Understanding the Cellular Ecology of Mtb Granulomas Using Single-Cell Sequencing

Understanding the Cellular Ecology of Mtb Granulomas Using Single-cell Sequencing

Abstract

Infection with *Mycobacterium tuberculosis* (Mtb) results in the formation of pulmonary granulomas, which are complex structures comprised of variable mixtures of stromal, parenchymal and immune cells. While Mtb granulomas within an individual have variable ability to control bacterial infection, we lack a complete understanding of the immunologic basis for bacterial control. Here, we develop and apply novel methods for high-throughput single-cell mRNA Sequencing (scRNA-Seq) to explore the relationship between the cellular composition of Mtb granulomas and their ability to control bacterial replication.

We initially report the development, optimization and application of Seq-Well, a portable, low-cost platform for high-throughput single-cell mRNA sequencing designed for use in BSL3 facilities. We further developed an optimized protocol that significantly improves the per-cell information content of scRNA-Seq data generated with Seq-Well. Using this improved technique, we went on to construct an atlas of multiple inflammatory skin conditions including acne, alopecia, granuloma annulare, leprosy, and psoriasis.

We next used Seq-Well to understand the relationship between the cellular and molecular features of Mtb granulomas and bacterial control and performed scRNA-Seq on a total of 26 non-human primate granulomas from 4 cynomolgus macaques. In low-burden lesions, we observe an overall expansion of T cells, particularly a group of T1-T17 T cells, while in high-burden lesions, there is an expansion of mast and plasma cells. Using longitudinal PET-CT imaging to track the development of granulomas throughout the course of infection, we observe that lesions that form later in infection are better able to control bacterial infection following the onset of adaptive immunity. Collectively, these data reveal a nuanced relationship between cellular composition, timing of granuloma formation and bacterial control.

Finally, we profile bronchoalveolar lavage of rhesus macaques following intravenous-BCG vaccination (IV-BCG), which results in robust protection against subsequent Mtb challenge. Here, we find a correlated expression program that includes genes associated with Th1 and Th17 effector function following IV-BCG vaccination.

In summary, we have developed and deployed a novel technique or scRNA-Seq to explore immunologic control in Mtb infection. Our data reveal a high-resolution image of T-cell mediated control in the setting of Mtb infection and protective vaccination.

Table of Contents

Acknowledgements

I started in the Shalek Lab in 2014 as an HST medical student uncertain whether I would pursue a PhD, and I had the good fortune of meeting Alex just as he was starting his lab at MIT. From our first meeting onwards, I have felt incredibly fortunate to have discovered a supportive mentor and friend. I came to HMS with an interest in global health, genomics and single-cell sequencing, and we somehow managed to combine these interests in ways I couldn't have imagined. I decided to affiliate with MD-PhD program from the moment I knew that Seq-Well would work, and it has been a wild ride ever since. Thank you for letting us all fly around the world to science with amazing people, while trying to make the world a bit better than we found it. While the lab has grown and your responsibilities as both a scientist and a father have expanded, your email response times have remained stunningly consistent (average under 5 minutes). Working together with you has truly been a life-changing experience, and I am eternally grateful for the way you have crafted the lab into a family. We have had some amazing times in and out of the lab, and I couldn't have asked for more in a mentor and friend. Alex, thank you for the things I can mention here, and even more for the things I shouldn't. It has truly been a pleasure to grow along with the lab, and I can't wait to see what comes next!

I began my tenure in the Shalek Lab as one of 8 first year graduate students, and in the ensuing six years the lab has grown to a point where I have lost count (many apologies to those I have left out, but I think there is a page limit here). I would like to begin by thanking Marc Wadsworth, my closest friend and co-conspirator in the Shalek Lab. Thank you for the adventures, the jokes and the good times through the trenches of grad school. Thanks to Jose Ordovas-Montanes for engaging conversations, good food, helpful feedback and always having a paper reference at the ready from your encyclopedic internal bibliography. Kelly Kolb for always brightening the lab with your acerbic Wisconsin wit and expert dinosaur drawings. Thank you to Aleth for making lab fun in a unicorn onesie and always reminding me when I was supposed to give group meeting – I seriously don't think I once knew when I was presenting otherwise. Thank

you to Sam Kazer for always being an encouraging friend and making sure the lab and the sequencers never fell apart throughout grad school – you the real MVP in the Shalek lab. Thank you to Jay Prakadan for always filling the lab with laughter and setting our feet to dancing. Thank you to Carly for your innovation and creativity in lab from spicy mezcal to space-cat – you are always an inspiration on the leading edge. Thank you to Alex Genshaft for making lab fun and all of the behind the scenes work – from building microscopes to making sure our sequencing runs never disappeared on the server, your efforts are much appreciated. Thank you to Nancy Tran for being the chief firefighter in the lab. Whether in ordering, organizing with collaborators, preparing and shipping arrays, I can't imagine grad school without you on board. Thank you to Sam "the Science-Mon" Allon for always being excited by crazy ideas. Your excitement and dedication is infectious, and I mean that in the best way possible as I write this amidst the 2020 coronavirus outbreak. Thank you to Conner Kummerlowe for going on a crazy trip to Zambia with me – next time I will see the smoke that thunders with you. A special thanks to Michelle Morrison for keeping me sane while sharing country music references and leaning into the occasional exaggerated Okie accent.

