring Harbor Laboratory, 658179
https://doi.org/10.1101/658179

Kevin Baßler, Jonas Schulte-Schrepping, Stefanie Warnat-Herresthal, Anna C. Aschenbrenner, and Joachim L. Schultze. 2019.
"*The Myeloid Cell Compartment—Cell by Cell*"
Annual Review of Immunology 37 (1): 269–93
https://doi.org/10.1146/annurev-immunol-042718-041728

Kevin Baßler, Patrick Günther, Jonas Schulte-Schrepping, Matthias Becker, and Paweł Biernat. 2019.
"*A bioinformatic toolkit for single-cell mRNA analysis*"
Methods in Molecular Biology, 1979:433–55. Humana Press
Inc. https://doi.org/10.1007/978-1-4939-9240-9_26

Anna C. Aschenbrenner, **Kevin Baßler**, Mirco Brondolin, Lorenzo Bonaguro, Pilar Carrera, Kathrin Klee, Thomas Ulas, Joachim L. Schultze, and Michael Hoch. 2017.

"*A Cross-Species Approach to Identify Transcriptional Regulators Exemplified for Dnajc22 and Hnf4a*"
Scientific Reports, no. May. Springer US: 1–15
https://doi.org/10.1038/s41598-017-04370-9

Co-author publications

Muus, Christoph, Malte D. Luecken, Gokcen Eraslan, Avinash Waghray, Graham Heimberg, Lisa Sikkema, Yoshihiko Kobayashi, […], **Kevin Baßler**, et al. 2020.
"*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 Cell*"
bioRxiv, April. Cold Spring Harbor Laboratory, 2020.04.19.049254
https://doi.org/10.1101/2020.04.19.049254

Stefanie Warnat-Herresthal, Konstantinos Perrakis, Bernd Taschler, Matthias Becker, **Kevin Baßler**, Marc Beyer, Patrick Günther, et al. 2020.
"*Scalable Prediction of Acute Myeloid Leukemia Using High-Dimensional Machine Learning and BloodTranscriptomics*"
iScience 23 (1): 100780
https://doi.org/10.1016/j.isci.2019.100780

Matthias Becker, Milind Chabbi, Stefanie Warnat-Herresthal, Kathrin Klee, Jonas Schulte-Schrepping, Pawel Biernat, Patrick Günther, **Kevin Baßler**, et al. 2019.
"*Memory-Driven Computing Accelerates Genomic Data Processing*"
bioRxiv, January. Cold Spring Harbor Laboratory, 519579
https://doi.org/10.1101/519579

Matthias Becker, Hartmut Schultze, Thomas Ulas, Sharad Singhal, Joachim L. Schultze, Milind Chabbi, Stefanie Warnat-Herresthal, […], **Kevin Baßler**, et al. 2019.
"*Accelerated Genomics Data Processing Using Memory-Driven Computing*"
In Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019, 1850–55. Institute of Electrical and Electronics Engineers Inc.
https://doi.org/10.1109/BIBM47256.2019.8983296

Lisa Schmidleithner, Yasser Thabet, Eva Schönfeld, Maren Köhne, Daniel Sommer, Zeinab Abdullah, Timothy Sadlon, […], **Kevin Baßler**, et al. 2019.
"*Enzymatic Activity of HPGD in Treg Cells Suppresses Tconv Cells to Maintain Adipose Tissue Homeostasis and Prevent Metabolic Dysfunction*"
Immunity 50 (5): 1232-1248.e14
https://doi.org/10.1016/j.immuni.2019.03.014

Eleni Kabrani, Van Trung Chu, Evangelia Tasouri, Thomas Sommermann, **Kevin Baßler**, Thomas Ulas, Thorsten Zenz, et. al. 2018.
"*Nuclear FOXO1 promotes lymphomagenesis in germinal center B cells"*
Blood 2018: blood-2018-06-856203
https://doi.org/10.1182/blood-2018-06-856203

Claus J. Scholz, Paweł Biernat, Matthias Becker, **Kevin Baßler**, Patrick Günther, Jenny Balfer, Henning Dickten, et al. 2018.
"*FASTGenomics: An Analytical Ecosystem for Single-Cell RNA Sequencing Data*"
bioRxiv, March. Cold Spring Harbor Laboratory, 272476
https://doi.org/10.1101/272476

Theodore S. Kapellos, **Kevin Baßler**, Anna C. Aschenbrenner, Wataru Fujii, and Joachim L. Schultze. 2018.
"*Dysregulated Functions of Lung Macrophage Populations in COPD*"
Journal of Immunology, Research 2018 (February). Hindawi: 1–19
https://doi.org/10.1155/2018/2349045

Anette Christ, Patrick Günther, Mario A.R. Lauterbach, Peter Duewell, Debjani Biswas, Karin Pelka, […], **Kevin Baßler**, et al. 2018.
"*Western Diet Triggers NLRP3-Dependent Innate Immune Reprogramming*"
Cell 172 (1–2): 162–175.e14
https://doi.org/10.1016/j.cell.2017.12.013

Marina Doebbeler, Christina Koenig, Lena Krzyzak, Christine Seitz, Andreas Wild, Thomas Ulas, **Kevin Baßler**, et al. 2018.
"*CD83 Expression Is Essential for Treg Cell Differentiation and Stability*"
JCI Insight 3 (11)
https://doi.org/10.1172/jci.insight.99712

Marc Beyer, Kristian Händler, Patrick Günther, **Kevin Baßler**, Thomas Ulas, Matthias Becker, Kathrin Klee, Joachim L. Schultze, and Andreas Schlitzer. 2017.

"*Navigating Disease Phenotypes – A Multidimensional Single-Cell Resolution Compass Leads the Way*"
Current Opinion in Systems Biology 3 (June). Elsevier: 147–53
https://doi.org/10.1016/J.COISB.2017.05.004.

Jil Sander, Susanne V. Schmidt, Branko Cirovic, Naomi McGovern, Olympia Papantonopoulou, Anna-Lena Hardt, […], **Kevin Baßler**, et al. 2017.
"*Cellular Differentiation of Human Monocytes Is Regulated by Time-Dependent Interleukin-4 Signaling and the Transcriptional Regulator NCOR2*"
Immunity 47 (6): 1051–1066.e12
https://doi.org/10.1016/j.immuni.2017.11.024

Summary

Chronic obstructive pulmonary disease (COPD) is ranked the third leading cause of death worldwide and hence constitutes a substantial medical and financial burden. The disease is characterized by a poorly reversible and progressive airway obstruction, which is mainly caused by chronic inflammation of the lung. However, the cellular mechanisms underlying COPD are not well understood and thus no effective therapy is currently available.

To broaden the current knowledge about the immune landscape in the lung, we characterized the cellular phenotypes within the alveolar space and peripheral blood of both COPD patients and control donors using the Seq-Well technology for single-cell RNA-sequencing (scRNA-seq). For analysis, we developed novel computational approaches, including cell type annotation (GenExPro), cell type classification based on machine learning, and identification of differentially expressed genes and functional differences between COPD and control cells. In addition, we applied state-of-the-art methods to assess the metabolic state of cells and modelled potential differentiation trajectories based on RNA velocity calculations.

The in-depth scRNA-seq analysis revealed a massive heterogeneity within immune cell types, especially among alveolar macrophages in the alveolar space and neutrophils in the blood. Interestingly, the alveolar macrophages in COPD showed an altered lipid metabolism with increased cellular cholesteryl ester content, reduced MHC I expression, defective chemotaxis towards CCL3, and a metabolic shift reminiscent of mitochondrial dysfunction. Furthermore, we found a shift in alveolar macrophage subtypes from COPD patients towards proliferating cells and cells expressing both monocyte and macrophage genes (referred to as ‘mono-like macrophages’). For the latter cells, we predicted blood monocytes as a putative source of replenishment. Intriguingly, the blood monocytes of COPD patients already showed a reduced MHC I expression similar to that of alveolar macrophages. In addition, using scRNA-seq data from a mouse model, we showed that some monocytes obtained from the lung exhibited an immature gene program reminiscent of the monocyte precursors in the bone marrow, suggesting that the blood monocyte population may be partially derived from an extramedullary hematopoiesis site in the lung. To identify signaling pathways that regulate differentially expressed genes between COPD and controls in human mono-like macrophages, we modelled the cell-cell communication of immune

cells in the alveolar space and identified in particular the TGF-ß signaling pathway, but also other pathways, including NOTCH, WNT and TNF signaling pathways, as potential DE gene regulators in COPD. An overview of the observed changes in the AM population is shown in **Figure 1**.

Taken together, scRNA-seq was used to provide a cellular and molecular framework for understanding and analyzing the pathophysiology of COPD, which is a prerequisite for the development of new diagnostic approaches, including molecular biomarkers, and causal therapies for this deadly disease.

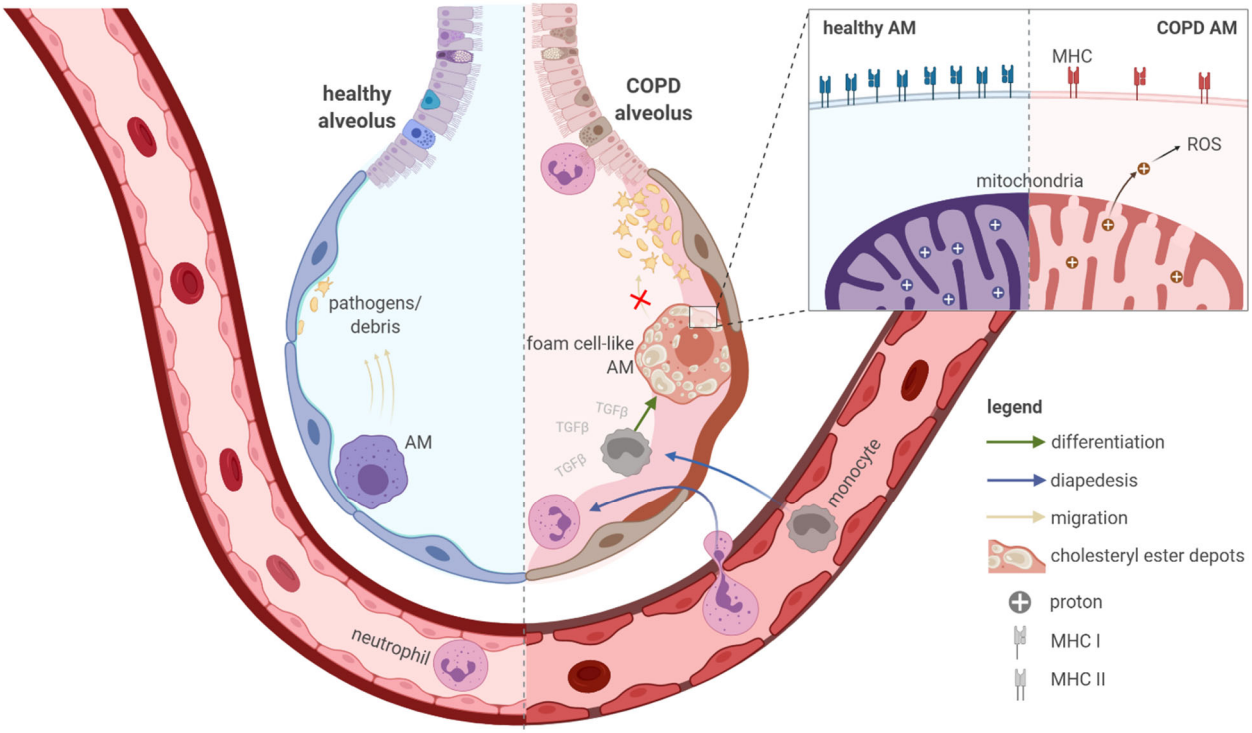


Figure 1. Schematic representation of the key findings of the present study

In healthy lungs. alveolar macrophages surveil the alveoli and remove pathogens and debris to enable proper gas exchange. In the alveoli of COPD patients, the alveolar macrophages accumulate cholesteryl esters and are likely to acquire a foam cell morphology. In addition, blood neutrophils and monocytes invade the alveoli and monocytes differentiate into alveolar macrophages. The transcriptome of COPD alveolar macrophages indicate TGF-β-associated cell signaling in the early stages of monocyte-to-macrophage differentiation. The alveolar macrophages in COPD show a reduced ability to migrate towards chemokine. Furthermore, they express fewer MHC molecules; especially MHC I. Together with the reduced phagocytosis of alveolar macrophages in COPD, the ability of these cells for immunosurveillance is severely limited during the disease. In addition, their

mitochondria are leaking (e.g. to protons) and therefore produce high amounts of reactive oxygen species. Taken together, the guardians of normal lung function (alveolar macrophages) are severely altered in COPD, preventing them from fulfilling their important physiological functions properly.

1. Introduction

1.1. The immune system

The defense against pathogenic cues is an inherent characteristic of living cells. In higher organisms, the protection against harmful influences, led to the emergence and evolution of an orchestrated network of cells, known as immune system. In all vertebrates, such as humans, the immune system can be divided into two cooperative branches, called innate and adaptive immune system that can be distinguished in terms of speed and specificity of the immune response. The innate immune system is older in evolutionary respect and fully functional from birth. The cells of the innate immune system are called myeloid cells and include monocytes, macrophages, dendritic cells (DCs), mast cells and granulocytes (comprising neutrophils, eosinophils and basophils) (**Figure 2**). Its immune protection is characterized by the non-specific sensing of pathogenic signals and the rapid and efficient elimination of any invading pathogenic material. In contrast, the adaptive immune system is highly specific for the detection of pathogens and develops and adapts throughout lifetime, which is why it is also known as the ‘specific immune system’ or ‘acquired immune system’. Cells of the adaptive immune response are commonly referred to as lymphocytes and includes T cells and B cells (**Figure 2**). However, innate lymphoid cells (ILCs), including NK cells, are also lymphocytes but belong to the innate arm of the immune system (Klose and Artis, 2016) (**Figure 2**). The specificity of T cells and B cells towards a pathogen comes from surface molecules that recognize certain molecular structures. In T cells, these surface molecules are T cell receptors (TCRs), which are produced in a wide range of antigen specificities so that each T cell carries TCRs of a unique antigen-specificity. The underlying principles that lead to the great variability of TCRs and thus to antigen-specificity are also found in immunoglobulins (also called antibodies) that are produced by B cells. Antibodies exist in a membrane-bound form, also called B cell receptors (BCRs), but can also be secreted by terminally differentiated B cells - the plasma cells. The molecular structures that are recognized by TCRs and BCRs are called antigens, which is a general term for any substance capable of triggering an adaptive immune response, but usually describes protein fragments or peptides. However, antigen recognition by TCR requires the presentation of the respective

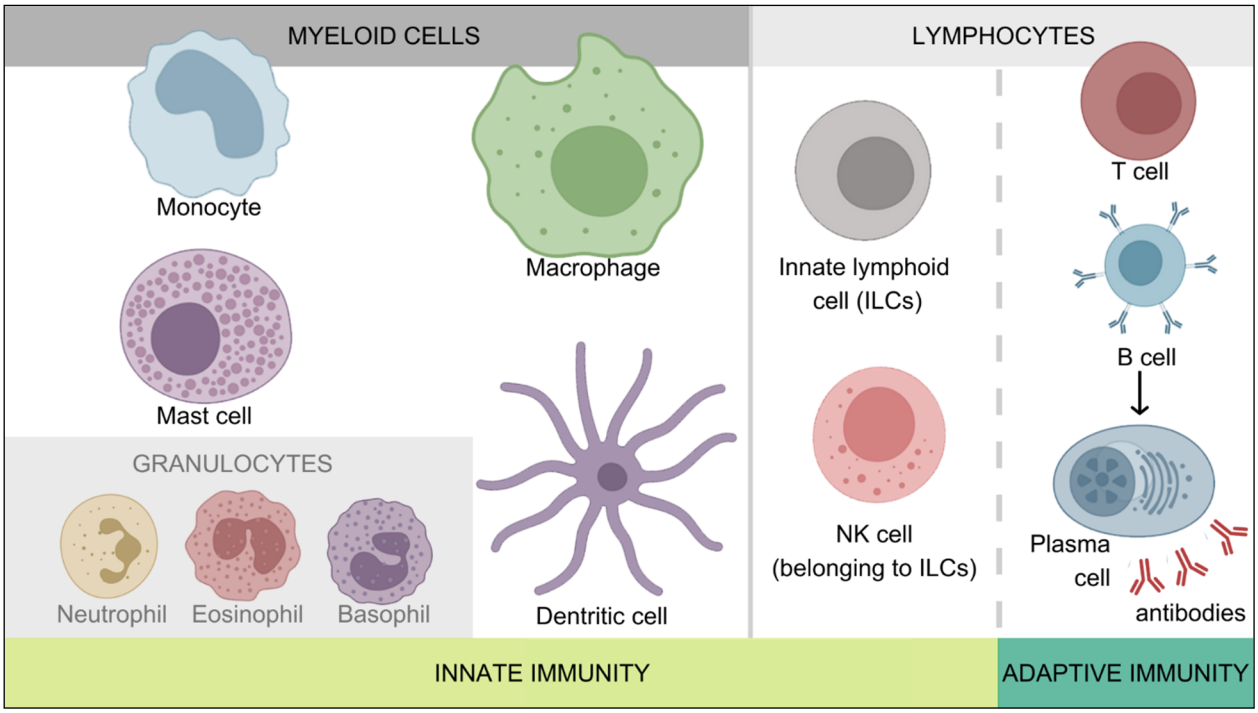


Figure 2. Overview of the cellular elements of the innate and adaptive immune response

antigen by a major histocompatibility complex (MHC), also referred to as human leukocyte antigen complex, on the surface of antigen-presenting cells (APCs), such as DCs, monocytes and macrophages, thereby linking the innate and adaptive immune systems. The MHC surface receptors for antigen presentation are classified into MHC class I (MHC I) and class II (MHC II) molecules. MHC II molecules are mainly expressed by APCs and present antigens derived from extracellular pathogens (Rock et al., 2016) acquired by phagocytosis. In contrast, MHC I molecules are expressed by all nucleated cells in the body and present antigens of intracellular pathogens during infection (Rock et al., 2016). However, during homeostasis, MHC I molecules present peptides derived from the normal cellular protein turnover. Furthermore, MHC I molecules are to some extent able to present antigens derived from extracellular pathogens in a process called ‘cross-presentation’ (Embgenbroich and Burgdorf, 2018; Joffre et al., 2012). The continuous presentation of loaded MHC I molecules on the cell surface is an immunological safety mechanism of the body, since the MHC I molecule serves as inhibitory ligand for NK cells. Some viruses (Hansen and Bouvier, 2009) and tumors (Garrido et al., 2016) cause a reduction in MHC I surface

expression, which in turn leads to an activation of NK cell-mediated killing of the presenting cell (Höglund and Brodin, 2010).

Taken together, the human immune system is characterized by a sophisticated and highly organized interaction network of specialized cells that aim to protect the body from pathogenic influences.

1.2. The architecture of the lung

The lung is the primary organ of the human respiratory system. It enables gas exchange by removing oxygen from the atmosphere and transferring it to erythrocytes and, in addition, by uptaking carbon dioxide from the blood and releasing it into the atmosphere. The human lung can be subdivided into two major parts, a right lung and a left lung, which are divided into air conducting areas (airways) and areas where the actual exchange of gas takes place (parenchyma) (Chaudhry and Bordoni, 2019). The airways inside the lungs comprise, ordered from large to small: the bronchi, the bronchioles and the terminal bronchioles (**Figure 3**). The terminal bronchioles divide further into respiratory bronchioles, which in turn divide into a series of alveolar ducts that terminate in alveolar sacs, where the alveoli are located, which are the parenchyma of the lung, since this is where gas exchange takes place. To keep the airways open during respiration, the bronchi and bronchioles are equipped with smooth muscles, while in the terminal bronchioles and the parenchyma the so-called surfactant takes over the task. The surfactant is produced by specialized epithelial cells (type II pneumocytes) and is a protein-lipid secretion that lines the air-liquid interface in the parenchyma and reduces surface tension. The lipid composition of surfactant is 90% phospholipids and 10% neutral lipids (especially cholesterol) (Nkadi et al., 2009).

1.3. The immune system of the lung

Among all areas in the body where pathogens can invade, the lung represents the greatest immunological challenge for the host. Not only is the lung one of the organs

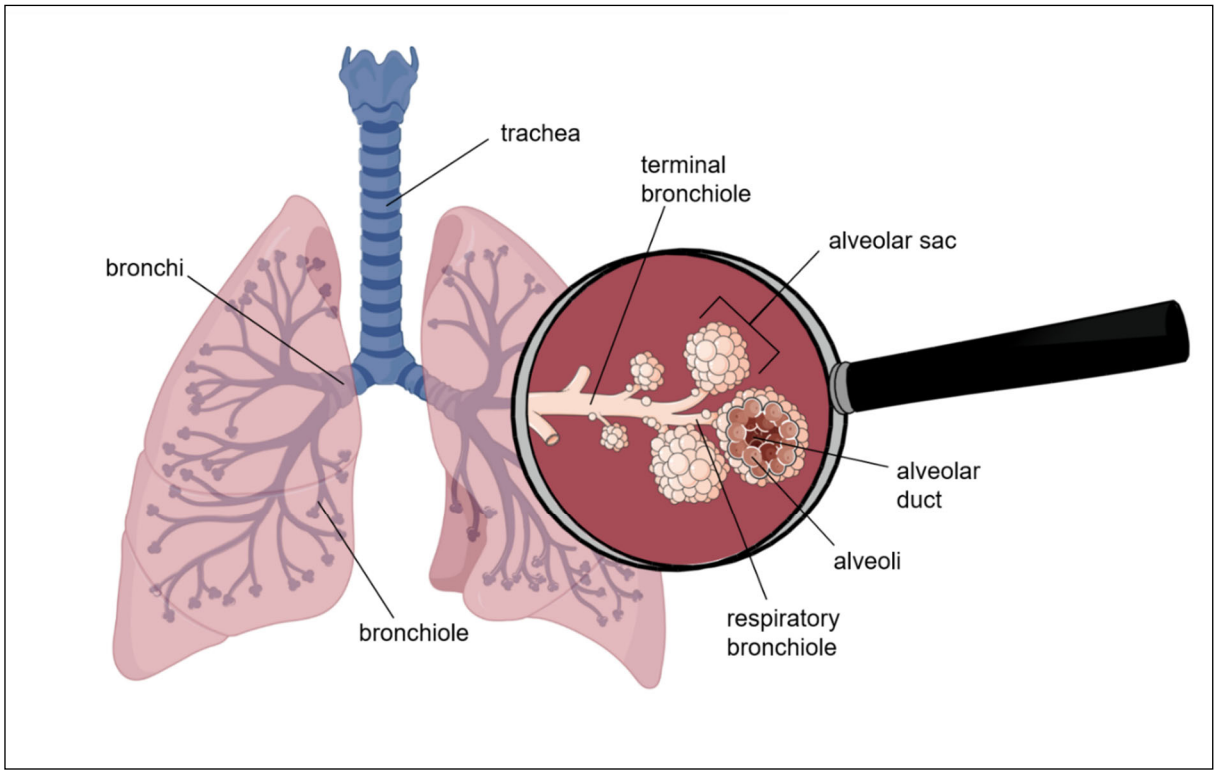


Figure 3. Anatomic architecture of the human lungs

that is in constant contact with the environment, but it is also the one that is most frequently attacked by pathogens. Moreover, due to its role as a gas exchange organ, the normal functioning of the lung is crucial for health and makes it extremely sensitive to damage. Therefore, the lung is equipped with a variety of different defense strategies. The frontline defense is predominantly represented by epithelial cells, which build a physical barrier by forming close connections to neighboring cells through cell-cell junctions (Denney and Ho, 2018; Nicod, 2005). In addition, the airways are covered with a film of mucous that traps and destroys pathogens, and the apical part of the epithelial cells is covered with motile cilia that have a continuous rhythmic beat, which transports the mucous towards the trachea and throat, where it is swallowed or expectorated. This mechanism is also known as "mucociliary escalator" (Bustamante-Marin and Ostrowski, 2017) and is found solely in the airways. Another important level of host defense is provided by immune cells that colonize the lung and cooperate closely with the epithelial cells (Lloyd and Marsland, 2017). This second tier of defense includes the coordinated activation of tissue- resident lymphoid cells, which patrol through the lung via the lymph system. Among the resident

pulmonary lymphoid cells, there are T cells and B cells, as well as ILCs (Lloyd and Marsland, 2017). However, as outlined above, the activation of T cells requires the presentation of antigens on MHC molecules. DCs are known APCs and under homeostatic conditions, a network of DCs traverses the airways. The pulmonary DCs can be divided into three different subtypes: plasmacytoid DCs (pDCs) and conventional DCs (cDC1 and cDC2) (Guilliams et al., 2013a; Neyt and Lambrecht, 2013). Other important APCs are pulmonary macrophages, which describes a heterogeneous cell population and exhibits high plasticity to ensure rapid adaptation to environmental changes (Garbi and Lambrecht, 2017; Hussell and Bell, 2014). In lung homeostasis, two main populations of macrophages are distinguished based on their localization: interstitial macrophages (IMs) are found in the interstitium, while airway macrophages are located in the lumen of the airways and parenchyma (Kapellos, Bassler et al., 2018). Airway macrophages are the most abundant immune cells in the lung (Joshi et al., 2018; Puttur et al., 2019) and the best known representatives are the alveolar macrophages (AMs) that are found in the luminal surface of the alveolar space. Morphologically, AMs are larger in size compared to IMs (Hoppstädter et al., 2010) and can be readily distinguished from other luminal immune cells by both high autofluorescence and HLA-DR expression (Bharat et al., 2016; van Haarst et al., 1994; Joshi et al., 2018; Vermaelen and Pauwels, 2004). In addition, CD206, CD169, CD163 and MARCO are also well-known markers for AMs in human (Joshi et al., 2018; Kapellos, Bassler et al., 2018). AMs play an important role in the removal of foreign material to ensure that the alveoli remain free of debris (Baharom et al., 2017). In rats, it has been estimated that the population of AMs can cope with up 10^9 bacteria injected into the lung before further layers of immune response, such as adaptive immunity, are induced (MacLean et al., 1996), highlighting the effective phagocytic and surveillance abilities of AMs. Interestingly, AMs are inherently suppressive as they play an immunoregulatory role in preventing overshooting reactions and thus mitigate tissue damage, while providing an adequate immune response to inhaled particles (Lambrecht, 2006). For this purpose, AMs are maintained in a quiescent state and produce TGF-β and IL-10 to attenuate the immune response during homeostasis (Garbi and Lambrecht, 2017). Moreover, AMs ensure surfactant homeostasis by contributing up to 40% of the total surfactant-lipid catabolism (Garbi and Lambrecht, 2017; Whitsett et al., 2010). Previous studies indicated that the AM population is highly heterogeneous (Kapellos, Bassler et al.,

2018). For example, density gradient centrifugation identified at least three AM subpopulations with distinct functionalities (Kapellos, Bassler et al., 2018; Spiteri et al., 1992). However, it is likely that higher resolution technologies will identify additional subpopulations and thus reveal the full spectrum of AM heterogeneity. The same is true for the IM population, whose heterogeneity has been demonstrated based on density gradient centrifugation, flow cytometry and single-cell RNA sequencing studies (Schyns et al., 2018, 2019). Interestingly, the morphology of IMs is reminiscent of monocytes, which is why these cells have long been regarded as an intermediate population in the differentiation process of infiltrating blood monocytes towards AMs (Kapellos, Bassler et al., 2018). Indeed, recent studies suggested that at least some IMs might be replenished by blood monocytes (Schyns et al., 2019).

Finally, mast cells and granulocytes are also found in homeostatic lungs, with a substantial number of marginating neutrophils within the lung microvasculature (Hidalgo et al., 2019).

Collectively, several layers of defense, involving both immune cells and epithelial cells, maintain the homeostasis of the lung. However, there are still some gaps in the understanding of the immune system in the lung, which necessitates high-resolution techniques, such as single-cell RNA sequencing to characterize and close these gaps.

1.4. Immunology in the era of single-cell genomics

This chapter emphasizes how single-cell RNA sequencing has changed our previous understanding of myeloid immunology.

As part of this partly cumulative thesis, this chapter is an extract from the following publication (Bassler et al., 2019):

The Myeloid Cell Compartment—Cell by Cell

Authors: *Kevin Bassler, Jonas Schulte-Schrepping, Stefanie Warnat-Herresthal, Anna C. Aschenbrenner and Joachim L. Schultze*

Publication details: *First published as a Review in Advance on January 16, 2019 in Annual Review of Immunology; 37:269–93; https://doi.org/10.1146/annurev-immunol-042718-041728*

To ease the tracking of the self-quotation, the citation is placed in quotation marks and the content is written in italics. In addition, shortenings of the original text are indicated by [...] and the citation style and numbering of the text and figures are adapted to the other texts of the dissertation.

"*The ongoing revolution in genomics allowing for single-cell resolution assessment of whole transcriptomes (Picelli et al., 2014; Ramsköld et al., 2012), genomes (Gawad et al., 2016; Navin et al., 2011), and epigenomes (Buenrostro et al., 2018; Cusanovich et al., 2015; Litzenburger et al., 2017) is currently reshaping our understanding of cellular constituents in every organ and cellular system. In particular, single-cell transcriptomics by single-cell RNA sequencing (scRNA-seq) allows for an unprecedented level of precision when describing cells (Shalek et al., 2013). Moreover, and in contrast to all other previous single-cell technologies, scRNA-seq can be applied in a completely unbiased fashion (Shapiro et al., 2013). This unique opportunity allows us to completely revise our understanding of all cellular components and their functionality in every organ and tissue and under any condition.*
Single-cell analysis has a long-standing track record not only in the immunological sciences (Hooke, 1665). Very early on, it became clear that many different immune cells exist. Cytological techniques based on classical light microscopy had already revealed that cells have many different shapes and sizes (Steinman and Cohn, 1973). Labeling of proteins with antibodies allowed for defining cells based on the expression of proteins found to be associated with a particular cell type. New technological fields emerged, including immunohistochemistry (Coons et al., 1941), immunofluorescence (Franke et al., 1978), and flow cytometry (Fulwyler, 1965) - all based on the usage of antibodies specifically binding to marker proteins. […] *These technical developments have been and are still extremely important in categorizing different cell types. With the newest technologies, including multicolor flow cytometry (MCFC) and cytometry by time of flight (CyTOF), one can measure more than 50 protein markers per cell (Bandura et al., 2009; Bendall et al., 2011; Guilliams et al., 2016); upcoming updates of these technologies suggest measurements of up to 100 markers. While this led to an incredible wealth of novel insights into immune cell biology and the stratification of*

subpopulations within the various immune cell types, all these methods still lack the ability to identify outright new cell types and cell states in an entirely unbiased fashion. […]*However,* […] *novel unbiased technologies hold great hope to directly address major unmet health needs. In fact, the single-cell genomics technologies should further propel the field of human immunology (Davis and Brodin, 2018), since measurements are finally possible on an unprecedented detailed level. Is there already sufficient evidence that such unbiased high-resolution technology can be successfully applied to answer burning questions about the myeloid cell compartment? Overall, several seminal papers have clearly demonstrated that single-cell genomics distinguishes immune cell types with remarkable resolution, leads to the identification and investigation of effects of the cellular environment, and can discover previously unknown types or states of immune cells (Gaublomme et al., 2015; Mass et al., 2016; Shalek et al., 2013, 2014), even when they are embedded in complex tissues (Jaitin et al., 2014; Gaublomme et al., 2015; See et al., 2017; Tirosh et al., 2016), and with spatial resolution of the immune response (Angelo et al., 2014; Liu et al., 2015).* […] *While we agree with many colleagues about the enormous excitement, the significant scientific insights, and the impact these technologies have on our understanding of myeloid cells in health and disease, we would like to firmly state that these technologies are far from routine, particularly in application to human samples. Moreover, some of the scRNA-seq techniques heavily advertised and used need to be approached with caution, considering the investigation of complex transcriptional and epigenetic processes and states including the plasticity of the myeloid cell compartment. […]*

1.4.1. New insights into the precursors of the myeloid cell compartment

There has been tremendous progress in the understanding of the ontogeny of tissue macrophages […] *and monocytes. While most tissue macrophages are derived from embryonic precursors in the yolk sac and fetal liver (Ginhoux and Jung, 2014; Varol et al., 2015), monocytes, granulocytes, and DCs originate from a multipotent stem cell, the so-called hematopoietic stem cell (HSC) (Laurenti and Göttgens, 2018). Over a century ago, scientists proposed a model that all cell lineages are like branches of a tree, with stem cells representing the trunk (Laurenti and Göttgens, 2018). This tree*

was later adopted to describe hematopoiesis and is still used in text- books. In fact, extensive work has been performed to define different developmental stages and lineages within the hematopoietic system, including defined precursors of the myeloid cell compartment […]. *Yet, novel single-cell analyses challenge this view on the hematopoietic system.* […] *HSCs may be a heterogeneous pool of multipotent cells that are precommitted to different extents. The term precommitment needs to be defined in this context. Assuming that the cellular program of HSCs can be seen as a balance with a myeloid lineage program on the one side and the lymphoid lineage program on the other side, commitment could mean that one of the programs weighs more (e.g., the myeloid lineage program). Thus, under normal circumstances, the HSC will prefer to develop into cells of the myeloid lineage. However, the balance can easily be shifted in the presence of certain stimuli (e.g., stress), and the lymphoid lineage program would hence be favored. Consequently, precommitment does not rule out plasticity of a progenitor cell but rather describes its preference for a certain lineage at a given time* […] *Precommitment might also become more pronounced during aging or could even be seen as an aging process of the immune system. This model would explain why aged HSCs and progenitor cells exhibit a marked bias toward the myeloid lineage along a decreased output of cells of the lymphoid lineages (Benz et al., 2012; Young et al., 2016) or the recently identified HSC bias toward platelet programming found using scRNA-seq (Grover et al., 2016). Whether this is accompanied by complete loss of plasticity requires further investigation. In addition to supporting findings on early precommitment, scRNA-seq also led the way to the development of tools that add a pseudotemporal layer to the analysis. Since the data generated by scRNA-seq methods present a snapshot of cells in different developmental, differentiation, or cell cycle states, the expression profiles can be used to computationally order the cells along a pseudo–time course in a trajectory (Ji and Ji, 2016; Moignard et al., 2015; Qiu et al., 2017a). Using such algorithms confirmed that hematopoiesis constitutes a continuum of differentiation (Nestorowa et al., 2016; Schultze and Beyer, 2016; Tusi et al., 2018; Xue et al., 2014). Thus, under this assumption, the previously described developmental interims might rather represent transitory states or branching points along the differentiation trajectory. To circumvent the limitations of the tree model, we introduce what we call the autobahn model, which highlights the immense dynamics and plasticity of cells within hematopoiesis* (**Figure 4**).

1.4.2. The monocyte compartment revisited

Monocytes are a major cellular compartment derived from HSCs. In humans, they are defined as blood-circulating, phagocytic immune cells. Based on the assessment of the expression of CD14 and CD16 on the cell surface, three different subsets (classical, intermediate, and non-classical monocytes) are currently distinguished (Ziegler-Heitbrock, 2015). In mice, only two subsets are regularly classified, the short-lived circulating Ly6C$^+$ monocytes (corresponding to CD14^{hi}CD16^{lo} monocytes in human) and Ly6C$^-$ monocytes (corresponding to CD14^{lo}CD16^{hi} monocytes in human). So far, no equivalent for the intermediate phenotype in humans has been observed in mice. In 1968, Van Furth introduced the concept of the mononuclear phagocyte system, stating that circulating monocytes continuously replenish tissue macrophages in homeostasis (van Furth and Cohn, 1968). Research over the last decade has revised this concept and consistently demonstrated that resident macrophages of most tissues are established prenatally by distinct waves of embryonic precursors from the yolk sac and fetal liver and maintained through adulthood by longevity and self-renewal independently of circulating monocytes (Ginhoux et al., 2010; Gomez Perdiguero et al., 2015; Guilliams et al., 2013b; Hoeffel et al., 2012, 2015; Merad et al., 2002; Schulz. Christian et al., 2012; Sheng et al., 2015; Yona et al., 2013).*However, under pathological conditions, monocytes constitute a major source of effector cells, as they can transform into cells with inflammatory macrophage- or DC-like phenotypes after infiltrating affected tissues. Most of our knowledge concerning the origin of monocytes is derived from studies in mice. Mainly based on MCFC analysis of bone marrow–derived cells, the progenitors of monocytes macrophage- or DC-like phenotypes after infiltrating affected tissues. Most of our knowledge concerning the origin of monocytes is derived from studies in mice. Mainly based on MCFC analysis of bone marrow–derived cells, the progenitors of monocytes were defined as macrophage and DC precursors (MDPs), which give rise to common monocyte progenitors (cMoPs) followed by differentiation into Ly6C$^+$ and Ly6C$^-$ monocytes.*

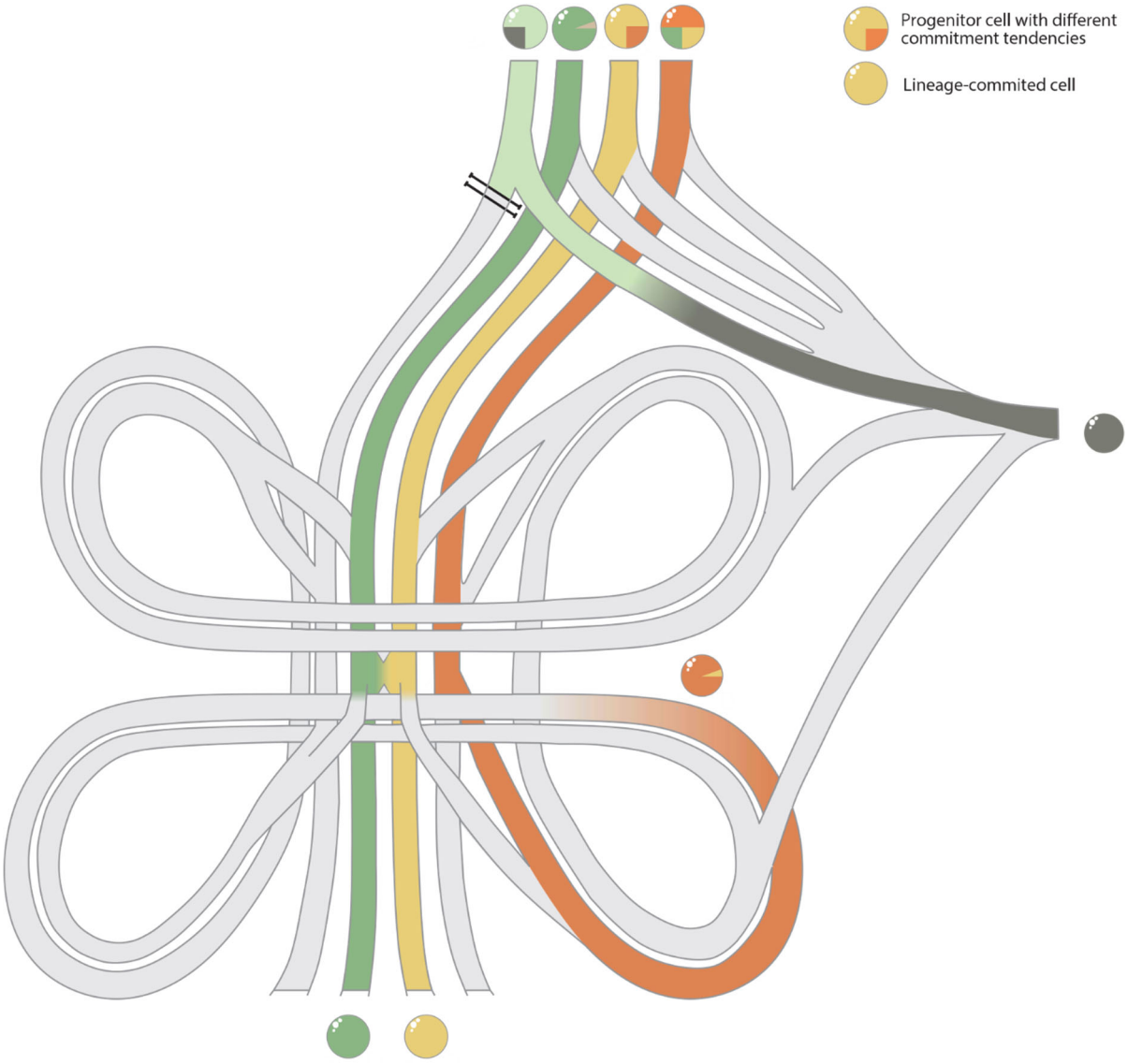


Figure 4. *The autobahn model of hematopoiesis and myeloid cell differentiation*

Since the understanding of early hematopoiesis and later differentiation into circulating and tissue-resident myeloid effector cells has changed dramatically during the recent past, we propose a new model describing the current understanding of its organization. […] The autobahn model aims at combining the concept of a continuous, hierarchical differentiation spanning multiple intermediate states with the cellular plasticity needed to explain the ability to cope with diverse physiological and pathological challenges. The autobahn model is an adaptation and extension of the classical Waddington landscape (Waddington, 1957), which is commonly used to describe cell differentiation. However, the autobahn model also takes into account the continuum of differentiation and both precommitment as well as the plasticity of a cell. We find the motorway junction presents a comprehensible analogy to the complex system of hematopoiesis. Traffic presents a continuous and active progress along straight paths to reach a predetermined destination while allowing course changes at multiple branches to respond to momentary situations, such as traffic jams. […] Similar

navigation system to find his or her way, epigenetic modifications direct a cell toward its destination and therefore predefine a cell's commitment toward a certain lineage. Additionally, the cell can dynamically respond to external signals, such as stress, inflammation, or other environmental factors, and integrate these with the internal disposition, enabling necessary deviations from its original path. In this scenario, autobahn interchanges present self-contained roads that, in principle, allow drivers to postpone their decision for the next route. Circulating in such autobahn interchanges would represent cellular dormancy and self-renewal of multipotent progenitor cells within the autobahn model. Accordingly, precursor cells persevere in specific self-contained niches to assure a steady supply of new cells following a certain path corresponding to required cellular phenotypes. Moreover, the autobahn interchange would be seen as an analog to the classically defined progenitor states, which should not be seen as discrete cell types but rather as transitory states. Finally, severe accidents or special events can cause one road to be impassable or jammed. Respectively, hematopoietic malignancies could be seen as the analogy of such events. They can affect specific lineages of the hematopoietic system and, for example, induce unrestricted overpopulation of certain cell types."

However, already in the early years of scRNA-seq, it was hypothesized that these proposed progenitor populations are actually heterogeneous (Ginhoux and Jung, 2014) and that scRNA-seq might shed light on the real developmental trajectories. For example, very recently a novel precursor monocyte population was identified in the bone marrow, positioned developmentally between the cMoP state and fully differentiated Ly6C+ monocytes - a transitional pre-monocyte population (Chong et al., 2016). […] *In a very recent study, it was proposed that there are human counterparts for cMoPs and the pre-monocytic population giving rise to fully differentiated monocytes (Kawamura et al., 2017).*

Whether the fully differentiated monocyte populations show further heterogeneity and whether this is dependent on environmental factors are other important questions being addressed by many groups. It was recently shown […] *that a subset of Ly6C$^+$ monocytes in the bone marrow seems to exist that is already predetermined to become either* […] *inflammatory macrophages or monocyte-derived DCs (moDCs) under inflammatory conditions (Menezes et al., 2016). Interestingly, another scRNA-seq study confirmed the presence of committed Ly6C$^+$ monocyte populations (e.g., moDC-primed monocytes) and found that they originate from different progenitors within the hematopoietic system (Yáñez et al., 2017).* […] *Collectively, all these studies favor a model of precommitted monocytes.*

However, alternative models have also been presented: It was recently hypothesized that all human blood-derived CD14$^+$ monocytes possess the potential to differentiate into inflammatory macrophages, whereas in the presence of certain environmental ligands […]*, monocytes mainly differentiate into moDCs (Goudot et al., 2017). Under steady-state conditions, CD14$^+$CD16$^-$ classical monocytes could not be further subdivided by scRNA-seq in this study. Furthermore, signatures derived from inflammatory macrophages were identified in all monocyte populations, including classical, intermediate, and non-classical subsets, suggesting that these cells are very homogeneous. Although intriguing, these interpretations should be taken with caution […], considering the sparsity of scRNA-seq data, confirming or rejecting the existence of precommitted monocytes might be hampered by missing data values. Clearly, this will require enrichment analyses with statistical models that enable the detection of small tendencies to identify subtle preferences as described in the precommitment model above.* […]

The definition of monocyte subsets was further complicated by a recent report suggesting four monocyte populations in human blood, including the classical and non-classical monocyte populations as well as two additional populations, with one expressing a cytotoxic gene signature (e.g., GNLY and CTSW) (Villani et al., 2017). Albeit interesting, these findings require further exploration […]*. Collectively, there is a strong necessity of future efforts to carefully link exciting new findings derived from scRNA-seq data to previous knowledge in the field. In the next iteration of defining the human myeloid cell atlas, we need to determine whether our previous classification of monocytes needs to be revisited and develop procedures that unequivocally define newly identified immune cell subsets in the framework of current knowledge.* […]

1.4.3. Tissue macrophages – one cell at a time

Macrophages constitute a broad family of tissue-resident, professional phagocytic cells, including […] *lung alveolar macrophages* […]*, that play an important role not only in defense against pathogens but also in tissue development and homeostasis (Ginhoux et al., 2016; Mass, 2018). Several studies have now been published using single-cell profiling to shed light on the ontogeny of these cells and their tissue-specific homeostatic and repair functions, as well as their diverse roles in immune responses.*

In contrast to the long-held view that circulating monocytes constitute the primary source for the replenishment of tissue-resident macrophages throughout life, we know today that resident macrophages of most tissues are established prenatally from embryonic precursors and, under homeostatic conditions, are maintained by self-renewal independently of monocytes (Ginhoux et al., 2010; Gomez Perdiguero et al., 2015; Guilliams et al., 2013; Hashimoto et al., 2013; Hoeffel et al., 2012, 2015; Merad et al., 2002; Schulz. Christian et al., 2012; Sheng et al., 2015; Yona et al., 2013). Under pathological conditions, monocytes infiltrate affected tissues and adapt inflammatory phenotypes and functions of macrophages and DCs, supporting efficient resolution of the inflammation (Ginhoux and Jung, 2014).How long these monocyte-derived macrophages (also referred to as inflammatory macrophages) persist within the tissues, adopt the phenotype of tissue-resident macrophages, and thus inconspicuously contribute to the population of tissue-resident macrophages is a topic of ongoing debate, as is how these compositional alterations of the resident macrophage population affect functional heterogeneity. Understanding the delicate mechanisms underlying macrophage ontogeny […] *and their heterogeneous homeostatic and immunological behavior require sophisticated approaches with single-cell resolution.*

Using scRNA-seq analysis of CD45^{+} cells […] *researchers under the leadership of Geissmann added an essential part to the puzzle of macrophage ontogeny: An intermediate cell type, labeled premacrophages,* […] *simultaneously colonize the whole embryo from embryonic day 9.5 while adapting a core macrophage expression program, which is diversified immediately after colonization by the expression of tissue-specific transcriptional regulators (Mass et al., 2016).* […]

Furthermore, […] *scRNA-seq studies focused on macrophages in the heart* […]*, in CNS* […] *and cardiovascular disease* [...]*. In fat tissue, adipose tissue macrophages (ATMs) constitute the predominant resident immune cells and accumulate in obesity, suggesting critical involvement in physiological and pathological processes. In a small-scale, explorative scRNA-seq analysis, two transcriptionally distinct ATM populations were identified (Hill et al., 2018). In follow-up experiments, these two subsets were characterized microscopically to be of unique morphology and tissue localization, and population-level omics analyses revealed distinct transcriptomic and epigenetic profiles.* […]

1.4.4. Defining single myeloid cells in major diseases

Cells of the myeloid compartment are of great clinical relevance, as the evidence for their implications in the etiology and pathophysiology of disease is ever growing. This is no coincidence: They function as tissue-resident sentinels to maintain tissue homeostasis or as patrolling cells ensuring the first line of defense for the immune system and reacting to pathogenic threats or endogenous inflammatory triggers. As described above, single-cell analyses have greatly advanced our understanding of the population structures of […] *monocytes and macrophages in different organs as well as delineated the respective subset characteristics - prerequisites to study these cells in pathophysiological contexts. Inflammation changes the composition of myeloid cell populations found in diseased tissue; for example, monocytes infiltrate and acquire characteristics similar to the tissue-resident cells depending on the tissue environment as well as disease conditions. Under these circumstances it is important to rely on more than a few markers for distinguishing cell types, which single-cell analyses clearly accomplish.* […]

Atherosclerosis is another pathophysiological condition in which myeloid cells play a crucial role in etiology. In two back-to-back studies the immune compartment within atherosclerotic plaques in a murine model was analyzed by scRNA-seq to describe the different leukocyte populations in the aorta and their changes in atherosclerosis, which nicely overlapped with data from a parallel CyTOF approach (Cochain et al., 2018; Winkels et al., 2018). In addition to the tissue-resident macrophages present in healthy and atherosclerotic aortic arches, plaques contained monocytes, moDCs, and two atherosclerosis-specific populations of inflammatory macrophages, one of which was newly characterized as Trem2+ (Cochain et al., 2018).

Monocytes are also recruited and beneficial during lung regeneration following pneumectomy, as characterization of the accumulating myeloid cell subsets by single-cell analysis determined them to promote alveolar epithelial stem cell proliferation (Lechner et al., 2017). On the other hand, a monocyte-derived macrophage population with a fibrosis-driving phenotype was specifically enriched in a mouse model for lung fibrosis (Aran et al., 2018). In the context of chronic kidney injury, it had been unclear which cell type contributes to the fibrosis in late stages of the disease. Kramann et al. (Kramann et al., 2018) determined this to be circulating monocytes by employing a

parabiosis model as well as scRNA-seq. […] *Collectively, these early examples applying novel single-cell technologies to define the heterogeneity of the myeloid cell compartment in major diseases give a first glimpse of the future potential of this approach for a better understanding of many chronic inflammatory conditions and hence the development of refined or novel therapeutic strategies. Finding alterations in the local immune environment at an early stage in disease as well as precise delineation of the subpopulation structure that may respond differentially to current therapies will allow the identification of new, effective therapeutic avenues.* […]

As outlined in this review, single-cell genomics - with scRNA-seq […] *- will provide us with novel molecular microscopes, allowing us to rewrite our understanding of myeloid cell biology. This will include all aspects of cell ontogeny, differentiation, homeostasis, and activation, including pathological activation during disease. These molecular microscopes with high-content information for each cell as well as spatial and temporal resolution will help us to better understand the cellular heterogeneity and plasticity of cells within the myeloid cell compartment.* […]

With the increasing evidence that the myeloid cell compartment is the major immune cell compartment involved in those chronic inflammatory diseases our communities are suffering from, we expect single-cell analysis of myeloid cells to become a driving force of biomarker exploration as well as diagnostic test and therapy development for all major diseases. […] *We strongly believe that single-cell genomics should lead the efforts in human immunology, as described recently by Davis and colleagues (Davis and Brodin, 2018).*

However, before we can routinely apply such sophisticated technology to medical questions, we must tackle numerous open questions and practical considerations […]*. In contrast to others in the field, we are not convinced that this will be a trivial task. Rather, this will require multi-science approaches beyond today's models of interdisciplinary research. Novel single-cell technologies often introduced by using cell lines or primary cells derived from animal models (Cao et al., 2017; Rosenberg et al., 2018) are repeatedly not tested for their applicability to human cells and - not surprisingly - often do not translate well to human cells under clinical conditions. In this respect, the myeloid cell compartment is particularly sensitive to technical procedures that can induce cell stress, thereby inducing either activation, cell death, or both (Orr et al., 2013; Zhang et al., 2008).*

The procurement and processing of fragile and precious human tissue samples constitute a major challenge in the effort to create comprehensive reference maps of all human cells. It is hardly feasible to obtain and process samples from multiple tissues in one experiment at one site, which leads to site- or experiment-specific technical differences prohibiting direct comparability. However, the integration of datasets from separate experiments is imperative for the comparison of human myeloid cell subtypes across different tissues. Besides rigorous standardization of experimental procedures, sophisticated computational approaches for data integration (Butler et al., 2018) are necessary to learn more about differences as well as conformities of heterogeneous myeloid cell populations across different tissues.

When analyzing human samples derived from blood or bronchoalveolar lavage, we need to ensure that certain fragile cell types including eosinophils […] *are not prone to loss by certain scRNA-seq technologies. Indeed, the impact of procurement and processing of human samples on data quality and data content is often underestimated. For example, while most of the droplet-based technologies perform very well with RNA-rich human or murine cell lines, measuring primary cells of the myeloid compartment that are easily activated and quickly undergo cell death is a much more challenging task. The results obtained by these technologies for these primary human cells are clearly inferior; for example, they result in significantly lower numbers of genes detected per cell or lower percentages of cells with sufficient numbers of genes detected. Therefore, additional methods, such as well-established MCFC, need to be included in the analysis of clinical samples to prevent over- or underestimation of particular cell types in clinical samples by single-cell genomics technologies.*

Also, when using clinically accessible tissues such as peripheral blood, certain aspects should be taken into consideration while designing experimental setups, since, for example, many studies aiming at identifying the cell population structure in peripheral blood use only PBMCs as the starting population. However, this excludes important cells of the myeloid compartment involved in many pathological processes. We speculate that the analysis of whole blood immune cells is more informative in chronic inflammatory diseases compared to focusing only on PBMCs. Furthermore, we will also need to learn how to best capture rare immune cell types, as these often are critical to certain pathological processes. This area requires substantial attention for

future studies in human immunology and in the Human Cell Atlas efforts (Regev et al., 2017). […]
Collectively, the revolution in single-cell genomics will bring new findings and medical applications to the field of myeloid cell biology, and we expect another decade of exciting research and development toward a better understanding of the role of these important immune cells in health and disease."

1.5. Analysis of scRNA-seq data

Since the data generated by scRNA-seq technologies comprise tens of thousands of genes in up to millions of cells, it can be categorized as 'big data'. Consequently, the scientific field is confronted with both analytical and computational challenges. This chapter is intended to provide an overview of the current analytical tools.

As part of this partly cumulative thesis, this chapter is an extract from the following publication (Baßler et al., 2019):

A bioinformatics toolkit for single-cell mRNA-Seq data analysis
Authors: *Kevin Baßler, Patrick Günther, Jonas Schulte-Schrepping, Matthias Becker, Paweł Biernat*

Publication details: *Published as book chapter in Single Cell Methods: Sequencing and Proteomics, Methods in Molecular Biology, vol. 1979; editor: Valentina Proserpio; Springer Nature 2019; Chapter 26; https://doi.org/10.1007/978-1-4939-9240-9_26*
Permission for use in the thesis: The permission was granted on 19.07.2019 by Patrick J. Marton, Executive Editor, Springer Protocols, Patrick.Marton@Springer.com

To ease the tracking of the self-quotation, the citation is placed in quotation marks and the content is written in italics. In addition, shortenings of the original text are indicated by [...] and the citation style and title numbering is adapted to the other texts of the dissertation. In addition, the term 'single-cell mRNA-Seq' is abbreviated as 'scRNA-seq' for consistency with the other texts.

"*The first step of every scRNA-Seq experiment is proper study design, which does not require sophisticated experimental or informatics skills but is nonetheless presumably the most important step. The quality of the resulting data strictly depends on the proper planning of the experiment, including the selection of the most suitable technology for the biological question of interest as well as an elaborated study design to minimize the influence of confounding factors. Once the experiment has been conducted and* […] *gene expression values have been extracted from the reads and normalized, the researcher has the agony of choosing between a plethora of analysis approaches to investigate diverse aspects of the single-cell transcriptomes, such as dimensionality reduction and clustering to explore cellular heterogeneity or trajectory analysis to model differentiation processes.* […] *Here, we introduce the main steps of a typical bioinformatics pipeline for the analysis of scRNA-seq data* […]. *Since each of these steps presents some downsides, we not only introduce algorithms, methods, and tools but also critically revise their applicability and limitations.* […]

1.5.1. Experimental planning

1.5.1.1. Choosing a scRNA-seq technology

One of the most important steps toward a successful application of scRNA-seq to a biological question is a detailed planning of the experiment. […]
Following the first description of single-cell mRNA sequencing in 2009 (Tang et al., 2009), a wide variety of scRNA-seq methods has been proposed. All methods have certain advantages, which demand an experimenter to choose a technique that is best suited for the biological question in mind. Regarding the gene body coverage of scRNA-seq data, two major protocol types have emerged. Full-length methods (e.g., SMART-Seq2 (Picelli et al., 2013) and Strt-Seq (Islam et al., 2011)) provide read coverage across the complete transcript allowing the investigation of, for example, alternative RNA processing. However, most available single-cell protocols (e.g., Drop-Seq (Macosko et al., 2015)*, Seq-Well (Gierahn et al., 2017) or sci-RNA-Seq (Cao et al., 2017)) sacrifice full-length coverage for the sake of early multiplexing, which minimizes the cost.*
Another important consideration during planning of a scRNA-seq experiment is the procedure of isolating single cells from a cell mixture. Early isolation protocols focus

on manual cell isolation techniques, such as micropipetting or laser capture microdissection. While these techniques gain spatial information about selected cells, their throughput is very low (Cadwell et al., 2016). Fluorescence-activated cell sorting (FACS) is a widely established technique that can be used for isolation of single cells. In addition, FACS records the protein expression of a cell, which allows to combine the proteome and transcriptome data derived from the same cell. It has been shown that this additional layer of information can be very valuable for characterization and investigation of cells of interest (Paul et al., 2015). However, the sorting procedure exerts stress to the cells in the form of high pressure and shear forces, which can change the transcriptome or even force the cells into apoptosis. Recently developed droplet-based isolation techniques, such as Drop-Seq (Macosko et al., 2015a) *and inDrop (Klein et al., 2015), have substantially decreased the cost while increasing the throughput. The same holds for cell isolation using microwell plates that allow for easy and fast separation of single cells into wells (Fan et al., 2015; Gierahn et al., 2017; Goldstein et al., 2017). Remarkably, very recent single-cell technologies do not rely on physical isolation/separation of single cells but rather perform each enzymatic step of scRNA-seq library preparation inside of a cell using a split-pool barcoding approach (Cao et al., 2017). These technologies rely on fixation of the cells, which might not be suited for all cell types with the existing protocols* […].

Another important question is the required number of cells that should be covered. Again, this is very much dependent on multiple factors, including cost, cell types, technology, and biological question. The number of required cells greatly depends on the assumed heterogeneity in the cell mixture. Since this is unknown for most of the experiments it helps to deploy other available resources (flow cytometry data, etc.), to estimate the expected heterogeneity. […] *As a rule of thumb, the less complex a heterogeneous cell mixture is, the higher the required information depth to detect heterogeneity, which can be increased by either increase of analyzed cells or by usage of a more sensitive approach (more genes detected). More cells analyzed means higher statistical power and lower impact of dropouts.* […]" A dropout event occurs when a transcript is not detected at all because of technical reasons. That is, some zeros in the count table do not mean that the respective gene is not expressed, but rather that its transcript was either lost during the library production or because of sequencing issues.

"***1.5.1.2. Reducing batch effects***

Since the starting material of single cells is very low, the generated gene expression measurements might be confounded because of differences in RNA extraction, enzyme activities, degradation/ fragment length, amplification, and sequencing depth. These effects become apparent when looking at batch effects. Batch effects are confounding factors, which occur because of, for example, different enzyme lots, and differences in personnel or preparation dates. However, batch effects can also occur within one experiment. Some of the available experimental protocols for single-cell genomics necessitate splitting of cells into different pools/batches during various steps of downstream processing. Although these pools are processed simultaneously, technical variation introduced due to processing in different batches is hard to avoid. A stringent study design may reduce their influence.

The easiest and most efficient way to account for batch effects is a proper experimental setup to begin with. To this end, it is important to balance biological conditions among batches to avoid a confounding study design. Ideally, all conditions (e.g., patient and control) should be represented and evenly distributed among all batches. […] *If processing of cells in batches cannot be avoided, it is important to include standards to estimate and ideally correct for the batch effects.* […]

Once the optimal technology and strategy have been found to answer the biological questions of interest and the respective dataset has been generated by trying to reduce the influence of potential batch effects, the actual analysis starts.

1.5.2. Computational aspects and challenges

1.5.2.1. Quality control

[…] *Gene expression data from single cells vary regarding the depth and quality of transcriptome information. It is important to account for such differences and it is crucial to remove cells from the analysis that are of low quality. Low quality data may be caused by a failure to capture a cell, capturing multiple cells, apoptotic cells, degrading RNA, low library complexity, or low sequencing depth. Dying cells have been shown to be associated with an increased ratio of reads mapping to*

mitochondrial genes and remaining endogenous genes. It is assumed that the membrane of apoptotic cells is leaking so that cytoplasmic RNA gets lost, but the mitochondrial RNA is retained causing an overrepresentation of mitochondrial transcripts (Ilicic et al., 2016). Since the ratio of mitochondrial genes to endogenous genes within a cell is highly dependent on the overall quality and experimental setup, the threshold should be determined in consideration of the distribution of this ratio among all cells within a dataset. However, usually the ratio is in a range of 5–20%. All cells exhibiting an exceptional high ratio should be considered apoptotic and removed from further analysis.

Moreover, cells with only a very low number of identified genes should also be removed from the analysis. However, it is important to keep in mind that different cells will vary in their number of identified genes, and great care must be taken to not bias the analysis by removal of certain cell types with lower intrinsic complexity.

Gene expression profiles generated from single cells contain a clear majority of zero measurements, either representing a failure of mRNA detection or a true missing of transcription of a gene. It was shown that variance is highly correlated with the mean expression (Grün et al., 2014). It is recommended to remove lowly expressed genes to limit their effect on the variance within a dataset. The identification of these genes can be performed based on the number of cells that express a certain gene. If a gene is expressed in less than 1% of the cells it is unlikely that this gene contributes to the overall variability. Since this data quality check is of very high importance, several pipelines and tools like Seurat (Butler et al., 2018) […] *have been developed. These tools suggest filtering low-quality cells by analyzing multiple QC parameters* […].

1.5.2.2. Normalization

An indispensable step for proper scRNA-seq analysis is the normalization of the data to make the transcriptome of the cells comparable to each other. Some of the normalization tools were initially developed for bulk data but have been successfully applied to single-cell data as well. Generally, one has to distinguish between within-sample normalization, which corrects for gene-specific biases and between-sample normalization, which adjusts for distributional differences across cells (e.g., read/transcript number). […]

A popular normalization method which is commonly applied to single-cell technologies with full-transcript coverage (such as SMART-Seq2 (Picelli et al., 2013)) is the TPM (transcripts per million kilobase) method. This method is used to normalize for differences in sequencing depth across samples (between-sample normalization) and is related to the RPM (reads per million) method whose principle is commonly applied to single-cell data. However, TPM also considers the gene length (within-sample normalization) and thus is very similar to RPKM (reads per million kilobase) and FPKM (fragments per kilobase million). However, for the sake of cell-to-cell comparison, TPM is more powerful. The main disadvantage of these estimates is that they can be dominated by a handful of highly expressed genes, which can bias the downstream analysis (Vallejos et al., 2017). […]

For technologies, which are capturing the 3'-end (e.g., Drop-Seq (Macosko et al., 2015)*, Seq-Well* (Gierahn et al., 2017) *or sci-RNA-Seq (Cao et al., 2017)), another type of normalization strategy can be introduced, namely via the usage of UMIs (unique molecular identifiers). These sequences barcode individual mRNA molecules and hence can be used to account for amplification bias which is a major source of technical variation. Because of technical limitations, UMIs are not usable for all single-cell technologies.* […]

1.5.2.3. Accounting for other unwanted sources of variation

Besides the aforementioned technical sources of variation, there are additional factors that might contribute substantially to the variability of gene expression. For some datasets, cell-to-cell variation can also reflect the cell cycle stage at which a cell was captured. In more detail, a proliferating cell upregulates its gene expression and hence will contribute more to the read pool in the single-cell library compared to a resting cell. Although normalization (RPM, RPKM, FPKM, and TPM) will account for some of this variability, it will not be able to remove all of the cell cycle-related variability.

