SINGLE CELL RIBONUCLEIC ACID SEQUENCING IN TUBERCULOSIS RESEARCH

By

Ian Maheti Mbano

A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy in the College of Health Sciences, University of KwaZulu-Natal

Supervisor

Dr Alasdair Leslie, Africa Health Research Institute

2021

Preface

The study described in this thesis was carried out at the Africa Health Research Institute (AHRI), Nelson R. Mandela School of Medicine, University of KwaZulu-Natal in Durban, South Africa between April 2018 and July 2021 under the supervision of Dr Alasdair Leslie.

The study described is an original work performed and reported by the author. The study has not been submitted in any form, by any person or submitted to any tertiary institution for award of a degree or diploma. Some of the work has been published in accredited journals in line with the UKZN thesis guidelines. Due acknowledgements have been accorded where other people's work has been used in the text, such as the seq-well platform created in Professor Alex Shalek and Christopher Love's laboratories at the Massachusetts Institute of Technology (MIT).

Declaration 1: Plagiarism

I **Ian Maheti Mbano** declare that,

(i) The research reported in this dissertation, except where otherwise indicated, is my original work.

(ii) This dissertation has not been submitted for any degree or examination at any other university.

(iii) This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

v) their words have been re-written, but the general information attributed to them has been referenced.

(vi) This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signed: 

Ian Maheti Mbano **Date:** 04 June 2021

Signed: 

Dr Alasdair Leslie **Date:** 04 June 2021

Declaration 2: Publications

The publication (published, in print and/or submitted) that constitute this thesis and the contribution I made to each of the manuscripts are presented here.

Publication 1

Carly G. K Ziegler, Samuel J. Allon, Sarah K. Nyquist, Ian M. Mbano *.et al* (2020). SARS-CoV-2 Receptor ACE2 is an Interferon- Stimulated Gene in Human Airway Epithelial Cells and is Detected in Specific Cell Subsets across Tissues. *Cell 181, 1016-1035.*

Author's contributions

I and my supervisor (Dr Alasdair Leslie) conceptualized and provided the analysis for the human lung data analysis, which formed the corner stone on which the study was based. We were also involved in drafting and critically reviewing the paper, earning me co-first authorship.

Publication 2

Ian M. Mbano, Tawanda Mandizvo, Jerome Rogich, Tafara T.R Kunota, Jared S. MacKenzie, Manormoney Pillay and Frederick K. Balagadde (2020). Light Forge: A Microfluidic DNA Melting based Tuberculosis Test. *Journal of Applied and Laboratory Medicine 5, 440-453*.

Author's contributions

I and my then supervisor (Dr Frederick Balagadde) and other co-author (Tawanda Mandizvo) conceptualised the idea. I did the experimentation, data analysis and developed the draft manuscripts. The supervisors and co-authors critically reviewed the manuscript.

Signed: 

Ian Maheti Mbano **Date:** 04 June 2021

Acknowledgements

I would like to take this opportunity to thank my supervisor, Dr Alasdair Leslie, who has patiently trained me in immunology, especially given that l had no experience in human biology. I would also like to thank the members of the Leslie and Kløverpris Labs, who have provided me with training and support throughout the duration of my PhD. Thank you for taking your time to teach me different experimental and analytical methods.

Thank you to the staff at the Africa Health Research Institute for their unending support, ranging from administrative functions such as facilitating my study permits to the clinical and bio-repository staff who seamlessly run the various studies. I would like to acknowledge the Council of Scientific and Industrial Research (CSIR) for graciously supporting me the first 2 years of this project.

I would like to thank our collaborators from the Massachusetts Institute of Technology (MIT), particularly Professor Alex K Shalek, the cardiothoracic surgeons at Inkosi Albert Luthuli Central Hospital and King Dinizulu Hospital together with their respective surgical teams. I would like to thank all my family members, notably my dad, Mr Fred Mbano for always believing in me. A huge thanks to my sisters Beatrice and Natasha for being a constant source of inspiration, together with all my wonderful nieces, nephews and everyone who constitutes my extended family.

I would also like to thank my Lord and Saviour Jesus Christ, for always strengthening me throughout this journey. There is no way l could have made it without the timely strengthening l got from the word of God and for that l am truly appreciative beyond words.

Finally, l would like to give a special note of gratitude to everyone who consented to be a participant to the study, without which new insights into our understanding of TB disease would not have been wrought.

Table of Contents

List of Acronyms

ACE2	Angiotensin Converting Enzyme 2
AIDS	Acquired Immune Deficiency Syndrome
BCG	Bacille Calmette-Guérin
CA12	Carbonic Anhydrase 12
CD69	Cluster Differentiation 69
CD103	Cluster Differentiation 103
COL1A1	Collagen Type 1 Alpha 1
COL1A2	Collagen Type 1 Alpha 2
COPD	Chronic Obstructive Pulmonary Disease
CTHRC1	Collagen Triple Helix Repeat Containing-1
CXCL13	C-X-C Motif Chemokine Ligand 13
FAP	Fibroblast Activated Protein
FACS	Fluorescence Activated Cell Sorter
GAS1	Growth Arrest Specific 1
H&E	Haemotoxylin and Eosin Stain
HIV	Human Immunodeficiency virus
IGRA	Interferon Gamma Release Assay
MMP1	Matrix Metalloproteinase-1
MMP3	Matrix Metalloproteinase-3
MMP9	Matrix Metalloproteinase-9
Mtb	Mycobacterium tuberculosis
NHP	Non-human primates
PBMCs	Peripheral blood mononuclear cells
PI15	Peptide Inhibitor 15
PDPN	Podoplanin
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
scRNAseq	Single cell ribonucleic acid sequencing
SIPR	Sphingosine 1 Phosphate Receptor
TB	Tuberculosis
Tcm	T Central Memory cells
TDO2	Tryptophan 2,3-dioxygenase

Tem	T Effector Memory cells
TIMP3	Tissue Inhibitor of Metalloproteinase 3
TLS	Tertiary Lymphoid Structure
TMPRSS2	Type 2 Transmembrane Serine Protease
Trm	Tissue resident Memory T cells
VATS	Video Assisted Thorascopic Surgery
UMAP	Uniform Manifold Approximation and Projection

Abstract

Tuberculosis (TB) remains a global challenge, with approximately 1,5 million deaths annually. Addressing deficits in our understanding of disease pathology and treatment is needed for the development of new treatment modalities. Despite much effort, prevalence of this disease remains high in resource limited regions, where research capacity is not sufficient to successfully combat the endemic. Research in developed countries has generally been constrained to animal models due lack of access to clinical samples from the site of TB disease, the human lung. Although these animal models have their utility, it is essential that findings from these systems be tested and validated in human tissue. In this thesis, I leveraged a relatively new technology called Seq-well, which is highly portable and low-tech single cell ribonucleic acid sequencing (scRNAseq) platform and access to TB infected lung tissue obtained from lung resections, to generate a single cell atlas of TB affected lung tissue. This involved processing the human tissue immediately post-surgery and loading unprocessed/neat cells or FACS sorted cells (tissue resident t cells) onto a microarray that allowed capture and subsequent sequencing of the cell transcriptomes. In the first part of the thesis, I identified and profiled cellular subsets from TB infected tissue, focussing on a subset of *FAP+PDPN +* fibroblasts associated with the organisation of tertiary lymphoid organs. I also demonstrated that this dataset can be useful in evaluating current and future TB biomarkers, by superimposing signatures from the literature onto the cellular subsets and localizing them to different parenchymal, stromal and immune cell types. I also profiled tissue resident CD4 T cells from the same lung tissue, identifying canonical marker genes (*ITGA1, PRF1*) in one specific cluster, together with naive (*CCR7, SELL*), regulatory (*RORA*) and activated/myeloid-like T cells (*LYZ, S100A9*) in separate clusters. Finally, I demonstrated the applicability of this dataset in research involving other pulmonary diseases, by identifying *ACE2+ TMPRSS2+* type 2 pneumocytes, a target of the SARS-CoV-2. Taken together, these findings provide new insights into the immunopathology of TB in the human lung together with the impact of HIV on specific immune subsets. It serves as a resource for cross validation of lung immune signatures generated in experimental infections of both mice and non-human primates, which is beneficial for scientists lacking access to the technology and/or tissue.

Iqoqa

Isifo sofuba (i-TB) silokhu siyinselelo emhlabeni jikelele, ngokufa okuhlobene naso okucishe kufike esigidini esi-1.5 njalo ngonyaka. Ukubhekana nokushoda ekuqondeni kwethu umumosakhiwo wesifo bese kuncishiswa ukufa. Ngaphandle kwemizamo emikhulu, ukudlanga kwalesi sifo kusalokhu kuphezulu ezifundeni ezintula imithombokusiza, lapho umthamokwenza wocwaningo unqindekile. Ucwaningo emazweni asethuthukile, ngakolunye uhlangothi, belwenzeka kuphela kumamodeli asebenzisa izilwane ngenxa yokuntuleka kokufinyelela amasampuleni okwelapha engxenyeni okuqubuke kuyo isifo sofuba, okuyiphaphu lomuntu. Nakuba kunamamodeli ezilwane anomsebenzi, kubalulekile ukuba okutholakele kulezo zinhlelo kuyohlolwa bese kuqinisekiswa ngesigqa somuntu ukuqinisekisa ubunjalo. Kule thesisi, ngiveze ubuchwepheshe obusha obungenayo obubizwa nge-Seq-well, iseli eyodwa e-low-tech ephathekayo ene-ribonucleic acid sequencing (scRNASeq) okuyindawo kanye nokufinyelela esicutshini sephaphu esitheleleke ngesifo sofuba esitholakale ekuhlukanisweni kabusha kwamaphaphu okukhonjwe ngokokwelapha, ukwakha iseli eyodwa yesicutshana sephaphu elitheleleke ngesifo sofuba.

Lokhu kwafaka ukusebenzakuhlola isicubu somuntu ngokushesha emva kokuhlinza nokufaka amaseli ahlanzekile angasetshenziwe noma amaseli ahleliwe angama-FACS (ama-T cells asesicutshini) ohlelweni lolibofuzo olwavumela ukufaka ohlwini nokulandelanisa okulandelayo womumofuzo oqondene nezicubu. Engxenyeni yokuqala yethesisi, amaqoqwana ahlonziwe nafakwe kwiphrofayli esicubini esitheleleke ngesifo sofuba kugxilwe eqoqweni le FAP+PDPN + amafayibhroplasti ahlobene nokuhlelwa kwezingxenye zomzimba ezinkulu zamalimfoyidi kanye nemichilwana yamafayibhrodi kanye noma igranyuloma yesifo sofuba. Ngivezile ukuthi lamadathasethi angaba nomsebenzi omkhulu ekuhlaziyeni amabhayomakha amanje nawasesikhathini esizayo esifo sofuba, ngokufaka izinkombabunjalo emaqoqweni amancane nokuwabeka ezinhlotsheni ezehlukene zamaseli angamapharenikhayma nangamastroma.

Ngiphinde ngachaza esizindeni sezicutshana ze-CD4 T esicutshini sephaphu elifanayo okuchaza ulibofuzo olukala amakhenoni (i-*ITGA1, PRF1*) eqoqweni elilodwa eliqondile, kanye namaseli angachazi lutho (*CCR7, SELL*), alawulayo (*RORA*) nama-T cell aqaliswe ukusebenza/efana ne-myeloid (*LYZ, S100A9*) emaqoqweni aseceleni. Okokugcina, ngiveze ukungena kwedathasethi ocwaningweni olufaka izifo zamaphaphu nokuphefumula ngokuhlonza i- *ACE2*+ *TMPRSS2*+ type 2 wama-pneumocytes, okuhlosiwe kwe-SARS-CoV-2. Uma kuhlanganisiwe, lokhu okutholakele kuletha imibono emisha yomumobugciwane bokutheleleka ngesifo sofuba ephashini lomuntu, umthelela we-HIV kokutholakele emumwenikuphila kwephaphu ekuthelelekeni okuyilinga kwakho kokubili amagundane kanye nalokho okungebona abantu.

Chapter 1: Single cell sequencing technology and its application in Tuberculosis (TB) research.

1.1 Introduction

Mycobacterium tuberculosis(Mtb), an ancient pathogen discovered in 1882, is the causative agent of the deadly human tuberculosis (TB) disease which primarily infects lung tissue [1]. This disease is the leading cause of death in the world by a single infectious agent, leading to 1.4 million deaths in 2019, with 251000 amongst people living with HIV/AIDS [2]. Untreated HIV/AIDS infection is a high risk factor for developing/contracting TB [3] . There is a strong correlation between TB incidence and the quality of the health delivery system as evidenced by 10 new cases per 100000 population in high income nations compared to 150 to 400 cases in resource constrained nations such as South Africa. However, the reduction in global mortality rates by 42% between 2008 to 2018 is encouraging [2]. Estimates suggest that 25% of the global population is infected with TB, with most individuals showing no clinical symptoms and only a tenth of these developing active disease at some point in their lifetime [4]. However, the human immune correlates of TB control or progression are not known, thus more basic science research is required to understand what factors of the immune system are uniquely reflective of active TB disease [5].

Correlates of protection are critical in the development of TB vaccines and diagnostics markers. Antibody levels and the activity of cytotoxic T cells have been proposed as effective surrogates for activity (even though they are not protective) against HIV [6]. These correlates have not been characterized in the TB vaccine *M. bovis* bacille Calmette–Guérin (BCG) induced response, due to inconsistent protective efficacies [7]. A quantitative review of a BCG trial showed

efficacy levels of 14% in a 10-year period [8], whilst another study on native American Indian and Alaskan populations reported efficacy of 82% in a 20 year period [9]. The large variance in the responses suggests that more studies centered around the host immune response are necessary amongst different population groups.

Our understanding of the immune response to TB is largely derived from animal models (mice, guinea pigs, rabbits and zebrafish) [10]. These different models have their respective strengths and weaknesses (Table1), continue to contribute immensely towards our understanding of a TB disease. Studies have shown that non-human primates (NHP) present with very similar TB immunopathology with humans, specifically in terms of the variety of granulomas and the infection spectrum (chronic and active) [11]. However, NHP require special units for husbandry within a biosafety level 3 (BSL3) facility, skilled personnel, large experimental numbers as they are essentially outbred, significantly increasing operational costs [12]. Other studies have focused on measurement of immune responses in human serum (from the blood) and bronchoalveolar lavage fluid (BALF) of TB infected individuals. It is assumed that both samples capture some of the temporal and spatial dynamics occurring within diseased lung tissue. Both these samples present with limitations as serum captures the overall systemic response, which is not necessarily reflective TB related pulmonary deterioration [13], whilst BALF represents the airway response as opposed to the parenchyma were the hallmarks of TB disease are observed [10].

Table 1: Animal models in TB research

Animal	Strain/Dose	Susceptibility to TB	Pulmonary Lesions			Pros	Cons	Application	Reference
			Necrotic	Caseation	Cavitation				
Zebrafish	*M.marinum*	none	-	-	-	1 Require less space and short experimental duration 2 Transluscent allowing visualization	1 Cannot be infected by *M.tuberculosis*. 2 Limited immune reagents constrains mechanistic studies	1 Latent infection model 2 Useful for drug efficacy assays and observing lesion formation	14
Mouse	*M.tuberculosis* H37Rv/10^5-10^7 CFU	low	Only in C3He/FeJ mice	no	no	1 Well studied model and abundant immune reagents 2 Small size, low cost and relatively easy to maintain 3 Good for studying the mechanistic consequence of genes using knockout mice	1 No clinical manifestations 2 Granulomans lack giant Langerhans and Epithelioid cells 3 Lack disseminated disease 4 Variation in infection outcome amongst individual mice	1 Good model for vaccines and mechanisms of TB responses 2 Characterization of immune response to TB infection	15
Guinea Pig	*M.tuberculosis* H37Rv/10^3-10^5 CFU	high	yes	yes	variable	1 Easily infected with TB 2 Granuloma like those in humans 3 Good response to TB vaccines and drugs	1 Limited immune reagents restrict research 2 No clinical manifestations 3 Does not develop latent infection	1 Vaccine safety and efficacy studies 2 Drug evaluations	16
Rabbit	*M.tuberculosis* H37Rv 10^8 CFU	low, more prone to *M.bovis*	yes	yes	yes	1 Granuloma very similar to those observed in humans 2 Allows for study of other types of TB disease	1 No obvious clinical manifestations 2 Limited immune reagents 3 Limited susceptibility to TB infection	1 Transmission studies 2 Model for cavitary, spinal, cutaneous and meningeal TB	17
Non-Human Primates	*M.tuberculosis* 10^2-$5*10^2$ CFU	high	yes	yes	yes	1 Mimic clinical manifestations of human TB 2 TB lesions are formed in the lung and other organs 3 Both pulmonary and extra-pulmonary TB 4 Granuloma similar to those in humans in terms of structure and cellular composition	1 Limited availability of monkeys and immune reagents 2 Expensive to house in BSL3 facilities	1 Drug and vaccine effectiveness 2 Immune response to TB and the associated pathological profile	18

Other studies have used healthy human lung explants to set up an *ex vivo* infection model, in an attempt to delineate the initial response to infection [19]. The tissue was obtained from participants undergoing surgery for non-pulmonary infections and was subsequently infected with a high concentration of *M.tuberculosis* (5 million colony forming units). Gene expression profiles generated using microarray and bulk rna sequencing were used to assess the response of the whole lung and fluorescent activated cell sorter (FACS) sorted cell types. These included alveolar macrophages, epithelial cells, endothelial cells, innate lymphoid cells (ILCs), γδ T cells and mucosal associated invariant T (MAIT) cells. These cell types were selected based on their ability to display a rapid, untrained (innate) response to infection. The myeloid cells showed a very strong inflammatory response marked with expression of IL1-β and IL23. Epithelial and endothelial cells showed the same with a significantly reduced magnitued. These inflammatory responses have been reported to also modulate ILCs, for example IL1-β and IL23 induce ILC3s [20] ,whereas TSLP and IL33 activate ILC2s [21]. Even though the model offers a unique way of studying the host response, limitations such the high infectious dose (likely not reflective of an infection event) minimize any extrapolation . The lack of blood circulation meant there was compromised migration of immune cells to the site of infection together with the disruption of tissue homeostasis. It is plausible that better insights into the immune response to TB disease will be generated by exploratory studies of lung infected tissue using tools that provide the most comprehensive gene expression profiles of the cells.

Single cell sequencing has the ability to profile unitary genomes [22], transcriptomes [23] or to a lesser extent the epigenome [24] or proteomes [25]. These technologies generate large amounts of data which provide analytical power that is unprecedented in the field of biology [26]. Single cell ribonucleic acid sequencing (scRNAseq), is most widely used and allows profiling of mRNA (transcriptome) from tissue or in-vitro experiments [27]. Several scRNAseq platforms have been

developed, which differ in the number of cells and number of genes that are recovered per run. For example, Smartseq2, a full length transcriptome sequencing approach, makes use of a high fidelity reverse transcriptase (yielding 7000 genes per cell) [28]. However, it has limited cell capacity (96 or 384 well plates in a single run [29]). The protocol is laborious, leading to high technical variability in the experimental outcomes as a result of pipetting errors in the sequencing library work up (robots are available but they are expensive). On the other hand, highly parallel approaches such as 10X (Chromium), that incorporate universal molecular identifiers for transcript enumeration can profile 100-100000 cells at a depth of 3000 genes per cell for primary cells [30]. Such methods often have less manual steps in the protocol and shorter turnaround times [28]. In practice, Smartseq2 is mostly used to characterize rare cell types due to the higher gene recovery [31], whilst massively parallel methods are applied in tissue phenotyping studies where the large number of cells are required to uncover low frequency, novel cell types [32]. Thus, scRNAseq represents a unique opportunity for accelerated discoveries in our understanding of the immune response to TB infection, improving our management of TB disease. In this review, we discuss aspects of the immune response to TB infection that will benefit from application of scRNAseq as a complement to existing approaches.

1.2 TB infected lung cellular diversity

Tuberculosis is generally believed to present as primary and post primary disease [33]. Primary TB is the disease without any TB history shortly after exposure to the *Mtb* pathogen, leading to a single lesion typically in the highly aerated areas of the lung such as the lower section of the upper lobe [33]. Post primary TB arises in individuals with previous TB history possibly due to reactivation (following an immune compromising event) of dormant bacteria. It largely occurs in the lung apices, but has also been observed in the hilum and hilar lymph nodes [34]. It either leads to necrotic cavitation when the lesion is located close to an airway or a fibrotic

scarring [35]. Amongst adults, post primary TB is suspected to be the most common, often with caseating granuloma , tissue cavitation and fibrosis at varying degrees [33].

Lung fibrosis observed in post primary TB is a consequence of anti-inflammatory responses that are important for prevention of host induced tissue damage [36]. Growth factors such as insulin-like growth factor-1 (IGF1), platelet derived growth factor (PDGF) and fibroblast growth factor-2 (FGF2) are implicated in driving fibrosis [36], together with migration of fibroblast like cells to the lung [37,38]. A highly activated subset of fibroblasts, known as myofibroblasts, is thought to drive the fibrotic phenotype by increasing extra cellular matrix deposition and shows aberrant contractile lead to dysfunctional wound healing [39]. Myofibroblast have been reported to be enriched by the expression of transforming growth factor beta (TGFB), tenascin-C (TNC), connective tissue growth factor (CTGF) and phosphate and tensin deleted on chromosome homolog 10 (PTEN) [37,40,41]. Cytokines associated with type 2 immune response (IL-4,IL-5,IL-9 and IL-13) [42] and vascularization (angiogenesis) [43], have been associated with a chronic fibroproliferative phenotype. It is important that we explore this phenotype in TB, considering reports of fibrosis occurring in older resolving TB granuloma and prominently occuring following anti-TB therapy [44,45].

Single cell RNA sequencing of TB infected lung tissue has the potential to provide unique insights into the different cell types (immune, stromal, epithelial and endothelial) associated with fibrosis. It can potentially highlight which cell types are enriched or lost as consequence of disease in a highly, bias-independent manner. Such ability was elegantly illustrated in a study with 8 idiopathic pulmonary fibrosis (IPF) patients and 8 lung transplant donors using scRNAseq uncovered a previously unknown populations of macrophages, airway stem and senescent cells that were enriched in the IPF lungs [46]. These observations were confirmed using

complementary technologies (bulk RNA-sequencing and immunohistochemistry), showing its unique ability to generate pathology specific insights which can guide mechanistic studies.

It was reported that HIV and TB positive individuals with suppressed CD4 cell counts had chest x-rays very divergent from TB only individuals, with attenuated tissue destruction [47]. This observation coincided with diminished levels of extracellular matrix (ECM) degradation enzymes known as matrix metalloproteinases (MMPs), in coinfected participants [48] . However, inconsistencies in the cytokine profile of individuals within and across studies complicates the interpretation of the these observations [49]. A study by Walker and colleagues compared the protein levels in plasma and sputum levels of MMPs of TB , TB/HIV and symptomatic respiratory cases [50]. TB infected participants had higher levels of MMP-1 and MMP-3 compared to the TB/HIV infected group, with lower frequency of cavities and inflammation. This observation suggests that TB/HIV individuals have reduced pulmonary damage, but it should be noted that the same study also reported high levels of ECM turnover product called Procollagen III N-terminal propeptide (PIIINP) in TB/HIV group. The authors speculated that this observation was a consequence of tissue damage in other organs due to the development of disseminated TB disease. Understanding the cellular diversity of lung tissue from TB in comparison to TB/HIV can provide insights into the mechanisms that drive tissue degradation. This has the potential to uncover cells which produce MMPs and their inhibitors; tissue inhibitors of metalloproteinases (TIMPs), together with genes that modulate tissue degradation such as growth factors, cytokines and hormones [51]. Single cell sequencing data lends itself to high throughput network interactions of each cell, thus it has the potential to also uncover cellular interactions that initiate and maintain this pathology [52].

Granulomas are the hallmark of the pulmonary response to TB disease [53]. It is believed that upon infection with *Mycobacterium tuberculosis (Mtb),* alveolar macrophages engulf the bacteria. This leads to production of chemokines by the macrophage which attracts other inflammatory cells such as neutrophils, NK cells, monocyte derived macrophages and γδ T cells [54-56]. Dendritic cells then migrate to the lymph nodes where they prime T cells to differentiate into th1 (helper t cells) and th17 (cytotoxic t cells) [57].These primed t cells are important for further activation of macrophages and their migration to the site of infection together with B cells. This creates a lymphocyte cuff surrounding the bacteria harboring macrophages [58].The histological presentation of granulomas varies from poorly structured containment units that are highly transmissive to protective often calcified units that suppress bacterial growth and transmission [59]. Studies in NHPs have shown that TB lesions within the same animal are heterogeneous, both in terms of histology and bacterial load, for both active and latently infected animals [60]. A study by Subbian and colleagues characterized the gene expression profile from a wide variety of lesions in individuals undergoing lobectomy due to TB related complications [13]. The authors reported host immune responses associated with inflammation (*CXCR4, CCL3, CXCL8*), tissue degradation (*MMP1, MMP9*) and downregulation of transcriptional regulators such as (*FOXC1, ERG, ATN1*) in lesions isolated from active TB participants. However, the low sample number and limitations inherent to microarray technology impeded firm conclusions, but that study represents a unique window into the granuloma gene expression profile [61]. Applying scRNAseq in the characterization of the transcriptional profiles of TB lesions at different stages has the potential to deconstruct host mechanisms that determine granuloma fate. In combination with histology, bacterial culture (colony forming units or chromosomal equivalents) and flow cytometry, this will improve our understanding of lesion biology. The ability to associate immune profiles at a cell level with

bacterial persistence will greatly improve our insight into latent TB, a major reservoir of the disease [4].

1.3 TB biomarker discovery and validation

Biomarkers are a necessary component of the successful development of new TB therapeutics and vaccines. Best performing biomarkers can detect either the pathogen together with its products (DNA, genes or proteins) or changes in the human immune system in response to pathology. The major sight of disease is the lung, thus pathogen directed tests use sputum, which has variable sensitivity, which is significantly lower in HIV positive individuals. To improve on this sensitivity, measured of host derived molecules that reflect disease has been pursued from blood and to a lesser extent urine. These tests often has a large number of molecules associated with the disease but variable accuracy and heterogenous immune response to the disease prevents such tests from being incorporated into the TB detection clinical algorithm.

For most individuals with TB, clinical detection occurs only after they present with significant symptoms. It has been suggested that the time from initial symptoms to the clinical diagnosis can be as much as 6 months due to inefficiencies of the currently available tools and health care delivery systems [62]. During this delay, the underlying biosignature that is the consequence or the driver of pathology is likely present at either the site of disease or the circulatory system [63]. It has been a focus of many researchers to discover this TB biosignature, particular one that can capture the full spectrum of disease from the point of infection, subclinical (latent) and active TB disease [64].

An elegant approach was reported by Singhania and colleagues, in which they developed a 20 gene signature that distinguished active from latent TB by utilizing bulk RNA sequencing of peripheral blood mononuclear cells (PBMC) [65]. They made use of weighted gene co-

expression network analysis [66], which produces related gene modules that allow insights into functional aspects of the biosignature. However, the signature was enriched with interferon inducible genes, which are also observed in viral infections such as influenza [67]. A chronic interferon response was associated with a high bacillary load and advanced pathology in animal models [65]. The overabundance of the interferon genes across different studies is likely a result of pathogen specific host responses and shared innate responses to other stimuli. In the same study, the biosignature was measured longitudinally in TB contacts tested using the interferon gamma release assay (IGRA). The contacts consisted of 31 (15 IGRA- and 16 IGRA+) healthy individuals, 9 contacts who developed culture positive disease and 5 individuals who were outliers at baseline (4 IGRA+ and 1 IGRA-). The signature was enriched in IGRA+ individuals in the different groups with significant variation within the groups. Two thirds of the 9 individuals with culture confirmed TB had very high baseline risk scores which amplified further after diagnosis, whereas the remaining 3 participants had a low-risk score from baseline to the time of TB culture confirmation. These household contacts comprised of individuals that did not express the signature, individuals with low levels of the signature and participants who expressed the signature at baseline and beyond the 4 months observation period.

Application of scRNAseq in the discovery of TB biomarkers has the potential to significantly refine our search for the biosignature as we can capture enrichment or depletion of cell types in a bias free manner. We speculate that characterizing disease specific tissue at single cell resolution provides us a good chance of uncovering subtle transcriptomic aberrations that can form the basis of a TB marker. This is important for probing the molecular mechanisms that lead to overabundance of the biosignature. Histo-cytometric analysis localizes particular biomarkers without compromising tissue architecture [68]. This is critical when investigating markers whose subcellular location provide information about activation, such as the

transcription factor nuclear factor B (NF-κβ). This transcription factor only becomes active when it is translocated to nucleus after its inhibitor, Iκβ is degraded [69].

In the clinical setting, where individuals with TB associated complications are undergoing lung resections, lung tissue as well as PBMC can be used to explore signatures observed from different studies. Cai and colleagues performed scRNAseq on PBMCs from healthy, latently infected and active TB cases on a droplet-based platform 10x Genomics [70]. The data from the study showed that active TB corresponds to loss of NK cells, which are restored with treatment. This observation was verified using flow cytometry and thus serves as a platform on which studies with mechanistic objectives can be pursued. NK populations across the spectrum of TB disease should be investigated for disease specific perturbations that can form the basis of a biosignature. NK cells have memory markers from previous antigen exposure, thus they mediated both innate and adaptive immune responses [71]. It was interesting to note that in the same study, all the participant groups showed varying levels of a subset of myeloid cells with stromal cell markers (*COL1A1*, *COL1A2)*, associated fibrocytes or fibroblasts in the lung [72]. Fibrocytes are derived from the bone marrow and serve as a progenitor for mesenchymal cells in lung disorders such as fibrosis [73]. They can differentiate into fibroblasts or myofibroblasts once they reach the tissue. It has been suggested that they drive the formation of the fibrotic foci, interstitial pneumonia reported in idiopathic pulmonary fibrosis (IPF) [73]. A study comparing IPF and control lungs using scRNAseq revealed that IPF lungs are enriched for myofibroblasts [74]. Due to the association of TB and fibrotic pulmonary impairment, it can be postulated that characterization of fibrocytes and myofibroblasts genes has the potential to provide a TB-associated fibrosis tissue biomarker.

1.4 Phenotyping tissue resident CD4 T cells

Bacille Calmette-Guérin (BCG), is the only licensed TB vaccine [75], with variable effectiveness across studies. It is effective in preventing TB meningitis in infants [76]. TB specific immune responses of blood cells are used as markers of efficacy, even though these immune responses are not protective [77]. A study compared the efficacy of oral and intravenous BCG vaccination by measuring the TB specific responses in mice [78]. The animals were challenged with aerosolized tuberculosis and the accumulation of CD8 T cells in infected tissue correlated with protection. The ability of CD4 T cells to produce interferon gamma was correlated to bacterial load. Another study tracked the blood intracellular cytokine profile of infants at the time of birth and 24 months after BCG vaccination [79]. Comparison of infants who developed culture confirmed TB and those with TB household contacts (with no disease) showed no differences in the frequencies TB specific CD4 t cells, CD8 t cells and γδ t cells. Such studies support the fact that the frequency and cytokine profile of circulatory t cells does not correlate to protection. These observations have shifted attention to the study of immune responses at the primary infection site (lung), where t cells have been shown to be critical for prevention of both primary and secondary TB [80,81]. Different subsets play a role in maintenance of long term immunity by recall of protective responses [5]. The full extent to which these subsets participate in controlling pathogens is yet to be fully explored. When T cells are exposed to an antigen, they proliferate and differentiate into central memory (Tcm), effector memory (Tem) and tissue resident memory (Trm) subsets [82]. For the immune system to be effectively primed, a memory phenotype should persist where the infection occurred and this can be achieved by the Trm subpopulation [83]. Lung tissue resident memory cells are specifically characterized by the expression of CD69 (which interacts with S1PR1 to prevent cells joining circulation) [84] and CD103 (which associates with integrin anchoring them on the epithelial barrier) [85], CD49a(adhesion) , CXCR6 (tissue localization), CD101(proliferation and activation), PD-

1(immune checkpoint) [77]. There are exceptions to these categorisations as highly activated CD69+ T cells in the circulatory system [86] and CD69- tissue resident T cells have been demonstrated [87]. Other researchers have obserbed that Trm populations might not permanent in the lung [88]. The phenotype and activity of Trm is highly dependent on the microenvironment,thus more research into lung Trm is necessary in the context of TB vaccines [89], considering that HIV associated depletion of CD4 T cells is associated with high incidences of TB disease [90].

Comprehensive studies on the transcriptional and functional capacities of Trm's have revealed a distinct phenotype, supporting the hypothesis that there are developmental/maintenance programs required for the persistence of this population [82]. In a study on human lung resections, Oja and colleagues analysed the transcriptional profile of both CD4 and CD8 Trm's together with matched blood. The gene expression of both Trm populations revealed chemokines (*XCL1,CXCL16,CCL4*), chemokine receptors (*CXCR3,CXCR6*), adhesion (*ICAM1,ITGAE,CD97*), effector (*GZMA,GZMB,PRF1*), immune checkpoint (*CTLA4*). Interestingly, the expression profiles of CD4 and CD8 Trm's were largely indistnguishable, highlighting the influence of the tissue micro-environment. Tissue resident t cells also expressed transcripts for *NOTCH1* and *EGR2*, which are associated with long term persistence of pathogen specific Trm [91]. Pathways associated with both genes may provide a therapeutic modulation route for a clinical intervention. The mechanisms by which this occurs are yet to be elucidated, but deletion of Notch1 and Notch2 in influenza infected mice resulted in fewer lung tissue resident T cells [91].

Single cell RNA sequencing (10X Chromium) was used to profile 6311 flow cytometry sorted tumour infiltrating lymphocytes from primary and metastatic breast cancer tumours, revealing

a cytotoxic CD8 Trm subset that was associated with favourable patient outcomes [92]. The same Trm signature was highly enriched in melanoma patients positively responding to immune inhibitor blockers at the initial stages of the treatment [93]. The Trm signature was composed of cytotoxic (*GZMB, PRF1*), adhesion (*VCAM1, ITGAE*), immune checkpoint (*HAVCR2, PDCD1, CTLA4, TIGIT, LAG3*), infiltration (*LAYN, KLRC1*), chemokine (*CXCL13, CCL3, CCL4*) and proliferation (*MKI67*). Another group profiled human lung memory t cells in HLA-disparate transplant donor/recipient pairs, sampling broncho alveolar lavage fluid (BALF) over a 15 month period demonstrated that donors with a persisting Trm signature where associated with lesser incidents of graft rejection [94]. Analysis of airway t cells showed that the donor t cells maintained the Trm phenotype, and the recipient lymphocytes gradually increased the expression of tissue residency markers. It is of interest to uncover how the Trm signature changes in the TB^-/HIV^-, TB^+ and TB^+/HIV^+ lung tissue, particularly by comparison of CD4+CD69+ lung T cells, CD4+CD69- lung T cells and CD4+CD69- blood T cells using scRNAseq (derived from matched individuals). Such an approach will allow profiling of these populations, stratifying them across a developmental trajectory of their cellular states using methods such as pseudotime or monocle [95]. Uncovering potential CD4 t cell subsets which are depleted with HIV infection will assist in further understanding the mechanisms of immune deficiency. A recent study suggested that CD4 t cell independent mechanisms play a role in suppression of latent tuberculosis reactivation in simian immunodeficiency virus (SIV) infected macaques [96].

1.5 Conclusion

Use of cutting-edge technology has the potential to advance our understanding of how to manage and ultimately eradicate TB disease. This will be achieved by robust approaches to uncover mechanisms that lead to failure of the immune response, as well as identifying molecular markers uniquely linked to this pathology. With the steady reduction in the cost of

running sequencing platforms, there is hope that high TB burden regions can begin large scale studies which will accelerate generation of data driven disease insights.

1.6 Research Problem and Significance

1.6.1 Statement of the Problem

Tuberculosis disease afflicts a significant portion of the global population and South Africa caries one of the highest burdens. Many aspects of the human immune response to TB infection are yet to be characterized, slowing down the development of new diagnostics, vaccines and therapeutic agents. Thus, for us to develop new solutions to address this endemic, more research is required which incorporates TB compromised tissue and cutting-edge technologies. One such technology is single cell ribonucleic acid sequencing (scRNAseq)

1.6.2 Hypothesis

TB related pulmonary pathology such as cavitation and fibrosis lead to the enrichment of lung tissue with cells involved in tissue destruction and collagen deposition, resulting in the skewing of stromal, immune and parenchymal cell types as the lung loses normal function.

1.6.3 Research Objectives

To date, the only TB study that leveraged scRNAseq compared peripheral blood mononuclear cells (PBMCs) from infected individuals to uninfected control groups. Whilst this study yielded informative insights such as initial depletion of NK cells, the pulmonary consequences of TB infection must be explored. The goal of this thesis is to characterize the different populations of cells enriched in TB and/or HIV infected tissue (1), explore correlations with already reported TB biomarkers (2), characterize tissue resident CD4 T, critical for a long-term memory response (3) and demonstrate the applicability of the dataset in other pulmonary diseases. To this end, several specific objectives were derived:

1.6.3.1 To define the cellular diversity in human lung tissue from individuals infected with tuberculosis (TB) and human immunodeficiency virus (HIV).

1.6.3.2 Explore if the tuberculosis blood transcriptional signatures can be localized and assigned to cell types at the site of TB disease using single cell sequencing.

1.6.3.3 Phenotype tissue resident CD4 T cells from TB/HIV coinfected lungs to assess their functional potential.

1.6.3.4 Demonstrate the usefulness of scRNAseq data in identifying potential target populations of SARS-CoV-2.

1.7 Research Methodology

1.7.1 Human Participants

Human lung tissue and blood samples were obtained from patients undergoing corrective surgery with previous TB episodes and other comorbidities such as (haemoptysis, cavitation, bronchiectasis, shrunken or collapsed lung). The surgery was done at the Department of Cardiothoracic Surgery at King Dinuzulu hospital in Durban, KwaZulu Natal and Inkosi Albert Luthuli Central Hospital in KwaZulu-Natal . We were unable to culture TB from the tissue and we suspect that low bacterial load and the growth suppressing influence of anti TB therapy led to this observation. All samples were collected with approval from the Biomedical Research Ethics Committee and written informed consents obtained from all subjects (BREC no 019/13)

1.7.2 Blood Processing

Blood was collected in BD Vacutainers (sodium heparin, BD), peripheral mononuclear cells were isolated using the ficoll-histopaque (Millipore Sigma) density gradient centrifugation method.

1.7.3 Lung Tissue Processing

The Lung tissue was processed within 5 hours of receipt as described [52]. Briefly, a piece of the lung tissue was cut for histology and placed in 4% Paraformaldehyde (PFA). The remaining tissue was dissected into small pieces (5x5x5 mm) and infiltrated with a collagenase (Sigma-Aldrich), DNase 1 (Sigma-Aldrich) in RPMI (Sigma-Aldrich) with 10% FBS (Hyclone) for 30 minutes. Mechanical digestion at room temperature using the Gentle MACS (Miltenyi Biotec) followed by agitation at 37°C for 30 minutes on a rotor ensued. The mechanical digestion and agitation were repeated once more, followed by filtration of the resulting cellular suspension using the 70 μm (Corning) and 40 μm (Corning) strainer. This was followed by the lysis of red blood cells. Cells were then stained with tryphan blue (Thermo Fischer) and enumerated using an automated cell counter (BioRad) or a manual counter (Kova).

1.7.4 Cell Staining Procedure for Flow Cytometry

The single cell suspensions were centrifuged at 800g for 5 minutes and the supernatant was discarded. The remaining pellet was stained with a monoclonal antibody cocktail containing Live/Dead (Life Technologies), CD45 (clone HI30, BD Biosciences), CD3 (clone UCHT1, BD Biosciences), CD4 (clone OKT4, Biolegend), CD8 (clone RPA-T8, BD Bioscience) , CD19 (clone SJ25C1, BD Bioscience) , CD69 (clone FNS0, Biolegend) , CD103 (clone Ber-ACT8, Biolegend), CD45RA (clone HI100, BD Bioscience), CCR7 (clone G043H7, Biolegend), PD-1 (clone EH12.1, BD Bioscience), CD154 (clone 24-31, Biolegend), CD27 (clone O323, Biolegend), CXCR3 (clone 1C6/CXCR3, BD Bioscience), CD25 (clone BC96, Biolegend). The cells were incubated for 20 minutes in the dark after which they were washed 2 times with PBS (Sigma-Aldrich) and suspended in 500 μl PBS containing 0.1% BSA (Separations). The CD4 T cell tissue resident population was isolated using the 5 laser fluorescence-activated cell sorting (FACS) Aria Fusion (BD Biosciences), 80μm nozzle and 40 psi pressure. FACS allows for the separation of heterogenous cells based on different light scattering and fluorescence

characteristics in a fast and quantitative manner. The gating strategy used was as follows : singlets CD45+CD3+CD19-CD4+CD69- or singlets CD45+CD3+CD19-CD4+CD69+. It should be noted the cells were sorted directly after the recovery of tissue suspension, as we had observed cell loses when freezing down for long term storage in pilot experiments. The flow cytometry data was analysed using FlowJo version 9.7.6 (TreeStar).

1.7.5 Single cell Rna sequencing (scRNAseq)

Seqwell was implemented as described [97]. Briefly, the single cell suspension was diluted to 15,000 cells in 200µL of RPMI (Sigma-Aldrich) + 10% FBS (Hyclone) and loaded onto a polymethylsiloxane (PDMS) array pretreated with the same solution for 15 minutes. The cells were allowed to settle into the microwells by gravity (by incubating for 20 minutes on a flat surface) and the array was washed with PBS (Sigma-Aldrich) and sealed with a plasma functionalized polycarbonate membrane (Sterlitech). The arrays were incubated at 37˚C for 40 minutes followed by a 20-minute incubation in lysis buffer containing guanidium thiocyanate (Sigma-Aldrich), EDTA (Thermo Fischer), 1% beta-mercaptoethanol (Sigma-Aldrich) and sarkosyl (Sigma-Aldrich) at room temperature. The arrays were then transferred to a hybridization buffer containing NaCl (Thermo Fischer), $MgCl_2$ (Sigma), 1X PBS (Thermo Fischer) and polyethylene glycol (Sigma-Aldrich) and were gently shaken at 60rpm for 40 minutes. The capture beads hybridized with the released mrna from the lysed cells were collected from the array by a series of wash steps with wash buffer containing NaCl (Thermo Fischer), $MgCl_2$ (Sigma), Tris-HCl (Thermo Fischer) and Water (Inqaba Biotech) . This was followed by centrifugation at 2500g for 5 minutes each iteration. The beads were resuspended in a master mix for reverse transcriptase containing Maxima H Minus Reverse Transcriptase, Maxima Buffer, dNTPs, RNAse inhibitor, a template switch oligonucleotide and PEG for 30 minutes at room temperature and overnight with end-to-end mixing at 52˚C. This was followed

by the standard exonuclease digestion and denaturation of cDNA hybridized to the beads by a 5-minute incubation in NaOH (Sigma-Alrich) and wash step with a solution containing Tris-HCl, EDTA and Tween-20 (Thermo Fischer). The beads were resuspended in a master mix containing Klenow Fragment (NEB), dNTPs, PEG and the dN-SMRT oligonucleotide, incubating for 45 minutes at 38˚C. PCR was performed as described in the protocol and the product was subjected to 2 rounds of AMPure XP SPRI (Agencourt) bead cleanup at 0.6x and 0.8x volumetric ratios respectively. The library size was analyzed using an Agilent Tapestation hsD5000 kit, ensuring that the expected product had an average size of ~1000bp and the absence of primer dimers especially below 200bp. The Qubit High Sensitivity DNA kit was used to quantify the libraries and they were prepared for Illumina sequencing using the Nextera XT DNA Sample Preparation kit (the sequencing and alignment was done by the Shalek Lab). A total of 900pg of the different libraries were added the tagmentation reaction. The amplified product was purified with the AMPure XP SPRI beads (0.8x ratio) and the libraries were pooled for loading. The libraries were sequenced on the NovaSeq 6000 using paired end read structure with custom read 1 primer: read 1:20 bases, read 2 : 50 bases, read 1 index: 8 bases.

1.7.6 Single cell rna sequencing (scRNAseq) data analysis

The sequencing data from the NovaSeq was aligned to the hg19 genome assembly and processed in accordance with the Drop-Seq Computational Protocol v2.0 (https://github.com/broadinstitute/Drop-seq). We used STAR alignment with the default parameters (genomeDir, runThreadN (n = 50), readFilesIn, mem = 100000) according to the alignment cookbook (https://github.com/broadinstitute/Drop-seq/blob/master/doc/Drop-seq_Alignment_Cookbook.pdf). On average, we observed saturation of sequencing reaction at 6000 detected genes and 50000 mapped reads. The data was then loaded to the Seurat R package v3.1.0 (https://satijalab.org/seurat/), transformed to $\log_e$(UMI + 1) followed by scaling by a factor of 10000. The overall quality was assessed by the distribution of reads, transcripts

and genes per cell recovered. Variable genes with an average expression > 0.1 log normalized UMI were used to compute the principal component analysis (PCA). The JackStraw function was used to identify 20 significant PCAs that were used for downstream analysis. For dimensionality reduction, we used a Uniform Manifold Approximation and Projection (UMAP) at "min_dist" of 0.5 and "neighbours" set to 30. Unsupervised clustering using the FindClusters was used to identify transcriptionally similar cells with parameters k.param set to 10 and resolution set to 0.5. The clusters were subclustered by using a differential expression test (FindMarkers implemented in Seurat, setting "test.use" to "bimod", Bonferroni-adjusted p value cutoff < 0.001). The cell types were annotated by cross-referencing canonical cluster defining genes with well curated lists, online databases such SaVant (http://newpathways.mcdb.ucla.edu/savant-dev/) and GSEA/MsigDB (https://www.gsea-msigdb. org/gsea/msigdb/index.jsp).

We used Monocle 2 to do trajectory analysis of a population of fibroblasts that we sub-clustered from our main dataset. Briefly, the program allowed us to compute pseudotime gene expressions changes. It used an inbuilt package called scEpath to divide the pseudotime into 10 different bins, then the expression of each gene was estimated by the trimean expression of the gene across all cell types. The genes were smoothened using the cubic regression splines. To identify pseudotime-dependent genes, we used the standard deviation of these genes and compared them to a 1000 randomly permuted genes per cell [98].

1.7.7 Lung tissue Histology

Multiplex fluorescent immunohistochemistry staining was performed on lung tissue sections using the Opal™ 4-Color Manual IHC Kit 50 Slides (PerkinElmer, USA) as directed by the manufacturers (a total of 3 primary antibodies and DAPI per run). For example, in a later section in chapter 2, we stained with 3 primary antibodies of interest (MMP1, MMP3 and COLLAGEN 1) , using the blue channel foDAPI for cellular localisation. The proceeding

description documents the sequential staining of the 3 primary antibodies and DAPI on a single slide. Briefly, lung tissue samples fixed in 4% formalin were paraffin-embedded. 4μm sections were cut on glass slides, allowed to dry for a minimum of 24 hours and the slides were baked at 60^{o}C overnight. Then, the combined process of deparaffinization, rehydration and antigen retrieval of the tissue sections was done using 1x Envision Target Retrieval Solution, High PH (Dako) in the PT-Link Pre-Treatment instrument (Dako). Thereafter, slides were incubated for 1 minute in distilled water and equilibrated in EnVision FLEX Wash Buffer (Dako) for 5 minutes at room temperature. Then, the slides were incubated in Peroxidase blocking solution (PerkinElmer) for 10 minutes and washed in wash buffer (Dako) immediately at room temperature. The slides were then incubated in Bloxall blocking solution (PerkinElmer) for 10 minutes, and then in primary antibody-1 for 30 minutes at room temperature. The platform allowed 3 primary antibodies of interest and reserved the blue channel for the nuclei stain DAPI useful for cellular localisation. For example, in a later section in chapter, we stained with 3 primary antibodies of interest (MMP1, MMP3 and COLLAGEN 1) , using the blue channel for DAPI. Slides were then washed for 5 minutes in wash buffer and incubated in Secondary Opal Polymer Horseradish Peroxidase (HRP) Mouse and Rabbit (PerkinElmer) for 30 minutes. Please note that this Opal Polymer HRP is recommended for human tissue sections with a mouse or rabbit primary antibody. Then, the slides were washed twice in wash buffer, drained and the sections were incubated in Opal Fluorophore (PerkinElmer) working solution for signal amplification at room temperature for 10 minutes. The slides were then washed for 5 minutes in wash buffer at room temperature. Afterwards, the antibody stripping via microwave treatment was done by placing the slides in a slide jar with pre-warmed buffer AR6 (PerkinElmer). The jar was loosely covered and placed in a microwave for 2 minutes at 100% power, 10 minutes at 50% power and 5 minutes at 20% power. Slides were cooled down in the dark by placing the slide jar on ice for 20 minutes and the slides were rinsed in distilled water,

followed by incubation in the wash buffer for 5 minutes to equilibrate slides. The microwave step strips the primary-secondary-HRP complex allows for the introduction of the next primary antibody. For the detection of the next target (primary antibody 2), the protocol was restarted at the blocking step using Bloxall blocking solution (PerkinElmer) for 10 minutes. After the third target was detected (primary antibody 3), a working solution of DAPI (PerkinElmer) was applied to the sections as the nuclear counterstain for 5 minutes in a humidity chamber. The slides were washed in wash buffer for 5 minutes, then in distilled water for 5 minutes and drained. Then, the sections were coverslip with Fluorescence Mounting Medium (Agilent Technologies, Inc.) and the edges of the coverslip were sealed with nail varnish. Slides were stored in a humidity chamber at 4^{o}C until images are acquired.

The unconjugated primary antibodies used are Anti-Collagen I (clone: ab34710, Abcam), Anti- Anti-CTHRC1 (clone: ab85739, Abcam), Anti-TDO2 (clone: OT14G2, Thermo Fisher Scientific), Anti-MMP9 (clone: EP1254, Abcam), Anti-PI15 (clone: PA5-52312, Thermo Fisher Scientific), Anti-TGFBR3 (clone: 1C5H11, Thermo Fisher Scientific), Anti-GAS1 (clone: 56-087, Thermo Fisher Scientific), Anti-EGFL6 (clone: PA5-51642, Thermo Fisher Scientific), Anti-CXCL13 (clone: H00010563bo2, Thermo Fisher Scientific), Anti-Carbonic Anhydrase 12/CA12 (clone: EPR14861, Abcam), Anti-TIMP3 (clone: AA-170-188, Thermo Fisher Scientific), Anti-MMP1 (clone: 3B6, Thermo Fisher Scientific), Anti- MCP-1/CCL2 (clone: 2D8, Thermo Fisher Scientific), Anti-MMP3 (clone: SB14d, LSBio), Anti-ACTA2 (clone: 1A4, LSBio). The primary antibodies were diluted in antibody diluent (PerkinElmer) as recommended by the antibody manufacturer, and the Opal fluorophores diluted in amplification diluent (PerkinElmer). The fluorophores used for signal generation in this study are FITC, Texas-Red and Cy5. Images were acquired on a Zeiss Axio Observer Z1 inverted microscope (Olympus) and analyzed with TissueFAXS imaging software (TissueGnostics).

Chapter 2: Cellular Heterogeneity of HIV/TB infected human lung tissue and cell type specific validation of TB biosignatures.

2.1 Introduction

A third of the global population is infected with *Mycobacterium tuberculosis*, with an estimated 10% developing active tuberculosis (TB) disease within their lifetime [2]. With 9 million new cases and 1,5 million deaths reported annually, TB remains one of the leading causes of death from an infectious agent. Availability of anti-TB drugs led to a cure rate of 85% in treatment of drug susceptible TB from 1995 to 2015 [99]. Despite such effectiveness, survivors of the disease often have pulmonary impairment and respiratory failure despite being culture negative [100-102]. HIV infection, a known risk factor for developing active TB, has been associated with less severe pulmonary impairment especially in individuals with low CD4 T cell counts, although 30% of the patients still presented with abnormal spirometry and severe bronchiectasis [103]. TB patients contribute substantially to global prevalence of chronic obstructive pulmonary disease (COPD) [104]; thus research focussed on the progression of pulmonary impairment from TB and the associated pathology can improve the treatment and subsequent quality of life of TB patients [105]. Another major challenge in TB control is the lack of biomarkers to capture the full spectrum of the disease (latent TB, incipient/subclinical TB and active TB), impeding the monitoring disease and patient outcomes [106]. There is need for better comprehension of how TB infection impacts the host immune system, particularly host immune cell composition and gene expression profiles [70]. The current approaches (bulk RNA sequencing or microarray) that have been used to generate TB specific biomarkers often have limited statistical power as the samples used are often orders of magnitude lower than the variables (gene expression), yielding 30% false positives after multiple corrections [107]. Thus,

there is a need to leverage different approaches to elucidate biologically representative TB biosignatures. In addition, these approaches have focused primarily on the peripheral blood, whilst the TB pathogen, for the most part, is restricted to the pulmonary system (except for extra pulmonary cases). Although consistent transcriptional differences have been observed in multiple studies [108], the biosignatures are dominated by leukocyte subsets, such as neutrophils [109], and are likely to reflect non-specific systemic effects of the innate response. Focusing on the host pathogen interaction in the lung may identify more subtle, disease specific differences. The advent of high throughput single cell RNA sequencing is improving our ability to analyse cell types, sub types and cell states [23]. This is critical when studying disease compromised tissue as it allows unbiased profiling of the cellular and functional consequences [46]. To illustrate this in the lung, scRNAseq generated a cellular atlas of pulmonary fibrosis, showing previously unknown skewing of alveolar macrophages and epithelial cells, linked with maintenance of the fibrotic phenotype [46]. ScRNAseq was also used to study differences in peripheral blood mononuclear cells (PBMC) from healthy, latent and active tuberculosis patients, showing an association between active disease and depletion of natural killer cells [70]. Whilst another study leveraged multimodal scRNAseq to profile memory t cells from a Peruvian cohort (n = 259) [110]. In human TB, to my best knowledge scRNAseq is yet to be applied to probe disease driven alterations in the lung. This due to the scarcity of fresh TB infected lung tissue or the ability to single cell sequence the material biosafety level 3 conditions.

Here, I characterize the cellular composition of the human lung tissue from individuals infected with *M. tuberculosis* and/or HIV together with non-TB cancer controls undergoing surgery due to pulmonary complications. We sequenced the cells using a high throughput scRNAseq platform known as Seq-well (S^3) [97]. I adapt this low-tech platform for the BSL3 laboratory and use it to generate a molecular atlas of the infected lung. This is then used to (1) determine the

cellular make up of TB diseased lung tissue, (2)identify cell cellular subsets involved in TB immunopathology, and (3) localize TB biosignatures to cell types. A larger proportion of the individuals in the study were co-infected with HIV, providing preliminary insight into the impact of HIV on the TB lung.

2.2 Results

2.2.1 Single Cell RNA sequencing of human lung tissue

The lung tissue was obtained from 13 participants (HIVTB (n=9); TB (n=2)); and 4 non-TB cancer controls (1 with HIV) undergoing lung resection surgery due to pulmonary complications ranging from fibrosis, cavitation, bronchiectasis, haemoptysis, adenopathy, nodules and cancerous nodules (Table 1). Following surgery, the lung tissue was homogenised into a single cell suspension using an optimised protocol [52]. After adjusting the cell concentration using a manual counter, 15000 cells were loaded onto a Seq-well microarray, preloaded with mRNA capture microspheres, in a bio safety level 3 laboratory and processed as described in the method section [97]. Data was aligned to the human genome, then subjected to rigorous quality controls to remove poorly sequenced cells and doublets. Samples were collected over a 3 year period. For participants, the libraries were generated with fresh tissue immediately after receiving the tissue to preserve the native transcriptional .

All samples were corrected for batch effects using harmony batch correction [111], prior to downstream analysis. RNA transcriptomes from 20962 cells were analysed using the seurat package[112]. To identify clusters of cells with similar transcriptomes we used uniform manifold approximation and projection (UMAP) shown in (Figure 1A). Segregation of cells by disease status or participant ID indicated the existence potential disease and/or participant specific .

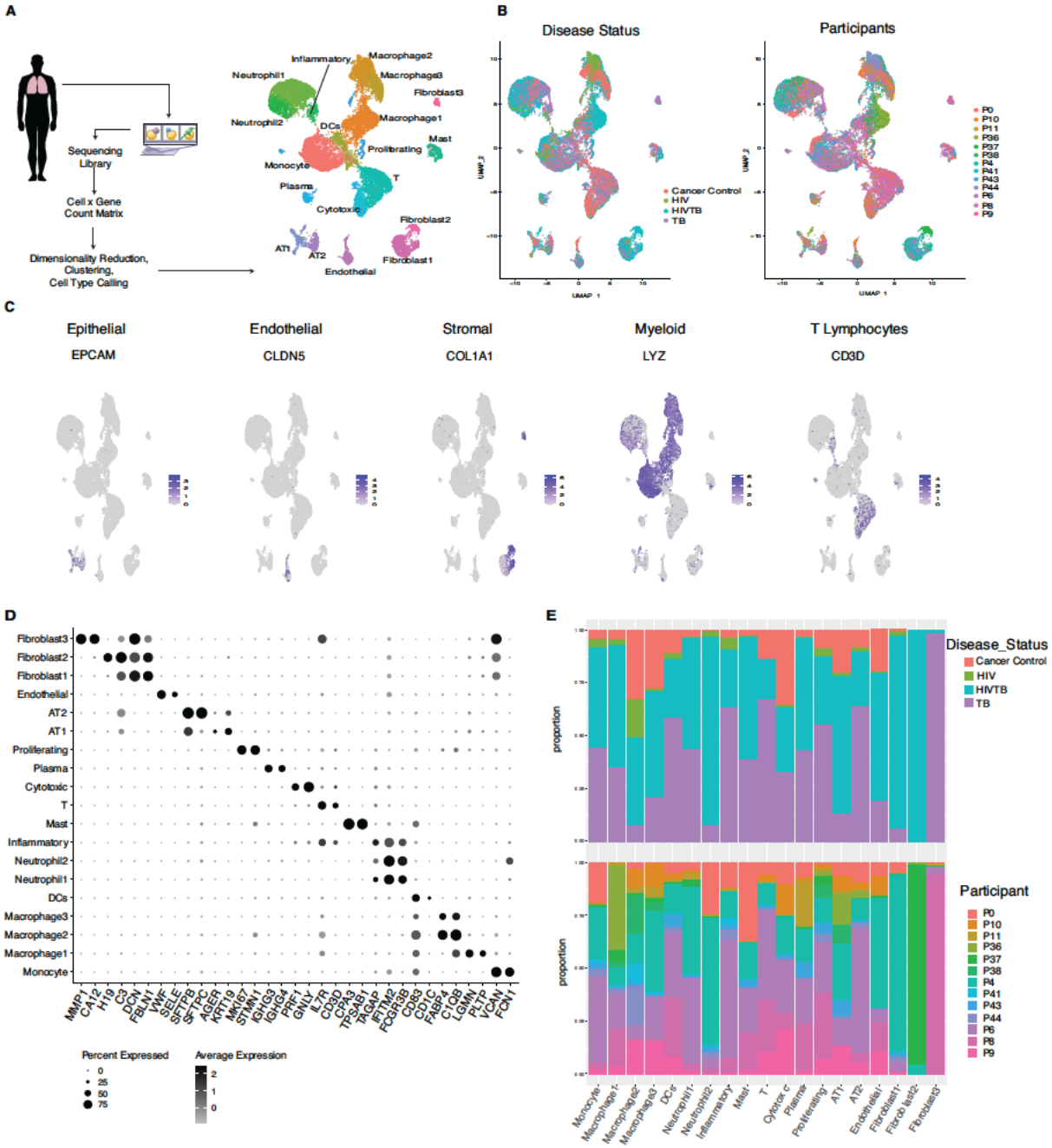


Figure 1: Overview of the lung tissue cells from TB and/or HIV and cancer patients.

(A) Schematic of the method used for the isolation of cells from human lung tissue, generation of single cell libraries using Seq-well (S^3) and the unbiased in silico analysis used to identify different cell types. Shown adjacent to the process flow is the UMAP projection of 20,962 cells from 13 donors (n = 7 HIVTB; n = 2 TB; n =1 HIV, n = 3 Cancer Control). The cells are represented by points and the colour represents the different cell types.

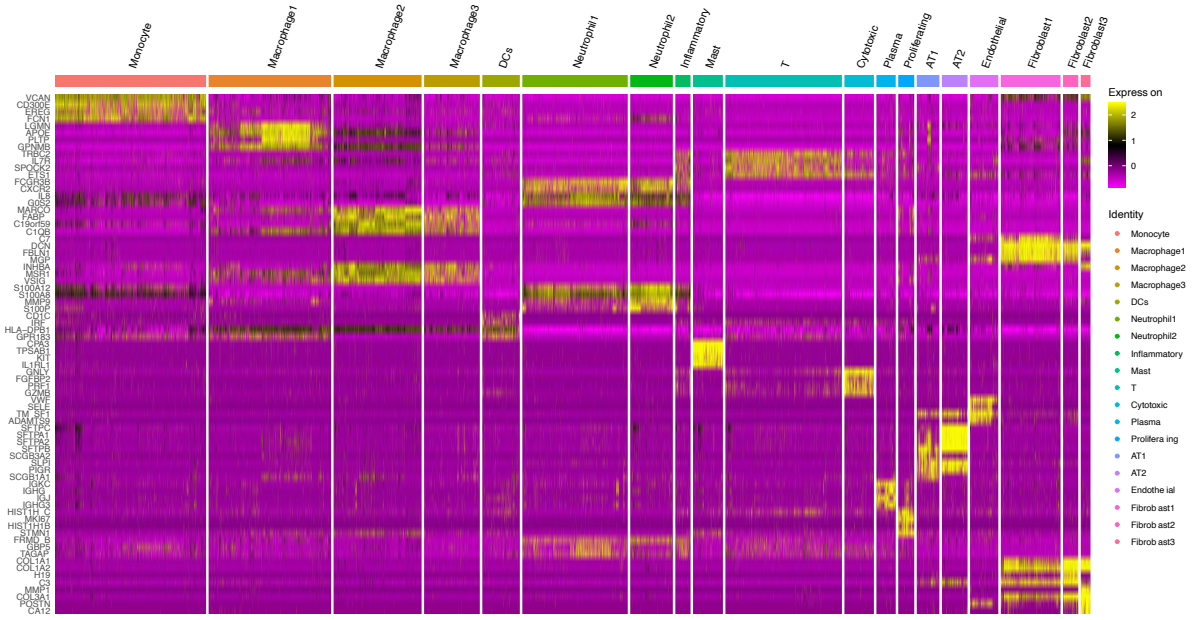
(B) UMAP projections of the cells, coloured according to the disease status or the corresponding study participant.