A very special thanks to Jack Strominger, with whom Melany and I lived with for a year during my graduate training. Jack, thank you for your guidance, advice and example of a full life in science. I will consider myself lucky in my career to have even a fraction of the passion and insatiable curiosity you had while running your lab at 94. Another very special thank you to Dr. Amr Sawalha. I am eternally grateful for your mentorship, guidance and the opportunities to publish and present that are rarely afforded to research technicians.

I would like to thank my many collaborators without whom this work would have been impossible. I would like to thank Todd Gierahn for teaching me to tinker and that the best protocols are written, remixed and revamped. Thank you to Chris Love for your support, encouragement, and courage completely opening up your lab to grad students from other labs. Thank you to Hannah Gideon for the incredible task of performing so many Seq-Well arrays in the BSL-3 facility

in Pittsburgh. Thank you for weathering all of the many Gates meetings with me and for introducing me to the many wonderful neighborhoods of Pittsburgh. I would like to thank Sarah Fortune for being a wonderful mentor and friend who has kept me focused on scientific rigor and openness throughout my research. Thank you to Robert Modlin for your mentorship and friendship. From introducing me to dermatology, the Hollywood Bowl and Iyengar yoga, you have influenced me in ways that will extend far beyond my PhD training. I would like to thank Al Leslie and Henrik Kloverpris for teaching me to surf in Durban and for helping us get Seq-Well off the ground in KZN. Finally, I would like to thank JoAnne Flynn for your support, constructive and insightful conversations throughout my graduate training.

Within the Immunology Program, I would like to thank Shiv Pillai and Wendy Garrett for their support and guidance through my PhD and Megan Eruzione for reminding me to turn in all of my forms and keeping me on track throughout grad school. I would like to thank the members of my PQE and DAC: Nir Hacohen, Jon Kagan, Barry Bloom and Sarah Fortune for your feedback, insights and support. Within the MD-PhD program and HST, I would like to thank my friends and classmates who have been a constant source of support and inspiration. Many thanks to the HST program directors Matthew Frosch, Rick Mitchell for their support, encouragement and guidance through this process. Many thanks to Patty Cunningham for your warmth andk kindness. On the MIT side of HST, I would like to thank Laurie Ward and Joe Stein for your friendship, kind words and help along the way. Finally at IMES, I would like to thank Hunter Lamere for keeping us safe from ourselves and our chemicals while talking about life and gardening and Tom Quinn for keeping the lights on at IMES and making sure I stay out of trouble.

Finally, I would like to thank Melany my wife and partner in life for your constant support through the trial and tribulations of graduate school. During my graduate training, our relationship has progressed through dating, engagement, marriage and beyond. You have inspired me to keep going when times have been tough, and you have been my source of joy in moments of despair. We have weathered the highs and lows of graduate school together, and I can’t imagine

this phase of life without you. I love you, and I am excited about what the future has in store for us together in life. I would also like to my parents Kevin and Carol Hughes for their love and support. The lessons I learned growing up are the ones that resonate most strongly and have propelled me forward through my academic journey.

Chapter 1

Introduction

1.1 Overview of Single Cell Sequencing

In recent years, single-cell genomics has fundamentally altered the resolution and perspective of biological inquiry [1-10]. Unlike traditional methods that rely on genomic material obtained from thousands of cells (e.g. DNA/RNA sequencing, microarray, etc.), single-cell genomics functions through the separate isolation and analysis of individual cells. Single-genomics has been used to interrogate numerous cellular analytes including surface protein expression, post-translational modification [11], genetic mutations [12, 13], and DNA methylation [14], but the most widely analyzed molecular species is messenger RNA (mRNA). Critically, single-cell mRNA sequencing (scRNA-Seq) provides a new perspective to biologic inquiry by enabling (1) broad interrogation of cellular identities and their relative proportions in complex tissue samples, (2) profiling of novel cell-types and phenotypic states, (3) ecosystems-level understanding of interactions between cell types.