One possibility to account for cell cycle-related effects is the usage of scLVM (Buettner et al., 2015), which builds on a latent-variable model based on Gaussian processes. Although scLVM was designed to account for cell cycle-induced variations, it can also be used to correct for other sources of variation, which can be modeled by latent variables. Inspired by scLVM, Satija and colleagues implemented the modeling of

latent variables in their single-cell analysis pipeline Seurat […] *and used them to account for differences in the alignment rate, expression differences, and mitochondrial gene expression. In principle, latent variable models can also be used to account for batch effects in datasets* […].

Batch effects are also a major problem of meta-analysis, when different single-cell datasets are combined in one analysis. Butler et al. recently developed an elegant method (Butler et al., 2018), which circumvents the limited comparability across datasets because of batch effects. This method relies on a strategy that identifies a shared structure (common sources of variation) between the different datasets based on canonical correlation analysis (CCA) followed by an alignment based on this structure. This CCA-based procedure […] *was successfully benchmarked in datasets confounded by different treatment conditions, technologies, and even species.*" Recently, an improved and further developed version of the CCA-based method has been published, which relies on an 'anchoring' strategy for the integration of different datasets (Stuart et al., 2019).

"*1.5.3. Exploring cellular heterogeneity*

1.5.3.1. Dimensionality reduction

A count table may contain counts for tens of thousands of different genes, more if we count different isoforms separately. Because every cell is characterized by a large number of values we say that the single-cell data is high-dimensional. This high-dimensional data often contains redundant information and can be summarized in a lower-dimensional space by applying a dimensionality reduction algorithm. The dimensionality reduction serve several purposes, firstly, it can be used to summarize the data by plotting it in a lower-dimensional space (2 or 3 dimensional). Secondly, the dimensionality reduction can be used as a preprocessing step before applying other algorithms (like clustering) to improve their efficiency (both computational-wise and by removing noise from the data).

[...] *The principal component analysis (PCA) (Pearson, 1901) is a fast and scalable algorithm that finds directions in the original space in which the data varies the most. These directions (principal components, or PCs) are sorted from the most to the least varying. Based on how much variation is captured in each of the PCs, we then specify*

how many of the PCs we wish to keep and discard the rest. Our data is then projected onto the remaining PCs in effect reducing the number of dimensions. Assuming that the biologically relevant information is responsible for the most variation in our data, by removing low varying components we discard only technical noise. The output of the PCA is therefore used as an initial denoising step and the resulting medium-dimensional data can be further analyzed. […]

The other goal of dimensionality reduction, visualization, is typically accomplished by much more sophisticated algorithms. The PCA is normally insufficient for this purpose, unless the data is extremely simple, because PCA is a linear transformation: it only shifts, rotates, and scales the original space. If the data […] *forms more complex structures, PCA alone* […] *cannot cope with simplifying these structures to present them in a two- or three-dimensional plot. Instead, nonlinear dimensionality reduction methods have to be applied, of which the most popular one is t-SNE (Maaten and Hinton, 2008). T-SNE computes the local relationships between points in the original high-dimensional space and places the points in a lower-dimensional space (normally two or three dimensional) in such a way as to preserve these local relations. Relying only on the local structure of the data allows it to simplify complex structures and lay them out in a 2D space in a clear fashion.*

Similar to other algorithms, there is a loss of information after dimensionality reduction, and t-SNE is no exception. It sacrifices the global structure of the data to preserve the local relationships. In effect, the t-SNE results can be difficult to interpret (Wattenberg et al., 2016). For example, if after performing t-SNE the cells are depicted as several separate clusters there is no way to tell how these clusters relate to each other in the original space. […] *The issues with t-SNE are common to other nonlinear dimensionality reduction algorithms* […]. *The danger here is two sided: we can lose the biologically relevant information after applying dimensionality reduction, or, perhaps even more dangerously, we can overinterpret the results and see structures, which are not really there.* […]

Aside from t-SNE other dimensionality-reduction algorithms are being applied to single-cell data, often offering better performance or interpretability. […] *The recently introduced UMAP (Becht et al., 2018; McInnes and Healy, 2018) could be used as a more scalable alternative to t-SNE.*" UMAP is the abbreviation for "Uniform Multiple Approximation and Projection" and describes a method for reducing dimensionality, in

which as much of the local and more of the global structure of high-dimensional data is retained as in t-SNE.

"***1.5.3.2. Developmental trajectories***

For an analyst of single-cell data, it is important to keep in mind that not all […] *algorithms are suitable for any biological problem of interest. For example, if one is interested in the transitional states of cells, it is not advisable to use t-SNE* […] *as dimensionality reduction method.* […] *To enable the inference of dynamic biological processes (e.g., cell cycle, cell activation or differentiation), a plethora of different approaches have been developed to model such trajectories. The assumption of these methods is that the recorded single cells are at different stages of the dynamic process and hence the trajectory can be computationally modeled by taking the information of all single cells into account. To this end, the cells are ordered along a pseudotime in a trajectory, which can have a simple linear shape, but also complex bifurcated structures like developmental trees are possible. By building a trajectory, the analyst can answer different biological questions, for example, in the context of differentiation trajectories, the identification of rare precursor cells, or the stage where a bifurcation occurs which means a stage in the pseudotime where cells undergo fate decision and branch into distinct differentiation directions.* […]

A recent paper by Saelens et al. comprehensively assessed the performance and robustness of different trajectory inference tools (Saelens et al., 2018). Most of the evaluated methods worked best for datasets containing topology type they were supposed to handle. For example, methods designed for linear trajectory types commonly performed best for datasets representing these types of structures. Consequently, an analyst needs to know a priori the underlying topological structure of the dataset, which is often difficult. Nevertheless, Saelens et al. provide guidance in form of a decision tree to help users decide which trajectory inference method is most suitable for the dataset of interest. For example, if the trajectory is expected to have a linear topology, SCORPIUS (Cannoodt et al., 2016) is the method of choice. Moreover, the authors recommend to use reCAT (https://github.com/tinglab/reCAT) for cycle topologies, Slingshot (Street et al., 2017) for bifurcated trajectories and Monocle DDRTree for complex tree trajectories. In the future, methods will be needed,

which are not designed for a specific topology and hence can efficiently model biological relevant trajectory without the necessity of a priori knowledge." An elegant alternative approach for trajectory inference has recently been introduced by the Kharchenko group (La Manno et al., 2018), based on unspliced and spliced mRNA information in single-cell data to model the transcriptional dynamics (RNA velocity) that can be used to predict the future state of individual cells in silico. A major drawback of trajectory analyses, however, is that current methods have not been developed to cope with batch effects, although this will be necessary for future large-scale clinical single cell studies.

"***1.5.3.3. Clustering***

Clustering algorithms serve to label similar cells in preparation for further analysis (counting, comparing differentially expressed genes, etc.). The clustering is normally performed on a 2D representation of the data, which means that the results rely heavily on the dimensionality-reduction algorithm. Clustering based on t-SNE […] *result is especially dangerous here, as it may result in artificial clusters which are not reflecting the complete underlying biology. Therefore, there is a recent trend in single-cell analysis to perform clustering on the high- or an intermediate-dimensional data […]. Graph-based clustering algorithms like Louvain clustering* (Blondel et al., 2008a) *(implemented in Seurat and scanpy* (Wolf et al., 2018)*)* […] *belong to this group.* […]

1.5.3.4. Identifying subpopulations

After identifying groups of cells that exhibit a high similarity in their gene expression profile, it can often be helpful to link the identity of these clusters to the established knowledge of cellular biology. Visualizing the expression of known marker genes in the respective clusters of cells in, for example, a violin plot or by color-coding the cells in their low-dimensional representation (PCA or t-SNE plot) is a quick and easy way to link a priori knowledge to the scRNA-seq data. However, it is important to keep in mind that especially lowly expressed genes might be affected by […] *technical noise […]. Therefore, rather than evaluating single marker genes, we recommend to use*

sets of genes, often referred to as cellular gene signatures, to assess the biological identity of the clusters. The current knowledge of cell types and their states is limited; thus, knowledge-driven classification must necessarily fail for undescribed cell types. However, due to its unbiased and encompassing nature, scRNA-seq presents unprecedented capabilities to readily identify novel cell types and thus expand the knowledge base (Mass et al., 2016).

One strategy to characterize groups of cells of unknown identity is based on the unbiased identification of marker genes specifically expressed in these clusters (often determined by differential gene expression analysis). Given such a list of genes, we can refer to the literature to further determine the characteristics of cells from this cluster. Moreover, the identified marker genes can be used for Gene Ontology enrichment analysis (GOEA) or gene set enrichment analysis (GSEA) to test whether identified marker genes significantly overlap with gene ontology terms or other gene sets […]. *This way, we can characterize the new cell type by its function and similarities to other cells.* […]

As repeatedly mentioned in the sections above, there is no one-fits-all solution to analyze any single-cell data. An analyst is demanded to carefully choose which methods and algorithms to use at various steps of analysis. It is very likely that an inexperienced user will be quickly overwhelmed. Although some tools have recently emerged that offer a guided analysis of data, they will reach the limits of the software's analytical capacities relatively fast. Therefore, we encourage any emerging analyst to learn the basics of a programming language, such as R, to open the door to a broader understanding of analysis and hence to exhaust the possibilities of scRNA-seq." In that way, the scRNA-seq technology and the analysis of the data generated from it will become a routine tool in laboratories and will allow us to expand our fundamental understanding of biological processes and the pathophysiology of diseases, such as cancer (Suvà and Tirosh, 2019) or chronic diseases, to an unprecedented extent.

1.6. Chronic obstructive pulmonary disease

Because of its central role in human physiology, lung dysfunction is associated with the manifestation of severe diseases. From a clinical-diagnostic perspective, a distinction is made between obstructive and restrictive (and sometimes

neuromuscular) ventilation disorders. Restrictive lung diseases (e.g. idiopathic pulmonary fibrosis) are characterized by a limited expansion capacity of the lungs during inhalation, while obstructive lung diseases comprise conditions that make it difficult to exhale all the air in the lungs, for example due to narrowing (obstruction) of the airways. Known examples of obstructive lung diseases are asthma and chronic obstructive pulmonary disease (COPD).

COPD is a progressive lung disease that is expected to become the third leading cause of death worldwide within 2020 (https://goldcopd.org/). The global prevalence of COPD is 10.1% (Celli and Wedzicha, 2019) and according to The Global Burden of Disease Study, it is estimated that 3.2 million deaths worldwide were caused by COPD in 2017, a number that is likely much higher because of un-reported cases and which is expected to increase to 4.4 million per year by 2040 (Rabe and Watz, 2017; Roth et al., 2018). The increase in mortality may be driven by factors such as the increase in noxious particles inhaled, e.g. from smoking or air pollution, the ageing of the world population, particularly in industrialized countries, and lower mortality from other diseases that are now better treatable. China and India account for more than 50% of all COPD cases, with the majority of COPD deaths occurring in low and middle-income countries (Celli and Wedzicha, 2019). A systematic review of COPD patient data collected over a decade in 28 countries showed a higher prevalence of the disease among smokers compared to non-smokers, among those who are ≥ 40 years old compared to < 40 years, and among men compared to women (Halbert et al., 2006). Besides the enormous medical burden, COPD is also associated with a considerable economic burden. In the EU, the annual cost of COPD is 38.6 billion euros, or 56% of the total annual health budget needed to treat lung diseases (https://www.who.int/gard/publications/The_Global_Impact_of_Respiratory_Disease.pdf).

This chapter summarizes the clinical aspects of COPD according to the 2020 report of the Global Initiative for Chronic Obstructive Lung Disease (GOLD) (https://goldcopd.org/) and the current scientific knowledge.

1.6.1. Definition of the disease

COPD is not considered a single disease but a clinical syndrome with structural lung abnormalities, impaired lung function, chronic respiratory symptoms or a combination of these (Celli and Agustí, 2018). The most common symptoms of COPD are shortness of breath (dyspnea), chronic coughing and increased sputum production. In addition, many patients also suffer from wheezing and chest tightness, especially during physical exertion. Since these symptoms are rather unspecific, they are sometimes not sufficiently reported by patients and complicate the diagnosis. A particular hallmark of this disease is that the clinically characteristic reduction in airflow caused by respiratory and/or parenchymal abnormalities (e.g. emphysema) and the associated symptoms are persistent and progressive. The idea that COPD is not a single disease is based on the observation that different clinical phenotypes have been observed and thus led to various taxonomies that have been proposed in recent decades and is still the subject of an ongoing debate (Al-Kassimi and Alhamad, 2013; Celli and Agustí, 2018; Makita et al., 2007; Mirza and Benzo, 2017; Pikoula et al., 2019). One of the most prominent taxonomies of COPD subtypes was first described in 1955 by Dornhorst (Dornhorst, 1955), who divided COPD patients into the emphysematic type, the so-called ‘pink puffer’, and the chronic bronchitis type, the so-called ‘blue bloater’ (Filley et al., 1968). Generalized and stereotyped, the latter group of patients is mainly defined by overweight people with bluish skin color due to shortness of breath and chronic bronchitis, while "pink buffers" breathe fast and suffer from shortness of breath and pursed lip breathing, due to emphysema, resulting in pink face color. However, this characterization at best describes COPD extrema, while most patients in the clinic tend to have a mixed form of chronic bronchitis with emphysema formation. Therefore, this taxonomy is no longer used in the clinic. All current classifications of COPD phenotypes are solely based on clinical parameters, which will be defined in more detail later in the text.

1.6.2. Factors contributing to the development and progression of COPD

The primary and most studied risk factor for the development of COPD is active and passive cigarette smoking (Laniado-Laborin, 2009; Putcha et al., 2016; Ramírez-

Venegas et al., 2014; Yin et al., 2007). However, there are also other risk factors, since epidemiological studies show that non-smokers may also develop COPD (Lamprecht et al., 2011). For example, occupational exposures, which are inhaled at workspace and comprises vapors, gases, dusts and fumes, are another important risk factor of COPD. A large study investigating the prevalence of developing COPD among non-smokers indicated that occupational exposure accounts for more than 30% of COPD cases (Hnizdo et al., 2002). Especially in developing countries, another source of particulate exposure and thus another risk factor exists, namely from the indoor burning of biomass in open fireplaces or poorly functioning stoves (Barnes et al., 2015). According to WHO, almost 3 billion people worldwide cook and heat with kerosene, biomass and coal (https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health), which leads not only to indoor air pollution but also to environmental pollution. However, the impact of outdoor air pollution on the development of COPD in adults remains controversial.

Like many other diseases, the development of COPD is the consequence of a complex interplay between environment and genes. Some genetic risk factors have already been identified, such as the hereditary deficiency of α1-antitrypsin, an inhibitor of serine proteases encoded by the *SERPINA1* gene. One of the main functions of α1-antitrypsin is the protection of the lung from proteolytic damage as a result of inflammation, so that a deficiency in this serine protease inhibitor due to gene mutation can lead to emphysema formation and the development of COPD (Stoller and Aboussouan, 2005). Other genetic mutations that were associated with increased risk of COPD, were identified at the gene locus coding for the glutathione-S-transferase (Ding et al., 2019), which is important for the detoxification of reactive oxygen species (ROS). On the other hand, genetic changes associated with a reduced risk of developing COPD have also been found, for example in the genes encoding for matrix metalloproteinases (MMPs) (Hunninghake et al., 2009), which degrade extracellular matrix proteins and thus contribute to emphysema formation (Churg et al., 2012; Kapellos, Bassler et al., 2018). A recent large-scale genome-wide association study has identified more than 80 genomic loci associated with an increased risk of developing COPD (Sakornsakolpat et al., 2019), but further analyses are needed to assess whether the identified loci are actual causative factors of the disease.

Another risk factor, which is determined by the inherent characteristics of an individual, is ageing, as there is growing evidence of a close association between ageing and

COPD, as reflected in the abovementioned increasing prevalence of COPD in elderly people (Brandsma et al., 2017). Although normal physiological aging of the lungs is already characterized by tissue changes, the so-called ‘senile emphysema’ (Mercado et al., 2015), it is not known whether healthy aging over time leads in general to COPD. However, since COPD can also manifest already in middle-aged people, the anatomical changes and cellular hallmarks in the lungs are described as ‘accelerated lung aging’ (Mercado et al., 2015).

Interestingly, processes in the earliest phases of life can also determine whether there is an increased risk of developing COPD. Thus, any factors influencing lung growth and development in the time of gestation until childhood are associated with an increased risk of COPD. For example, a meta-analysis study found a positive association between birthweight and later lung function (Lawlor et al., 2005), which was confirmed by another study (Cai et al., 2016). Importantly, reduced lung function established during childhood could be a contributing factor in up to 50% of people, who develop COPD (Lange et al., 2015).

A factor that contributes to reduced lung function at various stages of life is airway hyper-reactivity, such as asthma. The European Community Respiratory Health Survey identified airway hyper-reactivity as the second most important risk factor for developing COPD (approx. 16% of COPD cases) directly after smoking and identified respiratory infections as another important denominator (De Marco et al., 2011).

1.6.3. Pathophysiology and diagnosis

Pathological changes in the lungs of COPD patients affect the airways, the parenchyma and the pulmonary blood vessels and are caused by chronic inflammation and associated structural changes (including emphysema formation) of the lung (Hogg and Timens, 2009). These structural changes increase with the severity of the disease and are currently not reversible.

Overall, the pathological changes in the lung of COPD patients lead to an airflow limitation during expiration, which can be measured clinically with spirometry and is currently the only empirical measure for diagnosis. Some of the standard diagnostic parameters obtained by spirometry comprise the volume of air that can be forcibly exhaled after maximal inspiration (referred to as forced vital capacity, FVC) and the

volume that can be forcibly exhaled within one second (referred to as forced expiratory volume in one second, FEV_1). The division of these two values (FEV_1/FVC) is known as Tiffeneau-Pinelli index or FEV_1 %, in which a value of < 0.7 indicates an obstructive airflow limitation (Lange et al., 2016). As mentioned above, COPD is persistent, which means that it can be distinguished from asthma, which is largely reversible after administration of bronchodilators. Therefore, spirometry is performed at least twice, once without and once with administration of a bronchodilator. A Tiffeneau-Pinelli index of the latter being below 0.7 is an indication of persistent airflow limitation and thus of COPD (Vestbo et al., 2013). However, the Tiffeneau-Pinelli index also has some limitations, e.g. it overestimates the presence of COPD in older patients (Lange et al., 2016). Therefore, the Tiffenau-Pinelli index of a patient is set in relation to the average Tiffenau-Pinelli index in the population of similar age, sex, race, and body weight, resulting in the so-called FEV_1% predicted (Pellegrino et al., 2005). Interestingly, it was found that the correlation between spirometry measure and the actual symptoms and associated health status of the patient is very weak (Jones, 2009). This is very interesting, because the health status of a patient is a better predictor for mortality than, for example, FEV_1 (Jones, 2009). For the assessment of symptoms, a questionnaire, such as the COPD Assessment Test (CAT) (Jones et al., 2009), is used, which is answered by COPD patients. The CAT questionnaire includes questions such as the frequency of coughing or the severity of breathlessness after exertion, and is scored from 0 to 5 by patients, whereby 5 represents severe symptoms. In addition to spirometric and symptomatic evaluation, the diagnosis of COPD also includes an assessment of the risk of exacerbation. An exacerbation is defined as a rapid worsening of the respiratory symptoms, which requires additional therapy with possible hospitalization. It is mainly caused by acute infections with viruses and bacteria of the respiratory tract that can trigger an increase of the chronic inflammation of the lung, which is associated with a substantial mortality risk (Wedzicha and Seemungal, 2007). According to recommendations of the 2020 report of the Global Initiative for Chronic Obstructive Lung Disease (GOLD) (https://goldcopd.org/), the diagnosis and stratification of COPD should include spirometry, symptom evaluation and assessment of exacerbation (**Figure 5**). Overall, COPD is identified with a post-bronchodilatory Tiffenau-Pinelli index below 0.7, classified (GOLD grades) by the severity of airflow limitation based on the FEV1%

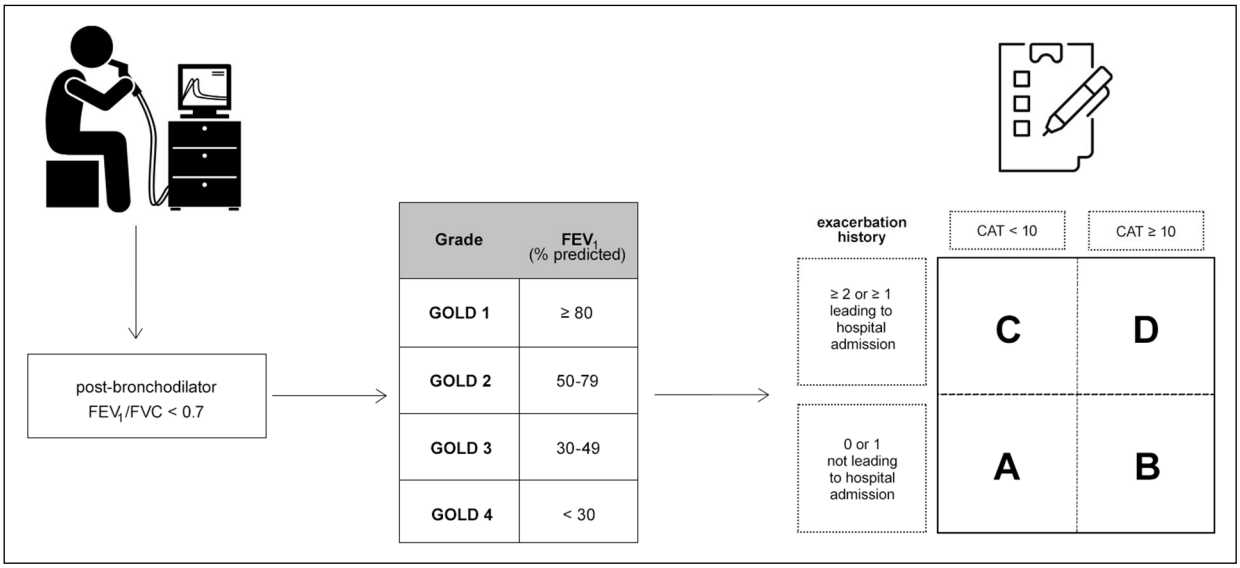


Figure 5. Current guidelines for diagnosis of COPD, classification by grade (GOLD1-2) and assessment of the severity of symptoms and the risk of exacerbation (ABCD assessment) (acc. to recommendations of the 2020 report of the Global Initiative for Chronic Obstructive Lung Disease (GOLD))

predicted value and sub-classified according to the severity of symptoms and the risk of exacerbation (ABCD assessment).

1.6.4. Pathogenesis

The pathogenesis of COPD is complex with numerous co-existing mechanisms and factors (Agustí and Faner, 2018; Agustí and Hogg, 2019), including increased apoptosis (Voelkel et al., 2004), failure of lung tissue maintenance (Tuder et al., 2006), protease/antiprotease imbalance (Abboud and Vimalanathan, 2008; Pandey et al., 2017), cellular senescence (Houssaini et al., 2018) and immunosenescence (Barnes, 2017). Moreover, there is increasing evidence that oxidative stress and the associated oxidative tissue damage is an important and predisposing factor in the development of COPD (Bowler et al., 2004; Domej et al., 2014; Kirkham and Barnes, 2013; McGuinness and Sapey, 2017). Oxidative stress is characterized by elevated ROS levels, which can cause damage to lipids, proteins and DNA and can be either derived from extrinsic, mainly from cigarette smoke, or from endogenous sources, such as from dysfunctional mitochondria (Wiegman et al., 2015) or activated inflammatory cells

(Rahman, 2005). However, the central factor and thus the determining mechanism of COPD pathogenesis is the chronic inflammation of the lung tissue (Barnes et al., 2015).

1.6.5. Immune cells in COPD

COPD is associated with enhanced recruitment, infiltration and activation of immune cells in the lung tissue (Ni and Dong, 2018). For example, altered cell numbers in COPD have been described for $CD8^+$ T cells (O'Shaughnessy et al., 1997) and innate lymphoid cells (ILCs) (De Grove et al., 2016). However, especially neutrophils and macrophages were found to be increased in COPD and thus gained much attention in recent years (Barnes, 2019). The inhalation of cigarette smoke or other irritants leads to the activation of airway epithelial cells and AMs, which in turn release, for example, the chemokine IL-8 that binds to CXCR2 expressed on neutrophils (Barnes, 2008). Neutrophils are the first-line defense of innate immunity and perform their functions by releasing ROS and proteases such as neutrophil elastase and cathepsin, which exert an antibacterial function (Hoenderdos and Condliffe, 2013a). However, the inherent function of these cells might contribute to a pronounced oxidative stress and protease/antiprotease imbalance during the chronic infiltration of the lung with neutrophils, leading to the progression of COPD. Thus, it was shown that the number of neutrophils in the lungs of patients could serve as a predictor of COPD exacerbations (Kinose et al., 2016). In addition, a previous study used the neutrophil to lymphocyte ratio in peripheral blood as further predictor of exacerbation (Lee et al., 2016). However, the efficacy and robustness of neutrophils as a biomarker needs to be evaluated in larger studies in the future. Interestingly, despite the central role in innate immunity and increased numbers in the lungs of COPD patients, therapeutic strategies targeting neutrophils, including CXCR2 antagonists, have not been clinically effective (Barnes, 2013), indicating that a better understanding of potential neutrophil subtypes and of the general immune response that orchestrates COPD is still needed. Another type of granulocyte, namely eosinophils, has also become the focus of many studies in recent years (Saha and Brightling, 2006; Tashkin and Wechsler, 2018). In the past, it was assumed that elevated eosinophils were more characteristic of asthma, while COPD is a disease characterized by neutrophils (Barnes, 2008). However,

studies have shown that an increased number of eosinophils can also be detected in the lungs of COPD patients (Berg and Wright, 2016; Saha and Brightling, 2006). Moreover, clinical observations suggested that COPD patients with eosinophilia in sputum or blood react better to inhaled corticosteroids, a treatment that otherwise seems to be ineffective in COPD (Agusti et al., 2018). Thus, eosinophils as biomarker for the prediction of corticosteroid therapy efficacy is now even included in the recommendations of the 2020 report of the Global Initiative for Chronic Obstructive Lung Disease (GOLD) (https://goldcopd.org/).

Although the inflammation underlying COPD is complex and involves a plethora of different immune cells, it is now recognized that especially AMs play a central role in the pathogenesis of COPD. The current knowledge about these cells in the context of COPD is discussed in the next section.

1.6.6. Alveolar macrophages as orchestrators of COPD

As the lungs need to adapt to ever-changing environmental challenges, so do the AMs constantly adapt to satisfy the needs of the tissue and thus ensure lung maintenance. For example, the O_2 partial pressure in the alveolar space can exhibit strong fluctuations, so that hypoxia of some parts of the lung is a quite common situation even under healthy conditions (Hussell and Bell, 2014), which in AMs requires a high degree of diversification to enable quick adaptation.

Interestingly, this remarkable diversity of AM responses, and the cellular and immune effects mediated by them, can in principle explain most of the known features of COPD (Barnes, 2004). For example, some studies reported that, upon COPD, AMs transform their rather quiescent phenotype into an activated/pro-inflammatory one by producing more cytokines (such as TNF-α) (Mukhopadhyay et al., 2006) and chemokines (such as IL-8) (Barnes, 2004), which lead to the recruitment of other immune cells and thus contribute to the inflammatory profile in COPD (Kapellos, Bassler et al., 2018). However, interestingly, the expression of TGF-β, which is also a known key homeostatic molecule of the lung (Garbi and Lambrecht, 2017), is also upregulated in COPD (Barnes, 2004), indicating either a mixed phenotype of AMs in COPD or AM subtypes with different immunophenotypes. The COPD-dependent functional change in AMs is also reflected in the pronounced upregulation of secreted extra cellular

matrix-degrading enzymes such as MMPs and cathepsins (Barnes, 2004; Kapellos, Bassler et al., 2018; Russell et al., 2002), which contribute to a protease/antiprotease imbalance and hence to tissue damage. Another mechanism by which AMs switch from lung tissue preserving to tissue destroying function is through the production of ROS. Upon activation, macrophages typically produce ROS to kill bacteria and other invading microorganisms, however, in COPD, AM-mediated ROS production seems to be deregulated and causes oxidative stress in the lung tissue (Kirkham and Barnes, 2013). A possible mechanism of increased endogenous ROS production in COPD is associated with mitochondrial dysfunction (Boukhenouna et al., 2018). A recent study associated this mitochondrial dysfunction in AMs of COPD patients with a reduced phagocytosis capacity of these cells due to an impaired mitochondrial membrane potential causing a deficit of the energy required for proper phagocytosis (Belchamber et al., 2019). In the context of tissue homeostasis and immune surveillance, phagocytosis of pathogens and debris, as well as the clearance of accumulating apoptotic cells via efferocytosis plays a central role in AMs (Vandivier et al., 2006). Interestingly, a reduction of phagocytosis and efferocytosis, is a well-known feature of AMs from COPD patients (Kapellos, Bassler et al., 2018) and some scientists consider them to be the key components in the development of exacerbations (Han et al., 2017; Hurst et al., 2010; Jubrail et al., 2017).

Several of the more recent scientific efforts focus on the metabolic state of lung cells, including AMs, in COPD (Chen et al., 2019). Although some interesting observations have already been made in this area, such as the systemic increase in cholesterol levels in COPD patients (Zafirova-Ivanovska et al., 2016), further analysis is needed to determine the metabolic state in COPD, especially in AMs, which are responsible for surfactant homeostasis.

Since many of the described defects and deregulation in AMs can also lead to cell death, there seems to be a discrepancy regarding the observed upregulation in macrophage numbers in COPD lungs (Barnes, 2019). A hypothesis for the increase in AMs is an elevated self-replenishment of macrophages by proliferation (Barnes, 2004), but proof for this is still pending. Another potential source for AMs are circulating monocytes, which infiltrate the lung and differentiate into macrophages. Indeed, mouse models have shown that circulating monocytes can readily differentiate to AMs *in vivo* (Landsman and Jung, 2007). In addition, a mouse smoke model, mimicking COPD, showed that infiltrating monocytes not only give rise to AMs but may also be

responsible for the observed upregulation of TNF-α and MMPs in COPD patients (Pérez-Rial et al., 2013). The idea of an involvement of systemically circulating cells to COPD pathogenesis is also in agreement with the hypothesis that COPD might be the pulmonary manifestation of a chronic systemic inflammation (Fabbri and Rabe, 2007).

Taken together, the overall phenotype of AMs in COPD is altered, and AMs appear to loose full functionality. Although AMs clearly play a key role in COPD, no detailed definition of potential subtypes or analysis of potential different functions has been performed so far. In principle, the investigation of AMs is straightforward, as the cells are obtained from bronchoalveolar lavage fluid (BALF), which is obtained during bronchoscopy of the lung, and isolation requires neither enzymatic digestion nor longer incubation periods, which can lead to changes in the phenotype and function of the cells (Garbi and Lambrecht, 2017).

1.6.7. Therapy

There is currently no effective therapy available to cure COPD, but there are measures that can help alleviate the symptoms. Among the most effective therapies is to support patients in smoking cessation, which can substantially reduce COPD-associated mortality (Tashkin, 2015). In addition, according to the 2020 report of the Global Initiative for Chronic Obstructive Lung Disease (GOLD) (https://goldcopd.org/), the most commonly used pharmacological therapies focus on relieving symptoms and comprise the administration of both long-acting muscarinic antagonists (LAMAs) and long-acting β2 agonists (LABAs) or a combination of both. In patients with recurrent exacerbations or severe dyspnea, triple therapy with inhaled corticosteroids (ICS) is also frequently used. However, the general efficacy of corticosteroids in COPD patients is rather low. Therefore, novel anti-inflammatory therapies are currently tested, including broad-spectrum anti-inflammatory treatments (such as phosphodiesterase-4 inhibitors), cytokine and chemokine inhibitors (such as anti-TNF and anti-CXCR2), anti-proteases, antioxidants, kinase-inhibitors (such as phosphatidylinositol-kinase inhibitors), or drugs aiming to reverse corticosteroid resistance (Barnes, 2013). In addition, there are non-pharmacological treatment strategies applied including oxygen therapy and ventilatory support or even surgical

interventions aimed at reducing lung volume to reduce hyperinflation due to the emphysematous tissue and thus improving the work of the remaining lung.

Nevertheless, there is an urgent need to develop therapies that reduce the progression of the disease, exacerbations, and comorbidities of COPD. However, currently no safe and effective treatments are available for the underlying chronic inflammation. It is therefore crucial to understand how the immune cells involved are linked to identify which of these cells and mediators are the most promising therapeutic targets.

2. Aim of the study

Due to its global spread and the high medical and financial burden without effective and safe available therapies, COPD represents a yet unmet challenge for modern society. An important and necessary step towards a better understanding of the disease is to increase the resolution during the analysis. The current knowledge is mainly based on information obtained at population level but not at the level of single cells. In recent years, the cellular components of the human lung in health and disease have been the target of an increasing number of studies, although so far, no study has focused on COPD. For example, scRNA-seq was used for the description of the parenchymal cell composition in the lung (Madissoon et al., 2020; Travaglini et al., 2019; Vieira Braga et al., 2019), the identification of novel cell types such as pulmonary ionocytes (Montoro et al., 2018; Plasschaert et al., 2018), the identification of ectopic and aberrant lung resident cell populations in idiopathic pulmonary fibrosis (Adams et al., 2019; Morse et al., 2019; Reyfman et al., 2019) or the investigation of the cellular contribution in lung cancer (Lambrechts et al., 2018; Lavin et al., 2017; Song et al., 2019; Zilionis et al., 2019). In addition, scRNA-seq was also used to investigate the COVID-19 disease (Liao et al., 2020; Ziegler et al., 2020). These studies in humans have been accompanied by scRNA-seq studies describing the cellular compositions in murine lung under homeostatic as well as stress conditions (Angelidis et al., 2019; Aran et al., 2019; McQuattie-Pimentel et al., 2019; Schyns et al., 2019; Strunz et al., 2019), and during development (Cohen et al., 2018; Guo et al., 2019). Together, these studies illustrate the enormous breadth of scRNA-seq technologies to describe the cellular composition of the lung and identify deviations from homeostasis in diseased organ tissues.

Encouraged by these seminal studies, we aimed to apply scRNA-seq on BALF samples obtained from both control donors and patient with an early-grade COPD (GOLD grade 2). At this early grade of COPD, disease progression is moderate and may therefore be a better stage of the disease for pharmacological intervention. We hypothesize that the high-resolution information obtained by scRNA-seq not only allows the characterization of the immune landscape in the alveolar space, but also contributes to a better understanding of the cellular and molecular factors involved in COPD pathogenesis.

To achieve this aim, we first needed to identify a clinically applicable scRNA-seq technology, develop a reliable cell-type annotation method and pursue a robust strategy for the identification of differentially expressed (DE) genes. Furthermore, the generation of a comprehensive understanding of the cells and mechanisms underlying COPD requires the construction of cell-to-cell interaction networks, the modelling of the metabolic landscape in AMs and the prediction of AM replenishment in COPD. The latter requires the generation of scRNA-seq data from blood of matching donors. Finally, the results obtained need to be validated experimentally, e.g. using flow cytometry.

3. Material and Methods

Table 1: Resource table (common laboratory materials (e.g. PBS) are not listed)

Reagent or resource	Source	Identifier
Antibodies		
PE-Cy7 anti-human CD45	Biolegend	Cat#304016; RRID: AB_314404
PE anti-human CD66b	Biolegend	Cat#305106; RRID: AB_2077857
APC-Cy7 anti-human CD3	Biolegend	Cat#300470; RRID: AB_2629689
APC-Cy7 anti-human CD19	Biolegend	Cat#302258; RRID: AB_2629691
APC-Cy7 anti-human CD56	Biolegend	Cat#362554; RRID: AB_2572105
APC anti-human HLA-A,B,C	Biolegend	Cat#311414; RRID: AB_493135
APC anti-human HLA-DR,DP,DQ	Biolegend	Cat#361704; RRID: AB_2563169
AF647 anti-human HLA-DR	Biolegend	Cat#307622; RRID: AB_493177
APC anti-human CD74	Biolegend	Cat#326812; RRID: AB_2564389
BV711 anit-human CD45	Biolegend	Cat#304050; RRID: AB_2563466
APC anti-human CD19	Biolegend	Cat#302212; RRID: AB_314242
APC-Cy7 anti-human CD14	Biolegend	Cat#325620; RRID: AB_830693
BV605 anti-human CD16	Biolegend	Cat#302039; RRID: AB_2561354
PerCP/Cy5.5 anti-human HLA-DR	Biolegend	Cat#307630; RRID: AB_893567
PE/Dazzle anti-human CD203c	Biolegend	Cat#324623; RRID: AB_2566234
PE/Cy7 anti-human Siglec-8	Biolegend	Cat#347112; RRID: AB_2629720
AF700 anti-human FcεRIα	Biolegend	Cat#334630; RRID: AB_2571902
FITC anti-human CD66b	Biolegend	Cat#305104; RRID: AB_314496
FITC anti-human CD11c	Biolegend	Cat#301604; RRID: AB_314174
PE-Cy7 anti-human CD3	Biolegend	Cat#300420; RRID: AB_439781
PerCP/Cy5.5 anti-human CD4	Biolegend	Cat#300530; RRID: AB_893322
PerCP/Cy5.5 anti-human CD4	Biolegend	Cat#344730; RRID: AB_2564510
APC anti-human CRTH2	Biolegend	Cat#350110; RRID: AB_11203707
AF647 anti-human CD3	BD	Cat#557706; RRID: AB_396815
AF647 anti-human CD56	BD	Cat#557711; RRID: AB_396820

BV421 anti-human CD117	BD	Cat#562434; RRID: AB_11154222
FITC anti-human CD33	BD	Cat#555626; RRID: AB_395992
FITC anti-human CD123	BD	Cat#558663; RRID: AB_1645485
BV510 anti-human CD56	BD	Cat#563041; RRID: AB_2732786
PE anti-human CD127	BD	Cat#557938; RRID: AB_2296056
FITC anti-human CD14	BD	Cat#130-110-520; RRID: AB_2655053
Bacterial strain		
Escherichia coli BL21 (DE3)	Merck	genotype: fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5; Cat#69450
Biological samples		
BALF	University Hospital Bonn	N/A
venipuncture blood	University Hospital Bonn	N/A
Chemicals and recombinant proteins		
CFSE	eBioscience	Cat#65-0850
Chitosan	Sigma	Cat#C3646
Chitin Resin	NEB	Cat#S6651L
Recombinant human CCL3	R&D	Cat#279-LD-010
Rotenone	Sigma	Cat#R8875-1G
Oligomycin	Sigma	Cat#75351-5MG
FCCP	Sigma	Cat#C2920-10MG
Antimycin A	Sigma	Cat#A8674-25MG
Crystal violet solution	Sigma	Cat#V5265
Cholesterol-d6	Avanti	Cat#700172
CL(56:0)	Avanti	Cat#710332
LPC(17:1)	Avanti	Cat#855677
Ceramide(17:0)	Avanti	Cat#860517
SM(17:0)	Avanti	Cat#860585
GlcCer(12:0)	Avanti	Cat#860543
GM3(18:0-d3)	Matreya	Cat#2052
PG(28:0)	Sigma	Cat#P6412
LPA(17:0)	Avanti	Cat#857127
LPE(17:1)	Avanti	Cat#856707
PA(31:1)	in-house synthesis (Christoph Thiele)	N/A
PC(31:1)	in-house synthesis (Christoph Thiele)	N/A
PE(31:1)	in-house synthesis (Christoph Thiele)	N/A
PI(34:0)	in-house synthesis (Christoph Thiele)	N/A

PS(31:1)	in-house synthesis (Christoph Thiele)	N/A
TG(47:1)	in-house synthesis (Christoph Thiele)	N/A
CE(17:1)	in-house synthesis (Christoph Thiele)	N/A
MAG(17:1)	in-house synthesis (Christoph Thiele)	N/A
DG(31:1)	in-house synthesis (Christoph Thiele)	N/A
Acyl-Carn(15:0)	in-house synthesis (Christoph Thiele)	N/A
Critical commercial assays and kits		
CD45 Microbeads, human	Miltenyi Biotec	Cat#130-045-801
StraightFrom™ Whole Blood CD66b MicroBeads, human	Miltenyi Biotec	Cat#130-104-913
Whole Blood Column Kit	Miltenyi Biotec	Cat#130-093-545
MACS Separation Columns LS	Miltenyi Biotec	Cat#130-042-401
MidiMACS Starting Kit (LS)	Miltenyi Biotec	Cat#130-042-301
APC Annexin V	Biolegend	Cat#640920
FcR Blocking Reagent, human	Miltenyi Biotec	Cat#130-059-901
LIVE/DEAD Fixable Yellow Dead Cell Stain Kit	Invitrogen	Cat#L34967
LIVE/DEAD Fixable Near-IR Dead Cell Stain Kit	Invitrogen	Cat#L34975
MinElute PCR Purification Kit	Qiagen	Cat#28004
Chromium™ Single Cell 3' Library and Gel Bead Kit v2	10x Genomics	Cat#120267
Chromium™ Single Cell A Chip Kit, 16 rxns	10x Genomics	Cat#1000009
Chromium™ i7 Multiplex Kit, 96 rxns	10x Genomics	Cat#120262
Quick Start™ Bradford 1x Dye Reagent	Bio-Rad	Cat#500-0205
Seahorse XF RPMI Medium pH 7.4	Agilent	Cat#103576-100
Seahorse XF96 Cell Culture Microplates	Agilent	Cat#101085-004
Seahorse XFe96 FluxPak	Agilent	Cat#102416-100
TapeStation HS D5000 Reagents (Sample Buffer & Ladder)	Agilent	Cat#5067-5593
High Sensitivity D5000 ScreenTape	Agilent	Cat#5067-5592
NEBNext High-Fidelity 2x PCR Master Mix	NEB	Cat#M0541L
Nextera XT DNA Library Preparation Kit (96 samples)	Illumina	Cat#FC-131-1096
2x Kapa Hifi HotStart Readymix	Kapa Biosystems	Cat#KK-2602
NextSeq® 500/550 High Output Kit v2.1 (75 cycles)	Illumina	Cat#20030410
NextSeq PhiX Control Kit	Illumina	Cat#FC-110-3002
NxGen™ RNase Inhibitor	Lucigen	Cat#F83923-1
Exonuclease I	NEB	Cat#M0293S
Maxima H Minus Reverse Transcriptase	Thermo Fisher	Cat#EPO0753
dNTP	NEB	Cat#N04465
Oligonucleotides		
Tn5ME-B Oligo: 5`-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3`	IDT	custom
Tn5MErev Oligo: 5`-[phos]CTGTCTCTTATACACATCT-3`	IDT	custom
TSO primer	IDT	custom
P5-SMART-PCR primer	IDT	custom
barcoded primer	IDT	custom
N70X Oligo	Illumina	custom

SMART PCR primer	Eurofins Scientific	Cat#74998995
Recombinant DNA		
plasmid: pTXB1	NEB	Cat#N6707S
Software and algorithms		
Agilent Seahorse Wave Software	Agilent	RRID: SCR_014526
biomaRt	Durinck et al., 2009	RRID: SCR_002987
flowCore	Ellis et al., 2019	RRID: SCR_002205
limma	Ritchie et al., 2015	RRID: SCR_010943
scran	Lun et al., 2016	RRID: SCR_016944
GSVA	Hänzelmann et al., 2013	https://bioconductor.org/packages/release/bioc/html/GSVA.html
AUCell	Aibar et al., 2017	https://bioconductor.org/packages/release/bioc/html/AUCell.html
Seurat	Butler et al., 2018	RRID: SCR_007322
umap	McInnes et al., 2018	https://cran.r-project.org/web/packages/umap/index.html
pheatmap	author: Raivo Kolde	RRID: SCR_016418
factoextra	authors: Alboukadel Kassambara and Fabian Mundt	RRID: SCR_016692
wordcloud	author: Ian Fellows	https://cran.r-project.org/web/packages/wordcloud/index.html
UpSetR	Conway et al., 2017	https://cran.r-project.org/web/packages/UpSetR/index.html
ggraph	author: Thomas Lin Pedersen	https://cran.r-project.org/web/packages/ggraph/index.html
ggplot2	Wickham, 2016	RRID: SCR_014601
Monocle	Trapnell et al., 2014	https://www.bioconductor.org/packages/release/bioc/html/monocle.html
g:Profiler	Raudvere et al., 2019	RRID: SCR_006809
R statistical programming	R Core Team	RRID: SCR_001905
RStudio	RStudio, Inc.	RRID: SCR_000432
CIBERSORT	Newman et al., 2015	RRID: SCR_016955
DoubletFinder	McGinnis et al., 2019	https://github.com/chris-mcginnis-ucsf/DoubletFinder
SingleR	Aran et al., 2019	https://github.com/dviraran/SingleR
dropSeqPipe	author: Patrick Rölli	https://github.com/Hoohm/dropSeqPipe
Rphenograph	Levine et al., 2015	RRID: SCR_016919
STAR	Dobin et al., 2013	RRID: SCR_015899
RSEM	Li and Dewey, 2011	RRID: SCR_013027

NIPY	NIPY community	RRID: SCR_002489
nichenetr	Browaeys et al., 2019	https://github.com/saeyslab/nichenetr
Scanpy	Wolf et al., 2018	RRID:SCR_018139
scIB	not published yet	https://github.com/theislab/scib
scVelo	Bergen et al., 2019	https://github.com/theislab/scvelo
Compass	Wagner et al., 2020	https://github.com/YosefLab/Compass
VISION	DeTomaso et al., 2019	https://github.com/YosefLab/VISION
scikit-learn	sckit-learn community	RRID: SCR_002577
LipidXplorer	Herzog et al., 2012	https://wiki.mpi-cbg.de/lipidx/LipidXplorer_Installation
CellPhoneDB	Efremova et al., 2020	RRID: SCR_017054
bcl2fastq2	Illumina	RRID: SCR_015058
Cytoscape	Shannon et al., 2003	RRID:SCR_003032
iRegulon	Janky et al., 2014	http://iregulon.aertslab.org/
ImageJ	Schneider et al., 2012	RRID: SCR_003070
FlowJo	Tree Star Inc.	RRID: SCR_008520
Other		
AMPure XP beads	Beckman Coulter	Cat#A63881
LifterSlip™	Electron Microscopy Science	Cat#72186-60
Polycarbonate (PCTE) membrane filters, 0.01 MICRON, 62MM X 22MM	Sterlitech	Cat#PCT00162X22100
mRNA Capture beads	Chemgenes	Cat#MACOSKO-2011-10
Dow SYLGARD™ 184 Silicone Encapsulant Clear 0.5kg kit	Dow	184 SIL ELAST KIT 0.5KG
Transwell polycarbonate membrane cell culture inserts	Corning	Cat#3422

Table 2: Clinical information of donors included in this study

identifier	diagnosis	phenotype	add. lung disease	COPD		sex	age	BMI	smok. history	FVC (%)	FEV1 (%)	FEV1/FVC (%)	RV
				grade	group								
Control 1	healthy					♂	27	N/A	current smoker	N/A	N/A	N/A	N/A
Control 2	healthy					♀	27	N/A	never-smoker	N/A	N/A	N/A	N/A
Control 3	chronic cough					♂	60	29	ex-smoker	107	120	88.04	87
Control 4	chronic cough					♀	68	27	never-smoker	99	98	85.34	101
Control 5	chronic cough					♀	56	23	never-smoker	81	92	99	121
Control 6	chronic cough					♀	49	23	never-smoker	107	111	82.04	139
Control 7	chronic cough					♂	58	22	never-smoker	119	132	82.93	130
Control 8	chronic cough					♀	55	21	never-smoker	101	88	76.79	109
Control 9	chronic cough					♂	60	29	never-smoker	91	99	81.07	101
Control 10	chronic cough					♂	53	30	never-smoker	107	111	82.03	109
Control 11	chronic cough					♀	60	30	never-smoker	98	90	82	126
Control 12	chronic cough					♀	58	23	never-smoker	106	102	84	129
Control 13	chronic cough					♂	58	33	ex-smoker	92	96	84	116
Control 14	chronic cough					♀	45	26	never-smoker	89	82	82.16	143
Control 15	chronic cough					♂	48	30	never-smoker	89	94	86.29	126
Control 16	chronic cough					♀	25	20	never-smoker	66	68	92.73	31
Control 17	chronic cough					♂	77	32	never-smoker	78	81	80.15	86
Control 18	chronic cough					♂	78	25	ex-smoker	89	91	75.65	140
Control 19	chronic cough					♀	66	24	N/A	92	86	71.98	100
Control 20	chronic cough					♂	78	26	ex-smoker	75	88	87.69	109
Control 21	chronic cough					♂	28	24	never-smoker	98	96	84.7	152
Control 22	chronic cough					♂	52	24	never-smoker	69	76	89.81	116
Control 23	chronic cough		lung cancer			♂	78	28	never-smoker	91	102	81.82	79
Control 24	chronic cough					♀	25	20	never-smoker	97	87	79.47	131
Control 25	chronic cough					♀	41	27	never-smoker	73	61	74.53	132
Control 26	chronic cough					♂	40	27	never-smoker	86	80	77.13	165
Patient 1	COPD	chronic bronchitis	embolism	2	B	♂	64	30	ex-smoker	53	57	68.15	156
Patient 2	CPFE	fibrosis + COPD				♂	80	26	ex-smoker	66	69	73.9	139
Patient 3	COPD	chronic bronchitis		2	B	♂	59	36	ex-smoker	81	68	67.35	89
Patient 4	COPD	emphysema		2	B	♂	68	34	ex-smoker	83	76	68.26	112
Patient 5	COPD	chronic bronchitis		2	D	♀	46	38	current smoker	68	53	69	117
Patient 6	COPD	chronic bronchitis	lung cancer	2	B	♀	66	20	current smoker	78	67	66.66	151
Patient 7	COPD	chronic bronchitis		2	D	♂	54	26	current smoker	79	70	68.7	169
Patient 8	COPD	chronic bronchitis		2	D	♀	53	28	ex-smoker	92	73	57.87	183
Patient 9	COPD	emphysema	lung cancer	2	B	♂	74	22	current smoker	63	59	64.34	162
Patient 10	COPD	emphysema		2	D	♂	55	34	current smoker	76	54	57.27	212
Patient 11	COPD	chronic bronchitis	lung cancer	3	B	♂	70	25	ex-smoker	58	46	60	73
Patient 12	COPD	chronic bronchitis		2	B	♂	64	36	ex-smoker	74	66	70	100
Patient 13	COPD	chronic bronchitis		2	B	♀	57	19	current smoker	92	73	57	206
Patient 14	COPD	chronic bronchitis		2	B	♂	58	24	current smoker	85	67	61.98	143
Patient 15	COPD	chronic bronchitis		2	B	♂	63	24	current smoker	65	78	59.1	225
Patient 16	COPD	chronic bronchitis		4	D	♀	65	23	current smoker	60	36	58.22	225
Patient 17	COPD	chronic bronchitis		3	D	♀	82	24	current smoker	58	39	50.69	258
Patient 18	COPD	chronic bronchitis		3	D	♀	60	21	ex-smoker	74	33	39.79	262
Patient 19	COPD	chronic bronchitis		3	D	♀	74	22	current smoker	47	43	40	N/A
Patient 20	COPD	chronic bronchitis		3	D	♂	68	31	ex-smoker	62	39	48.21	175
Patient 21	COPD	chronic bronchitis		2	D	♂	79	29	current smoker	58	51	67.38	145
Patient 22	COPD	chronic bronchitis		2	C	♂	70	29	ex-smoker	66	51	58.46	153
Patient 23	COPD	chronic bronchitis		3	D	♂	56	34	ex-smoker	49	38	62.73	219
Patient 24	COPD	chronic bronchitis		2	C	♂	60	25	ex-smoker	107	79	58.71	162
Patient 25	COPD	emphysema		2	D	♀	66	27	ex-smoker	81	54	55.64	189

Table 3: Medications from donors included in this study

identifier	corticosteroid		LABA	LAMA	PDE4 inhibitor	anti-hypertension	diuretics	anti-coagulant	anti-hyperlipidemia	anti-diabetic	anti-hyperuricemia	proton pump inhibitor	thyroid hormone	anti-depressant	antibiotics	analgesics
	oral	inhaled														
Control 1																
Control 2																
Control 3																
Control 4	+					+						+				+
Control 5																
Control 6																
Control 7																
Control 8														+		
Control 9																
Control 10																
Control 11		+				+		+	+							
Control 12		+														
Control 13		+	+	+		+	+	+	+			+				+
Control 14																
Control 15											+	+				
Control 16														+		
Control 17	+	+	+	+			+	+	+	+	+	+				+
Control 18		+	+	+		+	+				+		+			+
Control 19						+				+		+				+
Control 20		+							+							+
Control 21													+			
Control 22			+			+										
Control 23	+	+	+			+	+		+			+				+
Control 24		+	+				+					+				
Control 25															+	
Control 26				+											+	
Patient 1			+	+			+	+	+	+	+	+		+		+
Patient 2					+	+	+	+		+			+			
Patient 3					+	+		+	+							
Patient 4		+	+	+	+	+										
Patient 5	+	+	+	+	+	+		+	+	+	+	+				
Patient 6					+	+	+				+	+				
Patient 7		+	+	+				+	+							
Patient 8			+	+	+	+					+	+			+	
Patient 9			+	+	+	+		+			+	+				
Patient 10	+	+	+	+	+	+									+	
Patient 11			+	+		+	+	+		+	+	+	+		+	+
Patient 12	+		+	+		+			+	+		+				+
Patient 13		+	+	+												
Patient 14			+	+		+			+			+				
Patient 15			+	+												
Patient 16	+		+	+		+	+							+		+
Patient 17		+	+	+												
Patient 18			+	+	+		+	+						+		
Patient 19	+	+	+	+		+		+				+		+		+
Patient 20		+	+	+		+										
Patient 21			+	+		+	+					+				
Patient 22	+	+	+	+	+	+			+				+			+
Patient 23		+	+		+	+										+
Patient 24		+	+	+		+										
Patient 25		+	+	+												

Table 4: Overview of donors used for scRNA-seq and validation experiments (donors used solely for MCFC are not listed)

identifier	10x		Seq-Well		lipidomics	phenotyping	Olink	Seahorse	migration assay
	BALF	blood	BALF	blood					
Control 1		+		+					
Control 2		+		+					
Control 3	+	+	+	+					
Control 4	+		+						
Control 5			+						+
Control 6			+	+			+		
Control 7			+	+	+		+		
Control 8			+	+			+		
Control 9			+	+	+				
Control 10			+				+		+
Control 11									+
Control 12					+		+		+
Control 13					+		+		
Control 14							+		
Control 15							+		
Control 16							+		
Control 17						+			
Control 18						+			
Control 19						+			
Control 20						+		+	
Control 21						+			
Control 22						+			
Control 23						+			
Control 24						+			
Control 25						+			
Control 26								+	
Patient 1	+		+						
Patient 2			+				+		
Patient 3			+	+	+		+		
Patient 4			+	+			+		
Patient 5			+		+		+		
Patient 6			+	+	+		+		
Patient 7			+	+	+		+		
Patient 8			+				+		+
Patient 9			+	+			+		
Patient 10			+	+			+		+
Patient 11					+				
Patient 12					+				+
Patient 13							+		
Patient 14							+		
Patient 15							+		+
Patient 16						+			
Patient 17						+		+	
Patient 18						+			
Patient 19						+			
Patient 20						+			
Patient 21						+			
Patient 22						+			
Patient 23						+			
Patient 24								+	
Patient 25								+	

3.1. Human specimens

Human studies were approved by the ethics committees of the University of Bonn and University hospital Bonn (local ethics vote 076/16). All patients provided written informed consent according to the Declaration of Helsinki before specimens were collected. Each individual included in this study was diagnosed and the disease stage was stratified according to the recommendations of the global initiative for chronic obstructive lung disease (GOLD) (COPD recommendations, 2020), with a ratio of post-bronchodilator (salbutamol 400 µg) forced expiratory volume in 1 s (FEV1) to forced vital capacity (FVC) of less than 0.7, and moderate airflow limitation (50% <= FEV1 < 80%). For scRNA-seq, the eligible patients were aged 40 years or older and were either current or ex-smokers. Since COPD has recently been suggested to be a clinical syndrome rather than a single disease (Agustí and Hogg, 2019), we anticipated that despite the focus on GOLD 2 patients, the current study should include a spectrum of COPD patients (**Table 2-4**). For example, the generated dataset comprised COPD GOLD 2 patients with different emphysema proportions, exacerbation histories and even a patient suffering from combined pulmonary fibrosis and emphysema (CPFE). The latter patient was admitted based on an external diagnosis of COPD that was later diagnosed as CPFE. This disease type was first described by Cottin et al. (Cottin et al., 2005) and is defined radiologically by the presence of classical features of emphysema in the upper lobes and pulmonary fibrosis in the lower lobes and subnormal lung volumes and severe reduction of CO transfer. Irrespective of the expected heterogeneity within the COPD GOLD 2 patient cohort, stringent exclusion criteria for the current study were a primary diagnosis of asthma with a physician-judged need for oral corticosteroid therapy, clinically significant cardiovascular disorders or laboratory abnormalities and unstable concurrent disease (e.g. exacerbation of disease) that could have affected safety (as judged by the investigator). Individuals suffering from chronic cough without any signs of severe lung pathophysiology or subnormal lung functions served as control donors.

3.2. Isolation of cells from BALF

Human BALF was obtained from patients with or without COPD via bronchoscopy (at the University hospital Bonn). BALF was performed according to the official American

Thoracic Society guideline for interstitial lung disease patients to ensure highest quality of biospecimen material (Meyer et al., 2012). According to these guidelines, we excluded more than half of the clinical samples from further analyses because either the volume of saline solution recovered compared to the amount previously injected into the lungs during bronchoscopy was too low, or blood contamination or increased upper respiratory secretion was present. Each of these factors has an influence on the differential cell count of BALF samples and would have therefore had a negative effect on the analysis results. BALF samples fulfilling the quality criteria were once washed with PBS supplemented with 1 mM EDTA followed by washing with PBS supplemented with 2% fetal calf serum (FCS) and 1 mM EDTA. Throughout the isolation process, the samples were kept at 4°C and centrifugation steps performed at 300 g for 10 min. To exclude any macroscopic non-cellular particles and non-immune cells from further analyses, immune cells were enriched with MACS columns by using CD45 microbeads according to manufacturer's instructions.

3.3. Isolation of PBMC and blood granulocytes

For the assessment of relationship analysis of the myeloid cell compartment in BALF with cells from the systemic circulation, we obtained venipuncture blood on the day of bronchoscopy. PBMC were obtained by Pancoll density centrifugation (at 20°C and 700 g for 25min with centrifugation break was turned off) of the peripheral blood. After harvesting PBMC from the interphase, all further steps were conducted at 4°C. Granulocytes were recovered from the granulocyte/erythrocyte fraction using cold ACK (ammonium chloride potassium) lysing buffer (1.5M NH_4Cl, 0.1M $KHCO_3$ and 1mM EDTA in H_2O with pH 7.4 at 8°C) to lyse erythrocytes, followed by a washing step with PBS supplemented with 2% FCS and 1 mM EDTA. All centrifugation steps required for granulocyte isolation were performed with max. 300 g for 10 min. To assess the granulocyte fraction in further analyses (particularly in scRNA-seq experiments), it was mixed with the PBMC fraction in the ratio PBMC:granulocytes = 2:1. Finally, the PBMC/ granulocyte mix was stained with CD45 microbeads for 15 min in order to use a magnetic field in the cell loading of Seq-Well arrays (see below). This artificial ratio allowed to assess the granulocytes in addition to the PBMCs without

sequencing the majority of blood immune cells being granulocytes allowing sufficient granularity in the PBMC fraction.

3.4 Flow cytometric data generation

Cells were resuspended in PBS supplemented with 2% FCS and 1 mM EDTA for surface marker staining (**Table 5**). To distinguish live from dead cells, the cells were incubated with LIVE/DEAD Fixable Dead Cell Stain Kit (1:1000) at room temperature for 15 min protected from light. After washing, human FcR blocking reagent was included to reduce unspecific staining (incubation on ice for 15 min). Next, surface antibodies were added and after 30 min incubation at 4ºC in the dark, cells were washed and analyzed either on BD FACSAria III (Becton Dickinson) for acquisition and sorting or on BD FACSCanto II (Becton Dickinson) for acquisition only (**Table 5**). Fluorescence-minus-one (FMO) controls were prepared for non-lineage markers.

3.5. Flow cytometric data analysis

Preliminary data analysis was performed using FlowJo software (version 10). The package ‘flowCore’ (version 1.46.2, Ellis et al., 2019) was used to import the compensated data into R. For dimensionality reduction with UMAP implementation in R (version 0.2.1.0, McInnes et al., 2018), fluorescence parameters were transformed with logicleTransform (Becht et al., 2019; Parks et al., 2006). Subsequent clustering of the dataset was performed with the PhenoGraph algorithm implemented in the ‘Rphenograph’ package (version 0.99.1, Levine et al., 2015) by setting the number of nearest neighbors to 25. Based on marker detection, the major cell types in the BALF were defined as macrophages (Lin^- (including CD3, CD19 and CD56) $CD66b^-$ $HLA\text{-}DR^+$ $autofluorescence^+$), monocytes/DCs (Lin^- $CD66b^-$ $autofluorescence^-$ $HLA\text{-}DR^+$ and either
$CD14^+$, $CD16^+$ or $CD14^+$ $CD16^+$), granulocytes (Lin^- $HLA\text{-}DR^{low}$ $autofluorescence^{low}$ $CD66b^+$ and either $CD16^-$ $Siglec\text{-}8^+$, $CD16^+$ $Siglec\text{-}8^+$ and $CD16^+$ $Siglec\text{-}8^-$) and T cells/NK cells including a small fraction of B cells ($autofluorescence^-$ $CD14^-$ $CD66b^-$ Lin^+ and further resolved using the lymphoid panel (**Table 5**)).

Table 5: Overview of used FACS panels

usage	antigen/ detection target	conjugate	system
myeloid cell phenotyping	LIVE/DEAD Fixable Yellow Dead Cell Stain Kit	-	BD FACSAria III
	CD45	BV711	
	CD3	AF647	
	CD19	APC	
	CD56	AF647	
	CD14	APC-Cy7	
	CD16	BV605	
	HLA-DR	PerCP/Cy5.5	
	CD66b	PE	
	CD203c	PE/Dazzle	
	Siglec-8	PE/Cy7	
	CD117	BV421	
	FcεRIα	AF700	
lymphoid cell phenotyping	LIVE/DEAD Fixable Yellow Dead Cell Stain Kit	-	
	CD45	BV711	
	CD14	FITC	
	CD33	FITC	
	CD66b	FITC	
	CD11c	FITC	
	CD123	FITC	
	CD3	PE-Cy7	
	CD4	PerCP/Cy5.5	
	CD8	BV650	
	CD56	BV510	
	CD127	PE	
	CRTH2	APC	
	CD117	BV421	
	CD16	BV605	
Differential marker detection	LIVE/DEAD Fixable Near-IR Dead Cell Stain Kit	-	BD FACSCanto II
	CD45	PE-Cy7	
	CD66b	PE	
	CD3	APC-Cy7	
	CD19	APC-Cy7	
	CD56	APC-Cy7	
	HLA-A,B,C	APC	
	HLA-DR,DP,DQ	APC	
	HLA-DR	AF647	
	CD74	APC	
	Annexin V	APC	

In blood, the major cell types were defined as monocytes/DCs ($CD3^-$ $CD19^-$ $CD56^-$ $CD66b^-$ $HLA\text{-}DR^+$ and either $CD14^+$, $CD16^+$ or $CD14^+$ $CD16^+$), T cells/NK cells ($CD14^-$ $CD33^-$ $CD66b^-$ $CD11c^-$ $CD123^-$ $CD19^-$ and either $CD3^+$ $CD4^+$, $CD3^+$ $CD8^+$ or $CD56^+$),

granulocytes ($CD3^-$ $CD19^-$ $CD56^-$ HLA-DR^{low} $CD66b^+$ $CD16^+$ and either $CD16^-$ Siglec-8^+, $CD16^+$ Siglec-8^+ and $CD16^+$ Siglec-8^-) and B cells ($CD14^-$ $CD33^-$ $CD66b^-$ $CD11c^-$ $CD123^-$ $CD3^-$ $CD56^-$ $CD19^+$). According to these marker combinations, the identified clusters were annotated. To unify and simplify the analysis across multiple datasets, an annotated dataset was defined as the reference and the other flow cytometry datasets were projected onto its UMAP coordinates using the 'umap' object of the reference dataset and the logicle transformed flow cytometry data of the second dataset as input for the predict function in R. In addition, the same function was also used to predict the clusters of the remaining datasets with respect to the reference dataset. This step, together with the visualization of detected markers, made it possible to assess both the accuracy of the projection method and the cell-type annotation of the projected datasets.

