Line of Defense

Profiling single cells
to investigate disease
development

Integrating single-cell multi-omics to decipher host-pathogen interactions

Integration von Genomik Daten einzelner Zellen zur Entschlüsselung von Wirt-Pathogen Interaktionen

Doctoral thesis for a doctoral degree at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Infection and Immunity

submitted by

Oliver Dietrich

from

München

Würzburg, 2023

Submitted on: ....
Office stamp

Members of the Thesis Committee

Chairperson: Prof. Dr. Georg Gasteiger

Primary Supervisor: Prof. Dr. Antoine-Emmanuel Saliba

Supervisor (Second): Prof. Dr. Roy Gross

Supervisor (Third): Prof. Dr. Florian Erhard

Supervisor (Fourth): Prof. Dr. Leif Erik Sander

This document is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0): http://creativecommons.org/licenses/by/4.0 This CC license does not apply to third party material (attributed to another source) in this publication.

Abstract

Interactions between host and pathogen determine the development, progression and outcomes of disease. Medicine benefits from better descriptions of these interactions through increased precision of prevention, diagnosis and treatment of diseases. Single-cell genomics is a disruptive technology revolutionizing science by increasing the resolution with which we study diseases. Cell type specific changes in abundance or gene expression are now routinely investigated in diseases. Meanwhile, detecting cellular phenotypes across diseases can connect scientific fields and fuel discovery. Insights acquired through systematic analysis of high resolution data will soon be translated into clinical practice and improve decision making. Therefore, the continued use of single-cell technologies and their application towards clinical samples will improve molecular interpretation, patient stratification, and the prediction of outcomes.

In the past years, I was fortunate to participate in interdisciplinary research groups bridging biology, clinical research and data science. I was able to contribute to diverse projects through computational analysis and biological interpretation of sequencing data. Together, we were able to discover cellular phenotypes that influence disease progression and outcomes as well as the response to treatment. Here, I will present four studies that I have conducted in my PhD. First, we performed a case study of relapse from cell-based immunotherapy in Multiple Myeloma. We identified genomic deletion of the epitope as mechanism of immune escape and implicate heterozygosity or monosomy of the genomic locus at baseline as a potential risk factor. Second, we investigated the pathomechanisms of severe COVID-19 at the earliest stage of the COVID-19 pandemic in Germany in March 2020. We discovered that profibrotic macrophages and lung fibrosis can be caused by SARS-CoV-2 infection. Third, we used a mouse model of chronic infection with *Staphylococcus aureus* that causes Osteomyelitis similar to the human disease. We were able to identify dysregulated immunometabolism associated with the generation of myeloid-derived suppressor cells (MDSC). Fourth, we investigated *Salmonella* infection of the human small intestine in an *in vitro* model and describe features of pathogen invasion and host response.

Overall, I have been able to successfully employ single-cell sequencing to discover important aspects of diseases ranging from development to treatment and outcome. I analyzed samples from the clinics, human donors, mouse models and organoid models to investigate different aspects of diseases and managed to integrate data across sample types, technologies and diseases. Based on successful studies, we increased our efforts to combine data from multiple sources to build comprehensive references for the integration of large collections of clinical samples. Our findings exemplify how single-cell sequencing can improve clinical research and highlights the potential of mechanistic discoveries to drive precision medicine.

Kurzzusammenfassung

Interaktionen zwischen Wirt und Pathogen bestimmen die Entwicklung und den Verlauf von Erkrankungen als auch deren Ausgang. Die Medizin zieht Nutzen aus genaueren Beschreibungen von Krankheiten durch höhere Präzision von Prävention, Diagnose und Behandlung. Genomische Messungen in einzelnen Zellen werden durch innovative Technologien ermöglicht, welche die Wissenschaft revolutionieren indem sie die Auflösung erhöhen mit der wir Krankheiten untersuchen können. Inzwischen werden sowohl die Zusammensetzung von Zelltypen als auch Unterschiede in der Genexpression routinemäßig über Krankheiten hinweg untersucht. Der Einsatz von Technologien die einzelne Zellen untersuchen und ihre Anwendung auf klinische Proben wird die molekulare Interpretation, die Stratifizierung von Patienten und die Prognose des Ausgangs von Krankheiten verbessern.

In den letzten Jahren konnte ich mich an interdisziplinären Forschungsgruppen beteiligen und die Bereiche der Biologie, klinischer Forschung und Datenwissenschaften kombinieren. Ich war in der Lage zu unterschiedlichen Projekten beizutragen und eine führende Rolle in der Analyse und biologischen Interpretation von Daten aus Sequenzierungen zu übernehmen. Zusammen konnten wir zelluläre Phänotypen entdecken, die Entwicklung und Ausgang von Krankheiten sowie die Antwort auf Therapien beeinflussen. In dieser Arbeit werde ich vier Studien vorstellen, die ich während meiner Promotion durchgeführt habe. Zuerst haben wir einen Fall vom Rezidiv des Multiplen Myeloms nach zellulärer Immuntherapie untersucht. Dabei konnten wir feststellen, dass eine Deletion des genomischen Abschnitts für das immunogene Epitop dafür sorgte, dass die Krebszellen der Immunantwort entkommen konnten. Des weiteren konnten wir nachweisen, dass einige Patienten vor Beginn der Therapie nur eine Kopie des Gens besitzen und dadurch einen potentiellen Risikofaktor für ein Scheitern der Therapie. Zweitens haben wir im März 2020 die ersten Fälle von akutem Lungenversagen in COVID-19 und die Ursachen der Pathologie untersucht. Dabei haben wir festgestellt, das profibrotische Makrophagen und Lungenfibrose durch SARS-CoV-2 ausgelöst werden. Als Drittes haben wir Osteomyelitis in Mäusen untersucht, die von dem Bakterium *Staphylococcus aureus* ausgelöst wird und der Erkrankung im Menschen ähnlich ist. Wir konnten feststellen, dass deregulierter Metabolismus von Immunzellen der Enstehung von myeloiden Zellen mit T-Zell supprimierender Aktivität (MDSC) zugrunde liegt. Viertens haben wir die Infektion des humanen Dünndarms mit *Salmonella* in einem Organoidmodell untersucht und konnten Merkmale der Pathogeninvasion und der Wirtsantwort beschreiben.

Insgesamt konnte ich die Sequenzierung von RNAs in einzelnen Zellen nutzen um wichtige Aspekte in der Entwicklung, dem Verlauf und dem Ausgang von Erkrankungen zu entdecken. Ich konnte Proben aus der Klinik, von Donoren, Mausmodellen und Organoidmodellen analysieren und die Daten über die Art von Proben, Technologien und Krankheiten hinweg integrieren. Durch unsere erfolgreichen Studien konnten wir uns ambitioniertere Ziele setzen um Daten von verschiedenen Quellen in umfassenden Referenzen zusammenzuführen um große Kollektionen klinischer Proben gemeinsam zu untersuchen. Unsere Ergebnisse demonstrieren wie die Untersuchung einzelner Zellen die klinische Forschung verbessern kann und zeigt das Potential auf wie Entdeckungen in der Biomedizin zur Präzisionsmedizin beitragen können.

Acknowledgments

I want to thank my primary supervisor Antoine-Emmanuel Saliba and members of my thesis committee Roy Gross, Florian Erhard and Leif Erik Sander for the excellent guidance and moral support. I am grateful for the opportunity to perform research at the Helmholtz Institute for RNA-based Infection Research (HIRI) in Würzburg. As part of the Helmholtz Centre for Infection Research (HZI) I experienced an excellent environment to perform high-throughput sequencing studies and use very well established infrastructure. I am also thankful to belong to the graduate college (GRK) 2157 - 3D Infect at the University of Würzburg and appreciate the support from its head Thomas Rudel as well as the continued funding of our network by the German Research Association (DFG). I am much obliged to Emmanuel Saliba and Leif Sander as well as Malte Luecken and the Helmholtz Information and Data Science Academy (HIDA) for giving me the oppurtunity of research stays at the Charité Berlin and Helmholtz Munich to experience clinical science and data integration with machine learning from leading experts in the respective fields. I want to thank all my colleagues and collaborators working in the lab for all of our work builds on the biological samples, tissue models and experimental methods they establish, maintain and apply. Further I want to thank all colleagues that develop computational programs and share their work with all of us. My special appreciation goes to Tobi, Fabian and Nina for setting up and running our lab, to Matteo and Leo for motivating me with their enthusiasm, to Sophia, Anna and Daniel for the long nights in Berlin, to Tom and David for the opportunity to join their interesting work on organoids, to Kathrin and Mastura for the opportunity to set up experimental plans and Alex, Ariane and Vik for working together with me so well and staying to push it further, Christophe and Rob for the interesting discussions and especially Kathi for her never ending patience and support. I want to thank my lab and the whole institute for the great environment and hope that our social gatherings will continue. Finally, I want to thank my partner Katharina Wandera, my parents Ingrid & Frank Dietrich and siblings Isabella & Carina Dietrich, as well as all friends and colleagues without whom my life would be poorer and who kept me sane during times of hardship. In addition, I would like to acknowledge openAI’s tool DALL·E 2 for the creation of my cover art.

Danksagung

I möchte meinem Doktorvater Emmanuel Saliba und den Mitgliedern meines Dissertationskomitees Roy Gross, Florian Erhard und Leif Erik Sander für die exzellente Beratung und die moralische Unterstützung danken. Ich bin dankbar für die Gelegenheit meine Forschung am Helmholtz Zentrum für RNA-basierte Infektionsforschung (HIRI) in Würzburg durchzuführen. Als Teil des Helmholtz Zentrums für Infektionsforschung (HZI) habe ich eine hervorragende Umgebung genossen um Experimente durchzuführen die auf Sequenzierungen mit hohem Durchsatz basieren und konnte die sehr gut etablierte Infrastruktur nutzen. Ich bin dankbar dafür ein Teil des Graduiertenkolleg 2157 – 3D Infect an der Universität Würzburg zu sein und schätze besonders die Unterstützung der Vorsitzenden Thomas Rudel, sowie die beständige Finanzierung unseres Netzwerks durch die Deutsche Forschungsgemeinschaft (DFG). Ich bin Emmanuel Saliba und Leif Sander sowie Malte Luecken und der Helmholtz Informations- und Datenwissenschaften Akademie (HIDA) für die Möglichkeit von Forschungsaufenthalten an der Charité in Berlin und dem Helmholtz Zentrum München zu besonderem Dank verpflichtet. Dort konnte ich sowohl klinische Wissenschaft als auch Datenintegration durch maschinelles Lernen von führenden Experten in den jeweiligen Feldern lernen. I möchte all meinen Kollegen danken, die im Labor durch ihre Arbeit an biologischen Proben, Gewebemodellen und experimentellen Methoden die Grundlagen unserer gemeinsamen Forschung legen. Zudem möchte ich allen Kollegen danken, die an Computerprogrammen arbeiten und ihre Forschung mit uns allen teilen. Mein besonderer Dank geht an Tobi, Fabian und Nina, dass ihr unser Labor aufgebaut haben und es am Laufen haltet, an Matteo und Leo für die Motivation die ich aus eurem Enthusiasmus gezogen habe, an Sophia, Anna und Daniel für die langen Nächte in Berlin, an Tom und David für die Möglichkeit mich an eurer interessanten Arbeit mit Organoiden zu beteiligen, an Kathrin und Mastura für die Möglichkeit Projekte mit euch zu planen, Alex, Ariane und Vik für die gute Zusammenarbeit und dass ihr die Projekte weiter fortführt, Christophe und Rob für die interessanten Diskussionen und besonders Kathi für ihre unendliche Geduld und Hilfe. I möchte meiner ganzen Gruppe und dem ganzen Institut für die angenehme Umgebung danken und hoffe, dass unsere gemeinsamen Aktivitäten weiter bestehen. Zum Abschluss will ich meiner Partnerin Katharina Wandera, meinen Eltern Ingrid und Frank Dietrich und Schwestern Isabella und Carina Dietrich, sowie allen Freunden und Kollegen danken die mein Leben bereichern und die mich durch schwere Zeiten begleitet haben. Außerdem möchte ich openAI's DALL·E 2 für die Gestaltung meines Deckblatts anerkennen.

List of Figures

1.1 Precision medicine promises more effective treatments through identification of disease subgroups . . . 2
1.2 Schematic representation of the SARS-CoV-2 virion and structure of the RNA genome 3
1.3 Biofilm formation of *Staphylococci* . . . 4
1.4 Salmonella antigens that distinguish serovars . . . 4
1.5 Hallmarks of Cancer . . . 6
1.6 The gastrointestinal immune system . . . 7
1.7 Cell atlas of the human lung . . . 8
1.8 Advanced *in vitro* models of the small intestine serve as infection models . . . 9
1.9 Single-cell RNA-seq workflow . . . 10
1.10 Exponential scaling of scRNA-seq experiments . . . 11
1.11 Quality metrics vary between cell types . . . 12
1.12 Cell embeddings show similarity of cells . . . 13
1.13 Differential expression shows differences between cell types . . . 14
1.14 Milo reveals differential abundance of cell types . . . 15

3.1 Integration of intestinal tissue with intestinal organoids reveals maturation and fidelity 125
3.2 Longitudinal samples towards diverging outcomes can reveal distinguishing features 126
3.3 Sample embeddings reveal structure in data collections . . . 127
3.4 Reference design improves identification of disease states . . . 128
3.5 Disease dynamics can be discovered across patients and improve molecular interpretation, patient stratification, and outcome prediction . . . 129
3.6 Understanding cellular trajectories facilitates early disease detection . . . 131
3.7 Augmented reality brings interactivity to static documents . . . 132
3.8 Infection Atlas . . . 133

List of Tables

1.1 Fisher test reveals significantly enriched gene ontology (GO) categories 14

2.1 Research topics and cell sources for single-cell sequencing of the major projects. . 17

Contents

1 Introduction 1
- 1.1 Disease Research 1
  - 1.1.1 Hosts 2
  - 1.1.2 Pathogens 2
  - 1.1.3 Cancers 5
  - 1.1.4 Outcomes 5
- 1.2 Cell Sources 7
  - 1.2.1 Patients 7
  - 1.2.2 Donors 8
  - 1.2.3 Mouse Models 9
  - 1.2.4 Organoid Models 9
- 1.3 Single-cell technologies 10
  - 1.3.1 Microscopy 10
  - 1.3.2 Flow cytometry 10
  - 1.3.3 Sequencing 10
- 1.4 Data Analysis 11
  - 1.4.1 Generating the Count Matrix 11
  - 1.4.2 Quality Control 12
  - 1.4.3 Normalization 12
  - 1.4.4 Feature Selection and Summarization 12
  - 1.4.5 Data Integration 13
  - 1.4.6 Cell Type Annotation 13
  - 1.4.7 Differential Expression . . . 14
  - 1.4.8 Differential Abundance . . 15
- 1.5 Aims of the thesis 15

2 Results 17
- Homozygous *BCMA* gene deletion in response to anti-BCMA CAR T cells in a patient with multiple myeloma 17
- SARS-CoV-2 infection triggers profibrotic macrophage responses and lung fibrosis 34
- Dysregulated Immunometabolism Is Associated with the Generation of Myeloid-Derived Suppressor Cells in *Staphylococcus aureus* Chronic Infection 81
- A primary cell-based in vitro model of the human small intestine reveals host olfactomedin 4 induction in response to *Salmonella Typhimurium* infection 99

3 Discussion 123
- 3.1 Exploring Biological Systems . . . 123
  - 3.1.1 Defining Phenotypes 123
  - 3.1.2 Maturation & Fidelity 124
  - 3.1.3 Selecting Comparisons . . 125
  - 3.1.4 Determining Outcome . . . 125
  - 3.1.5 Functional Characterization 126
  - 3.1.6 Reproducibility 126
- 3.2 Genomic Mapmaking 127
  - 3.2.1 Reference Atlases 128
  - 3.2.2 Reference Usage 128
  - 3.2.3 Technological Advancements 129
- 3.3 Precision Medicine 130
  - 3.3.1 Defining Disease Phenotypes 130
  - 3.3.2 Predicting Disease Outcomes 130
  - 3.3.3 Targeting Treatments 131
- 3.4 Visualization & Design 131
  - 3.4.1 Infection Atlas 132
- 3.5 Solving Clinical Problems 132

Bibliography 135

Signed CV 143

Affidavit 145

Eidesstattliche Erklärung 145

Appendix 147
- Abbreviations 147
- Author contributions to manuscripts . . 149
- Author contributions to figures/tables . 151
- List of Publications 153

1 Introduction

The human body consists of approximately three trillion (3×10^{12}) nucleated cells which are outnumbered by roughly thirty-eight trillion (3.8×10^{13}) microbes [1]. Bacteria, the majority of microbes in the human body, are estimated to comprise more than 10 000 species [2]. While most bacteria live in symbiosis with their human host, an estimated 1513 species can cause disease in humans. These pathogens come from 24 different classes and show large genomic diversity [3]. In contrast, all cells from a human organism arise from a single cell passing on the genomic information to its progeny. The billions of cells arising from it, however, show remarkable phenotypic diversity and arrange in a complex anatomy of tissues and organs. Interactions between cells of the host, its natural microbiome and pathogens determine the maintenance of health and development of disease. Understanding these interactions can help us to build a healthier and safer world [4].

1.1 Disease Research

Disease Research bridges the fields of biology and medicine and strives to create better health for all humans. In addition to improving health, research should also serve society by catalyzing new scientific fields, tools and resources as well as in terms of the economy [5]. In the last years, the COVID-19 pandemic has highlighted that our society strongly depends on sophisticated means for the prevention, diagnosis and treatment of infectious diseases [6].

Prevention of infectious disease is the role of hygiene. This includes the identification and killing of pathogens, the disruption of transmission and protection by vaccination as well as the reduction of spread by education and law [7]. During the COVID-19 pandemic, masks were used to reduce transmission [8] and rapid vaccine development led to multiple highly effective vaccines, especially in the prevention of severe disease [9]. Moreover, governments restricted the free movement of people to limit exposure and prevent the exponential spread of disease [10].

Diagnostics refers to the identification of diseases. Some diseases can be identified based on clinical signs and symptoms (e.g. *Rubella, Measles, chicken pox*). Other diseases induce less distinct symptoms but lead to drastically different outcomes (e.g. Rhinoviruses and COVID-19) and thus require more sophisticated methods of identification [7]. Therefore, clinical chemistry (e.g. serum iron concentration), radiology (e.g. computed tomography, CT), pathology and microbiological diagnostics are often used to identify infectious diseases [7]. During the pandemic, three major methods have been used for the detection of SARS-CoV-2, the causative agent of COVID-19. Rapid detection of viral proteins as a lower-sensitivity assessment of the potential for transmission. Polymerase chain reaction (PCR)-based quantification of viral RNA with high sensitivity and specificity used for confirmation of diagnosis in symptomatic individuals and for activating public health measures. Antibody tests, measuring the immune response, with unclear association to the prevention of spread and disease. Combined use of those tests can inform public policy but requires careful consideration of purpose and resources [11].

Treatments for infectious diseases usually target structures and metabolic pathways of the pathogen that differ to those of the host. Bacterial infections are often treated with chemotherapy agents commonly referred to as antibiotics. A prominent example are synthetic antibiotics such as Sulfonamides, that inhibit folate synthesis in bacteria [7]. Therapeutic approaches for COVID-19 include antiviral chemotherapy, such as Remdesivir, a nucleotide analog that inhibits viral replication, but also immunotherapies have been proposed as treatment options [12]. Similar groups of therapeutics to the antibacterial and antiviral chemotherapy as well as immunotherapy are also used in the treatment of cancer. In analogy to Sulfonamides, cancer chemotherapy uses antifolates to inhibit cell division. Historically, surgery, chemotherapy and radiotherapy are the main pillars of cancer therapy but in recent years targeted and immunother-

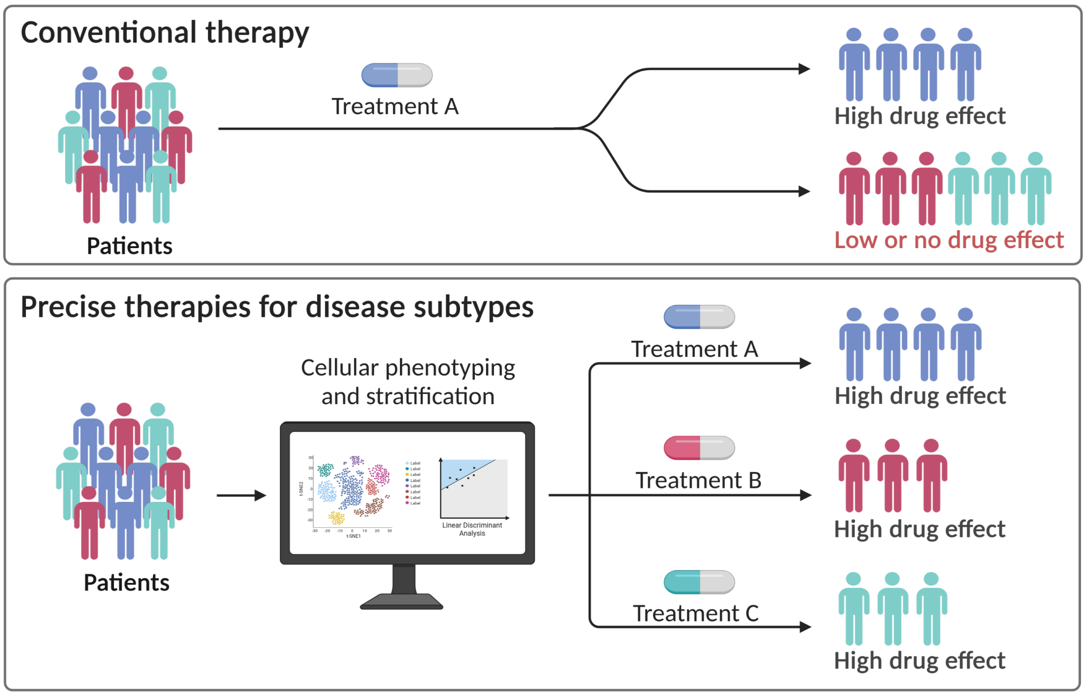

Conventional therapy
Patients
Treatment A
High drug effect
Low or no drug effect
Precise therapies for disease subtypes
Patients
Cellular phenotyping
and stratification
Linear Discriminant Analysis
Treatment A
High drug effect
Treatment B
High drug effect
Treatment C
High drug effect

Figure 1.1: Precision medicine promises more effective treatments through identification of disease subgroups. Adapted from "Precision Cancer Therapy" by `BioRender.com` (2023). Retrieved from `https://app.biorender.com/biorender-templates`.

apies have revolutionized the field [13].

Precision medicine, also referred to as targeted therapy or deep phenotyping, is based on patient stratification. Subgroups of patients that have a common basis of disease susceptibility and manifestation can receive more defined diagnoses and treatments (see figure 1.1) [14]. However, dividing diseases into subgroups requires an accurate description of the mechanistic principles that distinguish disease entities. Systematic analyses of host-pathogen interactions with high resolution can provide these descriptions, but they require detailed knowledge of hosts, pathogens, and potential outcomes.

1.1.1 Hosts

The primary host in focus of disease research are human individuals. While patients are studied with the intention of restoring health, donors support research through contribution of tissue samples. Studying animals as hosts for infections can be important in terms of disease reservoirs and the economic burden of disease. However, a major focus lies on the use of animals as disease models that facilitate understanding of the human disease [15].

The defensive capabilities of the host impact the clinical picture of disease and is influenced by genetic disposition, age, disease history, vaccination status, social status, comorbidities, and individual environmental factors. This explains why not every infection leads to infectious disease and how the same pathogen does not induce the same symptoms and affect the same organs in different patients [7]. While diagnosis of an infectious disease during an epidemic or pandemic is relatively easy, providing direct or indirect evidence for such a disease can require extensive clinical expertise and technical cost. In addition, detection of microbial agents is not necessarily a sign of involvement in the infectious process but might be a natural constituent of the host microbiome [7].

1.1.2 Pathogens

Infections can be caused by prions, viruses, bacteria, fungi, protozoa, helminths and arthropods. Infectious agents often enter via the mucosal surfaces of the eye, oropharynx, respiratory-, gastrointestinal- or urogenital tract. Moreover, arthropods can serve as vectors that transmit viruses, bacteria or protozoa into tissues or the bloodstream of their host. Pathogen quantity, pathogenicity and virulence impact the development of disease. Contagiousness refers to the ability of a pathogen to elicit an infection and can be measured using the minimal infectious dose. Spread of the pathogen depends both on host defense and virulence factors of the infectious agent, such as toxins or adhesion factors [7]. Further pathogen determinants are cell type tropism, the ability to circumvent innate immune response and antigenic immunodominance. While genetic instability of microorganisms, especially RNA viruses, facilitates rapid evolution and adaptation to changing environments [16].

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

A new virus species emerged in December 2019 in Wuhan, China and was termed SARS-CoV-2 due to its genetic similarity to severe acute respiratory virus coronavirus (SARS-CoV), a highly pathogenic coronavirus with zoonotic origin [17]. The coronavirus disease 2019 (COVID-19), caused by SARS-CoV-2, is a highly transmissible viral pneumonia which has caused a pandemic, marked by worldwide spread and a large number of infected individuals. Overall, estimates from November 2021 show a mortality rate of 2 % for infected individuals with significant morbidity arising from pneumonia, long term intensive care, and a range of symptoms referred to as long Covid/post Covid syndrome [7, 18].

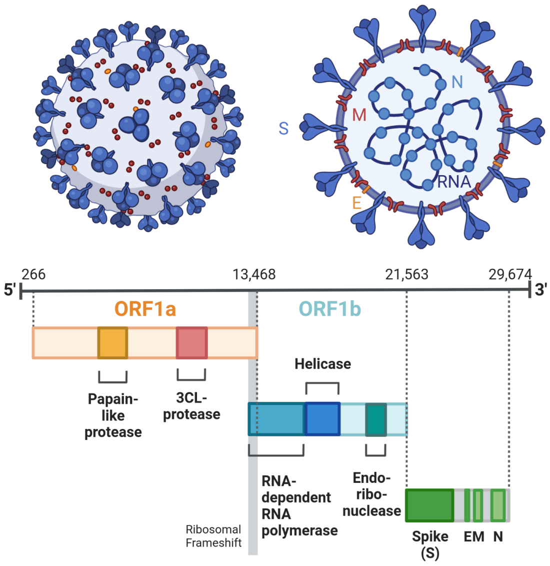


Figure 1.2: Schematic representation of the SARS-CoV-2 virion and structure of the RNA genome. Created with `BioRender.com`.

The fast evolution of this RNA virus (see figure 1.2) resulted in a number of variants that have led to concern of regulatory agencies. Changed properties of the pathogen include increased transmission and impact on immunity that governs the effectiveness of vaccines. For example, the delta variant shows mutations of the spike protein which increased transmission approximately 40-60 % and caused light reduction of vaccine efficacy compared to the earlier alpha variant [7]. However, attributing single mutations (e.g. D614G) to virulence is difficult and might not be associated with more severe disease [18].

SARS-CoV-2 pathogenesis depends on infection of epithelial cells in the upper and lower respiratory tract. While virus entry to epithelial cells is well described as a pathway dependent on ACE2 and TMPRSS2, it is important to note that vascular endothelial cells and pulmonary macrophages are also affected. In addition, extra-pulmonary manifestations have been described for numerous organs including the intestine, kidney, liver, vasculature, nervous system and heart [7].

Only a minority of infected people develop severe disease, characterized by the requirement of ventilation in an intensive care setting. Among those, the critical cases show respiratory failure, septic shock and/or multiple organ dysfunction/failure. While lymphopenia is common, bacterial superinfections remain rare and are overshadowed by a higher incidence of pulmonary aspergillosis [7, 18].

Treatment of COVID-19 focuses primarily on measures supporting vitality according to the severity of disease. Actions include administration of oxygen and compensating water and electrolyte balance [7]. Direct antiviral therapy focuses on the prevention of virus entry and suppression of virus replication facilitated through monoclonal antibodies and Remdesivir [7, 12].

Prevention of disease is favorable and can be achieved through preventive measures and vaccination. Multiple mRNA and vector-based vaccines for COVID-19 have been developed that are principally focused on the spike protein of SARS-CoV-2 [7].

Staphylococcus aureus

The *Staphylcoccus* genus consists of atrichous, non-motile, gram-positive cocci that form dense aggregates. Many species belong to the natural flora of the human skin and mucosa. *Staphylococci* have been recognized to cause nosocomial infections due to their ability to form biofilms (see figure 1.3), for example on catheters [7].

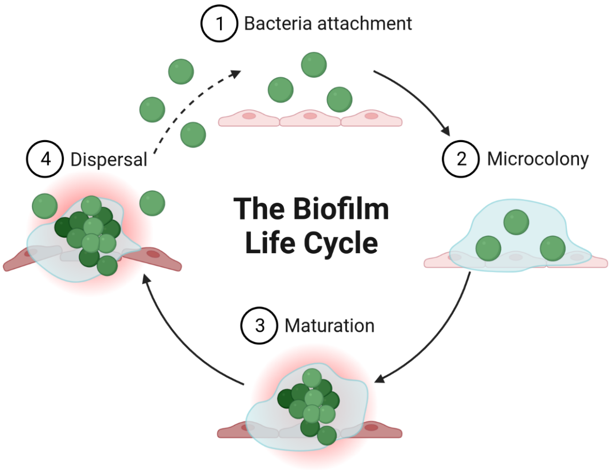


Figure 1.3: Biofilm formation of *Staphylococci*. Adapted from "The Biofilm Life Cycle" by `BioRender.com` (2023). Retrieved from `https://app.biorender.com/biorender-templates`.

S. aureus is generally distinguished from other *Staphylococci* based on the presence of coagulase, a virulence factor that induces clotting in human serum. Besides, many virulence factors exist in this species. A polysaccharide capsule and protein A can protect from phagocytosis while secreted proteins such as hemolysin and toxins induce tissue damage and shock syndromes. The repertoire of virulence factors in each strain determines its pathogenic potential and while some are mildly aggressive others can be extremely dangerous [7].

Antimicrobial resistance is widespread across strains. Methicillin resistant *S. aureus* (MRSA) are particularly important as causative agents of nosocomial infections. While skin infection leads to usually mild abscesses, systemic spread leading to infection of inner organs is possible and can result in osteomyelitis, endocarditis, septicemia or toxic shock [7].

Salmonella

The genus *Salmonella* consists of peritrichous, rod shaped, gram-negative enterobacteria. The species *Salmonella enterica* contains all medically relevant strains and consists of 2500 subtypes, named serovars. These serovars can be distinguished by patterns of surface (O), flagellar (H) and capsule (K) antigens (see figure 1.4). Medically relevant is the distinction in enteric and typhoid Salmonellosis based on the relatively harmless enteritis of the former and severe, septic, and cyclic infections of the latter [7].

Typhoid fever in immunocompetent hosts is caused only by the subspecies *enterica* serovars Typhi and Paratyphi A, B, and C. Infection occurs orally by uptake with water or food. Even a low infectious dose (>100 agents) can lead to invasion of the epithelium. From the epithelium, pathogens can migrate via lymph nodes to distant organs (e.g. bronchus, spleen, infectious metastases in the skin) or replicate in gut-associated lymphoid tissues, which can lead to necrosis and perforation of the intestine. Disease usually sets on after an incubation period of two weeks and ends after five weeks. Treatment, e.g. Cipro-floxacin and Ceftriaxon, can reduce the lethality of typhoid fever (e.g. pneumonia, myocarditis, toxic circulatory collapse) from 15 % to 1-2 %. However, there is significant risk of relapse, osteomyelitis or constant shedding of bacteria from reservoirs in the gall bladder. Vaccines are available but only confer weak immunity with uncertain protection [7].

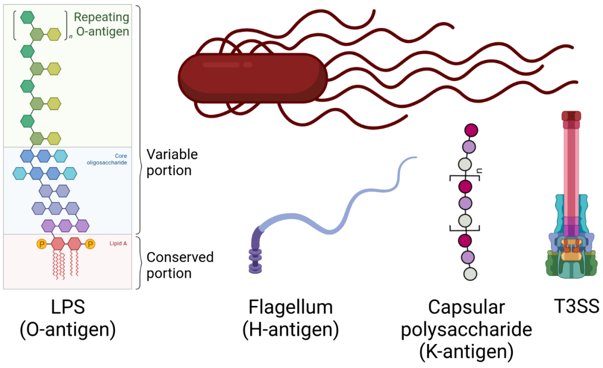


Figure 1.4: Salmonella antigens that distinguish serovars. The type III secretion system (T3SS) is an important virulence factor for the secretion of toxins. Created with `BioRender.com`.

Enteric Salmonellosis can be caused by all other *Salmonella* serovars (e.g. *S. enterica* serovar Enteritidis) which enter the host mostly through food with a high infectious dose ($>10 \times 10^5$ agents). The enteritis also arises from invasion of the intestinal epithelium but usually stays localized, is temporally more restricted and although hematological spread occurs in approximately 20 % of cases the untreated disease lethality is very low [7].

S. enterica serovar Typhimurium is a non-typhoidal serovar used as a model system for *Salmonella* research, especially in mouse models. While it causes gastroenteritis in humans, mice develop typhoid fever-like systemic disease. Many basic concepts of *Salmonella* pathogenesis have been established based on this model. Among them the role of the type III secretion system (T3SS) in inducing intestinal inflammation. Even though disease presentation is substantially different, at least some of the basic concepts should apply to the understanding of human disease [19].

1.1.3 Cancers

Cancer is a term for a large collection of diseases that arise from normal human cells through a multi-step process. A large body of research describes the identification of cancers, the mechanisms behind, and the development of therapies against them [13]. Conceptual progress came with definition of the six hallmarks of cancer. These distinctive and complementary capabilities comprise sustaining proliferative signaling, evading growth suppressors, activating invasion and metastasis, enabling replicative immortality, inducing angiogenesis, and resisting cell death. This view also embraced heterogeneity and described tumors as complex tissues [20]. Over time, emerging hallmarks such as the deregulation of cellular energetics, avoiding immune destruction, promoting inflammation, and genome instability and mutation were added. This highlighted both the impact of the tumor microenvironment and opportunities for therapeutic targeting (see figure 1.5) [21].

Cancers of the blood, also called hematological malignancies, were among the first to be targeted by immunotherapies due to their clearly defined populations distinguishable by epitopes referred to as clusters of differentiation (CD) [13]. Novel therapies for hematological malignancies have entered the clinics over the last decade and led to frequent shifts in the standard of care [22].

Multiple Myeloma

Multiple Myeloma accounts for approximately $10\,\%$ of hematological malignancies and is a cancer of plasma cells. Plasma cells are bone marrow derived lymphocytes that secrete antibodies as part of the adaptive immune system. The disease is defined by large expansion ($> 10\,\%$) of clonal plasma cells in the bone marrow and clinical symptoms such as hypercalcemia, renal failure, anemia, or lytic bone lesions [22, 23].

Multiple Myeloma has shown how therapeutic advances can significantly increase survival. However, the occurrence of relapse or refractory disease is common and highlights both difficulties in the choice of treatment regimens and opportunities of novel agents. Chimeric antigen receptor T cells (CAR-T) targeting B-cell maturation antigen (BCMA) have been proposed as an exciting new option to treat relapsed Multiple Myeloma [22].

1.1.4 Outcomes

Immunity refers to the protection from disease which is ensured through collective and coordinated response by the immune system [23]. The ideal outcome of disease is complete recovery of the damaged tissues and organs as well as the induction of lasting immunity to the causative agent.

Vaccination effectively induces immunity without causing disease and can prevent damage associated with the disease [23].

Mild disease usually weakens the host but does not cause symptoms that require clinical presentation. Tissue damage often recovers but pathogen persistence and predisposition to disturbance of organ function is possible. One example is the self-limiting enteritis caused by non-typhoidal *Salmonella* serovars that can lead to shedding of bacteria over years and irritable bowel syndrome (IBS) [7].

Severe disease is often life threatening and requires clinical presentation. Systemic spread and/or involvement of major organs are frequent and lead to high morbidity and mortality. While some patients die from disease others recover but the risk of long-term consequences is increased [7].

Deficient, hyperactive or deregulated immune systems are associated with multiple diseases and can increase susceptibility to infectious diseases. Hypersensitivity reaches from allergies (type I) over the deposition of immune complexes that can induce complement activation and tissue damage (type III) to T cell-mediated inflammation (type IV) and autoimmunity [7].

Diseases caused by the same pathogen do not necessarily develop in the same way. The disease can develop different between individuals, lead to varying outcomes, and potentially increase susceptibility to other conditions. Interactions between host cells and

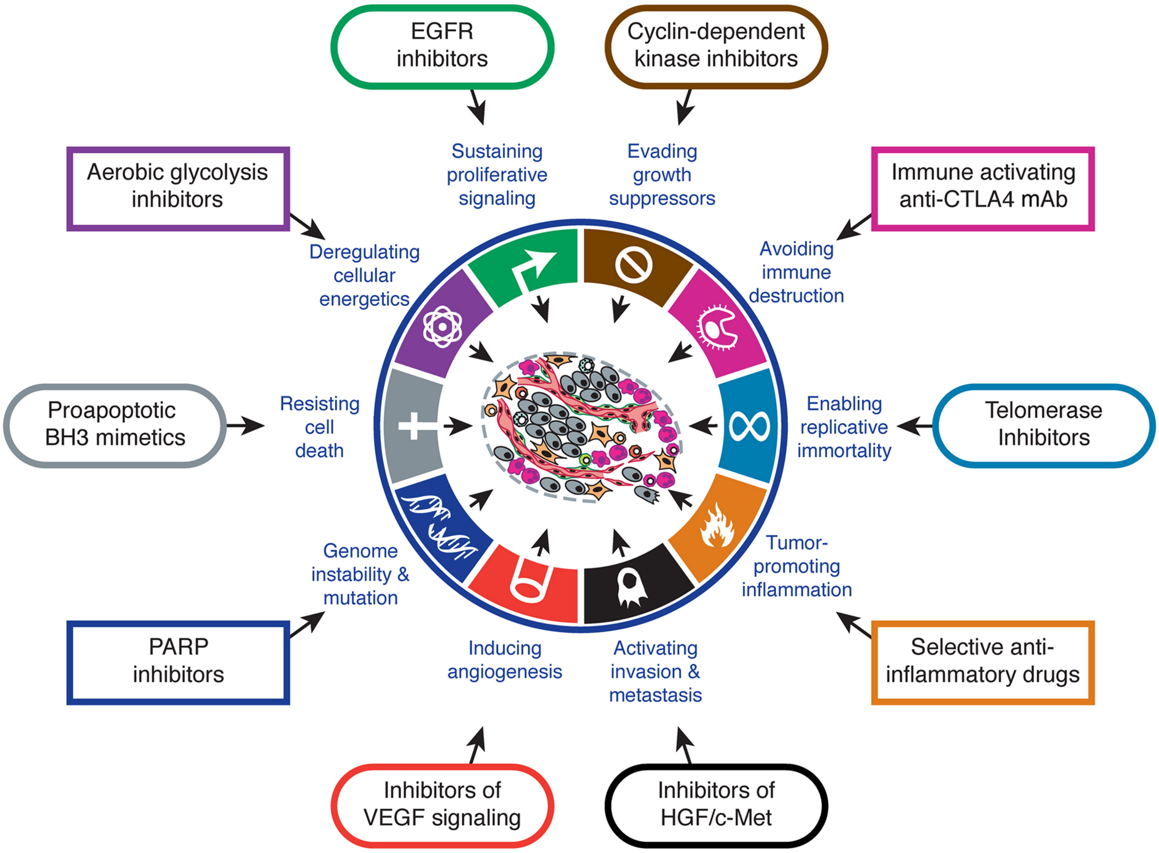

EGFR inhibitors
Cyclin-dependent kinase inhibitors
Aerobic glycolysis inhibitors
Sustaining proliferative signaling
Evading growth suppressors
Immune activating anti-CTLA4 mAb
Deregulating cellular energetics
Avoiding immune destruction
Proapoptotic BH3 mimetics
Resisting cell death
Enabling replicative immortality
Telomerase Inhibitors
Genome instability & mutation
Tumor-promoting inflammation
PARP inhibitors
Inducing angiogenesis
Activating invasion & metastasis
Selective anti-inflammatory drugs
Inhibitors of VEGF signaling
Inhibitors of HGF/c-Met

Figure 1.5: Hallmarks of Cancer highlight opportunities for therapeutic targeting. Reprinted from [21].

pathogens determine disease development and shape outcomes. A deeper understanding of these interactions can potentially reveal decision points that can be exploited for prevention, diagnosis or treatment. However, in order to truthfully interpret these interactions, it is crucial to understand the ontogeny of a cell and the environment it derives from.

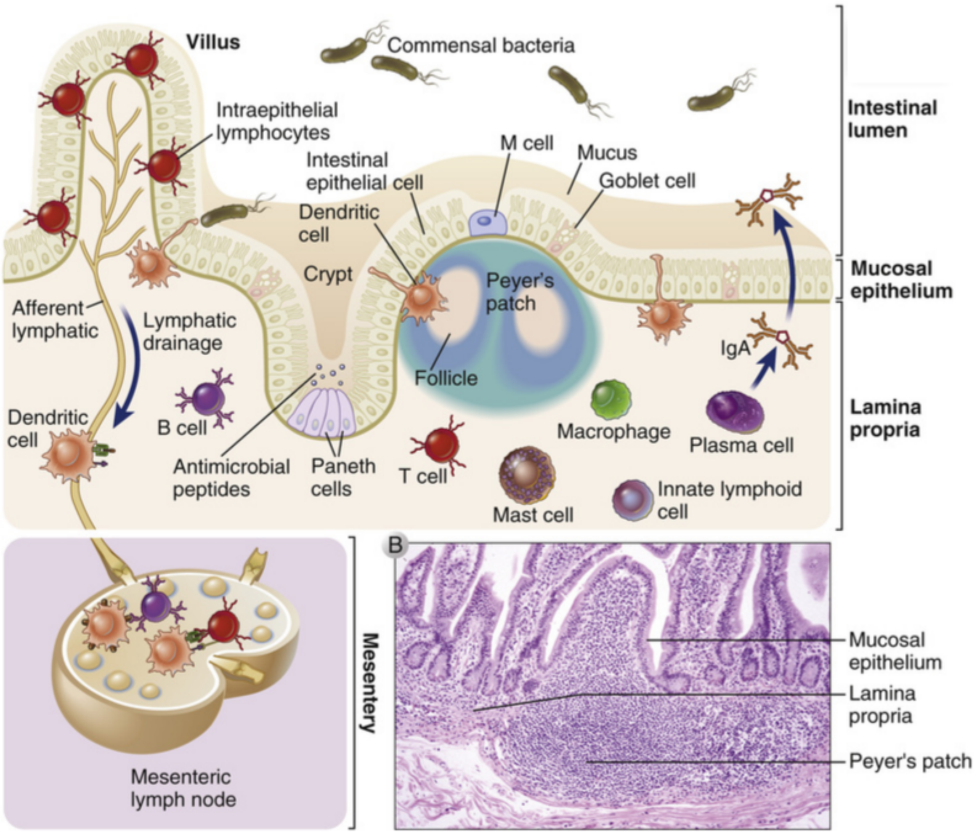

Villus
Commensal bacteria
Intestinal lumen
Intraepithelial lymphocytes
Intestinal epithelial cell
M cell
Mucus
Goblet cell
Dendritic cell
Crypt
Peyer's patch
Mucosal epithelium
Afferent lymphatic
Lymphatic drainage
IgA
Follicle
Lamina propria
Dendritic cell
B cell
Macrophage
Plasma cell
Antimicrobial peptides
Paneth cells
T cell
Mast cell
Innate lymphoid cell
Mesenteric lymph node
Mesentery
B
Mucosal epithelium
Lamina propria
Peyer's patch

Figure 1.6: The gastrointestinal immune system. Reprinted from [23].

1.2 Cell Sources

Infection and host response occur in highly complex tissues in which individual cells respond differently to the pathogens they encounter (see figure 1.6) [23]. Patients often present with a clinical picture that is characterized by the affected organs rather than the pathogenic agent and tissue samples are taken for diagnosis and identification of pathogens [7]. Observation of diseases in the clinical context is possible by investigating those samples. However, many tissues cannot be obtained and mild cases might not show the clinical need for such a procedure. Animals can develop diseases similar to those of humans. Therefore, modeling diseases in animals can be used to gain better access to tissues as well as testing and monitoring of novel treatments [24]. Recapitulating tissues in advanced culture systems opens up the possibility to study host-pathogen interactions in a closely defined cellular environment [25].

The study of diseases usually entails comparisons between cases and controls. However, cells have a certain state dependent on their species, tissue origin or culture condition. Knowledge about these factors is essential to determine how a useful control is obtained and which disease conditions should be investigated. In the following, I want to highlight the context from which biological samples are derived.

1.2.1 Patients

Recovery of clinical samples is generally performed for microbiological examination to clarify diagnosis. Blood, urine, stool, pus, exudate, transudate and tissue biopsies are regularly taken for analytical tests. The sample origin, time, amount, sterility, storage, transport and identification impact the overall usefulness. Selecting a sample type and processing workflow therefore depends on the requirements of analysis methods to obtain effective results [7].

Lung

Lung samples from infectious diseases can be obtained for microbial diagnostic in form of sputum or tracheal exudate, protected brush or bronchoalveolar lavage (BAL) [7]. Normal lung tissue can be obtained from uninvolved regions of surgical lung resection in cancer patients. A first atlas of the molecular cell states of the human lung was released in 2020 showing large epithelial, endothelial, stromal and immune cell diversity (see figure 1.7) [26].

Functional measurements and medical imaging techniques, such as computed tomography (CT), can inform further about the extent of disease-related changes [7].

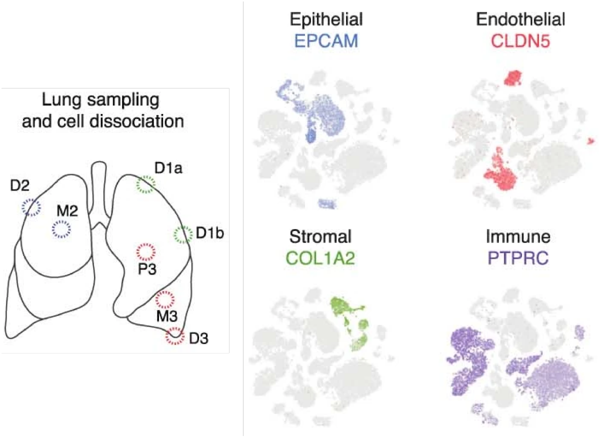


Figure 1.7: Cell atlas of the human lung. Adapted from [26].

1.2.2 Donors

Tissue samples from healthy individuals make the investigation of human cells possible in a controlled environment. Understanding the reaction of mature cells to infectious and non-infectious agents is important to put observations from patients into context. Obtaining donor material is possible but requires adherence to ethical standards and consent [27].

Blood

Healthy humans voluntarily donate blood both to help patients and for biomedical research [28]. Therefore, leukocytes can be used for research purposes. Gathering samples periodically from the same donors is possible and facilitates experimental setups with paired samples.

Neutrophils Neutrophils are the most abundant leukocyte in the human blood and can thus be isolated from peripheral blood samples. They are short-lived phagocytes with a variable nuclear morphology (polymorphonuclear leukocytes, PMN) that contain microbicidal substances [23]. Long considered a homogeneous and transcriptionally inactive population a recent analysis has revealed larger heterogeneity including the presence of a small subset (7 %) with higher baseline expression of interferon (IFN) inducible genes [29].

Monocytes & Macrophages Monocytes are after neutrophils and lymphocytes the third most abundant leukocytes. They are phagocytes that derive in the bone marrow, travel through blood and differentiate into macrophages when they migrate into tissues. Multiple subsets exists in the human blood that are phagocytic and produce inflammatory mediators (classical) or contribute to tissue repair (nonclassical) after recruitment to sites of infection or tissue injury. In contrast, tissue resident macrophages mostly derive from hematopoietic progenitors from fetal life. Macrophages are long-lived, ingest both microbes and necrotic cells, present antigens to lymphocytes, secrete cytokines to recruit monocytes and stimulate angiogenesis and fibrosis depending on their activation [23].

Intestine

The intestinal crypt represents the prototype of a stem cell compartment contained in a simple, repetitive architecture. Crypt base columnar (CBC) stem cells, fueled by Wnt signaling and epithelial growth factor (EGF), give rise to transit amplifying (TA) cells that largely determine crypt output. Among the progeny, secretory versus enterocyte fate is largely controlled by Notch signaling while the Bone Morphogenic Protein (BMP) signaling pathway acts as a negative regulator of crypts and thus drives differentiation. Enterocytes, absorptive and highly polarized columnar cells with an elaborate brush border, represent the default fate. Paneth cells are key determinants of stem cell niche and numbers and secrete bacteriocidal products. Goblet cells secrete mucus while enteroendocrine cells secrete a variety of hormones. Tuft cells are implicated in sensing of lumenal contents and microfold (M) cells reside in specialized lymphoid accumulations important for mucosal immunity [30].

Adult stem cells from the intestinal crypt can be isolated from surgically resected intestinal tissues or endoscopic biopsy samples from patients or during screenings. Normal tissue is taken from sites close to the disease site but separated from it by a certain margin (e.g. 3 cm) [27]. In addition, intestinal stem cells can

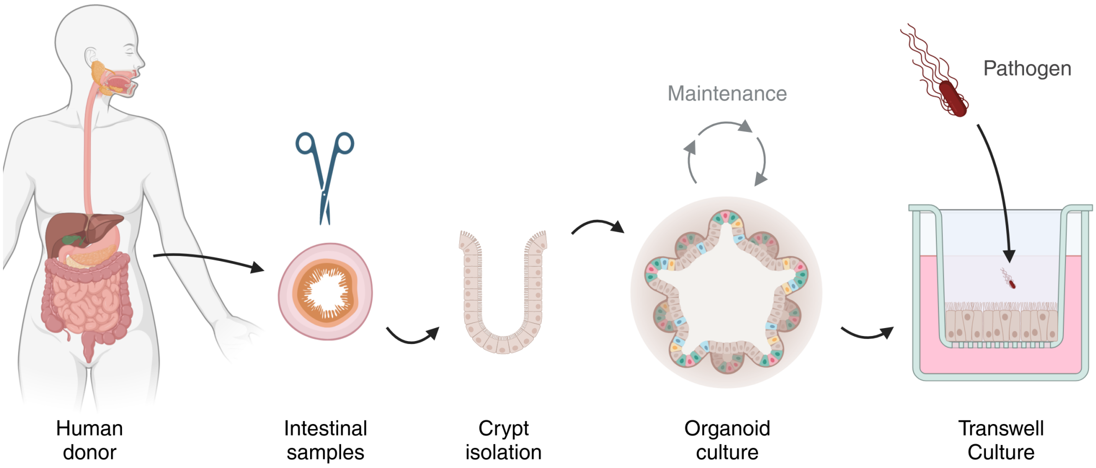


Figure 1.8: Advanced *in vitro* models of the small intestine serve as infection models. Created with `BioRender.com`.

also be derived from induced pluripotent stem cells (iPSC) [31].

However, reducing the gastrointestinal epithelial barrier to the epithelium would not give justice to the largest and most complex regional immune system of the human body (see figure 1.6) [23]. Choosing which components of a system to study is therefore an important task when working with biological models.

1.2.3 Mouse Models

Mice can be used as experimental models if disease pathogenesis closely mimics the human disease. For this, important features of the human disease must be precisely reproduced in the animal. The model can then be used for more thorough investigation of disease pathogenesis due to the availability of tissue samples and the application of treatments and novel imaging approaches [24]. If the murine disease does not capture the salient features of human disease closely enough, translational effectivity is low and carries large costs to the clinics [32].

1.2.4 Organoid Models

Recapitulating tissue architecture *in vitro* fuels understanding of homeostasis while facilitating experimentation for disease research. Early studies showed, that single Lgr5$^+$ CBC stem cells from the small intestine of mice can be cultured long term *in vitro* and build crypt-villus structures without a mesenchymal niche [33].

Organoids can be derived from both adult stem cells (ASC) and induced pluripotent stem cells (iPSC) [31]. However, recapitulating the intestinal crypt *in vitro* comes with reduced maturation and fidelity of cell types when compared to the tissue *in vivo* [34]. Cellular diversity and culture efficiency can potentially be improved by niche-inspired culture conditions for both ASC- and iPSC-derived organoids [35, 36].

The culture conditions for organoids depend on research use. Three-dimensional culture in a biological matrix supports long-term culture of cells. Seeding organoids in two dimensional culture plates, however, can make them more accessible to infection studies (see figure 1.8). Engineering systems for tissue culture can be used to further advance these models [25].

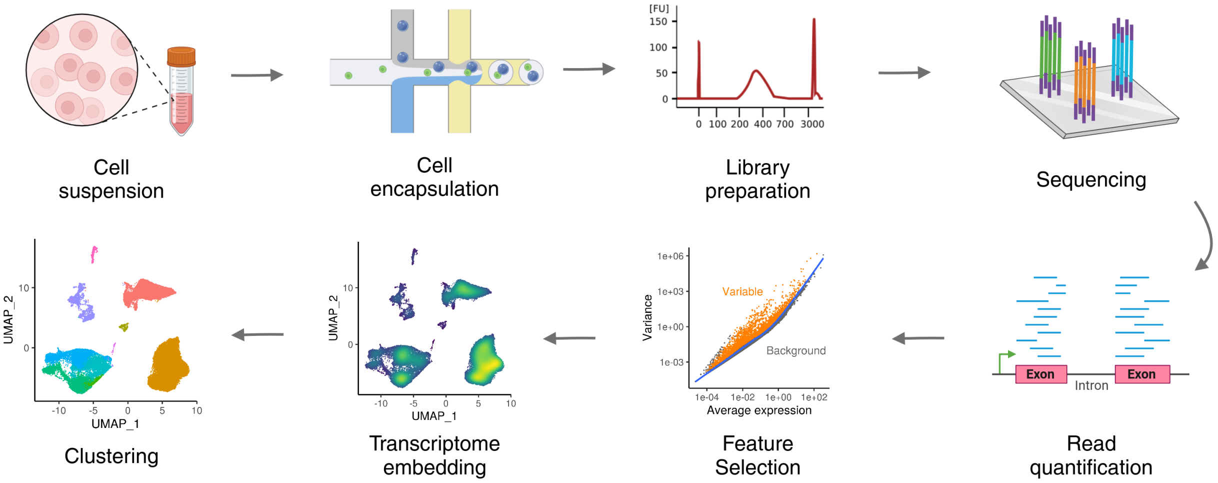

Cell
suspension
Cell
encapsulation
[FU]
150
100
50
0
0 100 200 400 700 3000
Library
preparation
Sequencing
UMAP_2
10
0
-10 -5 0 5 10
UMAP_1
Clustering
UMAP_2
10
0
-10 -5 0 5 10
UMAP_1
Transcriptome
embedding
1e+06
1e+03
1e+00
1e-03
Variance
Variable
Background
1e-04 1e-02 1e+00 1e+02
Average expression
Feature
Selection
Exon
Intron
Exon
Read
quantification

Figure 1.9: Single-cell RNA-seq workflow. Created with `BioRender.com`.

1.3 Single-cell technologies

Individual cells are the fundamental unit of biology. Whether they live in isolation, as colonies or form a single organism with their sister cells, the translation from genetic code into phenotypes occurs at the cellular level. Resolution matters when we want to ask questions about the behavior of cells. When cells of different types migrate through tissues, meet, and touch to exchange information, the ability to identify them is crucial to our understanding of the interactions.

Over the last decade, technological advances in microscopy, microfluidics and high-throughput sequencing have revolutionized the resolution with which we can interrogate single cells. The number of cells in scRNA-seq experiments has grown exponentially in the first decade [37]. Soon thereafter multiple modalities were sequenced and integrated during analysis [38]. Now, the potential to integrate single-cell sequencing data with spatial resolution promises advancement of our understanding of health and disease [39].

1.3.1 Microscopy

Light microscopy has evolved from the standard widefield and confocal microscopy techniques to breakthroughs of fluorescence microscopy with the invention of super-resolution microscopy. The most recent development of lattice light-sheet microscopy (LLSM) extends super resolution to volumetric samples [40]. In a similar fashion, the innovation of focused ion beam scanning electron microscopy (FIB-SEM) has addressed the limitation of three dimensional (3D) acquisition in transmission and scanning electron microscopy techniques [40]. Measurement of cellular phenotypes using spatial transcriptomics has been improved by high-throughput fluorescence in-situ hybridization (FISH) measuring 10 000 genes with single-molecule resolution [41, 42].

1.3.2 Flow cytometry

Individual cells in suspension can be labelled with fluorescently labelled probes specific for expressed molecules that inform about the tissue lineage, maturation stage, or activation status and measured using a flow cytometer. Relative amounts of molecules can thus be determined by the amount of emitted fluorescence. While permeabilization permits intracellular staining, epitopes can be stained on live cells that can be sorted using a fluorescence-activated cell sorter (FACS). The defined subpopulations obtained using FACS can be used for other applications, including sequencing [23].

1.3.3 Sequencing

Sequencing describes the identification of the order of monomers in a biopolymer, such as nucleotides in a nucleic acid. The three generations of high-throughput sequencing refer to Sanger dideoxy sequencing, sequencing by synthesis (next-generation sequencing, NGS), and long-read sequencing. Typical tradeoffs in the choice of sequencing platform include accuracy, number of reads, read length, single read or paired end, RNA or DNA, amount of material, cost, and time [43].

RNA-seq

RNA sequencing (RNA-seq) describes a collection of methods, experimental and computational, to determine the identity and abundance of nucleic acid sequences in biological samples. A common goal of RNA-seq is the quantification and comparison of gene expression, however, also the identification of gene structure, novel genes, expression quantitative trait loci (eQTL), gene variations, long and small non-coding RNAs are possible [43].

RNA-seq can be used to profile both the transcriptome of the host and pathogens at the same time. This makes it possible to study host-pathogen interactions with increased resolution [44]. However, transcripts are still measured over whole populations of cells, potentially obscuring biologically interesting variation.

scRNA-seq

Single-cell RNA-sequencing (scRNA-seq) is a revolutionary technology that made it possible to obtain genomic measurements from individual cells. Therefore, the heterogeneity of cell populations in biological samples can be assessed using transcriptional profiles. This has led to widespread adoption of the techniques exemplified by exponential growth of the number of profiled cells during the first decade [37].

Technological improvements have facilitated transcriptome-wide analyses of individual cells based on cell isolation, lysis, transcription of RNA into complementary DNA, amplification and generation of sequencing libraries (see figure 1.9). Multiple technologies with different cell isolation and library generation have been developed that differ in sensitivity and throughput [45]. While plate-based Smart-seq2 is the most sensitive method [46], droplet-based approaches like Chromium [47] have higher throughput (see figure 1.10). A recent benchmark of methods confirmed the high performance of both methods, especially in light of large integrated data sets and atlases [48].

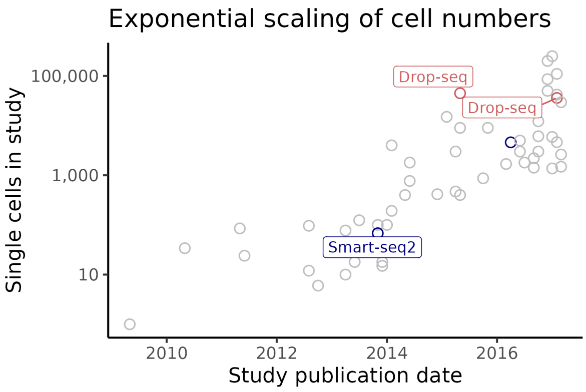


Figure 1.10: Exponential scaling of scRNA-seq experiments. Data retrieved from [37].

Multiple Modalities

Sequencing can measure multiple aspects of the cellular phenotype at the same time. Cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) captures both polyadenylated RNAs and DNA-barcoded antibodies to increase the information content and connect scRNA-seq to flow cytometry [49]. A similar approach can be used for sample multiplexing and doublet detection [50].

1.4 Data Analysis

Investigating biological processes through the lens of single-cell genomics requires extensive computational workflows that facilitate exploratory analysis of the data.

Kiselev, Andrews and Hemberg provided an early description of the challenges in unsupervised clustering of scRNA-seq data such as the curse of dimensionality, different clustering algorithms, discrete vs. continuous groupings, and batch effects as well as highlighting cell ontologies in atlas projects [51]. An updated version by Luecken and Theis revisits the aforementioned steps but adds discussion of downstream tasks such as trajectories and compositional analysis as well as common pitfalls and clear recommendations [52].

1.4.1 Generating the Count Matrix

The process of read demultiplexing, quality control (QC), alignment, counting and initial QC are generally performed using preset computational pipelines. The most popular tool, developed by 10x Genomics for their commercial single-cell sequencing solutions, is CellRanger [53]. An alternative that promises reduced runtime, extensions for new technologies and adjustments to read mapping has been built for the RNA-seq aligner STAR [54] while an addition to the pseudoaligner Salmon promises improvements in accuracy of gene abundance quantification combined with higher speed and less memory consumption [55].

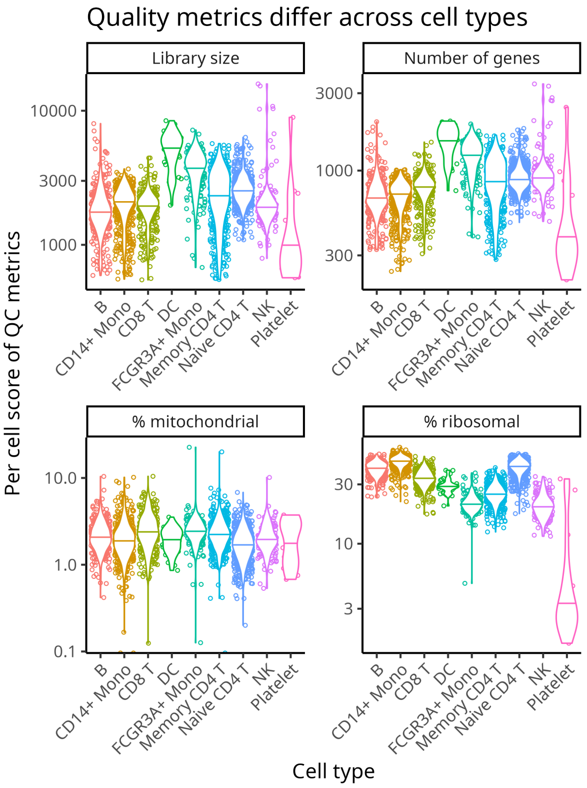


Figure 1.11: Quality metrics vary between cell types. Data retrieved from `https://satijalab.org/seurat/articles/pbmc3k_tutorial.html`.

1.4.2 Quality Control

Assessment of transcriptome quality is important in order to remove samples that would interfere with the analysis. In case of droplet-based single-cell transcriptomics there are usually two stages of quality control. The first stage comprises the removal of empty droplets while the second is more concerned with damaged cells and doublets.

Removal of empty droplets was initially performed using barcode ranks, which capture the total count distribution across droplets. Knee and inflection points on the barcode-rank plot should mark the transition between cell containing and empty droplets. EmptyDrops, a bioconductor R package, offers a more rigorous approach that tests the deviation of a transcriptome from the ambient RNA profile [56]. The cell calling algorithm of the CellRanger pipeline is based on EmptyDrops.

Characterization of damaged cells is primarily done using cell metrics such as the library size, number of expressed genes and percentage of mitochondrial or ribosomal genes. While algorithmic approaches for automated QC exist and cell type specific QC has been proposed (see figure 1.11), manually set thresholds prevail [57, 58].

Identification and removal of heterotypic doublets – formed between different cell types – is possible based purely on expression data using the simulation of doublets and measuring the similarity to those profiles as implemented in DoubletFinder [59].

1.4.3 Normalization

Normalization of count data usually involves two steps. The first adjusts for variable sampling efficiency while the second stabilizes variance across magnitude of expression. Size factors are used to adjust for sampling efficiency and are derived from each samples library size [60]. Variance stabilization using pearson residuals from regularized negative binomial regression aimed to improve downstream analytical tasks [61].

Systematic comparison of transformations for scRNA-seq data has shown that while model residuals, inferred latent expression state and factor analysis have appealing theoretical properties, the logarithm with a pseudocount followed by PCA performs very well [62].

1.4.4 Feature Selection and Summarization

Important parameters that impact the analysis of scRNA-seq data are the number of features that are considered when assessing the similarity of transcriptomes. The first step is the selection of features (i.e. genes) that are variable (highly variable genes, HVG), and thus informative, between samples. While statistical metrics might be used to determine the number of HVGs, a common implementation selects the 2000 features with the highest dispersion [63].

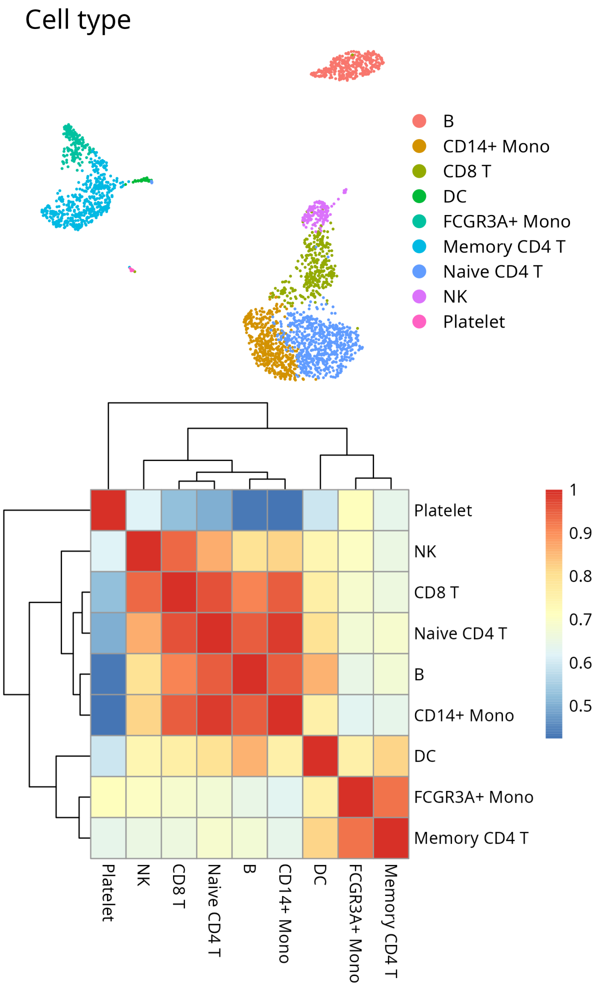


Figure 1.12: Cell embeddings show similarity of cells. Distances can be distorted and should be interpreted with care. Data retrieved from `https://satijalab.org/seurat/articles/pbmc3k_tutorial.html`.

Second, a dimensional reduction technique such as principal component analysis (PCA) is used to summarize the data. Alternatives exist but did not prevail in a recent benchmark [62]. While small data sets are often described well using the first two principal components (PC) alone, choosing the number of PCs is important for downstream analysis.

Embedding transcriptomes in a two-dimensional plane is common in scRNA-seq analysis and typically done by methods using gradient descent to group similar data points such as t-distributed stochastic neighbor embedding (t-SNE, [64]) or uniform manifold approximation and projection (UMAP [65]). While the resulting embeddings facilitate intuitive exploration of the data, distances between clusters should be assessed carefully using complementary methods (see figure 1.12).

Benchmarks that systematically evaluate the impact of feature selection are still lacking but they might have strong impact on data integration and interpretation.

1.4.5 Data Integration

High-throughput sequencing studies can be analyzed as an integrated whole instead of pairwise comparisons between conditions [66]. Large data sets usually contain systematic differences due to technology, time, place, and others that are referred to as batch effects and can be removed using correction algorithms [67]. With larger data sets and atlases the complexity of batch effects increases and analysis requires reliable integration. Important aspects of integration are both the removal of batch effects and the conservation of biological variance. Benchmarking tools across multiple integration tasks has shown that human immune cells are well integrated with fastMNN but scVI can perform better on more complicated tasks, especially in the presence of cell type labels (scANVI) [68].

1.4.6 Cell Type Annotation

Conventional classification of cell types is best exemplified by the surface antigen expression of immune cell subsets such as CD4 for T helpers and CD14 for classical monocytes [23]. For scRNA-seq data annotation is traditionally done manually by exploring marker gene expression across clusters. A hierarchical statistical framework was able to compute cell type probabilities for individual cells based on a marker gene dictionary [69] while another uses a reference of annotated cell types [70]. Guidelines for annotation recommend the use of automatic annotation when possible but emphasize the need of manual refinement, expert annotation and verification [71].

Clustering

Clustering algorithms directly or indirectly control the number of clusters, or resolution. While the choice of resolution can be guided by certain methods [72], significance analysis specifically designed for scRNA-seq data is available [73]. Alternatively, visualization of clustering across resolutions can inform the choice of resolution and guide identification [74]. Ultimately, clustering resolution depends on the type and structure of the data and the question that should be answered. Therefore, investigating differences between clusters is crucial.

1.4.7 Differential Expression

Statisticians have invented tests of significance to distinguish real differences from those occurring only by chance. The default assumption (null hypothesis) being that there is no real difference and the argument by contradiction (low probability, P) showing that this assumption should be rejected based on observed significance levels (see figure 1.13) [75].

In transcriptomics and other high-throughput sequencing (HTS) assays, the fundamental goal is determining the evidence for systematic changes across experimental conditions. For RNA-seq data the challenges come as small replicate numbers, discreteness, large dynamic range and the presence of outliers which has been addressed by DESeq2 through estimation and shrinkage of fold change and dispersion [76]. Meanwhile, McCarthy and Smith argue that statistical significance is not necessarily biologically meaningful and include thresholds for effect size in their method [77].

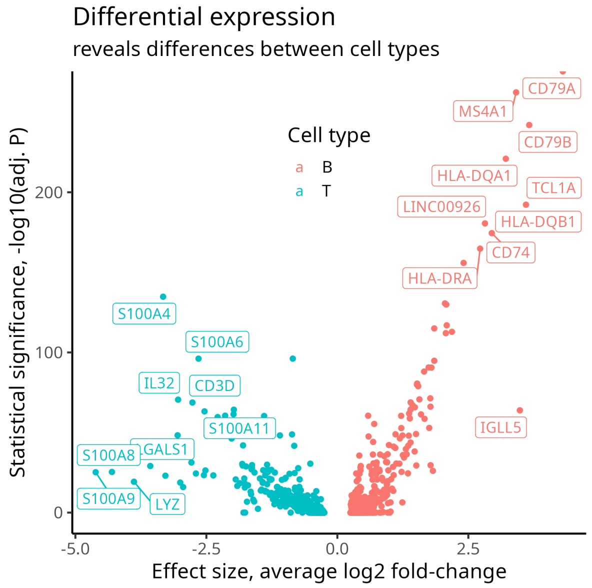


Figure 1.13: Differential expression shows differences between cell types. Data retrieved from https://satijalab.org/seurat/articles/pbmc3k_tutorial.html.

With the rise of single-cell genomics, specific tools for differential expression (DE) analysis of scRNA-seq data, characterized by low library sizes, high noise and dropout events were emerging. Comparison of methods for single-cell DE highlighted the need for prefiltering of genes but showed no benefit of methods specifically designed for scRNA-seq data [78]. Another systematic benchmark highlighted the requirement for replicates and the inevitable proneness to false discoveries if this variation is ignored [79]. Currently, the trend goes towards aggregation of single cell to pseudobulk transcriptomes and the application of RNA-seq DE methods, though some discussion is ongoing [80, 81]. Once a reliable set of differentially expressed genes (DEG) between conditions has been obtained, the next question is how they relate to function.

Gene Set Enrichment

Gene sets that are determined by differential expression analysis describe phenotypes of cells that should be informative but are often hard to understand. The Gene Ontology (GO) consortium strives to create a precisely defined vocabulary that describes the role of genes as either biological process (BP), molecular function (MF) or cellular compartment (CC) [82]. The GO resource now is the most comprehensive source of information about gene function which is consistently maintained and expanded [83]. A common approach to test for enrichment of DE genes in GO categories is over-representation analysis (ORA) based on the fisher test (see table 1.1) [84]. Gene set enrichment analysis (GSEA) is a more quantitative comparison between gene sets that can be applied to gene ontology and other resources [85, 86].

	DEG	No DEG
In GO cat.	7	10
Not in GO cat.	34	1057

Fisher's Exact Test - P: 8.326e-07

Table 1.1: Fisher test reveals significantly enriched gene ontology (GO) categories. GO category B-cell receptor signaling enriched in the differentially expressed genes (DEG) of B cells (see figure 1.13, FC >0 & P $<1 \times 10^{-50}$).

Transcription Factors & Regulons

Another approach to investigate gene function is the inference of regulators that drive expression, such as transcription factors.

Single-cell regulatory network inference and clustering (SCENIC) links gene expression measurements to regulatory sequences using transcription factor (TF) co-expression, motif analysis and activity inference [87].

Another approach relies on enrichment of transcription factors in databases of TF-gene co-expression from RNA-seq, TF-target association in ChIP-seq and other gene lists with TF-gene co-occurrence (ChIP-X Enrichment Analysis, ChEA3) [88].

1.4.8 Differential Abundance

The composition of cell types is different between organs and tissues. Differences in the cellular diversity of the same tissue can be associated with disease conditions. In analogy to differential expression, also differential abundance needs to be assessed with tests of significance. Two major approaches have been developed to address this issue.

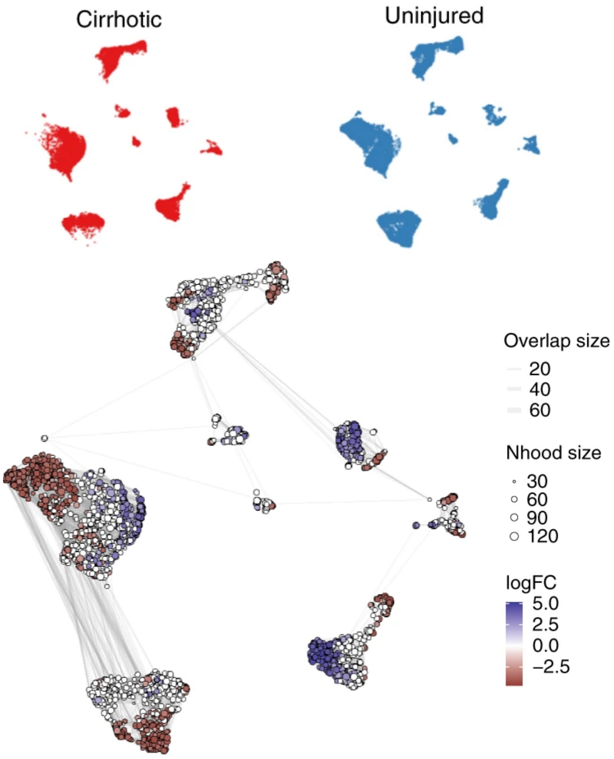


Figure 1.14: Milo reveals differential abundance of cell types. Adopted from [89].

Single-cell compositional data analysis (scCODA) determines the significance of compositional changes for each cluster of cells (e.g. cell type) [90]. In contrast, the statistical framework Milo (see figure 1.14) assigns cells to a large number of partially overlapping neighborhoods on a *k*NN-graph which removes the need for definition of clusters but relies on multiple neighborhoods showing similar compositional shifts [89].

1.5 Aims of the thesis

Single-cell genomics is a young field that has rapidly grown in the last years and revolutionized science by enabling the investigation of cellular heterogeneity. Infectious diseases are particularly promising targets for single-cell studies since differences in disease outcome must be understood in context of cellular interactions between tissues with large diversity. Pathogens infect different cell types, influence the host response, and trigger immune responses that can be deficient, hyperactive or deregulated. However, applications of single-cell technologies require a combination of experimental and computational methods and a well-designed setup in order to provide reliable insights about diseases.

Our primary goal was to introduce single-cell genomics to the clinics. Identification of cellular phenotypes, their interaction in disease and association with different outcomes should be able to offer diagnoses that are more precise. Further exploration and modelling of so-called disease endotypes will hopefully lead to better treatments. This includes identification of mechanisms that lead to refractory disease or relapse.

We were able to investigate relapse of Multiple Myeloma from BCMA-directed chimeric antigen receptor (CAR) T cell therapy [91]. This novel therapeutic, bb2121, was able to induce lasting responses in 85 % of patients that were already relapsed/refractory from heavy pretreatment [92]. I analyzed the transcriptomes of leukocytes from the bone marrow before therapy and after relapse and described both antigen loss and tumor evolution in the Myeloma cells after relapse. Further, we identified genomic deletion as the mechanism of immune escape and heterozygosity of the *TNFRSF17* locus as a potential risk factor for relapse.

With onset of the COVID-19 pandemic we were able to investigate the acute respiratory distress syndrome (ARDS) of patients with severe disease [93]. I was able to study the leukocyte populations in the human lung and compare transcriptional phenotypes across health and disease. With this investigation, we successfully connected the presence of profibrotic macrophages in the lung during severe COVID-19 ARDS to the development of lung fibrosis.

Second, we intended to investigate the impact of metabolism on cellular phenotypes, immune cell function and disease outcome. Insights into the pathobiological mechanism that connects immune cell metabolism, function, and disease outcome has the potential to improve treatment. In our study of chronic *S. aureus* infection in mice [94], I was able to investigate the diversity of leukocyte populations in health and disease. We identified phenotypic markers that distinguish mature from immature granulocytes and revealed the impact of metabolite availability on immune cell differentiation and bacterial load.

For our third goal, we set out to use *in vitro* models of human tissues to study the interactions between epithelia and pathogens in bacterial infections. Organoids and other types of advanced cell culture show larger heterogeneity than traditional cell culture but have not been extensively characterized as infection models. We decided to use single-cell genomics to characterize the models and aimed to investigate differences between highly and lowly infected cells. In our study of *Salmonella typhimurium* infections of an *in vitro* model of the human small intestine [95] we showed transcriptomic and morphological features of host cell differentiation and invasion characteristics of the pathogen.

Overall, the aim was to leverage computational advances in single-cell genomics to characterize cellular phenotypes and connect them to disease phenotypes in order to gain insights into the pathomechanism of infectious diseases for the development of more precise diagnoses and treatments.

2 Results

First author Year [Ref.]	Chumduri 2021 [96]	DaVia 2021 [91]	Wendisch 2021 [93]	Dietrich 2022 [94]	Däullary 2023 [95]
Infection			●	●	●
Cancer	●	●			
Patients		●	● ○		
Donors			○ ●		
Mouse model	●			●	
Organoid					●
Organ	Cervix	Bone Marrow	Lung & Blood	Spleen	Small Intestine
Tissue	Epithelia	Leukocytes	Leukocytes	Leukocytes	Epithelia

Table 2.1: Research topics and cell sources for single-cell sequencing of the major projects.

I analyzed single-cell RNA-seq data using computational methods for the interpretation of cellular phenotypes in biological samples. Beside knowledge in data science & analysis to establish analysis workflows, I needed to use high-performance computing resources and communicate results to biologists and clinicians.

Over the course of my doctoral research, I was able to contribute to multiple projects. In the major projects, I was involved as the leading analyst for single-cell genomics while in minor projects I contributed in terms of teaching and support or data visualization and presentation. Table 2.1 gives an overview of the different topics and cell sources for the major projects. While some of the projects focused on cancer research, the larger emphasis of the thesis was on host-pathogen interactions in infectious diseases. Cells were obtained from multiple sources including patients, donors, mouse models and organoids but leukocytes and epithelia represented the tissues of interest.

Analysis workflows were created individually for each project and tailored to the project needs. Data and programming scripts are published on GitHub (github.com/saliba-lab and github.com/OliverDietrich).

Data visualization and presentation of single-cell genomics is challenging and benefits from interactivity that facilitates exploration, especially of gene expression. Data accessibility is an important need of the research community in order to build a common foundation and facilitate the re-use of data. Interactive visualization and accessibility was facilitated through the development and deployment of web applications on the Infection Atlas (infection-atlas.org).

BRIEF COMMUNICATION
https://doi.org/10.1038/s41591-021-01245-5

nature medicine

Check for updates

Homozygous *BCMA* gene deletion in response to anti-BCMA CAR T cells in a patient with multiple myeloma

Matteo C. Da Vià[1], Oliver Dietrich[2], Marietta Truger[3], Panagiota Arampatzi[4], Johannes Duell[1], Anke Heidemeier[5], Xiang Zhou[1], Sophia Danhof[1], Sabrina Kraus[1], Manik Chatterjee[1], Manja Meggendorfer[3], Sven Twardziok[3], Maria-Elisabeth Goebeler[1], Max S. Topp[1], Michael Hudecek[1], Sabrina Prommersberger[1], Kristen Hege[6], Shari Kaiser[6], Viktoria Fuhr[7], Niels Weinhold[8], Andreas Rosenwald[7], Florian Erhard[9], Claudia Haferlach[3], Hermann Einsele[1], K. Martin Kortüm[1], Antoine-Emmanuel Saliba[2] and Leo Rasche[1,10] ✉

B cell maturation antigen (BCMA) is a target for various immunotherapies and a biomarker for tumor load in multiple myeloma (MM). We report a case of irreversible BCMA loss in a patient with MM who was enrolled in the KarMMa trial (NCT03361748) and progressed after anti-BCMA CAR T cell therapy. We identified selection of a clone with homozygous deletion of *TNFRSF17* (*BCMA*) as the underlying mechanism of immune escape. Furthermore, we found heterozygous *TNFRSF17* loss or monosomy 16 in 37 out of 168 patients with MM, including 28 out of 33 patients with hyperhaploid MM who had not been previously treated with BCMA-targeting therapies, suggesting that heterozygous *TNFRSF17* deletion at baseline could theoretically be a risk factor for BCMA loss after immunotherapy.

The *TNFRSF17* BCMA-encoding gene is located on the short arm of chromosome 16 and is expressed almost exclusively on plasma and B cells. γ-Secretase cleaves membrane-bound BCMA, and thus soluble sBCMA levels are elevated in the serum of patients with MM and correlate with disease activity[1]. BCMA-directed immunotherapies include chimeric antigen receptor (CAR) T cells, T cell engaging bispecific antibodies and antibody drug conjugates, which have shown promising activity in difficult-to-treat, relapsed/refractory MM. Idecabtagene-vicleucel (ide-cel, bb2121) is a BCMA-targeting CAR T cell therapy that showed an overall response rate of 85% in a dose-escalating phase 1 study, with 45% of patients achieving complete response[2]. The enrolled study population was heavily pretreated and enriched in patients suffering from extramedullary disease (EMD) and high-risk cytogenetics. Achieving deep responses by a single infusion is an important improvement in this population with poor-prognosis relapsed/refractory (R/R) MM. However, not all responses were durable, and the median progression-free survival was 11.8 months. Similarly, other anti-BCMA CAR T cell products have shown high response rates but most responders progress[3,4]. While decrease or loss of expression of the BCMA antigen has been described in a few patients[3,4], the tumor-intrinsic mechanism underlying relapse from BCMA-directed CAR T cell therapy has yet to be elucidated.

In this study, we present an in-depth analysis of BCMA loss in a 71-year-old male patient with immunoglobulin G (IgG)-κ MM enrolled in the ongoing KarMMa phase 2 trial (NCT03361748), which evaluates ide-cel in patients with R/R MM. The patient's characteristics are presented in Extended Data Fig. 1. Baseline whole-body diffusion-weighted magnetic resonance imaging (MRI) revealed multiple focal lesions located in the pelvis and spine as well as EMD in the left axillary region (Fig. 1a). After lymphodepletion with fludarabine and cyclophosphamide, the target dose of 450×10^6 CAR$^+$ T cells was infused and the patient experienced grade I° cytokine release syndrome. After 1 month, the M protein declined from $4.2\,\mathrm{g\,dl^{-1}}$ to $0.8\,\mathrm{g\,dl^{-1}}$ and minimal residual disease (MRD) was undetectable by next-generation sequencing at a threshold of 10^{-5} in a bone marrow aspirate. Likewise, diffusion-weighted MRI showed complete resolution of focal lesions and EMD. After 3 months of follow-up, the M protein further decreased to $0.2\,\mathrm{g\,dl^{-1}}$ and MRD remained undetectable, in line with very good partial response according to the current International Myeloma Working Group guidelines. However, MRI picked up a single lymph node, measuring 0.8 cm, located in the left axillary region, suspicious for an early yet localized relapse (Fig. 1a). Five months after ide-cel infusion, the patient presented with overt relapse, M protein levels of $3.8\,\mathrm{g\,dl^{-1}}$, bone marrow plasma cell infiltration (BMPC) of 90% and extensive EMD affecting lymph nodes, liver and spleen. Of note, sBCMA, usually a robust marker of tumor load in MM[1], was below the limit of detection despite the progression. Cellular kinetics analysis showed substantial expansion and persistence of CAR T cells, including at disease progression (Fig. 1b). Bortezomib-based salvage therapy was initiated, but the patient died two weeks later from refractory disease.