In healthy tissues, single-cell mRNA sequencing has been used to construct single cell "atlases" that enumerate constituent cell-types of complex tissues [2, 15-22]. Importantly, these efforts have uncovered novel cell types (e.g. pulmonary ionocyte) [23, 24], immune diversity across tissues, and cell-cell interactions between diverse cell types [18]. Reference atlases have now been constructed for multiple tissues from model organisms [21, 22, 25], and the Human Cell Atlas [26], an international collaborative effort to generate single-cell references for all human tissues, is currently underway [27].

Numerous studies have applied high-throughput scRNA-Seq to explore the cellular and molecular basis of human disease and immune responses in complex tissues [2, 12, 19, 20, 27, 28]. In atlases of healthy tissues, enumeration of cell-types relies on major differences in gene expression, while subtle differences in cellular phenotypes likely underlie biology of interest in disease biopsies. In the next section, we will explore common techniques for scRNA-Seq and the limitations presented by these approaches in understanding gene expression differences in human disease.

1.2 Methods for Single-cell Sequencing

Collectively, technologies for single-cell mRNA function by isolating individual cells so that cellular contents can be separately labeled and later combined for downstream processing [6]. Containment of individual cells with reverse transcription primers or mRNA capture beads has been achieved through a number of approaches including sorting single-cells into 96- or 384-well plates and reverse-emulsion droplets generated using microfluidics.

Plate-based Methods for Single-cell Sequencing

In plate-based methods for scRNA-Seq, isolation of cells is achieved by FACS sorting individual cells into 96-well or 384-well plates [3-5, 29-31]. In these methods, cells are sorted directly into a small volume of lysis buffer contained in a and are rapidly placed on dry ice for storage until later processing. Plate-based methods afford the highest sensitivity of transcript detection since all reactions are performed in solution phase, but this added sensitivity comes at the expense of scale.

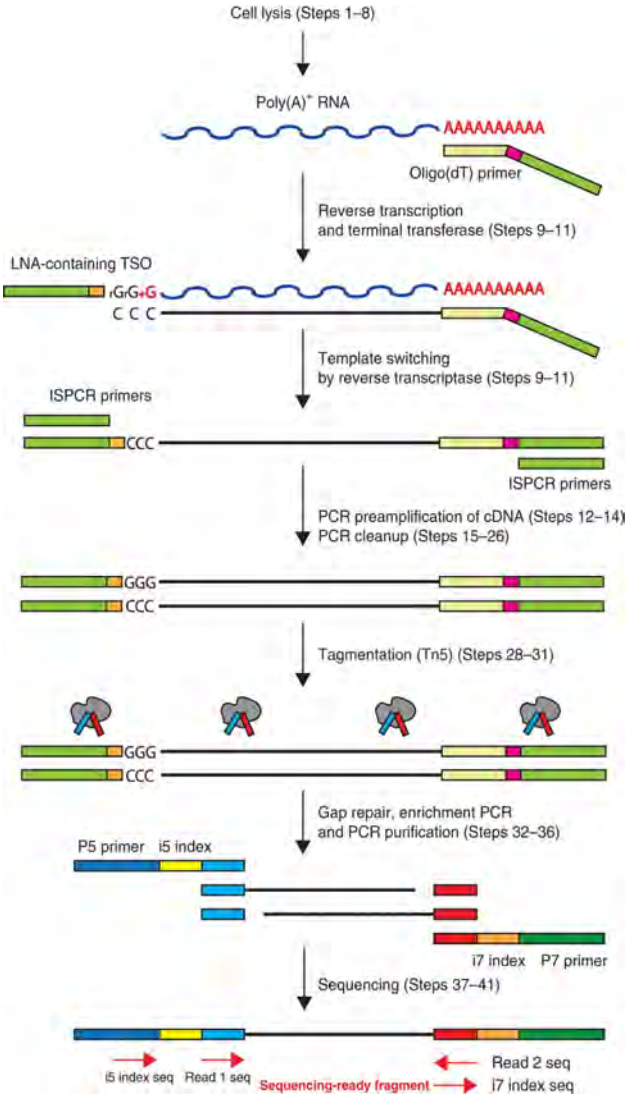


Figure 1.1 | Overview of Smart-Seq2 Pipeline. Adapted from Trombetta *et. al.* [3], cells are first lysed and mRNA transcripts captured with biotinylated poly-T SMART primers for reverse transcription (RT). Importantly, the RT enzyme deposits three ribosomal cytosines which are exploited to attach the TSO, generation full-length transcripts. Following complementary DNA (cDNA) generation, the product is amplified through PCR, tagmented by transposase (Illumina, Nextera) which deposits i7 and i5 adapters that serve as priming sights for the P7 and P5 primers. The quality of the library is validated using a BioAnalyzer (or Tape Station) and KAPA Quant. Once confirmed the library meets the necessary quality, it can be sequenced on an Illumina instrument.