We performed differential marker intensity measurements across individuals based on the Cohen's d definition of effect size as follows:

$$\text{effect size} = \frac{Mean_{comple\ stain} - Mean_{FMO}}{\sqrt{\frac{(SD_{complete\ stain}{}^2 + SD_{FMO}{}^2)}{2}}}$$

This procedure was followed since we observed strong variability in autofluorescence intensities of macrophages among donors, despite strictest standard operating procedure (SOP) compliance and the use of SOPs for application settings (BD Biosciences, 2012) during flow cytometry to minimize potential biases that can occur during sample-to-sample flow cytometry comparisons.

3.6. MitoStress assay on Seahorse

For the analysis of the metabolic state of donor-derived alveolar macrophages (AMs), freshly obtained BALF was centrifuged for 10 min at 300 g. Cell pellet was then washed carefully in PBS (supplemented with 0.02% EDTA) and finally resuspended in MACS buffer. Cell suspension was then stained for 15 min with CD66b microbeads and depleted from granulocytes according to manufacturer instructions. Granulocyte-depleted cell suspension was counted and seeded in Seahorse XF RPMI medium (supplemented with 2 mM L-glutamine, 1 mM sodium pyruvate, 10 mM glucose, adjusted to pH 7.4 prior to the assay) at a concentration of 200,000 cell per well; for each sample 2 to 4 technical replicates were performed. Cells were then incubated for 30 min in a 37°C incubator, washed two times with pre-warmed Seahorse XF RPMI

medium to remove all non-adherent cells and loaded onto the Seahorse XFe96 Analyzer (Agilent). After 3 cycles of baseline measurement, whereby one cycle is defined as 3 min of initial mixing and 3 min measurement, the cells were subsequently injected with Oligomycin (1:1000), FCCP (1:500) and finally a combination of Antimycin A and Rotenone (both 1:2000). Following each injection, oxygen consumption rate (OCR) was measured for 3 cycles.

After the assay, the relative cellular number was determined via crystal violet staining. Shortly, cells were fixed with 4% PFA for 5 min at room temperature and stained for 30 min with crystal violet (0.05% in H_2O). After two washes with H_2O the staining was air dried and the formed crystals were dissolved in 200 µL of methanol. Absorbance at 590 nm was measured and used to normalize the Seahorse assay within the Wave software (Agilent). The normalized data were finally exported, further analyzed and visualized in R, with values adjusted to the measured baseline (baseline-corrected).

3.7. Migration Assay

Remark: The migration assay was led and conducted by Wataru Fujii.

Migration was analyzed in 24 well transwell plate containing a 8 µm polycarbonate membrane. FACS sorted AMs were suspended in 300 µL starvation medium (RPMI 1640 medium supplemented with 0.5% FCS and 1% penicillin/streptomycin) and 50,000 AMs were seeded in each upper well, while the lower chamber was filled with 700 µL starvation medium only. After an incubation of 1 h in a 37°C incubator, the medium in the upper chamber was exchanged with 300 µL fresh starvation medium and the medium in the lower chamber with 700 µL starvation medium supplemented with 100 ng/mL recombinant human CCL3. The seeded AMs were incubated at 37°C overnight. Next, cells on the upper filter surface were removed with a cotton swab. Transmigrated cells on lower filter surface were incubated with 2 µM CFSE in 700 µL PBS for 10 min in a 37°C incubator. The transwell inserts were then transferred into wells containing 700 µL RPMI 1640 medium supplemented with 10% FCS and 1% penicillin/streptomycin and incubated for 10 min in a 37°C incubator. Finally, transwell inserts were washed with PBS and imaging of cells was performed using an inverted fluorescent microscope (Nikon) with a 10-fold objective and GFP filter. The number of migrated cells was quantified using ImageJ (version 2, Schneider et al., 2012).

3.8. Measurement of proteins in BALF

After isolation of cells (see above), the supernatant of BALF samples of both COPD patients and controls were collected and frozen at −80 °C before proteomics measurement. Protein levels from cell-free BALF samples were determined using the INFLAMMATION panel from Olink Proteomics, a commercial multiplex immunoassay for high-throughput detection of 92 inflammation-related protein biomarkers. The obtained normalized results were further analyzed in R, whereby proteins were kept for visualization that showed a statistically significant difference (Wilcoxon rank sum test-based p-value < 0.1) between COPD and control samples.

3.9. Lipidomics of macrophages in BALF

Remark: The spectrometric measurement of lipid species was conducted by Christoph Thiele.

AMs were sorted, washed with PBS and with 150 mM ammonium acetate in a glass tube, pelleted (300 g with slow brake), and frozen at -80°C until analysis. To the pellet, 500 µL of extraction mix ($CHCl_3$/MeOH 1/5 containing internal standards: 210 pmol PE(31:1), 396 pmol PC(31:1), 98 pmol PS(31:1), 84 pmol PI(34:0), 56 pmol PA(31:1), 51 pmol PG (28:0), 28 pmol CL(56:0), 39 pmol LPA (17:0), 35 pmol LPC(17:1), 38 pmol LPE (17:1), 32 pmol Cer(17:0), 99 pmol SM(17:0),55 pmol GlcCer(12:0), 14 pmol GM3 (18:0-D3), 359 pmol TG(47:1), 111 pmol CE(17:1), 64 pmol DG(31:1), 103 pmol MG(17:1), 724 pmol Chol(d6), 45 pmol Car(15:0)) were added and each sample sonicated for 2 min followed by centrifugation at 20,000 g for 2 min. The supernatant was collected into a new tube and 200 µL chloroform and 800 µL 1% AcOH in H_2O were added. The sample was then briefly shaken and spun for 2 min at 20,000 g for 2 min. 200 µL chloroform and 800 µL 1% AcOH in H_2O were added to the supernatant, briefly shaken and spun for 2 min at 20,000 g. The lower phase was transferred into a new tube and evaporated in a speed vac (45°C, 10 min). Spray buffer (500 µL of 8/5/1 2-propanol/MeOH/H_2O, 10 mM ammonium acetate) was added, sonicated for 5 min and infused at 10 µL/min into a Thermo Q Exactive Plus spectrometer (Thermo Fisher

Scientific) equipped with the HESI II ion source for shotgun lipidomics. MS1 spectra (res. 280000) were recorded in 100 m/z windows from 200 – 1200 m/z (pos.) and 200 – 1700 m/z (neg.) followed by recording MS/MS spectra (res. 70000) by data independent acquisition in 1 m/z windows from 200 – 1200 (pos.) and 200 – 1700 (neg.) m/z.

Raw files were converted to mzml files and imported into and analyzed by LipidXplorer (version 1.2.8, Herzog et al., 2012) software using custom mfql files to identify sample lipids and internal standards. For further data processing, absolute amounts were calculated using the internal standard intensities followed by normalization of the identified lipids on total lipid content. Lipid class sums were calculated for each donor and log_2-transformed. Differential lipid classes were calculated between COPD GOLD 2 vs control samples using the 'limma' package (version 3.42.2, Ritchie et al., 2015) under consideration of 'date of sampling'.

3.10. Nanodroplet-based scRNA-seq

For comparison of nanodroplet-based scRNA-seq with array-based scRNA-seq (Seq-Well technology, see below), cell preparations derived from three blood and three BALF donors were split in half to be further processed with the two different scRNA-seq technologies by two teams simultaneously. For each donor, 10,000 BALF or blood-derived cells were loaded onto the Chromium™ Controller instrument (10x Genomics) using the Chromium™ Single Cell A Chip Kit together with the Chromium™ Gel Bead Kit v2 following the manufacturer's recommendations. Libraries were prepared using Chromium™ Single Cell 3' Library Kit v2 according to manufacturer's recommendations and sequenced paired-end as followed: Read 1 26 cycles, i7 index 8 cycles and Read 2 56 cycles on a NextSeq500 instrument (Illumina) using High Output v2.1 chemistry. Single-cell data was demultiplexed and converted into fastq format using bcl2fastq2 (v2.20).

3.11. Preparation of Seq-Well arrays

Seq-Well arrays were prepared as described by Gierahn *et al*. (Gierahn et al., 2017). Briefly, Sylgard base and crosslinker were mixed at 10:1 ratio for 10 min, placed under vacuum pressure for 15 min to remove air bubbles and were next poured for a 2 h incubation at 70°C into a wafer with a mounted 86,000 well pattern-holding microscope slide. The arrays were then removed from the molds, excess silicone was cut off with a blade and were prepared for the functionalization process. This protocol adds chemical moieties to the surface of the arrays which facilitate the sealing of a semi-permeable polycarbonate membrane and the interchange of lysis and RNA hybridization buffers. Arrays were rinsed with EtOH, plasma treated for 10 min and successively submerged in APTES (0.05% APTES in 95% EtOH), acetone and PDITC buffers (0.2% PDITC, 10% pyridine, 90% DMF). Upon further washes with acetone, the arrays were spun and dried at 70°C for 2 h. Among the most critical steps in the protocol was the incubation of the arrays with 0.2% chitosan solution (pH=6.3) at 37°C for 1.5 h, after which an overnight incubation in PGA buffer (20 µg/mL polyglutamic acid, 2 M NaCl, 100 mM sodium carbonate (pH=10)) at room temperature under vacuum pressure followed. Finally, the arrays were removed from the vacuum and were rotated for 3 h at room temperature and subsequently moved to 4°C for at least 24 h before use.

3.12. Preparation of Seq-Well libraries and sequencing

Remark: The synthesis of in-house Tn5 was conducted by Ines Kaltheuner under the supervision of Matthias Geyer.

Apart from a few adjustments, which are listed in the following, will be listed, the Seq-Well libraries were generated as recently described by Gierahn *et al.* (Gierahn et al., 2017). After loading of the functionalized arrays with mRNA capture beads, 20,000 $CD45^+$ cells were applied that were previously coated with $CD45^+$ magnetic beads (see above) and suspended in RPMI 1640 medium supplemented with 10% FCS. During the incubation time of 10 min, the loaded arrays were placed on a strong magnetic plate to support the settling of the cells via a magnetic field. After repetitive washing with PBS and soaking with RPMI 1640 medium, the arrays were sealed using polycarbonate membranes that were 7 min treated with air plasma under mild vacuum

(Diener electronic). Following a 30 min incubation time in a 37°C cell culture incubator, the arrays were incubated in lysis buffer (5M guanidine thiocyanate, 1mM EDTA, 0.5% Sarkosyl and 1% β-mercaptoethanol in H_2O) for 20 min and then placed in hybridization buffer (2M NaCl, 3mM $MgCl_2$ and 0.5% Tween-20 in PBS) for 40 min. Next, the mRNA capture beads were washed from the arrays and collected using washing buffer (2M NaCl, 3mM $MgCl_2$ and 20mM Tris-HCl pH 8.0 in H_2O). The reverse transcription was performed on the bead pellet using a Maxima Reverse Transcriptase reaction (Maxima RT buffer, 4% Ficoll PM-400, 1mM dNTPs, 1U/µL RNase inhibitor, 2.5 µM template switch oligonucleotide (TSO) primer and 10U/µL Maxima Reverse Transcriptase in H_2O) for 30 min at room temperature followed by 90 min incubation at 52°C with end-over-end rotation. The reaction was stopped by washing the beads with TE buffer (10mM Tris-HCl pH 8.0 and 1mM EDTA in H_2O) supplemented with 0.1% Tween-20 (TE-TW) and TE buffer supplemented with 0.5% SDS (TE-SDS). After a washing step in 10mM TrisHCl pH 8.0, excess primers were digested in an exonuclease reaction (ExoI buffer and 1U/µL ExoI in H_2O) for 50 min at 37°C with end-over-end rotation and washed in TE-TW and TE-SDS. Beads were resuspended in 500 µL H_2O and counted with a Fuchs-Rosenthal cytometer in bead counting solution (10% PEG, 2.5 M NaCl). Pools of 5,000 beads (10 µL) were then added to 40 µL PCR reactions (2X KAPA HiFi Hotstart Readymix and 25 µM SMART PCR primer in H_2O) for the amplification of reverse transcribed cDNA libraries (95°C for 3 min, 4 cycles of 98°C for 20 s, 65°C for 45 s, 72°C for 3 min, 12 cycles of 98°C for 20 s, 67°C for 20 s, 72°C for 3 min and final extension of 72°C for 5 min). After PCR, 16,000-20,000 beads were combined (thereafter referred to as ‘pools’) and further processed. The pools were cleaned with 0.6x volumetric ratio AMPure XP beads (5 min incubation with beads, followed by 3 min on the magnet, two washes with 80% EtOH, 5 min dry-out, elution with 13 µL H_2O for 3 min, followed by 2 min on the magnet for collection of the eluent) and the library integrity was assessed using a High Sensitivity D5000 assay for the Tapestation 4200 (Agilent).

To reduce library costs, we produced homemade Tn5 transposase according to (Picelli et al., 2014b). Briefly, the Tn5 coding sequence (tnpA gene from *Escherichia coli*, Uniprot accession number: Q46731, residues 1-476) was purchased as a synthesized gene containing the mutations E54K and L372P for hyperactivation of the enzyme. Overhangs with the restriction sites *Xba*I and *Spe*I were used for cloning into pTXB1 vector, generating a Tn5-Intein-CBD fusion construct. The Tn5 coding sequence was

validated by Sanger sequencing. Next, the pTXB1-Tn5-Mxe-CBD plasmid was transformed into the *E.coli* strain BL21. Cells were grown in LB media supplemented with ampicillin at 37°C to an OD_{600} ~0.8. The temperature was then lowered to 10°C and protein expression was induced by addition of 0.25 mM IPTG. After incubation at 23°C for 4 h cells were harvested by centrifugation at 15,000 rpm on a JA 25.50 rotor (Beckman) for 20 min at 10°C. The cell pellet was resuspended in running buffer (20 mM Hepes-KOH, 0.8 M NaCl, 1 mM EDTA, 10% glycerol, 0.2% Triton-X 100) supplemented with 1 mM PMSF and disrupted by sonication. After centrifugation of cell debris at 15,000 rpm on a JA 25.50 rotor (Beckman) for 30 min at 10°C, residual nucleic acid contaminations from *E.coli* were precipitated by dropwise addition of polyethyleneimine pH 7.5 to a final concentration of 0.3%. The lysate was cleared by centrifugation at 12,000 rpm on a JA 25.50 rotor (Beckman) for 10 min at 4°C. Chitin resin (10 mL) was equilibrated with running buffer and then incubated with the prepared lysate for 1 h at 4°C. Beads were washed with 10 column volumes of running buffer. For elution by self-cleavage via the intein-tag, the Tn5-loaded resin was incubated overnight at 4°C in 3 mL elution buffer (20 mM Hepes-KOH, 0.8 M NaCl, 1 mM EDTA, 10% glycerol, 0.2% Triton-X 100, 100 mM DTT), followed by dialysis at 4°C overnight in dialysis buffer (100 mM Hepes-KOH, 0.2 M NaCl, 0.2 mM EDTA, 2 mM DTT, 0.2% Triton-X 100, 20% glycerol). The protein concentration was determined using Bradford Assay. Glycerol was added to a final concentration of 50% to the protein sample.

To load Tn5 with linker oligonucleotides (Tn5ME-B/Tn5MErev (Tn5ME-B: 5`-TCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3`; Tn5MErev: 5`-[phos]CTGTCTCTTATACACATCT-3`;)), single-stranded oligonucleotides were mixed in a 1:1 ratio. For pre-annealing, 2 µL of the oligonucleotide solution was mixed with 8 µL of H_2O and incubated in a thermocycler (95°C for 3 min, 70°C for 3 min and 45 cycles of temperature reduction (-1°C per 30 s)). The annealed oligonucleotides (0.25 vol.) were added to 0.1 vol. Tn5 solution and supplemented with •0.4 vol. glycerol (100%), 0.12 vol. dialysis buffer and 0.13 vol. H_2O. After incubation for 60 min at room temperature, the protein was stored at -20°C.

The cDNA libraries (1 ng) were tagmented with the prepared single-loaded Tn5 transposase in TAPS-DMF buffer (50mM TAPS-NaOH (pH 8.5), 25mM $MgCl_2$, 50% DMF in H_2O) for 10 min at 55°C and the tagmented products were cleaned with the

MinElute PCR kit following the manufacturer's instructions. Finally, a master mix was prepared (2X NEBNext High Fidelity PCR Master Mix, 2.5 µM barcoded index primer, 2.5 µM P5-SMART-PCR primer) and added to the samples to attach the Illumina indices to the tagmented products in a PCR reaction (72°C for 5 min, 98°C for 30 s, 15 cycles of 98°C for 10 s, 63°C for 30 s, 72°C for 1 min). The pools were cleaned with 0.8 x volumetric ratio AMPure XP beads, were run with a High Sensitivity DNA5000 assay on a Tapestation 4200 (Agilent), and quantified using the Qubit high-sensitivity dsDNA assay. Seq-Well libraries were equimolarly pooled and clustered at 1.4pM concentration with 10% PhiX using High Output v2.1 chemistry on a NextSeq500 system. Sequencing was performed paired-end as followed: custom Drop-Seq Read 1 primer for 21 cycles, 8 cycles for the i7 index and 61 cycles for Read 2. Single-cell data were demultiplexed using bcl2fastq2 (v2.20).

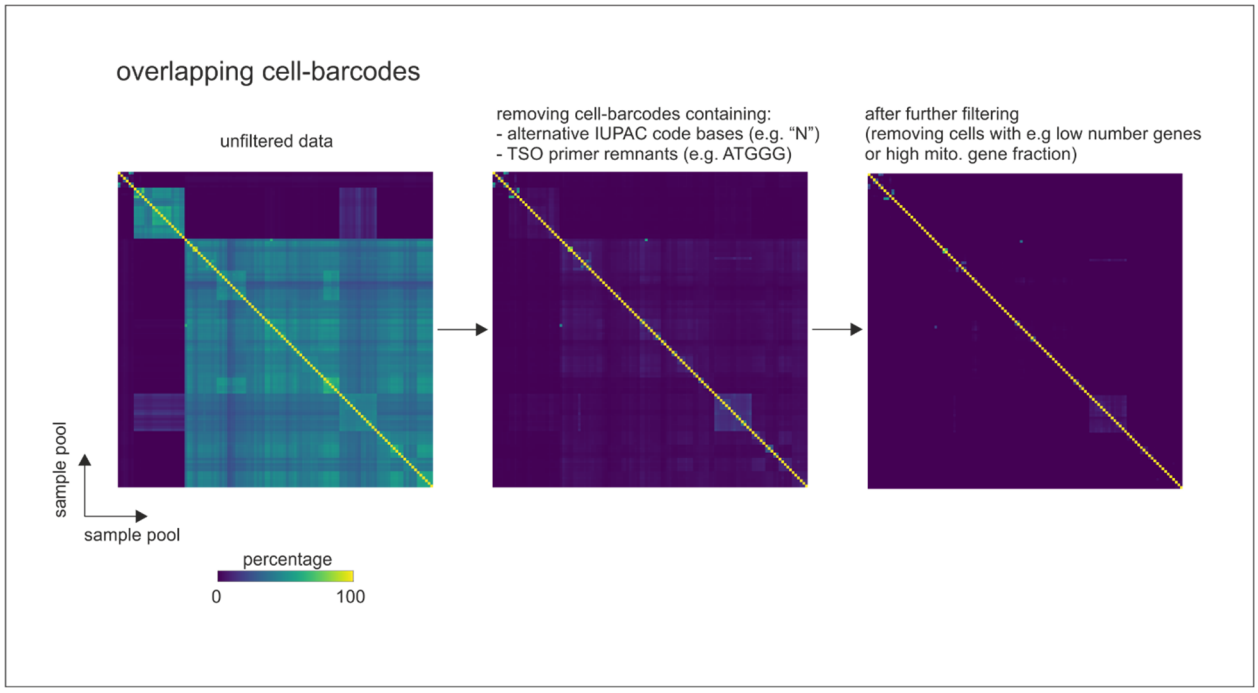


Figure 6. Percentage of overlapping cell barcodes across generated Seq-Well libraries.

3.13. Processing of scRNA-seq raw data

For preprocessing, the generated fastq files from both Chromium™ and Seq-Well were loaded into a data pre-processing pipeline (version 0.31, available at https://github.com/Hoohm/dropSeqPipe) that relies on Drop-seq tools provided by the McCarroll lab (Macosko et al., 2015b). STAR alignment within the pipeline was performed using the human GENCODE reference genome and transcriptome hg38 release 27 (Harrow et al., 2012). The resulting datasets were imported into R for further analyses. Interestingly, a remarkably high number of overlapping cell barcode sequences were detectable across pools (**Figure 6**) of the same sample. Further investigations showed that almost all overlapping cell barcodes started with the sequence 5'-ATGGGG-3'. It is important to note that some TSO primer used in the current study was based on the Smart Seq2 protocol (Picelli et al., 2014c), which means that two bases near the 3' end of the TSO primer sequence differed from the TSO primers used in the original Seq-Well protocol. Consequently, the final PCR during the generation of Seq-Well libraries led to the amplification of tagmentation fragments that did not contain cell barcode information (**Figure 7**). The primer annealing in this step also resulted in the first five bases in this erroneous cell barcode being 5'-ATGGG-3'. To account for this artefact, all putative cells containing cell barcodes, starting with either the sequence 5'-ATGGG-3' or 5'-GGG-3', and additionally cell barcodes with a Hamming-distance of 1 to 5'-ATGGG-3' were excluded. This greatly reduced the number of overlapping cell barcode sequences in the Seq-Well dataset (**Figure 6**).

Next, datasets were examined for content of mitochondrial ribosomal transcripts. For further downstream analyses, the highly abundant mitochondrial transcripts *MT-RNR1* and *MT-RNR2* were excluded. The resulting datasets were then imported into the R package ‘Seurat’ (v.3.0.0, Butler et al., 2018) for downstream analyses.

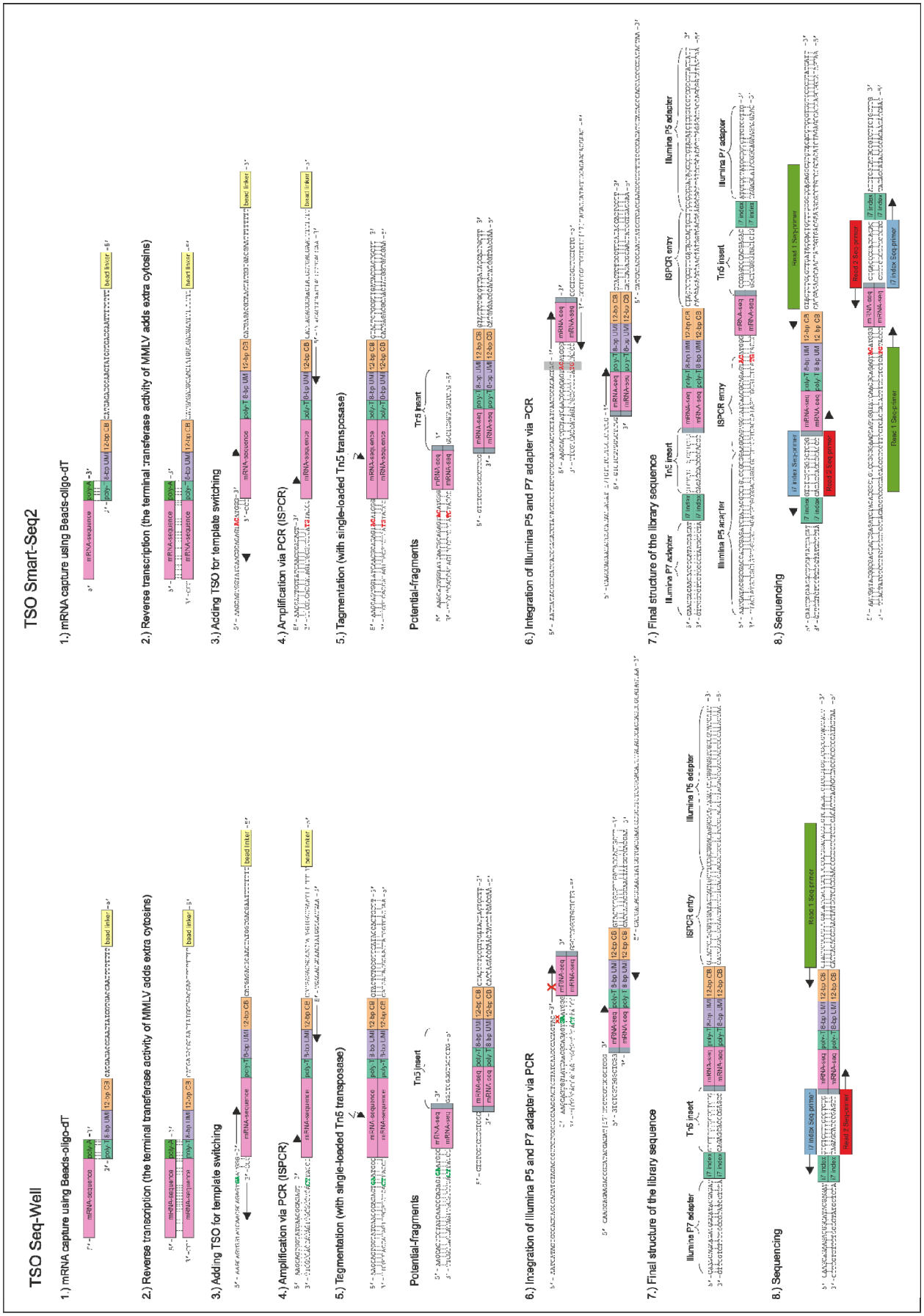

TSO Seq-Well
1.) mRNA capture using Beads-oligo-dT
2.) Reverse transcription (the terminal transferase activity of MMLV adds extra cytosins)
3.) Adding TSO for template switching
4.) Amplification via PCR (ISPCR)
5.) Tagmentation (with single-loaded Tn5 transposase)
Potential-fragments
Tn5 insert
6.) Integration of Illumina P5 and P7 adapter via PCR
7.) Final structure of the library sequence
Illumina P7 adapter
Illumina P5 adapter
ISPCR entry
8.) Sequencing
TSO Smart-Seq2
1.) mRNA capture using Beads-oligo-dT
2.) Reverse transcription (the terminal transferase activity of MMLV adds extra cytosins)
3.) Adding TSO for template switching
4.) Amplification via PCR (ISPCR)
5.) Tagmentation (with single-loaded Tn5 transposase)
Potential-fragments
Tn5 insert
6.) Integration of Illumina P5 and P7 adapter via PCR
7.) Final structure of the library sequence
Illumina P7 adapter
Illumina P5 adapter
ISPCR entry
Tn5 insert
Illumina P5 adapter
Illumina P7 adapter
8.) Sequencing

Figure 7. Artifacts in Seq-Well libraries (e.g. overlapping cell barcodes as shown in **Figure 6**) introduced by different TSO primers.

3.14. Quality control of scRNA-seq data

We defined cells and genes to be included for further analyses by the following criteria for each donor separately: (1) Only genes that were found in at least 3 cells were kept; (2) To retain granulocytes that contain only very limited number of transcripts, a relatively low threshold of 100 expressed genes was used to keep cells for further analyses; (3) With regard to the rate of endogenous-to-mitochondrial counts per cell, blood cells with a rate > 5% and lavage cells with a rate >10% were excluded. For the comparison of scRNA-seq methods for clinical applications, these quality control filters resulted in a Chromium™ dataset of 13,909 cells (BALF = 7,960 cells; blood = 5,949 cells) across 22,701 genes and a Seq-Well dataset comprised of 34,622 cells (BALF = 20,106 cells; blood = 14,516 cells) across 21,644 genes. For the integrated analysis of Seq-Well data from COPD GOLD 2 patients and control donors, we obtained a Seq-Well dataset of 60,925 lavage cells across 25,348 genes and 54,569 blood cells across 23,056 genes (**Table 6+7**).

Table 6: Cell numbers per annotated BALF cell type according to the four-step cell-type annotation approach

annotation	annotated states	total cell number	control cell numbers						COPD cell numbers								
			Control 5	Control 6	Control 7	Control 8	Control 9	Control 10	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
Airway epithelial cell		287	38	3	6	14	23	12	16	75	8	0	80	0	8		4
cDC1		255	9	12	117	2	10	5	13	16	54	1	1	12	3	0	0
cDC2		223	5	18	50	6	18	2	27	15	36	4	2	26	9	4	1
ILC		46	1	2	18	9	1		9	0	6	0	0	0	0	0	0
pDC		97	6	0	36	1	1	1	17	0	17	1	1	6	10	0	0
Macrophage 0	common Macrophage	14098	634	1426	2681	777	585	771	783	2213	953	1431	243	506	416	257	422
Macrophage 1	intermediate Macrophage	8990	525	691	1657	399	390	160	385	2872	669	397	123	294	73	191	164
Macrophage 2	PPBP+ Macrophage	6816	92	29	87	133	2	4	375	5809	22	34	124	52	7	0	46
Macrophage 3	C1Q+ Macrophage	4290	162	499	924	275	185	107	308	665	334	228	129	150	124	86	114
Macrophage 4	senescent Macrophage	4288	300	298	313	439	183	162	319	908	221	278	315	207	77	100	168
Macrophage 5	monocyte-like Macrophage	3797	131	326	734	128	125	70	366	574	480	253	64	300	77	86	83
Macrophage 6	HLA-DR+ Macrophage	2128	62	196	421	165	66	19	111	816	41	106	60	39	3	8	15
Macrophage 7	IFIT+ Macrophage	1331	40	74	396	116	21	32	96	357	64	28	31	33	20	6	17
Macrophage 8	proliferating Macrophage	1051	34	33	108	29	19	40	72	376	78	123	14	78	15	14	18
Macrophage 9	HLA-DQ+ Macrophage	958	12	132	331	35	49	17	18	231	16	56	24	21	2	3	11
Macrophage 10	Monocyte	463	37	21	75	10	50	19	38	22	38	20	9	105	9	4	6
Macrophage 11	ILC-like Macrophage	372	11	22	174	11	2	1	12	122	5	5	3	3	0	0	1
Macrophage 12	Macrophage/ erythrocyte	208	43	26	10	6	0	2	82	5	3	4	13	0	4	0	10
Mast cell		264	18	2	15	1	13	0	102	12	51	10	12	6	11	5	6
Neutrophil 1		284	9	5	11	16	10	5	117	0	58	5	1	22	8	1	16
Neutrophil 2		477	22	13	3	23	43	24	95	18	79	10	3	71	28	8	37
Neutrophil 3		471	13	11	18	33	45	21	106	13	104	34	10	23	21	6	13
Eosinophil		752	58	16	164	0	28	3	203	0	147	9	14	27	64	12	7
T cell 1		1964	38	66	779	115	148	35	172	8	217	6	41	118	136	82	3
T cell 2		1509	64	100	425	99	111	74	112	54	176	22	48	70	86	56	12
T cell 3		285	13	6	84	13	25	6	27	0	43	0	9	29	19	11	0
mixed cells		5221	128	307	1958	118	200	139	311	800	322	281	18	250	135	140	114

T**able 7:** Cell numbers per annotated blood cell type according to the four-step cell-type annotation approach

annotation	annotated states	total cell number	control cell numbers				COPD cell numbers					
			Control 6	Control 7	Control 8	Control 9	Patient 3	Patient 4	Patient 6	Patient 7	Patient 9	Patient 10
activated T cell/ NK cell		3177	393	214	295	102	608	403	130	151	342	539
proliferating T cell/ NK cell		45	1	7	3	2	4	11	7	4	1	5
CD4+ T cell		2750	219	296	354	119	220	290	211	423	186	432
CD8+ T cell		3915	186	407	270	109	842	649	453	375	214	410
NK cell		1197	70	101	308	55	104	122	63	84	96	194
DC		232	29	27	24	18	11	40	21	23	10	29
pDC		109	12	10	35	10	12	4	2	12	2	10
B cell		1244	86	105	241	30	216	128	15	112	85	226
Plasma cell		133	1	0	1	1	3	10	4	3	16	94
Eosinophil		282	8	21	34	13	15	58	43	21	14	55
immature Neutrophil		172	16	8	18	8	15	28	5	3	20	51
Neutrophil 1		5126	794	229	455	113	468	803	53	122	904	1185
Neutrophil 2		5618	725	102	358	98	1325	901	46	88	929	1046
Neutrophil 3		2733	562	88	168	63	230	162	216	43	745	456
Neutrophil 4		1570	154	31	55	30	137	167	31	38	413	514
Neutrophil 5		1211	185	26	88	21	116	244	10	26	203	292
Neutrophil 6		1202	179	27	85	21	194	136	20	19	238	283
Neutrophil 7		1176	133	34	65	15	114	141	13	13	315	333
Neutrophil 8		598	67	14	31	7	67	107	16	11	58	220
CD14+ Monocyte	classical Monocyte	6871	604	417	830	208	757	698	157	395	451	2354
CD14+ CD16+ Monocyte	intermediate Monocyte	1417	169	98	119	61	184	213	75	74	183	241
CD16+ Monocyte	non-classical Monocyte	1154	63	81	63	26	92	103	56	32	44	594
IFIT+ Monocyte		252	30	13	21	10	20	41	15	12	1	89
Erythrocyte		815	144	14	87	68	312	50	2	14	11	113
Megakaryocyte		246	62	5	25	4	26	23	17	6	43	35
mixed cell		3255	556	145	355	64	447	245	251	153	101	938
ribosomal-high mixed cell		6588	1106	323	944	340	2649	274	124	287	15	526
Doublet		1481	107	50	260	68	268	172	28	28	14	486

3.15. Dataset integration and dimensionality reduction of scRNA-seq data

If not stated otherwise, all following steps were conducted using the single-cell analysis pipeline Seurat. To account for variations in sequencing depth across cells, we applied a log-normalization strategy using CPM-normalization with a scale factor of 10,000. Next, the genes with the highest cell-to-cell variability in the dataset were determined by calculating the top 2,000 most variable genes by selecting the 'vst' method of the 'FindVariableFeatures' function in Seurat. For the comparison of scRNA-seq methods, the variable genes were determined separately for each technology, while for the integrated analysis of Seq-Well data from COPD GOLD 2 patients and control donors, variable genes were calculated separately for each donor. To analyze the data without having any influence of batch effects resulting from either different donors or technologies, an integration approach based on ‘anchors’ across batches (Stuart et al., 2019) was used to harmonize and integrate the different datasets by using the Seurat implementation with the default settings. After linear transformation of the remaining genes (scaling) to ensure homoscedasticity, the dimensionality of the data was reduced to 30 principal components (PCs) that was used as input for UMAP representation.

Next, doublet cells were identified utilizing the R package ‘DoubletFinder’ (version 2.0.2, McGinnis et al., 2019) by using the first 30 principal components of the non-integrated datasets, assuming a doublet formation rate of 10% and leaving all other parameters unaltered. The alleged duplicate cells were not removed from the dataset, but accumulations of these cells were highlighted and named accordingly. This procedure revealed, for example, that none of the identified AM clusters was defined by doublet cells.

3.16. Cell-type annotation based on reference transcriptomic datasets

Remark: The implementation of SingelR in Python and the development of GenExPro were realized in cooperation with the Comma Soft AG.

For the comparison of the datasets generated by the two different scRNA-seq technologies (see above), we developed a slightly modified Python implementation of SingleR (Aran et al., 2019) (commit a4afed8, https://github.com/dviraran/SingleR). The SingleR method iteratively computes the bivariate correlation between the respective cluster expression vector and each reference gene expression vector based on a set of differentially expressed (DE) genes. In each iteration, every cell type in the reference dataset is assigned a score based on this bivariate correlation. The cell type with the lowest score is dropped and the DE genes among the remaining cell types are computed and, based on these genes, the bivariate correlations are computed again. This procedure thereby iteratively reduces the number of cell types until only one best fitting cell type is retained. We reimplemented the SingleR functionality to assign cell types per cluster in Python to use in our framework and in addition to the original algorithm, we included a threshold for the bivariate correlation score based on tests with randomized reference data. This made it possible to label cell clusters as "unknown" if the bivariate correlation score of the best fitting reference cell type was below 0.1 and thus no cell type could be assigned. As a reference for SingleR, we used data from both Blueprint+ENCODE (Dunham et al., 2012; Stunnenberg et al., 2016) and the Human Primary Cell Atlas (HPCA) (Mabbott et al., 2013). In addition to the implementation of the SingleR algorithm in Python, we also modified the reference datasets by reducing the reference to immune cells and lung tissue cells. Furthermore, based on the experimental setting of the reference dataset, we adapted some cell labels, e.g. the neutrophils were divided into mature, immature and inflammatory neutrophils, whereas the original annotation had designated all these cells as neutrophils.

As another cell annotation approach, we developed the tool GenExPro (Gene Expression Profiler) that will be published elsewhere in detail. Similar to the implemented SingleR method, the basic idea of GenExPro is to compare the mean vector of gene expressions from a cluster of cells in the single-cell dataset to the expression profiles of a reference dataset of expression profiles with annotated cell types. Briefly, the GenExPro method fits a multiple linear regression for each cluster expression vector. The covariates in this regression are the reference expression vectors for each cell type that were obtained from the CIBERSORT algorithm (Newman et al., 2015). The more similar the cluster expression vector is to one of the reference expression vectors, the higher the regression coefficient for the respective

reference vector. If the highest regression vector is positive and above an uncorrected significance threshold of α = 0.05, the cluster is assigned the respective cell-type label of this reference cell type, otherwise, the cluster is labeled "Unassigned".

Technologically, we used the OLS (Ordinary least squares) regression model implemented in NIPY (version 0.4.2, https://github.com/nipy/nipy) to calculate the regression coefficients. As reference data, we used the leukocyte expression dataset LM22 (Newman et al., 2015). Similar to the SingleR references, the LM22 dataset was slightly modified, e.g. by subdividing neutrophils according to their activation state. We calculated the reference expression vectors by running CIBERSORT (version 1.06) on the modified LM22 dataset, leaving the default settings unchanged and setting the option "Filter non-hematopoietic genes from the signature matrix during construction". The obtained signature genes (derived from the calculated support vectors) were almost completely (>99 %; *data not shown*) contained in the signature genes of the original CIBERSORT publication (Newman et al., 2015).

Although both SingleR and GenExPro can be applied also to vectors of single-cell expressions, we applied it to the mean of expression vectors within a cluster for more robust results. Since both GenExPro and the modified SingleR are Python implementations, we performed clustering using the Louvain-clustering (Blondel et al., 2008) function of Scanpy (Wolf et al., 2018) by setting the number of neighbors to 24 and leaving the remaining parameters unaltered.

To assess the uncertainty of the annotation results, we added bootstrapping to GenExPro and SingleR. The basic principle of bootstrapping is to create an artificial dataset by sampling subjects, in our case cells, with replacement such that in the resulting artificial dataset some cells will be excluded, whereas others will be included more than once. The analyses are then repeated on multiple of these artificial datasets resulting in somewhat different results. For robust and certain patterns, different bootstrapped datasets generate similar results, while for random fluctuations different bootstraps result in highly different outcomes. Here, we conducted all cell typing analyses using 100 bootstrapped datasets.

3.17. Four-step strategy of cell-type annotation

For the cell-type annotation of the integrated 61K lavage and the 55K blood dataset we used a four-step strategy to annotate cells as well as to identify and finally remove cells of inferior quality (**Table 6+7**). The steps of the strategy include cell-type annotation, clustering and marker gene identification and the corresponding analytical details are described in the following three sections.

3.18. Consolidation of cell-type annotation using machine learning

Remark: The machine learning-based cell-type annotation was realized in cooperation with the Comma Soft AG.

To aggregate and consolidate the initial cell-type annotation, we trained a Gradient Boosting Classifier on the combined data of all datasets to classify each cell into a cell type. Gradient Boosting is a machine learning technique that combines multiple classification trees in order to assign an input to different classes. This method is highly flexible and robust in the classification task and has high predictive power. We used an implementation of the Gradient Boosting algorithm from scikit-learn (version 0.19.1, Pedregosa et al., 2012), the leading machine learning library for Python. For training the model, we used the raw gene expression matrix of each cell as input feature for the classification. We additionally extracted features from the data such as the type of tissue, the number of genes per cell, counts per cell, and the percentage of mitochondrial gene expression per cell. The training target of this model were the three cell-type labels from GenExPro and SingleR (Blueprint+Encode and HPCA). For this, we triplicated the data such that each cell with its feature vector was included three times, each with one label of the three cell-type annotations. Our aim was to apply the classifier to all cells in our data. However, as no distinct training data was available, we conducted a 3-fold cross-validation. In this procedure, two random thirds of a dataset were used as training data, and the model assigned cell type names to the remaining cells. Importantly, a cell with all three cell-type labels was only assigned either to the test or the training dataset. A major advantage of this machine learning method is that the classifier learns the specific expression profile of cell types and can take any cell-type annotation as input, independent of techniques such as bulk RNA-seq or microarray used as initial cell-type annotation reference. In addition, we were

able to apply the classifier at the single-cell level instead of the cluster mean expression level and thus achieved a higher resolution to exploit the full potential of scRNA-seq. This also allowed us to detect cell types with very low frequency in individual patients. Normally, these cells might end up in larger clusters with a different cell type and are therefore not detected. For all these reasons, this machine learning-based cell-type annotation is unbiased, reliable, reproducible and scalable.

3.19. Clustering of the integrated scRNA-seq datasets

The cellular heterogeneity of the integrated datasets was determined using a shared nearest neighbor (SNN)-graph based clustering algorithm implemented in the Seurat pipeline. For both the BALF and the blood data, we used the first 30 principle components as input and set the resolution to 0.7 and 0.6, respectively. The default setting for number of neighbors were used (k=20).

3.20. Marker gene identification of scRNA-seq data

DE genes between identified cell types/clusters (referred to as marker genes) were defined using a Wilcoxon rank sum test for differential gene expression implemented in Seurat. The significance threshold for marker genes were set to an adjusted p-value smaller than 0.001 and the logarithmic fold change cutoff to at least 0.4. In addition, the detected marker genes should have been expressed in at least 50% of the cells within the respective cell types/clusters. Visualization of the obtained marker genes were mainly done using Seurat functions, such as dot plot representation of cell types-/cluster-specific marker gene expression or heat map representation of marker genes across single cells. A more global overview of the expression profiles was obtained by calculating the mean expression values of marker genes per clusters, followed by scaling and centering of these values and representing them in a heatmap graph using the R package ‘pheatmap’ (version 1.0.12, https://CRAN.R-project.org/package=pheatmap), in which the genes were clustered according to the ‘ward.D’ agglomeration method.

3.21. 'Gene set distance' analysis of annotated cell types (GO-shuffling)

Remark: This approach was developed together with Stefanie Warnat-Herresthal.

Gene set annotations were downloaded from the Molecular Signatures Database v7.0 (MSigDB) and comprised gene sets from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa, 2019) database, the Pathway Interaction Database (PID) (Schaefer et al., 2009), the Reactome Pathway database (Fabregat et al., 2018), Hallmark gene sets (Liberzon et al., 2015), BioCarta Pathways (Nishimura, 2001) and Gene Ontology (GO) (Ashburner et al., 2000; Carbon et al., 2019). In addition, we retrieved gene sets from WikiPathways (Slenter et al., 2018). This search strategy resulted in a list of 12,755 gene sets, each containing a unique gene set term and a set of associated gene symbols.

As input, normalized scRNA-seq data was used, in which the cells were annotated according to the four-step cell-type annotation approach described above. Cell types containing at least 10 cells for each patient were retained and genes expressed in less than 5% of the cells in the respective cell type were excluded.

For each of the 12,755 gene sets, the "gene set distance" was calculated as follows for each cell type: Gene sets were taken into account that were present with a minimum of 3 genes. For each gene set, the Euclidean distance between all donors was calculated using the get_dist function from the R package 'factoextra' (version 1.0.5). Next, the mean distance of COPD patients, the mean distance of controls and the overall mean distance was calculated. The "gene set distance" was then defined as the overall mean distance divided by the mean distance of COPD patients plus the mean distance of control patients.

$$gene\ set\ distance = \frac{dist_{overall}}{dist_{COPD} + dist_{CTRL}}$$

This metric allows to determine for which gene set the quotient takes a value close to or greater than 1, which means that the distance within the groups (COPD ($dist_{COPD}$) or control ($dist_{CTRL}$)) is smaller than the overall distance ($dist_{overall}$) and consequently the distance is mainly defined by the difference between the groups. Since the Euclidean distance metric is prone to be affected by outliers in higher dimensions, we also tested this approach by using the Manhattan distance and got comparable results.

For each cell type, we ranked the gene sets by their gene set distance. Visualization of the most frequent terms contained in the upper percentile of the predicted gene sets in the AM clusters was performed using the R package 'wordcloud' (version 2.6), in which filler and connective words were excluded. Alternatively, the gene sets in the upper percentile were filtered for association with 'NOTCH', 'lipidomics', or 'immune system', whereby the latter was visualized in a bar plot showing the proportion of 'immune system'-related gene sets among all gene sets found in the upper percentile.

3.22. Modeling of metabolic pathways based on scRNA-seq data

The metabolic landscape of AMs was modelled using the Compass method (version 0.9.5, Wagner et al., 2020; Wang et al., 2020) by leaving the standard settings unaltered (model: RECON2 (Thiele et al., 2013); lambda: 0; media: media1, which represents a rich extracellular medium as defined in the Compass manuscript). As input, we simplified the single-cell data of the AMs by using the 'applyMicroClustering' function of the R package 'VISION' (version 2.1.0, DeTomaso et al., 2019), resulting in approximately 20 microclusters per patient. Next, we applied Compass to the microclusters for each donor separately. The output tables representing Compass scores for single reactions and synthesis of single metabolites of the individual donors were imported into R. They were concatenated and finally transformed as described in the Compass manuscript, except for disabling the division into meta-reactions. In detail, the concatenated output table x was first negatively log-transformed ($y = -\log(1+x)$), the global minimum value of table y was subtracted from the values ($z = y - \min(y)$) and the resulting table z was then used for further analysis. To determine which reactions and metabolites are significantly different between control donors and COPD patients, with the differences being reproducible in the COPD population, we performed Wilcoxon rank sum tests on Compass scores. We firstly computed the Wilcoxon p-value for every patient separately against all controls, took the median of these p-values, and kept reactions/metabolites for which -log10(median p-value) ≥ 2.5. We derived a second list of reactions and metabolites by similarly comparing control donors separately against all patients. The reactions and metabolites that have significant differences are the union of these two lists. Next, we excluded reactions with the lowest confidence score in the metabolic reconstruction (Thiele and Palsson,

2010), i.e., we discarded reactions with a confidence score of 1 and kept confidence scores of 2-4 (as well as 0 which is reserved for unannotated confidence). We also excluded metabolites that localize to cellular compartments other than the cytoplasm [c], extracellular space [e] or mitochondria [m]. Finally, the remaining reactions and metabolites were annotated using the Virtual Metabolic Human (VMH) database (Noronha et al., 2019) and visualized in a heat map.

3.23. Cell cycle state analysis of scRNA-Seq data

To categorize the cells within the AM clusters into the respective cell cycle states, we applied the ‘CellCycleScoring’ function of Seurat and substantiated the results using the ‘cyclone’ function (Scialdone et al., 2015) implemented in the R package ‘scran’ (version 1.10.2, Lun et al., 2016).

3.24. Gene set variation analysis

To predict the functions of the AM states, we performed gene set variation analysis (GSVA) (Hänzelmann et al., 2013a) by using the R package ‘GSVA’ (version 1.30.0) and defining ‘Poisson’ for the non-parametric estimation of the cumulative distribution function of expression levels across donors. For the GSVA input expression table, we calculated the sum of the expression of normalized scRNA-seq data for each patient in any AM states. As gene sets we used the gene set collection described in the section ‘GO-shuffling’ and additionally included the 'ImmuneSigDB' collection of MsigDB, whereby this collection was reduced to gene sets that had one of the following terms in the gene set description: 'Mono', 'Macro', 'MDC', 'MDM', 'Dend' and 'DC'. This resulted in 14,160 gene sets. Similar to GO-shuffling, we filtered this collection for gene sets that were present with a minimum of 3 genes in a respective AM state. We applied an additional filter step to increase the stringency of the analysis. Therefore, we retained only gene sets in which the sum of the genes contained in the set were expressed in more than 30% of an AM state. The GSVA results per donor were combined for the respective AM state using a Borda rank and the top 250 ranked gene

sets per subtype were visualized in an UpSet plot using the R package 'UpSetR' (version 1.3.3, Conway et al., 2017)).

3.25. AUCell for gene set enrichment analysis

Enrichment of gene sets was performed using the 'AUCell' method (Aibar et al., 2017) implemented in the package (version 1.4.1) in R. We set the threshold for the calculation of the area under the curve (AUC) to the top 3% of the ranked genes and normalized the maximum possible AUC to 1. The resulting AUC values were subsequently visualized in a violin plot. This approach was used, for example, in **Figure 32** to assess the enrichment of monocyte-derived macrophage signature genes provided by Wohnhaas (unpublished results). This signature was obtained from scRNA-seq data of monocyte-derived macrophages that were identified in BALF of a murine 12-week smoke model. Human orthologues (obtained from BioMart (version 2.42.0, Durinck et al., 2005)) of the murine marker genes were used for the enrichment analysis. In a similar way, we also performed the enrichment of monocyte-derived macrophage signatures obtained by Jaitin et al. (Jaitin et al., 2019) and Kim et al. (Kim et al., 2018).

3.26. Distribution-free DE analysis across patient groups

Remark: This approach was conceived by Jan Hasenauer and realized by Erika Dudkin.

To analyze the differences between patient and control cohort we developed a distribution-free test, which preserves patient and cell information. In contrast to available methods, it avoids the use of mini-bulk, the pooling of cells from different patients, and distribution assumptions. As input, we use the afore-computed cell cluster information and the normalized single-cell data.

For each AM cluster a differential expression between patient and control cohort was performed. Therefore, individuals not possessing cells in a cluster – which happened in a few cases – and genes expressed in less than 10% of cells were disregarded for the analysis of this cluster. For each gene, the differences between all possible pairs

of patients and controls was assessed using the nonparametric Wilcoxon rank sum test. The Wilcoxon rank sum test does not rely on a specific distribution assumption. This is beneficial as the distribution of single-cell expressions is often skewed or shows multiple modes. Furthermore, scRNA-seq data are characterized by multiple zero counts, which the Wilcoxon rank sum test takes into account. To assess the differences between patient groups, the median Wilcoxon score of the pair-wise tests was used considered as test statistic. This test statistic can be negative, describing here an overexpression of a gene in the patient cohort, and positive, relating here to an overexpression of a gene in the control cohort.

To assess if the observed value of the test statistic was significant, the probability of observing an equally or more extreme value of the test statistic under the null hypothesis was evaluated. The null hypothesis was that there is no difference between the two groups. The exact null distribution was evaluated with the permutation test, taking all possible permutations into account. A permutation is a rearrangement of patients into the opposing group, which conserves the group sizes. For all possible permutations the afore-described test statistic – the median Wilcoxon score – was evaluated. The distribution of the test statistic over all permutations provided the null distribution, since reshuffling of patients should not be significant under the null hypothesis. The p-value for the observed group assignment was then the fraction of permutations that led to an equal or more extreme value of the test statistic than the value of the test statistic of the observed patient arrangement.

3.27. Application of the novel DE analysis approach and GSEA

DE analysis was performed for all AM clusters and the results are provided in **Table S4**. For the classification of genes being significantly DE, a test statistic cutoff of 0.75 was chosen. Additionally, for each AM cluster, the DE genes were sorted ascendingly according to their p-values and the 300 top ranked genes were chosen. The visualization of which DE genes are found and shared in which AM clusters was performed using the UpSetR package in R.

Gene set enrichment analysis (GSEA) was performed to identify shared common biological functions by groups of DE genes. The web-tool ‘g:Profiler’ (version e98_eg45_p14_ce5b097, Reimand et al., 2007) was used to perform the functional

profiling of the DE genes of interest (genes fulfilling the cutoff criteria for DE genes in >2 AM clusters). As multiple-testing correction method, g:Profiler's in-house g:SCS algorithm was chosen, which corrects for multiple tests that are dependent on each other, which holds true for the hierarchically arranged GO terms. The analysis was done using the Gene Ontology (Ashburner et al., 2000; Carbon et al., 2019) database as well as biological pathway databases, like KEGG (Kanehisa, 2019), Reactome (Fabregat et al., 2018) and WikiPathways (Slenter et al., 2018).

3.28. Transcription factor prediction using iRegulon

The Cytoscape (version v3.7.1, Shannon et al., 2003) plug-in iRegulon (version 1.3, Janky et al., 2014) uses a large motif-database, consisting of up to ten thousand position weight matrices collected from various species, to predict the transcription factors (TFs) upstream of an input gene set. In the present study, we predicted the upstream regulators of calculated DE genes within an AM state between COPD patients and control donors. We restricted the prediction to AM states with a minimum of 30 DE genes. The genomic regions for TF-motif search was limited to 10kbp around the respective transcriptional start sites. For each AM state, we predicted TFs upstream of its DE-genes by using the default settings of iRegulon. The predicted TFs were visualized in an UpSet plot. The orthologous expression of a selection of the predicted TFs was further investigated in a mouse lung cell dataset (Angelidis et al., 2019) .

3.29. Cell-to-cell communication

Remark: This analysis part was conducted in cooperation with Nico Reusch and Patrick Günther.

Potential cell-cell-interactions were inferred using 'CellPhoneDB' (version 2.1.1, Efremova et al., 2019; Vento-Tormo et al., 2018). As input, we used the normalized gene expression matrix of control and COPD patients that was filtered separately for cell types, which were defined by the four-step cell-type annotation approach and identified in at least three patients of any group (COPD or control) and contained ≥ 10 cells per patient. Genes were filtered for being expressed in ≥ 5% of a respective cell

type. To run CellPhoneDB, the following parameters were set: --iterations=1,000 --pvalue=0.1 --result-precision=10.

In order to visualize the cell-cell communication, we filtered for significant interactions (adjusted (Holm) p-value < 0.05) and summarized the interactions per cell-type pair. Network visualizations was done with the 'ggraph' package (version 1.0.2) setting the layout to "fr". Next, sender and receiver interactions were split and the difference of significant interactions between COPD and control was depicted as a heat map. To visualize single receptor-ligand pairs, we filtered for group-specific interactions (-log10(p-value) > 1) and visualized the resulting interactions for control and COPD.

To evaluate the downstream transcriptomic changes caused by cell-cell-interactions, we applied 'NicheNet' (version 0.1.0, Bonnardel et al., 2019; Browaeys et al., 2019). As the CellPhoneDB analysis revealed a central role of the mono-like macrophages in the cellular communication in BALF, we focused on these cells for the subsequent analysis. As the model in NicheNet is based on a different collection of databases than CellPhoneDB, we defined potential sender cell-receiver cell interactions independent of CellPhoneDB. As potential ligands, we accepted all genes that were expressed in >5% of any cell type within the COPD group and which matched at least one receptor from the genes expressed in > 5% of the mono-like macrophages in the COPD group. As input genes to infer the ligand activity score from, we defined all DE genes with a median Wilcoxon score < (-0.75) and p value of the median Wilcox score <0.05. As background genes, we defined all genes that are not DE in mono-like macrophages and expressed in > 5% of mono-like macrophages. For ligand prioritization, we selected the top 20 genes with the highest PCC or AUPR resulting in 26 top ligands. To not miss out any cell type-specific cell-cell-interaction, we additionally used every cell type separately as sender cell and chose the top 7 genes according to the PCC and added these to the top ligands resulting in 32 top genes.

The expression of these ligands for each cell type was visualized in a heat map scaled by each gene. The genes and cell types were sorted by hierarchical clustering. This way, each gene could be assigned one main sender cell. The target genes of all top ligands were visualized in a heat map with their regulatory potential score for each ligand and their mean expression in every patient (scaled by gene and sorted by hierarchical clustering of the patients). The genes were manually assigned to corresponding pathways. The connection of ligands and target genes was additionally displayed in a circos plot connecting all ligands with target genes for which any

connection was defined in the ligand-target matrix. To further decipher the exact connection between the ligand and the target genes, we visualized the transcriptional network based on which NicheNet associated the target genes with *TGFB1* in a network with free topology. This network was subdivided into receptors for *TGFB1*, transcriptional regulators between *TGFB1* and the target genes. The connections were subdivided into signaling (which does not induce a direct transcriptional change) and transcriptional regulation.

3.30. Monocyte-to-macrophage trajectory analysis

Remark: This analysis part was conducted in cooperation with Malte Lücken under the supervision of Fabian J. Theis.

To generate a joint embedding of BAL and blood samples, the data were jointly pre-processed using 'Scanpy' (version 1.4.3 commit 0075c62, Wolf et al., 2018) on AnnData (version 0.6.22.post2 commit 72c2bde). In concordance with previous analysis, cells from BALF were filtered out if the fraction of mitochondrial reads exceeded 0.1, and a threshold of 0.05 was used for blood samples. Genes that were expressed in fewer than 200 cells were also filtered out. Following previously published best-practices (Luecken and Theis, 2019) we used scran normalization via the computeSumFactors function on the joint object. Spliced and unspliced counts were mapped to this object using scVelo (version 0.1.24 commit e45a65a, Bergen et al., 2019). Quality control for spliced and unspliced counts was performed by removing cells with fewer than 20 spliced and/or 10 unspliced counts. Subsequent normalization by total counts and log-transformation was performed via the filter_and_normalize function from scVelo. Subsetting to only relevant monocyte and macrophage populations from blood and BAL datasets (**Figure 35A**) resulted in a dataset of 57,280 cells and 11,530 genes.

The joint embedding of BAL and blood cells was generated by taking the top 4000 highly variable genes (HVGs) that were shared by most batches. This was done using the hvg_batch function from the single-cell data integration benchmarking package 'scIB' (https://github.com/theislab/scib; published separately). This function computes the top 4000 HVGs per batch (here: donor) using Scanpy's highly_variable_genes function with method cell_ranger and ranks these by the number of batches these

genes are highly variable in, and by their mean dispersion over all batches. In this list, the top 4000 genes are selected. On this gene set, we computed the top 50 principle components and used Euclidean distance on these to compute a kNN graph with a k=15. 'UMAP' (version 0.3.9, McInnes et al., 2018) was used to visualize the results.
Due to an observed batch effect when performing RNA velocity analysis across patients, we ran scVelo per patient and aggregated the individual patient velocities to create a joint velocity embedding. For each donor spliced and unspliced counts were smoothed using the moments function, velocity genes were selected by a stringent log likelihood threshold of 0.1 (between 45 and 172 genes per donor), and the dynamical scVelo model was fit. The resulting inferred single-cell velocities were projected onto the joint UMAP computed from all donors by running velocity_graph on the concatenated object.
Furthermore, partition-based graph abstraction ('PAGA', Wolf et al., 2019) was used to assess the connectivity of cell identity clusters that were suggested to show transitions by RNA velocity. To robustly assess the connectivity of cell identity clusters across donors, we performed PAGA analysis per donor. We computed a kNN graph with Scanpy's neighbors function (k=15) per donor using the joint PCA embedding across donors and ran the paga function on this graph. We used the resulting PAGA connectivities as a statistical test of kNN-graph connectivity between clusters. The median of PAGA connectivities over all donors with both blood and BAL samples was used as a PAGA distance metric.

3.31. Trajectory analysis of murine tissue monocyte differentiation

Remark: *The experimental part was conducted by Andreas Schlitzer and the scRNA-seq libraries were generated in Singapore at the A*STAR Immunology Network.*
MDPs and cMoPs were isolated from the bone marrow and Ly6c$^+$ monocytes were isolated from bone marrow, peripheral blood, lung, skin and intestine of mice. Full-length single-cell transcriptome data of these cells were generated using the C1 system (Fluidigm) followed by sequencing. The fastq files were aligned using STAR (version 2.5.1b) and quantified using RSEM (version 1.2.28) by using the mouse reference genome (mm10) that was used in the original Drop-Seq publication (Macosko et al., 2015b) and which is accessible via the GEO identification number

GSE63472. Genes, whose name began with 'GM' and was followed by a number, contributed to the instability in the subsequent analyses (*data not shown*) and were therefore removed from the trajectory construction together with the ribosomal genes. Cells were retained if more than 75% of the expressed genes were not mitochondrial and the respective cell expressed more than 1,000 genes. After this filtering step, mitochondrial genes were also excluded from further analyses, which resulted in a dataset comprising 16,433 genes across 323 cells. The cell-cycle states of the cell populations were determined using the cyclone' function of the scran package. Afterwards, cellular trajectories were derived using Monocle 2 (version 2.8.0, Qiu et al., 2017a, 2017b; Trapnell et al., 2014), whereby the TPM data were computational converted into quasi-transcript counts using Census (Qiu et al., 2017b). Subsequently, the Monocle object was generated by setting the expected data distribution parameter to negative binomial. Next, genes for building the trajectory were determined by using a slightly modified version of the dbFeature strategy that is described in the Monocle tutorial (http://cole-trapnell-lab.github.io/monocle-release/docs/#constructing-single-cell-trajectories). Briefly, the genes, which defined the highest variance in the dataset and were therefore included in the top PCs, were determined not only for a single 'num_cells_expressed' value, but for a series of values between 5 and 12. For each value, the top 100 genes for the first three PCs were determined, combined, and the 100 most frequent genes were identified by Borda rank. We excluded PC1 according to the recommendation of the Monocle developers, who claimed that this PC is often driven by technical artifacts. This approach resulted in 189 genes that was used as input for the 'setOrderingFilter' function of Monocle, followed by dimensionality reduction to three dimensions using the 'DDRTree' method. The trajectory construction revealed at least 6 cell states/branches, but to enable a possible investigation of early and late cell differentiation, we further separated the skin and intestinal states into two sub-states. DE genes between the identified cellular states were calculated using the Monocle function 'differentialGeneTest' and further filtered for genes, which were expressed in more than 50% of the cells in the respective state, had a maximal expression of larger than 10, and exhibited a q-value < 0.01.

3.32. Data visualization

In general, Seurat and the ggplot2 package (version 3.1.0, Wickham, 2016) was used to generate figures. For the monocyte-to-macrophage analysis Scanpy, UMAP and scVelo packages were used to generate figures.

3.33. Statistical analysis

If not otherwise stated, the statistical evaluation was carried out in relation to the total sample size *n*. A t-test was used for $n \leq 10$, otherwise a Wilcoxon rank-sum test was used.

4. Results

4.1. Using MCFC to profile human BALF immune cells

4.1.1. Characterization of the general immune landscape in the alveolar space

BALF is a biospecimen, which is used in the clinics for the diagnosis of various lung diseases and in the scientific world for the examination of e.g. immune cells. BALF provides information about the processes in the lumen of the distal airways and in particular of the alveolar space. In order to gain a general overview of the composition of immune cells in the alveolar space, we first designed an MCFC panel, which enabled us to identify both myeloid cells (monocytes, alveolar macrophages, DCs, mast cells, neutrophils, and eosinophils) and lymphoid cells (T cells, B cells, NK cells and ILCs) (**Table 5, Material and Methods**). We also included markers for basophils, but none of the samples analyzed showed a definite population. To ensure that the MCFC analysis results are not skewed by low quality BALF samples, we defined strict quality criteria (Meyer et al., 2012), which led to the inclusion of 29 control donors and 17 COPD GOLD grade 2 patients. A classical gating strategy (**Figure 8A**) of the MCFC data showed that most cells in the alveolar space were mainly HLA-DR^{+}, CD14^{low} and highly autofluorescent cells, which - based on prior knowledge (Bharat et al., 2016; van Haarst et al., 1994; Vermaelen and Pauwels, 2004) - defined these cells as AMs. In addition to AMs, granulocytes (mainly neutrophils), monocytes, DCs, NK cells, T cells, and a few B cells and ILCs were detected, which thus together form the immune compartment in the alveolar space.