To elucidate the mechanism underlying relapse from anti-BCMA CAR T cell therapy, we investigated paired bone marrow samples

---

[1]Department of Internal Medicine 2, University Hospital of Würzburg, Würzburg, Germany. [2]Helmholtz Institute for RNA-based Infection Research, Helmholtz-Center for Infection Research, Würzburg, Germany. [3]Munich Leukemia Laboratory, Munich, Germany. [4]Core Unit Systems Medicine, University of Würzburg, Würzburg, Germany. [5]Department of Radiology, University Hospital of Würzburg, Würzburg, Germany. [6]Bristol Myers Squibb, San Francisco, CA, USA. [7]Institute of Pathology, University of Würzburg, Würzburg, Germany. [8]Medizinische Klinik 5, University Hospital of Heidelberg, Heidelberg, Germany. [9]Institute for Virology and Immunobiology, University of Würzburg, Würzburg, Germany. [10]Mildred Scheel Early Career Center, University Hospital of Würzburg, Würzburg, Germany. ✉e-mail: Rasche_L@ukw.de

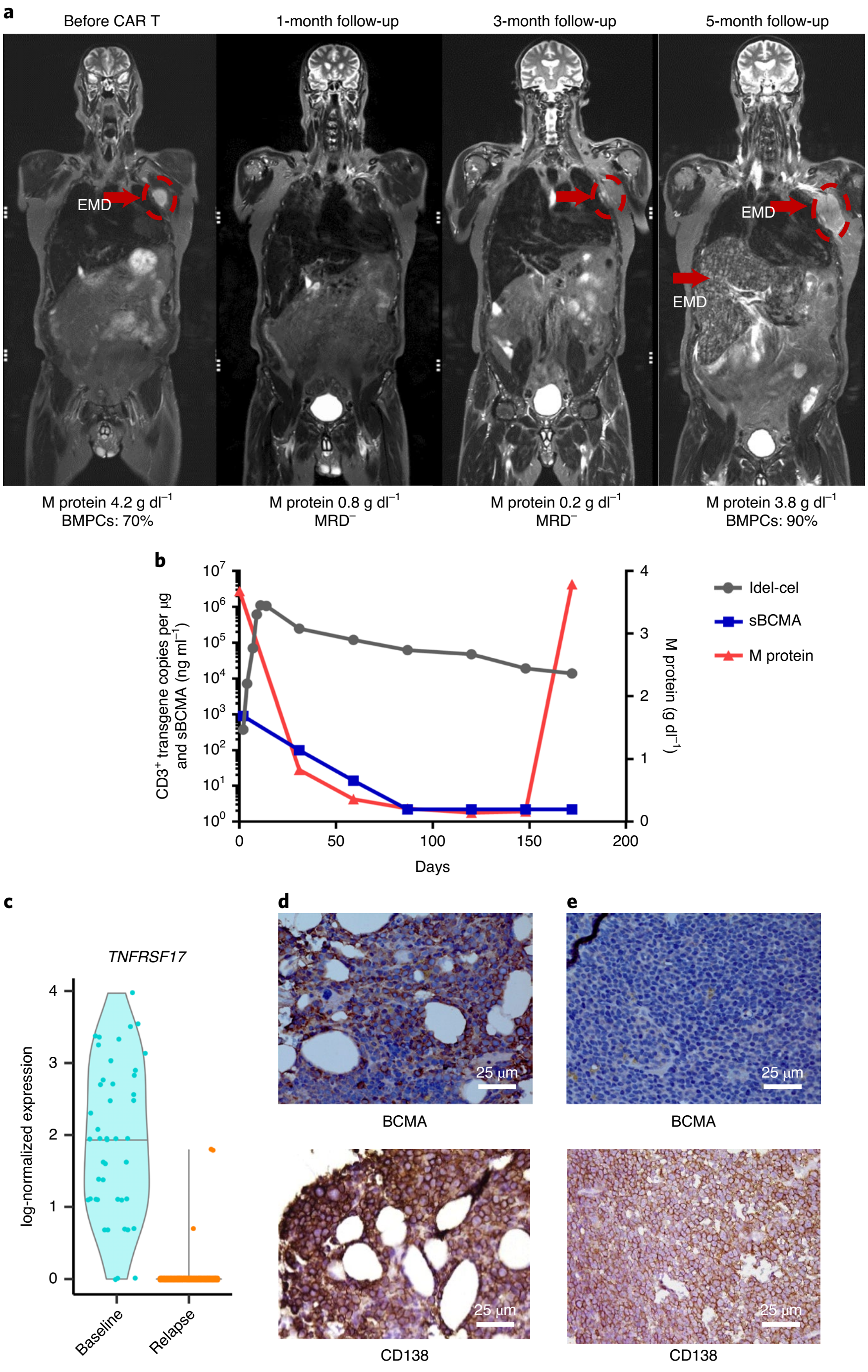

a
Before CAR T
1-month follow-up
3-month follow-up
5-month follow-up
EMD
EMD
EMD
M protein 4.2 g dl^{-1}
BMPCs: 70%
M protein 0.8 g dl^{-1}
MRD^{-}
M protein 0.2 g dl^{-1}
MRD^{-}
M protein 3.8 g dl^{-1}
BMPCs: 90%
b
CD3^{+} transgene copies per µg and sBCMA (ng ml^{-1})
10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7
M protein (g dl^{-1})
0
1
2
3
4
0
50
100
150
200
Days
Idel-cel
sBCMA
M protein
c
TNFRSF17
log-normalized expression
0
1
2
3
4
Baseline
Relapse
d
25 µm
BCMA
25 µm
CD138
e
25 µm
BCMA
25 µm
CD138

Fig. 1 | Clinical characteristics and BCMA expression from baseline to relapse. **a**, Coronal T2 MRI images showing resolution and reoccurrence of EMD (red arrow) at baseline and during follow-up. **b**, Ide-cel expansion, sBCMA levels and M protein over time. Pharmacokinetics analysis showed good persistence of CAR T cells in this patient, even at progression (gray curve). On the other hand, the sBCMA level decreased after CAR T cell treatment and was not measurable at relapse (blue curve). **c**, Violin plot showing the log-normalized expression of *TNFRSF17* in 45 MM cells detected at baseline and 2,426 MM cells detected at relapse. The horizontal line shows the median of the expression value. **d**,**e**, BCMA protein expression determined by immunohistochemistry on formalin-fixed paraffin-embedded bone marrow sections obtained at baseline (**d**) and relapse (**e**). The figures show representative images from three independent experiments.

collected at baseline and progression using single-cell RNA sequencing (scRNA-seq; Methods and Extended Data Fig. 2a–e). *TNFRSF17* was strongly expressed in myeloma cells at baseline but expression was virtually undetectable at progression ($P < 6.2 \times 10^{-94}$, two-sided Wilcoxon rank-sum test; Fig. 1c and Extended Data Fig. 3a–d). Gene expression results were confirmed at the protein level using immunohistochemistry (Fig. 1d,e). Other plasma cell markers and potential immunotherapy targets, such as *CD38* and *SLAMF7*, remained stably expressed at progression (Extended Data Fig. 3e,f), suggesting the selection of a BCMA^{-} MM clone by CAR T cell therapy. While all cells at progression were *TNFRSF17*$^{-}$, we still noted heterogeneity at the transcriptomic level with MM cells falling into

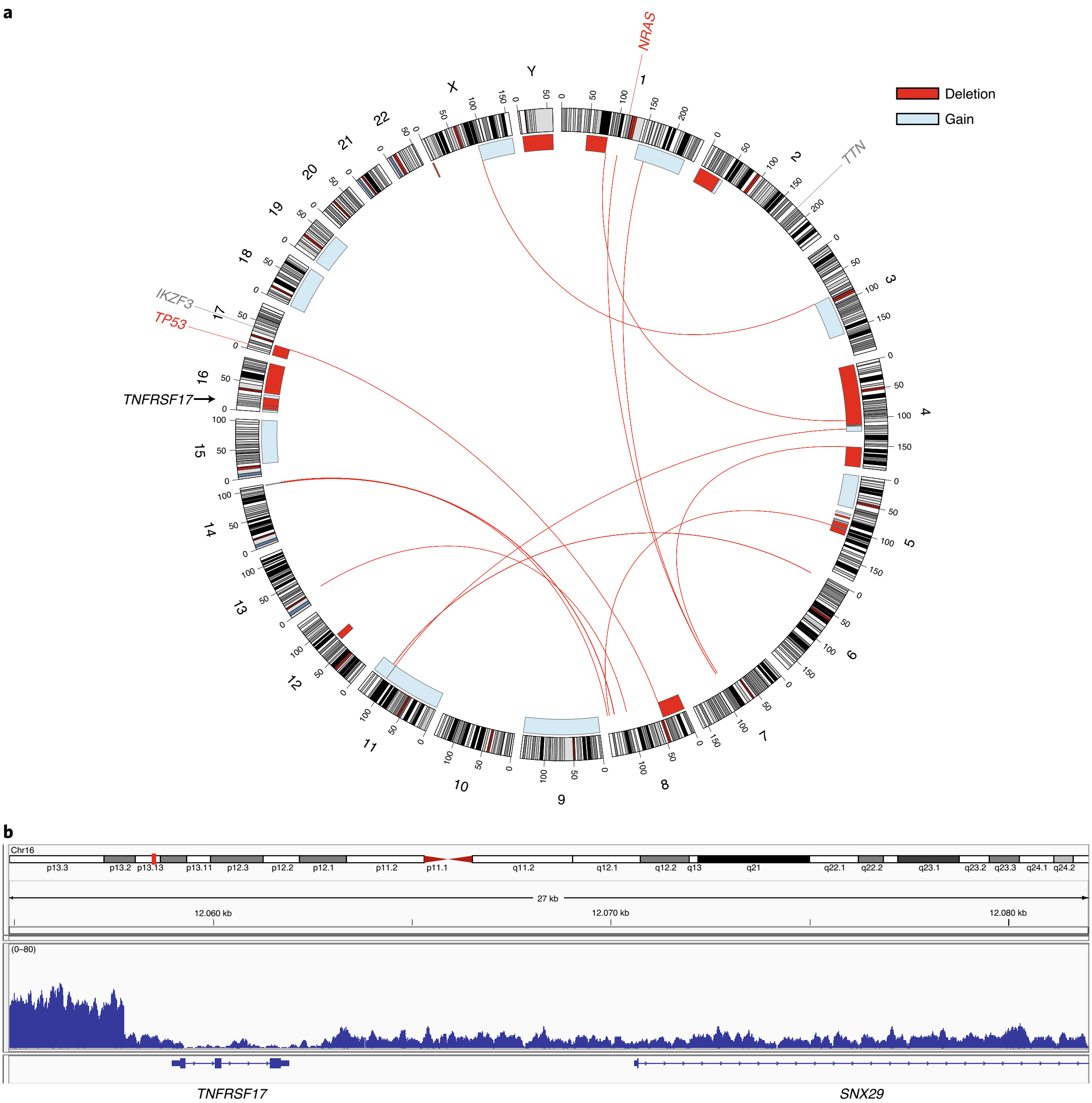


Fig. 2 | WGS and read coverage of CD138$^+$ purified plasma cells at progression from CAR T cell therapy. a, Circos plot with copy number variations, structural variations and single-nucleotide variations based on WGS data at progression after anti-BCMA CAR T cell therapy. Outer track runs clockwise from chromosome 1 to Y. The inner track shows gains >1 Mb in blue and losses >1 Mb in red. The red lines inside the circle represent interchromosomal reciprocal translocations with a VAF > 0.1. Genes with mutations (*NRAS* and *TP53*) are depicted in red and genes with detected variants of unknown significance (*IKZF3* and *TTN*) are shown in gray. **b**, Read coverage visualized with the Integrative Genomics Viewer illustrates copy number variations in the short arm of chromosome 16 consisting of heterozygous deletion of 21.3 Mb from 5,308,001 to 26,651,000 bp and homozygous deletion of 91 kb from 12,58,001 to 12,149,000 (hg19 reference genome). Read coverage is lowest across the homozygous deleted genomic region including *TNFRSF17* (*BCMA*).

eight clusters, such as highly proliferative cells (*MKI67*) or *CD27*$^+$ cells (Extended Data Fig. 3c,d). Furthermore, we noted increased expression of exhaustion markers in bone marrow-infiltrating T cells at relapse compared to baseline (Extended Data Fig. 4a–d).

To further investigate the molecular mechanism leading to total loss of *TNFRSF17* gene expression, we performed whole-genome sequencing (WGS) of MM cells at progression, which revealed numerous genomic aberrations including amplification of 1q, deletion of 1p, biallelic inactivation of *TP53* and a complex IGH-MYC rearrangement (Fig. 2a). We observed a large heterozygous deletion of 21.3 megabases (Mb) affecting chromosome 16 and a homozygous deletion of 91 kilobases (kb) (from 12,058,001 to 12,149,000), resulting in biallelic loss of the *TNFRSF17* gene at 16p13.13 (Fig. 2b). These data indicate a genomic mechanism for the observed BCMA$^-$ progression after CAR T cell therapy.

Homozygous deletions usually require two independent hits[5]. Therefore, we investigated whether this patient already had a heterozygous deletion of *TNFRSF17* before CAR T cell treatment. Using WGS and scRNA-seq, we detected no *TNFRSF17* gene deletion before treatment (Extended Data Fig. 5 and 6). However, we found a focal gain of 16p13.2–p13.13 including the *TNFRSF17* locus, next to a deletion of 16p13.3–p13.2, suggesting preexisting genomic instability at that region. Differences between the genomic profiles at baseline and progression, such as the 20p12.2–q11.21 deletion, which is only present at baseline, argue for branching evolution during treatment. The branch emerging at progression was not detectable in the baseline sample. Thus, we cannot exclude the possibility that a heterozygous 16p deletion was already present in a subclone at baseline in this patient.

To further investigate if heterozygous deletions on chromosome 16p are present at baseline in anti-BCMA immunotherapy-naïve patients, we performed WGS in a group of 50 patients with R/R MM from our institution (patient characteristics are presented in Supplementary Table 1). We identified three patients with heterozygous deletions encompassing the *TNFRSF17* locus, who could potentially be at risk for developing homozygous *TNFRSF17* loss during BCMA-targeted therapy (Supplementary Table 2 and Extended Data Fig. 7). We examined whether heterozygous deletions could impact *TNFRSF17* gene expression and analyzed a subgroup of 28 patients with R/R MM, including two patients with heterozygous 16p deletions, for whom bulk RNA-seq data were available. *TNFRSF17* gene expression levels were similar across these 28 patients with R/R MM, suggesting no impact of heterozygous deletions on *TNFRSF17* gene expression (Extended Data Fig. 8). However, the biological significance of these deletions has yet to be determined since we did not investigate protein expression levels, downstream signaling and potential cell survival advantages. In this context, *TNFRSF17* downregulation decreased cell proliferation rate and altered signaling in the MM cell lines[6]. Further analysis of three publicly available datasets[7–9] confirmed the presence of heterozygous *TNFRSF17* gene deletions or monosomy 16 in both newly diagnosed and relapsed patients at a low frequency (4 out of 41 newly diagnosed, 2 out of 44 relapsed, 28 out of 33 with hyperhaploid MM) (Supplementary Table 3). Hyperhaploid MM is a high-risk entity characterized by several chromosomal monosomies including monosomy 16; however, overall this subgroup is rare, constituting no more than 5% of all cases with MM[8]. When excluding patients with hyperhaploid MM, the prevalence of heterozygous *TNFRSF17* loss or monosomy 16 in newly diagnosed and relapsed patients in our institution dataset and the three publicly available datasets was 6.7% (9 out of 135 patients).

Antigen loss relapses pose a challenge to targeted immunotherapies since they represent the ultimate adaptation of a cancer cell to the selective pressure of CAR T cell therapies. CD19 loss has been observed in up to 40% of patients with B cell acute lymphoblastic leukemia treated with different CAR19 products; point mutations affecting *CD19* impacting protein anchoring to the cell membrane have been described[10]. In diffuse large B cell lymphoma, the frequency of CD19 loss was up to 30%[11]. In addition, CD20 or CD22 losses have occurred after respective targeted immunotherapies[12,13]. In MM, BCMA loss was observed in two early CAR T cell studies with frequencies around 4–9% but the underlying mechanism was not elucidated[4,14]. As a disease associated with a high frequency of copy number variations, (sub)clonal chromosomal aberrations are common in MM and may affect genes encoding the targets of immunotherapies[15]. In this context, we identified heterozygous deletions of chromosome 16p encompassing the *TNFRSF17* gene locus, which in theory could represent a first hit toward acquisition of resistance. However, this form of immune escape is not common and the clinical implications of 16p heterozygous loss before anti-BCMA therapies are unknown.

In summary, our study reveals the rare existence of chromosome 16 aberrations before and after CAR T cell therapy, linking genomic instability impacting the *TNFRSF17* locus to escape from BCMA-targeting immunotherapies. Therapeutically targeting a combination of different MM targets may overcome the substantial genomic heterogeneity in MM, avoiding clonal selection based on the loss of one particular antigen. There is a rich pipeline of multiantigen-targeting approaches including bispecific CAR T products[16], trispecific antibodies[17] and combination therapies with monotargeted immunotherapies. Beyond BCMA, G protein-coupled receptor class C group 5 member D[18], CD319 (ref. [19]) and CD44v6 (ref. [20]) are some plasma cell targets under clinical investigation and may be used in combination therapies to prevent BCMA escape in future studies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-021-01245-5.

Received: 12 August 2020; Accepted: 12 January 2021;
Published online: 22 February 2021

References

1. Shah, N., Chari, A., Scott, E., Mezzi, K. & Usmani, S. Z. B-cell maturation antigen (BCMA) in multiple myeloma: rationale for targeting and current therapeutic approaches. *Leukemia* **34**, 985–1005 (2020).
2. Raje, N. et al. Anti-BCMA CAR T-Cell therapy bb2121 in relapsed or refractory multiple myeloma. *N. Engl. J. Med.* **380**, 1726–1737 (2019).
3. Cohen, A. D. et al. B cell maturation antigen-specific CAR T cells are clinically active in multiple myeloma. *J. Clin. Invest.* **129**, 2210–2221 (2019).
4. Brudno, J. N. et al. T cells genetically modified to express an anti-B-cell maturation antigen chimeric antigen receptor cause remissions of poor-prognosis relapsed multiple myeloma. *J. Clin. Oncol.* **36**, 2267–2280 (2018).
5. Cheng, J. et al. Pan-cancer analysis of homozygous deletions in primary tumours uncovers rare tumour suppressors. *Nat. Commun.* **8**, 1221 (2017).
6. Tai, Y.-T. et al. APRIL and BCMA promote human multiple myeloma growth and immunosuppression in the bone marrow microenvironment. *Blood* **127**, 3225–3236 (2016).
7. Rasche, L. et al. Spatial genomic heterogeneity in multiple myeloma revealed by multi-region sequencing. *Nat. Commun.* **8**, 268 (2017).
8. Sawyer, J. R. et al. Hyperhaploidy is a novel high-risk cytogenetic subgroup in multiple myeloma. *Leukemia* **31**, 637–644 (2017).
9. Weinhold, N. et al. Clonal selection and double-hit events involving tumor suppressor genes underlie relapse in myeloma. *Blood* **128**, 1735–1744 (2016).
10. Orlando, E. J. et al. Genetic mechanisms of target antigen loss in CAR19 therapy of acute lymphoblastic leukemia. *Nat. Med.* **24**, 1504–1506 (2018).
11. Neelapu, S. S. et al. Axicabtagene ciloleucel CAR T-Cell therapy in refractory large B-cell lymphoma. *N. Engl. J. Med.* **377**, 2531–2544 (2017).
12. Bannerji, R. et al. Clinical activity of REGN1979, a bispecific human, anti-CD20 × anti-CD3 antibody, in patients with relapsed/refractory (R/R) B-cell non-Hodgkin lymphoma (B-NHL). *Blood* **134**, 762 (2019).
13. Paul, M. R., Wong, V., Aristizabal, P. & Kuo, D. J. Treatment of recurrent refractory pediatric pre-B acute lymphoblastic leukemia using inotuzumab ozogamicin monotherapy resulting in CD22 antigen expression loss as a mechanism of therapy resistance. *J. Pediatr. Hematol. Oncol.* **41**, e546–e549 (2019).
14. Martin, N. et al. Correlation of tumor BCMA expression with response and acquired resistance to idecabtagene vicleucel in the KarMMa study in relapsed and refractory multiple myeloma. *HemaSphere* https://doi.org/10.1097/HS9.0000000000000404 (2020).
15. Rustad, E. H. et al. Revealing the impact of structural variants in multiple myeloma. *Blood Cancer Discov.* **1**, 258–273 (2020).
16. Fernández de Larrea, C. et al. Defining an optimal dual-targeted CAR T-cell therapy approach simultaneously targeting BCMA and GPRC5D to prevent BCMA escape-driven relapse in multiple myeloma. *Blood Cancer Discov.* **1**, 146–154 (2020).
17. Garfall, A. L. & June, C. H. Trispecific antibodies offer a third way forward for anticancer immunotherapy. *Nature* **575**, 450–451 (2019).
18. Pillarisetti, K. et al. A T-cell-redirecting bispecific G-protein-coupled receptor class 5 member D × CD3 antibody to treat multiple myeloma. *Blood* **135**, 1232–1243 (2020).
19. Gogishvili, T. et al. SLAMF7-CAR T cells eliminate myeloma and confer selective fratricide of SLAMF7^{+} normal lymphocytes. *Blood* **130**, 2838–2847 (2017).
20. Casucci, M. et al. CD44v6-targeted T cells mediate potent antitumor effects against acute myeloid leukemia and multiple myeloma. *Blood* **122**, 3461–3472 (2013).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Trial design. The KarMMa trial (NCT03361748) is an open label, single-arm, phase 2 trial. It evaluates the security and effectiveness of anti-BCMA-directed CAR T cells in patients with R/R MM. In this study, we report on a patient enrolled in the KarMMa trial. Additional bone marrow samples were obtained before lymphodepletion therapy, infusion of CAR T cells and at the time of progression. Analysis included WGS, scRNA-seq, pharmacokinetics data, medical imaging and MRD results, and were reported in compliance with the CARE guidelines. A group of 50 patients with R/R MM underwent WGS of CD138 purified plasma cells to further investigate the *TNFRSF17* locus. All patients gave their written informed consent for scientific evaluations. The study was approved by the internal review board of the University of Würzburg (reference 309/17 for the KarMMa trial and KFO216 for the molecular analysis) and adhered to the tenets of the Declaration of Helsinki of 2008.

Sample preparation and scRNA-seq. Unselected bone marrow cells were collected from bone marrow aspirates at baseline (before CAR T cell infusion) and relapse by bone marrow biopsy. Mononucleated cells were purified and separated with Ficoll density gradient separation solution (Sigma-Aldrich) and washed twice with 0.04% PBS-BSA. Then, the percentage of $CD138^+$ (CD138 antibody provided by BioLegend; fluorochrome: PerCP-Cyanine 5.5; clone MI15; 1:50 dilution) cells was detected by fluorescence-activated cell sorting performed by Accuri C6 (BD Biosciences). Due to technical reason, the relapse sample was reloaded after thawing a frozen aliquot. The concentration of the single-cell suspension was adjusted to 700 cells μl^{-1}. Chromium Controller was used to partition single cells into nanoliter-scale Gel Bead-In-EMulsions and Single Cell 3′ reagent kit v2 or v3 for reverse transcription, complementary DNA amplification and library construction (10x Genomics). The detailed protocol was provided by 10x Genomics. The SimpliAmp Thermal Cycler was used for the amplification and incubation steps (Applied Biosystems). Libraries were quantified by Qubit 3.0 Fluorometer (Thermo Fisher Scientific) and quality was checked using the 2100 Bioanalyzer with High Sensitivity DNA kit (Agilent Technologies). Baseline and relapse sample libraries were sequenced with an S1 or S2 100 base pair (bp) flow cell using the NovaSeq 6000 platform (Illumina), leading to an average of 235,572 and 99,693 reads per cell, respectively.

Bioinformatics analysis of scRNA-seq data. The Cell Ranger v.3.0.1 software suite was obtained from 10x Genomics (https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest). The raw sequencing data was first demultiplexed and quality-checked using the Cell Ranger mkfastq script. For all sample libraries, alignment and transcript quantification were performed with the standard Cell Ranger count script against the GRCh38 human genome assembly. All samples were aggregated using the Cell Ranger aggr script with the default normalization step (by downsampling) disabled.

The expression matrices were filtered based on individual thresholds. The baseline sample was filtered to remove barcodes with more than 3,000 genes, 20,000 unique molecular identifiers and 12.5% mitochondrial genes detected. The relapse sample was filtered to remove barcodes with more than 7,500 genes, 65,000 unique molecular identifiers and 20% mitochondrial genes detected. Genes that were not detected for any barcode were also removed.

Downstream analysis was performed using the R package Seurat v.3.0.2 (ref. [21]). Count data was log-normalized and 2,000 variable features were selected using the Seurat functions NormalizeData and FindVariableFeatures using default settings. Principal component analysis was performed on the scaled, log-normalized data limited to the 2,000 most variable features. The first 10 principal components were selected for dimensional reduction and clustering of the regular datasets and 15 components were used for the subset of MM cells. A two-dimensional representation was computed by the uniform manifold approximation and projection algorithm[22]. For clustering, a shared nearest neighbor graph was constructed using the Seurat function FindNeighbors. Clusters were identified for the resolutions 0.1, 0.5 and 5 using the FindClusters function with the smart local moving algorithm[23].

Cell cycle stage annotation was performed based on a list of cell cycle markers from Kowalczyk et al.[24]. The Seurat function CellCycleScoring was used to compute the quantitative scores for G_2M and S phases and assign qualitative labels to each barcode.

Differential gene expression analysis was performed using the Wilcoxon rank-sum test through the Seurat function FindAllMarkers with default values.

No correction for batch effects was performed since we did not infer differences between the groups. Comparisons between the conditions (baseline and relapse) were limited to the MM subset using population summaries instead of distance. Interpretations based on distance were limited to the relapse dataset.

Copy number variant analysis using scRNA-seq. To directly compare the scRNA-seq data to the WGS data, we rerun Cell Ranger versus Ensembl v.75/GRCh37. We extracted all genes located within the heterozygous deletion region from the GTF file downloaded from Ensembl. The gene expression quantifications filtered by Cell Ranger were used for the baseline/relapse scatter plot for myeloma cells and monocytes (cell subsets as defined above; Extended Data Fig. 2c). For the location specific analyses, we clustered overlapping reads along chromosome 16 extracted from the BAM file generated by Cell Ranger and estimated the $\log_2$(fold change) with credible intervals between baseline and relapse using PsiLFC (v0.2.1)[25].

WGS. WGS was performed on CD138 purified cells derived from the baseline and relapse samples. Library preparation was performed using the TruSeq DNA PCR-Free HT Sample Preparation Kit (Illumina) according to manufacturer's protocol and 151 bp paired-end sequences were generated on NovaSeq 6000 sequencing instruments (Illumina) with 100× coverage. Since matched-normal samples were not available, a mixture of genomic DNA from multiple anonymous donors was used as normal controls. Read mapping and tumor/normal variant calling were performed using Illumina's WGS app v.5.0 and the tumor normal app v.3.0 for analysis of genomic data. WGS reads were mapped to the human reference genome (Ensembl GRCh37) using Ilumina's Isaac Aligner (iSAAC-03.16.02.19) (ref. [26]). Structural variations were called using Manta v.0.28.0 (ref. [27]); only passed variants with >2 paired tumor reads were considered for further analysis. Copy number variation (CNV) calling was performed using the Genome Analysis Toolkit following Broad's best practices recommendation. To illustrate structural variation and CNV via circos plot, only interchromosomal fusions with a variant allele frequency (VAF) > 0.1 and CNV > 1 Mb were included. Strelka2 variant caller (v2.4.7)[28] was used for single-nucleotide variants. To remove potential germline variants, each variant was queried against the Genome Aggregation Database (v2.1.1); variants with global population frequencies >1% were excluded.

Bulk RNA-seq. For bulk RNA-seq, 250 ng of total RNA per sample were used to produce stranded RNA libraries (TruSeq Total Stranded RNA; Illumina); 2 × 100 bp paired-end reads were sequenced on the NovaSeq 6000 system with a median of 64 million reads per sample. Gene counts were normalized by applying the trimmed mean of M-values normalization method; the resulting $\log_2$ counts per million were used as a proxy for gene expression. Gene expression differences were assessed using the edgeR package (v3.28.1) with false discovery rate correction for multiple testing. Genes with a false discovery rate <0.05 and an absolute log(fold change) >1.5 were considered differentially expressed.

Statistics. The nonparametric Wilcoxon rank-sum test was used to compare gene expression values between baseline and progression (*TNFRSF17* and T cell activation and exhaustion markers). $P < 0.05$ was considered statistically significant.

MRD test. MRD negativity was evaluated sequentially in bone marrow aspirates by next-generation sequencing (clonoSEQ; Adaptive Biotechnologies) with a minimum cutoff of 10^{-5} nucleated cells.

Ide-cel pharmacokinetics analysis. Expansion and persistence of CAR+ T cells were analyzed to determine the cellular pharmacokinetic profile of ide-cel. $CD3^+$ cells were purified from whole blood and their DNA was purified as described previously.[3] The time course of vector transgene copies per μg of genomic DNA, as measured by quantitative PCR, was assessed. Using the pharmacokinetic data, noncompartmental analysis was performed to calculate parameters such as time of maximum observed transgene level (T_{max}), maximum transgene level occurring at T_{max} (C_{max}), time of last measurable transgene level (T_{last}) and area under the curve using the software program Phoenix WinNonlin v.8.1.

Immunostaining on bone marrow paraffin sections. BCMA protein expression was determined by immunohistochemistry using a polyclonal goat anti-BCMA antibody (1:10 dilution, target retrieval pH 6.1; catalog no. AF193; R&D Systems) on paraffin-embedded bone marrow sections according to standard procedures.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All requests for raw and analyzed data and materials will be promptly reviewed by University Hospital Würzburg and Münchner Leukämielabor to verify if the request is subject to any confidentiality and data protection obligations. Any data and materials that can be shared will be released via a material transfer agreement. All raw and analyzed scRNA-seq data can be found in the National Center for Biotechnology Information Gene Expression Omnibus (accession no. GSE143317). Adherence to the Datenschutz-Grundverordnung (https://dsgvo-gesetz.de/ and https://data.consilium.europa.eu/doc/document/ST-5419-2016-REV-1/en/pdf) is mandatory for sharing WGS data. Thus, the complete whole-genome raw dataset of a single patient cannot be shared according to European law. Processed data from which the identification of a patient is not possible can be made available. WGS data derived from public datasets are deposited in the European Genome-phenome Archive with the references EGAS00001002111 (ref. [7]) and EGAS00001001810 (ref. [9]). Source data are provided with this paper.

References

21. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
22. McInnes, L. et al. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).

23. Waltman, L. & van Eck, N. J. A smart local moving algorithm for large-scale modularity-based community detection. *Eur. Phys. J. B* **86**, 471 (2013).
24. Kowalczyk, M. S. et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* **25**, 1860–1872 (2015).
25. Erhard, F. Estimating pseudocounts and fold changes for digital expression measurements. *Bioinformatics* **34**, 4054–4063 (2018).
26. Raczy, C. et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* **29**, 2041–2043 (2013).
27. Chen, X. et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
28. Kim, S. et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594 (2018).

Acknowledgements

We thank S. Roth for her excellent technical support. We acknowledge the work and contribution of T. Mack from Bristol Myers Squibb. L.R. was supported by the German Cancer Aid via the MSNZ program and IZKF Würzburg. H.E and M.H. were supported by the DFG via SFB/TR 221 project A3. A.-E.S. and O.D. were supported by the German Research Society graduate college 2157. Bristol Myers Squibb provided clinical trial data related to the presented case but no financial support for the study.

Author contributions

L.R., M.C.D.V., K.H., A.-E.S. and H.E. contributed to study design and manuscript preparation. K.M.K., M.G., M.C., J.D., S. Kraus, A.H., X.Z., S.D., M.S.T., M.H. and S.P. contributed to the collection of clinical specimens. A.R. and V.F. performed the immunohistochemistry analysis. M.C.D.V., O.D., P.A., V.F. and A.-E.S. performed the scRNA-seq. O.D. performed the scRNA-seq data analysis. M.T., S.M., S.T., M.M. and C.H. performed whole-genome and bulk RNA sequencing. F.E. contributed to the bioinformatics analysis. N.W. analyzed the additional patent cohorts. S. Kaiser and K.H. provided clinical trial data. All authors approved the manuscript.

Competing interests

K.H. and S.K. are employees of Bristol Myers Squibb. M.H. is listed as an inventor on patent applications and granted patents related to CAR T cell therapy and CAR T cell technologies; he has received speaker honoraria from Celgene/Bristol Myers Squibb, Janssen and Kite/Gilead. H.E. received research funding from Janssen, Bristol Myers Squibb/Celgene, Amgen, Novartis, Sanofi, GSK and Janssen and has received speaker honoraria from Bristol Myers Squibb/Celgene, Amgen, Novartis, Takeda, Sanofi and GSK. L.R. has received speaker honoraria from Bristol Myers Squibb/Celgene, Sanofi, GSK, Oncopeptides and Janssen. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-021-01245-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-021-01245-5.

Correspondence and requests for materials should be addressed to L.R.

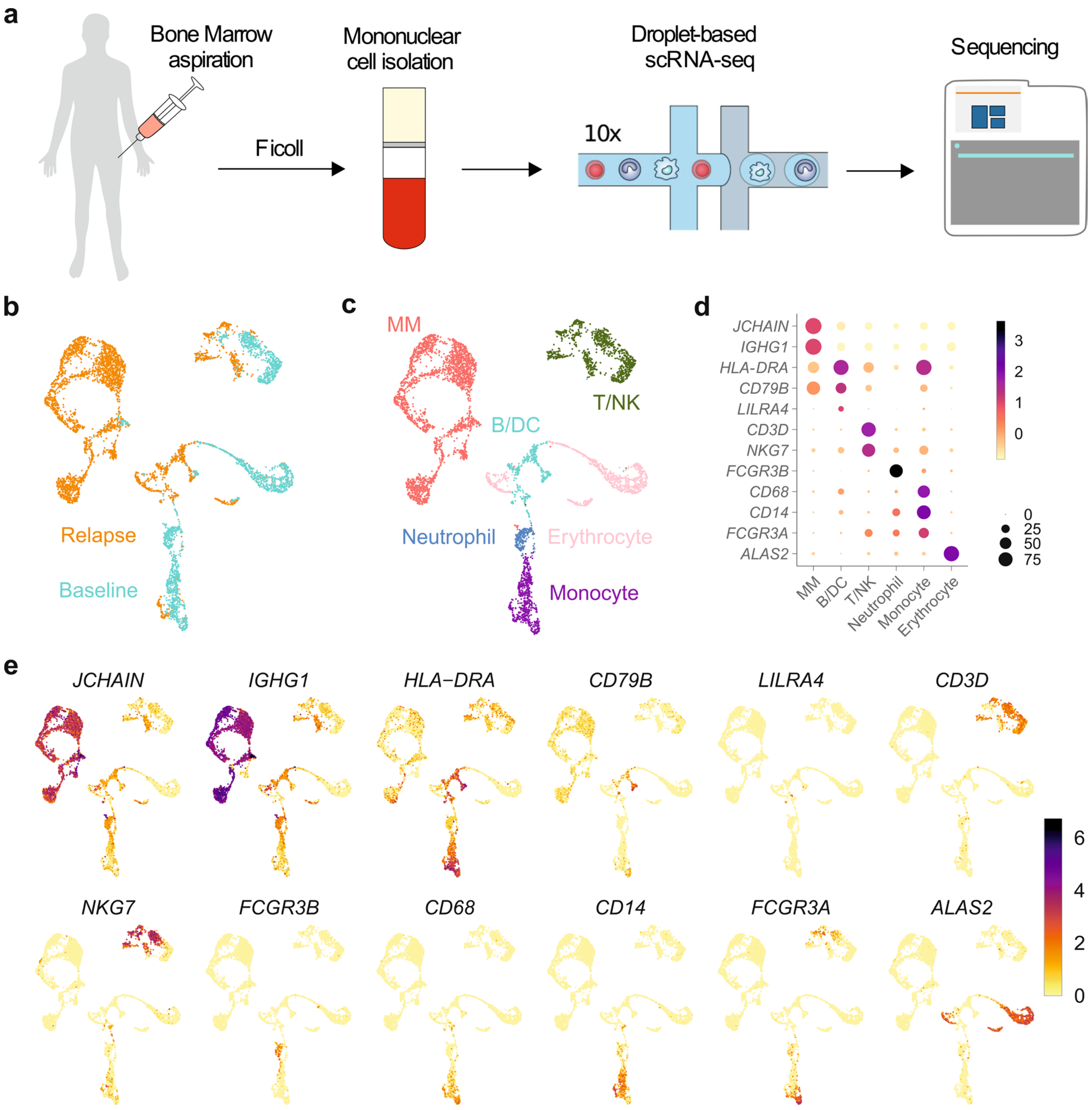
Peer review information *Nature Medicine* thanks Eric Smith, Premal Lulla and Alfred Garfall for their contribution to the peer review of this work. Saheli Sadanand was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

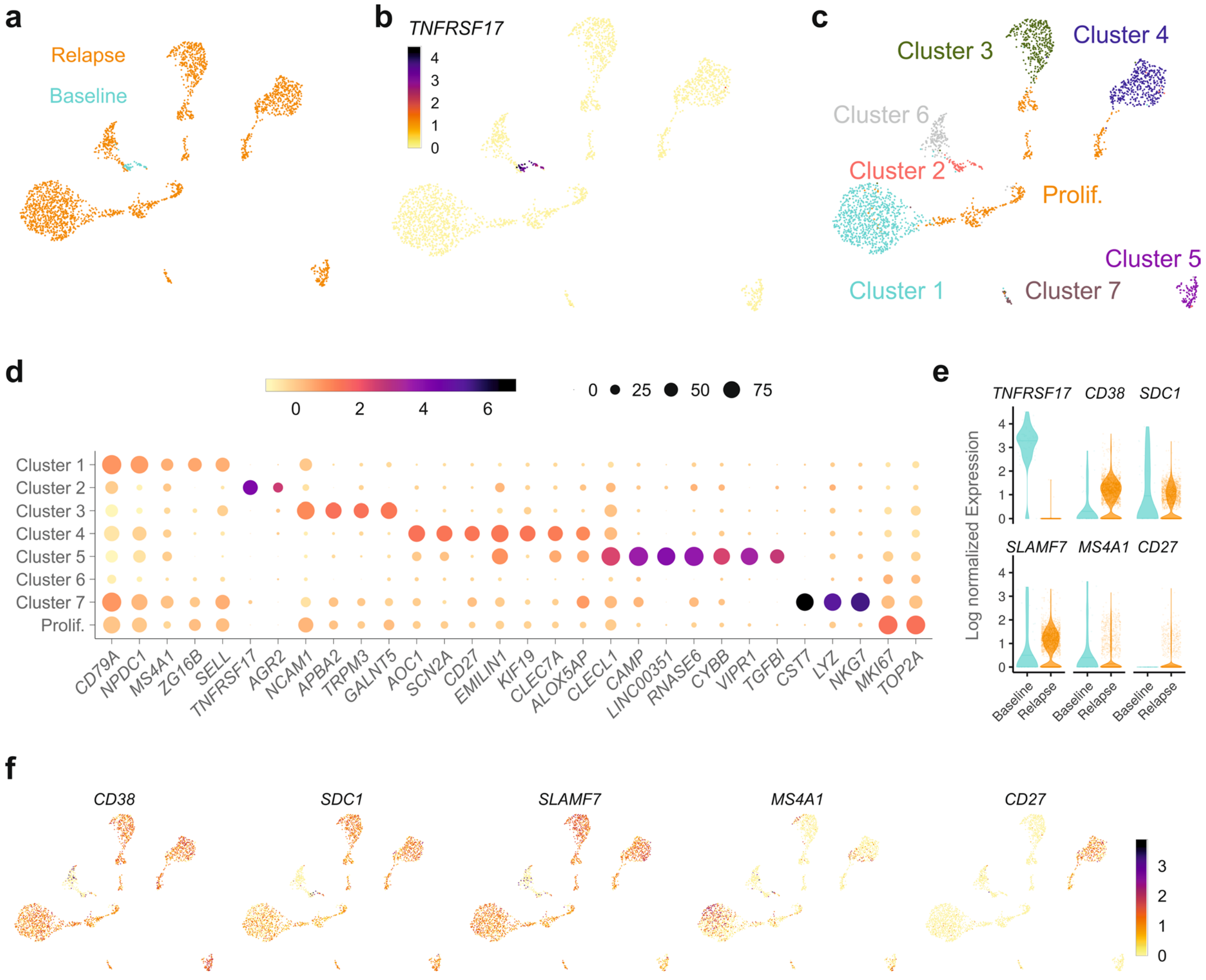
Baseline characteristics	**Diagnosis: 05/2016**	**ISS stage: II**	**Cytogenetics: t(8:14); del(17p), del(1p32), gain(1q21)**
Previous therapies			
1. Line	Induction	4 x PAD-Rev-Dara	Bortezomib 1,3mg/m2
			Doxorubicin 4mg/ m2
			Dexamethason 20mg
			Revlimid 15mg
	SC mob	CE-elderly	Cyclophosphamide 1250mg/m2 iv
			Etoposid 100mg/m2 iv
	1. auto SCT	Melphalan	200mg/m2
	2. auto SCT	Melphalan	200mg/m2
	Maintenance	Ixazomib, Revlimid und Dexamethason	
2 Line	Salvage	6 x PomPAD-Dara	Bortezomib 1,3mg/m²
			Doxorubicin 3mg/ m²
			Dexamethason 20mg
			Pomalidomid 2mg
			Daratumumab 16mg/kg
3. Line	Salvage	2 x Carf-Cyclo-Dexa+Daratumumab	Carfizomib 27mg/m²
			Cyclophosphamid 300mg/m²
			Dexamethason 40mg
			Daratumumab 16mg/kg
4th Line	3. auto SCT	Melphalan	70mg/m2
5th Line	Salvage	PomPAD-Dara	Bortezomib 1,3mg/m²
			Doxorubicin 3mg/ m²
			Dexamethason 20mg
			Pomalidomid 2mg
			Daratumumab 16mg/kg
6th Line	Bridging to CAR T	Carf-Thal-PAD/Daratumumab	

Extended Data Fig. 1 | See next page for caption.

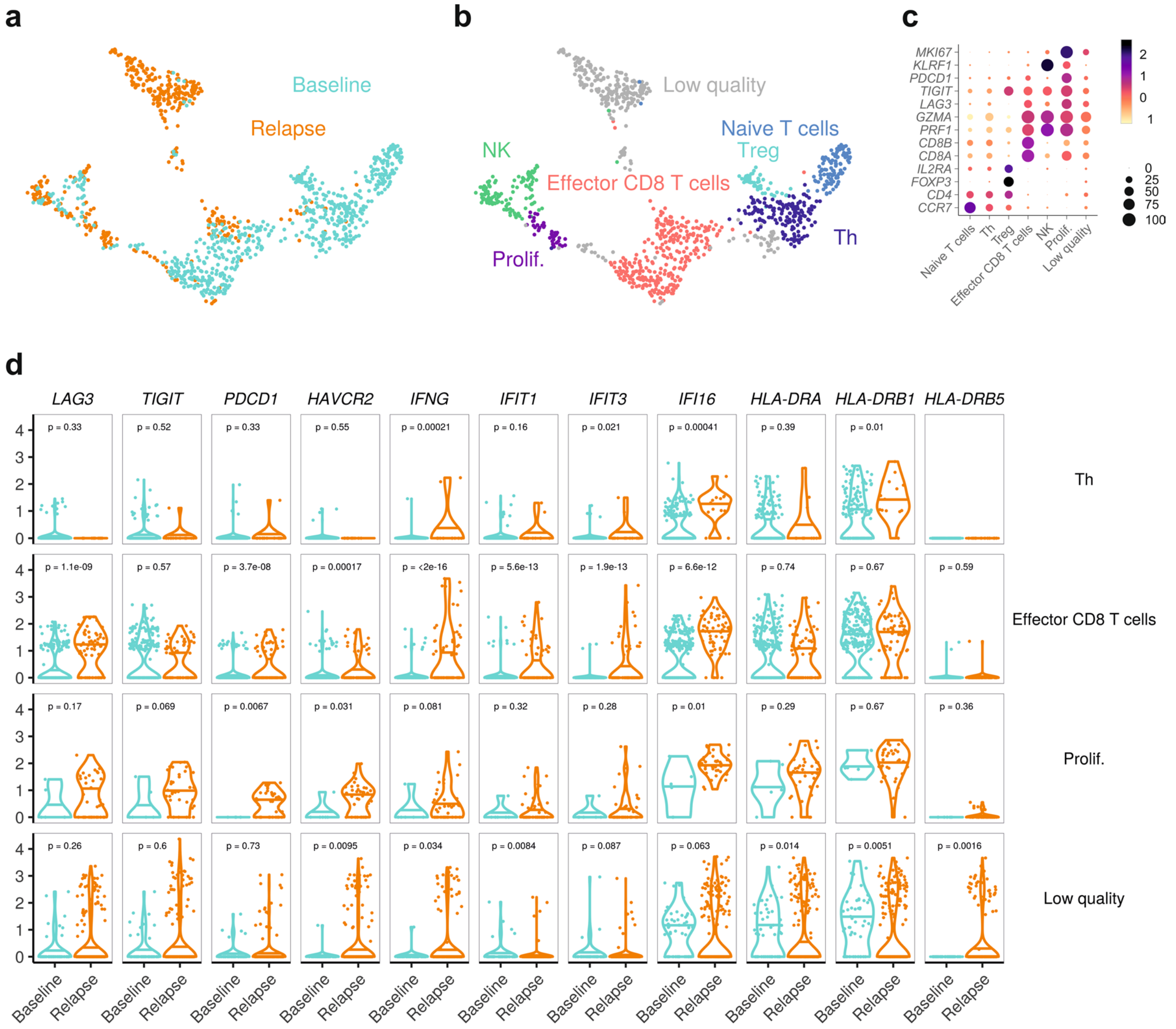
Extended Data Fig. 1 | Baseline characteristics and previous therapies. Summary of patient baseline characteristics and previous therapies before inclusion on the KarMMa trial. Abbreviations: Rev, Lenalidomide; Dara, Daratumumab; SC mob, stem cell mobilization; auto SCT, autologous stem cell transplantation; Pom, Polmalidomide; Carf, Carfilzomib; Cyclo, Cyclophosphamide; Dexa, Dexamethasone; Thal, Thalidomide.



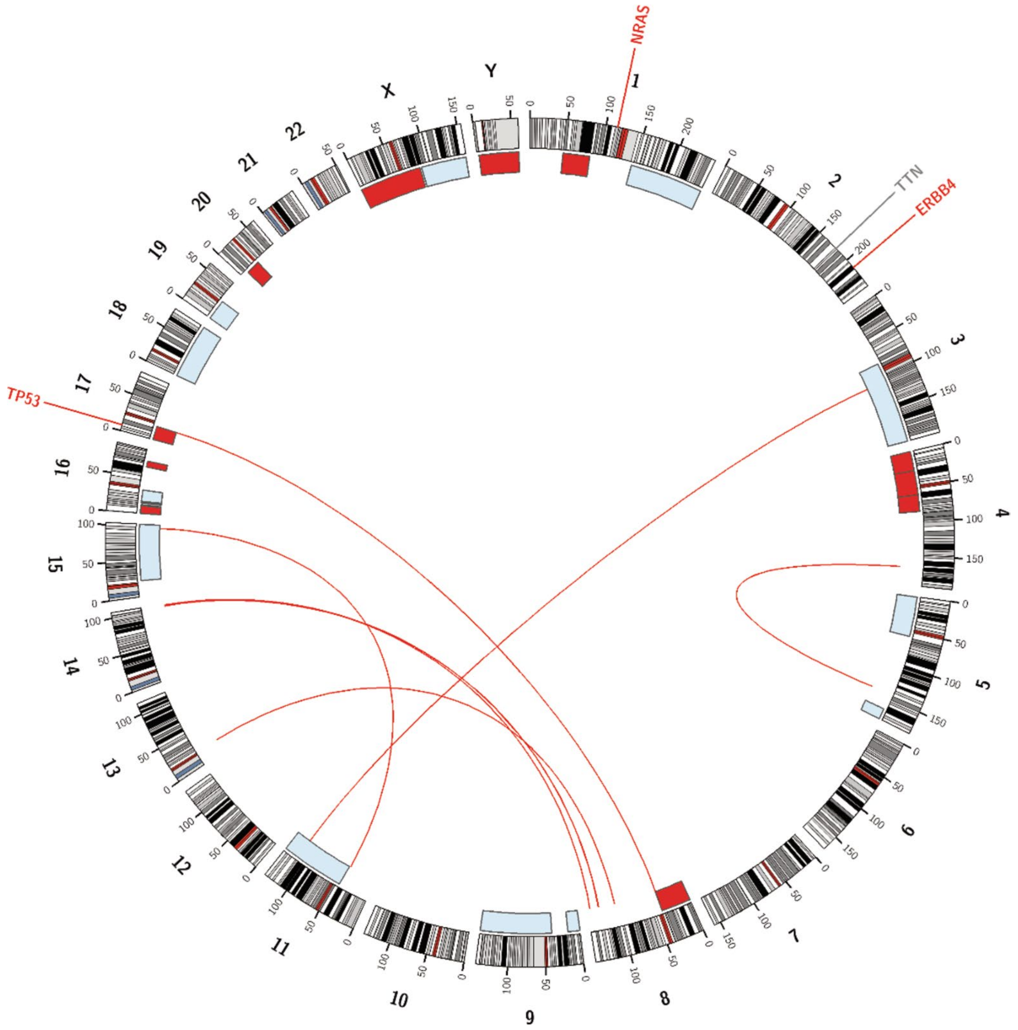
Extended Data Fig. 2 | Cell type annotation in bone marrow biopsies. a, Schematic of the experimental workflow. **b**, UMAP embedding of 5,592 single-cell transcriptomes from bone marrow biopsies before treatment (Baseline) and after relapse (Relapse). **c**, as b showing the annotated cell types. **d**, Dotplot showing the mean (of scaled log-normalized counts) expression (color) and the proportion in percentage of positive (non-zero expression) cells per cluster (size). MM, Multiple Myeloma. **e**, Log-normalized expression of marker genes specific for the identified cell types (MM: *IGHG1*+; B/DC: *HLA-DRA*+, *CD79B*+, *LILRA4*+; T/NK: *CD3D*+, *NKG7*+; Neutrophil: *HLA-DRA*-, *FCGR3B*+; Monocyte: *HLA-DRA*+, *CD68*+; Ery: *ALAS2*+) color coded on the UMAP embedding as panels **b** and **c**.



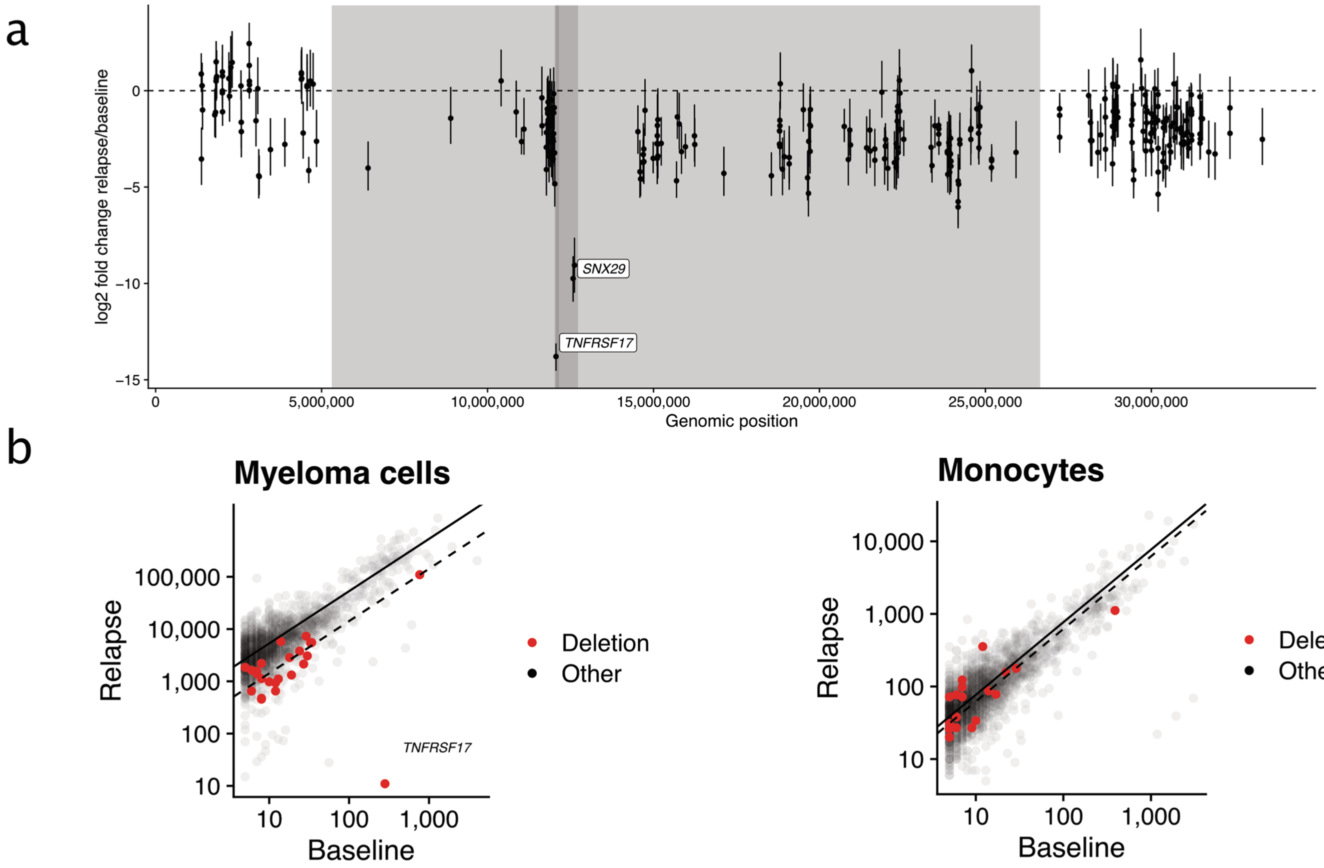
Extended Data Fig. 3 | *TNFRSF17* (*BCMA*) expression from baseline to relapse. a, UMAP embedding of 2,471 single-cell transcriptomes of Multiple Myeloma cells (MM cells from Extended Data Fig. 2 extracted and re-analyzed) from bone marrow biopsies before treatment (45 MM single cells, Baseline, blue) and after relapse (2,426 MM single cells, Relapse, orange). **b**, As panel **a** colored by the log-normalized expression of *TNFRSF17* (*BCMA*). **c**, As panel **a** colored by cluster. **d**, Dotplot of scaled, log-normalized expression (color) of marker genes identified for the different clusters shown in panel **d** and the proportion in percentage of positive (non-zero expression) cells per cluster (size). **e**, Violin plot showing the expression (log-normalized) of target antigens of immunotherapy between baseline and relapse. Horizontal line shows median of expression value. **f**, Expression of indicated genes shown in **e** color coded on top of the UMAP embedding as in panels **a** and **c**.



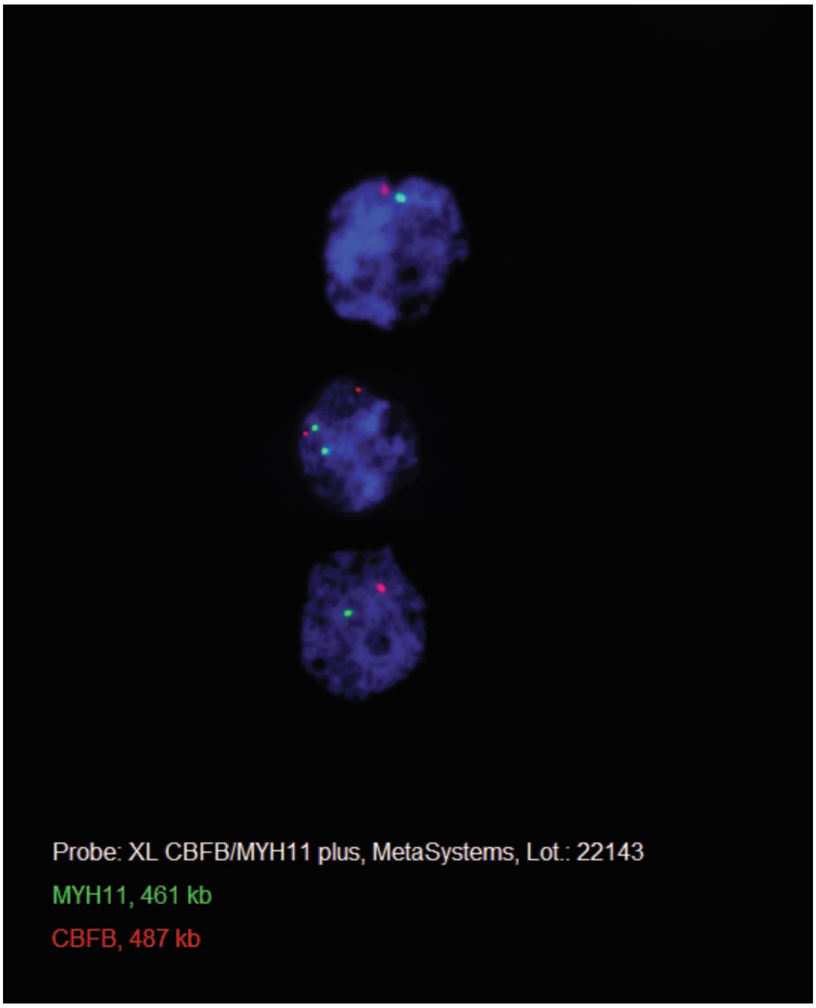
Extended Data Fig. 4 | T cell state characterization at baseline and relapse. **a**, UMAP embedding of 1,044 T cells from bone marrow biopsies before (635, blue dots) and at relapse (409, orange dots). **b**, UMAP embedding (as panel **a**) showing T cell subsets. **c**, Dotplot showing the scaled, log-normalized expression (color) of canonical marker genes for each cell subset and the proportion in percentage of positive (non-zero expression) cells per cluster (size). **d**, Violin-plots showing log-normalized expression of T cell exhaustion (*LAG3*, *TIGIT*, *PDCD1*, *HAVCR2*) and activation markers (*IFNG*, *IFIT1*, *IFIT3*, *IFI16*, *HLA-DRA*, *HLA-DRB1*, *HLA-DRB5*) between baseline (orange) and relapse (blue) across different cell subsets. P-values were calculated using the two-sided wilcoxon rank sum test, no adjustment for multiple comparisons was performed. Horizontal line shows median of expression value. Abbreviations: Th: Helper T cells; Treg: Regulatory T cells; NK: Natural killer cells; Prolif: Proliferating T cells.



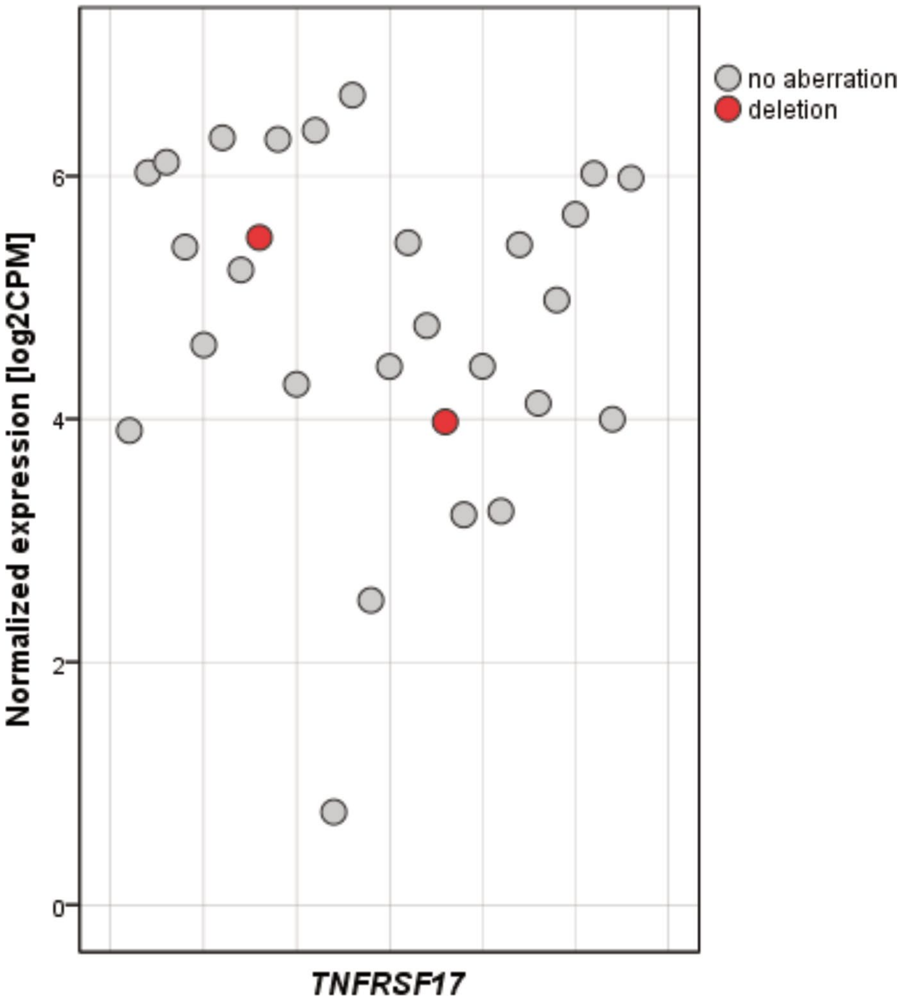
Extended Data Fig. 5 | Genomic architecture of baseline sample. Circos plot of CD138+ selected cells from the baseline sample including copy number variations, structural variants, and single nucleotide variations based on WGS data. Outer track runs clockwise from chromosome 1 to Y. Inner track shows gains > 1 Mb in blue and losses > 1 Mb in red. Red lines inside the circle represent interchromosomal reciprocal translocations with vaf > 0.1. Genes with mutations (*NRAS, ERBB4 and TP53*) are depicted in red and genes with detected variants of unknown significance (*TTN*) are shown in grey.



Extended Data Fig. 6 | Copy number variant analysis. a, Normalized Log2-ratio of gene expression within the cytogenetic band 16p13.13 between baseline and progression (n = number MM cells=45 MM cells at baseline and n = 2,426 MM cells at relapse). Error bars represent 90 % credible intervals. **b**, Total UMI count in non-malignant control cells at baseline vs. progression. We focused on the changes of expression levels from baseline to progression for genes located in the heterozygous deletion region. If the patient acquired the heterozygous deletion after treatment, systematic changes of gene expression are expected in this region. In contrast, this is not expected with the heterozygous deletion already present at baseline. In our patient, there were strong expression changes for genes in this region in malignant plasma cells (median > 3.7-fold difference) but not in monocytes (median < 1.3-fold difference) that we analyzed as negative control.


Probe: XL CBFB/MYH11 plus, MetaSystems, Lot.: 22143
MYH11, 461 kb
CBFB, 487 kb

Extended Data Fig. 7 | Interphase FISH of the 16p heterozygous deletion. Interphase Fluorescent-in-situ-hybridization (FISH) using the *CBFB/MYH11* probe set in a patient with heterozygous del(16p) and concomitant del(16q). Two aberrant cells show only 1 signal instead of 2 for the respective probes in line with two deletions encompassing the *MYH11* (located on 16p13.1) and CBFB (located on 16q22) locus.



Extended Data Fig. 8 | *TNFRSF17* RNA-seq expression data from 28 relapsed/refractory multiple myeloma (R/R MM) patients. *TNFRSF17* (*BCMA*) expression levels in patients with and w/o *BCMA* heterozygous deletions. Bulk RNAseq data for 28 R/R MM patients with availability of WGS data. Two patients showed a heterozygous *TNFRSF17* deletion (red dots), whereas the remaining patients did not show *BCMA* aberrations (grey dots). Expression levels did not differ between the two groups.

Cell

CellPress

Article

SARS-CoV-2 infection triggers profibrotic macrophage responses and lung fibrosis

Daniel Wendisch,[1,36] Oliver Dietrich,[2,36] Tommaso Mari,[3,36] Saskia von Stillfried,[4,36] Ignacio L. Ibarra,[5] Mirja Mittermaier,[1,6] Christin Mache,[7] Robert Lorenz Chua,[8] Rainer Knoll,[9,10] Sara Timm,[11] Sophia Brumhard,[1] Tobias Krammer,[2] Henrik Zauber,[3] Anna Luisa Hiller,[1] Anna Pascual-Reguant,[12,13] Ronja Mothes,[13,14] Roman David Bülow,[4] Jessica Schulze,[7] Alexander M. Leipold,[2] Sonja Djudjaj,[4] Florian Erhard,[15] Robert Geffers,[16] Fabian Pott,[6,17] Julia Kazmierski,[6,17] Josefine Radke,[6,14] Panagiotis Pergantis,[1] Kevin Baßler,[9,10]

(Author list continued on next page)

[1]Charité - Universitätsmedizin Berlin, Department of Infectious Diseases and Respiratory Medicine, Charité, Universitätsmedizin Berlin, Berlin, Germany
[2]Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz-Center for Infection Research (HZI), Würzburg, Germany
[3]Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Proteome Dynamics, Berlin, Germany
[4]Institute of Pathology, RWTH Aachen University Hospital, Aachen, Germany
[5]Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[6]Berlin Institute of Health (BIH), Berlin, Germany
[7]Unit 17 Influenza and other Respiratory Viruses, Robert Koch Institute, Berlin, Germany
[8]Center for Digital Health, Berlin Institute of Health (BIH) and Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany
[9]Systems Medicine, Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Bonn, Germany
[10]Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, University of Bonn, Germany
[11]Core Facility Electron Microscopy, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
[12]Charité - Universitätsmedizin Berlin, Department of Rheumatology and Clinical Immunology, Charité
[13]Deutsches Rheumaforschungszentrum, Immunodynamics, a Leibniz Institute, Berlin, Germany
[14]Charité - Universitätsmedizin Berlin, Department of Neuropathology, Berlin, Germany
[15]Institute for Virology and Immunobiology, Julius-Maximilians-University Würzburg, Würzburg, Germany
[16]Genome Analytics, Helmholtz-Center for Infection Research (HZI), Braunschweig, Germany

(Affiliations continued on next page)

SUMMARY

COVID-19-induced “acute respiratory distress syndrome” (ARDS) is associated with prolonged respiratory failure and high mortality, but the mechanistic basis of lung injury remains incompletely understood. Here, we analyze pulmonary immune responses and lung pathology in two cohorts of patients with COVID-19 ARDS using functional single-cell genomics, immunohistology, and electron microscopy. We describe an accumulation of CD163-expressing monocyte-derived macrophages that acquired a profibrotic transcriptional phenotype during COVID-19 ARDS. Gene set enrichment and computational data integration revealed a significant similarity between COVID-19-associated macrophages and profibrotic macrophage populations identified in idiopathic pulmonary fibrosis. COVID-19 ARDS was associated with clinical, radiographic, histopathological, and ultrastructural hallmarks of pulmonary fibrosis. Exposure of human monocytes to SARS-CoV-2, but not influenza A virus or viral RNA analogs, was sufficient to induce a similar profibrotic phenotype *in vitro*. In conclusion, we demonstrate that SARS-CoV-2 triggers profibrotic macrophage responses and pronounced fibroproliferative ARDS.

INTRODUCTION

“Severe acute respiratory syndrome coronavirus-2" (SARS-CoV-2), the causative agent of “coronavirus disease 2019” (COVID-19), initially infects and replicates in epithelial cells of the upper respiratory tract (Walls et al., 2020; Hoffmann et al., 2020; Wölfel et al., 2020). While SARS-CoV-2 infection causes mild respiratory disease in the majority of individuals, approximately 5% of patients develop acute respiratory distress syndrome (ARDS), which requires prolonged respiratory support and is associated with high mortality (Osuchowski et al., 2021; Richardson et al., 2020). ARDS is a clinical syndrome defined

© 2021 Published by Elsevier Inc.

Claudia Conrad,[1] Anna C. Aschenbrenner,[9,10,18,19] Birgit Sawitzki,[20] Markus Landthaler,[21] Emanuel Wyler,[21] David Horst,[22] Deutsche COVID-19 OMICS Initiative (DeCOI), Stefan Hippenstiel,[1,23] Andreas Hocke,[1,23] Frank L. Heppner,[14,24,25] Alexander Uhrig,[1] Carmen Garcia,[1] Felix Machleidt,[1] Susanne Herold,[23,26,27] Sefer Elezkurtaj,[22] Charlotte Thibeault,[1] Martin Witzenrath,[1,23] Clément Cochain,[28,29] Norbert Suttorp,[1,23] Christian Drosten,[17,30] Christine Goffinet,[6,17] Florian Kurth,[1,31,32] Joachim L. Schultze,[9,10,18] Helena Radbruch,[14] Matthias Ochs,[11,23,33] Roland Eils,[8] Holger Müller-Redetzky,[1] Anja E. Hauser,[12,13] Malte D. Luecken,[5] Fabian J. Theis,[5,34] Christian Conrad,[8] Thorsten Wolff,[7,37] Peter Boor,[4,37] Matthias Selbach,[3,35,37] Antoine-Emmanuel Saliba,[2,37,*] and Leif Erik Sander[1,23,37,38,*]

[17]Charité - Universitätsmedizin Berlin, Institute of Virology, Charité Universitätsmedizin Berlin, Berlin, Germany
[18]PRECISE Platform for Genomics and Epigenomics at Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), and University of Bonn, Bonn, Germany
[19]Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, the Netherlands
[20]Charité, Universitätsmedizin Berlin, Institute of Medical Immunology, Charité, Universitätsmedizin Berlin, Berlin, Germany
[21]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany
[22]Charité - Universitätsmedizin Berlin, Institute of Pathology, Charité - Universitätsmedizin Berlin, Berlin, Germany
[23]German Center for Lung Research (DZL), Germany
[24]Cluster of Excellence, NeuroCure, Berlin, Germany
[25]German Center for Neurodegenerative Diseases (DZNE) Berlin, Berlin, Germany
[26]Division of Infectious Diseases, Pulmonary and Critical Care Medicine, Department of Internal Medicine II, Universities of Giessen and Marburg Lung Center, Giessen, Germany
[27]Institute of Lung Health (ILH), Giessen, Germany
[28]Comprehensive Heart Failure Center Wuerzburg, University Hospital Würzburg, Germany
[29]Institute of Experimental Biomedicine, University Hospital Würzburg, Würzburg, Germany
[30]German Center for Infection Research (DZIF), Braunschweig, Germany
[31]Department of Tropical Medicine, Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany
[32]I. Department of Medicine, University Medical Center, Hamburg-Eppendorf, Hamburg, Germany
[33]Institute of Functional Anatomy, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
[34]Department of Mathematics, Technische Universität München, Garching bei München, Germany
[35]Charité-Universitätsmedizin Berlin, Berlin, Germany
[36]These authors contributed equally
[37]Senior author
[38]Lead contact
*Correspondence: emmanuel.saliba@helmholtz-hiri.de (A.-E.S.), leif-erik.sander@charite.de (L.E.S.)
https://doi.org/10.1016/j.cell.2021.11.033

as a combination of acute hypoxemia and bilateral radiographic opacities not explained by cardiac dysfunction or volume overload, occurring within seven days after a known clinical insult or onset of new respiratory symptoms (Ferguson et al., 2012). The pathophysiology of ARDS is typically viewed as a bi- or triphasic response to lung injury. The initial "exudative phase" is characterized by injury-induced and myeloid-cell-propagated diffuse alveolar damage, barrier breakdown, and edema. It is followed by a second "proliferative phase," marked by epithelial cell repair, reabsorption of fluid, and restoration of alveolar integrity. The third "fibrotic phase" only occurs in a subset of patients, and it is associated with respiratory failure and high mortality (Thompson et al., 2017).

COVID-19-induced ARDS requires protracted mechanical ventilation or extracorporeal membrane oxygenation (ECMO), and it is associated with high mortality (Barbaro et al., 2020; Hasan et al., 2020; Henry and Lippi, 2020). Respiratory failure and ARDS typically develop in the second week after symptom onset, in spite of declining viral loads (Barbaro et al., 2020; He et al., 2020; Jones et al., 2021; Wölfel et al., 2020). The delayed onset suggests that ARDS in COVID-19 is driven by secondary events, including inappropriate immune responses (Carsana et al., 2020; D'Alessio and Heller, 2020; Polak et al., 2020). Supporting this notion, dysregulated systemic and mucosal immune responses, particularly within the myeloid cell compartment, have been observed in severe COVID-19 (Bharat et al., 2020; Carsana et al., 2020; Chua et al., 2020; D'Alessio and Heller, 2020; Grant et al., 2021; Liao et al., 2020; Messner et al., 2020; Schulte-Schrepping et al., 2020; Szabo et al., 2021). It has been repeatedly suggested that a subset of COVID-19 patients develops a detrimental hyperinflammatory condition, and monocytes and macrophages have been proposed as critical mediators of this inflammatory syndrome (Blanco-Melo et al., 2020; Mehta et al., 2020; Webb et al., 2020). In contrast, systematic analysis of severe COVID-19 showed that the levels of inflammatory cytokines were one or two orders of magnitude lower than in comparable cohorts of non-COVID-19-induced ARDS (Calfee et al., 2014; Leisman et al., 2020; Sinha et al., 2020), calling into question the concept of a "cytokine storm" or "inflammatory macrophage" syndrome (Remy et al., 2020). In line with this observation, monocyte and neutrophil populations in the peripheral blood of critically ill COVID-19 patients show a suppressive phenotype (Schulte-Schrepping et al., 2020). However, anti-inflammatory treatments like dexamethasone and anti-interleukin 6 (IL-6)/anti-IL-6R antibodies or janus kinase (JAK) inhibitors improve clinical outcomes when administered in the early phase of severe COVID-19 (Guimarães et al., 2021; Patel et al., 2021; Horby et al., 2021), indicating that inflammatory mediators contribute to early organ injury in COVID-19.

Besides their role in host defense and inflammation, macrophages and monocytes are important mediators of tissue repair, remodelling, and fibrosis (Adler et al., 2020; Henderson et al., 2020; Wynn and Vannella, 2016). Distinct macrophage phenotypes have been implicated in the different phases of ARDS. Inflammatory macrophages are viewed as key propagators of lung injury during the exudative phase, whereas regulatory macrophages, previously referred to as "alternatively activated" or "M2" macrophages, have been associated with the proliferative and fibrotic phase of ARDS (Herold et al., 2015; Thompson et al., 2017). Several studies have reported an accumulation of distinct monocytes and macrophages in COVID-19 (Bharat et al., 2020; Grant et al., 2021; Liao et al., 2020), yet their functional contribution to ARDS remains unknown.

Here, we investigate pulmonary immune responses in severe COVID-19 at single-cell resolution. We describe a pronounced infiltration of monocyte-derived macrophages that acquired transcriptional signatures reminiscent of profibrotic macrophages identified in idiopathic pulmonary fibrosis (IPF). Alongside the potentially profibrotic program in pulmonary macrophages, we observed a pronounced expansion and activation of myofibroblasts and fibroblasts and their engagement in extensive intercellular communication networks with pulmonary macrophages. The IPF-like transcriptional profile was induced by the exposure of monocytes from healthy donors to SARS-CoV-2, but not to influenza A virus (IAV). Consistent with the observed macrophage and mesenchymal cell responses, clinical, histological, and ultrastructural analyses revealed extensive fibrotic tissue remodelling, indicative of exacerbated fibroproliferative response in COVID-19-associated ARDS.

RESULTS

Pulmonary CD163$^+$ macrophages accumulate in COVID-19 ARDS

We analyzed tissue samples and bronchoalveolar lavage (BAL) fluid samples from 47 patients with severe COVID-19 using functional single-cell genomics (single-cell RNA-sequencing [scRNA-seq] and single-nucleus RNA-sequencing [snRNA-seq]), multi-epitope ligand cartography (MELC), immunofluorescence (IF) microscopy, immunohistochemistry (IHC), RNA-fluorescence *in situ* hybridization (RNA-FISH), and transmission electron microscopy (EM), complemented by clinical evaluations, including lung mechanics and computed tomography (CT) imaging (Figure 1A and S1A; Table S1). In line with previous reports, histopathological analysis of lung autopsy samples revealed diffuse alveolar damage, fibroproliferative responses, and features of organizing pneumonia (Bharat et al., 2020; Carsana et al., 2020; Osuchowski et al., 2021; Speranza et al., 2021) (Figure 1B). Accumulation of macrophages in the damaged lung tissue was detected by IHC and IF microscopy of CD68$^+$ and hemoglobin-haptoglobin scavenger receptor CD163$^+$ cells (Figures 1B and 1C). Quantification of CD68$^+$ cells revealed a significant increase of macrophage density and a higher proportion of macrophages expressing CD163 in COVID-19 (Figure 1D). SARS-CoV-2 RNA was detected within epithelial cells and macrophages by RNA-FISH and consecutive IHC (Figures 1E and S1B). For a detailed analysis of pulmonary immune cell populations, we performed MELC (Schubert et al., 2006). We stained lung tissue samples of 9 autopsy cases of fatal COVID-19 with a panel of 22 immune cell markers (Figures S1C–S1E; STAR Methods) revealing a predominance of myeloid cells (CD45$^+$, CD3$^-$, CD4$^-$, CD20$^-$, and variable expression of CD11b, CD14, CD16, CD66b, CD68, CD84, HLA-DR, TREM1) with prominent clusters of macrophages (Figures 1F and 1G), in line with previous reports (Bharat et al., 2020; Carsana et al., 2020; Chua et al., 2020; Speranza et al., 2020). MELC confirmed the accumulation of CD163$^+$ macrophages, some of which co-expressed chemokine receptor CXCR3 and complement factor C1Q (Figures 1F, 1G, S1E, and S1F). In addition, we noted a prominent deposition of collagen (Figure S1C).

Monocyte-derived macrophages adopt a damage response signature in severe COVID-19

In order to gain a higher resolution of pulmonary immune response during severe COVID-19, we analyzed single-cell transcriptomes of BAL cells in patients with COVID-19-associated ARDS (Figures 2A and S2A–S2D; Table S1). Consistent with postmortem analysis of lung tissue (Figure 1) and in agreement with previous reports (Grant et al., 2021; Liao et al., 2020; Szabo et al., 2021), we identified a dominant proportion of myeloid cells, particularly neutrophils and monocytes/macrophages (Figures 2A and S2E; Table S2). SARS-CoV-2 transcripts were primarily detected within monocytes/macrophages (Figure S2C).

A detailed analysis revealed six monocyte/macrophage populations with distinct gene expression profiles (Figures 2B–2D, S2F, and S2G; Table S2). "*FCN1*-Monocytes" (*FCN1*-Mono) expressed high levels of ficolin-M (*FCN1*) and *CD14* but low levels of *FCGR3A*, encoding for CD16a (Figures 2C and 2D). *FCN1*-Mono were marked by a high expression of alarmins (*S100A8*, *S100A12*), selected inflammatory cytokines (*IL1B, IL6, CXCL8*), and *CCR2,* the receptor for monocyte chemoattractant protein-1 (MCP-1/CCL2) (Figures 2C, 2D, and S2F). Notably, *FCN1*-Mono also expressed *TGFB1*, encoding for TGF-β, a master regulator of wound healing and repair, and its downstream target *TGFBI*. The adjacent "Mono/Mϕ" population showed a less distinct phenotype, indicative of a transitory differentiation state. (Figures 2D and S2F). Mono/Mϕ and the adjacent monocyte-derived macrophage population expressed high levels of *SPP1*, encoding for Osteopontin, a multifunctional matricellular protein and cytokine expressed in macrophages in various pathologies (Rittling, 2011). The *SPP1*$^+$ macrophage population was defined by high expression of *CD163* and *LGMN*, encoding for Legumain, and is henceforth referred to as "*CD163/LGMN*-Mϕ" (Figures 2C and 2D). We also identified three types of alveolar macrophages (AMϕ1, AMϕ2, and proliferating AMϕ), characterized by high expression of *FABP1*, *FABP4*, and *MARCO* (Figure 2D), characteristic for AMϕ (Arredouani et al., 2005). AMϕ-2 expressed high levels of TGF-β family member "inhibin beta A" (*INHBA*), while "proliferating-AMϕ" expressed cell-cycle-related genes (*MKI67*, *TOP2A*, *NUSAP1*) (Figures 2C and 2D).

Transcriptional analysis indicated that infiltrating monocytes (*FCN1*-Mono) differentiated along a curved trajectory (Figure 2B) toward *CD163/LGMN*-Mϕ, AMϕ1, and AMϕ2 (Figure 2B). The

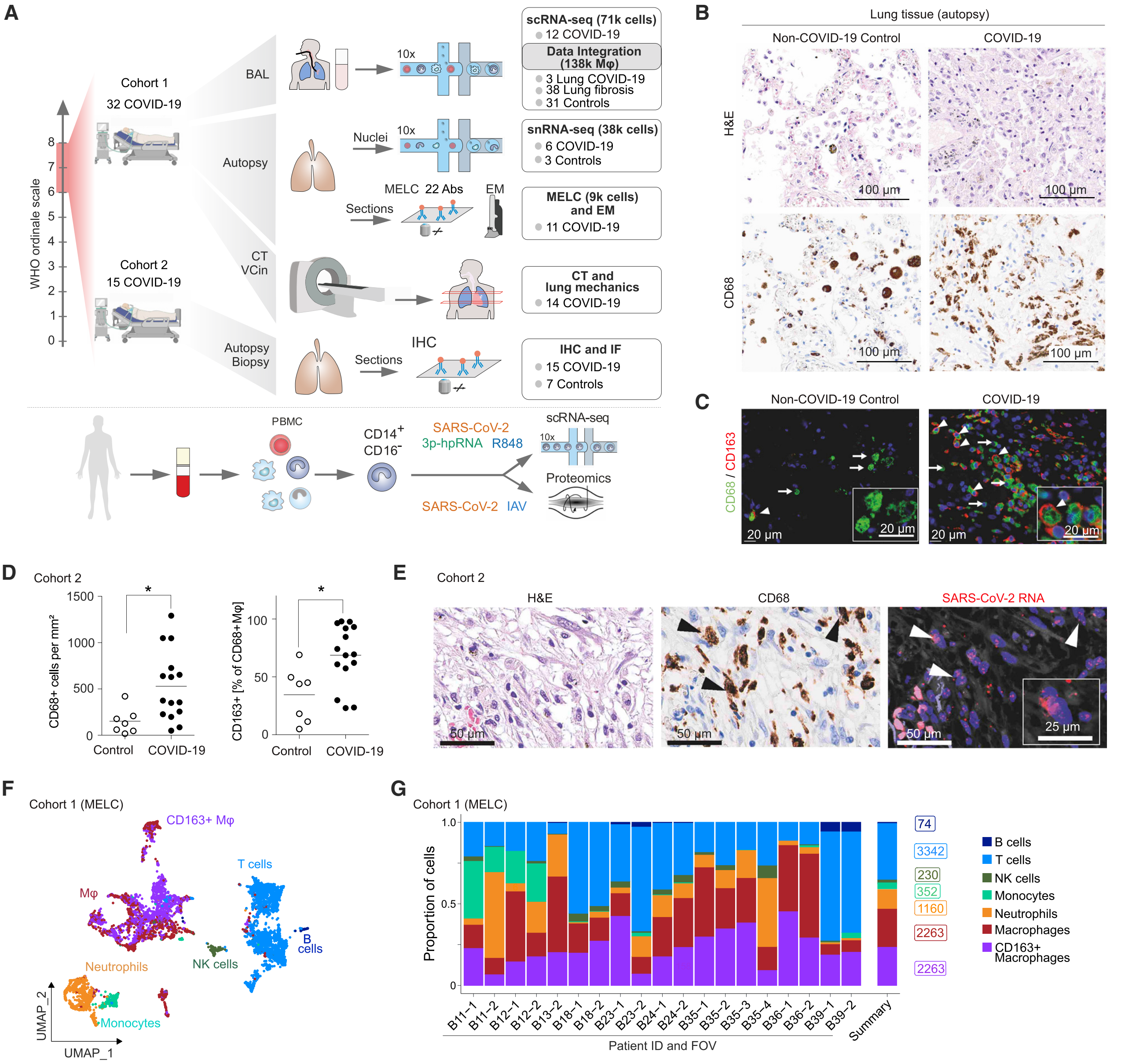

A
Cohort 1
32 COVID-19
Cohort 2
15 COVID-19
WHO ordinale scale
8
7
6
5
4
3
2
1
0
BAL
Autopsy
CT
VCin
Autopsy
Biopsy
10x
Nuclei
MELC 22 Abs
EM
Sections
IHC
scRNA-seq (71k cells)
12 COVID-19
Data Integration
(138k Mφ)
3 Lung COVID-19
38 Lung fibrosis
31 Controls
snRNA-seq (38k cells)
6 COVID-19
3 Controls
MELC (9k cells)
and EM
11 COVID-19
CT and
lung mechanics
14 COVID-19
IHC and IF
15 COVID-19
7 Controls
PBMC
CD14+
CD16−
SARS-CoV-2
3p-hpRNA
R848
SARS-CoV-2
IAV
scRNA-seq
Proteomics
B
Lung tissue (autopsy)
Non-COVID-19 Control
COVID-19
H&E
CD68
100 μm
C
Non-COVID-19 Control
COVID-19
CD68 / CD163
20 μm
D
Cohort 2
CD68+ cells per mm²
1500
1000
500
0
*
Control
COVID-19
CD163+ [% of CD68+Mφ]
100
50
0
E
Cohort 2
H&E
CD68
SARS-CoV-2 RNA
50 μm
25 μm
F
Cohort 1 (MELC)
CD163+ Mφ
Mφ
T cells
B cells
NK cells
Neutrophils
Monocytes
UMAP_2
UMAP_1
G
Cohort 1 (MELC)
Proportion of cells
1.0
0.5
0
B11-1
B11-2
B12-1
B12-2
B13-2
B18-1
B18-2
B23-1
B23-2
B24-1
B24-2
B35-1
B35-2
B35-3
B35-4
B36-1
B36-2
B39-1
B39-2
Summary
Patient ID and FOV
74
3342
230
352
1160
2263
2263
B cells
T cells
NK cells
Monocytes
Neutrophils
Macrophages
CD163+
Macrophages

Figure 1. CD163$^+$ macrophages accumulate in the lung in severe COVID-19

(A) Overview of study design and analyses. CT, computed tomography; BAL, bronchoalveolar lavage; scRNA-seq, single-cell RNA sequencing; snRNA-seq, single-nucleus RNA sequencing; IHC, immunohistochemistry; IF, immunofluorescence microscopy; MELC, multi-epitope ligand cartography; EM, electron microscopy; VCin, inspiratory vital capacity; PBMC, peripheral blood mononuclear cells; IAV, Influenza A virus.