(C) Expression of marker genes for the cells in different tissue compartments (epithelial, endothelial, stromal, myeloid and T lymphocytes).

(D) Dot plot of a minimum of 2 canonical genes for each of the various cell types (FDR-adjusted $p < 0.001$, the size of the dot represents the proportion of cells within a cluster that express a particular gene, with the intensity of the dots representing the binned count-based expression (log scaled UMI + 1).

(E)Bar plots showing the fractional distribution across cell types as a function of the disease status and participants.

effects in the relative frequency of cells, with minimal evidence of technical artefacts (Figure 1B). Unsupervised clustering of the cells revealed 19 distinct clusters, which were subsequently defined by making use of a combination of manual curation of canonical markers from previous studies and reference gene expression from the SaVant database [46,113]. As expected from complex tissue samples, we cells from the epithelial (*EPCAM*), endothelial (*CLDN5*), stromal (*COL1A1*) compartments, together with immune cells such as myeloid (*LYZ*) and lymphocytes (*CD3D*) (Figure 1C) [46]. Figure 1D shows the canonical genes expressed in the different clusters; monocytes (*VCAN, FCN1*), 3 macrophages subsets (*C1QB, FABP4, LGMN,PLTP*), dendritic cells (*CD83, CD1C*), neutrophil 1 (*S100A8, S10012*), neutrophil 2 (*ARG1, MMP9*), inflammatory cells (*TAGAP*), mast (*TPSAB1*), t cells (*CD3D*), cytotoxic cells (*GNLY*), plasma b cells (*IGHG4*), proliferating cells (*MKI67*), type 1 alveolar pneumocytes (*AGER*), type 2 alveolar pneumocytes (*SFTPA1*), endothelial cells (*VWF*) and 3 fibroblast subsets (defined by expression *of DCN, H19, MMP1*). Figure 1F shows the distribution of the top 4 genes per cell



(F)Heatmap depicting the relative expression (normalized and scaled) of the top 4 canonical marker genes of the 19 cell clusters.

We identified these clusters by making use of unsupervised clustering via the FindClusters tool within the seurat package with default parameters, k.pram set to 10 and resolution set to 0.5. The top4 genes shown were the canonical markers per cluster from the FindClusters output.

cluster in a heatmap. Expression of these transcripts correlated with protein expression data from previous human TB studies , such as elevated levels of *S100A9* and *STPA1*[114,115]. We observed a high degree of heterogeneity in the frequency of cell types identified per participant, which we expected due to varying clinical presentation (Table 1). Consistent with literature, we observed that a greater proportion of the t cells came from the TB only group. We believe this observation captured the cd4 t cell depletion associated with acute immune deficiency syndrome (AIDS) [116]. Interestingly, we observed an even distribution of cytotoxic cells (mixture of cd8 t and nk cells) in all disease groups, despite cancer patients contributing less than 11% of the total cells. Cytotoxic cd8 t cells infiltrate tumours in breast cancer patients, but the prognostic value of this phenotype is dependent on the density and location of these cells [117]. This suggests that our clustering captured disease driven skewing of cell populations. In conclusion, the identified clusters were very similar to cell types observed in studies that have performed scRNAseq on fibrotic lung tissue, specifically fibrotic tissue [46]. This data provides an overview of the stromal, immune and parenchymal subsets detected in TB lung tissue.

Table 1 : Patient metadata of selected lung study participants

Patient	Sex	Age	Previous TB	HIV Status	Pulmonary Complication	Surgery	Cell Yield
P0	Male	37	n/a	positive	Cavitation and Fibrosis	Right Upper Lobectomy	1751
P4	Female	49	2013	positive	Cavitation, Bronchiectasis, Nodules	Left Pneumonectomy	5486
P6	Female	21	2017	negative	Cavitation, Haemoptysis, Dyspnoea, Mycetoma	Right Pneumonectomy	5442
P8	Female	58	2007	negative	Haemoptysis, Cavitation, destroyed upper lobe	Left Pneumonectomy	2214
P9	Female	56	none	negative	Pulmonary Nodules	Rights VATS biopsy	1294
P10	Female	61	none	negative	Mediastinal Adenopathy	Rights VATS biopsy	654
P11	Female	32	2013	positive	Minor Haemoptysis	Left Pneumonectomy	363
P36	Male	41	2019	positive	Massive Haemoptysis	Left Pneumonectomy	1153
P37	Male	53	2019	positive	Left Chest pain	Left Bullectomy	632
P38	Female	58	2019	positive	Massive Haemoptysis	Left Pneumonectomy	552
P41	Male	45	none	negative	n/a	Left Upper Lobectomy	332
P43	Female	44	2019	positive	n/a	Right Pneumonectomy	392
P44	Female	36	none	positive	Minor Haemoptysis	n/a	697
P53	**Female**	**44**	**active TB**	**n/a**	**n/a**	**n/a**	**n/a**
P54	**Male**	**67**	**active TB**	**n/a**	**n/a**	**n/a**	**n/a**

*VATS-Video assisted thorascopic surgery
The patients are colour coded according to the disease status HIV (lime), HIVTB (blue), TB (purple), red (Cancer). It should be noted that we could not confirm TB disease with the time of surgery, so we relied on clinical history for the groupings.
P9 has a history of breast carcinoma, presented with invasive tumour
P10 was diagnosed with Stage 2a cervical cancer, with nodules and metastatic disease
P53 and P54 are individuals with well characterised, end stage TB granuloma (their lung tissue was not sequenced) that we used for the Immunohistochemistry. These individuals were selected based on their distinct TB pathology. We did not definitively ascertain the stage of TB development but assumed they represent post treatment chronic TB.

2.2.2 Skewed immune and pro fibrotic profile of compromised lung tissue.

To investigate the skewing of cellular subsets in TB lung, the relative proportion of each cell was calculated as a percentage of total assigned cells for each participant (Figure 2A). The frequencies of each cell subset were then compared between TB samples and non-TB controls, irrespective of HIV status. This analysis suggested a skewing of innate immune cells in TB disease, highlighted by a significant reduction of the macrophage 2 subset (p=0.01 by Kolmogorov-Smirnov; uncorrected), neutrophils (p=0.03, combined), and mast cells (p=0.04). The enrichment of mast cells in TB lungs mirrors a recent study applying the same technology to individual granuloma from TB infected non-human primates [118], with this cell type was enriched in granuloma with the highest *Mtb* burden. The same study also observed a significant association between high burden granuloma and plasma cells, which could not be shown in our data as the differences in cellular proportions were not statistically significant. The enrichment of neutrophils in the TB group is consistent with other studies that reported neutrophil infiltration in TB [119].

In addition to these, fibroblast subsets 2 and 3 were only observed in TB diseased individuals (8/9 vs 0/4; p=0.004). This is consistent with the pulmonary remodelling associated with tuberculosis disease due to cavitation, fibrosis and bronchiectasis [120]. To examine the lung architecture of the TB and non-TB samples, we performed haematoxylin and eosin (H &E) staining of paraformaldehyde fixed tissue. Figure 2B shows the 2 participants with pathology that captured the gross histological spectrum we observed. P11 presented with evidence of inflammation and fibrosis induced anatomical distortion. P6 displayed evidence of lung parenchymal haemorrhage, interstitial and confluent fibrosis.

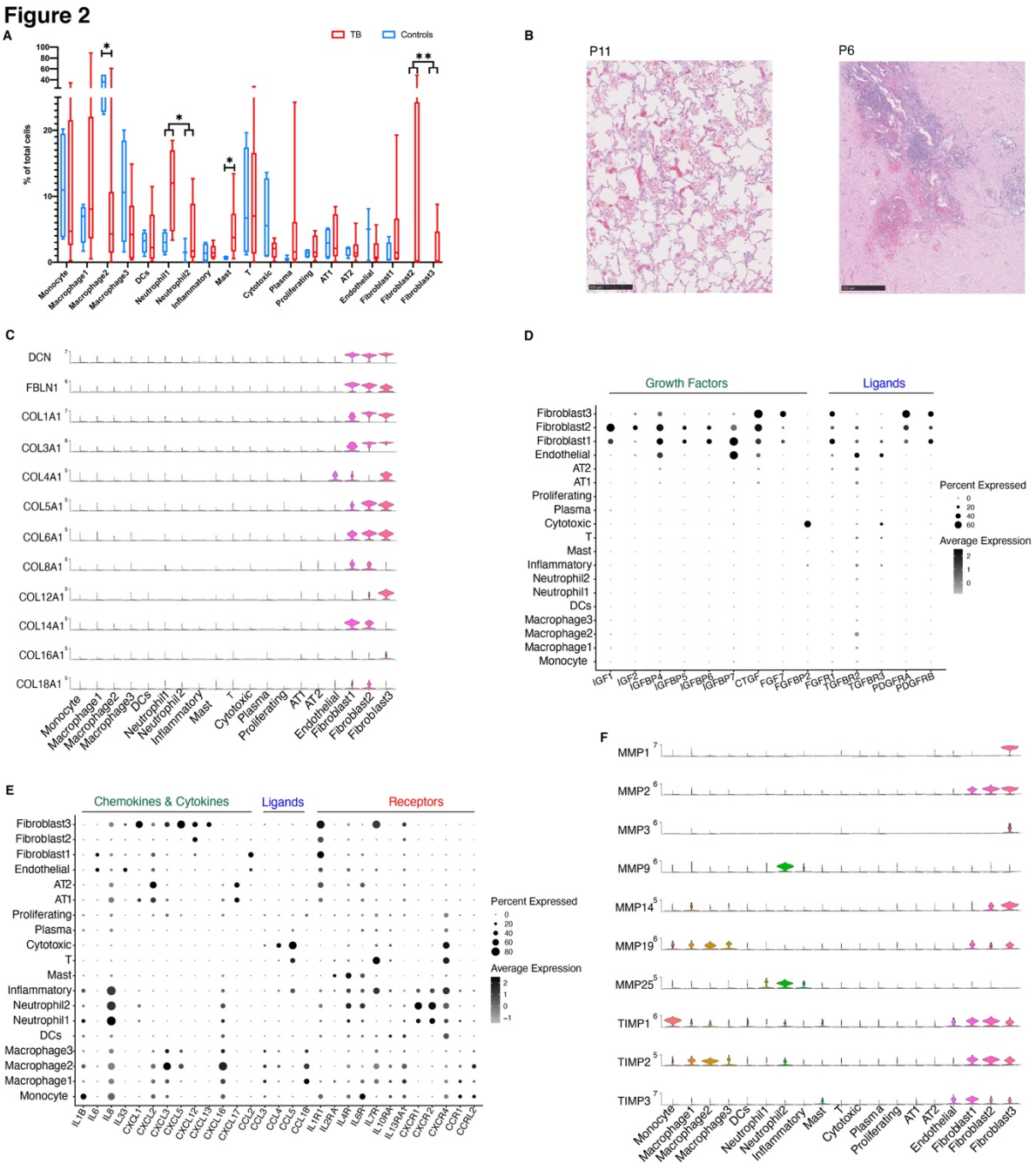


Figure 2: Spectrum of fibrotic lung tissue damage.

(A) The distribution of the different cell types when comparing TB infected tissue to cancer controls. For each cluster we performed a paired t-test between the 2 conditions and the * represent differences that were statistically significant. (B) Hematoxylin and Eosin images of lung tissue from Participant 11 (HIVTB) and Participant 6 (TB). The image from P11 show the initial stages of the fibrotic response in which alveoli are still visible whereas the image from P6 shows the terminal stage where the alveoli are infiltrated by fibrous tissue. Scale bars 500µm.
(C) Stacked Violin Plots showing Decorin (*DCN*), Fibronectin (*FBLN*) and Collagen (*COL)* expression across the clusters.
(D) Dot Plot showing expression levels of growth factors and ligands associated with a fibrotic response. The intensity of the dot corresponds to level of expression and the size of the dot shows the percentage of cells in the cluster expressing the gene.
(E) Dot Plot showing the relative expression of chemokines, cytokines, ligands and receptors in the lung cells. The intensity corresponds to the level of expression and the size to the proportion of cells expressing the gene in the cluster.
(F) Stacked violin plot showing the expression of Matrix Metalloproteinase (MMPs) and their inhibitors, Tissue Inhibitor of Metalloproteinases (TIMPs).

To investigate which cell types are involved in the fibrosis observed by histology, we assessed each cluster for the expression of genes associated with extra cellular matrix (ECM) synthesis such as decorin (*DCN*), fibronectin (*FBLN1*) and collagens (*COL1A1, COL3A1, COL4A1, COL5A1, COL6A1, COL8A1, COL12A1, COL14A1, COL16A1, COL18A1*). Consistent with their role in ECM production, these genes were highly restricted to the 3 fibroblast subtypes. The only notable exception being expression of *COL4A1* in endothelial cells, a form of collagen previously associated with lung endothelia [121]. Interestingly, the expression of these molecules varied greatly between these fibroblast populations, suggesting potential differences in fibrotic activity (Figure 2C). By contrasts *COL8A1* and *COL14A1*, are expressed by fibroblast 1 and 2, but not fibroblast 3, whereas *COL12A1* was highly expressed by fibroblast 3 [122].

Unregulated deposition of the ECM is associated with loss of lung function in patients with IPF [123]. Growth factors have been associated with fibrosis, utilizing autocrine signalling for epithelial cell development and paracrine signalling of fibroblast proliferation, migration and extracellular matrix synthesis [124]. We proceeded to investigate the data for cell types producing growth factors and their receptors. As shown in Figure 2D, fibroblasts were the major contributors of the growth factors and their high affinity, chaperone binding proteins involved in signal transduction insulin growth factor 1/ insulin growth factor binding protein 4 (*IGF1,IGFBP4*). The observations support robust autocrine capability in all fibroblast subsets via the fibroblast growth factor 7 / fibroblast growth factor receptor 1 (*FGF7/ FGFR1)* (false discovery rate (FDR)-adjusted p-value , $p < 1E-300$ and $p = 2E-96$ respectively) [125]. Strikingly, in line with the above observation, fibroblast 3 population expresses the highest levels of *CTGF*, which is known to be essential for fibrotic activity of transforming growth factor beta (*TGF-β)* [126] and is associated with IPF in humans [127]. Fibroblasts expressed the highest levels of *PDGFRA*, the receptor of *PDGF*, another important molecule in pulmonary fibrosis. Lineage

tracing experiments in the mouse model of IPF show platelet growth factor receptor alpha (*PDGRFA)* expressing fibroblasts are the key driver [128]. Interestingly, (*FGF7)*, which is also highly expressed in the fibroblast 3 subset, has been shown to be reduced in areas undergoing active remodelling, and inversely correlates with disease severity [129]. Overall, these data are consistent with the skewing of fibroblasts towards a profibrotic phenotype in TB patients.

2.2.3 Fibroblasts important drivers of immune cell recruitment and cavitation in TB disease

In general, fibrosis is thought to occur due to abberant wound healing associated with inflammation and angiogenesis [130]. To investigate this, we examined the expression of chemokines, cytokines, and their receptors. Of the transcripts detected within the dataset, fibroblasts expressed pro-angiogenesis chemokines *(CXCL1, CXCL2, CXCL3, CXCL5, CXCL12, CXCL13)* [43], the pro-fibrotic chemokine, (*CCL18)* [131], in addition to low levels of *IL-8* and *IL-33* (Figure 2E). *CXCL1* and *CXCL5* are known to play an essential role in the recruitment of neutrophils to the lung during TB infection [132]; whilst both *CXCL12* and *CXCL13* are involved in lymphocyte recruitment and formation of lymphoid follicles in the lung [133] [134]. We also observed within the fibroblast population 3 high levels of *IL7R (CD127)* transcripts, primarily associated with homeostatic signalling and cell survival in lymphocytes [135,136]. Fibroblasts and macrophages isolated from human subjects with rheumatoid arthritis have been shown to express high levels of *IL7R* [137].

Pulmonary cavitation is a hallmark of tuberculosis disease, as it facilitates parenchymal damage[138]. This process is not fully understood but is thought to be driven by the dysregulation of the matrix metalloproteinases (MMPs), enzymes that degrade the ECM, and their antagonists, tissue inhibitor of metalloproteinases (TIMPs). We therefore investigated the expression of these molecules in this dataset. Consistent with published data, we observed

several cell specific associations, including *MMP9, MMP25* with neutrophils, and *MMP19* with monocytes and macrophage [139]. Interestingly, *MMP1, MMP2, MMP3* and 14 were predominantly expressed by fibroblast (Figure 2F). In the case of *MMP1*, expression was only detected in fibroblast subset 3, in which subset it is ubiquitously expressed at a high level. Several studies indicate that *MMP1* is crucial for cavitation in human TB; bioengineering human *MMP1* into mice, which otherwise lack this key enzyme, leads to caseous necrosis [140], a canonical feature of human TB that missing from the mouse model. In humans, *MMP-1* transcript abundance was much greater in TB granuloma compared to Sarcoidosis granuloma. Sarcoidosis is a non-infectious granulomatous disease with none caseating lesions [141]. *MMP1* expression is also driven by hypoxic conditions generated within TB lesions [142]. In addition, the fibroblast 3 subset expressed lower levels of the three TIMPs and expressed very little *TIMP3*, the inhibitor with the broadest activity against MMPs [143].

2.2.4 Fibroblast subtypes and enrichment by disease status

Having observed the potential role of fibroblasts in TB immunopathology, we sought to investigate this cell type further. First, we sub-clustered all fibroblast subtypes (1792 cells), revealing 5 distinct clusters (Figure 3A). Of these, cluster 3 was uniquely expressed in individuals with TB alone (P6 and P8), and was associated with a distinct gene expression profiles (Figure 3B and C). Cluster 3 (207 cells) most closely resembled the fibroblast 3 population identified in the preceding sections, characterised by expression of collagen (*COL4A1 and COL12A1*), chemokines (*CXCL1, CXCL3, CXCL5 and CXCL12*), and *IL7R*. Focusing on the top 5 differentially expressed genes , cluster 3 showed elevated levels of *MMP1*, *MMP3*, CXCL*5* and *CXL13*, together with a specific repertoire of marker genes including *CA12*, *PDPN*, *TDO2 and FAP* (Figure 3B and Figure 3C). *TDO2* encodes tryptophan

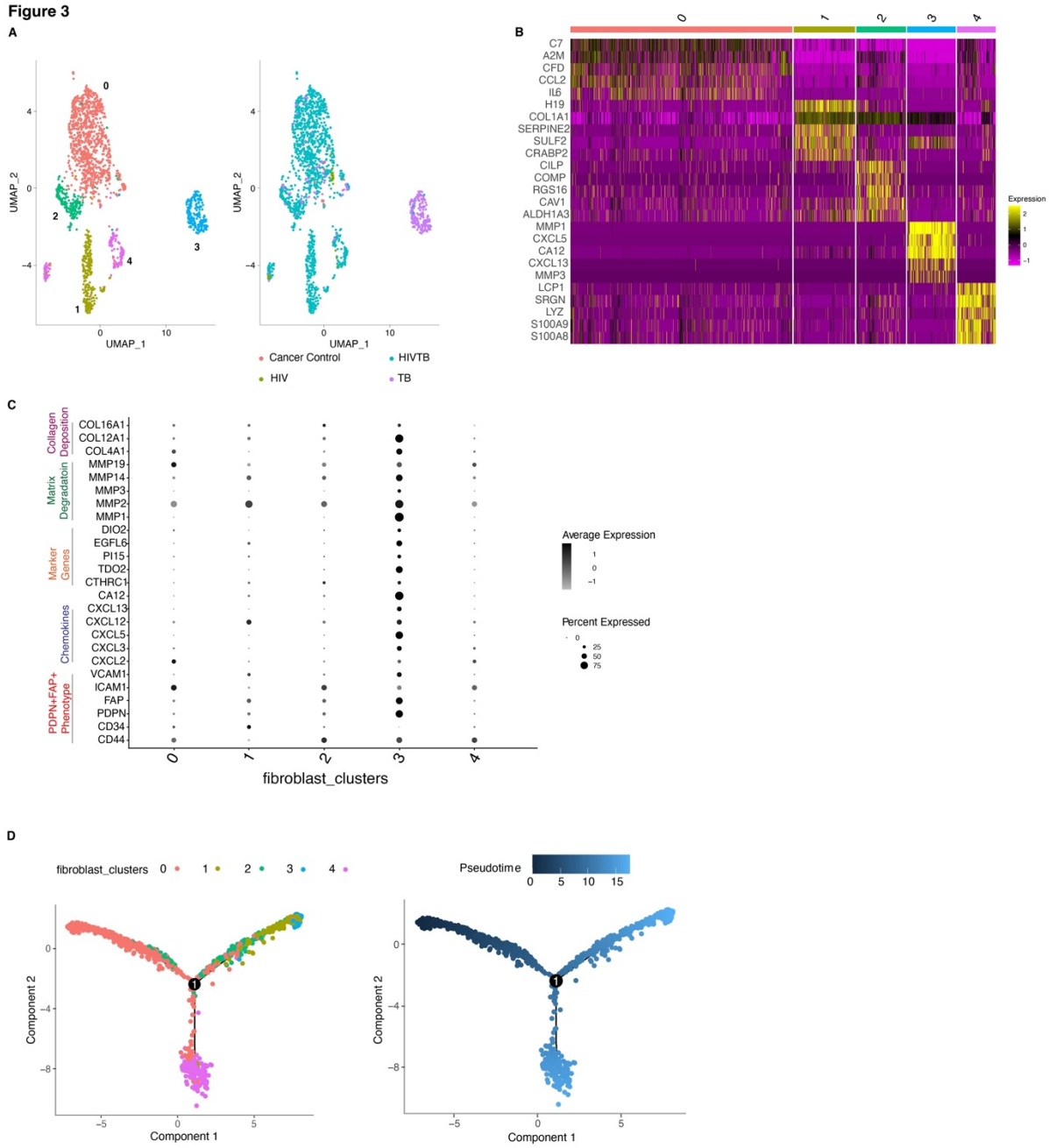


Figure 3: Fibroblast phenotypic diversity

(A) UMAP projections of 1792 fibroblasts (Fibroblast 1, Fibroblast 2 and Fibroblast 3 were combined and re-clustered), showing the distribution of the cells into generic clusters (left) and the distribution of the clusters based on disease status (right).
(B) Heatmap showing the top 7 marker genes for each of the generic fibroblast clusters.
(C) Dot Plot showing the relative expression of transcripts associated with immune-fibroblasts [144], chemokines, cell specific marker genes, extracellular matrix destruction and collagen accumulation. The intensity corresponds to the level of expression and the size to the proportion of cells expressing the gene in the cluster.
(D) Trajectory analysis of the 5 fibroblast clusters using monocle 2, showing bifurcation towards cluster 3 and 4. Most cells in cluster 3 and 4 occupy the terminal ends of each respective branch, suggesting that they are distinct cellular states.

2,3-dioxygenase, a rate limiting enzyme which, together with indoleamine 2,3-dioxygenase (IDO1), catabolizes the breakdown of tryptophan to kynurenine, which is involved in neurotransmission and immune regulation [145,146]. IDO1 is upregulated in the TB granuloma, where it is thought to play a key role in suppressing the T-cell immune response, and it's inhibition improves T-cell penetration in granuloma [147]. *TDO2* is upregulated in lung cancer associated fibroblasts and it's inhibition improves T-cell function [148]. Podoplanin (*PDPN*), together with fibroblast activating protein (*FAP*), are expressed by immune-fibroblasts and are required for organisation of tertiary lymphoid organs at sites of chronic inflammation, critical for establishment and maintenance of fibrotic foci [144].

The other clusters identified showed distinct gene expression profiles, including complement 7 (C7) and complement factor D (CFD) expression in cluster 0 (946 cells). CFD is associated with senescent fibroblasts [149]. Cluster 1 (262 cells) fibroblast uniquely express *(H19)*, a long non-coding RNA which induces proliferation of fibroblasts and is a potential driver of IPF [150] and *SERPINE2*, known to promote collagen deposition in scleroderma [151]. Cluster 2 expressed cartilage oligo matrix protein (*COMP*), associated with lung tissue stiffening whereas cluster 4 (165cells) had a myeloid-like expression profile (*LYZ, HLA-DRA, S100A9*). In an attempt to uncover the developmental association of these fibroblasts, we superimposed the 5 clusters onto a pseudo time trajectory using monocle 2 [95]. As illustrated in Figure 3D, the trajectory starts off with cells in cluster 0, which separates into two branches (at point labelled 1 in the Figure 3D), with one branch terminating at cluster 4 whilst the other terminating at cluster 3. This suggests that these two cell states represent committed paths along the developmental path of transcriptionally distinct fibroblasts.

To validate our observations, we explored the correlation of transcript abundance with protein expression, using immune fluorescence imaging fixed human tissue. We curated the markers that were part of the top 100 genes for fibroblast 3, selecting markers that were distinctly expressed in the fibroblast 3 cluster (*MMP1, MMP3, CA12, CTHRC1, GAS1, CXCL13, TDO2, PI15)*. *COL1A1* and *COL1A2* were selected to show areas of collagen deposition and *TIMP3* and *MMP9* were selected as their expression was observed in fibroblast 1/endothelial cell and neutrophils, respectively (Figure 4A). We stained the tissue with distinct evidence of fibrosis, cavitation and granuloma formation. Figure 4B-4E show representative staining of multiple granuloma-like structures from 2 participants (from the study database, cells were not sequenced). We observed that fibroblast 3 marker genes were highly expressed by cells that formed a circular cuff around necrotic/fibrotic zones, with distinct profiles associated with collagen 1 (Figure 4B-D). The localization of *CA12, TDO*, and *PI15* suggests that this fibroblast subset is associated with the organisation of TB granuloma. We further stained *MMP1* and *MMP3* on two separate cases, confirming abundant expression of these markers around the granuloma (Figure 5) and a stronger association with blood vessels surrounded by fibrotic tissue (Figure 6). The co-staining of these markers in concordance with our transcriptomic data supports the assertion that fibroblasts are expressing these proteins. Taken together, these findings suggest that diverse fibroblast populations play an important and understudied role in recruiting immune cells in the lung tissue, potentially organising the granuloma.

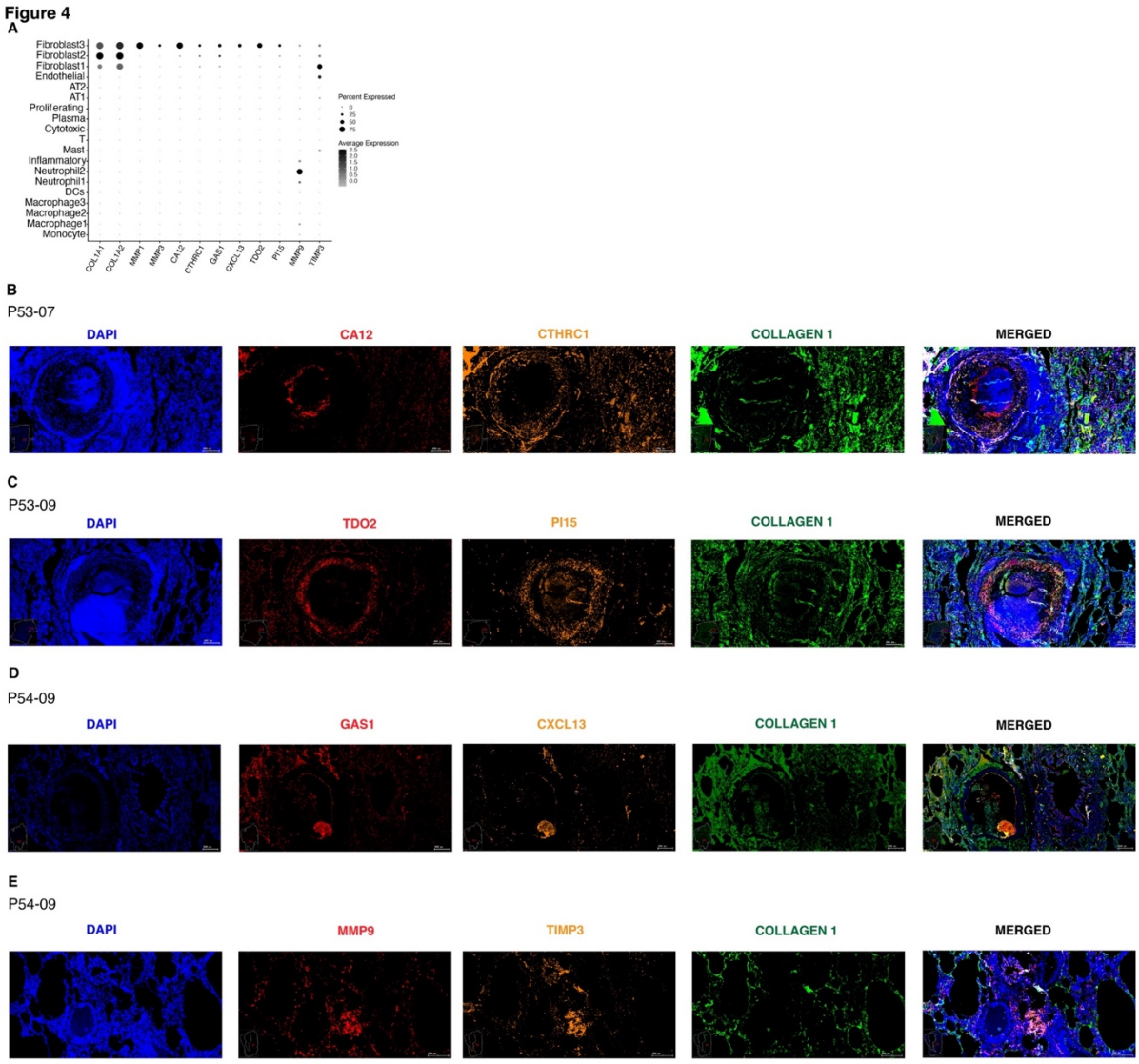


Figure 4: Immunohistochemistry of TB infected tissue.

(A) Dot plot showing the distribution of marker genes associated with fibroblast 3 (*MMP1, MMP3, CA12, CTHRC1, GAS1, CXCL13, TDO2, PI15*) across the data set, together with markers common for all fibroblast subsets (*COL1A1, COL1A2*), neutrophil 2 (*MMP9*) and fibroblast 1/endothelial cells (*TIMP3*).

(B)-(E) Representative fluorescent immunohistochemistry imaging of areas with severe fibrotic lesions in the human lung to visualize the localization of fibroblast 3 markers with collagen 1; (B) showing for P53-07 DAPI (blue), *CA12* (red), *CTHRC1* (orange) and COLLAGEN 1(polyclonal antibody for both alpha 1 and alpha 2) (green), it shows a fibrotic lesion from lung tissue; (C) showing a fibrotic lesion from P53-09 lung, DAPI (blue), *TDO2* (red), *PI15* (orange) and COLLAGEN 1 (green); (D) showing an airway adjacent to fibrotic lesion for P53-09 DAPI (blue), *GAS1* (red), *CXCL13* (orange) and COLLAGEN 1 (green). (E) showing for P54-09 DAPI (blue), *MMP9* (red), *TIMP3* (orange) and COLLAGEN 1 (green). The bars indicate 200µm for all images. These patients (P53 and P54) represented previously characterised cases of TB which showed distinct granuloma within the tissue. Due to the nature and timing of the sampling, we could not definitively determine the stage of granuloma formation.

Figure 5

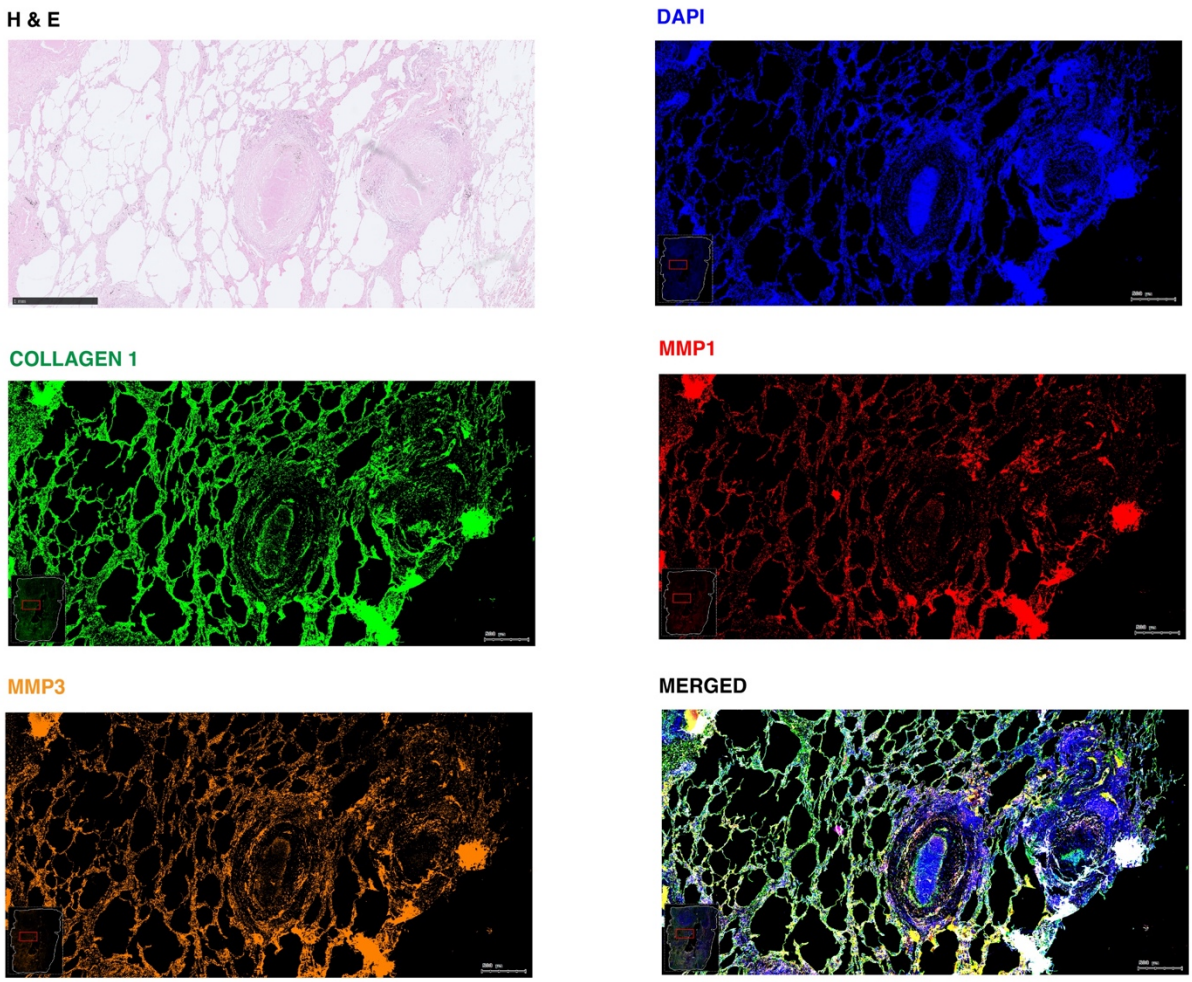


Figure 5: Immunohistochemistry of TB infected tissue showing colocalization of fibroblast 3 markers with TB granuloma-like fibrotic lesions.

Representative hematoxylin and eosin (H &E) images of lung tissue from participant 53 (culture positive TB) showing fibrotic lesions surrounded by thickened alveoli due to hyaline deposition. Scale bars 1mm. Fluorescent immunohistochemistry imaging of the fibrotic lesions in the human lung to visualize the localization of fibroblast 3 markers with collagen 1; DAPI (blue), COLLAGEN 1 (green), *MMP1* (red), *MMP3* (orange) and THE MERGED image for all 4 colours, scale bars 500µm for all images.

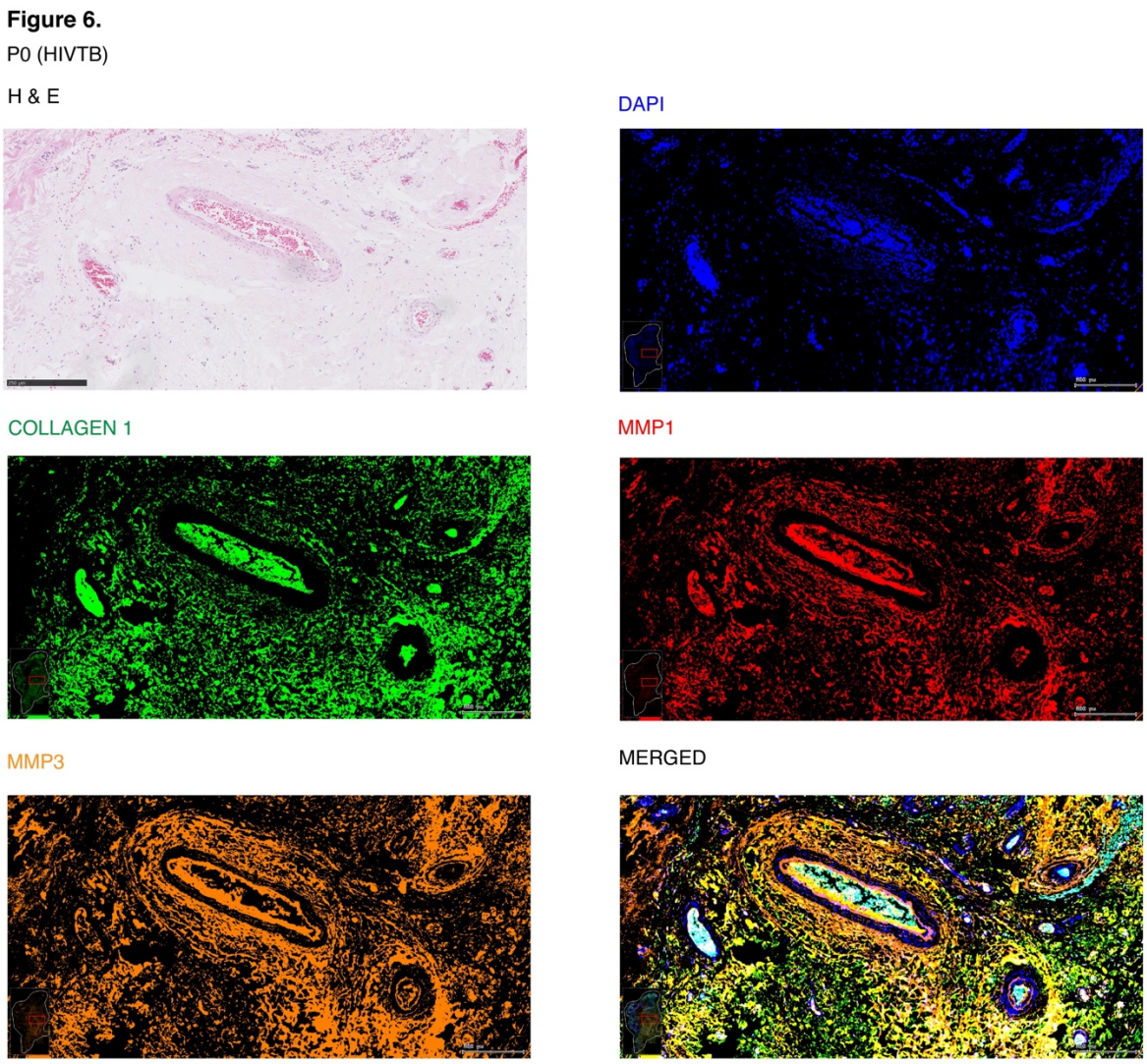


Figure 6: Immunohistochemistry of TB infected tissue showing association of fibroblast 3 markers with blood vessels.

Representative hematoxylin and eosin (H &E) images of lung tissue from participant 0 (HIVTB) showing blood vessels surrounded by fibrotic tissue. Scale bars 250µm. The red staining in the center of the vessels show red blood cells. Fluorescent immunohistochemistry imaging of the blood vessels in the human lung to visualize the localization of fibroblast 3 markers with collagen 1: DAPI (blue), COLLAGEN 1 (green), *MMP1* (red), *MMP3* (orange) and THE MERGED image for all 4 colours. Scale bars 200µm. Red blood cells, in the centre of the blood vessels have been reported to have a high degree of autofluorescence, shown here by their positivity for all stains [152].

2.2.5 Endothelial and Proliferating cells enrichment by disease status

Having observed the additional cellular sub-structure revealed by sub clustering fibroblasts, we repeated the process for endothelial (598) and the proliferating cells (339). Endothelial cells resolved into 5 generic clusters (Figure 7A). After re-clustering, we attempted to identify marker genes associated with each of these clusters (Figure 7B), revealing lymphatic (cluster 4, *PDPN*) and vascular endothelial cells (cluster 3, *PXDN*). Cluster 0 expressed apolipoprotein L domain containing, also known as VERGE (*APOLD1*), which plays a role in endothelial signalling and vascular function [153]. Cluster 1 expressed high levels of thioredoxin interacting protein (*TXNIP)*, which has been reported to induce inflammation, fibrosis and molecular damage by oxygen free radicals [154]. We further visualized these clusters by the disease status (Figure 7C), revealing that cluster 0 was enriched in both HIV and/or TB participants whilst cluster 1 was highly enriched in the non-TB control group. The lymphatic endothelial cells (cluster 4) signature was distributed throughout the patient groups whereas the vascular endothelial cells (cluster 3) was overrepresented in TB participants. This is consistent with *in vitro* and clinical evidence of HIV-1 induced vascular endothelial cell dysfunction [155]. Sub-clustering of the proliferating cells, revealing 3 distinct (Figure 7D) clusters of myeloid cells (cluster 0, *C1QB, MARCO, LYZ*), t cells (cluster 1, *CD3D, TRBC2*) and plasma cells (*IGHG1, IGHG2*) shown in (Figure 7E). Visualizing the cell types by disease revealed that macrophages were present in both TB and HIVTB groups, whereas t and b cells enriched in the TB group (Figure 7F). This observation was consistent with the understanding that HIV-1 infection leads to depletion CD4 t cells, leading to immune failure and the rise of opportunistic infections [156] , together with reports of increased b cell fractions in individuals with culture confirmed active TB [157,158]. Macrophages have been reported to proliferate at the sites of inflammation, primarily directed by the T helper 2 cytokine interleukin 4 (IL4) [159].

Figure 7

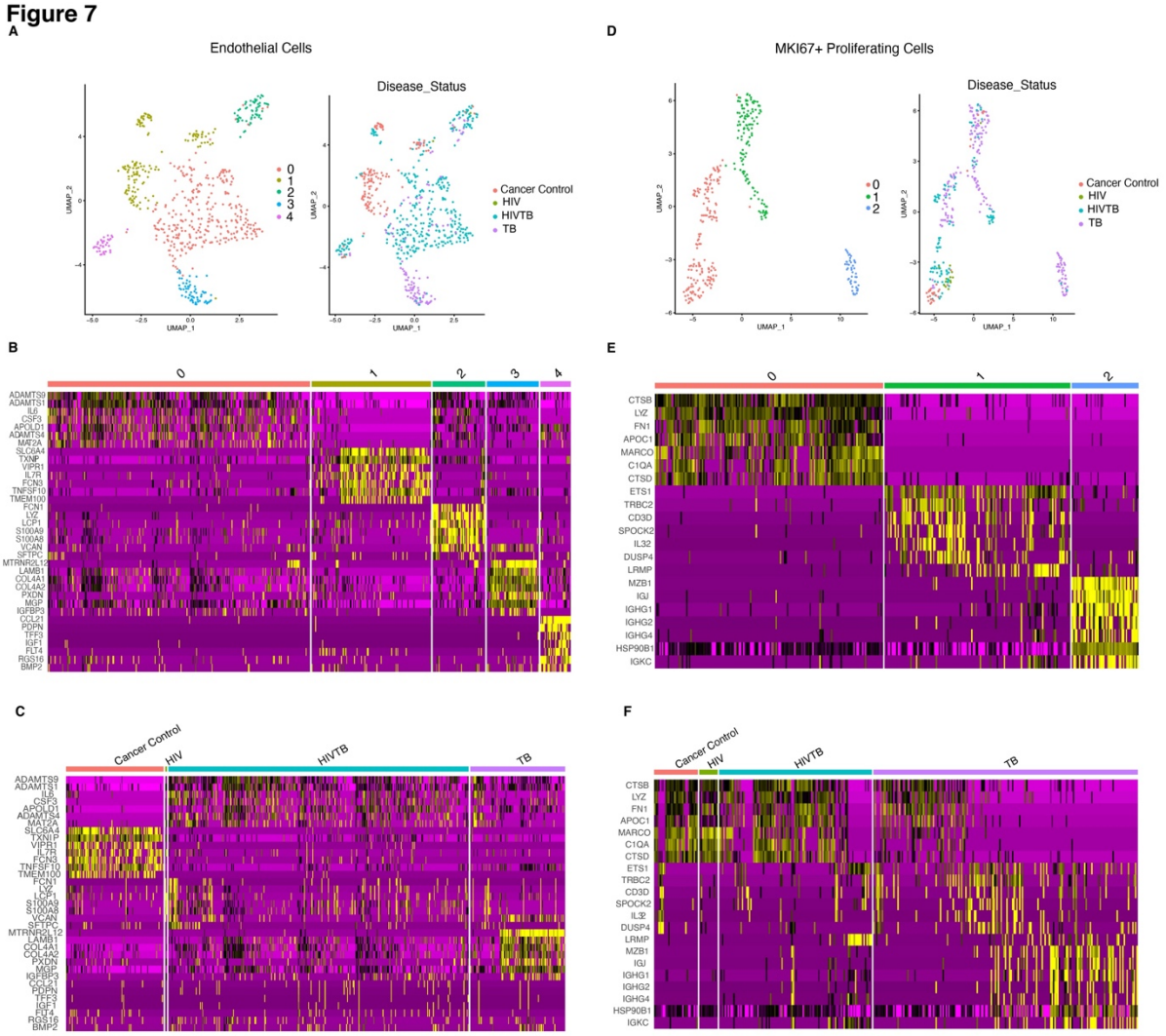


Figure 7: Cell types showing enrichment of specific clusters in the diseased tissue.

(A-C) An outline of the visualization of 598 endothelial cells enrichment based on the disease status of the participants.
(A) UMAP plots showing the clustering of the endothelial cells as a functional of the revealed 5 generic clusters and the disease status.
(B) Heatmap of the top 7 marker genes of the 5 generic clusters.
(C) Heatmap showing the enrichment of the generic clusters in different disease states.

(D-F) A Schematic showing the visualization of 339 *MKI67* Proliferating cells based on participant disease status.
(D) UMAP plots showing the distribution of the 3 generic subclusters and their distribution as a function of the disease status of the participants.
(E) Heatmap showing the top 7 markers in each 3 generic clusters.
(F) Heatmap of the generic clusters, showing the enrichment of the different clusters by disease status.

2.2.6 Monocyte, neutrophil and alveolar pneumocyte distribution by disease status

We further subclustered the abundant myeloid and epithelial cell populations, revealing additional population structure for monocytes (3198), neutrophils (3132) and alveolar type 1 (AT1) / type 2 pneumocytes (AT2) (1022). Monocytes resolved into 5 clusters, and we set to identify marker genes (Figure 8A). When manually curating marker genes for each cluster, we

used genes that appeared in at least 25% of the cells of the same cluster . We then used these genes to compute the differentially expressed genes for each cluster vs the rest of the cells at log fold difference > 0.25. This formed the basis by which we defined subclusters; smaller populations of cells with a similar transcriptomic profile. Cluster 0, expressed *CD55*, upregulated with bacterial infections and the long non-coding RNA *NEAT1*, involved in inflammasome activation. Cluster 1 expressed *DDX3Y*, a functional homolog of *DDX3X* that mediates the innate response to microbial infection; *HMOX1*, which suppresses the pro-inflammatory phenotype and heat shock 70 kD protein *HSPA1A/HSPA1B*, a cellular response to physiological stress. Cluster 2 expressed cytotoxic genes (*GNLY*), cluster 3 expressed pro-inflammatory (*ETS1*) and cell motility (*RHOC*) genes and cluster 4 expressed high levels of the activation marker *CD69* and the *IL-33* receptor, *IL1RL1*. Cluster 0 was enriched across all group whereas cluster 1 was prominent in HIVTB participants (Figure 8B). [160].

Neutrophils resolved into 3 distinct cell types (Figure 8C). Cluster 0 expressed neutrophil metalloprotease, *MMP9,* which degrades collagen IV in the basement membrane. Cluster 1 expressed genes associated with polymorphonuclear myeloid derived suppressor cells (*OLR1*)[161], IL-17 suppression (*TAOK1*), neutrophil recruitment (*GBP5*) and mesenchymal stem cell activation (*GBP1*). Cluster 2 expressed heat shock proteins like the monocyte sub population highlighted previously. Cluster 0 was enriched in participants with TB only, cluster 1 in HIVTB participants with cluster 3 common to all groups (Figure 8D). Type 1 (AT1) and type 2 (AT2) alveolar pneumocytes to were sub-clustered together, revealing the 4 distinct cell subtypes (Figure 8E). We observed AT2 (*SFTPC*), secretory club cells (*SCGB1A1*), macrophages (*C1QA, APOE)* and AT1 (*AGER*). Secretory cells, type 1 pneumocytes and macrophages were enriched in the HIVTB group, whereas type 2 pneumocytes were represented in both TB and HIVTB groups (Figure 8F).

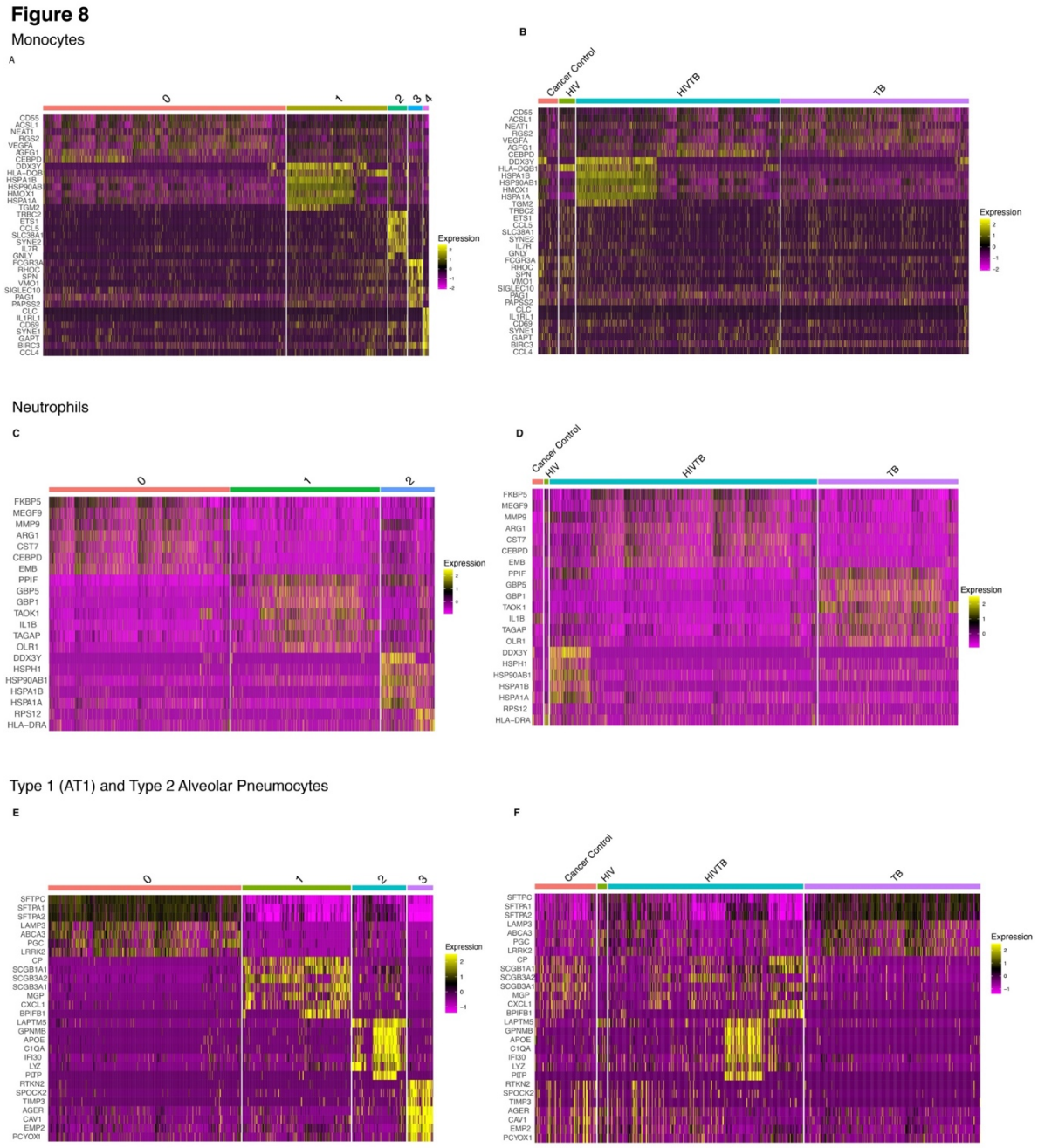


Figure 8: Cell subtype enrichment as a function of disease status.

(A-B) Shows the outline of the heatmap visualization used to show differences in monocyte (3198) subtype composition as determined by the disease status.
(A) Heatmap showing the separation of the monocyte cluster into 5 generic clusters.
(B) Heatmap showing the distribution of the clusters according to the disease status of the participants.

(C-D) Genes defining different neutrophil (3132) subsets and distribution across participants by disease status.
(C) Heatmap of the 3 generic sub-clusters derived from the neutrophil cluster.
(D) Heatmap showing the enrichment of these clusters in different disease status of the participants.

(E-F) Visualization methods used for (1022) Type 1 (AT1) and Type 2 (AT2) Alveolar Pneumocytes.
(E) Heatmap showing the top 7 highly expressed genes in the 4 generic clusters derived from the pneumocyte population.
(F) Heatmap showing the enrichment of these different pneumocyte clusters in the different disease states.

2.2.7 Evaluation of peripheral blood-derived TB signatures

We were interested in exploring how genes identified as blood biomarkers of TB were expressed within our lung data. We selected gene lists from the following publications Zak *et al*., 2016 [162] , Maertzdorf *et al*., 2016 [163] , Cliff *et al*., 2013 [164] , Singhania *et al*., 2018 [165] and Maertzdorf *et al*., 2011 [166]. In these studies, gene signatures of active pulmonary TB were all derived from PBMC using either bulking RNA sequencing or microarray hybridization technology. We computed the average enrichment score of each gene signature for the 19 cell clusters and visualized the enrichment scores normalized to random control feature genes (Figure 9A-E). The most striking enrichment of signature genes was that of the Maertzdorf *et al*., 2016 signature in neutrophils and inflammatory cells. Interestingly, the other signatures were not particularly elevated as the neutrophils/inflammatory cells signal. This was consistent with observations by Berry et al ., 2010 [167], who reported a neutrophil signature induced by type 1 and type 2 interferons [109]. The remaining signatures appeared to be generally enriched in myeloid cells, which often have higher transcriptional activity than other cell types such as lymphocytes [168]. However, the signature reported by Singhania *et al*., 2018 does appear to be more lung tissue specific, being enriched in type 1 pneumocytes, type 2 pneumocytes, endothelial cells and fibroblasts (Figure 9D). Interestingly, this signature was generated using a modular approach that specifically attempted to reduce dominant effect of highly upregulated interferon gene signatures by using weighted gene co-expression network analysis. Based on our observations, this signature may be able to capture the pulmonary remodelling associated with TB, but more testing is required with other pulmonary diseases that induce cavitation, fibrosis and granuloma formation.

Figure 9

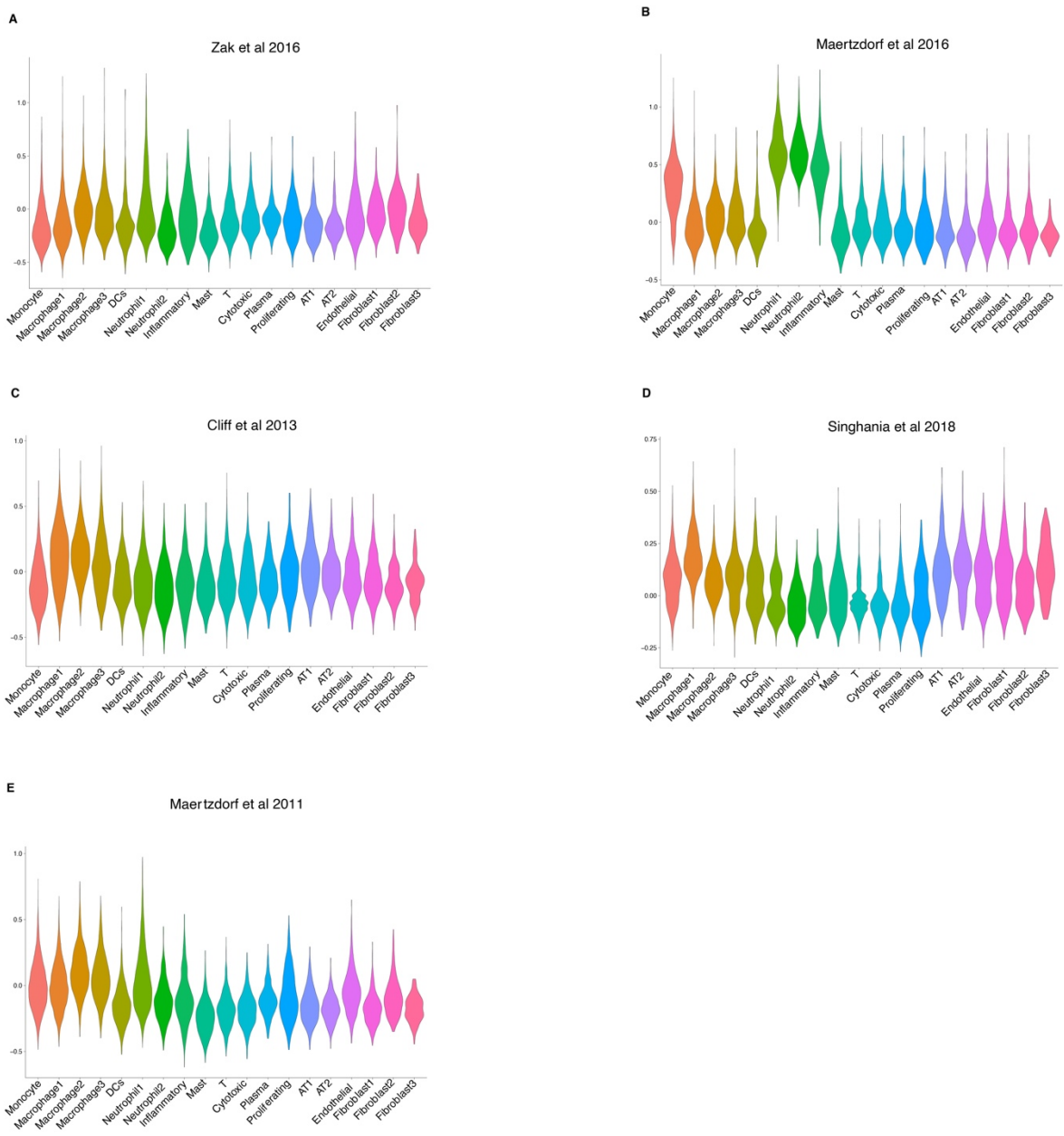


Figure 9: Evaluation of TB Blood signature enrichment in cell clusters from human lung tissue.

(A) Violin Plots depicting the module scores computed for each lung cell clusters using the 16 gene signature Zak *et al.,* 2016 [162] ,(B) 18 gene signature Maertzdorf *et al*., 2016 [163] ,(C) the 19 gene signature Cliff *et al*., 2013 [164] ,(D) the 20 gene signature Singhania *et al*., 2016 [65] ,(E) the 29 gene signature Maertzdorf *et al*., 2011 [166].

2.3 Discussion

Here we present an unbiased analysis of lung cells in resected tissue obtained from participants with TB and non-TB controls. By describing key molecular differences that are driven by the

disease, we confirmed several significant observations from previous *in vitro*/*ex vivo* studies, animal models, highlighting key areas for further studies in HIV/TB pathology. By identifying unique cell types and aberrant pathways, this data is useful in investigating therapeutic alternatives for management of tuberculosis disease. In the right context, this data can be useful in the validation of already existing research questions or leverage its unbiased nature to create unique thought paradigms.

A recent scRNAseq study on 3 participants reported 59 cell types from 75,000 cells using plate and droplet based RNA sequencing platforms, purifying compartment specific cell types using magnetic activated sorting (MACS) and fluorescence activated sorting (FACS) of immune (CD31+CD45+), epithelial (EPCAM+) and stromal (EPCAM-CD31-CD45-) [169]. In our study, we recovered 19 cell types from 20,962 cells from 13 participants. We observed fewer cell types due to the lower cell numbers and possible loss of cells from the homogenisation of the lung tissue. We speculate that the compartment specific enrichment used to reach 59 cell types allowed for a more targeted enrichment which was not feasible in our experiments [170]. The analysis identified stromal cell types such as fibroblasts, endothelial cells, immune cells (monocytes, macrophages, neutrophils, dendritic cells, t cells, plasma cells, mast cells) together with cell states (cytotoxic, proliferating and inflammatory). Neutrophils resolved into 2 separate clusters. Interestingly, neutrophil 2 expressed *ARG1* more abundantly, which has been linked with the down regulation of t cell activation and apoptosis of cancer cells [171]. It was interesting to note that with the exception of dendritic cells, greater than 50% of the myeloid cells were derived from the HIVTB group. In particular, macrophage 2 (*FABP4, C1QB*) and neutrophil2 (*IFITM2, FCGR3B*) showed lower proportions from TB group, suggesting disease specific depletion. FABP4 is involved in lipid metabolism as it relates to inflammation and macrophages are reportedly involved in inflammation of adipose tissue in HIV infected

individuals [172]. Enrichment of dysfunctional neutrophils at mucosal surfaces has been associated with HIV infection [173].We did not explore this dysfunction in our study as we lacked robust controls, but it remains interesting in the context of disease .

We observed a significant pro-fibrotic phenotype in the 3 distinct fibroblast cell types, with significant expression of collagens, growth factors and immune modulatory molecules. Focussing on the fibroblasts revealed an enrichment of a unique population (Figure 3C & D) in the TB group. These cells displayed a transcriptomic profile similar to podoplanin *(PDPN*), fibroblast activating protein (*FAP*) immune fibroblasts, which have been shown to coordinate the organisation of tertiary lymphoid structures (TLS) in response to inflammation [144]. Our exploration of the topographical positioning of these cells within TB granulomatous tissue suggested that these cells formed part of the fibrotic cuff surrounding fibrotic regions filled with dead cells or deposits of ECM (Figure 4). Fibrosis has been reported in chronic TB [174], with increased incidence in patients post anti-tuberculosis treatment [45]. In general, fibrosis occurs either at the periphery or in the centre of the lesion, but the prevailing view is that fibrotic granuloma are more protective as they can effectively contain the bacteria, although they inadvertently make it difficult for drugs to reach the pathogen during treatment [175]. The association of fibroblast 3 markers primarily with the cuff of the lesions suggests that these cells play a role in granuloma organisation. We suggest further probing, particularly of TB granuloma with these markers to explore the hypothesis that immune fibroblasts are involved in the formation and maintanace of the TB hallmark.