Since the classical gating strategy can be very cumbersome and the designed panels defined a high-dimensional space, we attempted a more convenient and unbiased way for MFCF data analysis by using UMAP to reduce the dimensionality of the flow cytometry data. As input for the UMAP calculation, we focused on CD45^{+} cells and the resulting structures within the UMAP plot (**Figure 8B and C**) showed a high degree of conformance to the gates defined in the classical analysis approach (**Figure 8A**), indicating that UMAP-

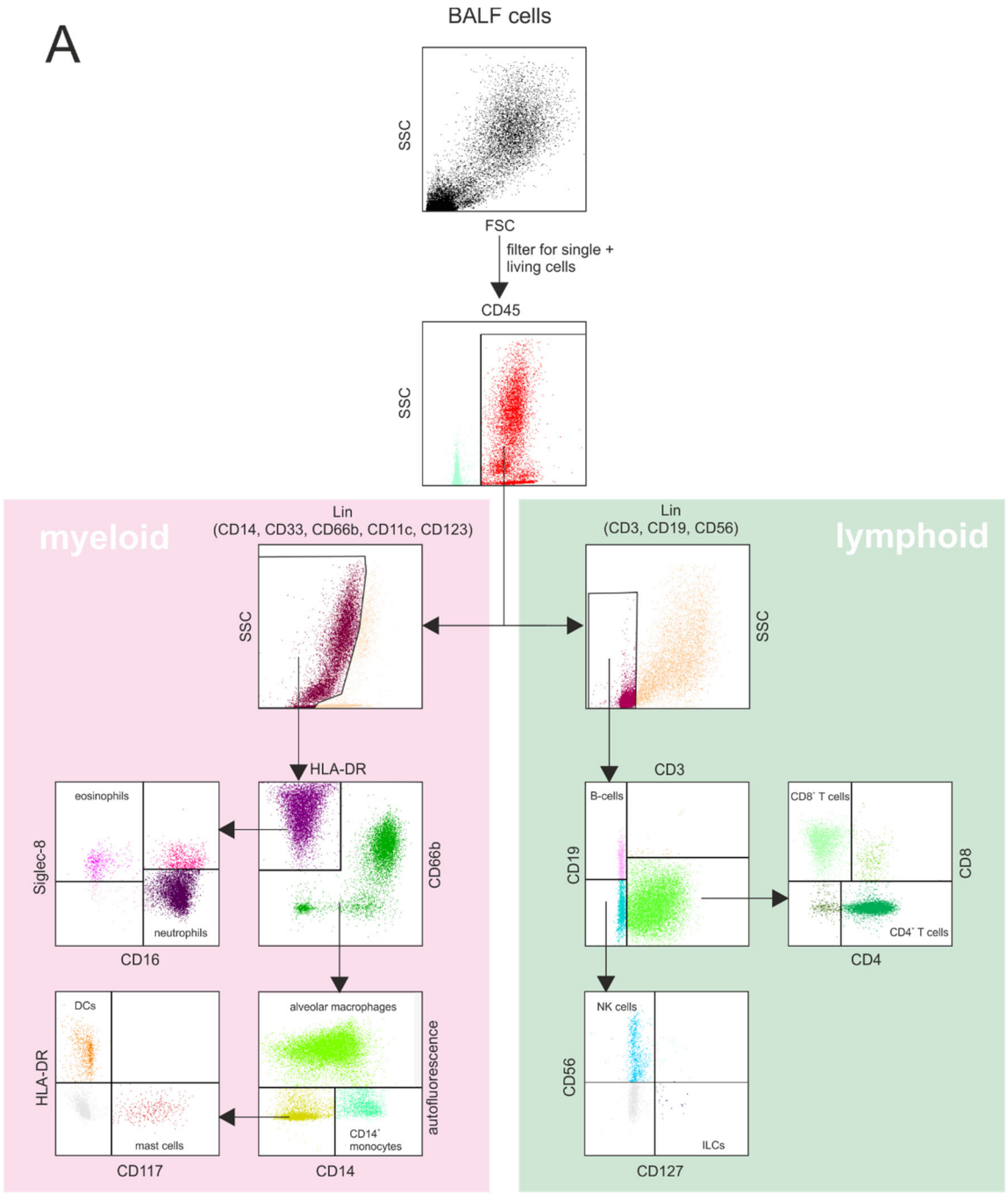

A
BALF cells
SSC
FSC
filter for single + living cells
CD45
SSC
myeloid
Lin
(CD14, CD33, CD66b, CD11c, CD123)
SSC
HLA-DR
eosinophils
Siglec-8
neutrophils
CD16
CD66b
DCs
HLA-DR
mast cells
CD117
alveolar macrophages
CD14+ monocytes
CD14
autofluorescence
lymphoid
Lin
(CD3, CD19, CD56)
SSC
CD3
B-cells
CD19
CD8+ T cells
CD8
CD4+ T cells
CD4
NK cells
CD56
ILCs
CD127
B
CD45+ Lin- myeloid cells
UMAP2
UMAP1
coloring acc. to gate color in A
C
CD45+ Lin- lymphoid cells
UMAP2
UMAP1
coloring acc. to gate color in A

Figure 8. Analysis of MCFC data of BALF immune cells

(A) Gating strategy to identify major immune cells in the alveolar space. **(B+C)** UMAP representation of $CD45^+$ Lin^- cells. Coloring according to the gates in (A).

based analysis of MCFC provides a fast, robust, and unbiased alternative for MCFC data analysis.

4.1.2. Increased numbers of neutrophils in the alveolar space of COPD patients

Next, we sought to determine whether there is an alteration in the relative population sizes of the identified immune cell types in BALF of COPD patients. For this purpose, we quantified the relative frequencies of immune cells for the COPD and control samples based on MCFC analysis (**Figure 9**). Our data revealed that the most significant change in cell type frequencies between COPD and control occurred in neutrophils (**Figure 9A**), which is one of the most common cell types in the alveolar space. In addition, there was a clear trend towards an increase in the relative cell numbers of eosinophils and AMs in COPD patients. In the lymphoid space, the strongest, although not significant, difference was found for the ILC population, which again showed a trend towards an increased cell number in COPD patients (**Figure 9B**).

Taken together, the AM population in the alveolar space represents the most abundant cell type in both COPD patients and control donors, whereby the largest relative change in the cell population in COPD was found in the neutrophil population.

Although MCFC was sufficient to describe the major cell types and disease-related changes in the population, it lacked the resolution to identify cellular subtypes that might be the actual underlying denominators of COPD pathogenesis.

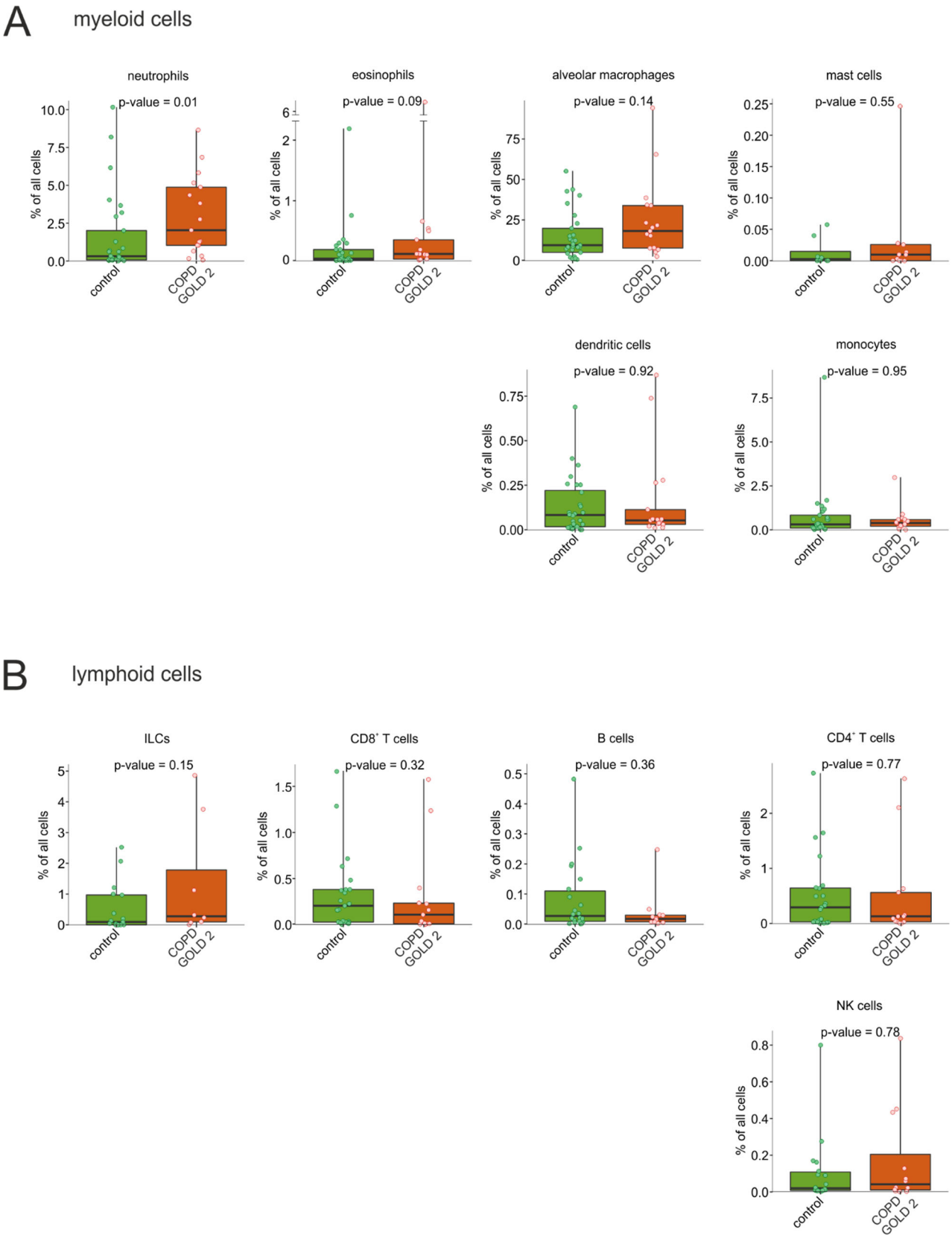

A
myeloid cells
neutrophils
p-value = 0.01
eosinophils
p-value = 0.09
alveolar macrophages
p-value = 0.14
mast cells
p-value = 0.55
dendritic cells
p-value = 0.92
monocytes
p-value = 0.95
% of all cells
control
COPD GOLD 2
B
lymphoid cells
ILCs
p-value = 0.15
CD8+ T cells
p-value = 0.32
B cells
p-value = 0.36
CD4+ T cells
p-value = 0.77
NK cells
p-value = 0.78

Figure 9. Quantification of the relative population sizes of BALF immune cells based on MCFC data

Population sizes are represented as cell type frequencies calculated per donor and visualized in box plots (in total: control n = 29, COPD n = 17). **(A)** Relative population sizes of myeloid cells. **(B)** Relative population sizes of lymphoid cells.

4.2. Assessment of scRNA-seq technologies for the investigation of BALF cells

4.2.1. Defining a ground truth about the distribution of cell types using MCFC

For an unbiased and in-depth characterization of the immune landscape in the alveolar space, we aimed for applying scRNA-seq (**Figure 10A**). However, the available scRNA-seq methods differ in terms of throughput and information content (Ding et al., 2020). In addition, some technologies even have a bias towards the cell types they detect, which can be particularly detrimental when investigating disease-associated specimens, as important cell types can get lost. Therefore, a suitable scRNA-seq method for use on clinically relevant biospecimens has to fulfil a series of criteria, including: 1) capture of the major cell types, 2) adequate scalability, and 3) sufficient feature recognition.

Before we evaluated scRNA-seq technologies for BALF cells, we defined a ground truth about the major cell types present in BALF samples by performing MCFC on the same samples that were used for scRNA-seq in the later steps (**Figure 10A**). As mentioned above, the analysis of BALF MCFC data can be performed using a UMAP-based approach, since the resulting structures resemble the classical gating strategies (**Figure 10B and C**). Based on this finding, we developed an analysis strategy that is basically relying on the MCFC analysis of a single sample, whose results can then be transferred to the remaining samples (**Figure 10B**). In particular, we first defined a sample as a reference case on which the two-dimensional UMAP topology of the cell space was predicted, followed by a clustering of the cells (**Figure 10B**). Next, the identified clusters

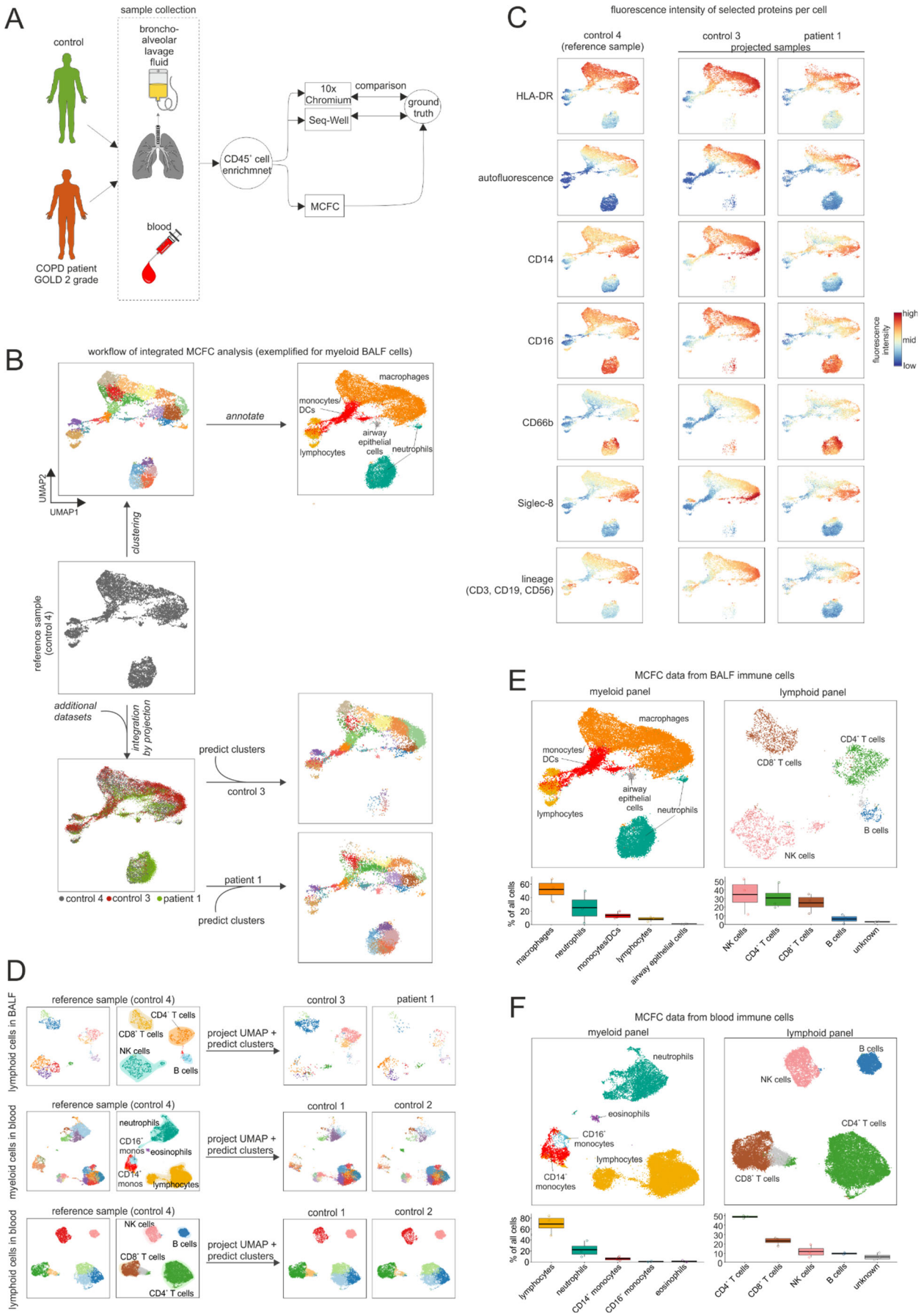

A
sample collection
control
broncho-alveolar lavage fluid
blood
COPD patient GOLD 2 grade
CD45+ cell enrichmnet
10x Chromium
Seq-Well
comparison
ground truth
MCFC
B
workflow of integrated MCFC analysis (exemplified for myeloid BALF cells)
annotate
macrophages
monocytes/ DCs
airway epithelial cells
lymphocytes
neutrophils
UMAP2
UMAP1
clustering
reference sample (control 4)
additional datasets
integration by projection
predict clusters
control 3
patient 1
control 4 control 3 patient 1
C
fluorescence intensity of selected proteins per cell
control 4 (reference sample)
control 3
patient 1
projected samples
HLA-DR
autofluorescence
CD14
CD16
CD66b
Siglec-8
lineage (CD3, CD19, CD56)
fluorescence intensity
high
mid
low
D
reference sample (control 4)
lymphoid cells in BALF
CD4+ T cells
CD8+ T cells
NK cells
B cells
project UMAP + predict clusters
control 3
patient 1
myeloid cells in blood
neutrophils
CD16+ monos
eosinophils
CD14+ monos
lymphocytes
control 1
control 2
lymphoid cells in blood
E
MCFC data from BALF immune cells
myeloid panel
lymphoid panel
% of all cells
macrophages
neutrophils
monocytes/DCs
lymphocytes
airway epithelial cells
unknown
F
MCFC data from blood immune cells
CD16+ monocytes
CD14+ monocytes

Figure 10. Using MCFC to generate a ground truth for comparing scRNA-seq technologies

(A) Schema describing the workflow of the comparison. **(B)** Workflow of MCFC analysis with data obtained from different patients, starting with the UMAP representation of a compensated reference sample, through clustering of the data and annotation of the clusters based on the protein fluorescence intensities (top panel). The other samples are then projected onto the UMAP of the reference sample and the accuracy of the projection is evaluated by predicting and comparing clusters (bottom panel). This workflow is exemplarily shown for the myeloid compartment in BALF. **(C)** Feature-plot representation of fluorescence intensities measured by MCFC. This information formed the basis for the assignment of clusters to cell types. **(D)** Application of the MCFC analysis described in (B) to the myeloid compartment in blood and the lymphoid compartment in BALF and blood. **(E+F)** UMAP representation of MCFC data obtained from BALF **(E)** and blood **(F)** immune cells. The relative proportion of cell types identified in the three donor samples displayed as boxplots. These values served a ground truth for later scRNA-seq technology comparisons.

were assigned to cell types based on the fluorescence intensities of the respective proteins (**Figure 10B**). Finally, the remaining samples were projected onto the UMAP plot of the reference case. In order to verify the validity of the projection, we checked whether cells with comparable fluorescence intensities were projected in a topologically similar way (**Figure 10C**). In addition, clusters for the remaining samples were predicted based on the reference case and their topological distribution was again evaluated (**Figure 10B+D**).

Applying this novel MCFC analysis strategy to BALF samples revealed that AMs, granulocytes, monocytes, DCs, NK cells, T cells and a few B cells define the immune compartment in the alveolar space, with AMs and neutrophils being the most abundant cell types (**Figure 10B-E**). These results are in accordance with the results from the classical gating strategy described in **Figure 8 and 9.** We also applied this analysis strategy to immune cells of the blood (**Figure 10D+F**), where neutrophils were found to be the second most common cell type. However, this does not reflect the physiological distribution of the cell types, as during isolation the blood granulocytes were mixed with the remaining PBMCs in a 1:2 ratio (**Material and Methods**).

4.2.2. Comparison of scRNA-seq technologies

Some of the identified cell types in the alveolar space, especially granulocytes, do not withstand cryopreservation (Boonlayangoor et al., 1980), which is why we decided to

conduct the COPD study on freshly isolated BALF and blood samples. For the determination of a suitable scRNA-seq technology, we compared Chromium (10x Genomics; Zheng et al., 2017) with the Seq-Well method (Gierahn et al., 2017). We applied $CD45^+$ immune cells derived from the same samples that were used for MCFC (**Figure 10**), which resulted in a scRNA-seq dataset comprising 48,531 cells (28,066 BALF cells and 20,465 blood cells). In a first comparative evaluation step, we investigated standard quality parameters of scRNA Seq datasets, which revealed a slightly higher rate of uniquely aligned reads (**Figure 11A**) and a tendency towards the detection of more cells (**Figure 11B**) in the Seq-Well method, whereas the Chromium technology exhibited a substantially higher rate of reads/cell (**Figure 11C**), transcripts/cell (**Figure 11D**), genes/cell (**Figure 11E**), but also of mitochondrial genes (**Figure 11F**).

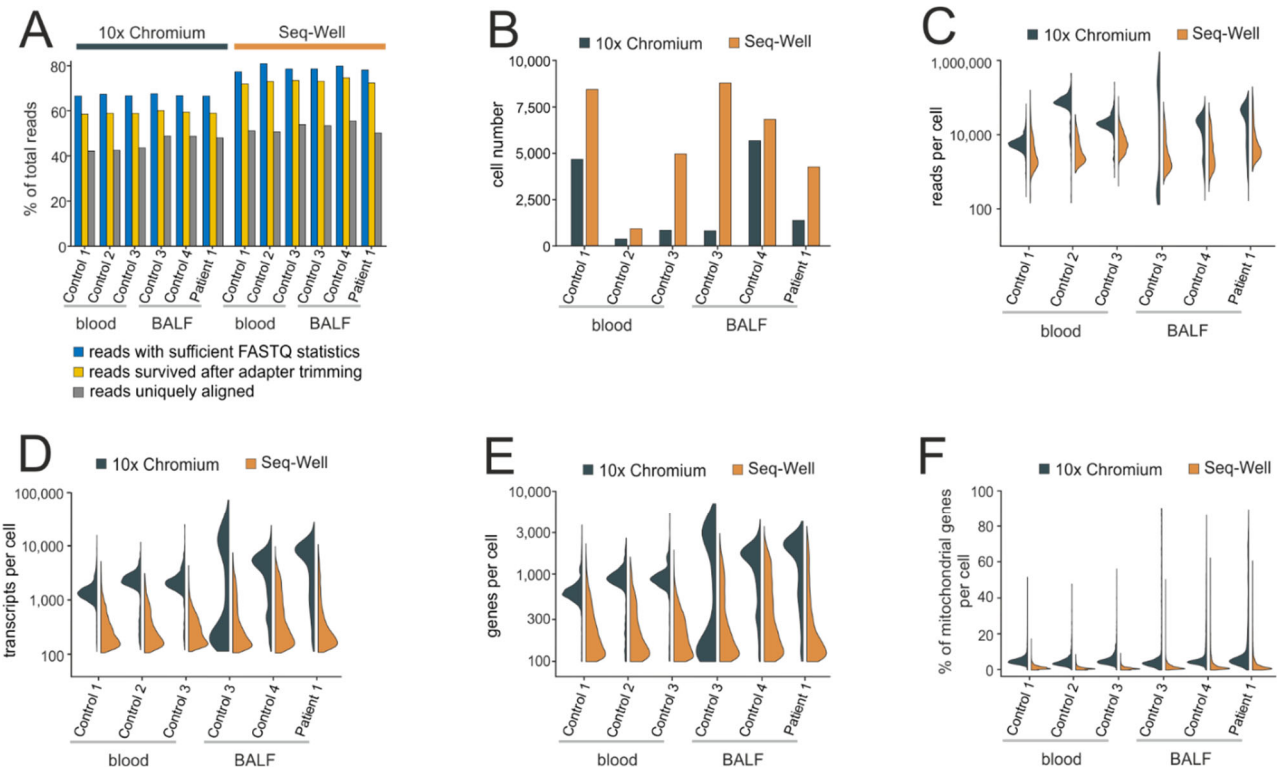


Figure 11. Comparison of Seq-Well and 10x Chromium (v2) in terms of detected information at cell level

(A) Alignment and quantification statistics. **(B)** Number of cells that exhibited at least 100 expressed genes. **(C-F)** scRNA-seq library statistics represented as split violin plots.

Next, we investigated which cell types can be detected by the respective single cell technologies. For this purpose, we integrated the scRNA-seq data of all samples using an anchor-based strategy (Stuart et al., 2019) and visualized them via UMAP (**Figure 12A**). The dimensional reduction revealed that some BALF cells were found in UMAP regions, which are mainly defined by immune cells derived from blood. However, the majority of immune cells in BALF did not show topological overlap with blood cells. The annotation of the individual cells using the SingleR method (Aran et al., 2019) and the Human Primary Cell Atlas (HPCA) (Mabbott et al., 2013) as reference dataset identified the cells found exclusively in BALF as macrophages with some contaminating $CD45^{+}$ respiratory epithelial cells (**Figure 12B**). Overall, the cell-type annotation showed that we could identify all major immune cell types defined by MCFC (**Figure 10E**) also by scRNA-seq (**Figure 12B**). To assess how the cell-type detection behaved across the used scRNA-seq technologies, we determined the cell-type distribution for the droplet- and well-based scRNA-seq methods independently. Remarkably, based on the cell-type annotation, it became evident that the entire granulocyte population was almost completely lost in the scRNA-seq data obtained using the Chromium technology (**Figure 12C**). This finding was corroborated by the UMAP topologies, where the granulocyte-defined region (**Figure 12B**) was almost completely unoccupied in the Chromium-derived scRNA-seq data (**Figure 12D**).
Since the observation of reduced granulocyte numbers in the Chromium technology is mainly based on the cell-type annotation by a single method, we wanted to assess the robustness of the findings by additional cell-type annotation approaches. Therefore, we used the Blueprint (Stunnenberg et al., 2016) and ENCODE (Dunham et al., 2012) datasets as reference for SingleR. As a complementary approach, we aimed at annotating cells using gene signatures of an immune cell dataset (LM22) that have been described to robustly differentiate immune cell types (Newman et al., 2015). To enable

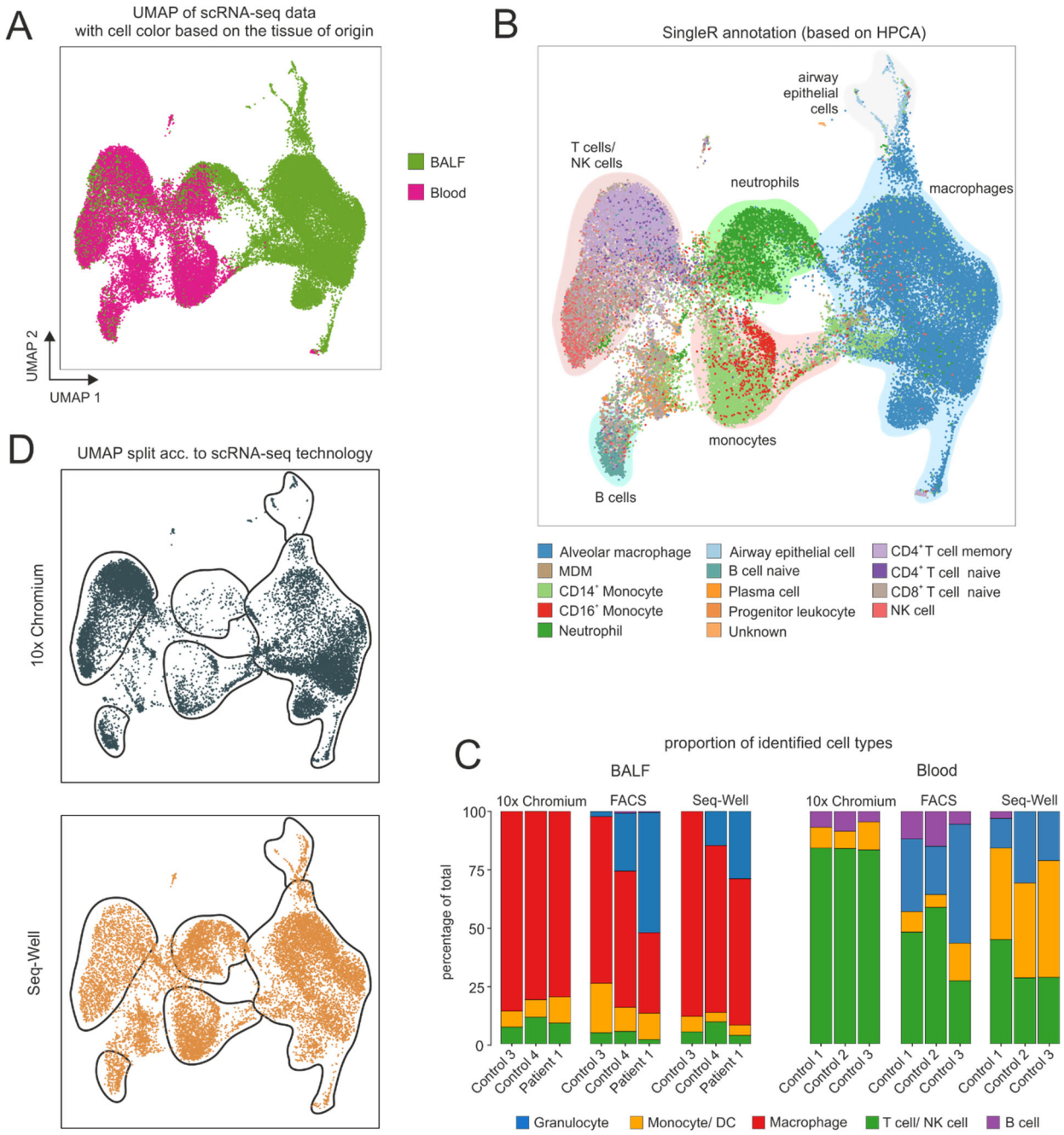


Figure 12. Comparison of Seq-Well and 10x Chromium (v2) on cell-type level

(A) UMAP representation of integrated blood and BALF data from different patients and the two scRNA-seq technologies. **(B)** Cell annotation based on the SingleR method with the HPCA dataset as reference (SingleR (HPCA)). Accumulation of cells with the same identity were highlighted with colored background and labeled accordingly. **(C)** Stacked bar plots of the relative cell-type proportions for MCFC, which served as ground truth, and cell-type proportions, as predicted by SingleR (HPCA), of the two scRNA-seq technologies. **(D)** UMAP representation of integrated data split by the two scRNA-seq technologies. The borders represent the areas of cell accumulation with the same identity as defined in (B).

annotation of single cells based on predefined gene signatures, we developed GenExPro, which uses a linear regression approach for cell annotation (**Material and Methods**). Importantly, the underlying reference datasets for the three cell-type annotation approaches (SingleR (HPCA), SingleR (Blueprint+ENCODE), and GenExPro (LM22)) collectively comprised a comprehensive collection of the immune cells present in lung tissue and blood (**Figure 13A**). However, not all immune cell types were included in all datasets, and in addition, some reference datasets were derived from microarrays, while others were derived from RNA-Seq experiments (**Figure 13A**). Nevertheless, the comparison of the results of cell-type annotation derived with the three approaches showed that there is a general, albeit not complete, concordance for the cell-types included in all reference datasets (**Figure 13B)**. While cell type annotation approaches showed slight differences, all indicated the almost complete absence of granulocytes in scRNA-seq data generated with the Chromium technology (**Figure 13C**). In addition, the cell-type annotation approaches also showed that the Seq-Well technology is prone to the reduction of B cell numbers and a shift in the blood population distribution, however, it closely resembled the MCFC-based ground truth of the most abundant cell types in BALF samples (**Figure 12C and 13C**).

Taken together, the comparison of the scRNA-seq technologies showed that the Chromium technology can generally provide more information per cell (indicated e.g. by the genes/cell statistics), but is inferior in the detection of one of the major cell types in BALF. For the latter reason, Seq-well is the preferred choice for the determination and characterization of immune cells in BALF and hence was used in the further COPD study.

4.3. Robust classification of immune cell types in the human alveolar space

4.3.1. Cell-type annotation using machine learning-based strategy

For in-depth investigation of COPD-associated cellular and molecular patterns, we generated a second, larger Seq-Well based scRNA-seq dataset from BALF samples

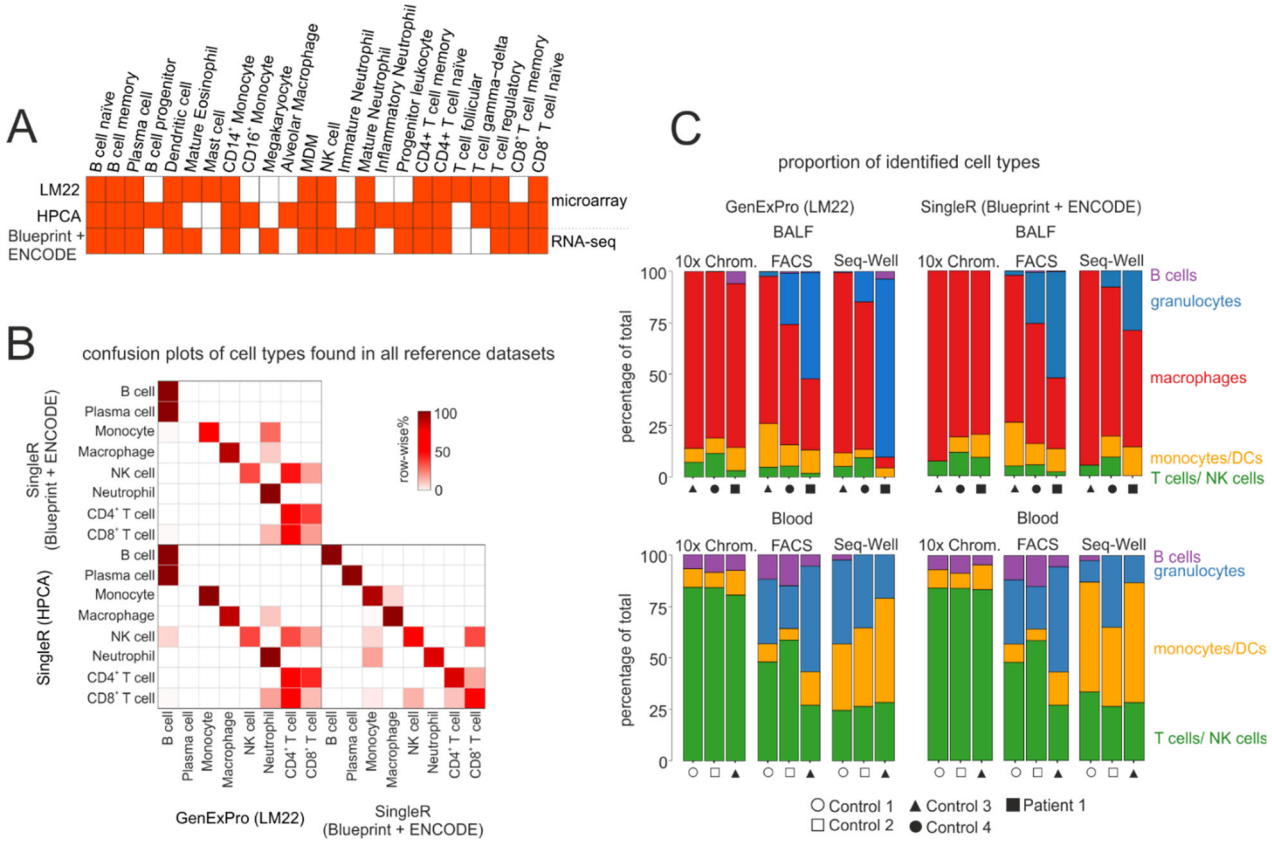


Figure 13. Confirmation of differences on cell-type level between Seq-Well and 10x Chromium (v2) using different references and cell type annotation approaches

(A) Overview of the cell types contained in the reference files used for cell-type annotation. The orange color indicates that the respective cell type is included in the reference file. **(B)** Confusion plots showing the concordance between the respective cell-type annotations across the different annotation methods. Only cell types that can be found in all reference files as shown in (A) are displayed. **(C)** Stacked bar plots of the relative cell-type proportions for MCFC, which served as ground truth, and cell type-proportions of the two scRNA-seq technologies, as predicted by SingleR with the Blueprint + ENCODE dataset as reference and the novel cell-type annotation method GenExPro.

derived from a cohort of nine patients with early-stage COPD (GOLD 2) and six control donors (**Table 2**). In the first step of the analysis, we aimed to robustly assign the cells to the respective cell types. However, as outlined above, the different cell-type annotation approaches tested and the underlying reference files showed some differences (**Figure 13A and B**), which made it difficult to decide on a particular cell annotation strategy. Therefore, instead of focusing on a single cell annotation approach and reference dataset, we aimed to combine the results of all three

annotation approaches (SingleR (HPCA), SingleR (Blueprint+ENCODE) and GenExPro (LM22)) to obtain the best possible cell type assignment. For this purpose, we developed a machine learning-based strategy to aggregate and consolidate the results of the three-cell type annotation approach and to assign the most likely cell type to each cell in the dataset (**Figure 14A**). To assess the validity of this cell type annotation based on machine learning, we used a benchmarking dataset generated by extracting cells with unequivocal expression of known cell type-specific markers from a scRNA-seq blood dataset (**Figure 14B**). Neither one of the SingleR annotations nor GenExPro was able to correctly annotate all cells within the benchmarking dataset (**Figure 14C**), since none of the underlying reference datasets included all major immune cell types present in the blood (**Figure 13A**). In contrast, the machine learning-based cell annotation was successful in aggregating the annotation results and thus resolving the different immune cell types (**Figure 14C**).

4.3.2. Development of a four-step strategy for robust cell-type annotation

Next, we applied the machine learning-based cell-type annotation approach to the integrated BALF dataset, which led to the identification of all major immune cell types (**Figure 15A**) that were also found by MCFC (**Figure 8+9**). Importantly, cells assigned to the same cell types accumulated in certain UMAP areas (**Figure 15A**). These areas were also defined by cell clustering as distinct clusters and overall the dataset could be partitioned into 18 main clusters. (**Figure 15B**). To further investigate the concordance between the machine learning-based cell-type annotation and cell clustering, we defined marker genes for each main cluster (**Figure 15C, heatmap**). In addition, we determined which cell type occurred most frequently per main cluster according to the machine learning-based annotation (**Figure 15D and Figure 15C, bottom annotation**). As exemplified for main cluster 13 (**Figure 15B**), we found a strong concordance between the calculated marker genes and the predicted most frequent cell type, since the cells

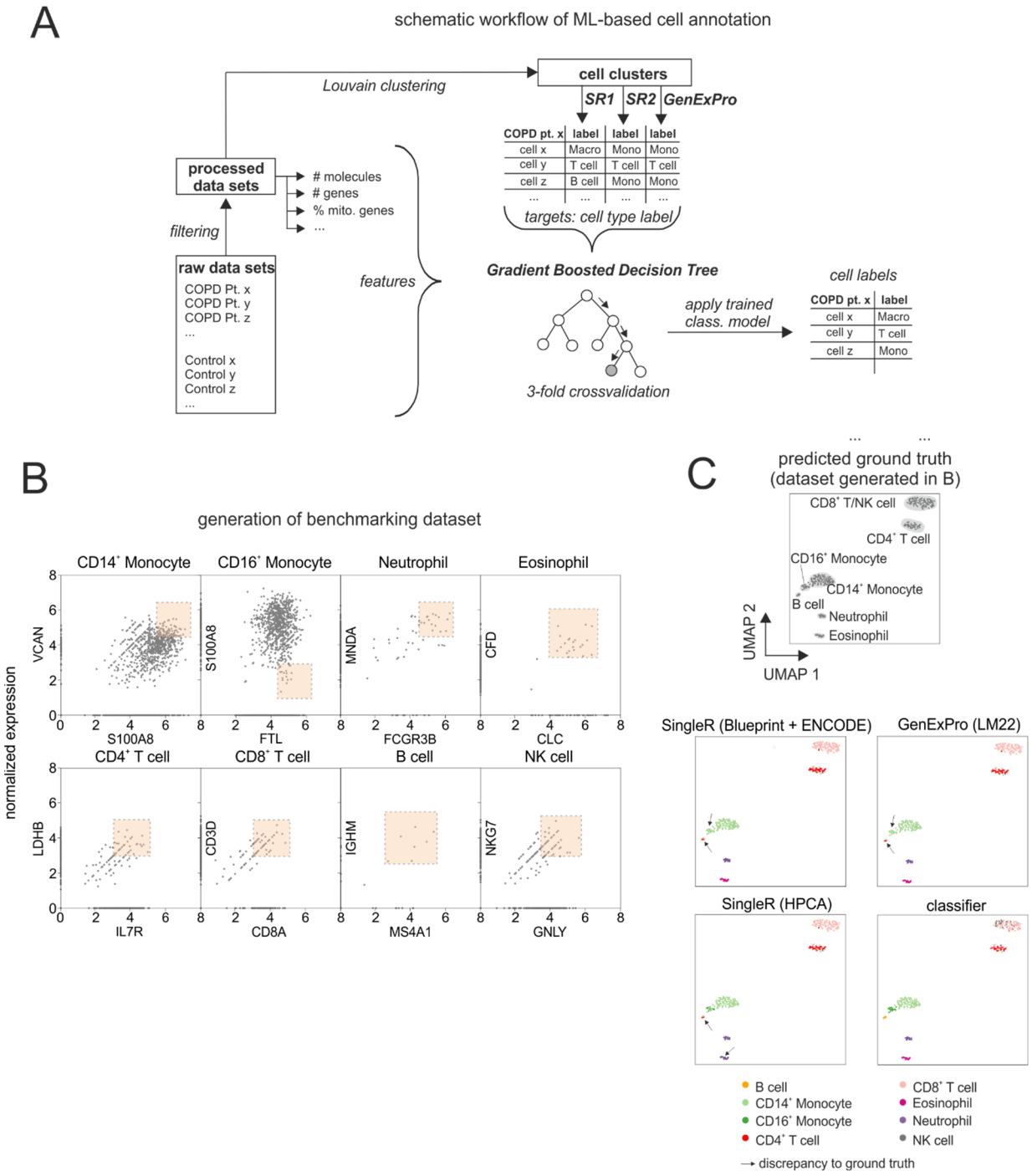


Figure 14. Generation and benchmarking of a machine learning-based cell-type annotation

(A) Scheme of a gradient boosted decision tree-based machine learning-approach for cell-type annotation. **(B)** Generation of a 'clean' Seq-Well dataset for benchmarking the machine learning-based cell-type annotation by selecting cells from a blood dataset (patient 6 acc. to **Table 2**) that expressed unique markers of a certain cell type. The selected cells, which are combined in a benchmarking dataset, are marked by the dashed rectangles with a reddish back-

ground. (C) UMAP representation of the benchmarking dataset (acc. to (B)) and coloring of the cells according to the cell annotation methods. The ground truth is derived based on the unique cell-type marker gene expression in (B). Accumulation of cells that are annotated by the respective annotation methods, but show a deviation in the annotation with respect to the ground truth, are marked with an arrow.

within this cluster expressed the neutrophil markers *IL8*, *FCGR3B* and *CXCR2* and most of the cells within the clusters were also annotated as neutrophils (**Figure 15C+D**). However, we also found some cells that were annotated e.g. as neutrophils by the machine learning-based approach, but scattered away from the other neutrophils and thus fell into non-neutrophil clusters (**Figure 15E**). Therefore, we further investigated the neutrophil-containing main cluster 13, where it became evident that although most cells were predicted by the classifier as neutrophils, a few cells were classified as eosinophils, T cells or macrophages (**Figure 15F**). Interestingly, the latter cells also formed distinct subclusters within cluster 13 (**Figure 15G**), in which they also expressed known markers for the predicted cell types (macrophage = *MARCO*, eosinophils = *CLC*, T cells = *CD3D*) (**Figure 15H**). However, they also expressed typical neutrophil markers (**Figure 15I**), suggesting these cells (hereinafter referred to as "mixed cells") could either be putative cell doublets, whereby corresponding tools did not identify these cells as doublets (*data not shown*), or could represent cells of low quality and were therefore excluded from further analysis. The remaining cells within main cluster 13 were analyzed in more detail by calculating marker genes of the three subclusters, which were referred to as ‘neutrophil clusters’ (**Figure 15K**). Intriguingly, cells within neutrophil cluster 2 expressed *CD63* (**Figure 15K**), which has been previously linked to airway neutrophils in cystic fibrosis (Tirouvanziam et al., 2008). Moreover, the same cluster showed increased expression of activation markers, such as *CD69*, *FCER1G* and *GBP1* (**Figure 15K**), which is consistent with the machine learning-based cell annotation, which assigned almost half of the cells within this cluster to inflammatory neutrophils (**Figure 15J**).

These results showed that the proposed cell annotation using the machine learning-based strategy can reliably identify the cell types in BALF. However, the power of this

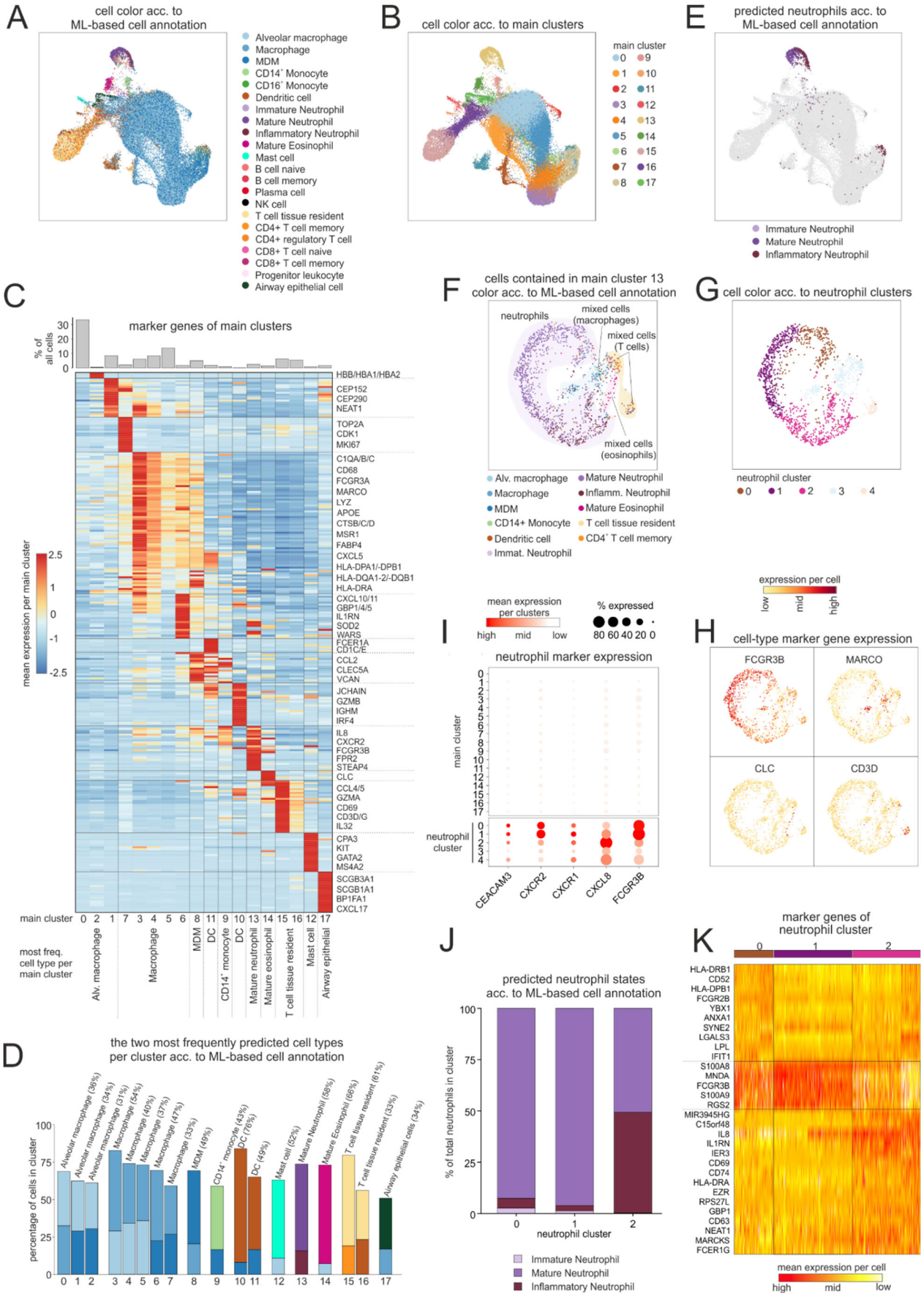

A
cell color acc. to ML-based cell annotation
Alveolar macrophage
Macrophage
MDM
CD14+ Monocyte
CD16+ Monocyte
Dendritic cell
Immature Neutrophil
Mature Neutrophil
Inflammatory Neutrophil
Mature Eosinophil
Mast cell
B cell naive
B cell memory
Plasma cell
NK cell
T cell tissue resident
CD4+ T cell memory
CD4+ regulatory T cell
CD8+ T cell naive
CD8+ T cell memory
Progenitor leukocyte
Airway epithelial cell
B
cell color acc. to main clusters
main cluster
E
predicted neutrophils acc. to ML-based cell annotation
Immature Neutrophil
Mature Neutrophil
Inflammatory Neutrophil
C
marker genes of main clusters
% of all cells
mean expression per main cluster
main cluster
most freq. cell type per main cluster
Alv. macrophage
Macrophage
MDM
DC
CD14+ monocyte
Mature neutrophil
Mature eosinophil
T cell tissue resident
Mast cell
Airway epithelial
F
cells contained in main cluster 13 color acc. to ML-based cell annotation
neutrophils
mixed cells (macrophages)
mixed cells (T cells)
mixed cells (eosinophils)
G
cell color acc. to neutrophil clusters
neutrophil cluster
D
the two most frequently predicted cell types per cluster acc. to ML-based cell annotation
percentage of cells in cluster
I
mean expression per clusters
% expressed
neutrophil marker expression
main cluster
neutrophil cluster
CEACAM3
CXCR2
CXCR1
CXCL8
FCGR3B
H
expression per cell
cell-type marker gene expression
FCGR3B
MARCO
CLC
CD3D
J
predicted neutrophil states acc. to ML-based cell annotation
% of total neutrophils in cluster
neutrophil cluster
K
marker genes of neutrophil cluster
mean expression per cell

Figure 15. Machine learning-based cell type annotation of BALF immune cells

(A) UMAP representation of integrated BALF data obtained from different COPD patients and control donors. Coloring according to the machine learning-based cell-type annotation. **(B)** UMAP representation of integrated BALF data with coloring according to identified main clusters. **(C)** Heat map of the calculated marker genes per main cluster with a bar chart representation of the relative cell-type proportions at the top. The marker gene expression per cluster is represented as a z-transformed value (across all clusters). Rows of the heat map are sorted by hierarchical clustering. At the bottom of the plot, the main cell type according to the machine learning-based cell-type annotation is displayed, which is contained in the respective main cluster (acc. to (D)). **(D)** Bar plot displaying the two most common cell types for each main cluster according to machine learning-based cell-type annotation, in color-coded form (acc. to (A)). In addition, the most frequently occurring cell type is written above the respective cluster together with its relative frequency. **(E)** UMAP representation of the integrated dataset with the coloring of different neutrophil states as predicted by the machine learning-based cell-type annotation. Non-neutrophils are colored light gray. **(F)** UMAP representation of cells contained in main cluster 13 (acc. to (B)). Coloring according to the machine learning-based cell-type annotation. Accumulation of cells with the same identity were highlighted with colored background and labeled accordingly. **(G)** UMAP representation and clustering of the cells contained in main cluster 13. **(H)** Feature plots showing the expression of cell type-specific markers (FCGR3B = neutrophils; MARCO = macrophages; CLC = eosinophils; CD3D = T cells). **(I)** Expression of known neutrophil-associated genes represented in a dot plot. Main clusters are shown, with the main cluster 13 subdivided into the neutrophil clusters 0-4 (acc. to (G)). **(J)** Bar chart showing the proportions of different neutrophil states in the neutrophil clusters 0-2 (acc. to (G)) according to the machine learning-based cell-type annotation. **(K)** Heat map of markers genes for neutrophil clusters 0-2, which were predicted to contain mainly neutrophils according to the machine learning-based cell-type annotation (acc. to. (F)). Rows of the heat map are sorted by hierarchical clustering.

method is even enhanced in combination with cell clustering, as this not only allows cells to be annotated, but also enables cells with inferior or ambiguous information (here: ‘mixed cells’) to be identified and excluded from further analysis. For this reason, we wanted to extend the described analysis steps to all cell types in BALF and therefore developed a four-step cell-annotation approach (**Figure 16**) comprising 1) the application of the newly generated machine learning-based classifier, 2) cell clustering, followed by 3) a manual classifier-to-cluster comparison and 4) cluster-level marker gene analysis including cleanup. The steps 1-3 were already carried out in this analysis (**Figure 15A-D**), so that only the fourth step was missing, which had to be performed separately for each cell type/cluster.

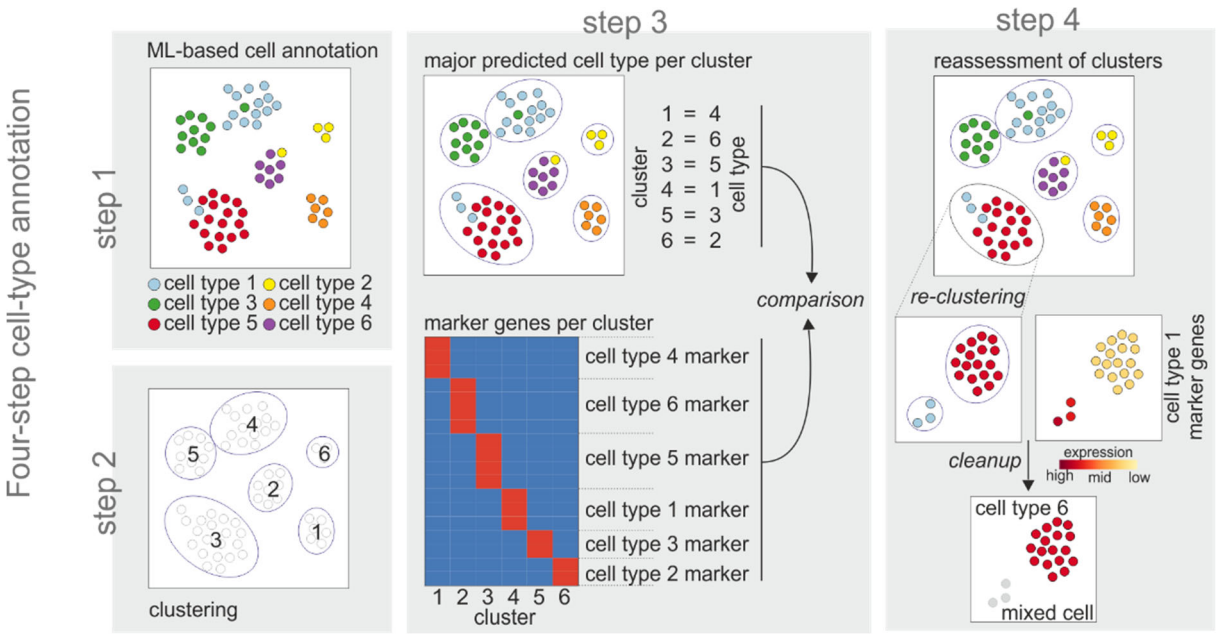


Figure 16. Strategy of the four-step cell annotation approach

Schematic workflow of the four-step annotation approach, including machine learning-based cell-type annotation, clustering, assignment and subsequent confirmation of a cluster to a cell type according to the machine learning-based cell-type annotation, and identification of 'contaminating' cells (referred to as 'mixed cells').

4.3.3. Application of the four-step cell-type annotation approach to the DC space

To further underpin the power and robustness of the four-step cell-type annotation approach, we applied the strategy to the main clusters 10 and 11 (**Figure 15B**) that were associated with DCs. Again, we found an aggregation of 'mixed cells' (DC-cluster 0 = macrophage-associated, DC cluster 1 = monocyte-/MDC-associated) (**Figure 17A**), which were also identified by cell clustering (**Figure 17B**). The remaining cells, which were annotated as DCs by the machine learning-based strategy, formed distinct clusters (called 'DC clusters') that expressed marker genes for different DC subtypes (*FCER1A* and *CD1C* = cDC2; *CADM1* and *CLEC9A* = cDC1; *JCHAIN* = pDC) (**Figure 17A-C**). Indeed, marker gene calculation revealed that the DC clusters represented cDC1, cDC2, and pDCs, respectively (**Figure 17D**). However, we also found one cluster (DC cluster 5) that showed expression of *IL7R* and *CCL22* (**Figure 17D**) but not *CD3D* (**Figure 17C**), which identified these cells most likely as rare ILCs

contaminating the DC population, further illustrating the power of our four-step cell-type annotation approach to even identify rare cell types.

4.3.4. Extending the four-step cell-type annotation approach to all cells in BALF

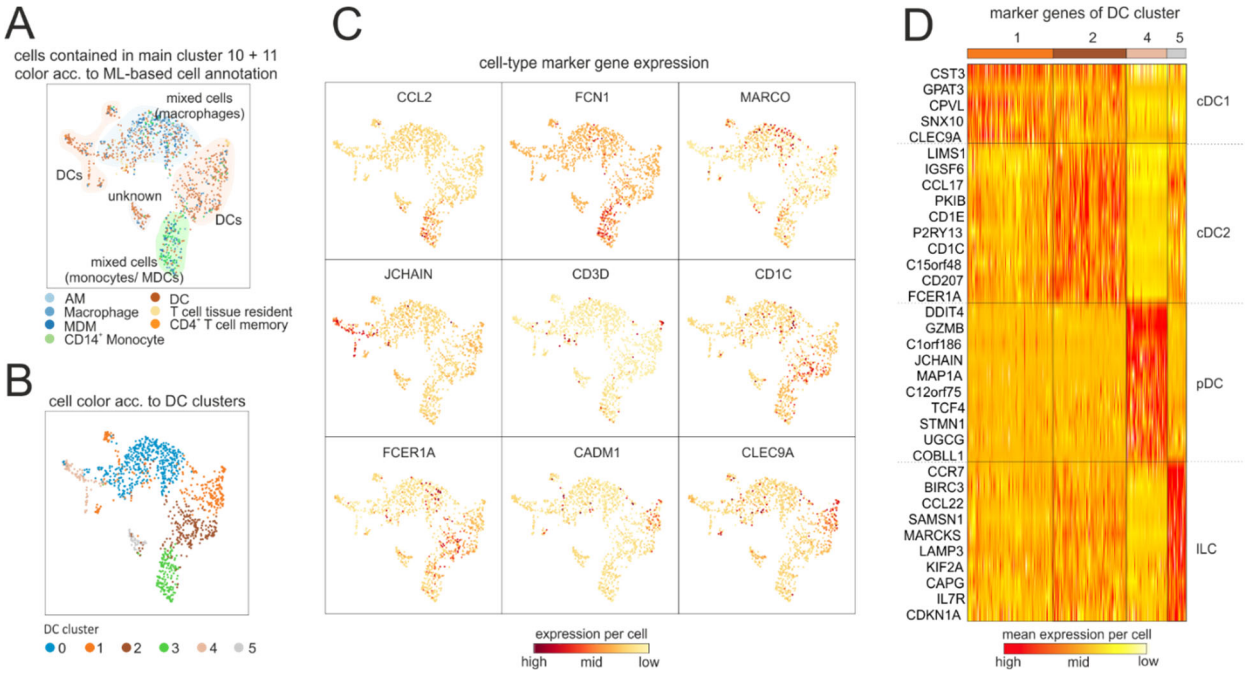


Figure 17. Application of the four-step cell annotation approach to the DC compartment in BALF

(A) UMAP representation of cells contained in main clusters 10 and 11 (acc. to **Figure 15B**). Coloring according to the machine learning-based cell-type annotation. Accumulation of cells with the same identity were highlighted with colored background and labeled accordingly. **(B)** UMAP representation and clustering of the cells contained in main clusters 10 and 11. **(C)** Feature plots showing the expression of cell type-specific markers (*CCL2* and *FCN1* = monocytes; *MARCO* = macrophages; *CD3D* = T cells; *FCER1A*, *JCHAIN*, *CD1C*, *CADM1* and *CLEC9A* = DCs and DC subtypes). **(D)** Heat map of markers genes for DC clusters 1, 2, 4, and 5 (acc. to (B)), which were predicted to contain mainly DCs according to the machine learning-based cell-type annotation (acc. to. (A)). Rows of the heat map are sorted by hierarchical clustering.

Finally, we applied the annotation approach to all 18 main clusters in the integrated BALF dataset, which led to a detailed resolution of the immune landscape in the alveolar space (**Figure 18**). In line with the MCFC analysis of the BALF cells (**Figure 8+9**), we found that AMs constituted the most abundant immune cell population and together with monocytes, DCs, mast cells, T cells, eosinophils and neutrophils, they form the major immune cell types in the alveolar space (**bar plot in Figure 15C, Figure 18, and Table 6**).

Collectively, our four-step cell-type annotation approach allowed us to robustly annotate the BALF cells in the scRNA-seq dataset, remove cells with ambiguous information content (about 9% were assigned as "mixed cells" and thus removed from further analyses) and led to the joint identification of cellular subpopulations and rare cell types in the 15 patients. Thus, the robust cell annotation presented here constituted an essential and integral step for further analyses with the aim to answer COPD-related questions.

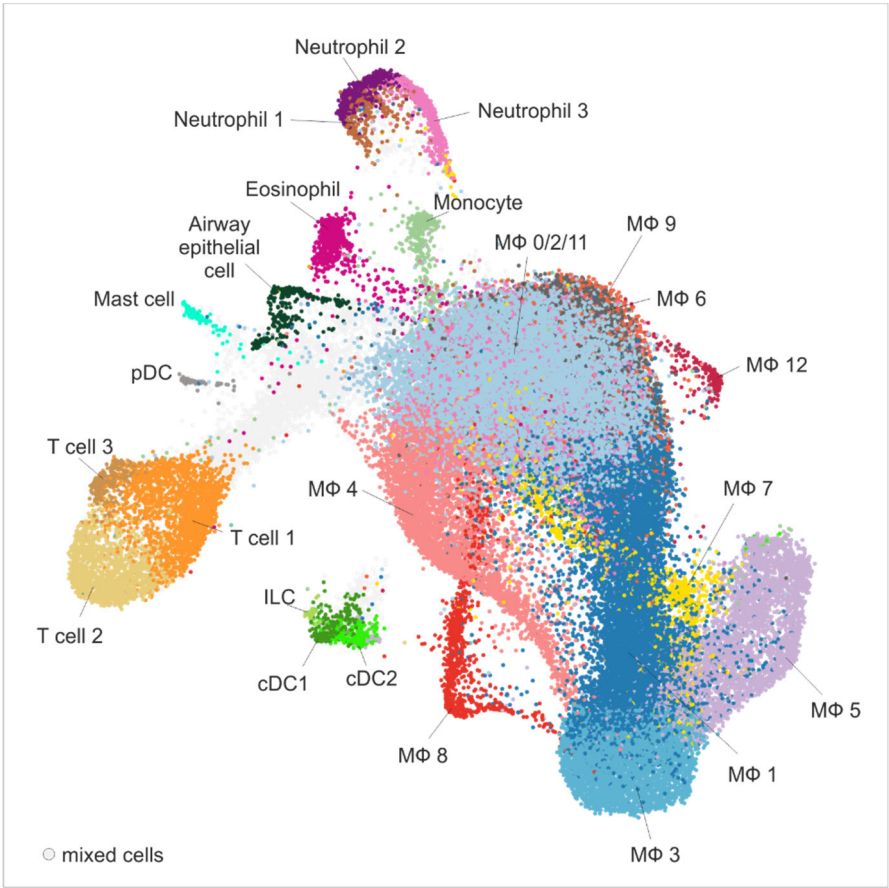


Figure 18. Identified BALF cell-types according to the four-step cell annotation approach

4.4. Modelling of metabolic alterations in AMs of COPD patients

4.4.1. COPD-mediated changes related to cellular functions

Both MCFC and scRNA-seq data analysis identified the AM compartment as being by far the largest cell population in the alveolar space, which is why we aimed to study these cells in more detail in the context of COPD. Therefore, we focused on the main cluster 0 – 9 (**Figure 15B**) and following the four-step cell-type annotation approach described above, we identified 13 subclusters (hereinafter referred to as 'AM clusters'), (**Figure 19A**) of which one was clearly a monocyte cluster (**Figure 19B**) (corresponding to main cluster 9 (**Figure 15B**)). Before we defined cluster-specific changes between COPD patients and control donors, we searched for changes associated with cellular functions that were carried by multiple AM clusters. For this purpose, we developed the so-called 'GO-shuffling' approach. The basic idea of this approach is to identify functional gene sets, such as those based on gene ontology (GO) or pathway annotations, which explain the strongest separation of COPD patients from controls in the Euclidean space (**Figure 19C**). To obtain a better overview about the GO-shuffling results of the AM clusters, we visualized the functional terms within the upper percentile of the functional gene sets with the highest potential to separate COPD patients from control donors in a wordcloud (**Figure 19D**). Interestingly, 'metabolic'-associated terms were strongest enriched in the wordcloud along with other terms such as 'protein', 'activation', 'morphogenesis', or 'chemotaxis', but also 'NOTCH' signaling (**Figure 19D**). A heat map representation of the genes contained in the NOTCH-associated gene sets confirmed that, for example, the expression profiles of metalloproteases of the ADAM family (*ADAM17* and *ADAM9*) and the components of the γ-secretase complex (*APH1B*, *APH1A*, *PSEN1*, *PSENEN*, and *NCSTN*) are able to separate COPD patients from control donors (**Figure 19E**).