Reverse transcription is performed using a primer containing a poly-T repeat sequence and PCR priming sequence. In some protocols, reverse transcription primers incorporate a unique molecular identifier (UMI) for unique labeling of individual mRNA molecules. Other methods rely on “spike-in” of known standardized mRNA transcripts of known size at a known concentration as a standardized estimate of mRNA capture [32]. During reverse transcription, many methods use “template switching” to append a second PCR priming sequence to the end of new synthesized cDNA molecules (**Figure 1.1**). Specifically, these methods rely on the terminal transferase activity of reverse transcriptase enzymes that add a terminal 3-base (CCC) overhang that can be used a priming site for additional extension. Incorporation of a template switching oligo (TSO) during reverse transcription in combination with the poly-T containing reverse transcription primer enables addition of complimentary primer sequences on both ends of cDNA molecules [33, 34]. Importantly, this enables integrated amplification of cDNA molecules using a single PCR primer, while suppressing

amplification of small fragments through self-dimerization. Following amplification, sequencing libraries are generated using tagmentation and low-round PCR amplification using indexed primers that append P5 and P7 sequences, which are compatible with Illumina sequencing (**Figure 1.1**).

High-throughput single-cell mRNA Sequencing

In recent years, there has been a significant increase in the scale of scRNA-Seq experiments [9, 10]. This increase in the scale of experimental processing has been realized through a fundamental advance: early bead-based mRNA capture and barcoding (**Figure 1.2**). Collectively, these techniques utilize microparticles (e.g. polymer microspheres or polyacrylamide hydrogels) decorated with barcoded, mRNA capture oligos to hybridize cellular mRNA. These mRNA captures sequences are covalently attached to solid support and contain key elements including (1) a PCR priming site or T7 promoter, (2) a cell barcode whose sequence is conserved within a microparticle but distinct from others within a single experiment, (3) a unique molecular identifier (UMI) that enables molecule-level enumeration of transcript counts, and (4) a poly-T capture sequence to hybridize poly-adenylated mRNA species (**Figure 1.2A**). Combinatorial diversity among cell barcode sequences is generated through the use of split-pool synthesis in which populations of beads are randomly split into 4 populations, a single base (either A,C,T or G) is synthesized, and beads are recombined. This process is repeated sequentially to obtain the desired number of combinatorial barcodes to avoid barcode collisions (**Figure 1.2B**). UMI sequences are generated through random synthesis of 8 bases so that each molecule on a given bead contains a random sequence that can be used to count individual transcripts within a given gene (**Figure 1.2C**). Using uniquely barcoded mRNA capture beads, single cells can be confined with a single mRNA capture bead to enable priming and reverse transcription of single cell transcriptomes.

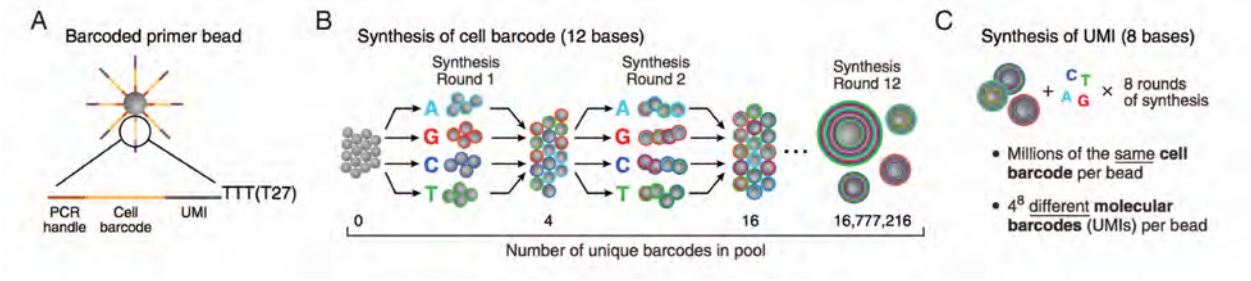


Figure 1.2 | Base-based Barcoding. Figure is adapted from Macosko et al. Cell 2015 [9]. (**A)** Structure of bead-bound oligonucleotide sequences including a PCR handle, 12 base-pair cell barcode, and 8 base-pair UMI. (**B)** Graphical overview of split-pool synthesis approach by which cell barcodes are constructed to yield a total of 16,777,216 combinatorial barcodes. (**C)** Graphical overview of the random synthesis of unique molecular identifier (UMI) sequences.