We observed an upregulation of *MMP1* and *MMP3* proteins within the granulomatous lesions as well as in association with blood vessels in highly fibrotic lung tissue (Figure 5 and Figure 6 respectively). A recent preprint reported an increase in the proportion of endothelial cells in

TB granuloma in comparison with granuloma from sarcoidosis patients [176], with another suggesting that elevated levels of endothelial cell progenitors in TB patients promotes bacterial dissemination [177]. Further investigations with these marker genes is necessary on TB granuloma from different patients to understand biological and more crucially clinical consequences of such an association. The fact that fibroblast 3 appeared to be overrepresented in TB patients who are HIV negative is interesting. HIVTB infected individuals present with fewer cavities and one can speculate that they lack these driving fibroproliferative and cavity inducing cell types.The TB profibrotic phenotype requires investigation in the context of idiopathic pulmonary and cystic fibrosis to better clarify biological elements unique to each disease. Additionally, better curated TB and HIVTB lung samples will be required to test this hypothesis. Another approach is to use flow cytometry for enrichment of the immune fibroblasts using the methodology outlined by Nayar *et al*., 2019 [144] ,with $CD45^{-}CD235a^{-}$ $CD11b^{-}$ $EpCAM^{-}CD31^{-}PDPN^{+}CD34^{-}$ as the gating strategy. Unfortunately, due to the closure of our clinical study, additional samples were not available to conduct these assays.

Many studies have highlighted how HIV infection can lead to the dysfunction of immune [178], endothelial [179], epithelial [180] and stromal cells. We observed a similar trend when we examined different clusters for gene expression strongly across the participant groups. The pro-inflammatory (*S100A8/A9, ADAMTS9/ADAMTS1*) together with lymphatic endothelial subsets (*PDPN*) were present in both patient groups. Vascular endothelial cells (*PXDN*) were enriched in the TB group, supporting reports that viral proteins lead to a pro-inflammatory, vasoconstriction and vascular endothelial cell apoptosis [181].We also observed that the endothelial cells from the cancer participants had high expression of markers reported in non-small cell lung cancer and lung carcinoma (*SLC6A4, TXNIP, VIPR, IL7R*) [182-184]. The *MKI67* proliferating cluster resolved into immune cells (myeloid, T and plasma cells), with the last 2

cell types enriched in the TB group. Both neutrophils and monocytes from HIVTB expressed heat shock proteins associated with cellular stress , consistent with virus induced dysregulation (neutropenia and monocytopenia) [185]. A scRNAseq study on blood monocytes, identified a subset of cytotoxic monocytes resembling natural killer dendritic cells [160], which we observed in all disease groups in Figure 8 (cluster 2, *GNLY*). We also observed that HIVTB patients were enriched for a highly inflammatory subtype of neutrophils (*FKBP5 , MEGF9 , MMP9 , ARG1 , CEBPD*) [186-190] whereas the TB group was enriched for a polymorphonuclear myeloid derived suppressor cell phenotype (*OLR1 , TAOK1*) [161]. Analysis of the alveolar epithelial compartment showed an enrichment of secretory cells (*SCGB1A1*), alveolar macrophages (*APOE*) and alveolar type 1 pneumocytes (*AGER*) in the HIVTB group. Whether the differences we observed are the consequence of TB or/and HIV disease remains to be elucidated. The enrichment of the subpopulations could be due to differences in the sequence of HIV and TB infection events as this might affect the cellular composition of the lung tissue differentially. Nevertheless, the observed correlation with existing literature suggests that some differences in cell proportions reflect the underlying biology and warrant further exploration.

Lastly, we explored how previous TB signatures align with our data by computing enrichment scores for each signature and super impose the scores on each of our 19 cell clusters. Most signature genes were highly expressed by myeloid cells , consistent with the dominant innate immune interferon-induced signalling. The notable exception, however, is the signature developed by Singhania *et al*. These authors reasoned that the majority of TB gene signatures were constructed using gene reduction methodologies that prioritize the most abundant transcripts [191]. The authors pointed out that this leads to a highly correlated gene set with a wider immunological focus due to the dominant, conserved properties of the innate immune response. Therefore, it is not surprising that these, and other TB signatures largely identify the

same genes, resulting in similar diagnostic utility [192]. Singhania *et al*., by contrast, sought to broaden their signature by taking a modular approach involving selection discriminant genes across the whole transcriptome. It is striking, therefore, that this signature was enriched in endothelial cells, pneumocytes and fibroblasts. We speculate that their approach allowed the signature to capture a component of the tissue remodelling. We suggest defining the role each cell type plays during the abnormal pulmonary wound healing especially in the late / chronic stage of TB disease using scRNAseq. This will help refine our diagnostic tools by producing biologically sensitive ,specific and functionally relevant markers [102]. It is also key that we include a protein expression panel for the refined biomarkers to see the degree to which it correlates gene expression. A recent study by Cai *et al*., 2020 made use of scRNAseq and compared PBMC from healthy, latent TB and active TB participants [70]. The results revealed a depletion of NK cells in the active TB participants, consistent with low NK cell frequencies in newly diagnosed patients [193]. NK cell exhaustion has been reported in cancer and other chronic diseases, thus further exploration is required to ascertain whether this observation is a general feature of the immune system or a TB specific phenotype [194]. We speculate the scRNAseq will prove useful in conjunction to bulk rna-sequencing in localizing transcripts to cell type, which informs us on the cell types to perturb for mechanism of disease progression. We suggest that future studies should include blood, bal and lung compartments across the full spectrum of TB infection. Identification of a TB specific population of cells involved in lung granuloma formation should be the ultimate goal of such endeavours. The high resolution capacity of scRNAseq allows us to reduce the false positive rates observed in bulk sequencing biomarker studies. This is due to the independent nature of transcript detection in scRNAseq compared to bulk RNAseq, which detects all transcripts as an averaged signal, missing genes from low abundance cell types [195].

Taken together, this chapter highlights the power of single cell sequencing in generating research questions specific to human TB disease in a highly unbiased, data driven approach. Further research is required to explore the significance of the activated fibroblast 3 population in the TB only participants together with formulating a biosignature that captures the TB specific tissue remodelling during and post anti-TB treatment. This will require a stringent criteria for patient selection to ensure robust TB case definition. We did not have this luxury in our study, thus we leaned heavily on clinical records and histological examinations in defining our groups. We speculate that expanding our dataset to 100,000 cells, whilst maintaining our current sequencing depth of 50000 reads/cell will provide greater statistical power for observation of disease driven skewing and cells types which are less than 1% of the population [196]. This can be followed up by population specific experiments, at a higher depth of 500000 reads/per cell (costs allowing), which likely capture more minute biological differences in the transcriptome that are important in defining the functionality of new cell types [197]. The dataset we presented was aligned with the hg19 human genome. Recently, the hg38 genome has become available and empirical studies suggests it leads to more annotated genes without changing the overall structure of the dataset (unpublished data). There is a need for studies to compare the 2 genomes and determine which performs better in the context of cell identification in human tissue.

The manuscript for work presented in Chapter 2 of this thesis in the advanced stage of preparation and will be presumptively submitted in the next 8 weeks.

Chapter 3: Single cell sequencing in profiling lung CD69+ CD4 T cells in tuberculosis infected individuals.

3.1 Introduction

Bacillus Calmette-Guerin (BCG) remains the only licensed TB vaccine since its introduction in 1921. It consists of attenuated *Mycobacterium bovis*, priming the immune system to respond to *Mtb* antigens [198]. Unfortunately, BCG has varying efficacy based on age, gender, geographical location [199] and ethnicity [200]. BCG provides effective protection to children [201] from TB meningitis, miliary TB and pulmonary TB [202]. This protection is believed to last for approximately 20 years , after which young adults disproportionately develop diseases in endemic areas [203]. Thus there is a pressing need for alternative vaccines to prevent disease development in adults.

T lymphocytes (T cells) play a critical role in the prevention of TB after primary *Mtb* infection as part of the adaptive immune response [204]. Human immunodeficiency virus (HIV) increases the risk of developing disseminated TB especially at low cd4 t cell levels [205]. In cases where individuals received antiretroviral therapy (ART) corresponding to increased cd4 t cells, they remained at greater risk of developing TB compared to the HIV negative group. This suggests that other cells are responsible for controlling infection or that ART does not fully restore the cell repertoire required to prevent TB disease.

T cells migrate to infected lung tissue in response to chemo-attractants and remain there as long lived memory cells, allowing for an optimum secondary reaction [77]. These so called “tissue resident memory” cd4 t cells (Trm) are derived from central ($SELL^{hi}/CCR7^{hi}$) and effector memory ($SELL^{lo}/CCR7^{lo}$) t cells. Trm are phenotypically heterogenous and express a variety

of surface markers associated with tissue retention, including CD69 and CD103. T-cell receptor (TCR) upregulation leads to the triggering of the CD69 [206], which can make the use of the latter as a marker of residency complicated. However, once upregulated, CD69 competitively binds to the S1PR on the surface of cells [207]. This prevents migration of the cells back to the circulation following the S1P gradient. CD103 is involved in the maintenance of t cells within tissue, particularly at the epithelial barrier, binding, in conjunction with the Beta 7 integrin (ITGB7) bind to E-cadherin [208]. Despite their wide use as canonical Trm markers, neither CD69 or CD103 perfectly corresponds to residency [209]. Trm are emerging as an important target for novel TB vaccine studies, with most of the work done in animal models.

As discussed previously, (scRNAseq) is a valuable tool in profiling cells at a cellular resolution, allowing advancement in our understanding cell specific responses to disease [210]. It has provided insight into the functional aspects of T cells, demonstrated when it identified a tumour infiltrating cd8 Trm population expressing both effector and suppressor transcripts in breast cancer patients [92]. In this chapter, we first applied scRNAseq to peripheral blood mononuclear cells (PBMC) of TB infected and healthy individuals in an attempt to reproduce the cellular subtypes obtained using a different scRNAseq platform (10X) by Cai and colleagues [70]. We then compared from the same individual, CD4 T cells from the PBMC and CD69+ / CD69^{-} CD4 T cells from the lung. We used CD69 as a marker for CD4 Trm as opposed to CD103 as the latter shown to be better associated with CD8 Trm cells [209].

3.2 Results

3.2.1 Single cell RNA sequencing of peripheral blood mononuclear cell (PBMC) from active TB and healthy individuals

PBMC were obtained from 3 HIV negative active TB patients with abnormal chest x-rays ,a positive interferon gamma release assay (IGRA+) and 3 healthy donors (IGRA-). The low

patient size (n = 6) was due to the fact that this was a pilot study and we did not have resources to follow up on. PBMC were isolated using density centrifugation from a previously reported protocol [52]. A total of 15000 cells per participant were loaded onto seq-well microarray and processed as discussed in the previous chapter [52]. Data from the sequencer was aligned to the human genome (hg19), followed by exhaustive quality check measures to remove low quality cells (defined as containing >5% mitochondrial genes, < 200 transcripts and > 2500 transcripts as set in seurat). This yielded 2998 cells (Figure 1A), distributed across the active TB (subjects P6033 (n = 338), P6077 (n = 297) and P6110 (n = 105)) and healthy controls (subjects P8346 (n = 1034), P8347 (n = 466), P8349 (n = 758)). We observed higher cell yield from the healthy controls compared to the TB group (Figure 1B). We expected even distribution between the groups and speculate that the extraction protocol performed sub optimally. Unsupervised clustering of the cells revealed 9 distinct clusters, which included red blood cells (*HBA2, HBB*), neutrophils (*S100A8, S100A9*), monocytes (*FCN1, VCAN*), cd8 t cells (*CD8A, GNLY*), neutrophil2 (*FCGR3B, G0S2*), cd4 t cells (*IL7R, SPOCK2*), plasma cells (*IGKC, IGHG4*), b cells (*MS4A1, BANK1*) and platelets (*TUBB1, PPBP*) as shown in Figure 1C. The presence of red blood cells supports the speculation that the extraction protocol under performed as these cells are supposedly removed by density centrifugation. It is possible that human error in enumerating the PBMC could have led to reduced cell yield. Figure 1B suggests that monocytes, t lymphocytes (cd4 and cd8) and platelets were comparatively more abundant across the disease groups. A study showed a significant increase in monocyte absolute counts in individuals with active TB compared to health individuals [211]. The same study found significant differences in the frequency of lymphocytes between active TB and latently infected patients. Other groups have associated higher platelet counts and thrombocytosis with advanced pulmonary TB [212] , with Fox et al., 2018 associating platelets with pulmonary inflammation and tissue destruction [213].

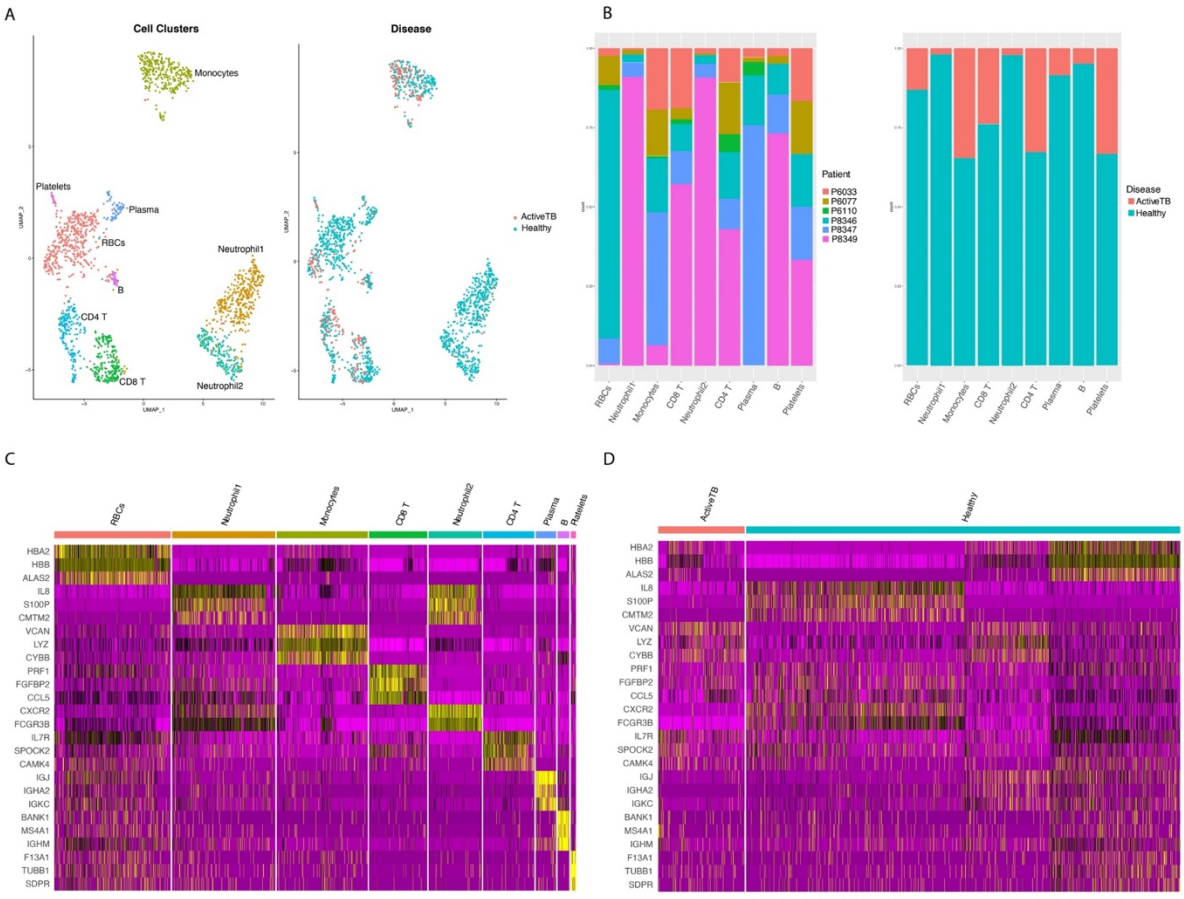


Figure 1 : Overview of PBMC extracted from healthy individuals and active tuberculosis individuals.

(A) UMAP projection showing 1590 PBMC from healthy and active TB patients coloured according to the different cell lineage clusters and disease status (red (active TB) and blue (healthy).

(B) Bar plots showing the cell yield per individual in accordance to the clusters and the disease status.

(C) Heatmap of the 3 highest expressed genes in each cluster, showing the distribution of different canonical markers.

(D) Heatmap showing the distribution of the clusters as a function of disease status. We identified these clusters by making use of unsupervised clustering via the FindClusters tool within the seurat package with default parameters, k.pram set to 10 and resolution set to 0.5.

It should be noted that other studies that have performed single cell sequencing do not report the presence of neutrophils [214], which are separated from the PBMC fraction during extraction. However, Figure 1B shows that the majority of both neutrophil 1 and 2 populations derive from a single individual, P8349. Neutrophils can separate out with the PBMC, particularly low density neutrophils (LDN). These LDN be elevated in individuals with active pulmonary TB [215]. However, it is possible that downstream processes such the library work up and sequencing might have been compromised, leading to the erroneous detection of neutrophils. The S100A9

gene observed in the neutrophils is also expressed in monocytes and dendritic cells, thus there is a possibility of erroneous cluster identification [216]. Stratification of the heatmaps according to the disease status did not reveal conclusive differences in gene expression due to the discrepant cell yield between the groups (Figure 1D).

3.2.2 Analysis of PBMC derived from HIVTB Participant 36

We enrolled 5 participants (4 HIVTB (participant 31, 36, 37 and 38) and 1 TB (participant 35), performing PBMC extraction and scRNAseq. We recovered usable data from participant 36 (Figure 2B), PMBC (n =1319), shown in Figure 2A-2C. We observed 7 representative clusters, neutrophils (*FCGR3B, NAMPT*), cd4 t cells (*CD3D, IL7R*), cd8 t cells (*CD8A, GNLY*), monocytes (*FCN1, VCAN*), b cells (*CD79A, MS4A1*), inflammatory cells (*CLC, CAT*) and platelets (*PPBP, TUBB1*). We proceeded to subset the cd4 t cells and re-cluster these cells to assess the underlying cell types (Figure 3C). We observed 2 distinct clusters, with cluster 0 expressing ribosomal genes (*RPL5, RPL34, RPS15A*), important in normal physiology and some pathologies [217]. This cluster also expressed genes associated with activation (CD69) [218] and regulation of T cells (*FOXP1*, *RUNX3* ,*RORA, CD6*) [219-222]. Cluster 1 expressed genes that are associated with naïve T cells (*FPR1, C5AR1 , GLUL, G0S2*) [223-225] and regulatory T cells (*MXD1 , PLAU , IFITM2*) [226-228]. We simultaneously sorted cd4 t cells from the PBMC fraction for the same participants according to the gating strategy in Figure 3A. We could not recover usable sequencing data and we believe there were issues with the FACS.

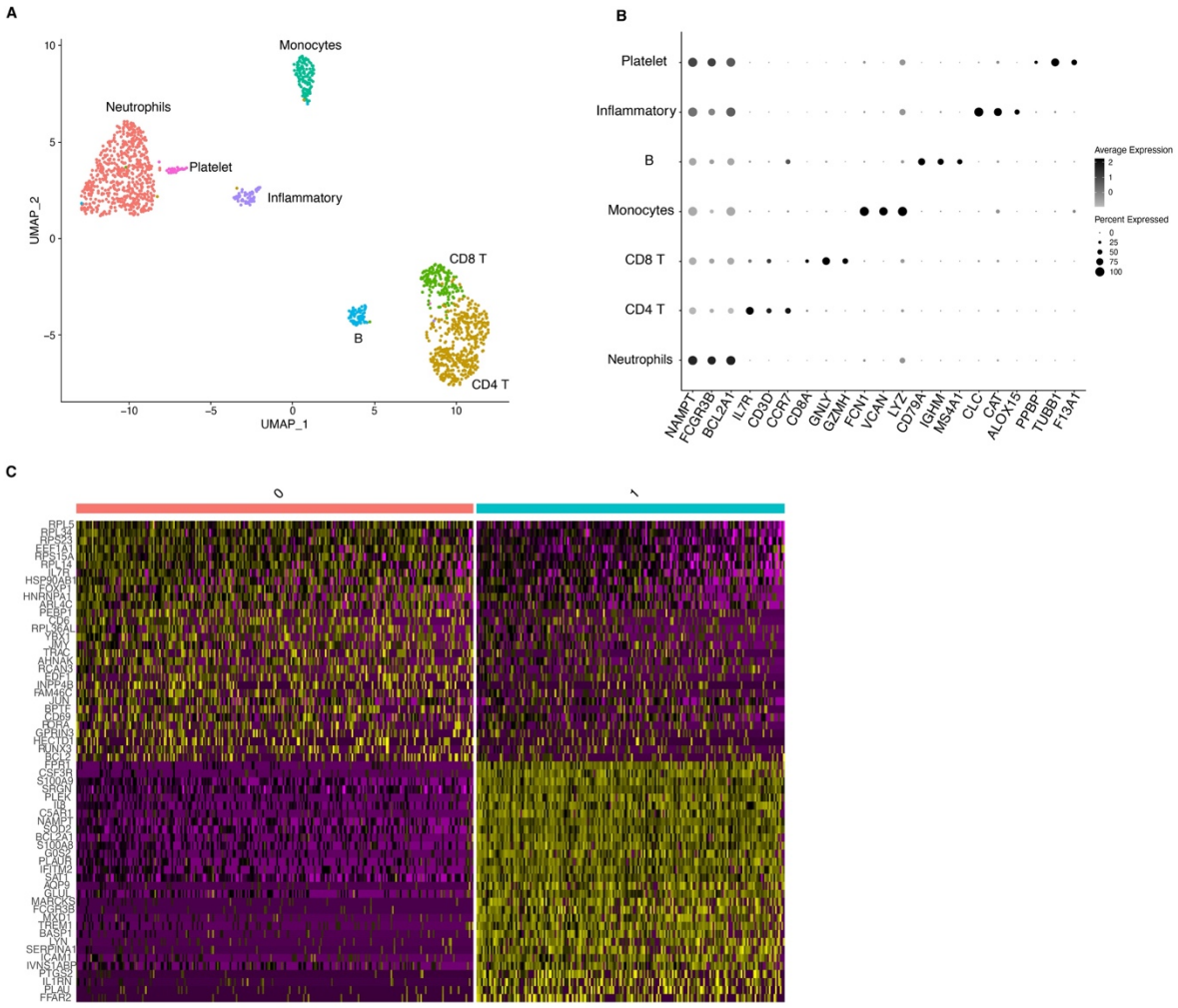


Figure 2 : PBMC isolated from Participant 36 and the gene expression profile of CD4 T cells.

(A) UMAP projection of 1319 cells from the PBMCs fraction, with colours representing different canonical clusters.

(B) Dot plot of PBMCs from Participant 36, showing expression of marker genes in each generic cluster.

(C) Heatmap showing the gene expression profile of CD4 T cells from the PBMC fraction of Participant 36.

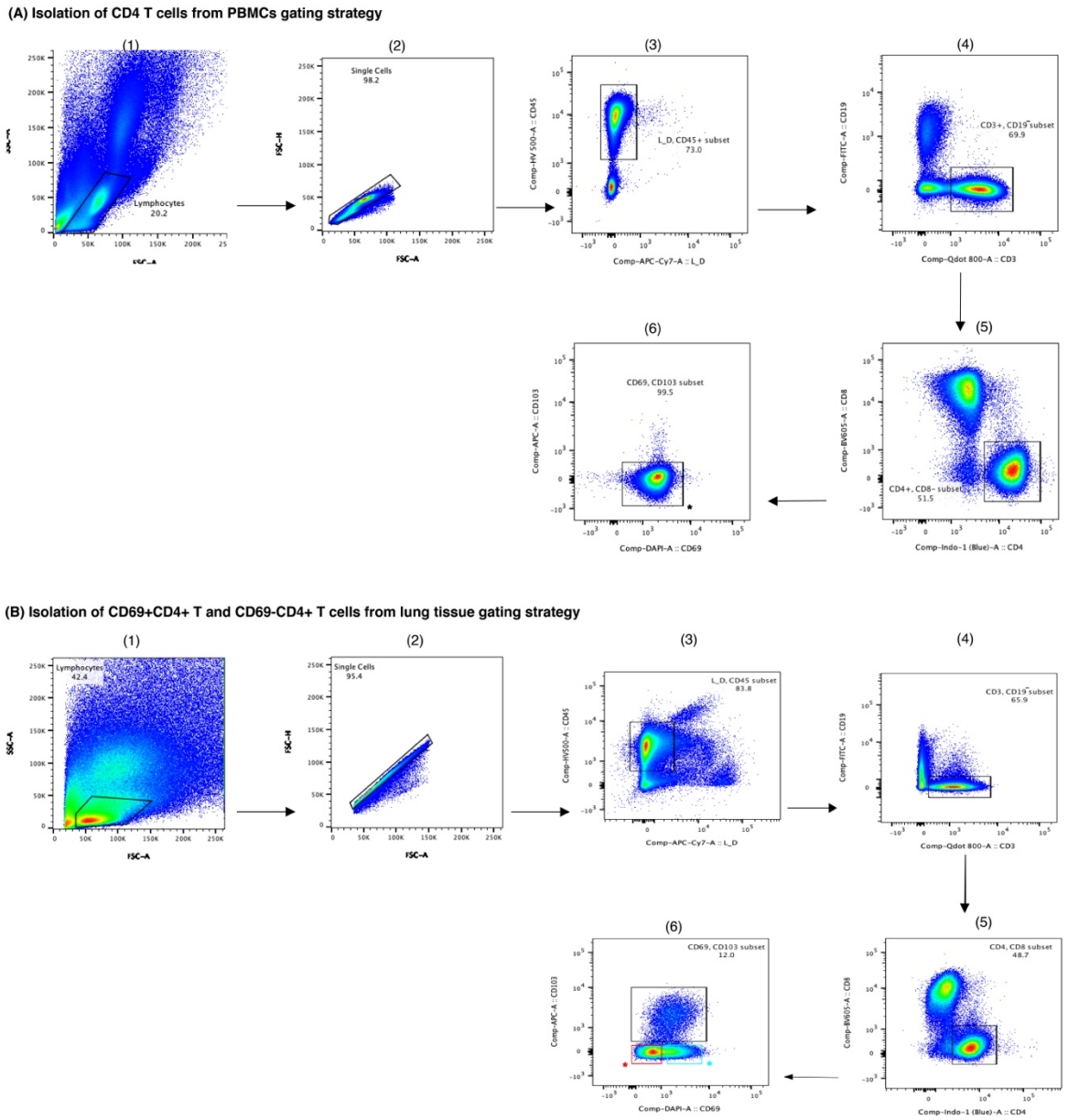


Figure 3 : Flow panel gating strategy used for isolating CD4 T cells from PBMC and CD4+CD69-/CD4+CD69+ T cell populations

(A)The gating strategy for T cells isolation starts with the lymphocytes gate (1), singlets (2), live CD45+ cells (3), CD3+CD19- (4), CD4+CD8- (5) and CD4+CD69- (6) population in the box adjacent to the *.

(B) Gating strategy for isolation of CD4 tissue resident T cell subpopulations. Similar to the PBMCs , the following population of cells from the lymphocytes gate (1-5) singletsCD45+ L/D- CD3+CD19- CD4+CD8- T cells. We then further gated out the cells as (6) CD69- (red box &*), CD69+ (turquoise box & *) and CD103+ (black box & *).

3.2.3 Comparison of Lung CD4+CD69- to CD4+CD69+ T cells from HIVTB Participant 36

To explore lung derived CD4 CD69- to CD4CD69+ cells, we sorted matched tissue single cell suspension from the same participants, recovering usable data from P36. It should be noted that the sorting was done on different days for each participant and sample type. We speculate that instrumental issues led to the poor yield. We also speculate that infection might have depleted cd4 t cells, impeding our efforts to sufficiently isolate them. For P36, we sorted cells into 2 different populations L/D- CD45+ CD3+CD19- CD4+ CD103- CD69- (~33000 events) and L/D- CD45+ CD3+CD19- CD4+ CD103- CD69 + (~27000 events) as shown in Figure 3B. We proceeded to load these populations on to 2 separate arrays per population and performed scRNAseq. We aligned the subsequent reads to the human genome (hg19) and recovered a total of 728 cells from the 2 populations (Figure 4), much lower than the expected 500-3000 cells per run (according to the historical performance of seqwell). This suggests that either enough cells were recovered or the library generation was suboptimal. Uniform manifold approximation and projection of the cells revealed 5 clusters (Figure 4), identified as naïve cd4 t cells (*CCR7, SELL/CD62L, TCF7*) [229-231]; cd4 t regulatory cells [cluster 1 and 2] (*KLRB1, PRDM1 , CD96/TACTILE*) [232-234] ; myeloid-like activated t cells expressing genes associated with regulation (*NR4A2, CD300E*) [235-237], activation (*TIMP1*) [238] together with cytotoxic cd4 t cells (*PRF1, GZMK, NKG7*) [239-241]. Examining the distribution of cells showed a relative enrichment of naïve T-cell in the CD69 negative fraction (Figure 4A). However, due to a failure of sequencing in the other participants, the analysis could not be extended beyond our first pilot, preventing statistically significant observations. We explored methanol cryopreservation of cells according to the protocol [242], but we could not recover enough material for the sequencing step so we used fresh cells for all experiments. Unfortunately, due to COVID-19 restrictions in 2020, the lung study was halted for almost 12 months, meaning that we could

not address these technical issues. Thus, the data presented here serves as a guide for future research in this area within Leslie lab and beyond.

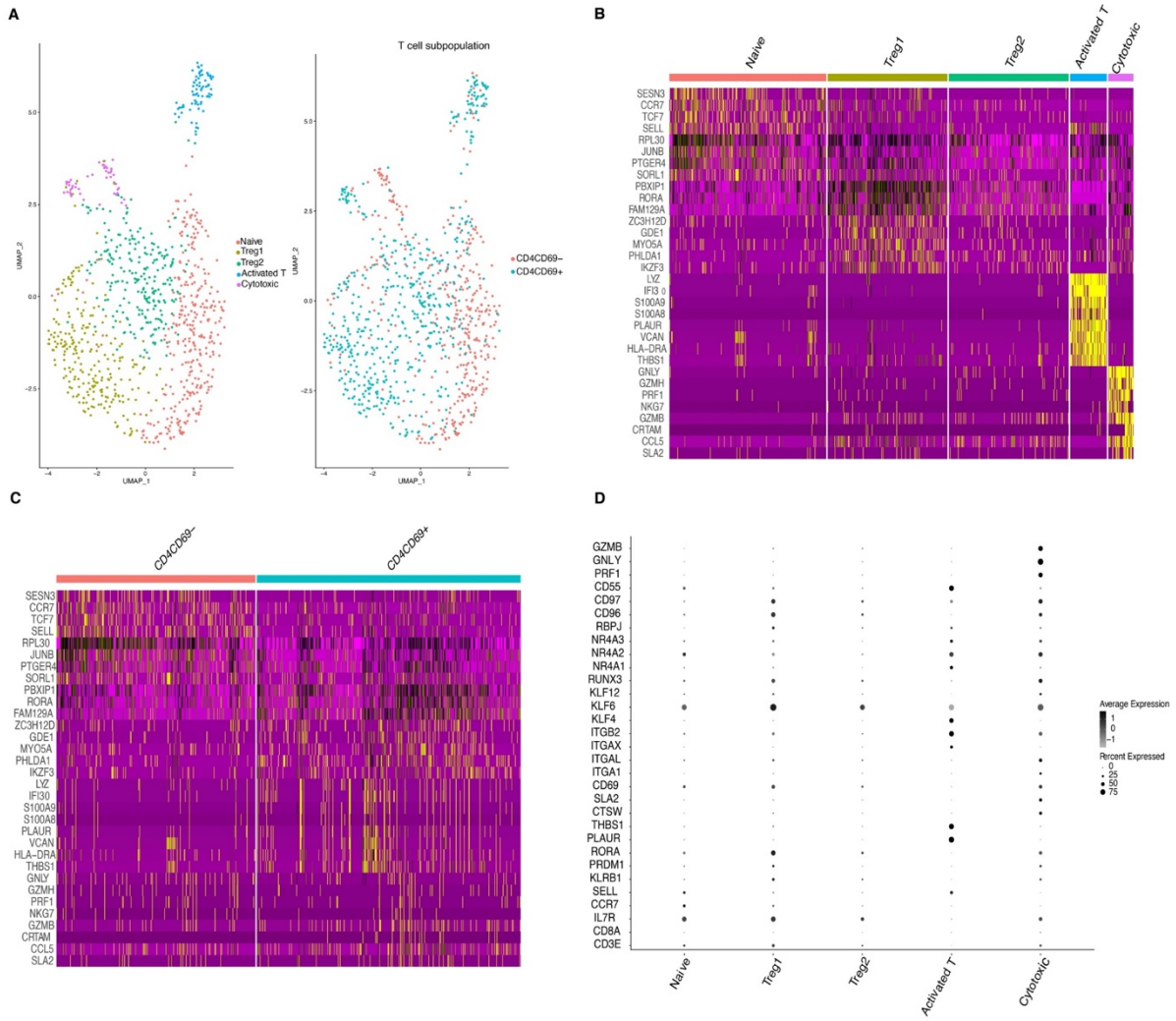


Figure 4 : Single cell RNA sequencing of lung CD4+CD69+ and CD4+CD69- T cells from Participant 36 lung tissue.

(A) UMAP projection of 728 cells, separating into generic clusters identified by the expression of marker genes. The cells are also coloured based on whether they are from the CD69+ sort (blue) or CD69- sort (red).

(B) Heatmap showing the expression of the top 8 marker genes of the generic clusters from Participant 36. We identified these clusters by making use of unsupervised clustering via the FindClusters tool within the seurat package with default parameters, k.pram set to 10 and resolution set to 0.5.

(C) Heatmap of the 728 cells stratified according to the sorting panels used, red (CD4+CD69-) and blue (CD4+CD69-). We identified these clusters by making use of unsupervised clustering via the FindClusters tool within the seurat package with default parameters, k.pram set to 10 and resolution set to 0.5. The top 11 genes shown were the topmost genes per cluster from the FindClusters output.

(D) Dot plot showing gene expression from the different cell clusters, from the naïve, Treg(T regulatory cells), activated and cytotoxic cells.

3.3 Discussion

In this chapter, we first explore differences in the PBMC from individuals with active TB and healthy controls. This was done to compare the phenotypic diversity of our PBMC compared to a recent study which generated a single cell atlas of PBMC [70]. We also compared, from one individual, cd4 t cells from the PBMC fraction to lung tissue derived CD4 CD69+ and CD4 CD69- T cells with the aim of identifying phenotypic diversity between the blood and lung compartments. In describing the phenotypes of both PBMC and tissue resident cd4 t cells, we demonstrated the applicability of scRNAseq in uncovering the underlying cellular sub-structure.

Our analysis the PBMC produced 9 clusters (resolution = 0.5 in FindClusters), showing greater granularity of our dataset compared to the study referenced above that only showed 3 clusters for 62000 cells. We observed red blood cells and neutrophils, which are typically depleted during the isolation of PBMC. This suggests that the extraction process was compromised, particularly for P8346 (RBCs) and P8349 (neutrophil 1 & 2) which produced > 80% of these cells respectively. The remaining cell types we observed (monocytes, cd4 t cells, cd8 t cells, plasma, b cells and platelets) have been reported in other studies, showing that seq-well can characterise similar cell types to those reported in the literature.

Blood cd4 t cells from P36 showed 2 distinct clusters, as illustrated in Figure 2C. We speculate that cluster 0 corresponds to an activated effector t cell due to the expression of ribosomal genes (*RPL5*), *CD69, JUN, RORA and CD6* [206,243-245]. Cluster 1 expressed genes associated with activation and immune response (*FPR1, SOD2, GLUL*), which we expect as this individual had both HIV and active TB [223,246,247]. We observed different cell subtypes when we combined CD4CD69- cells with the CD4CD69+ cells from participant 36 (Figure 4A-C). The cells subtypes separated into naïve t cells (*CCR7, SELL*), regulatory T cells (*RORA*), activated t cells

that expressed myeloid like markers (Figure 4D), together with cytotoxic t cells. Tissue resident, *ITGA1* expressing t cells have also been observed to exhibit a highly cytotoxic phenotype (*PRF1, IFNY*), with *ITGA1* negative cells expressing *IL-17* [91,248]. The observation that both the blood and tissue compartments had regulatory and activated cd4 t cells suggests that regulatory/effector subsets in both compartments have shared function. The lung microenvironment has a significant impact on the phenotype of tissue resident cells. We also observed expression of transcription factors associated with the tissue resident populations, such as *RUNX3, NR4A1-3*, but the expression was distributed throughout the clusters suggesting that they might be shared across different populations [249,250]. Expression of myeloid cell markers on t cells is consistent with reports of CD11b and CD11c on antiviral cd8 t cells in a mouse model [251]. The aberrant expression of t cell markers has been reported in patients with acute leukaemia [252]. We speculate that HIV/TB disease might impart similar lineage ambiguous phenotypes but we require more robust studies to explore this hypothesis. We also observed an enrichment of CD11a (*ITGAL*) in the cytotoxic cluster, a transcript which has been reported to be elevated in cd4 Trm relative to CD103 [253]. We observed *CD300E* expression in the activated T cell cluster (not shown), which belongs to a family of surface receptors that negatively regulate the ability of t cells to present antigens via *STAT1* pathway [237]. *CD300A* was upregulated in HIV positive ART naïve participants when [254]. A subset of CD300a+PD1+CD38+ of cd4 t cells which might be implicated in immune exhaustion. Participant 36 was HIV positive thus we speculate that this subset is a consequence of HIV driven dysregulation.

Taken together, the results shown in this section demonstrated the potential application of scRNAseq data in improving our understanding of the phenotypic/functional aspects of tissue resident cd4 t cells. Seq-well (S^3) incorporates second strand synthesis in addition to the normal

reverse transcriptase step, which increases recovery of low abundance transcripts. Thus we expected to recover more cells with greater annotation of transcription factors, cytokines and cytokine receptors [255]. We did not manage to recover cells from other participants and thus no conclusions can be made from this data. We suspect that this poor recovery is a function of low starting cell counts of the Trm's due to the highly compromised nature of the tissue and low frequency of this cell type. In some cases, the tissue we received from the surgery was mainly composed of cartilage (especially from bullectomy) or originated from HIV infected individuals (associated with CD4 T cell depletion), thus limiting cell recovery. Other failure points could have been reagents or instruments (PCR machines, sequencers etc) along our experimental pipeline. As discussed, it was not possible to obtain additional samples, due to cohort and lab closures during the last 12 months of my PhD study. In future studies, we must incorporate more phenotypic markers in defining Trm, such CD11a (*ITGAL*) to assess its applicability in differentiating cd4 from cd8 t cells. It has been reported that t cell receptor (TCR) specificity elicits Trm potential in t cells in responding to influenza infection [256,257]. The seq-well (S^3) platform used in this experiment is also capable of performing parallel sequencing of the TCR, thus in theory any population exhibiting the Trm phenotype should have limited clonal diversity [258]. Other research groups have suggested using t cells subsets such as CD1 restricted T cells, mucosal associated invariant T cells (MAIT) and γδ T cells [259] as vaccine candidates. They postulated that these cells are ideal for TB host directed t cell therapy due to low donor diversity coupled with conserved TB specific epitopes. To date, the only clinical studies that have been performed on the efficacy of unconventional γδ T cells involved t cell treatment for antitumor therapy [260].

Chapter 4: Applying TB scRNAseq data in COVID-19 research.

4.1 Introduction

Over the past decade, zoonotic transmissions from animals to humans have led to the emergence of human coronaviruses (CoVs) such as middle east respiratory syndrome (MERS)-CoV [261] and severe acute respiratory syndrome (SARS)-CoV [262]. A novel SARS-CoV-2, first reported in Wuhan City, China, December 2019, causes the disease COVID-19. It had led to 3,68 million deaths worldwide at the time of writing (June 2021). Prior work on SARS-CoV revealed the host angiotensin-converting enzyme 2 (*ACE2*) as a critical receptor for binding of the viral spike (S) protein with a high affinity [263]. In addition, type 2 transmembrane serine protease (*TMPRSS2*) has been shown to facilitate activation of the spike and initial viral entry in target cells [264]. Many other host proteases such as furin have been thought to also play a role in promoting uptake of the virus but the exact site and mechanism(s) by which they process the S protein are yet to be elucidated [265-267]. Expression of *ACE2* within tissue has been observed in human lung and small intestine epithelial cells by histological staining [268]. However, the specific cell subsets that expressed *ACE2* in the human lung were unknown during the early stages of the outbreak. Identifying cells that are *ACE2* positive together with *TMPRSS2* is important for understanding the potential mechanism of viral entry and the overall pathogenesis.

The advent of high throughput single cell RNA sequencing (scRNAseq) provides a powerful tool to characterize cell types that are associated with pathology by comparison of healthy and diseased tissue [269]. Profiling of cell subsets and states reveals dynamic variations in gene expression in *in vitro* and *in vivo* viral infection assays [270,271]. In this chapter, we utilized

scRNAseq from a selected lung samples described in chapter 2 (3 HIVTB, 3 TB and 2 Cancer patients) to investigate the expression of *ACE2* and *TMPRSS2* and identify potential virus targets within the lung tissue. The TB and HIVTB donors had a history of TB but were TB culture negative at the time of surgery. We converted the whole tissue to a single cell suspension as reported by Ardain *et al.,* 2018 [52] and used the seq-well protocol [272] . We then used the single cell libraries created to create a single cell atlas of lung tissue, shown in Figure 4A.

4.2 Results

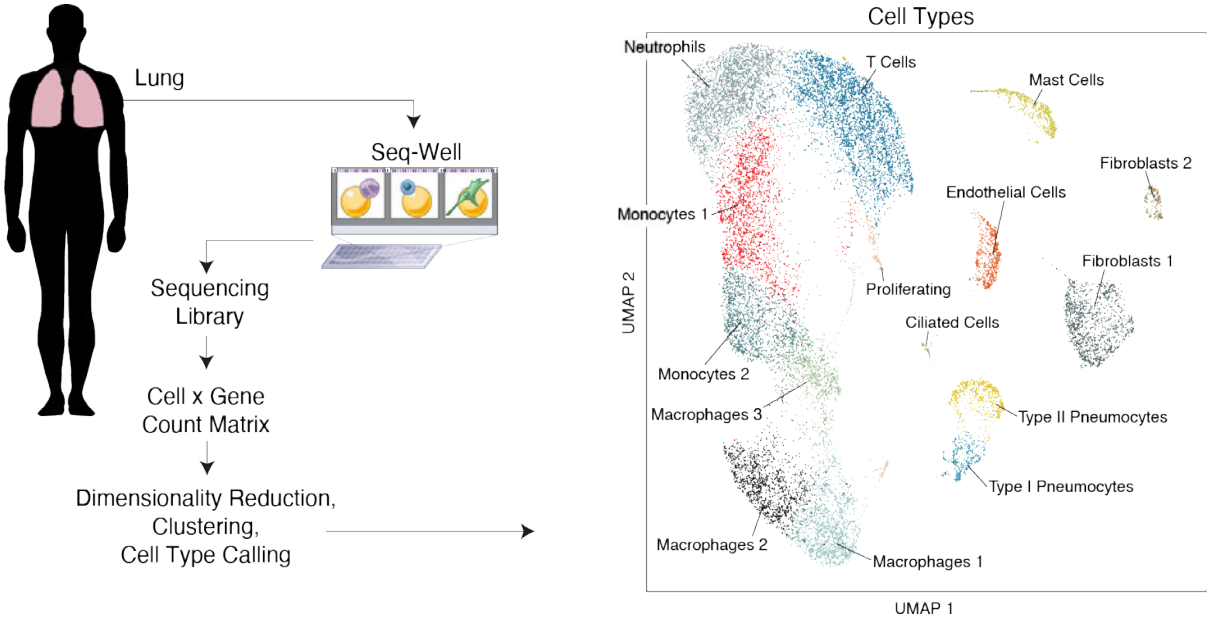


Figure 4A: Overview of the Workflow.

Isolation of human lung tissue from tissue resections, synthesis of a single cell library by Seq-well V3 (S^3) and the computational workflow used for analysis of different cell types. On the right is a UMAP visualization of 18915 cells derived from 8 donors (n = 3 TB/HIV, n = 3 TB and n =2 Cancer controls). The different cell types are visualized using the highlighted colour code.

Unsupervised clustering of the single cell library from these participants revealed multiple cell types and cell states in the lung tissue, like those discussed in chapter 2 (Figure 4A). It was interesting to note that in this chapter, we did not observe any b cells from the data which we attribute to lower patient numbers (n = 8). We would expect to observe b cells in abundance as they have been shown to be enriched in TB patients, playing a role in prevention of

disseminated disease [273]. B cells have been associated with TB granuloma, where they are thought to limit the disease via IL10 and IL21, thus their absence is rather peculiar [274]. We suspect that the absence of b cells is due to the relatively low cells numbers in this experiment. B cells have been shown in a separate scRNAseq study on disease human lung tissue to constitute less than 5% of the total cells thus it is possible to not detect them [46].

The cell types were identified based on the expression of canonical markers as shown (Figure 4B**)**. *ACE2* and *TMPRSS2* were expressed (albeit with low abundance) by type 2 pneumocytes and ciliated cells **(**Figure 4C**).** Type 2 pneumocytes were identified by the expression of surfactant proteins *SFTPA2, SFTPB.* These proteins, in combination with phospholipids, constitute the secretions necessary for the reduction of surface tension and maintenance of alveoli architecture [275]. In this cell type, 1.4% expressed *ACE2* (false discovery rate (FDR)-adjusted p = 1.35E-21), 34.2% expressed *TMPRSS2* (FDR p < 1E -300) and 0.8% expressed both receptors. Both receptor genes were also expressed in ciliated lung cells; 7% of ciliated cells expressed *ACE2* (FDR-adjusted p = 3.8 E-30), 24.2% expressed *TMPRSS2* (FDR adjusted p = 3.25 E-7 and 5.3 % expressed both receptors. Ciliated cells play a critical role in removal of harmful material from the lung airway via coordinated beating [276].

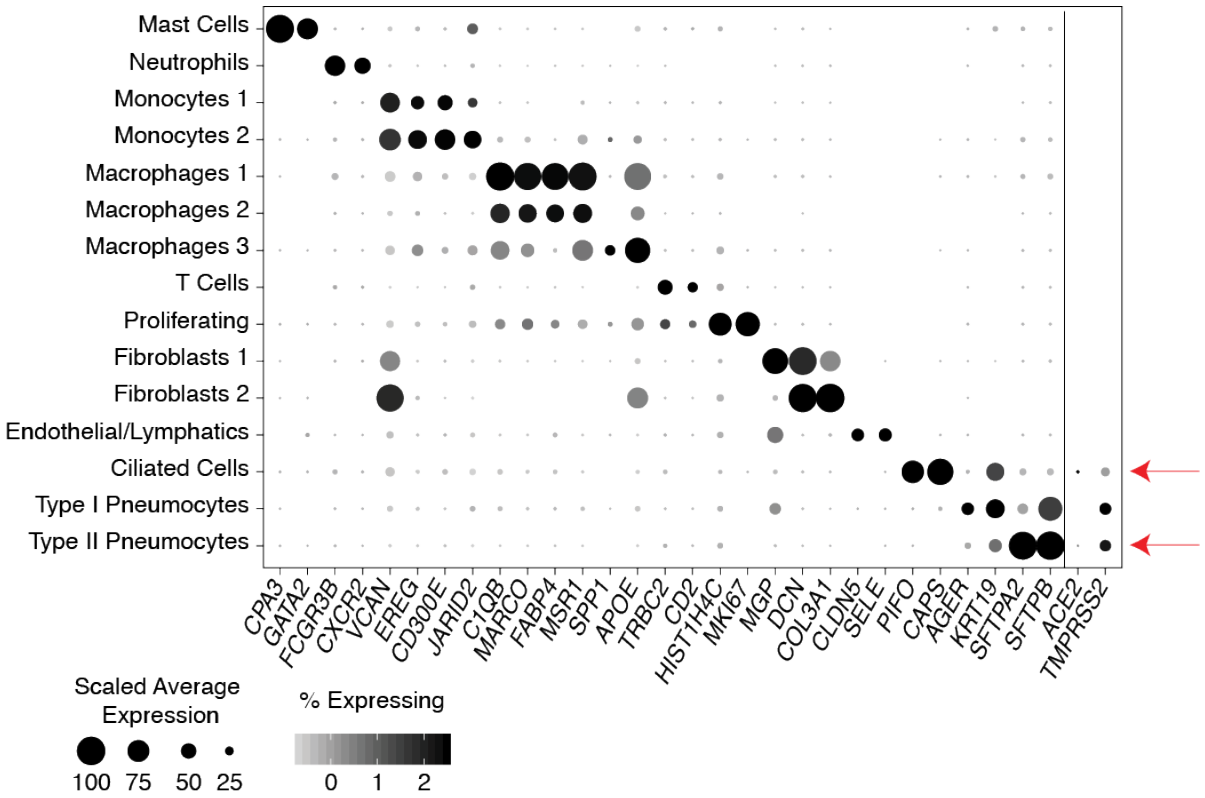


Figure 4B: Dot plot showing canonical marker genes for each cell type (FDR-Adjusted p< 0.001) as well as ACE2 and TMPRSS2 expression.

The size of the dots represents the corrected average expression (log (scaled UMI +1)) of a given gene and the color intensity represents the count-based expression amount and the red arrows highlight the cells expressing the highest proportion of *ACE2*.

To establish differences between cells that consistently express the target molecules for SARS-CoV-2, we computed the differentially expressed genes between *ACE2*$^{+}$*TMPRSS2*$^{+}$ type 2 pneumocytes and type 2 pneumocytes lacking the expression these receptors (Figure 4D). Dual expression of *ACE2* and *TMPRSS22* was associated with several genes, such as forkhead box J2 (*FOXJ2*) POZ/BTB and AT Hook Containing Zinc Finger 1 (*PATZ1*) and tetraspanin 7 (*TSPAN7*), all of which influence epithelial-mesenchymal cell transition (EMT)[277,278]. *RNF41* is essential for determining epithelial cell polarity i.e. having apical, lateral and basal domains allowing these cells to perform their homeostatic function [279]. *FOXJ2* has been implicated in the inhibition of transforming growth factor beta 1 (*TFG-β1*) induced (EMT) whereas *CASC7* has been reported to suppress cell proliferation, invasion, migration and amplifies cell apoptosis. Tripartite motif containing-28 (*TRIM28*) was also upregulated in this population of

pneumocytes and it has interestingly been associated with regulating the interferon gamma (IFN) response [280].

To assess whether the observations from the human lung cells were similar in other animal models, we analyzed data from non-human primate (*Macaca mulatta*) lung tissue collected at necropsy and analyzed using the seq-well protocol [272]. Consistent with the human data, *ACE2*+ *TMPRSS2*+ were concentrated in the epithelial cell, type 1 and type 2 pneumocytes, club cells and ciliated cells. Co-expression of these genes was concentrated in 3,8% type 2 pneumocytes. In the same animal model, we explored the expression of *ACE2*+ and *TMPRSS2*+ in tissue from the gastrointestinal tract as a follow up to reports suggesting digestive system impairment [281]. We analyzed cells from the ileum, jejunum, liver and colon using the seq-well platform, with 62% of absorptive enterocytes expressing *ACE2*+. Other data sets we analyzed included another non-human primate model (*Macaca fascicularis*) infected with TB until granuloma developed. The animals were subsequently sacrificed, granuloma and adjacent uninvolved tissue isolated, processed and the seq-well protocol was comparatively run. Once again, type 2 pneumocytes from the TB granuloma were enriched for *ACE2*+ *TMPRSS2*+ type 2 pneumocytes largely derived from the granulomas. Type 1 pneumocytes, club, secretory cells and ciliated cells were also positive for both these markers and were enriched in the granulomas. There is a need to explore the consequences of SARS-CoV-2 and TB co-infection primarily on tissue remodeling associated with granuloma formation, focusing on aberrant, stem cell like epithelial cell precursors [282].

As highlighted prior, the cells displayed were recovered from 3 HIVTB, 3 TB and 2 cancer controls (negative for both pathogens) donors. Interestingly, all the pneumocytes which were expressing *ACE2* were recovered from the HIVTB donors (Figure 4E), despite an equal

recovery rate of alveolar pneumocytes (likely-hood ratio test, p = 0.009). This suggests that concurrent HIV infection might lead to the upregulation of *ACE2* receptor in lung epithelial cells. Subsequent work by our collaborators on this study (not shown here) confirmed that *ACE2* is upregulated in SIV infected non-human primates. We also studied virus target cells at the primary site of viral exposure, the nasal cavities, by making use of scRNAseq datasets from human upper respiratory tract (inferior turbinate and ethmoid sinus mucosa from healthy individuals and individuals with chronic rhinosinusitis [269]. We observed enrichment of the *ACE2+ TMPRSS2+* in the secretory epithelial cells, with a dominant IFN-α induced gene signatures [283]. Goblet cells also displayed an upregulation of *ADAR, GBP2, OAS1, JAK1* and *DUOX2*, genes associated with interferon signaling [284]. We further demonstrated that IFN-α and IFN-γ upregulated the expression of *ACE2* in primary human epithelial cells in a dose dependent manner by using bulk RNA sequencing of the cytokine treated cells. This upregulation was positively correlated to the increase in the expression of canonical interferon stimulated genes (ISGs) such as *STAT1, BST2, XAF1, IFI35, MX1 and GBP2*. These observations support the hypothesis that there is an association of canonical IFN response and the *ACE2* response in human epithelial cells. It is significant to note that exposing mouse basal cells from trachea to mouse IFN-α , IFN-γ and IFN-β did not robustly lead to upregulation of *Ace2*, despite upregulation of mouse ISGs such as *Stat1* and *Gbp5.* A publication containing all the detailed analysis summarized above is attached in the annex section at the end of this chapter.

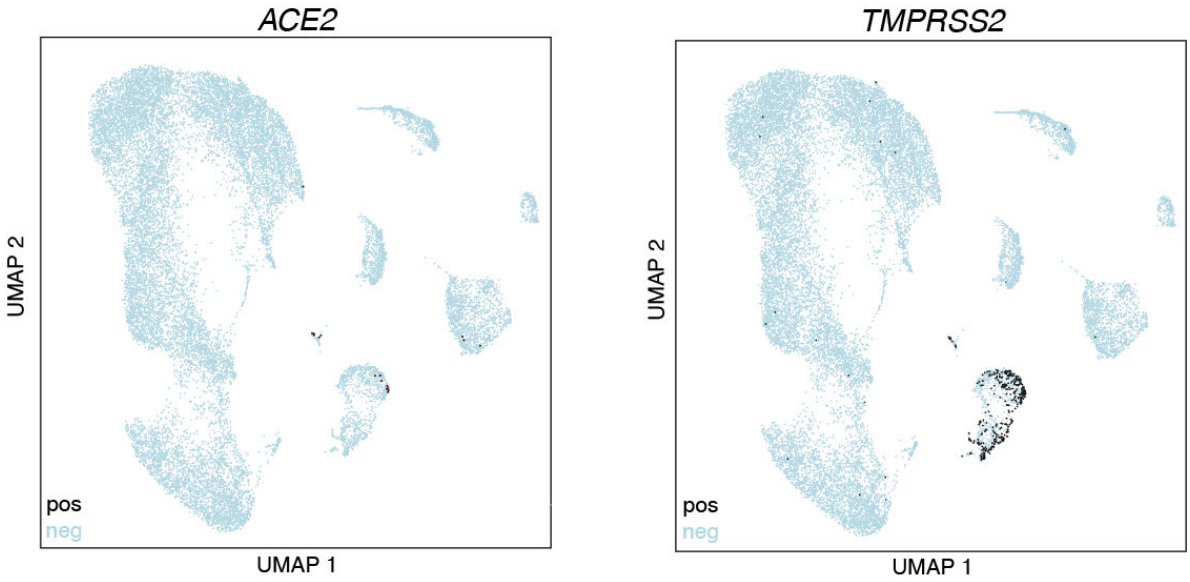


Figure 4C : UMAP projections after dimensional reduction.
The points are coloured based on detection of *ACE2* and *TMPRSS2*. Black being positive and blue negative for gene specific rna.

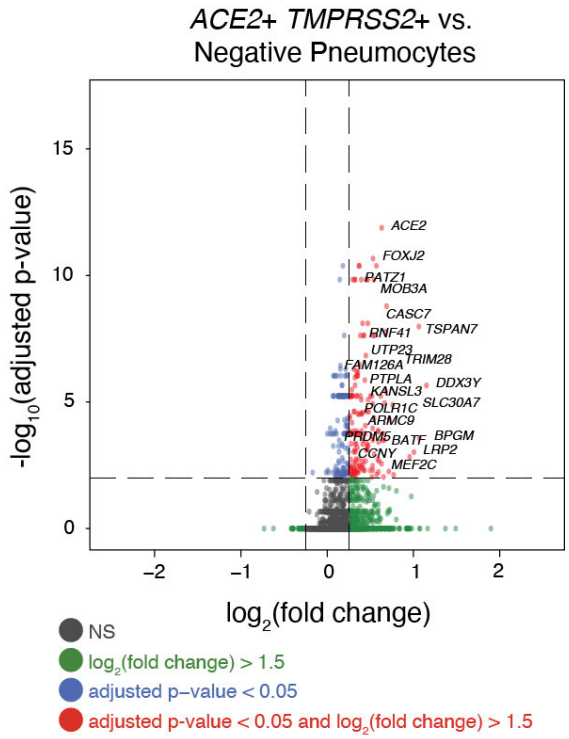


Figure 4D: Plot showing genes upregulated in *ACE2*$^{+}$*TMPRSS2*$^{+}$ pneumocytes in comparison to the rest of the pneumocytes.
The genes marked in red represents genes with FDR-adjusted p < 0.05 and $\log_2$(fold change) > 1.5.

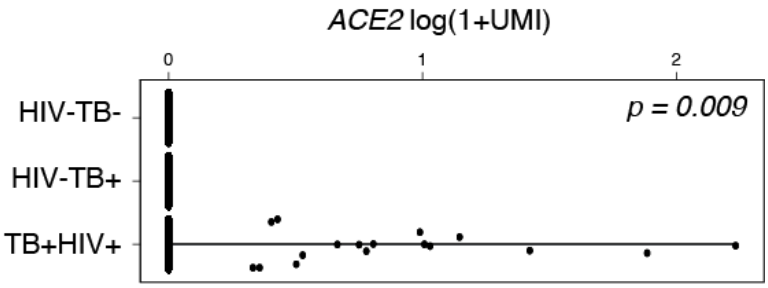


Figure 4E: Expression of ACE2 across the human donors by HIV and TB status (p = 0.009 likelihood-ratio test).

4.3 Discussion

In this chapter, we used our scRNAseq to investigate the potential cellular targets of SARS-CoV-2 infection. We found expression of *ACE2* and *TMPRSS2* to be significantly enriched in type 2 alveolar pneumocytes and ciliated cells. SARS-CoV-2 has been shown to have a high affinity to *ACE2*, which is thought to lead to its rapid spread in humans [285]. Interestingly, *ACE2* downregulates the renin-angiotensin-aldosterone system (RAAS), converting angiotensin II to angiotensin (I to VII) [286], which is beneficial for lung tissue repair. At the time of writing this section, no research group had managed to clinically link increased *ACE2* expression with higher rates of susceptibility and mortality of SARS-CoV-2 patients. We also observed that the cells producing most of the transcripts were recovered from HIVTB individuals. Subsequent data on morbidity and mortality of SARS-CoV2 in South Africa has shown that HIV infected individuals are 3 times more likely to have severe or fatal COVID compared to no HIV controls [287]. Recent studies have suggested that COVID-TB patients are more likely to have fatal outcomes [288]. There are many potential reasons behind this, and the immunosuppression associated with HIV is likely to impact the immune response to this infection. However, it is biologically plausible that an increase in SARS-CoV2 receptor expression associated with HIV may contribute to amplification of viremia and disease severity in some individuals. Additional work is needed to investigate this further.

It was interesting to note that ACE2 and TMPRSS2 co-expression was conserved in epithelial cells with similar cellular identities and frequencies across human and non-human primate cohorts. This was striking as the different data sets were curated from studies that used different protocols for tissue processing, collection and sequencing depth (potentially influencing cell recovery). The study also showed that type 1 and type 2 interferons (IFN) upregulate the expression of ACE2, as evidenced by increased expression in human nasal epithelial cell cultures together in both lower and upper respiratory tract tissue from both humans and non-human primates (NHP). It was interesting to note that ACE2 expression was absent from peripheral blood mononuclear cells in both human and NHP datasets [289]. ACE2 expression exerts a tissue protective function as host defense strategy to viral infection [290]. The discovery in our study that ACE2 is an ISG in human epithelial cells and its necessity for the virus to infect these cells, point towards the virus exploiting the IFN response to aid its establishment and transmission within the host.

In addition, it should be noted that low abundance transcripts such as *ACE2* and *TMPRSS2* tend to be under represented due to inefficiencies and biases that are inherent to scRNAseq methods that utilize universal molecular identifiers (UMI) [24]. Follow up studies which can confirm the protein levels of these 2 targets in the human lung tissue and their co-localization with canonical markers for type 2 pneumocytes (*SFTPC*) and ciliated cells (tubulin polymerization-promoting protein family member 3, *TPPP3*) are critical to confirm the differences driven by comorbidities such as HIV and TB. Preliminary COVID-19 studies have shown increased mortality rates of HIV infected individuals in the United Kingdom [287]. Type 2 pneumocytes have been shown as targets for the avian H5N1 flu [291] and ciliated cells have been preferentially targeted by the human parainfluenza virus type 3 [292], thus there is a precedent for these cells being targets of the SARS-CoV-2. It is critical that the immunofluorescence approach be

utilized to colocalize viral surface markers such as the spike protein and host cellular markers of these two cell types to verify preferential infection. Binding of SARS-CoV to mouse Ace2 leads to reduced expression of this target gene and acute lung failure [293]. Interestingly, the lung failure could be mitigated by blocking the renin-angiotensin pathway. This study was also key in highlighting the increased viral load and lung tissue damage when wild type mice were compared to *Ace2* knockout mutants. Experimental treatments using human recombinant ACE2 to competitively bind the virus or suppressors of ACE2 expression in individuals with the virus or those who are at high risk requires carefully considered animal models and human clinical trials [294,295]. Camostat mesylate, a protease inhibitor, attenuates the activity of TMPRSS2 , thus providing another potential treatment strategy [296].