(B) Postmortem analysis of consecutive histological sections of non-COVID-19 (left) and COVID-19 autopsy lung samples (right) by hematoxylin and eosin (H&E; top) and CD68 IHC (bottom). Scale bar, 100 μm.

(C) IF of CD68 (green) and CD163 (red) in lung tissue autopsy samples of COVID-19 patients and non-COVID-19 controls. Arrows indicate CD68$^+$CD163$^-$ macrophages, and arrowheads indicate CD68$^+$CD163$^+$ macrophages. Scale bar, 20 μm.

(D) Quantification of CD68$^+$ macrophage density (left) and the proportion of CD163$^+$ macrophages (right) in lung autopsy samples from fifteen donors (as in C). Mann-Whitney test; $^*p < 0.05$.

(E) Representative images of consecutive histological sections of lung autopsy samples. H&E (left), CD68 IHC (middle), and SARS-CoV-2 RNA-FISH (right). Arrowheads indicate SARS-CoV-2 RNA-positive macrophages. Scale bars, 50 μm, 25 μm. RNA-FISH, RNA-fluorescence *in situ* hybridization.

(F) Lung autopsy samples of 9 COVID-19 patients were analyzed by MELC with a panel of 22 markers on 19 fields of view (FOVs). Two-dimensional embedding computed by UMAP on 9,684 computationally identified CD45 positive cells (T cells, CD3$^+$; B cells, CD20$^+$; NK cells, CD56$^+$; neutrophils, MRP14$^+$/CD66b$^+$; monocytes, MRP14$^+$/CCR2$^+$; macrophages, MRP14$^+$/HLA-DR$^+$).

(G) Relative proportion (of total CD45$^+$ cells) of cell types in all 19 FOVs (left), and average cell numbers (summary, right).

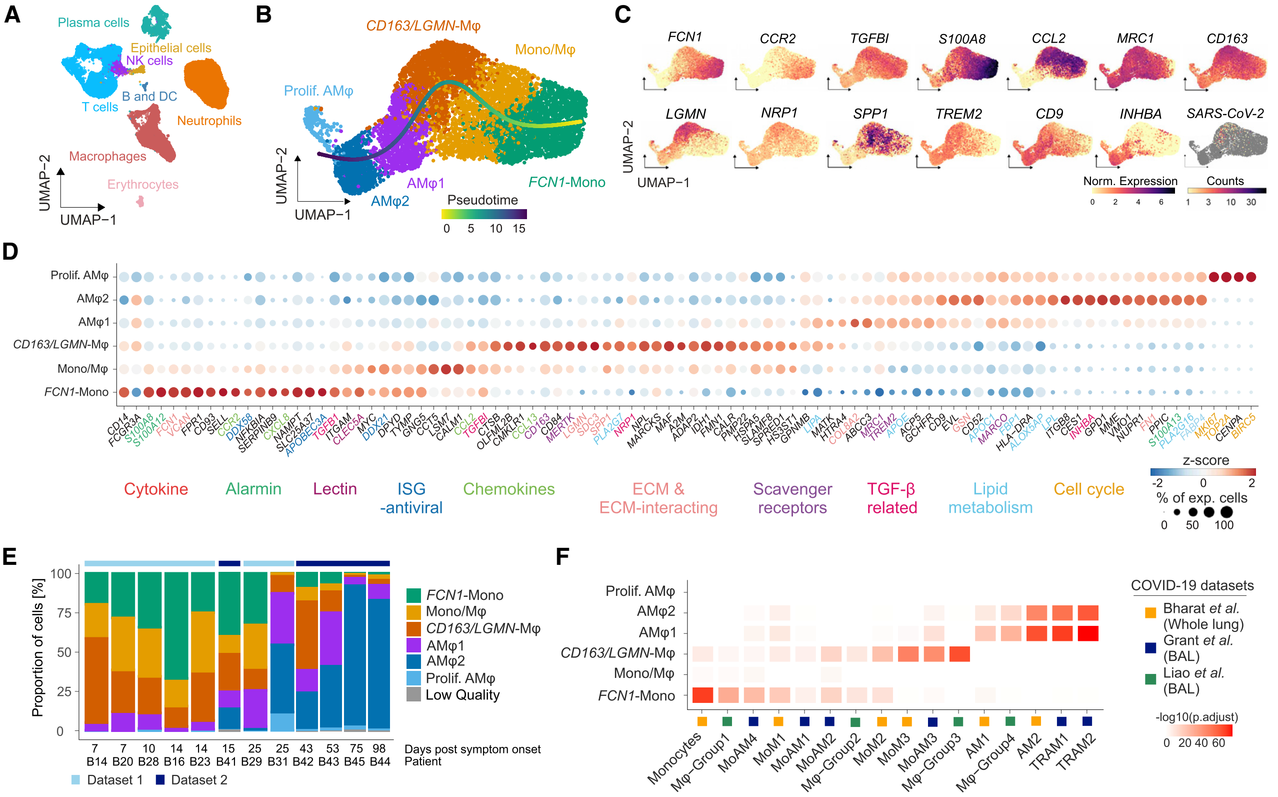


Figure 2. Monocyte-derived macrophages adopt a damage response signature in severe COVID-19

(A) UMAP (Uniform Manifold Approximation and Projection) embedding of 46,060 single-cell transcriptomes in the BAL fluid of patients with severe COVID-19 ARDS. Cell-type annotation was based on expression of canonical marker genes.

(B) UMAP embedding and slingshot trajectory of 7,816 transcriptomes of monocytes/macrophages identified in (A). Clusters were defined by comparing gene expression patterns of Leiden clusters (Mono, monocytes; Mono/Mϕ, transitory monocyte-macrophages; AMϕ, alveolar macrophages).

(C) Marker gene expression and SARS-CoV-2 mRNA counts, color-coded and projected onto the UMAP embedding in (B). Statistical significance of differential expression for each gene per cluster shown in Table S2.

(D) Dot plot of scaled, log-normalized expression of marker genes of the clusters in (B). Gene names color-coded by functional categories. Dot size indicates percentage of cells per cluster with any mRNAs detected, and color shows *Z*-scores of log-normalized mRNA counts. Statistics in Table S2.

(E) Relative proportions of cell types across all BAL scRNA-seq samples derived from (B) and Figure S2M ordered by increasing days post symptom onset.

(F) Heatmap displaying -log10 transformed adjusted p values (one-sided Fisher's exact test) assessing the overlap between gene sets from COVID-19-associated monocyte/macrophage clusters identified in (B) (y axis) and published transcriptional signatures of COVID-19-associated monocytes/macrophages (cluster names and reference studies indicated; Table S3).

prominent expression of monocyte chemoattractant *CCL2* in *FCN1*-Mono, Mono/Mϕ, and *CD163/LGMN*-Mϕ suggested a feedforward loop of monocyte recruitment and subsequent macrophage differentiation (Figure 2D).

Macrophage populations were characterized by the expression of genes related to TGF-β signaling (*TGFB, TGFBI, INHBA, NRP1*), scavenger receptors and molecules associated with apoptotic cell uptake (*MRC1, CD163, MERTK, TREM2, MARCO*), lipid handling and -metabolism (*PLA2G7, APOC, APOE, LIPA, LPL, FBP1, FABP4*), and extracellular matrix (ECM) components or molecules involved in ECM interaction and breakdown (*FN, VCAN, SPP1, LGMN, MMP9, SDC3*) (Figures 2C and 2D). Transcription factor enrichment analysis using the ChEA3 tool (Keenan et al., 2019) predicted *SPI1* (encoding PU.1) as regulatory in *FCN1*-Mono and transcription factor EC (TFEC) and glycosylated lysosomal membrane protein (GLMP) for *CD163/LGMN*-Mϕ, while nuclear receptor peroxisome proliferator-activated receptor gamma (PPARγ), involved in alveolar macrophage differentiation, lipid homeostasis, and repression of inflammatory macrophage responses (Kidani and Bensinger, 2012; Schneider et al., 2014), was predicted to control transcription in AMϕ1 and AMϕ2 (Figure S2H; Table S2).

We corroborated these findings on a second set of BAL samples from patients with severe COVID-19 ARDS, revealing similar macrophage populations (Figures S2I–S2O). *FCN1*-Mono, Mono-Mϕ, and *CD163/LGMN*-Mϕ were dominant in the first four weeks of COVID-19 ARDS, and AMϕ1 and AMϕ2 repopulated the alveolar compartment in the later stages of the disease (Figure 2E). For further validation, we compared the transcriptional phenotypes of macrophages identified in our cohort with previously published transcriptomes of pulmonary macrophages in COVID-19 (Grant et al., 2021; Liao et al., 2020). We

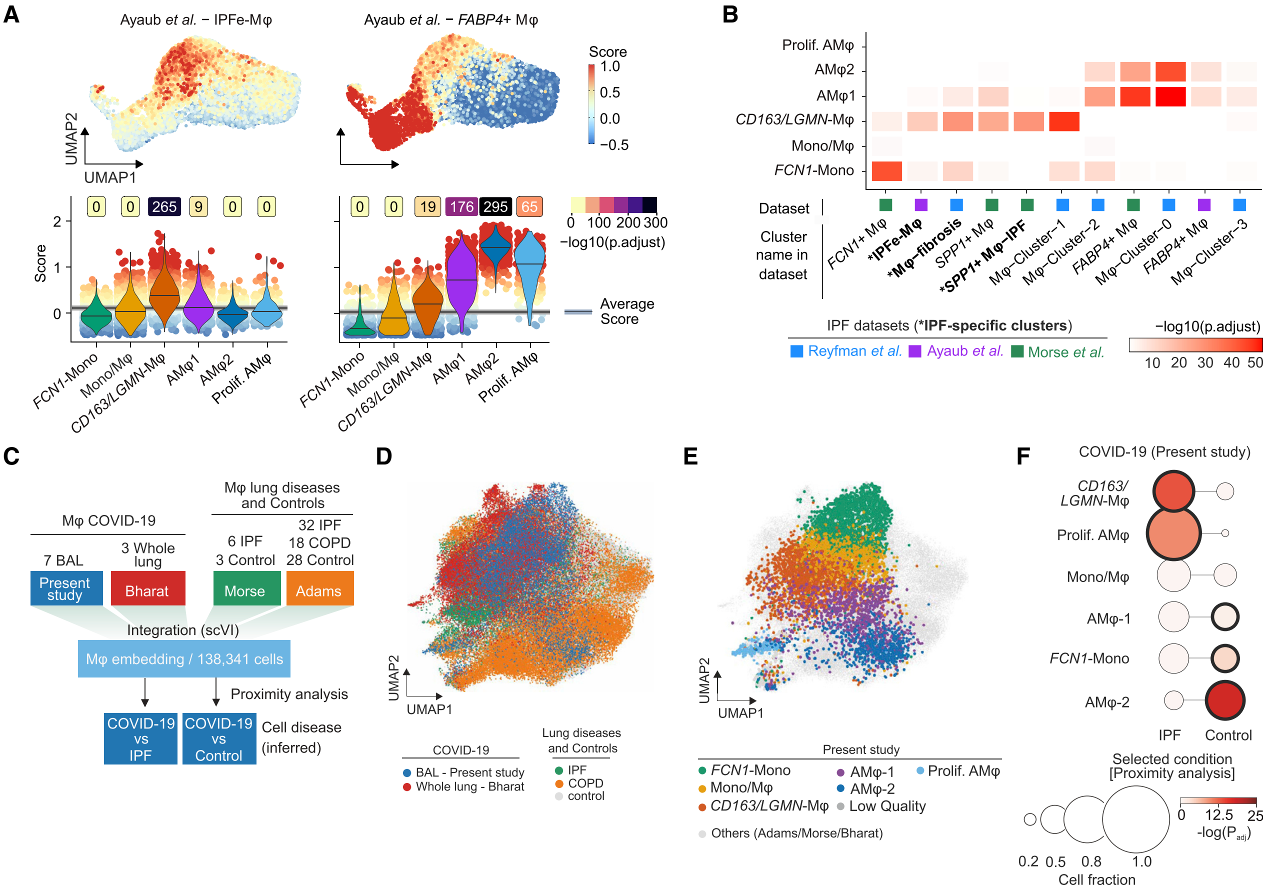


Figure 3. Gene set enrichment and data integration reveals a profibrotic phenotype of COVID-19-associated macrophages

(A) Gene set module score of “IPF-expanded macrophages” (IPFe-Mϕ) and alveolar *FABP4*+Mϕ (Ayaub et al., 2021), calculated based on single transcriptomes. Projected onto the UMAP embedding (top) and plotted as violin plots (bottom) across the monocyte/macrophage clusters (annotated in Figure 2B). Dot color indicates signature module score. Violin colors show cluster identity, numbers indicate -log10 transformed adjusted p values (one-sided wilcoxon test compared to average), and lines in violins indicate median scores per cluster.

(B) Heatmap representing -log10 transformed adjusted p values (one-sided Fisher’s exact test) assessing the overlap of gene sets from monocyte/macrophage clusters identified in Figure 2B (y axis) and published transcriptional signatures of monocyte/macrophage clusters derived from the indicated IPF datasets (cluster names and reference studies indicated on the x axis; Table S3).

(C) Schematic depicting monocyte/macrophage data integration from present study and Bharat et al. (2020) with two human lung fibrosis reference datasets (Adams et al., 2020; Morse et al., 2019) via scVI. COVID-19 macrophages were mapped to IPF or control macrophages based on a kNN (k-nearest neighbor)-proximity mapping.

(D) UMAP of 138,341 cells derived from all four datasets based on integrated scVI embedding.

(E) UMAP as in (D) highlighting COVID-19-associated macrophage clusters annotated in Figure 2B. Cells from reference datasets shaded in gray.

(F) Proximity analysis of macrophage clusters annotated in Figure 2B and macrophages identified in IPF and healthy controls, respectively. Circle size shows cell fraction, color codes indicate the -log10 transformed adjusted p values, and bold black circle indicates statistical significance (adjusted p < 0.0001, Fisher’s exact test, one-tailed with Benjamini-Hochberg correction).

found highly similar macrophage profiles in all three COVID-19 datasets (Figure 2F), confirming that the accumulation of monocyte-derived macrophages with damage response- and tissue repair signatures is a common feature of COVID-19 ARDS.

Pulmonary macrophages in COVID-19 adopt a profibrotic phenotype

Misguided damage repair- and wound healing responses by monocytes and macrophages play a key role in tissue remodeling, scarring, and fibrosis (Adler et al., 2020; Henderson et al., 2020; Misharin et al., 2017; Satoh et al., 2017). Macrophages within fibrotic niches in pulmonary fibrosis express shared transcriptional programs (Aran et al., 2019; Ayaub et al., 2021; Joshi et al., 2020; Morse et al., 2019; Reyfman et al., 2019). We therefore assessed the transcriptional similarity between monocytes and macrophages in severe COVID-19 and macrophages from four published datasets of pulmonary fibrosis (Adams et al., 2020; Ayaub et al., 2021; Reyfman et al., 2019; Morse et al., 2019). We computed cell-based scores based on gene set expression and assessed differences across populations (Figures 3A and S3A), in addition to overrepresentation analysis between gene sets (Figure 3B). These comparisons indicated that the *CD163/LGMN*-Mϕ population closely resembles IPF-specific macrophage phenotypes, while *FCN1*-Mono and AMϕ1 and AMϕ2 were more similar to homeostatic monocytes and alveolar macrophages, respectively (Figures 3A, 3B, and S3A; Table S3).

In order to directly compare the cellular transcriptomes of COVID-19-associated and IPF-associated macrophages, we integrated our scRNA-seq data along with another COVID-19 dataset (Bharat et al., 2020) with two scRNA-seq datasets containing IPF, COPD (chronic obstructive pulmonary disease), and control samples (Adams et al., 2020; Morse et al., 2019) using single-cell variational inference (scVI) (Lopez et al., 2018; Wolf et al., 2018) (Figures 3C–3F). The joint embedding of 138,341 macrophage transcriptomes revealed significant similarity of *CD163/LGMN*-Mϕ and proliferating-AMϕ populations with IPF-associated macrophages (Figures 3D–3F, S3B, and S3C). Both IPF- and COVID-19-associated macrophages expressed genes with well-known pathogenic functions in fibrosis, such as *SPP1, TGFB1, TGFBI, LGMN,* and *CCL18* (Figure S3D). Proximity analysis based on a k-nearest neighbors classifier revealed a significant similarity of *CD163/LGMN*-Mϕ and proliferating-AMϕ with IPF-associated macrophages (relative fractions 0.66 and 0.83, respectively; adjusted $p < 0.0001$; Fisher's exact test, one-tailed) (Figure 3F). AMϕ2 (0.63) and *FCN1*-Mono (0.47), which resemble steady-state alveolar macrophages and monocytes, were embedded in proximity to macrophages from healthy lungs (Figures 3F, S3B, and S3C). Similar results were obtained analyzing a published COVID-19 dataset (Bharat et al., 2020), confirming that monocyte-derived macrophages (MoM3), which corresponded to *CD163/LGMN*-Mϕ (Figure 2F), showed a high similarity with IPF-associated macrophages (Figures S3C and S3E).

Gene set overrepresentation, scRNA-seq data integration, and proximity analyses revealed that pulmonary *CD163/LGMN*-Mϕ in COVID-19 showed significant transcriptional similarity with macrophages found in IPF. The data indicate that newly recruited monocyte-derived macrophages in COVID-19 adopt a fibrosis-associated phenotype.

Interaction of pulmonary macrophages and mesenchymal cells in COVID-19 ARDS

To further investigate the association of pulmonary macrophage populations and fibrotic responses, we analyzed postmortem lung tissue samples from patients with fatal COVID-19. Using snRNA-seq, we identified 15 distinct populations of lung cells based on canonical marker genes (Figure 4A) (Lukassen et al., 2020). We found macrophage phenotypes similar to those described in BAL (Figures 4B and S4A), as well as pericytes, smooth muscle cells, fibroblasts, and myofibroblasts (Figures 4C and S4B). Fibroblasts and myofibroblasts showed strong upregulation of ECM protein-encoding genes in COVID-19, particularly in later stages of the disease (days 34–82), indicating a strong fibrotic response (Figure S4C). We next inferred communication networks between macrophages and mesenchymal cells using the ligand-receptor interaction tool CellChat (Jin et al., 2021). The analysis revealed strong interactions of *CD163/LGMN*-Mϕ and, to a lesser extent, Mono/Mϕ with myofibroblasts, fibroblasts, and pericytes (Figure 4D). The interaction strength was increased at later time points (days 34–82) relative to earlier stages of the disease (days 7–21) (Figure 4D). Information flow analysis showed an involvement of potent profibrotic pathways, including *Col*, *FGF* (fibroblast growth factor), *TGFB1,* and *SPP1*, among others (Figure 4E).

We next assessed tissue distribution of SM22$^+$ myofibroblasts and CD68$^+$ macrophages by IF in autopsy lung samples. Myofibroblasts were expanded compared to non-COVID-19 controls, and macrophages were frequently found in close proximity to SM22$^+$ loci (Figures 4F and S4D). We also noted a marked co-localization of CD163$^+$ macrophages and collagen deposits (Figures 4G, S4E, and S4F); however, expansion of both collagen areas and macrophages increases the probability of colocalization.

In summary, we found that pulmonary macrophages colocalized with collagen and myofibroblast loci and engaged in strong signal interactions with myofibroblasts, fibroblasts, and pericytes. Interactions between *CD163/LGMN*-Mϕ and mesenchymal cells may thus contribute to a highly profibrotic milieu in COVID-19 ARDS.

COVID-19 is associated with pronounced fibroproliferative ARDS

Identification of fibrosis-associated transcriptional signatures in macrophages and their interactions with mesenchymal cells in COVID-19 ARDS prompted us to assess evidence of lung fibrosis in COVID-19. We analyzed a cohort of 16 patients with severe COVID-19-induced ARDS (60.5 years old [IQR 16.3], 94% male), defined by the requirement of veno-venous (vv) ECMO (Table S4). The mean duration of vvECMO therapy was 36.5 days (IQR 57), the median time from start of mechanical ventilation to initiation of vvECMO was 12.5 days (IQR 10), 68.8% of patients could be weaned from vvECMO, and the overall mortality was 56.3% (Table S4). Patients in this cohort had a P_aO_2 /F_iO_2 ratio of 98.3 mmHg (IQR 56.9) (Figure S5A) and a median partial pressure of P_aCO_2 of 71.5 mmHg (IQR 15.9) (Figure S5B), measured 2 to 4 h prior to the initiation of vvECMO, indicating severe ARDS. The elevated P_aCO_2, while the patients were ventilated with a supranormal minute volume (8.7 l/min; IQR 3.5), indicated a pathologically increased dead space. This is consistent with fibroproliferative tissue remodeling in ARDS (Hendrickson et al., 2015). Inspiratory vital capacity (VCin) decreased continuously in severe ARDS on vvECMO support, indicating a progressive, restrictive ventilatory defect (Figure 5A). To assess radiographic correlates of these restrictive defects, we compared the first available CT scan to images during severe ARDS on vvECMO treatment and to the last available CT scan. The majority of ARDS patients showed multilocular bilateral ground glass opacities (GGOs) and consolidations in the first available CT scan, typical for acute COVID-19 pneumonia (Figures 5B and S5C). Over the course of the disease, CT imaging revealed progressive consolidation and reticulation, indicative of fibroproliferative ARDS. Patients who died on vvECMO failed to resolve consolidations and fibrous stripes, whereas patients who could be successfully weaned from vvECMO and ultimately recovered showed a gradual resolution of fibrosis with residual reticulations in the last available CT (Figures 5B and S5C), which was also reflected in a normalized P_aCO_2 and P_aO_2/F_iO ratio (Figure S5B).

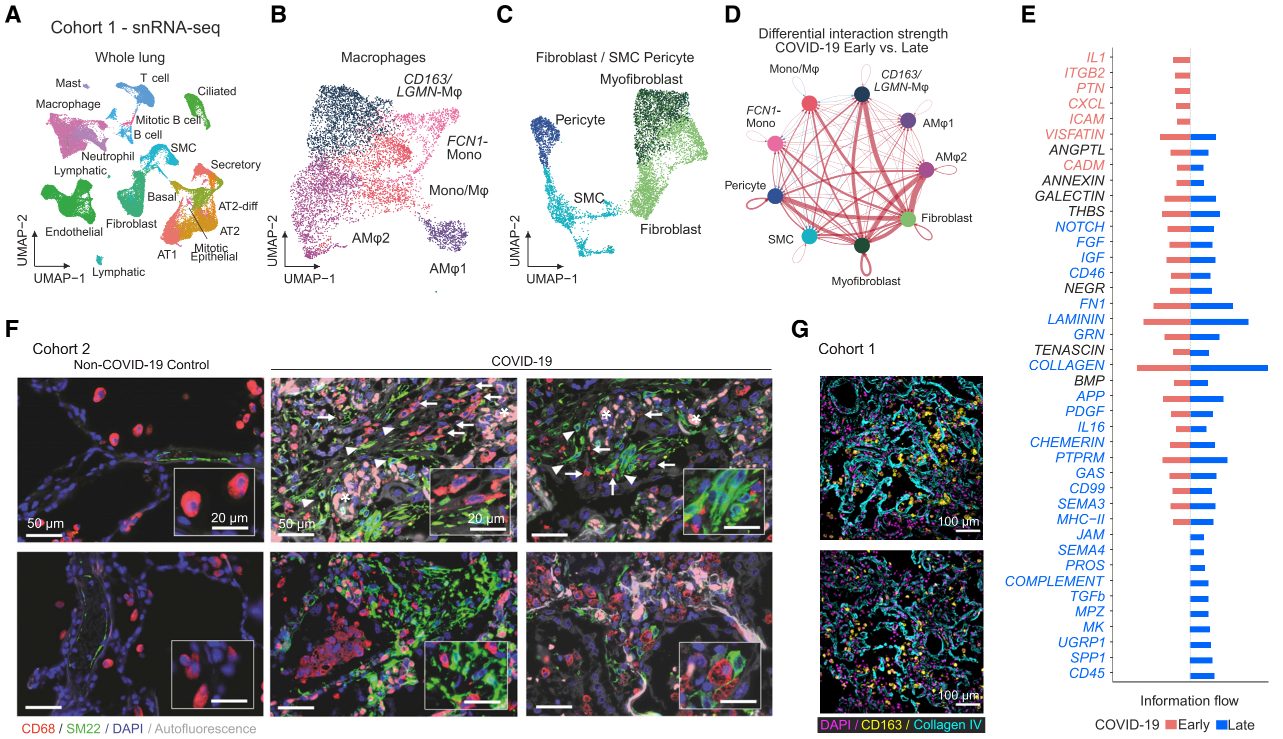


Figure 4. Macrophage-fibroblast interactions in COVID-19 lungs

(A) UMAP embedding of 48,656 snRNA-seq transcriptomes of lung tissue of six patients with fatal COVID-19 and three non-COVID-19 controls. Cell-type annotation based on expression of canonical marker genes.

(B) UMAP embedding of 7,504 macrophages identified in (A).

(C) UMAP embedding of 7,492 fibroblasts, smooth muscle cells (SMCs), and pericytes identified in (A).

(D) Circle plot showing cell-cell interaction strength between macrophage, fibroblast, SMC, and pericyte clusters predicted by CellChat. Each circle represents one cell type, edges between circles represent intracellular signaling between cell types, and edge thickness reflects interaction strength, while the colored edges show differential interaction strength, where red represents increased interaction strength in late (n = 3) versus early (n = 3) samples.

(E) Signaling pathways ranked by differential overall information flow of inferred interactions in early (red) and late (blue) samples.

(F) IF of lung tissue stained for macrophages (CD68, red) and myofibroblasts (SM22, green), nuclei (DAPI, blue), and autofluorescence visible as faint gray. Macrophages are indicated by arrows, expanded SM22 foci are indicated by arrowheads, and asterisks denote erythrocyte filled capillaries in alveolar septa (scale bar, 50 μm; insert scale bar, 20 μm).

(G) Two representative MELC FOVs showing CD163$^+$ macrophages (yellow), collagen (cyan), and nuclei (DAPI, magenta). Scale bar, 100 μm.

To assess the extent of fibrotic tissue remodelling, we analyzed 14 autopsy samples and one non-autopsy sample from 15 patients with COVID-19 ARDS (cohort 2) and seven non-COVID-19 samples from five autopsies and two non-autopsy cases (controls) (Figure S1A). Histopathological evaluation (H&E) revealed extensive diffuse alveolar damage and edema early on and increasing fibroproliferative tissue remodeling and fibrotic foci at later time points after the onset of ARDS (Figures 5C and S5D). Pulmonary fibrosis, scored by two independent experienced pathologists using a well-established semiquantitative fibrosis score (Ashcroft et al., 1988), was significantly increased in COVID-19 ARDS (Figure 5D). Pulmonary fibrosis is characterized by interstitial fibroblast proliferation and deposition of ECM proteins, particularly collagen. Here, we found extensive interstitial collagen type 1, 3, and 4 deposition in COVID-19-associated ARDS (Figures 5C, 5D, and S5D–S5F).

Transmission EM analysis of autopsy lung samples revealed thickened alveolar septa due to interstitial edema and an accumulation of connective tissue fibers, in particular, collagen fibrils and elastic fibers (Figure 5E). The alveolar epithelium was partly desquamated, resulting in a denuded alveolar epithelial basal lamina. Infoldings of the denuded basal laminae represent an ultrastructural characteristic of alveolar collapse and collapse induration, and these were frequently found in severe COVID-19 (Ochs et al., 2021). Collapse indurations were surrounded by deposits of collagen fibrils and elastic fibers. Interstitial cells within the thickened septa had a foamy appearance, containing membrane-bound vesicles of varying size. The vesicle contents appeared largely homogeneous, with low to moderate electron density. A distinction between interstitial macrophages and activated fibroblasts is difficult based solely on morphology, which is why we refer to these cells simply as interstitial cells. Alveolar macrophages also had a foamy appearance, but their vesicles also contained stacks of lipid lamellae, a typical feature in alveolar macrophages.

In conclusion, we revealed exacerbated fibroproliferative responses with clear ventilatory, radiographic, histological, and

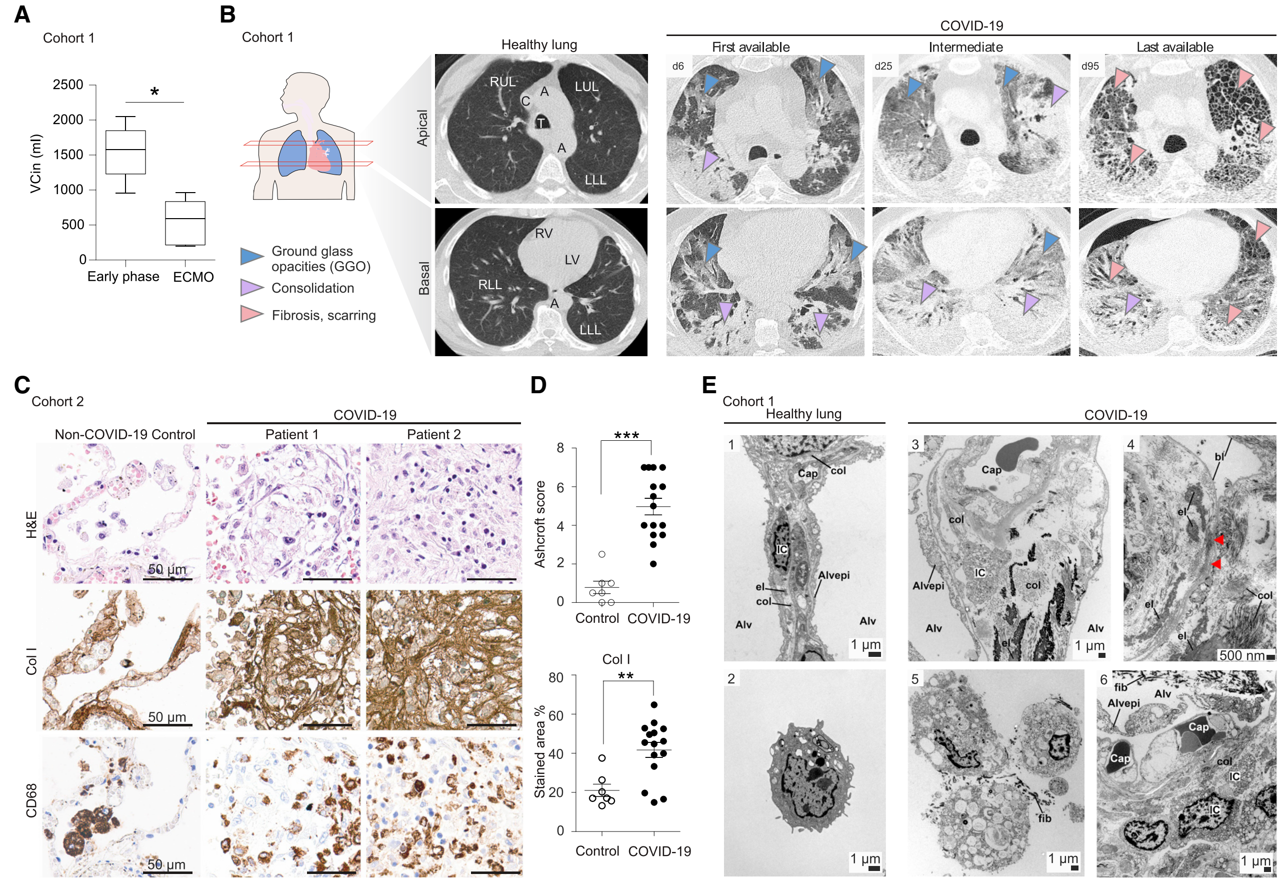

A
Cohort 1
2500
2000
1500
1000
500
0
VCin (ml)
*
Early phase
ECMO
B
Cohort 1
Healthy lung
Apical
Basal
RUL
LUL
C
A
T
LLL
RV
LV
RLL
A
LLL
Ground glass opacities (GGO)
Consolidation
Fibrosis, scarring
COVID-19
First available
Intermediate
Last available
d6
d25
d95
C
Cohort 2
Non-COVID-19 Control
COVID-19
Patient 1
Patient 2
H&E
Col I
CD68
50 µm
50 µm
50 µm
D
Ashcroft score
8
6
4
2
0
***
Control
COVID-19
Col I
Stained area %
80
60
40
20
0
**
Control
COVID-19
E
Cohort 1
Healthy lung
COVID-19
1
Cap
col
IC
el
col
Alvepi
Alv
Alv
1 µm
2
1 µm
3
Cap
col
Alvepi
IC
col
Alv
Alv
el
1 µm
4
bl
el
col
el
el
500 nm
5
fib
1 µm
6
fib
Alv
Alvepi
Cap
Cap
col
IC
IC
1 µm

Figure 5. Severe COVID-19 induces pronounced fibroproliferative ARDS

(A) Inspiratory vital capacity (VCin) in early phase, and acute vvECMO phase (paired t test $^{*}p < 0.05$; $^{**}p < 0.01$).

(B) (Left) Schematic representation indicating imaging planes of CT. (Middle) Healthy lung and denomination of anatomical structures. RLL, right lower lobe; RUL, right upper lobe; LUL, left upper lobe; LLL, left lower lobe; A, aorta; T, trachea; C, vena cava; RV, right ventricle; LV, left ventricle. (Right) Representative images from a case of severe COVID-19 ARDS, representing the first available (left column), one intermediate (middle column), and the last available (right column) CT scan.

(C) Histopathology of autopsy lung tissue of fatal COVID-19. High-power images of consecutive histological sections stained with H&E (top) and chromogenic IHC for collagen I (middle) and CD68 (bottom). Scale bar, 50 μm.

(D) Quantification of pulmonary fibrosis (Ashcroft score) and collagen-I-stained area. Dots represent individual autopsies (line at mean with SEM), and significance of population shift of COVID-19 compared to control assessed by Mann Whitney Test ($^{**}p < 0.01$; $^{***}p < 0.001$).

(E) Transmission EM of healthy (1–2) and COVID-19 (3–6) autopsy lungs. 1: Alveolar septum between two alveolar lumina (Alv) with capillary (Cap), interstitium, and alveolar epithelium (Alvepi). The interstitium with interstitial cells (ICs) and a connective tissue network of collagen fibrils (col) and elastic fibers (el). 2: Alveolar macrophage with lysosomal vesicles. 3: Alveolar septum containing Cap and interstitium. The alveolar epithelium is only partly present, leaving the alveolar epithelial basal lamina denuded toward the alveolar lumen at sites of detachment. The septum is thickened due to swelling of the interstitium, containing cells, collagen fibrils, elastic fibers, and homogeneous matrix. ICs contain high numbers of vesicles. 4: Infolding of denuded alveolar epithelial basal lamina (bl) with collapsed alveolar lumen and partly “glued” opposing basal lamina (red arrowheads), features of collapse induration. 5: Foamy alveolar macrophages containing vesicles of varying size and content. Fibrin accumulations (fib) in close proximity. 6: Thickened alveolar septum containing capillaries with swollen endothelium. The alveolar epithelium is desquamated toward the alveolar lumen containing fibrin. Note vesicle-filled ICs with foamy appearance.

ultrastructural features of pulmonary scarring and fibrosis during severe COVID-19-associated ARDS.

SARS-CoV-2 triggers a fibrosis-associated transcriptional profile in monocytes

Key mediators of wound healing and fibrosis, including TGF-β, are induced in phagocytes upon tissue damage and uptake of apoptotic cells (Huynh et al., 2002). Profibrotic damage response signatures in macrophages may therefore result from extensive lung injury during ARDS and ventilator-induced lung injury. As SARS-CoV-2 transcripts were found in macrophages (Figures 1E and S2C), and particularly in *CD163/LGMN*-Mϕ (Figure 2C), we tested whether viral contact might directly elicit fibrosis-associated signatures in monocytes. We stimulated classical ($CD14^{+}CD16^{-}$) monocytes isolated from healthy donors with SARS-CoV-2 for 18 h and analyzed the transcriptional responses by scRNA-seq with multiplexing of experimental conditions (by hashtag oligos [HTOs]) and donors (by single nucleotide

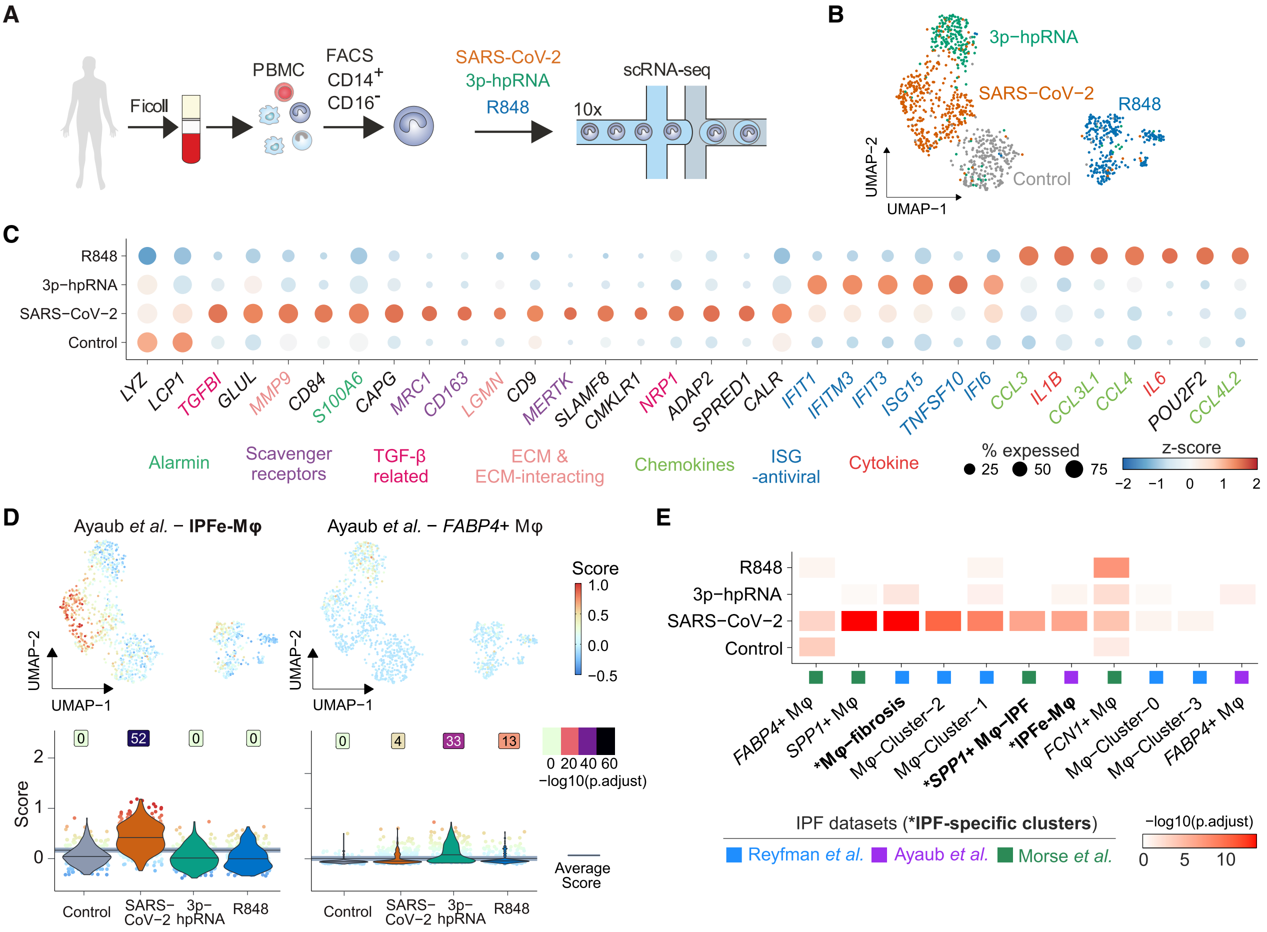


Figure 6. SARS-CoV-2 induces profibrotic programs in classical monocytes *in vitro*

(A) Schematic depiction of the experimental layout.

(B) UMAP embedding of 1,123 quality-filtered transcriptomes of human monocytes stimulated as outlined in (A).

(C) Dot plot displaying differentially expressed (DE) genes in the indicated stimulation conditions. Label color indicates gene categories. Adjusted p values are available in Table S5.

(D) Signature module score of IPF-expanded macrophages (IPFe-Mϕ) and alveolar *FABP4*+Mϕ (Ayaub et al., 2021) projected onto the UMAP embedding (top) and plotted as violin plots (bottom) across the clusters of stimulated monocytes. Numbers above violins show -log10 transformed adjusted p values (one-sided Wilcoxon test compared to average). Lines indicate median scores per cluster.

(E) Heatmap displaying -log10 transformed adjusted p values (one-sided Fisher's Exact Test) comparing overlap between gene sets from stimulated monocytes with published transcriptional signatures of IPF-associated monocytes/macrophages. Cluster names and reference studies are indicated on the x axis; Table S3.

polymorphism [SNP]) to minimize batch effects (Figures 6A, S6A, and S6B). We included ligands of viral RNA sensors, retinoic acid–inducible gene I (RIG-I), and melanoma differentiation-associated protein 5 (MDA5) agonist 3′-ppp-hairpin-RNA (3p-hpRNA) and R848, a dual agonist of Toll-like receptor (TLR)-7/-8, for comparison. We observed distinct responses to 3p-hpRNA, R848, and SARS-CoV-2 (Figure 6B; Table S5). The monocyte response to SARS-CoV-2 was distinct, however, partially overlapping with 3p-hpRNA stimulation, indicating the potential of SARS-CoV-2 to elicit modest type-I interferon (IFN) responses via RIG-I-like receptors (Figures 6C, S6B, and S6C). Notably, several genes characteristic of pulmonary macrophages identified both in COVID-19 and in lung fibrosis, including *MRC1*, *CD163*, *MERTK*, *LGMN*, and *MMP9*, were specifically induced upon exposure to SARS-CoV-2 (Figures 6C and S6D; Table S5). SARS-CoV-2 also induced TGF-β family genes *NRP1* and *TGFBI* (Figures 6C and S6D). NRP1 promotes host cell binding and entry of SARS-CoV-2 (Cantuti-Castelvetri et al., 2020; Daly et al., 2020); yet, despite abundant viral transcripts in SARS-CoV-2-stimulated monocytes, we found no evidence of productive infection (unpublished data; Figure S6A). In contrast to R848, SARS-CoV-2 stimulation induced only limited expression of *IL1B* and no *IL6* (Figures 6C and S6D). Transcription factor predictions revealed an overlap of SARS-CoV-2-stimulated monocytes with pulmonary *CD163/LGMN*-Mϕ macrophages in COVID-19 ARDS, including a predicted involvement of TFEC, GLMP, and HLX (H2.0-like homeobox protein) (Figure S6E; Table S5).

Given the overlap of gene expression in SARS-CoV-2-exposed monocytes and *CD163/LGMN*-Mϕ, we assessed the potential enrichment of IPF-associated macrophage gene sets. Fibrosis-associated macrophage gene signatures were specifically enriched in SARS-CoV-2-exposed monocytes, but not in 3′-hpRNA-stimulated, R848-stimulated, or unstimulated control monocytes (Figure 6D and S6F). We found a high similarity of gene expression in SARS-CoV-2-stimulated monocytes and IPF-associated macrophages, including IPF-specific clusters (Figure 6E). These results indicate that SARS-CoV-2 directly triggers a transcriptional profile in human monocytes that resembles fibrosis-associated pulmonary macrophages.

SARS-CoV-2 triggers a profibrotic proteome profile in monocytes

To validate and specify these findings, we performed multiplexed quantitative shotgun proteomics, which can provide data with high relevance for cellular phenotypes (Buccitelli and Selbach, 2020). Monocytes isolated from healthy donors were stimulated with SARS-CoV-2 or IAV (H3N2), which has been well-studied at the proteome level (Bogdanow et al., 2019; Sadewasser et al., 2017) (Figure 7A). Analysis at 1, 3, and 18 h post infection (hpi) reproducibly quantified 6,951 proteins and 5,299 phosphorylation sites in 2 replicates from 4 donors (Figure S7A). Consistent with the ability of IAV to productively infect monocytes (Cline et al., 2017), we found an increase of IAV proteins over time (Figure 7B). In contrast, SARS-CoV-2 proteins remained constant or decreased (Figure 7B). Only SARS-CoV-2 M protein (VME1) showed a moderate increase, suggesting some residual viral transcription and translation (Figures 7B and S7B).

SARS-CoV-2 and IAV induced distinct changes in the host proteome, particularly at 18 hpi (Figure 7C; Table S6). To identify SARS-CoV-2-specific responses, we performed gene set enrichment analysis (GSEA) on SARS-CoV-2 over IAV protein ratios (Subramanian et al., 2005). SARS-CoV-2-specific gene sets were related to innate immunity, antiviral defense, and RIG-I-like receptor signaling, while IAV induced viral gene expression- and replication-associated genes (Figure 7D; Table S6). Importantly, genes related to wound healing and fibrosis were upregulated in response to SARS-CoV-2 (Figure 7D).

For visualization, we mapped the proteomic data to antiviral and profibrotic pathways, revealing marked differences in the responses to IAV and SARS-CoV-2 (Figures 7E and S7C). SARS-CoV-2 induced upregulation of RIG-I-like receptors and double-stranded RNA (dsRNA) sensors, as well as scavenger receptors and related membrane proteins. Additionally, we observed an upregulation of transcription factors related to inflammation (NFkB, IRF7) and macrophage differentiation (MAFB, CEBPB) (Figures 7E and S7C). MAFB was highly expressed in *CD163/LGMN*-Mϕ (Figure 2D), and it has been identified as a specific marker of macrophages in pulmonary fibrosis (Aran et al., 2019). CEBPB activation was predicted in SARS-CoV-2-stimulated monocytes (Figure S6E), and it has been shown to license differentiation of profibrotic macrophages (Satoh et al., 2017). Additionally, we observed upregulation of the transcriptionally active longer isoforms of the liver-enriched activator protein (LAP*/LAP) and downregulation of the shorter inhibitory of ther liver-enriched inhibitory protein (LIP) isoform of CEBPB (Figure S7D). This isoform switch occurs via alternative translation initiation from the same mRNA and is only detectable at the protein level (Descombes and Schibler, 1991). An increased LAP:LIP ratio mediates macrophage differentiation (Calkhoven et al., 2000; Huber et al., 2012). Translation of the LAP*/LAP isoform is induced upon activation of PKR, consistent with its induction by SARS-CoV-2 (Figure 7E). We further identified increased phosphorylation of specific sites on CEBPB and IRF7, suggesting an enhanced transcriptional activity (Figures 7E, S7E, and S7F).

Consistent with *in vivo* transcriptomic data (Table S2), we detected the induction of myeloid-cell-attracting and profibrotic chemokines CCL2, CCL8, CCL24, and CXCL8 (Figures 7E and S7G). Several secretory proteins involved in tissue remodeling and fibrosis were upregulated, including proteases (LGMN, MMP9, MMP14, CTSL), protease inhibitors (TIMP1), phospholipase PLA2G7, transglutaminase TGM2, and TGF-β downstream target TGFBI (Figure 7E). Similar to the transcriptomic data, IPF-specific macrophage signatures "IPFe-Mϕ" (Ayaub et al., 2021), "*SPP1*+Mϕ-IPF" (Morse et al., 2019), and "Mϕ-fibrosis" (Reyfman et al., 2019) were highly enriched in monocytes stimulated with SARS-CoV-2 for 18 h, but not in IAV-infected cells (Figure 7F and S7H). Thus, detection of SARS-CoV-2, but not IAV, by human classical monocytes constitutes a trigger of fibrosis-associated differentiation programs.

DISCUSSION

Pathomechanisms of severe COVID-19-associated ARDS remain incompletely understood (Fan et al., 2020). Here, we report the accumulation of monocyte-derived macrophages with an enrichment of fibrosis-associated gene signatures and significant similarity to macrophage populations found in IPF in the lung during severe COVID-19 (Adams et al., 2020; Ayaub et al., 2021; Morse et al., 2019; Reyfman et al., 2019). Notably, SARS-CoV-2 was sufficient to induce a similar differentiation program in classical monocytes *in vitro*, indicating that viral contact may constitute a trigger for profibrotic macrophage reprogramming. These findings were corroborated by quantitative proteomics and extended to posttranslational alterations including the CEBPB isoform ratio, which has also been associated with fibrosis. In line with these findings, we observed restrictive ventilatory defects and radiographic signs of consolidation and fibrotic remodelling, and histopathology revealed myofibroblast and fibroblast expansion and pronounced ECM deposition. These findings aligned with the clinical observation that patients with COVID-19 ARDS require protracted respiratory support and ECMO therapy and show increased mortality rates compared to other forms of ARDS. Our study describes a predominantly profibrotic profile of pulmonary macrophages in severe COVID-19, accompanied by profound fibrotic lung tissue remodeling. This is in line with previous reports of organizing pneumonia, scarring, and fibrosis in patients with COVID-19 ARDS and even in individuals with initially mild or moderate disease (Bharat et al., 2020; Combet et al., 2020; Pan et al., 2020; Schwensen et al., 2020; Spagnolo et al., 2020).

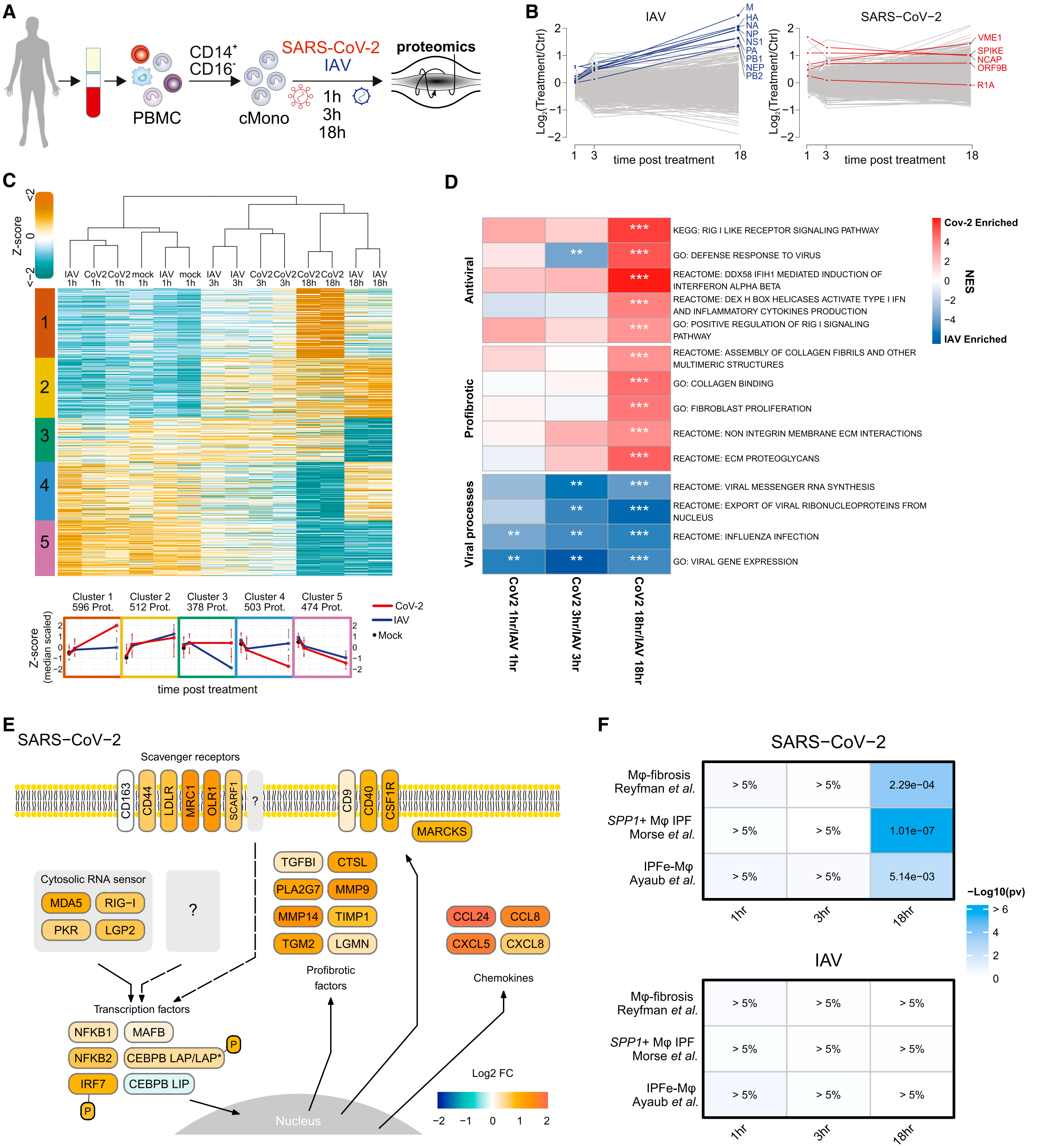


Figure 7. Proteomic analyses of SARS-CoV-2-induced profibrotic phenotype in classical monocytes

(A) Schematic depiction of the experimental layout.

(B) Protein log-2-fold-changes over time for IAV (left, blue) and SARS-CoV-2 (right, red) and host proteins (gray).

(C) Heatmap of DE host proteins (ANOVA test, filtered by Benjamini-Hochberg adjusted p value < 5%). Protein clusters obtained by fuzzy-c-means clustering of *Z*-scored protein intensities are indicated in the figure, and corresponding profiles are reported below the heatmap.

(D) GSEA of protein intensity ratios of SARS-CoV-2 over IAV infection, calculated for the host proteome dataset. *FDR < 10%; **FDR < 5%; ***FDR < 1%.

(E) Schematic presentation of selected proteins regulated by SARS-CoV-2 stimulation in monocytes, color-coded by log2-fold changes (infection versus control, 18 hpi).

(F) Heatmap representation of p-values (one sided Wilcoxon signed-rank test) for the enrichment of the indicated reference gene sets calculated by eCDF.

While IPF is characterized by chronically progressive, irreversible fibrosis, COVID-19-induced fibrotic tissue remodeling occurs rapidly and is at least partially reversible in the majority of survivors. Yet, both conditions share similar epidemiological risk factors, including older age, male sex, history of cigarette smoking, and other comorbidities. Viral infections have also been repeatedly implicated in the pathogenesis of IPF and as triggers of acute exacerbations (Meneghin and Hogaboam, 2007; Molyneaux and Maher, 2013; Naik and Moore, 2010; Sheng et al., 2020; Wootton et al., 2011). Moreover, acute exacerbations of IPF are characterized by bilateral GGOs and evidence of diffuse alveolar damage, which is suggestive of viral infection and subsequently enhanced fibrosis (Collard et al., 2007, 2016; Wootton et al., 2011). Our data reveal unexpected, common features of IPF and severe COVID-19 ARDS, namely, aberrant macrophage activation and inappropriate fibroproliferative responses in susceptible individuals (Morse et al., 2019; Schupp et al., 2015; Thompson et al., 2017). Monocytes and profibrotic macrophages play a critical role in the pathogenesis and progression of both ARDS and lung fibrosis (Gibbons et al., 2011; Herold et al., 2015; Misharin et al., 2017; Nouno et al., 2019; Scott et al., 2019; Thompson et al., 2017). Circulating monocyte numbers have been proposed as prognostic biomarkers in IPF, and genetic deletion of CCR2 or depletion of monocyte-derived macrophages protects mice from drug-induced lung fibrosis (Misharin et al., 2017; Moore et al., 2001; Brody et al., 2021; Teoh et al., 2020). The accumulation of CD163$^+$ macrophages has been associated with poor outcomes in IPF (Brody et al., 2021; Nouno et al., 2019). *CD163/LGMN*-Mϕ expressed high levels of Osteopontin (encoded by *SPP1*), a secreted ECM phosphoglycoprotein (Liaw et al., 1995), which also doubles as a cytokine that stimulates collagen-I production in fibroblasts and exerts profibrotic functions in IPF (Pardo et al., 2005; Urtasun et al., 2012). *TGFB1* (encoding TGF-β), a master regulator of wound healing and organ fibrosis (Frangogiannis, 2020; Morse et al., 2019), and several TGF-β-related genes, including *TGFBI* and *NRP1*, were also highly expressed in *CD163/LGMN-M*ϕ. Macrophage-derived TGFBI promotes collagen production in fibroblasts and inhibits collagen turnover by suppressing antifibrotic metalloproteinase 14 (MMP14) (Frangogiannis, 2020; Morse et al., 2019; Nacu et al., 2008). Legumain (*LGMN*) is an asparaginyl endopeptidase involved in MMP activation, TGF-β signaling, and ECM deposition and plays a prominent role in organ fibrosis (Bai et al., 2019; Ren et al., 2020). *LGMN* is highly expressed in *CD163/LGMN*-Mϕ and in profibrotic macrophages in IPF (Ayaub et al., 2021; Morse et al., 2019), and it was strongly induced in monocytes upon exposure to SARS-CoV-2.

We demonstrated that exposure to SARS-CoV-2, but not IAV, induces transcriptome and proteome profiles with high similarity to those of IPF-associated macrophages, and we also detected SARS-CoV-2 transcripts in pulmonary macrophages. These findings suggest that SARS-CoV-2 may directly contribute to the profibrotic macrophage phenotype in severe COVID-19. Other factors likely propagate fibrogenic macrophage responses at later stages, but fibrotic tissue states can be triggered by strong initial impulses into pathological stability and typically resolve slowly (Adler et al., 2020). Recognition of tissue damage and uptake of apoptotic cells are known stimuli of wound healing and profibrotic responses, through the induction of TGF-β (Huynh et al., 2002). Pulmonary macrophages in COVID-19 expressed a range of scavenger receptors and proteins involved in efferocytosis, including *MRC1*, *CD163*, *TREM2*, and *MERTK*, among others, and high levels of *TGFB1* and *TGFBI*. Hence, SARS-CoV-2 infection induces expression of genes that may directly and indirectly promote profibrotic functions of macrophages. These damage repair responses may be beneficial to control inflammatory tissue damage. Yet, unchecked and aberrant, they may cause enhanced fibroproliferation and protracted respiratory failure in susceptible individuals. In this regard, it will be important to assess the presence of *CD163/LGMN*-Mϕ and their fate in patients with milder forms of COVID-19.

Pulmonary sequelae of COVID-19, including fibrosis, have been previously reported, but the exact disease burden remains unclear. A recent 6-month follow up of 1,733 patients previously hospitalized for COVID-19 revealed a reduced median 6-min walking distance and decreased diffusion capacity and total lung capacity, indicative of restrictive ventilatory defects. This was more pronounced following severe disease, excluding patients on ECMO or organ support (Huang et al., 2021). In addition, there are several reports of severe, progressive lung fibrosis following mild to moderate SARS-CoV-2 infection (Arjun et al., 2020; Combet et al., 2020; Schwensen et al., 2020). It will be important to identify patients at risk of developing fibrotic complications of COVID-19 and to devise early intervention strategies, including potentially antifibrotic therapies (George et al., 2020). Blockade of monocyte influx may also present an attractive strategy. On the other hand, it is clear that fibrotic lesions resolve, or partially resolve, over time in patients who survive COVID-19 ARDS. The recovery phase may therefore provide a unique window to investigate molecular mechanisms of fibrosis resolution.

Limitations of the study

We combined *in vivo* and *in vitro* analyses of transcriptional and proteomic profiles in COVID-19-associated macrophages and compared these to macrophage phenotypes in IPF. Causal links between the accumulation of specific macrophage populations and fibrotic tissue remodeling is difficult to establish in an observational study. The snRNA-seq analysis and cell communication inference indicated profibrotic functions of COVID-19-associated macrophages. Additionally, recent modeling of macrophage-fibroblast cell circuits during fibrosis predicted excessive macrophage tissue influx and a profibrotic macrophage setpoint to cause increased fibroproliferation as well as pathological fibrosis (Adler et al., 2020). We demonstrate direct effects of SARS-CoV-2 on profibrotic macrophage programs *in vitro*, but the relevance of virus-macrophage interactions during COVID-19 ARDS *in vivo* is challenging to determine in human studies. The molecular mechanisms underlying the induction of profibrotic genes by SARS-CoV-2 remain to be investigated in greater detail.

In conclusion, we describe a profound fibroproliferative tissue response in severe COVID-19 ARDS, associated with an accumulation of monocyte-derived macrophages with significant transcriptional similarities to profibrotic macrophages in IPF. We propose that SARS-CoV-2 promotes fibrosis-associated genetic programs in macrophages, which are further sustained and enhanced by responses to extensive tissue damage. It will be important to dissect the molecular mechanisms linking viral recognition to profibrotic macrophage responses in ARDS and in chronic organ fibrosis, as these may provide new targets for therapeutic intervention.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
    - Lead contact
    - Materials availability
    - DeCOI consortium members
    - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
    - Cohort 1 - Berlin cohort
    - Cohort 2 - Aachen cohort
    - Additional datasets used to perform data integration and snRNA-seq
- METHOD DETAILS
    - Clinical investigation
    - Viral Stocks
    - Histology, Immunohistochemistry, Immunofluorescence and SARS-CoV-2 RNA in situ hybridization
    - Bronchoalveolar lavage (BAL)
    - Monocytes isolation and infection
    - Multi-epitope-ligand cartography (MELC)
    - Transmission electron microscopy (TEM)
    - Proteomics
    - ELISA
- QUANTIFICATION AND STATISTICAL ANALYSIS
    - MELC data analysis
    - Single-cell RNA-seq data analysis
    - Single-nucleus RNA-seq
    - Data integration
    - Proteomics data analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.11.033.

ACKNOWLEDGMENTS

The members of the DeCOI network are listed in the supplemental information. The authors are grateful to all patients and their relatives for consenting to biosampling and data collection. COVID-19 research at Charité - Universitätsmedizin Berlin was facilitated by the PA-COVID-19 study group (Kurth et al., 2020). Autopsy studies were facilitated by the Biobank of the Department of Neuropathology at Charité and by the Institute of Pathology, RWTH Aachen University Hospital, Aachen, Germany. We thank Achim Leutz (MDC) for valuable comments and Philipp Mertins and the MDC/BIH proteomics core unit team for supporting proteomic sample preparation.

The authors are indebted to the Clinical Study Center (CSC) at the Berlin Institute of Health (BIH) and Charité, particularly Christof von Kalle, Alexander Krannich, Uwe Behrens, Sein Schmidt, Chantip Dang-Heine, Saskia Zvorc, and Maria Rönnefahrt; and to the Central Biobank of the BIH (ZeBanC), particularly Michael Hummel, Denise Treue, and Dana Briesemeister, for ongoing support of the PA-COVID-19 Study. We thank Petra Schrade for expert technical assistance.

This work was supported by the German Ministry of Education and Research (BMBF) and the German Network University Medicine NUM FKZ 01KX2021 (COVIM to L.E.S., F.K., B.S., N.S., J.L.S., and S. Herold; DEFEAT PANDEMIcs to P.B. and F.L.H.; OrganoStrat to A.-E.S., M.L., E.W., S.H., A.H., F.L.H. and C.G.; NAPKON to M.W.); the BMBF CAPSyS- (01ZX1304B to M.W.); CAPSyS-COVID- (01ZX1604B to M.W.); SYMPATH- (01ZX1906A to M.W.); PROVID- (01KI20160A to M.W., C.D., L.E.S., S. Hippenstiel); RECAST- (01 KI20337 to B.S. and L.E.S.); and RAPID-consortia (01KI2006F to T.W.); the national research node MSTARS (031L0220B to M.S.); and the Comprehensive Heart Failure Center Würzburg (BMBF 01EO1504 to C.C. and A.-E. S.); BMBF (01DG14009 to M.O.); BMBF (01GM1901A to P.B.); BMBF (CompLS HOPARL 031L0289B to F.J.T. and A.-E.S.) and Deutsche Forschungsgemeinschaft (DFG): GRK 2157 (to O.D. and A.-E.S.); SFB-TR84 114933180 (B2 to T.W. and S. Herold, C6+C9 to M.W., C8+C10 to L.E.S., B6 to S.H. and A.H., and B9 to S. Herold); SPP1937 (HA5354/8-2 HA5354/10-1 to A.E.H.); SFB1021 (C5 to S. Herold); KFO309 (P2/P8 to S. Herold); TRR130 (TP17 and C01 to A.E.H .and H.R.); SFB-TRR 167 and HE 3130/6-1 to F.L.H.; SFB1449 (B2 to M.W.); Germany's Excellence Strategy (EXC2151 – 390873048 to J.L.S., INST 37/1049-1, INST 216/981-1, INST 257/605-1, INST 269/768-1, INST 217/988-1, and INST 217/577-1 to J.L.S.; EXC - 2049-390688087 to F.L.H., EXC2026 – 390649896 to S. Herold); SFB/TRR219 (Project-ID 322900939) and Project-ID 454024652 to P.B.; DJ100/1-1 to S.D. (Project-ID 432698239); Project-ID 445703531 to S.D. and P.B.; the EU Horizon 2020 grant 874656-DiscovAIR to J.L.S.; European Research Council (ERC CoG-101001791 to P.B.); the Berlin University Alliance (BUA: 501_Elektronenmikroskopie to M.O.); the German Registry of COVID-19 Autopsies (DeRegCOVID); the Federal Ministry of Health (BMG) (ZMVI1-2520COR201 to P.B.); and the Medical Faculty of RWTH Aachen University (START 125/17 to S.v.S.). The clinical study is supported by the Berlin Institute of Health (BIH) PA-COVID-19 study. This work was also supported by the Department of Genomics & Immunoregulation at LIMES Institute (to A.C.A.), IZKF at University Hospital Würzburg (IZKF-E-353 to C.C.), and the government of Bavaria (Bayerische Staatsregierung, FOR-COVID to A.E.S.). A.L.H. and D.W. were supported by the Jürgen Manchot Foundation, and M.M. is supported by the BIH-Charité Digital Clinician Scientist Program.

AUTHOR CONTRIBUTIONS

Conceptualization, D.W., O.D., T.M., S.v.S., M.D.L., P.B., M.S., A.-E.S., and L.E.S.; methodology, D.W., O.D., T.M., S.v.S., I.L.I., S.B., T.K., A.P.R., R.G., A.E.H., and M.D.L.; software/data analysis, D.W., O.D., T.M., S.v.S., I.L.I., R.L.C., H.Z., A.L.H, A.P.R., R.M., R.D.B., A.M.L., F.E., A.E.H., M.D.L., F.J.T., C.C., M.S., A.-E.S., and L.E.S.; resources, S.v.S., M.M, C.M., A.P.R., R.M., R.D.B., J.S., S.D., F.P., J.K., J.R., P.P., K.B., C. Conrad, A.C.A., B.S., M.L., E.W., D.H., S. Hippenstiel, A.H., F.L.H., A.U., C.G., F.M., S. Herold, S.E., C.T., M.W., C. Cochain, N.S., C.D., C.G., F.K., J.L.S., H.R., R.E., H.M.R., T.W., P.B., S.T., and M.O.; supervision, P.B., M.S., A.-E.S., and L.E.S.; writing – original draft, M.S., A.-E.S., and L.E.S.; writing – review & editing, D.W., O.D., T.M., S.v.S., A.C.A., J.L.S., M.S., A.-E.S., and L.E.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 30, 2020
Revised: July 28, 2021
Accepted: November 23, 2021
Published: November 27, 2021

REFERENCES

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. (2020). Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. Sci. Adv. *6*, eaba1983.

Adler, D., Murdoch, D., Nenadic, O., Urbanek, S., Chen, M., Gebhardt, A., and Senger, A. (2019). rgl: 3D Visualization Using OpenGL. R package version 0.100.19.

Adler, M., Mayo, A., Zhou, X., Franklin, R.A., Meizlish, M.L., Medzhitov, R., Kallenberger, S.M., and Alon, U. (2020). Principles of Cell Circuits for Tissue Repair and Fibrosis. iScience *23*, 100841.

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.

Arjun, S., Patel, D., Sanivarapu, R., Iqbal, J., and Anjum, F. (2020). Case report of severe pulmonary fibrosis as a sequelae of COVID-19 infection. Chest *158*, A433–A434.

Arredouani, M.S., Palecanda, A., Koziel, H., Huang, Y.-C., Imrich, A., Sulahian, T.H., Ning, Y.Y., Yang, Z., Pikkarainen, T., Sankala, M., et al. (2005). MARCO is the major binding receptor for unopsonized particles and bacteria on human alveolar macrophages. J. Immunol. *175*, 6058–6064.

Ashcroft, T., Simpson, J.M., and Timbrell, V. (1988). Simple method of estimating severity of pulmonary fibrosis on a numerical scale. J. Clin. Pathol. *41*, 467–470.

Ayaub, E.A., Poli, S., Ng, J., Adams, T., Schupp, J., Quesada-Arias, L., Poli, F., Cosme, C., Robertson, M., Martinez-Manzano, J., et al. (2021). Single cell RNA-seq and mass cytometry reveals a novel and a targetable population of macrophages in idiopathic pulmonary fibrosis. bioRxiv. https://doi.org/10.1101/2021.01.04.425268.

Bai, P., Lyu, L., Yu, T., Zuo, C., Fu, J., He, Y., Wan, Q., Wan, N., Jia, D., and Lyu, A. (2019). Macrophage-Derived Legumain Promotes Pulmonary Hypertension by Activating the MMP (Matrix Metalloproteinase)-2/TGF (Transforming Growth Factor)-β1 Signaling. Arterioscler. Thromb. Vasc. Biol. *39*, e130–e145.

Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: Open source software for digital pathology image analysis. Sci. Rep. *7*, 16878.

Barbaro, R.P., MacLaren, G., Boonstra, P.S., Iwashyna, T.J., Slutsky, A.S., Fan, E., Bartlett, R.H., Tonna, J.E., Hyslop, R., Fanning, J.J., et al.; Extracorporeal Life Support Organization (2020). Extracorporeal membrane oxygenation support in COVID-19: an international cohort study of the Extracorporeal Life Support Organization registry. Lancet *396*, 1071–1078.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. B *57*, 289–300.

Berg, S., Kutra, D., Kroeger, T., Straehle, C.N., Kausler, B.X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., et al. (2019). ilastik: interactive machine learning for (bio)image analysis. Nat. Methods *16*, 1226–1232.

Bharat, A., Querrey, M., Markov, N.S., Kim, S., Kurihara, C., Garza-Castillon, R., Manerikar, A., Shilatifard, A., Tomic, R., Politanska, Y., et al. (2020). Lung transplantation for patients with severe COVID-19. Sci. Transl. Med. *12*, eabe4282.

Blanco-Melo, D., Nilsson-Payant, B.E., Liu, W.-C., Uhl, S., Hoagland, D., Møller, R., Jordan, T.X., Oishi, K., Panis, M., Sachs, D., et al. (2020). Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. Cell *181*, 1036–1045.e9.

Bogdanow, B., Wang, X., Eichelbaum, K., Sadewasser, A., Husic, I., Paki, K., Budt, M., Hergeselle, M., Vetter, B., Hou, J., et al. (2019). The dynamic proteome of influenza A virus infection identifies M segment splicing as a host range determinant. Nat. Commun. *10*, 5518.

Brody, S.L., Gunsten, S.P., Luehmann, H.P., Sultan, D.H., Hoelscher, M., Heo, G.S., Pan, J., Koenitzer, J.R., Lee, E.C., Huang, T., et al. (2021). Chemokine Receptor 2-targeted Molecular Imaging in Pulmonary Fibrosis. A Clinical Trial. Am. J. Respir. Crit. Care Med. *203*, 78–89.

Brunson, J. (2020). ggalluvial: Layered Grammar for Alluvial Plots. J. Open Source Softw. *5* (49), 2017. https://doi.org/10.21105/joss.02017.

Buccitelli, C., and Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene expression control. Nat. Rev. Genet. *21*, 630–644.

Calfee, C.S., Delucchi, K., Parsons, P.E., Thompson, B.T., Ware, L.B., and Matthay, M.A.; NHLBI ARDS Network (2014). Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. Lancet Respir. Med. *2*, 611–620.

Calkhoven, C.F., Müller, C., and Leutz, A. (2000). Translational control of C/EBPalpha and C/EBPbeta isoform expression. Genes Dev. *14*, 1920–1932.

Campitelli, Elio (2021). ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'. R package version 0.4.5.. https://CRAN.R-project.org/package=ggnewscale.

Cantuti-Castelvetri, L., Ojha, R., Pedro, L.D., Djannatian, M., Franz, J., Kuivanen, S., van der Meer, F., Kallio, K., Kaya, T., Anastasina, M., et al. (2020). Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. Science *370*, 856–860.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. *7*, R100.

Carsana, L., Sonzogni, A., Nasr, A., Rossi, R.S., Pellegrinelli, A., Zerbi, P., Rech, R., Colombo, R., Antinori, S., Corbellino, M., et al. (2020). Pulmonary post-mortem findings in a series of COVID-19 cases from northern Italy: a two-centre descriptive study. Lancet Infect. Dis. *20*, 1135–1140.

Chua, R.L., Lukassen, S., Trump, S., Hennig, B.P., Wendisch, D., Pott, F., Debnath, O., Thürmann, L., Kurth, F., Völker, M.T., et al. (2020). COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. Nat. Biotechnol. *38*, 970–979.

Cline, T.D., Beck, D., and Bianchini, E. (2017). Influenza virus replication in macrophages: balancing protection and pathogenesis. J. Gen. Virol. *98*, 2401–2412.

Collard, H.R., Moore, B.B., Flaherty, K.R., Brown, K.K., Kaner, R.J., King, T.E., Jr., Lasky, J.A., Loyd, J.E., Noth, I., Olman, M.A., et al.; Idiopathic Pulmonary Fibrosis Clinical Research Network Investigators (2007). Acute exacerbations of idiopathic pulmonary fibrosis. Am. J. Respir. Crit. Care Med. *176*, 636–643.

Collard, H.R., Ryerson, C.J., Corte, T.J., Jenkins, G., Kondoh, Y., Lederer, D.J., Lee, J.S., Maher, T.M., Wells, A.U., Antoniou, K.M., et al. (2016). Acute Exacerbation of Idiopathic Pulmonary Fibrosis. An International Working Group Report. Am. J. Respir. Crit. Care Med. *194*, 265–275.

Combet, M., Pavot, A., Savale, L., Humbert, M., and Monnet, X. (2020). Rapid onset honeycombing fibrosis in spontaneously breathing patient with COVID-19. Eur. Respir. J. *56*, 2001808.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

D'Alessio, F.R., and Heller, N.M. (2020). COVID-19 and myeloid cells: complex interplay correlates with lung severity. J. Clin. Invest. *130*, 6214–6217.

Daly, J.L., Simonetti, B., Klein, K., Chen, K.-E., Williamson, M.K., Antón-Plágaro, C., Shoemark, D.K., Simón-Gracia, L., Bauer, M., Hollandi, R., et al. (2020). Neuropilin-1 is a host factor for SARS-CoV-2 infection. Science *370*, 861–865.

Descombes, P., and Schibler, U. (1991). A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA. Cell *67*, 569–579.

Dowle, M., and Srinivasan, A. (2019). data.table: Extension of 'data.frame'. R package version 1.13.0..

Fan, E., Beitler, J.R., Brochard, L., Calfee, C.S., Ferguson, N.D., Slutsky, A.S., and Brodie, D. (2020). COVID-19-associated acute respiratory distress syndrome: is a different approach to management warranted? Lancet Respir. Med. *8*, 816–821.

Ferguson, N.D., Fan, E., Camporota, L., Antonelli, M., Anzueto, A., Beale, R., Brochard, L., Brower, R., Esteban, A., Gattinoni, L., et al. (2012). The Berlin definition of ARDS: an expanded rationale, justification, and supplementary material. Intensive Care Med. *38*, 1573–1582.

Frangogiannis, N. (2020). Transforming growth factor-β in tissue fibrosis. J. Exp. Med. *217*, e20190103.

Garnier, S. (2018). viridis: Default Color Maps from "matplotlib". R package version 0.5.1.

Gassen, N.C., Papies, J., Bajaj, T., Emanuel, J., Dethloff, F., Chua, R.L., Trimpert, J., Heinemann, N., Niemeyer, C., Weege, F., et al. (2021). SARS-CoV-2-mediated dysregulation of metabolism and autophagy uncovers host-targeting antivirals. Nat. Commun. *12*, 3818.

Gayoso, A., Lopez, R., Xing, G., Boyeau, P., Wu, K., Jayasuriya, M., Mehlman, E., Langevin, M., Liu, Y., Samaran, J., et al. (2021). Scvi-tools: A library for deep probabilistic analysis of single-cell omics data. bioRxiv. https://doi.org/10.1101/2021.04.28.441833.

George, P.M., Wells, A.U., and Jenkins, R.G. (2020). Pulmonary fibrosis and COVID-19: the potential role for antifibrotic therapy. Lancet Respir. Med. *8*, 807–815.

Gibbons, M.A., MacKinnon, A.C., Ramachandran, P., Dhaliwal, K., Duffin, R., Phythian-Adams, A.T., van Rooijen, N., Haslett, C., Howie, S.E., Simpson, A.J., et al. (2011). Ly6Chi monocytes direct alternatively activated profibrotic macrophage regulation of lung fibrosis. Am. J. Respir. Crit. Care Med. *184*, 569–581.

Grant, R.A., Morales-Nebreda, L., Markov, N.S., Swaminathan, S., Querrey, M., Guzman, E.R., Abbott, D.A., Donnelly, H.K., Donayre, A., Goldberg, I.A., et al.; NU SCRIPT Study Investigators (2021). Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia. Nature *590*, 635–641.

Guimarães, P.O., Quirk, D., Furtado, R.H., Maia, L.N., Saraiva, J.F., Antunes, M.O., Kalil Filho, R., Junior, V.M., Soeiro, A.M., Tognon, A.P., et al.; STOP-COVID Trial Investigators (2021). Tofacitinib in Patients Hospitalized with Covid-19 Pneumonia. N. Engl. J. Med. *385*, 406–415.

Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. Nature *585*, 357–362.

Hasan, S.S., Capstick, T., Ahmed, R., Kow, C.S., Mazhar, F., Merchant, H.A., and Zaidi, S.T.R. (2020). Mortality in COVID-19 patients with acute respiratory distress syndrome and corticosteroids use: a systematic review and meta-analysis. Expert Rev. Respir. Med. *14*, 1149–1163.

He, X., Lau, E.H.Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y.C., Wong, J.Y., Guan, Y., Tan, X., et al. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat. Med. *26*, 672–675.

Heaton, H., Talman, A.M., Knights, A., Imaz, M., Gaffney, D.J., Durbin, R., Hemberg, M., and Lawniczak, M.K.N. (2020). Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. Nat. Methods *17*, 615–620.

Henderson, N.C., Rieder, F., and Wynn, T.A. (2020). Fibrosis: from mechanisms to medicines. Nature *587*, 555–566.

Hendrickson, C.M., Crestani, B., and Matthay, M.A. (2015). Biology and pathology of fibroproliferation following the acute respiratory distress syndrome. Intensive Care Med. *41*, 147–150.

Henry, B.M., and Lippi, G. (2020). Poor survival with extracorporeal membrane oxygenation in acute respiratory distress syndrome (ARDS) due to coronavirus disease 2019 (COVID-19): Pooled analysis of early reports. J. Crit. Care *58*, 27–28.

Herold, S., Becker, C., Ridge, K.M., and Budinger, G.R.S. (2015). Influenza virus-induced lung injury: pathogenesis and implications for treatment. Eur. Respir. J. *45*, 1463–1478.

Hijmans, R.J. (2020). Geographic Data Analysis and Modeling. R package raster version 3.4-5.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., Nitsche, A., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell *181*, 271–280.e8.

Holzwarth, K., Köhler, R., Philipsen, L., Tokoyoda, K., Ladyhina, V., Wählby, C., Niesner, R.A., and Hauser, A.E. (2018). Multiplexed fluorescence microscopy reveals heterogeneity among stromal cells in mouse bone marrow sections. Cytometry A *93*, 876–888.

Horby, P., Lim, W.S., Emberson, J.R., Mafham, M., Bell, J.L., Linsell, L., Staplin, N., Brightling, C., Ustianowski, A., Elmahi, E., et al.; RECOVERY Collaborative Group (2021). Dexamethasone in Hospitalized Patients with Covid-19. N. Engl. J. Med. *384*, 693–704.

Huang, C., Huang, L., Wang, Y., Li, X., Ren, L., Gu, X., Kang, L., Guo, L., Liu, M., Zhou, X., et al. (2021). 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. Lancet *397*, 220–232.

Huber, R., Pietsch, D., Panterodt, T., and Brand, K. (2012). Regulation of C/EBPβ and resulting functions in cells of the monocytic lineage. Cell. Signal. *24*, 1287–1296.

Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. *9*, 90–95.

Huynh, M.-L.N., Fadok, V.A., and Henson, P.M. (2002). Phosphatidylserine-dependent ingestion of apoptotic cells promotes TGF-beta1 secretion and the resolution of inflammation. J. Clin. Invest. *109*, 41–50.

Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat. Commun. *12*, 1088.

Jones, T.C., Biele, G., Mühlemann, B., Veith, T., Schneider, J., Beheim-Schwarzbach, J., Bleicker, T., Tesch, J., Schmidt, M.L., Sander, L.E., et al. (2021). Estimating infectiousness throughout SARS-CoV-2 infection course. Science *373*, eabi5273.

Joshi, N., Watanabe, S., Verma, R., Jablonski, R.P., Chen, C.-I., Cheresh, P., Markov, N.S., Reyfman, P.A., McQuattie-Pimentel, A.C., Sichizya, L., et al. (2020). A spatially restricted fibrotic niche in pulmonary fibrosis is sustained by M-CSF/M-CSFR signalling in monocyte-derived alveolar macrophages. Eur. Respir. J. *55*, 1900646.

Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0., https://CRAN.R-project.org/package=ggpubr.

Keenan, A.B., Torre, D., Lachmann, A., Leong, A.K., Wojciechowicz, M.L., Utti, V., Jagodnik, K.M., Kropiwnicki, E., Wang, Z., and Ma'ayan, A. (2019). ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic Acids Res. *47* (W1), W212–W224.

Kelly, T. (2019). leiden: R implementation of the Leiden algorithm. R package version 0.3.3. https://github.com/TomKellyGenetics/leiden.

Kidani, Y., and Bensinger, S.J. (2012). Liver X receptor and peroxisome proliferator-activated receptor as integrators of lipid homeostasis and immunity. Immunol. Rev. *249*, 72–83.

Korsunsky, I., Millard, N., Fan, J., et al. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.

Kurth, F., Roennefarth, M., Thibeault, C., Corman, V.M., Müller-Redetzky, H., Mittermaier, M., Ruwwe-Glösenkamp, C., Heim, K.M., Krannich, A., Zvorc, S., et al. (2020). Studying the pathophysiology of coronavirus disease 2019: a protocol for the Berlin prospective COVID-19 patient cohort (Pa-COVID-19). Infection *48*, 619–626.