4.4.2. Modelling of deregulated metabolism in AMs of COPD patients

Especially striking in the GO-shuffling analysis was that many of the terms, which most strongly separated COPD patients from controls, were associated with metabolism

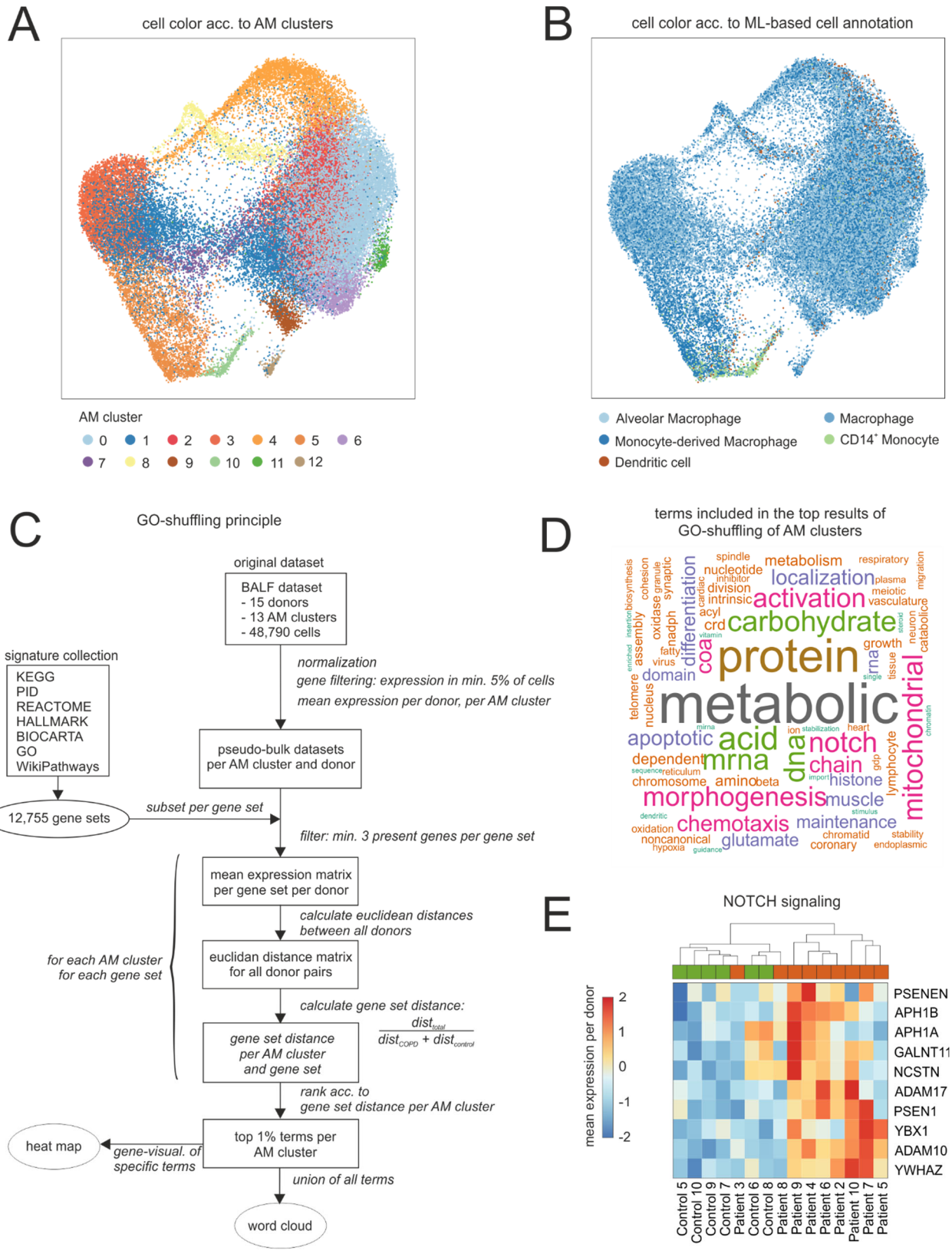

A
cell color acc. to AM clusters
AM cluster
0 1 2 3 4 5 6
7 8 9 10 11 12
B
cell color acc. to ML-based cell annotation
Alveolar Macrophage
Macrophage
Monocyte-derived Macrophage
CD14+ Monocyte
Dendritic cell
C
GO-shuffling principle
original dataset
BALF dataset
- 15 donors
- 13 AM clusters
- 48,790 cells
normalization
gene filtering: expression in min. 5% of cells
mean expression per donor, per AM cluster
signature collection
KEGG
PID
REACTOME
HALLMARK
BIOCARTA
GO
WikiPathways
pseudo-bulk datasets
per AM cluster and donor
12,755 gene sets
subset per gene set
filter: min. 3 present genes per gene set
mean expression matrix
per gene set per donor
calculate euclidean distances
between all donors
for each AM cluster
for each gene set
euclidan distance matrix
for all donor pairs
calculate gene set distance:
gene set distance
per AM cluster
and gene set
rank acc. to
gene set distance per AM cluster
heat map
gene-visual. of
specific terms
top 1% terms per
AM cluster
union of all terms
word cloud
D
terms included in the top results of
GO-shuffling of AM clusters
metabolic
protein
mitochondrial
carbohydrate
activation
localization
metabolism
differentiation
mrna
acid
notch
dna
chain
histone
morphogenesis
muscle
chemotaxis
maintenance
glutamate
apoptotic
dependent
chromosome
amino
oxidation
noncanonical
hypoxia
coronary
endoplasmic
E
NOTCH signaling
mean expression per donor
2
1
0
-1
-2
PSENEN
APH1B
APH1A
GALNT11
NCSTN
ADAM17
PSEN1
YBX1
ADAM10
YWHAZ
Control 5
Control 10
Control 9
Control 7
Patient 3
Control 6
Control 8
Patient 8
Patient 9
Patient 4
Patient 6
Patient 2
Patient 10
Patient 7
Patient 5

Figure 19. Identification of COPD-associated changes in AMs using GO-shuffling

A) UMAP representation and clustering of cells contained in main clusters 0-9, which are annotated as monocytes or macrophages (acc. to **Figure 15B**). **(B)** UMAP representation of cells contained in main clusters 0-9. Coloring according to the machine learning-based cell-type annotation. **(C)** Schematic workflow of the GO-shuffling approach. **(D)** Word cloud of the most common words in the top predicted terms of the GO-shuffling approach across all AM clusters. **(E)** Heat map of NOTCH-signaling associated genes predicted by the GO-shuffling approach. The mean gene expression per donor is represented as a z-transformed value (across all donors). Columns and rows of the heat map are sorted by hierarchical clustering.

(**Figure 19D**). To assess the possibility for deregulated metabolic pathways in COPD patients, we utilized the recently published Compass algorithm (Wagner et al., 2020; Wang et al., 2020), by which we comprehensively modelled the metabolic differences between COPD and control AMs by means of the transcriptome (**Figure 20**). When grouping the signaling pathways and metabolites found as DE into metabolic categories, it became apparent that most differences were found mainly in the amino acid- and lipid-associated metabolism (**Figure 20, pie chart**), with an overall higher predicted metabolic activity in the COPD patients (**Figure 20, heat map**). Especially the predicted deregulation in lipid metabolism is of great interest considering the essential function of AMs in surfactant homeostasis. Among the differential lipid-associated metabolites and reactions, phosphorylation of inositol was most prominent, but we also found differential metabolites and reactions indicating increased transport (monoacylglycerol), synthesis (phospholipids and cholesterol) and degradation (β-oxidation) of lipids in COPD AMs.

4.4.3. Altered lipid metabolism in AMs of COPD patients

To evaluate the reliability of the Compass based prediction, we extracted the part of the phospholipid metabolism from the KEGG database (**Figure 21A, pathway scheme**) which, according to the Compass results, had a higher predicted metabolic activity in COPD than in the controls and therefore the expression of the underlying

enzymes should be upregulated (**Figure 20**). Interestingly, we did not detect increased mean expression,

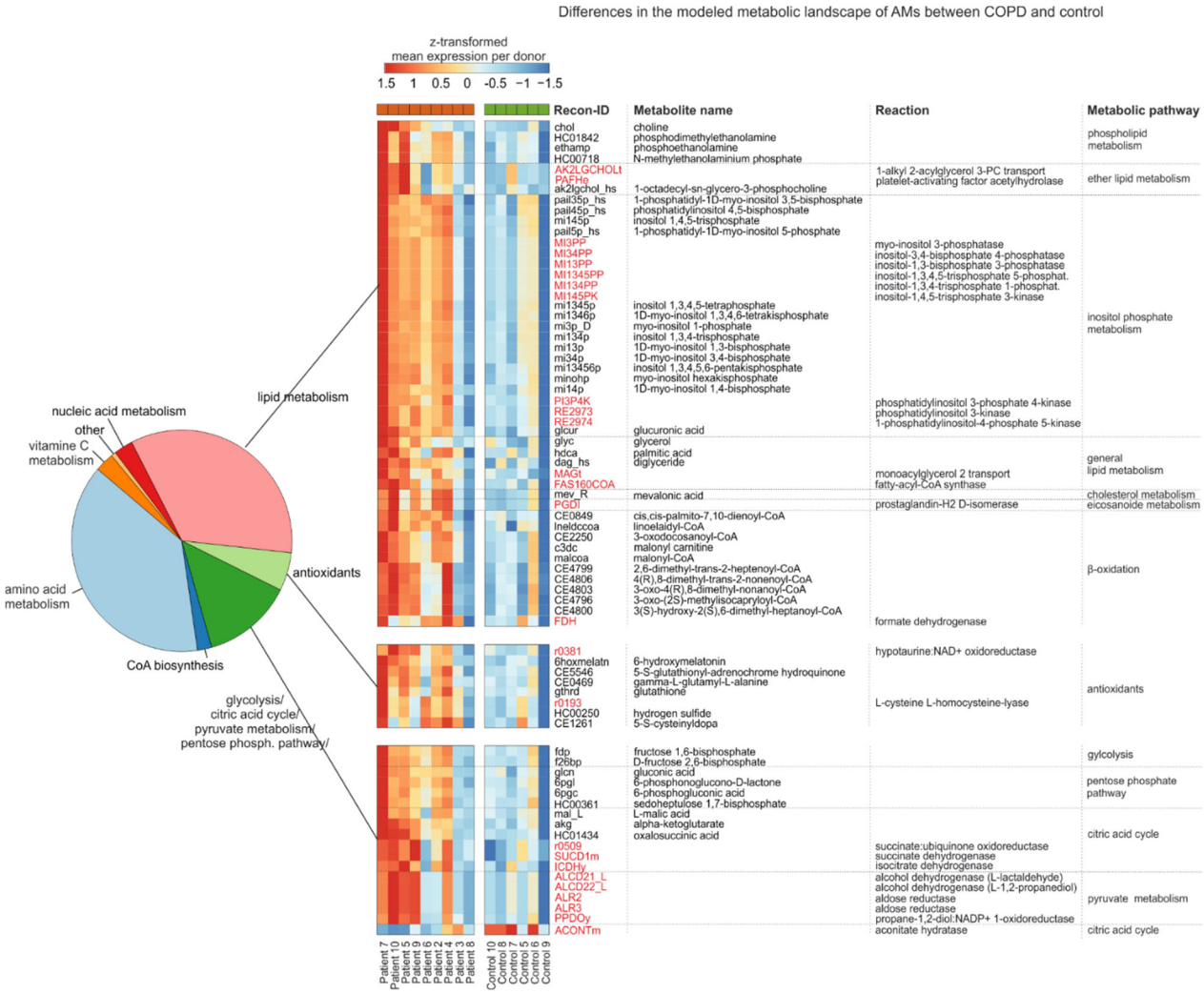


Figure 20. Modelling of metabolic changes in AMs of COPD patients using Compass

Compass results of the modelled metabolic landscape in AMs. The Venn diagram summarizes and categorizes the predicted metabolites and pathways that are significantly different between COPD and control. Heat map shows the predicted pathways and metabolites associated with lipid metabolism, antioxidants and energy metabolism. Recon2-ID (Thiele et al., 2013) of metabolites is shown in black and reactions in red. Columns and rows of the heat map are sorted by hierarchical clustering.

but remarkably, we found that more AMs in COPD express the genes encoding the enzymes involved in the predicted part of the phospholipid pathway (**Figure 21A, violin plots**), confirming the higher mean metabolic activity in COPD patients as indicated by Compass. Moreover, we found additional evidence of COPD-mediated alterations in lipid metabolism when visualizing the expression of genes found in lipid-associated gene sets contained in the upper percentile of the functional gene sets with the highest separation potential according to GO-shuffling (**Figure 21B**). Intriguingly, among these genes we found several that encoded for receptors of cholesterol uptake (*CD36*, *LDLR*, *MSR1*, and *TREM2*) and for proteins involved in cholesterol ester synthesis (*ACAT1/2* and *SOAT1*) which mediates cholesterol storage, but also genes that encoded cholesterol ester hydrolases (*LIPA* and *NCEH1*) (**Figure 21B**). These findings are in agreement with the Compass-predicted upregulation of mevalonate in COPD patients (**Figure 20**), which is one of the key metabolites in the synthesis of isoprenoids, to which cholesterol belongs. To gain a better understanding of whether and how the expression of genes encoding essential enzymes of the cholesterol metabolism in the identified AM cluster (**Figure 21C**) differed between COPD and controls, we visualized their respective median expressions (**Figure 21D**). While the smaller clusters (cluster 9 – 12, **Table 6**) showed hardly any differences, we detected a higher expression of cholesterol-associated genes in cells of COPD patients in most of the larger clusters (cluster 0 – 8). However, we also found a few exceptions, such as the cholesterol acyltransferase *ACAT2*, whose gene expression was downregulated in clusters 5, 6, and 8 in COPD, but showed either no or an upregulation in the other clusters (**Figure 21D**).

Next, we assessed the validity of the *in silico* predicted alterations in lipid metabolism of COPD patients by isolating AMs from COPD GOLD 2 patients and control donors and using them as input for an in-depth lipidomics analysis, which comprised the detection of 229 lipid species belonging to 16 different lipid classes. We observed a clear trend towards higher levels of the monoacylglycerol class in COPD patients (**Figure 21E**), but the biggest difference in lipid classes between COPD and control samples was found in cholesteryl ester, which was significantly higher in COPD AMs than in controls (**Figure 21F**). The detected accumulation of cholesterol within AMs is reminiscent of the pulmonary foam cell-like phenotypes of AMs reported for other lung diseases, such as

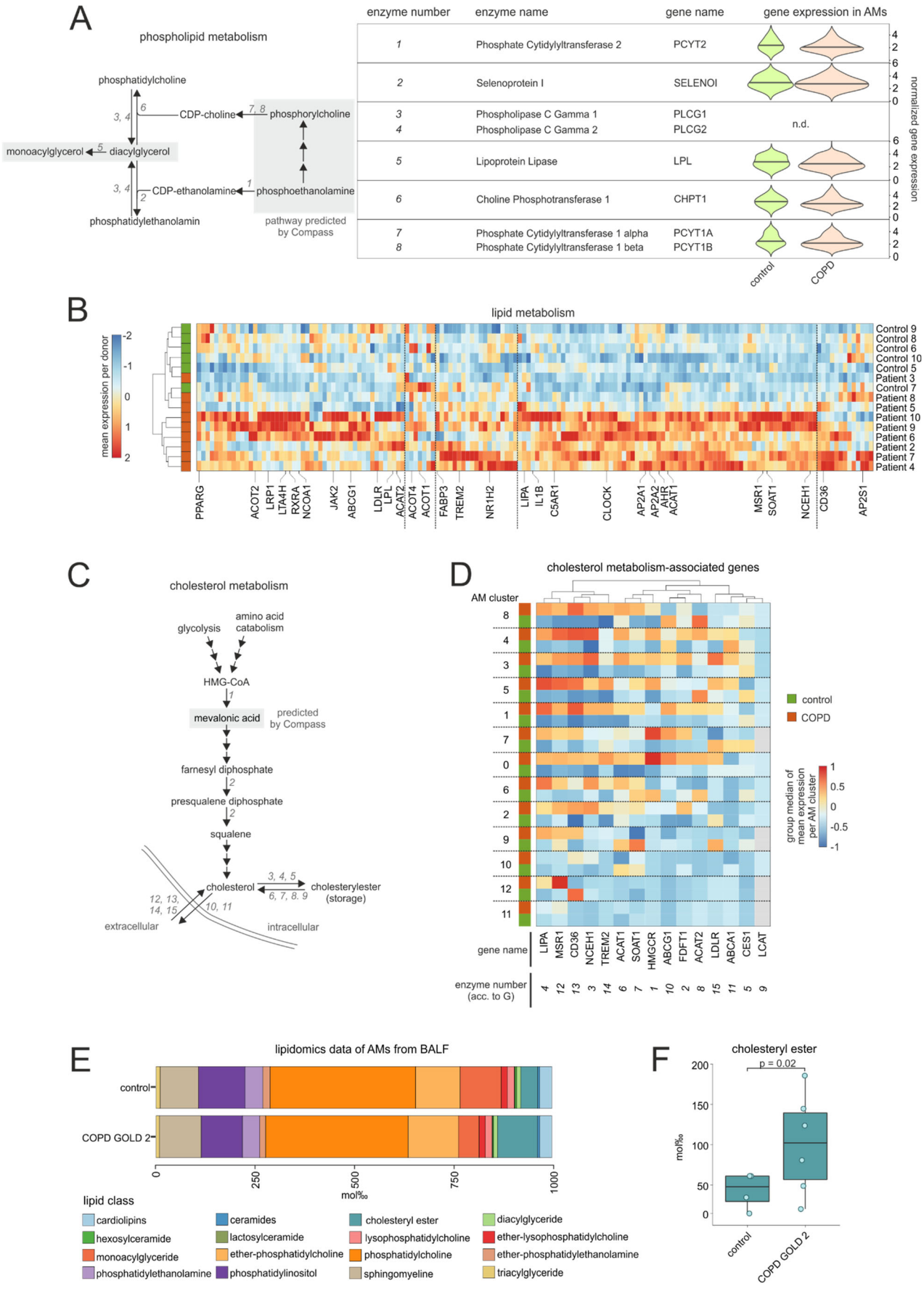

A
phospholipid metabolism
phosphatidylcholine
CDP-choline
phosphorylcholine
monoacylglycerol
diacylglycerol
CDP-ethanolamine
phosphoethanolamine
phosphatidylethanolamin
pathway predicted by Compass
enzyme number
enzyme name
gene name
gene expression in AMs
1 Phosphate Cytidylyltransferase 2 PCYT2
2 Selenoprotein I SELENOI
3 Phospholipase C Gamma 1 PLCG1
4 Phospholipase C Gamma 2 PLCG2
n.d.
5 Lipoprotein Lipase LPL
6 Choline Phosphotransferase 1 CHPT1
7 Phosphate Cytidylyltransferase 1 alpha PCYT1A
8 Phosphate Cytidylyltransferase 1 beta PCYT1B
normalized gene expression
control
COPD
B
lipid metabolism
mean expression per donor
Control 9
Control 8
Control 6
Control 10
Control 5
Patient 3
Control 7
Patient 8
Patient 5
Patient 10
Patient 9
Patient 6
Patient 2
Patient 7
Patient 4
PPARG
ACOT2
LRP1
LTA4H
RXRA
NCOA1
JAK2
ABCG1
LDLR
LPL
ACAT2
ACOT4
ACOT1
FABP3
TREM2
NR1H2
LIPA
IL1B
C5AR1
CLOCK
AP2A1
AP2A2
AHR
ARNT
MSR1
SOAT1
NCEH1
CD36
AP2S1
C
cholesterol metabolism
glycolysis
amino acid catabolism
HMG-CoA
mevalonic acid
predicted by Compass
farnesyl diphosphate
presqualene diphosphate
squalene
cholesterol
cholesterylester (storage)
extracellular
intracellular
D
cholesterol metabolism-associated genes
AM cluster
control
COPD
group median of mean expression per AM cluster
gene name
enzyme number (acc. to G)
LIPA MSR1 CD36 NCEH1 TREM2 ACAT1 SOAT1 HMGCR ABCG1 FDFT1 ACAT2 LDLR ABCA1 CES1 LCAT
E
lipidomics data of AMs from BALF
control
COPD GOLD 2
mol‰
lipid class
cardiolipins
ceramides
cholesteryl ester
diacylglyceride
hexosylceramide
lactosylceramide
lysophosphatidylcholine
ether-lysophosphatidylcholine
monoacylglyceride
ether-phosphatidylcholine
phosphatidylcholine
ether-phosphatidylethanolamine
phosphatidylethanolamine
phosphatidylinositol
sphingomyeline
triacylglyceride
F
cholesteryl ester
p = 0.02
mol‰
control
COPD GOLD 2

Figure 21. Characterization of altered lipid metabolism in AMs of COPD patients

(A) Scheme of the part of the phospholipid metabolism containing signaling pathways and metabolites for which Compass predicted a difference between COPD and control**.** Pathways and metabolites predicted by Compass are highlighted with a gray background. Enzymes involved in metabolism are abbreviated with a number that identifies them in the table on the left. Violin plots in the table are displaying the gene expression of the respective enzymes. The plots show the expression across the donors, whereby the donors were downsampled to the same number of cells, followed by downsampling to the same number of cells between COPD and control. The plots display cells with an expression > 0. The areas of the violin plot are scaled proportionally to the number of observations. **(B)** Heat map of lipid metabolism-associated genes predicted by the GO-shuffling approach. The mean gene expression per donor is represented as a z-transformed value (across all donors). Columns and rows of the heat map are sorted by hierarchical clustering. Names of selected genes are depicted at the bottom of the plot. **(C)** Schema of the key steps in cholesterol metabolism and storage. Metabolites predicted by Compass are highlighted with a gray background. Enzymes involved in metabolism are abbreviated with a number that identifies them also in (D). **(D)** Heat map representation of cholesterol metabolism-associated genes (acc. to (C)) across the identified AM clusters (acc. to **Figure 19A**). Depicted is the group median (group = COPD or control) of the z-transformed mean expression data per donor and AM cluster across all AM clusters. Columns and rows of the heat map are sorted by hierarchical clustering. **(E)** Stacked bar plot displaying the mean proportions (represented in mol permille) of lipid classes obtained by lipidomics analysis (control n = 4, COPD n = 6). **(F)** Box plot of cholesteryl ester proportions (acc. to (E)) with the representation of individual donors.

lipoid pneumonia (Collins et al., 1995) or vaping-related lung injury (Maddock et al., 2019). These phenotypes are characterized by the cells being predominantly cholesterol-laden. In pulmonary alveolar proteinosis (PAP), increased cholesterol/lipid accumulation within AMs can be mediated by defective GM-CSF signaling and, as a consequence, reduced PPARG and cholesterol transporter (ABCG1) expression (De Aguiar Vallim et al., 2017; Sallese et al., 2017; Trapnell et al., 2019). However, this mechanism is unlikely in COPD AMs since we did not observe clear downregulation of either *PPARG* (**Figure 21B**) or *ABCG1* (**Figure 21D**).

Taken together, we found metabolic changes in the AMs of COPD, which is partially characterized by an accumulation of cholesteryl ester in the diseased cells.

4.5. Identification of AM cluster-level differences using a novel DE analysis approach

4.5.1. Characterization of identified AM clusters

To gain a better understanding about the functionalities of the identified AM clusters (**Figure 19A and 22A**), we calculated the respective marker genes (**Figure 22B**). Collectively, all AM clusters clearly belonged to the macrophage cell lineage as defined by the expression of typical signature genes (*MSR1*, *MRC1, MARCO*) with the exception of cluster 10, which, as mentioned above, represented monocytes (**Figure 19B**). In addition to the macrophage signature, cluster 8 was also characterized by proliferation-associated genes (*MKI67*, *TOP2A*, and *NUSAP1*) as well as by increased expression of histones (*HIST1H4C* and *HIST1H1D*) (**Figure 22B**). Furthermore, the majority of the cells within cluster 8 could be assigned to the G2/M cell cycle phase (**Figure 22C**), strongly supporting that this cluster represented proliferating AMs. Clusters 9 and 6 were highly enriched for the expression of MHC class II molecules, namely *HLA-DQ* and *HLA-DR* respectively, while cluster 12 carried hemoglobin genes (*HBA2*, *HBA1*, and *HBB*) either due to engulfed erythrocytes, transcriptional mixture of erythrocytes and AMs or induction of hemoglobin genes in macrophages (Liu et al., 1999). We also examined whether the AM clusters were formed uniformly by all donors or whether a cluster was defined by the overrepresentation of an individual donor (**Figure 22D**). Exclusively for cluster 12 (MΦ/erythrocyte), 2 and 11, we found a donor effect, with the latter being characterized by the expression of the T cell-associated genes *CD2* and *CCL5*, which led us to label this cluster as ‘ILC-like’ macrophages. Interestingly, cluster 5 exhibited relatively strong expression of the monocyte-associated genes VCAN and S100A8 together with the monocyte attractant CCL2 and the late monocyte-to-macrophage differentiation marker CHIT1 and was therefore defined as ‘mono-like’ macrophages. This cluster also shared some markers with cluster 7, which was additionally high in interferon-response genes (*IFIT1* and *IFIT2*), and cluster 3 that showed increased expression of complement components (*C1QA-C*) and α1-antitrypsin (*SERPINA1*). The remaining cluster 4 contained a relatively large number of specific marker genes, but it was not possible to

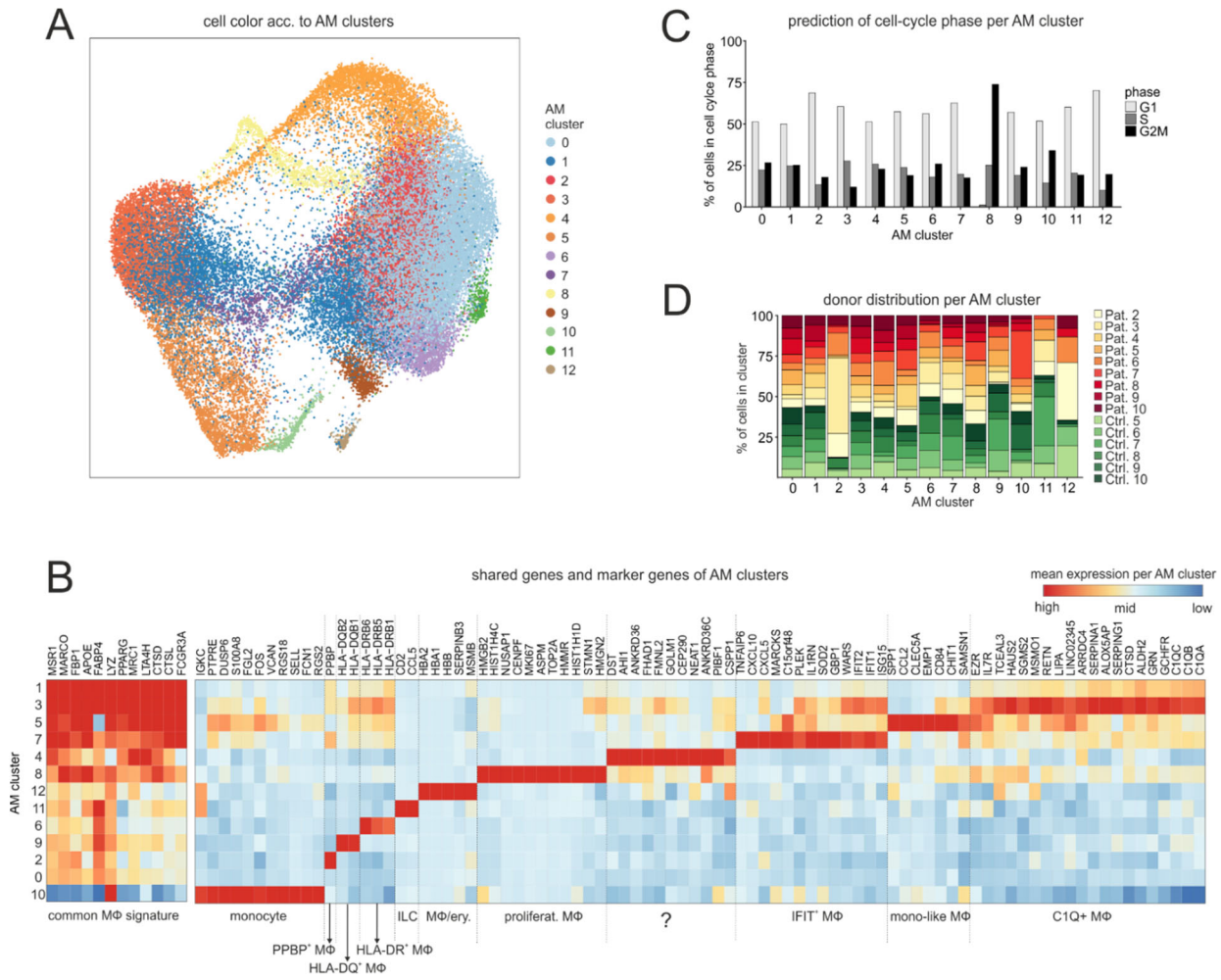


Figure 22. Characterization of identified AM clusters

A) UMAP representation and clustering of cells contained in main clusters 0-9, which are annotated as monocytes or macrophages (acc. to **Figure 15B**). **(B)** Heat map of calculated marker genes per AM cluster (acc. to (A)). The marker gene expression is represented as a z-transformed value across all AM clusters. On the left side of the heat map, conserved macrophage markers are depicted. Columns and rows of the heat map are sorted by hierarchical clustering. Suggested nomenclature for each AM cluster is depicted at the bottom of the plot. **(C)** Bar plot representation of the proportion of cells in the respective cell cycle states per AM cluster. **(D)** Stacked bar plot showing the proportion of individual donors in each AM cluster.

draw conclusions about the functionality of the cluster based on these genes (marked by '?' in **Figure 22B**).

4.5.2. Prediction of the functionalities of AM clusters

To shed further light on the potential functions of the identified AM clusters, we performed a gene set variation analysis (GSVA) (Hänzelmann et al., 2013) based on pseudo bulk samples per cluster and donor generated from scRNA-seq data (**Figure 23A**) and visualized the identified terms in an UpSet plot (Conway et al., 2017) (**Figure 23B**). This type of visualization provided the advantage of determining which cellular processes and functions were shared by the clusters and which were cluster-specific. Among the shared terms, we found enrichment of 'antigen presentation', 'endocytosis', 'oxidative phosphorylation' and 'β-oxidation', which represented some of the major cellular functions of macrophages in the alveolar space. Intriguingly, the cells of cluster 4, whose potential functionalities were difficult to identify on the basis of marker gene inspection (**Figure 23B**), showed a specific enrichment of the mTOR pathway, which was recently associated with the induction of cellular senescence in non-immune cells of the lung (Barnes, 2017; Houssaini et al., 2018). To determine whether cluster 4 cells are in a cellular senescent state, we performed an enrichment analysis on AMs using senescence-associated gene sets (**Figure 23C**). Indeed, we found an enrichment of cell ageing and mitochondrial genes in cluster 4, but without an enrichment of apoptosis and WNT signaling genes, which made a potential senescent state of the cells likely. Moreover, cluster 4 showed also a specific downregulation of genes described as downregulated in aged immune cells (IMM-AGE signature (Alpert et al., 2019)), which supported the characterization of cluster 4 as senescent AMs.
Taken together, the combination of marker gene inspection and GSVA enabled the characterization of the AM cluster, with the more functional-/marker-based labeling of the clusters being hereinafter referred to as 'AM states' (**Figure 23D**).

4.5.3. Novel DE analysis approach revealed major changes in most AM clusters of COPD patients

As the next layer of analysis to determine differences caused by COPD, we examined each identified AM cluster for statistically significant differences in the transcriptome

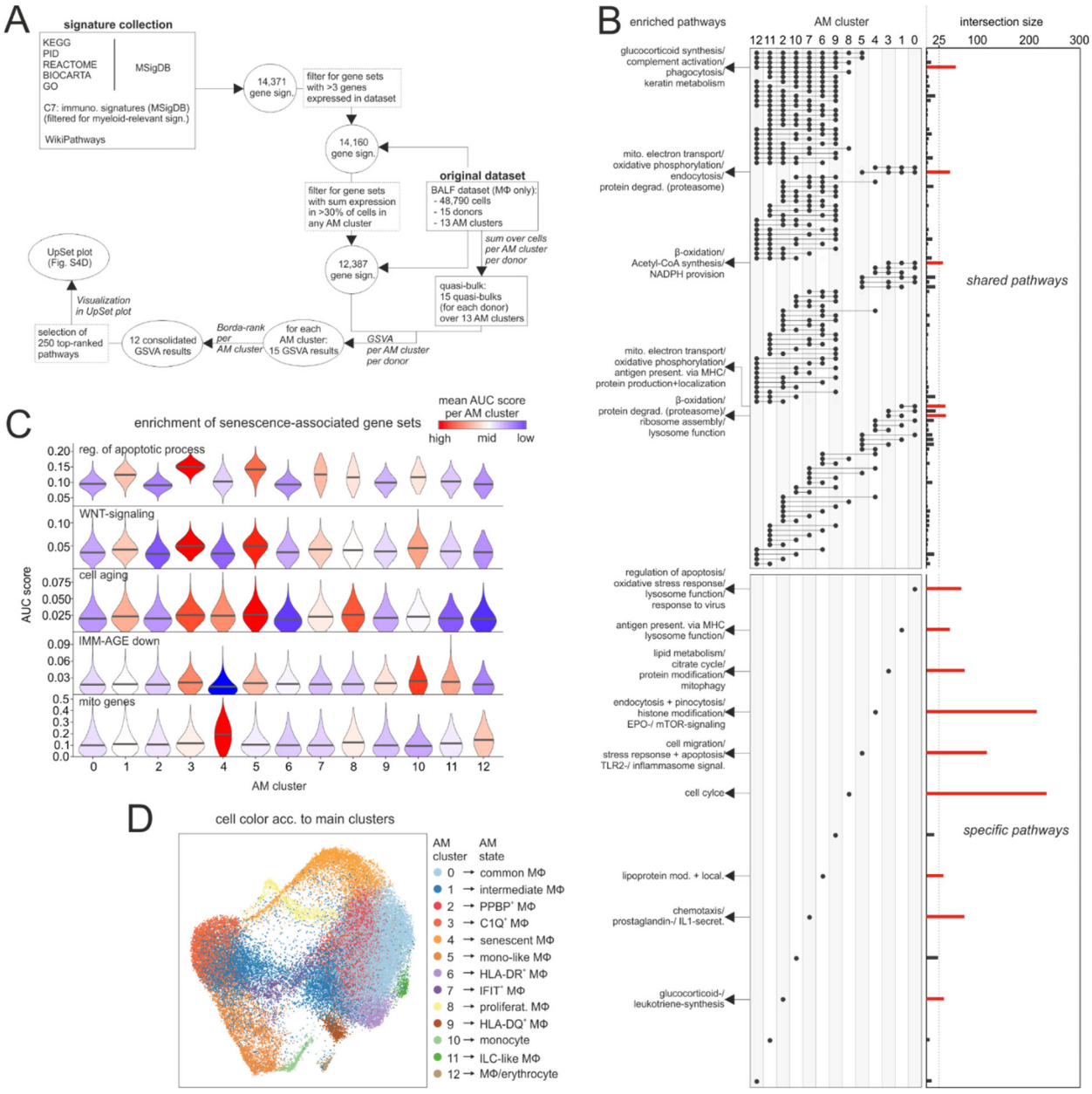


Figure 23. Association of AM clusters with functional terms

(A) Schematic workflow to predict the cellular functions of each AM cluster based on GSVA. **(B)** UpSet plot of the GSVA results (acc. to (A)). Terms of cellular functions found in the same clusters are grouped into bins and the size of the bins is represented as a bar plot on the right, with bins containing more than 25 terms (dashed line) colored red. On the left side, dots indicate which clusters contain and share the binned terms. Frequently occurring terms of cellular functions within the bins containing more than 25 terms are shown. **(C)** Violin plots displaying enrichment of different gene sets across clusters based on 'Area Under the Curve' (AUC). **(D)** UMAP representation of integrated macrophages and monocytes with coloring according to identified clusters (acc. to **Figure 22A**). The clusters are labeled based on information from marker genes (acc. to **Figure 22B**) and functional association (acc. to (B+C)).

between COPD and controls. The structure and characteristics of the dataset, on which the present COPD study is based, represent a form that is currently being increasingly generated in the scRNA-seq field, especially in studies to answer clinically relevant questions, namely a dataset with replicated multiple conditions. To meet the needs of the analysis of such scRNA seq data, we developed a DE-analysis approach based on the Wilcoxon rank sum test between COPD and control cells in combination with a robustness measure using permutation tests, which also considers possible individual donor effects (**Figure 24A**). The application of the novel DE analysis approach and the visualization of genes, which showed a statistically significant difference, in an UpSet plot indicated that the majority of DE genes were cluster-specific (**Figure 24B**). However, since the selection of DE genes is based on strong statistical differences and excludes transcriptional tendencies, we next investigated whether we could confirm cluster-specific differences in a heatmap plot showing the relative difference between controls and COPD per AM cluster (**Figure 24C**). This kind of data representation revealed that the putative cluster-specific DE genes also exhibited differences with similar directions between COPD and controls in other AM clusters. In addition, by focusing on DE genes found to be statistically significant in only a single cluster (**Figure 24B**) and plotting the corresponding calculated Wilcoxon scores for all AM clusters, it became evident that the direction of the scores was the same in all clusters, mainly towards increased expression in COPD (**Figure 24D**). Therefore, although the DE analysis revealed in which AM clusters the greatest difference between COPD and control for a given gene was found, this difference was not necessarily cluster-specific because the same direction of the difference could also been observed in several other clusters. To evaluate the validity of our novel DE analysis approach, we selected two genes (*ITIH5*, *COLEC12*) characterized by a particularly high Wilcoxon score in a single AM cluster (**Figure 24D**) and visualized their mean expression across all clusters (**Figure 24E**). The visualization confirmed that the difference between COPD and control for the selected genes was most pronounced in the AM clusters, which also exhibited the highest Wilcoxon score, thus confirming the reliability of our DE analysis approach.

After confirmation of the validity of the DE analysis approach and identification of conserved directions in the transcriptional differences between COPD and controls, we

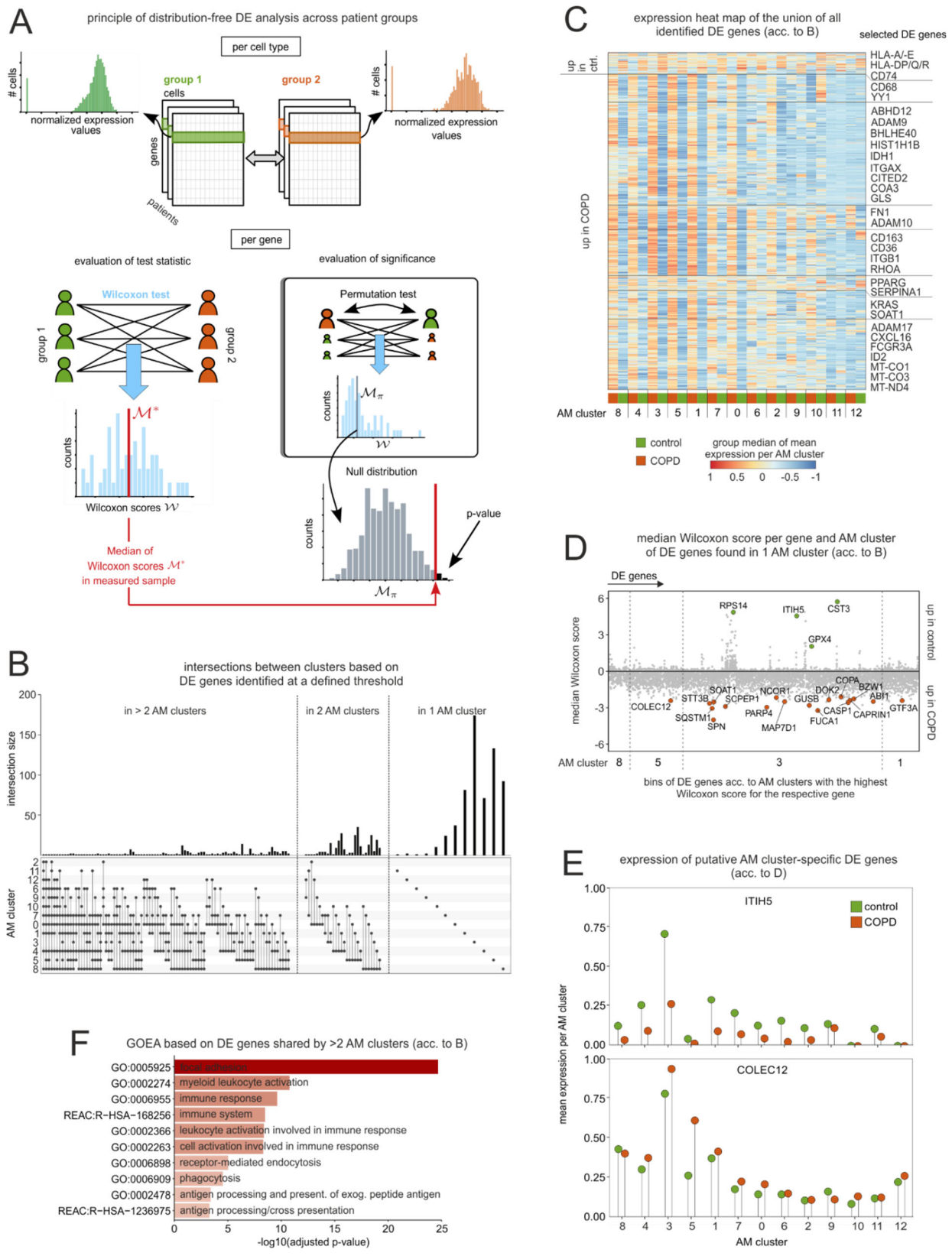

A
principle of distribution-free DE analysis across patient groups
per cell type
group 1
group 2
cells
cells
normalized expression values
genes
patients
per gene
evaluation of test statistic
evaluation of significance
Wilcoxon test
Permutation test
group 1
group 2
counts
Wilcoxon scores $\mathcal{W}$
Median of Wilcoxon scores $\mathcal{M}^*$ in measured sample
Null distribution
p-value
B
intersections between clusters based on DE genes identified at a defined threshold
in > 2 AM clusters
in 2 AM clusters
in 1 AM cluster
intersection size
AM cluster
C
expression heat map of the union of all identified DE genes (acc. to B)
selected DE genes
up in ctrl.
up in COPD
HLA-A/-E
HLA-DP/Q/R
CD74
CD68
YY1
ABHD12
ADAM9
BHLHE40
HIST1H1B
IDH1
ITGAX
CITED2
COA3
GLS
FN1
ADAM10
CD163
CD36
ITGB1
RHOA
PPARG
SERPINA1
KRAS
SOAT1
ADAM17
CXCL16
FCGR3A
ID2
MT-CO1
MT-CO3
MT-ND4
AM cluster 8 4 3 5 1 7 0 6 2 9 10 11 12
control
COPD
group median of mean expression per AM cluster
1 0.5 0 -0.5 -1
D
median Wilcoxon score per gene and AM cluster of DE genes found in 1 AM cluster (acc. to B)
DE genes
median Wilcoxon score
up in control
up in COPD
RPS14
ITIH5
CST3
GPX4
COPA
BZW1
ABI1
GTF3A
DOK2
GUSB
CASP1
CAPRIN1
FUCA1
NCOR1
SCPEP1
SOAT1
STT3B
COLEC12
PARP4
SQSTM1
SPN
MAP7D1
AM cluster 8 5 3 1
bins of DE genes acc. to AM clusters with the highest Wilcoxon score for the respective gene
E
expression of putative AM cluster-specific DE genes (acc. to D)
ITIH5
COLEC12
control
COPD
mean expression per AM cluster
AM cluster
F
GOEA based on DE genes shared by >2 AM clusters (acc. to B)
GO:0005925 focal adhesion
GO:0002274 myeloid leukocyte activation
GO:0006955 immune response
REAC:R-HSA-168256 immune system
GO:0002366 leukocyte activation involved in immune response
GO:0002263 cell activation involved in immune response
GO:0006898 receptor-mediated endocytosis
GO:0006909 phagocytosis
GO:0002478 antigen processing and present. of exog. peptide antigen
REAC:R-HSA-1236975 antigen processing/cross presentation
-log10(adjusted p-value)

Figure 24. Distribution-free DE-gene analysis of identified AM clusters

(A) Schematic workflow of the distribution-free permutation test-based DE analysis approach. **(B)** UpSet plot of calculated DE genes across AM clusters. DE genes found in the same AM clusters are binned and the size of the bins is represented as a bar chart. At the bottom, dots indicate which AM clusters contained and shared these DE genes. **(C)** Heat map representation of the union of all DE genes found in the AM cluster. Depicted is the group median (group = COPD or control) of the z-transformed mean expression data per donor and AM cluster across all AM clusters, and the names of some selected DE genes are shown on the right side of the plot. Columns and rows of the heat map are sorted by hierarchical clustering. **(D)** Dot plot for the assessment of the AM cluster specificity of DE genes. DE genes found in only one cluster according to UpSet plot in (B) are depicted on the x-axis and the respective Wilcoxon scores for each AM cluster on the y-axis. Dots of DE genes are highlighted (green = up in control; orange = up in COPD) and the respective gene name is shown if the p-value for the significance test of the DE gene in the respective AM cluster is < 0.01 and its Wilcoxon score has a difference of ≥ 2 to the median Wilcoxon score of the remaining AM clusters. In addition, DE genes with the highest Wilcoxon score in the same AM cluster are binned. The corresponding AM cluster number is shown at the bottom of the plot and dashed lines separate the bins. **(E)** Lollipop plot of two selected DE genes from (D)**.** **(F)** Selected functional gene sets from GSEA based on DE genes that reach the defined significance cutoffs in more than two AM clusters (acc. to (B)).

next examined the DE results. Interestingly, among the DE genes, we identified NOTCH-associated genes such as *YY1* and the metalloendopeptidases *ADAM9*, *ADAM10* and *ADAM17* (**Figure 24C**). Moreover, in accordance with the Compass analysis, lipid metabolism-associated genes (*e.g. CD36*, *COLEC12*, *SOAT1*, and *PPARG*) were upregulated in COPD, which was also true for genes associated with oxidative phosphorylation (*e.g. COA13*, *MT-CO2*, *MT-ND2*, and *MT-ATP6*) (**Figure 24C-E**). We also found an increased expression of the surface molecule *CD163*, which has been previously described in the context of AMs in COPD by immunohistochemistry (Kaku et al., 2014; Kunz et al., 2011).

Finally, in order to put the DE genes, which showed a strong difference in several AM clusters and thus are likely representatives of strong changes in the AM population, into a more functional context, we performed a gene set enrichment analysis (GSEA) on genes found as DE in at least three AM clusters (**Figure 24B**). This analysis revealed an enrichment of terms related to focal adhesion and immune response such as antigen processing and presentation (**Figure 24F**).

Collectively, we were able to characterize the identified AM clusters, which showed that this cell population exhibits high transcriptional plasticity. Furthermore, DE analysis showed that the majority of transcriptional differences are attributable to increased expression in COPD. In addition, we found that for the majority of the identified DE genes, the direction of the difference between COPD and control AMs was shared across several clusters, indicating a global and similar impact of the disease on different AM states in the alveolar space.

4.6. Validation of *in silico* findings

4.6.1. Reduced MHC I expression on the surface of AMs from COPD patients

Since the DE analysis with subsequent GSEA indicated an enrichment of focal adhesion and antigen presentation in AMs (**Figure 24F**), we further investigated the underlying DE genes. First, we visualized the expression of antigen presentation-associated DE genes, which showed again that the majority of genes were higher expressed in COPD, except for the human leukocyte antigen (HLA) genes *HLA-E* and *HLA-DRB1*, and the HLA-associated genes *B2M* and *CD74*, which exhibited higher expression in control samples (**Figure 15A**). To assess whether the altered expression could also be observed in other HLA genes, we comprehensively evaluated the expression profiles of MHC II and I cell surface receptors of AMs. The most abundant HLA gene in the dataset was the MHC II-encoding *HLA-DRA* gene (**Figure 25B**). Moreover, among the top expressed HLA genes, we found further genes of the MHC II class (*HLA-DRB1*, *HLA-DRB5*, *HLA-DPA1*, *HLA-DPB1* and *HLA-DQB1*) together with MHC I-encoding genes (*HLA-A*, *HLA-B*, *HLA-C* and *HLA-E*). Intriguingly, when plotting the expression of these genes for control and COPD samples separately (**Figure 25C+D**), we found that the HLA genes were consistently downregulated in the COPD samples, with the only exception of HLA-DRA and HLA-DRB5, which showed no difference in expression (**Figure 25D**). To assess whether the

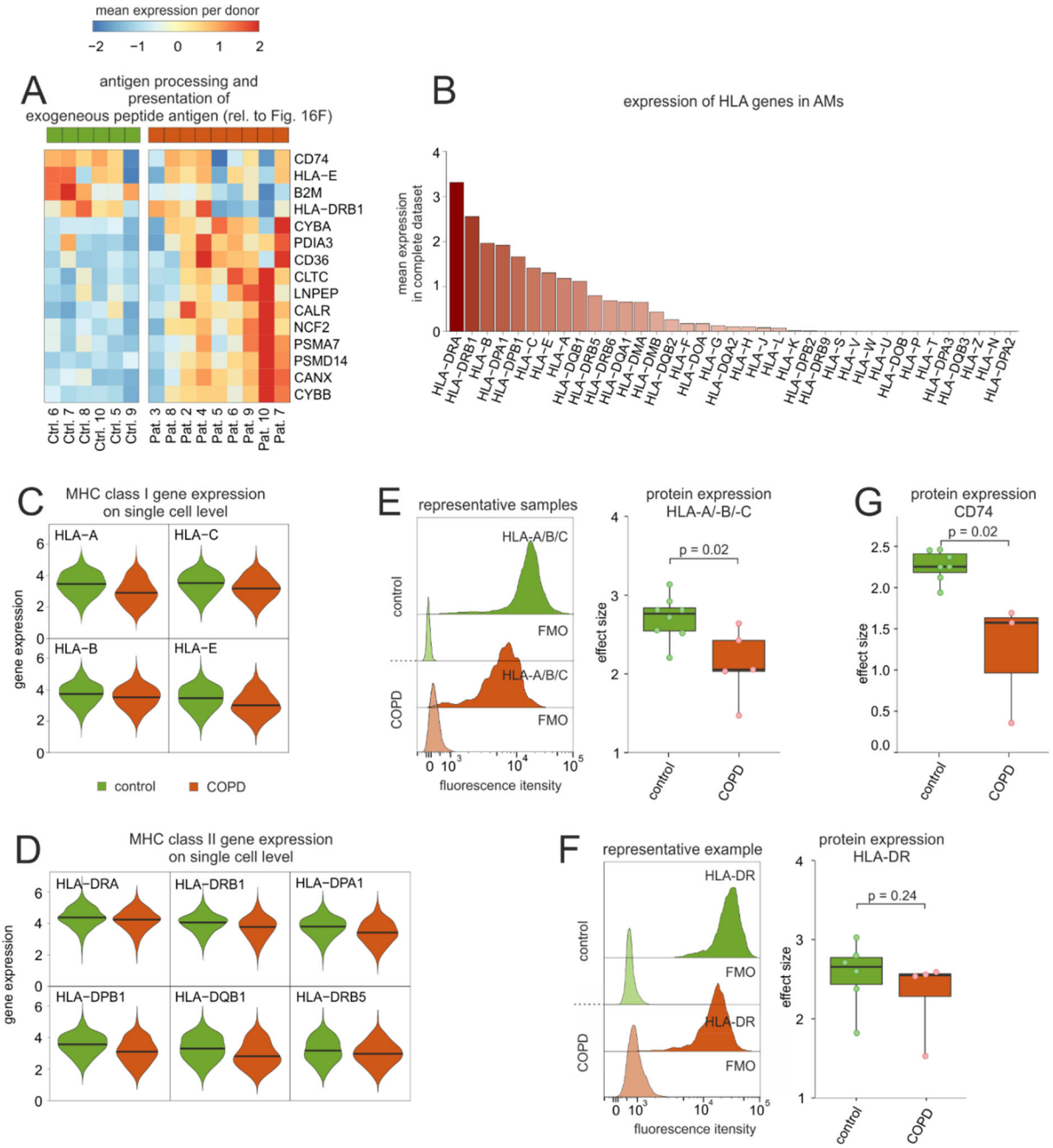


Figure 25. Alteration of MHC expression in AMs of COPD patients

(A) Heat map of DE genes, which according to GSEA are enriched in the GO term ‘antigen processing and presentation of exogenous peptide antigen’ (acc. to **Figure 24F**). The mean gene expression per donor is represented as a z-transformed value (across all donors). Rows of the heat map are sorted by hierarchical clustering. **(B)** Bar plots showing the mean expression of various MHC genes in complete AM dataset. (C+D) Violin plots of MHC I (C) and MHC II (D) gene expression in AMs based on scRNA-seq data. The plots show the expression across the donors, whereby the donors were downsampled to the same number

of cells, followed by downsampling to the same number of cells between COPD and control. The plots display cells with an expression > 0. **(E)** Fluorescence intensity histograms on the left showing representative samples of flow cytometric analysis of HLA-A/-B/-C expression on the cell surface of isolated AMs (FMO = fluorescence minus one). On the right side, box plots of the calculated effect sizes of HLA-A/-B/-C expression in COPD and control with the representation of individual donors (control n = 8, COPD n = 5). **(F)** Fluorescence intensity histograms on the left showing representative samples of flow cytometric analysis of HLA-DR expression on the cell surface of isolated AMs. On the right side, box plots of the calculated effect sizes of HLA-DR expression in COPD and control with representation of individual donors (control n = 6, COPD n = 4). **(G)** Box plots of the calculated effect sizes of CD74 expression in COPD and control with representation of individual donors (control n = 7, COPD n = 3).

observed differences between COPD and controls at the transcriptional level were also translated to the protein level, we isolated AMs from additional patients and measured protein levels of MHC class I (HLA-A/-B/-C) (**Figure 25E**) and class II (HLA-DR) (**Figure 25F**). In accordance with the transcriptome data, MHC class I was significantly reduced on the surface of AMs from COPD patients, while MHC class II molecules showed no clear difference (**Figure 25E+F**). As mentioned above, we have also observed transcriptional downregulation of *CD74*, which has an ambivalent function. Intracellularly, it serves as an invariant chain to support the loading of MHC class II molecules, but CD74, which is expressed on the cell surface, acts as an important receptor to trigger immune-regulatory responses. Interestingly, we observed a significant decrease in cell surface expression of CD74 on AMs of COPD patients (**Figure 25G**).

4.6.2. COPD-associated decrease in migration of AMs towards CCL3

Next, we plotted DE genes associated with focal adhesion according to GSEA (**Figure 24F**). The DE genes associated with focal adhesion and that were elevated in COPD, included cell adhesion molecules (e.g. ITGB1 or ALCAM) and cytoskeleton organizing molecules (e.g. PARVG or ARPC2/5) (**Figure 26A**), pointing towards either increased motility, or increased local adhesion strength of these cells. To assess the migratory and

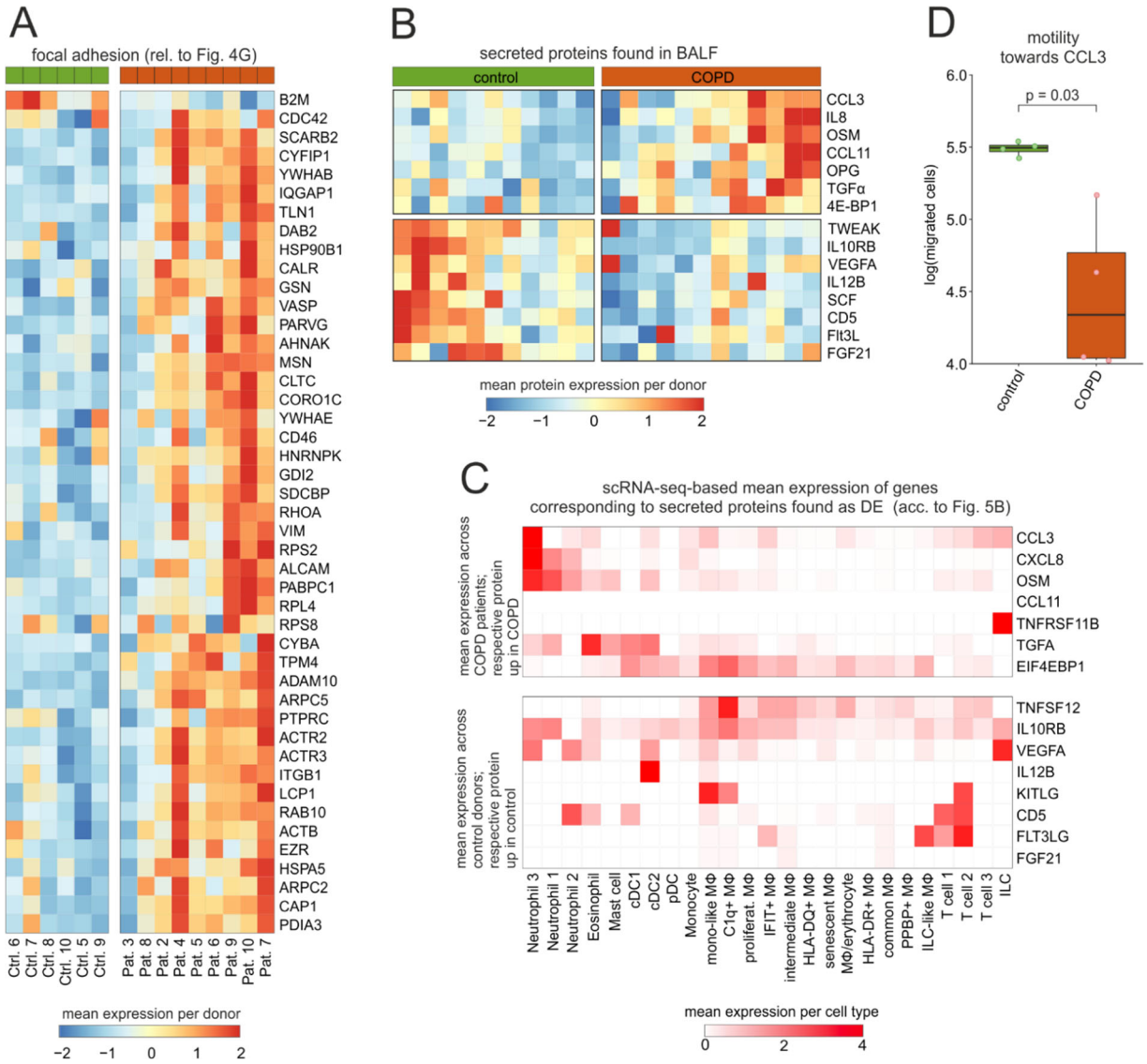


Figure 26. Reduced migration of AMs from COPD patients towards CCL3

(A) Heat map of DE genes, which according to GSEA are enriched in the GO term associated with focal adhesion (acc. to **Figure 24F**). The mean gene expression per donor is represented as a z-transformed value (across all donors). Rows of the heat map are sorted by hierarchical clustering. **(B)** Heat map representation of proteins detected in BALF with a p-value < 0.1 according to the Wilcoxon rank sum test between COPD patients and control donors (control n = 11, COPD n = 12). The mean protein expression (identified by Olink Proteomics) per donor is represented as a z-transformed value (across all donors). Columns of the heat map are sorted by hierarchical clustering. **(C)** Heat map representing the BALF cell type-dependent expression of genes whose protein counterparts were found in BALF (acc. to (B)). If the corresponding protein is upregulated in the control group, the mean gene expression was calculated across the control donors and represented as a z-transformed value (across all control donors) (top panel). Similarly, mean expression was calculated only in COPD patients when the corresponding protein is upregulated in COPD (bottom panel). **(D)** Quantification of the migratory capability of AMs towards CCL3 displayed in a box plot with the representation of individual donors (control n = 4, COPD n = 4).

chemotactic ability of AMs, we first aimed to select a chemokine gradient that most closely reflected the situation in the alveolar space. For this purpose, we examined BALF for cytokines and chemokines and found a significantly elevated protein concentrations of chemotactic factors that recruit neutrophils (IL8) and eosinophils (CCL11) in COPD patients (**Figure 26B**). This is in accordance with our MCFC-based quantification of immune cells in the alveolar space of COPD patients and control donors, which revealed a significant increase of neutrophils and a tendency towards elevated eosinophil numbers in COPD patients (**Figure 9**). Our BALF screen also revealed a COPD-associated elevation of the CCL3 level (**Figure 26B**), which is a known chemokine involved in motility and migration of macrophages (Opalek et al., 2007). As indicated by scRNA-seq, CCL3 can be derived from numerous cell types in the alveolar space of COPD patients, especially from neutrophils, but also from T cells, $IFIT^{+}$ and mono-like macrophages (**Figure 26C**). However, despite the increase of CCL3 in BALF, the intrinsic property of AMs from COPD patients was an overall reduced migratory response towards CCL3 (**Figure 26D**), possibly due to increased adhesive properties.

4.6.3. Increased proton leakage of mitochondria in AMs of COPD patients

Along with DE genes associated with antigen presentation and cell adhesion, we also found a significantly increased expression of mitochondrial genes in COPD AMs (**Figure 24C**). Given that we have already excluded cells from further analysis that possessed a high relative mitochondrial gene proportion, indicating apoptotic cells (**Material and Methods**), the upregulation of mitochondrial genes in the remaining AMs could be an indication of a process that may not be directly related to apoptosis. Thus, the increased mitochondrial gene expression might represent a cellular adaptation to elevated metabolic activity as indicated by the Compass analysis (**Figure 20**). In addition, it has been described that smoke exposure can lead to increased mitochondrial dysfunction in cells and thus to increased production of ROS (Hoffmann et al., 2013). Since the main cause of COPD in industrialized countries is cigarette abuse and the observed upregulation of mitochondrial genes could indicate

a compensation mechanism for mitochondrial dysfunction, we investigated the mitochondrial functionality of AMs in more detail by using the Seahorse technology (**Figure 27A**). Indeed, we identified increased baseline respiration in AMs of COPD patients (**Figure 27B+C**), which reflects an elevated energy demand. We also found a significant increase in proton leakage in AMs of COPD patients, which is indicative for increased ROS production in COPD (Boukhenouna et al., 2018; Cheng et al., 2017; McGuinness and Sapey, 2017) and mitochondrial dysfunction (Eapen et al., 2019; Hoffmann et al., 2019; Ng Kee Kwong et al., 2017).