It is critical that the transcriptional response in both animal models and human infection to the virus be properly characterized to establish a baseline disease phenotype. Such studies would need to factor in co-variates such as age, sex and the presence of other co-morbidities. In this chapter, we highlighted the power of scRNAseq datasets, exploring different hypotheses of relevance in human disease in an unbiased manner that informs mechanistic follow up studies.

Annex 1

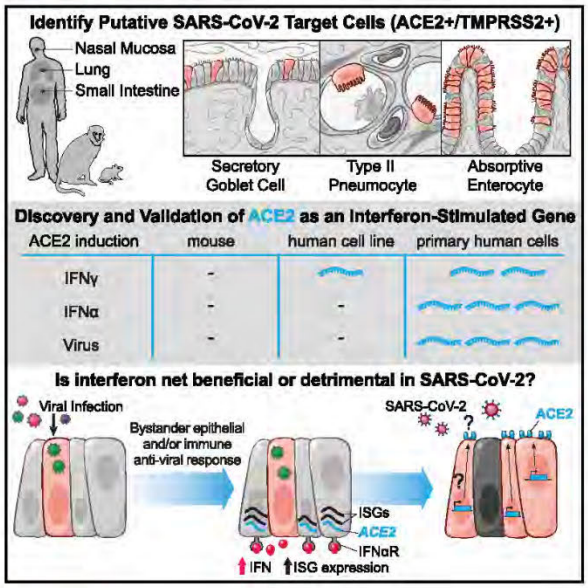
Publication 1: SARS-CoV-2 Receptor ACE2 is an Interferon- Stimulated Gene in Human Airway Epithelial Cells and is Detected in Specific Cell Subsets across Tissues

Cell

Article

SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues

Graphical Abstract



Highlights

- Meta-analysis of human, non-human primate, and mouse single-cell RNA-seq datasets for putative SARS-CoV-2 targets
- Type II pneumocytes, nasal secretory cells, and absorptive enterocytes are *ACE2*$^{+}$*TMPRSS2*$^{+}$
- Interferon and influenza increase *ACE2* in human nasal epithelia and lung tissue
- Mouse *Ace2* is not upregulated by interferon, raising implications for disease modeling

Authors

Carly G.K. Ziegler, Samuel J. Allon, Sarah K. Nyquist, ..., Alex K. Shalek, Jose Ordovas-Montanes, HCA Lung Biological Network

Correspondence

shalek@mit.edu (A.K.S.), jose.ordovas-montanes@childrens.harvard.edu (J.O.-M.), lung-network@humancellatlas.org (HCA Lung Biological Network)

In Brief

Analysis of single-cell RNA-seq datasets from human, non-human primate, and mouse barrier tissues identifies putative cellular targets of SARS-CoV-2 on the basis of *ACE2* and *TMPRSS2* expression. *ACE2* represents a previously unappreciated interferon-stimulated gene in human, but not mouse, epithelial tissues, identifying anti-viral induction of a host tissue-protective mechanism, but also a potential means for viral exploitation of the host response.

Ziegler et al., 2020, Cell *181*, 1016–1035
May 28, 2020 © 2020 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2020.04.035

CellPress

CellPress

Cell

Article

SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues

Carly G.K. Ziegler,[1,2,3,4,5,6,50] Samuel J. Allon,[2,4,5,7,50] Sarah K. Nyquist,[2,4,5,8,9,50] Ian M. Mbano,[10,11,50] Vincent N. Miao,[1,2,4,5] Constantine N. Tzouanas,[1,2,4,5] Yuming Cao,[12] Ashraf S. Yousif,[4] Julia Bals,[4] Blake M. Hauser,[4,13] Jared Feldman,[4,13,14] Christoph Muus,[5,15] Marc H. Wadsworth II,[2,3,4,5,7] Samuel W. Kazer,[2,4,5,7] Travis K. Hughes,[1,4,5,16] Benjamin Doran,[2,4,5,7,17,18] G. James Gatter,[2,4,5] Marko Vukovic,[2,3,4,5,7] Faith Taliaferro,[5,18] Benjamin E. Mead,[2,3,4,5,7] Zhiru Guo,[12] Jennifer P. Wang,[12] Delphine Gras,[19] Magali Plaisant,[20] Meshal Ansari,[21,22,23] Ilias Angelidis,[21,22] Heiko Adler,[22,24] Jennifer M.S. Sucre,[25] Chase J. Taylor,[26] Brian Lin,[27] Avinash Waghray,[27] Vanessa Mitsialis,[18,28] Daniel F. Dwyer,[29] Kathleen M. Buchheit,[29] Joshua A. Boyce,[29] Nora A. Barrett,[29] Tanya M. Laidlaw,[29] Shaina L. Carroll,[30]

(Author list continued on next page)

[1]Program in Health Sciences & Technology, Harvard Medical School & Massachusetts Institute of Technology, Boston, MA 02115, USA
[2]Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA 02139, USA
[5]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[6]Harvard Graduate Program in Biophysics, Harvard University, Cambridge, MA 02138, USA
[7]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[8]Program in Computational & Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[9]Computer Science & Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[10]Africa Health Research Institute, Durban, South Africa
[11]School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa
[12]University of Massachusetts Medical School, Worcester, MA 01655, USA
[13]Department of Microbiology, Harvard Medical School, Boston, MA 02115, USA

(Affiliations continued on next page)

SUMMARY

There is pressing urgency to understand the pathogenesis of the severe acute respiratory syndrome coronavirus clade 2 (SARS-CoV-2), which causes the disease COVID-19. SARS-CoV-2 spike (S) protein binds angiotensin-converting enzyme 2 (ACE2), and in concert with host proteases, principally transmembrane serine protease 2 (TMPRSS2), promotes cellular entry. The cell subsets targeted by SARS-CoV-2 in host tissues and the factors that regulate *ACE2* expression remain unknown. Here, we leverage human, non-human primate, and mouse single-cell RNA-sequencing (scRNA-seq) datasets across health and disease to uncover putative targets of SARS-CoV-2 among tissue-resident cell subsets. We identify *ACE2* and *TMPRSS2* co-expressing cells within lung type II pneumocytes, ileal absorptive enterocytes, and nasal goblet secretory cells. Strikingly, we discovered that *ACE2* is a human interferon-stimulated gene (ISG) *in vitro* using airway epithelial cells and extend our findings to *in vivo* viral infections. Our data suggest that SARS-CoV-2 could exploit species-specific interferon-driven upregulation of *ACE2*, a tissue-protective mediator during lung injury, to enhance infection.

INTRODUCTION

Human coronaviruses (CoVs) are single-stranded positive-sense RNA viruses that can cause mild to severe respiratory disease (Fung and Liu, 2019). Over the past two decades, zoonotic transmission events have led to the emergence of two highly pathogenic CoVs: severe acute respiratory syndrome (SARS)-CoV and Middle East respiratory syndrome (MERS)-CoV. SARS-CoV-2, which causes the disease known as COVID-19, was first reported in late 2019 (Coronaviridae Study Group of the International Committee on Taxonomy of, 2020; Lu et al., 2020; Paules et al., 2020). COVID-19 is characterized by pneumonia, fever,

CellPress

Lucrezia Colonna,[31] Victor Tkachev,[17,32,33] Christopher W. Peterson,[34,35] Alison Yu,[17,36] Hengqi Betty Zheng,[31,36] Hannah P. Gideon,[37,38] Caylin G. Winchell,[37,38,39] Philana Ling Lin,[38,40,41] Colin D. Bingle,[42] Scott B. Snapper,[18,28] Jonathan A. Kropski,[43,44,45] Fabian J. Theis,[23] Herbert B. Schiller,[21,22] Laure-Emmanuelle Zaragosi,[20] Pascal Barbry,[20] Alasdair Leslie,[10,11,46] Hans-Peter Kiem,[34,35] JoAnne L. Flynn,[37,38] Sarah M. Fortune,[4,5,47] Bonnie Berger,[9,48] Robert W. Finberg,[12] Leslie S. Kean,[17,32,33] Manuel Garber,[12] Aaron G. Schmidt,[4,13] Daniel Lingwood,[4] Alex K. Shalek,[1,2,3,4,5,6,7,8,16,33,49,51,52,*] and Jose Ordovas-Montanes[5,16,18,49,51,52,53,*] HCA Lung Biological Network*

[1]Program in Virology, Harvard Medical School, Boston, MA 02115, USA
[2]John A. Paulson School of Engineering & Applied Sciences, Harvard University, Cambridge, MA 02138, USA
[3]Program in Immunology, Harvard Medical School, Boston, MA 02115, USA
[4]Division of Pediatric Hematology/Oncology, Boston Children's Hospital, Boston, MA 02115, USA
[5]Division of Gastroenterology, Hepatology, and Nutrition, Boston Children's Hospital, Boston, MA 02115, USA
[6]Aix-Marseille University, INSERM, INRA, C2VN, Marseille, France
[7]Université Côte d'Azur, CNRS, IPMC, Sophia-Antipolis, France
[8]Comprehensive Pneumology Center & Institute of Lung Biology and Disease, Helmholtz Zentrum München, Munich, Germany
[9]German Center for Lung Research, Munich, Germany
[10]Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany
[11]Research Unit Lung Repair and Regeneration, Helmholtz Zentrum München, Munich, Germany
[12]Division of Neonatology, Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN 37232, USA
[13]Divison of Allergy, Pulmonary, and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA
[14]Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
[15]Division of Gastroenterology, Brigham and Women's Hospital, Boston, MA 02115, USA
[16]Division of Allergy and Clinical Immunology, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA
[17]University of California, Berkeley, CA 94720, USA
[18]University of Washington, Seattle, WA 98195, USA
[19]Dana Farber Cancer Institute, Boston, MA 02115, USA
[20]Harvard Medical School, Boston, MA 02115, USA
[21]Stem Cell & Gene Therapy Program, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[22]Department of Medicine, University of Washington, Seattle, WA 98195, USA
[23]Division of Gastroenterology and Hepatology, Seattle Children's Hospital, Seattle, WA 98145, USA
[24]Department of Microbiology & Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA 15219, USA
[25]Center for Vaccine Research, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA
[26]Division of Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA
[27]UPMC Children's Hospital of Pittsburgh, Pittsburgh, PA 15224, USA
[28]Department of Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, PA 15224, USA
[29]Department of Infection, Immunity & Cardiovascular Disease, The Medical School and The Florey Institute for Host Pathogen Interactions, University of Sheffield, Sheffield, S10 2TN, UK
[30]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA
[31]Department of Cell and Developmental Biology, Vanderbilt University Medical Center, Nashville, TN 37240, USA
[32]Department of Veterans Affairs Medical Center, Nashville, TN 37212, USA
[33]Department of Infection & Immunity, University College London, London, UK
[34]Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
[35]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[36]Harvard Stem Cell Institute, Cambridge, MA 02138, USA
[37]These authors contributed equally
[38]These authors contributed equally
[39]Senior author
[40]Lead Contact
*Correspondence: shalek@mit.edu (A.K.S.), jose.ordovas-montanes@childrens.harvard.edu (J.O.-M.), lung-network@humancellatlas.org (HCA Lung Biological Network)
https://doi.org/10.1016/j.cell.2020.04.035

cough, and occasional diarrhea (Guan et al., 2020; Holshue et al., 2020; Huang et al., 2020), and SARS-CoV-2 RNA has been reliably detected in nasopharyngeal swabs, sputum, and stool samples (Wang et al., 2020; Wölfel et al., 2020; Zou et al., 2020). As of April 19, 2020, SARS-CoV-2 continues to spread worldwide, and there are over 2,401,379 confirmed cases, 165,044 deaths, and 623,903 recovered individuals in 185 countries and regions (Dong et al., 2020a). Early models of COVID-19 transmission dynamics estimate one infectious individual infects slightly over two individuals; travel restrictions reduce that spread to one individual, although these figures might evolve as more accurate epidemiological data become available (Kucharski et al., 2020).

Work during the first SARS-CoV epidemic identified the human host factor angiotensin-converting enzyme 2 (ACE2) as the receptor for SARS-CoV (Li et al., 2003). SARS-CoV-2 spike (S) protein has been experimentally shown to bind ACE2 on host cells with significantly higher affinity than SARS-CoV-S (Hoffmann et al., 2020; Wrapp et al., 2020). The main host protease that mediates S protein activation on primary target cells and initial viral entry is the type II transmembrane serine protease

TMPRSS2 (Glowacka et al., 2011; Hoffmann et al., 2020; Iwata-Yoshikawa et al., 2019; Matsuyama et al., 2010; Shulla et al., 2011; Walls et al., 2020). Other host proteases, such as furin, have also been suggested to promote the pathogenesis of this pandemic SARS-CoV-2 clade, but when and where they process S protein remains to be determined (Böttcher-Friebertshäuser et al., 2013; Bugge et al., 2009; Coutard et al., 2020; Walls et al., 2020). Binding of SARS-CoV-S to ACE2 results in receptor-mediated internalization (Grove and Marsh, 2011; Kuba et al., 2005). Importantly, ACE2 functions as a key tissue-protective component during severe acute lung injury (Imai et al., 2005; Kuba et al., 2005).

A tissue-level basis for understanding SARS-CoV tropism was proposed based on ACE2 histological staining and expression in human epithelia of the lung and small intestine (Hamming et al., 2004; Harmer et al., 2002; Jonsdottir and Dijkman, 2016). However, unlike the specific expression of CDHR3 (the rhinovirus-C receptor), which is resolved to ciliated epithelial cells of the upper airway (Griggs et al., 2017), the specific cell subsets within each tissue that express *ACE2* remain unknown. Identifying the cell subsets targeted by SARS-CoV-2 (ACE2$^+$) and those at greatest risk of direct infection (ACE2$^+$TMPRSS2$^+$) is critical for understanding and modulating host defense mechanisms and viral pathogenesis.

After cellular detection of viral entry into a host cell, interferon (IFN) induction of interferon-stimulated genes (ISGs) is essential for host antiviral defense in mice, non-human primates (NHPs), and humans (Bailey et al., 2014; Deeks et al., 2017; Dupuis et al., 2003; Everitt et al., 2012; Schneider et al., 2014; Utay and Douek, 2016). There are three distinct types of IFNs: type I IFNs (IFN-α and IFN-β), type II IFNs (IFN-γ), and type III IFNs (IFN-λ) (Broggi et al., 2020; Müller et al., 1994; Stetson and Medzhitov, 2006). Each appears to converge on almost indistinguishable responses, mediated through the binding of STAT1 homodimers or STAT1/STAT2 heterodimers to ISGs. However, mounting evidence suggests that each type of IFN might have a non-redundant role in host defense or immunopathology, particularly at epithelial barriers (Broggi et al., 2020; Iwasaki et al., 2017; Iwasaki and Pillai, 2014; Jewell et al., 2010).

Although the host response to SARS-CoV highlighted a role for IFNs, most studies assessed the effect of IFN restriction in cell lines that might not fully recapitulate the repertoire of ISGs present in primary human target cells (Bailey et al., 2014; de Lang et al., 2006; Sainz et al., 2004; Zheng et al., 2004). One study of SARS-CoV suggested the timing of the type I IFN response was critical *in vivo* (Channappanavar et al., 2016). Clinical therapy using approved IFNs has been attempted for SARS-CoV, MERS-CoV, and SARS-CoV-2 in the absence of a controlled trial to mixed effect, resulting in anecdotal evidence suggesting either rapid improvement or worsening of symptoms (Dong et al., 2020b; Lei et al., 2020; Li and De Clercq, 2020). Elucidating tissue- and cell-type-specific ISGs and their activity is essential for understanding the role of IFNs in host defense during human SARS-CoV-2 infection.

Massively parallel single-cell RNA-sequencing (scRNA-seq) is transforming our ability to comprehensively map the cell types, subsets, and states present during health and disease in barrier tissues (Ordovas-Montanes et al., 2020; Ordovas-Montanes et al., 2018; Smillie et al., 2019). This has been particularly evident in the elucidation of novel human epithelial and stromal cell subsets and states (Ordovas-Montanes et al., 2018; Regev et al., 2017; Ruiz Garcia et al., 2019; Schiller et al., 2019; Smillie et al., 2019; Vieira Braga et al., 2019). Recently, scRNA-seq has been applied to better understand the cellular variation present during viral infection *in vitro* and *in vivo* (Russell et al., 2018; Steuerman et al., 2018). Global single-cell profiling efforts such as the Human Cell Atlas (HCA) initiative are ideally poised to rapidly share critical data and enhance our understanding of disease during emergent public health challenges (Sungnak et al., 2020).

Here, using published and unpublished datasets (all from non-SARS-CoV-2-infected samples), we analyze human, NHP, and mouse tissues that have been clinically identified to harbor virus in patients exhibiting COVID-19 symptoms. We provide a cautionary note on the interpretation of the scRNA-seq data presented below, given that many factors such as dissociation, profiling method, and sequencing depth can influence results (STAR Methods). Here, we focus our analysis and discussion on the specific subsets where *ACE2* and *TMPRSS2* are enriched and on relative comparisons *within* each dataset, rather than *between* datasets or equivalence to absolute numbers of total cells. Across several studies of human and NHP tissues, we found ISGs upregulated in *ACE2*-expressing cells.

Strikingly, by treating primary human upper airway basal cells with distinct types of inflammatory cytokines, we demonstrate that IFN-α drives *ACE2* expression. Human influenza infection also induces broader expression of *ACE2* in upper airway epithelial cells and is corroborated by publicly available databases. Overall, our data provide motivation to better understand the trade-offs of antiviral and/or IFN therapy in humans infected with SARS-CoV-2 in order to balance host restriction, tissue tolerance, and viral enhancement mechanisms (Davidson et al., 2015; Fung and Liu, 2019; Imai et al., 2005; Iwasaki et al., 2017; Kuba et al., 2005; Lei et al., 2020; Medzhitov et al., 2012; Zou et al., 2014). Importantly, although our findings identify similar cell subsets enriched for *Ace2* in mice, neither *in vitro* nor *in vivo* IFN-stimulation nor *in vivo* viral challenge substantially alter *Ace2* expression levels. The dynamic, species-specific and multifaceted role of IFN raises implications for pre-clinical COVID-19 disease modeling.

RESULTS

Lung Epithelial Cell Expression of Host Factors Used by SARS-CoV-2 in Non-Human Primates and Humans

To investigate which cells within human and NHP tissues represent likely SARS-CoV-2 targets, we analyzed new and existing scRNA-seq datasets to assess which cell types express *ACE2*, alone or with *TMPRSS2*. In a previously unpublished dataset consisting of NHP (*Macaca mulatta*) lung tissue collected after necropsy of healthy adult animals and analyzed by using Seq-Well v1 (Gierahn et al., 2017), we recovered at least 17 distinct major cell types, including various lymphoid, myeloid, and stromal populations (Figures 1A–1C; Table S1; STAR Methods). *ACE2* and *TMPRSS2* were primarily expressed in epithelial cells, with 6.7% of type II pneumocytes expressing *ACE2* and 3.8% co-expressing *ACE2* and *TMPRSS2* (Figures 1B and 1C).

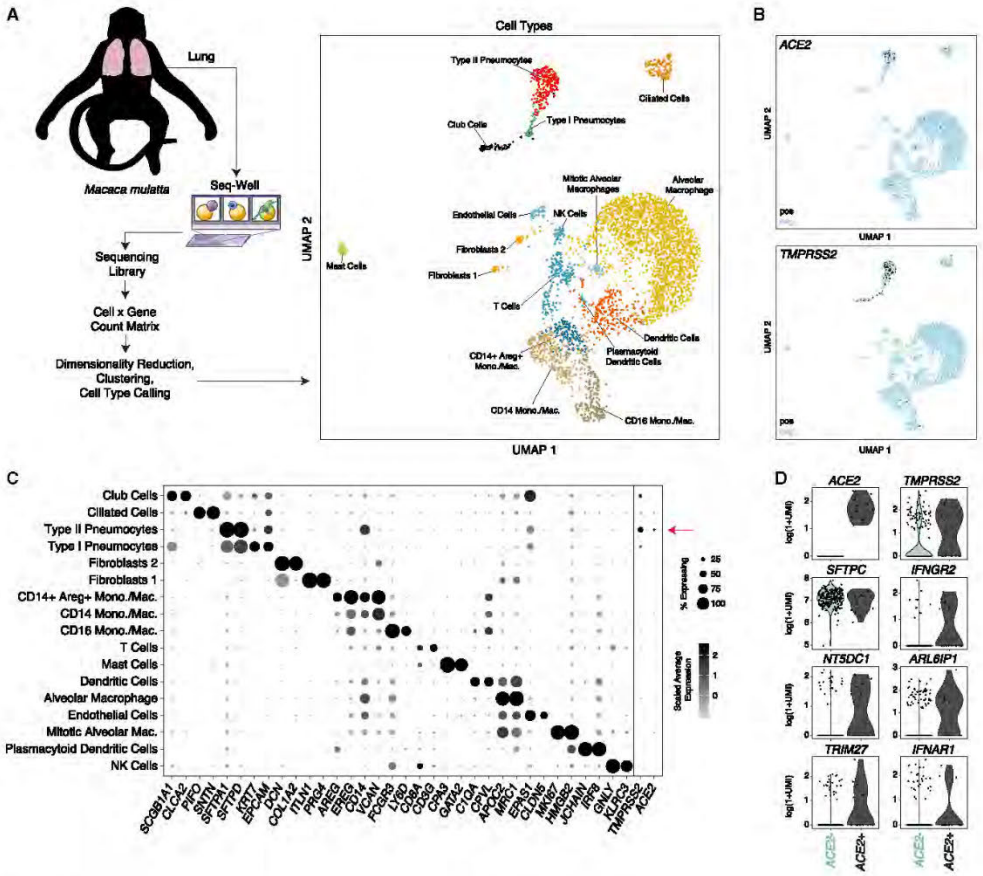


Figure 1. Expression of *ACE2* in Type II Pneumocytes in Healthy Lungs of Non-human Primates

(A) Schematic of protocol for isolation of lung tissue at necropsy from healthy non-human primates (*M. mulatta*, n = 3), creation of scRNA-seq libraries by using Seq-Well v1, and computational analysis to identify cell types by using unbiased methods. UMAP projection of 3,793 single cells, points colored by cell identity (see STAR Methods).

(B) Uniform manifold approximation and projection (UMAP) as in (A), points colored by detection of *ACE2* (coronavirus receptor, top) or *TMPRSS2* (coronavirus S protein priming for entry, bottom). Color coding is as follows: black, RNA positive; blue, RNA negative.

(C) Dot plot of 2 defining genes for each cell type (Table S1) (Bonferroni-adjusted $p < 0.001$) and *ACE2* and *TMPRSS2*. Dot size represents fraction of cells within that type expressing a given gene, and color intensity represents binned count-based expression amount (log(scaled UMI+1)) among expressing cells. *ACE2* is enriched in type II pneumocytes (6.7% expressing, Bonferroni-adjusted $p = 8.62E{-}33$), as is *TMPRSS2* (29.5% expressing, Bonferroni-adjusted $p = 8.73E{-}153$). Of all type II pneumocytes, 3.8% co-express *ACE2* and *TMPRSS2* (Table S9). Red arrow indicates cell type with largest proportion of *ACE2*$^+$*TMPRSS2*$^+$ cells.

(D) Genes differentially expressed among *ACE2*$^+$ and *ACE2*$^-$ type II pneumocytes. (SCDE package, FDR-adjusted $p < 0.05$ for *IFNGR2*, *NT5DC1*, *ARL6IP1*, and *TRIM27*; full results can be found in Table S1).

See also Table S1.

Notably, the only double-positive cells observed were classified within the type II pneumocyte population; however, we also identified *TMPRSS2* expression within club cells, ciliated epithelial cells, and type I pneumocytes, albeit at diminished abundance and frequency compared with type II pneumocytes (Figure 1C; Table S1).

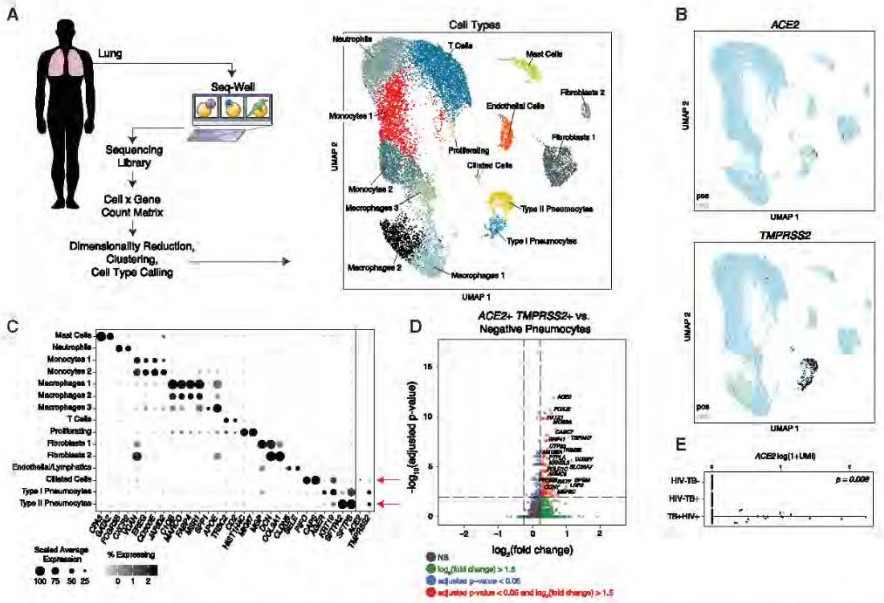


Figure 2. Select Lung Epithelial Cells from Control, HIV-1-Infected, and *Mycobacterium-tuberculosis*-Infected Human Donors Co-Express *ACE2* and *TMPRSS2*

(A) Schematic of protocol for isolation of human lung tissue from surgical excess, creation of scRNA-seq libraries by using Seq-Well S^3, and computational analysis to identify cell types by using unbiased methods. Shown on the right is a UMAP projection of 18,915 cells across 8 donors (n = 3 TB$^+$HIV$^+$; n = 3 TB$^+$; n = 2 non-infected patients). Cells represented by points, colored according to cell type (see STAR Methods).

(B) UMAP projection as in (A), points colored by detection of *ACE2* (top) or *TMPRSS2* (bottom). Color coding is as follows: black, RNA positive; blue, RNA negative.

(C) Dot plot of 2 defining genes for each cell type (FDR-adjusted $p < 0.001$), and *ACE2* and *TMPRSS2*; dot size represents fraction of cells within cell type expressing a given gene, and color intensity represents binned count-based expression amount (log(scaled UMI+1)) among expressing cells. All cluster-defining genes are provided in Table S2. Red arrow indicates cell types with largest proportion of *ACE2*$^+$*TMPRSS2*$^+$ cells.

(D) Volcano plot identifying significantly upregulated genes in *ACE2*$^+$*TMPRSS2*$^+$ pneumocytes compared with all remaining pneumocytes. Red points represent genes with a FDR-adjusted $p < 0.05$, and $\log_2$(fold change) >1.5. Text highlighting specific genes; the full list is available in Table S2.

(E) Expression of *ACE2* across human donors by HIV and TB status ($p = 0.009$ by likelihood-ratio test).

See also Table S2.

Next, we compared *ACE2*$^+$ with *ACE2*$^-$ type II pneumocytes to explore broader gene programs that differentiate putative SARS-CoV-2 target cells from cells of a similar phenotype and ontogeny (Figure 1D; Table S1). Among genes significantly upregulated in *ACE2*$^+$ type II pneumocytes, we observed *IFNGR2* (false discovery rate [FDR]-adjusted $p = 0.022$), a receptor for type II IFNs. Notably, previous work has demonstrated limited anti-viral potency of IFN-γ for SARS-associated coronaviruses, compared with that of type I IFNs, at least *in vitro* (Sainz et al., 2004; Zheng et al., 2004). Other co-regulated genes of potential interest include *TRIM27* (FDR-adjusted $p = 0.025$), as well as *NT5DC1* (FDR-adjusted $p = 0.003$) and *ARL6IP1* (FDR-adjusted $p = 0.047$), which were upregulated in the A549 adenocarcinoma alveolar basal epithelial cell line after exposure to IFN-α and IFN-γ for 6 h (Sanda et al., 2006). We found *IFNAR1* consistently expressed among both *ACE2*$^+$ type II pneumocytes and *ACE2*$^+$*TMPRSS2*$^+$ co-expressing type II pneumocytes, but its level of upregulation compared with all remaining pneumocytes did not meet statistical significance (FDR-adjusted $p = 0.11$). This analysis finds *ACE2*$^+$ cells enriched within a rare fraction of secretory cells in NHPs and that *ACE2* expression is co-regulated with genes involved in IFN responses.

To assess whether the findings from NHP lung cells were similarly present in humans, we analyzed a previously unpublished scRNA-seq dataset derived from surgical resections of fibrotic lung tissue collected with Seq-Well S^3 (Hughes et al., 2019). Unsupervised analysis identified multiple cell types and subtypes of immune cells (Figures 2A–2C; STAR Methods), as defined by the genes displayed in Figure 2C (full lists available in Table S2). Here, we found that *ACE2* and *TMPRSS2* were primarily expressed within type II pneumocytes and ciliated cells, in line with our analysis of the NHP-derived cells (Figures 1 and 2A,

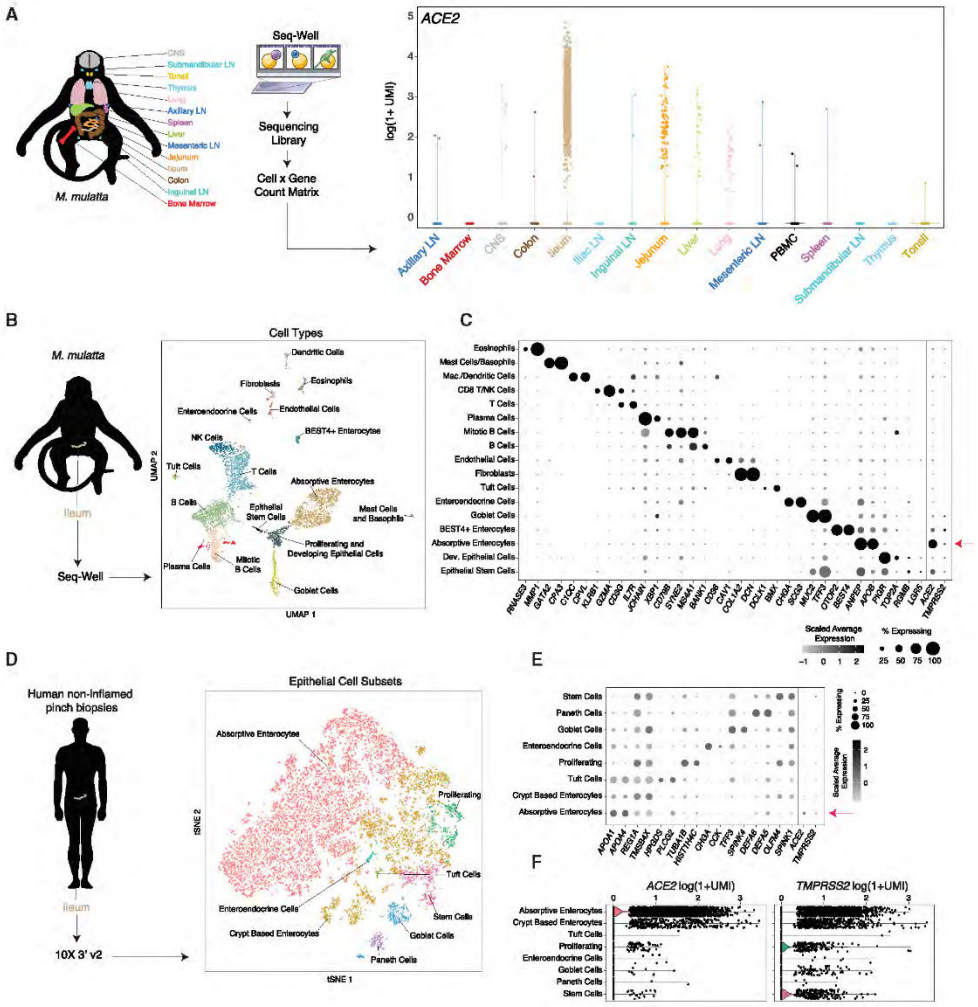


Figure 3. NHP and Human Ileal Absorptive Enterocytes Co-Express *ACE2* and *TMPRSS2*

(A) Expression *ACE2* across diverse tissues in healthy NHPs (n = 3 animals; 52,858 cells).

(B) Schematic of protocol for isolation of NHP ileum (n = 5) at necropsy for scRNA-seq using Seq-Well v1, and computational pipeline to identify cell types by using unbiased methods. Shown on the right is a UMAP projection of 4,515 cells colored by cell type.

(C) Dot plot of 2 defining genes for each cell type, with *ACE2* and *TMPRSS2*. Dot size represents fraction of cells within cell type expressing a given gene, and color intensity represents binned count-based expression amounts (log(scaled UMI+1)) among expressing cells. All cluster defining genes are provided in Table S4. Red arrow indicates cell type with largest proportion of *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells.

(D) Schematic of protocol for isolation of human ileal cells from endoscopic pinch biopsies in non-inflamed regions (n = 13). Shown on the right is a tSNE plot of 13,689 epithelial cells selected from original dataset generated by 10x 3′ *v2* (see Figure S2), colored by cellular subsets.

(legend continued on next page)

2B). In type II pneumocytes (identified by unique expression of surfactant proteins *SFTPC*, *SFTPB*, and *SFTPA1*), we found 1.4% of cells expressing *ACE2* (FDR-adjusted p = 1.35E−21), 34.2% expressing *TMPRSS2* (FDR-adjusted p < 1E−300), and 0.8% co-expressing both. In ciliated cells, we found 7% were *ACE2*$^+$ (FDR-adjusted p = 5E−64), 24.6% were *TMPRSS2*$^+$ (FDR-adjusted p = 3.8E−30), and 5.3% co-expressed both.

As above, to assess for cellular pathways significantly co-expressed within putative target cells for SARS-CoV-2, we computed differentially expressed genes between *ACE2*$^+$*TMPRSS2*$^+$ type II pneumocytes and all other type II pneumocytes (Figures 2C and 2D; Table S2). We found significant enrichment of *BATF* among *ACE2*$^+$*TMPRSS2*$^+$ cells (FDR-adjusted p = 3.25E−7), which has been demonstrated previously to be upregulated by type I and type II IFNs (Murphy et al., 2013). Of note, we also observed *TRIM28* co-expressed with *ACE2* and *TMPRSS2* among type II pneumocytes in this dataset (FDR-adjusted p = 2.34E−9), which might play a role in potentiating an IFN response in lung epithelial cells (Krischuns et al., 2018). Within this cohort of donors, 3 individuals were human immunodeficiency virus (HIV)$^+$ and diagnosed with active tuberculosis, 3 donors had active tuberculosis and were HIV , and 2 were negative for both pathogens. Surprisingly, we found that all of the *ACE2*$^+$ cells across all cell types were derived from HIV$^+$ *Mycobacterium tuberculosis* (Mtb)$^+$ donors despite approximately equivalent recovery of epithelial cell types from all donors (likelihood-ratio test, p = 0.009) (Figure 2E). Given limited cell and patient numbers combined with potential sampling biases, we caution that this observation requires much broader cohorts to validate a potential role for co-infections; still, we note our observation is suggestive of a role for chronic IFNs in the induction of *ACE2*, given that HIV infection is associated with persistent upregulation of ISGs, and we observed elevated amounts of *IFNAR2*, *IFI30*, and *IKBKB* (Utay and Douek, 2016) (FDR-adjusted p = 1.1E−6, 8.8E−9, 1.57E−7, respectively; HIV$^+$ versus HIV epithelial cells).

Next, using a previously unpublished scRNA-seq dataset consisting of granuloma and adjacent, uninvolved lung samples from Mtb-infected NHPs (*Macaca fascicularis*) collected with Seq-Well S^3, we identified subsets of epithelial cells expressing *ACE2* and *TMPRSS2* (Figure S1; Table S3; STAR Methods). The majority of *ACE2*$^+$*TMPRSS2*$^+$ cells were, once again, type II pneumocytes (22%) and type I pneumocytes (9.7%) and were largely enriched within granulomatous regions compared with those in adjacent uninvolved lung (Figures S1B and S1C) (p = 0.006, Fisher Exact Test). *ACE2*$^+$*TMPRSS2*$^+$ type II pneumocytes expressed significantly higher amounts of antimicrobial effectors such as *LCN2* compared with remaining type II pneumocytes (Figure S1D). Cells with club cell/secretory, type I pneumocyte, and ciliated cell types also contained some *ACE2*$^+$*TMPRSS2*$^+$ cells, but we did not have sufficient power to detect significantly differentially expressed genes between these cells and other cells within those clusters. Altogether, we identify *ACE2*$^+$*TMPRSS2*$^+$ cells in lower airways of humans and NHPs with consistent cellular phenotypes and evidence supporting a potential role for IFN-associated inflammation in upregulation of *ACE2*.

Ileal Absorptive Enterocytes Express Host Factors Used by SARS-CoV-2

Next, we examined several other tissues for *ACE2*-expressing cells on the basis of the location of hallmark symptoms of COVID-19, focusing on the gastrointestinal tract due to reports of clinical symptoms and viral shedding (Xiao et al., 2020). Leveraging a previously unpublished scRNA-seq atlas of NHP *(M. mulatta)* tissues collected with Seq-Well v1, we observed that the majority of *ACE2*$^+$ cells reside in the small intestine, principally within the ileum, jejunum, and, to a lesser extent, the liver and colon (Figure 3A; STAR Methods). Critically, we note that, in this experiment, the dissociation method used on each tissue was optimized to preserve immune cell recovery, and therefore under-sampled stromal and epithelial populations, as well as neurons from the brain. Within the ileum, we identified *ACE2*$^+$ cells as absorptive enterocytes on the basis of specific expression of *ACE2* within cells defined by *APOA1*, *SI*, *FABP6*, and *ENPEP*, among others, by a likelihood-ratio test (Figures 3B and 3C) (p < 1E−300, 62% of all absorptive enterocytes; see Table S4). All other epithelial subtypes expressed *ACE2* to a lesser extent, and variably co-expressed *ACE2* with *TMPRSS2* (see Table S4 for full statistics).

Persistent viral RNA in rectal swabs has been detected in pediatric infection, even after negative nasopharyngeal tests (Xu et al., 2020). In an additional dataset consisting of endoscopic biopsies from the terminal ileum of a human pediatric cohort (n = 13 donors, ranging in age from 10 to 18 years old), collected with 10X 3′ v2, we confirmed a large abundance of *ACE2*$^+$ cells with selective expression within absorptive enterocytes (29.7% *ACE2*$^+$, FDR-adjusted p = 2.46E−100) (Figures 3D and 3E; Table S5; STAR Methods). Furthermore, we identified a subset (888 cells, ~6.5% of all epithelial cells) that co-express both genes (Figures S2A–S2C). We performed differential expression testing and GO-term enrichment using these cells relative to matched non-expressers to highlight putative biological functions enriched within them, such as metabolic processes and catalytic activity, and to identify shared phenotypes of *ACE2*$^+$*TMPRSS2*$^+$ ileal cells across both human and NHP cohorts (Table S5). We speculate that viral targeting of these cells, taken from patients without overt clinical viral infection, might help explain intestinal symptoms. Finally, we compared ileal absorptive enterocytes from healthy NHPs and NHPs infected with simian-human immunodeficiency virus (SHIV) and then treated for 6 months with antiretroviral therapy (animal and infection characteristics published in Colonna et al., 2018) (STAR Methods). We found significant upregulation of *ACE2*, *STAT1*, and *IFI6* within the absorptive

(E). Dot plot of 2 defining genes for each cell type, with *ACE2* and *TMPRSS2*. Dot size represents fraction of cells within cell type expressing a given gene, and color intensity represents binned count-based expression amounts (log(scaled UMI+1)) among expressing cells. All cluster defining genes are provided in Table S5. Red arrow indicates cell type with largest proportion of *ACE2*$^+$*TMPRSS2*$^+$ cells.
(F). Expression of *ACE2* (left) and *TMPRSS2* (right) among all epithelial subsets from human donors.
See also Figure S2 and Tables S4 and S5.

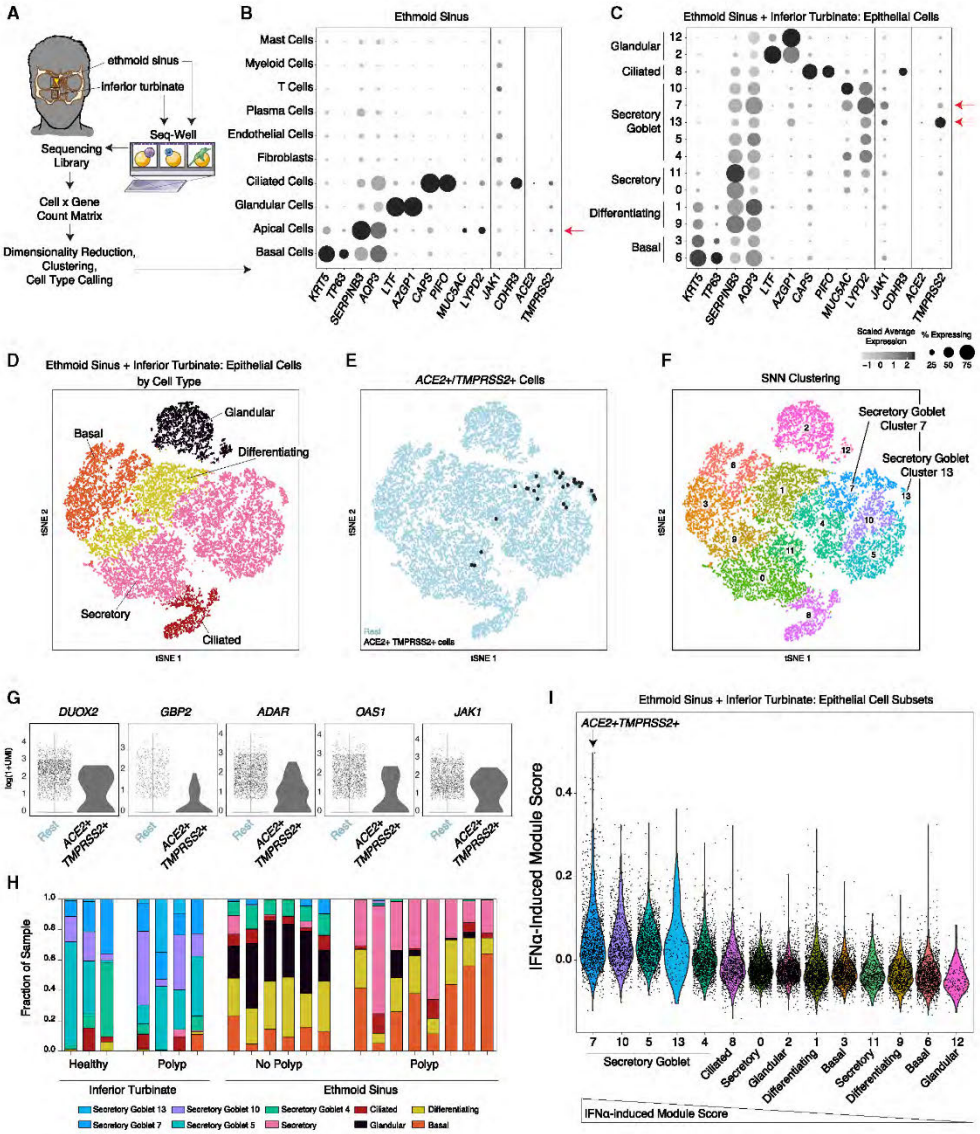


Figure 4. Healthy and Allergic Inflamed Human Nasal Mucosa Co-Express *ACE2* and *TMPRSS2* in a Subset of Goblet Secretory Cells

(A) Schematic for sampling of n = 12 ethmoid sinus surgical samples and n = 9 inferior turbinate nasal scrapings to generate scRNA-seq libraries by using Seq-Well v1. See Ordovas-Montanes et al., (2018).

(B) Dot plot of all cell types from ethmoid-sinus-derived cells (n = 6 non-polyp CRS samples, n = 6 polyp CRS samples). Two defining genes for each cell type, in addition to *CDHR3* (rhinovirus receptor), *ACE2*, *TMPRSS2*, and *JAK1*. Dot size represents fraction of cells within that type expressing a given gene, and color

(legend continued on next page)

enterocytes of SHIV-infected animals (which maintain chronically elevated amounts of IFNs and ISGs) compared with those of uninfected controls (FDR-adjusted p < 2E-7) (Figure S2D) (Deeks et al., 2017; Utay and Douek, 2016).

Upper Airway Expression of Host Factors Used by SARS-CoV-2

To identify potential viral target cells in nasal and sinus tissue, two regions that are frequently primary sites of exposure for coronaviruses, we analyzed existing scRNA-seq datasets from the human upper airway (inferior turbinate and ethmoid sinus mucosa) across a spectrum of healthy donors and individuals with allergic inflammation due to chronic rhinosinusitis (CRS) collected with Seq-Well v1 (Figure 4A; STAR Methods) (Ordovas-Montanes et al., 2018). We had previously noted a significantly enriched IFN-dominated gene signature in inferior turbinate secretory epithelial cells from both healthy and CRS donors compared with CRS samples from the ethmoid sinus, which were significantly enriched for interleukin-4 (IL-4)/IL-13 gene signatures (Giovannini-Chami et al., 2012; Ordovas-Montanes et al., 2018). We speculate that these cells, taken from clinically non-virally infected patients, yet constantly exposed to environmental viruses, might provide one of the earliest locations for coronaviruses to infect before spreading to other tissues. We observed significant enrichment of *ACE2* expression in apical epithelial cells and, to a lesser extent, ciliated cells compared with all cell types recovered from surgically resected mucosa (1% of apical epithelial cells, FDR-adjusted p = 4.55E−6, n.s. in ciliated cells) (Figure 4B; Table S6).

To better map putative SARS-CoV-2 targets among epithelial subsets, we employed a finer-grained clustering method applied to both ethmoid sinus surgical specimens and scrapings from the inferior turbinate and ethmoid sinus (Figures 4C–4F). Once again, we observed selective expression of *ACE2* within a minority of cell types, with 1.3% of all secretory cells expressing *ACE2* (Figure 4C) (FDR-adjusted p = 0.00023), specifically sub-clusters 7 and 13, which represent two varieties of secretory epithelial cell (Figures 4C, 4F, and 4G). Cluster 7 secretory cells are marked by *S100P*, *LYPD2*, *PSCA*, *CEACAM5*, and *STEAP4*; encompass some *MUC5AC* goblet cells; and contain the most significantly enriched *ACE2* and *TMPRSS2* expression (4% express *ACE2*, FDR-adjusted p = 7.32E−28; 28% express *TMPRSS2*, FDR-adjusted p = 2.15E−132; Table S6). We next explicitly gated cells by their *TMPRSS2* and *ACE2* expression, identifying a rare subset that co-expresses both, the majority of which fall within the "Secretory Cluster 7" cell type (Figures 4E and 4F) (30 cells, ~0.3% of all upper airway secretory cells, 1.6% of goblet "Secretory Cluster 7"). These findings are aligned with concurrent work by the HCA Lung Biological Network on human nasal scRNA-seq data, which identified nasal secretory cells to be enriched for *ACE2* and *TMPRSS2* expression (Sungnak et al., 2020).

Although we identified co-expression of *ACE2 and TMPRSS2* in few airway cells overall, we detected *ACE2* and *TMPRSS2* single- and double-positive cells in over 20 donors and thus posit that these genes are enriched in secretory cells and are not a product of individual-patient-driven variability (Figure S3A). Inferior turbinate scrapings collected on Seq-Well S^3, which increases the resolution of lower-abundance transcripts compared with Seq-Well v1, revealed consistent and specific expression restricted to goblet secretory cells, but at a greater detection frequency in samples from the same donors (Figure S3B) (*ACE2*+ from 4.7% v1 to 9.8% S^3; *ACE2*+*TMPRSS2*+ from 1.9% v1 to 4% S^3) (Hughes et al., 2019). Using the gated *ACE2*+*TMPRSS2*+ cells, we tested for differentially expressed genes compared to the remaining secretory epithelial cells (full results provided in Table S6). Notably, we observed significant upregulation of *ADAR*, *GBP2*, *OAS1*, *JAK1*, and *DUOX2* (FDR adjusted, all p < 0.02) within *ACE2*+*TMPRSS2*+ cells, potentially indicative of IFN signaling (Figure 4G). Almost all "Secretory Cluster 7" cells were from inferior turbinate scrapings of healthy and allergically inflamed individuals, few cells were from the ethmoid sinus tissue of patients with chronic rhinosinusitis without nasal polyps, and no cells were detected in polyp tissue (Figure 4H). Gene Ontology (GO) analysis of enriched genes in double-positive cells include processes related to intracellular cytoskeleton and macromolecular localization and catabolism, potentially involved in viral particle entry, packaging, and exocytosis (Fung and Liu, 2019).

We next utilized IFN-inducible gene sets of relevance to human airway epithelial cells, which we derived from a prior study by performing differential expression on a published dataset

intensity represents binned count-based expression amounts (log(scaled UMI+1)) among expressing cells (see Table S6 for statistics by subset). Red arrow indicates cell types with largest proportion of *ACE2*+*TMPRSS2*+ cells.

(C) Dot plot for 2 defining genes for each cell type identified from granular clustering of epithelial cells (18,325 single cells) derived from both ethmoid sinus and inferior turbinate sampling (healthy inferior turbinate [3,681 cells; n = 3 samples], polyp-bearing patient inferior turbinate [1,370 cells; n = 4 samples], non-polyp ethmoid sinus surgical samples [5,928 cells; n = 6 samples], and polyp surgical and scraping samples directly from polyp in ethmoid sinus [7,346 cells; n = 8 samples]). Red arrow indicates cell type with largest proportion of *ACE2*+*TMPRSS2*+ cells.

(D) tSNE of 18,325 single epithelial cells from inferior turbinate and ethmoid sinus (omitting immune cells). Colored by cell types 3,152 basal, 3,089 differentiating, 8,840 secretory, 1,105 ciliated, and 2,139 glandular cells.

(E) tSNE as in (D), identifying epithelial cells co-expressing *ACE2* and *TMPRSS2* (30 cells, black points).

(F) tSNE as in (D), colored by detailed cell types with higher granularity, as in (C).

(G) Individual differentially expressed genes between *ACE2*+*TMPRSS2*+ cells and all other secretory epithelial cells (see Table S6 for full gene list with statistics). Bonferroni-adjusted likelihood-ratio test p < 0.02 for all genes displayed.

(H) Stacked bar plot of each subset of epithelial cells among all epithelial cells by donor (each bar) and sampling location (noted below graph) (unpaired t test p < 0.00035 for Secretory Goblet 7 inferior turbinate versus ethmoid sinus; see Table S6 for raw values).

(I) Violin plot of cell clusters in respiratory epithelial cells (from Figures 4C and 4F) ordered by average expression of IFN-α-induced gene signatures, presented as a gene module score; non-normal distribution by Lilliefors test, Mann-Whitney U-test p = 2.2E−16, 1.21 effect size, IFN-α signature for Secretory Goblet Cluster 7 versus all epithelial cells. Arrow indicates cluster containing majority *ACE2*+*TMPRSS2*+ cells.

See also Figure S3 and Table S6.

where air-liquid interface cultures from primary human nasal epithelial cells were treated with IFN-αA/D, IFN-β1a, IFN-γ, IL-4, or IL-13 (Giovannini-Chami et al., 2012; Ordovas-Montanes et al., 2018). Using these gene lists, we scored the human nasal epithelial cells analyzed by scRNA-seq described in Figures 4C and 4F and found significant concomitant upregulation of the IFN-α-stimulated gene set within *ACE2*$^{+}$*TMPRSS2*$^{+}$ secretory goblet cluster 7 (Figure 4I).

Type I Interferon IFN-α Drives *ACE2* Expression in Primary Human Nasal Epithelial Cells

The meta-analysis described above consistently identified an association between *ACE2* expression and canonical ISGs or components of the IFN-signaling pathway. This prompted us to investigate whether IFNs might play an active role in regulating *ACE2* expression levels in specific target cell subsets, thus potentially allowing for a tissue-protective host response or increased viral binding of SARS-CoV-2 through ACE2. Our initial literature search indicated that IFN-γ and IL-4 downregulate the SARS-CoV receptor ACE2 in Vero E6 cells (African green monkey kidney epithelial cells [de Lang et al., 2006]), appearing to invalidate this hypothesis. Relatedly, *in vitro* stimulation of A549 cells, a commonly used cell line model for lung epithelia, with IFN-α, IFN-γ, and IFN-α+IFN-γ for 24 h did not identify *ACE2* as an ISG (Russell et al., 2018). This is potentially explained by recent work that aimed to understand SARS-CoV-2 receptor usage by performing screening studies within cell line models and found that A549 cells did not express *ACE2* and therefore represents a poor model to understand regulation of this gene (Letko et al., 2020). While conducting experiments to directly test the hypothesis that *ACE2* is an ISG, we noted in our own gene lists used for scoring from Ordovas-Montanes et al., 2018 and in a supplementary extended table available from Giovannini-Chami et al., 2012 that *ACE2* was in upregulated gene lists after exposure to Type I IFN.

We directly tested whether IFN-α induces *ACE2* in primary human upper airway epithelial cells in greater detail. We cultured human primary basal (stem and progenitors) epithelial cells to confluence and treated them with increasing doses (0.1–10 ng/mL) of IFN-α2, IFN-γ, IL-4, IL-13, IL-17A, or IL-1β for 12 h and then performed bulk RNA-seq (Figure S3C). Only IFN-α2 and IFN-γ led to upregulation of *ACE2* over the time period tested, and compared with all other cytokines, IFN-α2 lead to greater and more significant upregulation over all doses tested (Figure S3D,Wilcoxon test: IFN-α2 FDR-adjusted $p = 4.1E{-}07$; IFN-γ $p = 9.3E{-}03$,Figures S3E and S3F, all statistical tests compared with 0 ng/mL dose). We confirmed substantial and dose-dependent induction of canonical members of the interferon response after IFN-α2 and IFN-γ (Figures S3G and S3H). Conversely, we found that IFN-γ, relative to IFN-α2, induced potent upregulation of *GBP5*, a GTPase-like protein thought to act as a viral restriction factor through inhibiting furin-mediated protease activity, which could limit viral processing from infected cells, whereas IFN-α2 more robustly induced *IFITM1* (Figure S3G–S3K) (Braun and Sauter, 2019).

To further extend and substantiate these findings, as above, we stimulated primary mouse tracheal basal cells, the commonly used human bronchial cell line BEAS-2B, and upper airway basal cells from two human donors (Figure 5A-D). We confirmed appropriate induction of an IFN response in each cell type by performing differential expression testing between untreated cells and IFN-treated cells for each condition (Table S7). Within each cell type, stimulation with IFN-α2, IFN-γ, or IFN-β resulted in dose-dependent upregulation of canonical ISGs, including *STAT1/Stat1*, *BST2/Bst2*, *XAF1/Xaf1*, *IFI35/Ifi35*, *MX1/Mx1*, and *GBP2/Gbp2*. Notably, *Ace2* expression was not robustly induced in basal cells derived from healthy mouse trachea under any interferon stimulation condition (Figure 5A). The magnitude of *ACE2* upregulation was diminished in BEAS-2B cells compared to that in our original findings in primary human upper airway epithelial cells, but reached statistical significance compared with that of the untreated condition after IFN-γ exposure (Figure 5B). In primary basal cells derived from healthy nasal mucosa, we confirmed significant induction of *ACE2* after IFN-α2 stimulation and, to a lesser extent, after stimulation with IFN-γ (IFN-α2-stimulated: both Bonferroni-adjusted $p < 0.001$; IFN-γ-stimulated: both Bonferroni-adjusted $p < 0.05$) (Figures 5C and 5D). Expression of *ACE2* was significantly correlated with expression of *STAT1* in all human cell types, with a larger effect size and correlation coefficient in primary human basal cells (Figure 5E-H). These experiments support a relationship between induction of the canonical IFN response, including key transcription factors and transcriptional regulation of the *ACE2* locus. Finally, among primary human samples, we confirmed the dose-dependence of *ACE2* upregulation after IFN-α2 or IFN-γ treatment and significant induction of *ACE2* after IFN-α2 stimulation at concentrations as low as 0.1–0.5 ng/mL (Figure 5I-L).

Next, using a publicly available resource (interferome.org) that hosts genomic and transcriptomic data from cells or tissues treated with IFN, we queried *ACE2* expression within human and mouse cells, searching for datasets with a $\log_2$-fold-change of >1 or < −1 compared with untreated samples, including all IFN types (Rusinova et al., 2013). We recovered 21 datasets spanning 8 distinct primary tissues or cell lines with non-trivial changes in *ACE2* expression after both type I and type II IFN treatment (Figure S4A). We observed substantial upregulation of *ACE2* in primary skin and primary bronchial cells treated with either type I or type II IFN (> 5-fold upregulation compared with that in untreated cells), in strong support of our *in vitro* data (Figures 5C, 5D, 5G–5L, and S3D–S3F). Immune cell types, such as CD4 T cells and macrophages, were noticeably absent from datasets with a significant change in *ACE2* expression after IFN stimulation or were even found to downregulate *ACE2* (e.g., primary CD4 T cells + type I IFN) (Figure S4A, and in our analysis of scRNA-seq peripheral blood mononuclear cell data from Butler et al., (2018); data not shown).

Given that the majority of cells robustly upregulating *ACE2* were epithelial, this observation potentially explains why previous analyses to define canonical ISGs within immune populations did not identify *ACE2* as an induced gene. Furthermore, using both Transcription Factor database (TRANSFAC) data hosted by the interferome database, as well as chromatin immunoprecipitation sequencing (ChIP-seq) data (provided by the ENCODE Factorbook repository), we found evidence for STAT1, STAT3, IRF8, and IRF1 binding sites within −1500–

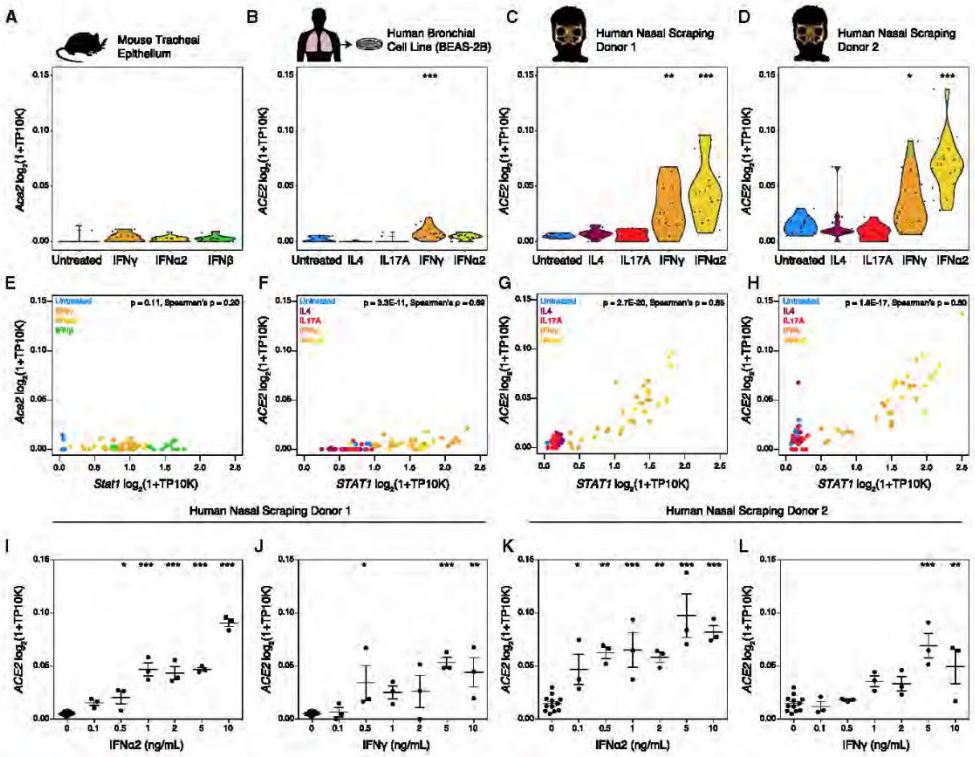


Figure 5. *ACE2* is an Interferon-Stimulated Gene in Primary Human Barrier Tissue Epithelial Cells

(A–D) Basal epithelial cells from distinct sources were cultured to confluence and treated with increasing doses (0.1–10 ng/mL) of IFN-α2, IFN-γ, IL-4, IL-17A, and/or IFN-β for 12 h and bulk RNA-seq analysis was performed. Expression of *ACE2* (human) or *Ace2* (mouse) by cell type and stimulation condition. (A) Primary mouse basal cells from tracheal epithelium are shown. (B) BEAS-2B human bronchial cell line is shown. (C) Primary human basal cells from nasal scraping, Donor 1, is shown. (D) Primary human basal cells from nasal scraping, Donor 2. Abbreviation is as follows: TP10K, transcripts per 10,000 reads. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, Bonferroni-corrected t test compared with untreated condition.

(E–H) Co-expression of *STAT1*/*Stat1* and *ACE2*/*Ace2* by cell type. (E) Primary mouse basal cells from tracheal epithelium are shown. (F) BEAS-2B human bronchial cell line is shown. (G) Primary human basal cells from nasal scraping, Donor 1, are shown. (H) Primary human basal cells from nasal scraping, Donor 2 are shown. Abbreviation is as follows: TP10K, transcripts per 10,000 reads. Statistical significance assessed by Spearman's rank correlation.

(I–L) Expression of *ACE2* in primary human basal cells from nasal scrapings across a range of concentrations of IFN-γ or IFN-α2. (I) IFN-α2 dose response in Donor 1 ($p < 0.001$ by one-way ANOVA) is shown. (J) IFN-γ dose response in Donor 1 ($p < 0.01$ by one-way ANOVA) is shown. (K) IFN-α2 dose response in Donor 2 ($p < 0.001$ by one-way ANOVA) is shown. (L) IFN-γ dose response in Donor 2 ($p < 0.001$ by one-way ANOVA). Abbreviation is as follows: TP10K, transcripts per 10,000 reads. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, Bonferroni-corrected post hoc testing compared with 0 ng/mL condition.

See also Figures S3 and S4 and Table S7.