Leisman, D.E., Ronner, L., Pinotti, R., Taylor, M.D., Sinha, P., Calfee, C.S., Hirayama, A.V., Mastroiani, F., Turtle, C.J., Harhay, M.O., et al. (2020). Cytokine elevation in severe and critical COVID-19: a rapid systematic review, meta-analysis, and comparison with other inflammatory syndromes. Lancet Respir. Med. *8*, 1233–1244.

Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., Cheng, L., Li, J., Wang, X., Wang, F., et al. (2020). Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nat. Med. *26*, 842–844.

Liaw, L., Skinner, M.P., Raines, E.W., Ross, R., Cheresh, D.A., Schwartz, S.M., and Giachelli, C.M. (1995). The adhesive and migratory effects of osteopontin are mediated via distinct cell surface integrins. Role of alpha v beta 3 in smooth muscle cell migration to osteopontin in vitro. J. Clin. Invest. *95*, 713–724.

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., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M.F., Strobl, D.C., Zappia, L., Dugas, M., Colomé-Tatché, M., et al. (2020). Benchmarking atlas-level data integration in single-cell genomics. bioRxiv. https://doi.org/10.1101/2020.05.22.111161.

Lukassen, S., Chua, R.L., Trefzer, T., Kahn, N.C., Schneider, M.A., Muley, T., Winter, H., Meister, M., Veith, C., Boots, A.W., et al. (2020). SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. EMBO J. *39*, e105114.

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. F1000Res. *5*, 2122.

Matthaei, M., Budt, M., and Wolff, T. (2013). Highly pathogenic H5N1 influenza A virus strains provoke heterogeneous IFN-α/β responses that distinctively affect viral propagation in human cells. PLoS ONE *8*, e56659.

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. J. Open Source Softw. *3*, 861.

Mehta, P., McAuley, D.F., Brown, M., Sanchez, E., Tattersall, R.S., and Manson, J.J.; HLH Across Speciality Collaboration, UK (2020). COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet *395*, 1033–1034.

Meinhardt, J., Radke, J., Dittmayer, C., Franz, J., Thomas, C., Mothes, R., Laue, M., Schneider, J., Brünink, S., Greuel, S., et al. (2021). Olfactory transmucosal SARS-CoV-2 invasion as a port of central nervous system entry in individuals with COVID-19. Nat. Neurosci. *24*, 168–175.

Melville, J. (2020). uwot: The Uniform Manifold Approximation and Projection (UMAP) Method for Dimensionality Reduction. R package version 0.1.8., https://CRAN.R-project.org/package=uwot.

Meneghin, A., and Hogaboam, C.M. (2007). Infectious disease, the innate immune response, and fibrosis. J. Clin. Invest. *117*, 530–538.

Mertins, P., Tang, L.C., Krug, K., Clark, D.J., Gritsenko, M.A., Chen, L., Clauser, K.R., Clauss, T.R., Shah, P., Gillette, M.A., et al. (2018). Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. Nat. Protoc. *13*, 1632–1661.

Messner, C.B., Demichev, V., Wendisch, D., Michalick, L., White, M., Freiwald, A., Textoris-Taube, K., Vernardis, S.I., Egger, A.-S., Kreidl, M., et al. (2020). Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. Cell Syst. *11*, 11–24.e4.

Misharin, A.V., Morales-Nebreda, L., Reyfman, P.A., Cuda, C.M., Walter, J.M., McQuattie-Pimentel, A.C., Chen, C.-I., Anekalla, K.R., Joshi, N., Williams, K.J.N., et al. (2017). Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. J. Exp. Med. *214*, 2387–2404.

Molyneaux, P.L., and Maher, T.M. (2013). The role of infection in the pathogenesis of idiopathic pulmonary fibrosis. Eur. Respir. Rev. *22*, 376–381.

Moore, B.B., Paine, R., 3rd, Christensen, P.J., Moore, T.A., Sitterding, S., Ngan, R., Wilke, C.A., Kuziel, W.A., and Toews, G.B. (2001). Protection from pulmonary fibrosis in the absence of CCR2 signaling. J. Immunol. *167*, 4368–4377.

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*, 1802441.

Nacu, N., Luzina, I.G., Highsmith, K., Lockatell, V., Pochetuhen, K., Cooper, Z.A., Gillmeister, M.P., Todd, N.W., and Atamas, S.P. (2008). Macrophages produce TGF-beta-induced (beta-ig-h3) following ingestion of apoptotic cells and regulate MMP14 levels and collagen turnover in fibroblasts. J. Immunol. *180*, 5036–5044.

Naik, P.K., and Moore, B.B. (2010). Viral infection and aging as cofactors for the development of pulmonary fibrosis. Expert Rev. Respir. Med. *4*, 759–771.

Niemeyer, D., Mösbauer, K., Klein, E.M., Sieberg, A., Mettelman, R.C., Mielech, A.M., Dijkman, R., Baker, S.C., Drosten, C., and Müller, M.A. (2018). The papain-like protease determines a virulence trait that varies among members of the SARS-coronavirus species. PLoS Pathog. *14*, e1007296.

Nouno, T., Okamoto, M., Ohnishi, K., Kaieda, S., Tominaga, M., Zaizen, Y., Ichiki, M., Momosaki, S., Nakamura, M., Fujimoto, K., et al. (2019). Elevation of pulmonary CD163$^+$ and CD204$^+$ macrophages is associated with the clinical course of idiopathic pulmonary fibrosis patients. J. Thorac. Dis. *11*, 4005–4017.

Ochs, M., Timm, S., Elezkurtaj, S., Horst, D., Meinhardt, J., Heppner, F.L., Weber-Carstens, S., Hocke, A.C., and Witzenrath, M. (2021). Collapse induration of alveoli is an ultrastructural finding in a COVID-19 patient. Eur. Respir. J. *57*, 2004165.

Ooms, J. (2014). The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. arXiv, 1403.2805. https://arxiv.org/abs/1403.2805.

Osuchowski, M.F., Winkler, M.S., Skirecki, T., Cajander, S., Shankar-Hari, M., Lachmann, G., Monneret, G., Venet, F., Bauer, M., Brunkhorst, F.M., et al. (2021). The COVID-19 puzzle: deciphering pathophysiology and phenotypes of a new disease entity. Lancet Respir. Med. *9*, 622–642.

Pan, Y., Guan, H., Zhou, S., Wang, Y., Li, Q., Zhu, T., Hu, Q., and Xia, L. (2020). Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): a study of 63 patients in Wuhan, China. Eur. Radiol. *30*, 3306–3309.

Pardo, A., Gibson, K., Cisneros, J., Richards, T.J., Yang, Y., Becerril, C., Yousem, S., Herrera, I., Ruiz, V., Selman, M., and Kaminski, N. (2005). Up-regulation and profibrotic role of osteopontin in human idiopathic pulmonary fibrosis. PLoS Med. *2*, e251.

Pascual-Reguant, A., Köhler, R., Mothes, R., Bauherr, S., Hernández, D.C., Uecker, R., Holzwarth, K., Kotsch, K., Seidl, M., Philipsen, L., et al. (2021). Multiplexed histology analyses for the phenotypic and spatial characterization of human innate lymphoid cells. Nat. Commun. *12*, 1737.

Patel, J., Beishuizen, A., Ruiz, X.B., Boughanmi, H., Cahn, A., Criner, G.J., Davy, K., de-Miguel-Díez, J., Fernandes, S., François, B., et al. (2021). A randomized trial of otilimab in severe COVID-19 pneumonia (OSCAR). medRxiv. https://doi.org/10.1101/2021.04.1421255475.

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal *10* (1), 439–446. https://doi.org/10.32614/RJ-2018-009.

Pertuz, S., Puig, D., Garcia, M.A., and Fusiello, A. (2013). Generation of all-in-focus images by noise-robust selective fusion of limited depth-of-field images. IEEE Trans. Image Process. *22*, 1242–1251.

Polak, S.B., Van Gool, I.C., Cohen, D., von der Thüsen, J.H., and van Paassen, J. (2020). A systematic review of pathological findings in COVID-19: a pathophysiological timeline and possible mechanisms of disease progression. Mod. Pathol. *33*, 2128–2138.

R Core Team (2019). R: A language and environment for statistical computing.

R Core Team (2020). R: A language and environment for statistical computing. https://www.R-project.org/.

Raivo, Kolde (2019). pheatmap: Pretty Heatmaps. R package version 1.0.12. https://CRAN.R-project.org/package=pheatmap.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. *2*, 1896–1906.

Remy, K.E., Mazer, M., Striker, D.A., Ellebedy, A.H., Walton, A.H., Unsinger, J., Blood, T.M., Mudd, P.A., Yi, D.J., Mannion, D.A., et al. (2020). Severe immunosuppression and not a cytokine storm characterizes COVID-19 infections. JCI Insight *5*, e140329.

Ren, Y.-C., Zhao, Q., He, Y., Li, B., Wu, Z., Dai, J., Wen, L., Wang, X., and Hu, G. (2020). Legumain promotes fibrogenesis in chronic pancreatitis via activation of transforming growth factor β1. J. Mol. Med. (Berl.) *98*, 863–874.

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. Am. J. Respir. Crit. Care Med. *199*, 1517–1536.

Richardson, S., Hirsch, J.S., Narasimhan, M., Crawford, J.M., McGinn, T., Davidson, K.W., Barnaby, D.P., Becker, L.B., Chelico, J.D., Cohen, S.L., et al.; the Northwell COVID-19 Research Consortium (2020). Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA *323*, 2052–2059.

Rittling, S.R. (2011). Osteopontin in macrophage function. Expert Rev. Mol. Med. *13*, e15.

Sadewasser, A., Paki, K., Eichelbaum, K., Bogdanow, B., Saenger, S., Budt, M., Lesch, M., Hinz, K.-P., Herrmann, A., Meyer, T.F., et al. (2017). Quantitative Proteomic Approach Identifies Vpr Binding Protein as Novel Host Factor Supporting Influenza A Virus Infections in Human Cells. Mol. Cell. Proteomics *16*, 728–742.

Satoh, T., Nakagawa, K., Sugihara, F., Kuwahara, R., Ashihara, M., Yamane, F., Minowa, Y., Fukushima, K., Ebina, I., Yoshioka, Y., et al. (2017). Identification of an atypical monocyte and committed progenitor involved in fibrosis. Nature *541*, 96–101.

Schapiro, D., Jackson, H.W., Raghuraman, S., Fischer, J.R., Zanotelli, V.R.T., Schulz, D., Giesen, C., Catena, R., Varga, Z., and Bodenmiller, B. (2017). histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. Nat. Methods *14*, 873–876.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schneider, C., Nobs, S.P., Kurrer, M., Rehrauer, H., Thiele, C., and Kopf, M. (2014). Induction of the nuclear receptor PPAR-γ by the cytokine GM-CSF is critical for the differentiation of fetal monocytes into alveolar macrophages. Nat. Immunol. *15*, 1026–1037.

Schubert, W., Bonnekoh, B., Pommer, A.J., Philipsen, L., Böckelmann, R., Malykh, Y., Gollnick, H., Friedenberger, M., Bode, M., and Dress, A.W.M. (2006). Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. Nat. Biotechnol. *24*, 1270–1278.

Schulte-Schrepping, J., Reusch, N., Paclik, D., Baßler, K., Schlickeiser, S., Zhang, B., Krämer, B., Krammer, T., Brumhard, S., Bonaguro, L., et al.; Deutsche COVID-19 OMICS Initiative (DeCOI) (2020). Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. Cell *182*, 1419–1440.e23.

Schupp, J.C., Binder, H., Jäger, B., Cillis, G., Zissel, G., Müller-Quernheim, J., and Prasse, A. (2015). Macrophage activation in acute exacerbation of idiopathic pulmonary fibrosis. PLoS ONE *10*, e0116775.

Schwensen, H.F., Borreschmidt, L.K., Storgaard, M., Redsted, S., Christensen, S., and Madsen, L.B. (2020). Fatal pulmonary fibrosis: a post-COVID-19 autopsy case. J. Clin. Pathol. *74*, 400–402.

Scott, M.K.D., Quinn, K., Li, Q., Carroll, R., Warsinske, H., Vallania, F., Chen, S., Carns, M.A., Aren, K., Sun, J., et al. (2019). Increased monocyte count as a cellular biomarker for poor outcomes in fibrotic diseases: a retrospective, multicentre cohort study. Lancet Respir. Med. *7*, 497–508.

Sheng, G., Chen, P., Wei, Y., Yue, H., Chu, J., Zhao, J., Wang, Y., Zhang, W., and Zhang, H.-L. (2020). Viral Infection Increases the Risk of Idiopathic Pulmonary Fibrosis: A Meta-Analysis. Chest *157*, 1175–1187.

Sinha, P., Matthay, M.A., and Calfee, C.S. (2020). Is a “Cytokine Storm” Relevant to COVID-19? JAMA Intern. Med. *180*, 1152–1154.

Spagnolo, P., Balestro, E., Aliberti, S., Cocconcelli, E., Biondini, D., Casa, G.D., Sverzellati, N., and Maher, T.M. (2020). Pulmonary fibrosis secondary to COVID-19: a call to arms? Lancet Respir. Med. *8*, 750–752.

Speranza, E., Williamson, B.N., Feldmann, F., Sturdevant, G.L., Pérez-Pérez, L., Mead-White, K., Smith, B.J., Lovaglio, J., Martens, C., Munster, V.J., et al. (2020). SARS-CoV-2 infection dynamics in lungs of African green monkeys. bioRxiv. https://doi.org/10.1101/2020.08.20.258087.

Speranza, E., Williamson, B.N., Feldmann, F., Sturdevant, G.L., Pérez-Pérez, L., Meade-White, K., Smith, B.J., Lovaglio, J., Martens, C., Munster, V.J., et al. (2021). Single-cell RNA sequencing reveals SARS-CoV-2 infection dynamics in lungs of African green monkeys. Sci. Transl. Med. *13*, eabe8146.

Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics *19*, 477.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Szabo, P.A., Dogra, P., Gray, J.I., Wells, S.B., Connors, T.J., Weisberg, S.P., Krupska, I., Matsumoto, R., Poon, M.M.L., Idzikowski, E., et al. (2021). Longitudinal profiling of respiratory and systemic immune responses reveals myeloid cell-driven lung inflammation in severe COVID-19. Immunity *54*, 797–814.e6.

Teoh, A.K.Y., Jo, H.E., Chambers, D.C., Symons, K., Walters, E.H., Goh, N.S., Glaspole, I., Cooper, W., Reynolds, P., Moodley, Y., and Corte, T.J. (2020). Blood monocyte counts as a potential prognostic marker for idiopathic pulmonary fibrosis: analysis from the Australian IPF registry. Eur. Respir. J. *55*, 1901855.

Thompson, B.T., Chambers, R.C., and Liu, K.D. (2017). Acute Respiratory Distress Syndrome. N. Engl. J. Med. *377*, 562–572.

Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233.

Trachsel, C., Panse, C., Kockmann, T., Wolski, W.E., Grossmann, J., and Schlapbach, R. (2018). rawDiag: An R Package Supporting Rational LC-MS Method Optimization for Bottom-up Proteomics. J. Proteome Res. *17*, 2908–2914.

Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. *11*, 2301–2319.

Urtasun, R., Lopategi, A., George, J., Leung, T.-M., Lu, Y., Wang, X., Ge, X., Fiel, M.I., and Nieto, N. (2012). Osteopontin, an oxidant stress sensitive cytokine, up-regulates collagen-I via integrin α(V)β(3) engagement and PI3K/pAkt/NFκB signaling. Hepatology *55*, 594–608.

Van Rossum, G., and Drake, F.L. (2009). Python 3 Reference Manual CreateSpace (Scotts Valley).

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al.; SciPy 1.0 Contributors (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272.

Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell *181*, 281–292.e6.

Waltman, L., and van Eck, N.J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. Eur. Phys. J. B *86*, 471.

Waskom, M. (2021). seaborn: statistical data visualization. J. Open Source Softw. *6*, 3021.

Webb, B.J., Peltan, I.D., Jensen, P., Hoda, D., Hunter, B., Silver, A., Starr, N., Buckel, W., Grisel, N., Hummel, E., et al. (2020). Clinical criteria for COVID-19-associated hyperinflammatory syndrome: a cohort study. Lancet Rheumatol. *2*, e754–e763.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (New York: Springer-Verlag).

Wickham, Hadley (2019). Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0.. https://CRAN.R-project.org/package=stringr.

Wickham, Hadley (2020). Hadley Wickham (2020). httr: Tools for Working with URLs and HTTP. R package version 1.4.2.. https://CRAN.R-project.org/package=httr.

Wickham, H. (2020). Tidy Messy Data. R package tidyr version 1.0.2.

Wickham, H., François, R., Henry, L., and Müller, K. (2020). A Grammar of Data Manipulation. R package dplyr version 1.0.2.

Wickham, Hadley, and Hester, Jim (2018). Read Rectangular Text Data. R package readr version 1.4.0.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15.

Wölfel, R., Corman, V.M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M.A., Niemeyer, D., Jones, T.C., Vollmar, P., Rothe, C., et al. (2020). Virological assessment of hospitalized patients with COVID-2019. Nature *581*, 465–469.

Wootton, S.C., Kim, D.S., Kondoh, Y., Chen, E., Lee, J.S., Song, J.W., Huh, J.W., Taniguchi, H., Chiu, C., Boushey, H., et al. (2011). Viral infection in acute exacerbation of idiopathic pulmonary fibrosis. Am. J. Respir. Crit. Care Med. *183*, 1698–1702.

Wynn, T.A., and Vannella, K.M. (2016). Macrophages in Tissue Repair, Regeneration, and Fibrosis. Immunity *44*, 450–462.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies: MELC antibodies (Cohort 1)		
DAPI	Roche	Cat# 10236276001, N/A
CCR2-PE	Miltenyi Biotec	Cat# 130-118-338, RRID:AB_2751486
CD163-PE	Biolegend	Cat# 333605, RRID:AB_1134005
CD56-PE	Miltenyi Biotec	Cat# 130-098-137, RRID:AB_2661200
CD3-PE	Miltenyi Biotec	Cat# 130-113-139, RRID:AB_2725967
CD169-PE	Miltenyi Biotec	Cat# 130-104-953, RRID:AB_2655537
CD14-PE	Miltenyi Biotec	Cat# 130-113-709, RRID:AB_2726250
CD45-PE	Miltenyi Biotec	Cat# 130-113-118, RRID:AB_2725946
CXCR3-PE	Miltenyi Biotec	Cat# 130-101-379, RRID:AB_2655734
CD16-PE	Miltenyi Biotec	Cat# 130-113-955, RRID:AB_2726428
CD4-PE	Miltenyi Biotec	Cat# 130-113-214, RRID:AB_2726025
TREM1-PE	Miltenyi Biotec	Cat# 130-101-033, RRID:AB_2657706
CD20-PE	Miltenyi Biotec	Cat# 130-113-374, RRID:AB_2726143
CD11b-PE	Miltenyi Biotec	Cat# 130-110-553, RRID:AB_2654665
CD8-PE	Miltenyi Biotec	Cat# 130-113-720, RRID:AB_2726261
CD1c-PE	Miltenyi Biotec	Cat# 130-113-864 RRID: AB_2726358
CD68-PE	Miltenyi Biotec	Cat# 130-118-486, RRID:AB_2784270
CD127-PE	Miltenyi Biotec	Cat# 130-113-414, RRID:AB_2733759
CD11c-PE	Miltenyi Biotec	Cat# 130-113-580, RRID:AB_2726180
CD57-PE	Miltenyi Biotec	Cat# 130-111-963, RRID:AB_2658747
HLA-DR-PE	Miltenyi Biotec	Cat# 130-120-715, RRID:AB_2752176
CD66b-PE	Miltenyi Biotec	Cat# 130-122-922, N/A
COL-IV-FITC	Antibodies-Online	Cat# ABIN376119, RRID:AB_10763557
MRP14-PE	Miltenyi Biotec	Cat# 130-114-516, RRID:AB_2726684
C1q-FITC	DAKO	Cat# F0254, RRID:AB_2335713
Antibodies: Immunohistochemistry (Cohort 2)		
Goat Anti-Type I Collagen	Southern Biotech	Cat# 1310-01; RRID:AB_2753206
Goat Anti-Type III Collagen	Southern Biotech	Cat# 1330-01; RRID:AB_2794734
Goat Anti-Type IV Collagen	Southern Biotech	Cat# 1340-01; RRID:AB_2721907
Biotinylated rabbit anti-goat	Vector	Cat# BA-5000; RRID:AB_2336126
Mouse anti-CD68	Agilent	Cat# M0876; RRID:AB_2074844
Rabbit anti-sm22	Abcam	Cat# Ab14106; RRID:AB_443021
Goat anti-rabbit Alexa 647	Abcam	Cat# Ab150079; RRID:AB_2722623
Mouse anti-CD163	Cell Marque	Cat# 163M-17; RRID:AB_1159119
Mouse anti-CD16	Santa Cruz	Cat# sc-20052; RRID:AB_626925
Antibodies: Antibodies for sorting		
Brilliant Violet 785 anti-human HLA-DR Antibody	Biolegend	Cat# 307642
FITC anti-human CD14 Antibody	Biolegend	Cat# 367116
CD16 APC B73.1	Biolegend	Cat# 360705
PerCP anti-human CD19 Antibody SJ25C1	Biolegend	Cat# 363013
CD3 PerCP UCHT1	Biolegend	Cat# 300427
CD56 PerCP 5.1 h11	Biolegend	Cat# 362526

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CD304 (BDCA-4)-PE-Vio770, human, AD5-17F6	Miltenyi	Cat# 130-113-518
BV510 Mouse Anti-Human CD141 Clone 1A4 (RUO)	BD	Cat# 563298
CD1c PE AD5-8E7	Miltenyi	Cat# 130-113-302
Virus strains		
BetaCoV/Munich/BavPat1/2020 (passage 2, no second site mutations)	This study	GISAID accession: EPI_ISL_406862
A/Panama/2007/1999	This study	N/A
Chemicals, peptides, and recombinant proteins		
KAPA HiFi HotStart Ready Mix	Roche	KK2601
Human Tru Stain FcX	Biolegend	422301
TE Buffer	Thermo Fisher	120900115
SPRIselect Reagent	Invitrogen	AM9937
10% Tween 20	BIO-RAD	1662404
Buffer EB	QIAGEN	19086
Ethanol, Absolute	Fisher Bioreagents	BP2818-500
Glycerol, 85%	Merck	1040941000
Bovine Serum Albumin	Jackson Immuno Research	001-000-161
RBC Lysis Buffer (10X)	Biolegend	1662404
TMTpro reagents	Thermo Fisher Scientific	A44520
Complete Protease Inhibitor Cocktail	Roche	11697498001
Phosphatase Inhibitor Cocktail 2	Sigma-Aldrich	P5726
Phosphatase Inhibitor Cocktail 3	Sigma-Aldrich	P0044
Lysyl Endopeptidase, Mass Spectrometry Grade (Lys-C)	FUJIFILM Wako Pure Chemical Corporation	125-05061
Sequencing Grade Modified Trypsin	Promega	V5113
Critical commercial assays		
Qubit dsDNA HS Assay Kit	ThermoFisher	Q32854
Chromium Next GEM Single Cell 3′ GEM, Library & Gel Bead Kit v3.1	10x genomics	1000121
Chromium Next GEM Chip G Single Cell Kit	10x genomics	1000120
Single Index Kit T Set A	10x genomics	1000213
High Sensitivity DNA Kit	Agilent Technologies	5067-4626
Pierce BCA Protein Assay Kit	Thermo Fisher Scientific	23225
EnV FLEX, High pH, (Link)	Agilent Technologies	K800021-2
EnVision FLEX/ HRP goat anti-mouse (ready-to-use)	Agilent Technologies	K8000
EnVision FLEX, High pH (Link), HRP. Rabbit/ Mouse	Agilent Technologies	DM802
Opal 690	Akoya Biosciences	FP1497001KT
Opal 650 TSA Plus	Akoya Biosciences	FP1496001KT
Opal 620 TSA Plus	Akoya Biosciences	FP1495001KT
Spectral DAPI	Akoya Biosciences	FP1490
Human CCL24/Eotaxin-2/MPIF-2 DuoSet	R&D Systems	DY343
Human CCL8/MCP-2 DuoSet ELISA	R&D Systems	DY281
Human IL-6 DuoSet ELISA	R&D Systems	DY206
Human CCL2/MCP-1 DuoSet	R&D Systems	DY279
Human IL-1 beta/IL-1F2 DuoSet	R&D Systems	DY201
Human CXCL5/ENA-78 DuoSet	R&D Systems	DY254
Human IFN-beta DuoSet	R&D Systems	DY814-05
RNAscope probe V-nCoV2019-S	Advanced Cell Diagnostics	#848561-C1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
UniProt Human protein database	Uniprot	https://www.uniprot.org/downloads
Uniprot SARS-CoV-2 protein database	Uniprot	https://www.uniprot.org/downloads
Uniprot Influenza A (Panama) protein database	Uniprot	https://www.uniprot.org/downloads
MSigDB v7.0	Broad Institute	https://www.gsea-msigdb.org/gsea/index.jsp
scRNA-seq raw data	This paper	EGAS00001004928 EGAS00001005634
snRNA-seq raw data	(Gassen et al., 2021)	EGAS00001004689
RAW proteomics and phosphoproteomics data	This paper	PXD022709
Experimental models: Cell lines		
Vero E6 cells	ATCC	No. CRL-1586
MDCKII cells	ATCC	No. CRL-2936
Oligonucleotides		
SI-PCR primer	IDT	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC*T*C
HTO additive primer	IDT	GTGACTGGAGTTCAGACGTGTGC*T*C
D701_S	IDT	CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGT*G*C
D702_S	IDT	CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTGACTGGAGTTCAGACGTGT*G*C
D703_S	IDT	CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCAGACGTGT*G*C
D705_S	IDT	CAAGCAGAAGACGGCATACGAGATTTCTGAATGTGACTGGAGTTCAGACGTGT*G*C
Software and algorithms		
MaxQuant 1.6.10.43	Cox and Mann, 2008	https://www.maxquant.org/
R 3.6	R Core Team, 2019	https://www.r-project.org/
GSEA 2.0	Subramanian et al., 2005	https://www.gsea-msigdb.org/gsea/index.jsp
R version 3.6.3	R Core Team, 2020	https://cran.r-project.org/
R package Seurat version 3.2.2	Stuart et al., 2019	https://cran.r-project.org/web/packages/Seurat/index.html
R package leiden version 0.3.3	Traag et al., 2019; Kelly, 2019	https://cran.r-project.org/web/packages/leiden/index.html
R package scran version 1.14.6	Lun et al., 2016	https://bioconductor.org/packages/release/bioc/html/scran.html
R package ggplot2 version 3.3.2	Wickham, 2016	https://cran.r-project.org/web/packages/ggplot2/index.html
R package dplyr version 1.0.2	Wickham et al., 2020	https://cran.r-project.org/web/packages/dplyr/index.html
R package uwot version 0.1.8	Melville, 2020	https://cran.r-project.org/web/packages/uwot/index.html
R package clusterProfiler version 3.14.3	Yu et al., 2012	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
R package ggpubr version 0.4.0	Kassambara, 2020	https://cran.r-project.org/web/packages/ggpubr/index.html
R package tidyr version 1.1.2	Wickham, 2020	https://cran.r-project.org/web/packages/tidyr/index.html
R package slingshot version 1.4.0	Street et al., 2018	https://bioconductor.org/packages/release/bioc/html/slingshot.html
R package rgl version 0.100.19	Adler et al., 2019	https://cran.r-project.org/web/packages/rgl/index.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R package readr version 1.4.0	Wickham and Hester, 2018	https://cran.r-project.org/web/packages/readr/index.html
R package ggalluvial version 0.12.2	Brunson, 2020	https://cran.r-project.org/web/packages/ggalluvial/index.html
R package pheatmap version 1.0.12	Raivo, 2019	https://cran.r-project.org/web/packages/pheatmap/index.html
R package httr version 1.4.2	Wickham, 2020a	https://cran.r-project.org/web/packages/httr/index.html
R package jsonlite version 1.7.1	Ooms, 2014	https://cran.r-project.org/web/packages/jsonlite/index.html
R package stringr version 1.4.0	Wickham, 2019	https://cran.r-project.org/web/packages/stringr/index.html
R package ggnewscale version 0.4.5	Campitelli, 2021	https://cran.r-project.org/web/packages/ggnewscale/index.html
R package viridis version 0.5.1	Garnier, 2018	https://cran.r-project.org/web/packages/viridis/index.html
R package grid version 3.6.3	R Core Team, 2020	https://www.R-project.org/
R package raster version 3.4	Hijmans, 2020	https://cran.r-project.org/web/packages/raster
R package sf version 0.9	Pebesma, 2018	https://cran.r-project.org/web/packages/sf
Python 3.7.8	Van Rossum and Drake, 2009	https://www.python.org/
SCANPY version 1.7.2	Wolf et al., 2018	https://scanpy.readthedocs.io/en/stable/
scVI version 0.6.7	Gayoso et al., 2021	https://scvi-tools.org/
Python package seaborn version 0.10.1	Waskom, 2021	https://seaborn.pydata.org/
Python package scipy version 1.5.2	Virtanen et al., 2020	https://scipy.org/
Python package numpy version 1.20.3	Harris et al., 2020	https://numpy.org/
Python package matplotlib version 3.3.3	Hunter, 2007	https://matplotlib.org/
Other		
QExactive HF-x Orbitrap MS	Thermo Fisher Scientific	IQLAAEGAAPFALGMBFZ
Waters XBridge Peptide BEH C18 (130A, 3.5µm; 2.1mm x 250mm)	Waters	186003566
Bravo Automated Liquid Handling Platform	Agilent	G5409-90006
AssayMAP Fe(III)-NTA cartridges	Agilent	G5496-60085
EASY-nLC 1200	Thermo Fisher Scientific	LC140
Image Cycler MM3 (TIC)	MelTec GmbH & Co.KG	N/A
QuPath	Bankhead et al., 2017	0.2.3
ZEN 3.0 black edition	Carl Zeiss AG	N/A
InForm	Akoya Biosciences	N/A
GraphPad Prism	GraphPad Software	Version 5.01

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Leif E. Sander (leif-erik.sander@charite.de).

Materials availability

This study did not generate new unique reagents.

DeCOI consortium members

Janine Altmüller, Angel Angelov, Anna C Aschenbrenner, Robert Bals, Alexander Bartholomäus, Anke Becker, Mattias Becker, Michael Beckstette, Daniela Bezdan, Michael Bitzer, Helmut Blum, Conny Blumert, Ezio Bonifacio, Peer Bork, Bunk Boyke, Nicolas

Casadei, Thomas Clavel, Maria Colome-Tatche, Markus Cornberg, Inti Alberto De La Rosa Velázquez, Andreas Diefenbach, Alexander Dilthey, Nicole Fischer, Konrad Förstner, Sören Franzenburg, Julia-Stefanie Frick, Gisela Gabernet, Julien Gagneur, Tina Ganzenmüller, Marie Gauder, Janina Geißert, Alexander Goesmann, Siri Göpel, Adam Grundhoff, Hajo Grundmann, Torsten Hain, Frank Hanses, Ute Hehr, André Heimbach, Marius Höper, Friedmann Horn, Daniel Hübschmann, Michael Hummel, Thomas Iftner, Angelika Iftner, Thomas Illig, Stefan Janssen, Jörn Kalinowski, René Kallies, Birte Kehr, Andreas Keller, Oliver T. Keppler, Sarah Kim-Hellmuth, Christoph Klein, Michael Knop, Oliver Kohlbacher, Karl Köhrer, Jan Korbel, Peter G. Kremsner, Denise Kühnert, Ingo Kurth, Markus Landthaler, Yang Li, Kerstin U. Ludwig, Oliwia Makarewicz, Manja Marz, Alice McHardy, Christian Mertes, Maximilian Münchhoff, Sven Nahnsen, Markus Nöthen, Francine Ntoumi, Peter Nürnberg, Uwe Ohler, Stephan Ossowski, Jörg Overmann, Silke Peter, Klaus Pfeffer, Isabell Pink, Anna R Poetsch, Ulrike Protzer, Alfred Pühler, Nikolaus Rajewsky, Markus Ralser, Kristin Reiche, Olaf Rieß, Stephan Ripke, Ulisses Rocha, Philip Rosenstiel, Antoine-Emmanuel Saliba, Leif Erik Sander, Birgit Sawitzki, Simone Scheithauer, Philipp Schiffer, Jonathan Schmid-Burgk, Wulf Schneider, Eva-Christina Schulte, Joachim L. Schultze, Nicole Schulz, Alexander Sczyrba, Mariam L. Sharaf, Yogesh Singh, Michael Sonnabend, Oliver Stegle, Jens Stoye, Fabian J. Theis, Thomas Ulas, Janne Vehreschild, Thirumalaisamy P. Velavan, Jörg Vogel, Sonja Volland, Max von Kleist, Andreas Walker, Jörn Walter, Dagmar Wieczorek, Sylke Winkler, John Ziebuhr.

Data and code availability

scRNA-seq data generated during this study are deposited at the European Genome-phenome Archive (EGA) under the accession numbers EGAS00001004928 and EGAS00001005634, which is hosted by the EBI and the CRG. snRNA-seq data generated previously (Gassen et al., 2021) are accessible under the accession number EGAS00001004689. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier: PXD022709.

R code used for the analysis of scRNA-seq data has been deposited on GitHub: https://github.com/OliverDietrich/SARS-CoV-2-infection-triggers-profibrotic-macrophage-responses-and-lung-fibrosis

Count matrices and Seurat objects have been deposited via Nubes: https://nubes.helmholtz-berlin.de/s/XrM8igTzFTFSoio.

Python code used for scRNA-seq data integration has been deposited on GitHub: https://github.com/theislab/covid_macrophages_integration

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cohort 1 - Berlin cohort

ICU cohort Berlin

Patients treated at a single ICU with COVID-19 associated acute respiratory distress syndrome (ARDS) between March 17th, 2020 and March 17th, 2021 were included in this analysis. This cohort represents a sub-cohort of the Pa-COVID-19 study, a prospective observational cohort study assessing pathophysiology and clinical characteristics of patients with COVID-19 treated at Charité Universitätsmedizin Berlin (Kurth et al., 2020). The study was approved by the Institutional Review board of Charité (EA2/066/20). Written informed consent was provided by all patients or legal representatives for participation in this study.

Patients treated with severe COVID-19 ARDS with requirement for veno-venous vvECMO treatment, qualified for inclusion for the assessment of CT images and pulmonary gas exchange. Out of 18 identified patients, two patients were excluded from the analysis due to death less than 36h after vvECMO initiation. Information on age, sex, medication, commorbidities and outcome is provided in Table S1.

Pathology cohort Berlin

The autopsy study was approved by the Ethics Committee of the Charité (EA 1/144/13, EA2/066/20 and EA1/075/19) and was performed in compliance with the Declaration of Helsinki. For histology, we analyzed cryopreserved lung tissue from deceased patients with COVID-19. Patients were selected from an autopsy cohort at Charité described in Meinhardt et al. (2021). Inclusion criteria were presence of cryopreserved material and detectable SARS-CoV-2 RNA load in the lung tissue. Patients with detectable tumor infiltration and one case with graft-versus-host reaction after stem cell therapy were excluded from the analysis. Information on age, sex, medication, commorbidities is listed in Table S1.

Cohort 2 - Aachen cohort

The study was approved by the local ethics committee (EK 304/20, EK 119/20, and EK 092/20). We included 15 consecutive clinical autopsies of COVID-19 positive patients between March 9th, 2020 and January 1st, 2021. Each patient had a positive clinical SARS-CoV-2 PCR test from upper or lower respiratory tract prior to autopsy. Consent to autopsy was obtained by the legal representatives of the deceased patients. The autopsies were performed in two steps according to a modified standard protocol to further increase employee safety and sample acquisition (developed in the frame of the German Registry of COVID-19 autopsies – DeRegCOVID). As a control, five non-COVID clinical autopsy lung tissues from 2013 to 2015 were included. Additionally, one non-autopsy lung tissue from a COVID-19 positive patient and two non-autopsy lung tissues from non-COVID-19 patients for routine diagnostic histological assessment of surgical specimens were included into histological analyses.

Additional datasets used to perform data integration and snRNA-seq

For the proximity analysis the BAL scRNA-seq macrophage data of this study was integrated together with data originating from three previously published datasets (Adams et al., 2020) (GSE136831), (Morse et al., 2019) (GSE128033) and (Bharat et al., 2020) (GSE158127). A detailed cohort description can be found in the original publications. Briefly, the Adams et al. dataset encompasses lung sample data of 32 IPF, 28 smoker and non-smoker controls, and 18 chronic obstructive pulmonary disease (COPD). Morse et al. includes two samples each of explanted lungs of three IPF patients from upper and lower lobes that represent early and late disease progression respectively, as well as three healthy controls. (Bharat et al., 2020) comprises overall three whole lung single-cell RNA-seq datasets, one originating from a patient that underwent bilateral lung transplantation ('case 1' in the in the original paper), and two originating from post-mortem lung biopsies from two patients who had died from severe COVID-19 ('PMB 1 and 2′ in the original paper). Raw snRNA-seq data (6 COVID-19 and 1 control) originates from Gassen et al. (2021) accessible under the repository EGAS00001004689. Two additional controls were added from Lukassen et al. (2020) accessible under the repository EGAS00001004419 (Patient ID: JVV9L8ng/SAMEA6848761; S4ECX8ng/SAMEA6848765).

METHOD DETAILS

Clinical investigation

VCin measurement

An automated inspiratory/expiratory pressure volume curve was performed using the ventilator (S1, Hamilton Medical, Bonaduz, Switzerland). Pressure was increased from 0 mbar up to 45 mbar and then released to 0 mbar again in steps of 2 mbar/second. The inflated volume at 45 mbar was defined as the inspiratory vital capacity (VCin). All measurements were performed in supine position under deep sedation thereby avoiding spontaneous breathing attempts of the patients during the maneuver. In five patients of cohort 1 with VCin measurements available, the highest VCin in the early phase (from intubation until day 7 of vvECMO support) and the lowest during the late vvECMO phase (> 7 days of vvECMO support) (acute vvECMO phase) was determined.

CT Scans

Computed tomographic (CT) scanning was performed in supine position. When available one CT scan in the acute phase within 7 days after submission to intensive care, one around the time point of vvECMO initiation and the last one available of the patient (either before death, dismission from the hospital or end of the follow up period). CT scans were evaluated and annotated by two board certified pulmonologists.

Viral Stocks

SARS-CoV-2 stock for scRNASeq

100 μl of passage 0 virus isolate of the BetaCoV/Munich/BavPat1/2020 EPI_ISL_406862 strain was diluted in 20 mL Dulbecco modified Eagle medium (DMEM) containing 1% sodium pyruvate, 1% non-essential amino acids and 2% fetal calf serum (FCS, GIBCO). Approximately 1x10^6 VeroE6 cells (ATCC CRL-1586) were infected with 20 mL of the diluted virus. At 3 dpi, supernatant was harvested and the virus was purified by membrane ultracentrifugation (Vivaspin 100 kDa MWCO, GE Healthcare). Briefly, the column was equilibrated with 10 mL PBS, followed by centrifugation for 10 min and 3.000 g. Flow-through was discarded and 20 mL of virus-containing supernatant was added to the column and centrifuged for 60 min at 3.000 g. The concentrated virus (approximately 0.5 ml) was resuspended in 3 mL OptiPro serum-free medium (GIBCO), which was then further diluted 1:2 in virus preservation medium (OptiPro containing 0.5% gelatin), aliquoted and stored at −80°C. To determine virus titer, all stocks were plaque titrated in three independent experiments as described previously (Niemeyer et al., 2018). The absence of any second site mutations was confirmed by next generation sequencing.

SARS-CoV-2 stock for proteomics

Approximately 1x10^7 VeroE6 cells (*ATCC* No. CRL-1586) were infected with BetaCoV/Munich/BavPat1/2020 strain (GISAID accession: EPI_ISL_406862) passage 1 at an MOI of 0,01 in 12 mL Dulbecco modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin, 1x non-essential amino acids and 1 mM sodium pyruvate and incubated at 37°C and 5% CO_2 for 48 h. Following centrifugation at 3.500 rpm for 5 min. to remove cell debris, virus-containing supernatant was added to Amicon Ultra Centrifugal Filters (100 kDa MWCO, Merck, Germany). Spin filters were centrifuged at 4.000 g for 15 min and subsequently culture medium was added to the concentrated virus suspension to a final volume of 4 ml. Viral stocks were re-sequenced after passaging to rule out cell culture adaptation mutations.

Influenza A/Panama/2007/1999

Virus stock was grown on MDCKII cells (ATCC No. CRL-2936) in minimum essential medium (MEM) supplemented with 0,2% BSA, 2mM L-glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin and 1 μg/ml TPCK-treated trypsin. Cells were infected at an MOI of 0,01 and incubated for 48h at 37°C and 5% $CO_{2.}$ Virus-containing supernatant was centrifuged for 5 min at 3.500 rpm. Virus titers were determined by plaque assay on Vero E6 cells for SARS-CoV 2 and MDCKII cells for Influenza A/Panama/2007/1999 using Avicel overlay as described previously (Matthaei et al., 2013; Niemeyer et al., 2018). Virus stocks were stored at −80°C.

Histology, Immunohistochemistry, Immunofluorescence and SARS-CoV-2 RNA in situ hybridization

Histological analysis

The samples of the central and peripheral lung from each lobe, 4% formaldehyde-fixed and paraffin-embedded (FFPE), were further processed and histologically evaluated. FFPE blocks were cut to 1 μm thick sections, deparaffinized, rehydrated, and stained with hematoxylin-eosin (H&E) staining according to standard protocol using an automated staining system (Tissue-Tek Prisma® Plus & Tissue-Tek Film®, Sakura).

Semiquantitative estimation of pulmonary fibrosis

To estimate pulmonary fibrosis, a semiquantitative score (Ashcroft Score) was used as previously described (Ashcroft et al., 1988). Briefly, an H&E section of the lung was examined systematically at 10x magnification. Each patient was assigned a score between 0 and 8 using a predetermined scale of severity (0 = normal lung, 8 = total fibrous obliteration of the field), based on the predominant degree of fibrosis on the lung section. Fields predominantly occupied by large bronchi or vessels, or by malignant tumor deposits were not counted. Inflammatory cells in airspaces were ignored, but organized exudate was treated as fibrosis. Two experienced pathologists scored each case independently and a mean value was calculated from the two scores for each patient.

Immunohistochemistry and immunofluorescence

For immunohistochemistry (collagens I, III, IV, CD68), FFPE blocks were cut to 1 μm thin sections, deparaffinized, rehydrated and washed in phosphate-buffered saline (PBS). Slides underwent heat-induced epitope retrieval in citrate buffer (pH 6.0, Antigen Unmasking Solution, Citric Acid-Based (Vector, H-3300)), were quenched with 3% H_2O_2, incubated in a humidified chamber with primary antibodies (Key resources table) (dilutions: anti- type I collagen, 1:100; anti- type III collagen, 1:200; anti- type IV collagen, 1:250, anti-CD68, 1:100) for one hour followed by incubation with a secondary antibody (dilution 1:300 or ready-to-use, Key resources table) (30 minutes), followed by ABC complex (30 minutes), followed by 3-3′-Diaminobenzidine (DAB) for 10 minutes, and counterstained with methyl green. CD68-/Sm22-costaining was performed accordingly with the following differences: no H_2O_2-blocking, ABC complex and DAB incubation were performed and slides were not counterstained with methyl green. Dilutions of primary antibodies were 1:100. After incubation of the secondary antibody (ready-to-use or dilution 1:300, Key resources table) DAPI was used to highlight nuclei.

To analyze Macrophage subtypes, we used the VECTRA automated quantitative pathology imaging system, and stained using our pre-established protocol as followed: slides underwent antigen retrieval in citrate buffer (EnVision FLEX TARGET RETRIEVAL SOLUTION LOW pH, from Agilent: K8005) using the pT-Link module (Agilent, Santa Clara, USA). After fixation in 4% formaldehyde for 10 min, slides were washed and blocking was performed with H_2O_2 (DAKO REAL PEROXIDASE-BLOCKING SOLUTION, Agilent, Santa Clara, USA: S2023) followed by 30 min incubation with antibody diluent (DAKO REAL ANTIBODY DILUENT, Agilent, Santa Clara, USA: S2022). Immunofluorescence multiplex staining was performed with Opal 7-Color Manual IHC Kit (AKOYA Biosciences, Menlo-Park, USA: NEL811001KT). The slides were incubated for 1 hour with primary antibodies: CD68 (Agilent, Santa Clara, USA: M0876), CD163 (Cell Marque: 163M-17) and CD16 (Santa Cruz Biotechnology, Dallas, USA: DJ130c; Key resources table), followed by incubation with EnVision FLEX HRP (Agilent, Santa Clara: DM802) and visualized with Opal 690 TSA Plus, using Opal 650 TSA Plus, and Opal 620 TSA Plus, respectively (all from AKOYA Biosciences, Marlborough, USA). The nuclei were counterstained using Spectral DAPI (AKOYA Biosciences, Marlborough, USA).

Analysis of immunohistochemistry and immunofluorescence

Immunohistochemistry slides were scanned with the 40x objective using an Aperio AT2 whole Slide Scanner (Leica Biosystems, Wetzlar, Germany). Immunohistochemistry whole slide images were evaluated using the latest stable QuPath (Bankhead et al., 2017) release (i.e., version 0.2.3) by first applying automated tissue detection and subsequently assessing the tissue area that showed 3,3’-Diaminobenzidine (DAB) positivity. DAB positive tissue was determined by applying a tissue classifier that counted all pixels as DAB-positive that showed a value of higher than 0.45 in the DAB channel. This threshold was used on all slides. The final readout was the proportion of DAB positive area of the complete tissue section. To quantify the amount of CD68-positive macrophages, we used QuPaths in-built positive cell detection algorithm with a single threshold. The final readout was the ratio between CD68 positive cells and all detected cells.

To subtype CD68 positive macrophages into CD163 positive and negative macrophages we scanned three regions of interest of lung tissue per sample using the 40x objective, corresponding to a tissue area of 753x103 μm^2 each. Scanning was performed using the VECTRA automated quantitative pathology imaging system (Perkin Elmer, Waltham, USA). After deploying automated cell detection using the InForm Software (Akoya Biosciences, Marlborough, USA), we trained an in-built cell phenotyping algorithm to detect CD68+/CD163- and CD68+/CD163+ cells in our cohorts. From each cohort, one slide was used for training and manually annotated using the above mentioned phenotypes. This algorithm was then deployed on all samples to detect the above mentioned phenotypes. Measurement outputs of the inForm-Software were analyzed using the phenoptr and phenoptrReports packages in R version 4.0.3. Our final readout was the proportion of CD163 positive and negative cells within the CD68 positive cell proportion. Representative immunofluorescence pictures were obtained with a Zeiss Axio Imager 2 microscope using 40x objective and image analysis software ZEN 3.0 black edition (both Carl Zeiss AG, Oberkochen, Germany). Statistical analysis was performed using Mann Whitney Test on GraphPad Prism Version 5.01 (GraphPad Software, Inc).

SARS-CoV-2 RNA in situ hybridization

FISH was performed on 1 μm thin FFPE sections with the RNAscope® Multiplex Fluorescent Reagent Kit v2 assay (Advanced Cell Diagnostics, Inc., Hayward, USA). Briefly, we incubated the tissue sections with H_2O_2, performed heat-induced target retrieval followed by protease incubation with the reagents provided. RNA sequence of SARS-CoV-2 S gene was hybridized using RNAscope® probe V-nCoV2019-S (#848561-C1). After the amplifier steps according to the manual, Opal^{TM} 650 fluorophore (Perkin Elmer, Waltham, USA) was applied to the tissues. Finally, nuclei were labeled with DAPI and the slides were mounted with ProLong^{TM} Gold antifade reagent (Invitrogen, Waltham, USA). Representative images were obtained with a Zeiss Axio Imager 2 microscope using the 40x objective and the image analysis software ZEN 3.0 black edition, (both Carl Zeiss AG, Oberkochen, Germany). Colocalization with CD68 was performed by analyzing serial sections stained with CD68 and RNA Scope for the SARS-CoV-2 S gene.

Bronchoalveolar lavage (BAL)

Bronchoalveolar lavage (BAL) samples were obtained from mechanically ventilated COVID-19 patients at the Department of Infectious Diseases and Respiratory Medicine, Charité - Universitätsmedizin Berlin. BAL fluid was filtered through a 70 μm mesh and centrifuged (400 g, 10 min, 4°C). The supernatant was removed and cells were washed once with DPBS (GIBCO). Erythrocytes were then removed using the Red Blood Cell (RBC) lysis buffer (Biolegend). The cells were washed twice and resuspended in DPBS (GIBCO) and cells were passed through a 40 μm mesh (Flowmi^{TM} Cell Strainer, Merck). The cell suspension was then adjusted to a concentration of 700-1200 cells/μl to load a total of 16.500 cells/reaction into the 10x Genomics Chromium controller for scRNA-seq. Single Cell 3′ reagent kit v3.1 was used for reverse transcription, cDNA amplification and library construction following the detailed protocol provided by 10x Genomics. Libraries were quantified by QubitTM 2.0 Fluorometer (ThermoFisher) and quality was checked using 2100 Bioanalyzer with High Sensitivity DNA kit (Agilent). Sequencing was performed in paired-end mode with S1 and S2 flow cells (2 × 50 cycles kit) using NovaSeq 6000 sequencer (Illumina).

Monocytes isolation and infection

Isolation of primary classical monocytes

Peripheral blood mononuclear cells were isolated by gradient centrifugation (1.077 g/ml Pancoll, PAN Biotech) from fresh EDTA blood or buffy coats (German Red Cross Blood Transfusion Service, Berlin) of healthy donors, followed by immunomagnetic depletion of CD3$^+$/CD19$^+$/CD20$^+$/CD56$^+$/CD235a$^+$ cells using biotinylated antibodies (Biolegend) and MagniSort Streptavidin Negative Selection Beads (Invitrogen) (Key resources table). Subsequently monocyte subsets were sorted using a BD FACSAria SORP cell sorter (BD Biosciences) starting with HLA-DR$^+$, CD3$^-$/CD19$^-$/CD20$^-$/CD56$^-$ cells following diverse gating strategies: classical monocytes (CD14$^+$, CD16$^-$), non-classical monocytes (CD14^{dim} CD16$^+$), myeloid dendritic cells (cDC2) (CD14$^-$/CD16$^-$/CD141$^-$/CD304$^-$, CD1c$^+$), and plasmacytoid dendritic cells (CD14$^-$/CD16$^-$/CD141$^-$/CD1c$^-$, CD304$^+$). Cells were washed in RPMI 1640 (GIBCO) supplemented with 10% (v/v) FCS (Sigma), 1% (v/v) non-essential amino acid solution (Sigma), 1% (v/v) HEPES (Sigma), 1% (v/v) Glutamine solution (GIBCO) and 1% (v/v) sodium pyruvate (GIBCO).

Infection assays

Sorted monocytes were washed and resuspended in complete medium (RPMI 1640, GIBCO), 10% (v/v) heat-inactivated FCS (Sigma), 1% (v/v) non-essential amino acid solution (Sigma), 1% (v/v) HEPES (Sigma), 1% (v/v) Glutamin solution (GIBCO), 1% (v/v) Sodium Pyruvate (GIBCO). For the scRNA-Seq experiments, 1x10^5 cells were seeded per well into a 96-well plate. For the proteomics experiment 1x10^6 cells/well were seeded into a 12 well plate. The cells were rested for one hour at 37°C and 5% CO_2. For the scRNA-seq experiments, the concentrated 2020 EPI_ISL_406862 SARS-CoV-2 isolate (see 'Viral Stock' section) was added at a ratio of cells to virus (plaque forming units; PFUs) of 5 (MOI = 5) and adjusted to a total volume of 100 μl. For the proteomic experiments SARS-CoV-2 (EPI_ISL_406862, passage 2) and Influenza A (passage 3, isolate: Panama/2007/1999) were added at a multiplicity of infection of 30, and 10 respectively in a total volume of 1,5 ml. For scRNA-Seq experiments R848 (InvivoGen) was added at a final concentration of 1,2 ug/ml and pre-complexed 3p-hpRNA (LyoVec) was added with a concentration of 16 ng/ml. Infected/stimulated cells were incubated at 37°C and 5% CO_2 before being harvested by rinsing of with ice-cold PBS supplemented with 10 mM EDTA at 16 h for scRNA-seq or 1 h, 3 h and 18 h for proteomics.

scRNA-seq library preparation and sequencing

Primary cells were prepared and stimulated as described above. Afterward, each condition was hashtagged with TotalSeq-A antibodies (Biolegend) according to the manufacturer's protocol for TotalSeq-A antibodies and cell hashing with 10X Single Cell 3′ Reagent Kit v3.1. 50 μL cell suspension with 1x10^6 cells/ml were resuspended in staining buffer (2% BSA, Jackson Immuno Research; 0,01% Tween-20, Sigma-Aldrich; 1x DPBS, GIBCO). 1 μg unique TotalSeq-A antibody in 50 ul staining buffer was added to each sample and incubated for 20 minutes at 4°C. After the incubation 1,5 mL staining buffer was added and centrifuged for 5 minutes at 350 g and 4°C. Washing was repeated for a total of 3 washes. Subsequently, the cells were resuspended in an appropriate volume of 1x DPBS (GIBCO), passed through a 40 μm mesh (Flowmi^{TM} Cell Strainer, Merck) and counted, using a Neubauer counting chamber (Marienfeld). Cell counts were adjusted and hashtagged cells were pooled equally. The cell suspension was then loaded in the Chromium^{TM} Controller (10x). Single Cell 3′ reagent kit v3.1 was used for reverse transcription, cDNA amplification and library construction following the detailed protocol provided by 10x Genomics. Hashtag libraries were prepared according to the cell hashing protocol for 10x Single Cell 3′ Reagent Kit v3.1 provided by Biolegend, including primer sequences and reagent specifications. A Biometra Trio Thermal Cycler was used for amplification and incubation steps (Analytik Jena). Libraries were quantified by Qubit^{TM} 2.0 Fluorometer (ThermoFisher) and quality was checked using 2100 Bioanalyzer with High Sensitivity DNA kit (Agilent). Sequencing was performed in paired-end mode with S1 and S2 flow cells (2 × 50 cycles kit) using NovaSeq 6000 sequencer (Illumina).

Multi-epitope-ligand cartography (MELC)

Tissue preparation for MELC

Fresh frozen tissue was cut in 5 µm sections with a MH560 cryotome (ThermoFisher, Waltham, Massachusetts, USA) on cover slides (24 × 60 mm; Menzel-Gläser, Braunschweig, Germany) that had been coated with 3-aminopropyltriethoxysilane (APES). Samples were fixed for 10 minutes at room temperature (RT) with 2% paraformaldehyde (methanol- and RNase-free; Electron Microscopy Sciences, Hatfield, Philadelphia, USA). After washing, samples were permeabilized with 0.2% Triton X-100 in PBS for 10 min at room temperature and unspecific binding was blocked with 10% goat serum and 1% BSA in PBS for at least 20 minutes. Afterward, a fluid chamber holding 100 µL of PBS was created using ‘‘press-to-seal’’ silicone sheets (Life technologies, Carlsbad, California, USA; 1.0 mm thickness) with a circular cut-out (10 mm diameter), which was attached to the coverslip surrounding the sample.

MELC image acquisition

MELC image acquisition was performed as previously shown (Holzwarth et al., 2018; Pascual-Reguant et al., 2021). We generated the multiplexed histology data on a modified Toponome Image Cycler® MM3 (TIC) originally produced by MelTec GmbH & Co.KG Magdeburg, Germany (Schubert et al., 2006). The ImageCycler is a robotic microscopic system with 3 main components: (1) an inverted widefield (epi)fluorescence microscope Leica DM IRE2 equipped with a CMOS camera and a motor-controlled XY-stage, (2) CAVRO XL3000 Pipette/Diluter (Tecan GmbH, Crailsheim, Germany), and (3) a software MelTec TIC-Control for controlling microscope and pipetting system and for synchronized image acquisition. The MELC run is a sequence of cycles, each containing the following four steps: (i) pipetting of the fluorescence-coupled antibody onto the sample, incubation and subsequent washing; (ii) cross-correlation auto-focusing based on phase contrast images, followed by acquisition of the fluorescence images 3D stack (+/− 5 z-steps) ; (iii) photo-bleaching of the fluorophore; and (iv) a second auto-focusing step followed by acquisition of a post-bleaching fluorescence image 3D stack (+/− 5 z-steps). In each four-step cycle up to three fluorescence-labeled antibodies were used, combining PE, FITC and DAPI. After the sample was labeled by all antibodies of interest as described above, the experiment is completed. The antibodies used for multiplexed immunofluorescence histology of lung samples are listed in the Key resources table. The antibodies were stained in the indicated order.

Transmission electron microscopy (TEM)

Samples for TEM from two deceased patients that were tested positive for SARS-CoV-2 RNA via PCR were processed as described previously (Ochs et al., 2021). Briefly, samples were fixed with 3% formaldehyde 0.1 M cacodylate buffer for 12 h followed by changing the fixative for another 12 h. Smaller tissue samples were cut for TEM and fixed with 1.5% glutaraldehyde / 1.5% paraformaldehyde in 0.15 M HEPES buffer and post-fixed with 1% OsO4 for 1 h followed by 0.1% tannic acid for 0.5 h. Samples were finally dehydrated in ethanol and embedded in Epon. Ultrathin sections were counterstained with lead citrate and examined with a Leo 906 TEM (Zeiss, Oberkochen, Germany).

Proteomics

Sample preparation

Proteomics and phosphoproteomics were prepared using a strategy adapted from Mertins et al. (Mertins et al., 2018). Lysates were cleared via centrifugation (20,000 g, 15min, 4°C) and protein content in the supernatant was measured via BCA assay (Thermo Fisher Scientific, 23225). Samples were diluted to the same protein concentration with lysis buffer (8M urea, 150mM NaCl, 50mM Tris HCl pH 8.0, 1mM EDTA, cOmplete Protease Inhibitor Cocktail (Roche, 11697498001), Phosphatase Inhibitor Cocktail 2 and 3 (Sigma-Aldrich, P5726 and P0044)) and subsequently reduced and alkylated via incubation with 5mM DTT and 10mM iodoacetamide, respectively. Samples were diluted to 2M Urea with 50Mm Tris pH 8.0, then pre-treated with LysC (Wako Chemicals) at 1:50 (w/w) ratio for 2 hours at room temperature, then trypsin (Promega) was added at 1:50 (w/w) ratio and samples were digested overnight at room temperature. Digests were acidified with formic acid (FA) and centrifuged (20,000 g, 15min) to remove the precipitated urea. Approximately 30 µg of protein digest were desalted using STop-And-Go Extraction tips (Stage-tips) (Rappsilber et al., 2007). 3 disks of C18 (3M Empore) material were inserted in a 200 µL p pipette tip, activated via methanol which was subsequently washed away with a solution of 50% acetonitrile (ACN) and 0.1% FA, followed by a wash with 1% FA. Samples were loaded onto the Stage-tips and the retained peptides were washed twice with 0.1% trifluoroacetic acid (TFA) followed by a wash with 1% FA. Finally, peptides were eluted from the C18 material with a solution of 50% ACN / 0.1%FA. Desalted digests were dried and resuspended in 50mM HEPES pH 8.5. Peptide concentration was evaluated with BCA assay, and equal amounts of peptides were brought to the same concentration using 50mM HEPES pH 8.5. Prior to TMT labeling, we randomly assigned a TMTpro channel per experiment as follows: CoV2 1hr rep A- > 126; R848 1hr rep B- > 127N; IAV 1hr rep A- > 127C; CoV2 3hr rep A- > 128N; mock 1hr rep A- > 128C; CoV2 18hr rep A- > 129N; IAV 1hr rep B- > 129C; CoV2 18hr rep B- > 130N; R848 1hr rep A- > 130C; IAV 18hr rep B- > 131N; CoV2 3hr rep B- > 131C; IAV 18hr rep A- > 132N; IAV 3hr rep A- > 132C; CoV2 1hr rep B- > 133N; mock 1hr rep B- > 133C; IAV 3hr rep B- > 134N. Samples were then labeled using TMTpro reagents (Thermo Fisher Scientific; product number A44520, lot number UL297970) with a 1:10 peptide weight to TMT reagents weight ratio and approx. 17% ACN concentration, for 1 hour at room temperature, and the reactions were quenched by addition of 1M Tris pH 8.0 to a final concentration of 5mM. All the reactions were combined with equivalent peptide amounts and desalted via SepPak (Waters) tC18 column. The C18 material was first activated with 100% ACN, then the solvent was removed via washes with 50% ACN / 0.1% FA, followed by washes with 0.1% TFA. Samples were loaded onto the column and desalted via washes with 0.1% TFA followed by washes with 0.1%FA. Finally, TMT-labeled peptides were eluted from the C18 material with 50% ACN / 0.1% FA. The labeled peptides were dried and resuspended in high pH buffer A (5mM ammonium formate, 2% ACN) prior to offline high pH Reverse phase fractionation by HPLC on an Agilent 1290 Infinity II. The separation was performed on a XBridge Peptide BEH C18 (130Å, 3.5µm; 2.1mm x 250mm) column (Waters) on a multi-step gradient from 0 to 60% High pH buffer B (5mM ammonium formate, 90% ACN) 96 minutes long and collected in 96 fractions (1 fraction/min). The fractions were then manually pooled into 29 fractions as follows: the first 12 fractions were pooled together, while every other x fraction was pooled with x+28 and x+56. Of each pooled fraction approximately 1µg of peptide was subjected to mass spectrometric (MS) analysis for total proteome measurement. The remaining amounts were further pooled into 5 fractions and used as input for a phosphopeptide enrichment via immobilized metal affinity chromatography (IMAC), which was performed by the Bravo Automated Liquid Handling Platform (Agilent) with AssayMAP Fe(III)-NTA cartridges. The flow-through of the first IMAC enrichment was further pooled into 2 fractions and subjected to a second IMAC enrichment with the same system.

Liquid chromatography mass spectrometry

All mass spectrometry samples were online-fractionated on a EASY-nLC 1200 and acquired on a Q-Exactive HFx (Thermo Fisher Scientific) on a profile-centroid mode. Peptides were separated on a fused silica, 25cm long column packed in-house with C18-AQ 1.9µm beads (Dr. Maisch Reprosil Pur 120) kept to a temperature of 45°C. Mobile phase A consisted of 0.1% FA and 3% ACN in water, while mobile phase B consisted of 0.1% FA and 90% ACN. After quilibritrating the column with 5µl mobile phase A, peptides were separated with a 250µl/min flow on a 110min gradient: mobile phase B increased from 4% to 30% in the first 88 minutes, followed by an increase to 60% in the following 10 minutes, to then reach 90% in one minute, which was held for 5 minutes. For total proteome analysis, the MS was operated in data dependent acquisition, with MS1 scans from 350 to 1500 m/z acquired at a resolution of 60,000 (measured at 200 m/z), maximum injection time (IT) of 10ms and an automatic gain control (AGC) target value of 3×10^6. The 20 most intense precursor ion peaks with charges from +2 to +6 were selected for fragmentation, unless present in the dynamic exclusion list (30 s). Precursor ions were selected with an isolation window of 0.7 m/z, fragmented in an HCD cell with a normalized collision energy of 30% and analyzed in the detector with a resolution of 45,000 m/z (measured at 200 m/z), AGC target value of 10^5, maximum IT of 86 ms. For phosphoproteome analysis, the MS was operated using the same parameters with the exception of MS2 maximum IT that was set to 240ms.

ELISA

Supernatants of infected or stimulated monocytes were analyzed using DuoSet ELISA Kits for human CCL24, CCL8, IL-6, MCP-1, IL1-β, CXCL5 and IFNβ (R&D Systems) according to manufacturer's instructions. The concentrations were calculated with GraphPad Prism using the protein standard included in the ELISA kits.

QUANTIFICATION AND STATISTICAL ANALYSIS

MELC data analysis

Image pre-processing

All images were aligned by cross-correlation based on the reference phase contrast image taken at the beginning of the measurement. Afterward, each fluorescence MELC image was processed by background subtraction and illumination correction, based on the signal of the bleaching images (Schubert et al., 2006). In order to account for slice thickness, an "Extended Depth of Field" algorithm was applied on the 3D fluorescence stack in each cycle (Pertuz et al., 2013). Images were then normalized in Fiji (Schindelin et al., 2012), where a rolling ball algorithm was used for background estimation, edges were removed (accounting for the maximum allowed shift during the autofocus procedure) and fluorescence intensities were stretched to the full intensity range (16 bit = > 2^{16}). The 2D fluorescence images generated in this way were subsequently segmented and analyzed.

Cell segmentation and single-cell feature extraction

Segmentation was performed in a two-step process, a signal-classification step using Ilastik 1.3.2 (Berg et al., 2019) followed by an object-recognition step using CellProfiler 3.1.8 (Carpenter et al., 2006), as described elsewhere (Schapiro et al., 2017). Ilastik was used to classify pixels into three classes (nuclei, membrane, and extracellular matrix –ECM-) and to generate probability maps for each class. Classification of images regarding membranes and ECM was performed by summing up a combination of images, using markers expressed in the respective compartments, while only the DAPI signal was used to classify nuclei. The random forest algorithm (machine-learning, Ilastik) was trained by manual pixel-classification in a small region of each data-set (approx. 6% of the image). CellProfiler was subsequently used to segment the nuclei and membrane probability maps and to generate nuclei and cellular binary masks, respectively. These masks were superimposed on the individual fluorescence images acquired for each marker, in order to extract single-cell information, i.e., mean fluorescence intensity (MFI) of each marker per segmented cell.

Data analysis

Mean fluorescence intensities were normalized to the full 16-bit range in Fiji, brought to a 0 to 1 scale in CellProfiler and transformed using the hyperbolic arcsine (cofactor/scale argument = 0.2) prior to clustering analysis. All CD45 expressing cells were selected and data was imported into R version 3.6.3 (https://cran.r-project.org/src/base/R-3/). A total of 22 markers per cell were included in the panel, normalized by the total fluorescence intensity across all markers per cell, scaled (z-scores), and used to compute a two dimensional Uniform Manifold Approximation and Projection (McInnes et al., 2018) embedding using the R package uwot (n,neighbors = 50, n_epochs = 500, n_trees = 100, init = "pca," min.dist = −0.1, metric = "euclidean"). For unsupervised clustering, a shared nearest neighbor tree was calculated using the scran (Lun et al., 2016) function buildSNNGraph. The Leiden algorithm (Traag et al., 2019; Kelly, 2019) was applied on the graph object with the resolution parameter 0.9 and seed 1993. Unless otherwise stated, all algorithms were used with default settings.

For Macrophage-Collagen IV distance measurements Collagen 4 layer .tif-files were converted into raster objects using the r library "raster." Euclidean distances of Cellprofiler cell coordinates to the closest Collagen IV structures were calculated with the R package "sf." Cells were grouped into the groups "in collagen," "adjacent to collagen," and "out of collagen" by clustered pixel distances that best represented the visual localizations over all fields of view. Proportions of CD163$^+$ and CD163$^-$ macrophages per field of view and localization were computed. Population mean rank differences were tested by paired, two-sided Wilcoxon signed rank test adjusted for multiple testing by the Bonferroni correction.

Single-cell RNA-seq data analysis

Cell Ranger

The Cell Ranger version 3.0.1 software suite was obtained from 10x Genomics. Raw sequencing data was first de-multiplexed and quality-checked using the Cell Ranger 'mkfastq' script. For all sample libraries, alignment and transcript quantification was performed with the standard Cell Ranger 'count' script against a custom made genome reference containing the GRCh38 human genome assembly and the SARS-CoV-2 genome. Multiple datasets belonging to the same experiment were combined with the Cell Ranger 'aggr' script.

Quality control

Count matrices were loaded into R and low quality transcriptomes (e.g., dead or ruptured cells) were removed. Thresholds for the number of genes, number of unique RNA molecules (UMI) and percent mitochondrial genes for each dataset are available in Table S3. In addition, cells were removed after clustering (Waltman and van Eck, 2013) based on low amount of genes and high percentage of mitochondrial genes per cell across clusters (code is available in the GitHub repository).

Normalization and feature selection

Expression matrices were separated from antibody-derived counts and treated separately. For the BAL datasets, the Seurat functions NormalizeData, FindVariableFeatures (n.features = 3000) and ScaleData were used. For the APC datasets, the Seurat function SCTransform (variable.features.n = 3000) was used. Viral mRNA counts were stored in a separated assay.

Demultiplexing of experimental conditions

Hashtag counts were normalized using the Seurat function NormalizeData(method = "CLR"). Cutoffs were chosen by k-means clustering of cells for each hashtag. Uniquely called hashtags were annotated with the experimental condition, double stained (doublet) and unstained (negative) cells were removed (code is available on the GitHub repository).

Demultiplexing of donors

Donors of PBMCs were demultiplexed using the souporcell algorithm (Heaton et al., 2020) based on the BAM files & barcodes exported by CellRanger count.

Dimensional reduction and clustering

Principal component analysis (PCA) was calculated for each dataset using the Seurat function RunPCA based on the highly variable genes. Batch correction (if applied) was performed during the PCA using the batchelor (Haghverdi et al., 2018) function fastMNN using patient identifiers as batch covariate. Uniform manifold approximation and projection (McInnes et al., 2018) embeddings were computed based on differing numbers of principal components using the Seurat function RunUMAP. Nearest neighbor graphs were constructed based on differing numbers of principal components using the Seurat function FindNeighbors. Clusters were identified using the SLM or Leiden algorithms with different resolution parameters using the Seurat function FindClusters (method = igraph) based on the nearest neighbor graph. The specific settings used for each dataset are available in Table S3.

scRNA-seq analysis of BAL data from recovering patients

Cellranger output files were loaded into R (4.0.3) using the Read10X function from Seurat. Low quality transcriptomes were removed from further analysis by filtering for cells with at least 150 unique features and a mitochondrial gene percentage lower than 30%. Subsequently, normalization, variable feature detection and scaling of the data were performed by using the Seurat implemented functions LogNormalization, FindVariableFeatures function (n.features = 2000) and ScaleData. For dimensional reduction, the RunPCA function was used based on the highly variable genes. To account for the batch-effect observed by sample, the "harmony" algorithm (RunHarmony function) (Korsunsky et al., 2019) was applied. For two-dimensional data visualization, UMAP was performed based on the first 50 principal components of the "harmony" data reduction. Subsequently, the cells were clustered using the Louvain algorithm based on the first 30 "harmony" dimensions with a resolution of 0.4. Resulting clusters were annotated for cell types using literature-based markers. For analysis of the monocyte/Mϕ compartment, the cells in the respective clusters were subsetted and basic steps (NormalizeData, FindVariableFeatures, ScaleData, RunPCA, RunHarmony, RunUMAP) and Louvain clustering with a resolution of 1.5 was performed. Resulting clusters were annotated to the respective monocyte/Mϕ subtype according to the previously identified markers.

Gene set enrichment

Scores for gene expression programs were computed using the Seurat function AddModuleScore (seed = 1993). The top 50 genes of different reference datasets were used, if the list contained less than 50 genes all of them were used as input. Statistical significance of population shifts between clusters were assessed by pairwise, one-sided (alternative = "greater"), two-sample Wilcoxon rank sum test (wilcox.test) of each cluster compared to the average.

Overrepresentation analysis of genes between two gene sets was performed by one-sided (alternative = "greater") Fisher's exact test (fisher.test). P values were adjusted for multiple comparisons (p.adjust). All functions were used with default settings unless specified otherwise.

Differential gene expression (DE) analysis

Statistical analysis of differential gene expression was performed using the scran function findMarkers (pval.type = "some," test.-type = "wilcox," direction = "up"), the block argument was used for patients in the BAL dataset and donors in the stimulated monocytes. Genes with a FDR below 10^{-15} were considered differentially expressed.

Trajectory inference and pseudotime

Trajectories with pseudotime were computed using the slingshot (Street et al., 2018) R package (version 1.4.0). Input were the UMAP coordinates and cluster annotations (Settings: allow.breaks = TRUE).

Transcription factor enrichment analysis (ChEA3)

Transcription factor (TF) enrichment was performed based on the ChEA3 (Keenan et al., 2019) query in R. The differentially expressed genes (FDR < 1e-15) for each cluster were used as input. The mean rank score was used for TF selection (cutoffs are specified in the figure legends and TFs highlighted in Tables S2 and S5) as recommended by the authors.

Data visualization

Plots were generated in R (version 3.6.3) using custom code based on the ggplot2 version 3.3.2 (https://cran.r-project.org/web/packages/ggplot2/index.html) and pheatmap version 1.0.12 (https://cran.r-project.org/web/packages/pheatmap/index.html). Data wrangling was performed using the dplyr (1.0.2), tidyr (1.1.2) and stringr (1.4.0) packages.

Single-nucleus RNA-seq

Refined cell type annotation was performed on the macrophage, and mesenchymal (fibroblast, SMC, and pericyte) clusters. They were each subsetted from the entire dataset and re-processed to find underlying subclusters by repeating the previous steps of CCA, PCA, UMAP, and clustering on the selected cells. Markers from the scRNA-seq BAL dataset were used to annotate the different macrophages, while mesenchymal markers from previously published scRNA-seq lung atlases were applied to identify fibroblasts, SMC, and pericytes (Adams et al., 2020; Reyfman et al., 2019).

Potential cell-cell interactions between the different subclusters of macrophages and mesenchymal cells were identified using CellChat v0.5.5 (Jin et al., 2021). Differential interactions between patient groups were calculated using mergeCellChat() and compareInteractions() functions iteratively per comparison. Intercellular communication networks are weighted directed graphs composed of significant interactions between cell groups, where "interaction strength" is defined as the communication probability of the computed networks (Jin et al., 2021). Differentially enriched interaction pathways were determined for early and late postmortem SARS-CoV-2 cases by using the rankNet() function. Significance was determined by the function by performing a paired Wilcoxon test. Information flow is the overall communication probability, where it is the summation of the probability among all pairs of cell groups in the inferred network.

Data integration

Single cells from this study and three previous studies were integrated into a single embedding. Briefly, macrophage cells from (Adams et al., 2020) and (Morse et al., 2019) were selected using their published annotations (Adams: all cells with the keyword "Macrophage" in the column "Subclass_Cell_Identity"; Morse: SPP1hi Macrophages, and FABP4hi Macrophages; (Bharat et al., 2020): AM1/2, MoM1/2/3, and Monocytes), and integrated with monocyte-derived cells annotated in this work (annotations: FCN1 Mono, Mono/Mϕ, *SPP1*/LGMN-Mϕ, *SPP1*/*TREM2*-Mϕ, *INHBA*-AMϕ, and Prolif-AMϕ, Low Quality). To minimize cell number bias possibly linked to high cell numbers in Adams, we downsampled cells from Adams to maintain an equal number of IPF, control and COPD cells (30,159 in each category). This gives a total of 90,477 cells from Adams (18 COPD, 32 IPF and 28 control), 17,551 from Morse (6 IPF and 3 control), and 22,810 cells from (Bharat et al., 2020) (1 case and 2 postmortem biopsies of patients that underwent lung transplantation after COVID-19 infections), with macrophage annotations and balanced condition, respectively. These cells were integrated with 7,503 monocyte-derived cells from this study. The integration of the three datasets was done using single-cell variational inference (scVI) as implemented scvi-tools (version 0.6.7) (Lopez et al., 2018), using patient identifiers in Adams (one per patient), sample identifiers in Morse (one or more per patient), and patient identifiers in (Bharat et al., 2020) (one per patient) as batch covariates, respectively. scVI was chosen for integration since it was a scalable top performer in a recent batch integration benchmark (Luecken et al., 2020). The network architecture had the following parameters: n_latent = 30, n_hidden = 128, and n_layers = 2. We trained this network for 400 epochs and used the latent representation as a low dimensional embedding to compute a k-nearest neighbor graph for the integrated cells with k = 15, using SCANPY (Wolf et al., 2018) (version 1.6.0; anndata version 0.7.4) and calculate a UMAP layout (McInnes et al., 2018).

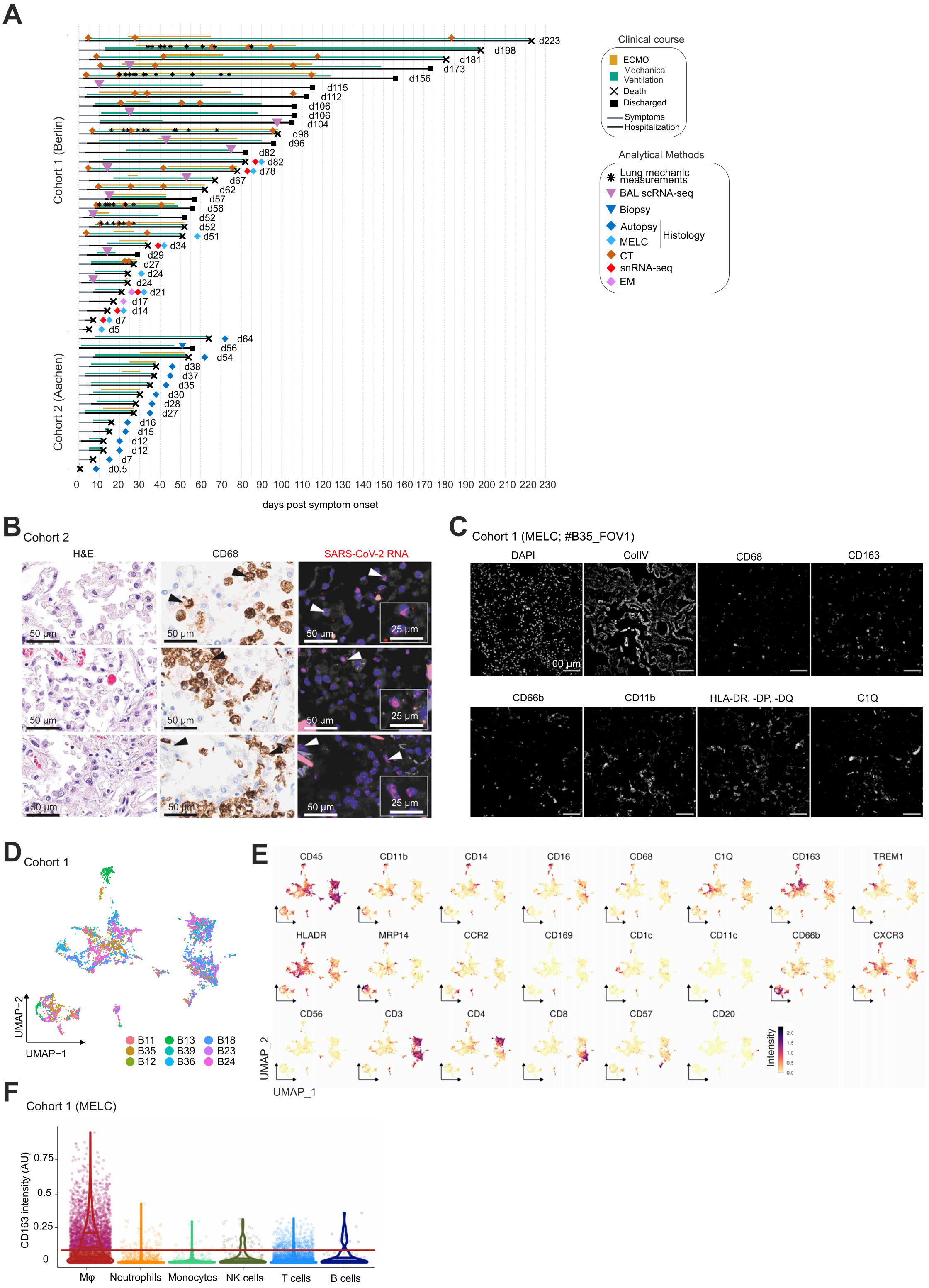
To assess the similarity of monocyte-derived macrophages from patients with COVID-19 to these cells from other diseases, we mapped condition labels (i.e., fibrosis, COPD, and control) from our reference to the COVID-19 monocyte-derived macrophages. To reach a robust result, we treated the (Bharat et al., 2020) study as a replicate of this study and thus repeated the same experiment with each dataset. Thus, cells from (Bharat et al., 2020), were held out for the analysis of cells from this study, and vice versa for the analysis of cells from (Bharat et al., 2020). Specifically, we mapped condition labels onto monocyte-derived macrophages from either COVID-19 study via local majority voting: briefly, based on the kNN graph built from the joint embedding, all nearest neighbors from other non-COVID-19 studies were retrieved for each monocyte-derived cell from the study. The most common condition label (IPF/control/COPD) across these neighbors was then projected onto the cell. We did not consider the monocyte-derived macrophage ''Low Quality'' category for this analysis, to allow easier interpretation. Furthermore, as cells from COPD were underrepresented in the integrated dataset despite previous down-sampling, we did not consider cells assigned to COPD for further analysis as these would be underrepresented purely due to the background distribution. To ensure a completely separate analysis of both COVID-19 studies, we annotated cells from our study after removing (Bharat et al., 2020) cells from the integrated embedding and recalculated nearest neighbors, and vice versa for mapping cells in (Bharat et al., 2020).

To assess the enrichment of specific condition labels for each monocyte-derived macrophage subtype, we calculated the enrichments of each monocyte-derived macrophage and condition combination using 2 × 2 contingency tables. Significance of observed odds ratios were calculated using Fisher's exact tests, one-tailed, and adjusting p values using Benjamini Hochberg's procedure (Benjamini and Hochberg, 1995).

Proteomics data analysis

RAW files were analyzed using MaxQuant (Tyanova et al., 2016) v1.6.10.43, where TMTpro was manually included as a fixed modification and quantitation method. Correction factors for each TMT channel were added to account for channel spillage and minimum reporter precursor intensity fraction was set to 0.5. The MS scans were searched against human, influenza A and SARS-CoV-2 uniprot databases (Jan 2020, Apr 2020 and Mar 2020 respectively) using the Andromeda search engine. FDR was calculated based on searches on a pseudo-reverse database and set to 0.05. The search included as fixed modifications carbamidomethylation of cysteine and as variable modifications methionine oxidation, N-terminal acetylation, and asparagine and glutamine deamidation. Trypsin/P was set as protease for in-silico digestion. Total proteome and IMAC-enriched phosphopeptides samples were analyzed in the same MaxQuant run in separate parameter groups with the same settings, except for the IMAC-enriched samples also Phospho (STY) was added as variable modification. Contaminants, hits in the reverse database, only identified by modified site and identified by less than two peptides of which one unique were removed from the ProteinGroups result table. Phosphosites were filtered by hits in the reverse database, potential contaminants and sites with localization probability lower than 50%. Differences in protein intensities across the samples were evaluated using an ANOVA test and results were filtered for Benjamini-Hochberg adjusted p values lower than 5%. Proteins passing this filtering were clustered using fuzzy-c-means clustering. Gene Set Enrichment Analysis was performed with the R GSEA suite (v2.0) of the Broad Institute using the Molecular Signature Database (MSigDB, v7.0) with the ontology and immunologic gene set collections. Data analysis was done using custom scripts and with the following packages: rawDiag (Trachsel et al., 2018) and data.table (Dowle and Srinivasan, 2019). Proteomic data was mapped to three different gene signatures and analyzed using the stat_ecdf function as implemented in the stats R-package. The signatures we tested originated from: genes identified as upregulated from scRNA-seq of macrophages in patients with pulmonary fibrosis (listed with a FC > 0) in supplementary_table_3_human_sc_de.xlsx table from (Reyfman et al., 2019); *SPP1*-expressing macrophages identified in idiopathic pulmonary fibrosis (Morse et al., 2019); genes characterizing IPF-expanded macrophages, identified from scRNA-seq analysis of patients with idiopathic pulmonary fibrosis, taken from Figure 3A of (Ayaub et al., 2021). All gene set distributions were tested for average upregulation for each time point compared to all other quantified proteins using a one-sided wilcoxon rank sum test using the wilcox.test function implemented in R (Version 3.6.3).

Supplemental figures



(legend on next page)

Figure S1. Study cohorts and (immuno-)histological analysis of lung tissues, related to Figure 1

(A) Schematic overview of all patients enrolled in the two cohorts. Clinical characteristics, course of disease, treatments, analysis time points, and outcomes are indicated.