Taken together, these validation studies verified the results obtained from scRNA-seq data. In detail, we confirmed a decreased expression of MHC I cell surface receptors, a reduced migratory ability and an increased mitochondrial dysfunction in AMs of COPD patients. Together with the aforementioned results, a picture of a profound metabolic and functional alteration of AMs in COPD patients emerges.

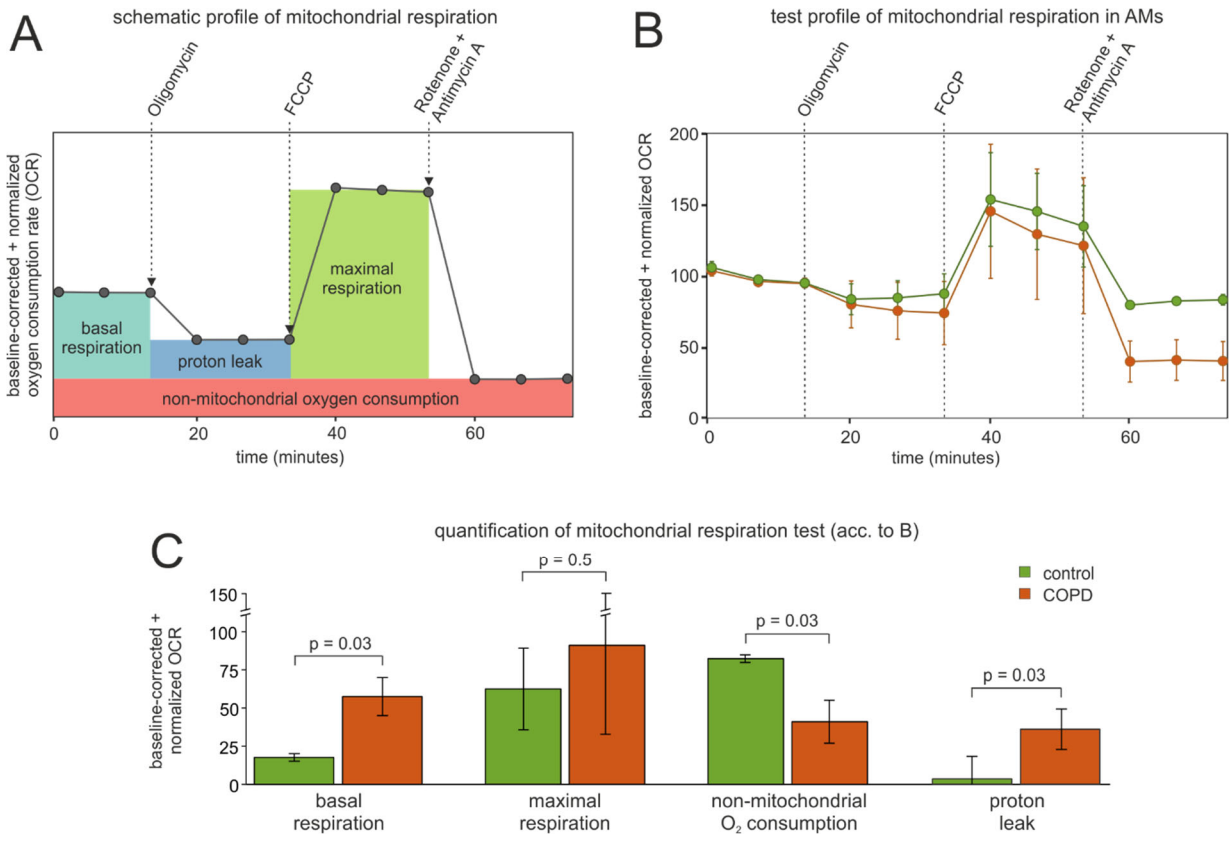


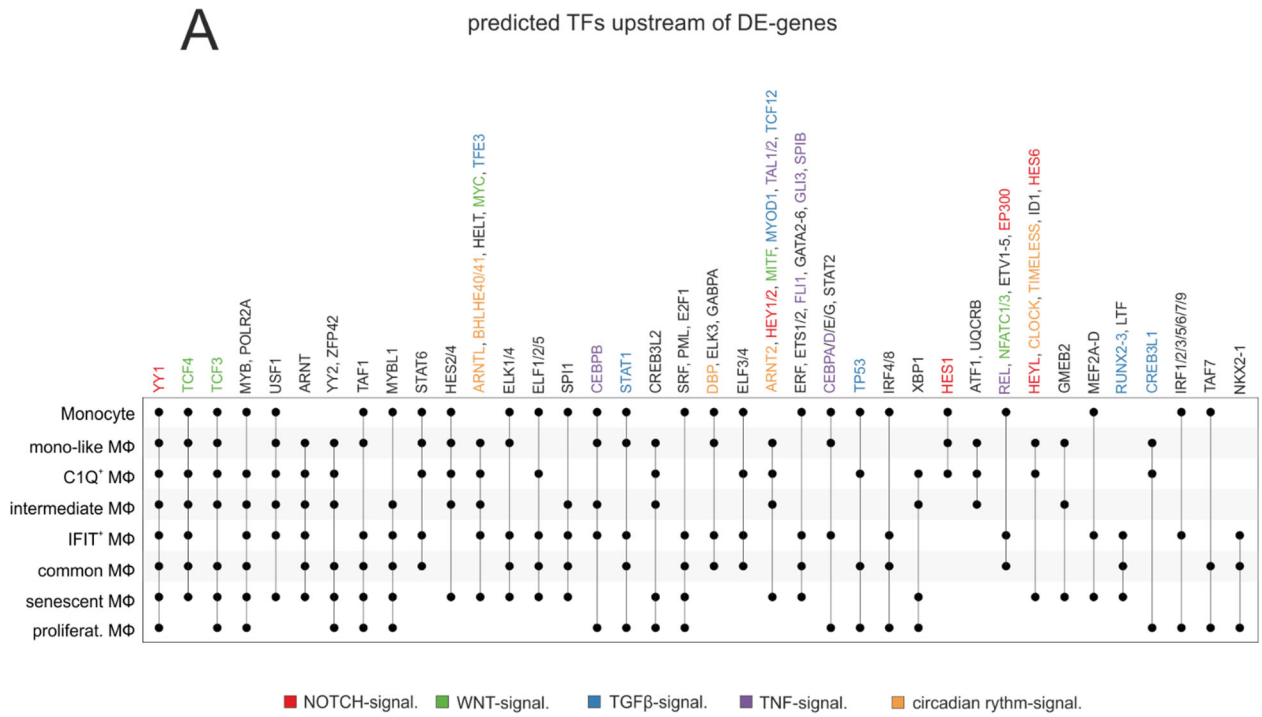
Figure 27. Mitochondrial alterations in AMs from COPD patients

(A) Schema of the time-dependent course of the oxygen consumption rate (OCR) and the inferred mitochondrial parameters based on the injection of different compounds (shown at the top of the plot). **(B)** Evaluation of mitochondrial function via the time-dependent course of the oxygen consumption rate (OCR) in AMs, using baseline-corrected values. Error bars indicate the standard deviations derived from the measurements of several donors (control n = 2, COPD n = 3). Dashed arrows represent the injection of various compounds (shown at the top of the plot) used to assess different aspects of mitochondrial function (acc. to (A)). **(C)** Bar plots showing quantifications of different aspects of mitochondrial function inferred from the OCR measurement in (B).

4.7. DE-gene regulation by complex cell-to-cell communication networks

4.7.1. Prediction of TFs regulating the expression of DE genes

After identification and characterization of transcriptional changes mediated by COPD, the next step was to understand the regulation of the identified DE genes in AMs. For this purpose, we focused on AM states with a minimum of 30 DE genes between COPD patients and control donors and predicted potential upstream transcriptional regulators of the DE genes. To determine which predictions were shared by several AM subtypes, we visualized TFs found in at least three AM states in an UpSet plot (**Figure 28A**). This representation of the data revealed that only a single TF was shared among all AM states, namely *YY1* that is an important modulator of NOTCH signaling (Liao et al., 2007; Yeh et al., 2003). The importance of the NOTCH signaling in the regulation of DE genes was further supported by the identification of the TFs *HES1* and *HEY1*, which are known to be induced by NOTCH signaling and whose co-regulation was predicted only in mono-like and C1Q^{+} macrophages. In addition to NOTCH signaling, we also found transcriptional regulators associated with WNT signaling (e.g. *TCF3/4*, *MYC*, and *NFATC1/3*), TNF/ NF-κB signaling (e.g. *CEBPB* and *REL*) and TGFβ signaling (e.g. *TFE3* and *MYOD1*), as well as TFs involved in the regulation of the circadian rhythm (e.g. *BHLHE40/41*, *CLOCK*, and


A
predicted TFs upstream of DE-genes
YY1
TCF4
TCF3
MYB, POLR2A
USF1
ARNT
YY2, ZFP42
TAF1
MYBL1
STAT6
HES2/4
ARNTL, BHLHE40/41, HELT, MYC, TFE3
ELK1/4
ELF1/2/5
SPI1
CEBPB
STAT1
CREB3L2
SRF, PML, E2F1
DBP, ELK3, GABPA
ELF3/4
ARNT2, HEY1/2, MITF, MYOD1, TAL1/2, TCF12
ERF, ETS1/2, FLI1, GATA2-6, GLI3, SPIB
CEBPA/D/E/G, STAT2
TP53
IRF4/8
XBP1
HES1
ATF1, UQCRB
REL, NFATC1/3, ETV1-5, EP300
HEYL, CLOCK, TIMELESS, ID1, HES6
GMEB2
MEF2A-D
RUNX2-3, LTF
CREB3L1
IRF1/2/3/5/6/7/9
TAF7
NKX2-1
Monocyte
mono-like MΦ
C1Q+ MΦ
intermediate MΦ
IFIT+ MΦ
common MΦ
senescent MΦ
proliferat. MΦ
NOTCH-signal.
WNT-signal.
TGFβ-signal.
TNF-signal.
circadian rythm-signal.

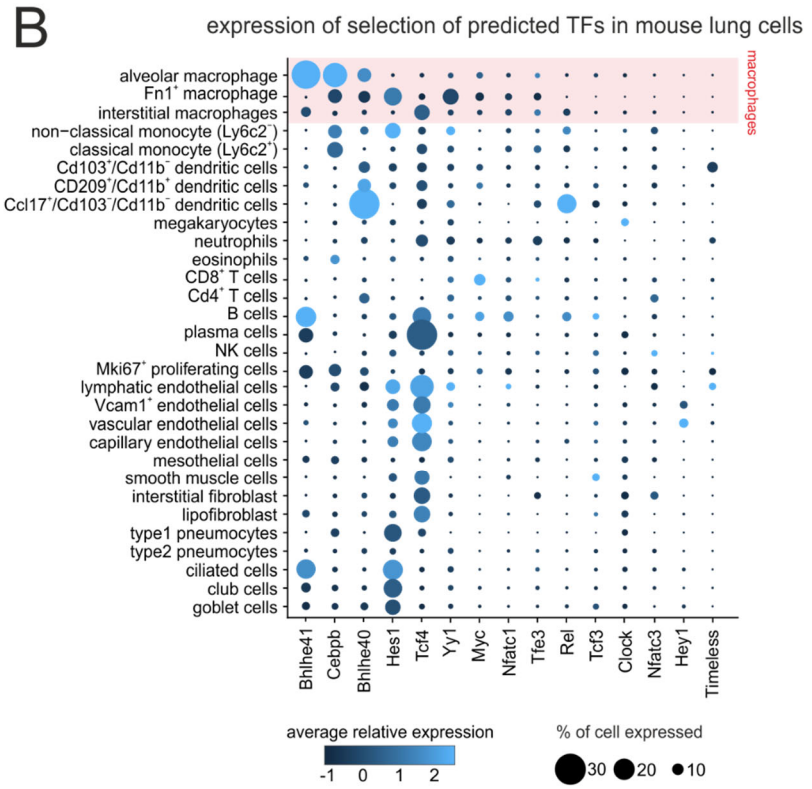

B
expression of selection of predicted TFs in mouse lung cells
alveolar macrophage
Fn1+ macrophage
interstitial macrophages
non−classical monocyte (Ly6c2−)
classical monocyte (Ly6c2+)
Cd103+/Cd11b− dendritic cells
CD209+/Cd11b+ dendritic cells
Ccl17+/Cd103−/Cd11b− dendritic cells
megakaryocytes
neutrophils
eosinophils
CD8+ T cells
Cd4+ T cells
B cells
plasma cells
NK cells
Mki67+ proliferating cells
lymphatic endothelial cells
Vcam1+ endothelial cells
vascular endothelial cells
capillary endothelial cells
mesothelial cells
smooth muscle cells
interstitial fibroblast
lipofibroblast
type1 pneumocytes
type2 pneumocytes
ciliated cells
club cells
goblet cells
macrophages
Bhlhe41
Cebpb
Bhlhe40
Hes1
Tcf4
Yy1
Myc
Nfatc1
Tfe3
Rel
Tcf3
Clock
Nfatc3
Hey1
Timeless
average relative expression
-1 0 1 2
% of cell expressed
30 20 10

Figure 28. Prediction of TFs upstream of DE-genes in AMs

(A) UpSet plot of predicted transcriptional regulators of DE genes. Dots indicate which clusters contain and share predicted transcriptional regulators. The names of selected regulators are shown on the bottom of the plot with the font color indicating the association with NOTCH, WNT, TGF-β, TNF or circadian rhythm signaling. **(B)** Dot plot representation of the expression of selected TFs (acc. to (A)) in lung-derived cells from mice (Angelidis et al., 2019). The macrophage populations are highlighted with a pale red background.

TIMELESS). Importantly, some predicted TFs are known to be not associated with a single signaling pathway, but rather represent a connecting point of several pathways, such as *MYC*, which plays a central role in WNT, NOTCH and TGFβ signaling and has therefore been predicted in almost all AM states. Next, we investigated whether a selection of the predicted TFs are expressed at all in lung macrophages. Since many TFs are rather lowly expressed and thus often not sufficiently detected by some scRNA-seq technologies, we examined the expression in an additional dataset obtained from aged mouse lungs (Angelidis et al., 2019b). Many of the predicted TFs are indeed expressed in lung macrophages, including *Bhlhe40/41*, *Cebpb*, *Yy1*, *Hes1*, *Tcf4*, *Myc*, *Nfatc1*, and *Tfe3* (**Figure 28B**).

4.7.2. Construction of cell-to-cell communication networks of BALF cells

Since the induction and regulation of specific signaling pathways is always dependent on the microenvironment and the associated intercellular communication, we next investigated cell-cell interactions between immune cells in the alveolar space. Therefore, we applied a recently introduced model for cell-to-cell communication based on known receptor-ligand interactions (CellPhoneDB (Efremova et al., 2019, 2020; Vento-Tormo et al., 2018)). The network construction of the identified cell-to-cell interactions from control samples revealed that mono-like and C1Q^{+} macrophages were the major hubs within the network (**Figure 29A, left network**). Upon COPD, the overall cell-to-cell communication was strongly increased (**Figure 29A, right network**). To determine differences between the control and COPD network, we visualized variations in connectivity in a heat map, which showed that in COPD the

highest overall increase in cell-to-cell interactions was found in mono-like, C1Q$^+$, and IFIT$^+$ macrophages (**Figure 29B**). Next, we focused our analyses on one of the hub cell-types within the networks, namely mono-like macrophages. Among the predicted mono-like macrophage interactions that possessed a clear difference between COPD and control, we identified several receptor-ligand pairs associated with the TNF superfamily (**Figure 29C**). Furthermore, we found an increased likelihood of interaction between the ligand TGFB1 and the receptor TGFBR1 in COPD. Taken together, the cell-to-cell interaction network construction revealed a strong increase in the communication of immune cells in the alveolar space in COPD, with mono-like and C1Q$^+$ macrophages being the most strongly connected cell types. In addition, we found that many predicted receptor-ligand interactions, which were found exclusively in COPD, were associated with TNF signaling.

4.7.3. Modeling of intracellular regulation of DE genes in mono-like macrophages

While CellPhoneDB predicts potential receptor-ligand interactions based on their expression on sender and receiver cells, it does not model the downstream transcriptional effects of these interactions. Thus, to enable the integration of downstream effects in our analysis, we applied the NicheNet algorithm (Bonnardel et al., 2019; Browaeys et al., 2019) and focused the analysis again on mono-like macrophages. In a first step, we identified ligands, whose ligand activities could best define the DE genes in mono-like macrophages (**Figure 30A**). Importantly, among the top-ranked ligands, we again found *TGFB1*. To facilitate the downstream interpretation of the DE genes, which might be regulated by the predicted upstream ligands, we grouped them according to the cellular processes in which they are involved including lipid metabolism, immune system process, cell adhesion, and differentiation (**Figure 30A, bottom panel**). Notably, some of these DE genes were also associated with cell signaling pathways containing TFs that were predicted as potential upstream transcriptional regulators (**Figure 28**), such as NOTCH signaling (*RBPJ* and *ADAM9*) and circadian rhythm (*BHLHE40/41*). Next, the linkage

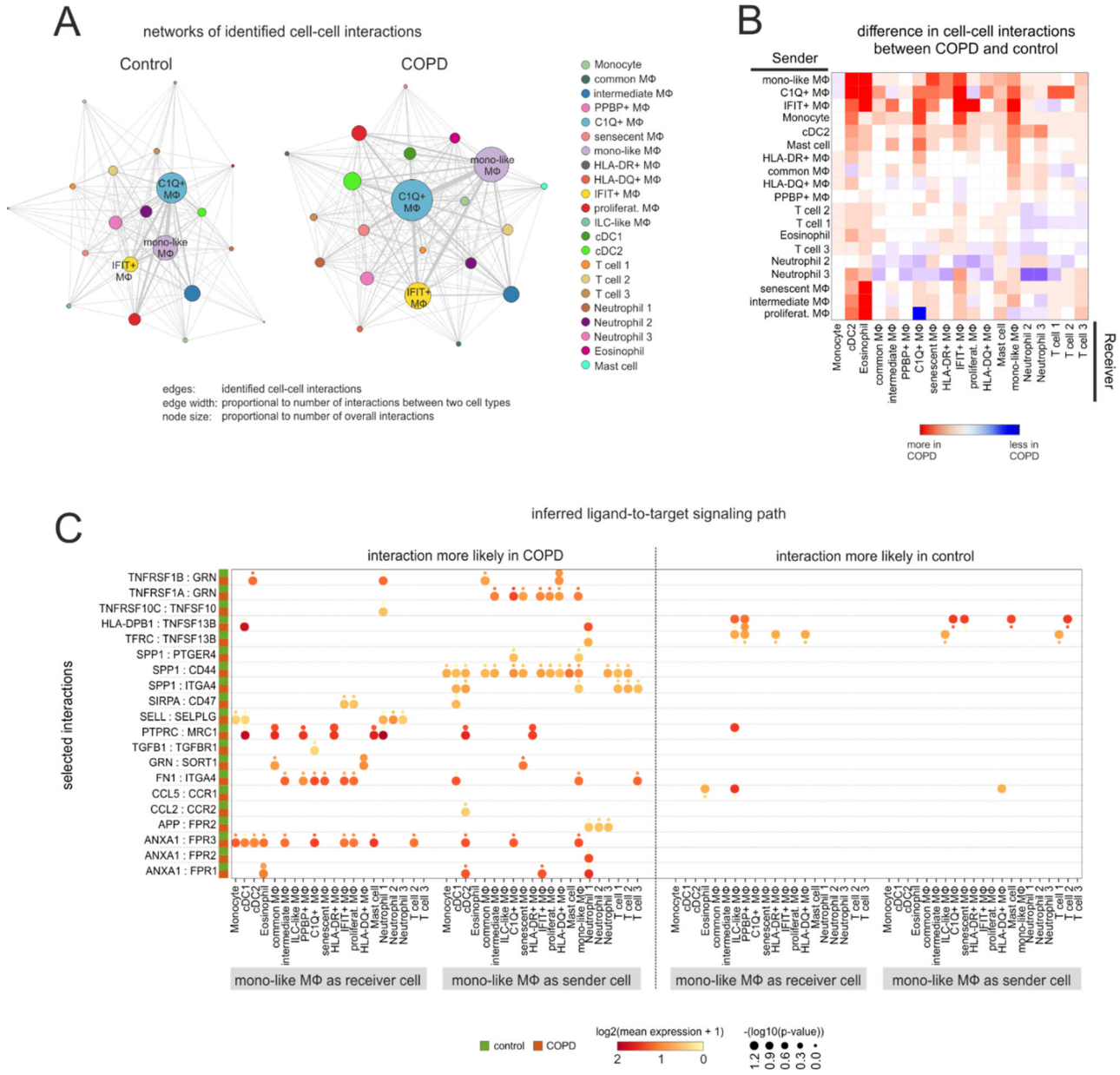


Figure 29. Construction of cell-to-cell communication networks of BALF immune cells

(A) Network representation of predicted cell-to-cell interactions derived from CellPhoneDB. The names of the three most interconntected cell types are displayed. **(B)** Heat map representing differences in the cell-to-cell connections between the networks in (A). Columns and rows of the heat map are sorted by hierarchical clustering. **(C)** Dot plot representation of mono-like macrophage-dependent ligand-receptor interactions predicted by CellPhoneDB that show significant enrichment (represented by the p-value) of the interacting pair in the interacting cell types either in COPD or in the control. Depicted are only selected interactions.

between target genes and the predicted upstream ligands were visualized in a Circos plot, revealing the highly complex and strongly interwoven nature of the cell-to-cell communication network for DE gene regulation (**Figure 30B**). In addition, the Circos plot representation also provided an overview of the potential sender cells of the predicted ligands. Thus, mast cells and the macrophage population were found to be the primary sender cells of *TGFB1*. This became even more evident when the mean expression of the predicted ligands per cell type and condition (COPD and control) was plotted, which showed that especially mono-like and C1Q$^+$ macrophages together with mast cells expressed *TGFB1*, whereby the respective mean expressions were elevated in COPD (**Figure 30C**). To investigate whether this observed increase of *TGFB1* expression in COPD patients was also translated to an elevated protein level, we examined the BALF of COPD patients and control donors for the latency-associated peptide TGF-β1 (LAP TGF-β1), which served as a surrogate for the TGF-β1 protein level. Although not significant, we detected a clear tendency towards increased LAP TGF-β1 levels in COPD (**Figure 30D**). Next, we further characterized the *TGFB1*-mediated regulation of identified DE genes by inferring signaling paths between *TGFB1* and its predicted target genes. The nodes in the constructed paths were colored according to the expression fold change (FC) between COPD and control, whereby, among the transcriptional regulators found to be upregulated in COPD, we identified the classical TGF-β signaling mediator *SMAD4* (**Figure 30E**). In addition, we also found some TFs that were already predicted as potential upstream transcriptional regulators of DE genes in AMs (**Figure 28**), which showed increased FCs in this analysis and thus indicated an up-regulation in COPD. Among these TFs were e.g. *EP300* and *MYC*, which are not only involved in TGFβ but also in other signaling pathways, which in turn illustrated again the complex and interconnected DE-gene regulation in mono-like macrophages.

In summary, within a highly complex network of DE-gene regulation , we predicted TGF-β signaling as a prominent regulator in mono-like macrophages but also other pathways including NOTCH signaling.

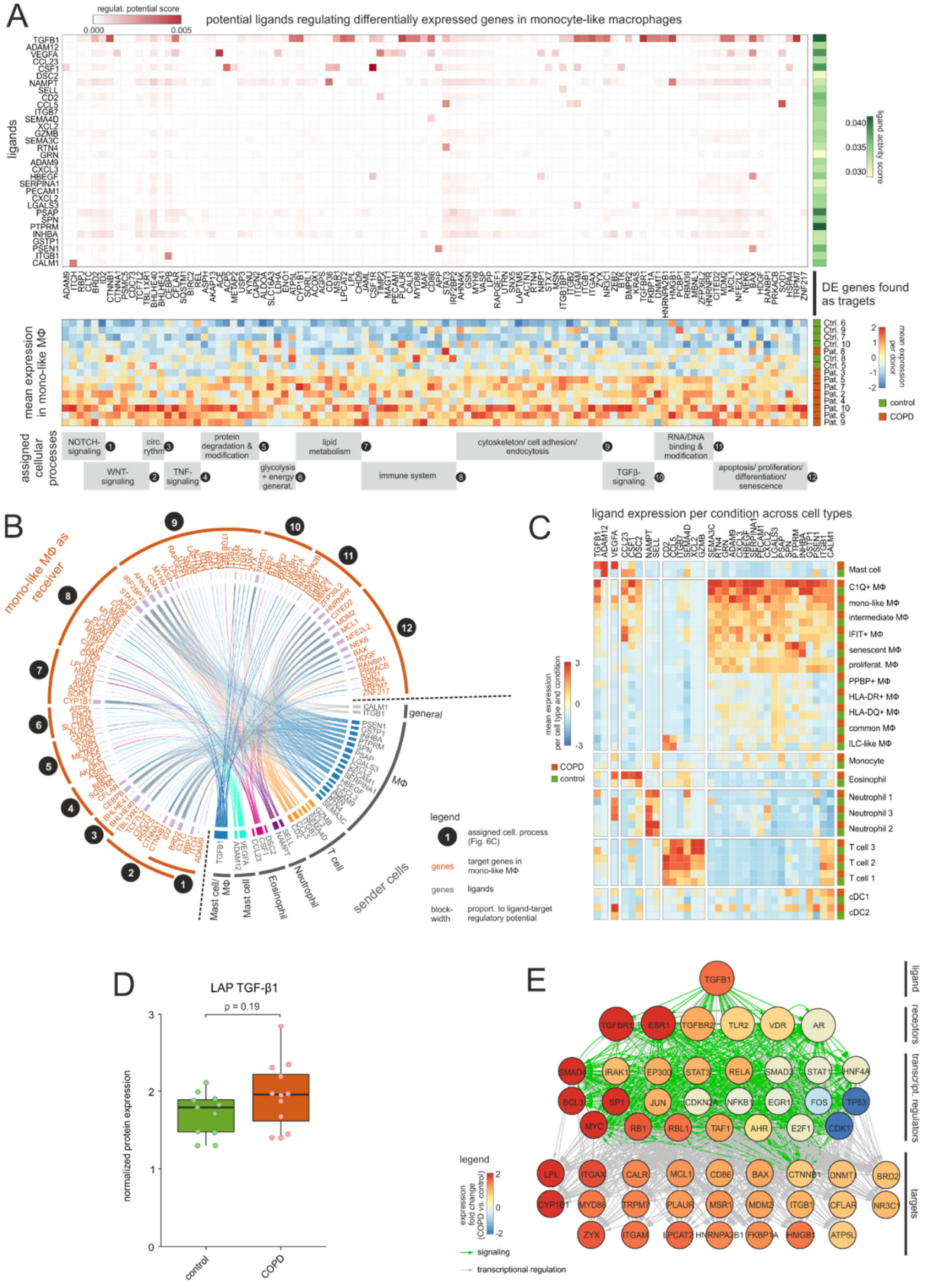

A
potential ligands regulating differentially expressed genes in monocyte-like macrophages
regulat. potential score
ligands
ligand activity score
DE genes found as tragets
mean expression in mono-like MΦ
mean expression per donor
control
COPD
assigned cellular processes
NOTCH-signaling
circ. rythm
protein degradation & modification
lipid metabolism
cytoskeleton/ cell adhesion/ endocytosis
RNA/DNA binding & modification
WNT-signaling
TNF-signaling
glycolysis + energy generat.
immune system
TGFβ-signaling
apoptosis/ proliferation/ differentiation/ senescence
B
mono-like MΦ as receiver
general
MΦ
T cell
Neutrophil
Eosinophil
Mast cell/ MΦ
Mast cell
sender cells
legend
assigned cell. process (Fig. 6C)
target genes in mono-like MΦ
ligands
proport. to ligand-target regulatory potential
C
ligand expression per condition across cell types
mean expression per cell type and condition
COPD
control
Mast cell
C1Q+ MΦ
mono-like MΦ
intermediate MΦ
iFIT+ MΦ
senescent MΦ
proliferat. MΦ
PPBP+ MΦ
HLA-DR+ MΦ
HLA-DQ+ MΦ
common MΦ
ILC-like MΦ
Monocyte
Eosinophil
Neutrophil 1
Neutrophil 3
Neutrophil 2
T cell 3
T cell 2
T cell 1
cDC1
cDC2
D
LAP TGF-β1
p = 0.19
normalized protein expression
control
COPD
E
ligand
receptors
transcript. regulators
targets
TGFB1
TGFBR1
ESR1
TGFBR2
TLR2
VDR
AR
SMAD3
IRAK1
EP300
STAT2
RELA
SMAD2
STAT1
HNF4A
BCL3
SP1
JUN
CDKN2A
NFKB1
EGR1
FOS
TP53
MYC
RB1
RBL1
TAF1
AHR
E2F1
CDK1
LPL
ITGAX
CALR
MCL1
CD86
BAX
CTNNB1
DNMT1
BRD2
CYP1B1
MYD88
TRPM7
PLAUR
MSR1
MDM2
ITGB1
CFLAR
NR3C1
ZYX
ITGAM
IPCAT2
HNRNPA2B1
FKBP1A
HMGB1
ATP5L
legend
expression fold change (COPD vs. control)
signaling
transcriptional regulation

Figure 30. Investigating cell-to-cell interactions to infer important signaling pathways in AMs

The focus of this analysis is on mono-like macrophages. **(A)** The top heat map represents the NicheNet analysis showing the potential of predicted upstream ligands (on the y-axis) to regulate downstream targets derived from DE genes (on the x-axis). The ligand activity score is depicted as the color-coded area under the precision recall curve (AUPR) on the right side of the plot. The ligands were grouped and ordered based on the cell types from which they are most likely expressed (acc. to (C)). The mean gene expression of the target genes across donors is shown in the lower heat map. The mean gene expression per donor is represented as a z-transformed value (across all donors). Rows of the heat map are sorted by hierarchical clustering. Target genes are ordered and grouped according to cellular functions and signaling pathways, as illustrated at the bottom of the plot. **(B)** Circle plot showing possible regulatory connections between the ligands expressed on different sender cells (acc. to (C)) and the downstream target genes. Target genes are ordered according to (A) as indicated by the numbering. **(C)** Heat map representation of the mean expression of predicted upstream ligands (acc. to (A)) in controls and COPD patients across cell types in BALF. Columns and rows are sorted by hierarchical clustering. **(D)** Box plot of the measured protein expression (by Olink Proteomics) in BALF of LAP TGF-β1 in COPD and control with representation of individual donors (control n = 11, COPD n = 12). **(E)** Representation of inferred ligand-to-target signaling path for TGF-β1 derived from the NicheNet analysis (targets: acc. to (A)). Coloring of nodes according to the mean expression fold change between mono-like macrophages from COPD and control.

4.8. Supply of the AM pool by blood monocytes

4.8.1. Identification of COPD-associated changes in cell-type proportions

To determine the difference between COPD and controls not only at the gene level, but also to investigate a possible redistribution in the population sizes of immune cells in the alveolar space upon disease, we next visualized the relative proportions of the identified immune cell types and subtypes (**Figure 31**). Among the AM states, which showed the largest population difference between COPD and controls, we found proliferating and mono-like macrophages as increased in COPD while common and HLA-DQ+ macrophages were decreased (**Figure 31A**). Intriguingly, together with T cell 3 **(Figure 31C**), the proliferating and mono-like macrophages were also the populations that

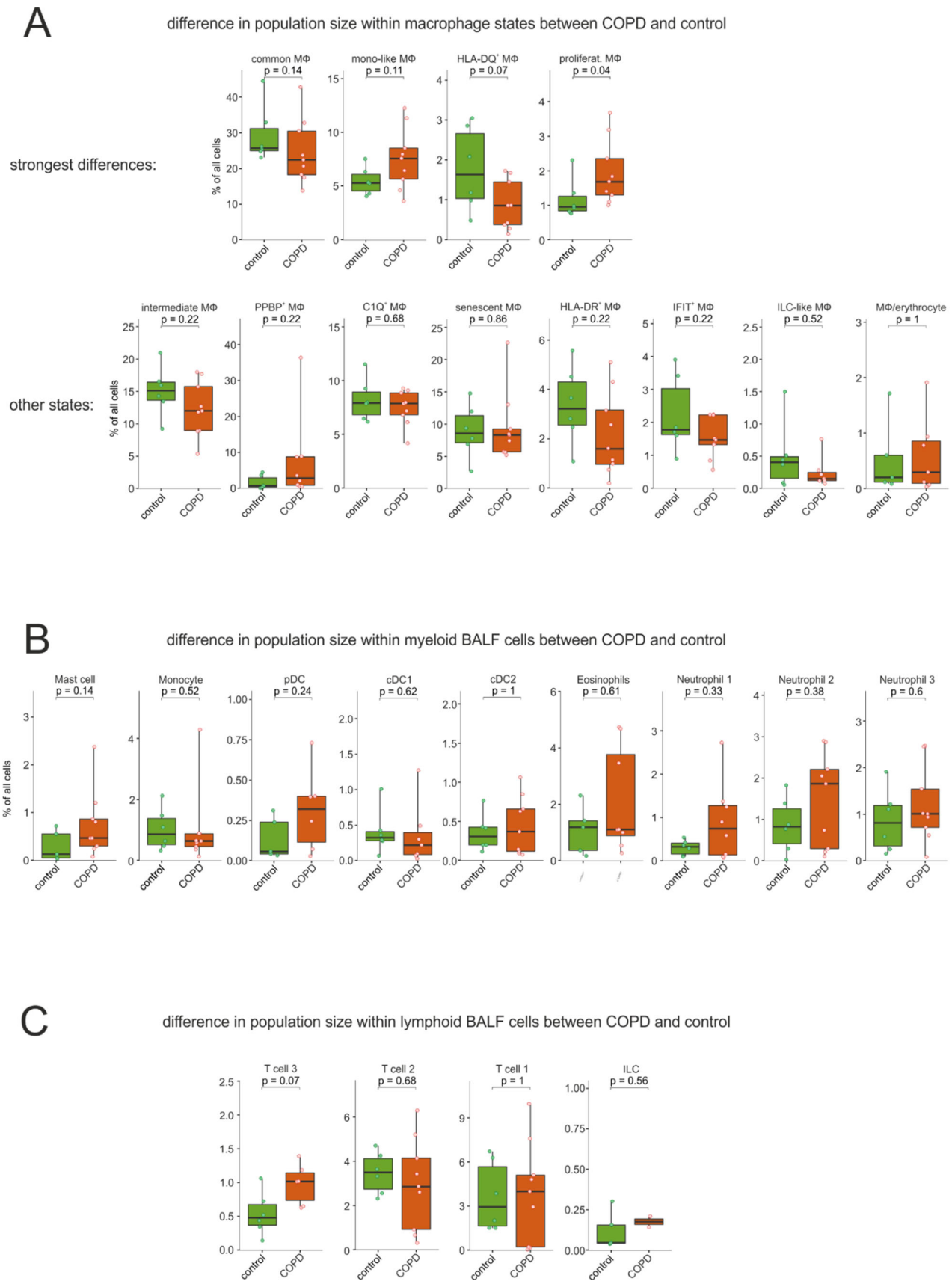

A
difference in population size within macrophage states between COPD and control
strongest differences:
other states:
common MΦ p = 0.14
mono-like MΦ p = 0.11
HLA-DQ+ MΦ p = 0.07
proliferat. MΦ p = 0.04
intermediate MΦ p = 0.22
PPBP+ MΦ p = 0.22
C1Q+ MΦ p = 0.68
senescent MΦ p = 0.86
HLA-DR+ MΦ p = 0.22
IFIT+ MΦ p = 0.22
ILC-like MΦ p = 0.52
MΦ/erythrocyte p = 1
% of all cells
control
COPD
B
difference in population size within myeloid BALF cells between COPD and control
Mast cell p = 0.14
Monocyte p = 0.52
pDC p = 0.24
cDC1 p = 0.62
cDC2 p = 1
Eosinophils p = 0.61
Neutrophil 1 p = 0.33
Neutrophil 2 p = 0.38
Neutrophil 3 p = 0.6
% of all cells
control
COPD
C
difference in population size within lymphoid BALF cells between COPD and control
T cell 3 p = 0.07
T cell 2 p = 0.68
T cell 1 p = 1
ILC p = 0.56
control
COPD

Figure 31. Quantification of the relative population sizes of BALF immune cells based on scRNA-seq data

Population sizes are represented as cell type frequencies calculated per donor and visualized in box plots. **(A)** Relative population sizes of AM subtypes. **(B)** Relative population sizes of BALF myeloid cells other than AMs. **(C)** Relative population sizes of BALF lymphoid cells.

showed the most significant increase in the relative proportion size of all immune cells of the alveolar space in COPD. Of the remaining cells, we found a tendency towards an

increased number of mast cells and pDCs in COPD patients; a trend that was also followed by neutrophils (**Figure 31B**).

Taken together, the strongest increase in the population size of immune cells derived from BALF was found in the T cell 3, proliferating macrophage, and mono-like macrophage populations.

4.8.2. Identification of AM states with enrichment of monocyte-derived macrophage signatures

As described above, mono-like macrophages expressed transcriptional markers reminiscent of monocytes (**Figure 22B**) and were increased in the relative population size upon COPD (**Figure 31A**). In addition, some of the identified pathways regulating the DE genes in this AM state (*e.g.* TGF-β and NOTCH signaling) (**Figure 30**) are known orchestrators of cell differentiation. These aspects are particularly interesting in the context of tissue macrophage replenishment that has been linked to the proliferation of tissue-resident cells (Hashimoto et al., 2013) but also influx and subsequent differentiation of monocyte-derived cells from the circulation (Guilliams and Scott, 2017). Therefore, we next investigated whether the mono-like AM state might represent an early stage of monocyte-to-macrophage differentiation. For this purpose, we used a gene signature of murine monocyte-derived macrophages (MDM) from the lungs of smoke-exposed mice (Wohnhaas et al., *unpublished data*) and assessed the enrichment of orthologous genes in the identified human AM states

(**Figure 32**). The strongest enrichment of the MDM signature was found in mono-like and C1Q⁺ AMs. As an additional gene signature, we used orthologous genes derived from murine lipid-associated macrophages (LAM), which were detected in human adipose tissue and were shown to be monocyte-derived by lineage tracing (Jaitin et al., 2019). We again found the strongest enrichment in mono-like and C1Q⁺ AMs. The same was also true for orthologous genes derived from Trem2⁺ foam cells of atherosclerotic plaques in mice (Kim et al., 2018), which were found by Lin et al. to originate from monocytes (Lin et al., 2019), supporting the hypothesis that mono-like and C1Q⁺ macrophages are actually derived from monocytes.

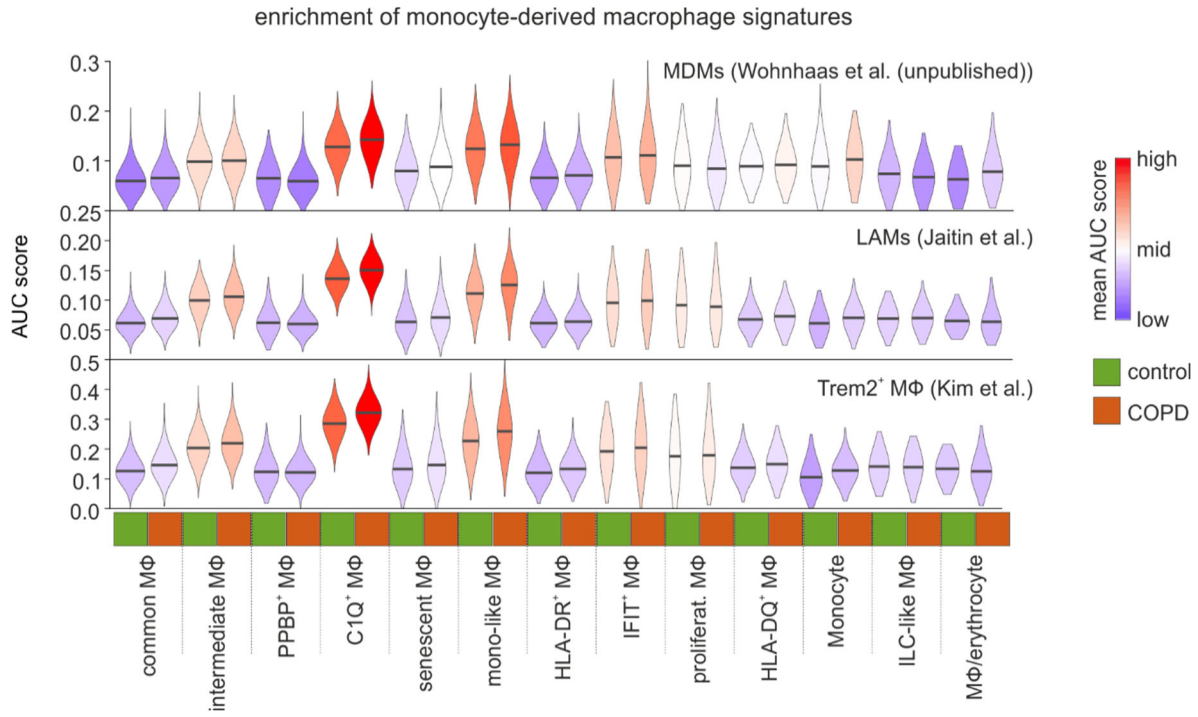


Figure 32. Enrichment of monocyte-derived macrophage signatures in AMs

Violin plots displaying enrichment of human orthologues of several murine monocyte-derived macrophage gene signatures across AM states in COPD and control based on 'Area Under the Curve' (AUC).

4.8.3. Generating a blood scRNA-seq dataset

The most likely source of monocytes that give rise to AMs are blood monocytes. Therefore, we performed scRNA seq of blood immune cells (n = 54,569 cells) from donors from whom we also obtained scRNA seq data of BALF cells (**Table 2**). Similar to the analysis of immune cells in the alveolar space (**Figure 15**), we annotated the cells according to the four-step cell-type annotation strategy (**Figure 16**), starting from the identification of main clusters (**Figure 33A**) and the assignment of the cells to the cell types using the machine learning-based cell-type annotation (**Figure 33B**). Next, we defined marker genes for the identified main clusters and determined which cell type occurred most frequently per main cluster according to the machine learning-based annotation (**Figure 33C**). Interestingly, main clusters 2, 6, and 13 were annotated as ‘Progenitor leukocytes’. Further characterization of these clusters according to step four of the annotation approach (**Figure 16**) revealed that these clusters mainly consisted of cells with high ribosomal gene content. However, cluster 6 was additionally characterized by the expression of the T-cell markers *IL7R*, *CD3D* and *CD3G*, but also by NK-cell markers (e.g. *NKG7* and *KLRG1*), suggesting that this cluster rather contains T and NK cells that represented an activated phenotype reflected by the upregulation of ribosomal genes. In contrast, cluster 13 contained erythrocytes, as indicated by the expression of hemoglobin genes (e.g. *HBA1*, *HBA2*, and *HBB*), whereas cluster 2 did not carry specific markers but was located in between several cell types (**Figure 33A+B**), which is why we annotated them as 'ribosomal-high mixed cells'.

The extension of the in-depth analysis to all identified main clusters according to the four-step cell-type annotation strategy led to a high-resolution annotation of the immune cells in the blood (**Figure 33D**).

4.8.4. Characterization of neutrophil subtypes in the blood

The analysis of scRNA-sq data identified monocytes, T cells, B cells, eosinophils and neutrophils as the main cell types in the blood, with the latter being characterized by high heterogeneity as indicated by cell clustering (**Figure 33A and Figure 34A**). Since the MCFC analysis showed a significant increase of neutrophils in the alveolar space

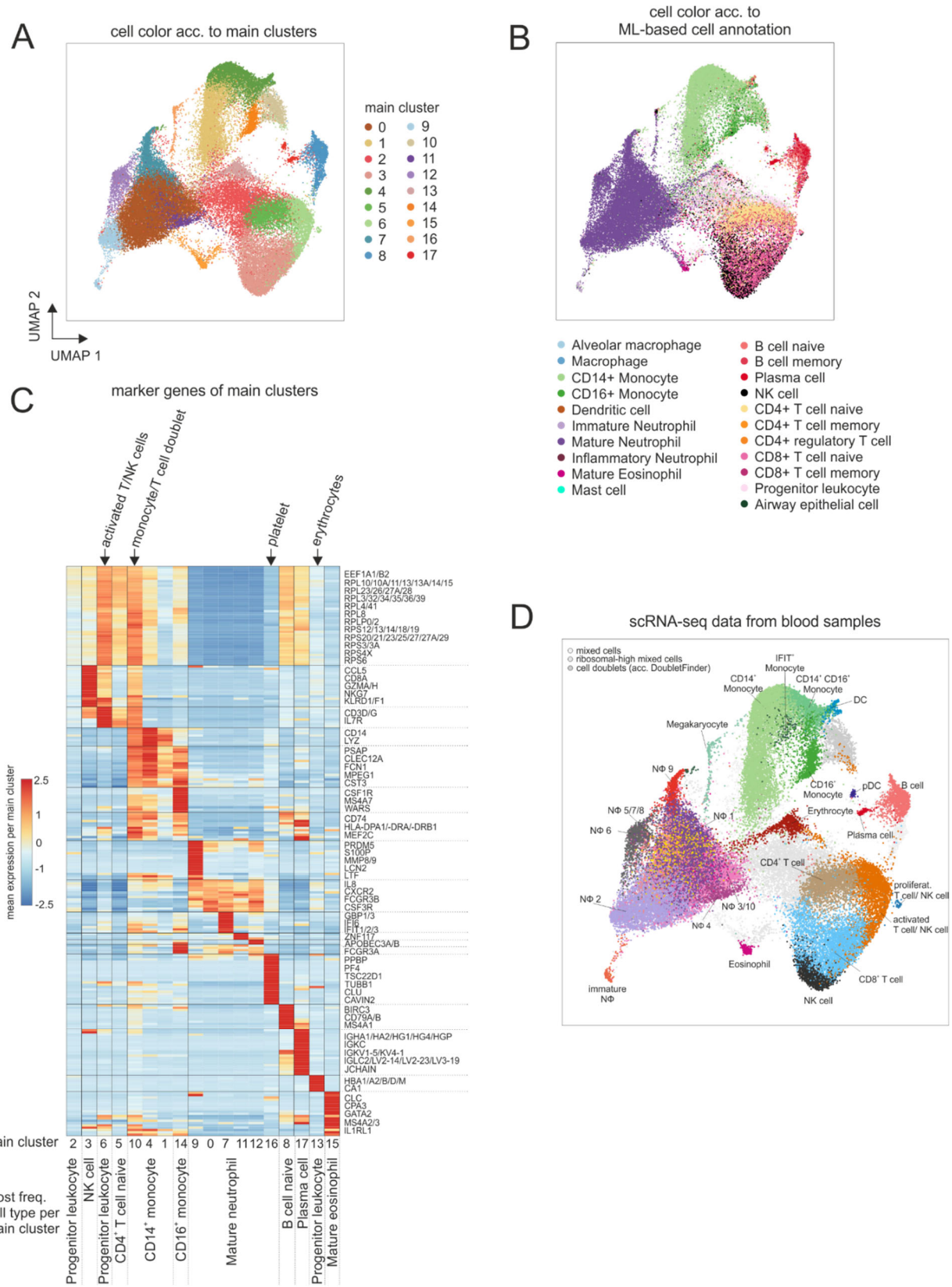

A
cell color acc. to main clusters
main cluster
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
UMAP 2
UMAP 1
B
cell color acc. to
ML-based cell annotation
Alveolar macrophage
Macrophage
CD14+ Monocyte
CD16+ Monocyte
Dendritic cell
Immature Neutrophil
Mature Neutrophil
Inflammatory Neutrophil
Mature Eosinophil
Mast cell
B cell naive
B cell memory
Plasma cell
NK cell
CD4+ T cell naive
CD4+ T cell memory
CD4+ regulatory T cell
CD8+ T cell naive
CD8+ T cell memory
Progenitor leukocyte
Airway epithelial cell
C
marker genes of main clusters
activated T/NK cells
monocyte/T cell doublet
platelet
erythrocytes
mean expression per main cluster
2.5
1
0
-1
-2.5
main cluster
2 3 6 5 10 4 1 14 9 0 7 11 12 16 8 17 13 15
most freq.
cell type per
main cluster
Progenitor leukocyte
NK cell
Progenitor leukocyte
CD4+ T cell naive
CD14+ monocyte
CD16+ monocyte
Mature neutrophil
B cell naive
Plasma cell
Progenitor leukocyte
Mature eosinophil
D
scRNA-seq data from blood samples
mixed cells
ribosomal-high mixed cells
cell doublets (acc. DoubletFinder)
IFIT+ Monocyte
CD14+ Monocyte
CD14+ CD16+ Monocyte
DC
Megakaryocyte
NΦ 9
CD16+ Monocyte
pDC
B cell
NΦ 5/7/8
NΦ 1
Erythrocyte
NΦ 6
Plasma cell
CD4+ T cell
proliferat. T cell/ NK cell
NΦ 2
NΦ 3/10
activated T cell/ NK cell
NΦ 4
Eosinophil
CD8+ T cell
immature NΦ
NK cell

Figure 33. Four-step cell-type annotation of blood immune cells

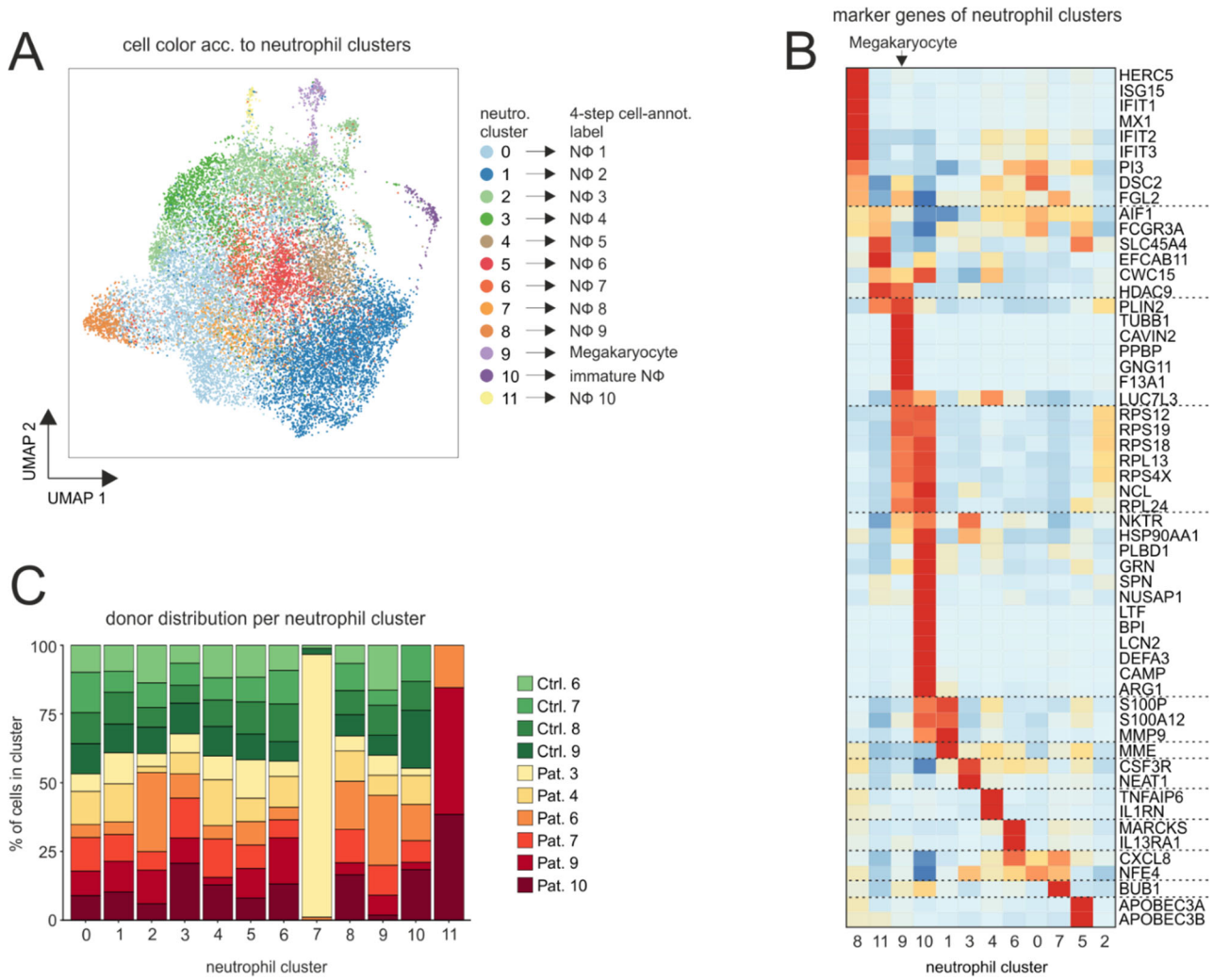
(A) UMAP representation of integrated blood data obtained from different COPD patients and control donors. Coloring according to identified main clusters. **(B)** UMAP representation of integrated blood data with coloring according to the machine learning-based cell-type annotation. **(C)** Heat map of the calculated marker genes per main cluster. The marker gene expression per cluster is represented as a z-transformed value (across all clusters). Rows of the heat map are sorted by hierarchical clustering. At the top of the plot, additional information on main clusters revealed by a combination of marker gene inspection and sub-clustering of cell types according to the four-step cell type annotation strategy (**Figure 16**). At the bottom of the plot, the main cell type of the respective main cluster (acc. to the machine learning-based cell-type annotation) is displayed. **(D)** Identified blood cell-types according to the four-step cell-type annotation approach.

of COPD patients (**Figure 9A**) and a clear description of the neutrophil subtypes in the blood is still pending, we used the in-depth analysis of the neutrophil clusters according to the four-step cell-type annotation strategy as an opportunity to further characterize these blood cells. In total, we found 12 neutrophil clusters, whereby cells within neutrophil cluster 9 expressed markers of platelets (e.g. *TUBB1*) and were thus identified as megakaryocytes/ platelets (**Figure 34B**). Interestingly, neutrophil cluster 2 showed no specific marker expression, while most other neutrophil clusters were characterized by only a few markers, such as neutrophil cluster 5, whose cells specifically expressed *APOBEC3A* and *APOBEC3B*. In contrast, neutrophil cluster 4 showed expression of genes associated with the IL-1 signaling pathway (*TNFAIP6* and *IL1RN*), while cluster 6 was associated with IL-4/IL-13 signaling (*IL13RA1*). In addition, cluster 6 shared the expression of the chemokine IL8 (*CXCL8*), with the neutrophil cluster 0 and 7. However, the latter had a clear donor bias, which also held true for neutrophil cluster 11 (**Figure 34C**). Neutrophil cluster 10 carried the largest number of specific markers, which comprised lysosomal maturation genes (*GRN*) and genes encoding proteins contained in neutrophil granules (*LTF*, *BPI* and *DEFA3*). Moreover, this cluster shared some marker genes with neutrophil cluster 1, including the matrix metalloprotease *MMP9*. Interestingly, according to the machine learning-based cell-type annotation, the neutrophil cluster 10 cells were assigned to immature neutrophils, while cluster 3 was annotated as mature neutrophils although its cells expressed the granulocyte maturation marker *CSF3R*, indicating an additional function of this receptor in mature neutrophils. Finally, the neutrophil cluster 8 (corresponded

to neutrophil 9 according to the four-step cell annotation) was characterized by high expression of interferon-associated genes (*IFIT1-3*). In addition, it was found that this cluster, together with the neutrophil cluster 3 (corresponded to neutrophil 4 according to the four-step cell annotation) and plasma cells, had the strongest tendency towards elevated relative cell numbers among all blood cells in COPD (**Figure 34D**).

4.8.5. Modelling the trajectory of blood monocytes into the alveolar space

The motivation for generating scRNA-seq blood data was to model and study the potential of blood monocytes to replenish AMs. Based on the four-step cell-type annotation strategy, we identified in our blood data the three known monocyte populations comprising classical monocytes ($CD14^+$ monocytes), intermediate monocytes ($CD14^+CD16^+$ monocytes) and non-classical monocytes ($CD16^+$ monocytes) along with a small monocyte population that expressed high numbers of interferon-associated genes ($IFIT^+$ monocytes) (**Figure 33D**). To evaluate a potential link between blood monocyte populations and alveolar space-derived monocytes and macrophages, we next modelled the possible trajectories of these cells by calculating the RNA velocities (Bergen et al., 2019; La Manno et al., 2018), which is a measure of expression dynamics in scRNA-seq data inferred from unspliced and spliced mRNAs detected by scRNA-seq technologies such as Seq-Well. However, current methods of RNA velocity analysis do not take into account potential batches within the dataset, and in addition, current integration methods have not yet been evaluated for use in RNA velocity analysis. Therefore, we used a strategy of combining the blood and BALF data while simultaneously considering donor batches by performing joint embedding based on highly variable genes shared by all patients (**Figure 35A**). While this approach allowed the combination of blood and BALF data and at the same time avoided potential over-fitting with current integration methods, we observed a reduced resolution of the defined AM states. Thus, we simplified the annotation by combining


A
cell color acc. to neutrophil clusters
neutro. cluster
4-step cell-annot. label
0 → NΦ 1
1 → NΦ 2
2 → NΦ 3
3 → NΦ 4
4 → NΦ 5
5 → NΦ 6
6 → NΦ 7
7 → NΦ 8
8 → NΦ 9
9 → Megakaryocyte
10 → immature NΦ
11 → NΦ 10
UMAP 2
UMAP 1
B
marker genes of neutrophil clusters
Megakaryocyte
HERC5
ISG15
IFIT1
MX1
IFIT2
IFIT3
PI3
DSC2
FGL2
AIF1
FCGR3A
SLC45A4
EFCAB11
CWC15
HDAC9
PLIN2
TUBB1
CAVIN2
PPBP
GNG11
F13A1
LUC7L3
RPS12
RPS19
RPS18
RPL13
RPS4X
NCL
RPL24
NKTR
HSP90AA1
PLBD1
GRN
SPN
NUSAP1
LTF
BPI
LCN2
DEFA3
CAMP
ARG1
S100P
S100A12
MMP9
MME
CSF3R
NEAT1
TNFAIP6
IL1RN
MARCKS
IL13RA1
CXCL8
NFE4
BUB1
APOBEC3A
APOBEC3B
8 11 9 10 1 3 4 6 0 7 5 2
neutrophil cluster
C
donor distribution per neutrophil cluster
% of cells in cluster
100
75
50
25
0
0 1 2 3 4 5 6 7 8 9 10 11
neutrophil cluster
Ctrl. 6
Ctrl. 7
Ctrl. 8
Ctrl. 9
Pat. 3
Pat. 4
Pat. 6
Pat. 7
Pat. 9
Pat. 10

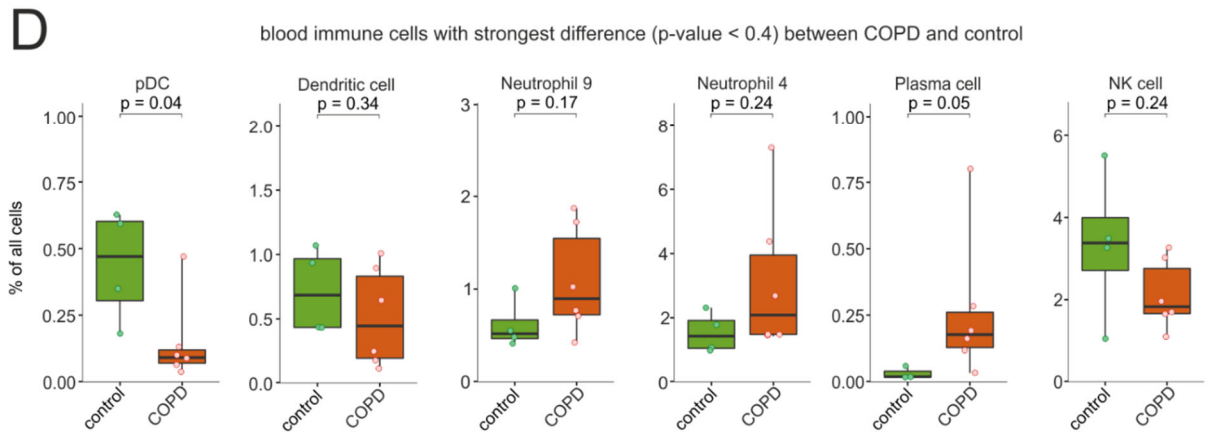

D
blood immune cells with strongest difference (p-value < 0.4) between COPD and control
% of all cells
pDC
p = 0.04
Dendritic cell
p = 0.34
Neutrophil 9
p = 0.17
Neutrophil 4
p = 0.24
Plasma cell
p = 0.05
NK cell
p = 0.24
control
COPD

Figure 34. Investigation of the blood neutrophil heterogeneity

(A) UMAP representation and clustering of cells found in main clusters 0, 7, 9, 11, 12, and 16 of the blood dataset (acc. to **Figure 33A**), which were annotated as neutrophil-containing clusters. In addition to the cluster numbers, the labels of the neutrophil clusters, as used in **Figure 33D**, are also displayed. **(B)** Heat map of calculated marker genes per neutrophil cluster (acc. to (A)). The marker gene expression is represented as a z-transformed value across all neutrophil clusters. Columns and rows of the heat map are sorted by hierarchical clustering. **(C)** Stacked bar plot showing the proportion of individual donors in each neutrophil cluster. **(D)** Box plots showing the frequencies of blood cell types, with the strongest difference in relative population size between COPD and control.

transcriptional similar AMs (**Figure 35A, dendrogram and combined labels**), resulting in an annotation that could well describe the structure of AMs in the embedded dataset (**Figure 35B, heatmaps**). Projection of the main average vector flow visualized by RNA-velocity streamlines calculated by the scVelo method (Bergen et al., 2019) onto the embedded data in a batch-corrected manner revealed a clear motion of blood monocytes towards the AMs (**Figure 35B**), further supporting circulating monocytes as precursors of AMs in the alveolar space. Since RNA velocity visualization on the UMAP did not provide sufficient resolution to reveal a clear link between individual AM states and blood monocyte populations, we calculated a higher-order representation using partition-based graph abstraction analysis (PAGA) (Wolf et al., 2019) (**Figure 35C, network**). As expected, the alveolar immune cells with the strongest connections to blood monocytes were alveolar space-derived monocytes. To evaluate the connectivity of the network more precisely, we next used the connectivity matrix of PAGA as a test statistic to assess whether two cell types from BAL and blood were connected more than expected at random. The monocytes within the alveolar space served as positive controls indicating very high relationships. However, importantly, among the AM states, monocyte-like macrophages possessed the highest test scores indicating the strongest connections to blood monocytes, with $CD16^{+}$ monocytes as the top-predicted connection partner in the blood (**Figure 35C, table**).

In summary, the generated models suggested that blood monocytes represented the precursor pool of monocyte-like macrophages in the alveolar space.