Multiple methods have been developed for high-throughput scRNA-Seq using early bead-based barcoding in combination with droplet-microfluidics (**Figure 1.3**) [9, 10, 35]. Here, individual cells are confined with barcoded mRNA capture beads or hydrogels containing mRNA capture sequences in reverse emulsion oil droplets. In these techniques, oil droplets are formed around an aqueous phase containing lysis buffer, cellular contents along with a single mRNA capture bead or hydrogel. Capture of mRNA molecules based on complementarity to a poly-dT capture sequence takes place within the confines of droplets, eliminating cross-contamination between droplets. After all input cells have been co-encapsulated in reverse emulsion droplets, oil droplets are broken and beads/hydrogels now decorated in DNA-RNA hybrid molecules are collected and reverse transcription is performed *en masse*. Following reverse transcription, cDNA libraries are amplified through either PCR or in vitro transcription. Finally, sequencing libraries are generated using whole-transcriptome amplification (WTA) products for sequencing using Illumina short-read sequencing.

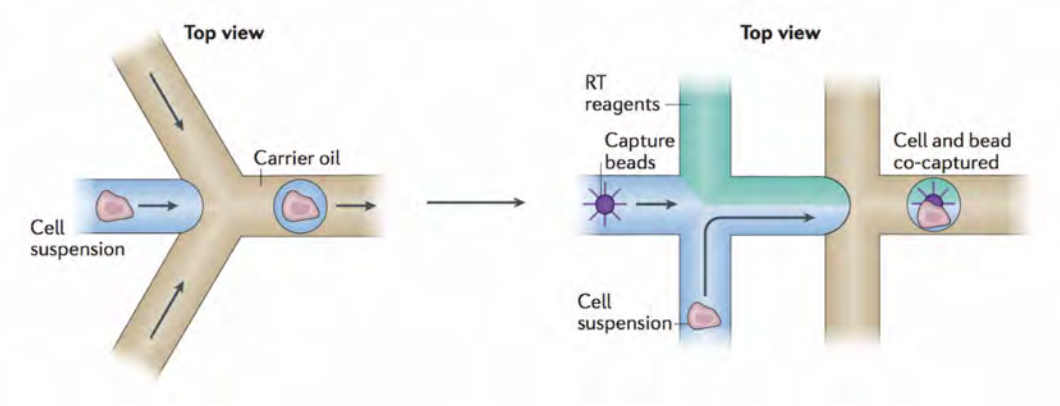


Figure 1.3 | Droplet-based single-cell mRNA sequencing. Figure is adapted from Prakadan et al. [1]. Droplet-based strategies for scRNA-Seq rely on flowing single cells in aqueous medium into a co-flowed oil phase. Small aqueous droplets containing cells are formed through encapsulation with oil. Individual droplets serve to physically segregate cellular material from adjacent droplets, enabling capture of an individual cell mRNA by a single, uniquely-barcoded capture bead.

1.4 Seq-Well: Designing Single-cell technologies to study Mtb Infection

Initially, we sought to adapt available scRNA-Seq technologies to profile Mtb granulomas. However, we quickly realized that existing technologies were limited in key ways. Specifically, plate-based scRNA-Seq methods (e.g. Smart-Seq2) are limited by the number of cells and cell-types that can be simultaneously profiled. Here, the small number of cells obtained from individual Mtb granulomas (100,000 – 500,000 total cells) severely limits the ability to accurately collect and profile rare populations across lesions. Further, the requirement to select marker panels prior to sorting would limit our ability to profile unanticipated cell populations. Finally, FACS sorting within BSL3 facilities presents numerous logistical issues, particularly in low-resource setting.

Droplet-based scRNA-Seq methods (e.g. Drop-Seq, InDrops, and 10x genomics), while not limited by the need for initial FACS sorting, present their own challenges. First, in droplet-based technologies, cells obtained from individual sample are processed in series rather than in parallel, leading to phenotypic changes between the first and last cells processed. Moreover, in many droplet microfluidic approaches, samples must be processed sequentially using separate microfluidic channels, which can lead to lengthy experiments and loss of sample viability over time. In experiments involving granulomas, time is a particularly important variable as cellular

viability and tissue necrosis are already present at baseline. Second, gentle lysis conditions are used in droplet microfluidic systems to maintain integrity of reverse-emulsion oil droplets. Harsh lysis buffers suitable for inactivation of pathogens like Mtb (e.g. 5M guanidine thiocyanate) are incompatible with the physical integrity of reverse emulsion droplets. The inability to inactivate the pathogen during sample processing would require downstream processing to occur within BSL3 facilities, where additional equipment may not be available.

Based on the aforementioned limitations of existing scRNA-Seq technologies, we set out to design a scRNA-Seq platform with the following objectives:

(1) *Minimal Sample Processing* – Methods that rely on pre-enrichment (e.g. FACS sorting or magnetic bead separation) would lead to loss of already limited cellular input (~100,000-1,000,000 cells/ granuloma).

(2) *Pathogen Inactivation* – The technique should be compatible with harsh lysis buffers capable of inactivation of Mtb and other pathogens.

(3) Portability *and Affordability* – A technology needs to perform equally well in high and low-resource setting. Whether in a BSL3 facility in the United States or South Africa, bulky and expensive peripheral equipment needed to be minimized.