500 bp of the transcription start site of *ACE2* (all in human studies, Figure S4B) (Gerstein et al., 2012; Matys et al., 2003; Wang et al., 2012; Wang et al., 2013). This finding is supportive of our current hypothesis that *ACE2* represents a previously unappreciated ISG in epithelial cells within barrier tissues.

Given minimal upregulation of *Ace2* among primary mouse basal cells *in vitro*, we were curious as to whether *Ace2* represented a murine ISG *in vivo*. We treated two mice intranasally with saline and two mice intranasally with 10,000 units of IFN-α (Guerrero-Plata et al., 2005). After 12 h, we isolated the nasal mucosa, consisting of both respiratory and olfactory epithelium, with underlying lamina propria, and performed scRNA-seq using Seq-Well S^3 (Figure S5A). We collected from both tissue sites because of early reports of anosmia in COVID-19 (Lechien et al., 2020). We recovered 11,358 single cells, including epithelial, stromal, neuronal, and immune cell types, generating the largest single-

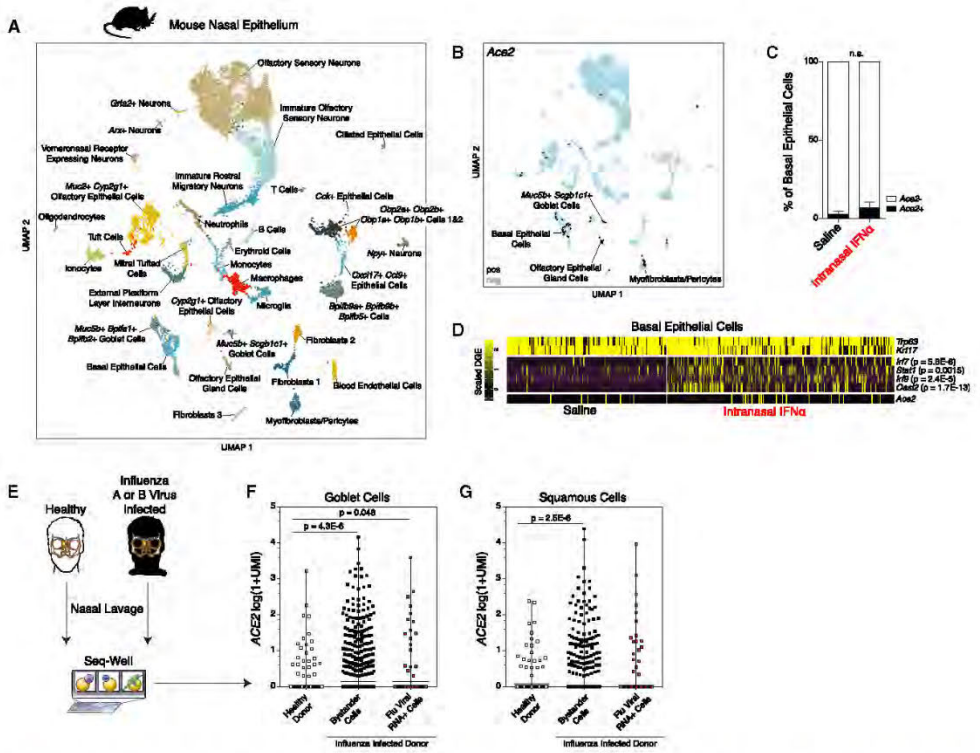


Figure 6. *In Vivo* Administration of Interferons in Mice Does Not Induce *Ace2*, and *ACE2* Is Induced in Goblet Secretory Cells during Human Influenza Infection

(A) UMAP of 11,358 single cells from mouse nasal epithelium (n = 4).

(B) UMAP projection as in (A), points colored by detection of *Ace2* (SARS-CoV-2 receptor homolog). Color coding is as follows: black, RNA positive; blue, RNA negative.

(C) Percent of *Ace2*$^+$ cells by treatment condition (n = 4 arrays per condition; n = 2 arrays per mouse). Black bars indicate *Ace2*$^+$ cells; white bars indicate *Ace2* cells. p = 0.4 by Student's t test.

(D) Heatmap of cell-type-defining genes (*Trp63* and *Krt17*), interferon-induced genes (*Irf7*, *Stat1*, *Irf9*, and *Oasl2*), and *Ace2* among basal epithelial cells, separated by cells derived from saline-treated mice (left) and IFN-α-treated mice (right). Statistical significance by likelihood-ratio test with Bonferroni correction is shown. A full list of differentially expressed genes can be found in Table S8.

(E) Schematic for sampling cells derived from nasal washes of n = 18 human donors with and without current influenza A or B infection for Seq-Well v1 (35,840 single cells). See Cao et al., (2020).

(F and G) *ACE2* expression among goblet cells (F) and squamous cells (G) by infection status. Shown are Healthy Donor cells from influenza-negative donors (white); Bystander Cells from influenza A (IAV)- or influenza B (IBV)-infected donors, no intracellular viral RNA detected (black); Flu Viral RNA$^+$ Cells with detectable intracellular influenza A or B viral RNA (red). Statistical significance by Wilcoxon test with Bonferroni correction, n.s. for Bystander versus Flu Viral RNA$^+$.

See also Figure S5 and Tables S6 and S8.

cell atlas of mouse respiratory and olfactory mucosa to date (Figures 6A and S5B). We annotated all 36 clusters, focusing our attention on epithelial cell clusters, given that we noted enrichment for *Ace2* and *Tmprss2* within epithelial cell subsets, consistent with our human and NHP results (Table S8). Specifically, we found *Ace2* enriched within olfactory epithelial gland cells, *Muc5b*$^+$*Scgb1c1*$^+$ goblet cells, basal epithelial cells, and myofibroblasts/pericytes (Bonferroni-corrected p < 0.01) (Figures 6B and S5B) (Brann et al., 2020; Dear et al., 1991; Montoro et al., 2018; Tepe et al., 2018). Notably, *Furin* was enriched within olfactory epithelial gland cells (Table S8). Next, we asked whether a 12 h stimulation with IFN-α would upregulate *Ace2 in vivo*. Focusing on basal epithelial cells, which contain the highest abundance of *Ace2*$^+$ cells, we found that despite robust upregulation of canonical murine ISGs, *Ace2* expression was only slightly elevated after IFN-α treatment (Figures 6C, 6D, S5C, and S5D).

This observation was supported by analysis of scRNA-seq data from 5,558 epithelial cells from the lungs of mice 3–6 days after intranasal infection with murine gamma herpesvirus-68 (MHV68) (Figure S5E). Here, we found significant enrichment of *Ace2*$^{+}$ cells within type II pneumocytes, in line with our data from NHP and human lungs (Figures S5F). We did not observe changes in *Ace2* expression among viral-transcript-positive cells or "bystander" type II pneumocytes (those without detectable cell-associated viral RNA in MHV68-infected animals), nor did we see significant alterations in *Ace2*$^{+}$ cell abundance among MHV68-infected mice lacking IFN-γR (Figure S5G and S5H). These observations were in agreement with our *in vitro* murine basal cell assay (Figure 5A and 5E).

Finally, we sought to validate our hypothesis that *ACE2* is upregulated in human epithelial cells during upper airway viral infections, which are known to induce a robust IFN response (Bailey et al., 2014; Everitt et al., 2012; Iwasaki and Pillai, 2014; Jewell et al., 2010; Russell et al., 2018; Steuerman et al., 2018). We reanalyzed a publicly available dataset of RNA-seq from human lung explants isolated after surgical resections that were infected with influenza A virus *ex vivo* for 24 h. Here, we found that *ACE2* expression was significantly correlated with that of *SFTPC*, supporting our hypothesis that *ACE2* is expressed within type II pneumocytes (Figures 1C, 2C, S5I, and S5J) (Matos et al., 2019). Furthermore, although the abundance of *SFTPC* was not significantly altered by influenza A virus infection, *ACE2* expression was significantly upregulated after viral exposure ($p = 0.0054$, ratio paired t test) (Figures S5K and S5L). This suggests that influenza A virus infection increases *ACE2* expression. Nevertheless, these population-level analyses are not able to definitively resolve specific cell subsets of relevance, nor whether they are directly infected cells or bystanders of infection.

In order to address these questions, we leveraged an ongoing scRNA-seq study of nasal washes from 18 individuals with confirmed influenza A virus or influenza B virus infection or healthy controls collected with Seq-Well v1, which yielded 35,840 cells resolved into 17 distinct cell types (Figure 6E; STAR Methods) (Cao et al., 2020). We investigated the cell types with greatest enrichment for *ACE2* and *TMPRSS2* in non-infected controls and individuals with influenza A and B. Strikingly, *ACE2* was most upregulated in samples from influenza-virus-infected individuals within bystander goblet or squamous cells not directly infected by virus (Figures 6F and 6G). *ACE2*$^{+}$*TMPRSS2*$^{+}$ goblet cells during influenza infection exhibited enrichment for canonical ISGs such as the *CXCL9/CXCL10/CXCL11* gene cluster; correspondence with *ACE2*$^{+}$*TMPRSS2*$^{+}$ goblet cells in healthy and allergic nasal scrapings; and a shared overlap in ISGs including *GBP2*, *ZNFX1*, *ADAR*, and *ACE2* (significantly differentially expressed gene lists) (Table S6). Together, our data suggest that *ACE2* is an ISG *in vitro* and *in vivo* in human primary upper airway epithelial basal cells, but that the murine homolog *Ace2* is not in airway epithelial basal cells or pulmonary epithelial cells *in vitro* or *in vivo*. Collectively, our findings suggest that careful considerations of animal and cellular models will be needed for assessing therapeutic interventions targeting the IFN system when studying *ACE2/Ace2*-associated biology.

Finally, because our *in vivo* and *in vitro* work indicate that IFN might promote human cellular targets for SARS-CoV-2 infection in the human upper airway by inducing *ACE2*, we attempted to extend our transcriptomic data on IFN-driven expression of *ACE2* to protein-level induction of ACE2. As testing of various commercially available polyclonal antibody preparations found broad evidence for non-specific or inconclusive staining in histological immunofluorescent based readouts (data not shown), we assessed whether IFN-γ-stimulated human bronchial air-liquid interface cultures induced ACE2 within 24 h. Our results show that cells from one patient robustly induced ACE2 (+2.02x), cells from another mildly induced ACE2 (+1.21x) and two patient's cells showed minor changes (+/−1.12x) (Figure S5M). We provide a note of caution as these cells were derived from asthmatic patients, and the overall changes did not reach significance. Furthermore, we could not determine cell surface localization of ACE2 but do note that these results align with our transcriptomic data.

DISCUSSION

Here, we utilize scRNA-seq across various barrier tissues and model organisms to identify the potential initial cellular targets of SARS-CoV-2 infection. To review the data presented: (1) we found that expression of the cellular entry receptor for SARS-CoV-2, *ACE2*, is primarily restricted to type II pneumocytes in the lung, absorptive enterocytes within the gut, and goblet secretory cells of the nasal mucosa; (2) *ACE2* and *TMPRSS2* co-expression in respiratory tissues is consistently found only among a rare subset of epithelial cells; (3) we observed similarities in the cellular identities and frequencies of putative SARS-CoV-2 target cells across human and NHP cohorts; (4) we observe increased expression of *ACE2* during SHIV and TB infection of NHPs, and HIV/TB co-infection and influenza infection of humans compared with that in matched controls but caution that none of the datasets presented here were designed to answer this specific query. Specific targeting of these cell subsets has only been described for a handful of viruses, including the following: goblet cells by human adenovirus-5p and enterovirus 71, type II pneumocytes by H5N1 avian influenza, and absorptive enterocytes by rotavirus (Fleming et al., 2014; Good et al., 2019; Holly and Smith, 2018; Weinheimer et al., 2012).

Additionally, we provide an overall note of caution when interpreting scRNA-seq data for low abundance transcripts like *ACE2* and *TMPRSS2* because detection inefficiencies might result in an underestimation of the actual frequencies of *ACE2*$^{+}$ or *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells in a tissue. Moreover, the protein amounts of each might differ from their mRNA abundances (Genshaft et al., 2016; Jovanovic et al., 2015; Rabani et al., 2011; Shalek et al., 2013). We also present datasets separately, given that each study differed in its methods of tissue processing and collection, which can influence the frequency of recovered cell subsets (STAR Methods). We provide Table S9 as a summary of *ACE2*$^{+}$ and *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells across various datasets. Moreover, we present Figure S6, which describes statistical modeling and power calculations underlying detection and dropout of *ACE2*, to help guide interpretation of these data. This includes an examination of the probability to detect a lowly expressed transcript like *ACE2* within a cell, as well as upper bound estimates on the percentage of positive cells within a

cluster, considering the effects of transcript counts, sequencing depth, and cell numbers in these calculations (STAR Methods).

Whether ACE2 and TMPRSS2 are needed on the same cell or soluble proteases can activate SARS-CoV-2 S protein to invade ACE2 single-positive cells is an area of active inquiry (Coutard et al., 2020; Letko et al., 2020). Importantly, rapidly evolving literature has identified that SARS-CoV-2-S might have a furin cleavage site, leading to a broader set of host proteases that could mediate S protein activation (Bugge et al., 2009; Coutard et al., 2020; Walls et al., 2020). However, because an active S protein has a finite lifetime to find a target cell membrane, the timing and cellular location of S protein activation is key to consider. Activation events proximal to the plasma membrane have been shown to be most effective for SARS-CoV entry (Shulla et al., 2011).

Our study finds that type I IFNs, and to a lesser extent type II IFNs, upregulate *ACE2*. This is based on several lines of evidence: (1) we identified a human goblet secretory cell subset in upper airway nasal epithelium enriched for *ACE2* expression to have the highest IFN-α-induced gene signature; (2) we found that IFN-α, and to a lesser extent IFN-β or IFN-γ, induced *ACE2* expression in a published dataset of air-liquid interface cultures derived from human nasal epithelial cells (Giovannini-Chami et al., 2012; Ordovas-Montanes et al., 2018); (3) we extended our search through the Interferome database (Rusinova et al., 2013) and found that, in epithelial barrier tissues, type I IFNs upregulate *ACE2* in multiple studies, especially in primary bronchial cells and keratinocytes (Rusinova et al., 2013); (4) we found two STAT1 binding sites in the promoter of *ACE2*; (5) in our unpublished atlas of SHIV-infected macaques, known to have elevated amounts of chronic IFN signaling, we found *ACE2* upregulation in absorptive enterocytes; (6) we directly provided evidence for IFN-α, and to some extent IFN-γ, inducing *ACE2* expression in primary human upper airway basal cells; and (7) influenza infection in humans, a known inducer of the IFN pathway, leads to increased *ACE2* expression in goblet secretory cells of the nasal epithelium (Cao et al., 2020).

Altogether, our own and publicly available data highlight that *ACE2* might have been missed as a canonical ISG because of its notable absence in peripheral blood mononuclear cell datasets and in lung-derived transformed cell lines such as the A549 cell line (Butler et al., 2018; Letko et al., 2020; Rusinova et al., 2013). Importantly, other groups have independently analyzed publicly available datasets, some referenced in our work, and observed ACE2's behavior as an ISG (Wang and Cheng, 2020). Furthermore, we found weak IFN- or virally driven induction of *Ace2* in murine cells and tissues. This highlights the importance of studying primary human epithelial cells and the careful consideration of appropriately selected gene lists and *in vitro* models of *in vivo* cellular systems for understanding human biology (Jonsdottir and Dijkman, 2016; Mead and Karp, 2019; Regev et al., 2017).

As SARS-CoV-S leads to ACE2-receptor-mediated internalization, the host IFN response could thus promote the ability for SARS-CoV and SARS-CoV-2 to maintain cellular targets in neighboring human upper airway epithelial cells. Altogether along with a study of HCoV-OC43, which co-opts IFN-inducible transmembrane 2 (IFITM2) and IFITM3 to promote viral entry, this adds to the growing evidence that coronaviruses, as well as other viruses, have evolved to leverage features of the human IFN pathway (Fung and Liu, 2019; Mar et al., 2018; Zhao et al., 2014). Whether type I IFNs are net protective or detrimental to the host might depend on the stage of infection; cell subsets in question; the SARS viral clade (Channappanavar et al., 2016; Channappanavar et al., 2019; Channappanavar and Perlman, 2017; Davidson et al., 2015); and other factors such as co-infection, age, gender, and co-morbidities, among others. Understanding the specific host restriction factors targeting SARS-CoV-2 and identifying specific drivers of these genes in the absence of *ACE2* upregulation might provide strategies to dissociate the dual roles of IFN in certain coronavirus infections. Whether IFNs upregulate ACE2 in putative target cell subsets *in vivo* will be of significant interest to define in future work once current COVID-19-related restrictions on basic scientific inquiry are lifted (Qian et al., 2013).

ACE2 is a central component of the renin-angiotensin system, which has emerged as a key regulator of sterile- or microbially induced lung pathology (Imai et al., 2005). In brief, ACE cleaves angiotensin I to generate angiotensin II (Skeggs et al., 1980). Angiotensin II then acts to drive acute lung injury through various mechanisms, including increased vascular permeability (Imai et al., 2005). Amounts of angiotensin II in humans and mice are elevated during influenza infection, and ACE2 exerts tissue-protective functions by reducing amounts of angiotensin II (Zou et al., 2014). Binding of SARS-CoV-S to mouse ACE2 *in vivo* reduced ACE2 expression leading to acute acid-aspiration-induced lung failure (Kuba et al., 2005). Depending on the questions asked in future work, there are mouse models available on the basis of transgenic expression of human ACE2 (required for overt infectious pathology of SARS-CoV in mice), there are established NHP models available of SARS-CoV infection in *M. fascicularis* and *C. aethiops*, and early reports suggest symptomatic infection in *M. mulatta* and *M. fascicularis* models for SARS-CoV-2 (Bao et al., 2020; McCray et al., 2007; Munster et al., 2020; Rockx et al., 2020; Smits et al., 2011). For example, examining the efficacy of recombinant human ACE2 to act as a decoy receptor or the effect of "ACE inhibitors" in patients with, or at risk for, COVID-19 will require careful experimentation in appropriate models together with well-controlled clinical trials (Hofmann et al., 2004; Monteil et al., 2020; Vaduganathan et al., 2020).

IFN responses that induce ISGs are essential for host antiviral defense in mice, NHPs, and humans (Bailey et al., 2014; Dupuis et al., 2003; Everitt et al., 2012). Canonical ISGs function by directly restricting viruses and reducing burden (Schneider et al., 2014). More recently, disease tolerance to equivalent pathogen burden by factors that increase the ability of the host to tolerate tissue damage has been identified as part of a combined host defense strategy (Iwasaki et al., 2017; Iwasaki and Pillai, 2014; Medzhitov et al., 2012; Schneider and Ayres, 2008). Disease tolerance factors in the lung include IL-22 and amphiregulin (Iwasaki et al., 2017). During acute infection in the respiratory system, ACE2 is critical for early tissue tolerance responses to respiratory infection, including H5N1 influenza (Huang et al., 2014; Zou et al., 2014). However, our discovery that *ACE2* is an ISG in human epithelial cells, along with SARS-CoV-2 utilizing host ACE2

to gain entry to cells, suggests that SARS-CoV and SARS-CoV-2 might exploit the ACE2-mediated tissue-protective response to provide further cellular targets for entry. This potential strategy employed by SARS-CoV-2 could present a unique challenge for the human host and is distinct from HCoV-OC43, which targets the two restriction factors IFITM2 and IFITM3 (Zhao et al., 2014). Our study provides motivation to understand the specific role and balance of type I and type II IFNs, as well as type III IFNs, in tissue protection during, and host restriction of, SARS-CoV-2 infection. Key experiments to understand *ACE2* as an ISG in tissue protection or genuine tolerance will require the appropriate mouse, NHP, or other model in BSL3 or BSL4 facilities to execute SARS-CoV-2 viral infections and measure host tissue health along with viral loads. Further work will also be needed to understand how co-infections, as well as other host factors, might affect both the susceptibility to, and dynamics of, host SARS-CoV-2 infection. Moreover, carefully controlled clinical trials will be essential to determine the overall effects of different IFNs (Prokunina-Olsson et al., 2020).

Altogether, we anticipate that comprehensive characterization of the putative cellular targets of SARS-CoV-2 will be critical to understand basic mechanisms of viral tropism and disease pathophysiology, inform differential susceptibility among vulnerable populations, and potentially suggest unanticipated targets for drug inhibitors of viral infection. The cellular targets we nominate will need to be confirmed by specific reagents for SARS-CoV-2, as done for SARS-CoV (Ding et al., 2004). Furthermore, the transcriptional response to the virus will need to be rigorously characterized in appropriate *in vitro* and *in vivo* model systems (Blanco-Melo et al., 2020). We provide gene lists associated with target cells in specific tissues and diseases to aid the community in understanding this emergent disease. A concurrent HCA Lung Biological Network study assessing *ACE2* and *TMPRSS2* across more tissues also identified enrichment in nasal goblet and ciliated cells (Sungnak et al., 2020). Other studies are considering additional tissues; co-variates such as age, sex, and co-infection state; and represent a large coordinated international effort to the ongoing crisis (Pinto et al., 2020). One study in particular identified upregulation of *ACE2* by respiratory viruses and *TMPRSS2* by IL-13 in a pediatric cohort, suggesting further links to how underlying allergic conditions or co-infections might modulate these two SARS-CoV-2-related host factors (Sajuthi et al., 2020).

During the preparation of this manuscript, several papers have been posted to bioRxiv assessing patterns of *ACE2*$^{+}$ and *TMPRSS2*$^{+}$ cells in barrier tissues (Brann et al., 2020; Lukassen et al., 2020; Qi et al., 2020; Wu et al., 2020; Zhang et al., 2020). At a high level, these studies are largely in agreement with our report. Furthermore, another study appeared on medRxiv profiling bronchoalveolar lavage fluid from 3 severe and 3 mild COVID-19 patients, though they were unable to profile sufficient numbers of epithelial cells (Liao et al., 2020).

Our study highlights the power of scRNA-seq datasets, both existing and novel, to derive hypotheses relevant to human disease that might differ from paradigms established by using cell lines. Further work will be critical to determine how SARS-CoV-2 influences temporal dynamics of host responses at single-cell resolution and which host factors might affect this (Kazer et al., 2020). Given the unappreciated complexities of host-pathogen interactions between humans and SARS-CoV-2, the best measures to combat this pandemic continue to be surveillance and avoidance—especially given that a deep understanding of the full spectrum of resistance and tolerance mechanisms will require the concerted efforts of scientists around the globe (Amanat et al., 2020; Chu et al., 2020; Hadfield et al., 2018). Here, we seek to share our initial findings and data so that other groups might build on this discovery of *ACE2* as an ISG and further consider the careful balance between tissue tolerance and viral infection needed at the human airway epithelium.

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
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human Intestinal Biopsies
  - Human Lungs, Surgical Excess
  - Human Nasal Polyps and Scrapings
  - Human Nasal Washes, Healthy and Influenza Infected
  - Cell Culture of Primary Basal Cells and Cell Lines
  - Non-Human Primates (*M. mulatta*)
  - Non-Human Primates (*M. fascicularis*)
  - Mouse Nasal and Olfactory Epithelium and Tracheal Cells
  - Mouse Lungs, MHV68 Infection
- METHOD DETAILS
  - Methods of Sample Collection and Tissue Preparation for Single-Cell RNA-Seq
  - Methods to Generate Single-Cell and Bulk RNA-seq Libraries
  - Human and Mouse Basal Cell Cytokine Stimulation
  - Western blot for human ACE2
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Non-Human Primate Lung and Ileum
  - Human Lung Tissue
  - Human Ileum
  - Human Adult Nasal Mucosa
  - Granulomatous Tissue from Mycobacterium Tuberculosis Infected NHPs
  - Basal Cell Cytokine Stimulation
  - Interferon Treatment of Mouse Nasal Mucosa
  - Lung from MHV68-Infected WT and IFNγR KO Mice
  - Nasal Washes during Influenza Infection
  - Power Calculations for Detection of Rare Transcripts
  - Statistical Testing

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cell.2020.04.035.

CONSORTIA

The members of HCA Lung Biological Network are Nicholas E. Banovich, Pascal Barbry, Alvis Brazma, Tushar Desai, Thu Elizabeth Duong, Oliver Eickelberg, Christine Falk, Michael Farzan, Ian Glass, Muzlifah Haniffa, Peter Horvath, Deborah Hung, Naftali Kaminski, Mark Krasnow, Jonathan A. Kropski, Malte Kuhnemund, Robert Lafyatis, Haeock Lee, Sylvie Leroy, Sten Linnarson, Joakim Lundeberg, Kerstin B. Meyer, Alexander Misharin, Martijn Nawijn, Marko Z. Nikolic, Jose Ordovas-Montanes, Dana Pe'er, Joseph Powell, Stephen Quake, Jay Rajagopal, Purushothama Rao Tata, Emma L. Rawlins, Aviv Regev, Paul A. Reyfman, Mauricio Rojas, Orit Rosen, Kourosh Saeb-Parsy, Christos Samakovlis, Herbert Schiller, Joachim L. Schultze, Max A. Seibold, Alex K. Shalek, Douglas Shepherd, Jason Spence, Avrum Spira, Xin Sun, Sarah Teichmann, Fabian Theis, Alexander Tsankov, Maarten van den Berge, Michael von Papen, Jeffrey Whitsett, Ramnik Xavier, Yan Xu, Laure-Emmanuelle Zaragosi, and Kun Zhang. Pascal Barbry, Alexander Misharin, Martijn Nawijn, and Jay Rajagopal serve as the coordinators.

ACKNOWLEDGMENTS

We are grateful to the study participants who made this work possible. We would like to thank Bruce Horwitz, Ivan Zanoni, Matt Sampson, Michael Retchin, Peter Winter, Andrew Navia, Jamie Cohen, and Audrey Sporrij for discussions. Mengyang (Vicky) Li Horst, Timothy Tickle, Jonathan Bistline, Jean Chang, Eric Weitz, Eno-Abasi Augustine-Akpan, and Devon Bush for development and support of the Broad Institute Single Cell Portal. This work was supported in part by the Searle Scholars Program, the Beckman Young Investigator Program, the Pew-Stewart Scholars Program for Cancer Research, a Sloan Fellowship in Chemistry, the MIT Stem Cell Initiative through Fondation MIT, the NIH (5U24AI118672 and BAA-NIAID-NIHAI201700104), and the Bill and Melinda Gates Foundation to A.K.S., as well as NIH R56 AI139053 to J.L.F and P.L.L., and the Aeras Foundation to J.L.F. B.B. and S.K.N. are partially supported by NIH 5R01GM081871. We acknowledge support from the Damon Runyon Cancer Research Foundation (DRG-2274-16) and Richard and Susan Smith Family Foundation to J.O.-M; from a National Science Foundation Graduate Research Fellowship (1122374) to S.K.N., S.J.A., and C.N.T.; from a Fannie and John Hertz Foundation Fellowship to C.N.T.; by T32GM007753 from the National Institute of General Medical Sciences to C.G.K.Z. This work was further supported by the UMass Center for Clinical and Translational Science Project Pilot Program; and the Office of the Assistant Secretary of Defense for Health Affairs, through the Peer Reviewed Medical Research Program (W81XWH-15-1-0317) to R.W.F. We also acknowledge support from NIH grants AI078908, HL111113, HL117945, R37AI052353, R01AI136041, R01HL136209, and U19AI095219 to J.A.B.; by grants from the NIH and National Heart, Lung, and Blood Institute (U19 HL129902) to H.P.K and L.S.K; from National Institute of Allergy and Infectious Diseases (UM1 AI126623) to H.P.K.; and to P.B. from the Fondation pour la Recherche Médicale (DEQ20180339158), and the Agence Nationale pour la Recherche (ANR-19-CE14-0027); and by the following grants to L.S.K: NIH/NIAID U19 AI051731, NIH/NHLBI R01 HL095791 NIH/NIAID R33-AI116184, NIH/NIAID U19 AI117945, and DHHS/NIH 1UM1AI126617. B.E.M. was supported by the Massachusetts Institute of Technology - GlaxoSmithKline (MIT-GSK) Gertrude B. Elion Postdoctoral Fellowship; T.M.L. by the NIH/NHLBI 1R01HL128241-01, K.M.B. by NIH/NIAID K23AI139352; and D.L. by NIH R01AI137057, DP2DA042422, and R01AI124378. This publication is part of the Human Cell Atlas (www.humancellatlas.org/publications).

AUTHOR CONTRIBUTIONS

Document S1 details contributions of all authors.

DECLARATION OF INTERESTS

A.R. is an SAB member of ThermoFisher Scientific, Neogene Therapeutics, Asimov, and Syros Pharmaceuticals; a co-founder of and equity holder in Celsius Therapeutics; and an equity holder in Immunitas Therapeutics. A.K.S. reports compensation for consulting and/or SAB membership from Merck, Honeycomb Biotechnologies, Cellarity, Cogen Therapeutics, Orche Bio, and Dahlia Biosciences. L.S.K. is on the SAB for HiFiBio; she reports research funding from Kymab Limited, Bristol Meyers Squibb, Magenta Therapeutics, BlueBird Bio, and Regeneron Pharmaceuticals and consulting fees from Equillium, FortySeven, Inc, Novartis, Inc, EMD Serono, Gilead Sciences, and Takeda Pharmaceuticals. A.S. is an employee of Johnson and Johnson. N.K. is an inventor on a patent using thyroid hormone mimetics in acute lung injury that is now being considered for intervention in COVID-19 patients. J.L. is a scientific consultant for 10X Genomics, Inc. O.R.R, is a co-inventor on patent applications filed by the Broad Institute to inventions relating to single-cell genomics applications, such as in PCT/US2018/060860 and US Provisional Application No. 62/745,259. S.T. in the last three years was a consultant at Genentech, Biogen, and Roche and is a member of the SAB of Foresite Labs. M.H.W. is now an employee of Pfizer. F.J.T. reports receiving consulting fees from Roche Diagnostics GmbH and ownership interest in Cellarity, Inc. P.H. is a co-inventor on a patent using artificial intelligence and high-resolution microscopy for COVID-19 infection testing based on serology.

Received: March 13, 2020
Revised: April 3, 2020
Accepted: April 20, 2020
Published: April 27, 2020

REFERENCES

Adler, H., Messerle, M., Wagner, M., and Koszinowski, U.H. (2000). Cloning and mutagenesis of the murine gammaherpesvirus 68 genome as an infectious bacterial artificial chromosome. J. Virol. *74*, 6964–6974.

Amanat, F., Nguyen, T., Chromikova, V., Strohmeier, S., Stadlbauer, D., Javier, A., Jiang, K., Asthagiri-Arunkumar, G., Polanco, J., Bermudez-Gonzalez, M., et al. (2020). A serological assay to detect SARS-CoV-2 seroconversion in humans. medRxiv. https://doi.org/10.1101/2020.03.17.20037713.

Ardain, A., Domingo-Gonzalez, R., Das, S., Kazer, S.W., Howard, N.C., Singh, A., Ahmed, M., Nhamoyebonde, S., Rangel-Moreno, J., Ogongo, P., et al. (2019). Group 3 innate lymphoid cells mediate early protective immunity against tuberculosis. Nature *570*, 528–532.

Bailey, C.C., Zhong, G., Huang, I.C., and Farzan, M. (2014). IFITM-Family Proteins: The Cell's First Line of Antiviral Defense. Annu. Rev. Virol. *1*, 261–283.

Bao, L., Deng, W., Gao, H., Xiao, C., Liu, J., Xue, J., Lv, Q., Liu, J., Yu, P., Xu, Y., et al. (2020). Reinfection could not occur in SARS-CoV-2 infected rhesus macaques. bioRxiv. https://doi.org/10.1101/2020.03.13.990226.

Blanco-Melo, D., Nilsson-Payant, B.E., Liu, W.-C., Møller, R., Panis, M., Sachs, D., Albrecht, R.A., and tenOever, B.R. (2020). SARS-CoV-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems. bioRxiv. https://doi.org/10.1101/2020.03.24.004655.

Böttcher-Friebertshäuser, E., Klenk, H.D., and Garten, W. (2013). Activation of influenza viruses by proteases from host cells and bacteria in the human airway epithelium. Pathog. Dis. *69*, 87–100.

Brann, D., Tsukahara, T., Weinreb, C., Logan, D.W., and Datta, S.R. (2020). Non-neural expression of SARS-CoV-2 entry genes in the olfactory epithelium suggests mechanisms underlying anosmia in COVID-19 patients. bioRxiv. https://doi.org/10.1101/2020.03.25.009084.

Braun, E., and Sauter, D. (2019). Furin-mediated protein processing in infectious diseases and cancer. Clin. Transl. Immunology *8*, e1073.

Broggi, A., Granucci, F., and Zanoni, I. (2020). Type III interferons: Balancing tissue tolerance and resistance to pathogen invasion. J. Exp. Med. *217*, e20190295.

Bugge, T.H., Antalis, T.M., and Wu, Q. (2009). Type II transmembrane serine proteases. J. Biol. Chem. *284*, 23177–23181.

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

Cao, Y., Guo, Z., Vangala, P., Donnard, E., Liu, P., McDonel, P., Ordovas-Montanes, J., Shalek, A.K., Finberg, R.W., Wang, J.P., et al. (2020). Single-cell

analysis of upper airway cells reveals host-viral dynamics in influenza infected adults. bioRxiv. https://doi.org/10.1101/2020.04.15.042978.

Channappanavar, R., and Perlman, S. (2017). Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. Semin. Immunopathol. *39*, 529–539.

Channappanavar, R., Fehr, A.R., Vijay, R., Mack, M., Zhao, J., Meyerholz, D.K., and Perlman, S. (2016). Dysregulated Type I Interferon and Inflammatory Monocyte-Macrophage Responses Cause Lethal Pneumonia in SARS-CoV-Infected Mice. Cell Host Microbe *19*, 181–193.

Channappanavar, R., Fehr, A.R., Zheng, J., Wohlford-Lenane, C., Abrahante, J.E., Mack, M., Sompallae, R., McCray, P.B., Jr., Meyerholz, D.K., and Perlman, S. (2019). IFN-I response timing relative to virus replication determines MERS coronavirus infection outcomes. J. Clin. Invest. *130*, 3625–3639.

Chu, H.Y., Boeckh, M., Englund, J.A., Famulare, M., Lutz, B.R., Nickerson, D.A., Rieder, M.J., Starita, L.M., Adler, A., Brandstetter, E., et al. (2020). The Seattle Flu Study: a multi-arm community-based prospective study protocol for assessing influenza prevalence, transmission, and genomic epidemiology. medRxiv. https://doi.org/10.1101/2020.03.02.20029595.

Colonna, L., Peterson, C.W., Schell, J.B., Carlson, J.M., Tkachev, V., Brown, M., Yu, A., Reddy, S., Obenza, W.M., Nelson, V., et al. (2018). Evidence for persistence of the SHIV reservoir early after MHC haploidentical hematopoietic stem cell transplantation. Nat. Commun. *9*, 4438.

Coronaviridae Study Group of the International Committee on Taxonomy of V. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat. Microbiol. https://doi.org/10.1038/s41564-020-0695-z.

Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N.G., and Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res. *176*, 104742.

Davidson, D.J., Gray, M.A., Kilanowski, F.M., Tarran, R., Randell, S.H., Sheppard, D.N., Argent, B.E., and Dorin, J.R. (2004). Murine epithelial cells: isolation and culture. J. Cyst. Fibros. *3 (Suppl 2)*, 59–62.

Davidson, S., Maini, M.K., and Wack, A. (2015). Disease-promoting effects of type I interferons in viral, bacterial, and coinfections. J. Interferon Cytokine Res. *35*, 252–264.

de Lang, A., Osterhaus, A.D., and Haagmans, B.L. (2006). Interferon-gamma and interleukin-4 downregulate expression of the SARS coronavirus receptor ACE2 in Vero E6 cells. Virology *353*, 474–481.

Dear, T.N., Boehm, T., Keverne, E.B., and Rabbitts, T.H. (1991). Novel genes for potential ligand-binding proteins in subregions of the olfactory mucosa. EMBO J. *10*, 2813–2819.

Deeks, S.G., Odorizzi, P.M., and Sekaly, R.P. (2017). The interferon paradox: can inhibiting an antiviral mechanism advance an HIV cure? J. Clin. Invest. *127*, 103–105.

Derr, A., Yang, C., Zilionis, R., Sergushichev, A., Blodgett, D.M., Redick, S., Bortell, R., Luban, J., Harlan, D.M., Kadener, S., et al. (2016). End Sequence Analysis Toolkit (ESAT) expands the extractable information from single-cell RNA-seq data. Genome Res. *26*, 1397–1410.

Ding, Y., He, L., Zhang, Q., Huang, Z., Che, X., Hou, J., Wang, H., Shen, H., Qiu, L., Li, Z., et al. (2004). Organ distribution of severe acute respiratory syndrome (SARS) associated coronavirus (SARS-CoV) in SARS patients: implications for pathogenesis and virus transmission pathways. J. Pathol. *203*, 622–630.

Dong, E., Du, H., and Gardner, L. (2020a). An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis., S1473-3099(20)30120-1.

Dong, L., Hu, S., and Gao, J. (2020b). Discovering drugs to treat coronavirus disease 2019 (COVID-19). Drug Discov. Ther. *14*, 58–60.

Dunston, D., Ashby, S., Krosnowski, K., Ogura, T., and Lin, W. (2013). An effective manual deboning method to prepare intact mouse nasal tissue with preserved anatomical organization. J. Vis. Exp. (78) https://doi.org/10.3791/50538.

Dupuis, S., Jouanguy, E., Al-Hajjar, S., Fieschi, C., Al-Mohsen, I.Z., Al-Jumaah, S., Yang, K., Chapgier, A., Eidenschenk, C., Eid, P., et al. (2003). Impaired response to interferon-alpha/beta and lethal viral disease in human STAT1 deficiency. Nat. Genet. *33*, 388–391.

Everitt, A.R., Clare, S., Pertel, T., John, S.P., Wash, R.S., Smith, S.E., Chin, C.R., Feeley, E.M., Sims, J.S., Adams, D.J., et al.; GenISIS Investigators; MOSAIC Investigators (2012). IFITM3 restricts the morbidity and mortality associated with influenza. Nature *484*, 519–523.

Fleming, F.E., Böhm, R., Dang, V.T., Holloway, G., Haselhorst, T., Madge, P.D., Deveryshetty, J., Yu, X., Blanchard, H., von Itzstein, M., and Coulson, B.S. (2014). Relative roles of GM1 ganglioside, N-acylneuraminic acids, and α2β1 integrin in mediating rotavirus infection. J. Virol. *88*, 4558–4571.

Fung, T.S., and Liu, D.X. (2019). Human Coronavirus: Host-Pathogen Interaction. Annu. Rev. Microbiol. *73*, 529–557.

Genshaft, A.S., Li, S., Gallant, C.J., Darmanis, S., Prakadan, S.M., Ziegler, C.G., Lundberg, M., Fredriksson, S., Hong, J., Regev, A., et al. (2016). Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. Genome Biol. *17*, 188.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. Nature *489*, 91–100.

Gierahn, T.M., Wadsworth, M.H., 2nd, Hughes, T.K., Bryson, B.D., Butler, A., Satija, R., Fortune, S., Love, J.C., and Shalek, A.K. (2017). Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat. Methods *14*, 395–398.

Giovannini-Chami, L., Marcet, B., Moreilhon, C., Chevalier, B., Illie, M.I., Lebrigand, K., Robbe-Sermesant, K., Bourrier, T., Michiels, J.F., Mari, B., et al. (2012). Distinct epithelial gene expression phenotypes in childhood respiratory allergy. Eur. Respir. J. *39*, 1197–1205.

Glowacka, I., Bertram, S., Müller, M.A., Allen, P., Soilleux, E., Pfefferle, S., Steffen, I., Tsegaye, T.S., He, Y., Gnirss, K., et al. (2011). Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. J. Virol. *85*, 4122–4134.

Good, C., Wells, A.I., and Coyne, C.B. (2019). Type III interferon signaling restricts enterovirus 71 infection of goblet cells. Sci Adv. https://doi.org/10.1126/sciadv.aau4255.

Griggs, T.F., Bochkov, Y.A., Basnet, S., Pasic, T.R., Brockman-Schneider, R.A., Palmenberg, A.C., and Gern, J.E. (2017). Rhinovirus C targets ciliated airway epithelial cells. Respir. Res. *18*, 84.

Grove, J., and Marsh, M. (2011). The cell biology of receptor-mediated virus entry. J. Cell Biol. *195*, 1071–1082.

Guan, W.J., Ni, Z.Y., Hu, Y., Liang, W.H., Ou, C.Q., He, J.X., Liu, L., Shan, H., Lei, C.L., Hui, D.S.C., et al.; China Medical Treatment Expert Group for Covid-19 (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. N. Engl. J. Med. https://doi.org/10.1056/NEJMoa2002032.

Guerrero-Plata, A., Baron, S., Poast, J.S., Adegboyega, P.A., Casola, A., and Garofalo, R.P. (2005). Activity and regulation of alpha interferon in respiratory syncytial virus and human metapneumovirus experimental infections. J. Virol. *79*, 10190–10199.

Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., and Neher, R.A. (2018). Nextstrain: real-time tracking of pathogen evolution. Bioinformatics *34*, 4121–4123.

Hamming, I., Timens, W., Bulthuis, M.L., Lely, A.T., Navis, G., and van Goor, H. (2004). Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. J. Pathol. *203*, 631–637.

Harmer, D., Gilbert, M., Borman, R., and Clark, K.L. (2002). Quantitative mRNA expression profiling of ACE 2, a novel homologue of angiotensin converting enzyme. FEBS Lett. *532*, 107–110.

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.

Hofmann, H., Geier, M., Marzi, A., Krumbiegel, M., Peipp, M., Fey, G.H., Gramberg, T., and Pöhlmann, S. (2004). Susceptibility to SARS coronavirus S protein-driven infection correlates with expression of angiotensin converting enzyme 2 and infection can be blocked by soluble receptor. Biochem. Biophys. Res. Commun. *319*, 1216–1221.

Holly, M.K., and Smith, J.G. (2018). Adenovirus Infection of Human Enteroids Reveals Interferon Sensitivity and Preferential Infection of Goblet Cells. J. Virol. *92*, e00250-18.

Holshue, M.L., DeBolt, C., Lindquist, S., Lofy, K.H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A., et al.; Washington State 2019-nCoV Case Investigation Team (2020). First Case of 2019 Novel Coronavirus in the United States. N. Engl. J. Med. *382*, 929–936.

Huang, F., Guo, J., Zou, Z., Liu, J., Cao, B., Zhang, S., Li, H., Wang, W., Sheng, M., Liu, S., et al. (2014). Angiotensin II plasma levels are linked to disease severity and predict fatal outcomes in H7N9-infected patients. Nat. Commun. *5*, 3595.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet *395*, 497–506.

Hughes, T.K., Wadsworth, M.H., Gierahn, T.M., Do, T., Weiss, D., Andrade, P.R., Ma, F., de Andrade Silva, B.J., Shao, S., Tsoi, L.C., et al. (2019). Highly Efficient, Massively-Parallel Single-Cell RNA-Seq Reveals Cellular States and Molecular Features of Human Skin Pathology. bioRxiv. https://doi.org/10.1101/689273.

Imai, Y., Kuba, K., Rao, S., Huan, Y., Guo, F., Guan, B., Yang, P., Sarao, R., Wada, T., Leong-Poi, H., et al. (2005). Angiotensin-converting enzyme 2 protects from severe acute lung failure. Nature *436*, 112–116.

Iwasaki, A., and Pillai, P.S. (2014). Innate immunity to influenza virus infection. Nat. Rev. Immunol. *14*, 315–328.

Iwasaki, A., Foxman, E.F., and Molony, R.D. (2017). Early local immune defences in the respiratory tract. Nat. Rev. Immunol. *17*, 7–20.

Iwata-Yoshikawa, N., Okamura, T., Shimizu, Y., Hasegawa, H., Takeda, M., and Nagata, N. (2019). TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. J. Virol. *93*, e01815-18.

Jewell, N.A., Cline, T., Mertz, S.E., Smirnov, S.V., Flaño, E., Schindler, C., Grieves, J.L., Durbin, R.K., Kotenko, S.V., and Durbin, J.E. (2010). Lambda interferon is the predominant interferon induced by influenza A virus infection in vivo. J. Virol. *84*, 11515–11522.

Jonsdottir, H.R., and Dijkman, R. (2016). Coronaviruses and the human airway: a universal system for virus-host interaction studies. Virol. J. *13*, 24.

Jovanovic, M., Rooney, M.S., Mertins, P., Przybylski, D., Chevrier, N., Satija, R., Rodriguez, E.H., Fields, A.P., Schwartz, S., Raychowdhury, R., et al. (2015). Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens. Science *347*, 1259038.

Kazer, S.W., Aicher, T.P., Muema, D.M., Carroll, S.L., Ordovas-Montanes, J., Miao, V.N., Tu, A.A., Ziegler, C.G.K., Nyquist, S.K., Wong, E.B., et al. (2020). Integrated single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. Nat. Med. *26*, 511–518.

Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. Nat. Methods *11*, 740–742.

Krischuns, T., Günl, F., Henschel, L., Binder, M., Willemsen, J., Schloer, S., Rescher, U., Gerlt, V., Zimmer, G., Nordhoff, C., et al. (2018). Phosphorylation of TRIM28 Enhances the Expression of IFN-β and Proinflammatory Cytokines During HPAIV Infection of Human Lung Epithelial Cells. Front. Immunol. *9*, 2229. https://doi.org/10.3389/fimmu.2018.02229.

Kuba, K., Imai, Y., Rao, S., Gao, H., Guo, F., Guan, B., Huan, Y., Yang, P., Zhang, Y., Deng, W., et al. (2005). A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. Nat. Med. *11*, 875–879.

Kucharski, A.J., Russell, T.W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., and Eggo, R.M.; Centre for Mathematical Modelling of Infectious Diseases COVID-19 working group (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect. Dis. https://doi.org/10.1016/S1473-3099(20)30144-4.

Lechien, J.R., Chiesa-Estomba, C.M., De Siati, D.R., Horoi, M., Le Bon, S.D., Rodriguez, A., Dequanter, D., Blecic, S., El Afia, F., Distinguin, L., et al. (2020). Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. Eur. Arch. Otorhinolaryngol. https://doi.org/10.1007/s00405-020-05965-1.

Lei, J., Li, J., Li, X., and Qi, X. (2020). CT Imaging of the 2019 Novel Coronavirus (2019-nCoV) Pneumonia. Radiology *295*, 18.

Letko, M., Marzi, A., and Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat. Microbiol. *5*, 562–569.

Li, G., and De Clercq, E. (2020). Therapeutic options for the 2019 novel coronavirus (2019-nCoV). Nat. Rev. Drug Discov. *19*, 149–150.

Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C., et al. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature *426*, 450–454.

Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., Chen, L., Li, J., Wang, X., Wang, F., et al. (2020). The landscape of lung bronchoalveolar immune cells in COVID-19 revealed by single-cell RNA sequencing. medRxiv. https://doi.org/10.1101/2020.02.23.20026690.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet *395*, 565–574.

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 predominantly expressed in a transient secretory cell type in subsegmental bronchial branches. bioRxiv. https://doi.org/10.1101/2020.03.13.991455.

Lun, A.T., 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. https://doi.org/10.12688/f1000research.9501.2.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Mar, K.B., Rinkenberger, N.R., Boys, I.N., Eitson, J.L., McDougal, M.B., Richardson, R.B., and Schoggins, J.W. (2018). LY6E mediates an evolutionarily conserved enhancement of virus infection by targeting a late entry step. Nat. Commun. *9*, 3603. https://doi.org/10.1038/s41467-018-06000-y.

Martin, C.J., Cadena, A.M., Leung, V.W., Lin, P.L., Maiello, P., Hicks, N., Chase, M.R., Flynn, J.L., and Fortune, S.M. (2017). Digitally Barcoding *Mycobacterium tuberculosis* Reveals *In Vivo* Infection Dynamics in the Macaque Model of Tuberculosis. MBio *8*, e00312-17.

Matos, A.D.R., Wunderlich, K., Schloer, S., Schughart, K., Geffers, R., Seders, M., Witt, M., Christersson, A., Wiewrodt, R., Wiebe, K., et al. (2019). Antiviral potential of human IFN-α subtypes against influenza A H3N2 infection in human lung explants reveals subtype-specific activities. Emerg. Microbes Infect. *8*, 1763–1776.

Matsuyama, S., Nagata, N., Shirato, K., Kawase, M., Takeda, M., and Taguchi, F. (2010). Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. J. Virol. *84*, 12658–12664.

Matys, V., Fricke, E., Geffers, R., Gössling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V., et al. (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res. *31*, 374–378.

McCray, P.B., Jr., Pewe, L., Wohlford-Lenane, C., Hickey, M., Manzel, L., Shi, L., Netland, J., Jia, H.P., Halabi, C., Sigmund, C.D., et al. (2007). Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus. J. Virol. *81*, 813–821.

Mead, B.E., and Karp, J.M. (2019). All models are wrong, but some organoids may be useful. Genome Biol. *20*, 66. https://doi.org/10.1186/s13059-019-1677-4.

Mead, B.E., Ordovas-Montanes, J., Braun, A.P., Levy, L.E., Bhargava, P., Szucs, M.J., Ammendolia, D.A., MacMullan, M.A., Yin, X., Hughes, T.K., et al. (2018). Harnessing single-cell genomics to improve the physiological fidelity of organoid-derived cell types. BMC Biol. *16*, 62. https://doi.org/10.1186/s12915-018-0527-2.

Medzhitov, R., Schneider, D.S., and Soares, M.P. (2012). Disease tolerance as a defense strategy. Science *335*, 936–941.

Monteil, V., Kwon, H., Prado, P., Hagelkryus, A., Wimmer, R.A., and al, e. (2020). Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. Cell. https://doi.org/10.1016/j.cell.2020.04.004.

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

Müller, U., Steinhoff, U., Reis, L.F., Hemmi, S., Pavlovic, J., Zinkernagel, R.M., and Aguet, M. (1994). Functional role of type I and type II interferons in antiviral defense. Science *264*, 1918–1921.

Munster, V.J., Feldmann, F., Williamson, B.N., van Doremalen, N., Pérez-Pérez, L., Schulz, J., Meade-White, K., Okumura, A., Callison, J., Brumbaugh, B., et al. (2020). Respiratory disease and virus shedding in rhesus macaques inoculated with SARS-CoV-2. bioRxiv. https://doi.org/10.1101/2020.03.21.001628.

Murphy, T.L., Tussiwand, R., and Murphy, K.M. (2013). Specificity through cooperation: BATF-IRF interactions control immune-regulatory networks. Nat. Rev. Immunol. *13*, 499–509.

Ordovas-Montanes, J., Dwyer, D.F., Nyquist, S.K., Buchheit, K.M., Vukovic, M., Deb, C., Wadsworth, M.H., 2nd, Hughes, T.K., Kazer, S.W., Yoshimoto, E., et al. (2018). Allergic inflammatory memory in human respiratory epithelial progenitor cells. Nature *560*, 649–654.

Ordovas-Montanes, J., Beyaz, S., Rakoff-Nahoum, S., and Shalek, A.K. (2020). Distribution and storage of inflammatory memory in barrier tissues. Nat. Rev. Immunol. https://doi.org/10.1038/s41577-019-0263-z.

Paules, C.I., Marston, H.D., and Fauci, A.S. (2020). Coronavirus Infections-More Than Just the Common Cold. JAMA. https://doi.org/10.1001/jama.2020.0757.

Pinto, B.G., Oliveira, A.E., Singh, Y., Jimenez, L., Goncalves, A.N., Ogava, R.L., Creighton, R., Peron, J.P., and Nakaya, H.I. (2020). ACE2 Expression is Increased in the Lungs of Patients with Comorbidities Associated with Severe COVID-19. medRxiv. https://doi.org/10.1101/2020.03.21.20040261.

Prokunina-Olsson, L., Alphonse, N., Dickenson, R.E., Durbin, J.E., Glenn, J.S., Hartmann, R., Kotenko, S.V., Lazear, H.M., O'Brien, T.R., Odendall, C., et al. (2020). COVID-19 and emerging viral infections: The case for interferon lambda. J. Exp. Med. *217*, e20200653. https://doi.org/10.1084/jem.20200653.

Qi, F., Qian, S., Zhang, S., and Zhang, Z. (2020). Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. bioRxiv. https://doi.org/10.1101/2020.02.16.951913.

Qian, Z., Travanty, E.A., Oko, L., Edeen, K., Berglund, A., Wang, J., Ito, Y., Holmes, K.V., and Mason, R.J. (2013). Innate immune response of human alveolar type II cells infected with severe acute respiratory syndrome-coronavirus. Am. J. Respir. Cell Mol. Biol. *48*, 742–748.

Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., et al. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat. Biotechnol. *29*, 436–442.

Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al.; Human Cell Atlas Meeting Participants (2017). The Human Cell Atlas. eLife *6*, e27041.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rockx, B., Kuiken, T., Herfst, S., Bestebroer, T., Lamers, M.M., Oude Munnink, B.B., de Meulder, D., van Amerongen, G., van den Brand, J., Okba, N.M.A., et al. (2020). Comparative pathogenesis of COVID-19, MERS, and SARS in a nonhuman primate model. Science. https://doi.org/10.1126/science.abb7314.

Rodriguez, A., and Laio, A. (2014). Machine learning. Clustering by fast search and find of density peaks. Science *344*, 1492–1496.

Ruiz García, S., Deprez, M., Lebrigand, K., Cavard, A., Paquet, A., Arguel, M.J., Magnone, V., Truchi, M., Caballero, I., Leroy, S., et al. (2019). Novel dynamics of human mucociliary differentiation revealed by single-cell RNA sequencing of nasal epithelial cultures. Development *146*, dev177428. https://doi.org/10.1242/dev.177428.

Rusinova, I., Forster, S., Yu, S., Kannan, A., Masse, M., Cumming, H., Chapman, R., and Hertzog, P.J. (2013). Interferome v2.0: an updated database of annotated interferon-regulated genes. Nucleic Acids Res. *41*, D1040–D1046.

Russell, A.B., Trapnell, C., and Bloom, J.D. (2018). Extreme heterogeneity of influenza virus infection in single cells. eLife *7*, e32303. https://doi.org/10.7554/eLife.32303.

Sainz, B., Jr., Mossel, E.C., Peters, C.J., and Garry, R.F. (2004). Interferon-beta and interferon-gamma synergistically inhibit the replication of severe acute respiratory syndrome-associated coronavirus (SARS-CoV). Virology *329*, 11–17.

Sajuthi, S.P., DeFord, P., Jackson, N.D., Montgomery, M.T., Everman, J.L., Rios, C.L., Pruesse, E., Nolin, J.D., Plender, E.G., Wechsler, M.E., et al. (2020). Type 2 and interferon inflammation strongly regulate SARS-CoV-2 related gene expression in the airway epithelium. bioRxiv. https://doi.org/10.1101/2020.04.09.034454.

Sanda, C., Weitzel, P., Tsukahara, T., Schaley, J., Edenberg, H.J., Stephens, M.A., McClintick, J.N., Blatt, L.M., Li, L., Brodsky, L., and Taylor, M.W. (2006). Differential gene induction by type I and type II interferons and their combination. J. Interferon Cytokine Res. *26*, 462–472.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Schiller, H.B., Montoro, D.T., Simon, L.M., Rawlins, E.L., Meyer, K.B., Strunz, M., Vieira Braga, F.A., Timens, W., Koppelman, G.H., Budinger, G.R.S., et al. (2019). The Human Lung Cell Atlas: A High-Resolution Reference Map of the Human Lung in Health and Disease. Am. J. Respir. Cell Mol. Biol. *61*, 31–41.

Schneider, D.S., and Ayres, J.S. (2008). Two ways to survive infection: what resistance and tolerance can teach us about treating infectious diseases. Nat. Rev. Immunol. *8*, 889–895.

Schneider, W.M., Chevillotte, M.D., and Rice, C.M. (2014). Interferon-stimulated genes: a complex web of host defenses. Annu. Rev. Immunol. *32*, 513–545.

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236–240.

Shulla, A., Heald-Sargent, T., Subramanya, G., Zhao, J., Perlman, S., and Gallagher, T. (2011). A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. J. Virol. *85*, 873–882.

Skeggs, L.T., Dorer, F.E., Levine, M., Lentz, K.E., and Kahn, J.R. (1980). The biochemistry of the renin-angiotensin system. Adv. Exp. Med. Biol. *130*, 1–27.

Smillie, C.S., Biton, M., Ordovas-Montanes, J., Sullivan, K.M., Burgin, G., Graham, D.B., Herbst, R.H., Rogel, N., Slyper, M., Waldman, J., et al. (2019). Intra-

and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell. https://doi.org/10.1016/j.cell.2019.06.029.

Smits, S.L., van den Brand, J.M., de Lang, A., Leijten, L.M., van Ijcken, W.F., van Amerongen, G., Osterhaus, A.D., Andeweg, A.C., and Haagmans, B.L. (2011). Distinct severe acute respiratory syndrome coronavirus-induced acute lung injury pathways in two different nonhuman primate species. J. Virol. *85*, 4234–4245.

Stetson, D.B., and Medzhitov, R. (2006). Type I interferons in host defense. Immunity *25*, 373–381.

Steuerman, Y., Cohen, M., Peshes-Yaloz, N., Valadarsky, L., Cohn, O., David, E., Frishberg, A., Mayo, L., Bacharach, E., Amit, I., et al. (2018). Dissection of Influenza Infection In Vivo by Single-Cell RNA Sequencing. Cell Syst. https://doi.org/10.1016/j.cels.2018.05.008.

Sungnak, W., Huang, N., Becavin, C., Berg, M., and Network, H.L.B. (2020). SARS-CoV-2 Entry Genes Are Most Highly Expressed in Nasal Goblet and Ciliated Cells within Human Airways. Nat. Med. https://doi.org/10.1038/s41591-020-0868-6.

Svensson, V. (2020). Droplet scRNA-seq is not zero-inflated. Nat. Biotechnol. *38*, 147–150.

Tepe, B., Hill, M.C., Pekarek, B.T., Hunt, P.J., Martin, T.J., Martin, J.F., and Arenkiel, B.R. (2018). Single-Cell RNA-Seq of Mouse Olfactory Bulb Reveals Cellular Heterogeneity and Activity-Dependent Molecular Census of Adult-Born Neurons. Cell Rep. https://doi.org/10.1016/j.celrep.2018.11.034.

Trombetta, J.J., Gennert, D., Lu, D., Satija, R., Shalek, A.K., and Regev, A. (2014). Preparation of Single-Cell RNA-Seq Libraries for Next Generation Sequencing. Curr Protoc Mol Biol. https://doi.org/10.1002/0471142727.mb0422s107.

Utay, N.S., and Douek, D.C. (2016). Interferons and HIV Infection: The Good, the Bad, and the Ugly. Pathog. Immun. *1*, 107–116.

Vaduganathan, M., Vardeny, O., Michel, T., McMurray, J.J.V., Pfeffer, M.A., and Solomon, S.D. (2020). Renin-Angiotensin-Aldosterone System Inhibitors in Patients with Covid-19. N. Engl. J. Med. https://doi.org/10.1056/NEJMsr2005760.

Vieira Braga, F.A., Kar, G., Berg, M., Carpaij, O.A., Polanski, K., Simon, L.M., Brouwer, S., Gomes, T., Hesse, L., Jiang, J., et al. (2019). A cellular census of human lungs identifies novel cell states in health and in asthma. Nat. Med. *25*, 1153–1163.

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.

Wang, P.-H., and Cheng, Y. (2020). Increasing Host Cellular Receptor—Angiotensin-Converting Enzyme 2 (ACE2) Expression by Coronavirus may Facilitate 2019-nCoV Infection. bioRxiv. https://doi.org/10.1101/2020.02.24.963348.

Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y., et al. (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res. *22*, 1798–1812.

Wang, J., Zhuang, J., Iyer, S., Lin, X.Y., Greven, M.C., Kim, B.H., Moore, J., Pierce, B.G., Dong, X., Virgil, D., et al. (2013). Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. Nucleic Acids Res. *41*, D171–D176.

Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., and Tan, W. (2020). Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA. https://doi.org/10.1001/jama.2020.3786.

Weinheimer, V.K., Becher, A., Tönnies, M., Holland, G., Knepper, J., Bauer, T.T., Schneider, P., Neudecker, J., Rückert, J.C., Szymanski, K., et al. (2012). Influenza A viruses target type II pneumocytes in the human lung. J. Infect. Dis. *206*, 1685–1694.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

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. https://doi.org/10.1038/s41586-020-2196-x.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science *367*, 1260–1263.

Wu, C., Zheng, S., Chen, Y., and Zheng, M. (2020). Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCoV, in the nasal tissue. medRxiv. https://doi.org/10.1101/2020.02.11.20022228.

Xiao, F., Tang, M., Zheng, X., Liu, Y., Li, X., and Shan, H. (2020). Evidence for gastrointestinal infection of SARS-CoV-2. Gastroenterology. https://doi.org/10.1053/j.gastro.2020.02.055.

Xu, Y., Li, X., Zhu, B., Liang, H., Fang, C., Gong, Y., Guo, Q., Sun, X., Zhao, D., Shen, J., et al. (2020). Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. Nat. Med. *26*, 502–505.

Young, M.D., and Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet based single-cell RNA sequencing data. bioRxiv. https://doi.org/10.1101/303727.

Zhang, H., Kang, Z., Gong, H., Xu, D., Wang, J., Li, Z., Cui, X., Xiao, J., Meng, T., Zhou, W., et al. (2020). The digestive system is a potential route of 2019-nCov infection: a bioinformatics analysis based on single-cell transcriptomes. bioRxiv. https://doi.org/10.1101/2020.01.30.927806.

Zhao, X., Guo, F., Liu, F., Cuconati, A., Chang, J., Block, T.M., and Guo, J.T. (2014). Interferon induction of IFITM proteins promotes infection by human coronavirus OC43. Proc. Natl. Acad. Sci. USA *111*, 6756–6761.

Zheng, B., He, M.L., Wong, K.L., Lum, C.T., Poon, L.L., Peng, Y., Guan, Y., Lin, M.C., and Kung, H.F. (2004). Potent inhibition of SARS-associated coronavirus (SCOV) infection and replication by type I interferons (IFN-alpha/beta) but not by type II interferon (IFN-gamma). J. Interferon Cytokine Res. *24*, 388–390.

Zou, Z., Yan, Y., Shu, Y., Gao, R., Sun, Y., Li, X., Ju, X., Liang, Z., Liu, Q., Zhao, Y., et al. (2014). Angiotensin-converting enzyme 2 protects from lethal avian influenza A H5N1 infections. Nat. Commun. *5*, 3594.

Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., Yu, J., Kang, M., Song, Y., Xia, J., et al. (2020). SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. N. Engl. J. Med. *382*, 1177–1179.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
M. mulatta lung, bone marrow, brain, colon, ileum, jejunum, liver, lung, peripheral blood, spleen, thymus, tonsil, and lymph nodes from various sites	Washington National Primate Research Center	N/A
Human lung tissue from surgical excess	University of KwaZulu-Natal	IRB Code: BE024/09
Human non-inflamed ileal pinch biopsies	Multi-center clinical study, approved by the Institutional Review Board at Boston Children's Hospital	IRB Code: IRB-P00030890
Human nasal lavage	University of Massachusetts Medical School	N/A
Human nasal scraping, polyp scrapings, ethmoid sinus surgical tissue samples	Partners HealthCare Institute	N/A
M. fascicularis lung and granulomatous tissue	University of Pittsburgh School of Medicine	N/A
Antibodies		
anti-ACE2 human antibody, goat polyclonal	R&D	Cat#AF933
Bacterial and Virus Strains		
MHV-68	Adler et al., 2000	N/A
Mycobacterium Tuberculosis, Modified Erdman Strain	Martin et al., 2017	N/A
Chemicals, Peptides, and Recombinant Proteins		
2-Mercaptoethanol	Sigma	Cat#M3148-25ML
RLT Buffer	QIAGEN	Cat#79216
dNTP	New England BioLabs	Cat#N0447L
RNase Inhibitor	Fisher Scientific	Cat#AM2696
Maxima RNaseH-minus RT Enzyme	Fisher Scientific	Cat#EP0753
$MgCl_2$	Sigma	Cat#63069-100ML
Betaine	Sigma	Cat#B0300-5VL
AMPure RNAClean XP RNA-SPRI beads	Beckman Coulter	Cat#A63987
AMPure XP SPRI beads	Beckman Coulter	Cat#A63881
Guanidinium thiocyanate	Sigma	Cat#AM9422
Sarkosyl	Sigma	Cat#L7414
Exonuclease I	New England BioLabs	Cat#M0293S
Klenow Fragment	New England BioLabs	Cat#M0212L
DNase I	Roche	Cat#10104159001
Collagenase IV	Life Technologies	Cat#17104019
Collagenase D	Roche	Cat#11088858001
Liberase TM	Roche	Cat#5401119001
TrypLE	Thermo Fisher	Cat#12604013
ACK Buffer	Thermo Fisher	Cat#A1049201
IFN-α	Biolegend	Cat#752802
Dispase II	Thermo Fisher	Cat#17105041
Elastase	Worthington Biochem	Cat#LS002292
Pneumacult-Ex serum-free media	StemCell Technologies, Inc.	Cat#05040

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
IL-4, human	Biolegend	Cat#574002
IL17A, human	Biolegend	Cat#570502
IFNγ, human	Biolegend	Cat#570202
IFNγ, mouse	Peprotech	Cat#315-05
IFNα, human	Biolegend	Cat#592702
IFNα, mouse	Biolegend	Cat#752802
IFNβ, mouse	R&D Systems	Cat#8234-MB-010
Critical Commercial Assays		
Nextera XT DNA Library Preparation Kit	Illumina	Cat#FC-131-1096
High Sensitivity D5000 ScreenTape	Agilent	Cat#5067-5592
Qubit dsDNA High-Sensitivity kit	ThermoFisher	Cat#Q32854
NextSeq 500/550 High Output v2 (75 cycles)	Illumina	Cat#FC-404-2005
NovaSeq 6000 S2 (100 cycles)	Illumina	Cat#20012862
Kapa HiFi HotStart ReadyMix	Kapa Biosystems	Cat#KK2602
MACOSKO-2011-10 mRNA Capture Beads	ChemGenes	Cat#NC0927472
Tumor Dissociation Kit, Human	Miltenyi Biotec	Cat#130-095-929
Chromium Single Cell 3′ v2	10X Genomics	Cat#120237
Deposited Data		
scRNA-seq Processed Data	This paper	https://singlecell.broadinstitute.org/single_cell?scpbr=the-alexandria-project
scRNA-seq Processed Data	This paper	https://drive.google.com/drive/folders/1bxClqNeZ7wLuVOT16gphwj98_cc9KhrV?usp=sharing
scRNA-seq Processed Data	This paper	https://chanzuckerberg.github.io/cellxgene/posts/cellxgene_cziscience_com
scRNA-seq Processed Data	This paper	https://singlecell.broadinstitute.org/single_cell/covid19
scRNA-seq Processed Data (all species) and FASTQ files (for NHP and murine datasets)	This paper	GEO: GSE148829
scRNA-seq data from human nasal mucosa	Ordovas-Montanes et al., 2018	https://singlecell.broadinstitute.org/single_cell/study/SCP253/allergic-inflammatory-memory-in-human-respiratory-epithelial-progenitor-cells
Human reference genome NCBI build 38 (GRCh38)	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
Human reference genome NCBI build 19	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
Mouse reference genome NCBI build 10	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/mouse/
Macaca mulatta reference genome assembly 8.0.1, annotation 102	NCBI Eukaryotic Genome Annotation Pipeline	https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Macaca_mulatta/102/
Macaca fascicularis reference genome assembly 5, annotation 101	NCBI Eukaryotic Genome Annotation Pipeline	https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Macaca_fascicularis/101/
Interferome Database	Rusinova et al., 2013	http://www.interferome.org/
RNA-seq from human lung explants ± *ex vivo* IAV infection	Matos et al., 2019	GEO: GSE135069
RNA-seq from human nasal epithelial cells	Giovannini-Chami et al., 2012	GEO: GSE19190, GSE22147

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Cell Lines		
Human: Passage 4 BEAS-2B	ATCC	CRL-9609
Experimental Models: Organisms/Strains		
Mouse: C57BL/6J	The Jackson Laboratory	Cat#000664
Mouse: C57BL/6, IFNγR−/− B6.129S7-*Ifngr1*^{tm1Agt}/J	The Jackson Laboratory	Cat#003288
Oligonucleotides		
SMART-seq2 2 3′ Oligo-dT Primer: /5Biosg/ AAG CAG TGG TAT CAA CGC AGA GTA CTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TVN	Integrated DNA Technologies	N/A
SMART-seq2 5′ TSO: AAG CAG TGG TAT CAA CGC AGA GTA CAT rGrGrG	Integrated DNA Technologies	N/A
SMART-seq2 and Seq-Well ISPCR: AAG CAG TGG TAT CAA CGC AGA GT	Integrated DNA Technologies	N/A
Custom Read 1 Primer: GCC TGT CCG CGG AAG CAG TGG TAT CAA CGC AGA GTA C	Integrated DNA Technologies	N/A
Seq-Well 5′ TSO: AAG CAG TGG TAT CAA CGC AGA GTG AAT rGrGrG	Integrated DNA Technologies	N/A
Seq-Well Custom P5-SMART PCR hybrid oligo: AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG CCT GTC CGC GGA AGC AGT GGT ATC AAC GCA GAG TAC	Integrated DNA Technologies	N/A
Seq-Well dN-SMRT oligo: AAG CAG TGG TAT CAA CGC AGA GTG ANN NGG NNN B	Integrated DNA Technologies	N/A
Software and Algorithms		
R	R Core Team	https://www.r-project.org
R package – Seurat v2.3.4 and v3.1.0	Github	https://github.com/satijalab/seurat
Scanpy	Wolf et al., 2018	https://github.com/theislab/scanpy
R package – SCDE	Bioconductor	http://bioconductor.org/packages/scde/
Prism 6	GraphPad Software	https://www.graphpad.com/scientific-software/prism/
STAR	Github	https://github.com/alexdobin/STAR
Uniform Manifold Approximation and Projection	Github	https://github.com/lmcinnes/umap
Rtsne	CRAN	https://cran.r-project.org/web/packages/Rtsne/

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by Dr. Jose Ordovas-Montanes (jose.ordovas-montanes@childrens.harvard.edu).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

In Table S9, we provide a guide to all datasets analyzed in this paper as well as links to each individual dataset for download with the main landing page here: https://singlecell.broadinstitute.org/single_cell?scpbr=the-alexandria-project. To download the data from the portal, follow the link to the visualization page, sign in a free account in the portal using a Google apps enabled email address,

and select the 'Download' tab in the study. Downloadable datasets include both raw and normalized cell x gene matrices, as well as relevant metadata. These datasets are additionally available here to facilitate downloading: https://drive.google.com/drive/folders/1bxCIqNeZ7wLuVOT16gphwj98_cc9KhrV?usp=sharing. We have also posted these cell x gene matrices to Chan Zuckerberg Initiative cellxgene (https://chanzuckerberg.github.io/cellxgene/posts/cellxgene_cziscience_com) and the Broad Institute Single Cell COVID-19 portal (https://singlecell.broadinstitute.org/single_cell/covid19) as leading community efforts. FASTQ files and cell x gene matrices for NHP and murine datasets, and cell x gene matrices for human datasets, are available at GEO: GSE148829.

In this same table, we further highlight four access types. 1. published datasets where everything is available (1 study); 2. unpublished datasets where everything is available (2 studies, 19,670 new cells for download), 3. unpublished datasets where *ACE2*+ cell subsets, and the necessary subsets to contextualize those cells (i.e., epithelial cells for type II pneumocytes) are fully available (5 studies, 17,986 new cells for download); and, 4. those unpublished datasets where expression is shared for *ACE2*/*TMPRSS2* (2 studies, 9,112 new cells). For those unpublished datasets where only specific subsets of cells or genes are available, full expression matrices are available upon request for COVID-19 related questions.

All data included in the present study can be visualized using the following web viewer:
https://singlecell.broadinstitute.org/single_cell?scpbr=the-alexandria-project.

As we gain further insight and feedback from our own groups, collaborators, and investigators, we will continue to provide updates on our resource websites, including the utility of *in vitro* systems, such as organoids (Mead et al., 2018), for the study of SARS-CoV-2: http://shaleklab.com/resource/covid-19-resources/ and www.ordovasmontaneslab.com/covid-19-resources/. We also note that there are several ongoing efforts unified together through the HCA Lung Biological Network group that we will reference and to which we will link as they become available.

No custom code was used to analyze these data and all methods and packages used are cited in the Method Details section.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Intestinal Biopsies

For human intestinal biopsies from the terminal ileum, the subjects were enrolled on a multi-center clinical study, which was approved by the Institutional Review Board at Boston Children's Hospital (protocol number: IRB-P00030890). Full information related to subject age/developmental stage and sex found in metadata associated with provided raw datasets.

Human Lungs, Surgical Excess

Samples were obtained through indicated lung lobe resection or diagnostic procedures in collaboration with clinicians at the Department of Cardiothoracic Surgery at Inkosi Albert Luthuli Central Hospital in Durban, South Africa. Informed consent was obtained from each participant. The study protocol was approved by the University of KwaZulu-Natal Institutional Review Board (approval BE024/09). Full information related to subject age/developmental stage and sex found in metadata associated with provided raw datasets.

Human Nasal Polyps and Scrapings

For inferior turbinate nasal scrapings, polyp scrapings, and ethmoid sinus surgical tissue samples, the Partners HealthCare Institutional Review Board (Boston, Massachusetts), approved the study and all subjects provided written informed consent (Ordovas-Montanes et al., 2018). Full information related to subject age/developmental stage and sex found in metadata associated with provided raw datasets.

Human Nasal Washes, Healthy and Influenza Infected

The Institutional Review Board of the University of Massachusetts Medical School (Worcester, Massachusetts) approved the study and all subjects provided written informed consent.

Cell Culture of Primary Basal Cells and Cell Lines

Human basal cells from non-polyp surgical resections from ethmoid sinus, BEAS-2B cells (ATCC), or mouse tracheal basal cells were placed into culture at a number of 10,000 cells seeded at passage 3 and cultured at confluence in 96 well flat-bottom collagen-coated tissue culture plates (Corning 3596) for 48 h in Pneumacult-Ex serum-free media (StemCell Technologies, Inc.). All cells were incubated at 37°C and 5% CO2.

Non-Human Primates (*M. mulatta*)

Healthy and SHIV-infected non-human primate (*M. mulatta*) work was conducted at the Washington National Primate Research Center (WaNPRC), an AAALAC accredited program, in accordance with the regulations detailed in the U.S. Department of Agriculture Animal Welfare Act and in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. It was approved by University of Washington Institutional Animal Care and Use Committee. Expanded cohort characteristics described previously (Colonna et al., 2018). Full information related to subject age/developmental stage and sex found in metadata associated with provided raw datasets.

Non-Human Primates (*M. fascicularis*)

Tissues from *Mycobacterium tuberculosis*-infected non-human primates (*M. fascicularis*) were conducted at the University of Pittsburgh School of Medicine, an AAALAC accredited program, in accordance with the regulations detailed in the U.S. Department of Agriculture Animal Welfare Act and in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. Full information related to subject age/developmental stage and sex found in metadata associated with provided raw datasets.

Mouse Nasal and Olfactory Epithelium and Tracheal Cells

C57BL/6J mice purchased from Jackson laboratory (Bar Harbor, ME, USA) were maintained within Ragon Institute's HPPF barrier facility and all experiments were conducted with institutional IACUC approval. In this study, mice were 8-10 weeks of age, representing male and female animals.

Mouse Lungs, MHV68 Infection

C57BL/6 mice were purchased from Charles River Laboratories (Sulzfeld, Germany). IFNγR−/− mice on C57BL/6 background (C57BL/6, IFNγR$^{-/-}$ B6.129S7-*Ifngr1*^{tm1Agt}/J) were originally obtained from the Jackson Laboratory (Bar Harbor, ME, USA) and subsequently bred and propagated under SPF conditions at the Helmholtz Zentrum München. Animals with different genotypes were kept in the same animal room for the time of the experiment including an adaptation period prior to the start of the experiment. All animal experiments were in compliance with the German Animal Welfare Act (German Federal Law §8 Abs. 1 TierSchG), and the protocols were approved by the local Animal Care and Use Committee.

METHOD DETAILS

Methods of Sample Collection and Tissue Preparation for Single-Cell RNA-Seq

NHP Ileum, Jejunum, Colon, Liver, Tonsil, Thymus, and Lung Tissue

Animals were perfused with 0.5 L of PBS/kg immediately following euthanasia, tissues were isolated and placed in RPMI + 10% FBS and kept on ice until dissociation. Tissue sections were digested by mincing and incubating with collagenase IV (Life Technologies) and DNase I (Roche) at 37°C for 1 h with agitation. Digested tissue was passed through a 100 μm metal strainer, cells were pelleted by centrifugation at 300 g, rinsed with RPMI + 10% FBS, counted, and prepared as a single cell suspension for scRNA-seq using Seq-Well v1 (see below).

NHP Lymphoid Organs, Bone Marrow, PBMCs

All lymph nodes, spleen, and bone marrow were ground through a metal strainer, transferred to a conical in RPMI + 10% FBS, and pelleted by centrifugation at 400 g x 10 min. LN-derived cells were resuspended in RPMI + 10% FBS, counted and prepared as a single cell suspension. Spleen, bone marrow, and PBMCs were subjected to ACK lysis for 10 min at room temperature, quenched with RPMI + 10% FBS. PBMCs and bone marrow derived cells were purified over a ficoll gradient (GE Healthcare) by centrifuging at 400 g for 20 min at room temperature with no brake. Cells were then resuspended in RPMI + 10% FBS, counted, and diluted for scRNA-seq using Seq-Well v1 (see below).

NHP Tuberculosis Infected Lung and Granuloma

Ten *Mycobacterium tuberculosis* infected (Martin et al., 2017) adult non-human primates (*M. fascicularis*) were included in this study. A piece of lung tissue (without any grossly visible pathology) and 4 individual TB lung granulomas per animal were excised at necropsy and enzymatically dissociated using the GentleMacs system (Tumor dissociation kit, human; Miltenyi Biotec). Single cell suspensions were resuspended in RPMI + 10% FBS, counted and diluted for scRNA-seq using Seq-Well S^3 (see below).

Human Lung Tissue

Surgical samples from diseased lung tissue (n = 3 TB$^+$HIV$^+$; n = 3 TB$^+$; n = 2 non-infected patients) were processed as described in (Ardain et al., 2019). Briefly, each sample was collected into cold RP-10 (RPMI (Sigma-Aldrich) + 10% FBS), minced, and incubated for 25-30 min at 37°C with digestion buffer containing collagenase D (Sigma-Aldrich), DNase I (Sigma-Aldrich) in RPMI 1640 (Sigma-Aldrich) with 10% FBS (Hyclone). Following incubation, samples were homogenized using a GentleMACS, filtered using a 70 μm metal strainer, and pelleted by centrifugation at 400 g for 5 min. After obtaining the pellet, cells were resuspended in RP-10, passed through another 70μm strainer (Corning), stained with trypan blue, and then counted and diluted for scRNA-seq using Seq-Well S^3 (see below).

Human Ileum

Single-cell suspensions were collected from biopsies as described (Smillie et al., 2019). Briefly, biopsies were rinsed in cold PBS, the epithelial layer was separated from the underlying lamina propria by end over end rotation for 15 min. The lamina propria and epithelial fractions were digested separately, using Liberase TM (Roche) and DNase I (Roche) for the lamina propria, and TrypLE (ThermoFisher) for the epithelial fraction. Following digestion, cells were pelleted by centrifugation, subjected to ACK lysis for 3 min, and filtered through a 40 μm strainer. Following centrifugation, the cells were counted and prepared as a single cell suspension for scRNA-seq using 10X 3′ v2 (10X Genomics).

Nasal Mucosa and Nasal Scrapings

Surgical samples from ethmoid sinus and nasal scraping of the inferior turbinate were processed as described (Ordovas-Montanes et al., 2018). Briefly, each sample was collected into cold RPMI (Corning), minced and incubated for 30 min (15 min for nasal

scrapings) at 37°C with digestion buffer containing collagenase IV (Worthington), DNase I (Roche) in RPMI with 10% FBS. Samples were triturated and digestion quenched with EDTA. Cells were filtered using a 70 μm metal strainer and pelleted by centrifugation at 500 g, rinsed with PBS, and subjected to red blood cell (RBC) lysis using ACK buffer (ThermoFisher) for 3 min on ice, and finally pelleted prepared as a single cell suspension for scRNA-seq using Seq-Well v1 or S^3 (see below).

Interferon Treatment of Mouse Nasal Mucosa

Mice received either 200ng of IFNα (Biolegend 752802) or saline intranasally (each group n = 2 mice), and were sacrificed 12 h later. Respiratory and olfactory mucosa were isolated as in (Davidson et al., 2004; Dunston et al., 2013). Briefly, using surgical tools under a dissecting microscope, the skull bones surrounding the nasal tissue of skinned mouse heads were removed. The respiratory and olfactory mucosa were collected in RPMI media with 10% FBS. Cells were digested in media containing Liberase TM (Roche) and DNase I (Roche) for 30 min at 37°C with agitation. Cells were filtered using a 70 μm strainer, washed with EDTA-containing media to quench enzymatic digestion, and pelleted by centrifugation. RBCs were lysed using ACK buffer (ThermoFisher) for 2 min, cells were again pelleted, counted, and prepared as a diluted single cell suspension for scRNA-seq using Seq-Well S^3.

MHV68 Infected Mouse Lung

Mice were housed in individually ventilated cages during the MHV68 infection period. MHV68 stocks were grown and quantified by plaque assay as previously described (Adler et al., 2000). Mice were infected intranasally (i.n.) with 5 × 10^4 plaque forming units of MHV68 diluted in PBS in a total volume of 30 μl. Prior to i.n. infection, mice were anesthetized with medetomidine–midazolam–fentanyl. At the predetermined time points, mice were sacrificed by cervical dislocation and lung tissue was processed for subsequent experiments. All lobes were removed, minced and transferred for mild enzymatic digestion for 20-30 min at 37°C in an enzymatic mix containing Dispase (50 caseinolytic U/mL), Collagenase (2 mg/mL), Elastase (1 mg/mL), and DNase I (30 μg/mL). Single cells were harvested by straining the digested tissue suspension through a 70μm strainer. After centrifugation at 300 x g for 5 min, single cells were counted, and prepared as a single cell suspension. For Drop-seq, cells were aliquoted in PBS supplemented with 0.04% of bovine serum albumin at a final concentration of 100 cells/μl.

Nasal Washes during Influenza Infection

Nasal washes were obtained from adult healthy controls and from adults with diagnosis of acute influenza A or B by rapid antigen test (Flu A or B antigen, direct fluorescence antigen test) and/or by respiratory virus panel (PCR testing for influenza A, influenza A H1, influenza A H3, influenza B, adenovirus, metapneumovirus, respiratory syncytial virus A, respiratory syncytial virus B, rhino/enterovirus, parainfluenza 1, parainfluenza 2, parainfluenza 3), who show symptoms up to seven days (Cao et al., 2020). Samples were obtained by irrigation of each naris with up to 10 mL of saline, and collected in a single container. The sample was then transported to the research laboratory for processing. Upon receipt, the sample was immediately stored on ice and 10 mL cell growth media (DMEM or RPMI1640 with 10% fetal bovine serum) was added. The material was strained using a 40 μm nylon cell strainer (Corning) into a 50 mL centrifuge tube. Cells were pelleted at 1300 rpm for 10 min at 4°C. All but 1 mL of supernatant was discarded, the pellet resuspended in the remaining 1 mL of supernatant, and material was transferred to an Eppendorf tube and pelleted at 2000 rpm for 5 min. If the pellet contained visible blood, 200 μL of RBC lysis solution (ACK buffer, Thermo Fisher) was added to resuspend the pellet and incubated at room temperature for 2 min, after which 1 mL of cell media was added, and the cells were pelleted at 2000 rpm for 5 min. The final pellet was resuspended in up to 1 mL of media and quantified before performing scRNA-seq with Seq-Well v1.

Methods to Generate Single-Cell and Bulk RNA-seq Libraries

Seq-Well v1

Seq-Well was performed as described (Gierahn et al., 2017). Single cells were diluted to 15,000 cells in 200 μL RPMI + 10% FBS and deposited onto a pre-functionalized PDMS array. 15,000 cells were deposited onto the top of each PDMS array and let settle by gravity into distinct wells. The array was gently washed with PBS, and sealed using a functionalized polycarbonate membrane. Seq-Well arrays were sealed in a dry 37°C oven for 40 min, and submerged in a lysis buffer containing guanidium thiocyanate (Sigma), EDTA, 1% beta-mercaptoethanol and sarkosyl (Sigma) for 20 min at room temperature. Arrays were transferred to hybridization buffer containing NaCl (Fisher Scientific) and agitated for 40 min at room temperature, mRNA capture beads with mRNA hybridized were collected from each Seq-Well array, and beads were resuspended in a master mix for reverse transcription containing Maxima H Minus Reverse Transcriptase and buffer, dNTPs, RNase inhibitor, a 5′ template switch oligonucleotide, and PEG for 30 min at room temperature, and overnight at 52°C with end-over-end rotation. Exonuclease digestion and PCR were carried out as described. Post-whole transcriptome amplification workup involved AMPure XP SPRI bead cleanup occurred at a 0.6 x volume ratio, followed by 0.8x. Library size was analyzed using an Agilent Tapestation hsD5000 kit, confirming the expected peak at ~1000 bp, and absence of smaller peaks corresponding to primer. Libraries were quantified using Qubit High-Sensitivity DNA kit and prepared for Illumina sequencing using Nextera XT DNA Sample Preparation kit using 900 pg of cDNA library as input to tagmentation reactions. Amplified final libraries were purified twice with AMPure XP SPRI beads as before, with a volume ratio of 0.6x followed by 0.8x. Libraries from 2-3 Seq-Well arrays were pooled and sequenced together using a NextSeq 500/550 High Output v2 kit (75 cycles) using a paired end read structure with custom read 1 primer: read 1: 20 bases, read 2: 50 bases, read 1 index: 8 bases.

Seq-Well S^3

Seq-Well S^3 modified the following protocol steps from v1, above (Hughes et al., 2019). First, hybridization buffer was supplanted with 8% (v/v) polyethylene glycol (PEG, Sigma). Second, after exonuclease digestion, bead-associated cDNA was denatured for 5 min in 0.2 mM NaOH with end over end rotation. Next, beads were washed with TE + 0.01% tween-20, and second strand synthesis

was carried out by resuspending beads in a master mix containing Klenow Fragment (NEB), dNTPs, PEG, and the dN-SMRT oligonucleotide to enable random priming off of the beads.

10X v2 3′

Single cells were loaded onto 3′ library chips as per the manufacturers protocol for Chromium Single Cell 3′ Library (v2) (10X Genomics). Each biopsy was sequenced on two channels of the 10X Chromium Single Cell Platform, one for the epithelial fraction and the other for the lamina propria fraction in order to recover sufficient numbers of epithelial and lamina propria cells for downstream analyses. An input of 6,000 single cells was added to each channel with a recovery rate of approximately 2,000 cells.

Drop-seq

Drop-seq experiments were performed according to the original protocol (Macosko et al., 2015). Briefly, single cells (100/μl) were co-encapsulated in droplets with barcoded beads (120/μl, ChemGenes) at rates of 4000 μl/h. Droplet emulsions were collected for 10-20 min/each prior to droplet breakage by perfluorooctanol (Sigma-Aldrich). After breakage, beads were harvested and the hybridized mRNA transcripts reverse transcribed (Maxima RT, Thermo Fisher). Exonuclease digestion and PCR were carried out as described (12 PCR cycles). For each sample, 1 ng of pre-amplified cDNA from an estimated 1000 cells was tagmented by Nextera XT (Illumina) with a custom P5-primer (Integrated DNA Technologies). Single-cell libraries were sequenced in a 100 bp paired-end run on the Illumina HiSeq4000.

Smart-Seq2 for Bulk RNA-Seq

Population RNA-seq was performed as described (Ordovas-Montanes et al., 2018; Trombetta et al., 2014). Briefly, RNA from population lysates was purified using AMPure RNA Clean Spri beads (Beckman Coulter) at a 2.2x volume ratio, and mixed with oligo-dT primer, dNTPs (NEB), and RNase inhibitor (Fisher Scientific) at 72°C for 3 min on a thermal cycler to anneal the 3′ primer to polyadenylated mRNA. Reverse transcription was carried out in a master mix of Maxima RNaseH-minus RT enzyme and buffer (Fisher Scientific), $MgCl_2$ (Sigma), Betaine (Sigma), RNase inhibitor, and a 5′ template switch oligonucleotide, and PCR was carried out using KAPA HiFi HotStart ReadyMix (Kapa Biosystems) and IS PCR primer and amplified for 18 cycles. Libraries were purified using AMPure XP SPRI beads at a volume ratio of 0.8x followed by 0.9x. Library size was assessed using a High-Sensitivity DNA chip (Agilent Bioanalyzer), confirming the expected size distribution of ~1,000-2,000 bp. Tagmentation reactions were carried out with the Nextera XT DNA Sample Preparation Kit (Illumina) using 250 pg of cDNA per single cell as input, with modified manufacturer's instructions as described. Libraries were purified twice with AMPure XP SPRI beads at a volume ratio of 0.9x, size distribution assessed using a High Sensitivity DNA chip (Agilent Bioanalyzer) and Qubit High-Sensitivity DNA kit (Invitrogen). Libraries were pooled and sequenced using NextSeq500/550 High Output v2 kits (75 cycles, Illumina) using 30-30 paired end sequencing with 8-mer dual indexing.

Human and Mouse Basal Cell Cytokine Stimulation

Data represented in Figures 5A–5L: Cytokines were added for 12 h overnight at increasing doses (0, 0.1, 0.5, 1, 2, 5, 10 ng/mL) of IL-4 (human: Biolegend 574002), IL-17A (human: Biolegend 570502), IFNγ (human: Biolegend 570202; mouse: Peprotech 315-05), IFNα (human: Biolegend 592702; mouse: Biolegend 752802), or IFNβ (mouse: R&D Systems 8234-MB-010). Each condition was run as a biological triplicate. Data represented in Figure S3C–K: cytokines were added for 12 h overnight at increasing doses (0, 0.1, 0.5, 1, 5, 10 ng/mL) of human IL-4 (Biolegend 574004), IL-13 (Biolegend 571104), IFNα (Biolegend 592704), IFNγ (Biolegend 570204), IL-17A (Biolegend 570504), or IL-1β (Biolegend 579404) (each condition run as a biological quadruplicate). All populations were lysed in 50 μL lysis buffer (RLT + 1% BME, QIAGEN and Sigma, respectively) and snap frozen on dry ice. Bulk RNA-seq was performed as described previously and summarized above (Ordovas-Montanes et al., 2018). Populations were sequenced to an average ± SEM read depth of 3.95 ± 0.11 million reads per sample, with an average ± SEM alignment percentage to either hg19 or mm10 reference transcriptomes of 71 ± 0.3%. All samples met quality thresholds regarding genomic and transcriptomic alignment.

Western blot for human ACE2

Established air-liquid interface cultures from bronchial brushings of four asthmatic patients were treated with 10ng/μL of human IFNγ for 24 h. Protein lysates were prepared, and anti-ACE2 human antibody (AF933 R&D goat polyclonal) was used to probe for ACE2 by western blot. Bands were normalized to GAPDH as loading control, and fold change was computed based on normalized ACE2 values.

QUANTIFICATION AND STATISTICAL ANALYSIS

Non-Human Primate Lung and Ileum

Libraries corresponding to 7 animals (variable number of tissues per animal) were sequenced using Illumina NextSeq. Reads were aligned to the *M. mulatta* genome assembly 8.0.1 annotation version 102 and processed according to the Drop-Seq Computational Protocol v2.0 (https://github.com/broadinstitute/Drop-seq). Data was normalized and scaled using the Seurat R package v2.3.4 (https://satijalab.org/seurat/): transforming the data to $\log_e$(UMI+1) and applying a scale factor of 10,000. To identify major axes of variation within our data, we first examined only highly variable genes across all cells, yielding approximately 1,000-3,000 variable genes with average expression > 0.1 log-normalized UMI across all cells. An approximate principal component analysis was applied to the cells to generate 100 principal components (PCs). Using the JackStraw function within Seurat, we identified significant PCs to be used for subsequent clustering and further dimensionality reduction. For 2D visualization and cell type clustering, we used a Uni-

form Manifold Approximation and Projection (UMAP) dimensionality reduction technique (https://github.com/lmcinnes/umap) with "min_dist" set to 0.5 and "n_neighbors" set to 30. To identify clusters of transcriptionally similar cells, we employed unsupervised clustering as described above using the FindClusters tool within the Seurat R package with default parameters and k.param set to 10 and resolution set to 0.5. Each cluster was sub-clustered to identify more granular cell types, requiring each cell type to express > 25 significantly upregulated genes by differential expression test (FindMarkers implemented in Seurat, setting "test.use" to "bimod," Bonferroni-adjusted p value cutoff < 0.001). Differential expression tests between cells from *ACE2*$^{+}$ versus *ACE2*$^{-}$ Type II Pneumocytes were conducted using the SCDE R package with default parameters (Kharchenko et al., 2014). Expression data for epithelial cells and enterocytes included in this dataset can be visualized and downloaded here: https://singlecell.broadinstitute.org/single_cell/study/SCP807?scpbr=the-alexandria-project#study-summary.

Human Lung Tissue

Libraries corresponding to 8 donors were sequenced using Illumina NextSeq. Reads were aligned to the hg19 genome assembly and processed according to the Drop-Seq Computational Protocol v2.0 (https://github.com/broadinstitute/Drop-seq). Data was normalized and scaled using the Seurat R package v3.1.0 (https://satijalab.org/seurat/), transforming the data to $\log_e$(UMI+1) and applying a scale factor of 10,000. For each array, we assessed the quality of constructed libraries by examining the distribution of reads, genes and transcripts per cell. Variable gene selection, principal components analysis, and selection of significant principal components was performed as above. We visualized our results in a two-dimensional space using UMAP (https://github.com/lmcinnes/umap), and annotated each cluster based on the identification of highly expressed genes. To further characterize substructure within cell types (for example, T cells), we performed dimensionality reduction (PCA) and clustering over those cells alone. Sub-clusters (i.e., clusters within broad cell type classifications) were annotated by cross-referencing cluster-defining genes with curated gene lists and online databases SaVanT (http://newpathways.mcdb.ucla.edu/savant-dev/) and GSEA/MsigDB (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp). Proliferating cells from the human lung (Figure 2C) express high levels of mitotic markers, such as *MKI67*, and represent primarily T cells (*CD3D, CD3E*), B cells/antibody-secreting cells (*IGJ, MZB1, IGHG1*), and myeloid cells (*CD14, APOE*) and represent a composite cell cluster. Differential expression analysis between *ACE2+ TMPRSS2+* and negative type II pneumocytes was performed in Seurat using a likelihood-ratio test (FindMarkers implemented in Seurat, setting "test.use" to bimod). Expression data for epithelial cells included in this dataset can be visualized and downloaded here: https://singlecell.broadinstitute.org/single_cell/study/SCP814?scpbr=the-alexandria-project#study-summary.

Human Ileum

Libraries corresponding to 13 donors were sequenced using Illumina NovaSeq S2 with a Read 1 26bp, Read 2 91bp, Index 1 8bp configuration before reads were aligned to GRCh38. Each sample was filtered individually for low quality cells and genes by analyzing distributions of reads, transcripts, percent reads mapped to mitochondrial genes, and complexity per cell, then merged as an outer join to create a single dataset. Clustering and differential expression tests were processed using Seurat v3.1.0 (https://satijalab.org/seurat/). Normalization and variable gene selection was processed with SCTransform (https://github.com/ChristophH/sctransform). Clustering for major cell types was performed using Louvain clustering on dimensionally reduced PCA space with resolution set via grid search optimizing for maximum average silhouette score. Due to the scale of the dataset, a randomized subsampling from across the dataset was used to calculate the silhouette score. We annotated clusters based on highly expressed genes, then sub-clusters were characterized by performing PCA dimensionality reduction and clustering over those cells alone, and annotated based on highly expressed genes found via one-versus-rest differential expression test (Wilcoxon) within the major cell type. Differential expression analysis between *ACE2*$^{+}$*TMPRSS2*$^{+}$ and negative epithelial cells was performed in Seurat using a Wilcoxon test and Bonferroni p value correction. Expression data for epithelial cells included in this dataset can be visualized and downloaded here: https://singlecell.broadinstitute.org/single_cell/study/SCP812?scpbr=the-alexandria-project#study-summary.

Human Adult Nasal Mucosa

Sample processing, sequencing, and analysis was performed as in (Ordovas-Montanes et al., 2018). Briefly, scRNA-seq cell suspensions were freshly processed using Seq-Well v1 and Seurat v2.3.4 was utilized for computational analyses presented here (Butler et al., 2018; Satija et al., 2015). Cell by gene matrix and R code for initialization of object available to download as Supplemental Data and Supplementary Tables here https://www.nature.com/articles/s41586-018-0449-8 and here: http://shaleklab.com/resource/mapping-allergic-inflammation/? and visualized here: https://singlecell.broadinstitute.org/single_cell/study/SCP253?scpbr=the-alexandria-project#study-summary. Scores for various cytokines acting on human airway epithelial cells were calculated based on gene lists derived for (Ordovas-Montanes et al., 2018), calculated using AddModuleScore function Seurat, and effect size calculated by Cohen's d, as previously reported.

Granulomatous Tissue from Mycobacterium Tuberculosis Infected NHPs

Libraries corresponding to 10 animals (variable number of tissues/animal) were sequenced using Illumina NovaSeq S2. Data was aligned using the Dropseq-tools pipeline on Terra (app.terra.bio) to *M. fascicularis* reference genome assembly 5, annotation version 101. Clustering was performed using Leiden clustering in the Scanpy (scanpy.readthedocs.io) package (Wolf et al., 2018). Cell type labels were assigned using known marker genes. In this analysis, we include all epithelial cell subsets (secretory, multiciliated, type II

pneumocytes, and type I pneumocytes) from all samples. Differential expression between *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells and other cells of the matched cell subtype (e.g., Secretory Cells) were performed using the "bimod" likelihood-ratio test within each cell subtype and filtered on Benjamini-Hochberg-corrected p value < 0.05. Expression data for epithelial cells included in this dataset can be visualized and downloaded here:

https://singlecell.broadinstitute.org/single_cell/study/SCP806?scpbr=the-alexandria-project#study-summary.

Basal Cell Cytokine Stimulation

Libraries corresponding to 279 populations were sequenced using Illumina NextSeq. Reads were aligned to the hg19 or mm10 genome assembly using the cumulus platform https://cumulus-doc.readthedocs.io/en/0.12.0/smart_seq_2.html and output as TPM using RSEM v1.3.2. Populations were transformed to transcripts per 10K reads and log2(1+TP10K) transformed. *ACE2* expression by stimulation condition and dose were assessed using one-way ANOVA with post hoc testing using a Bonferroni correction. Plots were generated using ggplot2, and transcriptome-wide differential expression was calculated using the Seurat R package v3.1.0 (https://satijalab.org/seurat), function FindMarkers with test.use = "bimod." Expression data can be visualized and downloaded here:

https://singlecell.broadinstitute.org/single_cell/study/SCP822?scpbr=the-alexandria-project.

Interferon Treatment of Mouse Nasal Mucosa

Libraries corresponding to 4 mice, with 2 Seq-Well arrays per mouse were sequenced using Illumina NextSeq as described (Gierahn et al., 2017; Hughes et al., 2019). Reads were aligned to the mm10 genome and processed according to the Drop-Seq Computational Protocol v2.0 (https://github.com/broadinstitute/Drop-seq). Data was normalized and scaled using the Seurat R package v2.3.4 (https://satijalab.org/seurat/): transforming the data to $\log_e$(UMI+1) and applying a scale factor of 10,000. Cells with fewer than 1000 UMIs and 500 unique genes were removed. To identify major axes of variation within our data, we first examined only highly variable genes across all cells, yielding approximately 5,000 variable genes. An approximate principal component analysis was applied to the cells to generate 200 principal components (PCs). Using a combination of the Jackstraw function in Seurat and observing the "elbow" of the standard deviations of PCs, we chose the top 70 PCs for subsequent clustering and visualization. For 2D visualization, we used a Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction technique (https://github.com/lmcinnes/umap) with "min_dist" set to 0.3 and "n_neighbors" set to 50. To identify clusters of transcriptionally similar cells, we employed unsupervised clustering as described above using the FindClusters tool within the Seurat R package with default parameters and k.param set to 10. Resolution was chosen based on maximization of the average silhouette width across all cells. Clusters were merged if a cell type expressed fewer than 25 significantly upregulated genes by differential expression test (FindAllMarkers implemented in Seurat, setting "test.use" to "bimod," Bonferroni-adjusted p value cutoff < 0.001). Differential expression tests between cells from saline-treated or IFNa-treated mice were assessed using the FindMarkers function with "test.use" set to "bimod. This dataset can be visualized and downloaded here:

https://singlecell.broadinstitute.org/single_cell/study/SCP832?scpbr=the-alexandria-project#study-summary.

Lung from MHV68-Infected WT and IFNγR KO Mice

Libraries corresponding to 14 mice were aligned to a custom reference genome encompassing both murine (mm10) and herpes virus genes: 84 known genes from MHV68 were retrieved from NCBI (NCBI: txid33708) and added to the mm10 mouse genome. Reads were aligned to the custom joint genome and processed according to the Drop-Seq Computational Protocol v2.0 (https://github.com/broadinstitute/Drop-seq). Barcodes with < 200 unique genes, > 20,000 UMI counts, and > 30% of transcript counts derived from mitochondrially encoded genes were discarded. Data analysis was performed using the Scanpy Package following the common procedure, the expression matrices were normalized using *scran*'s size factor based approach and log transformed via scanpy's pp.log1p() function (Lun et al., 2016; Wolf et al., 2018). SoupX was utilized to reduce ambient RNA bias, using default parameters with pCut set to 0.3, and was applied to each sample before merging the count matrices (Young and Behjati, 2020). UMI per cell and cell cycle were regressed out. Highly variable genes were selected by running pp.highly_variable_genes() for each sample separately, returning the top 4,000 variable genes per sample, and genes identified in variable in > 5 samples were retained, yielding 14,305 genes. Next, only *Epcam*+ cells were considered, principal components (PCs) were calculated using only the selected variable genes, and 6 PCs were used to perform unsupervised Louvain clustering. Type I Pneumocytes were excluded from this analysis based on uniformly negative expression of *Ace2*, resulting in a final dataset subset of 5,558 cells. Cells were identified as infected if at least one viral read was detected.

Nasal Washes during Influenza Infection

Sample processing, sequencing, and analysis was performed as in (Cao et al., 2020). Reads were aligned to the GRCh37 reference genome combined with influenza genomes. Mapped reads from each sample were then corrected for Drop-seq barcode synthesis error using the Drop-seq core computational tools developed by the McCarroll Lab (Macosko et al., 2015). Genes were quantified using End Sequence Analysis Toolkit (ESAT, github/garber-lab/ESAT) with parameters *-wlen 100 -wOlap 50 -wExt 0 -scPrep* (Derr et al., 2016). Finally, UMIs that likely result from sequencing errors were corrected by merging any UMIs that were observed only once and have 1 hamming distance from a UMI detected by two or more aligned reads. Only cell barcodes with more than

1,000 UMIs were analyzed. Cell barcodes with mostly erythrocyte genes (*HBA, HBB*) were removed. From here on, the remaining cell barcodes in the matrix would be referred to as cells. The final gene by cell matrix was normalized using the scran package v3.10 (Lun et al., 2016). The normalized matrix was used for dimensionality reduction by first selecting variable genes that had a high coefficient of variance (CV) and were expressed (> = 1 UMI) by more than three cells. Influenza viral genes, interferon stimulated genes, and cell cycle related genes were removed from the variable gene list in order to minimize the impact of viral responses and mitosis on clustering and cell type identification. This resulted in the selection of 2484 variable genes. t-distributed stochastic neighbor embedding (tSNE) was applied to the first ten principal components (PCs), which explained 95% of the total data variance. Density clustering (Rodriguez and Laio, 2014) was performed on the resulting tSNE coordinates and identified four major clusters: epithelial cells, neutrophils, macrophages and leukocytes. The epithelial cell cluster and the leukocyte cluster were then re-clustered independently, as described above, to identify populations within each metacluster. Specifically, the epithelial cell cluster was re-embedded using 2629 variable genes selected by the same criteria mentioned in the previous section and 13 PCs that explained 95% of the variance. Density clustering over the epithelial cell subset revealed ten clusters. Differential gene expression analysis using edgeR (Robinson et al., 2010) was performed to identify marker genes for each cluster. Influenza-infected and bystander cells were identified after correcting for sample-specific distribution of ambient influenza mRNA contamination and predicted cells most likely to be infected identified using a hurdle zero inflated negative binomial (ZINB) model and a support vector machine (SVM) classifier.

Power Calculations for Detection of Rare Transcripts

We conducted the following statistical analysis to estimate the effects of various factors on our ability to make confident claims regarding the presence/absence of transcripts of interest (e.g., *ACE2*), both within individual cells and clusters (Figure S6). Specifically, we investigated the roles of capture/reverse transcription efficiency, *ACE2* expression level, sequencing depth, and cell numbers. Taken together, the results of this power analysis are in agreement with other efforts to model biological and technical sources of zero-inflation within scRNA-seq data (e.g., https://satijalab.org/howmanycells and Kharchenko et al., 2014; Svensson, 2020).

We began by quantifying how likely we are to capture and transcribe at least one *ACE2* mRNA molecule, as a function of the number *ACE2* mRNA molecules per cell and a protocol's efficiency (Figure S6A). Drop-Seq has a capture/transcription efficiency of ~10% (as estimated using ERCC spike ins; see (Macosko et al., 2015), and the experimental platforms used in this study are either equivalent (e.g., Seq-Well v1, (Gierahn et al., 2017) or superior (e.g., 10-fold better unique molecule detection, 5-fold better gene detection using Seq-Well S^3;(Hughes et al., 2019)). Most relevant to this context, inferior turbinate scrapings were processed using both Seq-Well v1 and Seq-Well S^3 (Figure S3B). Importantly, Seq-Well S^3 provided > two-fold increase in the detection frequency of rare *ACE2* transcripts (i.e., *ACE2*+: 4.7% for v1 versus 9.8% for S^3), making it reasonable to expect that such improvements in single-cell experimental technologies have yielded corresponding improvements in capture and transcription efficiency. Based on Drop-Seq's 10% efficiency, even if *ACE2* is expressed at the low level of 5 mRNA molecules per cell (a reasonable order-of-magnitude estimate, given that non-human primate ileum cells had a maximum of 10 *ACE2* unique molecules per cell observed via sequencing and an average of 1.93 molecules per cell in expressing cells, see Figures 3B and 3C), our experimental platforms have a minimum likelihood of 41% to capture and reverse transcribe at least one *ACE2* mRNA molecule in any given individual cell. This likelihood rapidly increases if we estimate higher efficiencies for improved scRNA-seq technologies (e.g., 67% likelihood within any individual cell at 20% capture/transcription efficiency, 76% likelihood at 25% efficiency, Figure S6A). Thus, while transcript drop-out may reduce the fraction of positive cells, with the capture and transcription efficiencies of improved single-cell technologies, the impact is likely to be minor (reads are likely underestimated by up to a factor of ~2.5x), given a sufficient depth of sequencing (see below). We note that this impacts both clusters deemed to contain and not contain *ACE2*+ cells, and suggests our percentages are likely lower bounds for true expression (within a factor of ~2.5x).

Next, we examined the probability of sequencing an *ACE2* transcript as a function of read depth and *ACE2*'s fractional abundance in each single cell within our sequencing libraries. First, across two different tissues (non-human primate ileum and lung, representing a high expresser of *ACE2* and low expresser, respectively), we calculated the proportion of unique *ACE2* molecules in our *ACE2*+ cells (defined as any cell with at least 1 UMI aligning to *ACE2*) as a fraction of total reads within individual cells to provide an order-of-magnitude estimate for average *ACE2* abundance in our single-cell sequencing libraries (i.e., the probability that a read within a cell corresponds to a unique molecule of *ACE2*, Figure S6B). We highlight that by calculating probabilities based on *ACE2* unique molecules divided by an individual cell's total reads, we are providing a conservative estimate for the probability of observing *ACE2* as a function of sequencing depth (e.g., as compared to basing these probabilities on *ACE2* non-UMI-collapsed reads divided by total reads). Next, we obtained information on the number of reads in these cell populations to provide estimates of average sequencing depths (Figure S6C). Using the mean fractional abundances of ACE2 from each tissue (Figure S6B) and the mean read depths for all genes (Figure S6C), we calculated the probability of detecting at least 1 *ACE2* molecule (i.e., P(detecting > 0 ACE2 molecules) = $1 - (1 - \text{ACE2 fractional abundance})^{\text{Read depth}}$). This results in a 93.7% probability in ileum-derived cell libraries that contain *ACE2*, and a 76.0% probability for lung-derived cell libraries, indicating that our sequencing depths are sufficient to detect *ACE2*+ cells (Figure S6D).

To further evaluate whether our ability to detect *ACE2*+ cells was an artifact of sequencing depth, we compared the number of *ACE2*+ cells in a cluster to the mean number of reads across all cells in that same cluster (Figure S6E). We did not observe any significant correlation: the ileum cell cluster with the highest number of *ACE2*+ cells had the lowest sequencing depth of all ileum clusters, and the lung cell cluster with the highest number of *ACE2*+ cells was approximately average in its read depth (on a log-log scale,

Pearson's r = −0.31, non-significant). Further, when comparing *ACE2*+ cells to *ACE2*- cells within a given tissue, we did not observe a positive correlation between read depth and *ACE2* status (i.e., mean ± standard error of the mean, SEM, reads among all lung cells = 28,512 ± 344; mean ± SEM reads among *ACE2*+ lung cells = 28,553 ± 2,988; mean ± SEM reads among all ileum cells = 14,864 ± 288; mean ± SEM reads among *ACE2*+ ileum cells = 10,591 ± 441, full statistics on cell depth among *ACE2*+ cells compared to *ACE2*- cells of the same cell type can be found in Table S9). Thus, we can be confident that the observed differences in *ACE2*+ proportions across clusters are not driven by differences in sequencing depth.

Finally, we investigated how observed differences in *ACE2*+ proportions across clusters might be affected by cell sampling. Using the proportion of *ACE2*+ cells in a "typical" cluster annotated as being *ACE2* positive (i.e., 6.8% in non-human primate type II pneumocytes, Figure 1), we calculated the cluster sizes needed to be confident that the probability of observing zero to a few positive cells is unlikely to have arisen by random chance (probabilities calculated under a negative binomial distribution with parameter p = 0.068, Figure S6E). We found that as cluster sizes approach and exceed 100 cells, the probability of observing zero to a few positive cells rapidly approaches zero, if we assume 6.8% of cells are positive. Further, to examine our confidence in estimating an approximate upper bound (ignoring the impact of protocol inefficiencies discussed above) for the fraction of cells positive in a cluster as a function of the number of cells in that cluster, we also calculated the probability of observing zero (and its complement, probability of observing at least 1) *ACE2*+ cells as a function of cluster size across true positive proportions ranging from 0.1% to 10% (probabilities calculated under a negative binomial distribution with parameter p = 0.001 to 0.1, representing hypothetical proportions of *ACE2*+ cells Figure S6F). Given our typical cluster sizes (on the order of hundreds of cells, exact values provided in Table S9), we find that for us to observe 0 *ACE2*+ cells in a cluster due to sampling artifacts, the fraction of true positives must be ~1% or less. Thus, these complementary approaches demonstrate that our observed variations in *ACE2*+ cell proportions across clusters likely reflect underlying biological differences, rather than random chance.

Statistical Testing

Parameters such as sample size, number of replicates, number of independent experiments, measures of center, dispersion, and precision (mean ± SEM) and statistical significances are reported in Figures and Figure Legends. A p value less than 0.05 was considered significant. Where appropriate, a Bonferroni or FDR correction was used to account for multiple tests, alternative correction methods are noted in the figure legends or Methods. All statistical tests corresponding to differential gene expression are described above and completed using R language for Statistical Computing.

Supplemental Figures

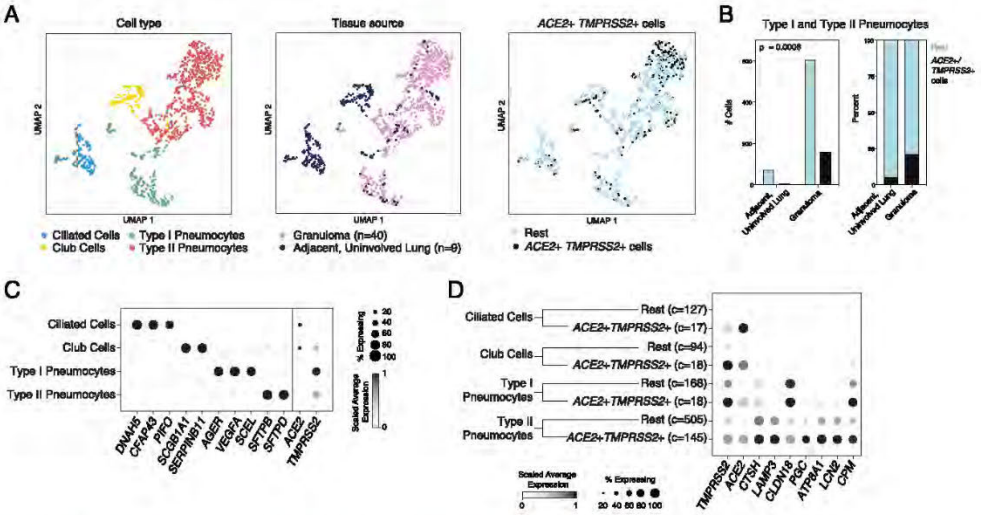


Figure S1. NHP Tuberculosis Infected Lung and Granuloma, Related to Figures 1 and 2

(A). UMAP projection of epithelial cells (1,099 cells) colored by annotated cell type, tissue source, and gating as *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells. *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells comprise 11% of ciliated cells, 16% of club cells, 10% type I pneumocytes, and 22% type II pneumocytes. Data generated using Seq-Well S^{2} (Table S3).

(B). Number of cells (left) and % (right) *ACE2*$^{+}$*TMPRSS2*$^{+}$ cells by tissue source (granuloma versus uninvolved lung) and cell type. Ciliated cells and club cells were omitted from this analysis as we detected too few cells (< 7 total cells) belonging to these clusters in the granulomas. Statistical significance assessed by Fisher Exact Test (Table S3).

(C). Dot plot of top cluster defining genes for each epithelial cell type and *ACE2* and *TMPRSS2*. Dot size represents fraction of cells expressing, and color intensity represents average log(normalized UMI + 1) among all cells in each group scaled between 0 and 1 by gene. *ACE2* expression is enriched in club cells (Bimodal test, Bonferroni-corrected $p < 0.001$), ciliated cells ($p < 0.005$), and type I pneumocytes ($p < 0.001$). *TMPRSS2* expression is enriched in type I pneumocytes ($p < 0.001$) and ciliated cells ($p < 0.001$) (Table S3).

(D). Dot plot of genes differentially expressed between *ACE2*$^{+}$*TMPRSS2*$^{+}$ epithelial cells versus rest (Bimodal test, Bonferroni-corrected $p < 0.01$, log fold change > 0.5). (Table S3, c = number of cells, n = number of animals).

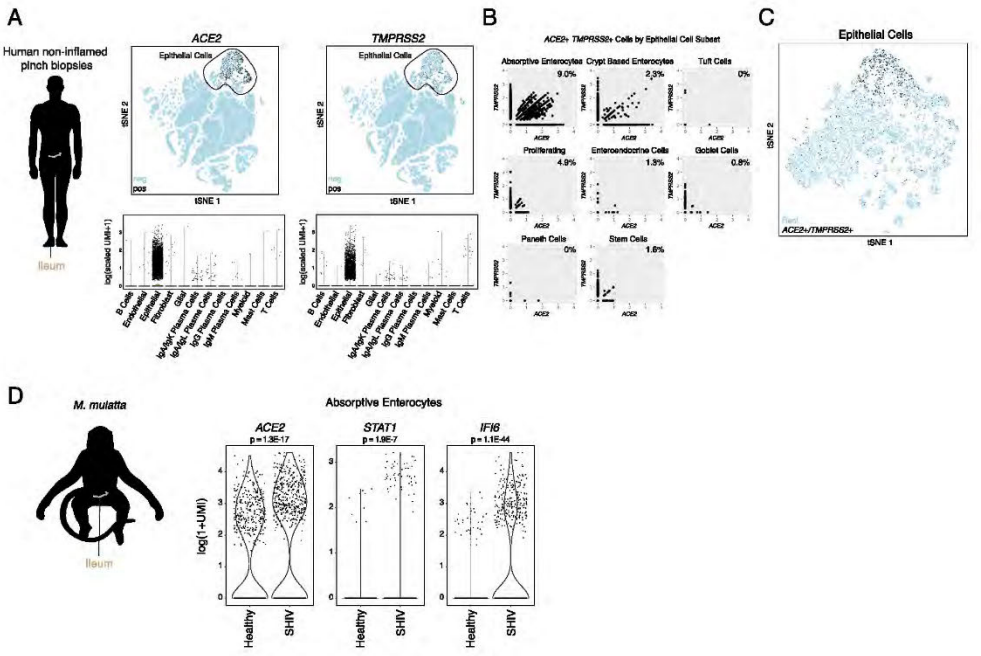


Figure S2. Human and NHP Ileum, Related to Figure 3

(A). Top: tSNE projection of all cells from healthy pediatric human ileum within a previously-unpublished 10x 3′ v2 dataset (115,569 cells). Black: higher expression of *ACE2* (left), *TMPRSS2* (right). Bottom: Corresponding violin plots of expression values for *ACE2* (left) and *TMPRSS2* (right). Solid line: epithelial cells.

(B). Co-expression of *ACE2* and *TMPRSS2* by epithelial cell subset. Number indicates % of *ACE2*$^{-}$*TMPRSS2*$^{+}$ cells by cell subset.

(C). tSNE projection of 13,689 cells as in Figure 3D, cells colored by co-expression of *ACE2* and *TMPRSS2* (black).

(D). Expression of *ACE2* and canonical interferon-responsive genes among absorptive enterocytes from Healthy (n = 2) and SHIV-infected, anti-retroviral treated animals (n = 3). Bonferroni-adjusted p-values by Wilcoxon test (healthy: 510 cells, SHIV-infected: 636 cells).

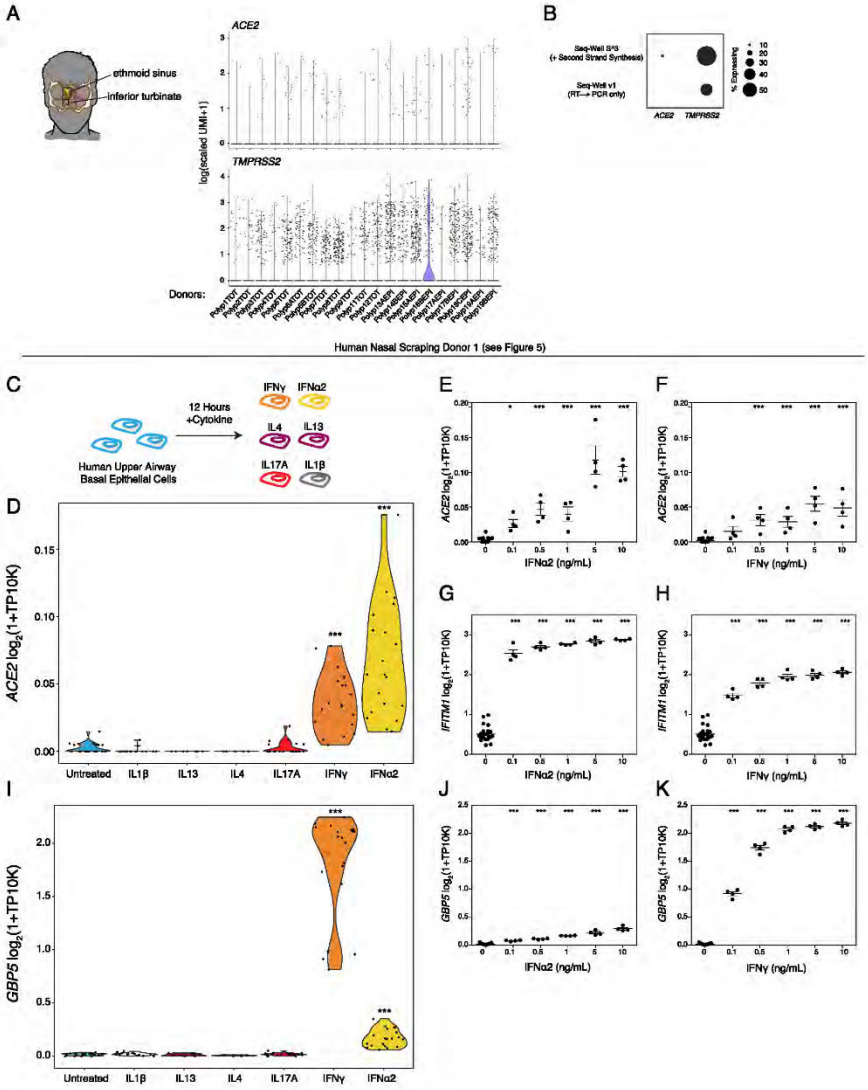


Figure S3. Nasal and Sinus Mucosa, Related to Figures 4 and 5

(A). Expression of *ACE2* and *TMPRSS2* across donors.

(B). Enhanced capture of *ACE2* mRNA with second strand synthesis protocol employed in Seq-Well S^3. Dot size represents fraction of cells expressing.

(C). Cultured human primary basal epithelial cells at confluence were treated with increasing doses (0.1 to 10ng/mL) of IFNα2, IFNγ, IL-4, IL-13, IL-17A, and IL-1B for 12 h and bulk RNA-seq analysis was performed (Replicate experiment using Human Donor 1 as in Figure 5)

(D). *ACE2* expression by stimulation condition. Wilcoxon test between each cytokine (combined doses) versus rest: IFNα Bonferroni-adjusted p = 4.1E-07; IFNγ Bonferroni-adjusted p = 9.3E-03; all else n.s. *** $p < 0.001$.

(E). *ACE2* expression by IFNα2 dose. Bonferroni-corrected t-test compared to 0 ng/mL condition: *** $p < 0.001$, * $p < 0.05$.

(legend continued on next page)

(F). *ACE2* expression by IFNγ dose. Bonferroni-corrected t-test compared to 0 ng/mL condition: *** $p < 0.001$, * $p < 0.05$.
(G). *IFITM1* expression by IFNα2 dose. Bonferroni-corrected t-test compared to 0 ng/mL condition: *** $p < 0.001$.
(H). *IFITM1* expression by IFNγ dose. Bonferroni-corrected t-test compared to 0 ng/mL condition: *** $p < 0.001$.
(I). *GBP5* expression among cultured human primary basal epithelial cells. Wilcoxon test: IFNα versus IFNγ Bonferroni-adjusted p = 2.94E-07; IFNγ Bonferroni-adjusted p = 9.3E-03. TP10K: transcripts per 10,000 reads. *** $p < 0.001$.
(J). *GBP5* expression by IFNα2 dose. Bonferroni-corrected t-test compared to 0 ng/mL condition: *** $p < 0.001$.
(K). *GBP5* expression by IFNγ dose. Bonferroni-corrected t-test compared to 0 ng/mL condition: *** $p < 0.001$.