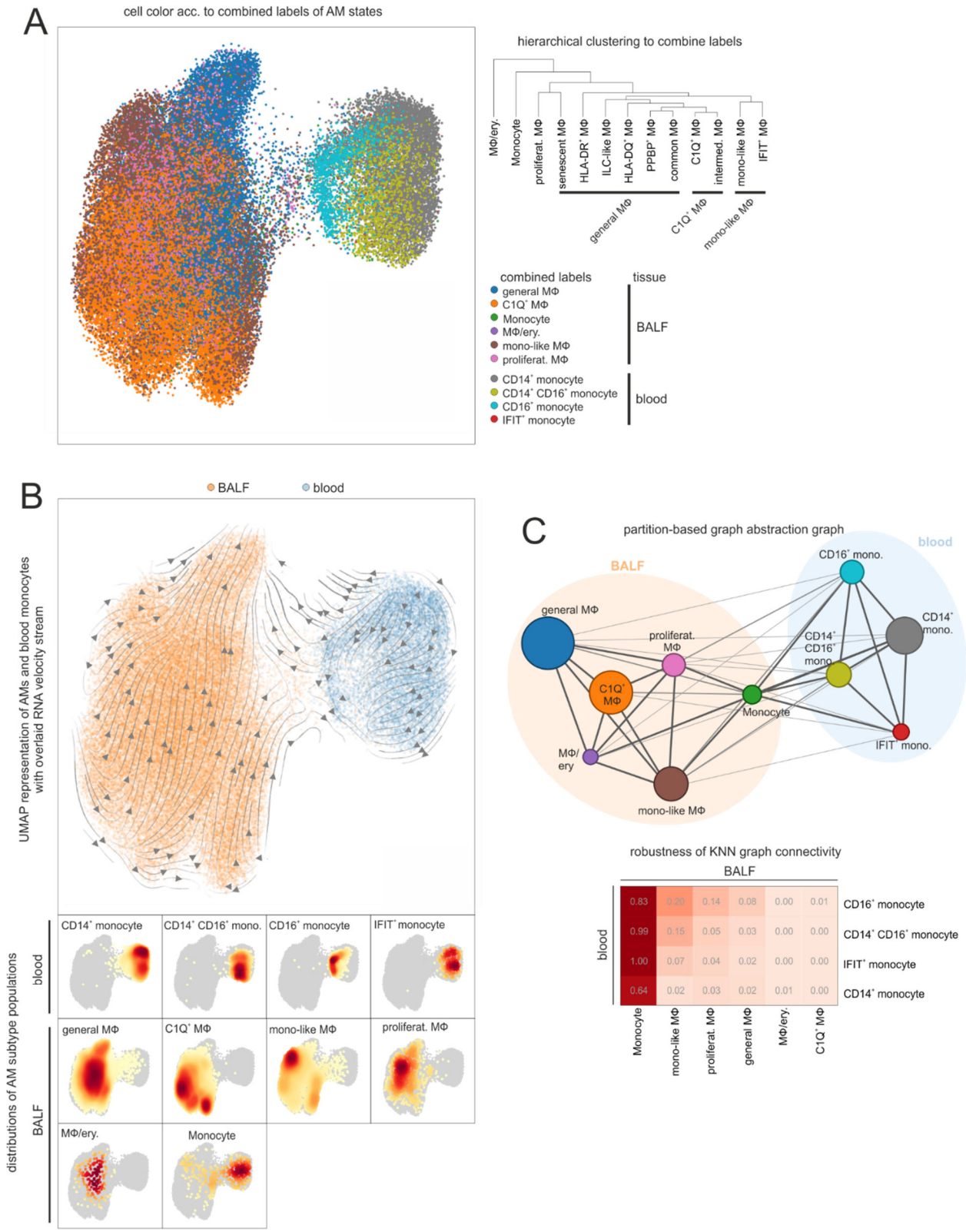

A
cell color acc. to combined labels of AM states
hierarchical clustering to combine labels
MΦ/ery.
Monocyte
proliferat. MΦ
senescent MΦ
HLA-DR+ MΦ
ILC-like MΦ
HLA-DQ+ MΦ
PPBP+ MΦ
common MΦ
C1Q+ MΦ
intermed. MΦ
mono-like MΦ
IFIT+ MΦ
general MΦ
C1Q+ MΦ
mono-like MΦ
combined labels
general MΦ
C1Q+ MΦ
Monocyte
MΦ/ery.
mono-like MΦ
proliferat. MΦ
CD14+ monocyte
CD14+ CD16+ monocyte
CD16+ monocyte
IFIT+ monocyte
tissue
BALF
blood
B
BALF
blood
UMAP representation of AMs and blood monocytes with overlaid RNA velocity stream
distributions of AM subtype populations
blood
BALF
CD14+ monocyte
CD14+ CD16+ mono.
CD16+ monocyte
IFIT+ monocyte
general MΦ
C1Q+ MΦ
mono-like MΦ
proliferat. MΦ
MΦ/ery.
Monocyte
C
partition-based graph abstraction graph
BALF
blood
CD16+ mono.
CD14+ mono.
general MΦ
proliferat. MΦ
CD14+ CD16+ mono.
C1Q+ MΦ
Monocyte
MΦ/ ery
IFIT+ mono.
mono-like MΦ
robustness of KNN graph connectivity
BALF
blood
0.83 0.20 0.14 0.08 0.00 0.01
0.99 0.15 0.05 0.03 0.00 0.00
1.00 0.07 0.04 0.02 0.00 0.00
0.64 0.02 0.03 0.02 0.01 0.00
CD16+ monocyte
CD14+ CD16+ monocyte
IFIT+ monocyte
CD14+ monocyte
Monocyte
mono-like MΦ
proliferat. MΦ
general MΦ
MΦ/ery.
C1Q+ MΦ

Figure 35. Modelling of differentiation trajectories between blood monocytes and AMs

(A) UMAP of embedded macrophages/monocytes from BALF and blood monocytes with coloring according to the cell types derived from the combined labels. The dendrogram on the right side illustrates the transcriptional relationship between the macrophage subtypes and shows how several subtypes were summarized in the combined labels. **(B)** UMAP of embedded macrophages/monocytes from BALF and blood monocytes. Inferred main average vector flow is indicated by velocity streamlines that are projected as vectors. The heat maps at the bottom indicate locations of the main cell types (acc. to the combined labels from (A)) in the UMAP. **(C)** PAGA graph derived from embedded BALF and blood data (acc. to (B)). The weight of an edge, which reflects a statistical measure of connectivity, is represented as the edge width. The table below summarizes the results of the PAGA connectivity calculation, where a value of 1 indicates a strong connection and 0 indicates a weak connection between two cell types.

4.8.6. Transcriptional alterations of blood monocytes from COPD patients

After identifying blood monocytes as a potential source of AM replenishment, we next investigated whether the COPD-associated transcriptional changes observed in AMs were restricted to cells within the lung or whether we could observe changes also in the blood. For this purpose, we again applied the GO-shuffling analysis, which showed that among the top genesets with the highest ability to separate control donors from COPD patients, $CD16^+$ monocytes contained the most immune-related terms (**Figure 36A**). Intriguingly, some of these genesets were associated with cellular extravasation and leukocyte adhesion to vascular endothelial cells, which was also reflected by the heat map representation of the expressed genes within the identified genesets, which revealed upregulation of integrins (*ITGB1*, *ITGB2*, *ITGA4* and *ITGAL*) in COPD (**Figure 36B**).

Our next step was to investigate whether the identified DE genes in AMs from COPD patients were already altered in blood monocytes (**Figure 36C**). Clearly, in blood monocytes derived from COPD patients, we detected a strong enrichment of DE genes found as upregulated in COPD AMs, while blood monocytes of controls showed an enrichment of downregulated genes (**Figure 36C**). Interestingly, we observed the strongest enrichments of DE genes, which were found as upregulated in COPD, in $CD14^+CD16^+$ and $CD16^+$ monocytes. In addition, visualization of MHC I, MHC II and mitochondrial genes, which were found to be transcriptionally different between COPD

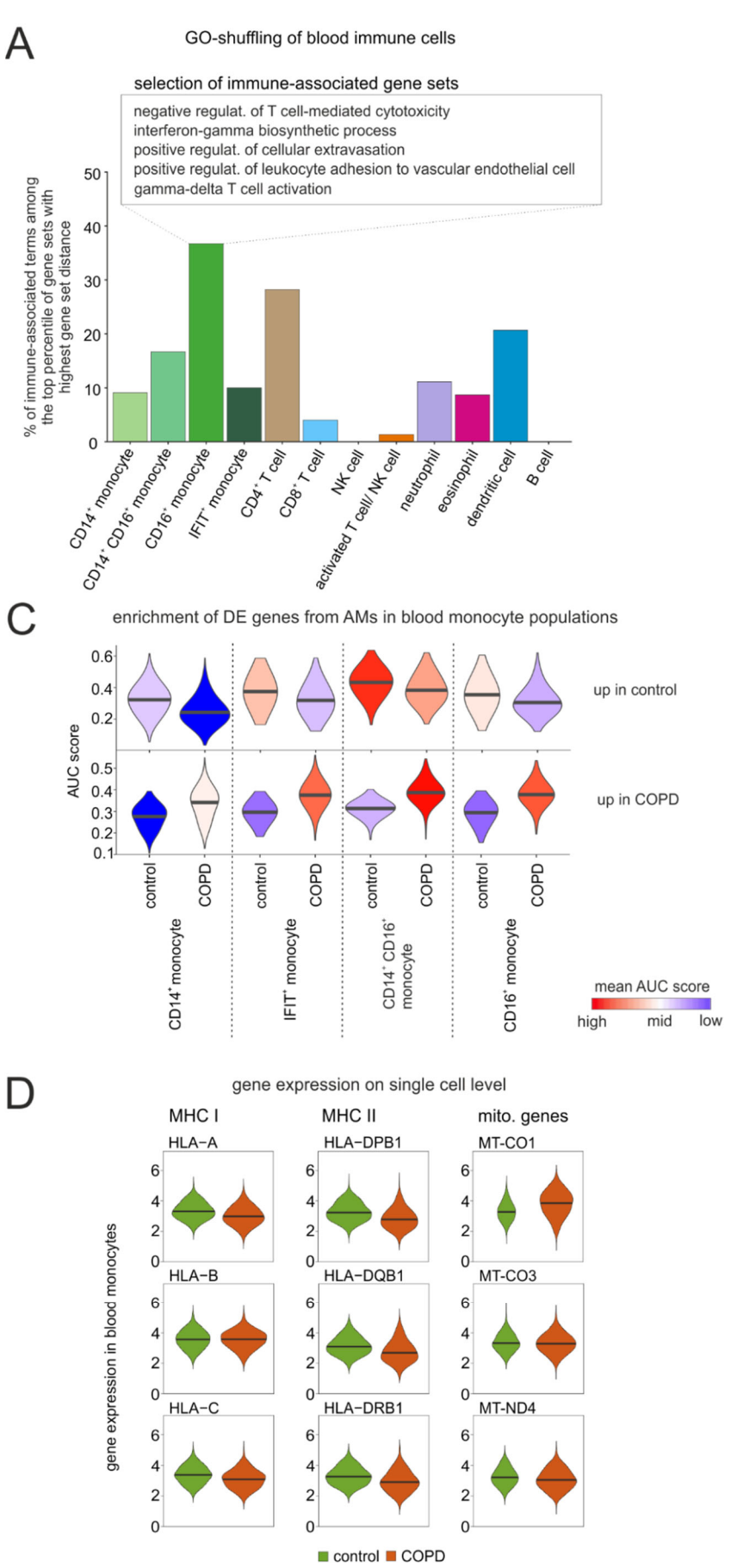

A
GO-shuffling of blood immune cells
selection of immune-associated gene sets
negative regulat. of T cell-mediated cytotoxicity
interferon-gamma biosynthetic process
positive regulat. of cellular extravasation
positive regulat. of leukocyte adhesion to vascular endothelial cell
gamma-delta T cell activation
% of immune-associated terms among the top percentile of gene sets with highest gene set distance
CD14+ monocyte
CD14+ CD16+ monocyte
CD16+ monocyte
IFIT+ monocyte
CD4+ T cell
CD8+ T cell
NK cell
activated T cell/ NK cell
neutrophil
eosinophil
dendritic cell
B cell
B
genes involved in immune-associated gene sets in blood CD16+ monocytes (acc. to A)
ETS1
TGFB1
HLA-DMB
PTGER4
BTN3A2
PTPRJ
RAPGEF1
HSPD1
HMOX1
BST2
LILRB2
CEBPG
HLA-F
TGFBR2
TLR8
IFIH1
TMBIM6
ITGA4
LILRB1
MAPK1
RIPOR2
IRAK3
PRKDC
PTPRC
LYN
RHOA
PTAFR
STAT6
ITGAL
NDFIP1
CD4
VIM
CD86
TLR4
BST1
TRIM27
CD36
JAML
IL27RA
NCKAP1L
RIF1
DDX58
SYK
GBP5
RAP1A
CX3CR1
ITGB2
BTN3A1
DHX9
DUSP22
IL7R
STAT5B
LGALS3
ITGB1
LGALS9
Ctrl. 6
Ctrl. 7
Ctrl. 8
Ctrl. 9
Pat. 7
Pat. 3
Pat. 4
Pat. 9
Pat. 6
Pat. 10
mean expression per donor
control
COPD
C
enrichment of DE genes from AMs in blood monocyte populations
AUC score
up in control
up in COPD
control
COPD
CD14+ monocyte
IFIT+ monocyte
CD14+ CD16+ monocyte
CD16+ monocyte
mean AUC score
high
mid
low
D
gene expression on single cell level
MHC I
MHC II
mito. genes
HLA-A
HLA-B
HLA-C
HLA-DPB1
HLA-DQB1
HLA-DRB1
MT-CO1
MT-CO3
MT-ND4
gene expression in blood monocytes
control
COPD

Figure 36. Assessment of immune-related alterations in blood monocytes from COPD patients

(A) Bar plot showing for blood immune cell types the proportion of predicted immune-related GO terms within the upper percentile of GO terms with the highest ability to separate COPD patients from control donors according to GO-shuffling. A selection of immune-related GO terms found in $CD16^+$ monocytes is depicted in the box at the top. **(B)** Heat map of immune system-associated genes in $CD16^+$ monocytes predicted by the GO-shuffling approach. The mean gene expression per donor is represented as a z-transformed value (across all donors). Columns and rows of the heat map are sorted by hierarchical clustering. **(C)** Violin plots displaying enrichment of AM-related DE genes (acc. to **Figure 24B+C**) in blood monocytes based on the 'Area Under the Curve' (AUC). **(D)** Violin plots of the expression of genes, found as DE in AMs, in blood monocytes based on scRNA-seq data. The plots show the expression across the donors, whereby the donors were downsampled to the same number of cells, followed by downsampling to the same number of cells between COPD and control. The plots display cells with an expression > 0.

and control AMs (**Figure 24-26**), confirmed that blood monocytes at least show tendencies for similar transcriptional differences (**Figure 36D**).

Taken together, the investigation of transcriptional changes in blood monocytes revealed changes similar to AMs, suggesting a systemic component of COPD that leads to transcriptional changes in circulating immune cells such as monocytes.

4.9. Identification of monocyte progenitor cells in murine lungs

The observation of similar COPD-associated transcriptional changes in monocytes and AMs raises the question of how chronic inflammation of the lung can become systemic. To gain a better overview at which point of the monocytes lifecycle the transcriptional changes might occur, we investigated scRNA-seq data obtained from mouse monocyte progenitors and $ly6c^+$ monocytes (murine counterpart of human $CD14^+$ monocytes) isolated from different tissues (bone marrow, blood, lung, intestine, and skin) (**Figure 37A**). When analyzing the cell cycle states of the isolated cell populations, it became evident that the cells isolated from the bone marrow had a relative high percentage of cells in pro-mitotic phases (S, G2 or M phase), whereas monocytes isolated from tissues were exclusively non-proliferative (G1 phase) (**Figure 37B**). Intriguingly, the only exception to this observation were some monocytes isolated from mouse lungs, which were also in mitotic phases. Next, we constructed

the developmental pathway of monocytes using Monocle (Trapnell et al., 2014), which showed that the trajectory was characterized by several branches, with MDPs being the starting point of monocyte development, while skin monocytes lined up at the other extreme (**Figure 37C**). In total, we were able to identify 8 cellular states, which were mainly defined by the branches. For example, cellular state 1 included MDPs and cMoPs, where it was remarkable that there was a considerable gap between these two populations within the trajectory. Interestingly, skin and intestinal monocytes were enriched in one or two branches, while lung monocytes, together with blood and BM monocytes, could not be assigned to a specific branch, but rather represented a connection between different branches. Importantly, we also found some lung monocytes that aligned with cMoPs in cell state 1, which was striking because cells within this cell state carried clear precursor cell markers such as *Cd34*, *Kit*, *Myb*, and *Myc* (**Figure 37D**). Indeed, when plotting the lung monocytes aligning with cMoPs, we found a strong enrichment of *Cd34* and *Kit*, indicating that the lung contains monocyte progenitor cells (**Figure 37E**).

Based on these findings, it is possible that the identified extramedullary pool of monocyte progenitor cells directly at the lung is altered during COPD pathogenesis leading to the maturation of monocytes with transcriptional changes.

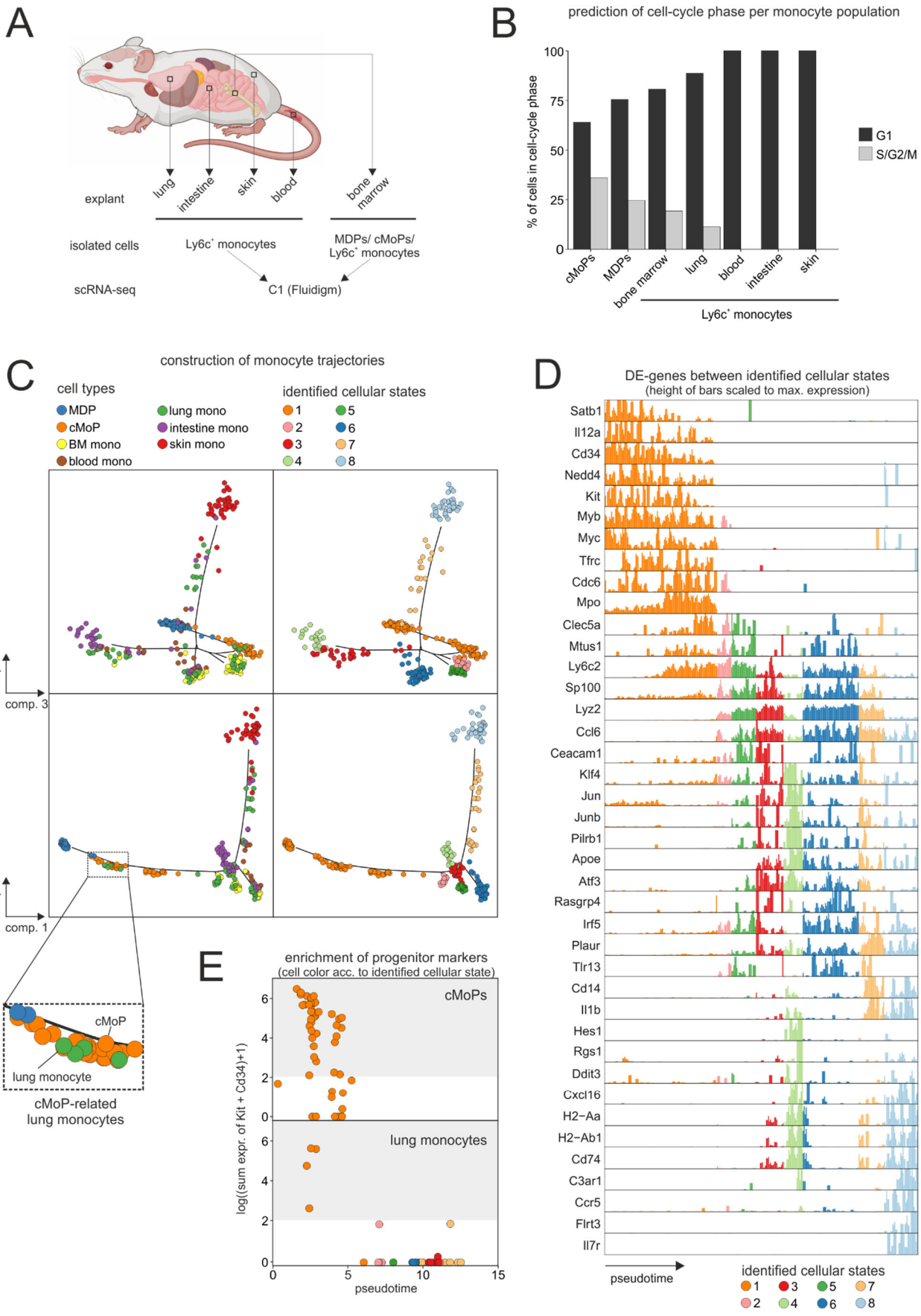

A
explant
lung
intestine
skin
blood
bone marrow
isolated cells
Ly6c+ monocytes
MDPs/ cMoPs/ Ly6c+ monocytes
scRNA-seq
C1 (Fluidigm)
B
prediction of cell-cycle phase per monocyte population
% of cells in cell-cycle phase
100
75
50
25
0
G1
S/G2/M
cMoPs
MDPs
bone marrow
lung
blood
intestine
skin
Ly6c+ monocytes
C
construction of monocyte trajectories
cell types
MDP
cMoP
BM mono
blood mono
lung mono
intestine mono
skin mono
identified cellular states
1
2
3
4
5
6
7
8
comp. 2
comp. 3
comp. 2
comp. 1
cMoP
lung monocyte
cMoP-related lung monocytes
D
DE-genes between identified cellular states
(height of bars scaled to max. expression)
Satb1
Il12a
Cd34
Nedd4
Kit
Myb
Myc
Tfrc
Cdc6
Mpo
Clec5a
Mtus1
Ly6c2
Sp100
Lyz2
Ccl6
Ceacam1
Klf4
Jun
Junb
Pilrb1
Apoe
Atf3
Rasgrp4
Irf5
Plaur
Tlr13
Cd14
Il1b
Hes1
Rgs1
Ddit3
Cxcl16
H2−Aa
H2−Ab1
Cd74
C3ar1
Ccr5
Flrt3
Il7r
pseudotime
identified cellular states
1
2
3
4
5
6
7
8
E
enrichment of progenitor markers
(cell color acc. to identified cellular state)
cMoPs
lung monocytes
log((sum expr. of Kit + Cd34)+1)
6
4
2
0
6
4
2
0
0
5
10
15
pseudotime

Figure 37. Identification of monocyte progenitors in murine lungs

(A) Schematic representation of the mouse tissues from which monocytes and monocyte precursors (MDPs and cMoPs) were isolated and subsequently used for scRNA-seq. **(B)** Bar plot representation of the proportion of cells in the respective cell cycle states per isolated precursor and monocyte population. **(C)** Differentiation trajectory of monocytes constructed by Monocle (Trapnell et al., 2014). Top panel shows the second and third component of the trajectory, whereas bottom panel shows the first and second component. Cell coloring on the left side according to the isolated precursor and monocyte populations. Cell coloring on the right side corresponding to identified cell states, which are mainly determined by branching points. A zoom into an area containing both cMoPs and lung monocytes is shown in the box at the bottom of the plot. **(D)** Expression of genes found as being DE along the pseudotime of the constructed differentiation trajectory (acc. to (C)). The cells were ordered along the pseudotime and the y-axis represents the expression of each gene as bars, whereby each bar represents an individual cell. Coloring of the bars corresponding to the identified cell states (acc. to (C)). **(E)** Expression of the cumulative expression of the precursor-associated genes Kit and Cd34 in cMoPs and lung monocytes. Cells were ordered along the pseudotime.

5. Discussion

COPD is a lung disease with a high global spread, increasing incidence, prevalence, morbidity and mortality, mainly due to rising air pollution and high smoking rates worldwide (Rabe and Watz, 2017). However, the cellular and molecular mechanisms underlying the pathophysiology of this disease are poorly understood (Rabe and Watz, 2017), thus leading to a lack of effective therapies (Barnes et al., 2015). In the present study, we investigated the immune system in the alveolar space and the blood of COPD patients using high-resolution single-cell technologies. We focused on patients who were diagnosed with an early stage of the disease (GOLD 2) because it was hypothesized that pharmacological interventions at earlier stages have the highest chance of beneficial clinical outcomes. (Sun and Zhou, 2019). To enable the best possible characterization of COPD-associated changes, we used a scRNA-seq technology capable of detecting the major immune cell types in alveolar space and blood, and developed or adapted advanced computational approaches including cell-type classification based on machine learning, prediction of metabolic changes, identification of differently expressed genes and functionalities, and modeling of cell trajectories within a patient cohort. These approaches revealed in AMs of COPD patients: alterations in lipid metabolism, reduced expression of MHC class I molecules, and potential TGF-β involvement in DE-gene regulation. In addition, we found that proliferating and mono-like macrophages were elevated in COPD, with evidence that the latter were derived from blood monocytes. Finally, we investigated the development of monocytes isolated from different tissues of mice, which showed that a fraction of the monocytes from the lung showed progenitor signatures.

5.1. Comparison of the Chromium and Seq-Well technology

To obtain the best possible representation of the immune compartment in the alveolar space on the transcriptional level, we evaluated two different scRNA-seq technologies, namely the array-based Seq-Well technology (Gierahn et al., 2017) and the widely used 10x Chromium (v2), which is a droplet-based method (Zheng et al., 2017). In accordance with the statistics of our data, a recently performed systematic comparison

of different scRNA-seq methods (Ding et al., 2020) revealed that, among all high-throughput technologies, 10x Chromium (v2) detected the most UMIs and genes per cell and took the least time for library generation. In contrast, the Seq-Well technology detected fewer UMIs and genes, but also had by far the lower costs per cell. Moreover, an advantage of the Seq-well technology is that the cells are exposed to less stress during loading of the arrays (Gierahn et al., 2017) compared to the shear forces experienced when loading onto the microfluidic instrument of the 10x Chromium technology. Interestingly, considering the MCFC data, we found that Seq-Well reliably detected the major cell types in the alveolar space, while 10x Chromium (v2) performed inferiorly in detecting granulocytes. This finding is consistent with the observations of Travaglini et al., who claimed that neutrophils in the lungs cannot be detected with 10x Chromium (Travaglini et al., 2019). On the other hand, one of the first published high-throughput scRNA-seq datasets of human lungs was also generated by the Chromium technology (v2) and reported on neutrophil populations (Vieira Braga et al., 2019). However, these neutrophils were annotated based on the expression of genes such as *FCN1*, *VCAN* and *S100A8*, which we found to define monocytes rather than neutrophils, suggesting that the putative neutrophils were incorrectly annotated. On the other hand, several scRNA-seq papers reported on neutrophils in mouse using 10 Chromium (v2) (Park et al., 2018; Ponzetta et al., 2019; Vafadarnejad et al., 2019; Xie et al., 2019). This apparent discrepancy could be explained by the fact that the granulocyte compartment between mice and humans differs substantially, which is reflected, for example, in the percentage of neutrophils in the blood, which is 50-70% in humans but only 10-25% in mice (Mestas and Hughes, 2004). In addition, human neutrophils are a rich source of defensins, whereas defensins are not expressed by neutrophils in mice (Eisenhauer and Lehrer, 1992). It is therefore likely that the content or activity of RNAses in the granules of neutrophils also differs between the two species, which in turn might lead to the different efficiency in scRNA-seq profiling of human or murine neutrophils. In fact, the recently published 10x Chromium (v3) technology showed a higher sensitivity in transcript detection compared to its predecessor (Ding et al., 2020b), which, according to the manufacturer, was achieved by changing the chemistry of the reactions that probably also includes a higher content of RNases in the enzyme mix. Indeed, newer attempts using 10x Chromium (v3) for BALF samples suggest that human neutrophils can be

detected (Liao et al., 2020), but it remains unclear, whether all granulocytes were truly captured in this study.

5.2. Neutrophil compartment resolved by Seq-Well

The apparent inability to detect granulocytes with 10x Chromium (v2) is extremely problematic in the study of COPD as our MCFC analysis clearly showed that neutrophils are the only cell type in which the relative cell number in the BALF of COPD patients is significantly increased, and the importance of neutrophils in this disease has been reported in several studies before (Barnes, 2019; Hoenderdos and Condliffe, 2013; Kinose et al., 2016; Lee et al., 2016; Noguera et al., 2001). In contrast, we reliably detected granulocytes with Seq-Well, which revealed that especially in the blood the neutrophil population is characterized by an unprecedented high heterogeneity. For example, we discovered immature neutrophils, which we have also recently detected in increased numbers in COVID-19 patients with severe disease course, indicating emergency myelopoiesis (*unpublished data*). However, in the blood of non-exacerbating COPD patients we found no change in the cell numbers of immature neutrophils compared to controls, but in contrast, we observed an increase in a population of neutrophils associated with an interferon-associated gene program. The interferon signature in neutrophils was recently linked to activation of these cells (Rocha et al., 2015) and was associated with elevated ROS production (Wright et al., 2008). Moreover, type I interferon-signaling pathways are found to be associated with neutrophil-extracellular trap (NET) formation (Pylaeva et al., 2016), which is in agreement with the detected enrichment of NETs in neutrophils from COPD patients (Uddin et al., 2019). Interestingly, NETs can readily form a platform for thrombus formation (de Bont et al., 2019), which is especially important in the context of thromboembolism, which is a major threat to patients during acute exacerbations (Ambrosetti et al., 2003), where it is likely that the production of type I interferon is increased in response to an elevated viral burden. To better understand this potential relationship, it is important to study neutrophils associated with an interferon-associated gene program during exacerbations in future scRNA-seq studies. In the present study, we found that the identified neutrophil populations in the lung were

associated with the expression of genes for antigen presentation (e.g. *CD74*, *HLA-DRB1*, *HLA-DPB1*, *HAL-DRA*), which is particularly interesting in the context of increased neutrophil numbers in the BALF of COPD patients and the observed downregulation of molecules for antigen presentation in AMs. This observation is in agreement with the emerging body of evidence that neutrophils can acquire the function of antigen-presenting cells under pathological conditions (Lin and Loré, 2017). However, the importance of antigen-presenting neutrophils in the pathophysiology of COPD remains elusive.

5.3. Addressing analytical challenges in scRNA-seq data of COPD patients

To investigate, for example, the neutrophils in detail, the first step of the analysis was to identify the cells belonging to the respective cell types. This cell-type annotation step is of particular importance when it is intended to answer clinically relevant questions using scRNA-seq data, since contamination of a cell population by incorrectly annotated cells would make downstream analysis difficult or even mislead the results. However, the correct annotation of cell types is still a major problem in the field of scRNA-seq analysis (Abdelaal et al., 2019). To overcome this analytical limitation, several methods based on machine learning have recently been developed that were successfully used for annotating cells (Alquicira-Hernandez et al., 2019; Hou et al., 2019; Lopez et al., 2018; Song et al., 2019). Surprisingly, a comparison showed that the different machine learning methods hardly represented a qualitative difference, because most of them performed similarly well (Abdelaal et al., 2019; Köhler et al., 2019). Still, a major problem in cell-type classification is the access to well curated and reliably annotated training data. To address these limitations, we combined several methods based on different reference datasets to obtain multiple cell-type labels per cell. These multi-labeled cells were then used in a cross-validation approach to train a gradient boosting classifier to predict the most likely cell type for each cell. During the analysis of the present study, the idea of combining the results of several cell labels in a single annotation was also addressed by other groups and is, for example, provided in the latest SingleR version (Aran et al., 2019). However, as

with all classification methods, the classification accuracy of current cell-type annotation methods is hardly 100%, especially when it is based on complex, sparse and noisy data such as those obtained by scRNA-seq. Therefore, the incorrectly classified cells could still tamper with subsequent analysis results. In the present study, we circumvent these limitations by using a simple four-step cell-type annotation approach, in which the classification results are combined with unsupervised clustering of scRNA-seq data. It is important to note that clustering of data is also an area that is heavily debated and continuously leads to new and improved methods, some of which are particularly well suited to certain types of scRNA-seq data (Kiselev et al., 2019). Moreover, especially in the case of poorly defined biospecimens, such as BALF, it is difficult to decide at what point a cluster actually represents a new cell type or only a cellular state (Kiselev et al., 2019). However, in combination with machine learning-based cell-type classification, also known as supervised clustering (Lee and Hemberg, 2019), this uncertainty can be overcome and simultaneously, as done in this study, misclassified cells can be identified and excluded from further analysis.

Another remaining challenge in the scRNA-seq field is the identification of DE genes within a cohort setting, here COPD and controls, with multiple biological samples (patients) per group. Comparisons of current DE-Gene calculation tools showed that methods specifically designed for scRNA-seq data do not show significantly better performance compared to methods designed for bulk RNA-seq data (Wang et al., 2019). Therefore, many current methods are based on the generation of "mini-bulk" gene expression values for defined clusters of cells, which are then used to compare two groups, e.g. COPD and controls. Although this approach is computationally straightforward and thus offers a fast runtime, a major disadvantage of these methods is the loss of the single-cell gene information. Moreover, it is difficult to determine appropriate significance levels for these methods because there are several sources of variability between biological samples that are difficult to emulate *in silico* when working with ‘mini-bulks’. To address these limitations, we developed a novel DE-analysis approach that considered the influence of individual biological samples, exploited the information of single cells, but avoided assumptions about data distribution and determined the significance of the observed effects using a permutation test. Such tests require large parameter spaces and were therefore not considered in previous benchmark studies for comparisons with multiple samples

(Crowell et al., 2019). However, the sufficiently large number of biological samples (patients) included in our study enabled and facilitated the use of such a method (Ernst, 2004).

5.4. Cholesteryl ester accumulation in AMs of COPD patients

By applying the aforementioned computational innovations, we were able to resolve the transcriptional heterogeneity of the AM population, which constituted the most prominent cell type in the alveolar space. It is important to note that the term 'alveolar macrophage' is a generic term, because although bronchoalveolar lavage mainly includes the collection of cells from the alveolar space, it also collects some cells from the bronchioles (Heron et al., 2012). However, for the sake of simplicity, the macrophage compartment of BALF is collectively referred to as AMs in the present study. The transcriptional analysis revealed an altered lipid metabolism in AMs of COPD patients with COPD GOLD grade 2, which was particularly characterized by an increased accumulation of cholesteryl ester. As already outlined in the introduction, all tissue macrophages possess ‘accessory’ functions by which they support the parenchymal cells (Okabe and Medzhitov, 2016). In the case of AMs, this ‘accessory’ function includes surfactant homeostasis in the lung (Remmerie and Scott, 2018), which is why these cells are highly associated with lipid metabolism. A malfunction of lipid metabolism in AMs is associated with the clinical manifestation of PAP, in which the accumulation of cholesterol leads to a morphological change of AMs, whereby the macrophages become foam cell-like (Sallese et al., 2017). These pulmonary foam cells were also detected in other chronic lung diseases (Basset-Léobon et al., 2009) and recent studies found also a link between lipid accumulation in lung macrophages and smoking (Morissette et al., 2015; Wilson et al., 2011). Although, as already outlined above, the molecular pathways might be different compared to PAP, the described accumulation of cholesteryl ester in AMs indicates increased pulmonary foam cell formation also in COPD patients. A recently performed genome-wide association study identified mutations in a gene associated with the high-density lipoprotein (HDL) complex, which transports cholesterol in the blood, and found clear correlation between high HDL levels, lower Tiffeneau-Pinelli index, and more

extensive emphysema formation among COPD patients (Burkart et al., 2014). In contrast, other studies claimed that rather the low-density lipoprotein (LDL) complex in the blood of COPD patients is elevated (Krakhmalova and Hetman, 2013; Zafirova-Ivanovska et al., 2016). Nevertheless, all studies together point towards an increase in cholesterol transport in COPD and thus indicate a higher systemic cholesterol level. However, it is not yet known whether this increase in cholesterol is also one of the driving forces for the formation of pulmonary foam cells. Cholesterol accumulation and foam cell formation are also a hallmark of macrophages in atherosclerosis, and interestingly, antiatherosclerotic treatment with statins also showed positive effects in patients suffering from both atherosclerosis and COPD, with a reduction in COPD-associated exacerbations and mortality (Ingebrigtsen et al., 2015). This observation is particularly interesting since atherosclerosis is a common comorbidity in COPD (Sevenoaks and Stockley, 2006). Therefore, it is important to assess in future projects whether the well-studied molecular pathways underlying foam cell formation in atherosclerotic plaques could also explain the observed alterations in alveolar macrophages upon COPD.

5.5. Alveolar macrophage replenishment by blood monocytes

Another aspect that COPD and atherosclerosis have in common is the increased recruitment of monocytes to the site of inflammation. A recently proposed model (Guilliams and Scott, 2017) suggested that under homeostatic conditions the population of tissue-resident macrophages is maintained by self-renewal via proliferation, while the recruitment of monocytes from the circulation is rather limited (Scott et al., 2016). However, during inflammation, the pool of self-renewing macrophages in tissue is additionally supported by a substantial proportion of blood-derived monocytes to maintain the tissue macrophage population (Hashimoto et al., 2013b). In accordance with this model, we still detected local proliferation of a small fraction of AMs in COPD, but additionally, we found evidence of AM replenishment by blood-derived monocytes. More specifically, mono-like and C1Q^{+} AMs showed a strong enrichment of monocyte-derived macrophage signatures, and RNA velocity-based analysis supported a differentiation process between blood monocytes,

especially the CD16$^+$ subset, and the mono-like macrophages. These findings are supported by a recently published study demonstrating that mouse CD16$^+$ blood monocytes can effectively differentiate into lung macrophages (Schyns et al., 2019). Another line of evidence comes from murine smoking models illustrating that Cx3cr1$^+$ macrophages expand in response to cigarette smoke (Lee et al., 2012) and exhibit a profibrotic function (Aran et al., 2019). Cx3cr1$^+$ cells are also found in atherosclerosis and have been shown to be associated in mice with Ly6C$^-$ monocytes, the murine counterpart of the human CD16$^+$ monocytes (Moore et al., 2013), which makes it possible that the Cx3cr1$^+$ macrophages in the lung also have a similar association. Interestingly, COPD patients with a α1-antitrypsin deficiency showed a strong decrease in the CD16$^+$ monocyte pool compared to the CD14$^+$ monocytes in the blood, but further analyses are needed to determine whether this reduction is associated with an increased migration of these cells into the lungs. (Stolk et al., 2019).

5.6. Identified deregulated paths and their potential associations

Our scRNA-seq analysis of mono-like macrophages identified increased NOTCH, WNT, TGF-β and TNF signaling in COPD, which again indicated increased differentiation of monocytes in AMs, as this process has been recently associated with TGF-β (Yu et al., 2017) and NOTCH signaling (Bonnardel et al., 2019). Elevated NOTCH signaling has already been observed in alveolar epithelial cells of patients with chronic respiratory diseases such as pulmonary fibrosis (Reyfman et al., 2019), whereas WNT signaling in human epithelial cells of COPD patients are rather downregulated, which is associated with a decrease in their repair capacity (Skronska-Wasek et al., 2017). In macrophages, the importance of NOTCH signaling in COPD pathogenesis was especially recognized in the context of the NOTCH ligand DNER, for which SNPs were found that were correlated with a higher risk of developing COPD (Busch et al., 2017; Hancock et al., 2012). Moreover, DNER induces IFN-γ production in both cigarette-exposed animals and COPD patients (Ballester-López et al., 2019) and could thus contribute to the inflammatory process in the lungs of COPD patients and the destruction of lung tissue. The upstream ligand of another predicted pathway, namely TGF-β, has also been described in the context of COPD. Some studies

reported on higher expression of this cytokine in diseased airway epithelial cells (de Boer et al., 1998), which supports our observed tendency for elevated protein detection of LAP TGF-β1 in the alveolar space. Furthermore, our analysis at the single cell level revealed that in the alveolar space some immune cells, namely mast cells, mono-like and C1Q^{+} macrophages, might also contribute to the increased expression levels of TGF-β. Importantly, TGF-β is a heavily debated molecule in the COPD field, and a recently published study reported on the overall downregulation of TGF-β in different lung segments of COPD patients (Di Stefano et al., 2018). However, the authors of this study also discussed their results in the context of previous publications and suggested that the different antibodies that were used and that have varying specificity in the detection of TGF-β could explain the discrepancies. Nevertheless, when considering the supposedly conflicting reports on downregulation of TGF-β expression in COPD lungs, the predicted TGF-β signaling could also be explained by an induction of TGF-β signaling already in blood monocytes that later differentiate into mono-like macrophages. Indeed, TGF-β levels in the blood of COPD patients are significantly elevated compared to healthy controls (Mak et al., 2009). The signaling of TGF-β was recently found to crosstalk with NOTCH and WNT pathways in non-small cell lung cancer (Li et al., 2011; Ohnuki et al., 2014), however, this has not yet been reported in the context of COPD.

5.7. Aged immune cells in COPD

Some of the pathways predicted to regulate DE genes in mono-like macrophages might also be involved in the induction of immunosenescence in COPD. For example, NOTCH, WNT and TNF signaling can induce mTOR signaling (Chan et al., 2007; Saxton and Sabatini, 2017), which was recently associated with cellular senescence in lung cells (Houssaini et al., 2018). As older people have a higher prevalence of developing COPD (Halbert et al., 2006), cellular aging, which in turn is associated with senescence, could also be a hallmark of the disease (Barnes, 2017). Features of cellular senescence include an increase in the number of mitochondria and mitochondrial dysfunction, which is characterized by increased proton leakage and an associated elevation in ROS production (Gorgoulis et al., 2019; Korolchuk et al.,

2017). Oxidative stress due to increased ROS production is known in COPD, and there is in fact evidence that this is partly attributable to mitochondrial dysfunction (Bowler et al., 2004; L et al., 2012; McGuinness and Sapey, 2017; Ryter et al., 2018). In line with cellular senescence, we found an increased expression of mitochondrial genes and proton leakage in the mitochondria of AMs in COPD patients. Additionally, the observed reduction in the chemotactic capacity of AMs in COPD might also be a result of cellular aging (Oishi and Manabe, 2016; Shaw et al., 2013; Solana et al., 2012). However, dysregulation of macrophage chemotaxis has also been described as a consequence of smoking (Berg et al., 2016) and in lung cancer (Lemarie et al., 1984). Reduced migration capacity of AMs can have deleterious consequences for the alveoli, as it impairs the efficient removal of pollutants from the alveolar space, which in turn can lead to cell death of alveolar cells and the induction of inflammation. In addition, the clearance of the alveolar space is further deteriorated due to the reduced phagocytosis ability of AMs in COPD (Taylor et al., 2010).

5.8. Reduced MHC expression and potential link to monocyte progenitor pool in the lung

In the present study, we could also identify another layer of macrophage dysfunction in COPD, namely a significant downregulation of molecules associated with antigen presentation, especially MHC class I. This finding is in accordance with previous studies that linked the downregulation of surface MHC class I in COPD to impaired immunoproteasome activity (Hodge et al., 2011; Kammerl et al., 2016). Depending on the magnitude, downregulation of MHC class I molecules on the surface of cells can trigger NK-dependent killing, but it has been shown that immunosenescent cells can escape this fate by surface expression of the non-classical MHC class I molecule HLA-E (Pereira et al., 2019). In the present study, although reduced compared to controls, we found HLA-E expression in all AM states of COPD patients. However, the decreased MHC I expression poses a critical functional problem, as it makes AMs less efficient in inducing an antiviral immune response, which may explain the high susceptibility of COPD patients to viral infections, one of the major causes of the disease's exacerbations (Woodhead et al., 2005). Importantly, the reduction in MHC

expression was not restricted to the alveolar space but was also observed on circulating blood monocytes, which underlines the systemic character of COPD (Agusti, 2005; Agusti and Soriano, 2008; Fabbri and Rabe, 2007). However, the question arises as to where and how the blood monocytes experience a COPD-dependent change. Numerous studies have reported on elevated blood plasma levels of pro-inflammatory cytokines, including TNFα, IL-6, IL-8 and C-reactive protein in stable COPD patients, but especially under exacerbations (Gan et al., 2004; Wei et al., 2015; Zhang and Bai, 2017). Interestingly, some of these cytokines can directly affect hematopoietic stem cells in the bone marrow (Kovtonyuk et al., 2016) and thus may lead to disease-related changes in immune cells in chronic situations (Pietras, 2017). Alternatively, the monocytes could also be altered during their transit through the pulmonary circulation (Agusti and Soriano, 2008). However, we propose also a third possibility how blood monocytes experience transcriptional changes similar to what we have found in macrophages of the alveolar space. We discovered that a population of monocytes isolated from mouse lungs expressed genes that are actually found in precursor cells. Interestingly, Lefrançais et al. recently also described that the lung contains a reservoir of hematopoietic progenitors that contribute up to 50% of the total platelet production in adult mice (Lefrançais et al., 2017). Moreover, the authors showed that these progenitor cells have at least the capability to differentiate also in granulocytes and lymphocytes. Our scRNA seq data analysis now provides evidence that monocyte precursors can be found in the lungs of mice as well and that blood monocytes can thus be produced by pulmonary extramedullary hematopoiesis. Assuming that precursor monocytes also exist in human lungs, it is very likely that they are exposed to the same pathogenic influences, such as cigarette smoke, which promote the pathogenesis of COPD. A recent study has shown that changes in airway stem cells indeed play a role in COPD (Rao et al., 2020). However, future studies are needed to investigate whether such disease-promoting alterations can also be found in the pulmonary monocyte progenitor pool.

5.9. Future perspectives

Collectively, this study provides a framework for a better understanding of the pathophysiology of COPD at the systemic level, by providing single-cell information on immune cells in two major compartments, the alveolar space and peripheral blood. It is important to emphasize that one of the main limitations of this study is that the analysis did not include lung epithelial cells. Many studies have shown that changes in epithelial cells are an important denominator in the pathogenesis and pathophysiology of COPD (Mostafaei et al., 2018; Rao et al., 2020; Yoshida et al., 2019). Therefore, in future work this dataset must be extended by the integration with all lung parenchyma cells. Subsequently, further levels of information need to be obtained, such as the study of COPD-dependent changes in different lung segments using spatial transcriptomic or imaging data (Berglund et al., 2018; Nichterwitz et al., 2016; Ståhl et al., 2016; Thrane et al., 2018). The investigation of the epigenetics of individual cells will also become an important component in the future (multimodal single cell information (Li et al., 2019)). In addition, the temporal component must also be taken into account by examining the cells of COPD patients longitudinally over a longer period of time, whereby it will be important to also include acute phases of exacerbations. The patient cohort must also be greatly enlarged in order to capture the complete heterogeneity of this complex disease. Basically, the future must be to study the disease by trying to create a COPD cell atlas, similar to what is already being done for healthy lung tissue in the HCA consortium (Schiller et al., 2019). This type of research will not only define the future of COPD, but for all complex chronic and acute diseases.

6. References

Abboud, R.T., and Vimalanathan, S. (2008). Pathogenesis of COPD. Part I. The role of protease-antiprotease imbalance in emphysema. Int. J. Tuberc. Lung Dis. *12*, 361–367.

Abdelaal, T., Michielsen, L., Cats, D., Hoogduin, D., Mei, H., Reinders, M.J.T., and Mahfouz, A. (2019). A comparison of automatic cell identification methods for single-cell RNA sequencing data. Genome Biol. *20*, 194.

Adams, T.S., Schupp, J.C., Poli, S., Ayaub, E.A., Neumark, N., Ahangari, F., Chu, S.G., Raby, B.A., DeIuliis, G., Januszyk, M., et al. (2019). Single Cell RNA-seq reveals ectopic and aberrant lung resident cell populations in Idiopathic Pulmonary Fibrosis. BioRxiv 759902.

Agusti, A.G.N. (2005). Systemic Effects of Chronic Obstructive Pulmonary Disease. Proc. Am. Thorac. Soc. *2*, 367–370.

Agusti, À., and Soriano, J.B. (2008). COPD as a systemic disease. COPD J. Chronic Obstr. Pulm. Dis. *5*, 133–138.

Agustí, A., and Faner, R. (2018). COPD beyond smoking: new paradigm, novel opportunities. Lancet Respir. Med. *6*, 324–326.

Agustí, A., and Hogg, J.C. (2019a). Update on the Pathogenesis of Chronic Obstructive Pulmonary Disease. N. Engl. J. Med. *381*, 1248–1256.

Agustí, A., and Hogg, J.C. (2019b). Update on the pathogenesis of chronic obstructive pulmonary disease. N. Engl. J. Med. *381*, 1248–1256.

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: Single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086.

Al-Kassimi, F.A., and Alhamad, E.H. (2013). A challenge to the seven widely believed concepts of COPD. Int. J. COPD *8*, 21–30.

Alpert, A., Pickman, Y., Leipold, M., Rosenberg-Hasson, Y., Ji, X., Gaujoux, R., Rabani, H., Starosvetsky, E., Kveler, K., Schaffert, S., et al. (2019). A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. Nat. Med. *25*, 487–495.

Alquicira-Hernandez, J., Sathe, A., Ji, H.P., Nguyen, Q., and Powell, J.E. (2019). scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. Genome Biol. *20*, 264.

Ambrosetti, M., Ageno, W., Spanevello, A., Salerno, M., and Pedretti, R.F.E. (2003). Prevalence and prevention of venous thromboembolism in patients with acute exacerbations of COPD. Thromb. Res. *112*, 203–207.

Angelidis, I., Simon, L.M., Fernandez, I.E., Strunz, M., Mayr, C.H., Greiffo, F.R., Tsitsiridis, G., Ansari, M., Graf, E., Strom, T.M., et al. (2019). An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. Nat. Commun. *10*, 963.

Angelo, M., Bendall, S.C., Finck, R., Hale, M.B., Hitzman, C., Borowsky, A.D., Levenson, R.M., Lowe, J.B., Liu, S.D., Zhao, S., et al. (2014). Multiplexed ion beam imaging of human breast tumors. Nat. Med. *20*, 436–442.

Aran, D., Looney, A.P., Liu, L., Fong, V., Hsu, A., Wolters, P.J., Abate, A., Butte, A.J., and Bhattacharya, M. (2018). Single-cell RNA-seq reveals profibrotic macrophages in lung fibrosis. BioRxiv 284604.

Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat. Immunol. *20*, 163–172.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: Tool for the unification of biology. Nat. Genet. *25*, 25–29.

Baharom, F., Rankin, G., Blomberg, A., and Smed-Sörensen, A. (2017). Human Lung Mononuclear Phagocytes in Health and Disease. Front. Immunol. *8*, 499.

Ballester-López, C., Conlon, T.M., Ertüz, Z., Greiffo, F.R., Irmler, M., Verleden, S.E., Beckers, J., Fernandez, I.E., Eickelberg, O., and Yildirim, A.Ö. (2019). The Notch ligand DNER regulates macrophage IFNγ release in chronic obstructive pulmonary disease. EBioMedicine *43*, 562–575.

Bandura, D.R., Baranov, V.I., Ornatsky, O.I., Antonov, A., Kinach, R., Lou, X., Pavlov, S., Vorobiev, S., Dick, J.E., and Tanner, S.D. (2009). Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. Anal. Chem. *81*, 6813–6822.

Barnes, P.J. (2004). Alveolar macrophages as orchestrators of COPD. COPD *1*, 59–70.

Barnes, P.J. (2008). Immunology of asthma and chronic obstructive pulmonary disease. Nat. Rev. Immunol. *8*, 183–192.

Barnes, P.J. (2013). New anti-inflammatory targets for chronic obstructive pulmonary disease. Nat. Rev. Drug Discov. *12*, 543–559.

Barnes, P.J. (2017). Senescence in COPD and Its Comorbidities. Annu. Rev. Physiol. *79*, 517–539.

Barnes, P.J. (2019). Inflammatory endotypes in COPD. Allergy Eur. J. Allergy Clin. Immunol. *74*, 1249–1256.

Barnes, P.J., Burney, P.G.J., Silverman, E.K., Celli, B.R., Vestbo, J., Wedzicha, J.A.,

and Wouters, E.F.M. (2015). Chronic obstructive pulmonary disease. Nat. Rev. Dis. Prim. *1*, 1–21.

Basset-Léobon, C., Lacoste-Collin, L., Aziza, J., Bes, J.C., Jozan, S., and Courtade-Saïdi, M. (2009). Cut-off values and significance of Oil Red O-positive cells in bronchoalveolar lavage fluid. Cytopathology *21*, 245–250.

Bassler, K., Schulte-Schrepping, J., Warnat-Herresthal, S., Aschenbrenner, A.C., and Schultze, J.L. (2019). The Myeloid Cell Compartment—Cell by Cell. Annu. Rev. Immunol. *37*, 269–293.

Baßler, K., Günther, P., Schulte-Schrepping, J., Becker, M., and Biernat, P. (2019). A bioinformatic toolkit for single-cell mRNA analysis. In Methods in Molecular Biology, (Humana Press Inc.), pp. 433–455.

BD Biosciences (2012). Standardizing Application Setup Across Multiple Flow Cytometers Using BD FACSDiva™ Version 6 Software. BD Biosci. Tech. Bull. 1–16.

Becht, E., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Evaluation of UMAP as an alternative to t-SNE for single-cell data. BioRxiv 298430.

Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. *37*, 38–47.

Belchamber, K.B.R., Singh, R., Batista, C.M., Whyte, M.K., Dockrell, D.H., Kilty, I., Robinson, M.J., Wedzicha, J.A., Barnes, P.J., and Donnelly, L.E. (2019). Defective bacterial phagocytosis is associated with dysfunctional mitochondria in COPD macrophages. Eur. Respir. J. *54*.

Bendall, S.C., Simonds, E.F., Qiu, P., Amir, E.D., Krutzik, P.O., Finck, R., Bruggner, R. V, Melamed, R., Trejo, A., Ornatsky, O.I., et al. (2011). Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. Science (80-.). *332*, 687–695.

Benz, C., Copley, M.R., Kent, D.G., Wohrer, S., Cortes, A., Aghaeepour, N., Ma, E., Mader, H., Rowe, K., Day, C., et al. (2012). Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. Cell Stem Cell *10*, 273–283.

Berg, K., and Wright, J.L. (2016). The pathology of chronic obstructive pulmonary disease: Progress in the 20th and 21st centuries. Arch. Pathol. Lab. Med. *140*, 1423–1428.

Berg, R.D., Levitte, S., O'Sullivan, M.P., O'Leary, S.M., Cambier, C.J., Cameron, J., Takaki, K.K., Moens, C.B., Tobin, D.M., Keane, J., et al. (2016). Lysosomal Disorders Drive Susceptibility to Tuberculosis by Compromising Macrophage Migration. Cell *165*, 139–152.

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2019). Generalizing RNA velocity to transient cell states through dynamical modeling. BioRxiv 820936.

Berglund, E., Maaskola, J., Schultz, N., Friedrich, S., Marklund, M., Bergenstråhle, J., Tarish, F., Tanoglidi, A., Vickovic, S., Larsson, L., et al. (2018). Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. Nat. Commun. *9*.

Bharat, A., Bhorade, S.M., Morales-Nebreda, L., McQuattie-Pimentel, A.C., Soberanes, S., Ridge, K., DeCamp, M.M., Mestan, K.K., Perlman, H., Budinger, G.R.S., et al. (2016). Flow cytometry reveals similarities between lung macrophages in humans and mice. Am. J. Respir. Cell Mol. Biol. *54*, 147–149.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008a). Fast unfolding of communities in large networks.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. (2008b). Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. *2008*.

de BOER, W.I., van SCHADEWIJK, A., SONT, J.K., SHARMA, H.S., STOLK, J., HIEMSTRA, P.S., and van KRIEKEN, J.H.J.M. (1998). Transforming Growth Factor β_1 and Recruitment of Macrophages and Mast Cells in Airways in Chronic Obstructive Pulmonary Disease. Am. J. Respir. Crit. Care Med. *158*, 1951–1957.

Bonnardel, J., T'Jonck, W., Gaublomme, D., Browaeys, R., Scott, C.L., Martens, L., Vanneste, B., De Prijck, S., Nedospasov, S.A., Kremer, A., et al. (2019). Stellate Cells, Hepatocytes, and Endothelial Cells Imprint the Kupffer Cell Identity on Monocytes Colonizing the Liver Macrophage Niche. Immunity *51*, 638-654.e9.

de Bont, C.M., Boelens, W.C., and Pruijn, G.J.M. (2019). NETosis, complement, and coagulation: a triangular relationship. Cell. Mol. Immunol. *16*, 19–27.

Boonlayangoor, P., Telischi, M., Boonlayangoor, S., Sinclair, T.F., and Millhouse, E.W. (1980). Cryopreservation of human granulocytes: study of granulocyte function and ultrastructure. Blood *56*, 237–245.

Boukhenouna, S., Wilson, M.A., Bahmed, K., and Kosmider, B. (2018a). Reactive Oxygen Species in Chronic Obstructive Pulmonary Disease. Oxid. Med. Cell. Longev. *2018*.

Boukhenouna, S., Wilson, M.A., Bahmed, K., and Kosmider, B. (2018b). Reactive oxygen species in chronic obstructive pulmonary disease. Oxid. Med. Cell. Longev. *2018*.

Bowler, R.P., Barnes, P.J., and Crapo, J.D. (2004). The role of oxidative stress in chronic obstructive pulmonary disease. COPD J. Chronic Obstr. Pulm. Dis. *1*, 255–277.

Brandsma, C.A., de Vries, M., Costa, R., Woldhuis, R.R., Königshoff, M., and Timens, W. (2017). Lung ageing and COPD: Is there a role for ageing in abnormal tissue

repair? Eur. Respir. Rev. *26*.

Browaeys, R., Saelens, W., and Saeys, Y. (2019). NicheNet: modeling intercellular communication by linking ligands to target genes. Nat. Methods *17*, 159–162.

Buenrostro, J.D., Corces, M.R., Lareau, C.A., Wu, B., Schep, A.N., Aryee, M.J., Majeti, R., Chang, H.Y., and Greenleaf, W.J. (2018). Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. Cell *173*, 1535-1548.e16.

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S. a, Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat. Biotechnol. *33*, 155–160.

Burkart, K.M., Manichaikul, A., Wilk, J.B., Ahmed, F.S., Burke, G.L., Enright, P., Hansel, N.N., Haynes, D., Heckbert, S.R., Hoffman, E.A., et al. (2014). APOM and high-density lipoprotein cholesterol are associated with lung function and per cent emphysema. In European Respiratory Journal, (European Respiratory Society), pp. 1003–1017.

Busch, R., Hobbs, B.D., Zhou, J., Castaldi, P.J., McGeachie, M.J., Hardin, M.E., Hawrylkiewicz, I., Sliwinski, P., Yim, J.J., Kim, W.J., et al. (2017). Genetic association and risk scores in a chronic obstructive pulmonary disease meta-analysis of 16,707 subjects. Am. J. Respir. Cell Mol. Biol. *57*, 35–46.

Bustamante-Marin, X.M., and Ostrowski, L.E. (2017). Cilia and mucociliary clearance. Cold Spring Harb. Perspect. Biol. *9*.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Cadwell, C.R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., Reimer, J., Shen, S., Bethge, M., Tolias, K.F., et al. (2016). Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. Nat. Biotechnol. *34*, 199–203.

Cai, Y., Shaheen, S.O., Hardy, R., Kuh, D., and Hansell, A.L. (2016). Birth weight, early childhood growth and lung function in middle to early old age: 1946 British birth cohort. Thorax *71*, 916–922.

Cannoodt, R., Saelens, W., Sichien, D., Tavernier, S., Janssens, S., Guilliams, M., Lambrecht, B.N., Preter, K. De, and Saeys, Y. (2016). SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. BioRxiv 079509.

Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science *357*, 661–667.

Carbon, S., Douglass, E., Dunn, N., Good, B., Harris, N.L., Lewis, S.E., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., et al. (2019). The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. *47*, D330–D338.

Celli, B.R., and Agustí, A. (2018). COPD: Time to improve its taxonomy? ERS Monogr. *4*.

Celli, B.R., and Wedzicha, J.A. (2019). Update on clinical aspects of chronic obstructive pulmonary disease. N. Engl. J. Med. *381*, 1257–1266.

Chan, S.M., Weng, A.P., Tibshirani, R., Aster, J.C., and Utz, P.J. (2007). Notch signals positively regulate activity of the mTOR pathway in T-cell acute lymphoblastic leukemia. Blood *110*, 278–286.

Chaudhry, R., and Bordoni, B. (2019). Anatomy, Thorax, Lungs (StatPearls Publishing).
Chen, H., Li, Z., Dong, L., Wu, Y., Shen, H., and Chen, Z. (2019). Lipid metabolism in chronic obstructive pulmonary disease. Int. J. COPD *14*, 1009–1018.

Cheng, J., Nanayakkara, G., Shao, Y., Cueto, R., Wang, L., Yang, W.Y., Tian, Y., Wang, H., and Yang, X. (2017). Mitochondrial proton leak plays a critical role in pathogenesis of cardiovascular diseases. In Advances in Experimental Medicine and Biology, (Springer New York LLC), pp. 359–370.

Chong, S.Z., Evrard, M., Devi, S., Chen, J., Lim, J.Y., See, P., Zhang, Y., Adrover, J.M., Lee, B., Tan, L., et al. (2016). CXCR4 identifies transitional bone marrow premonocytes that replenish the mature monocyte pool for peripheral responses. J. Exp. Med. *213*, 2293–2314.

Churg, A., Zhou, S., and Wright, J.L. (2012). Matrix metalloproteinases in COPD. Eur. Respir. J. *39*, 197–209.

Cochain, C., Vafadarnejad, E., Arampatzi, P., Pelisek, J., Winkels, H., Ley, K., Wolf, D., Saliba, A.-E., and Zernecke, A. (2018). Single-Cell RNA-Seq Reveals the Transcriptional Landscape and Heterogeneity of Aortic Macrophages in Murine Atherosclerosis. Circ. Res. *122*, 1661–1674.

Cohen, M., Giladi, A., Gorki, A.D., Solodkin, D.G., Zada, M., Hladik, A., Miklosi, A., Salame, T.M., Halpern, K.B., David, E., et al. (2018). Lung Single-Cell Signaling Interaction Map Reveals Basophil Role in Macrophage Imprinting. Cell *175*, 1031-1044.e18.

Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. Bioinformatics *33*, 2938–2940.

Coons, A.H., Creech, H.J., and Jones, R.N. (1941). Immunological Properties of an Antibody Containing a Fluorescent Group. Exp. Biol. Med. *47*, 200–202.

Cottin, V., Nunes, H., Brillet, P.Y., Delaval, P., Devouassaoux, G., Tillie-Leblond, I., Israel-Biet, D., Court-Fortune, I., Valeyre, D., Cordier, J.F., et al. (2005). Combined

pulmonary fibrosis and emphysema: A distinct underrecognised entity. Eur. Respir. J. *26*, 586–593.

Crowell, H.L., Soneson, C., Germain, P.-L., Calini, D., Collin, L., Raposo, C., Malhotra, D., and Robinson, M.D. (2019). On the discovery of population-specific state transitions from multi-sample multi-condition single-cell RNA sequencing data. BioRxiv 713412.

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910–914.

Davis, M.M., and Brodin, P. (2018). Rebooting Human Immunology. Annu. Rev. Immunol *36*, 843–864.

Denney, L., and Ho, L.P. (2018). The role of respiratory epithelium in host defence against influenza virus infection. Biomed. J. *41*, 218–233.

DeTomaso, D., Jones, M.G., Subramaniam, M., Ashuach, T., Ye, C.J., and Yosef, N. (2019). Functional interpretation of single cell similarity maps. Nat. Commun. *10*, 1–11.

Diego Jaitin, Ephraim Kenigsberg, Hadas Keren-Shaul, Naama Elefant, Franziska Paul, Irina Zaretsky, Alexander Mildner, Nadav Cohen, Steffen Jung, Amos Tanay, I.A. (2014). Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. Science (80-.).

Ding, J., Adiconis, X., Simmons, S.K., Kowalczyk, M.S., Hession, C.C., Marjanovic, N.D., Hughes, T.K., Wadsworth, M.H., Burks, T., Nguyen, L.T., et al. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat. Biotechnol. 1–10.

Ding, Z., Wang, K., Li, J., Tan, Q., Tan, W., and Guo, G. (2019). Association between glutathione S-transferase gene M1 and T1 polymorphisms and chronic obstructive pulmonary disease risk: A meta-analysis. Clin. Genet. *95*, 53–62.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dobnhoest, A.C. (1955). RESPIRATORY INSUFFICIENCY. Lancet *265*, 1185–1187.

Domej, W., Oettl, K., and Renner, W. (2014). Oxidative stress and free radicals in COPD-implications and relevance for treatment. Int. J. COPD *9*, 1207–1224.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber,

W. (2005). BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. Bioinformatics *21*, 3439–3440.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191.

Eapen, M.S., Sharma, P., and Sohal, S.S. (2019). Mitochondrial dysfunction in macrophages: A key to defective bacterial phagocytosis in COPD. Eur. Respir. J. *54*.

Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2019). CellPhoneDB v2.0: Inferring cell-cell communication from combined expression of multi-subunit receptor-ligand complexes. BioRxiv 680926.

Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat. Protoc. 1–23.