Based on these design objectives, I worked to co-develop Seq-Well [36], portable, low-cost platform for high-throughput scRNA-Seq (Chapter 2). In the Seq-Well platform, mRNA capture beads developed for the Drop-Seq protocol are loaded into a PDMS microwell device that contains 85,000 wells that are sized to accommodate only a single bead (**Figure 1.4**). Following bead loading, a single-cell suspension at a limiting dilution (10,000 – 20,000 cells) is applied to the surface of the device and cells fall into wells by gravity. After cell loading, the device is reversibly sealed with a semi-permeable polycarbonate membrane and cells are lysed using 5M guanidine thiocyante. Following hybridization of cellular mRNA within wells, beads are removed and reverse transcription is performed *en masse*. PCR is then performed using the beads as a template to amplify cDNA libraries from which sequencing libraries can then be prepared.

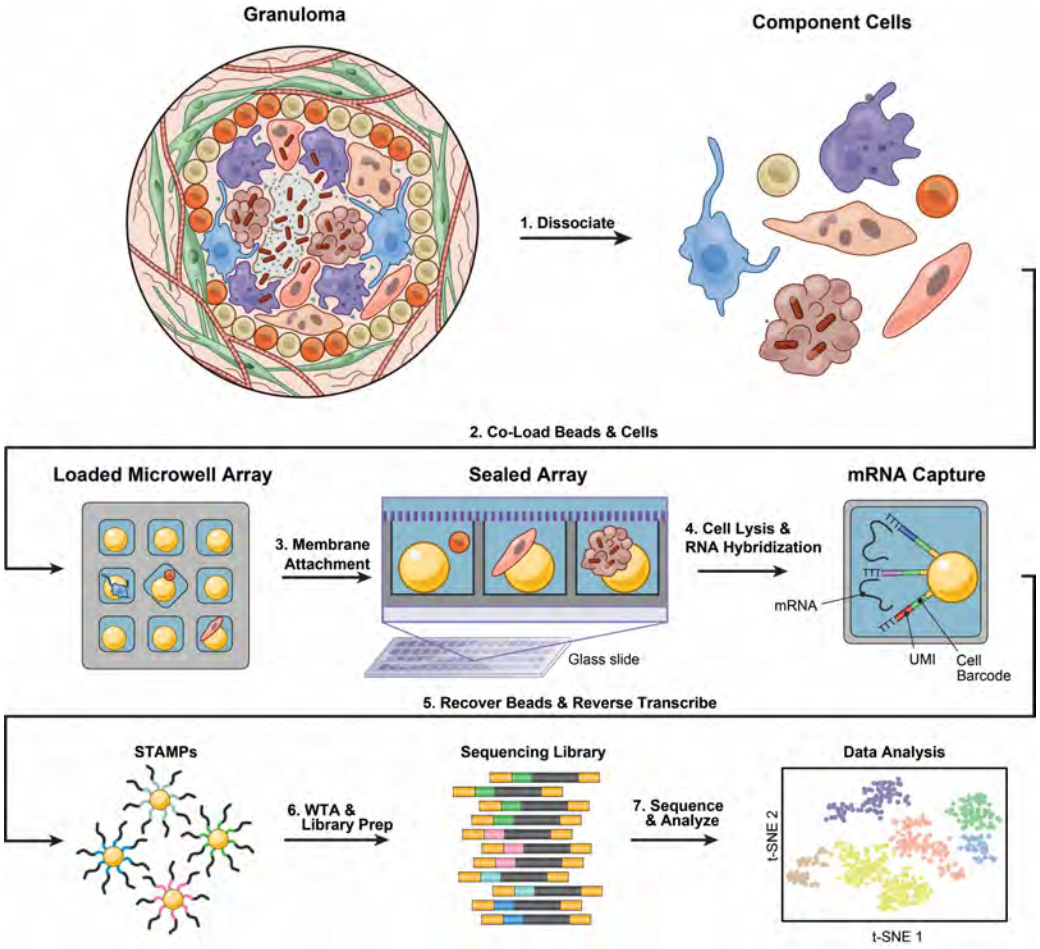


Figure 1.4 | Overview of Seq-Well Pipeline. Adapted from Gierahn *et. al.* [36], Following tissue dissociation and generation of a single cell suspension, uniquely-barcoded mRNA capture beads and single cells are loaded onto a functionalized Polydimethylsiloxane (PDMS) array. Critically, after beads and cells settle into wells, the array is sealed with a semi-permeable polycarbonate membrane that facilitates buffer exchange will keeping biological macromolecules (e.g. mRNA molecules) confined to their wells, preventing cross contamination. The mRNA molecules then hybridize to bead oligos, which are then recovered from the PDMS array and pooled for reverse transcription (RT). Following cDNA generation, the STAMPs (single-cell transcriptomes attached to microparticles) undergo subsequent processing (i.e. whole transcriptome amplification (WTA) and library preparation to generate sequencing libraries. Once libraries pass the necessary quality controls metrics, libraries are then sequenced on an Illumina instrument.