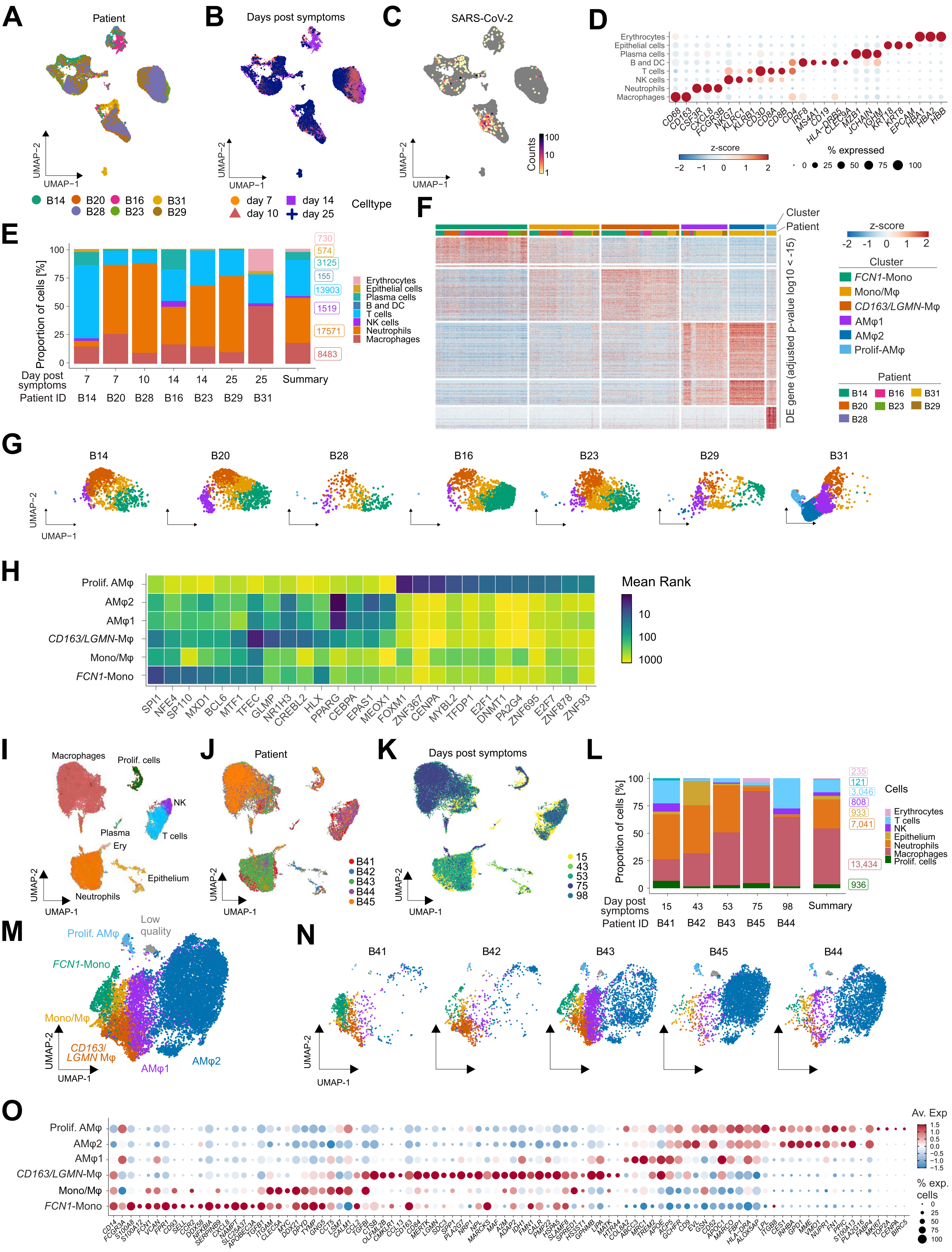
(B) Consecutive histological sections of COVID-19 lung tissue showing H&E (left), CD68 (middle) and SARS-CoV-2 RNA (right; scale bar, 50 μm; insert scale bar, 25 μm).

(C) MELC analysis of lung autopsy tissue showing collagen and immune cell staining.

(D) UMAP embedding as shown in Figure 1F color-coded by donor.

(E) UMAP embedding as in Figure 1F, color-coded arcsin-transformed mean fluorescence intensity across all epitopes measured by MELC.

(F) CD163 fluorescence intensity across the different cellular populations identified by MELC as presented in Figure 1F, every dot represents one cell. The line indicates the threshold of CD163+ and CD163- cells.



(legend on next page)

Figure S2. Monocyte/macrophage transcriptional profiles in BAL, related to Figure 2

(A–C) UMAP embedding (as in Figure 2A, dataset 1) of BAL scRNA-seq transcriptomes color-coded according to the patients of origin (A), by sampling time after symptom onset (B), and SARS-CoV-2 mRNA molecule counts (C).

(D) Dot plot displaying the expression of canonical marker genes delineates the cell types identified in BAL (Figure 2A). Dot size shows the percentage of cells with any mRNA counts, color shows the z-scores of log-normalized expression.

(E) Cellular composition of BAL fluid across patients by cell type according to scRNA-seq. Bar height shows proportion in percent, labels show the real cell numbers, color indicates the cell type. Summary shows average across patients.

(F) Heatmap displaying differential expressed (DE, FDR < 10e-15) genes between macrophage populations (as in Figure 2B) and across the different patients analyzed.

(G) UMAP as Figure 2B split by patient, color indicates macrophage clusters as in Figure 2B.

(H) Heatmap showing the mean rank of ChEA3 transcription factor enrichment. Clusters (y axis) as in Figure 2B, transcription factors (x axis) ordered by cluster and mean rank. Input to ChEA3 were the DE genes shown in F, TFs were selected by mean rank < 30.

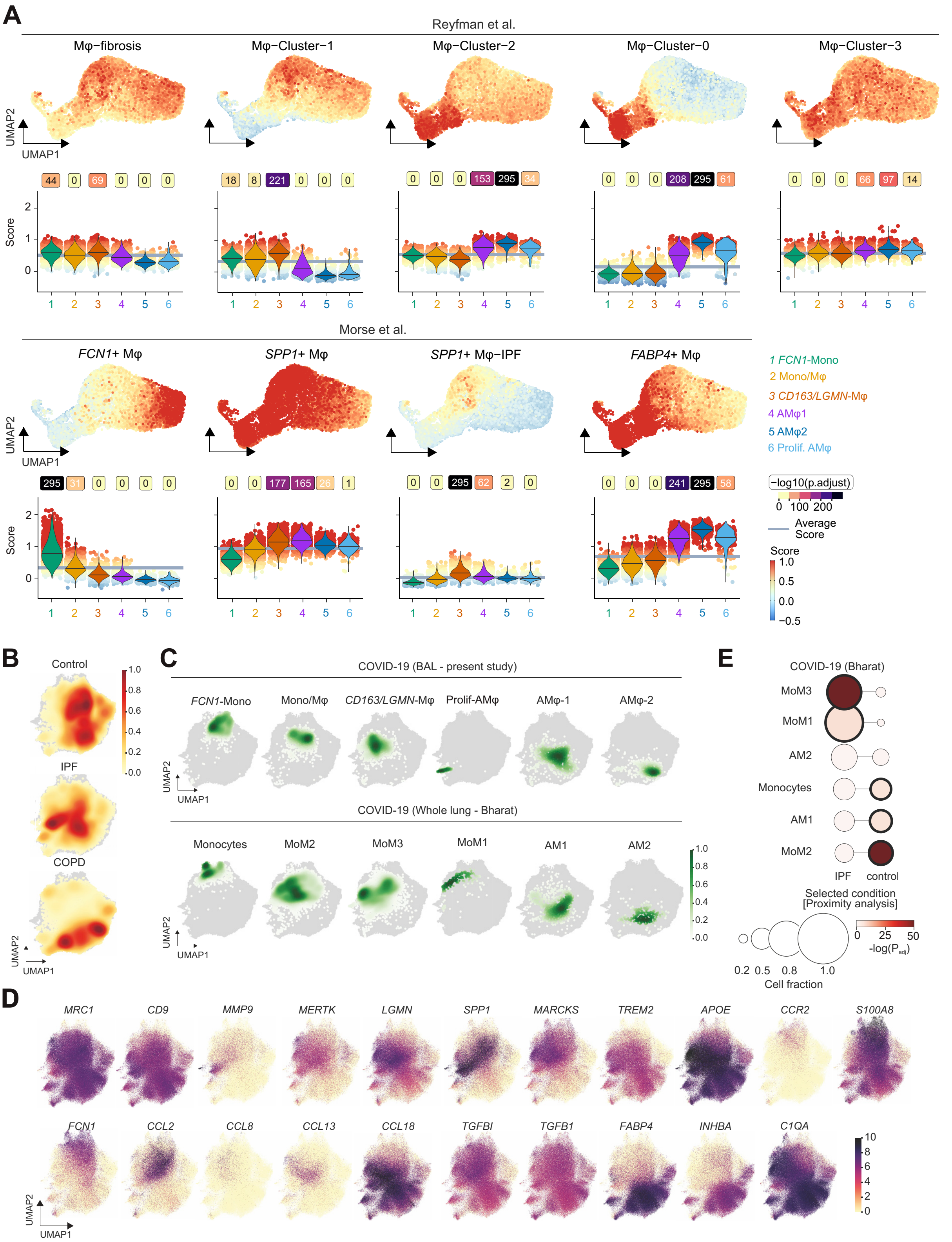
(I–K) UMAP embedding of 26,554 single-cell transcriptomes in the BAL fluid of severe COVID-19 patients at late stage of disease (dataset 2), color-coded according to identified cell types using canonical markers (I), patients of origin (J) and sampling time after symptom onset (K).

(L) Cellular composition of BAL fluid (dataset 2) across patients by cell type according to scRNA-seq. Bar height shows proportion in percent, labels show the total cell numbers, colors indicate the cell type. Summary shows average across patients.

(M) UMAP embedding of 12,712 transcriptomes of monocytes/macrophages in (I). Cell subtype labels were defined by cluster specific expression of previously identified BAL monocyte/macrophage markers (Figure 2F) (Mono; Monocytes, Mono/Mϕ; Monocyte-derived macrophages, AMϕ; Alveolar macrophages). Low quality refers to a cluster of cells with very high mitochondrial marker gene expression.

(N) UMAP from (M) split by patient, colors indicate macrophage clusters.

(O) Dot plot showing the previously identified monocyte/macrophage markers as in Figure 2F for the cell subtype labels from (M). Dot size shows the percentage of cells per cluster, color shows average expression of log-normalized mRNA counts.



(legend on next page)

Figure S3. Gene set enrichment analysis and data integration analysis with reference datasets, related to Figure 3

(A) Signature module scores of monocyte-macrophage clusters associated to idiopathic pulmonary fibrosis identified in two publically available datasets (Adams et al., 2020; Morse et al., 2019; Reyfman et al., 2019) projected on the UMAP embedding (top), and plotted as violin plots (bottom) across the clusters of monocyte-macrophage clusters of BAL scRNA-seq (annotation in Figure 2B). Violin plots are filled with color displaying cluster identity as in Figure 2B. Boxes above the violins show negative log10 transformed adjusted p values (one-sided Wilcoxon test compared to average). The lines in the violin plots represent the median of the respective scores per cluster.

(B) UMAP with kernel density overlay showing the density of cells from each condition (Control, IPF, and COPD) in the embedding (related to Figure 3D). Darker red indicates higher relative fractions of those cells in that UMAP region.

(C) Cell population density of macrophage clusters identified in this study (top) and in Bharat et al. (2020) (bottom). Kernel density overlay on UMAP embedding as in Figure 3D, color intensity shows relative fraction of cells.

(D) Marker gene expression projected on the UMAP of COVID-19/lung diseases integration analysis as presented in Figure 3D. Color shows normalized gene counts in ln(CPM+1). CPM: counts per million.

(E) Proximity analysis shows similarity of macrophage populations in COVID-19 (Bharat et al., 2020) to those in IPF and healthy patients (control). Circle size shows cell fraction, color codes the -log10 transformed adjusted p values, and bold black circle indicates statistical significance (adjusted $p < 0.0001$) (Fisher exact test, one-tailed with Benjamini-Hochberg correction).

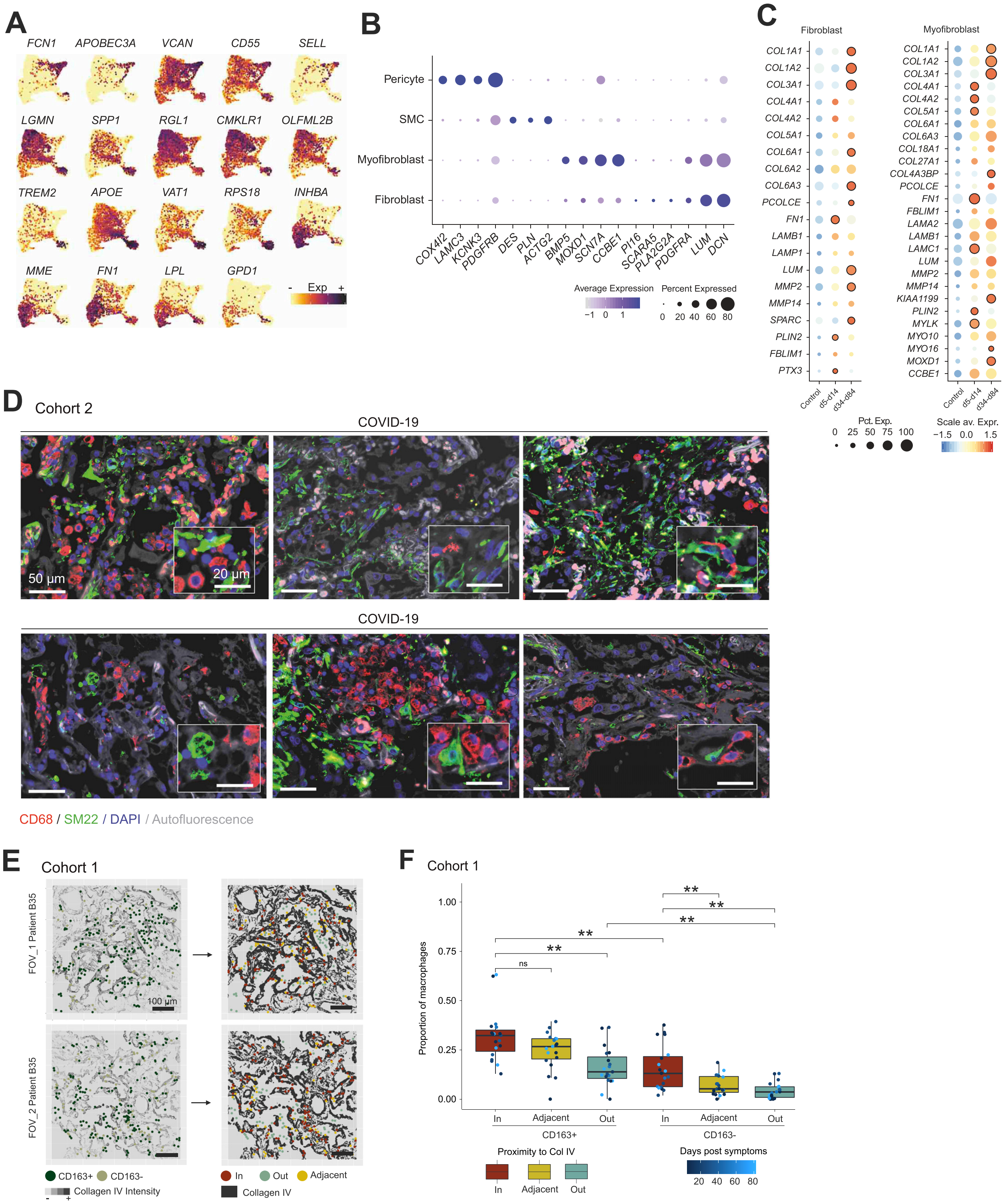


Figure S4. Macrophages-fibroblast interactions in COVID-19 lungs, related to Figure 4

(A) Marker gene expression delineates the macrophage embedding in Figure 4A. Color shows the normalized mRNA counts.

(B) Dot plot showing marker genes used to annotate the fibroblast, SMC and pericyte subclusters. Related to Figure 4A (right).

(legend continued on next page)

(C) Dot plot depicting scaled average expression of profibrotic factors in fibroblasts and myofibroblasts split according to control and disease duration. Scaled expression levels are color coded and the percentage of cells expressing the gene is size coded. Significant differences between early (d < 30) and late (d > 30) patients are highlighted by a black circle. Genes highlighted in early/late patients indicate th e condition where the gene is upregulated.

(D) Autopsy lung tissue reveals close association between macrophages (CD68, red) and fibroblasts (SM22, green) in COVID-19 compared to control (left). Cell nuclei are stained with DAPI (blue), autofluorescence is visible in faint gray (Scale bar, 50 μm; Insert scale bar, 20 μm).

(E) Analysis of MELC-imaging displayed in Figure 4G. (Left) Center coordinates of CD163+ (blue) and CD163- (red) macrophage localizations in respect to collagen IV staining. (Right) Segregation of macrophages into localization areas named '*in collagen IV*', '*adjacent to collagen IV*' or '*outside of collagen IV*'.

(F) Proportions of CD163^{+} and CD163^{-} macrophages per field of view of analyzed autopsy tissue localized 'in', 'adjacent' or 'outside' of collagen IV structures (** = Bonferroni corrected p value < 0.01, paired two-sided Wilcoxon signed rank test).

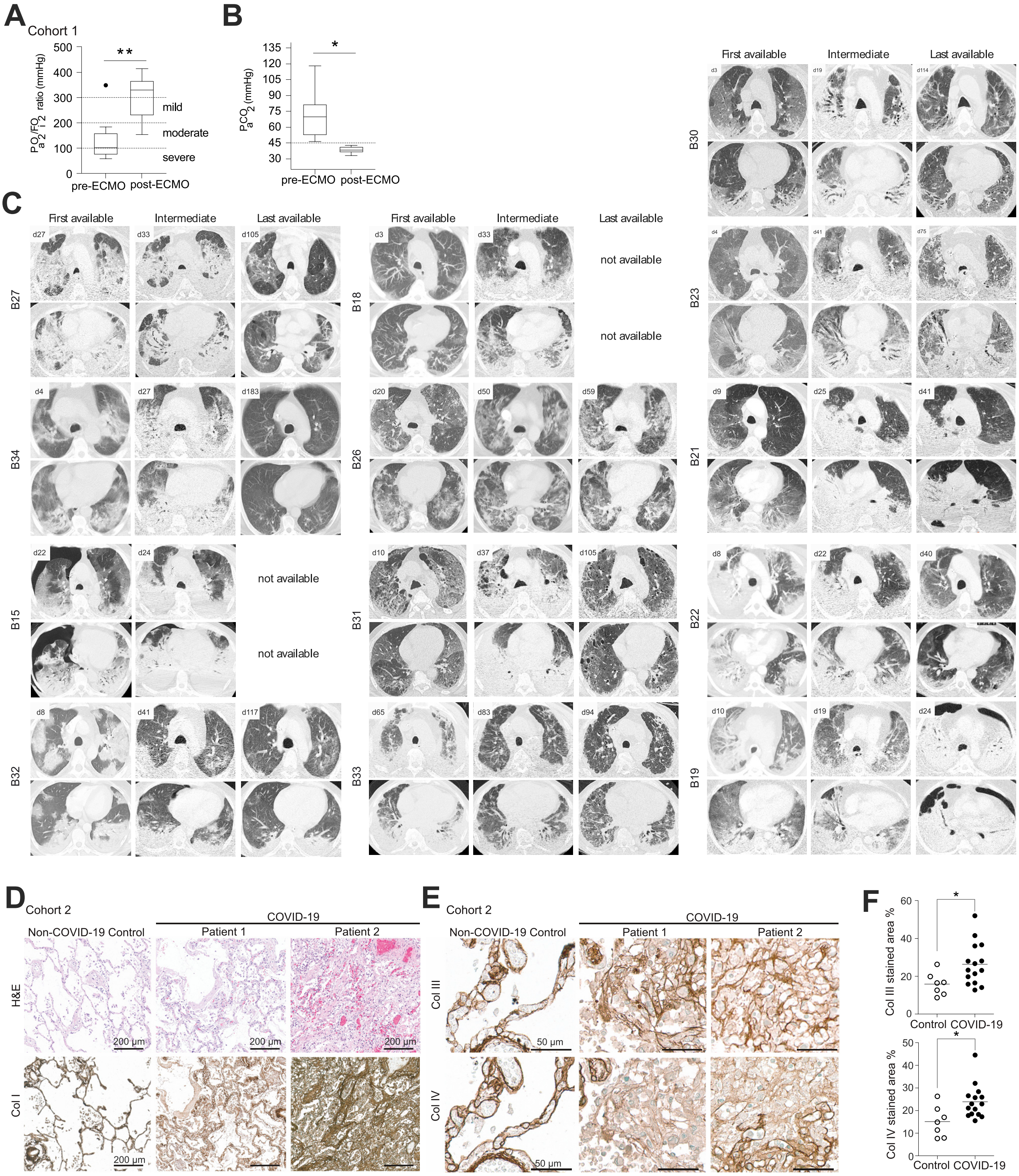

A
Cohort 1
**
P_aO_2/F_IO_2 ratio (mmHg)
500
400
300
200
100
0
mild
moderate
severe
pre-ECMO
post-ECMO
B
*
P_aCO_2 (mmHg)
135
120
105
90
75
60
45
30
pre-ECMO
post-ECMO
C
First available
Intermediate
Last available
B27
d27
d33
d105
B34
d4
d27
d183
B15
d22
d24
not available
not available
B32
d8
d41
d117
B18
d3
d33
not available
not available
B26
d20
d50
d59
B31
d10
d37
d105
B33
d65
d83
d94
B30
d3
d19
d114
B23
d4
d41
d75
B21
d9
d25
d41
B22
d8
d22
d40
B19
d10
d19
d24
D
Cohort 2
Non-COVID-19 Control
COVID-19
Patient 1
Patient 2
H&E
Col I
200 µm
E
Cohort 2
Non-COVID-19 Control
COVID-19
Patient 1
Patient 2
Col III
Col IV
50 µm
F
*
Col III stained area %
60
40
20
0
Control
COVID-19
*
Col IV stained area %
50
40
30
20
10
0
Control
COVID-19

Figure S5. CT imaging and histopathology analysis of severe COVID-19-associated ARDS, related to Figure 5

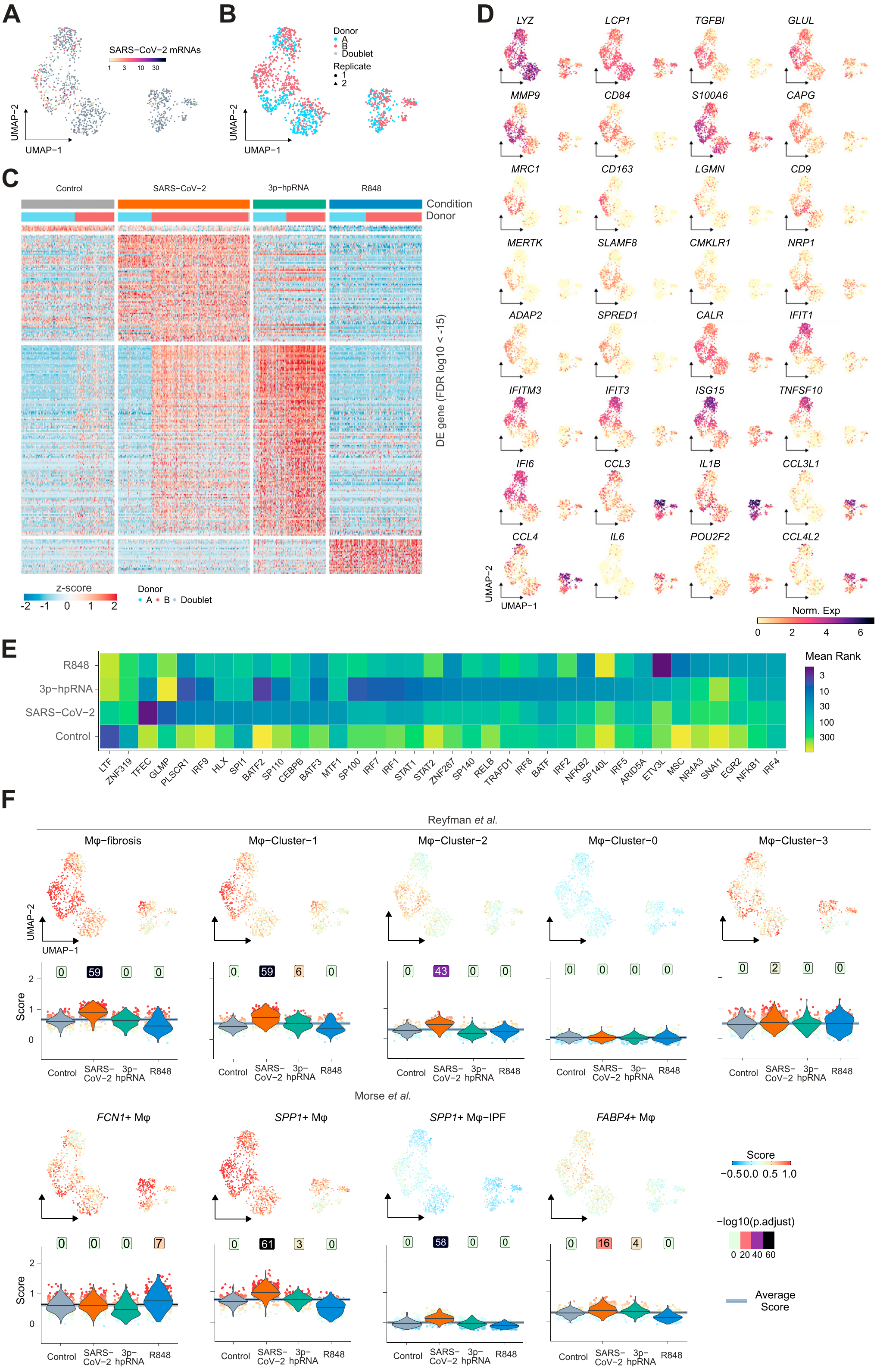
(A) P/F ratio (horowitz index) before and after vvECMO. ARDS severity is indicated by dashed lines. Statistical significance determined by paired t test (*p < 0.05; **p < 0.01).

(B) Arterial CO_2 partial pressure before and after initiation of vvECMO therapy. Upper limit of normal pCO2 range is depicted by a dashed line. Statistical significance determined by Mann Whitney Test (*p < 0.05; **p < 0.01).

(C) Representative computed tomography (CT) images of the apical (top row) and basal (bottom row) lung from 13 additional COVID-19 patients (cohort 1). Columns indicate the first, intermediate and last available images.

(legend continued on next page)

(D) Low power images of consecutive histological sections of autopsy lung tissue of fatal COVID-19 compared to control stained with H&E and chromogenic immunohistochemistry against collagen I. Scale bars represent 200 μm.
(E) High power images of consecutive histological sections (same field of view of Figure 5C) of autopsy lung tissue of fatal COVID-19 compared to control stained with chromogenic immunohistochemistry against collagen III and IV. Scale bars represent 50 μm.
(F) Quantification of collagen III and IV stained area in histological sections. Dots represent autopsy cases, significance of population shift of COVID-19 compared to control assessed by Mann Whitney Test (*p < 0.05; **p < 0.01; ***p < 0.001).



(legend on next page)

Figure S6. Monocyte gene expression after stimulation with SARS-CoV-2, 3p-hpRNA, and R848

(A) SARS-CoV-2 mRNA counts projected onto the UMAP embedding (Figure 6B).

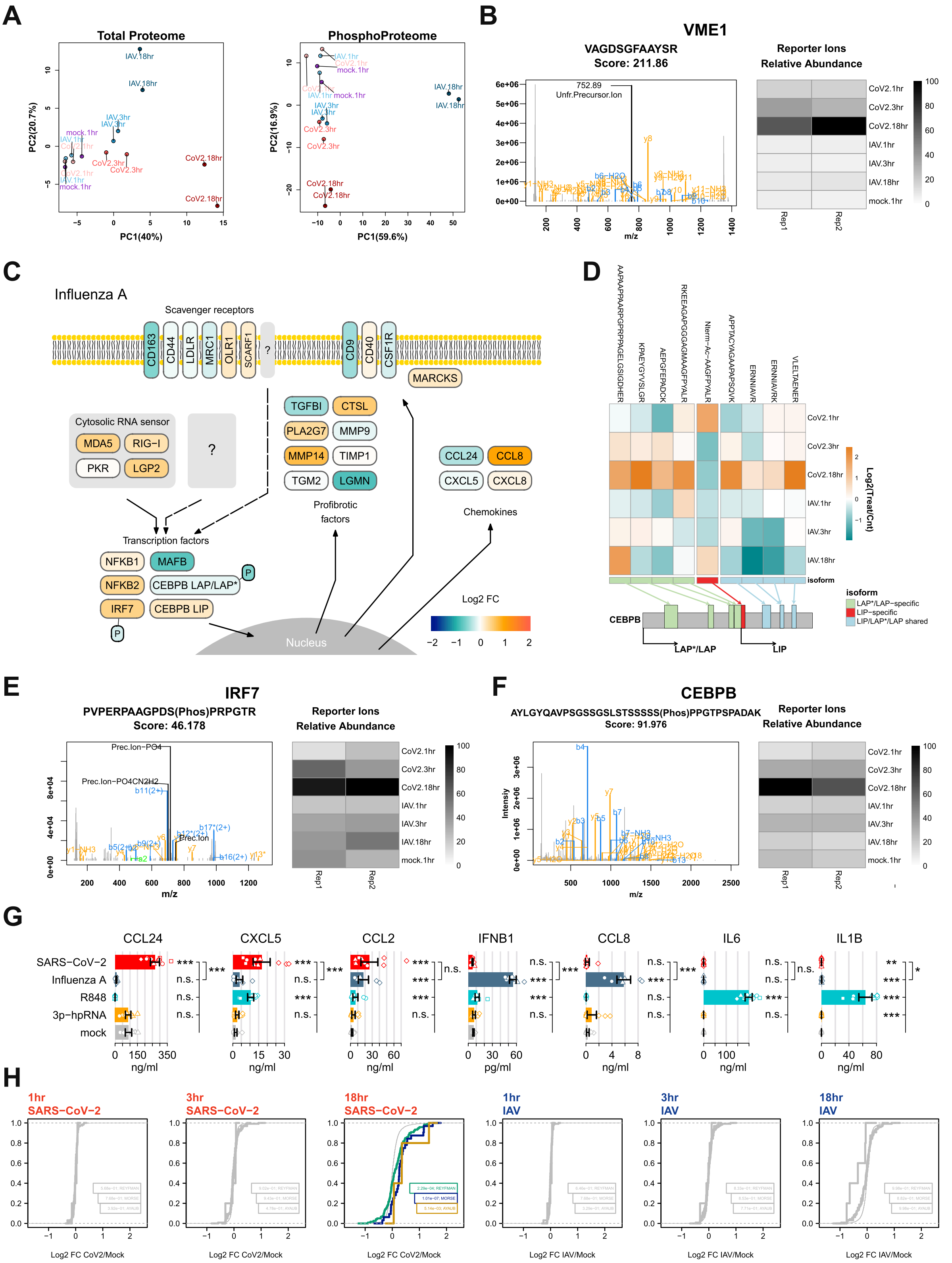
(B) Transcriptomes derived from two donors (indicated in blue and red) and two technical replicates (circles and triangles) are indicated in the UMAP embedding corresponding to Figure 6B.

(C) Heatmap displaying z-scores of log-normalized mRNA counts across all stimulation conditions. Differential expression (DE) cutoff was set at FDR of 1e-15.

(D) Marker gene expression projected onto the UMAP embedding as in Figure 6B.

(E) Heatmap showing the mean rank of ChEA3 transcription factor enrichment. Clusters (y axis) as in Figure 6B, transcription factors (x axis) ordered by cluster and mean rank. Input to ChEA3 were the DE genes shown in Figure S6C, TFs were selected by mean rank < 35.

(F) Signature module scores of IPF-associated monocyte/macrophage clusters derived from two published datasets (Morse et al., 2019; Reyfman et al., 2019) projected onto the UMAP embedding (top), and plotted as violin plots (bottom) across the clusters of stimulated monocytes (annotation in Figure 6B). Negative log10-transformed adjusted p values (one-sided wilcoxon test compared to average) are displayed above violins. Lines indicate median scores per cluster.



(legend on next page)

Figure S7. Quantitative shotgun proteomics and phosphoproteomics of SARS-CoV-2- and IAV-infected monocytes, related to Figure 7

(A) Principal component analysis of proteome and phosphoproteome for SARS-CoV-2, IAV and mock infection.

(B) Annotated MS2 spectrum of one peptide identified from SARS-CoV-2 M protein (left) and heatmap representing the TMT reporter ion relative intensities for the specified peptide (right).

(C) Schematic presentation of selected proteins involved in the inflammatory response pathways in monocytes, color-coded by log2-fold changes (IAV-infection versus control, 18h time point).

(D) Heatmap for all CEBPB identified peptides (top panel) and schematic representation of peptide location within the CEBPB sequence (bottom panel).

(E) Annotated MS2 spectrum of the phosphopeptide identified from IRF7 (left) and heatmap representing the TMT reporter ion relative intensities for the specified peptide (right).

(F) Annotated MS2 spectrum of the phosphopeptide identified from CEBPB (left) and heatmap representing the TMT reporter ion relative intensities for the specified peptide (right).

(G) Secretion of selected proteins quantified by ELISA. Bars represent the mean across all corresponding measurements. Error bars represent the standard deviation. Symbols depict donor-specific measurements. All experiments were tested against mock for significance (one-sided t test, on log transformed data). Differences between SARS-CoV-2 and IAV-stimulated cells were tested for significance using a two-sided t test. Significance reported in the figure corresponds to Benjamini-Hochberg adjusted p values of: *; $p < 10\%$, **; $p < 5\%$, ***; $p < 1\%$.

(H) Empirical cumulative distributions of gene sets depicted in Figure 7F. Log2-fold-change distributions of the gene sets were tested against all other proteins by one-sided Wilcoxon signed-rank tests. p values are depicted next to each distribution.

Journal of
Innate Immunity

Research Article

J Innate Immun
DOI: 10.1159/000519306

Received: May 18, 2021
Accepted: August 23, 2021
Published online: November 11, 2021

Dysregulated Immunometabolism Is Associated with the Generation of Myeloid-Derived Suppressor Cells in *Staphylococcus aureus* Chronic Infection

Oliver Dietrich[a] Alexander Heinz[b] Oliver Goldmann[c] Robert Geffers[d]
Andreas Beineke[e] Karsten Hiller[b, f] Antoine-Emmanuel Saliba[a] Eva Medina[c]

[a]Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), Wuerzburg, Germany; [b]Department of Bioinformatics and Biochemistry and Braunschweig Integrated Center of Systems Biology (BRICS), Technische Universität Braunschweig, Braunschweig, Germany; [c]Infection Immunology Research Group, Helmholtz Centre for Infection Research, Braunschweig, Germany; [d]Genome Analytics, Helmholtz Centre for Infection Research, Braunschweig, Germany; [e]Department of Pathology, University of Veterinary Medicine, Hannover, Germany; [f]Computational Biology of Infection Research, Helmholtz Centre for Infection Research, Braunschweig, Germany

Keywords
Myeloid-derived suppressor cells · *Staphylococcus aureus* · Immunometabolism · Single-cell RNA sequencing

Abstract

Myeloid-derived suppressor cells (MDSCs) are a compendium of immature myeloid cells that exhibit potent T-cell suppressive capacity and expand during pathological conditions such as cancer and chronic infections. Although well-characterized in cancer, the physiology of MDSCs in the infection setting remains enigmatic. Here, we integrated single-cell RNA sequencing (scRNA-seq) and functional metabolic profiling to gain deeper insights into the factors governing the generation and maintenance of MDSCs in chronic *Staphylococcus aureus* infection. We found that MDSCs originate not only in the bone marrow but also at extramedullary sites in *S. aureus*-infected mice. scRNA-seq showed that infection-driven MDSCs encompass a spectrum of myeloid precursors in different stages of differentiation, ranging from promyelocytes to mature neutrophils. Furthermore, the scRNA-seq analysis has also uncovered valuable phenotypic markers to distinguish mature myeloid cells from immature MDSCs. Metabolic profiling indicates that MDSCs exhibit high glycolytic activity and high glucose consumption rates, which are required for undergoing terminal maturation. However, rapid glucose consumption by MDSCs added to infection-induced perturbations in the glucose supplies in infected mice hinders the terminal maturation of MDSCs and promotes their accumulation in an immature stage. In a proof-of-concept in vivo experiment, we demonstrate the beneficial effect of increasing glucose availability in promoting MDSC terminal differentiation in infected mice. Our results provide valuable information of how metabolic alterations induced by infection influence reprogramming and differentiation of MDSCs.

© 2021 The Author(s).
Published by S. Karger AG, Basel

Introduction

MDSCs are an aberrant population of immature myeloid cells that fail to undergo terminal differentiation and accumulate during pathological conditions

Oliver Dietrich and Alexander Heinz contributed equally to this work.

karger@karger.com
www.karger.com/jin

© 2021 The Author(s).
Published by S. Karger AG, Basel

This is an Open Access article licensed under the Creative Commons Attribution-NonCommercial-4.0 International License (CC BY-NC) (http://www.karger.com/Services/OpenAccessLicense), applicable to the online version of the article only. Usage and distribution for commercial purposes requires written permission.

Correspondence to:
Eva Medina, eva.medina@helmholtz-hzi.de

such as cancer, chronic infection, and autoimmunity [1, 2]. In contrast to normal, mature myeloid cells, which play a pivotal role in host defense against pathogens and in the initiation of T-cell immunity, MDSCs exert immune regulatory functions and are potent suppressors of T-cell responses [3]. In humans and mice, MDSCs have been typically divided into 2 different subsets, monocytic and granulocytic, based on cell surface markers. In humans, granulocytic MDSCs are identified by the expression of $CD15^+CD11b^+CD33^+HLA\text{-}DR^-$ and monocytic MDSCs by the expression of $CD14^+CD11b^+CD33^+HLA\text{-}DR^-$ [4]. In mice, monocytic MDSCs express $CD11b^+Ly6C^+Ly6G^{low}$, while granulocytic MDSCs express $CD11b^+Ly6C^{low}L6G^+$ [4]. However, it has become clear that this classification is rather simplistic and does not recapitulate the high degree of phenotypic and functional heterogeneity of MDSCs [5, 6].

Although MDSCs have been extensively studied and characterized in the cancer setting, where they seem to play an important role in supporting tumor progression [7], mounting evidence indicates that MDSCs play also an important regulatory role in the immune response to pathogens [8]. MDSCs have been reported to play an important role in chronic infections caused by *S. aureus* [9–14], which is a major human pathogen that causes a wide variety of infections ranging from mild, self-limited infections to chronic and difficult-to-treat diseases including osteomyelitis, prosthetic joint infections, and biofilm-related infections [15]. We have previously reported the expansion of MDSCs in murine models of *S. aureus* chronic abscesses and bone infection where they induce progressive T-cell dysfunction and promote pathogen persistence [12]. In the same line, Heim and colleagues [9, 11, 14] demonstrated that MDSCs infiltrate the site of infection in a mouse model of *S. aureus* orthopedic implant infection, where they promote an anti-inflammatory environment that favore biofilm persistence. Accumulation of granulocytic MDSCs at the site of prosthetic joint infections has also been observed in humans [13]. The authors proposed that the accumulation of MDSCs could account for the chronicity of these infections [13]. All together, these observations indicate that MDSCs are an important element of the host response to *S. aureus* chronic infections, and therefore, targeting MDSCs may represent a promising therapeutic intervention to overcome immunosuppression and facilitate pathogen clearance by the immune system.

In cancer, several preclinical and clinical studies have shown the benefit of including MDSC-targeting approaches such as depletion of MDSCs or blockade of MDSCs migration in combination therapies to reduce tumor progression [16]. In the infection setting, however, considering that MDSCs also encompass populations of mature myeloid cells that are critical for the control of many pathogens, these strategies may have a profound negative effect in the course of infection. The development of such strategies will require a better understanding of how MDSCs are generated during chronic infection, which factors are involved in the process, and the mechanisms that prevent their maturation. In the current study, we used single-cell RNA sequencing (scRNA-seq) and metabolic profiling to investigate the origin, heterogeneity, molecular mechanisms, and pathways underlying the development and maintenance of MDSCs in a murine model of *S. aureus* chronic infection.

Materials and Methods

Bacterial Strains

S. aureus strains 6850 and SH1000 were grown to the mid-log phase in brain heart infusion medium (BHI, Roth) at 37°C with shaking (120 rpm), collected by centrifugation, washed with sterile PBS, and diluted to the required concentration. The number of viable bacteria was determined by tenfold serial dilution and colony count by plating on blood agar.

Mice and Infection Model

Pathogen-free 9- to 10-week-old C57BL/6 female mice were purchased from Envigo (The Netherlands) and maintained according to institutional guidelines in individually ventilated cages with food and water provided ad libitum. Mice were intravenously inoculated either with 10^6 CFU of *S. aureus* strain 6850 or with 4×10^7 CFU of *S. aureus* strain SH1000 in 100 μL of PBS via a lateral tail vein, and sacrificed by CO_2 asphyxiation at indicated times. Bacteria were counted in the tibia and spleen by preparing homogenates in PBS and plating tenfold serial dilutions on blood agar. In some experiments, infected mice were fed with water supplemented with 10% glucose during 10 days after bacterial inoculation. This time period was selected to minimize potential secondary metabolic alterations such as increased glucose intolerance and insulin resistance associated with long-term consumption of glucose-sweetened water. Blood glucose was measured using a Contour XT glucometer (Bayer).

Cell suspensions were prepared from the spleen of infected mice by gently teasing the spleen tissue through a 100-μm pore size nylon cell strainer and PBS+10% FCS. Splenocytes were spun down and erythrocytes were lysed after incubation for 5 min at RT in ammonium-chloride-potassium lysing buffer and then washed 3 times in PBS+10% FCS.

The bone marrow was flushed out of both tibia and femur from one hind limb nonaffected by the infection using a 21-gauge needle attached to a 5-mL syringe filled with PBS, followed by centrifugation and erythrocyte removal with ammonium-chloride-potassium.

Flow Cytometry Analysis

Cell suspensions were incubated with anti-mouse CD16/32 (eBioscience) for 5 min at RT to block Fc receptors and stained for 20 min at 4°C with antibodies against surface antigens. Cells were washed with PBS+10% FCS followed by fixation for 15 min with fixation buffer (BioLegend) and analyzed on a LSRII cytometer (Becton Dickinson).

For intracellular staining, cells were stained first against surface antigens as described above, fixed for 15 min at RT with fixation buffer, washed twice with permeabilization buffer (BioLegend), and stained for intracellular markers. After washing with permeabilization buffer, cells were analyzed on a LSRII cytometer. Data were analyzed using FlowJo v9.3 software.

The following antibodies have been used in this study: rat anti-mouse CD11b-PE/Cy7, rat anti-mouse CD45R/B220-PE, rat anti-mouse Ly6C-APC, rat anti-mouse Ly6G-PE, rat anti-mouse Ly6G-APC, anti-mouse CD117 (c-Kit)-FITC, rat anti-mouse Ly6A/E (Sca-1)-APC/Cy7, rat anti-mouse Lineage Cocktail-PB, and rat anti-mouse CD127 (IL-7Rα)-PE/Cy5, all from BioLegend; rat anti-mouse CD4-FITC, rat anti-mouse CD8-PE, and CD16/CD32 unconjugated from eBioscience; Armenian hamster anti-mouse CD3e purified and Syrian hamster anti-mouse CD28 purified from BD Pharmingen; and rat anti-mouse CCL6-Alexa Fluor 647 from R&D Systems.

Cell viability was determined by flow cytometry using propidium iodide solution following the manufacturer's recommendations (BioLegend).

Carboxyfluorescein Succinimidyl Ester Staining and Proliferation Assay

CD4$^+$ T cells were isolated from the spleen of uninfected mice using the mouse CD4$^+$ T Cell Isolation kit (Miltenyi Biotec), and Ly6C$^+$Ly6G$^+$ MDSCs were isolated from the spleen of *S. aureus*-infected mice at day 21 of infection using the mouse Myeloid-Derived Suppressor Cell Isolation Kit (Miltenyi Biotec) according to the manufacturer's instructions. Isolated CD4$^+$ T cells were then labeled with carboxyfluorescein succinimidyl ester (BioLegend) following the manufacturer's recommendations and cultured at 5×10^5 cells per well in complete RPMI-1640 medium (Gibco) supplemented with antibiotic-antimycotic (1:1,000) (VWR International), 4 mML-glutamine (Sigma-Aldrich), and 10% FCS and 2 μg/mL of Armenian hamster anti-mouse CD3ε plus 2 μg/mL of Syrian hamster anti-mouse CD28 antibodies (BD Pharmingen) at 37°C and 5% CO$_2$ for 72 h in the presence or absence of 5×10^5 per well of MDSCs isolated from the spleen of *S. aureus*-infected mice at a 1:1 ratio. Unstimulated CD4$^+$ T cells incubated in medium without anti-CD3ε and anti-CD28 antibodies were used as control. Proliferation was determined by flow cytometry analysis and dilution of CSFE as indication of cell division.

Cytokine Determination

IL-2 and IFN-γ levels were determined in the culture supernatant of CD4$^+$ T cells unstimulated or stimulated for 72 h with anti-CD3/anti-CD28 in the presence or absence of MDSCs isolated from the spleen of *S. aureus*-infected mice at day 21 of infection using mouse IL-2 and mouse IFN-γ ELISA sets according to the manufacturer's recommendations (BD Biosciences).

Histology

Spleens were removed from uninfected or *S. aureus*-infected mice at day 21 of infection, fixed in 10% formalin, and embedded in paraffin. Tissue section samples (2 μm thick) were stained with hematoxylin/eosin (Roth) and examined under a light microscope.

Single-Cell RNA Sequencing

Spleens isolated from 5 uninfected and 5 *S. aureus*-infected mice (day 21 of infection) were transformed into a single-cell suspension and pooled. The CD11b$^+$ populations in the infected and uninfected samples were sorted using a FACSAria(TM) SORP and approximately 4,000 cells loaded onto the 10x Genomics Chromium Controller following the single-cell 3′ v3 protocol (10x Genomics). Libraries were prepared from single-cell suspensions according to the 10x Genomics 3′ v3 protocol and sequenced using an Illumina NovaSeq 6000 sequencer (Illumina) with a sequencing depth of 200 million reads per sample.

In vivo 5-Ethynyl-2′-Deoxyuridine-Based Cell Proliferation Assay

5-Ethynyl-2′-deoxyuridine (EdU) (Thermo Fisher Scientific) was administered intraperitoneally (0.5 mg/mice) to uninfected or *S. aureus*-infected mice (day 21 of infection) 24 h before sacrifice. Spleens were removed, converted into a single-cell suspension, and stained for surface markers. EdU staining was performed with the Click-iT EdU AlexaFluor647 Flow Cytometry Assay Kit following the manufacturer's instructions (Thermo Fisher Scientific). Proliferating cells were determined by flow cytometry analysis.

In vitro Culture of MDSCs

Spleen cells isolated from *S. aureus*-infected mice (day 21 of infection) were cultured in vitro at a density of 5×10^6 cells/mL in complete RPMI-1640 medium at 37°C, 5% CO$_2$. Cells were harvested at the indicated times of in vitro culture, stained with antibodies against the surface marker Ly6G and with antibodies against the intracellular marker CCL6, and analyzed by flow cytometry.

In some experiments, spleen cells were cultured in complete RPMI-1640 medium containing different concentrations of glucose (0, 0.5, 1, and 2 mg/mL). To inhibit glycolysis, spleen cells were incubated in complete RPMI-1640 medium containing glucose (2 mg/mL) in the presence of 10 mM of the glycolysis inhibitor 2-deoxy-D-glucose (2-DG).

Cytospin

Cytospin were prepared using aliquots of in vitro-cultured MDSCs. The material was centrifuged at 500 rpm for 5 min in a Shandon cytocentrifuge (Cytospin 2, Shandon, UK). Slides were stained using May-Grünwald-Giemsa (Polysciences) and photographed with a light microscope.

Phagocytosis and Killing Assay

Ly6C$^+$Ly6G$^+$ cells were isolated from *S. aureus*-infected mice (day 21 of infection) and cultured for 96 h in complete RPMI medium. Cells were collected, washed, seeded in multi-well plates at 5×10^5 cells per well, and incubated with *S. aureus* at an MOI of 10:1 in the presence of 10% mouse serum. The plates were centrifuged at 700 *g* for 5 min and incubated at 37°C for 1 h to allow phagocytosis. Noningested extracellular bacteria were then killed by addition of 100 μg/mL gentamicin (Gibco) and 5 μg/mL lysostaphin (Sigma-Aldrich), and cells were washed and further incubated for 3 h at 37°C. Cells were then harvested, pelleted by centrifugation, and lysed with 0.1% Triton X-100 (Sigma), and CFU were enumerated by plating on blood agar.

Glucose Uptake Assay

Ly6C$^+$Ly6G$^+$ cells were isolated from *S. aureus*-infected mice (day 21 of infection) using the Myeloid-Derived Suppressor Cell Isolation Kit (Miltenyi Biotec) according to the manufacturer's instructions. Ly6C$^+$Ly6G$^+$ cells prior to (ex vivo) or after in vitro culture for 96 h (in vitro) were seeded in 48-well plates at 10^6 cells per well in glucose-free RPMI-1640 medium supplemented with 300 μM 2-(N-[7-nitrobenz-2-oxa-1,3-diazol-4-yl] amino)-2-deoxyglucose (Thermo Fisher Scientific) and incubated for 30 min at 37°C, 5% CO_2. Cells were washed and analyzed by flow cytometry.

Seahorse Extracellular Flux Analysis

Ly6C$^+$Ly6G$^+$ cells were isolated from *S. aureus*-infected mice (day 21 of infection) using the Myeloid-Derived Suppressor Cell Isolation Kit (Miltenyi Biotec) according to the manufacturer's instructions and used prior to (ex vivo) or after in vitro culture for 96 h (in vitro). Oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) of Ly6C$^+$Ly6G$^+$ cells were assessed using an Agilent Seahorse XF96 Analyzer (Agilent Technologies). One day prior to the assay, the Seahorse XF Utility Plate (Agilent Technologies) was hydrated by adding 200 μL of sterile Milli-Q H_2O to each well and incubated overnight in a non-CO_2 37°C incubator together with the XF sensor cartridge (Agilent Technologies). Before seeding the cells, the wells of a Seahorse 96-well XF cell culture plate (Agilent Technologies) were incubated with poly-L-lysin (Sigma-Aldrich) for 1 h at 37°C, extensively washed with Milli-Q H_2O after removing the poly-L-lysin, and left to dry for 30 min at 37°C. Cells were added to poly-L-lysin-coated plates at a concentration of 5×10^5 cells per well in 180 μL Seahorse RPMI medium supplemented with 10 mM glucose and 2 mM glutamine (pH 7.4) and centrifuged at 1,000 *g* for 5 min. The wells filled up with only assay medium were used as background control. Water was removed from the wells in the utility plate and 200 μL prewarmed (37°C) Seahorse XF calibrant solution (Agilent Technologies) was added to each well. The cell culture plate and utility plate with the sensor cartridge were equilibrated after incubation in a non-CO_2 37°C incubator for 1 h. The different inhibitors of the Seahorse XF Cell Glycolytic Rate Assay Kit (Agilent Technologies) were added to the corresponding ports of the sensor cartridge prior to starting the assay. Thus, 20 μL of 5 μM rotenone and 5 μM antimycin A solution was added to port A and 22 μL of 500 μM 2-DG solution to port B. The utility plate and the sensor cartridge were then placed into the XF96 analyzer and calibrated. After calibration, the utility plate was replaced by the cell culture plate, and cell respiration parameters were determined by stepwise injection of the different inhibitors. During each measurement cycle, the OCR and ECAR were determined 3 times including 3 min of mixing and 3 min of measurement.

Stable Isotope Labeling, Metabolite Extraction, GC-MS Measurement, and Data Processing

Ly6C$^+$Ly6G$^+$ cells isolated from the spleen of *S. aureus*-infected mice (day 21 of infection) prior to (ex vivo) and after in vitro culture for 96 h (in vitro) were seeded in 6-well plates at 8×10^6 cells per well in RPMI medium containing either 11 mM [U-$^{13}C_6$]-glucose (Cambridge Isotope Laboratories), 2 mM [U-$^{13}C_5$]-glutamine (Cambridge Isotope Laboratories), or 100 μM [U-$^{13}C_{16}$]-palmitate (Cambridge Isotope Laboratories), 10% FCS, 1% penicillin (10,000 IU/mL), and streptomycin (20 mg/mL). ^{12}C metabolites were added to each tracer to ensure equivalent nutrient state. Before the treatment, [U-$^{13}C_{16}$]-palmitate was noncovalently conjugated to fatty-acid-free BSA (Sigma-Aldrich) as previously described [17]. The cells were incubated at 37°C, 5% CO_2 for 4 h. Cells were then harvested and intracellular metabolites were extracted as previously described [19]. Briefly, cell suspensions were collected and centrifuged at 250 *g* for 5 min. The cell pellet was washed with 1 mL 0.9% NaCl, followed by centrifugation at 250 *g* for 5 min. The cells were immediately put on ice to quench the metabolism, and 250 μL ice-cold HPLC-grade methanol (Sigma-Aldrich), 250 μL Milli-Q H_2O with 1 μg/mL D6 glutaric acid (CDN isotopes) as internal standard, and 250 μL HPLC-grade chloroform (Sigma-Aldrich) were added. The cells were agitated at 4°C for 20 min at 1,400 rpm, followed by centrifugation at 17,000 *g* at 4°C for 5 min. After phase separation, 300 μL of the polar phase was transferred to a glass vial with a micro-insert and dried at 4°C under vacuum. Derivatization for gas chromatography was performed using a Gerstel MPS. Dried polar metabolites were dissolved in 15 μL of 20 mg/mL methoxyamine hydrochloride (Sigma-Aldrich) in pyridine (Roth) at 40°C while shaking for 90 min. An equal volume of N-tert-butyldimethylsilyl-N-methytrifluoroacetamide (Restek) was then injected, and the cells were further incubated for 60 min at 55°C under shaking. GC-MS measurement was performed on an Agilent 7890B GC coupled to an Agilent 5977B with extractor EI source (Agilent Technologies). Metabolites of interest were measured in selected ion monitoring mode. The Metabolite Detector software was used for the data analysis with the following settings: peak threshold, 5; minimum peak height, 5; bins per scan, 10; deconvolution width, 5 scans; no baseline adjustment; required peaks, 2; and no minimum required peak intensity. Retention index was calibrated based on the retention time. An in-house mass spectral library was used for compound identification. Mass isotopomer distributions were calculated by MetaboliteDetector's MID wizard. Fractional contribution of glutamine-, glucose-, and palmitate-derived carbon to total metabolite carbon was calculated by dividing the sum of the abundance of all isotopologs (except M0) by the total number of carbons in the respective metabolite.

Statistical Analysis

Single-Cell RNA-Seq Data Analysis

Sequencing data were demultiplexed using Cell Ranger software (version 2.0.2) (10x Genomics), and FASTQ files were generated. Reads were aligned to the UCSC mm10 reference genome (GRCm37) using Cell Ranger followed by quantification of gene expressions and generation of a gene-barcode matrix. Individual datasets were aggregated using the Cellranger aggr command and further analyzed using the R package Seurat (version 3.1.4) (https://cran.r-project.org/package=Seurat). The data were subjected to library-size normalization and log transformation, and the 3,000 most variable genes (based on variance-stabilizing transformation) in the dataset were used for downstream analysis. Principal component analysis (PCA) was used to reduce the dimensionality of the original matrix, and 10 principal components were used to calculate the Uniform Manifold Approximation and Projection (UMAP) and clusters. Genes that were differentially expressed using the FindMarkers function (default parameters) in Seurat and genes with *p* values <0.01 were considered as differentially expressed genes.

The raw expression matrix was subset by cluster annotation (classical monocytes, immature myeloid cells, and neutrophils) and normalized by SCTransform. The 3,000 most variable genes (as above) were chosen for downstream analysis. Data were scaled, and a PCA was calculated, of which the first 30 components were used for UMAP and clustering. Mutual nearest neighbor batch correction was performed on the low-dimensional representation (PCA) as recommended in the batchelor vignette by Aaron Lun (https://bioconductor.org/packages/release/bioc/vignettes/batchelor/inst/doc/correction.html). Pseudotemporal ordering of single cells was performed using monocle3 using the normalized data (preprocess_cds params: norm_method = "none," num_dim = 15) including mutual nearest neighbor batch correction (alignment group = sample). Cell cycle assignment was performed using the Seurat function CellCycleScoring using 20 bins and the genes previously reported by Kowalczyk et al. [20]. Over-representation of gene ontology (GO) categories was calculated using the R package clusterProfiler (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html). Visualizations were produced with the R package ggplot2 (https://cran.r-project.org/package=ggplot2).

Other Data Analysis

Comparisons between groups were made using a parametric ANOVA test with the Tukey posttest or a 2-way ANOVA test. p values <0.05 were considered significant. Heatmap of metabolite concentration was generated with R package "pheatmap" with a p value cutoff of 0.05, ANOVA, and z-score normalization. Results are presented as mean values ± SD of a minimum of 3 replicates, and all experiments were repeated at least 3 times.

Results

MDSCs Originate from Both Bone Marrow and Extramedullary Sites in S. aureus *Chronic Infection*

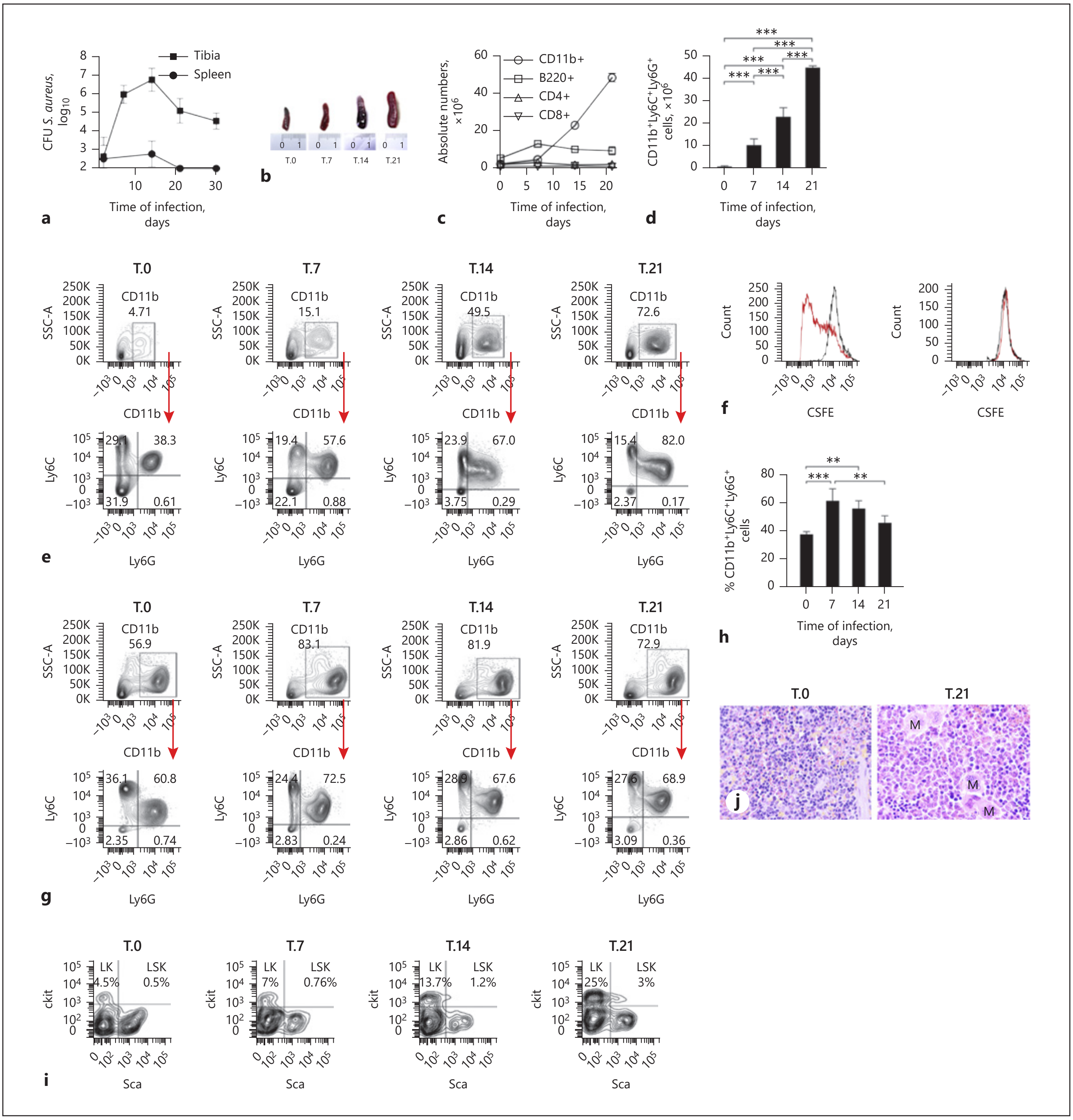
We used a previously described experimental model of *S. aureus* chronic infection [21] to investigate the origin and physiology of infection-driven MDSCs. C57BL/6 were infected intravenously with 10^6 CFU of *S. aureus* strain 6850, and bacterial loads were determined in the tibia and spleen at progressing times after bacterial inoculation. Consistent with previous observations [21], *S. aureus* was detectable in the tibia of infected mice for up to 30 days, but it was under detection levels in the spleen from day 20 onward (Fig. 1a). Infected mice developed pronounced splenomegaly with the progression of infection (Fig. 1b), which was largely due to a disproportionate accumulation of CD11b$^+$ cells (Fig. 1c) expressing the markers Ly6C and Ly6G (Fig. 1d, e; online suppl. Fig. 1; for all online suppl. material, see www.karger.com/doi/10.1159/000519306), typical of murine MDSCs [4]. Since the ability to suppress T-cell responses is the hallmark of MDSCs [4], we then assessed the capacity of the Ly6C$^+$Ly6G$^+$ cells accumulating in the spleen of infected mice to inhibit T-cell proliferation. For this purpose, mouse CD4$^+$ T cells isolated from the spleen of uninfected C57BL/6 mice and labeled with CSFE were stimulated with anti-CD3 and anti-CD28 antibodies and incubated in the presence or absence of Ly6C$^+$Ly6G$^+$ cells isolated from the spleen of *S. aureus*-infected mice at day 21 of infection. On day 3 of culture, proliferation of CD4$^+$ T cells was determined by flow cytometry. As shown in Figure 1f, Ly6C$^+$Ly6G$^+$ potently suppressed proliferation of CD4$^+$ T cells and therefore fulfilled the functional criteria for MDSCs. Furthermore, secretion of cytokines such as IL-2 and IFN-γ by anti-CD3/anti-CD28-stimulated CD4$^+$ T cells was significantly decreased in the presence of MDSCs (online suppl. Fig. 2).

We also demonstrated that the expansion of MDSCs during *S. aureus* chronic infection is not bacterial strain-specific since similar splenomegaly and accumulation of MDSCs with inhibitory effects on T-cell responses were observed in mice intravenously infected with *S. aureus* strain SH1000, a strain that causes chronic renal infection in mice [22] (online suppl. Fig. 3).

To investigate the origin of the MDSCs arising during *S. aureus* infection, we first focused on the bone marrow since this is the primary site where myeloid cells are produced. Flow cytometry analysis of the bone marrow isolated from *S. aureus*-infected mice at progressing times of infection showed a significant increase in the percentage of CD11b$^+$ cells predominantly expressing Ly6C and Ly6G at day 7 that gradually decreased at later times (Fig. 1g, h; online suppl. Fig. 4). Because the kinetic of CD11b$^+$Ly6C$^+$Ly6G$^+$ cells in the bone marrow (Fig. 1g, h; online suppl. Fig. 4) did not match the progressive increase of these cells observed in the spleen during the course of infection (Fig. 1d, e; online suppl. Fig. 1), we speculated that in addition to the bone marrow, MDSCs may also originate from other sites. In this regard, it has been reported that MDSCs can originate at extramedullary sites such as the spleen and liver during chronic inflammatory conditions as a consequence of extramedullary hematopoiesis [5]. To investigate if extramedullary hematopoiesis is occurring at peripheral sites during *S. aureus* infection, we determined the percentage of Lin$^-$IL-7Rα$^-$c-Kit$^+$Sca-1$^-$ lineage-committed progenitors (LK) and of Lin$^-$IL-7Rα$^-$Sca-1$^+$c-Kit$^+$ myeloid progenitors (LSK) in the spleen of infected mice at progressing times after bacterial inoculation. A time-dependent increase in the frequency of LK and of LSK was observed in the spleen of *S. aureus*-infected mice (Fig. 1i). Histological examination of the spleen tissue taken from *S. aureus*-infected mice at day 21 of infection showed the red pulp markedly expanded by numerous hematopoietic cells including erythroid and myeloid precursors as well as megakaryocytes, further confirming the occurrence of extramedullary hematopoiesis (Fig. 1j). Together, these results indicate that both the bone marrow and extramedullary hematopoiesis at peripheral sites may contribute to the expansion of MDSCs observed during *S. aureus* chronic infection.

High-Resolution Mapping of Infection-Driven MDSCs Determined by scRNA-Seq

To capture the phenotypic variation among MDSCs present in the spleen of *S. aureus*-infected mice at a high resolution, we performed scRNA-seq on sorted CD11b$^+$ cells at day 21 of infection. CD11b$^+$ cells isolated from the



1

(For legend see next page.)

spleen of uninfected control mice were included to determine the changes in cell composition specifically induced by infection. The scRNA-seq data acquired using the droplet-based 10x Genomics technology from both conditions were combined (1,897 cells from control and 1,497 cells from infected), and a 2-dimensional representation of the single cell transcriptomes was obtained using a UMAP (Fig. 2a). A total of 9 different cell clusters were identified according to the expression of known marker genes, including NK cells (*Nkg7*, *Klrb1c*, *Klre1*, *Klrk1*, *Klra7-9,* and *Klrd1*), B cells (*Cd79a*, *Cd79b*, *Cd19*, and *Cd74*), dendritic cells (*H2-Ab1*, *H2-Eb1*, *H2-Aa*, *Cd209a*, and *H2-DMb1*), classical- (*Ly6c2*, *Ccl9*, *Ccr2*, and *Cd68*) and nonclassical (*Fabp4*, *Cx3cr1*, and *Csf1r*) monocytes, plasma B cells (*Jchain* and *Sdc1*), T cells (*Cd3d*), neutrophils (*S100a8*, *Ccl6*, *Il1b*, *Ly6g*, and *Wfdc21*), and a cluster of immature myeloid cells expressing markers along the granulocytic differentiation axis (*Ly6c2*, *Ly6g*, *Chil3*, *Camp*, *Ltf*, *S100a8*, and *Wfdc21*) that we classified as MDSCs (Fig. 2b, c; online suppl. Table 1). The proportions of cell types between the 2 conditions are shown in Figure 2d. Notably, the majority of CD11b$^+$ cells from infected mice could be classified as immature myeloid cells, while no such population was present in the control sample.

To investigate the full extent of heterogeneity of MDSCs, we extracted all transcriptomes annotated as classical monocytes, immature myeloid cells, and neutrophils from the combined dataset and reanalyzed this subset. Conceptionally, the classical monocytes and neutrophils, present in both conditions, represent the typical myeloid cell populations under homeostatic conditions (compare Fig. 3a, b). In infected mice, however, a continuous spectrum of cells could be observed between these 2 populations representing the different stages of granulocyte differentiation as shown in Figure 3b.

Neutrophils contain 4 types of granules including primary (azurophilic), secondary (specific), tertiary (gelatinase), and ficolin-1-rich granules. These granules are produced stepwise during the different stages of maturation that start with promyelocytes followed by myelocytes, metamyelocytes, band cells, and end with terminally differentiated segmented neutrophils [23]. This process has been described as a targeting-by-timing model to explain the differences in protein contents among neutrophil granule subsets [23–26]. Accordingly, we used the expression levels of the genes encoding the different granule proteins to classify the spectrum of cell populations identified by scRNA-seq within the MDSCs into specific neutrophil differentiation categories. Azurophilic granule proteins including myeloperoxidase (encoded by *Mpo*), elastase (encoded by *Elane*), cathepsin G (encoded by *Ctsg*), and proteinase 3 (encoded by *Prtn3*) are produced only at the promyelocyte stage (Fig. 3c, d; online suppl. Table 2). Myelocytes were identified by the high expression of genes encoding secondary granule proteins such as lactoferrin (*Ltf*), cathelicidin (*Camp*), and neutrophil gelatinase-associated lipocalin (*Lcn2*) as well as by the expression of the gene encoding ficolin-1 (*Fcnb*), which originates during the transition from myelocytes to metamyelocytes (Fig. 3c, e; online suppl. Table 2). Metamyelocytes could be identified based on the high expression of the gene encoding the above-mentioned secondary granule proteins (*Ltf*, *Camp*, *Lcn2*, and *Fcnb*) and by the increased expression of the gene encoding Ly6G (*Ly6g*) (Fig. 3c–f; online suppl. Table 2). The expression of genes encoding tertiary granule proteins such

Fig. 1. MDSCs originate from both the bone marrow and extramedullary sites in *S. aureus* chronic infection. **a** Bacterial load in the tibia and spleen of mice after intravenous inoculation with 10^6 CFU of *S. aureus* 6850. **b** Photographs showing splenomegaly in *S. aureus*-infected mice. **c** Absolute numbers of CD11b$^+$, B cells (B220$^+$), and CD4$^+$ and CD8$^+$ T cells in the spleen of *S. aureus*-infected mice at progressing times after bacterial inoculation. **d** Absolute numbers of CD11b$^+$Ly6C$^+$Ly6G$^+$ in the spleen of *S. aureus*-infected mice. **e** Flow cytometry analysis showing the frequency of CD11b$^+$ cells (upper panels) and the frequency of Ly6C$^+$Ly6G$^+$ cells within the CD11b$^+$ population (lower panels) in the spleen of *S. aureus*-infected mice. Gating strategies are shown in online suppl. Figure 5. **f** Proliferation of CD4$^+$ T cells unstimulated (black line histograms) or stimulated with anti-CD3ε/anti-CD28 antibodies (red line histograms) for 72 h in the absence (left panel) or presence (right panel) of Ly6C$^+$Ly6G$^+$ cells isolated from the spleen of *S. aureus*-infected mice at day 21 of infection at a 1:1 ratio. **g** Flow cytometry analysis showing the frequency of CD11b$^+$ cells (upper panels) and the frequency of Ly6C$^+$Ly6G$^+$ cells within the CD11b$^+$ population (lower panels) in the bone marrow of *S. aureus*-infected mice. **h** Percentage of CD11b$^+$Ly6C$^+$Ly6G$^+$ cells in the bone marrow of *S. aureus*-infected mice. **i** Representative flow cytometry analysis showing LK and LSK in the spleen of uninfected (T.0) or *S. aureus*-infected mice at progressing times of infection. **j** Representative hematoxylin and eosin-stained histological sections of formalin-fixed and paraffin-embedded spleen sections from an uninfected (left) and from a *S. aureus*-infected (right) mouse at day 21 of infection (M indicates megakaryocytes). Original magnification ×40. Results are presented as the mean ± SD of biological replicates ($N = 5$ for **a** and $N = 4$ for **e–h**) and are representative of 3 independent experiments. Statistical significance: $^{**}p < 0.01$; $^{***}p < 0.001$. LK, Lin$^-$IL-7Rα$^-$c-Kit$^+$Sca-1$^-$ lineage-committed progenitors; LSK, Lin$^-$IL-7Rα$^-$Sca-1$^+$c-Kit$^+$ myeloid progenitors.

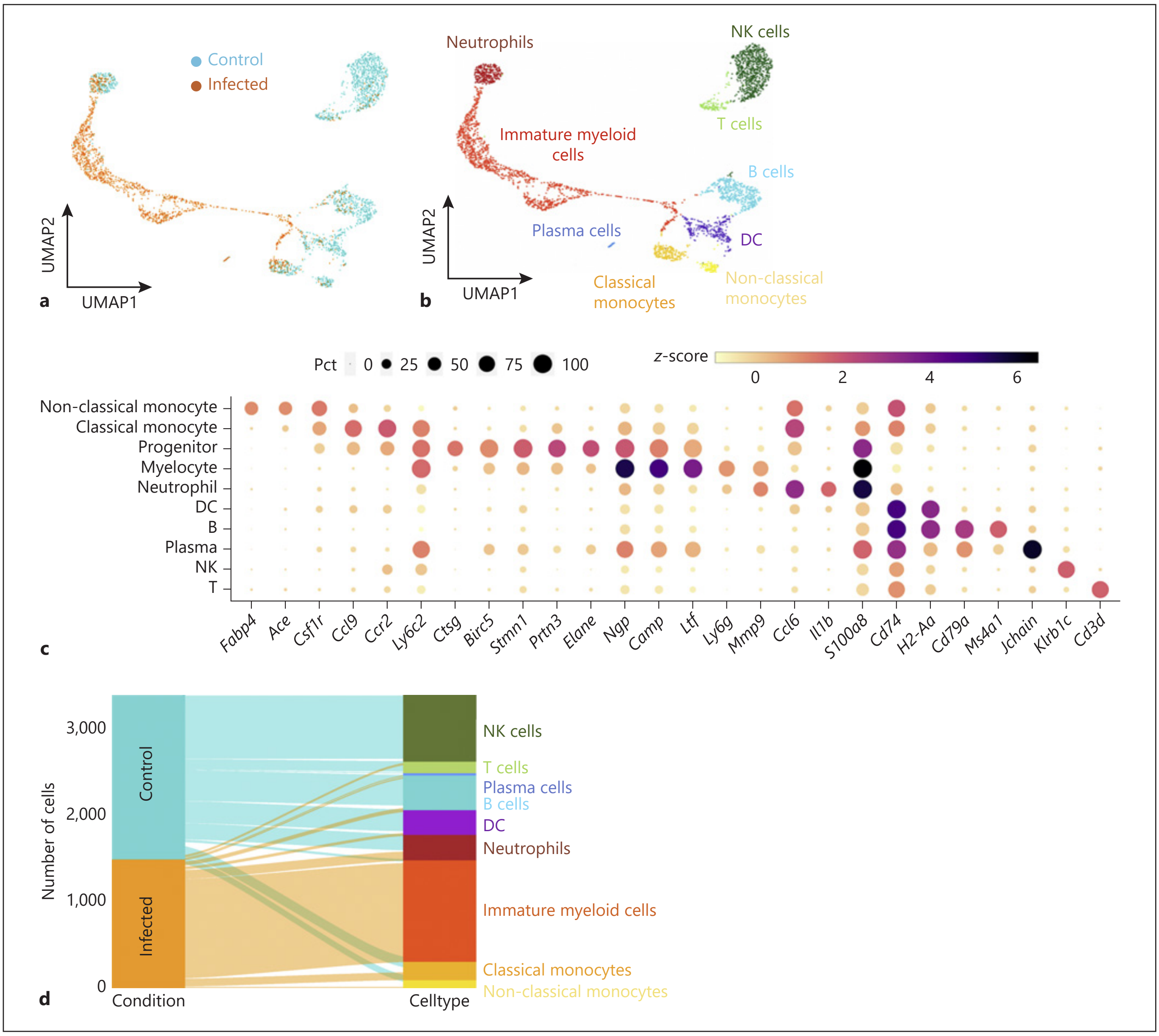
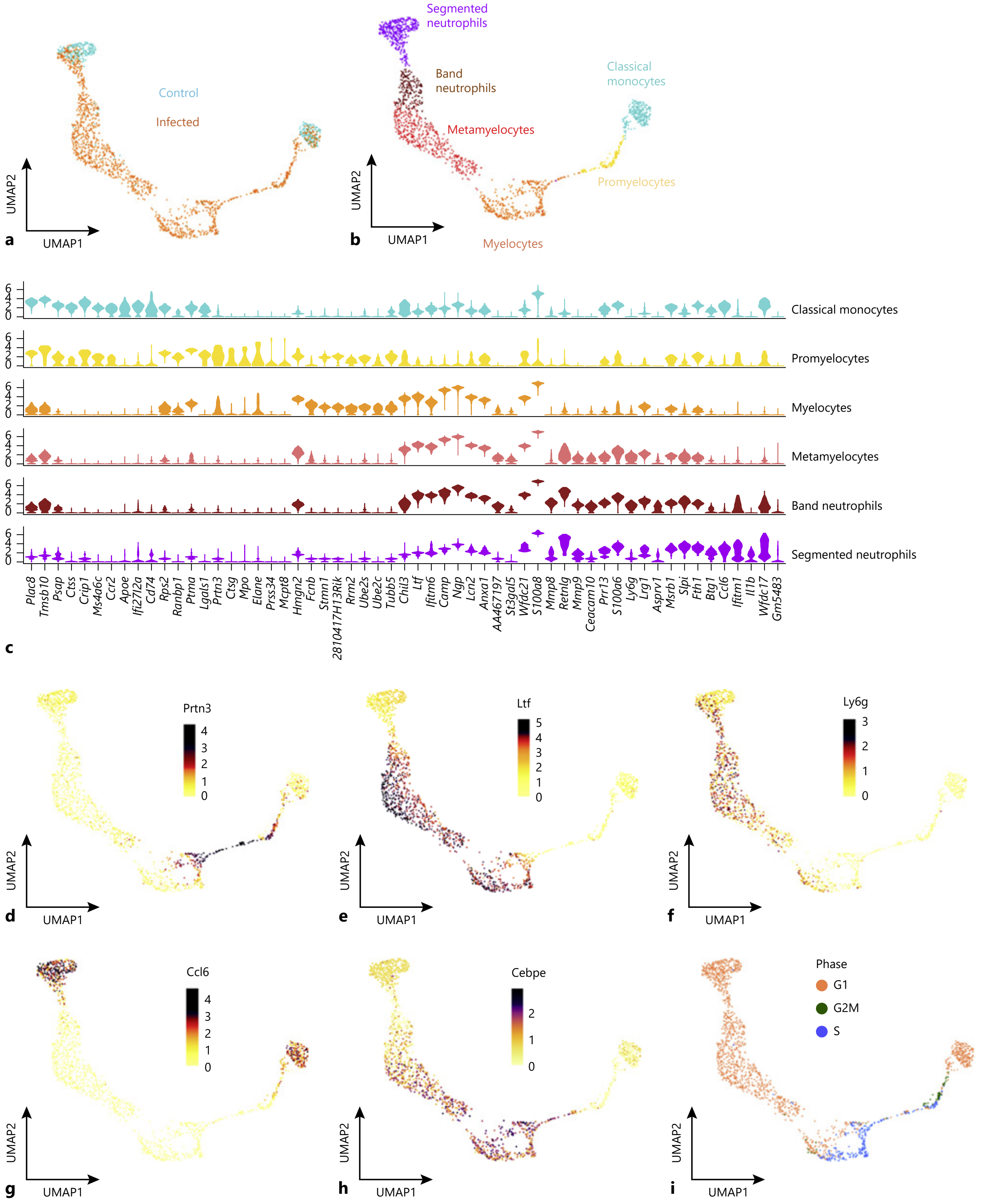


Fig. 2. High-resolution map of infection-driven MDSCs determined by scRNA-seq analysis. **a** UMAP embedding of 1,497 transcriptomes of CD11b$^+$ cells isolated from the spleen of *S. aureus*-infected mice at day 21 of infection (infected) and 1,897 transcriptomes of CD11b$^+$ cells isolated from uninfected mice (control). **b** UMAP embedding (as **a**) showing clusters of cells annotated by cell type. **c** Dot plot of marker genes underlying the cell type annotation. As indicated by the legend, the size of each dot reflects the proportion of cells of each type expressing the marker gene, and the color intensity reflects the mean expression of each marker gene across all cells in the respective cluster. **d** Sankey plot showing the proportion of cell types between control and infection. UMAP, Uniform Manifold Approximation and Projection.

Fig. 3. Single-cell RNA-seq identifies cell populations within infection-driven MDSCs covering the full spectrum of granulocyte differentiation. **a** UMAP embedding of 1,687 single-cell transcriptomes from the cell clusters *classical monocyte*, *immature myeloid cells*, and *neutrophils* extracted from the combined dataset shown in Figure 2b. **b** UMAP embedding (as **a**) showing clusters of cells annotated by cell type. **c** Violin plots showing the expression of selected marker genes of the different cell clusters depicted in **b**. Expression in each cell population is shown along with the probability density of gene expression symbolized by the shape of the plot. **d–h** UMAP embeddings (as **a**) showing the normalized expression of *Prtn3* (**d**), *Ltf* (**e**), *Ly6g* (**f**), *Ccl6* (**g**), and *Cebpe* (**h**) genes. **i** UMAP embedding showing the cell cycle phase.UMAP, Uniform Manifold Approximation and Projection.

(For figure see next page.)


a
Control
Infected
UMAP1
UMAP2
b
Segmented neutrophils
Band neutrophils
Classical monocytes
Metamyelocytes
Promyelocytes
Myelocytes
UMAP1
UMAP2
c
6
4
2
0
Classical monocytes
Promyelocytes
Myelocytes
Metamyelocytes
Band neutrophils
Segmented neutrophils
Plac8
Tmsb10
Psap
Ctss
Crip1
Ms4a6c
Ccr2
Apoe
Ifi27l2a
Cd74
Rps2
Ranbp1
Ptma
Lgals1
Prtn3
Ctsg
Mpo
Elane
Prss34
Mcpt8
Hmgn2
Fcnb
Stmn1
2810417H13Rik
Rrm2
Ube2s
Ube2c
Tubb5
Chil3
Ltf
Ifitm6
Camp
Ngp
Lcn2
Anxa1
AA467197
St3gal5
Wfdc21
S100a8
Mmp8
Retnlg
Mmp9
Ceacam10
Prr13
S100a6
Ly6g
Lrg1
Asprv1
Msrb1
Slpi
Fth1
Btg1
Ccl6
Ifitm1
Il1b
Wfdc17
Gm5483
d
Prtn3
4
3
2
1
0
UMAP1
UMAP2
e
Ltf
5
4
3
2
1
0
UMAP1
UMAP2
f
Ly6g
3
2
1
0
UMAP1
UMAP2
g
Ccl6
4
3
2
1
0
UMAP1
UMAP2
h
Cebpe
2
1
0
UMAP1
UMAP2
i
Phase
G1
G2M
S
UMAP1
UMAP2

3

as metalloproteinase 8 (*Mmp8*) and metalloproteinase 9 (*Mmp9*) identified band neutrophils (Fig. 3c; online suppl. Table 2). Last, terminally differentiated segmented neutrophils were identified based on the high expression of genes encoding markers such as colony-stimulating factor 3 receptor (*Csf3r*), IL1-β (*Il1b*), and CCL6 (*Ccl6*) (Fig. 3c, g; online suppl. Table 2). The expression pattern of *Cebpe*, which encodes the transcription factor CCAAT/enhancer binding protein-ε (C/EBP-ε) that is predominantly expressed during the myelocyte and metamyelocyte differentiation stages [27], confirmed the classification performed based on the expression of genes encoding granule proteins (Fig. 3h; online suppl. Table 2).

Mature neutrophils are mitotically inactive with cell cycle arrest occurring during the myelocyte to metamyelocyte transition [25]. To substantiate this in our experimental setting, we performed cell cycle analysis in the scRNA-seq data. The results indicated that both promyelocytes and myelocytes were actively proliferating as they were in phases S (DNA synthesis) and G2/M (cell division) of the cell cycle (Fig. 3i). On the other hand, metamyelocytes, band neutrophils, and segmented neutrophils were all in the postmitotic G1 phase (growth phase, Fig. 3i). Consistent with these results, gene ontology analysis identified the cell cycle to be an overrepresented functional category in promyelocytes and myelocytes, whereas functional categories associated with cell migration and host defense were overrepresented in the more mature populations such as band and segmented neutrophils (Fig. 4a). In addition, we performed functional analysis to determine the presence of actively proliferating cells in the spleen of *S. aureus*-infected mice in vivo using EdU, a thymidine analog that is incorporated into proliferating cells during DNA synthesis. A significantly higher number of proliferating cells were detected in the spleen of *S. aureus*-infected mice (day 21 of infection) in comparison to uninfected control mice (Fig. 4b, upper panels). More than 60% of the actively proliferating cells (EdU$^+$) were Ly6C$^+$/Ly6G$^-$ (Fig. 4b, lower panel) and most probably represented promyelocyte and myelocyte cell populations.

To further validate the hierarchy between the different myeloid cell populations within the MDSC cluster identified by the scRNA-seq data, we performed trajectory analysis based on pseudotime, where cells represent distinct stages in a continuous developmental process. This enables the association of specific cell types with the initial, intermediate, and terminal states of the trajectory [28]. The pseudotime analysis shown in Figure 4c recapitulated the trajectory of cell differentiation from promyelocytes (initial) to terminally differentiated segmented neutrophils (final) including several intermediate developmental states comprising myelocytes, metamyelocytes, and band neutrophils.