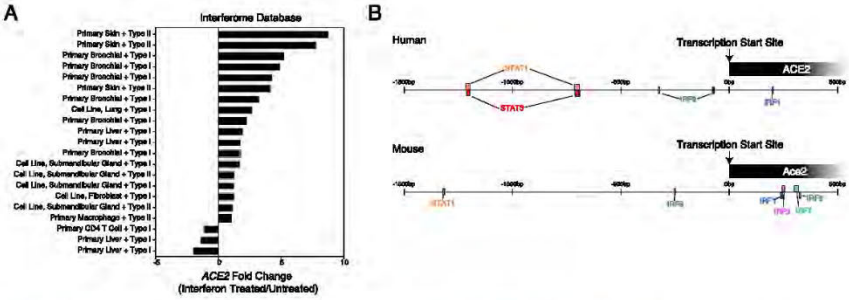
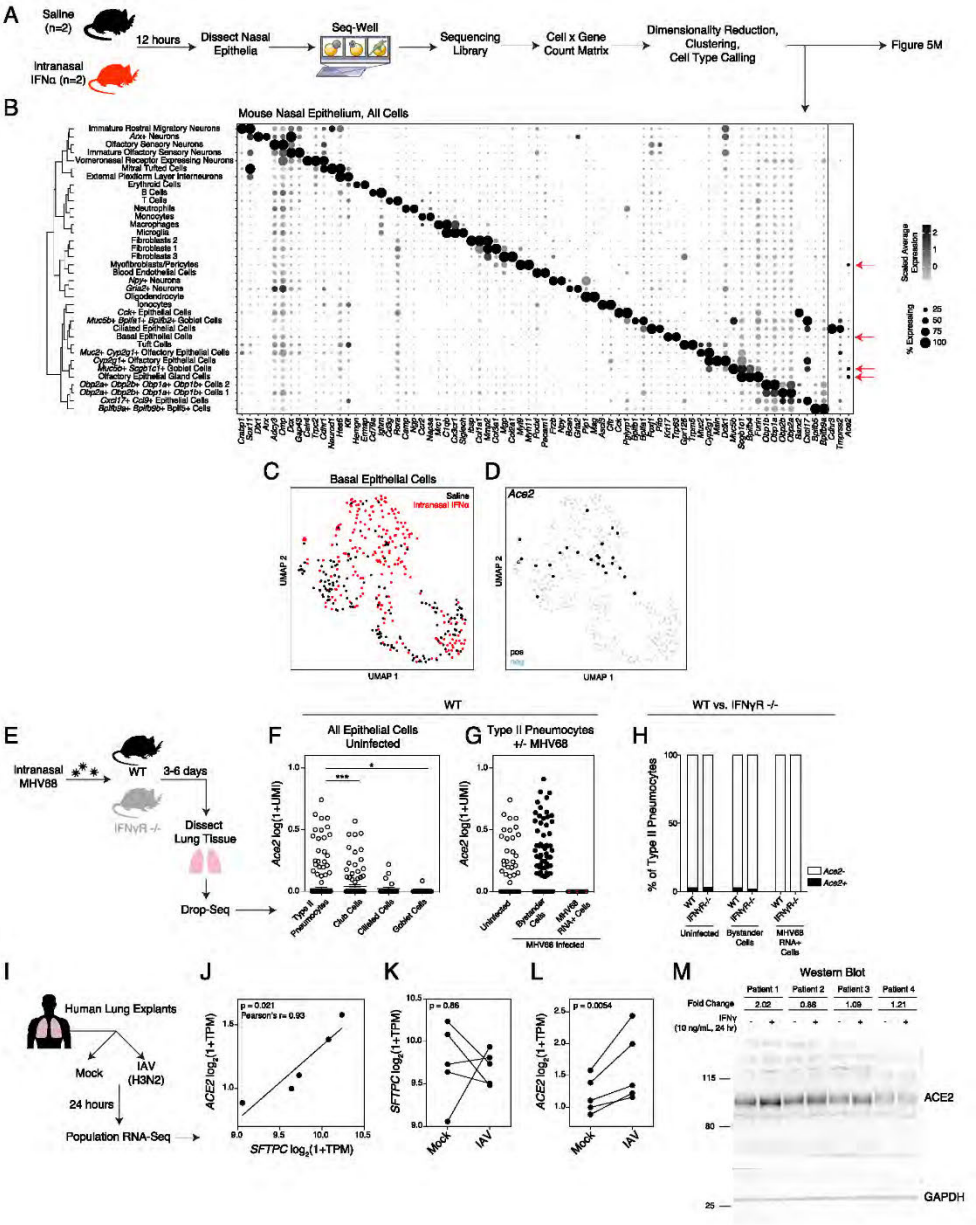


Figure S4. Published Studies of Epithelial Cells Following Interferon Treatment Related to Figure 5

(A). Fold change of *ACE2* expression among human or mouse datasets following Type I or Type II interferon treatment compared to untreated control. Generated from publicly available microarray data curated at interferome.org. Includes all studies with abs(fold-change) > 1.

(B). Location of transcription factors binding regions spanning −1500 bp to +500 bp from the transcription start site of *ACE2* (human, top) or *Ace2* (mouse, bottom). Generated from TRANSFAC data using the interferome.org database (Matys et al., 2003; Rusinova et al., 2013).



(legend on next page)

Figure S5. Mouse Nasal Epithelium Following Interferon-α Exposure Related to Figure 6

(A). Schematic: mice were exposed to 10,000 units of IFN-α or saline by intranasal application (n = 2 per group). After 12 h, animals were sacrificed and nasal epithelium was dissected and dissociated for scRNA-seq using Seq-Well S^3.

(B). Dot plot of 2 defining genes for each cell type, with *Ace2*, *Tmprss2*, and *Cdhr3*. Dot size represents fraction of cells within cell type expressing, and color intensity binned count-based expression level (log(scaled UMI+1)) among expressing cells. All cluster defining genes are provided in Table S8. Red arrows: cell types with largest proportion of *Ace2*+ cells. Dendrogram (left) by person correlation over differentially expressed genes with Ward clustering.

(C). UMAP of Basal Epithelial Cells (380 cells) across 4 mice. Black: Saline-treated mouse; red: IFN-α treated.

(D). UMAP of Basal Epithelial Cells as in C, points colored by detection of *Ace2*. Black: RNA positive, blue: RNA negative (6.6% *Ace2*$^+$, Bonferroni-adjusted p = 1.1E-10 for Basal Epithelial Cell expression versus all other cells).

(E). Schematic: wildtype (WT) and IFNγ-receptor knockout (IFNγR−/−) mice were infected intranasally with murine gamma-herpesvirus-68 (MHV68). Cells from whole lung were digested for scRNA-seq using Drop-seq (yielding 5,558 Epcam+ cells).

(F). Expression of *Ace2* by epithelial cell type, wild type (WT) mice. Statistical significance by Wilcoxon rank sum test with Bonferroni correction.

(G). Expression of *Ace2* among type II pneumocytes binned by infection status in WT mice. All pairwise comparisons non-significant (p > 0.05) by Wilcoxon rank sum test.

(H). Percent of *Ace2*$^+$ cells by infection condition (uninfected, bystander cells in MHV68-infected mouse, MHV68 RNA+ cells) and mouse genotype (WT, IFNγR −/−). Black bars: *Ace2*$^+$ positive cells; white bars: *Ace2*$^-$ cells.

(I). Schematic of RNA-Seq data from (Matos et al., 2019) of human lung explants (n = 5 donors) exposed to influenza A virus (IAV, H3N2) at 24 h post infection.

(J). Expression of *SFTPC* (surfactant protein C, a marker of type II pneumocytes) versus *ACE2* among mock-infected lung explants. Statistical significance assessed by Pearson's correlation, r = 0.93, p = 0.021. TPM: transcripts per million.

(K). *SFTPC* expression among matched donors following mock or IAV infection for 24 h. Statistical significance assessed by ratio paired t test, p = 0.86.

(L). *ACE2* expression among matched donors following mock or IAV infection for 24 h. Statistical significance assessed by ratio paired t test, p = 0.0054.

(M). Western blot of fully-differentiated air-liquid interface cultures from bronchial cells derived from 4 human donors with asthma. Cells from each donor were treated with 10 ng/mL IFNγ for 24 h, and compared to a matched untreated condition. ACE2 protein: AF933 (R&D). Fold changes quantified for IFNγ treated versus untreated for each patient donor following normalization to GAPDH.

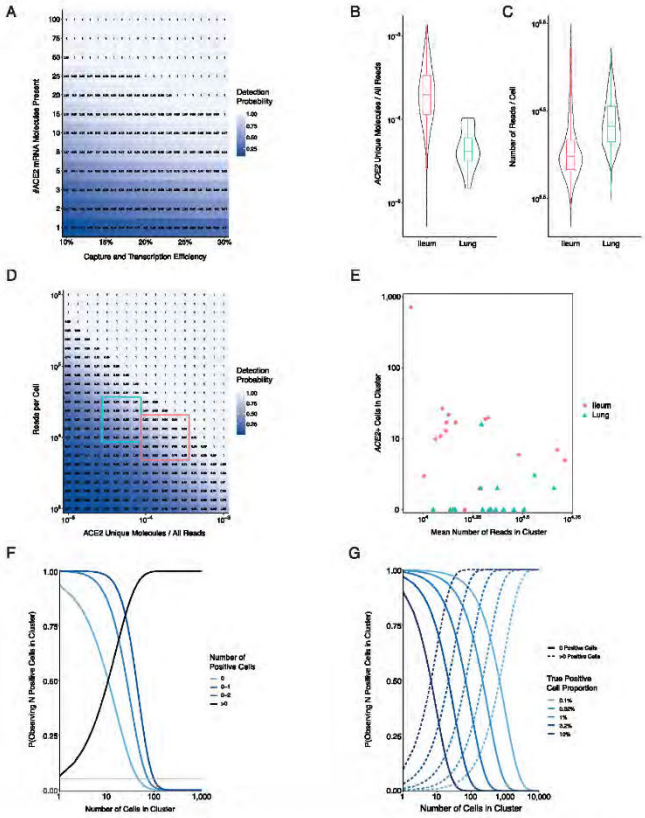


Figure S6. Power Calculations and Statistical Modeling of *ACE2* Capture and Dropout Related to STAR Methods

(A). Probability of capturing and transcribing at least 1 *ACE2* cDNA molecule, as a function of the capture/reverse transcription efficiency for a single molecule and the number of *ACE2* molecules expressed in an individual cell. Note that Drop-Seq provides a capture/transcription efficiency of approximately 11-13%, setting a floor on this parameter, and the experimental platforms used in this study are either equivalent or superior (Macosko et al., 2015).

(B). Distribution of *ACE2* fractional abundance within individual cells' cDNA libraries (i.e., *ACE2* UMIs / total number of reads), across non-human primate lung and ileum cell populations (see Figures 1 and 3). Mean fractional abundance among *ACE2*$^-$ lung cells = 5.0E-5; mean fractional abundance among *ACE2*$^+$ ileum cells = 2.7E-4.

(C). Distribution of the number of reads within non-human primate lung and ileum cell populations (see Figures 1 and 3). Mean ± SEM reads among all lung cells = 28,512 ± 344; *ACE2*$^+$ lung cells = 28,553 ± 2,988; all ileum cells = 14,864 ± 288; *ACE2*$^+$ ileum cells = 10,591 ± 441.

(D). Probability of observing at least one transcript for a gene of interest (e.g., *ACE2*) within an individual cell, as a function of sequencing depth and the gene's fractional abundance (i.e., *ACE2* reads / all reads) within the cell's cDNA library. Fractional abundance provides the probability that a single read corresponds to the gene of interest, and presented heatmap indicates the probability that at least one read in the total number of reads allocated to the cell (i.e., from 10^3 to 10^6) originates from the gene of interest. Mean read depths and *ACE2* fractional abundances for each tissue produce a 93.7% probability of detecting at least 1 *ACE2* read in ileum cells, and a 76.0% chance for lung cells. Outlined rectangles highlight the regimes where cells from lung (turquoise) and ileum (pink) samples typically lie.

(E). Number of *ACE2*$^-$ cells within each cluster, as a function of average read depth for all cells in that cluster. Number of cells detected as *ACE2*$^-$ is not correlated with read depth, even across relatively wide ranges of average read depths (Pearson's r = −0.31, n.s.).

(F). Probability of observing a particular number of cells positive for a gene of interest within a cluster, as a function of number of cells in the cluster. Probabilities were calculated under a negative binomial distribution with parameter p = 0.063 (the proportion of *ACE2*$^-$ cells among type II pneumocytes presented in Figure 1; STAR Methods). The horizontal gray line indicates the arbitrary cut-off value of p = 0.05.

(G). Given a population of cells with a known proportion that are positive for a gene of interest, probability of observing no positive cells (i.e., false negative identification of the cluster; solid lines) and probability of observing at least one positive cell as a function of cluster size.

Chapter 5: Final Discussion, Conclusion and Future research

5.1 Final Discussion and Conclusion

Tuberculosis (TB) continues to be one of the leading causes of death due to an infectious agent globally, with sub Saharan nations such as South Africa carrying a significant portion of the burden [297]. Unfortunately, many aspects of the human immune response to TB are yet to be elucidated [298], due to difficulty in accessing human infected tissue (lung in the case of pulmonary TB) and limited application of cutting edge technological advancements to this clinical material, as the regions most affected are often resource constrained [299]. This invariably limits the rate at which vaccines, therapeutics and diagnostics are developed, resulting in an endemic disease [300]. To address these challenges, we aimed to couple access to diseased human tissue with high throughput and locally available single cell sequencing technology to generate new and unbiased insights into TB disease. To this end, we managed to access human lung tissue from individuals undergoing corrective lung surgery due to TB associated pulmonary complications such as and not limited to fibrosis, cavitation, haemoptysis, nodules, bronchiectasis, mycetoma. Clinical records from many of the participants showed previous TB episodes with corresponding treatment. Therefore, we postulated that these individuals were being affected by post TB treatment pulmonary impairment. We then made use of scRNAseq, to profile this infected tissue as a means of unbiased/unsupervised classification of different cell types found in the tissue, probing their roles in the establishment of the TB pathology. We made use of a relatively new single cell sequencing platform known as seq-well (S^3), developed by the Shalek-Love laboratories at the Massachusetts Institute of Technology (MIT). This technology captured and barcoded transcripts of individual cells, generating a cellular atlas of

the TB lung [301]. This platform was comparable to 10X (Chromium Technologies) in terms of number of UMI and gene output whilst showing increased recovery of transcription factors, cytokines and cytokine receptors [255]. We explored the following aspects in TB research: (1) the cellular diversity of the resected lung tissue and the involvement of different cells in the immunopathology; (2) how transcriptional profiles in diseased lung tissue relate to published TB blood transcriptional and the association of these signatures with cell subsets; (3) the phenotypic diversity of lung resident CD4 T cells, as a step in understanding the different cellular subsets and their functional capacity. The latter aim was curtailed by the emergence of the SARS-CoV2 pandemic, which disrupted the core activities of the lung study. However, this prompted me to embark on an additional analysis (4) and use the single cell data generated prior to the pandemic to investigate the expression of the SARS-CoV-2 entry receptors in human lung. This opportunistic but highly informative study demonstrated how scRNAseq libraries are a valuable resource that can be useful in generating insights into other pulmonary diseases.

To our knowledge this was the first study to use scRNAseq to profile TB diseased lung tissue. For safety of personnel, this tissue had to be fully processed under strict biosafety level 3 (BSL3) conditions. This presented a considerable challenge, and involved many unsuccessful experiments not included here (~60% failure rate). In most cases, this appeared to be due to either the predominance of dead cells (% live cells ranged from 20-40% in the single cell suspensions) or a failure to obtain sequence products. Multiple wash steps with disinfectant were required at many stages of the processing of lung tissue, sample loading onto the seq-well arrays and subsequently processing steps, which might have negatively impacted the yield. Use of an automated, standardized system such as 10X can mitigate a myriad of experimental inconsistencies but the operation costs and rigid workflow were prohibitive. Trouble shooting

of this process continued throughout the thesis. As an illustration, in chapter 3, I was only able to present data from 1 of the 5 individuals from whom I sorted CD4 T-cells from processed lung tissue and PBMC.

Nevertheless, with over 3 years of effort I was able to obtain high quality sequences from the lung tissue of 13 participants. Unbiased analysis of the data identified 19 distinct cellular clusters, as shown in chapter 2. These cells included immune (myeloid, B and T cells, mast), stromal (fibroblasts), endothelial (vascular and lymphatic) together with epithelial cells. Franks and colleagues suggest that there are over 40 unique cell types throughout the respiratory tract [302], a number which we believe will increase with more scRNAseq studies [303]. It should be noted that our study managed to recover ~21,000 cells primarily from the lower respiratory tract. Our dataset yielded similar cell types (except basal cells) albeit in different proportions to a lower respiratory tract, scRNAseq dataset from 8 IPF and 8 healthy donors, in which macrophages and type pneumocytes were the most abundant cell types [46]. The researchers used droplet based scRNAseq, suggesting that cross platform datasets are comparable. The droplet vs microwell capture of cells requires further study, considering that cells in tissue come in different shapes and sizes. It has been shown that droplet based methods recover almost identical cellular populations in PBMC data, but verification is required in tissue [255]. Our dataset will be available to use online both as fastq files and processed data, allowing fellow scientists to further explore it. In future, distributing the sampling points throughout the entire length of the pulmonary system as well as increasing the cell yield by 5 or 10-fold can increase the likelihood of identifying new cell types and better approximately and exceed the reported 40 unique cell types.

The most striking observation we made was the differential regulation of fibroblast populations in TB diseased lung. The fibroblast 3 population was enriched for transcripts associated with immune cell recruitment (chemokines), parenchymal tissue consolidation (collagenases (COLs)) and cavitation (matrix metalloproteinases (MMPs) and their inhibitors(TIMPs)) [304]. We also observed that the fibroblast 3 cluster had transcripts which localized with TB granuloma, blood vessels and airways. The association between fibrosis and dysregulated vascular remodeling has been observed in idiopathic pulmonary fibrosis, suggesting a possible mechanism by which fibrotic foci can be established [305]. Previous studies have not adequately focused on the involvement of non-hematopoietic TB disease, yet these observations indicate they may play an important role in orchestrating cellular recruitment to the lung and pathological pulmonary remodeling. Our findings suggest that we need to consider stromal cells as a potential target in granuloma-directed TB therapy, which aim make the lesions less favorable to Mtb growth [306]. Our dataset also allowed us to explore possible enrichment of cell types driven by either TB or HIVTB. An interesting observation was that myeloid (monocytes and neutrophils) from HIVTB group expressed transcripts associated with heat shock proteins (physiological stress or death signature), a potential consequence of neutropenia and monocytopenia [307]. We also observed a depletion of alveolar type 1 (pneumocytes), alveolar macrophages and secretory cells in TB only participants, consistent with reports that individuals with TB only experience more severe parenchymal destruction and associated pulmonary impairment compared to HIVTB individuals [103].

The TB field is currently searching for biomarker signatures that can be used for evaluating the success of current as well as emerging therapies [308]. The ability to detect TB disease prior to the onset of clinical symptoms, and rapid the diagnosis of active TB, can significantly improve treatment outcomes. Studies evaluating these biosignatures have focused on blood as the

sampling site, due in part to the ease of access and standardized processing methods for extraction of PBMC. The literature is replete with studies that performed bulk RNA sequencing or microarray sequencing of PBMC from healthy, active TB and in some cases latent TB participants in an attempt to uncover biomarkers associated with TB disease [309]. We selected 5 of these signatures from the literature and super imposed them onto our scRNAseq library in chapter 2. We observed an enrichment in the neutrophil/inflammatory cell clusters of the signature proposed by Maertzdorf *et al.* in 2016 [163]. Neutrophils have been implicated in the inflammatory response to TB disease and consequent pulmonary impairment [119]. They are the most abundant immune subset in blood, and neutrophil derived signatures may be expected to overwhelm TB specific signals from low frequency cells. To correct for this effect, Singhania *et al*. (2018) used a modular approach that that allowed gene signatures from less dominant cellular subsets to emerge. This gene signature was the only one elevated in non-myeloid cell types such as fibroblast, endothelial, alveolar pneumocyte (type 1 and type 2) cells. Recent work on a 6 gene signature, derived from the original Zak signature used here, has shown that it is elevated in TB uninfected individuals responding to respiratory viral infections including influenza and non-pandemic corona viruses (Tom Scriba; unpublished data). This is likely to be a common feature of TB biomarker signatures that focus on the innate immune response of myeloid cells. However, the underlying lung destruction occurring in TB, which may be picked up by the Singhania *et al*. gene signature, can improve the discriminatory power compared to its more inflammatory counterparts [102]. These signatures will have to be compared to other diseases which cause lung injury to explore their discriminatory efficacy. Again, this highlights the potential value of unbiased profiling of TB diseased lung tissue in developing more targeted TB disease gene signatures.

We should explore fibroblast associated the marker genes and their role in the TB granuloma. Measurement of serum levels of *IDO* is already being explored as a potential simple biomarker of pulmonary TB [310]. Our data suggested, *TDO2*, a downstream partner of *IDO* in immune regulation, may also be associated with TB lung disease. Similarly, peptidase inhibitor 15 (*PI15*), was highly upregulated in the TB associated cell subsets . This molecule has not been widely studied, but recently published data suggests it shows promise as a potential blood diagnostic biomarker of cholangiocarcinoma [311]. The fact that it is upregulated within these tumors and in the blood of the same subjects, making it a good option for blood based signatures.

In chapter 3, we used scRNAseq to explore the phenotypic diversity of tissue resident cd4 t cells, as t cells that have been shown to localize in tissue to elicit a robust adaptive response to a secondary TB infection. We purified populations of CD4+CD69+ and CD4+CD69- T cells using fluorescence activated cell sorting (FACS) from an HIV positive participant. A single cell library of these populations resolved into 5 distinct clusters including: one naïve, two regulatory, an activated and one cytotoxic. The cytotoxic (*PRF1, GNLY*) and regulatory clusters (*TACTILE, PRDM1*) have been previously described in tissue resident cells [312]. The activated t cells had a similar phenotype to negative immune regulatory, *CD300A* t cells that have been reported in PBMC from HIV infected patients [254]. Taken together, scRNAseq of these cells uncovered the different cell types comprising the Trm's, however more cells are required from more participants to ensure that the full spectrum of these cells is captured. We did no recover usable cells from most of the participants (4/5) likely due to low starting T cell numbers that were observed with most lung samples. The FACS machine purifies multiple cell types simultaneously for heterogenous sample but it should be noted for populations of interest 1% or less of the total cell count, the sorting efficiency significantly decreases leading

to elevated cell loses [313]. We could not address this as the lung study was suspended due to the instituted COVID-19 lockdown regulations.

In chapter 4, we demonstrated the importance of generating single cell databases of human tissue in the context of other pulmonary diseases. Specifically, we managed to show that our dataset could be extended to the COVID-19 pandemic by using transcript capture technology to predict which cell type would make a suitable target for the SARS-CoV-2 virus based on preliminary studies. Angiotensin-converting enzyme 2 (*ACE2*) had been reported as a critical receptor through which the virus can infect cells [314], together with the Type 2 transmembrane serine protease (*TMPRSS2*), required to cleave the viral spike protein to facilitate cell entry [315]. We managed to show expression of *ACE2* and *TMPRSS2* transcripts in type 1 pneumocytes, type 2 pneumocytes and ciliated cells. Interestingly, the HIVTB group showed an upregulation of these markers in comparison to the TB only group, suggesting that HIV positive individuals could be more susceptible to SARS-CoV-2 infection. However, this has not been ascertained with clinical data and the expression of these viral target receptors is dependent on sex and age together with other ongoing inflammation inducing events, thus more research is required to demonstrate its clinical usefulness [316-318]. However, this section served as a step towards understanding the dynamics of SARS-CoV-2 infection and can be extended to the design and development of potential therapeutic interventions.

5.2 Future Research

5.2.1 Whilst we managed to profile 19 cell subsets from human whole lung tissue, there is a need to increase the number of cells. The target cell number is estimated to be around 100,000 [303], which will allow us to effectively profile more subsets and uncover cellular states are

affected by TB and HIVTB disease. Identifying the fibroblast 3 subset and localization of marker genes within the TB granuloma require a follow up study to confirm their applicability.

5.2.2 The section on applicability of the blood derived TB biosignatures showed to a moderate degree how the biosignatures mapped onto the scRNAseq. It has been well characterized that biomarker development studies are negatively impacted by the high rate of false positives and a poor statistical/analytical frameworks, coupled to shared immune responses to other bacterial, viral infections or inflammation inducing antigens [107]. We propose a bottom-up approach, in which we benchmark lung tissue cavitation and fibrosis as gross indications of active TB disease when coupled to chest-xrays, gene xpert, culture and smear microscopy. This will be proceeded by a cavitation and fibrosis signature in the blood with easily adaptable technologies such as PCR.

5.2.3 This study managed to demonstrate the phenotypic diversity in the tissue resident CD4 T cells derived from an HIVTB infected individuals. However, more data is required from participants with different disease statuses (TB only and healthy controls) to increase the resolution of the T cell subsets in the tissue resident fraction of the lung tissue. More consideration is required for the cell types such as MAIT and γδ T cells that are potentially sorted with the cd4 tissue resident t cell fraction [319,320].

5.2.4 The penultimate section of this study (Chapter 4) demonstrated the applicability of scRNAseq data derived from TB research in other diseases such as COVID-19. TB research needs such data in increasing quantities to improve the generation/validation of pertinent questions by taking advantage of the unparalleled depth and breadth of the cell/gene readouts. Due to the cost and complexity involved in this technology, we recommend that a bulk RNA

sequencing databases of the TB human lung tissue (showing the full spectrum of disease) be created, similar to those that exist in cancer research [168]. These can also be useful in establishing biomarker signatures for lung damage due to fibrosis and cavitation of the lung tissue.

I would also like to state that l embarked on this PhD journey in 2016, under the supervision of Dr Frederick Balagdde in the bioengineering department. However, due to unforeseen circumstances, he resigned from his position. I had already started many projects in the lab and had to join Dr Leslie's (my current supervisor). I managed to publish one paper from my previous lab during the course of my PhD, which was a proof of principle study, demonstrating the work flow of Light Forge, a real-time PCR based microfluidic device for detecting drug resistance linked mutations in the *rpoB* locus of *Mtb* , with comparable and in some cases superior performance compared to commercial machines such as the Light Cycler 96 (roche diagnostics) [321]. I have attached the publication at the end of this chapter in the annex section.

Annex 2

Publication 2 : Light Forge ; A Microfluidic DNA Melting-based Tuberculosis Test

ARTICLE

Light Forge: A Microfluidic DNA Melting-based Tuberculosis Test

Ian M. Mbano,[a*,†] Tawanda Mandizvo,[a*,†] Jerome Rogich,[b] Tafara T.R. Kunota,[a] Jared S. Mackenzie,[a] Manormoney Pillay,[c] and Frederick K. Balagaddé[a]

Background: There is a well-documented lack of rapid, low-cost tuberculosis (TB) drug resistance diagnostics in low-income settings across the globe. It is these areas that are plagued with a disproportionately high disease burden and in greatest need of these diagnostics.

Methods: In this study, we compared the performance of Light Forge, a microfluidic high-resolution melting analysis (HRMA) prototype for rapid low-cost detection of TB drug resistance with a commercial HRMA device, a predictive "nearest-neighbor" thermodynamic model, DNA sequencing, and phenotypic drug susceptibility testing (DST). The initial development and assessment of the Light Forge assay was performed with 7 phenotypically drug resistant strains of *Mycobacterium tuberculosis (M.tb)* that had their *rpoB* gene subsequently sequenced to confirm resistance to Rifampin. These isolates of *M.tb* were then compared against a drug-susceptible standard, *H37Rv*. Seven strains of *M.tb* were isolated from clinical specimens and individually analyzed to characterize the unique melting profile of each strain.

Results: Light Forge was able to detect drug-resistance linked mutations with 100% concordance to the sequencing, phenotypic DST and the "nearest neighbor" thermodynamic model. Researchers were then blinded to the resistance profile of the seven *M.tb* strains. In this experiment, Light Forge correctly classified 7 out of 9 strains as either drug resistant or drug susceptible.

Conclusions: Light Forge represents a promising prototype for a fast, low-cost diagnostic alternative for detection of drug resistant strains of TB in resource constrained settings.

[a]Africa Health Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu Natal, Durban, South Africa; [b]University of Massachusetts Medical School, Worcester, MA; [c]Medical Microbiology, School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa.

*Address correspondence to: I.M.M. at Africa Health Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu Natal, Durban, South Africa. Fax: +27312604302; e-mail ian.mbano@ahri.org. T.M. at Africa Health Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu Natal, Durban, South Africa. E-mail tawanda.mandizvo@ahri.org.

† These authors contributed equally to this work.

Received July 8, 2019; accepted October 15, 2019.

DOI: 10.1093/jalm/jfaa019

IMPACT STATEMENT

The information presented in this study is primarily positioned to benefit TB-infected individuals from resource limited regions such as Sub-Saharan Africa and South East Asia where affordable and accessible diagnostics are required. The evidence presented in this manuscript illustrates that combining DNA-melting analysis with microfluidics can be a foundational formulation for a diagnostic that will be cheap, parallel, and accessible in various clinical settings.

INTRODUCTION

Tuberculosis (TB) is a deadly infectious disease with 1.6 million deaths reported in 2017 *(1)*. The highest disease burden is seen in Africa and South East Asia, low income regions, often with poor healthcare delivery *(2)*. The inability to control infectious diseases adequately in these areas is rooted in poor diagnosis and treatment. In TB, this leads to increased infectivity, transmission, morbidity, and mortality. Further compounding this problem is the significant rise in the number of drug-resistant strains of *Mycobacterium tuberculosis (M.tb)*, which are associated with higher morbidity and mortality rates *(3)*. It would be possible to provide more efficient healthcare delivery by increasing access to more economical diagnostic devices. If *M.tb* infection and drug resistance profiles can be detected at a significantly lower cost to public health systems, treatment would be initiated earlier, with a substantial attainable decrease in disease incidence *(4)*. Unfortunately, despite increased academic and commercial interest in point-of-care diagnostics, few commercially available devices have managed to effectively deliver to this underserved demographic.

Creating diagnostics within this context requires an adept appreciation of the unique challenges and limitations in the development, production, and marketing of a diagnostic test for the developing world *(5)*. Notably, the most important considerations are device affordability and turnaround time *(6)*. Microfluidic technology can reduce the cost of diagnosis by precisely manipulating minute fluid volumes in parallel, thereby reducing the overall consumption of reagents whilst increasing diagnostic throughput. Engineering and refining these fine networks of micro-plumbing allows integration of many functional components onto a single device *(7)*. A sample can be partitioned into several independent fluid circuits in an efficient manner, allowing multiple assays to be carried out in parallel with minimal end-user intervention. Another cost saving strategy for TB diagnostic devices is the utilization of high-resolution melting analysis (HRMA), a post polymerase chain reaction (PCR) method used to detect single nucleotide polymorphisms (SNPs) *(8)*. It has the advantage of being a single step as well as a closed tube assay that allows for a rapid and reliable examination of PCR product. This method has been used to detect drug-resistant *M.tb*, with performance metrics comparable to the "gold standard" *M.tb* phenotypic drug susceptibility testing *(9, 10)*. HRMA has been reported to cost USD $0.30 per reaction, but performing it at volumes consistent with microfluidics will reduce reagent costs by approximately 1,000 *(11)*. We leveraged the efficiency and simplicity afforded by a microfluidics platform with the simple, linear workflow of HRMA to create Light Forge, a functional and foundational blueprint for a low-cost TB diagnostic for resource-limited settings.

In this study, we address the following research questions: (1) Are the performance characteristics

of Light Forge comparable to a commercial HRMA device, phenotypic drug susceptibility testing, a predictive nearest-neighbor thermodynamic model, and Sanger sequencing? (2) Can Light Forge provide similar specificity to traditional Sanger sequencing when blinded samples are analyzed? To answer these questions, we obtained seven rifampicin-resistant strains with known mutations in the rifampicin (RIF) resistance determining region (RRDR) *rpoB*. Custom primers were designed and the amplicons were melted to quantify the melting temperature compared to *H37Rv*. RIF resistance was selected as an appropriate target as sequence aberrations in and around the 81 bp fragment of the RRDR accounting for 95% of TB drug resistant cases *(12, 13)*. A similar approach was adopted for the subsequent blinded phase of the study.

MATERIALS AND METHODS

Mycobacteria tuberculosis Isolates

Eight DNA samples were used to develop the assay for subsequent experiments. These were obtained from 7 rifampicin-resistant laboratory strains (*Kzn605, R35, R271, Tkk-01-0039, Tkk-01-0043, Tkk-01-0050, Tkk-01-0062*), and *H37Rv*, used as a drug-susceptible standard throughout the study. The strains were sourced from Medical Microbiology, School of Laboratory Medicine and Medical Sciences at the Nelson R. Mandela School of Medicine (Durban, South Africa) and Dr. Alex Pym's laboratory at the African Health Research Institute (AHRI) (Durban, South Africa). The phenotypic drug-susceptibility profile of all the test strains confirmed rifampicin resistance.

To further validate the performance of the assay, genomic DNA from 9 clinical *M. tuberculosis* isolates (*Tkk-01-0011, Tkk-01-0030, Tkk-01-0032, Tkk-01-0061, Tkk-01-0078, Tkk-03-0082, Tkk-04-0006, Tkk-04-0030, Tkk-04-0048*) were also sourced from Dr. Alex Pym's laboratory at AHRI. The drug susceptibility profile of each of these strains was revealed after the Light Forge blinded study was concluded.

DNA Extraction

DNA was extracted from heat-killed cultures using the hexadecyltrimethylammonium bromide (CTAB) method *(14)*.

Sanger Sequencing of the *rpoB* Gene Region

PCR amplicons of the *rpoB* gene for each strain were sequenced by Inqaba Biotech Industries, using primer sets identical to the primers in the high-resolution melting (HRMA) assay (Table 1).

Real-Time PCR and High-Resolution Melting Analysis (HRMA) with Light Cycler®96

Polymerase chain reaction (PCR) was performed by first preparing a master mix through adding 25 µL of 2X Xtreme™ Buffer (Novagen), 10 µL of dNTPs (Novagen,), 5 µl of LC Green (Idaho Technology Inc.), 0.3 µM of each Forward and Reverse primer, 1 µL of KOD Xtreme™ Hot Start DNA polymerase (Novagen, Toyobo) and 5 µL of template DNA containing 200 ng of the DNA in a final volume of 50 µL. PCR was performed using the Light Cycler®96 (Roche Diagnostics, Switzerland). The initial denaturation temperature was 95 °C for 300 s, followed by 35 cycles of 95 °C for 10 s, 60 °C for 10 s, 72 °C for 10 s (fluorescence readout step). A nontemplate control (NTC) was included in all experiments, in which PCR Grade water (Life Technologies) was substituted for the DNA template.

For the HRMA, the following profile was used: 95 °C for 60 s, 40 °C for 60 s, 65 °C for 1 s then at a 0.07 °C/s ramp rate, acquiring 15 readings every degree until 97 °C. Difference plots were generated using *H37Rv* as the baseline signal. The test readings from the samples were then normalized to this standard. The total reaction time was 70 min.

Table 1. Primer sequences used for PCR amplification, Sanger sequencing and subsequent HRMA of the *rpoB* gene's region known to determine rifampicin resistance *(15)*.

Primer	Sequence	Annealing temperature(°C)	Product size(bp)	Nucleotide position	Accession number
rpoB-F	CGCGATCAAGGAGTTCTTC	65	118	2339 to 2357	L27989.1
rpob-R	TGACAGACCGCCGGGCCC			2456 to 2439	

Light Forge Microfluidic Chip Design and Fabrication

The Light Forge PCR microfluidic chip was fabricated out of the silicone elastomer polydimethylsiloxane (PDMS) (General Electric RTV 615) using multi-layer soft lithography, as described previously *(16)*. Up to 20 independent PCR reactions can run in parallel on each chip (Fig. S1).

Real-Time PCR and High-Resolution Melting with Light Forge

The real-time PCR and HRMA steps on the chip were captured using the Light Forge software, which was developed in house by Dr. Frederick Balagaddé. This software uses a feedback system that allows acquisition of fluorescence signals at easily programmable temperatures, whilst displaying the reaction progress.

The PCR master mix was prepared using 30 µL of 2X Xtreme^{TM} Buffer (Novagen), 15 µL of dNTPs (Novagen), 9 µL of LC Green (Idaho Technology Inc.), 0.46 µM of each primer (Life Technologies), 3 µL of KOD Xtreme^{TM} Hot Start DNA polymerase (Novagen), 0.08%(v/v) 1% Tween 20, and 3 µL of template DNA(~6ng) to make a total volume of 75 µL. For the 20-reactor chip, 8 reactors were used for *H37Rv*, 8 for the test strains (*R35, Kzn 605*). The final 4 reactors contained the nontemplate controls.

The thermal cycling was performed using a G-STORM GS1 (Somerten) thermocycler modified to house the microfluidic device. The fluorescence signal was acquired in real-time using an Olympus MVX10 (New York) Macro zoom microscope (Fig. S2). The thermal profile was 99 °C for 8 minutes, followed by 35 cycles of 99 °C for 65 s, 60 °C for 115 s, 74.5 °C for 130 s. The HRMA was performed by increasing the temperature from 75 °C to 94 °C at a ramp rate of 0.5 °C/s with 0.25 °C increments for each step. This resulted in a total reaction time of 180 min (Fig. S3).

Nearest-Neighbor (NN) Thermodynamic Model

The data from our experiments were compared to the theoretical relative melting temperature differences predicted by the mathematical nearest-neighbor (NN) thermodynamic model *(17)*. This model predicts the melting temperature (T_M) of a DNA strand based on the cumulative standard enthalpies (ΔH^o) and standard entropies (ΔS^o) of neighboring duplex base pairs as well as the total oligonucleotide strand concentration (C_t) as shown by Equation 1 below.

$$T_M = \frac{\sum_{i=1}^{i=n-1} \Delta H^o}{\sum_{i=1}^{i=n-1} \Delta S^o + \ln\left(\frac{C_t}{4}\right)} - 273.15 \qquad \text{(Eq. 1)}$$

Using the sequencing information from Sanger sequencing, the (ΔT_M) between the reference strain (*H37Rv*) and the test strains used for the initial evaluation were computed. This model was also used to compute the expected (ΔT_M) of the blinded samples. The computed prediction error of the NN model was ±1.2 ^{0}C *(18)*. The SNPs observed in resistant strains of *M.tb* result in

temperature deviations ranging from (<0.1 to 1.4 ^{0}C) *(19)*. Thus, the model cannot be used for predicting exact differences but remains useful in predicting whether the SNPs lead to a positive or negative ΔT_M.

Statistical Analysis

A two-way *t*-test was used to evaluate the significance of the differences, where a *P* value of 0.05 or less was considered statistically significant (GraphPad Prism 7).

RESULTS

rpoB Gene Mutations in Clinical *M.tb* Strains

Seven strains with rifampicin resistance and a reference *H37Rv* wild-type strain (rifampicin sensitive) were selected for the initial evaluation of the Light Forge. The 118 bp *rpoB* gene target of each of the isolates was Sanger-sequenced prior to HRMA. All the 7 sequence variants demonstrated SNPs associated with resistance to rifampicin. A total of 5 sequence SNPs present in 7 strains were identified, including *Kzn 605* (533 T → **C** and 516 A → **G**), *R35* (533 T → **C**), *R271* (531 C → **T**), *Tkk-01-0062* (516 A → **T**), *Tkk-01-0050* (526 C → **T**), *Tkk-01-0043* (531 C → **T**), and *Tkk-01-0039* (526 C → **T**).

Real-Time PCR and HRMA for Drug Susceptibility Testing on Test Strains

The *rpoB* PCRs for 7 clinical strains were run on the Light Cycler®96 and were compared to the Light Forge system (using the microfluidic chip shown in Fig. 1). Fig. 2 illustrates the fluorescence imaging of the Light Forge reactors at the 1st (A) and 25th (B) thermal cycles during real-time PCR. The A panel in Fig. 2 is a fluorescent image at 60 °C at the first PCR cycle, while the B panel is the fluorescent image at 60 °C after 25 PCR cycles. Each color-coded chamber contains ~1.5 nL volume of PCR master mix. This is a representative layout using a clinical mutant isolate (*Tkk-01-0050*) and the wild type reference control *H37Rv* used in the subsequent experiments. The test strain was amplified in 8 out of 20 reactors (red border), the reference wild type strain (*H37Rv*) was amplified in 8 out of 20 reactors (blue border), while the non-template controls were amplified in 4 of 20 reactors (yellow border). At the end of the amplification, the Light Forge analysis was immediately initiated to melt the amplicons and detect mutations. Mutations were identified as melting temperature (T_M) deviations relative to *H37Rv*. Each individual line indicates a melt-curve profile for an individual reactor.

The real-time PCR and HRMA are graphically illustrated in Fig. 3 a and b respectively. The HRMA profiles of *H37Rv* were used as the standard to which the 7 clinical strains were then compared. Mutant melting curves could be distinguished from the wild type melting curve in the normalized graphs (Fig. 3b) but were best differentiated in the negative 1st derivative plot (*-dRFU/dT*) shown in Fig. 3c. The average melting temperature was computed for both the test strain and H37Rv as shown in Fig. 3d. There was 100% concordance of the results from Light Cycler®96 and Light Forge.

Comparison of Light Forge HRMA with Sanger Sequencing, Roche Light Cycler®96 and the Nearest-Neighbor (NN) Model

This section refers to Fig. 4. The sequencing data for *Kzn 605* identified 2 positive class two mutations 533 T → **C** and 516 A → **G** within the *rpoB* region. The Light Forge system identified these *Kzn 605* SNPs as having a melting temperature difference (ΔT_M) of 0.95 ± 0.06 °C. The Light Cycler®96 detected a (ΔT_M) of (0.61 ± 0.02 °C). The NN model predicted a positive (ΔT_M) of +1.88 °C.

Sequencing the *R35* isolate revealed the positive class one mutation 533 T → **C** within the *rpoB* region. Consistently, Light Forge detected a (ΔT_M) of (0.30 ± 0.11 °C). For the same strain, Light

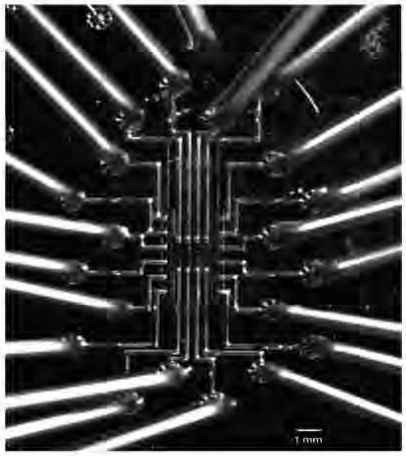


Fig. 1. The Light Forge microfluidic chip. Optical micrograph of the Light Forge microfluidic chip/cartridge for real-time PCR and high-resolution melting analysis of amplified samples. Scale bar, 1 mm. The chip/cartridge is made up of an elastomeric transparent polymer called polydimethylsiloxane and has 20 reactors, each linked to an independent input port. These reactors are 500 μm long, 200 μm wide, and 150 μm high, translating into a volume of 1.5 nL. The PCR master mix was loaded at 5 psi until the PCR chambers were full.

Cycler®96 detected a (ΔT_M) of (0.22 ± 0.02 °C). The NN model predicted a melting temperature of + 0.67 °C.

Sequencing the *R271* isolate revealed the presence of a negative class one SNP 531 C → **T** within the *rpoB* region. Light Forge detected a (ΔT_M) of −0.29 ± 0.09 °C. A (ΔT_M) of −0.23 ± 0.03 °C was detected for the same strain using the Light Cycler®96. Consistent with the two systems, the NN model detected a (ΔT_M) −0.23 °C.

The Light Forge system detected a (ΔT_M) of (0.22 ± 0.05 °C) when the *Tkk-01-0062* isolate was compared to the reference. Sequencing high lighted the presence of (1 positive) or (a positive) class four SNP 516 A → **T**. The Light Cycler®96 could not detect this SNP, shown by a (ΔT_M) of 0.013 ± 0.02 °C. A (ΔT_M) of + 0.20 °C was predicted by the NN model.

When the *Tkk-01-0050* isolate was run in the Light Cycler®96 system, a (ΔT_M) of −0.28 ± 0.02 °C was detected. This observation was consistent with sequencing findings, which denoted the presence of 1 negative energy class one SNP 526 C → **T**. Light Forge detected a (ΔT_M) of −0.52 ± 0.08 °C. The NN model predicted a (ΔT_M) of −0.84 °C, comparable with both the Light Cycler®96 and the Light Forge systems.

The sequencing data for the *Tkk-01-0043* isolate identified one negative energy class one SNP 531 C → **T**. Light Forge system detected a (ΔT_M) of −0.48 ± 0.09 °C, whereas the Light Cycler®96 system detected −0.24 ± 0.02 °C as the (ΔT_M). The NN model predicted a melting temperature difference of −0.23 °C.

Sanger sequencing the *Tkk-01-0039* isolate identified 1 negative energy class one SNP 526 C → **T**.

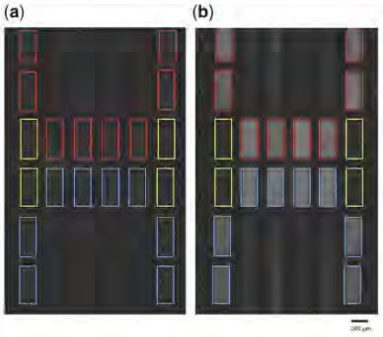


Fig. 2. DNA quantification and high-resolution melt analysis in Light Forge PCR reactors. The Light Forge chip/cartridge imaging area is shown in greater detail, demonstrating the visual accumulation of fluorescence signal (within each of the 20 reactors, linked to independent input ports) as the Light Forge system performs real-time PCR and High-Resolution Melting Analysis. An increased fluorescence signal intensity is demonstrated with an increase in PCR cycle number. Fluorescence images of Light Forge PCR reactors of the 1st (a) and 25th (b) thermal cycles during real-time PCR. The reactors in the image have been demarcated with color-coded boundaries depending on the genomic DNA amplified: drug-resistant mutants [*Tkk-01-0050* (526 C→T), red]; drug-susceptible strains (*H37Rv*, blue); and nontemplate controls (yellow). The amount of double-stranded DNA in each reactor corresponds to the level of fluorescence of LC Green—a reporter double-stranded DNA intercalating dye in the master mix. LC Green binds specifically to double-stranded DNA and fluoresces brightly when it binds. Unless specified otherwise, all images were taken during the 60°C step of each thermal cycle. We were detecting fluorescence—in real-time via a custom program that runs on a temperature feedback loop—at 470-520 nm, the emission wavelength range of the reporter double stranded DNA saturating dye LC Green.

Consistently, the Light Forge system detected a (ΔT_M) of -0.37 ± 0.16°C when this isolate was compared to the reference, whereas the Light Cycler®96 system detected a (ΔT_M) of -0.26 ± 0.02 °C. The NN model predicted a melting temperature difference of -0.84°C, which agreed with both the Light Cycler®96 and the Light Forge systems.

Validation of Light Forge with Blinded Samples

For further validation of the Light Forge assay, 9 blinded samples were analyzed. The (ΔT_M) melt temperature of the blinded samples was compared to *H37Rv*. Four (44.4%) of the 9 blinded isolates showed melting curve profiles that were synonymous to that of *H37Rv* (Fig. 5). However, 5 (55.5%) of the 9 blinded isolates were distinguishable from the *H37Rv* melting curve profile. In comparison to the culture-based RIF-susceptibility test (DST), Light Forge correctly identified 7 out of 9 of the isolates with 71.43% sensitivity and 100% specificity. The discrepancy was noted for *Tkk-04-0048*, which had a drug-susceptible profile on the Light Forge system but showed resistance in culture, confirmed with sequencing by the presence of a double mutation in the drug-resistance determining region of the *rpoB*. The isolate *Tkk-01-0030* also showed discrepancy with a drug-susceptible profile on the Light Forge but resistance in culture,

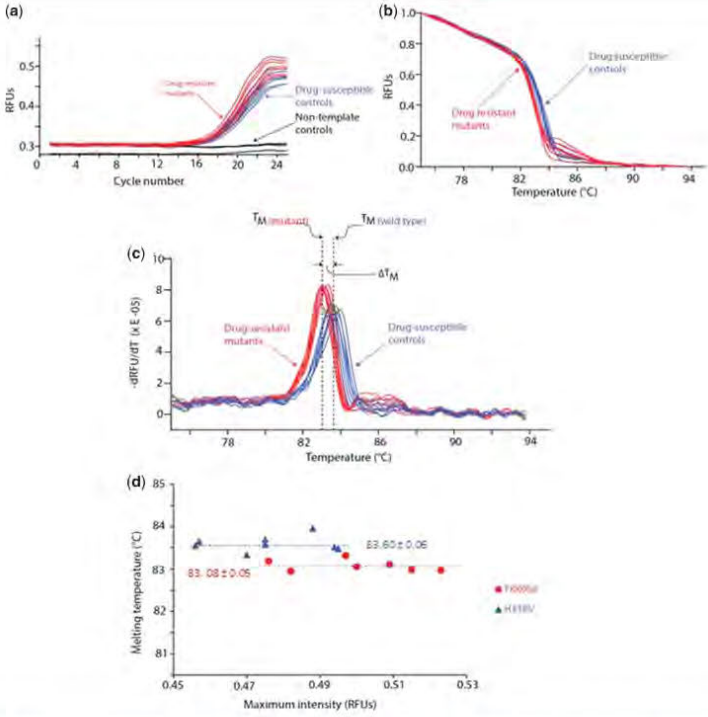


Fig. 3. The real-time PCR and HRMA data from the Light Forge system testing the *rpoB* gene of the [*Tkk-01-0050* (526 C→T)] clinical strain. (a) Real-time PCR amplification curves for the *rpoB* region. These data were all acquired simultaneously using the Light Forge microfluidic chip. For each reactor, a single data point was acquired during the 60 °C step. The relative fluorescence units (RFUs) corresponding to the concentration of double-stranded DNA was plotted against the PCR cycle number. Eight real-time amplification curves for drug-resistant [*Tkk-01-0050* (526 C→T)] mutant strain are shown in red, 8 curves for the drug-susceptible controls strains (*H37Rv*) in blue and four curves of the no template control reactors in black. Both drug-susceptible control and drug-resistant mutant strain reactors increased by more than 0.15 RFUs to over 0.45 RFUs, whereas the amplification signal for the nontemplate controls remained relatively suppressed. (b) Normalized high-resolution melt (HRM) curves for the *rpoB* PCR amplicons of the drug-resistant mutant strain [*Tkk-01-0050* (526C→T), red] and the drug-susceptible control strain (*H37Rv*, blue). (c) Negative first derivatives plots revealing the melt peaks. The melting temperature (T_M) of each strain was determined by taking the mean of the melt peak of every replicate for each individual strain. The average melting temperature for the drug-susceptible strain $T_{M(wild\ type)}$ was 83.60 + 0.06 °C and that of the drug-resistant mutant $T_{M(mutant)}$ was 83.08 + 0.05 °C. The relative difference of the melt peaks of these strains (ΔT_M) provides important clues to the nature of the mutation in the drug-resistant genome. (d) Peak intensity as a function of melting temperature. The average melting temperatures for the drug-resistant mutant (*Tkk-01-0050*, red circles) and drug-susceptible control (*H37Rv*, blue triangles) were 83.60 + 0.06 °C and 83.08 + 0.05 °C respectively. The melting temperatures were independent of the maximum fluorescence intensity (double-stranded DNA concentration) and thus independent of the amplification efficiency in each reactor.

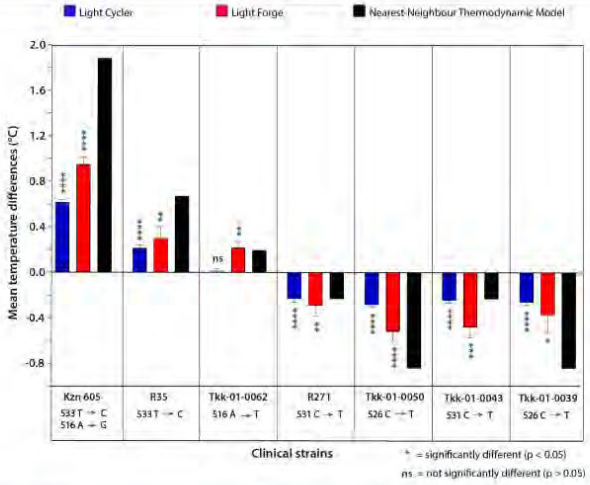


Fig. 4. Comparison of melting temperature differences from Light Forge and Roche Light Cycler®96 HRMA systems alongside the NN thermodynamic model T_M predictions. Relative melting temperature differences of the *rpoB* amplicon among 7 RIF-resistant strains and the reference RIF-susceptible strain *H37Rv* were plotted using data generated from Light Forge (red bars) and Light Cycler®96 (blue bars). The black bars represent the theoretical relative melting temperature differences predicted by the nearest neighbor (NN) thermodynamic model. The Light Forge data is consistent with Light Cycler®96 data as well as the theoretical model predictions. However, Light Forge successfully detected a low-energy class 4 mutation that eluded detection by Light Cycler®96 despite the nearest neighbor (NN) thermodynamic model having predicted that the mutation is detectable.

confirmed with sequencing by the presence of an (A to T) change.

DISCUSSION

Detecting drug-resistant strains of TB remains vital for the timely clinical management of the disease, with potential to significantly reduce transmission. This proof of concept (POC) study details the design, development and preliminary performance evaluation of Light Forge, a microfluidic device to detect RIF resistance linked mutations in *M.tb*.

In the initial phase of the study, 7 RIF-resistant strains of *M.tb* were used to bench mark the performance of Light Forge to the Roche Light Cycler®96. As shown in Fig. 4, Light Forge showed 100% concordance with Sanger sequencing compared to the 86% demonstrated by the Light Cycler®96. The same trend was observed when Light Forge was compared to the nearest neighbor (NN) thermodynamic model. An interesting observation from the panel was that the commercial device did not detect the presence of the mutation in the *Tkk-01-0062* strain. This strain harbored a class 4 transversion (516 A → **T**), which is difficult to detect as it results in a melting temperature difference of less than 0.4 °C *(15)*. This was consistent with a study that showed that HRMA had diminished ability to detect trans

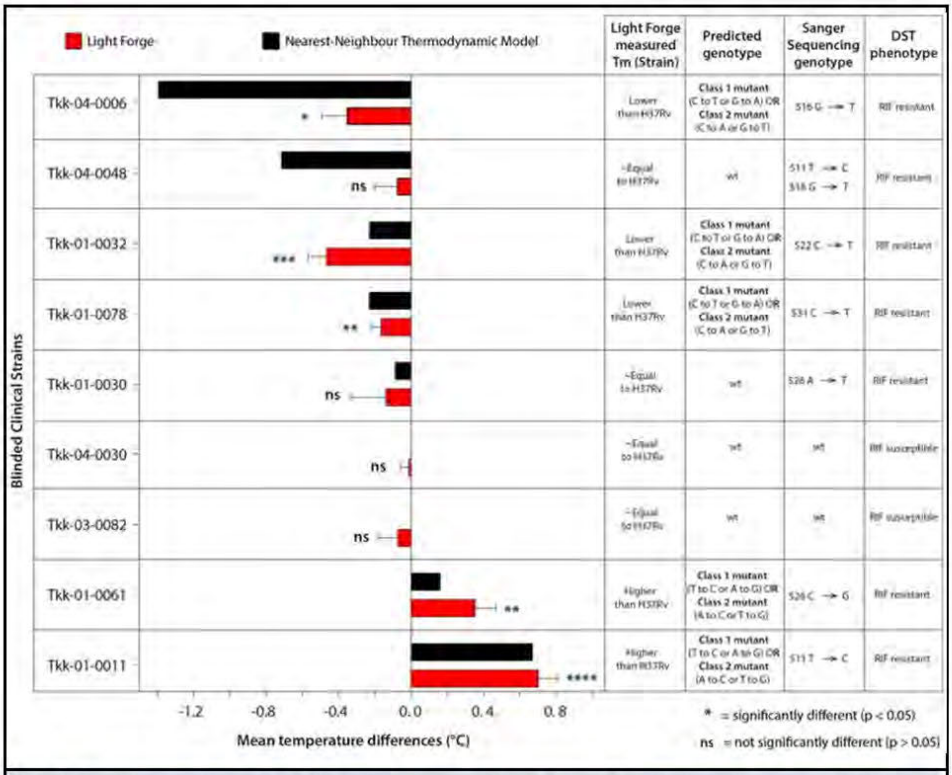


Fig. 5. Precision and reproducibility comparisons of Light Forge phenotypic resistance prediction on blinded TB clinical isolates using HRMA. These results were reproducible by two independent Light Forge experiment repeats with very small standard deviations.

versions when the reaction volume increased 5-fold *(20)*. The higher sensitivity is achieved through rapid heat transfer in microfluidic-based PCR due to the small reaction mass and the higher surface to volume ratio of the small reactor, leading to a more uniform temperature distribution *(21)*.

To circumvent subjectivity of the study *(22)*, the second phase of the experiments was performed with the scientists blinded to the resistance profile for all nine clinical isolates. Light Forge detected the presence of mutations in 7/9 isolates as shown in Fig. 5 (5 RIF-R and 3 RIF-S), when compared to sequencing and drug susceptibility profiles. Two RIF-resistant strains, *Tkk-04-0048* and *Tkk-01-0030*, were erroneously detected as wild type. *Tkk-04-0048* contained double mutations at positions (511 T → **C** and 516 G → **T**). These mutations were undetectable as the increase in melting temperature due to the mutation at position 511 was offset by a decrease in melting temperature

by the mutation at position 516. This observation is not unique to this study, as it has been reported that co-occurrence of class 3 and 4 mutations within the same amplicon leads to minimal melting temperature deviation *(23)*. *Tkk-01-0030* contained a single mutation (526 A → **T**). This specific mutation was located between two cytosine (C) nucleotides as shown in supplemental Fig. S4. The mutation appears to lead to minimal deviation in the melting temperature in comparison to H37Rv, consistent with the observations that (A/T) transversion have miniscule impact on the melting profile of the amplicons *(24–26)*. As stated previously, we speculate that reducing the volume of the microfluidic reactors by between 5- to 20-fold could potentially enhance the resolution power of the HRMA assay.

Whilst Light Forge has the potential to contribute significantly to improving health care delivery systems, particularly those in low income settings, it has limitations arising from its design based on molecular testing of TB and its reliance on microfluidic technology. First, HRMA can detect melting temperature aberrations that can be misinterpreted as the presence of mutations, but which do not confer any phenotypic drug resistance (including silent mutations). Thus, the assay can erroneously predict that a strain with the mutation (516 G → **C**) is drug resistant, contrary to its drug-sensitive profile *(27)*. A similar occurrence was reported when a silent mutation at codon 514 of the *rpoB* gene was misclassified as drug resistant using a commercial genotyping kit *(28)*. This is the major reason why sequencing and DST remain more precise tools for asserting clinical resistance of strains *(29)*. Nevertheless, with proper execution, Light Forge could be a useful tool for screening TB patients.

Light Forge was designed using principles adopted from microfluidic large-scale integration (MLSI), which allow for several hundred to thousands of reactors on a single device *(30)*. Some limitations of this study are that few isolates were used, and that only DNA from pure clinical cultures but no primary specimens such as blood or sputum were used. In addition, the assay would have benefited from a comparison with the well-established GeneXpert MTB/RIF Assay. However, these limitations do not detract from the value and significance of the POC Light Forge findings. To our knowledge, the Light Forge chip represents one of the few iterations that were performed at nanoliter scale. There are several modifications that can be incorporated into Light Forge to address possible errors that could emanate from heterogeneity of *M.tb* populations, which may be the case if primary specimens are used. We, therefore, propose a future Light Forge platform that interrogates a single copy of DNA in each reactor by combining limiting dilution and MLSI to acquire melting point data for each single genome amplicon. Further advancements are required for Light Forge to transition from proof-of-concept to a commercially viable product. The run time of 180 min requires reduction to make it feasible in the clinical setting, an improvement which will be easily implemented with more robust hardware. It is encouraging that several devices have completed the transition from the diagnostic development pipeline to commercial point-of-care microfluidic devices for tasks such as blood analysis and nucleic acid quantification, as well as identification of pathogens such as *M.tb* (Cepheid's GeneXpert System) *(31)*.

Light Forge successfully detected 14 out of 16 samples based on their drug-resistance profiles. Whilst this is a positive step towards creating an accurate low-cost test for TB drug resistance, the assay needs to account for difficult-to-detect mutations such as double mutations and silent mutations, as well as work toward the final development of a version ready for clinical testing. Possible improvements include use of a 21 megapixel camera, a fluorescent lamp, a thermal block,

and a simple computer interface to create a cost-sensitive device similar to that developed by Hatch and colleagues *(32)*. Creating a device that is easy for an end user to operate will allow rapid integration within affected countries. A simplistic design will also reduce the cost of purchase and maintenance of the device. Prior to implementation, Light Forge should be validated using a larger number of isolates as well as biological specimens from patients with and without HIV from different geographical regions *(33)*.

SUPPLEMENTAL MATERIAL

Supplemental material is available at *The Journal of Applied Laboratory Medicine* online.

Nonstandard Abbreviations: TB, tuberculosis; HRMA, high-resolution melting analysis; DST, drug susceptibility testing; M.tb, Mycobacterium tuberculosis; SNP, single nucleotide polymorphism; PCR, polymerase chain reaction; RIF, rifampicin; RRDR, rifampicin resistance determining region; AHRI, Africa Health Research Institute; NTC, non-template control; PDMS, polydimethylsiloxane; NN, nearest neighbor thermodynamic model; RFU, relative fluorescence units; T_M, melting temperature; ΔT_M, melting temperature change; ΔH^o, standard enthalpy change; ΔS^o, standard entropy change; C_t, total oligonucleotide strand concentration; RIF-R, rifampicin resistant; RIF-S, rifampicin susceptible; A, adenine; T, tyrosine; G, guanine; C, cytosine; POC, proof of concept.

Disclaimer: The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.*

I.M. Mbano, T. Mandizvo, J. Rogich, and F.K. Balagaddé developed the concepts and designed the study. JR was involved in the development and implementation of the technology. I.M. Mbano and T. Mandizvo performed the experiments. All the authors were involved in the analysis and interpretation of the results. I.M. Mbano, T. Mandizvo wrote the manuscript with feedback from all the authors.

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Authors' disclosures and/or potential conflicts of interest:*

Employment or Leadership: None declared. **Consultant or Advisory Role:** None declared. **Stock Ownership:** None declared.
Honoraria: None declared. **Research Funding:** The Howard Hughes Medical Institute (http://www.hhmi.org/), The National Research Foundation (https://www.nrf.ac.za/) NFS13091238686, The Cannon Collins Educational and Legal Assistance Trust (https://www.canoncollins.org.uk/ and The University of KwaZulu-Natal College of Health Sciences. The Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE), a DELTAS Africa Initiative [grant # DEL-15-006]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant # 107752/Z/15/Z] and the UK government. **Expert Testimony:** I.M. Mbano, Biophysical Society; T. Mandizvo, Biophysics in the Understanding, Diagnosis, and Treatment of Infectious Diseases, International Workshop on Microsystems Technologies for African Health. **Patents:** None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

Acknowledgments: We thank Dr. Alex S. Pym for *M.tb* clinical strains as well as helpful discussions. Further, we acknowledge the mechanical and computational assistance received from Engineer, Yashveer Ramlakhan.