Eisenhauer, P.B., and Lehrer, R.I. (1992). Mouse neutrophils lack defensins. Infect. Immun. *60*, 3446–3447.

Ellis, B., Haaland, P., Hahne, F., Le Meur, N., Gopalakrishnan, N., Spidlen, J., Jiang, M., and Finak, G. (2019). flowCore: Basic structures for flow cytometry data. R package version 1.50.0.

Embgenbroich, M., and Burgdorf, S. (2018). Current concepts of antigen cross-presentation. Front. Immunol. *9*.

Ernst, M.D. (2004). Permutation methods: A basis for exact inference. Stat. Sci. *19*, 676–685.
Fabbri, L.M., and Rabe, K.F. (2007). From COPD to chronic systemic inflammatory syndrome? Lancet *370*, 797–799.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The Reactome Pathway Knowledgebase. Nucleic Acids Res. *46*, D649–D655.

Fan, H.C., Fu, G.K., and Fodor, S.P. a. (2015). Combinatorial labeling of single cells for gene expression cytometry. Science (80-.). *347*, 1258367–1258367.

Filley, G.F., Beckwitt, H.J., Reeves, J.T., and Mitchell, R.S. (1968). Chronic obstructive bronchopulmonary disease. Am. J. Med. *44*, 26–38.

Franke, W.W., Schmid, E., Osborn, M., and Weber, K. (1978). Different intermediate-sized filaments distinguished by immunofluorescence microscopy. Proc. Natl. Acad. Sci. U. S. A. *75*, 5034–5038.

Fulwyler, M.J. (1965). Electronic separation of biological cells by volume. Science *150*, 910–911.

van Furth, R., and Cohn, Z.A. (1968). The origin and kinetics of mononuclear phagocytes. J. Exp. Med. *128*, 415–435.

Gan, W.Q., Man, S.F.P., Senthilselvan, A., and Sin, D.D. (2004). Association between chronic obstructive pulmonary disease and systemic inflammation: A systematic review and a meta-analysis. Thorax *59*, 574–580.

Garbi, N., and Lambrecht, B.N. (2017). Location, function, and ontogeny of pulmonary macrophages during the steady state. Pflugers Arch. Eur. J. Physiol. *469*, 561–572.

Garrido, F., Aptsiauri, N., Doorduijn, E.M., Garcia Lora, A.M., and van Hall, T. (2016). The urgent need to recover MHC class I in cancers for effective immunotherapy. Curr. Opin. Immunol. *39*, 44–51.

Gaublomme, J.T., Yosef, N., Lee, Y., Gertner, R.S., Yang, L.V., Wu, C., Pandolfi, P.P., Mak, T., Satija, R., Shalek, A.K., et al. (2015). Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity. Cell *163*, 1400–1412.

Gawad, C., Koh, W., and Quake, S.R. (2016). Single-cell genome sequencing: current state of the science. Nat. Rev. Genet. *17*, 175–188.

Gierahn, T.M., Wadsworth, M.H., Hughes, T.K., Bryson, B.D., Butler, A., Satija, R., Fortune, S., Love, J.C., and Shalek, A.K. (2017). Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat. Methods.

Ginhoux, F., and Jung, S. (2014). Monocytes and macrophages: developmental pathways and tissue homeostasis. Nat. Rev. Immunol. *14*, 392–404.

Ginhoux, F., Greter, M., Leboeuf, M., Nandi, S., See, P., Gokhan, S., Mehler, M.F., Conway, S.J., Ng, L.G., Stanley, E.R., et al. (2010). Fate mapping analysis reveals that adult microglia derive from primitive macrophages. Science *330*, 841–845.

Ginhoux, F., Schultze, J.L., Murray, P.J., Ochando, J., and Biswas, S.K. (2016). New insights into the multidimensional concept of macrophage ontogeny, activation and function. Nat. Immunol. *17*, 34–40.

Goldstein, L.D., Chen, Y.-J.J., Dunne, J., Mir, A., Hubschle, H., Guillory, J., Yuan, W., Zhang, J., Stinson, J., Jaiswal, B., et al. (2017). Massively parallel nanowell-based single- cell gene expression profiling. 1–10.

Gomez Perdiguero, E., Klapproth, K., Schulz, C., Busch, K., Azzoni, E., Crozet, L., Garner, H., Trouillet, C., de Bruijn, M.F., Geissmann, F., et al. (2015). Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. Nature *518*, 547–551.

Gorgoulis, V., Adams, P.D., Alimonti, A., Bennett, D.C., Bischof, O., Bishop, C., Campisi, J., Collado, M., Evangelou, K., Ferbeyre, G., et al. (2019). Cellular Senescence: Defining a Path Forward. Cell *179*, 813–827.

Goudot, C., Coillard, A., Villani, A.-C., Gueguen, P., Cros, A., Sarkizova, S., Tang-

Huau, T.-L., Bohec, M., Baulande, S., Hacohen, N., et al. (2017). Aryl Hydrocarbon Receptor Controls Monocyte Differentiation into Dendritic Cells versus Macrophages. Immunity *47*, 582-596.e6.

De Grove, K.C., Provoost, S., Verhamme, F.M., Bracke, K.R., Joos, G.F., Maes, T., and Brusselle, G.G. (2016). Characterization and quantification of innate lymphoid cell subsets in human lung. PLoS One *11*.

Grover, A., Sanjuan-Pla, A., Thongjuea, S., Carrelha, J., Giustacchini, A., Gambardella, A., Macaulay, I., Mancini, E., Luis, T.C., Mead, A., et al. (2016). Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. Nat. Commun. *7*.

Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. Nat. Methods *11*, 637–640.

Guilliams, M., and Scott, C.L. (2017). Does niche competition determine the origin of tissue-resident macrophages? Nat. Rev. Immunol. *17*, 451–460.

Guilliams, M., Lambrecht, B.N., and Hammad, H. (2013a). Division of labor between lung dendritic cells and macrophages in the defense against pulmonary infections. Mucosal Immunol. *6*, 464–473.

Guilliams, M., De Kleer, I., Henri, S., Post, S., Vanhoutte, L., De Prijck, S., Deswarte, K., Malissen, B., Hammad, H., and Lambrecht, B.N. (2013b). Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life via GM-CSF. J. Exp. Med. *210*, 1977–1992.

Guilliams, M., Dutertre, C.-A., Scott, C.L., McGovern, N., Sichien, D., Chakarov, S., Van Gassen, S., Chen, J., Poidinger, M., De Prijck, S., et al. (2016). Unsupervised High-Dimensional Analysis Aligns Dendritic Cells across Tissues and Species. Immunity *45*, 669–684.

Guo, M., Du, Y., Gokey, J.J., Ray, S., Bell, S.M., Adam, M., Sudha, P., Perl, A.K., Deshmukh, H., Potter, S.S., et al. (2019). Single cell RNA analysis identifies cellular heterogeneity and adaptive responses of the lung at birth. Nat. Commun. *10*.

van Haarst, J.M., Hoogsteden, H.C., de Wit, H.J., Verhoeven, G.T., Havenith, C.E., and Drexhage, H.A. (1994). Dendritic cells and their precursors isolated from human bronchoalveolar lavage: immunocytologic and functional properties. Am. J. Respir. Cell Mol. Biol. *11*, 344–350.

Halbert, R.J., Natoli, J.L., Gano, A., Badamgarav, E., Buist, A.S., and Mannino, D.M. (2006). Global burden of COPD: Systematic review and meta-analysis. Eur. Respir. J. *28*, 523–532.

Han, M.L.K., Quibrera, P.M., Carretta, E.E., Barr, R.G., Bleecker, E.R., Bowler, R.P., Cooper, C.B., Comellas, A., Couper, D.J., Curtis, J.L., et al. (2017). Frequency of exacerbations in patients with chronic obstructive pulmonary disease: an analysis of the SPIROMICS cohort. Lancet Respir. Med. *5*, 619–626.

Hancock, D.B., Artigas, M.S., Gharib, S.A., Henry, A., Manichaikul, A., Ramasamy, A., Loth, D.W., Imboden, M., Koch, B., McArdle, W.L., et al. (2012). Genome-Wide Joint Meta-Analysis of SNP and SNP-by-Smoking Interaction Identifies Novel Loci for Pulmonary Function. PLoS Genet. *8*.

Hansen, T.H., and Bouvier, M. (2009). MHC class i antigen presentation: Learning from viral evasion strategies. Nat. Rev. Immunol. *9*, 503–513.

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics *14*, 7.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: The reference human genome annotation for the ENCODE project. Genome Res. *22*, 1760–1774.

Hashimoto, D., Chow, A., Noizat, C., Teo, P., Beasley, M.B., Leboeuf, M., Becker, C.D., See, P., Price, J., Lucas, D., et al. (2013). Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. Immunity *38*, 792–804.

Heron, M., Grutters, J.C., Ten Dam-Molenkamp, K.M., Hijdra, D., Van Heugten-Roeling, A., Claessen, A.M.E., Ruven, H.J.T., Van den Bosch, J.M.M., and Van Velzen-Blad, H. (2012). Bronchoalveolar lavage cell pattern from healthy human lung. Clin. Exp. Immunol. *167*, 523–531.

Herzog, R., Schuhmann, K., Schwudke, D., Sampaio, J.L., Bornstein, S.R., Schroeder, M., and Shevchenko, A. (2012). Lipidxplorer: A software for consensual cross-platform lipidomics. PLoS One *7*.

Hidalgo, A., Chilvers, E.R., Summers, C., and Koenderman, L. (2019). The Neutrophil Life Cycle. Trends Immunol. *40*, 584–597.

Hill, D.A., Lim, H.-W., Kim, Y.H., Ho, W.Y., Foong, Y.H., Nelson, V.L., Nguyen, H.C.B., Chegireddy, K., Kim, J., Habertheuer, A., et al. (2018). Distinct macrophage populations direct inflammatory versus physiological changes in adipose tissue. Proc. Natl. Acad. Sci. U. S. A. 1–10.

Hnizdo, E., Sullivan, P.A., Bang, K.M., and Wagner, G. (2002). Association between chronic obstructive pulmonary disease and employment by industry and occupation in the US population: A study of data from the Third National Health and Nutrition Examination Survey. Am. J. Epidemiol. *156*, 738–746.

Hodge, S., Matthews, G., Mukaro, V., Ahern, J., Shivam, A., Hodge, G., Holmes, M., Jersmann, H., and Reynolds, P.N. (2011). Cigarette Smoke-Induced Changes to Alveolar Macrophage Phenotype and Function Are Improved by Treatment with Procysteine. Am. J. Respir. Cell Mol. Biol. *44*, 673–681.

Hoeffel, G., Wang, Y., Greter, M., See, P., Teo, P., Malleret, B., Leboeuf, M., Low, D., Oller, G., Almeida, F., et al. (2012). Adult Langerhans cells derive predominantly from embryonic fetal liver monocytes with a minor contribution of yolk sac–derived

macrophages. J. Exp. Med. *209*, 1167–1181.

Hoeffel, G., Chen, J., Lavin, Y., Low, D., Almeida, F.F., See, P., Beaudin, A.E., Lum, J., Low, I., Forsberg, E.C., et al. (2015). C-Myb+ Erythro-Myeloid Progenitor-Derived Fetal Monocytes Give Rise to Adult Tissue-Resident Macrophages. Immunity *42*, 665–678.

Hoenderdos, K., and Condliffe, A. (2013a). The neutrophil in chronic obstructive pulmonary disease: Too little, too late or too much, too soon? Am. J. Respir. Cell Mol. Biol. *48*, 531–539.

Hoenderdos, K., and Condliffe, A. (2013b). The neutrophil in chronic obstructive pulmonary disease: Too little, too late or too much, too soon? Am. J. Respir. Cell Mol. Biol. *48*, 531–539.

Hoffmann, R.F., Zarrintan, S., Brandenburg, S.M., Kol, A., de Bruin, H.G., Jafari, S., Dijk, F., Kalicharan, D., Kelders, M., Gosker, H.R., et al. (2013). Prolonged cigarette smoke exposure alters mitochondrial structure and function in airway epithelial cells. Respir. Res. *14*.

Hoffmann, R.F., Jonker, M.R., Brandenburg, S.M., de Bruin, H.G., ten Hacken, N.H.T., van Oosterhout, A.J.M., and Heijink, I.H. (2019). Mitochondrial dysfunction increases pro-inflammatory cytokine production and impairs repair and corticosteroid responsiveness in lung epithelium. Sci. Rep. *9*, 1–10.

Hogg, J.C., and Timens, W. (2009). The Pathology of Chronic Obstructive Pulmonary Disease. Annu. Rev. Pathol. Mech. Dis. *4*, 435–459.

Höglund, P., and Brodin, P. (2010). Current perspectives of natural killer cell education by MHC class i molecules. Nat. Rev. Immunol. *10*, 724–734.

Hooke, R. (1665). Micrographia: or some physiological descriptions of minute bodies made by magnifying glasses (London: J. Martyn and J. Allestry).

Hoppstädter, J., Diesel, B., Zarbock, R., Breinig, T., Monz, D., Koch, M., Meyerhans, A., Gortner, L., Lehr, C.M., Huwer, H., et al. (2010). Differential cell reaction upon Toll-like receptor 4 and 9 activation in human alveolar and lung interstitial macrophages. Respir. Res. *11*.

Hou, R., Denisenko, E., Forrest, A.R.R., and Kelso, J. (2019). ScMatch: A single-cell gene expression profile annotation tool using reference datasets. Bioinformatics *35*, 4688–4695.

Houssaini, A., Breau, M., Kebe, K., Abid, S., Marcos, E., Lipskaia, L., Rideau, D., Parpaleix, A., Huang, J., Amsellem, V., et al. (2018). mTOR pathway activation drives lung cell senescence and emphysema. JCI Insight *3*.

Hunninghake, G.M., Cho, M.H., Tesfaigzi, Y., Soto-Quiros, M.E., Avila, L., Lasky-Su, J., Stidley, C., Melén, E., Söderhäll, C., Hallberg, J., et al. (2009). *MMP12,* Lung Function, and COPD in High-Risk Populations. N. Engl. J. Med. *361*, 2599–2608.

Hurst, J.R., Vestbo, J., Anzueto, A., Locantore, N., Müllerova, H., Tal-Singer, R., Miller, B., Lomas, D.A., Agusti, A., MacNee, W., et al. (2010). Susceptibility to exacerbation in chronic obstructive pulmonary disease. N. Engl. J. Med. *363*, 1128–1138.

Hussell, T., and Bell, T.J. (2014). Alveolar macrophages: Plasticity in a tissue-specific context. Nat. Rev. Immunol. *14*, 81–93.

Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C., and Teichmann, S.A. (2016). Classification of low quality cells from single-cell RNA-seq data. Genome Biol. *17*, 29.

Ingebrigtsen, T.S., Marott, J.L., Nordestgaard, B.G., Lange, P., Hallas, J., and Vestbo, J. (2015). Statin use and exacerbations in individuals with chronic obstructive pulmonary disease. Thorax *70*, 33–40.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.-B., Lönnerberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. *21*, 1160–1167.

Jaitin, D.A., Adlung, L., Thaiss, C.A., Weiner, A., Li, B., Descamps, H., Lundgren, P., Bleriot, C., Liu, Z., Deczkowska, A., et al. (2019). Lipid-Associated Macrophages Control Metabolic Homeostasis in a Trem2-Dependent Manner. Cell *178*, 686-698.e14.

Janky, R., Verfaillie, A., Imrichová, H., van de Sande, B., Standaert, L., Christiaens, V., Hulselmans, G., Herten, K., Naval Sanchez, M., Potier, D., et al. (2014). iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections. PLoS Comput. Biol. *10*, e1003731.

Ji, Z., and Ji, H. (2016). TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. Nucleic Acids Res. *44*, e117.

Joffre, O.P., Segura, E., Savina, A., and Amigorena, S. (2012). Cross-presentation by dendritic cells. Nat. Rev. Immunol. *12*, 557–569.

Jones, P.W. (2009). Health status and the spiral of decline. COPD J. Chronic Obstr. Pulm. Dis. *6*, 59–63.

Jones, P.W., Harding, G., Berry, P., Wiklund, I., Chen, W.H., and Kline Leidy, N. (2009). Development and first validation of the COPD Assessment Test. Eur. Respir. J. *34*, 648–654.

Joshi, N., Walter, J.M., and Misharin, A. V. (2018). Alveolar Macrophages. Cell. Immunol. *330*, 86–90.

Jubrail, J., Kurian, N., and Niedergang, F. (2017). Macrophage phagocytosis cracking the defect code in COPD. Biomed. J. *40*, 305–312.

Kaku, Y., Imaoka, H., Morimatsu, Y., Komohara, Y., Ohnishi, K., Oda, H., Takenaka, S., Matsuoka, M., Kawayama, T., Takeya, M., et al. (2014). Overexpression of CD163,

CD204 and CD206 on alveolar macrophages in the lungs of patients with severe chronic obstructive pulmonary disease. PLoS One *9*.

Kammerl, I.E., Dann, A., Mossina, A., Brech, D., Lukas, C., Vosyka, O., Nathan, P., Conlon, T.M., Wagner, D.E., Overkleeft, H.S., et al. (2016). Impairment of Immunoproteasome Function by Cigarette Smoke and in Chronic Obstructive Pulmonary Disease. Am. J. Respir. Crit. Care Med. *193*, 1230–1241.

Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. Protein Sci. *28*, 1947–1951.

Kapellos, T.S., Bassler, K., Aschenbrenner, A.C., Fujii, W., and Schultze, J.L. (2018). Dysregulated functions of lung macrophage populations in COPD. J. Immunol. Res. *2018*.

Kawamura, S., Onai, N., Miya, F., Sato, T., Tsunoda, T., Kurabayashi, K., Yotsumoto, S., Kuroda, S., Takenaka, K., Akashi, K., et al. (2017). Identification of a Human Clonogenic Progenitor with Strict Monocyte Differentiation Potential: A Counterpart of Mouse cMoPs. Immunity *46*, 835-848.e4.

Kim, K., Shim, D., Lee, J.S., Zaitsev, K., Williams, J.W., Kim, K.W., Jang, M.Y., Jang, H.S., Yun, T.J., Lee, S.H., et al. (2018). Transcriptome analysis reveals nonfoamy rather than foamy plaque macrophages are proinflammatory in atherosclerotic murine models. Circ. Res. *123*, 1127–1142.

Kinose, D., Ogawa, E., Kudo, M., Marumo, S., Kiyokawa, H., Hoshino, Y., Hirai, T., Chin, K., Muro, S., and Mishima, M. (2016). Association of COPD exacerbation frequency with gene expression of pattern recognition receptors in inflammatory cells in induced sputum. Clin. Respir. J. *10*, 11–21.

Kirkham, P.A., and Barnes, P.J. (2013). Oxidative stress in COPD. Chest *144*, 266–273.

Kiselev, V.Y., Andrews, T.S., and Hemberg, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. Nat. Rev. Genet. *20*, 273–282.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell *161*, 1187–1201.

Klose, C.S.N., and Artis, D. (2016). Innate lymphoid cells as regulators of immunity, inflammation and tissue homeostasis. Nat. Immunol. *17*, 765–774.

Köhler, N.D., Büttner, M., and Theis, F.J. (2019). Deep learning does not outperform classical machine learning for cell-type annotation. BioRxiv 653907.

Korolchuk, V.I., Miwa, S., Carroll, B., and von Zglinicki, T. (2017). Mitochondria in Cell Senescence: Is Mitophagy the Weakest Link? EBioMedicine *21*, 7–13.

Kovtonyuk, L. V., Fritsch, K., Feng, X., Manz, M.G., and Takizawa, H. (2016). Inflamm-

aging of hematopoiesis, hematopoietic stem cells, and the bone marrow microenvironment. Front. Immunol. *7*, 502.

Krakhmalova, O., and Hetman, O. (2013). Lipid profile in COPD patients with coronary artery disease comorbidity. Eur. Respir. J. *42*.

Kramann, R., Schneider, R.K., Benjamin, D., Kramann, R., Machado, F., Wu, H., Kusaba, T., and Hoeft, K. (2018). Parabiosis and single-cell RNA sequencing reveal a limited contribution of monocytes to myofibroblasts in kidney fibrosis Find the latest version : Parabiosis and single-cell RNA sequencing reveal a limited contribution of monocytes to myofibroblasts in k. J. Clin. Investig. Insight *3(9)*.

Kunz, L.I.Z., Lapperre, T.S., Snoeck-Stroband, J.B., Budulac, S.E., Timens, W., van Wijngaarden, S., Schrumpf, J.A., Rabe, K.F., Postma, D.S., Sterk, P.J., et al. (2011). Smoking status and anti-inflammatory macrophages in bronchoalveolar lavage and induced sputum in COPD. Respir. Res. *12*, 34.

L, P.-M., A, T., A, L., J, de M., L, A.-S., F, G.-A., C, S., and A, A. (2012). Site of Mitochondrial Reactive Oxygen Species Production in Skeletal Muscle of Chronic Obstructive Pulmonary Disease and Its Relationship With Exercise Oxidative Stress. Am. J. Respir. Cell Mol. Biol. *47*.

Lambrecht, B.N. (2006). Alveolar Macrophage in the Driver's Seat. Immunity *24*, 366–368.

Lambrechts, D., Wauters, E., Boeckx, B., Aibar, S., Nittner, D., Burton, O., Bassez, A., Decaluwé, H., Pircher, A., Van den Eynde, K., et al. (2018). Phenotype molding of stromal cells in the lung tumor microenvironment. Nat. Med. *24*, 1277–1289.

Lamprecht, B., McBurnie, M.A., Vollmer, W.M., Gudmundsson, G., Welte, T., Nizankowska-Mogilnicka, E., Studnicka, M., Bateman, E., Anto, J.M., Burney, P., et al. (2011). COPD in never smokers: Results from the population-based burden of obstructive lung disease study. Chest *139*, 752–763.

Landsman, L., and Jung, S. (2007). Lung Macrophages Serve as Obligatory Intermediate between Blood Monocytes and Alveolar Macrophages. J. Immunol. *179*, 3488–3494.

Lange, P., Celli, B., Agustí, A., Jensen, G.B., Divo, M., Faner, R., Guerra, S., Marott, J.L., Martinez, F.D., Martinez-Camblor, P., et al. (2015). Lung-function trajectories leading to chronic obstructive pulmonary disease. In New England Journal of Medicine, (Massachussetts Medical Society), pp. 111–122.

Lange, P., Halpin, D.M., O'Donnell, D.E., and MacNee, W. (2016). Diagnosis, assessment, and phenotyping of COPD: beyond FEV_1. Int. J. Chron. Obstruct. Pulmon. Dis. *11 Spec*, 3–12.

Laniado-Laborin, R. (2009). Smoking and chronic obstructive pulmonary disease (COPD). Parallel epidemics of the 21st century. Int. J. Environ. Res. Public Health *6*, 209–224.

Laurenti, E., and Göttgens, B. (2018). From haematopoietic stem cells to complex differentiation landscapes. Nat. 2018 5537689 *553*, 418.

Lavin, Y., Kobayashi, S., Leader, A., Amir, E. ad D., Elefant, N., Bigenwald, C., Remark, R., Sweeney, R., Becker, C.D., Levine, J.H., et al. (2017). Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. Cell *169*, 750-765.e17.

Lawlor, D.A., Ebrahim, S., and Smith, G.D. (2005). Association of birth weight with adult lung function: Findings from the British Women's Heart and Health Study and a meta-analysis. Thorax *60*, 851–858.

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

Lee, J.S. (2012). Heterogeneity of lung mononuclear phagocytes in chronic obstructive pulmonary disease. J. Innate Immun. *4*, 489–497.

Lee, J.T.H., and Hemberg, M. (2019). Supervised clustering for single-cell analysis. Nat. Methods *16*, 965–966.

Lee, S.J., Lee, H.R., Lee, T.W., Ju, S., Lim, S., Go, S. Il, You, J.W., Cho, Y.J., Lee, G.W., Jeong, Y.Y., et al. (2016). Usefulness of neutrophil to lymphocyte ratio in patients with chronic obstructive pulmonary disease: A prospective observational study. Korean J. Intern. Med. *31*, 891–898.

Lefrançais, E., Ortiz-Muñoz, G., Caudrillier, A., Mallavia, B., Liu, F., Sayah, D.M., Thornton, E.E., Headley, M.B., David, T., Coughlin, S.R., et al. (2017). The lung is a site of platelet biogenesis and a reservoir for haematopoietic progenitors. Nature *544*, 105–109.

Lemarie, E., Carre, P., Legrand, M.F., Lavandier, M., Boissinot, E., Renoux, M., and Renoux, G. (1984). Alveolar macrophage dysfunction in malignant lung tumors. Thorax *39*, 448–452.

Levine, J.H., Simonds, E.F., Bendall, S.C., Davis, K.L., Amir, E.A.D., Tadmor, M.D., Litvin, O., Fienberg, H.G., Jager, A., Zunder, E.R., et al. (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell *162*, 184–197.

Li, B., and Dewey, C.N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, C., Zhang, Y., Lu, Y., Cui, Z., Yu, M., Zhang, S., and Xue, X. (2011). Evidence of the cross talk between Wnt and Notch signaling pathways in non-small-cell lung cancer (NSCLC): Notch3-siRNA weakens the effect of LiCl on the cell cycle of NSCLC cell lines. J. Cancer Res. Clin. Oncol. *137*, 771–778.

Li, G., Liu, Y., Zhang, Y., Kubo, N., Yu, M., Fang, R., Kellis, M., and Ren, B. (2019). Joint profiling of DNA methylation and chromatin architecture in single cells. Nat. Methods *16*, 991–993.

Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., Chen, L., Li, J., Wang, X., Wang, F., et al. (2020a). The landscape of lung bronchoalveolar immune cells in COVID-19 revealed by single-cell RNA sequencing. MedRxiv 2020.02.23.20026690.

Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., Cheng, L., Li, J., Wang, X., Wang, F., et al. (2020b). Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nat. Med. 1–3.

Liao, W.R., Hsieh, R.H., Hsu, K.W., Wu, M.Z., Tseng, M.J., Mai, R.T., Wu lee, Y.H., and Yeh, T.S. (2007). The CBF1-independent Notch1 signal pathway activates human c-myc expression partially via transcription factor YY1. Carcinogenesis *28*, 1867–1876.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst. *1*, 417–425.

Lin, A., and Loré, K. (2017). Granulocytes: New members of the antigen-presenting cell family. Front. Immunol. *8*.

Lin, J. Da, Nishi, H., Poles, J., Niu, X., Mccauley, C., Rahman, K., Brown, E.J., Yeung, S.T., Vozhilla, N., Weinstock, A., et al. (2019). Single-cell analysis of fate-mapped macrophages reveals heterogeneity, including stem-like properties, during atherosclerosis progression and regression. JCI Insight *4*.

Litzenburger, U.M., Buenrostro, J.D., Wu, B., Shen, Y., Sheffield, N.C., Kathiria, A., Greenleaf, W.J., and Chang, H.Y. (2017). Single-cell epigenomic variability reveals functional cancer heterogeneity. Genome Biol. *18*, 15.

Liu, L., Zeng, M., and Stamler, J.S. (1999). Hemoglobin induction in mouse macrophages. Proc. Natl. Acad. Sci. U. S. A. *96*, 6643–6647.

Liu, Z., Gerner, M.Y., Van Panhuys, N., Levine, A.G., Rudensky, A.Y., and Germain, R.N. (2015). Immune homeostasis enforced by co-localized effector and regulatory T cells. Nature *528*, 225–230.

Lloyd, C.M., and Marsland, B.J. (2017). Lung Homeostasis: Influence of Age, Microbes, and the Immune System. Immunity *46*, 549–561.

Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nat. Methods *15*, 1053–1058.

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

Lun, A.T.L., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-

level analysis of single-cell RNA-seq data with Bioconductor. F1000Research *5*, 2122.

Maaten, L. Van Der, and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Mabbott, N.A., Baillie, J., Brown, H., Freeman, T.C., and Hume, D.A. (2013). An expression atlas of human primary cells: inference of gene function from coexpression networks. BMC Genomics *14*, 632.

MacLean, J.A., Xia, W., Pinto, C.E., Zhao, L., Liu, H.W., and Kradin, R.L. (1996). Sequestration of inhaled particulate antigens by lung phagocytes: A mechanism for the effective inhibition of pulmonary cell-mediated immunity. Am. J. Pathol. *148*, 657–666.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015a). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell *161*, 1202–1214.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015b). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell *161*, 1202–1214.

Madissoon, E., Wilbrey-Clark, A., Miragaia, R.J., Saeb-Parsy, K., Mahbubani, K.T., Georgakopoulos, N., Harding, P., Polanski, K., Huang, N., Nowicki-Osuch, K., et al. (2020). scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. Genome Biol. *21*, 1.

Mak, J.C.W., Chan-Yeung, M.M.W., Ho, S.P., Chan, K.S., Choo, K., Yee, K.S., Chau, C.H., Cheung, A.H.K., and Ip, M.S.M. (2009). Elevated plasma TGF-β1 levels in patients with chronic obstructive pulmonary disease. Respir. Med. *103*, 1083–1089.

Makita, H., Nasuhara, Y., Nagai, K., Ito, Y., Hasegawa, M., Betsuyaku, T., Onodera, Y., Hizawa, N., Nishimura, M., Kawakami, Y., et al. (2007). Characterisation of phenotypes based on severity of emphysema in chronic obstructive pulmonary disease. Thorax *62*, 932–937.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

De Marco, R., Accordini, S., Marcon, A., Cerveri, I., Antó, J.M., Gislason, T., Heinrich, J., Janson, C., Jarvis, D., Kuenzli, N., et al. (2011). Risk factors for chronic obstructive pulmonary disease in a European cohort of young adults. Am. J. Respir. Crit. Care Med. *183*, 891–897.

Mass, E. (2018). Delineating the origins, developmental programs and homeostatic functions of tissue-resident macrophages. Int. Immunol.

Mass, E., Ballesteros, I., Farlik, M., Halbritter, F., Günther, P., Crozet, L., Jacome-Galarza, C.E., Händler, K., Klughammer, J., Kobayashi, Y., et al. (2016). Specification of tissue-resident macrophages during organogenesis. Science *353*, aaf4238.

McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. *8*, 329-337.e4.

McGuinness, A., and Sapey, E. (2017). Oxidative Stress in COPD: Sources, Markers, and Potential Mechanisms. J. Clin. Med. *6*, 21.

McInnes, L., and Healy, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.

McQuattie-Pimentel, A.C., Ren, Z., Joshi, N., Watanabe, S., Stoeger, T., Chi, M., Lu, Z., Sichizya, L., Piseaux, R., Chen, C.-I., et al. (2019). The Aging Microenvironment Shapes Alveolar Macrophage Identity in Aging. BioRxiv 717033.

Menezes, S., Melandri, D., Anselmi, G., Perchet, T., Loschko, J., Dubrot, J., Patel, R., Gautier, E.L., Hugues, S., Longhi, M.P., et al. (2016). The Heterogeneity of Ly6Chi Monocytes Controls Their Differentiation into iNOS+ Macrophages or Monocyte-Derived Dendritic Cells. Immunity *45*, 1205–1218.

Merad, M., Manz, M.G., Karsunky, H., Wagers, A., Peters, W., Charo, I., Weissman, I.L., Cyster, J.G., and Engleman, E.G. (2002). Langerhans cells renew in the skin throughout life under steady-state conditions. Nat. Immunol. *3*, 1135–1141.

Mercado, N., Ito, K., and Barnes, P.J. (2015). Accelerated ageing of the lung in COPD: New concepts. Thorax *70*, 482–489.

Mestas, J., and Hughes, C.C.W. (2004). Of Mice and Not Men: Differences between Mouse and Human Immunology. J. Immunol. *172*, 2731–2738.

Meyer, K.C., Raghu, G., Baughman, R.P., Brown, K.K., Costabel, U., Du Bois, R.M., Drent, M., Haslam, P.L., Kim, D.S., Nagai, S., et al. (2012). An official American Thoracic Society clinical practice guideline: The clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. Am. J. Respir. Crit. Care Med. *185*, 1004–1014.

Mirza, S., and Benzo, R. (2017). Chronic Obstructive Pulmonary Disease Phenotypes: Implications for Care. Mayo Clin. Proc. *92*, 1104–1112.

Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A.J., Tanaka, Y., Wilkinson, A.C., Buettner, F., Macaulay, I.C., Jawaid, W., Diamanti, E., et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat. Biotechnol. *33*, 269–276.

Montoro, D.T., Haber, A.L., Biton, M., Vinarsky, V., Lin, B., Birket, S.E., Yuan, F., Chen, S., Leung, H.M., Villoria, J., et al. (2018). A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature *560*, 319–324.

Moore, K.J., Sheedy, F.J., and Fisher, E.A. (2013). Macrophages in atherosclerosis: A dynamic balance. Nat. Rev. Immunol. *13*, 709–721.

Morissette, M.C., Shen, P., Thayaparan, D., and Stämpfli, M.R. (2015). Disruption of pulmonary lipid homeostasis drives cigarette smoke-induced lung inflammation in mice. Eur. Respir. J. *46*, 1451–1460.

Morse, C., Tabib, T., Sembrat, J., Buschur, K.L., Bittar, H.T., Valenzi, E., Jiang, Y., Kass, D.J., Gibson, K., Chen, W., et al. (2019). Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. Eur. Respir. J. *54*.

Mostafaei, S., Kazemnejad, A., Azimzadeh Jamalkandi, S., Amirhashchi, S., Donnelly, S.C., Armstrong, M.E., and Doroudian, M. (2018). Identification of Novel Genes in Human Airway Epithelial Cells associated with Chronic Obstructive Pulmonary Disease (COPD) using Machine-Based Learning Algorithms. Sci. Rep. *8*, 1–20.

Mukhopadhyay, S., Hoidal, J.R., and Mukherjee, T.K. (2006). Role of TNFα in pulmonary pathophysiology. Respir. Res. *7*, 125.

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D., et al. (2011). Tumour evolution inferred by single-cell sequencing. Nature *472*, 90–94.

Nestorowa, S., Hamey, F.K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N.K., Kent, D.G., and Göttgens, B. (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood *128*, e20-31.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457.

Neyt, K., and Lambrecht, B.N. (2013). The role of lung dendritic cell subsets in immunity to respiratory viruses. Immunol. Rev. *255*, 57–67.

Ng Kee Kwong, F., Nicholson, A.G., Harrison, C.L., Hansbro, P.M., Adcock, I.M., and Chung, K.F. (2017). Is mitochondrial dysfunction a driving mechanism linking COPD to nonsmall cell lung carcinoma? Eur. Respir. Rev. *26*.

Ni, L., and Dong, C. (2018). Roles of myeloid and lymphoid cells in the pathogenesis of chronic obstructive pulmonary disease. Front. Immunol. *9*, 1431.

Nichterwitz, S., Chen, G., Aguila Benitez, J., Yilmaz, M., Storvall, H., Cao, M., Sandberg, R., Deng, Q., and Hedlund, E. (2016). Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. Nat. Commun. *7*, 1–11.

Nicod, L.P. (2005). Lung defences: An overview. Eur. Respir. Rev. *14*, 45–50.

Nishimura, D. (2001). BioCarta. Biotech Softw. Internet Rep. *2*, 117–120.

Nkadi, P.O., Merritt, T.A., and Pillers, D.A.M. (2009). An overview of pulmonary surfactant in the neonate: Genetics, metabolism, and the role of surfactant in health and disease. Mol. Genet. Metab. *97*, 95–101.

Noguera, A., Batle, S., Miralles, C., Iglesias, J., Busquets, X., MacNee, W., and Agustí, A.G.N. (2001). Enhanced neutrophil response in chronic obstructive pulmonary disease. Thorax *56*, 432–437.

Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., Daníelsdóttir, A.D., Krecke, M., Merten, D., Haraldsdóttir, H.S., et al. (2019). The Virtual Metabolic Human database: Integrating human and gut microbiome metabolism with nutrition and disease. Nucleic Acids Res. *47*, D614–D624.

O'Shaughnessy, T.C., Ansari, T.W., Barnes, N.C., and Jeffery, P.K. (1997). Inflammation in bronchial biopsies of subjects with chronic bronchitis: Inverse relationship of CD8+ T lymphocytes with FEV1. Am. J. Respir. Crit. Care Med. *155*, 852–857.

Ohnuki, H., Jiang, K., Wang, D., Salvucci, O., Kwak, H., Sánchez-Martín, D., Maric, D., and Tosato, G. (2014). Tumor-infiltrating myeloid cells activate Dll4/Notch/TGF-β signaling to drive malignant progression. Cancer Res. *74*, 2038–2049.

Oishi, Y., and Manabe, I. (2016). Macrophages in age-related chronic inflammatory diseases. Npj Aging Mech. Dis. *2*, 1–8.

Okabe, Y., and Medzhitov, R. (2016). Tissue biology perspective on macrophages. Nat. Immunol. *17*, 9–17.

Opalek, J.M., Ali, N.A., Lobb, J.M., Hunter, M.G., and Marsh, C.B. (2007). Alveolar macrophages lack CCR2 expression and do not migrate to CCL2. J. Inflamm. *4*, 19.
Orr, J.S., Kennedy, A.J., and Hasty, A.H. (2013). Isolation of adipose tissue immune cells. J. Vis. Exp. e50707.

Pandey, K.C., De, S., and Mishra, P.K. (2017). Role of Proteases in Chronic Obstructive Pulmonary Disease. Front. Pharmacol. *8*, 512.

Park, J., Shrestha, R., Qiu, C., Kondo, A., Huang, S., Werth, M., Li, M., Barasch, J., and Suszták, K. (2018). Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. Science (80-.). *360*, 758–763.

Parks, D.R., Roederer, M., and Moore, W.A. (2006). A new "Logicle" display method avoids deceptive effects of logarithmic scaling for low signals and compensated data. Cytom. Part A *69A*, 541–551.

Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., et al. (2015). Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. Cell *163*, 1663–1677.

Pearson, K. (1901). LIII. *On lines and planes of closest fit to systems of points in space*. London, Edinburgh, Dublin Philos. Mag. J. Sci. *2*, 559–572.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., et al. (2012). Scikit-learn: Machine Learning in Python.

Pellegrino, R., Viegi, G., Brusasco, V., Crapo, R.O., Burgos, F., Casaburi, R., Coates, A., van der Grinten, C.P.M., Gustafsson, P., Hankinson, J., et al. (2005). Interpretative strategies for lung function tests. Eur. Respir. J. *26*, 948–968.

Pereira, B.I., Devine, O.P., Vukmanovic-Stejic, M., Chambers, E.S., Subramanian, P., Patel, N., Virasami, A., Sebire, N.J., Kinsler, V., Valdovinos, A., et al. (2019). Senescent cells evade immune clearance via HLA-E-mediated NK and CD8+ T cell inhibition. Nat. Commun. *10*, 1–13.

Pérez-Rial, S., Puerto-Nevado, L. del, Terrón-Expósito, R., Girón-Martínez, Á., González-Mangado, N., and Peces-Barba, G. (2013). Role of Recently Migrated Monocytes in Cigarette Smoke-Induced Lung Inflammation in Different Strain of Mice. PLoS One *8*, e72975.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014a). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Picelli, S., Björklund, A.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014b). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. *24*, 2033–2040.

Pietras, E.M. (2017). Inflammation: A key regulator of hematopoietic stem cell fate in health and disease. Blood *130*, 1693–1698.

Pikoula, M., Quint, J.K., Nissen, F., Hemingway, H., Smeeth, L., and Denaxas, S. (2019). Identifying clinically important COPD sub-types using data-driven approaches in primary care population based electronic health records. BMC Med. Inform. Decis. Mak. *19*, 86.

Plasschaert, L.W., Žilionis, R., Choo-Wing, R., Savova, V., Knehr, J., Roma, G., Klein, A.M., and Jaffe, A.B. (2018). A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. Nature *560*, 377–381.

Ponzetta, A., Carriero, R., Carnevale, S., Barbagallo, M., Molgora, M., Perucchini, C., Magrini, E., Gianni, F., Kunderfranco, P., Polentarutti, N., et al. (2019). Neutrophils Driving Unconventional T Cells Mediate Resistance against Murine Sarcomas and Selected Human Tumors. Cell *178*, 346-360.e24.

Putcha, N., Barr, R.G., Han, M.K., Woodruff, P.G., Bleecker, E.R., Kanner, R.E., Martinez, F.J., Smith, B.M., Tashkin, D.P., Bowler, R.P., et al. (2016). Understanding the impact of second-hand smoke exposure on clinical outcomes in participants with COPD in the SPIROMICS cohort. Thorax *71*, 411–420.

Puttur, F., Gregory, L.G., and Lloyd, C.M. (2019). Airway macrophages as the guardians of tissue repair in the lung. Immunol. Cell Biol. *97*, 246–257.

Pylaeva, E., Lang, S., and Jablonska, J. (2016). The essential role of type I interferons in differentiation and activation of tumor-associated neutrophils. Front. Immunol. *7*, 629.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017a). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*.

Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.-A., and Trapnell, C. (2017b). Single-cell mRNA quantification and differential analysis with Census. Nat Meth *14*.

Rabe, K.F., and Watz, H. (2017). Chronic obstructive pulmonary disease. Lancet *389*, 1931–1940.

Rahman, I. (2005). The role of oxidative stress in the pathogenesis of COPD: Implications for therapy. Treat. Respir. Med. *4*, 175–200.

Ramírez-Venegas, A., Sansores, R.H., Quintana-Carrillo, R.H., Velázquez-Uncal, M., Hernandez-Zenteno, R.J., Sánchez-Romero, C., Velazquez-Montero, A., and Flores-Trujillo, F. (2014). FEV1 decline in patients with chronic obstructive pulmonary disease associated with biomass exposure. Am. J. Respir. Crit. Care Med. *190*, 996–1002.

Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. *30*, 777–782.

Rao, W., Wang, S., Duleba, M., Niroula, S., Goller, K., Xie, J., Mahalingam, R., Neupane, R., Liew, A.A., Vincent, M., et al. (2020). Regenerative Metaplastic Clones in COPD Lung Drive Inflammation and Fibrosis. Cell *181*, 848-864.e18.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. *47*, W191–W198.

Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The human cell atlas. Elife *6*.

Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007). G:Profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. *35*, 193–200.

Remmerie, A., and Scott, C.L. (2018). Macrophages and lipid metabolism. Cell. Immunol. *330*, 27–42.

Reyfman, P.A., Walter, J.M., Joshi, N., Anekalla, K.R., McQuattie-Pimentel, A.C., Chiu, S., Fernandez, R., Akbarpour, M., Chen, C.I., Ren, Z., et al. (2019). Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*.

Rocha, B.C., Marques, P.E., Leoratti, F.M. de S., Junqueira, C., Pereira, D.B., Antonelli, L.R. do V., Menezes, G.B., Golenbock, D.T., and Gazzinelli, R.T. (2015). Type I Interferon Transcriptional Signature in Neutrophils and Low-Density Granulocytes Are Associated with Tissue Damage in Malaria. Cell Rep. *13*, 2829–2841.

Rock, K.L., Reits, E., and Neefjes, J. (2016). Present Yourself! By MHC Class I and MHC Class II Molecules. Trends Immunol. *37*, 724–737.

Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science *360*, 176–182.

Roth, G.A., Abate, D., Abate, K.H., Abay, S.M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., et al. (2018). Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet *392*, 1736–1788.

Russell, R.E.K., Thorley, A., Culpitt, S. V., Dodd, S., Donnelly, L.E., Demattos, C., Fitzgerald, M., and Barnes, P.J. (2002). Alveolar macrophage-mediated elastolysis: Roles of matrix metalloproteinases, cysteine, and serine proteases. Am. J. Physiol. - Lung Cell. Mol. Physiol. *283*.

Ryter, S.W., Rosas, I.O., Owen, C.A., Martinez, F.J., Choi, M.E., Lee, C.G., Elias, J.A., and Choi, A.M.K. (2018). Mitochondrial Dysfunction as a Pathogenic Mediator of Chronic Obstructive Pulmonary Disease and Idiopathic Pulmonary Fibrosis. Ann. Am. Thorac. Soc. *15*, S266–S272.

Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2018). A comparison of single-cell trajectory inference methods: towards more accurate and robust tools. BioRxiv 276907.

Saha, S., and Brightling, C.E. (2006). Eosinophilic airway inflammation in COPD. Int. J. Chron. Obstruct. Pulmon. Dis. *1*, 39–47.

Sakornsakolpat, P., Prokopenko, D., Lamontagne, M., Reeve, N.F., Guyatt, A.L.,

Jackson, V.E., Shrine, N., Qiao, D., Bartz, T.M., Kim, D.K., et al. (2019). Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. Nat. Genet. *51*, 494–505.

Sallese, A., Suzuki, T., McCarthy, C., Bridges, J., Filuta, A., Arumugam, P., Shima, K., Ma, Y., Wessendarp, M., Black, D., et al. (2017). Targeting cholesterol homeostasis in lung diseases. Sci. Rep. *7*.

Saxton, R.A., and Sabatini, D.M. (2017). mTOR Signaling in Growth, Metabolism, and Disease. Cell *168*, 960–976.

Schaefer, C.F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., and Buetow, K.H. (2009). PID: The pathway interaction database. Nucleic Acids Res. *37*, 674–679.

Schiller, H.B., Montoro, D.T., Simon, L.M., Rawlins, E.L., Meyer, K.B., Strunz, M., Vieira Braga, F., Timens, W., Koppelman, G.H., Budinger, G.R.S., et al. (2019). The Human Lung Cell Atlas - A high-resolution reference map of the human lung in health and disease. Am. J. Respir. Cell Mol. Biol. *61*, rcmb.2018-0416TR.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Schultze, J.L., and Beyer, M. (2016). Myelopoiesis Reloaded: Single-Cell Transcriptomics Leads the Way. Immunity *44*, 18–20.

Schulz. Christian, Gomez Perdiguero, E., Chorro, L., Szabo-Rogers, H., Cagnard, N., Kierdorf, K., Prinz, M., Wu, B., Jacobson, S.E.W., Pollard, J.W., et al. (2012). A Lineage of Myeloid Cells Independent of Myb and Hematopoietic Stem Cells. Science (80-.). *336*, 86–90.

Schyns, J., Bureau, F., and Marichal, T. (2018). Lung Interstitial Macrophages: Past, Present, and Future. J. Immunol. Res. *2018*.

Schyns, J., Bai, Q., Ruscitti, C., Radermecker, C., De Schepper, S., Chakarov, S., Farnir, F., Pirottin, D., Ginhoux, F., Boeckxstaens, G., et al. (2019). Non-classical tissue monocytes and two functionally distinct populations of interstitial macrophages populate the mouse lung. Nat. Commun. *10*.

Scialdone, A., Natarajan, K.N., Saraiva, L.R., Proserpio, V., Teichmann, S.A., Stegle, O., Marioni, J.C., and Buettner, F. (2015). Computational assignment of cell-cycle stage from single-cell transcriptome data. Methods *85*, 54–61.

Scott, C.L., Zheng, F., De Baetselier, P., Martens, L., Saeys, Y., De Prijck, S., Lippens, S., Abels, C., Schoonooghe, S., Raes, G., et al. (2016). Bone marrow-derived monocytes give rise to self-renewing and fully differentiated Kupffer cells. Nat. Commun. *7*, 1–10.

See, P., Dutertre, C.A., Chen, J., Günther, P., McGovern, N., Irac, S.E., Gunawan, M., Beyer, M., Händler, K., Duan, K., et al. (2017). Mapping the human DC lineage through the integration of high-dimensional techniques. Science (80-.). *356*.

Sevenoaks, M.J., and Stockley, R.A. (2006). Chronic obstructive pulmonary disease, inflammation and co-morbidity - A common inflammatory phenotype? Respir. Res. *7*, 70.
Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236–240.

Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363–369.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A software Environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat. Rev. Genet. *14*, 618–630.
Shaw, A.C., Goldstein, D.R., and Montgomery, R.R. (2013). Age-dependent dysregulation of innate immunity. Nat. Rev. Immunol. *13*, 875–887.

Sheng, J., Ruedl, C., and Karjalainen, K. (2015). Most Tissue-Resident Macrophages Except Microglia Are Derived from Fetal Hematopoietic Stem Cells. Immunity *43*, 382–393.
Skronska-Wasek, W., Mutze, K., Baarsma, H.A., Bracke, K.R., Alsafadi, H.N., Lehmann, M., Costa, R., Stornaiuolo, M., Novellino, E., Brusselle, G.G., et al. (2017). Reduced Frizzled Receptor 4 Expression Prevents WNT/β-Catenin–driven Alveolar Lung Repair in Chronic Obstructive Pulmonary Disease. Am. J. Respir. Crit. Care Med. *196*, 172–185.

Slenter, D.N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., Mélius, J., Cirillo, E., Coort, S.L., DIgles, D., et al. (2018). WikiPathways: A multifaceted pathway database bridging metabolomics to other omics research. Nucleic Acids Res. *46*, D661–D667.

Solana, R., Tarazona, R., Gayoso, I., Lesur, O., Dupuis, G., and Fulop, T. (2012). Innate immunosenescence: Effect of aging on cells and receptors of the innate immune system in humans. Semin. Immunol. *24*, 331–341.

Song, Q., Hawkins, G.A., Wudel, L., Chou, P.C., Forbes, E., Pullikuth, A.K., Liu, L., Jin, G., Craddock, L., Topaloglu, U., et al. (2019). Dissecting intratumoral myeloid cell plasticity by single cell RNA-seq. Cancer Med. *8*, 3072–3085.

Spiteri, M.A., Clarke, S.W., and Poulter, L.W. (1992). Isolation of phenotypically and functionally distinct macrophage subpopulations from human bronchoalveolar lavage. Eur. Respir. J. *5*, 717–726.

Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J.,

Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science (80-.). *353*, 78–82.

Di Stefano, A., Sangiorgi, C., Gnemmi, I., Casolari, P., Brun, P., Ricciardolo, F.L.M., Contoli, M., Papi, A., Maniscalco, P., Ruggeri, P., et al. (2018). TGF-β Signaling Pathways in Different Compartments of the Lower Airways of Patients With Stable COPD. Chest *153*, 851–862.

Steinman, R.M., and Cohn, Z.A. (1973). Identification of a novel cell type in peripheral lymphoid organs of mice. J. Exp. Med. *137*, 1142–1162.

Stolk, J., Aggarwal, N., Hochnadel, I., Wrenger, S., Martinez-Delgado, B., Welte, T., Yevsa, T., and Janciauskiene, S. (2019). Blood monocyte profiles in COPD patients with PiMM and PiZZ α1-antitrypsin. Respir. Med. *148*, 60–62.

Stoller, J.K., and Aboussouan, L.S. (2005). α1-antitrypsin deficiency. In Lancet, (Lancet), pp. 2225–2236.

Strategy, G., Obstructive, C., and Disease, P. (2020). Gold 2020.

Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2017). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. BioRxiv 128843.

Strunz, M., Simon, L.M., Ansari, M., Mattner, L.F., Angelidis, I., Christoph, H., Kathiriya, J., Yee, M., Ogar, P., Sengupta, A., et al. (2019). Longitudinal single cell transcriptomics reveals Krt8 + alveolar epithelial progenitors in lung regeneration.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e21.

Stunnenberg, H.G., Abrignani, S., Adams, D., de Almeida, M., Altucci, L., Amin, V., Amit, I., Antonarakis, S.E., Aparicio, S., Arima, T., et al. (2016). The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. Cell *167*, 1145–1149.

Sun, Y., and Zhou, J. (2019). New insights into early intervention of chronic obstructive pulmonary disease with mild airflow limitation. Int. J. COPD *14*, 1119–1125.

Suvà, M.L., and Tirosh, I. (2019). Single-Cell RNA Sequencing in Cancer: Lessons Learned and Emerging Challenges. Mol. Cell *75*, 7–12.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382.

Tashkin, D.P. (2015). Smoking Cessation in Chronic Obstructive Pulmonary Disease. Semin. Respir. Crit. Care Med. *36*, 491–507.

Tashkin, D.P., and Wechsler, M.E. (2018). Role of eosinophils in airway inflammation of chronic obstructive pulmonary disease. Int. J. COPD *13*, 335–349.

Taylor, A.E., Finney-Hayward, T.K., Quint, J.K., Thomas, C.M.R., Tudhope, S.J., Wedzicha, J.A., Barnes, P.J., and Donnelly, L.E. (2010). Defective macrophage phagocytosis of bacteria in COPD. Eur. Respir. J. *35*, 1039–1047.

Thiele, I., and Palsson, B. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. *5*, 93–121.

Thiele, I., Swainston, N., Fleming, R.M.T., Hoppe, A., Sahoo, S., Aurich, M.K., Haraldsdottir, H., Mo, M.L., Rolfsson, O., Stobbe, M.D., et al. (2013). A community-driven global reconstruction of human metabolism. Nat. Biotechnol. *31*, 419–425.

Thrane, K., Eriksson, H., Maaskola, J., Hansson, J., and Lundeberg, J. (2018). Spatially resolved transcriptomics enables dissection of genetic heterogeneity in stage III cutaneous malignant melanoma. Cancer Res. *78*, 5970–5979.

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science (80-.). *352*, 189–196.

Tirouvanziam, R., Gernez, Y., Conrad, C.K., Moss, R.B., Schrijver, I., Dunn, C.E., Davies, Z.A., Herzenberg, L.A., and Herzenberg, L.A. (2008). Profound functional and signaling changes in viable inflammatory neutrophils homing to cystic fibrosis airways. Proc. Natl. Acad. Sci. U. S. A. *105*, 4335–4339.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Travaglini, K.J., Nabhan, A.N., Penland, L., Sinha, R., Gillich, A., Sit, R. V, Chang, S., Conley, S.D., Mori, Y., Seita, J., et al. (2019). A molecular cell atlas of the human lung from single cell RNA sequencing. BioRxiv *7191*, 742320.

Tuder, R.M., Yoshida, T., Fijalkowka, I., Biswal, S., and Petrache, I. (2006). Role of lung maintenance program in the heterogeneity of lung destruction in emphysema. In Proceedings of the American Thoracic Society, pp. 673–679.

Tusi, B.K., Wolock, S.L., Weinreb, C., Hwang, Y., Hidalgo, D., Zilionis, R., Waisman, A., Huh, J.R., Klein, A.M., and Socolovsky, M. (2018). Population snapshots predict early haematopoietic and erythroid hierarchies. Nature *555*, 54–60.

Uddin, M., Watz, H., Malmgren, A., and Pedersen, F. (2019). NETopethic inflammation in chronic obstructive pulmonary disease and severe asthma. Front. Immunol. *10*.

Vafadarnejad, E., Rizzo, G., Krampert, L., Arampatzi, P., Nugroho, V.A., Schulz, D.,

Roesch, M., Alayrac, P., Vilar, J., Silvestre, J.-S., et al. (2019). Time-resolved single-cell transcriptomics uncovers dynamics of cardiac neutrophil diversity in murine myocardial infarction. BioRxiv 738005.

Vallejos, C.A., Risso, D., Scialdone, A., Dudoit, S., and Marioni, J.C. (2017). Normalizing single-cell RNA sequencing data: challenges and opportunities. Nat. Methods *14*, 565–571.

Vandivier, R.W., Henson, P.M., and Douglas, I.S. (2006). Burying the dead: The impact of failed apoptotic cell removal (efferocytosis) on chronic inflammatory lung disease. Chest *129*, 1673–1682.

Varol, C., Mildner, A., and Jung, S. (2015). Macrophages: Development and Tissue Specialization. Annu. Rev. Immunol.

Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B., Park, J.E., Stephenson, E., Polański, K., Goncalves, A., et al. (2018). Single-cell reconstruction of the early maternal–fetal interface in humans. Nature *563*, 347–353.

Vermaelen, K., and Pauwels, R. (2004). Accurate and simple discrimination of mouse pulmonary dendritic cell and macrophage populations by flow cytometry: Methodology and new insights. Cytometry *61A*, 170–177.

Vestbo, J., Hurd, S.S., Agustí, A.G., Jones, P.W., Vogelmeier, C., Anzueto, A., Barnes, P.J., Fabbri, L.M., Martinez, F.J., Nishimura, M., et al. (2013). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease GOLD executive summary. Am. J. Respir. Crit. Care Med. *187*, 347–365.

Vieira Braga, F.A., Kar, G., Berg, M., Carpaij, O.A., Polanski, K., Simon, L.M., Brouwer, S., Gomes, T., Hesse, L., Jiang, J., et al. (2019). A cellular census of human lungs identifies novel cell states in health and in asthma. Nat. Med. *25*, 1153–1163.

Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science *356*, eaah4573.

Voelkel, N.F., Taraseviciene-Stewart, L., and Tuder, R.M. (2004). Apoptosis and emphysema. In Lung Development and Regeneration, (CRC Press), pp. 395–409.

Wagner, A., Wang, C., DeTomaso, D., Avila-Pacheco, J., Zaghouani, S., Fessler, J., Akama-Garren, E., Pierce, K., Ron-Harel, N., Douglas, V.P., et al. (2020). In Silico Modeling of Metabolic State in Single Th17 Cells Reveals Novel Regulators of Inflammation and Autoimmunity. BioRxiv 2020.01.23.912717.

Wang, C., Wagner, A., Fessler, J., Avila-Pacheco, J., Karminski, J., Thakore, P., Zaghouani, S., Pierce, K., Bod, L., Schnell, A., et al. (2020). Metabolic and Epigenomic Regulation of Th17/Treg Balance by the Polyamine Pathway. BioRxiv 2020.01.23.911966.

Wang, T., Li, B., Nelson, C.E., and Nabavi, S. (2019). Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data. BMC Bioinformatics *20*, 40.

Wattenberg, M., Viégas, F., and Johnson, I. (2016). How to Use t-SNE Effectively. Distill *1*, e2.

Wedzicha, J.A., and Seemungal, T.A. (2007). COPD exacerbations: defining their cause and prevention. Lancet *370*, 786–796.

Wei, J., Xiong, X.F., Lin, Y.H., Zheng, B.X., and Cheng, D.Y. (2015). Association between serum interleukin-6 concentrations and chronic obstructive pulmonary disease: A systematic review and meta-analysis. PeerJ *2015*.

Whitsett, J.A., Wert, S.E., and Weaver, T.E. (2010). Alveolar Surfactant Homeostasis and the Pathogenesis of Pulmonary Disease. Annu. Rev. Med. *61*, 105–119.

Wickham, H. (2016). ggplot2 Elegant Graphics for Data Analysis (Springer International Publishing).

Wiegman, C.H., Michaeloudes, C., Haji, G., Narang, P., Clarke, C.J., Russell, K.E., Bao, W., Pavlidis, S., Barnes, P.J., Kanerva, J., et al. (2015). Oxidative stress-induced mitochondrial dysfunction drives inflammation and airway smooth muscle remodeling in patients with chronic obstructive pulmonary disease. J. Allergy Clin. Immunol. *136*, 769–780.

Wilson, A.M., Nair, P., Hargreave, F.E., Efthimiadis, A.E., Anvari, M., and Allen, C.J. (2011). Lipid and smoker's inclusions in sputum macrophages in patients with airway diseases. Respir. Med. *105*, 1691–1695.

Winkels, H., Ehinger, E., Vassallo, M., Buscher, K., Dinh, H., Kobiyama, K., Hamers, A., Cochain, C., Vafadarnejad, E., Saliba, A.-E., et al. (2018). Atlas of the Immune Cell Repertoire in Mouse Atherosclerosis Defined by Single-Cell RNA-Sequencing and Mass Cytometry. Circ. Res. *122*, CIRCRESAHA.117.312513.

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

Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. *20*, 59.

Woodhead, M., Blasi, F., Ewig, S., Huchon, G., Leven, M., Ortqvist, A., Schaberg, T., Torres, A., van der Heijden, G., and Verheij, T.J.M. (2005). Guidelines for the management of adult lower respiratory tract infections. Eur. Respir. J. *26*, 1138–1180.

Wright, H.J., Matthews, J.B., Chapple, I.L.C., Ling-Mountford, N., and Cooper, P.R. (2008). Periodontitis Associates with a Type 1 IFN Signature in Peripheral Blood Neutrophils. J. Immunol. *181*, 5775–5784.

Xie, X., Shi, Q., Wu, P., Zhang, X., Kambara, H., Su, J., Yu, H., Park, S.-Y., Guo, R., Ren, Q., et al. (2019). Single-cell transcriptome profiling reveals neutrophil heterogeneity and orchestrated maturation during homeostasis and bacterial infection. BioRxiv 792200.

Xue, J., Schmidt, S.V., Sander, J., Draffehn, A., Krebs, W., Quester, I., De Nardo, D., Gohel, T.D., Emde, M., Schmidleithner, L., et al. (2014). Transcriptome-Based Network Analysis Reveals a Spectrum Model of Human Macrophage Activation. Immunity *40*, 274–288.

Yáñez, A., Coetzee, S.G., Olsson, A., Muench, D.E., Berman, B.P., Hazelett, D.J., Salomonis, N., Grimes, H.L., and Goodridge, H.S. (2017). Granulocyte-Monocyte Progenitors and Monocyte-Dendritic Cell Progenitors Independently Produce Functionally Distinct Monocytes. Immunity *47*, 890-902.e4.

Yeh, T.S., Lin, Y.M., Hsieh, R.H., and Tseng, M.J. (2003). Association of Transcription Factor YY1 with the High Molecular Weight Notch Complex Suppresses the Transactivation Activity of Notch. J. Biol. Chem. *278*, 41963–41969.

Yin, P., Jiang, C., Cheng, K., Lam, T., Lam, K., Miller, M., Zhang, W., Thomas, G., and Adab, P. (2007). Passive smoking exposure and risk of COPD among adults in China: the Guangzhou Biobank Cohort Study. Lancet *370*, 751–757.

Yona, S., Kim, K.-W., Wolf, Y., Mildner, A., Varol, D., Breker, M., Strauss-Ayali, D., Viukov, S., Guilliams, M., Misharin, A., et al. (2013). Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. Immunity *38*, 79–91.

Yoshida, M., Minagawa, S., Araya, J., Sakamoto, T., Hara, H., Tsubouchi, K., Hosaka, Y., Ichikawa, A., Saito, N., Kadota, T., et al. (2019). Involvement of cigarette smoke-induced epithelial cell ferroptosis in COPD pathogenesis. Nat. Commun. *10*, 1–14.

Young, K., Borikar, S., Bell, R., Kuffler, L., Philip, V., and Trowbridge, J.J. (2016). Progressive alterations in multipotent hematopoietic progenitors underlie lymphoid cell loss in aging. J Exp Med *213*, 1–9.

Yu, X., Buttgereit, A., Lelios, I., Utz, S.G., Cansever, D., Becher, B., and Greter, M. (2017). The Cytokine TGF-β Promotes the Development and Homeostasis of Alveolar Macrophages. Immunity *47*, 903-912.e4.

Zafirova-Ivanovska, B., Stojkovikj, J., Dokikj, D., Anastasova, S., Debresliovska, A., Zejnel, S., and Stojkovikj, D. (2016). The level of cholesterol in COPD patients with severe and very severe stage of the disease. Open Access Maced. J. Med. Sci. *4*, 277–282.

Zhang, J., and Bai, C. (2017). Elevated serum IL-8: a biomarker indicating exacerbation-prone COPD. In European Respiratory Journal, (European Respiratory Society (ERS)), p. PA3601.

Zhang, X., Goncalves, R., and Mosser, D.M. (2008). The Isolation and Characterization of Murine Macrophages. In Current Protocols in Immunology, (Hoboken, NJ, USA: John Wiley & Sons, Inc.), pp. 14.1.1-14.1.14.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*.

Ziegler-Heitbrock, L. (2015). Blood monocytes and their subsets: Established features and open questions. Front. Immunol. *6*, 1–5.

Ziegler, C., Allon, S.J., Nyquist, S.K., Mbano, I., Miao, V.N., Cao, Y., Yousif, A.S., Bals, J., Hauser, B.M., Feldman, J., et al. (2020). SARS-CoV-2 Receptor ACE2 is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Enriched in Specific Cell Subsets Across Tissues. SSRN Electron. J.

Zilionis, R., Engblom, C., Pfirschke, C., Savova, V., Zemmour, D., Saatcioglu, H.D., Krishnan, I., Maroni, G., Meyerovitz, C. V., Kerwin, C.M., et al. (2019). Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. Immunity *50*, 1317-1334.e10.