Unlike droplet microfluidic systems that require extensive equipment, loading of beads and cells is achieved through gravity and can be performed using only a pipette and standard lab equipment (e.g., incubators and shakers). Importantly, Seq-Well is compatible with harsh lysis conditions, which are capable of neutralizing Mtb. After optimizing this technology platform, we went on to demonstrate its ability to profile Mtb-infected macrophages in a BSL3 facility [36], and Seq-Well has since been used to profile numerous states of health and disease around the world.

1.5 Principles of Granulomatous Inflammation

Mtb Granuloma Structure and Function

Tuberculosis remains the largest cause of death due to infectious disease with an estimated 1.5 million deaths annually, while an estimated 1/3 of the world's population is latently infected [37]. Tuberculosis is transmitted through respiratory aerosol droplets, and upon infection, Mtb is sequestered in immune aggregates composed of activated macrophages and lymphocytes called granulomas [38]. Activation of alveolar macrophages represents the earliest event in granuloma formation, after which macrophages recruit additional cell types to the granuloma microenvironment. These additional cell types include T cells, B cells, plasma cells, dendritic cells, mast cells, eosinophils and fibroblasts (**Figure 1.5**).

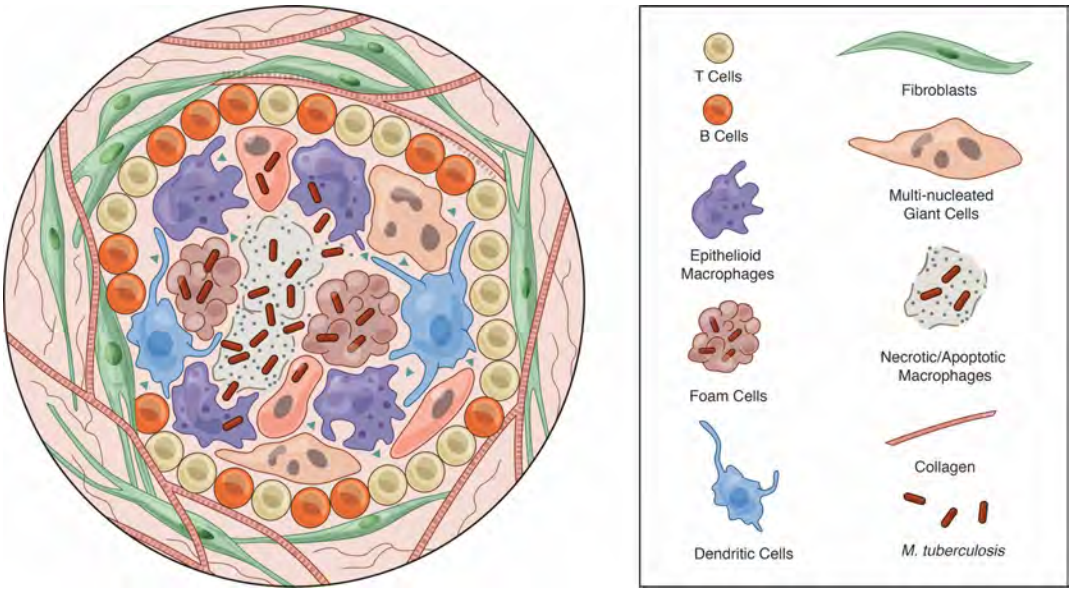


Figure 1.5 | Cellular Composition of Mtb Granulomas. Figure adapted from Pagan et al. [39]. Granulomas have a stereotyped spatial architecture. In caseating/necrotic granulomas, the center of lesions is made up of lipid-laden debris and necrotic material containing extracellular Mtb. Epithelioid macrophages are typically found in the center of lesions adjacent to the necrotic core. There is a lymphocytic cuff composed on T and B cells around the edge of lesions. Scattered dendritic cells can also be found within the lymphocytic region on the periphery of lesions. Finally, fibroblasts form around the edges of granulomas where they elaborate collagen fibers.

Canonically, Mtb granulomas display a stereotyped spatial architecture in which the center of lesions contains necrotic debris, often referred to as caseum due to its cheese-like histologic appearance. This necrotic core is surrounded by activated macrophages that adopt a unique,

epithelioid morphology in the granuloma microenvironment [40]. On the periphery of Mtb granulomas there is a pronounced lymphocytic cuff, which is composed primarily of T and B cells. The spatial organization of cells within granulomas is further reinforced by spatial localization of Inflammatory signaling [41, 42]. Mass spectroscopy of human Mtb granulomas localization of pro-inflammatory mediators areas of central necrosis surrounded by anti-inflammatory signals [41]. Targeted spatial sequencing of Mtb infection in C3HeB/FeJ mice further reveals elevated expression of Treg associated transripts (e.g. Il-10 and Foxp3) near the center of encapsulated granulomas [42].