REFERENCES

1. World Health Organization. Global tuberculosis report 2018. Geneva: World Health Organization; 2018.
2. World Health Organization. World health statistics 2018: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization; 2018.
3. Chung Delgado K, Guillen Bravo S, Revilla Montag A, Bernabe Ortiz A. Mortality among MDR TB cases: comparison with drug susceptible tuberculosis and associated factors. PloS One 2015;10:e0119332.
4. Jaramillo E. Guidelines for the programmatic management of drug resistant tuberculosis. Geneva: World Health Organization; 2008.
5. McNerney R. Diagnostics for developing countries. Diagnostics 2015;5:200 9.
6. Peeling R, McNerney R. Increasing access to diagnostics through technology transfer and local production; 2011.
7. Mark D, Haeberle S, Roth G, Von Stetten F, Zengerle R. Microfluidic lab on a chip platforms: requirements, characteristics and applications. In: Microfluidics based microsystems. Berlin: Springer; 2010. p. 305 76.
8. Reed GH, Kent JO, Wittwer CT. High resolution DNA melting analysis for simple and efficient molecular diagnostics; 2007.
9. Lee AS, Ong DC, Wong JC, Siu GK, Yam W C. High resolution melting analysis for the rapid detection of fluoroquinolone and streptomycin resistance in *Mycobacterium tuberculosis*. PLoS One 2012;7:e31934.
10. Wolff BJ, Thacker WL, Schwartz SB, Winchell JM. Detection of macrolide resistance in *Mycoplasma pneumoniae* by real time PCR and high resolution melt analysis. Antimicrob Agents Chemother 2008;52:3542 9.
11. Ong DC, Yam W C, Siu GK, Lee AS. Rapid detection of rifampicin and isoniazid resistant *Mycobacterium tuberculosis* by high resolution melting analysis. J Clin Microbiol 2010;48:1047 54.
12. Ramaswamy S, Musser JM. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. Tubercle Lung Dis 1998;79: 3 29.
13. Telenti A, Imboden P, Marchesi F, Schmidheini T, Bodmer T. Direct, automated detection of rifampin resistant *Mycobacterium tuberculosis* by polymerase chain reaction and single strand conformation polymorphism analysis. Antimicrob Agents Chemother 1993;37:2054 8.
14. Larsen MH, Biermann K, Tandberg S, Hsu T, Jacobs WR Jr. Genetic manipulation of *Mycobacterium tuberculosis*. Curr Protoc Microbiol 2007;6:10A.
15. Chen X, Kong F, Wang Q, Li C, Zhang J, Gilbert GL. Rapid detection of isoniazid, rifampin, and ofloxacin resistance in *Mycobacterium tuberculosis* clinical isolates using high resolution melting analysis. J Clin Microbiol 2011;49: 3450 7.
16. Thorsen T, Maerkl SJ, Quake SR. Microfluidic large scale integration. Science 2002;298:580 4.
17. SantaLucia J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest neighbor thermodynamics. Proc Natl Acad Sci 1998;95:1460 5.
18. Watkins NE Jr, SantaLucia J Jr. Nearest neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. Nucleic Acids Res 2005;33:6258 67.
19. Liew M, Pryor R, Palais R, Meadows C, Erali M, Lyon E, et al. Genotyping of single nucleotide polymorphisms by high resolution melting of small amplicons. Clin Chem 2004;50:1156 64.
20. Pholwat S, Liu J, Stroup S, Gratz J, Banu S, Rahman SM, et al. Integrated microfluidic card with TaqMan probes and high resolution melt analysis to detect tuberculosis drug resistance mutations across 10 genes. MBio 2015;6: e02273 14.
21. Park S, Zhang Y, Lin S, Wang T H, Yang S. Advances in microfluidic PCR for point of care infectious disease diagnostics. Biotechnol Adv 2011;29:830 9.
22. Day SJ, Altman DG. Blinding in clinical trials and other studies. BMJ 2000;321:504.
23. Ramirez MV, Cowart KC, Campbell PJ, Morlock GP, Sikes D, Winchell JM, et al. Rapid detection of multidrug resistant *Mycobacterium tuberculosis* by use of real time PCR and high resolution melt analysis. J Clin Microbiol 2010;48: 4003 9.
24. Hoek K, van Pittius NG, Moolman Smook H, Carelse Tofa K, Jordaan A, Van Der Spuy G, et al. Fluorometric assay for testing rifampin susceptibility of *Mycobacterium tuberculosis* complex. J Clin Microbiol 2008;46:1369 73.
25. Pietzka AT, Indra A, Stöger A, Zeinzinger J, Konrad M, Hasenberger P, et al. Rapid identification of multidrug resistant *Mycobacterium tuberculosis* isolates by rpoB gene scanning using high resolution melting curve PCR analysis. J Antimicrob Chemother 2009;63: 1121 7.
26. Yin X, Zheng L, Liu Q, Lin L, Hu X, Hu Y, et al. High resolution melting curve analysis for rapid detection of rifampin resistance in *Mycobacterium tuberculosis*: a meta analysis. J Clin Microbiol 2013;51:3294 9.
27. Nakata N, Kai M, Makino M. Mutation analysis of mycobacterial rpoB genes and rifampin resistance using recombinant *Mycobacterium smegmatis*. Antimicrob Agents Chemother 2012;56:2008 13.
28. Alonso M, Palacios JJ, Herranz M, Penedo A, Menéndez Á, Bouza E, et al. Isolation of *Mycobacterium tuberculosis* strains with a silent mutation in rpoB leading to potential misassignment of resistance category. J Clin Microbiol 2011;49:2688 90.
29. Moure R, Martín R, Alcaide F. Silent mutation in rpoB detected from clinical samples with rifampin susceptible *Mycobacterium tuberculosis*. J Clin Microbiol 2011;49: 3722 3722.
30. Whitesides GM. The origins and the future of microfluidics. Nature 2006;442:368 373.
31. Eicher D, Merten CA. Microfluidic devices for diagnostic

applications. Exp Rev Mol Diagn 2011;11:505 19.
32. Hatch AC, Fisher JS, Tovar AR, Hsieh AT, Lin R, Pentoney SL, et al. 1 Million droplet array with wide field fluorescence imaging for digital PCR. Lab Chip 2011;11: 3838 45.
33. Mahilum Tapay L, Laitila V, Wawrzyniak JJ, Lee HH, Alexander S, Ison C, et al. New point of care Chlamydia Rapid Test bridging the gap between diagnosis and treatment: performance evaluation study. BMJ 2007;335:1190 4.

References

1 Organization, W. H.

2 2019, G. T. R. Geneva: World Health Organization,2019. *License:CCBY-NC-SA3.0IGO* (2019).

3 Lawn, S. D., Kerkhoff, A. D., Vogt, M. & Wood, R. HIV-associated tuberculosis: relationship between disease severity and the sensitivity of new sputum-based and urine-based diagnostic assays. *BMC Medicine* **11**, 231, doi:10.1186/1741-7015-11-231 (2013).

4 Houben, R. M. & Dodd, P. J. The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. *PLoS medicine* **13**, e1002152 (2016).

5 Jeyanathan, M., Yao, Y., Afkhami, S., Smaill, F. & Xing, Z. New tuberculosis vaccine strategies: taking aim at un-natural immunity. *Trends in immunology* **39**, 419-433 (2018).

6 Ellner, J. J., Hirsch, C. S. & Whalen, C. C. Correlates of protective immunity to Mycobacterium tuberculosis in humans. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **30 Suppl 3**, S279-S282, doi:10.1086/313874 (2000).

7 Fine, P. BCG vaccination against tuberculosis and leprosy. *British medical bulletin* **44**, 691-703 (1988).

8 Sterne, J., Rodrigues, L. & Guedes, I. Does the efficacy of BCG decline with time since vaccination? *The international journal of tuberculosis and lung disease* **2**, 200-207 (1998).

9 Aronson, N. E. *et al.* Long-term efficacy of BCG vaccine in American Indians and Alaska Natives: a 60-year follow-up study. *Jama* **291**, 2086-2091 (2004).

10 Stek, C. *et al.* The immune mechanisms of lung parenchymal damage in tuberculosis and the role of host-directed therapy. *Frontiers in microbiology* **9**, 2603 (2018).

11 Flynn, J. L., Gideon, H. P., Mattila, J. T. & Lin, P. L. Immunology studies in non-human primate models of tuberculosis. *Immunological reviews* **264**, 60-73 (2015).

12 Patel, K., Jhamb, S. S. & Singh, P. P. Models of latent tuberculosis: their salient features, limitations, and development. *Journal of laboratory physicians* **3**, 75 (2011).

13 Subbian, S. *et al.* Lesion-specific immune response in granulomas of patients with pulmonary tuberculosis: a pilot study. *PLoS One* **10** (2015).

14 van Leeuwen, L. M., van der Sar, A. M. & Bitter, W. Animal models of tuberculosis: zebrafish. *Cold Spring Harbor perspectives in medicine* **5**, a018580 (2015).

15 Lanoix, J.-P., Lenaerts, A. J. & Nuermberger, E. L. Heterogeneous disease progression and treatment response in a C3HeB/FeJ mouse model of tuberculosis. *Disease models & mechanisms* **8**, 603-610 (2015).

16 Clark, S., Hall, Y. & Williams, A. Animal models of tuberculosis: Guinea pigs. *Cold Spring Harbor perspectives in medicine* **5**, a018572 (2015).

17 Rahyussalim, A. J. *et al.* New bone formation in tuberculous-infected vertebral body defect after administration of bone marrow stromal cells in rabbit model. *Asian Spine Journal* **10**, 1 (2016).

18 Phuah, J. *et al.* Effects of B cell depletion on early Mycobacterium tuberculosis infection in cynomolgus macaques. *Infection and immunity* **84**, 1301-1311 (2016).

19 Maertzdorf, J. *et al.* Mycobacterium tuberculosis invasion of the human lung: first contact. *Frontiers in immunology* **9**, 1346 (2018).
20 Eberl, G., Colonna, M., Di Santo, J. & McKenzie, A. (2015).
21 Sonnenberg, G. F. & Artis, D. Innate lymphoid cells in the initiation, regulation and resolution of inflammation. *Nature medicine* **21**, 698-708 (2015).
22 Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics* **17**, 175 (2016).
23 Cui, Y. *et al.* Single-cell transcriptome analysis maps the developmental track of the human heart. *Cell reports* **26**, 1934-1950. e1935 (2019).
24 Chen, W. *et al.* UMI-count modeling and differential expression analysis for single-cell RNA sequencing. *Genome biology* **19**, 70 (2018).
25 Slavov, N. Unpicking the proteome in single cells. *Science* **367**, 512-513 (2020).
26 Linnarsson, S. & Teichmann, S. A. (BioMed Central, 2016).
27 Ranzoni, A. M., Strzelecka, P. M. & Cvejic, A. Application of single-cell RNA sequencing methodologies in understanding haematopoiesis and immunology. *Essays in biochemistry* **63**, 217-225 (2019).
28 See, P., Lum, J., Chen, J. & Ginhoux, F. A single-cell sequencing guide for immunologists. *Frontiers in immunology* **9**, 2425 (2018).
29 Baran-Gale, J., Chandra, T. & Kirschner, K. Experimental design for single-cell RNA sequencing. *Briefings in functional genomics* **17**, 233-239 (2018).
30 Papalexi, E. & Satija, R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nature Reviews Immunology* **18**, 35 (2018).
31 Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature biotechnology* **30**, 777 (2012).
32 Angerer, P. *et al.* Single cells make big data: New challenges and opportunities in transcriptomics. *Current Opinion in Systems Biology* **4**, 85-91 (2017).
33 Hunter, R. L. Pathology of post primary tuberculosis of the lung: an illustrated critical review. *Tuberculosis* **91**, 497-509 (2011).
34 Miyagawa, Y. *et al.* A case of pulmonary tuberculosis with bilateral hilar lymphadenopathy diagnosed by sputum culture subsequent to open thoracic biopsy. *Respiration* **59**, 247-249 (1992).
35 Hunter, R. L., Actor, J. K., Hwang, S.-A., Karev, V. & Jagannath, C. Pathogenesis of post primary tuberculosis: immunity and hypersensitivity in the development of cavities. *Annals of Clinical & Laboratory Science* **44**, 365-387 (2014).
36 Astuti, T., Chozin, I. N., Damayanti, N. S. & Nugrahenny, D. The levels of pro-fibrotic cytokines in pulmonary tuberculosis with minimal and extensive lesions. *Lung India: official organ of Indian Chest Society* **35**, 204 (2018).
37 Todd, N. W., Luzina, I. G. & Atamas, S. P. Molecular and cellular mechanisms of pulmonary fibrosis. *Fibrogenesis & tissue repair* **5**, 1-24 (2012).
38 Choi, J.-E. *et al.* Insulin-like growth factor-I receptor blockade improves outcome in mouse model of lung injury. *American journal of respiratory and critical care medicine* **179**, 212-219 (2009).
39 Klingberg, F., Hinz, B. & White, E. S. The myofibroblast matrix: implications for tissue repair and fibrosis. *The Journal of pathology* **229**, 298-309 (2013).
40 Geng, J. *et al.* Phosphatase and tensin homolog deleted on chromosome 10 contributes to phenotype transformation of fibroblasts in idiopathic pulmonary

fibrosis via multiple pathways. *Experimental Biology and Medicine* **241**, 157-165 (2016).
41 Galloway, N. L. *et al.* Cell-to-cell transmission of HIV-1 is required to trigger pyroptotic death of lymphoid-tissue-derived CD4 T cells. *Cell reports* **12**, 1555-1563 (2015).
42 Gieseck III, R. L., Wilson, M. S. & Wynn, T. A. Type 2 immunity in tissue repair and fibrosis. *Nature Reviews Immunology* **18**, 62 (2018).
43 Strieter, R. M., Gomperts, B. N. & Keane, M. P. The role of CXC chemokines in pulmonary fibrosis. *The Journal of clinical investigation* **117**, 549-556 (2007).
44 Lin, P. L. *et al.* Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. *Nature medicine* **20**, 75-79 (2014).
45 DiFazio, R. M. *et al.* Active transforming growth factor-β is associated with phenotypic changes in granulomas after drug treatment in pulmonary tuberculosis. *Fibrogenesis & tissue repair* **9**, 1-11 (2016).
46 Reyfman, P. A. *et al.* Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *American journal of respiratory and critical care medicine* **199**, 1517-1536 (2019).
47 Kwan, C. K. & Ernst, J. D. HIV and tuberculosis: a deadly human syndemic. *Clinical microbiology reviews* **24**, 351-376 (2011).
48 Walker, N. F. *et al.* Matrix Degradation in Human Immunodeficiency Virus Type 1–Associated Tuberculosis and Tuberculosis Immune Reconstitution Inflammatory Syndrome: A Prospective Observational Study. *Clinical infectious diseases* **65**, 121-132 (2017).
49 Kassa, D. *et al.* The effect of HIV coinfection, HAART and TB treatment on cytokine/chemokine responses to Mycobacterium tuberculosis (Mtb) antigens in active TB patients and latently Mtb infected individuals. *Tuberculosis* **96**, 131-140 (2016).
50 Brilha, S. *et al.* Early secretory antigenic target-6 drives matrix metalloproteinase-10 gene expression and secretion in tuberculosis. *American journal of respiratory cell and molecular biology* **56**, 223-232 (2017).
51 Gaffney, J., Solomonov, I., Zehorai, E. & Sagi, I. Multilevel regulation of matrix metalloproteinases in tissue homeostasis indicates their molecular specificity in vivo. *Matrix Biology* **44**, 191-199 (2015).
52 Ardain, A. *et al.* Group 3 innate lymphoid cells mediate early protective immunity against tuberculosis. *Nature* **570**, 528-532 (2019).
53 Orme, I. M. & Basaraba, R. J. in *Seminars in immunology.* 601-609 (Elsevier).
54 Eum, S.-Y. *et al.* Neutrophils are the predominant infected phagocytic cells in the airways of patients with active pulmonary TB. *Chest* **137**, 122-128 (2010).
55 Feng, C. G. *et al.* NK cell-derived IFN-γ differentially regulates innate resistance and neutrophil response in T cell-deficient hosts infected with Mycobacterium tuberculosis. *The Journal of Immunology* **177**, 7086-7093 (2006).
56 Lockhart, E., Green, A. M. & Flynn, J. L. IL-17 production is dominated by γδ T cells rather than CD4 T cells during Mycobacterium tuberculosis infection. *The Journal of Immunology* **177**, 4662-4669 (2006).
57 Cooper, A. M. Cell-mediated immune responses in tuberculosis. *Annual review of immunology* **27**, 393-422 (2009).
58 North, R. J. & Jung, Y.-J. Immunity to tuberculosis. *Annu. Rev. Immunol.* **22**, 599-623 (2004).

59 Ehlers, S. & Schaible, U. E. The granuloma in tuberculosis: dynamics of a host–pathogen collusion. *Frontiers in immunology* **3**, 411 (2013).
60 Lin, P. L. *et al.* Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. *Nature medicine* **20**, 75 (2014).
61 Jaksik, R., Iwanaszko, M., Rzeszowska-Wolny, J. & Kimmel, M. Microarray experiments and factors which affect their reliability. *Biology direct* **10**, 46 (2015).
62 Sreeramareddy, C. T., Panduru, K. V., Menten, J. & Van den Ende, J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. *BMC infectious diseases* **9**, 91 (2009).
63 de Martino, M., Lodi, L., Galli, L. & Chiappini, E. Immune response to Mycobacterium tuberculosis: a narrative review. *Frontiers in Pediatrics* **7**, 350 (2019).
64 Petruccioli, E. *et al.* Correlates of tuberculosis risk: predictive biomarkers for progression to active tuberculosis. *European Respiratory Journal* **48**, 1751-1763 (2016).
65 Singhania, A. *et al.* A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. *Nature communications* **9**, 2308 (2018).
66 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559 (2008).
67 Parnell, G. *et al.* Aberrant cell cycle and apoptotic changes characterise severe influenza A infection–a meta-analysis of genomic signatures in circulating leukocytes. *Plos one* **6** (2011).
68 Al-Kofahi, Y. *et al.* Cell-based quantification of molecular biomarkers in histopathology specimens. *Histopathology* **59**, 40-54 (2011).
69 Baeuerle, P. A. & Baltimore, D. I kappa B: a specific inhibitor of the NF-kappa B transcription factor. *Science* **242**, 540-546 (1988).
70 Cai, Y. *et al.* Single-cell transcriptomics of blood reveals a natural killer cell subset depletion in tuberculosis. *EBioMedicine* **53**, 102686 (2020).
71 Garand, M. *et al.* Functional and phenotypic changes of natural killer cells in whole blood during Mycobacterium tuberculosis infection and disease. *Frontiers in immunology* **9**, 257 (2018).
72 Kleaveland, K. R. *et al.* Fibrocytes are not an essential source of type I collagen during lung fibrosis. *The Journal of Immunology* **193**, 5229-5239 (2014).
73 Maharaj, S., Shimbori, C. & Kolb, M. Fibrocytes in pulmonary fibrosis: a brief synopsis. *European Respiratory Review* **22**, 552-557 (2013).
74 Habermann, A. C. *et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Science Advances* **6**, eaba1972 (2020).
75 Luca, S. & Mihaescu, T. History of BCG Vaccine. *Maedica* **8**, 53-58 (2013).
76 Mangtani, P. *et al.* Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clinical infectious diseases* **58**, 470-480 (2014).
77 Ogongo, P., Porterfield, J. Z. & Leslie, A. Lung tissue resident memory T-cells in the immune response to Mycobacterium tuberculosis. *Frontiers in immunology* **10**, 992 (2019).
78 Mittrücker, H.-W. *et al.* Poor correlation between BCG vaccination-induced T cell responses and protection against tuberculosis. *Proceedings of the National Academy of Sciences* **104**, 12434-12439 (2007).

79 Kagina, B. M. *et al.* Specific T cell frequency and cytokine expression profile do not correlate with protection against tuberculosis after bacillus Calmette-Guerin vaccination of newborns. *American journal of respiratory and critical care medicine* **182**, 1073-1079 (2010).
80 Cooper, A. M. *et al.* Disseminated tuberculosis in interferon gamma gene-disrupted mice. *Journal of Experimental Medicine* **178**, 2243-2247 (1993).
81 O'Garra, A. *et al.* The immune response in tuberculosis. *Annual review of immunology* **31**, 475-527 (2013).
82 Oja, A. *et al.* Trigger-happy resident memory CD4+ T cells inhabit the human lungs. *Mucosal immunology* **11**, 654-667 (2018).
83 Sakai, S. *et al.* Cutting edge: Control of Mycobacterium tuberculosis infection by a subset of lung parenchyma–homing CD4 T cells. *The Journal of Immunology* **192**, 2965-2969 (2014).
84 Kumar, B. V. *et al.* Human tissue-resident memory T cells are defined by core transcriptional and functional signatures in lymphoid and mucosal sites. *Cell reports* **20**, 2921-2934 (2017).
85 Thome, J. J. *et al.* Spatial map of human T cell compartmentalization and maintenance over decades of life. *Cell* **159**, 814-828 (2014).
86 Collins, N. *et al.* Skin CD4+ memory T cells exhibit combined cluster-mediated retention and equilibration with the circulation. *Nature communications* **7**, 11514 (2016).
87 Steinert, E. M. *et al.* Quantifying memory CD8 T cells reveals regionalization of immunosurveillance. *Cell* **161**, 737-749 (2015).
88 Wu, T. *et al.* Lung-resident memory CD8 T cells (TRM) are indispensable for optimal cross-protection against pulmonary virus infection. *Journal of leukocyte biology* **95**, 215-224 (2014).
89 Gebhardt, T., Mueller, S. N., Heath, W. R. & Carbone, F. R. Peripheral tissue surveillance and residency by memory T cells. *Trends in immunology* **34**, 27-32 (2013).
90 Corleis, B. *et al.* HIV-1 and SIV infection are associated with early loss of lung interstitial CD4+ T cells and dissemination of pulmonary tuberculosis. *Cell reports* **26**, 1409-1418. e1405 (2019).
91 Hombrink, P. *et al.* Programs for the persistence, vigilance and control of human CD8+ lung-resident memory T cells. *Nature immunology* **17**, 1467-1478 (2016).
92 Savas, P. *et al.* Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nature medicine* **24**, 986-993 (2018).
93 Riaz, N. *et al.* Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* **171**, 934-949. e916 (2017).
94 Snyder, M. E. *et al.* Generation and persistence of human tissue-resident memory T cells in lung transplantation. *Science immunology* **4** (2019).
95 Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology* **32**, 381 (2014).
96 Foreman, T. W. *et al.* CD4+ T-cell–independent mechanisms suppress reactivation of latent tuberculosis in a macaque model of HIV coinfection. *Proceedings of the National Academy of Sciences* **113**, E5636-E5644 (2016).

97 Hughes, T. K. *et al.* Highly efficient, massively-parallel single-cell RNA-Seq reveals cellular states and molecular features of human skin pathology. *BioRxiv*, 689273 (2019).
98 Qiu, X. *et al.* Single-cell mRNA quantification and differential analysis with Census. *Nature methods* **14**, 309-315 (2017).
99 Organization, W. H. & Organization, W. H. (2015).
100 Ross, J., Ehrlich, R., Hnizdo, E., White, N. & Churchyard, G. Excess lung function decline in gold miners following pulmonary tuberculosis. *Thorax* **65**, 1010-1015 (2010).
101 Pasipanodya, J. G. *et al.* Pulmonary impairment after tuberculosis. *Chest* **131**, 1817-1824 (2007).
102 Ravimohan, S., Kornfeld, H., Weissman, D. & Bisson, G. P. Tuberculosis and lung damage: from epidemiology to pathophysiology. *European Respiratory Review* **27** (2018).
103 Ngahane, B. H. M. *et al.* Post-tuberculous lung function impairment in a tuberculosis reference clinic in Cameroon. *Respiratory medicine* **114**, 67-71 (2016).
104 Hnizdo, E., Singh, T. & Churchyard, G. Chronic pulmonary function impairment caused by initial and recurrent pulmonary tuberculosis following treatment. *Thorax* **55**, 32-38 (2000).
105 Plit, M. *et al.* Influence of antimicrobial chemotherapy on spirometric parameters and pro-inflammatory indices in severe pulmonary tuberculosis. *European Respiratory Journal* **12**, 351-356 (1998).
106 Wallis, R. S. *et al.* Tuberculosis biomarkers discovery: developments, needs, and challenges. *The Lancet infectious diseases* **13**, 362-372 (2013).
107 Colquhoun, D. An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society open science* **1**, 140216 (2014).
108 Kwan, P. K. W. *et al.* A blood RNA transcript signature for TB exposure in household contacts. *BMC Infectious Diseases* **20**, 1-8 (2020).
109 Berry, M. P. *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* **466**, 973-977 (2010).
110 Nathan, A. *et al.* Multimodal memory T cell profiling identifies a reduction in a polyfunctional Th17 state associated with tuberculosis progression. *bioRxiv* (2020).
111 Tran, H. T. N. *et al.* A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome biology* **21**, 1-32 (2020).
112 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nature biotechnology* **33**, 495-502 (2015).
113 Lopez, D. *et al.* SaVanT: a web-based tool for the sample-level visualization of molecular signatures in gene expression profiles. *BMC genomics* **18**, 1-9 (2017).
114 Gopal, R. *et al.* S100A8/A9 proteins mediate neutrophilic inflammation and lung pathology during tuberculosis. *American journal of respiratory and critical care medicine* **188**, 1137-1146 (2013).
115 Downing, J. F., Pasula, R., Wright, J. R., Twigg, H. L. & Martin, W. J. Surfactant protein A promotes attachment of Mycobacterium tuberculosis to alveolar macrophages during infection with human immunodeficiency virus. *Proceedings of the National Academy of Sciences* **92**, 4848-4852 (1995).
116 Doitsh, G. *et al.* Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. *Nature* **505**, 509-514 (2014).

117 Liu, F. *et al.* CD8+ cytotoxic T cell and FOXP3+ regulatory T cell infiltration in relation to breast cancer survival and molecular subtypes. *Breast cancer research and treatment* **130**, 645-655 (2011).
118 Gideon, H. P. *et al.* Single-cell profiling of tuberculosis lung granulomas reveals functional lymphocyte signatures of bacterial control. *bioRxiv* (2020).
119 Muefong, C. N. & Sutherland, J. S. Neutrophils in tuberculosis-associated inflammation and lung pathology. *Frontiers in Immunology* **11**, 962 (2020).
120 Dheda, K. *et al.* Lung remodeling in pulmonary tuberculosis. *The Journal of infectious diseases* **192**, 1201-1210 (2005).
121 Urushiyama, H. *et al.* Role of α1 and α2 chains of type IV collagen in early fibrotic lesions of idiopathic interstitial pneumonias and migration of lung fibroblasts. *Laboratory Investigation* **95**, 872-885 (2015).
122 Parker, M. W. *et al.* Fibrotic extracellular matrix activates a profibrotic positive feedback loop. *The Journal of clinical investigation* **124** (2014).
123 Burgstaller, G. *et al.* The instructive extracellular matrix of the lung: basic composition and alterations in chronic lung disease. *European Respiratory Journal* **50** (2017).
124 Allen, J. T. & Spiteri, M. A. Growth factors in idiopathic pulmonary fibrosis: relative roles. *Respiratory research* **3**, 1-9 (2001).
125 Zinkle, A. & Mohammadi, M. Structural biology of the FGF7 subfamily. *Frontiers in Genetics* **10**, 102 (2019).
126 Lipson, K. E., Wong, C., Teng, Y. & Spong, S. in *Fibrogenesis & tissue repair.* 1-8 (Springer).
127 Yanagihara, T. *et al.* Connective-Tissue Growth Factor (CTGF/CCN2) Contributes to TGF-β1-Induced Lung Fibrosis. *bioRxiv* (2020).
128 Li, R. *et al.* Pdgfra marks a cellular lineage with distinct contributions to myofibroblasts in lung maturation and injury response. *Elife* **7**, e36865 (2018).
129 El Agha, E. *et al.* Is the fibroblast growth factor signaling pathway a victim of receptor tyrosine kinase inhibition in pulmonary parenchymal and vascular remodeling? *American Journal of Physiology-Lung Cellular and Molecular Physiology* **315**, L248-L252 (2018).
130 Sahin, H. & Wasmuth, H. E. Chemokines in tissue fibrosis. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* **1832**, 1041-1048 (2013).
131 Lewis, C. *et al.* CCL18 as a Mediator of the Pro-Fibrotic Actions of M2 Macrophages in the Vessel Wall during Hypertension. *The FASEB Journal* **31**, 825.822-825.822 (2017).
132 Nouailles, G. *et al.* CXCL5-secreting pulmonary epithelial cells drive destructive neutrophilic inflammation in tuberculosis. *The Journal of clinical investigation* **124**, 1268-1282 (2014).
133 Fleige, H. *et al.* IL-17–induced CXCL12 recruits B cells and induces follicle formation in BALT in the absence of differentiated FDCs. *Journal of Experimental Medicine* **211**, 643-651 (2014).
134 Gopal, R. *et al.* Interleukin-17-dependent CXCL13 mediates mucosal vaccine–induced immunity against tuberculosis. *Mucosal immunology* **6**, 972-984 (2013).
135 Carrette, F. & Surh, C. D. in *Seminars in immunology.* 209-217 (Elsevier).
136 Passtoors, W. M. *et al.* IL7R gene expression network associates with human healthy ageing. *Immunity & ageing* **12**, 1-9 (2015).
137 Pickens, S. R. *et al.* Characterization of interleukin-7 and interleukin-7 receptor in the pathogenesis of rheumatoid arthritis. *Arthritis & Rheumatism* **63**, 2884-2893 (2011).

138 Ong, C. W., Elkington, P. T. & Friedland, J. S. Tuberculosis, pulmonary cavitation, and matrix metalloproteinases. *American journal of respiratory and critical care medicine* **190**, 9-18 (2014).
139 Savai, R. *et al.* Immune and inflammatory cell involvement in the pathology of idiopathic pulmonary arterial hypertension. *American journal of respiratory and critical care medicine* **186**, 897-908 (2012).
140 Al Shammari, B. *et al.* The extracellular matrix regulates granuloma necrosis in tuberculosis. *The Journal of infectious diseases* **212**, 463-473 (2015).
141 Narula, N. & Iannuzzi, M. Sarcoidosis: pitfalls and challenging mimickers. *Frontiers in Medicine*, 856 (2021).
142 Belton, M. *et al.* Hypoxia and tissue destruction in pulmonary TB. *Thorax* **71**, 1145-1153 (2016).
143 Fan, D. & Kassiri, Z. Biology of Tissue Inhibitor of Metalloproteinase 3 (TIMP3), and Its Therapeutic Implications in Cardiovascular Pathology. *Frontiers in Physiology* **11**, 661 (2020).
144 Nayar, S. *et al.* Immunofibroblasts are pivotal drivers of tertiary lymphoid structure formation and local pathology. *Proceedings of the National Academy of Sciences* **116**, 13490-13497 (2019).
145 Prendergast, G. C., Malachowski, W. J., Mondal, A., Scherle, P. & Muller, A. J. Indoleamine 2, 3-dioxygenase and its therapeutic inhibition in cancer. *International review of cell and molecular biology* **336**, 175-203 (2018).
146 Opitz, C. A. *et al.* The therapeutic potential of targeting tryptophan catabolism in cancer. *British journal of cancer* **122**, 30-44 (2020).
147 Gautam, U. S. *et al.* In vivo inhibition of tryptophan catabolism reorganizes the tuberculoma and augments immune-mediated control of Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences* **115**, E62-E71 (2018).
148 Hsu, Y.-L. *et al.* Lung cancer-derived galectin-1 contributes to cancer associated fibroblast-mediated cancer progression and immune suppression through TDO2/kynurenine axis. *Oncotarget* **7**, 27584 (2016).
149 Ezure, T., Sugahara, M. & Amano, S. Senescent dermal fibroblasts negatively influence fibroblast extracellular matrix-related gene expression partly via secretion of complement factor D. *BioFactors* **45**, 556-562 (2019).
150 Lu, Q. *et al.* The lncRNA H19 mediates pulmonary fibrosis by regulating the miR-196a/COL1A1 axis. *Inflammation* **41**, 896-903 (2018).
151 Showalter, K. *et al.* Machine learning integration of scleroderma histology and gene expression identifies fibroblast polarisation as a hallmark of clinical severity and improvement. *Annals of the Rheumatic Diseases* **80**, 228-237 (2021).
152 Stoya, G. *et al.* Determination of autofluorescence of red blood cells (RbCs) in uremic patients as a marker of oxidative damage. *Clinical nephrology* **58**, 198-204 (2002).
153 Regard, J. B. *et al.* Verge: a novel vascular early response gene. *Journal of Neuroscience* **24**, 4092-4103 (2004).
154 Kseibati, M. O., Shehatou, G. S., Sharawy, M. H., Eladl, A. E. & Salem, H. A. Nicorandil ameliorates bleomycin-induced pulmonary fibrosis in rats through modulating eNOS, iNOS, TXNIP and HIF-1α levels. *Life Sciences* **246**, 117423 (2020).
155 Margaritis, M. Endothelial dysfunction in HIV infection: experimental and clinical evidence on the role of oxidative stress. *Ann Res Hospitals* **3**, 1-11 (2019).

156 Okoye, A. A. & Picker, L. J. CD 4+ T-cell depletion in HIV infection: mechanisms of immunological failure. *Immunological reviews* **254**, 54-64 (2013).
157 Joosten, S. A. *et al.* Patients with tuberculosis have a dysfunctional circulating B-cell compartment, which normalizes following successful treatment. *PLoS pathogens* **12**, e1005687 (2016).
158 van Woudenbergh, E. *et al.* HIV Is Associated with Modified Humoral Immune Responses in the Setting of HIV/TB Coinfection. *Msphere* **5** (2020).
159 Jenkins, S. J. *et al.* Local macrophage proliferation, rather than recruitment from the blood, is a signature of TH2 inflammation. *science* **332**, 1284-1288 (2011).
160 Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356** (2017).
161 Condamine, T. *et al.* Lectin-type oxidized LDL receptor-1 distinguishes population of human polymorphonuclear myeloid-derived suppressor cells in cancer patients. *Science immunology* **1** (2016).
162 Zak, D. E. *et al.* A prospective blood RNA signature for tuberculosis disease risk. *Lancet (London, England)* **387**, 2312 (2016).
163 Maertzdorf, J. *et al.* Concise gene signature for point-of-care classification of tuberculosis. *EMBO molecular medicine* **8**, 86-95 (2016).
164 Cliff, J. M. *et al.* Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. *The Journal of infectious diseases* **207**, 18-29 (2013).
165 Singhania, A. *et al.* A modular transcriptional signature identifies phenotypic heterogeneity of human tuberculosis infection. *Nature communications* **9**, 1-17 (2018).
166 Maertzdorf, J. *et al.* Human gene expression profiles of susceptibility and resistance in tuberculosis. *Genes & Immunity* **12**, 15-22 (2011).
167 Berry, M. P. *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* **466**, 973 (2010).
168 Lambrechts, D. *et al.* Phenotype molding of stromal cells in the lung tumor microenvironment. *Nature medicine* **24**, 1277-1289 (2018).
169 Travaglini, K. J. *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619-625, doi:10.1038/s41586-020-2922-4 (2020).
170 Nguyen, Q. H., Pervolarakis, N., Nee, K. & Kessenbrock, K. Experimental considerations for single-cell RNA sequencing approaches. *Frontiers in cell and developmental biology* **6**, 108 (2018).
171 García-Navas, R., Gajate, C. & Mollinedo, F. Neutrophils drive endoplasmic reticulum stress-mediated apoptosis in cancer cells through arginase-1 release. *Scientific reports* **11**, 1-17 (2021).
172 Shikuma, C. M. *et al.* The role of HIV and monocytes/macrophages in adipose tissue biology. *Journal of acquired immune deficiency syndromes (1999)* **65**, 151 (2014).
173 Hensley-McBain, T. & Klatt, N. R. The dual role of neutrophils in HIV infection. *Current Hiv/aids Reports* **15**, 1-10 (2018).
174 Gil, O. *et al.* Granuloma encapsulation is a key factor for containing tuberculosis infection in minipigs. *PloS one* **5**, e10030 (2010).
175 Evans, S., Butler, J. R., Mattila, J. T. & Kirschner, D. E. Systems biology predicts that fibrosis in tuberculous granulomas may arise through macrophage-to-myofibroblast transformation. *PLOS Computational Biology* **16**, e1008520 (2020).

176 McCaffrey, E. F. *et al.* Multiplexed imaging of human tuberculosis granulomas uncovers immunoregulatory features conserved across tissue and blood. *bioRxiv* (2020).
177 Polena, H. *et al.* Mycobacterium tuberculosis exploits the formation of new blood vessels for its dissemination. *Scientific reports* **6**, 1-11 (2016).
178 Boasso, A., Shearer, G. & Chougnet, C. Immune dysregulation in human immunodeficiency virus infection: know it, fix it, prevent it? *Journal of internal medicine* **265**, 78-96 (2009).
179 Anand, A. R., Rachel, G. & Parthasarathy, D. HIV proteins and endothelial dysfunction: implications in cardiovascular disease. *Frontiers in Cardiovascular Medicine* **5**, 185 (2018).
180 Brune, K. A. *et al.* HIV impairs lung epithelial integrity and enters the epithelium to promote chronic lung inflammation. *PloS one* **11**, e0149679 (2016).
181 Katsuumi, G., Shimizu, I., Yoshida, Y. & Minamino, T. Vascular senescence in cardiovascular and metabolic diseases. *Frontiers in cardiovascular medicine* **5**, 18 (2018).
182 Lin, J. *et al.* Lung cancer transcriptomes refined with laser capture microdissection. *The American journal of pathology* **184**, 2868-2884 (2014).
183 Le Jan, S. *et al.* Characterization of the expression of the hypoxia-induced genes neuritin, TXNIP and IGFBP3 in cancer. *FEBS letters* **580**, 3395-3400 (2006).
184 Ming, J., Zhang, Q., Jiang, Y., Jiang, G. & Qiu, X. Interleukin 7 and its receptor promote cell proliferation and induce lymphangiogenesis in non-small cell lung cancer. *Zhonghua bing li xue za zhi= Chinese journal of pathology* **41**, 511 (2012).
185 Kuritzkes, D. R. Neutropenia, neutrophil dysfunction, and bacterial infection in patients with human immunodeficiency virus disease: the role of granulocyte colony-stimulating factor. *Clinical infectious diseases* **30**, 256-270 (2000).
186 Zannas, A. S. *et al.* Epigenetic upregulation of FKBP5 by aging and stress contributes to NF-κB–driven inflammation and cardiovascular risk. *Proceedings of the National Academy of Sciences* **116**, 11370-11379 (2019).
187 Brandt-Bohne, U., Keene, D. R., White, F. A. & Koch, M. MEGF9: a novel transmembrane protein with a strong and developmentally regulated expression in the nervous system. *Biochemical Journal* **401**, 447-457 (2007).
188 Zhang, H. *et al.* MMP9 protects against LPS-induced inflammation in osteoblasts. *Innate Immunity* **26**, 259-269 (2020).
189 You, J. *et al.* The oncogenic role of ARG1 in progression and metastasis of hepatocellular carcinoma. *BioMed research international* **2018** (2018).
190 Ko, C.-Y., Chang, W.-C. & Wang, J.-M. Biological roles of CCAAT/Enhancer-binding protein delta during inflammation. *Journal of biomedical science* **22**, 1-8 (2015).
191 Maertzdorf, J., Kaufmann, S. H. & Weiner, J. Toward a unified biosignature for tuberculosis. *Cold Spring Harbor perspectives in medicine* **5**, a018531 (2015).
192 Warsinske, H., Vashisht, R. & Khatri, P. Host-response-based gene signatures for tuberculosis diagnosis: A systematic comparison of 16 signatures. *PLoS medicine* **16**, e1002786 (2019).
193 Sia, J. K., Georgieva, M. & Rengarajan, J. Innate immune defenses in human tuberculosis: an overview of the interactions between Mycobacterium tuberculosis and innate immune cells. *Journal of immunology research* **2015** (2015).

194 Schafer, J., Muller-Trutwin, M. & Reeves, R. K. NK cell exhaustion: bad news for chronic disease? (2015).
195 Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nature communications* **10**, 1-9 (2019).
196 Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome medicine* **9**, 1-12 (2017).
197 Ziegenhain, C. *et al.* Comparative analysis of single-cell RNA sequencing methods. *Molecular cell* **65**, 631-643. e634 (2017).
198 Colditz, G. A. *et al.* Efficacy of BCG Vaccine in the Prevention of Tuberculosis: Meta-analysis of the Published Literature. *JAMA* **271**, 698-702, doi:10.1001/jama.1994.03510330076038 (1994).
199 Dockrell, H. M. & Smith, S. G. What Have We Learnt about BCG Vaccination in the Last 20 Years? *Front Immunol* **8**, 1134, doi:10.3389/fimmu.2017.01134 (2017).
200 Deshpande, S. A. Ethnic differences in the rates of BCG vaccination. *Archives of Disease in Childhood* **89**, 48-49 (2004).
201 Mangtani, P. *et al.* Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials. *Clin Infect Dis* **58**, 470-480, doi:10.1093/cid/cit790 (2014).
202 Ottenhoff, T. H. M. & Kaufmann, S. H. E. Vaccines against Tuberculosis: Where Are We and Where Do We Need to Go? *PLOS Pathogens* **8**, e1002607, doi:10.1371/journal.ppat.1002607 (2012).
203 Sterne, J. A., Rodrigues, L. C. & Guedes, I. N. Does the efficacy of BCG decline with time since vaccination? *Int J Tuberc Lung Dis* **2**, 200-207 (1998).
204 Loxton, A. G. Bcells and their regulatory functions during tuberculosis: latency and active disease. *Molecular immunology* **111**, 145-151 (2019).
205 Barnes, P. F., Lakey, D. L. & Burman, W. J. Tuberculosis in patients with HIV infection. *Infectious Disease Clinics* **16**, 107-126 (2002).
206 Cibrián, D. & Sánchez-Madrid, F. CD69: from activation marker to metabolic gatekeeper. *European journal of immunology* **47**, 946-953 (2017).
207 Baeyens, A. A. & Schwab, S. R. Finding a way out: S1P signaling and immune cell migration. *Annual review of immunology* **38**, 759-784 (2020).
208 Djenidi, F. *et al.* CD8+ CD103+ tumor–infiltrating lymphocytes are tumor-specific tissue-resident memory T cells and a prognostic factor for survival in lung cancer patients. *The Journal of Immunology* **194**, 3475-3486 (2015).
209 Shacklett, B. L., Ferre, A. L. & Kiniry, B. E. Defining T Cell Tissue Residency in Humans: Implications for HIV Pathogenesis and Vaccine Design. *Current HIV/AIDS Reports* **17**, 109-117 (2020).
210 Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202-1214 (2015).
211 La Manna, M. P. *et al.* Quantitative and qualitative profiles of circulating monocytes may help identifying tuberculosis infection and disease stages. *PloS one* **12**, e0171358 (2017).
212 Şahin, F., Yazar, E. & Yıldız, P. Prominent features of platelet count, plateletcrit, mean platelet volume and platelet distribution width in pulmonary tuberculosis. *Multidisciplinary respiratory medicine* **7**, 1-7 (2012).

213 Fox, K. A. *et al.* Platelets regulate pulmonary inflammation and tissue destruction in tuberculosis. *American journal of respiratory and critical care medicine* **198**, 245-255 (2018).
214 Zarinsefat, A. *et al.* Single-cell RNA sequencing of Tocilizumab-treated peripheral blood mononuclear cells as an in vitro model of inflammation. *Frontiers in genetics* **11**, 1695 (2021).
215 Deng, Y. *et al.* Low-density granulocytes are elevated in mycobacterial infection and associated with the severity of tuberculosis. *PloS one* **11**, e0153567 (2016).
216 Averill, M. M. *et al.* S100A9 differentially modifies phenotypic states of neutrophils, macrophages, and dendritic cells: implications for atherosclerosis and adipose tissue inflammation. *Circulation* **123**, 1216-1226 (2011).
217 Wang, W. *et al.* Ribosomal proteins and human diseases: pathogenesis, molecular mechanisms, and therapeutic implications. *Medicinal research reviews* **35**, 225-285 (2015).
218 Kilmartin, D. J. *et al.* CD69 expression on peripheral CD4+ T cells parallels disease activity and is reduced by mycophenolate mofetil therapy in uveitis. *Investigative ophthalmology & visual science* **42**, 1285-1292 (2001).
219 Garaud, S. *et al.* FOXP1 is a regulator of quiescence in healthy human CD4+ T cells and is constitutively repressed in T cells from patients with lymphoproliferative disorders. *European journal of immunology* **47**, 168-179 (2017).
220 Wong, W. F., Kohu, K., Chiba, T., Sato, T. & Satake, M. Interplay of transcription factors in T-cell differentiation and function: the role of Runx. *Immunology* **132**, 157-164 (2011).
221 Malhotra, N. *et al.* RORα-expressing T regulatory cells restrain allergic skin inflammation. *Science immunology* **3** (2018).
222 Hassan, N. J. *et al.* CD6 regulates T-cell responses through activation-dependent recruitment of the positive regulator SLP-76. *Molecular and cellular biology* **26**, 6727 (2006).
223 Lee, H. Y. *et al.* Intracellular formyl peptide receptor regulates naive CD4 T cell migration. *Biochemical and biophysical research communications* **497**, 226-232 (2018).
224 Hansen, C. B., Willer, A., Bayarri-Olmos, R., Kemper, C. & Garred, P. Expression of complement C3, C5, C3aR and C5aR1 genes in resting and activated CD4+ T cells. *Immunobiology* **224**, 307-315 (2019).
225 Lee, P. H. *et al.* G0S2 modulates homeostatic proliferation of naïve CD8+ T cells and inhibits oxidative phosphorylation in mitochondria. *Immunology and cell biology* **93**, 605-615 (2015).
226 del Carmen Lafita-Navarro, M. *et al.* MXD1 localizes in the nucleolus, binds UBF and impairs rRNA synthesis. *Oncotarget* **7**, 69536 (2016).
227 He, F. *et al.* PLAU inferred from a correlation network is critical for suppressor function of regulatory T cells. *Molecular systems biology* **8**, 624 (2012).
228 Yánez, D. C., Ross, S. & Crompton, T. The IFITM protein family in adaptive immunity. *Immunology* **159**, 365-372 (2020).
229 Bjorkdahl, O. *et al.* Characterization of CC-chemokine receptor 7 expression on murine T cells in lymphoid tissues. *Immunology* **110**, 170-179 (2003).

230 Yang, S., Liu, F., Wang, Q. J., Rosenberg, S. A. & Morgan, R. A. The shedding of CD62L (L-selectin) regulates the acquisition of lytic activity in human tumor reactive T lymphocytes. *PloS one* **6**, e22560 (2011).
231 Willinger, T. *et al.* Human naive CD8 T cells down-regulate expression of the WNT pathway transcription factors lymphoid enhancer binding factor 1 and transcription factor 7 (T cell factor-1) following antigen encounter in vitro and in vivo. *The Journal of Immunology* **176**, 1439-1446 (2006).
232 Rodriguez, M. W., Paquet, A. C., Yang, Y. H. & Erle, D. J. Differential gene expression by integrin β7+ and β7-memory T helper cells. *BMC immunology* **5**, 1-11 (2004).
233 Fu, S.-H., Yeh, L.-T., Chu, C.-C., Yen, B. L.-J. & Sytwu, H.-K. New insights into Blimp-1 in T lymphocytes: a divergent regulator of cell destiny and effector function. *Journal of biomedical science* **24**, 1-17 (2017).
234 Bernhardt, G. TACTILE becomes tangible: CD96 discloses its inhibitory peculiarities. *Nature immunology* **15**, 406-408 (2014).
235 Sekiya, T. *et al.* The nuclear orphan receptor Nr4a2 induces Foxp3 and regulates differentiation of CD4+ T cells. *Nature communications* **2**, 1-12 (2011).
236 Simhadri, V. R., Mariano, J. L., Zhou, Q., DeBell, K. E. & Borrego, F. Differential expression of CD300a/c on human TH 1 and TH 17 cells. *BMC immunology* **12**, 1-11 (2011).
237 Coletta, S. *et al.* The immune receptor CD300e negatively regulates T cell activation by impairing the STAT1-dependent antigen presentation. *Scientific reports* **10**, 1-13 (2020).
238 Adamson, A. *et al.* Tissue inhibitor of metalloproteinase 1 is preferentially expressed in Th1 and Th17 T-helper cell subsets and is a direct STAT target gene. *PLoS One* **8**, e59367 (2013).
239 Kozłowska, A., Hrycaj, P., Łącki, J. K. & Jagodziński, P. P. Perforin level in CD4+ T cells from patients with systemic lupus erythematosus. *Rheumatology international* **30**, 1627-1633 (2010).
240 Patil, V. S. *et al.* Precursors of human CD4+ cytotoxic T lymphocytes identified by single-cell transcriptome analysis. *Science immunology* **3** (2018).
241 Takeuchi, A. & Saito, T. CD4 CTL, a cytotoxic subset of CD4+ T cells, their differentiation and function. *Frontiers in immunology* **8**, 194 (2017).
242 Wang, X., Yu, L. & Wu, A. R. The effect of methanol fixation on single-cell RNA sequencing data. *BMC genomics* **22**, 1-16 (2021).
243 Koizumi, S.-i. *et al.* JunB regulates homeostasis and suppressive functions of effector regulatory T cells. *Nature communications* **9**, 1-14 (2018).
244 Haim-Vilmovsky, L. *et al.* Mapping Rora expression in resting and activated CD4+ T cells. *PloS one* **16**, e0251233 (2021).
245 Orta-Mascaró, M. *et al.* CD6 modulates thymocyte selection and peripheral T cell homeostasis. *Journal of Experimental Medicine* **213**, 1387-1397 (2016).
246 Case, A. J. *et al.* Elevated mitochondrial superoxide disrupts normal T cell development, impairing adaptive immune responses to an influenza challenge. *Free Radical Biology and Medicine* **50**, 448-458 (2011).
247 Sener, Z., Cederkvist, F. H., Volchenkov, R., Holen, H. L. & Skålhegg, B. S. T helper cell activation and expansion is sensitive to glutaminase inhibition under both hypoxic and normoxic conditions. *PLoS One* **11**, e0160291 (2016).

248 Cheuk, S. *et al.* CD49a expression defines tissue-resident CD8+ T cells poised for cytotoxic function in human skin. *Immunity* **46**, 287-300 (2017).
249 Collins, A., Littman, D. R. & Taniuchi, I. RUNX proteins in transcription factor networks that regulate T-cell lineage choice. *Nature Reviews Immunology* **9**, 106-115 (2009).
250 Rodríguez-Calvo, R., Tajes, M. & Vázquez-Carrera, M. The NR4A subfamily of nuclear receptors: potential new therapeutic targets for the treatment of inflammatory diseases. *Expert opinion on therapeutic targets* **21**, 291-304 (2017).
251 Lin, Y., Roberts, T. J., Sriram, V., Cho, S. & Brutkiewicz, R. R. Myeloid marker expression on antiviral CD8+ T cells following an acute virus infection. *European journal of immunology* **33**, 2736-2743 (2003).
252 Lewis, R. E., Cruse, J. M., Sanders, C. M., Webb, R. N. & Suggs, J. L. Aberrant expression of T-cell markers in acute myeloid leukemia. *Experimental and molecular pathology* **83**, 462-463 (2007).
253 Teijaro, J. R. *et al.* Cutting edge: Tissue-retentive lung memory CD4 T cells mediate optimal protection to respiratory virus infection. *The Journal of Immunology* **187**, 5510-5514 (2011).
254 Vitallé, J. *et al.* Altered expression of CD300a inhibitory receptor on CD4+ T cells from human immunodeficiency virus-1-infected patients: association with disease progression markers. *Frontiers in immunology* **9**, 1709 (2018).
255 Hughes, T. K. *et al.* Second-Strand Synthesis-Based massively parallel scRNA-seq reveals cellular states and molecular features of human inflammatory skin pathologies. *Immunity* **53**, 878-894. e877 (2020).
256 Pizzolla, A. *et al.* Influenza-specific lung-resident memory T cells are proliferative and polyfunctional and maintain diverse TCR profiles. *The Journal of clinical investigation* **128**, 721-733 (2018).
257 Yoshizawa, A. *et al.* TCR-pMHC encounter differentially regulates transcriptomes of tissue-resident CD8 T cells. *European journal of immunology* **48**, 128-150 (2018).
258 Tu, A. A. *et al.* TCR sequencing paired with massively parallel 3' RNA-seq reveals clonotypic T cell signatures. *Nature immunology* **20**, 1692-1699 (2019).
259 La Manna, M. P. *et al.* Harnessing unconventional T cells for immunotherapy of tuberculosis. *Frontiers in Immunology* **11**, 2107 (2020).
260 Lo Presti, E. *et al.* Current advances in γδ T cell-based tumor immunotherapy. *Frontiers in immunology* **8**, 1401 (2017).
261 Paules, C. I., Marston, H. D. & Fauci, A. S. Coronavirus infections—more than just the common cold. *Jama* **323**, 707-708 (2020).
262 Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* **395**, 565-574 (2020).
263 Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450-454 (2003).
264 Glowacka, I. *et al.* Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *Journal of virology* **85**, 4122-4134 (2011).
265 Böttcher-Friebertshäuser, E., Klenk, H.-D. & Garten, W. Activation of influenza viruses by proteases from host cells and bacteria in the human airway epithelium. *Pathogens and disease* **69**, 87-100 (2013).
266 Bugge, T. H., Antalis, T. M. & Wu, Q. Type II transmembrane serine proteases. *Journal of Biological Chemistry* **284**, 23177-23181 (2009).

267 Coutard, B. *et al.* The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral research* **176**, 104742 (2020).
268 Jonsdottir, H. R. & Dijkman, R. Coronaviruses and the human airway: a universal system for virus-host interaction studies. *Virology journal* **13**, 1-9 (2016).
269 Ordovas-Montanes, J. *et al.* Allergic inflammatory memory in human respiratory epithelial progenitor cells. *Nature* **560**, 649-654 (2018).
270 Russell, A. B., Trapnell, C. & Bloom, J. D. Extreme heterogeneity of influenza virus infection in single cells. *Elife* **7**, e32303 (2018).
271 Steuerman, Y. *et al.* Dissection of influenza infection in vivo by single-cell RNA sequencing. *Cell systems* **6**, 679-691. e674 (2018).
272 Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nature methods* **14**, 395 (2017).
273 Rijnink, W. F., Ottenhoff, T. H. & Joosten, S. A. B-cells and antibodies as contributors to effector immune responses in tuberculosis. *Frontiers in Immunology* **12**, 383 (2021).
274 Rahman, S. *et al.* Pulmonary tuberculosis patients with a vitamin D deficiency demonstrate low local expression of the antimicrobial peptide LL-37 but enhanced FoxP3+ regulatory T cells and IgG-secreting cells. *Clinical immunology* **156**, 85-97 (2015).
275 Pogoriler, J. & Husain, A. Pulmonary Development and Pediatric Lung Diseases. (2014).
276 Ostrowski, L., Andrews, K., Potdar, P. & Nettesheim, P. Ciliated-cell differentiation and gene expression. *Protoplasma* **206**, 245-248 (1999).
277 Wang, C. *et al.* High expression of Collagen Triple Helix Repeat Containing 1 (CTHRC1) facilitates progression of oesophageal squamous cell carcinoma through MAPK/MEK/ERK/FRA-1 activation. *Journal of Experimental & Clinical Cancer Research* **36**, 84 (2017).
278 Chen, L. *et al.* The long non-coding RNA CASC7 inhibits growth and invasion of non-small cell lung cancer cells through phosphatase and tensin homolog upregulation via sequestration of miR-92a. *International Journal of Oncology* **57**, 466-477 (2020).
279 Lewandowski, K. T. & Piwnica-Worms, H. Phosphorylation of the E3 ubiquitin ligase RNF41 by the kinase Par-1b is required for epithelial cell polarity. *Journal of cell science* **127**, 315-327 (2014).
280 Krischuns, T. *et al.* Phosphorylation of TRIM28 enhances the expression of IFN-β and proinflammatory cytokines during HPAIV infection of human lung epithelial cells. *Frontiers in immunology* **9**, 2229 (2018).
281 Wong, S. H., Lui, R. N. & Sung, J. J. Covid-19 and the digestive system. *Journal of gastroenterology and hepatology* **35**, 744-748 (2020).
282 Peng, Y. *et al.* Aberrant epithelial cell proliferation in peripheral airways in bronchiectasis. *Frontiers in cell and developmental biology* **8**, 88 (2020).
283 Giovannini-Chami, L. *et al.* Distinct epithelial gene expression phenotypes in childhood respiratory allergy. *European Respiratory Journal* **39**, 1197-1205 (2012).
284 Ziegler, C. G. *et al.* SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell* **181**, 1016-1035. e1019 (2020).

285 Liu, M.-Y., Zheng, B., Zhang, Y. & Li, J.-P. Role and mechanism of angiotensin-converting enzyme 2 in acute lung injury in coronavirus disease 2019. *Chronic Diseases and Translational Medicine* (2020).
286 Serfozo, P. *et al.* Ang II (angiotensin II) conversion to angiotensin-(1-7) in the circulation is POP (prolyloligopeptidase)-dependent and ACE2 (angiotensin-converting enzyme 2)-independent. *Hypertension* **75**, 173-182 (2020).
287 Bhaskaran, K. *et al.* HIV infection and COVID-19 death: a population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform. *The Lancet HIV* **8**, e24-e32 (2021).
288 Song, W.-m. *et al.* COVID-19 and Tuberculosis Coinfection: An Overview of Case Reports/Case Series and Meta-Analysis. *Frontiers in medicine* **8** (2021).
289 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology* **36**, 411-420 (2018).
290 Zou, Z. *et al.* Angiotensin-converting enzyme 2 protects from lethal avian influenza A H5N1 infections. *Nature communications* **5**, 1-7 (2014).
291 Weinheimer, V. K. *et al.* Influenza A viruses target type II pneumocytes in the human lung. *The Journal of infectious diseases* **206**, 1685-1694 (2012).
292 Zhang, L. *et al.* Infection of ciliated cells by human parainfluenza virus type 3 in an in vitro model of human airway epithelium. *Journal of virology* **79**, 1113-1124 (2005).
293 Kuba, K. *et al.* A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus–induced lung injury. *Nature medicine* **11**, 875-879 (2005).
294 Monteil, V. *et al.* Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. *Cell* (2020).
295 Vaduganathan, M. *et al.* Renin–angiotensin–aldosterone system inhibitors in patients with Covid-19. *New England Journal of Medicine* **382**, 1653-1659 (2020).
296 Kawase, M., Shirato, K., van der Hoek, L., Taguchi, F. & Matsuyama, S. Simultaneous treatment of human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute respiratory syndrome coronavirus entry. *Journal of virology* **86**, 6537-6545 (2012).
297 Naidoo, P. *et al.* The South African tuberculosis care cascade: estimated losses and methodological challenges. *The Journal of infectious diseases* **216**, S702-S713 (2017).
298 Cliff, J. M., Kaufmann, S. H., McShane, H., van Helden, P. & O'Garra, A. The human immune response to tuberculosis and its treatment: a view from the blood. *Immunological reviews* **264**, 88-102 (2015).
299 Konduri, N., Bastos, L. G. V., Sawyer, K. & Reciolino, L. F. A. User experience analysis of an eHealth system for tuberculosis in resource-constrained settings: a nine-country comparison. *International journal of medical informatics* **102**, 118-129 (2017).
300 Marinova, D., Gonzalo-Asensio, J., Aguilo, N. & Martin, C. MTBVAC from discovery to clinical trials in tuberculosis-endemic countries. *Expert review of vaccines* **16**, 565-576 (2017).
301 Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nature methods* **14**, 395-398 (2017).
302 Franks, T. J. *et al.* Resident cellular components of the human lung: current knowledge and goals for research on cell phenotyping and function. *Proceedings of the American Thoracic Society* **5**, 763-766 (2008).

303 Travaglini, K. J. *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619-625 (2020).
304 Nachiappan, A. C. *et al.* Pulmonary tuberculosis: role of radiology in diagnosis and management. *Radiographics* **37**, 52-72 (2017).
305 Barratt, S. & Millar, A. Vascular remodelling in the pathogenesis of idiopathic pulmonary fibrosis. *QJM: An International Journal of Medicine* **107**, 515-519 (2014).
306 Kiran, D., Podell, B. K., Chambers, M. & Basaraba, R. J. in *Seminars in immunopathology.* 167-183 (Springer).
307 Treacy, M., Lai, L., Costello, C. & Clark, A. Peripheral blood and bone marrow abnormalities in patients with HIV related disease. *British journal of haematology* **65**, 289-294 (1987).
308 Thompson, E. G. *et al.* Host blood RNA signatures predict the outcome of tuberculosis treatment. *Tuberculosis* **107**, 48-58 (2017).
309 MacLean, E. *et al.* A systematic review of biomarkers to detect active tuberculosis. *Nature microbiology* **4**, 748-758 (2019).
310 Adu-Gyamfi, C. G., Savulescu, D., George, J. A. & Suchard, M. S. Indoleamine 2, 3-Dioxygenase-Mediated Tryptophan Catabolism: A Leading Star or Supporting Act in the Tuberculosis and HIV Pas-de-Deux? *Frontiers in cellular and infection microbiology* **9**, 372 (2019).
311 Dunn, J. *et al.* Proteomic analysis discovers the differential expression of novel proteins and phosphoproteins in meningioma including NEK9, HK2 and SET and deregulation of RNA metabolism. *EBioMedicine* **40**, 77-91 (2019).
312 Steinbach, K., Vincenti, I. & Merkler, D. Resident-memory T cells in tissue-restricted immune responses: for better or worse? *Frontiers in immunology* **9**, 2827 (2018).
313 Liao, X., Makris, M. & Luo, X. M. Fluorescence-activated cell sorting for purification of plasmacytoid dendritic cells from the mouse bone marrow. *Journal of visualized experiments: JoVE* (2016).
314 Ni, W. *et al.* Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. *Critical Care* **24**, 1-10 (2020).
315 Bertram, S. *et al.* TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium. *Journal of virology* **87**, 6150-6160 (2013).
316 Kopel, J. *et al.* Racial and gender-based differences in COVID-19. *Frontiers in public health* **8**, 418 (2020).
317 Ortiz, M. E. *et al.* Heterogeneous expression of the SARS-Coronavirus-2 receptor ACE2 in the human respiratory tract. *EBioMedicine* **60**, 102976 (2020).
318 Braun, E. & Sauter, D. Furin-mediated protein processing in infectious diseases and cancer. *Clinical & translational immunology* **8**, e1073 (2019).
319 Carnero Contentti, E., Farez, M. F. & Correale, J. Mucosal-associated invariant T cell features and TCR repertoire characteristics during the course of multiple sclerosis. *Frontiers in immunology* **10**, 2690 (2019).
320 Guerra-Maupome, M., Palmer, M. V., Waters, W. R. & McGill, J. L. Characterization of γδ T cell effector/memory subsets based on CD27 and CD45R expression in response to Mycobacterium bovis infection. *Immunohorizons* **3**, 208-218 (2019).
321 Mbano, I. M. *et al.* Light Forge: A Microfluidic DNA Melting-based Tuberculosis Test. *The journal of applied laboratory medicine* **5**, 440-453 (2020).