Distinguishing between immature granulocytic precursors and mature segmented neutrophils has been very difficult, and no phenotypic marker has been identified so far that enables precise separation of these populations. The results of the scRNA-seq analysis performed in this study have revealed that the expression of both Ly6G and CCL6 markers may be suitable to separate mature neutrophils (Ly6G$^+$CCL6$^+$) from immature MDSC precursors (Ly6G$^+$CCL6$^-$). This was corroborated by flow cytometry analysis showing that whereas approximately 90% of neutrophils in the spleen (Fig. 4d, f) and blood (Fig. 4e, f) of uninfected mice were mature neutrophils (Ly6G$^+$CCL6$^+$), <10% of Ly6G$^+$ cells expressed CCL6 in the spleen (Fig. 4d, f) and blood (Fig. 4e, f) of *S. aureus*-infected mice.

In the cancer setting, it has been reported that MDSCs are not irreversibly arrested in an immature stage and could terminally differentiate after being removed from the tumor environment and cultured under in vitro conditions [29–31]. To investigate if this was also the case for infection-driven MDSCs, we determined the capacity of MDSCs isolated from the spleen of *S. aureus*-infected mice at day 21 of infection to undergo terminal maturation upon in vitro culture conditions. As the surface expression of Ly6G and the intracellular expression of CCL6 were revealed by the scRNA-seq analysis as markers of mature neutrophils, we monitored MDSC maturation by measuring the level of expression of these markers at increasing times of in vitro culture. Although cell viability slowly decreased with time, over 70% of Ly6G$^+$cells were still viable after 96 h of in vitro culture (online suppl. Fig. 7a, b). Flow cytometry analysis of Ly6G$^+$ at different times of in vitro culture showed a time-dependent gradual increase in the percentage of Ly6G$^+$CCL6$^+$ (Fig. 4g, h). After 96 h of in vitro culture, >80% of Ly6G$^+$ cells expressed CCL6, when only 10% of the Ly6G$^+$ expressed CCL6 prior to in vitro culture (0 h) (Fig. 4g, h). Morphological changes were also observed in in vitro-cultured MDSCs, which showed a transition from predominant immature myeloid cells including cells exhibiting round nuclei typical of promyelocytes, kidney-shaped nuclei typical of myelocytes and metamyelocytes, and band-like-shaped nuclei typical of band neutrophils prior to in vitro culture (Fig. 4i, upper panel) to cells with segmented nuclei morphology typical of mature neutrophils after 96 h of in vitro culture (Fig. 4i, lower panel). Furthermore, the in vi-

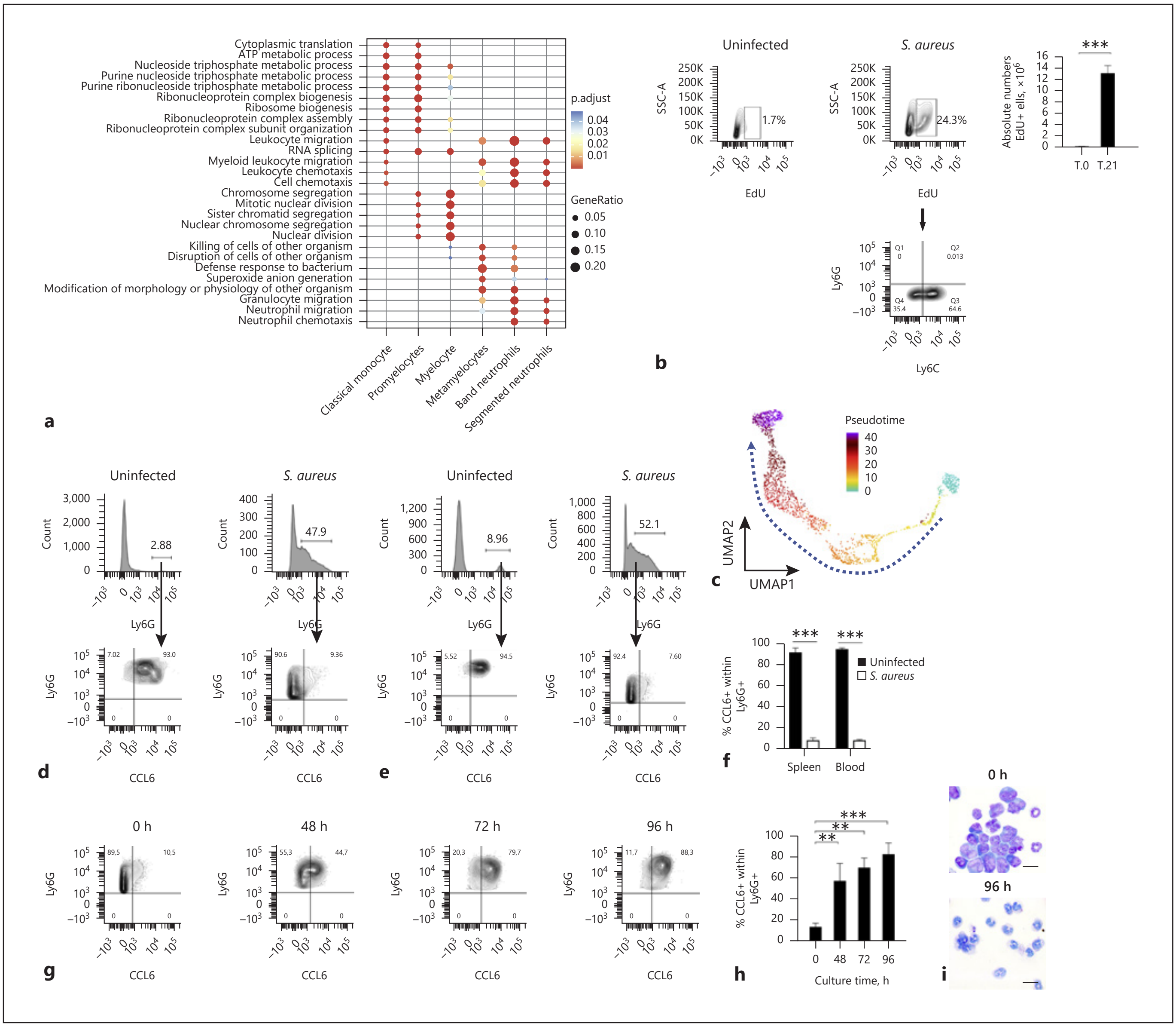


Fig. 4. Reconstruction of trajectories and transition states of infection-driven MDSCs. **a** Most enriched GO categories of the upregulated (ranked by log-fold change) genes for each cell cluster depicted in Figure 3b. **b** Representative flow cytometry contour plots showing the percentage of EdU$^+$ cells gated in the total spleen cell population of uninfected mice (left panel) and *S. aureus*-infected mice at day 21 of infection (middle panel). The absolute numbers of EdU$^+$ cells in the spleen of uninfected (T.0) and *S. aureus*-infected (T.21) mice are shown in the right panel. The expression of Ly6G and Ly6C within the EdU$^+$ cell population in the spleen of *S. aureus*-infected mice is shown in the contour plot depicted in the lower panel. The gating strategy is described in online suppl. Figure 6. **c** UMAP embedding showing granulocyte differentiation as a continuous spectrum of pseudotime. The dashed line indicates the progression toward terminal differentiation. **d**, **e** Flow cytometry analysis of CCL6 expression within the Ly6G$^+$ population in the spleen (**d**) and blood (**e**) of uninfected (left panels) or *S. aureus*-infected (right panels) mice at day 21 of infection. **f** Percentage of CCL6$^+$ cells within Ly6G$^+$ cell population in the spleen and blood on uninfected (black bars) or *S. aureus*-infected (white bars) mice. **g** Flow cytometry contour plots showing the expression kinetic of CCL6 within Ly6G$^+$ cells isolated from the spleen of *S. aureus*-infected mice at day 21 of infection and cultured in vitro for 96 h. The gating strategy is depicted in online suppl. Figure 8. **h** Percentage of CCL6$^+$ cells within the Ly6G$^+$ cell population at progressive times during in vitro culture. **i** May-Grünwald-Giemsa staining of cytospin preparations from Ly6G$^+$ spleen cells isolated from *S. aureus*-infected mice at day 21 of infection before (upper panel) and 96 h after (lower panel) in vitro culture. Scale bars represent 200 µm. Results are presented as the mean ± SD of 3 biological replicates ($N = 3$) and are representative of 3 independent experiments. Statistical significance: $^{**}p < 0.01$; $^{***}p < 0.001$. UMAP, Uniform Manifold Approximation and Projection; GO, gene ontology; EdU, 5-ethynyl-2′-deoxyuridine.

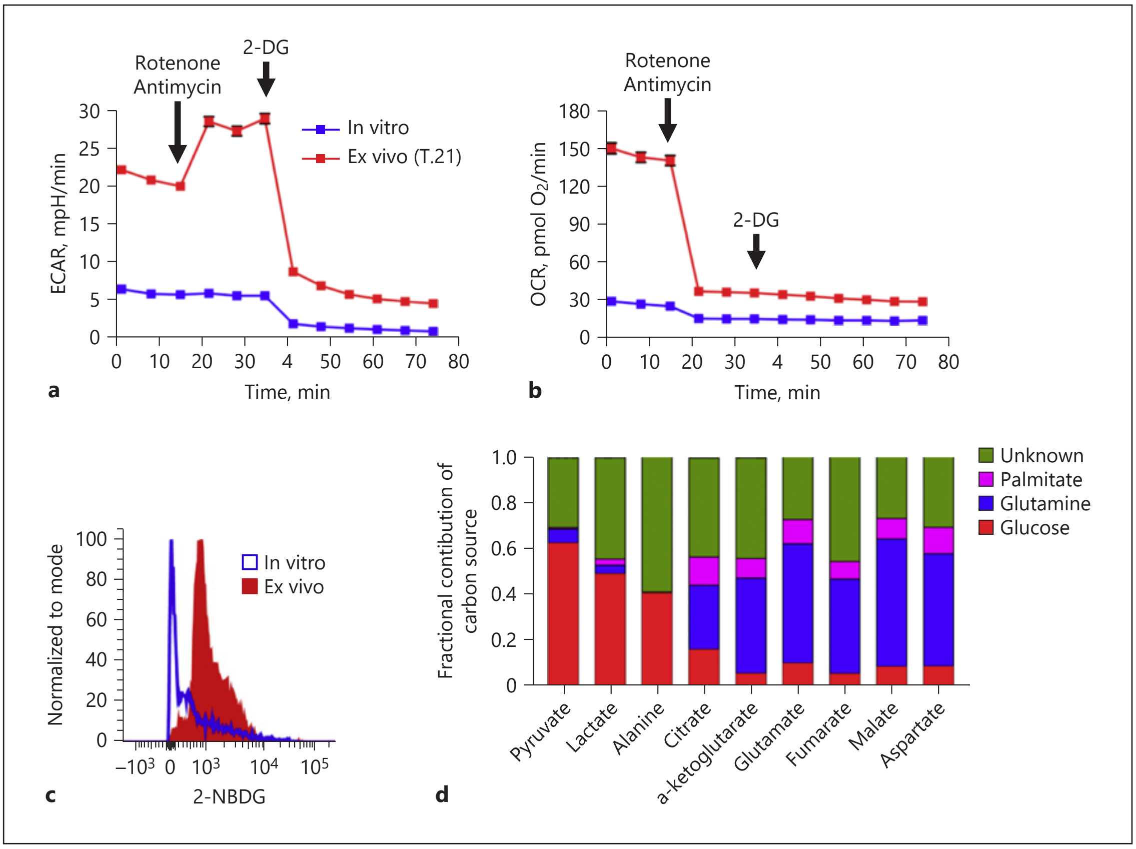


Fig. 5. Infection-driven MDSCs rely heavily on aerobic glycolysis and glucose uptake to meet their bioenergetic demands. **a, b** Extracellular flux measurements of ECAR (**a**) and OCR (**b**) in splenic Ly6C$^+$Ly6G$^+$ cells isolated from *S. aureus*-infected mice (day 21 of infection) prior to (ex vivo, red symbols) and after in vitro culture for 96 h (in vitro, blue symbols) determined by a Seahorse XF96 analyzer. **c** Representative histogram showing fluorescence 2-NBDG uptake by Ly6C$^+$Ly6G$^+$ cells isolated from *S. aureus*-infected mice (day 21 of infection) prior to (ex vivo, red histogram) and after in vitro culture for 96 h (in vitro, blue histogram). **d** Fractional contribution of different carbon sources to the total carbon of cellular metabolites in Ly6C$^+$Ly6G$^+$ cells isolated from the spleen of *S. aureus*-infected mice (day 21 of infection) determined by isotopic tracing with [U-^{13}C$_6$]-glucose, [U-^{13}C$_5$]-glutamine, or [U-^{13}C$_{16}$]-palmitate and GC-MS measurement. ECAR, extracellular acidification rate; OCR, oxygen consumption rate; 2-DG, 2-deoxy-D-glucose; 2-NBDG, 2-(N-[7-nitrobenz-2-oxa-1,3-diazol-4-yl] amino)-2-deoxyglucose.

tro-cultured Ly6C$^+$Ly6G$^+$ cells were capable of phagocytizing and killing internalized *S. aureus* (online suppl. Fig. 9). Together, these results indicate that *S. aureus* infection-driven MDSCs retained their capacity to terminally differentiate into mature myeloid cells and can undergo maturation after being removed from the spleen environment.

Infection-Driven Immature MDSCs Rely on Aerobic Glycolysis to Complete Their Maturation Process

As metabolism has been shown to influence immune cell differentiation and function [32], a better understanding of the metabolic pathways used by MDSCs to support both their energetic and biosynthetic demands could provide important information about their difficulties for undergoing terminal differentiation.

Therefore, we analyzed ECAR as a surrogate for glycolytic rate and OCR as an indicator of mitochondrial oxidative phosphorylation in immature Ly6C$^+$Ly6G$^+$ cells directly isolated from the spleen of *S. aureus*-infected mice at day 21 of infection (ex vivo) and in Ly6C$^+$Ly6G$^+$ cells after maturation in in vitro culture for 96 h (in vitro) using a Seahorse XF biochemical analyzer [33]. Ex vivo immature Ly6C$^+$Ly6G$^+$ cells showed increased basal ECAR (Fig. 5a) and OCR (Fig. 5b) as compared to in vitro-cultured mature Ly6C$^+$Ly6G$^+$ cells, indicating that immature Ly6C$^+$Ly6G$^+$ cells had increased energetic demands compared with mature Ly6C$^+$Ly6G$^+$ cells. Injection of a mixture of complex I inhibitor rotenone and complex III inhibitor antimycin A blocked mitochondrial respiration in both ex vivo immature and in vitro-cultured mature Ly6C$^+$Ly6G$^+$ cells, as evidenced by reduced OCR values (Fig. 5b). The decrease in OCR in ex vivo immature Ly6C$^+$Ly6G$^+$ cells was accompanied by an increase in ECAR (Fig. 5a), indicating that glycolysis is induced after inhibition of the electron transport chain to compensate ATP production and to meet their energy demand. Subsequent injection of 2-DG, a glucose analog that inhibits hexokinase, the first enzyme in the glycolysis pathway, resulted in substantial reduction of ECAR below the basal level in both ex vivo immature and in vitro-cultured mature Ly6C$^+$Ly6G$^+$ cells (Fig. 5a), confirming thus that acidification of the medium after inhibition of mitochondrial respiration was driven by glycolysis. These results imply that infection-driven immature MDSCs used both aerobic glycolysis and oxidative phosphorylation to support their bioenergetic demands.

In accordance with their high glycolytic activity, ex vivo immature Ly6C$^+$Ly6G$^+$ cells consumed greater amounts of glucose than in vitro-cultured mature Ly6C$^+$Ly6G$^+$ cells as determined by flow cytometry using the fluorescent D-glucose analog 2-(N-[7-nitrobenz-2-oxa-1,3-diazol-4-yl] amino)-2-deoxy-D-glucose (Fig. 5c). As ex vivo immature Ly6C$^+$Ly6G$^+$ MDSCs also utilized oxidative phosphorylation to fulfill their bioenergetic requirements, we also determined which carbon sources were used by these cells to support oxidative metabolism. For this purpose, ex vivo immature Ly6C$^+$Ly6G$^+$ cells were incubated with [U-$^{13}C_6$]-glucose, [U-$^{13}C_5$]-glutamine, or [U-$^{13}C_{16}$]-palmitate, and the labeling pattern of selected metabolites was determined by GC-MS measurement. As expected, glucose-derived carbon was incorporated not only into pyruvate but also into lactate, further corroborating the use of aerobic glycolysis by *S. aureus* infection-driven immature MDSCs (Fig. 5d). The flux of glutamine-derived carbon into TCA cycle intermediates showed that approximately 50% of carbon in the TCA cycle is derived from glutamine (Fig. 5d). This may indicate that due to the excessive conversion of glucose into lactate, MDSCs used anaplerosis of glutamine to replenish TCA cycle intermediates. The results depicted in Figure 5d also show that glucose and palmitate fueled the TCA cycle with acetyl-CoA at comparable levels.

Infection-Driven Immature MDSCs Are Reliant on Glucose Availability for Terminal Differentiation

We next investigated the reason why MDSCs failed to complete their maturation program and accumulated in the spleen of infected mice in an immature stage of differentiation. The metabolic analysis performed above indicated that MDSCs exhibited high glycolytic activity and high rate of glucose consumption that may support the elevated biosynthetic requirements associated with the maturation process. However, glucose may become rapidly depleted in the spleen microenvironment due to its rapid consumption by the increased proportion of MDSCs accumulating in this organ. Furthermore, glucose supply during infection may be insufficient as a consequence of reduced food intake by infected mice. This was particularly evident in our study since *S. aureus*-infected mice exhibited progressive weight loss during the course of infection (Fig. 6a) and exhibited reduced concentrations of glucose in blood (Fig. 6b). Based on these observations, we postulated that limited glucose availability during infection may pose a bottleneck for MDSCs to undergo complete maturation. To substantiate this assumption, we investigated if the level of glucose availability influenced the maturation status of MDSCs. For this purpose, we determined the capacity of immature MDSCs isolated from the spleen of *S. aureus*-infected mice to undergo terminal maturation under in vitro culture conditions in the presence or absence of glucose or after inhibition of glycolysis by measuring the level of expression of surface Ly6G and intracellular CCL6. The results show that approximately 90% of MDSCs underwent terminal maturation at 96 h of culture in the presence of glucose, but only 40% underwent terminal maturation in cultures where glucose was removed from the culture medium (Fig. 6c, d). Terminal differentiation of MDSCs was completely suppressed when the inhibitor of glycolysis 2-DG was added to the cultures (Fig. 6c, d). Cell viability was significantly lower in the absence of glucose or after inhibition of glycolysis with 2-DG than in cells cultured in the presence of glucose (online suppl. Fig. 10a).

The impact of glucose availability on MDSC maturation was further confirmed by a trend toward a reduction in the terminal maturation of MDSCs observed upon exposure to decreased glucose concentrations (online suppl. Fig. 10b, c). Furthermore, only MDSCs cultured in vitro for 96 h in the presence of 2 mg/mL of glucose exhibited significantly lower capacity to inhibit T-cell proliferation than immature MDSCs prior to culture (online suppl. Fig. 10d).

Based on these observations, we next investigated the effect of increasing glucose availability in vivo by supplementing *S. aureus*-infected mice with 10% glucose in drinking water for 10 days on the maturation of splenic

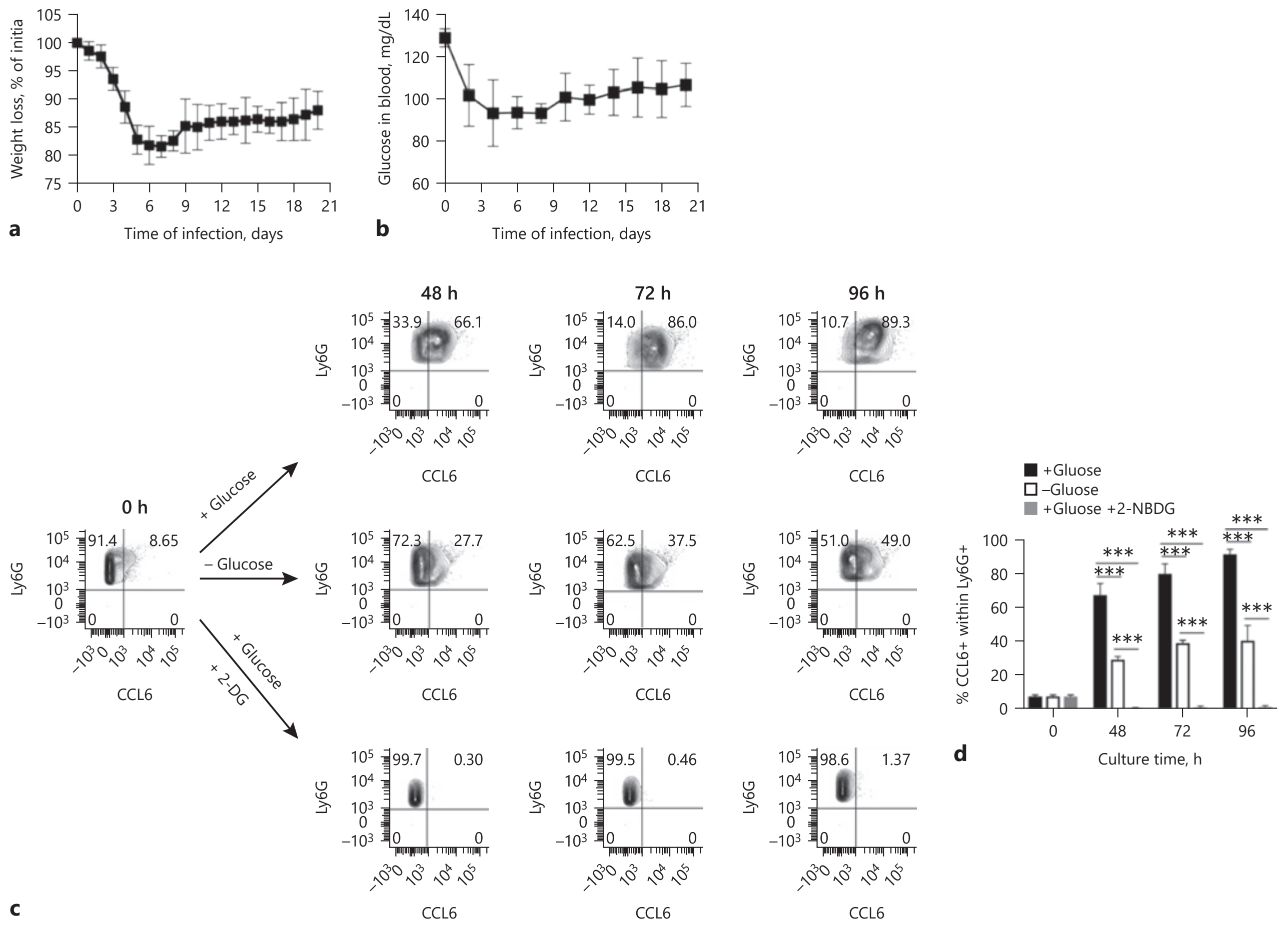


Fig. 6. Infection-driven MDSCs are reliant on glucose availability for terminal differentiation. **a** Percentage of weight loss of *S. aureus*-infected mice during the course of infection. **b** Glucose levels in the blood of *S. aureus*-infected mice during the course of infection. **c** Representative flow cytometry contour plots showing the expression of CCL6 within Ly6G$^+$ cells isolated from the spleen of *S. aureus*-infected mice at day 21 of infection and cultured in vitro either in the presence or absence of glucose or after inhibition of glycolysis by addition of 2-DG. **d** Percentage of CCL6$^+$ cells within the Ly6G$^+$ cell population at progressive times during in vitro culture either in the presence (black bars) or absence (white bars) of glucose or after inhibition of glycolysis with 2-DG (gray bars). Results are presented as the mean ± SD of biological replicates (N = 5 for **a**, **b** and N = 3 for **c**, **d**) and are representative of 3 independent experiments. Statistical significance: ***$p < 0.001$. 2-DG, 2-deoxy-D-glucose.

MDSCs. Glucose-treated mice exhibited significantly less body weight loss (Fig. 7a) and higher levels of glucose in blood (Fig. 7b) than untreated mice. Importantly, glucose-treated mice exhibited a significantly higher number of mature neutrophils (Ly6G$^+$CCL6$^+$) in the spleen than the untreated group (Fig. 7c, d). However, glucose supplementation did not affect the bacterial loads in infected organs (Fig. 7e).

Discussion

In this study, we integrated scRNA-seq analysis and functional metabolic profiling to gain a deeper understanding of the generation and physiology of MDSCs in the context of *S. aureus* chronic infection. The results of the scRNA-seq analysis emphasize the vast heterogeneity and functional diversity of infection-driven MDSCs, which comprise a continuous spectrum of cell popula-

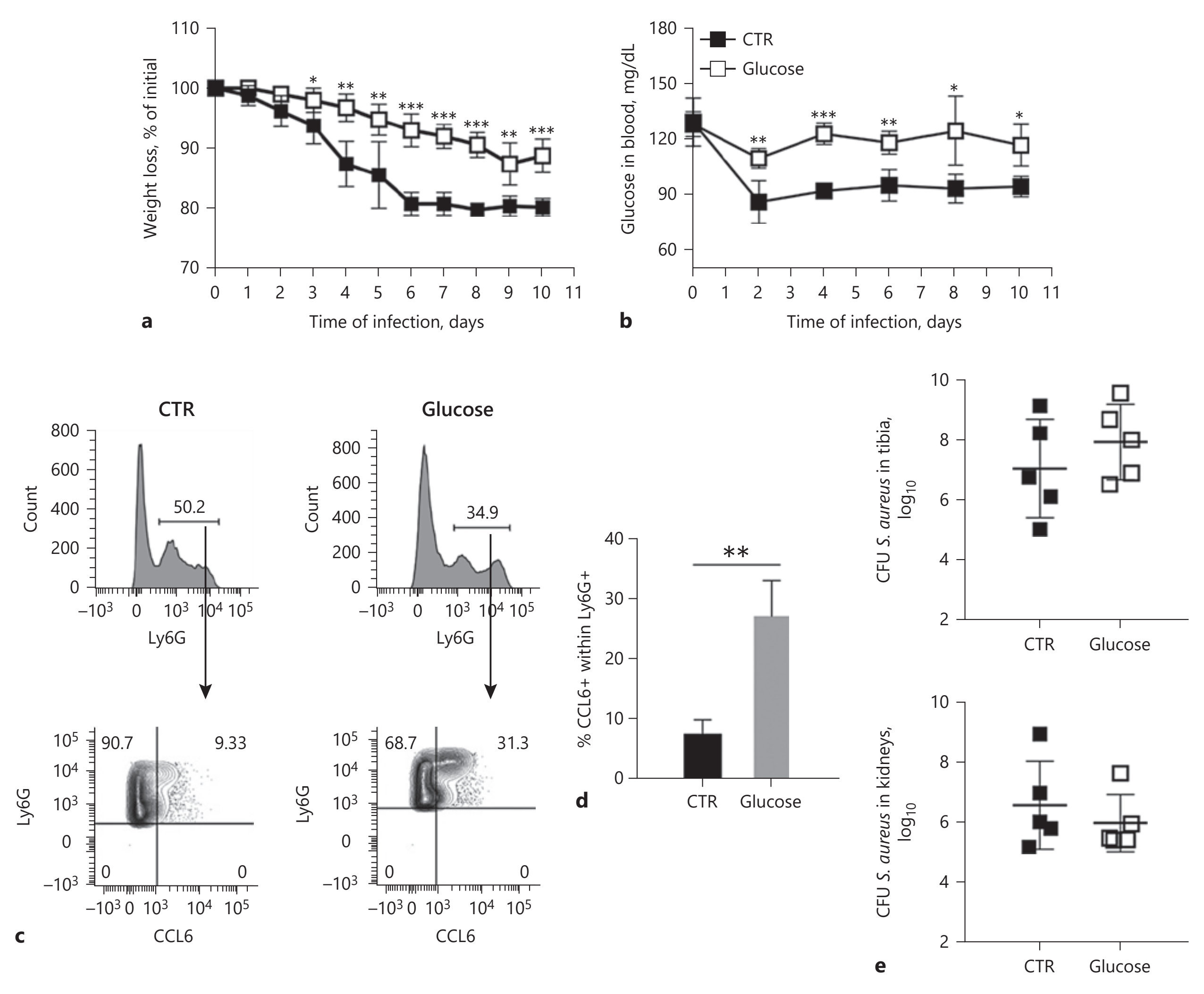


Fig. 7. Supplying *S. aureus*-infected mice with glucose accelerated differentiation of MDSCs in vivo. Percent weight loss (**a**) and glucose levels in the blood (**b**) of *S. aureus*-infected mice fed with glucose-supplemented (white symbols) or nonsupplemented (black symbols) drinking water for 10 days. **c** Representative flow cytometry analysis showing the expression of CCL6 within the Ly6G$^+$ cell population in the spleen of *S. aureus*-infected mice (day 10 of infection) fed with glucose-supplemented (right panels) or nonsupplemented (left panels) drinking water. **d** Percentage of CCL6$^+$ cells within the Ly6G$^+$ cell population in the spleen of *S. aureus*-infected mice (day 10 of infection) that received glucose-supplemented (gray bar) or nonsupplemented (black bar) drinking water. **e** Bacterial load in the tibia (upper panel) and kidneys (lower panel) of *S. aureus*-infected mice (day 10 of infection) that received glucose-supplemented (white symbols) or nonsupplemented (black symbols) drinking water. Results are presented as the mean ± SD of collated data from 3 independent experiments ($N = 9$). Statistical significance: $^{**}p < 0.01$; $^{***}p < 0.001$.

tions representing transitions between different states of granulocyte differentiation, ranging from promyelocytes to mature segmented neutrophils. In mice, the phenotypic distinction between mature neutrophils and immature progenitors has been difficult, and no phenotypic marker has been identified so far that enables to precisely separate these populations. The scRNA-seq analysis performed in our study has identified surface expression of Ly6G and intracellular expression of CCL6 as phenotypic markers that enable to distinguish mature neutrophils (Ly6G$^+$CCL6$^+$) from immature granulocytes precursors (Ly6G$^+$CCL6$^-$) by flow cytometry. However, CCL6 has the drawback to be intracellular and its detection requires cell fixation and therefore does not allow recovery of live cells for subsequent functional studies. To date, relatively little information is available about the function and expression of CCL6. In mice, CCL6 is expressed in cells of granulocyte and macrophage lineages and is highly induced upon stimulation with GM-CSF [34]. Other studies have reported a role for this chemokine in inflammation and tissue remodeling [18]. The human homolog of CCL6 has not yet been identified, an issue that deserves further attention in future studies.

We found that infection-driven MDSCs originate from both the bone marrow, most probably as a consequence of emergency granulopoiesis, and in situ within the spleen from extramedullary hematopoiesis. Emergency granulopoiesis induced by infection and the concomitant release of immature myeloid cells in the circulation seem to be a mechanism triggered to restore the neutrophil pool that is rapidly depleted from peripheral blood due to extravasation from the bloodstream into the sites of infection [35]. We also investigated the reason why infection-driven MDSCs fail to undergo terminal maturation and accumulate in an early stage of differentiation. We found that, similar to other pathological conditions [29–31], MDSCs in chronically infected mice are not irreversibly arrested in an immature stage and still retain their capacity to terminally differentiate into mature myeloid cells under in vitro culture conditions. Since metabolism plays an important role in immune cell differentiation and function [32], we explored a possible connection between the metabolic demands of MDSCs and their difficulties for undergoing terminal differentiation. The metabolic flux and isotope tracing analysis performed in our study indicate that infection-driven MDSCs use both aerobic glycolysis and oxidative phosphorylation for ATP production. The benefits of aerobic glycolysis for MDSCs may be both bioenergetics and biosynthesis. During glycolysis, 1 molecule of glucose is converted into 2 molecules of pyruvate with the concomitant generation of 2 molecules of ATP. Generally, pyruvate enters the mitochondria where it is converted into acetyl-CoA, which enters the TCA cycle. In certain circumstances, as those observed in our study in the infection-driven MDSCs, a proportion of pyruvate can be also converted into lactate in the cytosol by lactate dehydrogenase with concomitant regeneration of NAD^+ from NADH that keeps fueling the glycolytic pathway. Therefore, although the ATP generated per glucose molecule during aerobic glycolysis is rather low, a very high glycolytic flux like that detected in MDSCs from infected mice can produce high levels of ATP. Furthermore, in addition to ATP generation, glycolysis may provide biosynthetic intermediates to support the synthesis of important molecular building blocks required by MDSCs for undergoing differentiation and maturation.

Based on the results of the metabolic analysis, we speculated that MDSCs may rely on high glycolytic activity to complete their maturation process and that glucose limitation in the spleen microenvironment, possibly due to its rapid consumption by MDSCs and/or to a decline in glucose blood concentrations observed in infected mice, could prevent their complete maturation. This assumption proved to be true since MDSCs isolated from infected mice were capable of undergoing terminal differentiation under in vitro conditions when glucose was added to the culture medium, but differentiation was hampered in the absence of glucose or when glycolysis was inhibited. Furthermore, we could show that supplying *S. aureus*-infected mice with glucose in the drinking water resulted in improved blood glucose levels, ameliorated weight loss, and accelerated differentiation of immature myeloid cells in the spleen. However, the bacterial loads in the organs of infected mice were not affected by glucose supplementation. One possible explanation for this phenomenon could be that a proportion of MDSCs were still present in the glucose-treated mice that could interfere with effective T-cell responses. Furthermore, as glucose is the principal energy source of *S. aureus*, increasing glucose levels in treated mice could enhance *S. aureus* pathogenesis as reported by previous studies [36–39].

In summary, the results of our study have uncovered a link between metabolic alterations induced by infection and the accumulation of MDSCs.

Acknowledgments

The authors would like to thank S. Beyer and P. Hagendorff for excellent technical assistance.

Statement of Ethics

Animal experiments were performed in strict accordance with the German regulations of the Society for Laboratory Animal Science (GV-SOLAS) and the European Health Law of the Federation of Laboratory Animal Science Associations (FELASA). All experiments were approved by the ethical board Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit, Oldenburg, Germany (LAVES; permit Nos. 33.9-42502-04-10/0296 and 33.19-42502-04-19/3307).

Conflict of Interest Statement

The authors declare no competing interests.

Funding Sources

This work was supported in part by internal funding provided by the Helmholtz Centre for Infection Research.

O.D. and A.-E.S. are supported by the Deutsche Forschungsgemeinschaft (DFG) GRK 2157 3D Tissue Models for Studying Microbial Infections by Human Pathogens.

Author Contributions

O.G. and E.M. conceived and designed the study, performed the in vivo experiments, and analyzed the data. O.D., R.G., and A.-E.S. carried out the scRNA-seq procedure and analysis. A.H. and K.H. performed the metabolic analysis. A.B. performed histological evaluation. E.M. wrote the manuscript with input from all authors.

Data Availability Statement

The raw bulk RNA-seq data and raw single-cell RNA-seq data generated during this study are available at the European Nucleotide Archive (ENA) under the accession number: PRJEB38066.

References

1 Gabrilovich DI. Myeloid-derived suppressor cells. Cancer Immunol Res. 2017 Jan;5(1):3–8.
2 Veglia F, Perego M, Gabrilovich D. Myeloid-derived suppressor cells coming of age. Nat Immunol. 2018 Feb;19(2):108–19.
3 Gabrilovich DI, Nagaraj S. Myeloid-derived suppressor cells as regulators of the immune system. Nat Rev Immunol. 2009 Mar;9(3):162–74.
4 Bronte V, Brandau S, Chen SH, Colombo MP, Frey AB, Greten TF, et al. Recommendations for myeloid-derived suppressor cell nomenclature and characterization standards. Nat Commun. 2016 Jul 6;7:12150.
5 Cuenca AG, Delano MJ, Kelly-Scumpia KM, Moreno C, Scumpia PO, Laface DM, et al. A paradoxical role for myeloid-derived suppressor cells in sepsis and trauma. Mol Med. 2011 Mar–Apr;17(3–4):281–92.
6 Goldmann O, Beineke A, Medina E. Identification of a novel subset of myeloid-derived suppressor cells during chronic staphylococcal infection that resembles immature eosinophils. J Infect Dis. 2017 Dec 12;216(11):1444–51.
7 Gabrilovich DI, Ostrand-Rosenberg S, Bronte V. Coordinated regulation of myeloid cells by tumours. Nat Rev Immunol. 2012 Mar 22;12(4):253–68.
8 Medina E, Hartl D. Myeloid-derived suppressor cells in infection: a general overview. J Innate Immun. 2018;10(5–6):407–13.
9 Heim CE, Vidlak D, Scherr TD, Kozel JA, Holzapfel M, Muirhead DE, et al. Myeloid-derived suppressor cells contribute to Staphylococcus aureus orthopedic biofilm infection. J Immunol. 2014 Apr 15;192(8):3778–92.
10 Heim CE, Vidlak D, Kielian T. Interleukin-10 production by myeloid-derived suppressor cells contributes to bacterial persistence during Staphylococcus aureus orthopedic biofilm infection. J Leukoc Biol. 2015 Dec;98(6):1003–13.
11 Heim CE, Vidlak D, Scherr TD, Hartman CW, Garvin KL, Kielian T. IL-12 promotes myeloid-derived suppressor cell recruitment and bacterial persistence during Staphylococcus aureus orthopedic implant infection. J Immunol. 2015 Apr 15;194(8):3861–72.
12 Tebartz C, Horst SA, Sparwasser T, Huehn J, Beineke A, Peters G, et al. A major role for myeloid-derived suppressor cells and a minor role for regulatory T cells in immunosuppression during Staphylococcus aureus infection. J Immunol. 2015 Feb 1;194(3):1100–11.
13 Heim CE, Vidlak D, Odvody J, Hartman CW, Garvin KL, Kielian T. Human prosthetic joint infections are associated with myeloid-derived suppressor cells (MDSCs): implications for infection persistence. J Orthop Res. 2018 Jun;36(6):1605–13.
14 Heim CE, West SC, Ali H, Kielian T. Heterogeneity of Ly6G(+) Ly6C(+) myeloid-derived suppressor cell infiltrates during Staphylococcus aureus biofilm infection. Infect Immun. 2018 Dec;86(12):e00684-18.
15 Tong SY, Davis JS, Eichenberger E, Holland TL, Fowler VG Jr. Staphylococcus aureus infections: epidemiology, pathophysiology, clinical manifestations, and management. Clin Microbiol Rev. 2015 Jul;28(3):603–61.
16 Law AMK, Valdes-Mora F, Gallego-Ortega D. Myeloid-derived suppressor cells as a therapeutic target for cancer. Cells. 2020 Feb 27;9(3):561.
17 Pike LS, Smift AL, Croteau NJ, Ferrick DA, Wu M. Inhibition of fatty acid oxidation by etomoxir impairs NADPH production and increases reactive oxygen species resulting in ATP depletion and cell death in human glioblastoma cells. Biochim Biophys Acta. 2011 Jun;1807(6):726–34.
18 Coelho AL, Schaller MA, Benjamim CF, Orlofsky AZ, Hogaboam CM, Kunkel SL. The chemokine CCL6 promotes innate immunity via immune cell activation and recruitment. J Immunol. 2007 Oct 15;179(8):5474–82.
19 Sapcariu Sean C, Kanashova Tamara, Weindl Daniel, Ghelfi Jenny, Dittmar Gunnar, Hiller Karsten. Simultaneous extraction of proteins and metabolites from cells in culture. MethodsX. 2215-0161. 2014;1:74–80. 26150938.
20 Kowalczyk MS, Tirosh I, Heckl D, Rao TN, Dixit A, Haas BJ, et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome Res. 2015 Dec;25(12):1860–72.
21 Horst SA, Hoerr V, Beineke A, Kreis C, Tuchscherr L, Kalinka J, et al. A novel mouse model of Staphylococcus aureus chronic osteomyelitis that closely mimics the human infection: an integrated view of disease pathogenesis. Am J Pathol. 2012 Oct;181(4):1206–14.
22 Ziegler C, Goldmann O, Hobeika E, Geffers R, Peters G, Medina E. The dynamics of T cells during persistent Staphylococcus aureus infection: from antigen-reactivity to in vivo anergy. EMBO Mol Med. 2011 Nov;3(11):652–66.
23 Cowland JB, Borregaard N. Granulopoiesis and granules of human neutrophils. Immunol Rev. 2016 Sep;273(1):11–28.
24 Cowland JB, Borregaard N. The individual regulation of granule protein mRNA levels during neutrophil maturation explains the heterogeneity of neutrophil granules. J Leukoc Biol. 1999 Dec;66(6):989–95.
25 Theilgaard-Monch K, Jacobsen LC, Borup R, Rasmussen T, Bjerregaard MD, Nielsen FC, et al. The transcriptional program of terminal granulocytic differentiation. Blood. 2005 Feb 15;105(4):1785–96.
26 Rorvig S, Ostergaard O, Heegaard NH, Borregaard N. Proteome profiling of human neutrophil granule subsets, secretory vesicles, and cell membrane: correlation with transcriptome profiling of neutrophil precursors. J Leukoc Biol. 2013 Oct;94(4):711–21.

27 Bjerregaard MD, Jurlander J, Klausen P, Borregaard N, Cowland JB. The in vivo profile of transcription factors during neutrophil differentiation in human bone marrow. Blood. 2003 Jun 1;101(11):4322–32.

28 Griffiths JA, Scialdone A, Marioni JC. Using single-cell genomics to understand developmental processes and cell fate decisions. Mol Syst Biol. 2018 Apr 16;14(4):e8046.

29 Bronte V, Apolloni E, Cabrelle A, Ronca R, Serafini P, Zamboni P, et al. Identification of a CD11b(+)/Gr-1(+)/CD31(+) myeloid progenitor capable of activating or suppressing CD8(+) T cells. Blood. 2000 Dec 1;96(12):3838–46.

30 Kusmartsev S, Gabrilovich DI. Inhibition of myeloid cell differentiation in cancer: the role of reactive oxygen species. J Leukoc Biol. 2003 Aug;74(2):186–96.

31 Li Q, Pan PY, Gu P, Xu D, Chen SH. Role of immature myeloid Gr-1+ cells in the development of antitumor immunity. Cancer Res. 2004 Feb 1;64(3):1130–9.

32 Buck MD, Sowell RT, Kaech SM, Pearce EL. Metabolic instruction of immunity. Cell. 2017 May 4;169(4):570–86.

33 Zhang J, Nuebel E, Wisidagama DR, Setoguchi K, Hong JS, Van Horn CM, et al. Measuring energy metabolism in cultured cells, including human pluripotent stem cells and differentiated cells. Nat Protoc. 2012 May 10;7(6):1068–85.

34 Orlofsky A, Berger MS, Prystowsky MB. Novel expression pattern of a new member of the MIP-1 family of cytokine-like genes. Cell Regul. 1991 May;2(5):403–12.

35 Manz MG, Boettcher S. Emergency granulopoiesis. Nat Rev Immunol. 2014 May;14(5):302–14.

36 Rich J, Lee JC. The pathogenesis of Staphylococcus aureus infection in the diabetic NOD mouse. Diabetes. 2005 Oct;54(10):2904–10.

37 Waldrop R, McLaren A, Calara F, McLemore R. Biofilm growth has a threshold response to glucose in vitro. Clin Orthop Relat Res. 2014 Nov;472(11):3305–10.

38 Equils O, da Costa C, Wible M, Lipsky BA. The effect of diabetes mellitus on outcomes of patients with nosocomial pneumonia caused by methicillin-resistant Staphylococcus aureus: data from a prospective double-blind clinical trial comparing treatment with linezolid versus vancomycin. BMC Infect Dis. 2016 Sep 6;16:476.

39 Vitko NP, Grosser MR, Khatri D, Lance TR, Richardson AR. Expanded glucose import capability affords Staphylococcus aureus optimized glycolytic flux during infection. mBio. 2016 Jun 21;7(3):e00296-16.

GUT MICROBES
2023, VOL. 15, NO. 1, 2186109
https://doi.org/10.1080/19490976.2023.2186109

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

A primary cell-based *in vitro* model of the human small intestine reveals host olfactomedin 4 induction in response to *Salmonella* Typhimurium infection

Thomas Däullary[a,b], Fabian Imdahl[c], Oliver Dietrich[c], Laura Hepp[a], Tobias Krammer[c], Christina Fey[d], Winfried Neuhaus[e,f], Marco Metzger[a,d,g], Jörg Vogel[c,h], Alexander J. Westermann[c,h], Antoine-Emmanuel Saliba[c], and Daniela Zdzieblo[a,d,g]

[a]Chair of Tissue Engineering and Regenerative Medicine, University Hospital Würzburg (UKW), Würzburg, Germany; [b]Faculty of Biology, Biocenter, Chair of Microbiology, Julius-Maximilians-Universität Würzburg (JMU), Würzburg, Germany; [c]Helmholtz-Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), Würzburg, Germany; [d]Fraunhofer Institute for Silicate Research (ISC), Translational Center Regenerative Therapies (TLC-RT), Würzburg, Germany; [e]Austrian Institute of Technology (AIT), Vienna, Austria; [f]Department of Medicine, Faculty of Medicine and Dentistry, Danube Private University (DPU), Krems, Austria; [g]Fraunhofer Institute for Silicate Research, Project Center for Stem Cell Process Engineering, Würzburg, Germany; [h]Institute for Molecular Infection Biology (IMIB), University of Würzburg, Würzburg, Germany

ABSTRACT
Infection research largely relies on classical cell culture or mouse models. Despite having delivered invaluable insights into host-pathogen interactions, both have limitations in translating mechanistic principles to human pathologies. Alternatives can be derived from modern Tissue Engineering approaches, allowing the reconstruction of functional tissue models *in vitro*. Here, we combined a biological extracellular matrix with primary tissue-derived enteroids to establish an *in vitro* model of the human small intestinal epithelium exhibiting *in vivo*-like characteristics. Using the foodborne pathogen *Salmonella enterica* serovar Typhimurium, we demonstrated the applicability of our model to enteric infection research in the human context. Infection assays coupled to spatio-temporal readouts recapitulated the established key steps of epithelial infection by this pathogen in our model. Besides, we detected the upregulation of olfactomedin 4 in infected cells, a hitherto unrecognized aspect of the host response to *Salmonella* infection. Together, this primary human small intestinal tissue model fills the gap between simplistic cell culture and animal models of infection, and shall prove valuable in uncovering human-specific features of host-pathogen interplay.

ARTICLE HISTORY
Received 13 November 2022
Revised 3 February 2023
Accepted 22 February 2023

KEYWORDS
Intestinal enteroids; biological scaffold; *Salmonella* Typhimurium; OLFM4; NOTCH; filamentous *Salmonella* Typhimurium; bacterial migration; bacterial virulence; 3D tissue model; olfactomedin 4; infection

Introduction

The gastrointestinal mucosa forms a barrier that shields the human body from a variety of contaminants and pathogenic agents. However, enteric pathogens have evolved specific mechanisms to overcome this barrier[1]. A detailed understanding of the underlying virulence mechanisms is therefore important for the development of effective therapeutic treatments of infectious diseases. Importantly, the structural, biological, and physico-chemical properties of the gastrointestinal mucosa affect the infection process and the associated host responses. Therefore, the models used in infection research need to adequately recapitulate these aspects – ideally, in a human context.

In the last decade, pluripotent stem cell (PSC)-derived organoids or adult stem cell (ASC)-derived enteroids gained popularity for modeling the human intestinal epithelium *in vitro*[2–7]. Cultured in specified media and embedded in Matrigel® — a biological extracellular matrix (ECM) — they exhibit certain characteristics of the native tissue, yet show an inverted architecture with the apical cell surface facing the lumen of the organoid. Alternatively, epithelial stem cells can be seeded and grown as two-dimensional (2D) monolayers on top of synthetic scaffolds[8,9]. This enables apical and basolateral accessibility, but requires coating with an ECM-like substance such as Matrigel®, collagen, or gelatin, for cellular attachment and growth. Both enteroid/

CONTACT Daniela Zdzieblo daniela.zdzieblo@uni-wuerzburg.de Institute for Silicate Research - Project Center for StemCell Process Engineering and Translational Center Regenerative Therapies, University Hospital Würzburg - Chair of Tissue Engineering and Regenerative Medicine, Röntgenring 11, Würzburg 97070

Supplemental data for this article can be accessed online at https://doi.org/10.1080/19490976.2023.2186109

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

organoid or 2D monolayer-based models of the small intestinal epithelium have contributed significantly to our understanding of fundamental principles of gut biology in general and tissue-specific immunity in particular[8–10]. Likewise, when used as host models of infection, they improved our understanding of the pathogenesis of diverse viral and bacterial pathogens[9,11–15].

We have previously combined a biological, organ-specific ECM scaffold (small intestinal submucosa; SIS) with human primary tissue-derived enteroids in a Transwell®-like system[16]. The resulting *in vitro* model of the human small intestinal epithelium displays an *in vivo*-like compartmentalization with high physiological performance, recapitulating the apical-basolateral polarity of the native tissue[16]. However, the complex setup of the previously published model in a closed bioreactor with dynamic flow conditions restricts the versatile application in infectious disease research, since throughput and standardization are limited in addition to technical challenges.

In the present work, we therefore adapted the model setup to a static and open system with adjusted cultivation conditions, which facilitates handling and increases the throughput of the *in vitro* models. The optimized human small intestinal epithelium model (hITM) is characterized by a monolayer with appropriate barrier functions and resembles *in vivo*-like cellular phenotypes as demonstrated by immunohistochemistry and single-cell RNA-seq. Using *Salmonella enterica* serovar Typhimurium (*S*Tm) as a representative enteric pathogen, we demonstrated the suitability of the hITM as an *in vitro* infection model. Specifically, the hITM model was effectively infected by *S*Tm, and recapitulated key steps of the infection process. Besides, we revealed the enhanced expression of olfactomedin 4 (OLFM4) in *S*Tm-infected cells, a previously unrecognized aspect of the host cell response to this pathogen. Together, this work introduces a primary cell-based *in vitro* system of the human small intestinal epithelium and illustrates its use as a host model for enteric infection research.

Results

The hITM resembles morphological, morphometric, and structural key features of the human small intestinal epithelium

We introduced a revised version of our previously published Transwell®-like model of the hITM, combining a porcine-derived ECM scaffold (SIS) with intestinal epithelial cells (IECs) isolated from human enteroid cultures (Figure 1a). Major novelties in the set-up of the hITM are as follows: 1) the extended time frames for the controlled proliferation of IECs on the ECM scaffold, 2) the used culture medium during proliferation as well as 3) the prolonged differentiation time to derive functional cellular subtypes of the intestinal epithelium under static cell culture conditions. In addition, we also extended our analysis panel to comprehensively characterize the hITM regarding morphological, morphometric, functional, molecular, and cellular features. Human intestinal tissue biopsies or enteroid cultures grown in Matrigel® were used as references as indicated below.

To get first insights into the tissue structure of the hITM, histological analyses were performed. Native intestinal tissue samples served as controls. Alcian blue staining revealed typical compartmentalization of the native small intestinal epithelium characterized by alternating villus and crypt domains with a highly polarized columnar monolayer in the villus region and the existence of mucin positive cells (Figure 1 a1, **a2**). Similar, to the native tissue, a confluent and polarized IEC monolayer formed in the hITM (Figure 1**a3**) with a mucin layer covering its apical region (Figure 1 **a4**, **a5**). In comparison to the native tissue, mucin inclusion bodies and mucin positive cells appeared to be differently distributed in the hITM.

As cellular polarity is key for the functionality of the intestinal epithelium, we next evaluated this feature within the hITM by quantitatively determining the cell height based on F-actin cytoskeleton staining and observed a mean cell height of 32.76 µm (±5.99 µm) which is within the range of the cell height reported for the native tissue (30–43 µm, according to PT and MN, 1993[17] (Figure 1 **b**)). Furthermore, the polarized IEC monolayer developed a tight barrier with a mean transepithelial

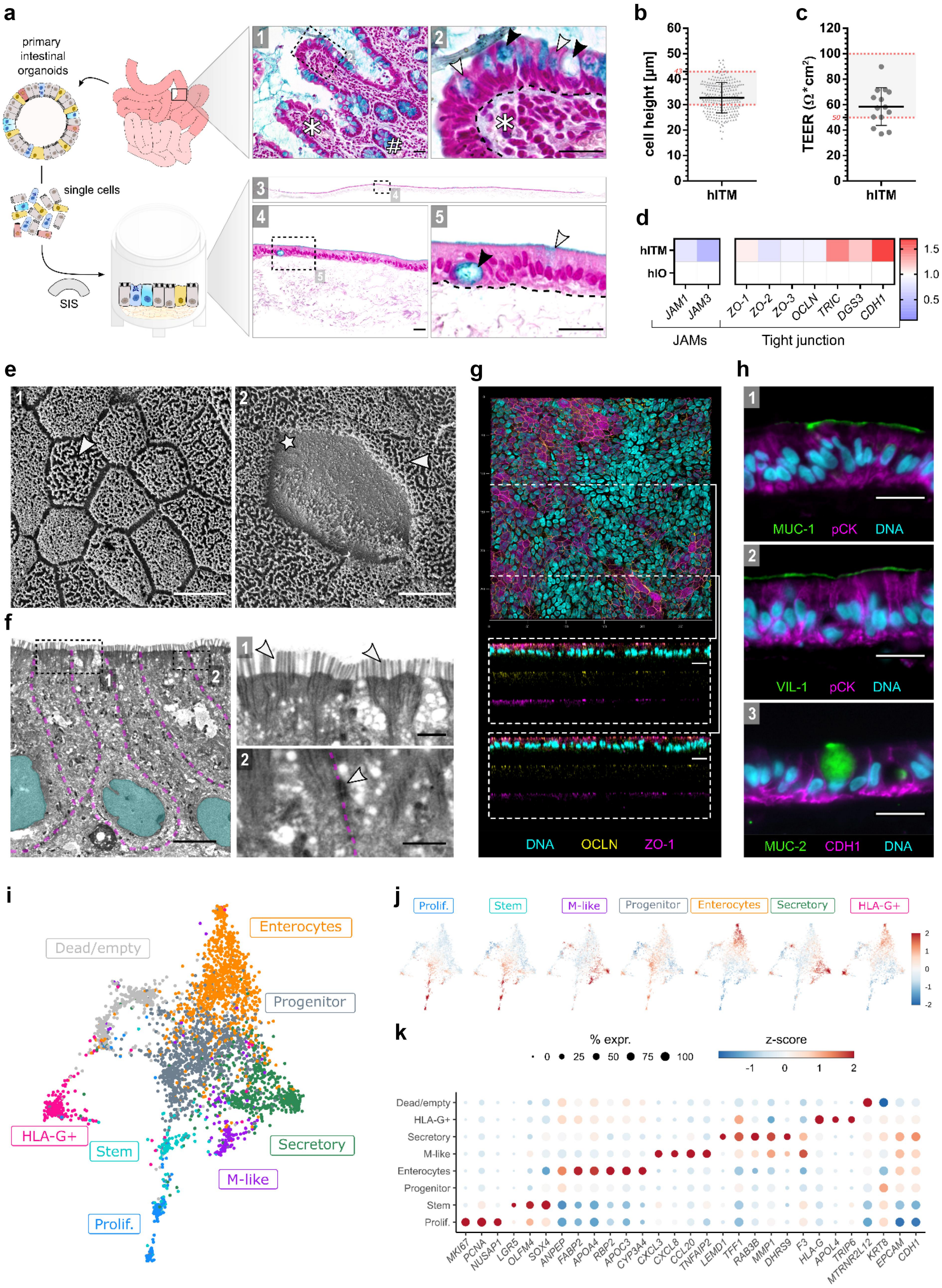

a
primary intestinal organoids
single cells
SIS
1
2
3
4
5
b
cell height [µm]
60
50
40
30
20
10
0
hITM
c
TEER (Ω*cm²)
120
100
80
60
40
20
0
hITM
d
hITM
hIO
JAM1
JAM3
ZO-1
ZO-2
ZO-3
OCLN
TRIC
DGS3
CDH1
1.5
1.0
0.5
JAMs
Tight junction
e
f
g
DNA
OCLN
ZO-1
h
MUC-1
pCK
DNA
VIL-1
pCK
DNA
MUC-2
CDH1
DNA
i
Enterocytes
Dead/empty
Progenitor
HLA-G+
Stem
Secretory
M-like
Prolif.
j
Prolif.
Stem
M-like
Progenitor
Enterocytes
Secretory
HLA-G+
2
1
0
-1
-2
k
% expr.
0
25
50
75
100
z-score
-1
0
1
2
Dead/empty
HLA-G+
Secretory
M-like
Enterocytes
Progenitor
Stem
Prolif.
MKI67
PCNA
NUSAP1
LGR5
OLFM4
SOX4
ANPEP
FABP2
APOA4
RBP2
APOC3
CYP3A4
CXCL3
CXCL8
CCL20
TNFAIP2
LEMD1
TFF1
RAB3B
MMP1
DHRS9
F3
HLA-G
APOL4
TRIP6
MTRNR2L12
KRT8
EPCAM
CDH1

Figure 1. IECs develop a monolayer on a biological scaffold with *in vivo*-like tissue characteristics in a Transwell®-like setup. **Aa left** Experimental workflow: IECs were isolated from small intestinal biopsies followed by expansion as enteroids in 3D Matrigel®-embedded culture as previously reported[7]. for hITM establishment, dissociated enteroids were seeded as single cells on a biological SIS scaffold in a Transwell®-like setup. **a right** Representative histochemical Alcian blue staining of the native human small intestinal mucosa (**top**, *n* = 3) in comparison to the epithelial monolayer in the hITM (**bottom**, *n* = 3). Glycosylated proteins, typically mucins, are stained in blue, whereas nuclei are stained in deep purple and cell cytoplasm in light purple. Scale bar: 25 µm. **a1** Cross-section of the native human small mucosa with compartmentalization into villus (*) and crypt (#) domains. Scale bar: 25 µm. **a2** Magnification of the villus region shown in a1 highlighting the columnar structure of the cells organized as epithelial monolayer. The

electrical resistance (TEER) value of 58.46 Ω*cm^2 (±18.83 Ω*cm^2) that is similar to the TEER value published for native tissue samples (50–100 Ω*cm^2, according to Srinivasan et al., 2015[18] (Figure 1 c)). In line with the TEER measurements, gene expression analysis of junction associated molecules (*JAMs*; *JAM 1*: 0,79-fold; *JAM 2*: 0.42-fold) and further tight junction (TJ) associated genes (Zonula Occludens, *ZO-1*: 1.06-fold; *ZO-2*: 0.76-fold, *ZO-3*: 0.89-fold; Occludin, *OCLN*: 0.90-fold) demonstrated comparable expression levels in cells isolated from the hITM and human enteroid samples that have been used as reference in the high-throughput qRT-PCR analysis. Of note, expression values obtained for Tricellulin (*TRIC*), Desmoglein-3 (*DSG3*) and Cadherin-1 (*CDH1*) were increased in hITM-derived cells (*TRIC*: 1.44-fold, *DSG3*: 1.21-fold *CDH1*: 1.68-fold) compared to enteroids of human origin.

Next, we performed electron microscopy (EM) to gain insights into the ultrastructural organization of the hITM (Figures 1 e, 1 e) and observed a polygonal cobblestone-like arrangement of the IEC monolayer (Figure 1 e1) and its decoration with typical microvilli structures (Figure 1 f1). Some cells lacked microvilli characterizing them as Microfold (*M*-) cells (Figure 1 e2)[19]. In addition, ultrastructural analyses highlighted desmosomes as well as TJ formations between cells in the apical region (Figure 1 f2), that were verified via immunohistochemical (IHC) stainings against ZO-1 and OCLN (Figure 1 g). Additionally, IHC analyses revealed a "honeycomb"-like pattern between individual cells and a concentrated localization of ZO-1 and OCLN signals at the apical cell borders (Figure 1 g).

In the final step of hITM characterization, we evaluated whether the adapted differentiation in the novel protocol leads to an *in vivo*-like cellular diversity in the hITM model. Thus, we first investigated the expression and localization of typical intestinal proteins via IHC (Figure 1 h) and could demonstrate positive signals for Mucin-1 (MUC-1; Figure 1 h1), a transmembrane mucin expressed in the gut epithelium on the apical side of individual cells and cell clusters[20], Villin-1 (VIL-1; Figure 1 h2), a brush border protein localized to the microvilli, broadly but exclusively expressed on the apical cell surface[21,22],

white arrowheads depict individual cells filled with mucins. The black arrowheads point toward mucin inclusion bodies and the dashed line indicates separation of the epithelial layer from the underlying connective tissue. Scale bar: 25 µm. **a3** Cross-section of the hITM showing the epithelial cell layer in deep purple and the underlying SIS scaffold in light purple. Scale bar: 25 µm. **a4** Magnification of the hITM demonstrating IECs lining the apical surface of the SIS as a confluent monolayer. Scale bar: 25 µm. **a5** Magnification of A4 highlighting the apical-basolateral polarity of the epithelial monolayer in the hITM, the presence of a mucin layer on top of the IECs as well as representative mucin+ cells (white arrowhead marks a cell filled with mucins; black arrowhead points at a cell with a mucin inclusion body). Scale bar: 25 µm. **b** Mean cellular height calculated from the measurements of 256 individual IECs within the hITM (*n* = 4, 64 cells per biological replicate); dotted red lines mark the physiological range of 30–43 µm, according to PT and MN, 1993[17]. **c** Mean TEER value measured in the hITM (*n* = 3, each with >4 independent technical replicates); dotted red lines indicate the physiological range of 50–100 Ω*cm^2 reported for the native human intestinal epithelium according to Srinivasan et al., 2015[18]. **d** Heat map showing the gene expression profile for *JAM1* and *3* as well as tight junction associated proteins (*ZO −1,-2,-3*, *OCLN*, *TRIC*, *DSG3*, *CDH1)*. Expression in the hITM was compared to the expression in human enteroids. Data were obtained by high-throughput qPCR and were normalized to the expression level observed for enteroid samples (*n* = 1). **e** Representative SEM images of the hITM apical surface revealing IECs decorated with microvilli (**e1**; white arrow head) as well as M-like cells characterized by typical folds and the absence of microvilli (**e2**; white star) (*n* = 2). Scale: 1 µm. **f** Representative TEM pictures of a hITM cross section. Cell borders are indicated by dashed lines in magenta, while nuclei are presented in cyan (*n* = 2). Scale bar: 5 µm. **f1** Magnification of the apical cell region with closely arranged, upright microvilli (white arrowhead). Scale bar: 1 µm. **f2** Magnification of the cell-cell border in the apical region with indicated desmosome formation (white arrowheads). Scale bar: 1 µm. **g** Top view of representative IHC images of the hITM stained for *OCLN* (yellow) and *ZO-1* (magenta) with side view of the cellular monolayer at the indicated regions (dashed line) (*n* = 3). DAPI counterstaining is shown in blue. Scale bar: 20 µm. **h** Representative IHC stainings of the hITM demonstrating the expression of characteristic intestinal proteins including MUC-1 (green), VIL-1 (green), MUC-2 (green), pCK (magenta) and CDH1 (magenta). Cell nuclei are visualized by DAPI counterstaining shown in blue (*n* = 3). Scale bar: 20 µm. **i** UMAP projection of 3,952 single-cell transcriptomes from two models color coded according to assigned cell types. **j** Cell embedding (as in **i**) showing scaled, log-normalized gene expression of cell type markers (see supplement table 1). **k** Dot plot showing the cluster average of scaled, log-normalized gene expression of cell type markers. IECs: intestinal epithelial cells, hITM: human small intestinal epithelial tissue model, SIS: small intestinal submucosa, TEER: transepithelial electrical resistance, JAM: junction associated molecule, ZO: Zona Occludens, OCLN: Occludin, TRIC: Tricellulin, DSG3: Desmoglein-3, CDH1: Cadherin-1, SEM: scanning electron microscopy, TEM: transmission electron microscopy, TJ: tight junction, IHC: immunohistochemistry, MUC-1: Mucin-1, VIL-1: Villin-1, MUC-2: Mucin-2, pCK: pan-Cytokeratin, M-like cells: microfold-like cells.

and Mucin-2 (MUC-2; Figure 1 h3), a mucin secreted by goblet cells, albeit rarely observed as an intracellular signal in only a few hITM cells.

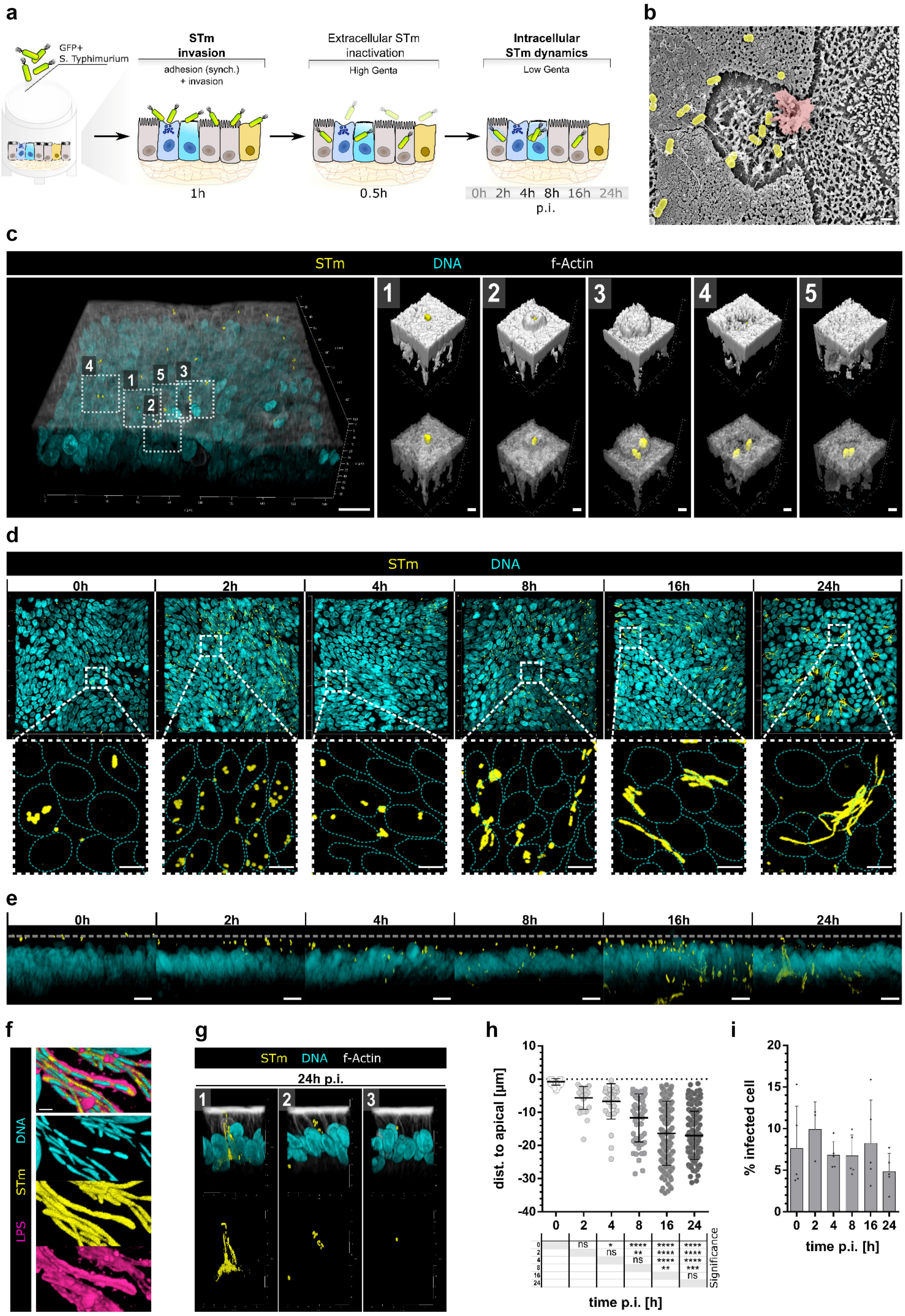
The hITM epithelium consists of a heterogeneous cell composition, resembling major cell types of the native tissue

In order to extend the cell-type classification within the *in vitro* model, we applied droplet-based single-cell RNA-seq. In total 3,360 cells were analyzed with 21,645 identified genes (4,814 median genes per cell, 23,123 median counts per cell). Unsupervised clustering of cells in the hITM (Figure 1 i) revealed eight populations with distinct transcriptional phenotypes. Based on cell type-specific gene expression signatures, five populations could be annotated to respective IEC subtypes (Figure 1 j-k). Transit amplifying (TA) cells were defined as *"proliferation cluster"* based on genes involved in cell cycle regulation (*MKI67: Marker Of Proliferation Ki-67, PCNA: Proliferating Cell Nuclear Antigen, NUSAP1: Nucleolar And Spindle Associated Protein 1*)[23–26] and represented 6.7% of the total cell number. The *"stem cluster"* comprises stem cells identified by the expression of typical intestinal stem cell markers (*LGR5: Leucine Rich Repeat Containing G-Protein Coupled Receptor 5, OLFM4: Olfactomedin-4, SOX4: SRY-Box Transcription Factor 4*)[23–26], accounting for 3.5% of all cells. The *"progenitor cluster"* constitutes 29.8% of all cells classified based on the relative absence or downregulation of marker genes specific for differentiated epithelial subtypes. Enterocytes (*enterocytes cluster*) accounted for of 28.9% of all cells. Immature enterocytes were defined by their expression of *ANPEP* (*Alanyl Aminopeptidase, Membrane*), *FABP2* (*Fatty Acid Binding Protein 2*) and *APOA4* (*Apolipoprotein 4*), while mature enterocytes were characterized based on their expression of *RBP2* (*Retinol Binding Protein 2*), *APOC3* (*Apolipoprotein C3*) and *CYP3A4* (*Cytochrome P450 Family3 Subfamily A Member 4*)[23–26]. Around 4% of all cellsbelonged to the *"M-like cluster"* expressing the corresponding cell type-specific genes (*CXCL3*: *C-X-C Motif Chemokine Ligand 3, CCL20 C-C-Motif Chemokine Ligand 20, TNFAIP2*: *TNF Alpha-Induced Protein 2*)[23,25,27,28] as well as immune-associated genes such as *CXCL8* (*C-X-C Motif Chemokine Ligand 8).* A sixth cluster was defined as secretory-like cells (*secretory cluster*), accounting for 16.4% of all cells expressing *LEMD1* (*LEM Domain Containing 1*), *TFF1* (*Trefoil Factor 1*)[24,29], *RAB3B* (*Member RAS Oncogene Family*)[30,31], *MMP1* (*Matrix Metalloproteinase 1*), *DHRS9* (*Dehydrogenase/Reductase 9*)[32] and *F3* (*Coagulation Factor III*). In addition, the combined expression of general epithelial cell markers along with *HLA-G* (*Major Histocompatibility Complex, Class I, G*), *APOL4* (*Apolipoprotein 4*), and *TRIP6* (*Thyroid Hormone Receptor Interactor 6*) defined a seventh population classified as *"HLA-G+ cell cluster"* (10.6%). The high expression of the mitochondrial gene *MTRNR2L12* (*MT-RNR2 Like 12*, pseudogene) in combination with a low number of genes/counts determined a low quality cell cluster representing dead/fragmented cells. All clusters were positive for the expression of the general epithelial markers *KRT8* (*Keratin 8*), *EPCAM* (*Epithelial Cell Adhesion Molecule*), and *CDH1*.

*S*Tm infection of the hITM recapitulates main stages of human infection

Having demonstrated that the hITM recapitulates key characteristics of the human intestinal epithelium, we next studied infection of this model by an exemplary enteropathogenic bacterium (Figure 2a). For that, we used a GFP-expressing *S*Tm strain that allowed us to track the bacteria within the tissue model during the infection process. The infection protocol (see material and methods for details) involved the administration of *S*Tm to the apical compartment of the hITM, an incubation for 1 h at 37°C during which the bacteria invaded the host cells, followed by a high-dose gentamicin treatment to inactivate the remaining extracellular bacteria. Thereafter, a low dose of gentamicin was constantly applied to prevent reinfection by cell-released bacteria (Figure 2a).

Via microscopic analysis, we observed *S*Tm attached to the apical brush border surface of the hITM at the 1 h time point (Figure 2 b). Membrane perturbations in the vicinity of extracellular *S*Tm (Figure 2 b) might indicate *S*Tm-induced membrane ruffling and ongoing *S*Tm invasion. In order to evaluate the membrane ruffling in more detail, we stained our infected hITM for the cytoskeletal protein F-actin and focused on structural


a
GFP+
S. Typhimurium
STm
invasion
adhesion (synch.)
+ invasion
1h
Extracellular STm
inactivation
High Genta
0.5h
Intracellular
STm dynamics
Low Genta
0h 2h 4h 8h 16h 24h
p.i.
b
c
STm
DNA
f-Actin
1
2
3
4
5
d
STm
DNA
0h
2h
4h
8h
16h
24h
e
0h
2h
4h
8h
16h
24h
f
DNA
STm
LPS
g
STm
DNA
f-Actin
24h p.i.
1
2
3
h
dist. to apical [µm]
10
0
-10
-20
-30
-40
0
2
4
8
16
24
time p.i. [h]
Significance
i
% infected cell
20
15
10
5
0
0
2
4
8
16
24
time p.i. [h]

	0	2	4	8	16	24
0		ns	*	****	****	****
2			ns	**	****	****
4				ns	****	****
8					**	***
16						ns
24						

Figure 2. Invasion characteristics of *Salmonella* Typhimurium in the hITM. **Aa** Schematic depiction of the *S*Tm infection process. hITMs were inoculated with *S*Tm expressing GFP by administration to the apical compartment. Incubation at 37°C allowed bacterial adhesion and invasion of epithelial cells. After 1 h, extracellular bacteria were killed by high-gentamicin treatment of the hITM (50 µg/ml) for 0.5 h, followed by a medium-exchange to low-gentamicin conditions (10 µg/ml) for the rest of the assay. Samples were taken and analyzed at 0 h, 2 h, 4 h, 6 h, 8 h, 16 h and 24 h p.i. **b** Representative SEM images of infected hITM 1 h p.i. highlighting invasion by

alterations in proximity of GFP-expressing *S*Tm (Figure 2 c). In doing so, we recapitulated the characteristic steps of *S*Tm invasion[33:] (1) adherence, (2) initiation of F-actin remodeling and (3) F-actin ruffling, (4) reestablishment of the F-actin layer and bacterial entry, and (5) restoration the F-actin cytoskeleton (Figure 2 c **1–5**).

We further evaluated spatial localization of the bacteria via confocal microscopy and 3D projection with subsequent spatial image analysis (Figure 2 d, **e**). Our data show that the bacteria were exclusively intracellular and that the hITM integrity was maintained after infection, as indicated by the comparable distribution of cellular nuclei (Figure 2 d, top row). Interestingly, morphologies and relative intracellular localizations of bacteria changed over time (Figure 2 d). Immediately after infection (0 h p.i.), mainly individual *S*Tm bacteria and only a few cluster-forming colonies were detected in infected IECs. With ongoing cultivation, increased frequencies of those clusters were observed, especially at 8 h post infection. Of note, a few structures that appeared more elongated with diffuse edges were visible at 8 h p.i. and became dominant from 16 h p.i. onwards. These structures likely reflect bacteria blocked in the cellular division process and hence adopting a filamentous morphology (Figure 2 f). Additionally, we observed infected cells, which carried singular or multiple non-filamentous bacteria at each individual time point, thereby indicating heterogeneity in intracellular *S*Tm populations.

Within their host cells, *S*Tm were detected apically of the nuclei, in close proximity to the apical membrane of the epithelium at 0 h p.i. (Figure 2 e, **g**). From 4 h p.i. onwards, *S*Tm were predominantly observed at the same horizontal positioning as the cellular nucleus and from 8 h p.i. on, also on the basolateral side of the epithelium. Simultaneously with the emergence of filamentous *S*Tm at 16 h p.i., *S*Tm-specific GFP signals were detected at higher frequencies at the basal site of infected cells, indicating the migration of certain bacteria within the IEC monolayer. Interestingly at 24 h p.i., the elongated *S*Tm structures stretched over the complete cell, from the apical to the basal side (Figure 2 e, **g1**), while *S*Tm without an elongated shape seemed to be randomly distributed along the apical-basolateral axis (Figure 2 g**2**, **g3**). We determined the localization of individual bacteria in relation to the apical cell membrane and revealed significant changes over time. The mean distance of *S*Tm from the apical to the basolateral side was −0.82 µm (±0.99 µm) at 0 h p.i., −5.64 µm (±3.44 µm) at 2 h p.i., −6.74 µm (±5.34 µm) at 4 h p.i., −11.70 µm (±7.23 µm) at 8 h p.i., −16.34 µm (±9.70 µm) at 16 h p.i. and −16.97 µm (±7.27 µm) at 24 h p.i. (Figure 2 h). Together, this indicates a time-dependent migration of *S*Tm toward the basolateral side of infected cells. In addition, flow

protuberances of the infected IEC membrane (red) and *S*Tm (yellow) on top of the IEC brush border in close proximity to the microvilli (*n* = 2). Scale bar: 1 µm. **c** Apical membrane area of the *S*Tm-infected hITM 1 h p.i. Representative 3D surface rendered images taken by confocal microscopy revealed sequential invasion steps characteristic for *S*Tm (*n* = 5). GFP expressing *S*Tm are depicted in yellow and F-actin stained by Phalloidin in gray. DAPI counterstaining is shown in cyan. Scale bar: 20 µm and scale bar in c1–5: 1 µm. **d** Representative microscope images (top view on 3D projection) of *S*Tm-infected hIITMs 0 h-24 h p.i. highlighting the existence of filamentous *S*Tm observed at 16 h and 24 h p.i. GFP expressing *S*Tm are shown in yellow and nuclear counterstaining by DAPI in cyan. Top panel represents merged images and the bottom panel visualizes *S*Tm-specific GFP signals in yellow as well as dotted cyan lines indicating nuclei. Scale bar: 5 µm (*n* = 3). **e** Side view on 3D projections from **d** showing intracellular migration of *S*Tm within the hITM over time. Dotted line indicates apical actin layer. Scale bar: 10 µm (*n* = 3). **f** Representative image of intracellular filamentous *S*Tm depicting bacterial DNA (cyan), bacterial cytoplasm with GFP (yellow), and the continuous bacterial membrane with LPS (magenta). Scale bar: 1 µm (*n* = 3). **g** Representative images from **e** of heterogeneous *S*Tm stages in individual host cells at 24 h p.i.: (1) formation of filamentous *S*Tm, (2) *S*Tm forming clusters, and (3) *S*Tm as single bacteria. Scale bar: 10 µm (*n* = 3). **h** Graph showing the quantification of the mean bacterial intracellular migration distance, given as negative distance from the apical membrane surface, of individual *S*Tm (analyzed number of bacteria per time point: 0 h: 43, 2 h: 34, 4 h: 32, 8 h: 52, 16 h: 204, 24 h: 185). Significance values are shown in the table below the graph. Significance was calculated by ordinary one-way ANOVAF and Tukey's multiple comparisons test with ****= $p \leq 0.0001$, ***= $p \leq 0.001$, **= $p \leq 0.01$, *= $p \leq 0.05$, ns= $p > 0.05$. Insignificant p-values are not displayed. **i** Bar diagram depicting the mean percentage with standard deviation of infected epithelial cells at 0 h, 2 h, 4 h, 8 h, 16 h, 24 h p.i. Infection rate was determined by measuring the relative proportion of GFP+ cells in the whole host cell population by flow cytometry (*n* = 3–5). *S*Tm: *Salmonella* Typhimurium, p.i.: post infection, SEM: scanning electron microscopy, IEC: intestinal epithelial cell, hITM: human intestinal tissue model.

cytometric detection of the GFP signal emitted by *S*Tm revealed a stable infection rate of 7.35% (±1.54%) within the hITM over the time course of 24 h (Figure 2 i).

*S*Tm infection in the hITM results in the elevated expression of OLFM4 in infected IECs

Next, we analyzed the response of our hITM model to *S*Tm infection. We profiled the expression of markers characteristic for intestinal cell types via qRT-PCR. Our data show that the mRNA levels of *VIL1*, a representative marker of mature enterocytes, and *LYZ*, a marker expressed by Paneth cells, were unchanged during the course of infection. Likewise, the expression of the canonical stem cell marker *LGR5* was not significantly altered upon *S*Tm infection. In contrast, we observed increased expression values for *MUC1*, a transmembrane mucin expressed in the gut epithelium, and *OLFM4*, another common ISC marker (Figure 3 a).

In light of the emerging role of OLFM4 in the context of infectious diseases (Liu and Rodgers, 2022), we sought to validate the *S*Tm-induced upregulation of *OLFM4* mRNA via an independent method. By HCR-FISH analysis, we compared *OLFM4* mRNA levels in *S*Tm-infected and uninfected hITMs. As shown in Figure 3 b, only few *OLFM4*+ cells were detected in the infected hITM as well as the mock control immediately after infection (0 h p.i.), presumably representing ISCs, which — according to our scRNA-seq data (Figure 1 i-k) — represent a minority in the model with a frequency of 3.5%. While the expression level of *OLFM4* remained unchanged over 24 h in the mock control, we observed an increased *OLFM4* expression in the infected hITM from 4 h p.i. onwards (Figures 3 b, c). Quantification by HCR-FlowFISH (Figure 3 c) revealed a significantly increased percentage of *OLFM4*+ cells in the *S*Tm-infected cell population from 4 h to 24 h p.i. (2.28% ± 1.45% at 0 h p.i.; 8.81% ± 5.18% at 4 h p.i.; 11.76% ± 4.60% at 8 h p.i.; 18.50% ± 5.08% at 16 h p.i.; 31.50% ± 8.01% at 24 h p.i.). In comparison, the percentages of *OLFM4*+ bystander cells were significantly lower (2.08% ± 1.32% at 0 h p.i.; 1.86% ± 0.71% at 4 h p.i.; 2.46% ± 0.78% at 8 h p.i.; 2.26% ± 1.57% at 16 h p.i.; 3.42% ± 0.53% at 24 h p.i.). Similarly, lower frequencies were observed for *OLFM4*+ cells in the uninfected mock control (1.32% ± 0.30% at 0 h p.i.; 0.34% ± 0.13% at 4 h p.i.; 0.42% ± 0.02% at 8 h p.i.; 2.81% ± 2.32% at 16 h p.i.; 3.44% ± 1.00% at 24 h p.i.). Interestingly, our HCR-FISH analysis demonstrated that the *OLFM4* levels correlate with bacterial numbers (GFP signal intensity per infected host cell), as shown by the representative microscope image depicted in Figure 3 d. In this context, we further measured an elevated median fluorescence intensity (MFI) of GFP, expressed by *S*Tm, at 16 h p.i. and 24 h p.i. in the *OLFM4*+ cells (Figure 3 e), indicating a link between the bacterial load of individual *S*Tm-infected cells and the upregulation of *OLFM4*. Along with the MFI of GFP, the MFI for *OLFM4* showed an increasing trend in *OLFM4*+ cells over time (Figure 3 f). Together, these data suggest a link between the bacterial load of individual *S*Tm-infected cells and the upregulation of *OLFM4* in the hITM.

*S*Tm-mediated OLFM4 induction depends on NOTCH signaling

In the healthy gut, intestinal *OLFM4* expression is predominantly restricted to the stem cell population residing in the crypt compartment, and regulated by the NOTCH pathway[35]. To confirm a role for the NOTCH pathway in the elevated expression of *OLFM4* in *S*Tm-infected hITM, we measured *OLFM4* expression in the infected hITM treated with DAPT, a γ-secretase inhibitor of the NOTCH pathway (Figure 4 a). For quantification, we normalized the HCR-FISH intensity against the number of nuclei (Figure 4 b).

Immediately after infection, we detected similar *OLFM4* signals in the infected hITM (MOI 10), the infected hITM treated with DAPT (MOI 10 + DAPT) and the uninfected control (mock) (Figure 4 a). In line with the data shown in Figure 3, we observed a significant increase of *OLFM4* levels in the infected model as early as 4 h p.i., with further upregulation at 16 h and 24 h after infection (Figure 4 b). From 16 h p.i. onwards, the *OLFM4* expression not only significantly increased in infected samples but also dropped when the infected hITM was treated with DAPT. Similar effects were seen at 24 h p.i. with a significantly

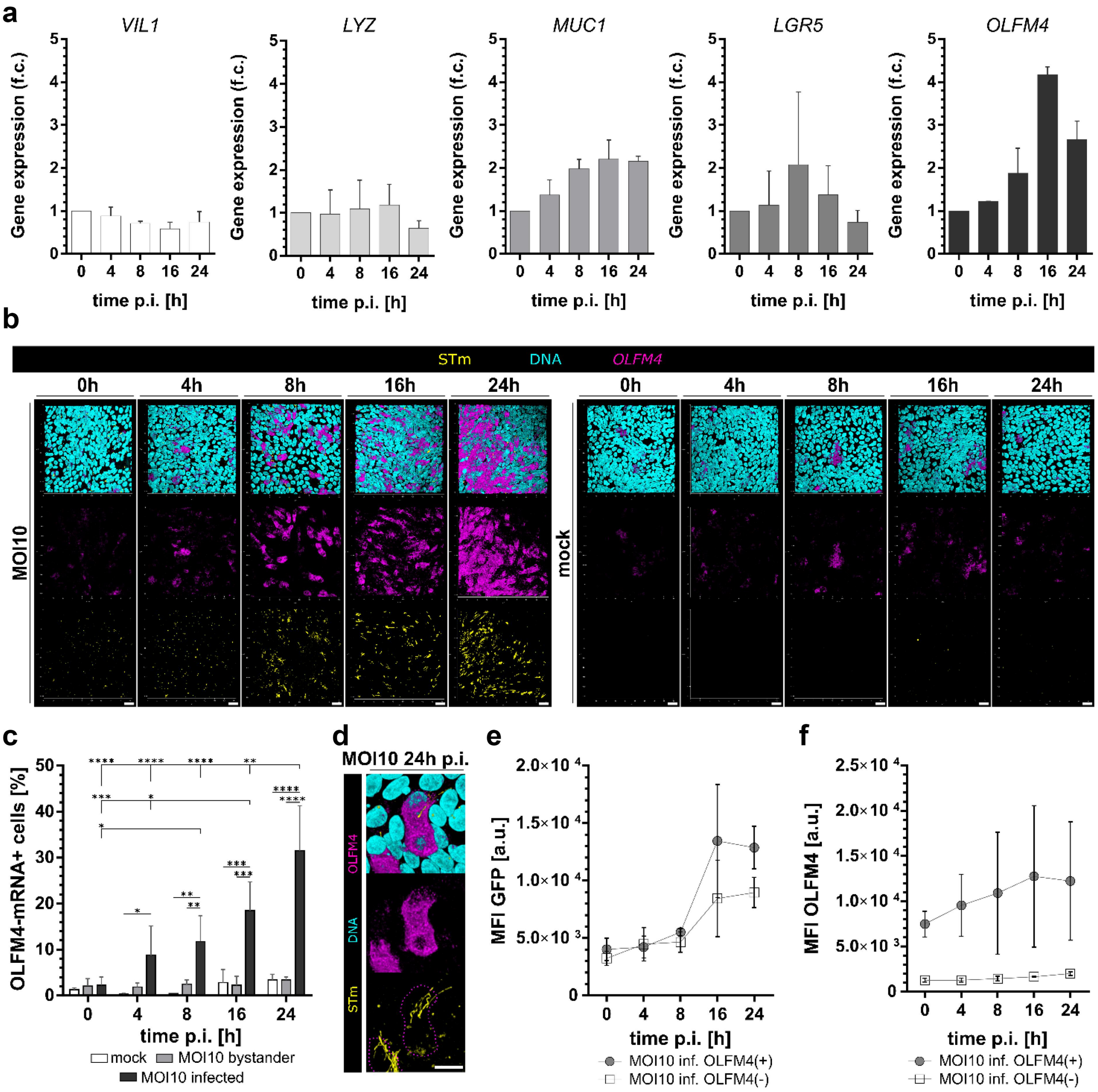


Figure 3. *OLFM4* is upregulated in *S*Tm-infected cells. **Aa** qRT-PCR-derived gene expression of *VIL1, LYZ, MUC1, LGR5*, and *OLFM4* in *S*Tm-infected tissue models from 0 to 24 h p.i. Expression is depicted as a fold change relative to 0 h p.i. ($n = 2$) and was calculated using the ΔΔCT method[34]. Human *Ef1a* was used as a reference mRNA. **b** Top view on 3D projections of representative fluorescent images of *S*Tm-infected hITM (MOI10, left) and uninfected hITM controls (mock, right). *OLFM4* expression (magenta) becomes visible at 4 h p.i. with ongoing increase of expression over time. GFP expressing *S*Tm are shown in yellow and DNA counterstaining in cyan ($n = 2$). **c** Frequency of *OLFM4*+ IECs in infected hITM (MOI10 infected) at the indicated time points compared to uninfected mock controls as well as in the population of non-infected bystander cells (MOI10 bystander), detected via HCR-Flowfish ($n = 3$; minimum 10,000 cells). Significance was calculated via matching two-way ANOVA with statistically significant differences and Tukey's multiple comparisons test. **d** Representative microscope pictures of *S*Tm-infected hITM (MOI10) at 24 h p.i., demonstrating the restriction of increased *OLFM4* expression to infected cells ($n = 3$). **e** MFI of GFP in infected *OLFM4*+ and *OLFM4*- IECs ($n = 3$). **f** MFI of *OLFM4* in infected *OLFM4*+ and *OLFM4*- IECs at 0 h-24 h p.i. measured via HCR-Flowfish ($n = 3$). Significance is indicated with ****= $p \leq 0.0001$, ***= $p \leq 0.01$ **= $p \leq 0.01$, *= $p \leq 0.05$. Insignificant p-values are not displayed. OLFM4: Olfactomedin 4, *S*Tm: *Salmonella* Typhimurium, p.i.: post infection, hITM: human intestinal tissue model, IEC: intestinal epithelial cell, MFI: median fluorescence intensity, HCR-FlowFISH: Hybridization Chain Reaction Fluorescent in-Situ Hybridization coupled with flow cytometry.

increased *OLFM4* expression in the infected hITM compared to DAPT-treated or uninfected models. Of note, the *OLFM4* signal intensities in infected and DAPT-treated models matched those of uninfected controls. Importantly, the observed patterns of *OLFM4* mRNA levels were reflected at the protein level, as revealed by immunofluorescence (Figure 4 d). The intracellular growth of *S*Tm – as inferred from the mean fluorescence intensity per infected IEC – was unaffected by the DAPT treatment (Figure 4 c). Interestingly, under infection conditions, the filamentous *S*Tm form was observed in cells with high OLFM4 protein levels (Figures 3 b and 4 e). DAPT treatment of the infected hITM prevented *S*Tm filamentation (Figure 4 d). Quantification of this observation revealed significantly decreased filament lengths in DAPT-treated hITM (MOI 10 + DAPT; median: 3.00, 25%/75% percentile: 2.40/4.05) compared to untreated hITM (MOI 10; median: 3.50, 25%/75% percentile: 2.54/5.49) at 24 h p.i. (Figure 4 **e**). Together, our data support that DAPT-mediated NOTCH inhibition interferes with *OLFM4* expression in the hITM epithelium and seems to counteract *S*Tm filamentation.

Discussion

Given the key role of the intestinal epithelium as the main barrier of the body to the environment[36], it is important to replicate physico-chemical, structural, and biological properties of this tissue *in vitro*. Here, we refined a previously published method from our group to engineer a Transwell®-like model of the human small intestinal epithelium grown under static cell culture conditions by combining a porcine-derived, organ-specific ECM and primary tissue-derived enteroids (referred to as hITM)[16,37]. Epithelial key features of the hITM were consistent with previous reports[38] and comparable to the native tissue[17,18]. The intestinal epithelial phenotype was validated by the expression of key markers including the transmembrane-expressed MUC1 and VIL1, expressed on the apical surface of intestinal epithelial cells[21,22]. Structural similarity of the hITM to the native intestinal epithelium is further highlighted by a distinct polygonal, cobblestone arrangement of individual cells that are covered by microvilli[36,39]. Tissue-specific epithelial junction complexes (JC) including TJ, adherens junctions (AJ), and desmosomes were formed between individual cells of the epithelial monolayer. Not only did we detect a tissue-specific JC expression profile, but also the distinct spatial arrangement of JC proteins in the model. For example, similar to the native tissue, the TJ-associated proteins ZO-1 and OCLN were observed in the apical region of the cells, whereas desmosomes were localized in the lower cell regions as well as laterally[36]. Interestingly, compared to Matrigel®-based enteroid cultures, hITM cells showed increased gene expression values for specific TJ proteins such as TRIC, a structurally specialized TJ protein as well as for CDH1, a major constituent of AJ, and for DSG3, a desmosomal cadherin. The increased expression of *TRIC*, *DSG3*, and *CDH1* could be related to the planar arrangement and the pronounced polarization of the cells grown on the biological ECM scaffold.

In addition to the structural features, we further demonstrated that the hITM is composed of cellular phenotypes similar to those observed in the native epithelium. RNA-seq analysis at the single-cell level showed that the hITM is composed of stem cells, highly proliferative TA progenitor cells, and differentiated cell types, such as enterocytes and cell types with a transcriptional profile similar to secretory cells. The cell types were classified according to specific gene expression patterns. An M-like cell cluster was defined by *CCL20*, *TNFAIP2*, *CXCL8*, and *CXCL3* expression[23–26,40], supporting the findings observed by EM analysis. According to Nakamura et al., 2018, mature M-cells express *SPIB* (Spi-B transcription factor) and *GP2* (Glycoprotein 2)[41]. However, none of these genes were detected in our scRNA-seq analysis, implying an immature M-cell phenotype in our hITM. In the native tissue, M-cell development is tightly regulated via RANKL signaling, a factor secreted by stromal cells[25,42,43]. The lack of stromal cells and the consequent absence or diminished RANKL signaling possibly explains this apparent immaturity of the M-like cells in the hITM. Further, in contrast to enterocytes or M-like cells, the secretory-like cell cluster showed a non-canonical gene expression profile lacking the expression of markers characterizing typical subtypes of secretory cells in the small intestine[23,26].

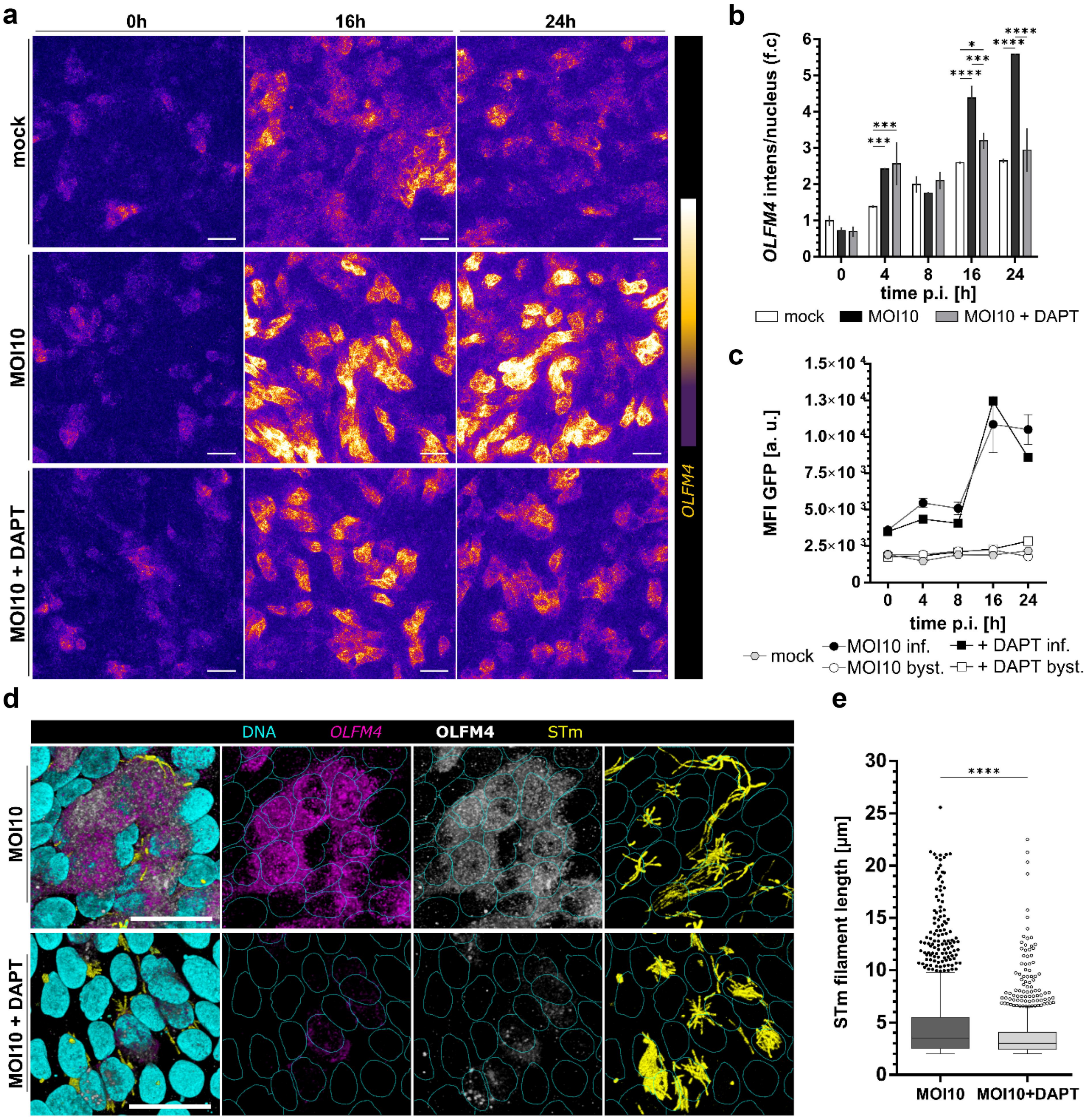

a
0h
16h
24h
mock
MOI10
MOI10 + DAPT
OLFM4
b
OLFM4 intens/nucleus (f.c)
6
5
4
3
2
1
0
***
***
*
***
****
****
****
0
4
8
16
24
time p.i. [h]
mock
MOI10
MOI10 + DAPT
c
1.5×10 4
1.3×10 4
1.0×10 4
7.5×10 3
5.0×10 3
2.5×10 3
0
MFI GFP [a. u.]
0
4
8
16
24
time p.i. [h]
mock
MOI10 inf.
MOI10 byst.
+ DAPT inf.
+ DAPT byst.
d
DNA
OLFM4
OLFM4
STm
MOI10
MOI10 + DAPT
e
30
25
20
15
10
5
0
****
STm filament length [µm]
MOI10
MOI10+DAPT

Figure 4. OLFM4 induction in *S*Tm-infected hITM depends on NOTCH signaling. **Aa** Representative fluorescent images of uninfected (mock) and *S*Tm-infected models, either treated (MOI 10 + DAPT) or not (MOI 10) with the NOTCH-specific γ-secretase-inhibitor DAPT at 0 h, 8 h, and 24 h post infection. *OLFM4* mRNA levels were detected with HCR-FISH and visualized as intensity range in maximum projection images (n = 2). Scale bar: 20 µm. **b** Quantification of *OLFM4* mRNA intensity per nucleus. *OLFM4* mRNA-specific signal intensity was measured in fluorescent images. The mean intensity value was normalized to the number of nuclei and is shown as fold change (f.c.) (n = 2). Significance was calculated comparing the conditions within each time point via ordinary two-way ANOVA with statistically significant differences and Tukey's multiple comparisons test with ****= $p \leq 0.0001$, ***= $p \leq 0.001$, **= $p \leq 0.01$, *= $p \leq 0.05$. Insignificant p-values are not displayed. **c**MFI of GFP per infected and uninfected IEC of mock- or *S*Tm-infected models, treated (+ DAPT) or not (MOI 10) with the NOTCH-specific γ-secretase-inhibitor DAPT at 0 h, 4 h, 8 h, 16 h, and 24 h post infection (n = 3 for mock and MOI 10, n = 1 for MOI 10 + DAPT; all technical duplicates). **d** 3D projection of representative images of HCR-FISH analyses at 24 h p. i. showing the expression of OLFM4 mRNA and protein in *S*Tm-infected cells in the untreated hITM (MOI 10) compared to DAPT-treated models (MOI 10 + DAPT). GFP expressing *S*Tm are shown in yellow, *OLFM4* mRNA signals in magenta, OLFM4 protein expression in white and nuclear counterstaining with DAPI in cyan(n = 2). Scale bar: 20 µm. **e** Bar diagram depicting *S*Tm filament lengths in the infected untreated hITM (MOI 10) compared to the DAPT-treated hITM infected with *S*Tm (MOI 10 + DAPT) at 24 h p.i. Filament length was determined in fluorescent images by automated filament recognition via Imaris; filaments with < 2 µm were excluded from the analysis (n = 2). Data are presented as box-plot with Tukey Whiskers and significance was calculated by Mann–Whitney test with ****= $p \leq 0.0001$; two tailed. OLFM4: Olfactomedin 4, *S*Tm: *Salmonella* Typhimurium, MOI: multiplicity of infection, p.i.: post infection, hITM: human intestinal tissue model, IEC: intestinal epithelial cell, MFI: median fluorescence intensity.

Among them are *CHGA* (Chromogranin A) and *NEUROG3* (Neurogenin 3) expressed by enteroendocrine cells, *DEFA5* (Defensin Alpha 5) and *REG3A* (Regenerating Family Member 3 Alpha) expressed by Paneth cells, *TFF3* (Trefoil Factor 3) and *SPINK4* (Serine Peptidase Inhibitor Kazal Type 4) expressed by mature goblet cells. Instead, we could define the secretory-like cell cluster based on non-canonical genes including *F3* or *RAB3B* expressed by enteroendocrine cells[30,31], *DHSR9* or *TFF1* associated with goblet cell identity[24,29,32], according to the online intestinal cell atlas[23]. In general, the gene expression profile of the secretory-like cell cluster does not show specificity for a secretory subtype, but rather a mixture of gene expression patterns of all secretory cells. This suggests that we possibly generated a precursor of the secretory lineage by our applied differentiation protocol. However, whether differentiated secretory subtypes derive from a specific precursor has not yet been fully proven and therefore requires further investigation, as shown by the non-canonical WNT regulation of Paneth and EEC development without goblet cell differentiation[44]. In this context, our hITM could represent a suitable *in vitro* platform enabling secretory lineage differentiation studies involving a bioartificial cell niche represented by the native ECM component. In addition to the secretory-like cell cluster, we defined a second, HLA-G cell cluster based on a non-canonical gene expression profile, predominantly characterized by elevated expression levels of the *HLA-G* gene, which encodes an immunomodulatory molecule, by intestinal epithelial cells[45,46]. The data confirm that the epithelium of the hITM contains key cell types of the native tissue.

To evaluate the hITM as a potential host model for infection experiments with enteric pathogens, we centered on the well-established model pathogen *S*Tm. Infection of our hITM recapitulated (1) *S*Tm adherence to the brush border, (2) actin remodeling, resulting in cell membrane protrusion and formation of the characteristic “donut” shape surrounding the bacteria, (3) *S*Tm encapsulation by the membrane, leading to membrane perturbations, (4) endocytosis of *S*Tm into the cytoplasm, and (5) restoration of the F-actin cytoskeleton[33,47]. In contrast to the conventional ruffle-induced invasion process typically observed with cell line-based models[48,49], our data support a recently proposed alternative *S*Tm invasion mechanism referred to as discrete invasion (DI)[48]. Specifically, using a mouse model, these authors showed that *S*Tm enter absorptive epithelial cells via a Rho GTPase-independent process, which does not entail actin ruffling. The hITM thus lends itself for mechanistic studies of the DI-induced invasion process of *S*Tm.

Following invasion, enteric pathogens often transmigrate to the basolateral side of the epithelium by hijacking host cellular transport machinery[50–52]. Indeed, we observed time-dependent *S*Tm transmigration from the apical membrane toward the basolateral region of the epithelium. Future experiments may address whether these processes are mediated by *Salmonella* pathogenicity island 2 (SPI-2) T3SS effector proteins, as recently suggested[53]. Transmigration was accompanied by morphological changes of the intracellular bacteria. In the early phases of infection (0-8 h p.i.), *S*Tm exhibited its typical rod-shaped morphology; however, adopted a more elongated shape at the later stages. This culminated at 24 h p.i., when *S*Tm formed multinucleated, elongated filaments. Bacterial filamentation can result from perturbed cell division and has been observed in a variety of species[54,55]. Several *Salmonella* enterica serovars are known to form this filamentous morphology extracellularly, in response to environmental factors, such as osmotic stress or temperature variations[56,57]. However, the significance of intracellular filamentous *S*Tm is currently poorly understood[58–61]. Of note, we are confident that filamentous *S*Tm inside IECs are not just technical artifacts of our infection protocol. For example, all our models were exposed to the same gentamicin concentration, whereas *S*Tm filamentation was exclusive to NOTCH-proficient cells yet absent when NOTCH signaling was inhibited.

We also measured the host responses to *S*Tm infection of the hITM. Among others, we observed an increase in *OLFM4* and *MUC1* expression following infection. The increased *MUC1* expression is consistent with a recently reported host defense response against enteric pathogens[62]. In contrast, *OLFM4* is predominantly expressed by intestinal stem cells[23,63], but was upregulated in the gastric mucosa of patients with *Helicobacter pylori*[64], in children infected with *Staphylococcus aureus*[65,66], in *Lawsonia intracellularis* infected intestinal porcine crypts[67], and linked to the infection of the oral mucosa with *Porphyromonas gingivalis*[68]. Our further characterization revealed the elevated expression of *OLFM4* to be triggered by NOTCH signaling in consistency with former reports[69]. Importantly, *OLFM4* expression did not correlate with the levels of *LGR5* mRNA, suggesting the upregulation of *OLFM4* as a potential novel host cell response to *S*Tm infection, independent of the stem cell context. Previous *S*Tm infection studies that were coupled to transcriptomics, were often based on immortalized cell-line models,[37,70] providing a possible explanation for why this aspect of the epithelial response to *S*Tm may have previously gone unnoticed. Consistent with recent studies, which listed increased *OLFM4* expression in *S*Tm-infected iPSC derived organoids[71,72], our study demonstrated that *OLFM4* is upregulated in individual infected epithelial cells during *S*Tm infection. Of note, *OLFM4* expression was found upregulated in IBD patient-derived intestinal epithelial cells[73,74], suggesting this to be a genuine inflammation-induced response of human primary intestinal tissues. Together, our present data imply a new role of OLFM4 in *Salmonella* infection. Yet further analysis is needed to evaluate the effect of OLFM4 on infection outcome and thus, to distinguish between a potential *S*Tm virulence versus host defense mechanism.

Outlook

Taken altogether, the here presented hITM represents a valuable tool for preclinical infection research, but still leaves room for future improvements. Although composed of several cellular entities of the native intestinal epithelium, our model does not yet include all mature intestinal cell types present *in vivo*. For example, we could not identify mature canonical goblet cells[75], mature Paneth cells[76], enteroendocrine cells (EEC; including I-, K-, L-, *M*-, *N*-, D-, enterochromaffin cells),[77] or Tuft cells[78]. Future studies will therefore focus on increasing the cellular complexity of the hITM. For example, refinements of the differentiation protocol could increase cellular diversity, e.g. by the administration of a MAP kinase inhibitor and/or bone morphogenesis protein-4 (BMP4), possibly triggering the differentiation of the enteroid-containing ISCs in the hITM toward a mature EEC phenotype[79,80]. The inclusion of additional cell types would also be relevant to assess the role of stromal, endothelial, and/or immune cells in *S*Tm infections. In addition, substituting the porcine matrix with a human matrix would bring the hITM even closer to the human situation.

Material & methods

Animal handling and biological matrix preparation

Animal research was performed according to the German law and institutional guidelines approved by the Ethics Committee of the District of Unterfranken, Würzburg, Germany (approval number 55.2-2532-2-256). The biological matrix (SIS; small intestinal submucosa) was prepared from jejunal segments explanted from young pigs (age: 6–8 weeks; weight: ~20 kg; provided by Niedermayer, Dettelbach, Germany). After explantation of jejunal segments, chemical decellularization was performed according to previously published protocols[81,82].

Human tissue

Human jejunal biopsies used for crypt isolation were obtained from obese patients undergoing gastric bypass surgery at the University Hospital Würzburg, surgery unit of PD Dr. med. C. Jurowich. Informed written consent was obtained beforehand. The use of human tissue was approved by the Institutional Ethics Committee on Human Research of the Julius-Maximilians University Würzburg (approval number: 280/18-sc).

Organoid culture

Small intestinal organoids were established from human tissue biopsies as previously described[7,16,83]. Briefly, after isolating the intestinal crypts from tissue biopsies, cells were diluted in ice cold Matrigel®, mixed in a 1:1 ratio with an expansion medium, and were seeded as 10–50 µl drops into 24 well tissue culture plates. After solidification of the Matrigel® drops at 37°C, cells were covered with 300 µl Expansion medium consisting of 75% conditioned LWRN medium (contains Wnt-3A, R-Spondin-1, and Noggin, produced by cell-line L-WRN (ATCC)[84]+25% Complex medium (Advanced DMEM F12 (Gibco) with 10 mM HEPES (Sigma Aldrich), 1× GlutaMax-I (Gibco), 1× Anti-Anti (Gibco), 1 mM N-Acetylcysteine (Sigma Aldrich), 1× N2-Supplement (Gibco), 1× B27-Supplement without vitamin A (Gibco)) supplemented with 0.05 µg/ml mEGF (Peprotech), 0.01 µM Leu-Gastrin (Sigma Aldrich), 10 µM Nicotinamide (Sigma Aldrich), 0.5 µM A83–01 (Tocris), 10 µM SB202190 (Sigma Aldrich), 0.5 µM LY2157299 (CAYMAN Chemical Company). Cells were cultured at 37°C, 95% humidity, 5% CO_2 and medium was changed every 2–3 days.

Passaging of small intestinal organoid cultures was performed every 5–6 days by enzymatic and mechanical dissociation. To this aim, the organoid-containing Matrigel® drops were collected with the Expansion medium, incubated on ice for 30 min, centrifuged at 350×g for 3 min, followed by dissociation of the pellet using 1–2 ml 1× TrypLE™ express incubated for up to 10 min at 37°C. Afterward, the pellet was gently pipetted to obtain single cells that were washed in a complex medium and subsequently reseeded in Matrigel® drop cultures with a cell density of 1000 IECs/µl. After solidification of the Matrigel® drops at 37°C, cells were covered with 300 µl Expansion medium supplemented with 10 µM Y-27632 (CAYMAN Chemical Company) and 10 µM JAG-1 (AnaSpec Inc.). After 2 days, the medium was changed to Expansion medium without Y-27632 and JAG-1.

hITM generation

The hITM model was established as recently described[16] with the following modifications: 1) small intestinal organoids were washed with 1× PBS without calcium or magnesium (PBS-) and dissociated into single cells by TrypLE™ express treatment, 2) 4 × 10^5 cells diluted in 300 µl Expansion medium supplemented with 10 µM Y-27632 (CAYMAN Chemical Company), and 10 µM JAG-1 (AnaSpec Inc.) were seeded per 0.54 cm^2 into the apical compartment of a iScript™-like cell crown system with the biological SIS matrix as scaffold, 3) the basolateral compartment was filled with 900 µl Expansion medium containing 10 µM Y-27632 (CAYMAN Chemical Company) and 10 µM JAG-1 (AnaSpec Inc.), 4) cells adhered after 2 days to the scaffold and tissue models were kept in Expansion Medium without Y27632 and JAG-2 for additional 3–4 days (= Proliferation Phase), 5) the Expansion medium was changed to Differentiation medium (25% conditioned Wnt-3A medium (produced by cell-line L-Wnt-3A (ATTC)[85]+75% Complex medium supplemented with 0.5 µg/ml R-Spondin-1 (Peprotech), 0.1 µg/ml rec mNoggin (Peprotech), 0.05 µ/ml mEGF (Peprotech), 0.01 µM Leu-Gastrin (Sigma Aldrich), 0.5 µM A83–01 (Tocris), 0.5 µM LY2157299 (CAYMAN Chemical Company)) to initiate the differentiation of intestinal stem cells toward distinct cellular entities of the intestinal epithelium, and 6) the differentiation was performed for 4 days (= Differentiation Phase).[86]

Salmonella enterica serovar Typhimurium culture and hITM infection

The constitutively GFP expressing the Wild-type derivative of *Salmonella* Typhimurium strain SL1344 (JVS-3858, (Papenfort et al. 2009)) was used in infection experiments. Bacteria were cultured in 5 ml Lennox broth (LB) at 37°C under constant agitation at 220 rpm (New Brunswick, Innova 44), overnight. For hITM infection, a 1:100 dilution of the overnight culture was grown to an OD of 2.0, pelleted by centrifugation (10,000×g for 5 min RT) and resuspended in a Complex medium without antibiotics to obtain the bacterial infection medium. For an MOI of 10, 4 × 10^ bacteria diluted in 300 µl Complex medium were applied to the apical compartment of the hITM. The basolateral compartment was filled with 900 µl Complex medium . Subsequently, for synchronized bacterial adhesion, the tissue models were centrifuged at 250×g for 10 min at RT and incubated at 37°C, 95% humidity, 5% CO_2 for 1 h. After adhesion, the medium (apical and basolateral) was exchanged to a Complex medium containing High Gentamicin (Gibco) (50 µg/ml), and models were incubated for 30 min at 37°C, 95%

humidity, 5% CO_2, in order to inactivate extracellular bacteria. Afterward, the High Gentamicin containing Complex medium was exchanged to a Low Gentamicin-Complex medium (10 µg/ml) applied for ongoing culturing of the models at 37°C, 95% humidity, 5% CO_2 to prevent reinfection of the hITM.

Model dissociation for analyses of infection characteristics

For analyses of the hITM 0 – 24 h post infection, supernatants of the apical and basolateral compartments were discarded, followed by washing the models with 1× PBS+1 mM EDTA (ThermoFisher) and incubation with Accutase (Gibco), added to both compartments, incubating for 10 min at 37°C. To detach the cells from the scaffold, the Accutase suspension in the apical compartment was vigorously resuspended and transferred to a 2 ml reaction tube. The Accutase solution of the basolateral compartment was discarded and the SIS scaffold was transferred to the 2 ml reaction tube containing the Accutase solution harvested from the apical site of the model with subsequent incubation on a shaker for 10 min, 37°C at 1000 rpm. After careful resuspension using a 1 ml pipet and 10 min incubation at RT to allow settling of the cells by gravitation, the scaffold was removed. Next, the cells were centrifuged at 500×g for 3 min, RT and the supernatant was discarded. Afterward, the pellet was washed twice with 1× PBS-. Finally, cells were resuspended in 1× PBS- and stored on ice for further analysis. Before flow cytometry and Drop-seq were performed, the cells were filtered through a 40 µm Strainer (Miltenyi).

Infection rates were determined by flow cytometry performed via BD FACS Aria III (BD Biosciences) gating for STm-expressed GFP. To this aim, single-cell suspensions were filtered through a 30 µm cell strainer (Miltenyi Biontech) and a 85 µm nozzle was applied for flow cytometric procedure. Flow cytometric laser and filter settings were adjusted on unstained and uninfected cells and at least 10,000 cells were analyzed. Cells were gated based on SSC-A and FSC-A, followed by doublet discrimination via SSC-A/SSC-H and FSC-A/FSC-H. Uninfected and bystander cells were determined as GFP(-) PerCP-Cy5–5-A(-), including a small population of GFP(+)/PerCP-Cy5–5-A(+) cells. Infected cells were determined as GFP(+)/PerCP-Cy5–5-A(-) cells.

Cell hashing and dropseq

Single cells were dissociated from the hITM as described. The cells of two different models were hashtagged with TotalSeq-A antibodies (Biolegend) following the manufacturer’s protocol for TotalSeqTM-A antibodies and cell hashing with 10× Single Cell 3’ Reagent kit v3.1 (10× Genomics). Approximately 400.000 cells per sample were resuspended in 100 µl Cell Staining Buffer (Biolegend) and 5 µl Human TruStain FcXTM FcBlocking (Biolegend) reagent were added. For the blocking reaction, the cells were kept at 4°C for 10 min. 1 µg of TotalSeqTM-Antibody was added to each sample, followed by 30-minutes incubation at 4°C. Afterward, cells were washed three times with 1 ml Cell Staining Buffer and spun down for 5 min at 350×g and 4°C. Finally, the cells were resuspended in an appropriate volume of PBS- and passed through a 40 µm cell strainer (FlowmiTM Cell Strainer, Merck). Cells were counted in a Neubauer Hemacytometer (Marienfeld) and concentration was adjusted to 1000 cells/µl with PBS-. The hashtagged cells were pooled equally and ~ 20.000 cells were loaded in the ChromiumTM Controller. The machine creates Gel Bead-In-Emulsions (GEMs) to separate single cells into a nanoliter compartment together with an individual barcode. Reverse transcription, cDNA amplification, and the construction of gene expression libraries were performed using the 10× Single Cell 3’ reaction kit v3.1 (10× Genomics). Incubation and amplification steps were carried out using a SimpliAmp Thermal Cycler (ThermoFisher). Library quantification and quality control was observed using a QubitTM 4.0 Fluorometer (ThermoFisher) and a 2100 Bioanalyzer with High Sensitivity DNA kit (Agilent). Sequencing was performed on a NextSeq 500 sequencer (Illumina).

Bioinformatic analysis

FASTQ files were aligned and counted using the CellRanger count pipeline against the GRCh38 human genome reference as well as the TotalSeqA Hashtags 1 and 2. The count matrix was imported into R for analysis using the Seurat framework. All R scripts used in the analysis process are available at https://github.com/saliba-lab/tissue-model-human-intestine.git. Briefly, the count matrix was split between counts for hashtags and gene expression. Hashtags were assigned by using count thresholds to distinguish replicates (Hashtag 1, 30; Hashtag 2, 50). Unstained (negative) and double positive (doublet) cells were removed. Gene expression counts were normalized (log1pCP10k), 5000 highly variable genes were selected, and 45 principle components (PCs) were computed for UMAP projection, SNN graph (k = 10, type = rank), and leiden clustering (resolution_parameter = 1.2, n_iterations = 5). Differential gene expression was assessed using scran:findMarkers(block=Replicate, pval.type=-some). Visualizations were created with custom code using the ggplot2, dplyr, tidyr, and pheatmap packages. The versions of all installed packages are specified in conda YML files in the GitHub repository (envs/default.yml).

Histological analyses

For histological analyses, tissue models were washed with 1× PBS–before fixation in 4% PFA, 2 h at RT. Afterward, tissue models were washed with 1× PBS-, covered with 70% EtOH and stored in the dark at 4°C until further processing.

Paraffin embedding was performed using a Microm STP 120 (ThermoFisher). Briefly, samples were first dehydrated by incubating in an ascending row of H_2O, 50% EtOH, 75% EtOH, 90% EtOH, 2-Propanol and Xylol with subsequent incubation in liquid paraffin.

Alcian blue staining was performed on 5 µm sections of paraffin embedded samples, cut with a Sliding Microtome RM 2255 (Leica). Tissue slices were deparaffinized at 60°C for 1 h followed by rehydration of the sections by incubating them in a descending row of Xylol, 90% EtOH, 75% EtOH, 50% EtOH, and H2O. Afterward, the sections were incubated in 3% acetic acid, 1% Alcian blue (Morphisto), and nuclear fast red solution (Morphisto) according to HADDOCK, 1948 before the samples were dehydrated and embedded in Entellan (Merck).

For immunohistological (IHC) analysis, paraffin-sections were dewaxed and rehydrated. After rehydration, heat-mediated antigen retrieval was performed for 20 min at 95°C in 1× citrate-buffer (Sigma). Following this, sections were permeabilized with 0.2% Triton X-100 (BioRad) in PBS-, blocked with 5% donkey serum in 1× PBS–containing 0.5% Tween 20 (PBS-T) and finally incubated in PBS-T with primary antibodies (MUC-1 (Abcam, ab109185), MUC-2 (Abcam, ab76774), VIL-1 (Santa Cruz, sc7672), pan-Cytokeratin (DAKO, Z0622)) overnight at 4°C. After washing the samples twice with PBS-T on the next day, secondary antibodies (Donkey anti-rabbit AF647, donkey anti-goat AF555 (Invitrogen)) were applied for 2 h at RT diluted in PBS-T. Stained samples were washed with PBS-T and embedded in Fluoromount G containing DAPI (Invitrogen) for nuclei staining.

For whole mount staining, fixed models were disassembled, transferred to a well plate and then permeabilized using 0.2% Triton X-100 (BioRad). After washing with PBS-T, unspecific-binding sites were blocked by incubation with 5% donkey serum (Biozol) diluted in PBS-T for 30 min at RT before incubation with primary antibodies ((ZO-1 (Ptglab 21,773), OCLN (ThermoFisher 33–1500), OLFM4 (Cell Signaling D1E4M)) diluted in PBS-T overnight at 4°C. After washing twice with PBS-T the next day, secondary antibodies (Donkey anti-rabbit AF647, donkey anti-mouse AF555 (Invitrogen)) diluted in PBS-T were applied for 2 h, RT. Models were washed once in PBS-T, incubated with Phalloidin (Abcam ab176756, ab176759) and/or DAPI (ThermoFisher) diluted in PBS-T for 20 min at RT and washed twice with PBS-T. Finally, models were embedded in Fluoromount G (Invitrogen).

Electron microscopy

For scanning electron microscopy (SEM) and transmission-electron microscopy (TEM), tissue models were first washed with 1× PBS-before fixation overnight at 4°C in 0.1 M phosphate buffer pH 7.4 (9.46 g/L Na_2HPO4, 9,078 g/L KH_2PO_4 in ddH2O) containing 6.25% glutaraldehyde (VWR). Fixed samples were then washed five times in 0.1 M phosphate buffer before sample preparation for SEM/TEM and imaging that was carried out in collaboration with Prof. Stigloher from the Biocenter Imaging Core Facility, University of Würzburg. For TEM, the samples were washed in 50 mM Cacodylate Buffer (pH 7.2, Roth) and further fixed with 2% buffered OsO_4 (ScienceServices). They were then contrasted with 0.5% watery Uranyl Acetate (Merck) and finally embedded in Epon (Serva) after EtOH-based dehydration. Cross-sections with 65 nm thickness were generated from the Epon embedded samples and contrasted with 2% Uranyl Acetate in EtOH and Reynolds stain[87]. The TEM imaging was performed on a JEM-2100 (JEOL Ltd.) with 200 kV using a TVIPS TemCam F416 for image registration. For SEM, the samples were dehydrated with acetone and further dried via critical point drying. Before imaging 10–20 nm gold/palladium (80/20) were applied on the sample in an argon bath using a BAL-TEC SCD 005 Sputter Coater (Leica Mikrosysteme). The SEM imaging was performed on a JEOL JSM-7500F scanning electron microscope (JEOL Ltd.).

Quantitative real-time PCR (qRT-PCR)

For RNA isolation, tissue models were washed with PBS- and directly frozen at −80°C. RNA was isolated using the RNeasy Micro Kit (Quiagen) according to the manufacturer’s protocol. Afterward, cDNA synthesis was performed with iScript™ cDNA Synthesis Kit (BioRad) following the manufacturer’s instructions within a thermocycler (Sigma) with 5 min at 25°C, 30 min at 42°C, 5 min at 95°C, and hold at 4°C. RT-qPCR was performed with 25 ng cDNA using the EvaGreen® Supermix (Bio-Rad) and a CFX 96 TouchTM Real-Time PCR Detection 395 System (Bio-Rad). Plates with technical duplicates were analyzed with the following reaction condition: 40 cycles of 95°C for 10 sec, 60°C 10 sec,72°C 30 sec. The following exon-spanning primer pair sequences were used: *OLFM4* (P1: 5’-ACTGTCCGAATTGACATCATGG–3’, P2: 5’-TTCTGAGCTTCCACCAAAACTC–3’[88]), *EF1a* (P1: 5′-AGGTGATTATCCTGAACCATCC-3′, P2: 5′-AAAGGTGGATAGTCTGAGAAG C-3′[16]), (P1: 5’-GCAGCATTACCTGCTCTACGT T-3’, P2: 5’-GCTTGATAAGCTGATGCTGTAA TTT-3’), (P1: 5’-CCGCTACTGGTGTAATGAT GG-3’, P2: 5’-CATCAGCGATGTTATCTTGC AG-3’), (P1: 5’-AGCTTCTACTCTGGTGCACA A-3’, P2: 5’-GGTGGCTGGGAATTGAGA-3’), (P1: 5’-TCACCTTCCCCAGGCCCCTTC-3’, P2: 5’-TGTTCACTGCTGCGATGACCCC-3’). Human *EF1a* was used as a reference gene. Fold changes of gene expression were calculated using the ΔΔCT method.

High-throughput qPCR barrier chip

Gene expression analysis via a high-throughput qPCR barrier chip was performed as previously reported[89–92]. In detail, 20 µl cDNA was produced from 250 ng RNA using the High Capacity cDNA Reverse Transcriptase Kit (Thermo Fisher Scientific). After preamplification of the targets, the high-throughput qPCR chip was performed with the preamplified cDNA in 96 × 96 chips using the Biomark$^{\text{TM}}$ system (Fluidigm®).

HCR-FISH

DNA probes for HCR-FISH were designed as previously described[93]. 25-nucleotide (nt) sequences of each gene were extracted. One probe consists of a pair of two 25-nt long oligos (25-nt encoding region, 2-nt spacer, and 18-nt initiator region) separated by a 2-nt gap. The oligo pairs were selected to have a Tm difference of<5°C. In addition, probe sequences were required to have a GC content within the 40–60% range. The NCBI database was used to look up mRNA sequences. Any probe sequence that contained five or more consecutive bases of the same kind was dropped. To ensure specificity, an NCBI BLAST query was run on each probe against the human transcriptome. BLAST hits on sequences other than the target gene with a 15-nt match were considered off-targets. For each gene, eight encoding probe pairs were designed. Read-out amplifiers B3 with AF546 fluorophores were ordered from Molecular Instruments.

HCR-FISH staining was performed with a HCR-FISH kit (Molecular Instruments) according to the manufacturer's protocol with minor modifications. Briefly, after fixation, the samples were permeabilized in 70% EtOH, washed once with 1× PBS-T and once with 5× saline sodium citrate buffer (Sigma) with 0.5%Tween (BioRad) (SSC-T). For detection, the samples were equilibrated with hybridization buffer and 5 pmol of the detection probe pool set was added. After incubation for 12 h at 37°C, samples were washed with Probe wash buffer at 37°C followed by washing with 5× SSC-T. For amplification, samples were equilibrated with the amplification buffer and incubated for 16 h at RT with 30 pmol of the amplification hairpin solution (AlexaFluor 546). The samples were washed with 5× SSC-T, stained with DAPI (ThermoFisher) diluted in PBS-T. Finally, models were embedded in Fluoromount G (Invitrogen).

For HCR-FlowFISH, dissociated single cells were fixed in 4% PFA and permeabilized with 70% EtOH. The cells were washed with 1× PBS-T and with 5× SSC-T. Afterward, 4×10^5 cells were equilibrated in hybridization buffer, 5 pmol of the detection probe pool set were added and incubated for 12 h at 37°C. After washing with Probe wash buffer and 5×-SSC-T, the cells were equilibrated in the amplification buffer followed by incubation with 30 pmol of the amplification hairpin solution (AlexaFluor 546) for 16 h at RT. Cells were washed using 5× SSC-T with subsequent dilution in 1× PBS- and flow cytometric analysis on a BD FACS Aria III (BD Biosciences). Cells (uninfected, bystander, or infected) were separated into OLFM4(+) and OLFM4(-) based on the AlexaFluor 546 signal. Gates were set discriminating between uninfected and unstained (no detection probes) cells.

Imaging

Imaging was performed with the Keyence B×810 widefield microscope (Keyence), the Leica SP8 confocal microscope (Leica), the Scanning Electron Microscope JSM-7500F (JEOL), and the Transmission Electron Microscope JEM-2100 (JEOL). The representative images were processed with Fiji (v1.51s), and 3D modeling was generated by LasX 3D Visualization.

Cell height measurement

Cell height measurements were performed on cross-section views of z-stack confocal images representative of whole tissue samples. Briefly, a defined grid of 9 × 9 fields was applied on the field of view (FOV). Cell height was determined by measuring the distance between the apical and the basolateral F-actin signal of individual cells located at the cross points of the grid lines. The cell height measurements were performed via Fiji (v1.51s).

Bacterial migration measurement

For the measurement of bacterial migration within the hITM, an automated batch analysis was performed with IMARIS (v8.4.2) on representative z-stack images in 3D projection. In a first step, the threshold determining the apical F-actin signal was set. Secondly, the centers of *S*Tm-expressing GFP signals were defined, and bacterial migration was determined as distance between the defined GFP spot and the apical F-actin signal.

OLFM4 intensity measurement

OLFM4 intensity was determined with Fiji (v1.51s) as follows: 1) z-stacks of images representative of the whole hITM were projected as the sum of all stacks, 2) the mean OLFM4 HCR-FISH signal intensity in the FOV was measured, 3) the number of nuclei per FOV were determined. The OLFM4 intensity was calculated as $OLFM4int = \frac{OLFM4meanintensity}{numberofnuclei}$.

*S*Tm filament length

Filament lengths were determined in whole tissue samples using the Imaris software (v8.4.2) as follows: 1) regions with GFP signals were identified and determined as volumes, 2) the pixel intensity of the GFP signal outside of the defined volumes was set to zero, 3) the inbuilt filament tool was applied in the FOV to identify filaments with 0.5 μm diameter and branching points at a filament length of >2 μm (= minimal length of one individual *S*Tm). Identified filaments with lengths<2 μm were not considered and excluded.

Statistical analysis

Statistical analysis was performed via GraphPad Prism (v6.02) with unpaired t-test, ordinary one-way ANOVA with Tukey's multiple comparison tests, two-ANOVA with Tukey's multiple comparison tests, and Mann-Whitney-U test.

Acknowledgments

We thank S. Reichardt, M. Krafft, and E. Reitenbach for their excellent technical support and assistance; C. Stigloher, C. Gehrig-Höhn, and D. Bunsen for the excellent support during SEM and TEM preparation and imaging; M. Heckmann and C. Werner for kindly provision of IMARIS. A. Gerhartl and A. Brachner for their support in high-throughput qPCR barrier chip measurement; Sara Giddins for editorial support.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The work was supported by the Deutsche Forschungsgemeinschaft (DFG) [270563345]. in the context of the GRK 2157. Furthermore, the DFG funded the SEM (JEOL JSM-7500F) [218894895] and the TEM (JEOL JEM 2100) [218894163]. T.D, A.-E.S., and O.D. thanks DFG for funding through GRK2157. A.-E.S. thanks DFG funding SFB1583 (DECIDE; Project B05). A.-E.S. thank the Single Cell Center Würzburg for support. T.K. thanks DFG funding via SFB1525 (Cardio-Immuno Interface: PS2 Project).

ORCID

Thomas Däullary http://orcid.org/0000-0001-5113-3768
Fabian Imdahl http://orcid.org/0000-0002-3796-2946
Oliver Dietrich http://orcid.org/0000-0002-4776-8960
Tobias Krammer http://orcid.org/0000-0002-7638-9476
Christina Fey http://orcid.org/0000-0001-6860-8184
Winfried Neuhaus http://orcid.org/0000-0002-6552-7183
Marco Metzger http://orcid.org/0000-0003-2053-9401
Jörg Vogel http://orcid.org/0000-0003-2220-1404
Alexander J. Westermann http://orcid.org/0000-0003-3236-0169
Antoine-Emmanuel Saliba http://orcid.org/0000-0001-8539-2784
Daniela Zdzieblo http://orcid.org/0000-0002-6606-2445

Data availability statement

Repository: https://www.ncbi.nlm.nih.gov/geo/ Accession number: GSE217976

References

1. Kolling G, Wu M, Guerrant RL. Enteric pathogens through life stages. Front Cell Infect Microbiol. 2012;2:114. doi:10.3389/fcimb.2012.00114.
2. McCracken KW, Catá EM, Crawford CM, Sinagoga KL, Schumacher M, Rockich BE, Tsai Y-H, Mayhew CN, Spence JR, Zavros Y, et al. Modelling human development and disease in pluripotent stem-cell-derived gastric organoids. Nat. 2014;516(7531):400–404. *20145167531*. doi:10.1038/nature13863.
3. McCracken KW, Howell JC, Wells JM, Spence JR. Generating human intestinal tissue from pluripotent stem cells in vitro. Nat Protoc. 2011;6(12):1920–1928. doi:10.1038/nprot.2011.410.
4. Miura S, Suzuki A. Generation of mouse and human organoid-forming intestinal progenitor cells by direct lineage reprogramming. Cell Stem Cell. 2017;21(4):456–471.e5. doi:10.1016/j.stem.2017.08.020.
5. Spence JR, Mayhew CN, Rankin SA, Kuhar MF, Vallance JE, Tolle K, Hoskins EE, Kalinichenko VV, Wells SI, Zorn AM, et al. Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. Nat. 2010;470(7332):105–109. *2010 4707332*. doi:10.1038/nature09691.
6. Miyoshi H, Stappenbeck TS. In vitro expansion and genetic modification of gastrointestinal stem cells in spheroid culture. Nat Protoc. 2013;8(12):2471–2482. doi:10.1038/nprot.2013.153.
7. Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, van Es JH, Abo A, Kujala P, Peters PJ, et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature. 2009;459(7244):262–265. doi:10.1038/nature07935.
8. Aguilar C, Alves da Silva M, Saraiva M, Neyazi M, Olsson IAS, Bartfeld S. Organoids as host models for infection biology – a review of methods. Exp Mol Med. 2021 *2021*310;53(10):1471–1482. doi:10.1038/s12276-021-00629-4.
9. Aguilar C, Pauzuolis M, Pompaiah M, Vafadarnejad E, Arampatzi P, Fischer M, Narres D, Neyazi M, Kayisoglu Ö, Sell T, et al. Helicobacter pylori shows tropism to gastric differentiated pit cells dependent on urea chemotaxis. Nat Commun. 2022;13(1):5878. doi:10.1038/s41467-022-33165-4.
10. Pinto N, et al. Markers and methods to study adult midgut stem cells. Methods Mol Biol. 2018;1842:123–137.

11. Saxena K, Blutt SE, Ettayebi K, Zeng X-L, Broughman JR, Crawford SE, Karandikar UC, Sastri NP, Conner ME, Opekun AR, et al. Human intestinal enteroids: a new model to study human rotavirus infection, host restriction, and pathophysiology. J Virol. 2016;90(1):43–56. doi:10.1128/JVI.01930-15.
12. Holly MK, Smith JG, Pfeiffer JK. Adenovirus infection of human enteroids reveals interferon sensitivity and preferential infection of goblet cells. J Virol. 2018;92(9). doi:10.1128/JVI.00250-18.
13. Han Y, Yang L, Lacko LA, Chen S. Human organoid models to study SARS-CoV-2 infection. Nat Methods. 2022 194 2022;19(4):418–428. doi:10.1038/s41592-022-01453-y.
14. Foulke-Abel J, In J, Kovbasnjuk O, Zachos NC, Ettayebi K, Blutt SE, Hyser JM, Zeng X-L, Crawford SE, Broughman JR, et al. Human enteroids as an ex-vivo model of host–pathogen interactions in the gastrointestinal tract. Exp Biol Med (Maywood). 2014;239(9):1124–1134. doi:10.1177/1535370214529398.
15. Nickerson KP, Llanos-Chea A, Ingano L, Serena G, Miranda-Ribera A, Perlman M, Lima R, Sztein MB, Fasano A, Senger S, et al. A versatile human intestinal organoid-derived epithelial monolayer model for the study of enteric pathogens. Microbiol Spectr. 2021;9(1). doi:10.1128/Spectrum.00003-21.
16. Schweinlin M, Wilhelm S, Schwedhelm I, Hansmann J, Rietscher R, Jurowich C, Walles H, Metzger M. Development of an advanced primary human in vitro model of the small intestine. Tissue Eng Part C Methods. 2016;22(9):873–883. doi:10.1089/ten.tec.2016.0101.
17. Crowe P, Marsh M. Morphometric analysis of small intestinal mucosa. IV. determining cell volumes. Virchows Arch A Pathol Anat Histopathol. 1993;422(6):459–466. doi:10.1007/BF01606454.
18. Srinivasan B, Kolli AR, Esch MB, Abaci HE, Shuler ML, Hickman JJ. TEER measurement techniques for in vitro barrier model systems. J Lab Autom. 2015;20(2):107–126. doi:10.1177/2211068214561025.
19. Owen RL, Jones AL. Epithelial cell specialization within human peyer's patches: an ultrastructural study of intestinal lymphoid follicles. Gastroenterology. 1974;66(2):189–203. doi:10.1016/S0016-5085(74)80102-2.
20. Grondin JA, Kwon YH, Far PM, Haq S, Khan WI. Mucins in intestinal mucosal defense and inflammation: learning from clinical and experimental studies. Front Immunol. 2020;11:2054. doi:10.3389/fimmu.2020.02054.
21. Revenu C, Ubelmann F, Hurbain I, El-Marjou F, Dingli F, Loew D, Delacour D, Gilet J, Brot-Laroche E, Rivero F, et al. A highlights from MBoC selection: a new role for the architecture of microvillar actin bundles in apical retention of membrane proteins. Mol Biol Cell. 2012;23(2):324. doi:10.1091/mbc.e11-09-0765.
22. Khurana S, George SP. Regulation of cell structure and function by actin-binding proteins: villin's perspective. FEBS Lett. 2008;582(14):2128–2139. doi:10.1016/j.febslet.2008.02.040.
23. Elmentaite R, Kumasaka N, Roberts K, Fleming A, Dann E, King HW, Kleshchevnikov V, Dabrowska M, Pritchard S, Bolt L, et al. Cells of the human intestinal tract mapped across space and time. Nature. 2021;597(7875):250–255. doi:10.1038/s41586-021-03852-1.
24. Wang Y, Song W, Wang J, Wang T, Xiong X, Qi Z, Fu W, Yang X, Chen Y-G. Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. J Exp Med. 2020;217(2):1–15. doi:10.1084/jem.20191130.
25. Haber AL, Biton M, Rogel N, Herbst RH, Shekhar K, Smillie C, Burgin G, Delorey TM, Howitt MR, Katz Y, et al. A single-cell survey of the small intestinal epithelium. Nature. 2017;551(7680):333–339. doi:10.1038/nature24489.
26. Burclaff J, Bliton RJ, Breau KA, Ok MT, Gomez-Martinez I, Ranek JS, Bhatt AP, Purvis JE, Woosley JT, Magness ST. A proximal-to-distal survey of healthy adult human small intestine and colon epithelium by single-cell transcriptomics. Cell Mol Gastroenterol Hepatol. 2022;13(5):1554–1589. doi:10.1016/j.jcmgh.2022.02.007.
27. Kimura S, Yamakami-Kimura M, Obata Y, Hase K, Kitamura H, Ohno H, Iwanaga T. Visualization of the entire differentiation process of murine M cells: suppression of their maturation in cecal patches. Mucosal Immunol. 2015 *201483*;8(3):650–660. doi:10.1038/mi.2014.99.
28. Mabbott NA, Donaldson DS, Ohno H, Williams IR, Mahajan A. Microfold (M) cells: important immunosurveillance posts in the intestinal epithelium. Mucosal Immunol. 2013;6(4):666–677. doi:10.1038/mi.2013.30.
29. Madsen J, Nielsen O, Tornøe I, Thim L, Holmskov U. Tissue localization of human trefoil factors 1, 2, and 3. J Histochem Cytochem. 2007;55(5):505–513. doi:10.1369/jhc.6A7100.2007.
30. Rupnik M, Kreft M, Nothias F, Grilc S, Bobanovic LK, Johannes L, Kiauta T, Vernier P, Darchen F, Zorec R. Distinct role of Rab3A and Rab3B in secretory activity of rat melanotrophs. Am J Physiol - Cell Physiol. 2007;292(1):98–105. doi:10.1152/ajpcell.00005.2006.
31. Ryu GR, Lee E, Kim JJ, Moon S-D, Ko S-H, Ahn Y-B, Song K-H. Comparison of enteroendocrine cells and pancreatic β-cells using gene expression profiling and insulin gene methylation. PLoS One. 2018;13(10):e0206401. doi:10.1371/journal.pone.0206401.
32. Parikh K, Antanaviciute A, Fawkner-Corbett D, Jagielowicz M, Aulicino A, Lagerholm C, Davis S, Kinchen J, Chen HH, Alham NK, et al. Colonic epithelial cell diversity in health and inflammatory bowel

disease. Nature. 2019;567(7746):49–55. doi:10.1038/s41586-019-0992-y.

33. Larock DL, Chaudhary A, Miller SI. Salmonellae interactions with host processes. Nat Rev Microbiol. 2015 *201534*;13(4):191–205. doi:10.1038/nrmicro3420.
34. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. Methods. 2001;25(4):402–408. doi:10.1006/meth.2001.1262.
35. VanDussen KL, Carulli AJ, Keeley TM, Patel SR, Puthoff BJ, Magness ST, Tran IT, Maillard I, Siebel C, Kolterud Å, et al. Notch signaling modulates proliferation and differentiation of intestinal crypt base columnar stem cells. Development. 2012;139(3):488–497. doi:10.1242/dev.070763.
36. Odenwald MA, Turner JR. The intestinal epithelial barrier: a therapeutic target? Nat Rev Gastroenterol Hepatol. 2016;14(1):9–21. doi:10.1038/nrgastro.2016.169.
37. Schulte LN, Schweinlin M, Westermann AJ, Janga H, Santos SC, Appenzeller S, Walles H, Vogel J, Metzger M. An advanced human intestinal coculture model reveals compartmentalized host and pathogen strategies during salmonella infection. MBio. 2020;11(1). doi:10.1128/mBio.03348-19.
38. Meran L, Massie I, Campinoti S, Weston AE, Gaifulina R, Tullie L, Faull P, Orford M, Kucharska A, Baulies A, et al. Engineering transplantable jejunal mucosal grafts using patient-derived organoids from children with intestinal failure. Nat Med. 2020;26(10):1593–1601. doi:10.1038/s41591-020-1024-z.
39. Siddiqui KM, Chopra DP. Primary and long term epithelial cell cultures from human fetal normal colonic mucosa. Vitro. 1984;20(11):859–868. doi:10.1007/BF02619632.
40. Kanaya T, Sakakibara S, Jinnohara T, Hachisuka M, Tachibana N, Hidano S, Kobayashi T, Kimura S, Iwanaga T, Nakagawa T, et al. Development of intestinal M cells and follicle-associated epithelium is regulated by TRAF6-mediated NF-κB signaling. J Exp Med. 2018;215(2):501–519. doi:10.1084/jem.20160659.
41. Nakamura Y, Kimura S, Hase K. M cell-dependent antigen uptake on follicle-associated epithelium for mucosal immune surveillance. Inflamm Regen. 2018;38(1):1–9. doi:10.1186/s41232-018-0072-y.
42. Knoop KA, Kumar N, Butler BR, Sakthivel SK, Taylor RT, Nochi T, Akiba H, Yagita H, Kiyono H, Williams IR. RANKL is necessary and sufficient to initiate development of antigen-sampling M cells in the intestinal epithelium. J Immunol. 2009;183(9):5738–5747. doi:10.4049/jimmunol.0901563.
43. Nagashima K, Sawa S, Nitta T, Tsutsumi M, Okamura T, Penninger JM, Nakashima T, Takayanagi H. Identification of subepithelial mesenchymal cells that induce IgA and diversify gut microbiota. Nat Immunol. 2017 *2017186*;18(6):675–682. doi:10.1038/ni.3732.
44. Böttcher A, Büttner M, Tritschler S, Sterr M, Aliluev A, Oppenländer L, Burtscher I, Sass S, Irmler M, Beckers J, et al. Non-canonical Wnt/PCP signalling regulates intestinal stem cell lineage priming towards enteroendocrine and Paneth cell fates. Nat Cell Biol. 2021;23(1):23–31. doi:10.1038/s41556-020-00617-2.
45. da Costa Ferreira S, Sadissou IA, Parra RS, Feitosa MR, Neto FSL, Pretti da Cunha Tirapelli D, Ramalho LNZ, Féres O, da Rocha JJR, Donadi EA, et al. Increased HLA-G expression in tissue-infiltrating cells in inflammatory bowel diseases. Dig Dis Sci. 2021;66(8):2610–2618. doi:10.1007/s10620-020-06561-3.
46. Onno M, Guillaudeux T, Amiot L, Renard I, Drenou B, Hirel B, Girr M, Semana G, Le Bouteiller P, Fauchet R. The HLA-G gene is expressed at a low mRNA level in different human cells and tissues. Hum Immunol. 1994;41(1):79–86. doi:10.1016/0198-8859(94)90089-2.
47. Malik-Kale P, Jolly CE, Lathrop S, Winfree S, Luterbach C, Steele-Mortimer O. Salmonella – at home in the host cell. Front Microbiol. 2011;2:125. doi:10.3389/fmicb.2011.00125.
48. Fattinger SA, Böck D, Di Martino ML, Deuring S, Samperio Ventayol P, Ek V, Furter M, Kreibich S, Bosia F, Müller-Hauser AA, et al. Salmonella typhimurium discreet-invasion of the murine gut absorptive epithelium. PLoS Pathog. 2020;16(5):e1008503. doi:10.1371/journal.ppat.1008503.
49. Fattinger SA, Sellin ME, Hardt WD. Salmonella effector driven invasion of the gut epithelium: breaking in and setting the house on fire. Curr Opin Microbiol. 2021;64:9–18. doi:10.1016/j.mib.2021.08.007.
50. Backert S, Boehm M, Wessler S, Tegtmeyer N. Transmigration route of campylobacter jejuni across polarized intestinal epithelial cells: paracellular, transcellular or both? Cell Commun Signal. 2013;11(1):72. doi:10.1186/1478-811X-11-72.
51. Burns JL, Griffith A, Barry JJ, Jonas M, Chi EY. Transcytosis of gastrointestinal epithelial cells by escherichia coli K1. Pediatr Res. 2001 *2001491*;49(1):30–37. doi:10.1203/00006450-200101000-00010.
52. Colonne PM, Winchell CG, Voth DE. Hijacking host cell highways: manipulation of the host actin cytoskeleton by obligate intracellular bacterial pathogens. Front Cell Infect Microbiol. 2016;6:107. doi:10.3389/fcimb.2016.00107.
53. Fulde M, van Vorst K, Zhang K, Westermann AJ, Busche T, Huei YC, Welitschanski K, Froh I, Pägelow D, Plendl J, et al. SPI2 T3SS effectors facilitate enterocyte apical to basolateral transmigration of Salmonella-containing vacuoles in vivo. Gut Microbes. 2021;13(1). doi:10.1080/19490976.2021.1973836.
54. Justice SS, Hunstad DA, Cegelski L, Hultgren SJ. Morphological plasticity as a bacterial survival strategy. Nat Rev Microbiol. 2008 *200862*;6(2):162–168. doi:10.1038/nrmicro1820.
55. Abell-King C, Costas A, Duggin IG, Söderström B, Coers J. Bacterial filamentation during urinary tract

infections. PLoS Pathog. 2022;18(12):e1010950. doi:10.1371/journal.ppat.1010950.

56. Lensmire JM, Pratt ZL, Wong ACL, Kaspar CW. Phosphate and carbohydrate facilitate the formation of filamentous salmonella Enterica during osmotic stress. Microbiol (United Kingdom). 2018;164 (12):1503–1513. doi:10.1099/mic.0.000731.
57. Mattick KL, Phillips LE, Jørgensen F, Lappin-Scott HM, Humphrey TJ. Filament formation by salmonella spp. inoculated into liquid food matrices at refrigeration temperatures, and growth patterns when warmed. J Food Prot. 2003;66(2):215–219. doi:10.4315/0362-028X-66.2.215.
58. Henry T, Garcia-Del Portillo F, Gorvel JP. Identification of Salmonella functions critical for bacterial cell division within eukaryotic cells. Mol Microbiol. 2005;56(1):252–267. doi:10.1111/j.1365-2958.2005.04540.x.
59. Humphrey S, MacVicar T, Stevenson A, Roberts M, Humphrey TJ, Jepson MA. SulA-induced filamentation in salmonella enterica serovar typhimurium: effects on SPI-1 expression and epithelial infection. J Appl Microbiol. 2011;111(1):185–196. doi:10.1111/j.1365-2672.2011.05022.x.
60. Martínez-Lorenzo MJ, Méresse S, De Chastellier C, Gorvel JP. Unusual intracellular trafficking of Salmonella typhimurium in human melanoma cells. Cell Microbiol. 2001;3(6):407–416. doi:10.1046/j.1462-5822.2001.00123.x.
61. Schuster EM, Epple MW, Glaser KM, Mihlan M, Lucht K, Zimmermann JA, Bremser A, Polyzou A, Obier N, Cabezas-Wallscheid N, et al. TFEB induces mitochondrial itaconate synthesis to suppress bacterial growth in macrophages. Nat Metab. 2022;4(7):856–866. doi:10.1038/s42255-022-00605-w.
62. Dhar P, McAuley J. The role of the cell surface mucin MUC1 as a barrier to infection and regulator of inflammation. Front Cell Infect Microbiol. 2019;9:117. doi:10.3389/fcimb.2019.00117.
63. van der Flier LG, Haegebarth A, Stange DE, van de Wetering M, Clevers H. OLFM4 is a robust marker for stem cells in human intestine and marks a subset of colorectal cancer cells. Gastroenterology. 2009;137 (1):15–17. doi:10.1053/j.gastro.2009.05.035.
64. Mannick EE, Schurr JR, Zapata A, Lentz JJ, Gastanaduy M, Cote RL, Delgado A, Correa P, Correa H. Gene expression in gastric biopsies from patients infected with helicobacter pylori. Scand J Gastroenterol. 2004;39(12):1192–1200. doi:10.1080/00365520410003588.
65. Liu W, Yan M, Sugui JA, Li H, Xu C, Joo J, Kwon-Chung KJ, Coleman WG, Rodgers GP. Olfm4 deletion enhances defense against Staphylococcus aureus in chronic granulomatous disease. J Clin Invest. 2013;123 (9):3751–3755. doi:10.1172/JCI68453.
66. Ramilo O, Allman W, Chung W, Mejias A, Ardura M, Glaser C, Wittkowski KM, Piqueras B, Banchereau J, Palucka AK, et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. Blood. 2007;109(5):2066–2077. doi:10.1182/blood-2006-02-002477.
67. Huan YW, Bengtsson RJ, MacIntyre N, Guthrie J, Finlayson H, Smith SH, Archibald AL, Ait-Ali T. Lawsonia intracellularis exploits β-catenin/wnt and notch signalling pathways during infection of intestinal crypt to alter cell homeostasis and promote cell proliferation. PLoS One. 2017;12(3):e0173782. doi:10.1371/journal.pone.0173782.
68. Fitzsimonds ZR, Liu C, Stocke KS, Yakoumatos L, Shumway B, Miller DP, Artyomov MN, Bagaitkar J, Lamont RJ. Regulation of olfactomedin 4 by Porphyromonas gingivalis in a community context. Isme J. 2021 *2021159*;15(9):2627–2642. doi:10.1038/s41396-021-00956-4.
69. Kuno R, Ito G, Kawamoto A, Hiraguri Y, Sugihara HY, Takeoka S, Nagata S, Takahashi J, Tsuchiya M, Anzai S, et al. Notch and TNF-α signaling promote cytoplasmic accumulation of OLFM4 in intestinal epithelium cells and exhibit a cell protective role in the inflamed mucosa of IBD patients. Biochem Biophys Reports, 2021;25:100906. doi:10.1016/j.bbrep.2020.100906.
70. Westermann AJ, Förstner KU, Amman F, Barquist L, Chao Y, Schulte LN, Müller L, Reinhardt R, Stadler PF, Vogel J. Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions. Nature. 2016;529 (7587):496–501. doi:10.1038/nature16547.
71. Abuaita BH, Lawrence ALE, Berger RP, Hill DR, Huang S, Yadagiri VK, Bons B, Fields C, Wobus CE, Spence JR, et al. Comparative transcriptional profiling of the early host response to infection by typhoidal and non-typhoidal Salmonella serovars in human intestinal organoids. PLoS Pathog. 2021;17(10):e1009987. doi:10.1371/journal.ppat.1009987.
72. Lawrence ALE, Abuaita BH, Berger RP, Hill DR, Huang S, Yadagiri VK, Bons B, Fields C, Wobus CE, Spence JR, et al. Salmonella enterica serovar typhimurium spi-1 and spi-2 shape the global transcriptional landscape in a human intestinal organoid model system. MBio. 2021;12(3). doi:10.1128/mBio.00399-21.
73. Shinozaki S, Nakamura T, Iimura M, Kato Y, Iizuka B, Kobayashi M, Hayashi N. Upregulation of Reg 1α and GW112 in the epithelium of inflamed colonic mucosa. Gut. 2001;48(5):623–629. doi:10.1136/gut.48.5.623.
74. Gersemann M, Becker S, Nuding S, Antoni L, Ott G, Fritz P, Oue N, Yasui W, Wehkamp J, Stange EF. Olfactomedin-4 is a glycoprotein secreted into mucus in active IBD. J Crohn's Colitis. 2012;6(4):425–434. doi:10.1016/j.crohns.2011.09.013.
75. Kim YS, Ho SB. Intestinal goblet cells and mucins in health and disease: recent insights and progress. Curr Gastroenterol Rep. 2010;12(5):319–330. doi:10.1007/s11894-010-0131-2.
76. Clevers HC, Bevins CL. Paneth cells: maestros of the small intestinal crypts. Annu Rev Physiol. 2013;75

(1):289–311. doi:10.1146/annurev-physiol-030212-183744.

77. Worthington JJ, Reimann F, Gribble FM. Enteroendocrine cells-sensory sentinels of the intestinal environment and orchestrators of mucosal immunity. Mucosal Immunol. 2018;11(1):3–20. doi:10.1038/mi.2017.73.
78. Gerbe F, Jay P. Intestinal tuft cells: epithelial sentinels linking luminal cues to the immune system. Mucosal Immunol. 2016;9(6):1353–1359. doi:10.1038/mi.2016.68.
79. Beumer J, Artegiani B, Post Y, Reimann F, Gribble F, Nguyen TN, Zeng H, Van den Born M, Van Es JH, Clevers H. Enteroendocrine cells switch hormone expression along the crypt-to-villus BMP signalling gradient. Nat Cell Biol. 2018 *2018208*;20(8):909–916. doi:10.1038/s41556-018-0143-y.
80. Pleguezuelos-Manzano C, Puschhof J, den Brink S, Geurts V, Beumer J, Clevers H. Establishment and culture of human intestinal organoids derived from adult stem cells. Curr Protoc Immunol. 2020;130(1):e106. doi:10.1002/cpim.106.
81. Jannasch M, Groeber F, Brattig NW, Unger C, Walles H, Hansmann J. Development and application of three-dimensional skin equivalents for the investigation of percutaneous worm invasion. Exp Parasitol. 2015;150:22–30. doi:10.1016/j.exppara.2015.01.005.
82. Pusch J, Votteler M, Göhler S, Engl J, Hampel M, Walles H, Schenke-Layland K. The physiological performance of a three-dimensional model that mimics the microenvironment of the small intestine. Biomaterials. 2011;32(30):7469–7478. doi:10.1016/j.biomaterials.2011.06.035.
83. Sato T, Stange DE, Ferrante M, Vries RGJ, van Es JH, van den Brink S, van Houdt WJ, Pronk A, van Gorp J, Siersema PD, et al. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology. 2011;141(5):1762–1772. doi:10.1053/j.gastro.2011.07.050.
84. VanDussen KL, Sonnek NM, Stappenbeck TS. L-WRN conditioned medium for gastrointestinal epithelial stem cell culture shows replicable batch-to-batch activity levels across multiple research teams. Stem Cell Res. 2019;37:101430. doi:10.1016/j.scr.2019.101430.
85. Vonk AM, van Mourik P, Ramalho AS, Silva IAL, Statia M, Kruisselbrink E, Suen SWF, Dekkers JF, Vleggaar FP, Houwen RHJ, et al. Protocol for application, standardization and validation of the forskolin-induced swelling assay in cystic fibrosis human colon organoids. STAR Protoc. 2020;1(1):100019. doi:10.1016/j.xpro.2020.100019.
86. Haddock NH. Alcian blue, a new phthalocyanine dyestuff. Research; a Journal of Science and Its Applications. 1948;1:685–689.
87. Reynolds ES. THE USE of LEAD CITRATE at HIGH pH as an ELECTRON-OPAQUE STAIN in ELECTRON MICROSCOPY. J Cell Biol. 1963;17(1):208. doi:10.1083/jcb.17.1.208.
88. Freire R, Ingano L, Serena G, Cetinbas M, Anselmo A, Sapone A, Sadreyev RI, Fasano A, Senger S. Human gut derived-organoids provide model to study gluten response and effects of microbiota-derived molecules in celiac disease. Sci Rep. 2019;9(1):1–15. doi:10.1038/s41598-019-43426-w.
89. Krasemann S, Haferkamp U, Pfefferle S, Woo MS, Heinrich F, Schweizer M, Appelt-Menzel A, Cubukova A, Barenberg J, Leu J, et al. The blood-brain barrier is dysregulated in COVID-19 and serves as a CNS entry route for SARS-CoV-2. Stem Cell Rep. 2022;17(2):307–320. doi:10.1016/j.stemcr.2021.12.011.
90. Lin GC, Leitgeb T, Vladetic A, Friedl H-P, Rhodes N, Rossi A, Roblegg E, Neuhaus W. Optimization of an oral mucosa in vitro model based on cell line TR146. Tissue Barriers. 2020;8(2):1748459. doi:10.1080/21688370.2020.1748459/SUPPL_FILE/KTIB_A_1748459_SM4220.ZIP.
91. Gerhartl A, Pracser N, Vladetic A, Hendrikx S, Friedl H-P, Neuhaus W. The pivotal role of micro-environmental cells in a human blood–brain barrier in vitro model of cerebral ischemia: functional and transcriptomic analysis. Fluids Barriers CNS. 2020;17(1):1–17. doi:10.1186/s12987-020-00179-3.
92. Ramme AP, Koenig L, Hasenberg T, Schwenk C, Magauer C, Faust D, Lorenz AK, Krebs A-C, Drewell C, Schirrmann K, et al. Autologous induced pluripotent stem cell-derived four-organ-chip. Futur Sci OA. 2019;5(8):413–2056. doi:10.2144/fsoa-2019-0065.
93. Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J, Cunha A, Pierce NA. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Dev. 2018;145(12). doi:10.1242/dev.165753.

3 Discussion

In the research projects presented above, I was able to discover relevant disease and model characteristics by focusing on the description of cellular phenotypes. The biological samples I investigated were obtained from patients, donors as well as mouse models and *in vitro* models of human tissue, all of which come with their own set of prospects and challenges. In the following discussion, I want to highlight similarities and differences between the observational studies we conducted with clinical specimens and experiments we performed on disease models. I want to emphasize the opportunity to combine information from multiple sources in order to improve our understanding of disease biology and most importantly to solve problems in the clinics.

3.1 Exploring Biological Systems

Modern Biology aims to understand different states and interactions within the tens of thousands of components of a biological system. The key challenge of such data is heterogeneity [97]. A major way to address this challenge is detailed knowledge about the biological systems and their components that were used to generate these data.

Simple cell culture systems usually contain only a single cell type that resembles a relatively homogeneous population. While these models are frequently used in biomedical research, they fail to mimic important tissue functions. This helps to explain that transferring results from research models into clinical applications shows a low success rate [25, 98].

Advanced cell culture systems show the potential to fill the gap between monotypic cells with little heterogeneity and complex organisms. Three-dimensional organoids build complex structures containing multiple cell types and can be cultivate over long periods of time. Furthermore, creating organoid-derived monolayers can be beneficial for infection research by providing access to the apical side of cells and facilitating readout by microscopy [25]. Cell culture media containing defined growth factors can drive differentiation of stem cells into specific cell types and thus facilitate interrogation of defined aspects of a biological system. The improvement of models through tissue engineering create the opportunity for more reliable research. However, the degree of maturation of cell types from stem cells and their fidelity to the *in vivo* counterparts are important characteristics the must be addressed in order to provide reliable insights [34].

Mouse models, or animal models in general, facilitate studies of infectious diseases within fully mature tissues and capture interconnections between organs and cell types such as epithelia and immune cells. While certain biological processes are conserved across species, and can thus be studied more efficiently in simpler animals, prediction of clinical outcomes using animal models is so far not accurate enough. This is demonstrated by the failure of more than 80 % of therapeutics in human trials after they have been suggested safe and effective after animal studies [32].

Clinical samples are directly associated to disease and should be informative about the development and outcomes of disease. However, most clinical research does not lead to favorable change in the decision making concerning the prevention, diagnosis, treatment or prognosis of disease [99].

High-throughput sequencing studies can be combined across studies based on quantification of the same features. The widespread application of these technologies therefore generates a large body of knowledge that can be re-used. Data integration can thus combine newly generated data with prior knowledge to provide more context for the analysis. Generating data sets that combine the fidelity of phenotypes from clinical data with the ease of experimentation in cell culture and the systemic response in animal models might mitigate the individual downsides and improve the overall usefulness of biomedical research.

3.1.1 Defining Phenotypes

Single-cell genomics is revolutionary because it can reveal unknown diversity in biological systems. Neutrophils have long been considered a homogeneous and transcriptionally inactive cell type. Nonetheless, single-cell sequencing has revealed a range of transcriptional states in the blood of human donors that potentially impact the characterization of this cell type in health and disease [29]. Annotation of cells based on well-characterized cell types, previously discovered subtypes or novel states is a requirement for further exploration. Cell type annotation is often performed manually based on canonical marker genes and domain knowledge. However, more reproducible approaches based on computational models and either marker genes or reference data are available. Especially for larger number of samples, automated or semi-automated annotation of cell types becomes increasingly important [71].

In our analyses of clinical data, we first annotated cell clusters based on the expression of canonical marker genes. We focused primarily on human leukocytes and were able to successfully distinguish the major cell types of the immune system [23]. Cell subtypes were assigned after extensive exploratory analysis and we performed systematic comparisons to other studies in order to validate the macrophage subsets we found in severe COVID-19. Both gene set enrichment and data integration with data sets from COVID-19 and idiopathic pulmonary fibrosis (IPF) showed conservation of the identified subsets across studies.

Previously published data sets from biological samples with similar cell types can serve as a reference. These can originate from the same condition but also from others such as the healthy tissue or related diseases. Integration of data with a reference can not only improve annotation but uncertainty in the transfer of labels from reference to query can even highlight areas of interest [100]. I intend to use our previously published data sets as well as public data sets in order to further improve cell annotation and the identification of novel states associated to disease.

3.1.2 Maturation & Fidelity

Biological models can reproduce features of a disease and therefore facilitate experiments that aim to discover mechanistic insights into disease development, progression, and outcomes. Identification of the target cells and strictness of cell type tropism of pathogens is an important aspect of infection research. However, reliable assessment relies on faithful recapitulation of the host cell phenotypes. For this, maturation and fidelity have been introduced as metrics to judge the similarity of cultured cells to their *in vivo* counterparts (see figure 3.1) [34].

Our characterization of an *in vitro* model of the human small intestine showed remarkable recapitulation of the differentiation trajectory from CBC stem cells, marked by LGR5 expression, via transit amplifying cells towards enterocytes. Secretory lineage differentiation, governed by the Notch pathway and identifiable by expression of *DLL1, DLL4 & NEURL1*, was detectable in the human intestine *in vivo* but not in the *in vitro* model [34, 95].

Studying *Salmonella typhimurium* infection in such a model has clear advantages over the use of monotypic cell culture, especially cancer cell lines, due to the greater phenotypic similarity to the human intestine on the morphological, structural and transcriptomic level [95]. Investigation of invasion, spatial localization and bacterial structures shows great potential for further experiments that study the infection dynamics of the intestine.

However, investigation of low, medium and high bacterial load across host cells by single-cell sequencing proved too complex to draw reliable conclusions. The excessive variability originates both from the library preparation of single-cell technologies and transcriptional noise in the differentiation of cells in the *in vitro* model. Furthermore, assessment of cell tropism of *S. typhimurium* was not possible due to the lack of secretory cells and the fidelity of M-like cells [95].

In spite of these difficulties, using organoids as model for *Salmonella* infection is promising. The different *Salmonella enterica* serovars Typhimurium, a model system, and Typhi, the causal agent of life-threatening disease, stimulate different transcriptomic responses in cultured epithelial cells [101]. A fundamental dichotomy lies between the systemic infection – typhoid fever – of *S. thyphi* in humans and *S. thyphimurium* in mice and the localized, self-limiting infection vice versa (e.g. *S. typhimurium* in humans). This difference is highly interesting and direct comparisons of the localized infection might only be possible in organoid models.

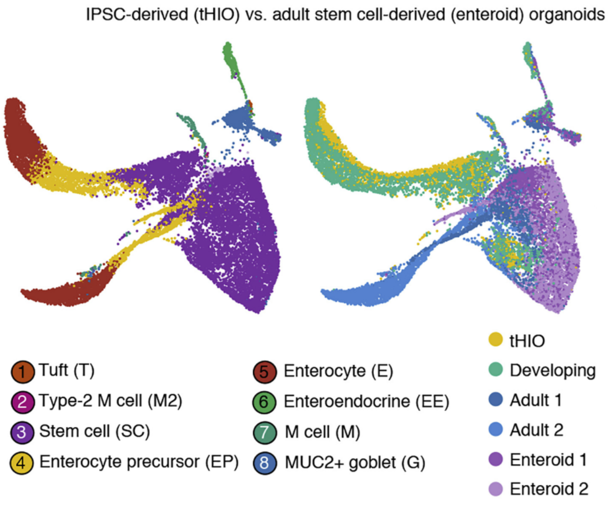


Figure 3.1: Integration of intestinal tissue with intestinal organoids reveals maturation and fidelity of the individual cell types. Adapted from [34].

3.1.3 Selecting Comparisons

In simple experiments, selecting appropriate controls delivers straightforward comparisons of a sample to the absence and presence of an effect (e.g. band on a gel) as negative and positive control, respectively.

High-throughput sequencing studies measure many more features but are often performed with a more reduced setup. A typical example consists of comparison between case and control with a small number of replicates, usually three. Differences between those two conditions are then considered informative about the case. In our study of *S. aureus* infection in mice we managed to demonstrate how such simple comparisons can provide useful results [94]. However, simple comparisons tend to require large effects such as the difference between mature and immature neutrophils in our study. Stimulating cells with agents that induce well-characterized pathways can add controls that facilitate interpretation of results. In our investigation of profibrotic macrophages we were able to compare monocytes from healthy donors to those that were stimulated *in vitro* with SARS-CoV-2, R848 or 3p-hpRNA. The presence of multiple conditions greatly improved the interpretability of results [93].

The analysis of clinical samples can suffer from the lack of controls, especially a reliable characterization of what is considered healthy and how it differs across individuals. While the comparison of cell types and states between patients is often reliable, differences between patients can be unreliable. Therefore, analyses within a data set can be insufficient. Comparison of cellular phenotypes across data sets, including multiple individuals and conditions, can mitigate the lack of information that comes from constraints in the study setup. Our investigation of macrophages across health and disease using both gene set over-representation and integration connected the disease phenotypes of severe COVID-19 ARDS and IPF.

Systematic comparisons between phenotypes across conditions require both evaluation of integration and methods to assess changes across both major cell types and more subtle cell states [68, 89]. The prospect of integrating data across studies and comparing conditions lies in the potential to definie features of the pathomechanism and detect host factors that determine different disease outcomes.

3.1.4 Determining Outcome

Diseases have multiple outcomes with the two extremes being complete recovery or death. Long term adverse effects of disease are possible and brought to public attention during the COVID-19 pandemic due to long COVID, a multisystemic condition more frequent than severe disease [102].

In our study of severe COVID-19, we identified profibrotic macrophages and showed that they are associated to the development of lung fibrosis as a consequence of SARS-CoV-2 infection [93]. However, the connection between lung fibrosis as a disease manifestation and the final outcomes of recovery and death are not sufficiently understood. Similarly, the connection between COVID-19 severity and long-term consequences such as long-COVID requires more rigorous investigation. In order to further investigate these disease outcomes I intend to analyze a collection of longitudinal BAL samples gathered from patients that received intensive care for severe COVID-19. Comparing samples during disease progression towards diverging outcomes should reveal distinguishing features that improves our understanding of the disease (see figure 3.2).

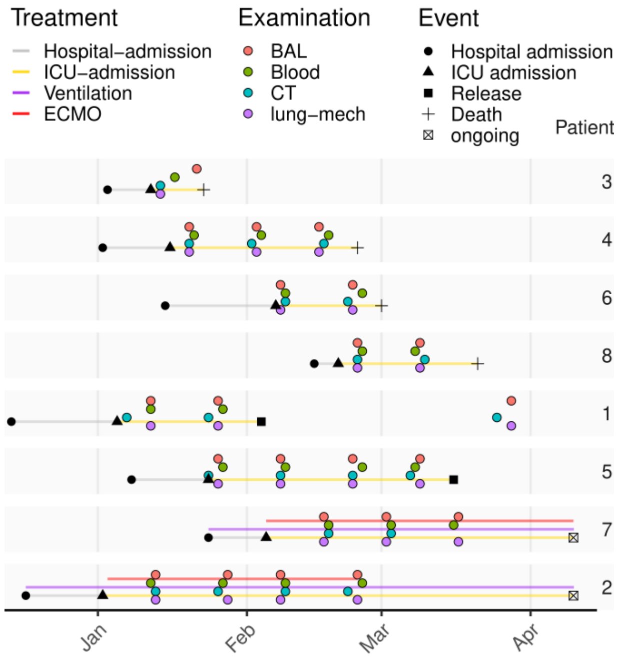


Figure 3.2: Longitudinal samples towards diverging outcomes can reveal distinguishing features.

BAL fluid (BALF) can only be obtained based on clinical need and is usually restricted to patients with severe disease. In contrast, blood is routinely collected even from individuals with mild disease or healthy donors. Analyses focused on leukocytes in the blood have revealed dysfunctional monocytes and neutrophils in severe COVID-19 [103]. Meanwhile, systematic comparisons to the healthy lung depend on comprehensive references such as the Human Lung Cell Atlas (HLCA) [104]. I intend to integrate cellular transcriptomes from both blood and BALF with the reference atlas of the human lung to better define phenotypes, assess migration patterns and investigate associations to clinical outcomes.

Clinical observations are limited when the target of investigation are isolated features of the pathomechanism or when manipulation is desired. Moreover, certain tissues are not accessible and might only be investigated post mortem. For this, biological models of the disease or certain features of it are useful or even necessary.

3.1.5 Functional Characterization

Phenotypes originate from the expression of genes but have behavioral and functional consequences that govern tissue homeostasis and response to infection. While scRNA-seq can distinguish cell populations based on RNA transcripts the more traditional description in immunology is based on epitopes, often called clusters of differentiation (CD) [23]. While cell populations can be sorted by FACS and subjected to scRNA-seq separately, CITE-seq, measuring both mRNA and epitopes, can directly connect the traditional description by flow cytometry with the high resolution of scRNA-seq [49]. Precise descriptions of cell states can facilitate isolation of particular populations for functional investigation. However, access to well defined populations is often impaired by the lack of specific epitopes or limited availability of cells, such as in clinical samples.

Another approach for functional investigation is the specific induction of phenotypes in cultured cells. In our study of profibrotic macrophages in severe COVID-19 we have already shown that stimulation of monocytes from healthy human donors with SARS-CoV-2 induces a transcriptional phenotype related to profibrotic macrophages in the lung [93]. Functional consequences of this phenotype on other cell types such as fibroblasts can likely be determined by using co-culture systems. Once such a system has been set up, it would be possible to study the effects of other pathogens on monocyte-fibroblast interactions.

The reaction and susceptibility of epithelial cell types to bacterial infections should be reflected in the functional features of the barrier, the secreted products or the permissiveness to pathogens. To this end, studies of epithelial cell types driven to distinct states using advanced culture systems should be able to reveal important aspects of tropism and innate immunity.

3.1.6 Reproducibility

Single-cell sequencing has the potential to advance our understanding of disease by dramatically increasing the resolution and thus explaining important aspects of the heterogeneity in biological systems. However, variability between samples is still driven by many factors, including inter-individual differences, culture conditions or randomness in the capture of cells.

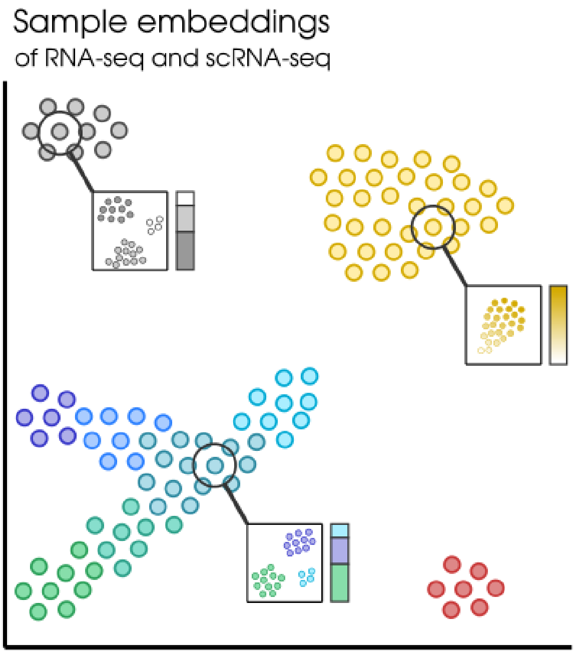


Figure 3.3: Sample embeddings reveal structure in data collections. Single-cell RNA-seq can reveal heterogeneity of cell types in RNA-seq samples using deconvolution.

Clinical data is heavily dominated by variability between patients. Integration of data between patients is therefore necessary but needs technical expertise and thorough investigation. Creating collections of samples that are well defined and interpretable is therefore a key challenge in the future of clinical single-cell analysis.

Characterization of research models is essential in order to judge the reproducibility of results and potential problems for translation [32]. Typically, a small number of replicates is used to approximate measurement error and infer significant differences. Unknown sources of variation can lead to inconsistent findings and thus impede analysis. In a study of *Bordetella pertussis* infection in an *in vitro* model of the human airway, we detected dramatic changes in the abundance of cell types. Reproducibility of cell culture therefore constitutes a major requirement for the use of advanced model systems for infection studies.

Embedding samples instead of cells can facilitate a higher-level summary of the data that captures reproducibility of models or patient-specific differences across diseases (see figure 3.3). Reversing that process, RNA-seq data can be deconvoluted using scRNA-seq measurements of similar samples in well-characterized conditions [105]. Accordingly, bulk assays have to potential to be used for characterization of models before performing single-cell assays. In addition, the use of flow cytometry or microscopy provides a possibility to assess reproducibility.

The transcriptional landscape of biological samples from donors, patients and models can highlight important aspects of disease. However, creating an interpretable view of this landscape is an ongoing challenge.

3.2 Genomic Mapmaking

Cartographers have created maps as graphic representations of parts of the Earth to communicate routes, locations, and hazards which played an important role in discovery and exploration [106]. Maps face the dilemma of projection of multi-dimensional data, such as the three dimensional irregular ellipsoid shape of earth, into a flat, two-dimensional surface. The systematic method of projecting Earth's latitude and longitude onto a plane creates inevitable distortions [107]. Use cases of the map, therefore, depend on context and need to be aware of these distortions. For example, using the Mercator projection for navigation charts works well due to the conservation of angles while judging areas creates misconceptions [108, 109].

Projection of transcriptomes in two dimensions is commonly performed using PCA and UMAP. While the input data can be derived from bulk or single-cell transcriptomes based on all or just a subset of genes, the overall goal is improved interpretability of the similarities and differences between samples. Some scientists fear that distortions of local and global structure may lead to contradictory interpretations [110] while others show that highlighting those distortions can reveal hidden characteristics of the data [111]. Undisputed, however, is the widespread adoption of these maps in the field of single-cell genomics.

Maps or embeddings of transcriptomes can be created for a single experiment and give a sense of distance between samples. Inclusion of more samples, conditions and controls makes the analysis more complex but often adds essential information. Designing maps with a clear landscape facilitates interpretation and usability. Therefore, comprehensive data sets need to include all necessary populations to clearly identify population shifts and gene expression changes associated with a condition.

3.2.1 Reference Atlases

Reference atlases represent integrated collections of data sets, typically of one organ, across multiple anatomical locations, individuals, technologies and centers. In addition to health, they should capture normal aging, exposures due to lifestyle and environment, genetic backgrounds and diseases [104].

Comprehensive references of the cellular landscape of multiple organs of humans [112] and mice [113] have been built over the last years in addition to countless studies investigating specific diseases in particular organs. Consortia such as the Human Cell Atlas and LifeTime Initiative are emerging that aim to define the cellular landscape of the human body in health and disease [114, 115]. At least two such atlases already exists for the human airway [26, 104] and the human gastrointestinal (GI) tract [34, 116]. In the future, the range of atlases for organs and tissues in health and disease will continue to expand and will have broad impact on the field of medicine and biomedical research [117].

Making use of atlases requires clear goals for the analysis and selection of a suitable reference. However, the choice of data sets can be difficult and involve curation of atlas-level data to design a suitable reference.

3.2.2 Reference Usage

References can be used for annotation, mapping or integration of a query. For annotation, cell type labels are transferred from the reference. Mapping data to a reference can position cells in the already existing embedding without re-computing the entire map. Most accurate but also most computationally expensive is *de novo* integration which calculates a new embedding and can transfer annotations by transferring majority labels of reference cells to clusters. Multiple methods exist that perform either one or multiple of these tasks. While SingleR only transfers annotations [70], Seurat version 4 [38] and scArches [118] are able to transfer annotations and perform reference mapping. Integration relies on the multitude of methods developed for correction of scRNA-seq data including canonical correlation analysis (CCA), mutual nearest neighbors (MNN), and single-cell variational inference (scVI) [63, 67, 119] which have been benchmarked for different applications [68].

Accuracy in the annotation of query data sets depends strongly on the reference design, referring to the selection of data sets included in the reference. Cell types that are less abundant or more related are predicted less accurately. However, prediction can be improved by adjusting the reference design to be more balanced compared to the composition of the query. The balance between reference and query can focus on cell types but also sex, ethnicity or other factors that impact cellular phenotypes. This is especially important since the confidence scores reported by the methods are also negatively impacted by the unbalanced design [120]. Harmonization of labels is part of atlas building and typically involves multiple levels of annotation [104]. However, the query structure might need different granularity of annotations than are present in the reference and have to be adjusted before transfer. For example, the presence of few epithelial cells in the presence of ambient RNA contamination might result in prediction errors when using very fine labels but higher accuracy on coarse labels when focusing on leukocyte populations in the lung. I intend to create a harmonized data set for the integration of clinical samples of BAL fluid to further study the cellular phenotypes associated to the outcomes of COVID-19. In the context of organoid models, cell ontology provides a conceptual framework that includes culture artifacts and distinguishes naive and cultured cells [121]. In analogy, atlases incorporating native and cultured tissues might be a requirement for reliable analysis of infection models from advanced tissue models.

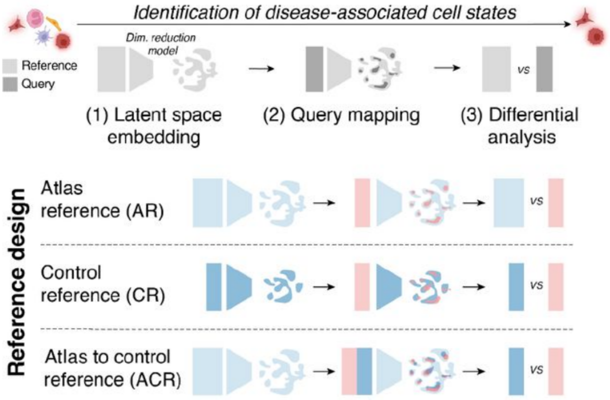

Identification of disease-associated cell states
Dim. reduction model
Reference
Query
(1) Latent space embedding
(2) Query mapping
(3) Differential analysis
vs
Reference design
Atlas reference (AR)
Control reference (CR)
Atlas to control reference (ACR)
vs
vs
vs

Figure 3.4: Reference design improves identification of disease states. Adopted from [122].

Integration of multiple query data sets with a reference can be useful but computationally demanding because of hardware limitations and run times. Reference mapping is more computationally efficient while having good accuracy on balanced references [118]. Matched samples from control and disease samples can be mapped together against a reference atlas. While the reference atlas does not replace the control samples, it increases sensitivity and precision of disease-state discovery and can reduce the number of control samples necessary to obtain sufficient statistical power (see figure 3.4) [122]

3.2.3 Technological Advancements

Representations of biological systems through maps derived from sequencing data are immensely useful. However, novel technologies help to further improve those representations by either connecting them to the spatial context of tissues and organs, adding temporal resolution to disease dynamics or increasing resolution of RNA structure in terms of differentially spliced isoforms or modifications.

Tissue Space

Microscopy can reveal host cells and their intracellular structures as well as bacteria, fungi, and viruses [7]. Connecting transcriptome-wide studies of cellular phenotypes with the tissue context capture by microscopy has been made possible through breakthroughs of both microscopy and sequencing techniques [41, 123, 124].

Integration between single-cell and spatial transcriptomics requires computational methods similar to those used in reference integration. However, those data show even larger differences between technologies. Frameworks for spatial data processing have been developed that tackle the multitude of problems from cell segmentation to intercellular communication(spatialdata.scverse.org, [125, 126]).

A spatially resolved atlas of the lung has already been published, but standards for the use of spatial reference atlases, especially for infection research, are still needed [127].

Disease Time

Critical aspects of disease development can occur at small timescales. Metabolic labelling can connect transcriptomic phenotypes across time points of multiple hours. Comparing RNA that has been synthesized before (old) and during (new) a labelling time with nucleoside analogs, that introduce specific mismatches, can connect transcriptional states and reveal temporal disease dynamics [129].

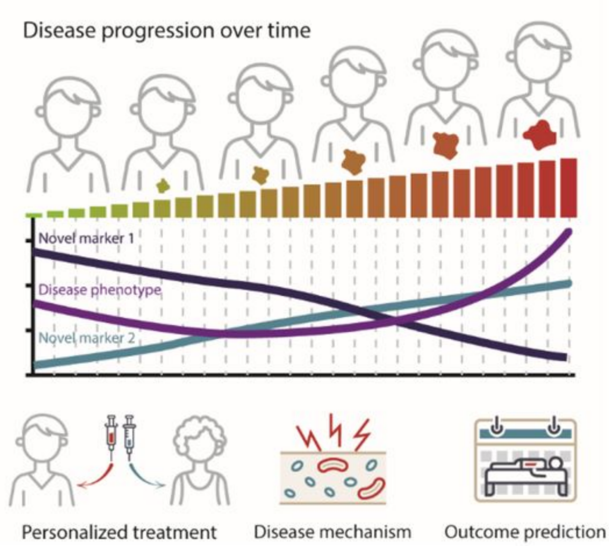


Figure 3.5: Disease dynamics can be discovered across patients and improve molecular interpretation, patient stratification, and outcome prediction. Adopted from [128].

Disease progression in patients is often assessed relative to symptom onset or hospital admission but well characterized disease stages might be a more relevant classification. Disease progression dynamics are important for prognosis and treatment and can possibly be inferred from high-dimensional short time-series data (see figure 3.5). While the requirement for multiple longitudinal samples from multiple individuals can be difficult to meet, it has the potential to highlight diverging disease dynamics and capture novel features of the pathomechanism [128].

Isoform Discovery

Long read sequencing has already revolutionized genome assembly [130] and has been applied for single-cell genomics [131]. It will continue to improve resolution from genes to transcript isoforms and therefore continue to revolutionize scientific research, for example by revealing receptor profiles of lymphocytes [132].

3.3 Precision Medicine

Defining groups of patients with similar disease characteristics that benefit from the same treatments is the principle of precision medicine. Diseases can have distinguishable phenotypes, also called endotypes, associated to differences in progression and outcomes. Molecular and cellular descriptions of disease phenotypes can reveal biomarkers that make the stratification of patients possible [14]

Cancers were long considered shared disease entities based on the affected organ. Identification of molecular mechanisms that determine cancer growth and metastasis has revealed subtypes that differ in clinical outcomes and benefit from different therapies [13]. In analogy, distinguishing disease phenotypes in acute respiratory failure is advancing based on molecular features involved in inflammation [133].

The search for biomarkers therefore depends on the resolution and accuracy of our descriptions of disease phenotypes.

3.3.1 Defining Disease Phenotypes

Disease phenotypes that are distinguishable in progression and outcome must differ in certain features of the pathomechanism. Single-cell sequencing improves detection of these features by revealing effects limited to a small part of the population. However, once a specific effect has been understood, it can be measured across patients to distinguish disease phenotypes and guide treatment.

Cell-based immunotherapy is targeted at a single epitope that, ideally, marks only one cell type in the human body. We were able to distinguish malignant plasma cells in the bone marrow of a patient with relapsed/refractory Multiple Myeloma (RRMM). Single-cell sequencing enabled us to assess expression of the *BCMA* gene within this cell type. Further investigation of the aberrant plasma cells revealed genomic deletion of *TNFRSF17* as the mechanism of antigen loss. Identification of the mechanism led to the assessment of heterogeneity at the genomic locus, which exists in the patient population and represents a potential risk factor for relapse. Genomic screening of cancer patients can therefore lead to patient stratification into subgroups with different treatment response.

Acute respiratory distress can be associated with a hyperinflammatory phenotype or non-hyperinflammatory phenotype that respond different to positive end-expiratory pressures (PEEP). However, translating phenotypic findings into clinical care is impeded by the limited availability of biomarker detection in the clinical setting and prospective validation [133]. It is possible that higher resolution descriptions of ARDS with single-cell sequencing can reveal features of the pathomechanism. Therefore, identification of the subpopulations and effects that drive the hyperinflammatory phenotype might lead to improvements in patient stratification and treatment.

Pulmonary fibrosis arises from abnormalities in the wound healing response. Currently, the approved therapies act broadly on multiple disease pathways and targeted therapies have not proven effective. Still, a personalized approach to IPF therapy is anticipated for the future [134]. The development of lung fibrosis as a consequence of acute respiratory distress in severe COVID-19 highlights the ill-understood interplay between diseases. Studying the connection between ARDS in severe COVID-19, the development of lung fibrosis and the outcomes of disease might be able to help disentangle the disease phenotypes and reveal distinguishing features.

3.3.2 Predicting Disease Outcomes

Accurate descriptions of disease progression make it possible to predict outcomes. Particular focus is needed on the decision points that determine different outcomes. Cellular trajectories between health and disease can reveal those points and highlight potential for interception (see figure 3.6). However, such descriptions require systematic assessment of the phenotypic landscape of disease.

Predicting treatment response becomes increasingly important in the field of cancer therapy. Relapse is frequent and multiple lines of treatment are used to increase survival [135]. Meanwhile, the number of treatments is increasing and immunotherapies are moving to the first line of therapy [136]. We were able to reveal heterozygosity or monosomy at the *TNFRSF17* locus as a potential risk factor for relapse and investigate the heterogeneity of tumor cells after relapse. The expression of genes such as *CD38* and *SLAMF7* indicates the potential to target the epitopes with other immunotherapies. However, how the cellular phenotypes relate to treatment outcome is difficult to assess. Identification of resistance pathways through single-cell sequencing will help to better define the treatment landscape and highlight therapeutic targets [137].

The presence of profibrotic macrophages in severe COVID-19 is associated to lung fibrosis. Therefore, the transcriptional phenotype is connected to a certain disease state but association to mortality has not been assessed. CMKLR1 is a precise marker for the profibrotic macrophages and has been detected by differential expression in our data set [93]. Recently, positron emission tomography (PET) of CMKLR1 has been developed that detects pulmonary inflammation during experimental acute lung injury in mice. CMKLR1-targeted PET shows potential to be used in the future to monitor the dynamics of lung inflammation and treatment response [138].

Host factors that are so far often not measured might also play an important role. Microbiota can exert both protective and compromising functions in the development of pneumonia [139]. Predictive models of diseases should consider the potential of missing data and uncertainties.

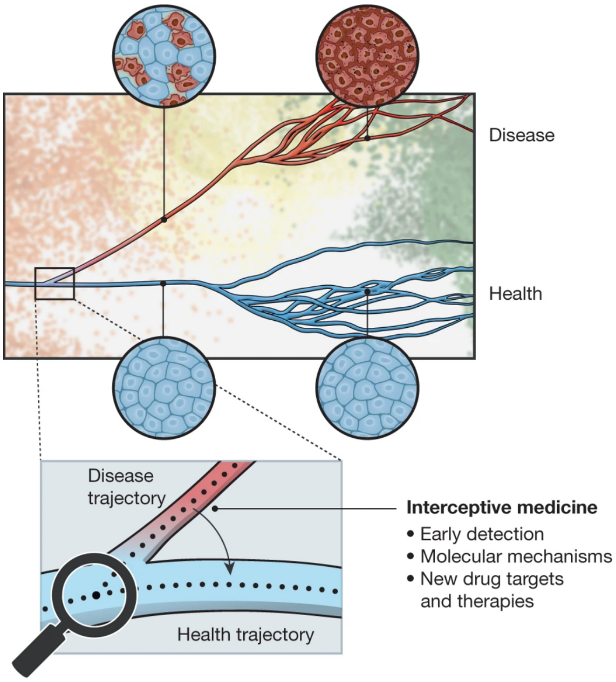


Figure 3.6: Understanding cellular trajectories facilitates early disease detection. Adopted from [140].

3.3.3 Targeting Treatments

A global landscape of disease progression and treatment response may be able to guide the selection of therapies for individual patients and the development of novel therapies. Single-cell multi-omics does provide the necessary resolution while artificial intelligence facilitates data integration and predictive modeling. Advanced model systems should be used to test treatments and investigate mechanisms [140]. However, such complex data needs to stay interpretable in order to guide clinical decisions. Comprehensive presentation as well as intuitive exploration might be required to drive the adoption of those modern technologies.

3.4 Visualization & Design

Data generation using single-cell and spatial genomics is increasing with many studies building and describing references that are waiting for applications [141]. Data is often deposited to archives that store hundreds of terabyte of data (nemoarchive.org). Interactive web applications are developed and deployed for those data sets and facilitate exploration of gene expression in the tissue space (braincelldata.org). Some applications are centralized and provide interactive exploration of curated data such as Nextstrain for tracking pathogen evolution (nextstrain.org). Others are build for specific data structures but generalize to any data set such as the single-cell browsers CellxGene (cellxgene.cziscience.com) and Interactive SummarizedExperiment Explorer (iSEE) [142].

Data presentation is used to convey findings from exploratory analyses in a coherent structure. While manuscripts and books use static images, web applications can introduce interactivity, which adds features of exploration to the presented data. Common frameworks that add interactivity to the programming environments of R and python are shiny and plotly [143, 144].

Other forms of interactivity can come in the form of virtual reality (VR) and augmented reality (AR, see figure 3.7). While those technologies can create more immersive experiences and be added to common formats such as static text, they can be distracting and their ideal use case will have to be determined [145].

Creating visualizations that are intuitive and easy to interpret is difficult and requires knowledge of design principles that influence readers and will become more important in the future [146, 147]. To this end, mockups help to review format, layout, and content, which improves the quality of the finished product and fuels progress.

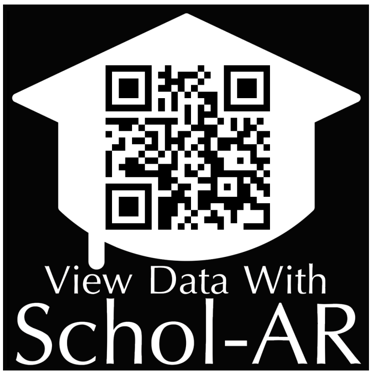


Figure 3.7: Augmented reality brings interactivity to static documents. Scanning this QR code via the Schol-AR application enables interactive features of this document. For example a video of the UMAP of macrophages in severe COVID-19 in Wendisch *et al.* 2021, figure 2B.

.

3.4.1 Infection Atlas

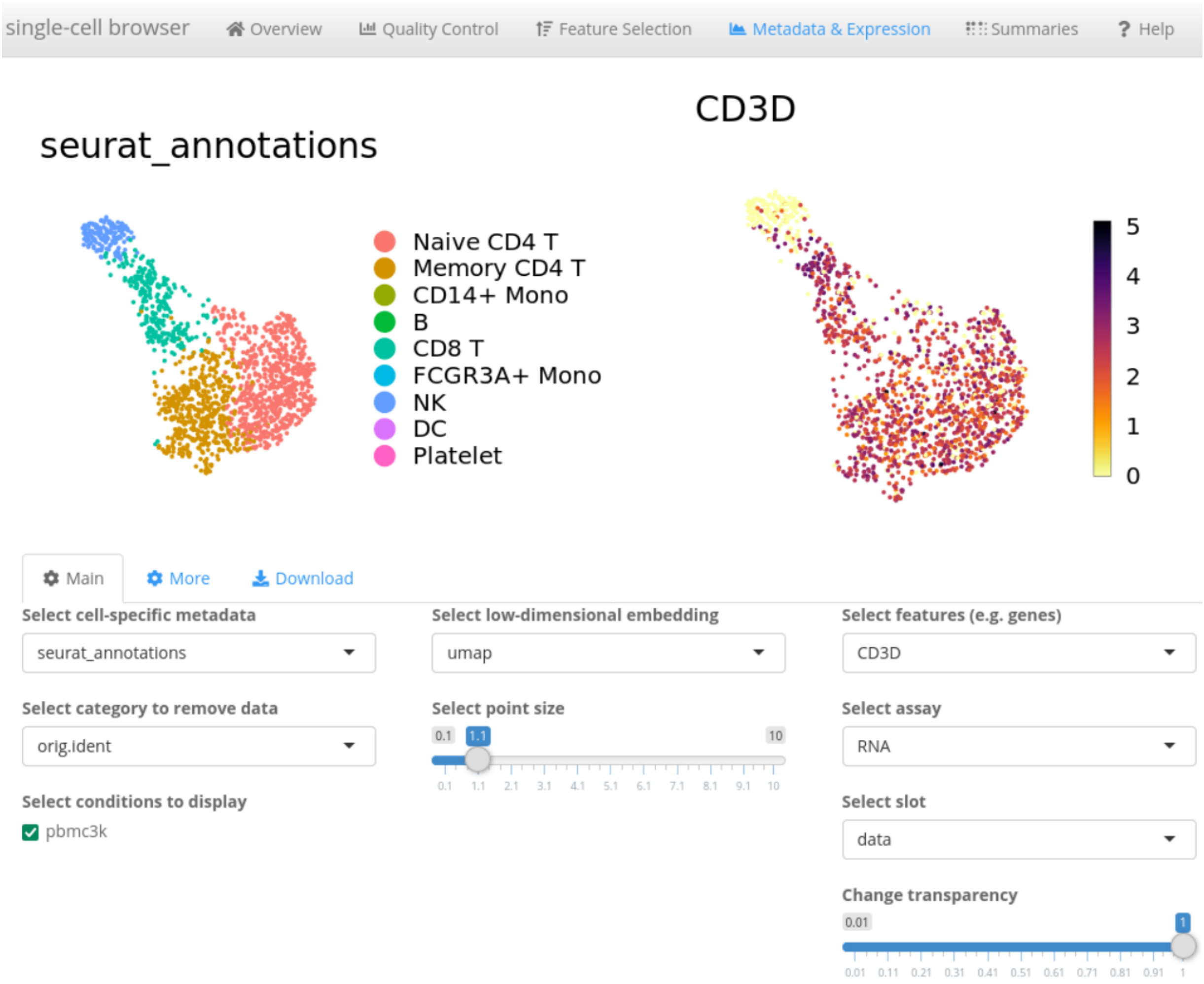
Web interfaces that facilitate the exploration of organ atlases are important resources for the research community that facilitate exploration as well as data accessibility. We have built the Infection Atlas as a community resource that gives accessibility of exploration to the researchers that were involved in creation of the data (see figure 3.8).

Over time, the tool evolved and data dissemination and presentation to outside collaborators gained importance. Currently, challenges lie in the performance of the web applications and interactivity of visualizations but also in the design of data sets and usability of resources across the community. However, the prospects to build well-integrated data sets for use in collaborative projects such as the CRC DECIDE justify those efforts.

3.5 Solving Clinical Problems

I highlighted the possibility of collecting tissue samples from different sources and employ multiple technologies to obtain measurements of genomic features in single cells. Advances in bioinformatics facilitate the analysis of large numbers of data sets together and remove technical variation computationally while retaining interesting biological signals. The precise measurement of phenotypes and their role in disease development should be entering clinical practice to improve the precision with which diseases are diagnosed and treated.

In the future, I want to combine data from clinical samples and advanced tissue culture to identify features of diseases that are associated to different outcomes and investigate responses to treatment. I will continue to improve our computational workflows to better evaluate the analysis output and work on the design of data visualization.


single-cell browser
Overview
Quality Control
Feature Selection
Metadata & Expression
Summaries
Help
seurat_annotations
Naive CD4 T
Memory CD4 T
CD14+ Mono
B
CD8 T
FCGR3A+ Mono
NK
DC
Platelet
CD3D
5
4
3
2
1
0
Main
More
Download

Select cell-specific metadata: seurat_annotations

Select category to remove data: orig.ident

Select conditions to display: ☑ pbmc3k

Select low-dimensional embedding: umap

Select point size: 1.1

Select features (e.g. genes): CD3D

Select assay: RNA

Select slot: data

Change transparency: 1

0.1
10
0.1 1.1 2.1 3.1 4.1 5.1 6.1 7.1 8.1 9.1 10
0.01
0.01 0.11 0.21 0.31 0.41 0.51 0.61 0.71 0.81 0.91 1

Figure 3.8: Infection Atlas. A web portal offers access to explore and analyze scRNA-seq data. Available at `https://www.infection-atlas.org/cell.browser/`

Bibliography

[1] Ron Sender, Shai Fuchs, and Ron Milo. "Revised Estimates for the Number of Human and Bacteria Cells in the Body". In: *PLOS Biology* 14.8 (Aug. 2016), pp. 1–14. DOI: `10.1371/journal.pbio.1002533`.

[2] National Human Genome Research Institute. *NIH Human Microbiome Project defines normal bacterial makeup of the body*. `https://www.genome.gov/27549144/2012-release-nih-human-microbiome-project-defines-normal-bacterial-makeup-of-the-body`. Accessed: 2023-04-13. 2012.

[3] Abigail Bartlett et al. "A comprehensive list of bacterial pathogens infecting humans". In: *Microbiology* 168.12, 001269 (2022). ISSN: 1465-2080. DOI: `https://doi.org/10.1099/mic.0.001269`.

[4] WHO. *Our work*. `https://www.who.int/our-work`. Accessed: 2023-04-13.

[5] National Institutes of Health. *Impact of NIH Research*. `https://www.nih.gov/about-nih/what-we-do/impact-nih-research`. Accessed: 2023-04-13.

[6] World Health Organization. *Coronavirus disease (COVID-19) pandemic*. `https://www.who.int/emergencies/diseases/novel-coronavirus-2019`. Accessed: 2023-04-13.

[7] Herbert Hof et al. *Medizinische Mikrobiologie*. 8th ed. Thieme, 2022. ISBN: 978-3-13-241038-1.

[8] Daniela Coclite et al. "Face Mask Use in the Community for Reducing the Spread of COVID-19: A Systematic Review". In: *Frontiers in Medicine* 7 (2021). ISSN: 2296-858X. DOI: `10.3389/fmed.2020.594269`.

[9] John S Tregoning et al. "Progress of the COVID-19 vaccine effort: viruses, vaccines and variants versus efficacy, effectiveness and escape". In: *Nature Reviews Immunology* 21.10 (Oct. 2021), pp. 626–636.

[10] The Lancet. "India under COVID-19 lockdown". In: *The Lancet* 395.10233 (Apr. 2020), p. 1315. ISSN: 0140-6736. DOI: `10.1016/S0140-6736(20)30938-7`.

[11] Rosanna W Peeling et al. "Diagnostics for COVID-19: moving from pandemic response to control". In: *The Lancet* 399.10326 (Feb. 2022), pp. 757–768.

[12] Joydeb Majumder and Tamara Minko. "Recent Developments on Therapeutic and Diagnostic Approaches for COVID-19". In: *The AAPS Journal* 23.1 (Jan. 2021), p. 14. ISSN: 1550-7416. DOI: `10.1208/s12248-020-00532-2`.

[13] Robert A. Weinberg. *The Biology of Cancer*. 2nd ed. W. W. Norton & Company, 2013. ISBN: 978-0815342205.

[14] Inke R. König et al. "What is precision medicine?" In: *European Respiratory Journal* 50.4 (2017). ISSN: 0903-1936. DOI: `10.1183/13993003.00391-2017`.

[15] World Health Organization. *One Health*. `https://www.who.int/health-topics/one-health`. Accessed: 2023-04-18.

[16] David M. Morens and Anthony S. Fauci. "Emerging Pandemic Diseases: How We Got to COVID-19". In: *Cell* 182.5 (Sept. 2020), pp. 1077–1092. ISSN: 0092-8674. DOI: `10.1016/j.cell.2020.08.021`.

[17] Alexander E. Gorbalenya et al. "The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2". In: *Nature Microbiology* 5.4 (Apr. 2020), pp. 536–544. ISSN: 2058-5276. DOI: `10.1038/s41564-020-0695-z`.

[18] Ben Hu et al. "Characteristics of SARS-CoV-2 and COVID-19". In: *Nature Reviews Microbiology* 19.3 (Mar. 2021), pp. 141–154. ISSN: 1740-1534. DOI: `10.1038/s41579-020-00459-7`.

[19] Jorge E. Galán. "Salmonella Typhimurium and inflammation: a pathogen-centric affair". In: *Nature Reviews Microbiology* 19.11 (Nov. 2021), pp. 716–725. ISSN: 1740-1534. DOI: `10.1038/s41579-021-00561-4`.

[20] Douglas Hanahan and Robert A. Weinberg. "The Hallmarks of Cancer". In: *Cell* 100.1 (Jan. 2000), pp. 57–70. ISSN: 0092-8674. DOI: `10.1016/S0092-8674(00)81683-9`.

[21] Douglas Hanahan and Robert A. Weinberg. "Hallmarks of Cancer: The Next Generation". In: *Cell* 144.5 (Mar. 2011), pp. 646–674. ISSN: 0092-8674. DOI: `10.1016/j.cell.2011.02.013`.

[22] S. Vincent Rajkumar. "Multiple myeloma: Every year a new standard?" In: *Hematological Oncology* 37.S1 (2019), pp. 62–65. DOI: `https://doi.org/10.1002/hon.2586`.

[23] Abul K. Abbas et al. *Cellular and Molecular Immunology*. 9th ed. Elsevier, 2018. ISBN: 978-0-323-47978-3.

[24] Sarah A. Horst et al. "A Novel Mouse Model of Staphylococcus aureus Chronic Osteomyelitis That Closely Mimics the Human Infection: An Integrated View of Disease Pathogenesis". In: *The American Journal of Pathology* 181.4 (Oct. 2012), pp. 1206–1214. ISSN: 0002-9440. DOI: `10.1016/j.ajpath.2012.07.005`.

[25] Carmen Aguilar et al. "Organoids as host models for infection biology – a review of methods". In: *Experimental & Molecular Medicine* 53.10 (Oct. 2021), pp. 1471–1482. ISSN: 2092-6413. DOI: `10.1038/s12276-021-00629-4`.

[26] Kyle J. Travaglini et al. "A molecular cell atlas of the human lung from single-cell RNA sequencing". In: *Nature* 587.7835 (Nov. 2020), pp. 619–625. ISSN: 1476-4687. DOI: `10.1038/s41586-020-2922-4`.

[27] Toshiro Sato et al. "Long-term Expansion of Epithelial Organoids From Human Colon, Adenoma, Adenocarcinoma, and Barrett's Epithelium". In: *Gastroenterology* 141.5 (2011), pp. 1762–1772. ISSN: 0016-5085. DOI: `https://doi.org/10.1053/j.gastro.2011.07.050`.

[28] Vera Raivola et al. "Blood donors' preferences for blood donation for biomedical research". In: *Transfusion* 58.7 (2018), pp. 1640–1646. DOI: `https://doi.org/10.1111/trf.14596`.

[29] Gustaf Wigerblad et al. "Single-Cell Analysis Reveals the Range of Transcriptional States of Circulating Human Neutrophils". In: *The Journal of Immunology* 209.4 (Aug. 2022), pp. 772–782. ISSN: 0022-1767. DOI: `10.4049/jimmunol.2200154`.

[30] Hans Clevers. "The Intestinal Crypt, A Prototype Stem Cell Compartment". In: *Cell* 154.2 (July 2013), pp. 274–284. ISSN: 0092-8674. DOI: `10.1016/j.cell.2013.07.004`.

[31] Jason R. Spence et al. "Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro". In: *Nature* 470.7332 (Feb. 2011), pp. 105–109. ISSN: 1476-4687. DOI: `10.1038/nature09691`.

[32] Steve Perrin. "Preclinical research: Make mouse studies work". In: *Nature* 507.7493 (Mar. 2014), pp. 423–425. ISSN: 1476-4687. DOI: `10.1038/507423a`.

[33] Toshiro Sato et al. "Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche". In: *Nature* 459.7244 (May 2009), pp. 262–265. ISSN: 1476-4687. DOI: `10.1038/nature07935`.

[34] Qianhui Yu et al. "Charting human development using a multi-endodermal organ atlas and organoid models". In: *Cell* 184.12 (June 2021), 3281–3298.e22. ISSN: 0092-8674. DOI: `10.1016/j.cell.2021.04.028`.

[35] Masayuki Fujii et al. "Human Intestinal Organoids Maintain Self-Renewal Capacity and Cellular Diversity in Niche-Inspired Culture Condition". In: *Cell Stem Cell* 23.6 (Dec. 2018), 787–793.e6. ISSN: 1934-5909. DOI: `10.1016/j.stem.2018.11.016`.

[36] Umut Kilik et al. "Maturation of human intestinal epithelium from pluripotency in vitro". In: *bioRxiv* (2021). DOI: `10.1101/2021.09.24.460132`.

[37] Valentine Svensson, Roser Vento-Tormo, and Sarah A. Teichmann. "Exponential scaling of single-cell RNA-seq in the past decade". In: *Nature Protocols* 13.4 (Apr. 2018), pp. 599–604. ISSN: 1750-2799. DOI: `10.1038/nprot.2017.149`.

[38] Yuhan Hao et al. "Integrated analysis of multimodal single-cell data". In: *Cell* 184.13 (June 2021), 3573–3587.e29. ISSN: 0092-8674. DOI: `10.1016/j.cell.2021.04.048`.

[39] Sarah Aldridge and Sarah A. Teichmann. "Single cell transcriptomics comes of age". In: *Nature Communications* 11.1 (Aug. 2020), p. 4307. ISSN: 2041-1723. DOI: `10.1038/s41467-020-18158-5`.

[40] Ana Teresa López-Jiménez and Serge Mostowy. "Emerging technologies and infection models in cellular microbiology". In: *Nature Communications* 12.1 (Nov. 2021), p. 6764. ISSN: 2041-1723. DOI: `10.1038/s41467-021-26641-w`.

[41] Chee-Huat Linus Eng et al. "Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+". In: *Nature* 568.7751 (Apr. 2019), pp. 235–239. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1049-y.

[42] Chenglong Xia et al. "Multiplexed detection of RNA using MERFISH and branched DNA amplification". In: *Scientific Reports* 9.1 (May 2019), p. 7721. ISSN: 2045-2322. DOI: 10.1038/s41598-019-43943-8.

[43] Eija Korpelainen et al. *RNA-seq Data Analysis*. Taylor Francis, 2015. ISBN: 978-1-4665-9500-2.

[44] Alexander J. Westermann, Stanislaw A. Gorski, and Jörg Vogel. "Dual RNA-seq of pathogen and host". In: *Nature Reviews Microbiology* 10.9 (Sept. 2012), pp. 618–630. ISSN: 1740-1534. DOI: 10.1038/nrmicro2852.

[45] Christoph Ziegenhain et al. "Comparative Analysis of Single-Cell RNA Sequencing Methods". In: *Molecular Cell* 65.4 (2017), 631–643.e4. ISSN: 1097-2765. DOI: https://doi.org/10.1016/j.molcel.2017.01.023.

[46] Simone Picelli et al. "Full-length RNA-seq from single cells using Smart-seq2". In: *Nature Protocols* 9.1 (Jan. 2014), pp. 171–181. ISSN: 1750-2799. DOI: 10.1038/nprot.2014.006.

[47] Grace X. Y. Zheng et al. "Massively parallel digital transcriptional profiling of single cells". In: *Nature Communications* 8.1 (Jan. 2017), p. 14049. ISSN: 2041-1723. DOI: 10.1038/ncomms14049.

[48] Elisabetta Mereu et al. "Benchmarking single-cell RNA-sequencing protocols for cell atlas projects". In: *Nature Biotechnology* 38.6 (June 2020), pp. 747–755. ISSN: 1546-1696. DOI: 10.1038/s41587-020-0469-4.

[49] Marlon Stoeckius et al. "Simultaneous epitope and transcriptome measurement in single cells". In: *Nature Methods* 14.9 (Sept. 2017), pp. 865–868. ISSN: 1548-7105. DOI: 10.1038/nmeth.4380.

[50] Marlon Stoeckius et al. "Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics". In: *Genome Biology* 19.1 (Dec. 2018), p. 224. ISSN: 1474-760X. DOI: 10.1186/s13059-018-1603-1.

[51] Vladimir Yu Kiselev, Tallulah S. Andrews, and Martin Hemberg. "Challenges in unsupervised clustering of single-cell RNA-seq data". In: *Nature Reviews Genetics* 20.5 (May 2019), pp. 273–282. ISSN: 1471-0064. DOI: 10.1038/s41576-018-0088-9.

[52] Malte D. Luecken and Fabian J. Theis. "Current best practices in single-cell RNA-seq analysis: a tutorial". In: *Molecular Systems Biology* 15 (6 2019). DOI: 10.15252/msb.20188746.

[53] 10x Genomics. *What is Cell Ranger*. https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger. Accessed: 2023-04-25.

[54] Benjamin Kaminow, Dinar Yunusov, and Alexander Dobin. "STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data". In: *bioRxiv* (2021). DOI: 10.1101/2021.05.05.442755.

[55] Avi Srivastava et al. "Alevin efficiently estimates accurate gene abundances from dscRNA-seq data". In: *Genome Biology* 20.1 (Mar. 2019), p. 65. ISSN: 1474-760X. DOI: 10.1186/s13059-019-1670-y.

[56] Aaron T. L. Lun et al. "EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data". In: *Genome Biology* 20.1 (Mar. 2019), p. 63. ISSN: 1474-760X. DOI: 10.1186/s13059-019-1662-y.

[57] ATL Lun, DJ McCarthy, and JC Marioni. "A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor [version 2; peer review: 3 approved, 2 approved with reservations]". In: *F1000Research* 5.2122 (2016). DOI: 10.12688/f1000research.9501.2.

[58] Ayshwarya Subramanian et al. "Biology-inspired data-driven quality control for scientific discovery in single-cell transcriptomics". In: *Genome Biology* 23.1 (Dec. 2022), p. 267. ISSN: 1474-760X. DOI: 10.1186/s13059-022-02820-w.

[59] Christopher S. McGinnis, Lyndsay M. Murrow, and Zev J. Gartner. "DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors". In: *Cell Systems*

8.4 (Apr. 2019), 329–337.e4. ISSN: 2405-4712. DOI: 10.1016/j.cels.2019.03.003.

[60] Simon Anders and Wolfgang Huber. "Differential expression analysis for sequence count data". In: *Genome Biology* 11.10 (Oct. 2010), R106. ISSN: 1474-760X. DOI: 10.1186/gb-2010-11-10-r106.

[61] Christoph Hafemeister and Rahul Satija. "Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression". In: *Genome Biology* 20.1 (Dec. 2019), p. 296. ISSN: 1474-760X. DOI: 10.1186/s13059-019-1874-1.

[62] Constantin Ahlmann-Eltze and Wolfgang Huber. "Comparison of transformations for single-cell RNA-seq data". In: *Nature Methods* (Apr. 2023). ISSN: 1548-7105. DOI: 10.1038/s41592-023-01814-1.

[63] Tim Stuart and Rahul Satija. "Integrative single-cell analysis". In: *Nature Reviews Genetics* 20 (2019), pp. 257–272. DOI: 10.1038/s41576-019-0093-7.

[64] Laurens van der Maaten and Geoffrey Hinton. "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research* 9 (2008), pp. 2579–2605.

[65] Leland McInnes, John Healy, and James Melville. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction". In: *arXiv* (2018).

[66] Matthew E. Ritchie et al. "limma powers differential expression analyses for RNA-sequencing and microarray studies". In: *Nucleic Acids Research* 43.7 (Jan. 2015), e47–e47. ISSN: 0305-1048. DOI: 10.1093/nar/gkv007.

[67] Laleh Haghverdi et al. "Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors". In: *Nature Biotechnology* 36.5 (May 2018), pp. 421–427. ISSN: 1546-1696. DOI: 10.1038/nbt.4091.

[68] Malte D. Luecken et al. "Benchmarking atlas-level data integration in single-cell genomics". In: *Nature Methods* 19.1 (Jan. 2022), pp. 41–50. ISSN: 1548-7105. DOI: 10.1038/s41592-021-01336-8.

[69] Allen W. Zhang et al. "Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling". In: *Nature Methods* 16.10 (Oct. 2019), pp. 1007–1015. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0529-1.

[70] Dvir Aran et al. "Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage". In: *Nature Immunology* 20.2 (Feb. 2019), pp. 163–172. ISSN: 1529-2916. DOI: 10.1038/s41590-018-0276-y.

[71] Zoe A. Clarke et al. "Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods". In: *Nature Protocols* 16.6 (June 2021), pp. 2749–2764. ISSN: 1750-2799. DOI: 10.1038/s41596-021-00534-0.

[72] Samuele Mazzanti. *Are You Still Using the Elbow Method?* Feb. 2023. URL: https://towardsdatascience.com/are-you-still-using-the-elbow-method-5d271b3063bd.

[73] Isabella N. Grabski, Kelly Street, and Rafael A. Irizarry. "Significance Analysis for Clustering with Single-Cell RNA-Sequencing Data". In: *bioRxiv* (2022). DOI: 10.1101/2022.08.01.502383.

[74] Luke Zappia and Alicia Oshlack. "Clustering trees: a visualization for evaluating clusterings at multiple resolutions". In: *GigaScience* 7.7 (July 2018). giy083. ISSN: 2047-217X. DOI: 10.1093/gigascience/giy083.

[75] David Freedman, Robert Pisani, and Roger Purves. *Statistics*. 4th ed. W. W. Norton & Company, 2007. ISBN: 0-393-93043-2.

[76] Michael I. Love, Wolfgang Huber, and Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biology* 15.12 (Dec. 2014), p. 550. ISSN: 1474-760X. DOI: 10.1186/s13059-014-0550-8.

[77] Davis McCarthy and Gordon K. Smyth. "Testing significance relative to a fold-change threshold is a TREAT". In: *Bioinformatics* 25 (6 2009), pp. 765–771. DOI: 10.1093/bioinformatics/btp053.

[78] Charlotte Soneson and Mark D. Robinson. "Bias, robustness and scalability in single-cell differential expression analysis". In: *Nature Methods* 15.4 (Apr. 2018), pp. 255–261. ISSN: 1548-7105. DOI: 10.1038/nmeth.4612.

[79] Jordan W. Squair et al. "Confronting false discoveries in single-cell differential expression". In: *Nature Communications* 12.1 (Sept. 2021), p. 5692. ISSN: 2041-1723. DOI: 10.1038/s41467-021-25960-2.

[80] Kip D. Zimmerman, Mark A. Espeland, and Carl D. Langefeld. "A practical solution to pseudoreplication bias in single-cell studies". In: *Nature Communications* 12.1 (Feb. 2021), p. 738. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21038-1.

[81] Alan E. Murphy and Nathan G. Skene. "A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis". In: *Nature Communications* 13.1 (Dec. 2022), p. 7851. ISSN: 2041-1723. DOI: 10.1038/s41467-022-35519-4.

[82] Michael Ashburner et al. "Gene Ontology: tool for the unification of biology". In: *Nature Genetics* 25.1 (May 2000), pp. 25–29. ISSN: 1546-1718. DOI: 10.1038/75556.

[83] The Gene Ontology Consortium. "The Gene Ontology resource: enriching a GOld mine". In: *Nucleic Acids Research* 49.D1 (Dec. 2020), pp. D325–D334. ISSN: 0305-1048. DOI: 10.1093/nar/gkaa1113.

[84] Guangchuang Yu et al. "clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters". In: *OMICS: A Journal of Integrative Biology* 16.5 (2012). PMID: 22455463, pp. 284–287. DOI: 10.1089/omi.2011.0118.

[85] Aravind Subramanian et al. "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles". In: *Proceedings of the National Academy of Sciences* 102.43 (2005), pp. 15545–15550. DOI: 10.1073/pnas.0506580102.

[86] Tianzhi Wu et al. "clusterProfiler 4.0: A universal enrichment tool for interpreting omics data". In: *The Innovation* 2.3 (Aug. 2021). ISSN: 2666-6758. DOI: 10.1016/j.xinn.2021.100141.

[87] Sara Aibar et al. "SCENIC: single-cell regulatory network inference and clustering". In: *Nature Methods* 14.11 (Nov. 2017), pp. 1083–1086. ISSN: 1548-7105. DOI: 10.1038/nmeth.4463.

[88] Alexandra B Keenan et al. "ChEA3: transcription factor enrichment analysis by orthogonal omics integration". In: *Nucleic Acids Research* 47.W1 (May 2019), W212–W224. ISSN: 0305-1048. DOI: 10.1093/nar/gkz446.

[89] Emma Dann et al. "Differential abundance testing on single-cell data using k-nearest neighbor graphs". In: *Nature Biotechnology* 40.2 (Feb. 2022), pp. 245–253. ISSN: 1546-1696. DOI: 10.1038/s41587-021-01033-z.

[90] M. Büttner et al. "scCODA is a Bayesian model for compositional single-cell data analysis". In: *Nature Communications* 12.1 (Nov. 2021), p. 6876. ISSN: 2041-1723. DOI: 10.1038/s41467-021-27150-6.

[91] Matteo C. Da Vià et al. "Homozygous BCMA gene deletion in response to anti-BCMA CAR T cells in a patient with multiple myeloma". In: *Nature Medicine* 27.4 (Apr. 2021), pp. 616–619. ISSN: 1546-170X. DOI: 10.1038/s41591-021-01245-5.

[92] Noopur Raje et al. "Anti-BCMA CAR T-Cell Therapy bb2121 in Relapsed or Refractory Multiple Myeloma". In: *New England Journal of Medicine* 380.18 (2019). PMID: 31042825, pp. 1726–1737. DOI: 10.1056/NEJMoa1817226.

[93] Daniel Wendisch et al. "SARS-CoV-2 infection triggers profibrotic macrophage responses and lung fibrosis". In: *Cell* 184 (26 Dec. 2021), 6243–6261.e27. ISSN: 00928674. DOI: 10.1016/j.cell.2021.11.033.

[94] O. Dietrich et al. "Dysregulated Immunometabolism Is Associated with the Generation of Myeloid-Derived Suppressor Cells in <i>Staphylococcus aureus</i> Chronic Infection". In: *Journal of Innate Immunity* 14.3 (2022), pp. 257–274. ISSN: 1662-811X. DOI: 10.1159/000519306.

[95] Thomas Däullary et al. "A primary cell-based in vitro model of the human small intestine reveals host olfactomedin 4 induction in response to Salmonella Typhimurium infection". In: *Gut Microbes*

15.1 (2023). PMID: 36939013, p. 2186109. DOI: 10.1080/19490976.2023.2186109.

[96] Cindrilla Chumduri et al. "Opposing Wnt signals regulate cervical squamocolumnar homeostasis and emergence of metaplasia". In: *Nature Cell Biology* 23.2 (Feb. 2021), pp. 184–197. ISSN: 1476-4679. DOI: 10.1038/s41556-020-00619-0.

[97] Wolfgang Huber and Susan Holmes. *Modern Statistics for Modern Biology*. 1st ed. deNBI – German Network for Bioinformatics Infrastructure, 2023.

[98] Andrew M. K. Law et al. "Advancements in 3D Cell Culture Systems for Personalizing Anti-Cancer Therapies". In: *Frontiers in Oncology* 11 (2021). ISSN: 2234-943X. DOI: 10.3389/fonc.2021.782766.

[99] John P. A. Ioannidis. "Why Most Clinical Research Is Not Useful". In: *PLOS Medicine* 13.6 (June 2016), pp. 1–10. DOI: 10.1371/journal.pmed.1002049.

[100] Jan Engelmann et al. "Uncertainty Quantification for Atlas-Level Cell Type Transfer". In: *arXiv* (Nov. 2022). DOI: 10.48550/arXiv.2211.03793.

[101] Sebastian Hannemann and Jorge E. Galán. "Salmonella enterica serovar-specific transcriptional reprogramming of infected cells". In: *PLOS Pathogens* 13.7 (July 2017), pp. 1–17. DOI: 10.1371/journal.ppat.1006532.

[102] Hannah E. Davis et al. "Long COVID: major findings, mechanisms and recommendations". In: *Nature Reviews Microbiology* 21.3 (Mar. 2023), pp. 133–146. ISSN: 1740-1534. DOI: 10.1038/s41579-022-00846-2.

[103] Jonas Schulte-Schrepping et al. "Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment". In: *Cell* 182.6 (Sept. 2020), 1419–1440.e23. ISSN: 0092-8674. DOI: 10.1016/j.cell.2020.08.001.

[104] L Sikkema et al. "An integrated cell atlas of the human lung in health and disease". In: *bioRxiv* (2022). DOI: 10.1101/2022.03.10.483747.

[105] Francisco Avila Cobos et al. "Benchmarking of cell type deconvolution pipelines for transcriptomics data". In: *Nature Communications* 11.1 (Nov. 2020), p. 5650. ISSN: 2041-1723. DOI: 10.1038/s41467-020-19015-1.

[106] Charles F. Fuechsel. *map*. https://www.britannica.com/science/map. Accessed: 15 May 2023.

[107] Johnny Harris. *Why all world maps are wrong*. Dec. 2016. URL: https://www.youtube.com/watch?v=kIID5FDi2JQ.

[108] The Editors of Encyclopedia Britannica. *Mercator projection*. https://www.britannica.com/science/cartography. Accessed: 15 May 2023.

[109] John Nelson. *Misconceptions*. URL: https://storymaps.arcgis.com/stories/1e7f582d478a4b99bd0c70fffeac4c8b.

[110] Tara Chari, Joeyta Banerjee, and Lior Pachter. "The Specious Art of Single-Cell Genomics". In: *bioRxiv* (2021). DOI: 10.1101/2021.08.25.457696.

[111] Svetlana Ovchinnikova and Simon Anders. "Exploring dimension-reduced embeddings with Sleepwalk". In: *Genome Research* (2020). DOI: 10.1101/gr.251447.119.

[112] The Tabula Sapiens Consortium and Stephen R Quake. "The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans". In: *bioRxiv* (2021). DOI: 10.1101/2021.07.19.452956.

[113] The Tabula Muris Consortium et al. "Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris". In: *Nature* 562 (2018), pp. 367–372. DOI: 10.1038/s41586-018-0590-4.

[114] *Human Cell Atlas*. URL: https://www.humancellatlas.org/.

[115] *LifeTime Initiative*. URL: https://lifetime-initiative.eu/.

[116] Rasa Elmentaite et al. "Cells of the human intestinal tract mapped across space and time". In: *Nature* 597.7875 (Sept. 2021), pp. 250–255. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03852-1.

[117] Jennifer E. Rood et al. "Impact of the Human Cell Atlas on medicine". In: *Nature Medicine* 28.12 (Dec. 2022), pp. 2486–2496. ISSN: 1546-170X. DOI: 10.1038/s41591-022-02104-7.

[118] Mohammad Lotfollahi et al. "Mapping single-cell data to reference atlases by transfer learning". In: *Nature Biotechnology* 40.1 (Jan. 2022), pp. 121–130. ISSN: 1546-1696. DOI: 10.1038/s41587-021-01001-7.

[119] Romain Lopez et al. “Deep generative modeling for single-cell transcriptomics”. In: *Nature Methods* 15.12 (Dec. 2018), pp. 1053–1058. ISSN: 1548-7105. DOI: 10.1038/s41592-018-0229-2.

[120] Carla Mölbert and Laleh Haghverdi. “Adjustments to the reference dataset design improve cell type label transfer”. In: *Frontiers in Bioinformatics* 3 (2023). ISSN: 2673-7647. DOI: 10.3389/fbinf.2023.1150099.

[121] Alexander D. Diehl et al. “The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability”. In: *Journal of Biomedical Semantics* 7.1 (July 2016), p. 44. ISSN: 2041-1480. DOI: 10.1186/s13326-016-0088-7.

[122] Emma Dann, Sarah A. Teichmann, and John C. Marioni. “Precise identification of cell states altered in disease with healthy single-cell references”. In: *bioRxiv* (2022). DOI: 10.1101/2022.11.10.515939.

[123] Kok Hao Chen et al. “Spatially resolved, highly multiplexed RNA profiling in single cells”. In: *Science* 348.6233 (2015), aaa6090. DOI: 10.1126/science.aaa6090.

[124] Patrik L. Ståhl et al. “Visualization and analysis of gene expression in tissue sections by spatial transcriptomics”. In: *Science* 353.6294 (2016), pp. 78–82. DOI: 10.1126/science.aaf2403.

[125] Giovanni Palla et al. “Squidpy: a scalable framework for spatial omics analysis”. In: *Nature Methods* 19.2 (Feb. 2022), pp. 171–178. ISSN: 1548-7105. DOI: 10.1038/s41592-021-01358-2.

[126] David S. Fischer, Anna C. Schaar, and Fabian J. Theis. “Modeling intercellular communication in tissues using spatial graphs of cells”. In: *Nature Biotechnology* 41.3 (Mar. 2023), pp. 332–336. ISSN: 1546-1696. DOI: 10.1038/s41587-022-01467-z.

[127] Elo Madissoon et al. “A spatially resolved atlas of the human lung characterizes a gland-associated immune niche”. In: *Nature Genetics* 55.1 (Jan. 2023), pp. 66–77. ISSN: 1546-1718. DOI: 10.1038/s41588-022-01243-4.

[128] Amit Frishberg et al. “Reconstructing disease dynamics for mechanistic insights and clinical benefit”. In: *bioRxiv* (2022). DOI: 10.1101/2021.11.17.468952.

[129] Florian Erhard et al. “scSLAM-seq reveals core features of transcription dynamics in single cells”. In: *Nature* 571.7765 (July 2019), pp. 419–423. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1369-y.

[130] Yafei Mao and Guojie Zhang. “A complete, telomere-to-telomere human genome sequence presents new opportunities for evolutionary genomics”. In: *Nature Methods* 19.6 (June 2022), pp. 635–638. ISSN: 1548-7105. DOI: 10.1038/s41592-022-01512-4.

[131] Aziz M. Al'Khafaji et al. “High-throughput RNA isoform sequencing using programmable cDNA concatenation”. In: *bioRxiv* (2021). DOI: 10.1101/2021.10.01.462818.

[132] Mandeep Singh et al. “High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes”. In: *Nature Communications* 10.1 (July 2019), p. 3120. ISSN: 2041-1723. DOI: 10.1038/s41467-019-11049-4.

[133] Judith A. Gilbert. “Advancing towards precision medicine in ARDS”. In: *The Lancet Respiratory Medicine* 6.7 (July 2018), pp. 494–495. ISSN: 2213-2600. DOI: 10.1016/S2213-2600(18)30156-5.

[134] T.M. Maher. “Precision medicine in idiopathic pulmonary fibrosis”. In: *QJM: An International Journal of Medicine* 109.9 (Aug. 2016), pp. 585–587. ISSN: 1460-2725. DOI: 10.1093/qjmed/hcw117.

[135] Sarah Goldman-Mazur et al. “Second- and third-line treatment strategies in multiple myeloma: a referral-center experience”. In: *Blood Cancer Journal* 12.12 (Dec. 2022), p. 164. ISSN: 2044-5385. DOI: 10.1038/s41408-022-00757-8.

[136] Leo Rasche, Michael Hudecek, and Hermann Einsele. “What is the future of immunotherapy in multiple myeloma?” In: *Blood* 136.22 (Nov. 2020), pp. 2491–2497. ISSN: 0006-4971. DOI: 10.1182/blood.2019004176.

[137] Yael C. Cohen et al. “Identification of resistance pathways and therapeutic targets in relapsed multiple myeloma patients through single-cell sequencing”. In: *Nature Medicine* 27.3 (Mar. 2021), pp. 491–503. ISSN: 1546-170X. DOI: 10.1038/s41591-021-01232-w.

[138] Philip Z. Mannes et al. “Molecular imaging of chemokine-like receptor 1 (CMKLR1) in experimental acute lung injury”. In: *Proceedings of the National Academy of Sciences* 120.3 (2023), e2216458120. DOI: 10.1073/pnas.2216458120.

[139] Charlotte Thibeault, Norbert Suttorp, and Bastian Opitz. “The microbiota in pneumonia: From protection to predisposition”. In: *Science Translational Medicine* 13.576 (2021), eaba0501. DOI: 10.1126/scitranslmed.aba0501.

[140] Nikolaus Rajewsky et al. “LifeTime and improving European healthcare through cell-based interceptive medicine”. In: *Nature* 587.7834 (Nov. 2020), pp. 377–386. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2715-9.

[141] Jonah Langlieb et al. “The cell type composition of the adult mouse brain revealed by single cell and spatial genomics”. In: *bioRxiv* (2023). DOI: 10.1101/2023.03.06.531307.

[142] K Rue-Albrecht et al. “iSEE: Interactive SummarizedExperiment Explorer [version 1; peer review: 3 approved]”. In: *F1000Research* 7.741 (2018). DOI: 10.12688/f1000research.14966.1.

[143] Hadley Wickham. *Mastering Shiny*. Accessed: 01 June 2023. O’Reilley, 2020.

[144] The Editors of Encyclopedia Britannica. *Plotly Open Source Graphing Library for Python*. https://plotly.com/python/. Accessed: 01 June 2023.

[145] Tyler Ard et al. “Integrating Data Directly into Publications with Augmented Reality and Web-Based Technologies – Schol-AR”. In: *Scientific Data* 9.1 (June 2022), p. 298. ISSN: 2052-4463. DOI: 10.1038/s41597-022-01426-y.

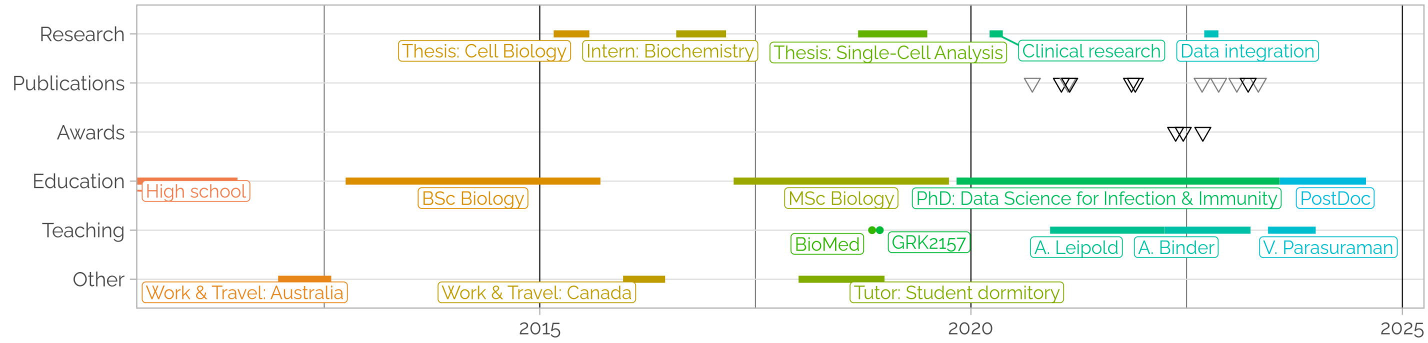
[146] Seán I. O’Donoghue et al. “Visualization of Biomedical Data”. In: *Annual Review of Biomedical Data Science* 1.1 (2018), pp. 275–304. DOI: 10.1146/annurev-biodatasci-080917-013424.

[147] Seán I. O’Donoghue. “Grand Challenges in Bioinformatics Data Visualization”. In: *Frontiers in Bioinformatics* 1 (2021). Specialty Grand Challenge. ISSN: 2673-7647.

Oliver Dietrich

14.03.1993
München

Dietrich.OliverDominik@gmail.com
Büttnerstraße 52, 97070 Würzburg



PROFILE

I am a doctoral candidate at the Helmholtz Institute for RNA-based Infection Research (HIRI) and the Graduiertenkolleg 2157 – 3D Infect – of the University of Würzburg. I work on the analysis and interpretation of single-cell sequencing data. My focus is on host-pathogen interactions in infection models and clinical samples. In order to provide access to data and analysis I have developed web-applications that are currently being hosted under Infection-Atlas.org

EDUCATION

Doctoral candidate, Infection & Immunity present
Julius-Maximilians-Universität Würzburg

Master of Science, Biosciences September 2019
Julius–Maximilians Universität Würzburg

Bachelor of Science, Biology September 2015
University of Regensburg

RESEARCH EXPERIENCE

Helmholtz Institute for RNA-based Infection Research
Dr. Antoine-Emmanuel Saliba Nov 2019 – present
Analysis of single-cell sequencing data from clinical samples and organoids

Helmholtz Munich, Institute for Computational Biology
Dr. Malte Lücken Sep 2022 – Nov 2022
Data integration and reference mapping using deep learning models

Charité, Immunology of Infectious Diseases & Vaccinology
Dr. Leif Erik Sander Mar 2020 – May 2020
Data analysis of clinical samples from patients with severe COVID-19

Helmholtz Institute for RNA-based Infection Research
Dr. Antoine-Emmanuel Saliba Sep 2018 – Jul 2019
Analysis of single-cell RNA-seq data for biomedical research

Fraunhofer Center for Molecular Biotechnology
Dr. Alexei Prokhnevsky Aug 2016 – Feb 2017
Purification strategy for enveloped virus-like particles for respiratory syncytial virus (RSV)

University of Regensburg, Dept. for Preclinical Medicine
Prof. Michael Krahn Mar – Sep 2015
Gene-knockouts in human cancer cell lines using CRISPR/Cas9

AWARDS

GRK2157 Start-Up PostDoc position	2023
Helmholtz Information and Data Science Academy (HIDA) Research Grant	2022
3rd International Conference 'Innate Immunity of the Lung – Improving Pneumonia Outcome', Best Abstract	2022
Helmholtz Centre for Infection Research (HZI), Paper of the Month	2021
GRK2157 PhD position	2019

Expertise
Host-Pathogen
Data Science & Viz
Clinics & Organoids

Languages
German, English,
French, R, Python,
bash (Linux), LaTeX

Ass. Peer Review
GigaScience, mBio
Nature Comms,
EMBO Journal

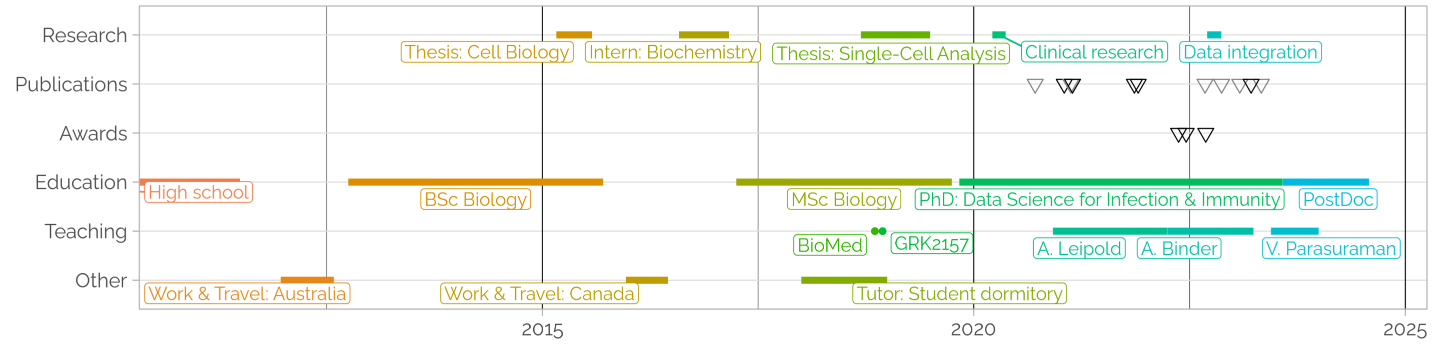
Tools
Seurat, scran,
scanpy, ggplot2,
tidyverse, shiny

GitHub.com
/OliverDietrich
/saliba-lab

Oliver Dietrich

14.03.1993
München

✉ Dietrich.OliverDominik@gmail.com
⌂ Büttnerstraße 52, 97070 Würzburg



TEACHING & SUPERVISION

Helmholtz Institute for RNA-based Infection Research
MSc supervision, Ariane Binder — APRIL 2022 – PRESENT
MSc supervision, Alexander Leipold — DEC 2020 – MAR 2022
scRNA-seq workshop, GRK2157 – 3D Infect — DEC 2018
scRNA-seq workshop, Biomedical Students Symposium — NOV 2018

PUBLICATIONS

(Co-)First Author

1. Wendisch, **Dietrich**, Mari, von Stillfried *et al.* SARS-CoV-2 infection triggers profibrotic macrophage responses and lung fibrosis. Cell (2021). 10.1016/j.cell.2021.11.033

2. **Dietrich O.**, Heinz A. *et al.* Dysregulated Immunometabolism Is Associated with the Generation of Myeloid-Derived Suppressor Cells in Staphylococcus aureus Chronic Infection. Journal of Innate Immunity (2021). 10.1159/000519306

Principal Analyst (scSeq)

3. Da Vià MC, **Dietrich O**, Truger M *et al.* Homozygous BCMA gene deletion in response to anti-BCMA CAR T cells in a patient with multiple myeloma. Nature Medicine (2021). 10.1038/s41591-021-01245-5
4. Chumduri C, Gurumurthy RK, Berger H, **Dietrich O** *et al.* Opposing Wnt signals regulate cervical squamocolumnar homeostasis and emergence of metaplasia. Nature Cell Biology (2021) 10.1038/s41556-020-00619-0
5. Däullary T, Imdahl F, **Dietrich O** *et al.* A primary cell-based in vitro model of the human small intestine reveals host OLFM4 induction in response to Salmonella Typhimurium infection. Gut Microbes (2023) 10.1080/19490976.2023.2186109

Co-Author

6. Leipold AM, Werner RA, Düll J, Jung P, John M, Stanojkovska E, Zhou X, Hornburger H, Ruckdeschel A, **Dietrich O** *et al.* Th17.1 cell driven sarcoidosis-like inflammation after anti-BCMA CAR T cells in multiple myeloma. Leukemia (2023) 10.1038/s41375-023-01824-0
7. Rizzo G, Gropper J, Piollet M, Vafadarnejad E, Rizakou A, Bandi SR, Arampatzi P, Krammer T, DiFabion N, **Dietrich O** *et al.* Dynamics of monocyte-derived macrophage diversity in experimental myocardial infarction. Cardiovascular Research (2022). 10.1093/cvr/cvac113
8. Friedrichs V, Toussaint C, Schäfer A, Rissmann M, **Dietrich O** *et al.* Landscape and age dynamics of immune cells in the Egyptian rousette bat. Cell Reports (2022) 10.1016/j.celrep.2022.111305
9. Schulte-Schrepping J, Reusch N, Paclik D, Baßler K, Schlickeiser S, Zhang B, Krämer B, Krammer T, Brumhard S, Bonaguro L, De Domenico E, Wendisch D, Grasshoff M, Kapellos TS, Beckstette M, Pecht T, Saglam A, **Dietrich O** *et al.* Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. Cell (2020) 10.1016/j.cell.2020.08.001
10. Fuhr V, Vafadarnejad E, **Dietrich O** *et al.* Time-Resolved scRNA-Seq Tracks the Adaptation of a Sensitive MCL Cell Line to Ibrutinib Treatment. International Journal of Molecular Sciences. (2021) 10.3390/ijms22052276

REFERENCES

Dr. Antoine-Emmanuel Saliba
Helmholtz Institute for RNA-based Infection Research, Single Cell Analysis
✉ EMMANUEL.SALIBA@HELMHOLTZ-HIRI.DE

Dr. Leif Erik Sander
Charité Berlin, Clinic for Infectiology and Intensive Care
✉ LEIF-ERIK.SANDER@CHARITE.DE

Expertise	**Languages**	**Ass. Peer Review**	**Tools**	**GitHub.com**
Host-Pathogen	German, English,	GigaScience, mBio	Seurat, scran,	/OliverDietrich
Data Science & Viz	French, R, Python,	Nature Comms,	scanpy, ggplot2,	/saliba-lab
Clinics & Organoids	bash (Linux), LaTeX	EMBO Journal	tidyverse, shiny	

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Graduate School
Life Sciences

Affidavit

I hereby confirm that my thesis entitled *Integrating single-cell multi-omics to decipher host-pathogen interactions* is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Place, Date Signature

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation *Integration von Genomik Daten einzelner Zellen zur Entschlüsselung von Wirt-Pathogen Interaktionen* eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Ort, Datum Unterschrift

Appendix

Abbreviations

ATAC-seq Assay for Transposase-Accessible Chromatin with high-throughput sequencing

AR Augmented reality

ARDS Acute respiratory distress syndrome

AUC Area under the curve

ASC Adult stem cell

BAL Bronchoalveolar lavage

BALF BAL fluid

BP Biological Process

CBC Crypt base columnar

CAR Chimeric antigen receptor

CC Cellular Compartment

CCA Canonical correlation analysis

ChEA3 ChIP-X enrichment analysis

CITE-seq Cellular Indexing of Transcriptomes and Epitomes by Sequencing

COVID-19 Coronavirus disease 2019

CT Computed tomography

DE Differential expression

DEG Differentially expressed gene

DNA Deoxyribonucleic acid

FACS Fluorescence-activated cell sorter

GI Gastrointestinal

GO Gene Ontology

GSEA Gene Set Enrichment Analysis

HCA Human Cell Atlas

HLCA Human Lung Cell Atlas

HTS High-throughput sequencing

HVG Highly variable genes

IBS Irritable bowel syndrome

IPF Idiopathic pulmonary fibrosis

iPSC Induced pluripotent stem cell

iSEE Interactive SummarizedExperiment Explorer

kNN *k*-nearest neighbors

LLSM Lattice light-sheet microscopy

MDSC Myeloid-derived suppressor cells

MNN Mutual nearest neighbor

ORA Over-representation analysis

PC Principal component

PCA Principal component analysis

PEEP Positive end-expiratory pressures

PET Positron emission tomography

QC Quality Control

RNA Ribonucleic acid

RNA sequencing RNA-seq

ROC Receiver operator characteristic

RRMM Relapsed/refractory Multiple Myeloma

scCODA Single-cell compositional data analysis

scRNA-seq Single-cell RNA-seq

scSeq Single-cell sequencing

scVI Single-cell variational inference

SARS-CoV Severe acute respiratory syndrome coronavirus

SCENIC Single-cell regulatory network inference and clustering

TF Transcription factor

tSNE t-distributed stochastic neighbor embedding

UMAP Uniform manifold approximation and projection

VR Virtual reality

WHO World Health Organization

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Graduate School
Life Sciences

Statement of individual author contributions and of legal second publication rights to manuscripts included in the dissertation

Homozygous *BCMA* gene deletion in response to anti-BCMA CAR T cells in a patient with multiple myeloma [91]

Participated in	**Author Initials**, Responsibility decreasing from left to right
Study Design Method Development	L.R., M.C.D.V., K.H., A.-E.S., H.E.
Data Collection	K.M.K., M.G., M.C., J.D., S. Kraus, A.H., X.Z., S.D., M.S.T., M.H., S.P., A.R., V.F, M.C.D.V., O.D., P.A., A.-E.S., M.T., S.M., S.T., M.M., C.H., N.W.
Data Analysis and Interpretation	O.D., F.E.
Manuscript Writing	L.R., M.C.D.V., K.H., A.-E.S., H.E., O.D.
Introduction	L.R., M.C.D.V., K.H., A.-E.S., H.E.
Materials & Methods	L.R., M.C.D.V., K.H., A.-E.S., H.E., O.D.
Discussion	L.R., M.C.D.V., K.H., A.-E.S., H.E.
First Draft	L.R., M.C.D.V., K.H., A.-E.S., H.E.

Explanations:

SARS-CoV-2 infection triggers profibrotic macrophage responses and lung fibrosis [93]

Participated in	**Author Initials**, Responsibility decreasing from left to right
Study Design	D.W., O.D., T.M., S.v.S., M.D.L., P.B., M.S., A.-E.S., and L.E.S
Method Development	D.W., O.D., T.M., S.v.S., I.L.I., S.B., T.K., A.P.R., R.G., A.E.H., M.D.L.
Data Collection	S.v.S., M.M, C.M., A.P.R., R.M., R.D.B., J.S., S.D., F.P., J.K., J.R., P.P., K.B., C. Conrad, A.C.A., B.S., M.L., E.W., D.H., S. Hippenstiel, A.H. F.L.H., A.U., F.M., S. Herold, S.E., C.T., M.W., C. Cochain, N.S., C.D., C.G., F.K., J.L.S., H.R., R.E., H.M.R., T.W., P.B., S.T., and M.O.
Data Analysis and Interpretation	D.W., O.D., T.M., S.v.S., I.L.I., R.L.C., H.Z., A.L.H, A.P.R., R.M., R.D.B., A.M.L., F.E., A.E.H., M.D.L., F.J.T., C.C., M.S., A.-E.S., and L.E.S
Manuscript Writing	D.W., O.D., T.M., S.v.S., A.C.A., J.L.S., M.S., A.-E.S., and L.E.S.
Introduction	D.W., O.D., T.M., S.v.S., A.C.A., J.L.S., M.S., A.-E.S., and L.E.S.
Materials & Methods	D.W., O.D., T.M., S.v.S., A.C.A., J.L.S., M.S., A.-E.S., and L.E.S.
Discussion	D.W., O.D., T.M., S.v.S., A.C.A., J.L.S., M.S., A.-E.S., and L.E.S.
First Draft	M.S., A.-E.S., and L.E.S.

Explanations:

Dysregulated Immunometabolism Is Associated with the Generation of Myeloid-Derived Suppressor Cells in *Staphylococcus aureus* Chronic Infection [94]

Participated in	**Author Initials**, Responsibility decreasing from left to right
Study Design Methods Development	O.G., E.M.
Data Collection	O.G., E.M., O.D., R.G., A.-E.S., A.H., K.H., A.B.
Data Analysis and Interpretation	O.D., A.E.S, A.H., K.H., A.B.
Manuscript Writing	E.M., O.G., O.D., A.H.
Introduction	E.M., O.G., O.D., A.H.
Materials & Methods	E.M., O.G., O.D., A.H.
Discussion	E.M., O.G., O.D., A.H.
First Draft	E.M.

Explanations:

A primary cell-based in vitro model of the human small intestine reveals host olfactomedin 4 induction in response to *Salmonella Typhimurium* infection [95]

Participated in	**Author Initials**, Responsibility decreasing from left to right
Study Design Methods Development	T.D., F.I., A.-E.S, D.Z.
Data Collection	T.D., F.I.
Data Analysis and Interpretation	O.D., T.D., F.I.
Manuscript Writing	T.D., F.I., O.D., D.Z., A.W.
Introduction	T.D., F.I., O.D., D.Z., A.W.
Materials & Methods	T.D., F.I., O.D.
Discussion	T.D., F.I., O.D., D.Z., A.W.
First Draft	T.D., F.I., O.D.

Explanations:

I confirm that I obtained the permission from both the publishers and co-authors for legal second publication.
I also confirm the correctness of the above mentioned information.

Oliver Dietrich	Date	Place	Signature
Antoine-Emmanuel Saliba	Date	Place	Signature

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Graduate School
Life Sciences

Statement of individual author contributions to figures/tables of manuscripts included in the dissertation

Homozygous BCMA gene deletion in response to anti-BCMA CAR T cells in a patient with multiple myeloma [91].

Figure	**Author Initials**, Responsibility decreasing from left to right
1	L.R., M.C.D.V., A.R., V.F., O.D.
2	M.T., S.M., S.T., M.M. and C.H.
S1	L.R., M.C.D.V., S. Kaiser, K.H.
S2	O.D.
S3	O.D.
S4	O.D.
S5	M.T., S.M., S.T., M.M. and C.H.
S6	F.E.
S7	A.R., V.F.
S8	M.T., S.M., S.T., M.M. and C.H.

Table	**Author Initials**, Responsibility decreasing from left to right
ST1	K.M.K., M.G., M.C., J.D., S. Kraus, A.H., X.Z., S.D., M.S.T., M.H., S.P
ST2	M.T., S.M., S.T., M.M., C.H.
ST3	N.W.
SourceEDF2	O.D., A.-E.S
SourceEDF3	O.D., A.-E.S
SourceEDF4	O.D., A.-E.S

SARS-CoV-2 infection triggers profibrotic macrophage responses and lung fibrosis [93].

Figure	**Author Initials**, Responsibility decreasing from left to right
1 & S1	D.W., L.E.S., O.D., A.E.S., S.v.S., A.P.R, R.M., A.E.H., H.R., S.D., P.B.
2 & S2	O.D., D.W., , A.E.S., L.E.S., T.K., S.B., M.M., R.K., H.M-R.
3 & S3	O.D., I.L.I., D.W., A.E.S., L.E.S., M.D.L., T.K., S.B., H.M-R.
4 & S4	R.L.C., D.W., S.D., C.C., P.B., A.P.R.
5 & S5	D.W., S.v.S., M.M., H.M-R. S.D., P.B., S.T., M.O.
6 & S6	O.D., D.W., A.E.S., L.E.S., F.P., J.K., C.G.
7 & S7	T.M., H.Z, M.S., C.M., J.S.

Table	**Author Initials**, Responsibility decreasing from left to right
S1	D.W., M.M., L.E.S.
S2	O.D., A.E.S.
S3	O.D., A.E.S.
S4	D.W., S.v.S.
S5	O.D., A.E.S.
S6	T.M., H.Z, M.S., C.M., J.S.

Dysregulated Immunometabolism Is Associated with the Generation of Myeloid-Derived Suppressor Cells in *Staphylococcus aureus* Chronic Infection [94]

Figure	**Author Initials**, Responsibility decreasing from left to right
1	E.M., O.G.
2	O.D., A.E.S.
3	O.D., A.E.S.
4	O.G, E.M., O.D., A.E.S., A.B.
5	A.H., K.H.
6	O.G., E.M.
7	O.G., E.M.
S1-10	O.G., E.M.

Table	**Author Initials**, Responsibility decreasing from left to right
S1	O.D., A.E.S.
S2	O.D., A.E.S.

A primary cell-based *in vitro* model of the human small intestine reveals host olfactomedin 4 induction in response to *Salmonella Typhimurium* infection [95]

Figure	**Author Initials**, Responsibility decreasing from left to right
1	T.D., O.D., F.I.
2	T.D., F.I.
3	T.D., F.I.
4	T.D., F.I.

Table	**Author Initials**, Responsibility decreasing from left to right
1	O.D.

I also confirm my primary supervisor's acceptance.

Oliver Dietrich	Date	Place	Signature

List of Publications

(Co-)First Author

1. Wendisch, **Dietrich**, Mari, von Stillfried *et al.* SARS-CoV-2 infection triggers profibrotic macrophage responses and lung fibrosis. Cell (2021). 10.1016/j.cell.2021.11.033

2. **Dietrich O.**, Heinz A. *et al.* Dysregulated Immunometabolism Is Associated with the Generation of Myeloid-Derived Suppressor Cells in Staphylococcus aureus Chronic Infection. Journal of Innate Immunity (2021). 10.1159/000519306

Principal Analyst (scSeq)

3. Da Vià MC, **Dietrich O**, Truger M *et al.* Homozygous BCMA gene deletion in response to anti-BCMA CAR T cells in a patient with multiple myeloma. Nature Medicine (2021). 10.1038/s41591-021-01245-5

4. Chumduri C, Gurumurthy RK, Berger H, **Dietrich O** *et al.* Opposing Wnt signals regulate cervical squamocolumnar homeostasis and emergence of metaplasia. Nature Cell Biology (2021) 10.1038/s41556-020-00619-0

5. Däullary T, Imdahl F, **Dietrich O** *et al.* A primary cell-based in vitro model of the human small intestine reveals host OLFM4 induction in response to Salmonella Typhimurium infection. Gut Microbes (2023) 10.1080/19490976.2023.2186109

Co-Author

6. Leipold AM, Werner RA, Düll J, Jung P, John M, Stanojkovska E, Zhou X, Hornburger H, Ruckdeschel A, **Dietrich O** *et al.* Th17.1 cell driven sarcoidosis-like inflammation after anti-BCMA CAR T cells in multiple myeloma. Leukemia (2023) 10.1038/s41375-023-01824-0

7. Rizzo G, Gropper J, Piollet M, Vafadarnejad E, Rizakou A, Bandi SR, Arampatzi P, Krammer T, DiFabion N, **Dietrich O** *et al.* Dynamics of monocyte-derived macrophage diversity in experimental myocardial infarction. Cardiovascular Research (2022). 10.1093/cvr/cvac113

8. Friedrichs V, Toussaint C, Schäfer A, Rissmann M, **Dietrich O** *et al.* Landscape and age dynamics of immune cells in the Egyptian rousette bat. Cell Reports (2022) 10.1016/j.celrep.2022.111305

9. Schulte-Schrepping J, Reusch N, Paclik D, Baßler K, Schlickeiser S, Zhang B, Krämer B, Krammer T, Brumhard S, Bonaguro L, De Domenico E, Wendisch D, Grasshoff M, Kapellos TS, Beckstette M, Pecht T, Saglam A, **Dietrich O** *et al.* Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. Cell (2020) 10.1016/j.cell.2020.08.001

10. Fuhr V, Vafadarnejad E, **Dietrich O** *et al.* Time-Resolved scRNA-Seq Tracks the Adaptation of a Sensitive MCL Cell Line to Ibrutinib Treatment. International Journal of Molecular Sciences. (2021) 10.3390/ijms22052276