Pathological examination reveals multiple states of Mtb granulomas including solid, caseous, cavitary, calcific lesions, which generally reflect different stages in the life-cycle of a granuloma. Solid lesions typically occur in earlier stages of disease when lesions are smaller in size. Granulomas with caseous necrosis form as lesions grow in size and death of macrophages and Mtb results in the central accumulation of necrotic debris. Cavitary lesions result from the growth of granulomas to a point of collision with an adjacent conducting airway. In this case, cavitary lesions lead to the spread of disease as the contents of granulomas containing bacteria are exposed to airways and spread upon coughing. Finally, calcific lesions are typically observed in later stages of disease and represent “healed” lesions. These lesions, however, can cause significant fibrosis and deformation of lung parenchyma leading to reduced pulmonary function.

Innate Immune Responses in Mtb Infection

In the early stages of Mtb infection, macrophages are the primary cell type that responds to infection [43, 44]. Intracellular infection of macrophages is the key feature of established Mtb infection, so understanding the factors that contribute to cell-autonomous elimination or tolerance of Mtb infection by macrophages is critical. Canonically, macrophages are known to be activated in 2 primary modes: (1) classical activation (M1) and (2) alternative activation (M2). Classical activation of macrophages occurs following exposure to instructive cytokines IFN-γ, while

alternative activation occurs in response to IL-4 or IL-13 [45]. While the M1/M2 framework has proved useful, this represents a vast over-simplification of possible macrophage states [28, 46]. Further, incomplete or frustrated phagocytosis is a central feature of granulomatous inflammation. Here, the inability of individual macrophages to completely engulf foreign material results in the formation of epithelioid macrophage and multi-nucleated giant cells – a cellular hallmark of granulomas [40]. In Mtb infection, multi-nucleated giant cells have been proposed to emerge through cell-cell fusion events [47, 48] or through mitotic defects in cytokinesis [49].

Recent studies have revealed a role for trained innate immunity in response to infection. Collectively, these studies suggest that the immunologic ground state of myeloid cells is not uniform. [50-52]. Inflammatory signals (e.g. IL-6, TNF) act to alter the trajectory of myeloid lineages during hematopoiesis [52] and reprogramming tissue-resident macrophages *in situ* [53]. For example, recent infection can result in in situ reprogramming of alveolar macrophages based on prior IFN-γ exposure [53]. Granulomas that are formed from trained myeloid cells are likely better equipped to control intracellular infection. Metabolic reprogramming is central feature of both innate immune training and control of intracellular infection [50]. Specifically, fatty acid oxidation (FAO) results in reduced ability of macrophages to eliminate bacteria, while increases in oxidative phosphorylation lead to improved control of Mtb [54-56].

Adaptive Immune Responses in Mtb Infection

Adaptive immunity is an essential feature of immune control in Mtb infection. In the absence of T cell and associated signaling molecules, control of Mtb infection is significantly impaired. Specifically, depletion of CD4+ T cells exacerbates early Mtb infection and can lead to re-activation of latent disease [57]. While antigen-specific T cells play a critical role in the control of tuberculosis infection, an appropriate balance of T cell phenotypes is critical [58]. Expression of Th1 cytokines (e.g. IFN-γ) is critically important for macrophage bacterial control of Mtb [59],

while type 2 cytokines (IL-4 and IL-13) inhibit control of vacuolar MTB [60]. Further, the presence of regulatory T cells can serve to limit protective, Mtb-specific T cell responses [61, 62].

Th1 cells are thought to play a major role in the control of Mtb infection. Th1 cells are marked by expression of key effector cytokines (e.g IFN-γ and TNF-α) involved in improved Mtb clearance in macrophages [63, 64]. Enhanced Th1 responses are associated with reduced Mtb burden in IL-10 deficient mice [65]. However, the appropriate distribution of Th1 cells is critical, as intravascular accumulation of Th1 effector cells impairs their ability to protect against Mtb within the lung parenchyma [66]. Intravital imaging of Mtb granulomas has revealed TNF-α mediated recruitment and retention of T cells in the granuloma microenvironment [67]. However, adoptive transfer and intravital imagine of antigen-specific T cells reveals a limited role for antigen presentation in the retention of T cells within the granuloma [68, 69].

Expression of TNF-α is a hallmark of TB-specific immune responses. Patients with active tuberculosis have a higher frequency of TNF-α positive CD4+ T cells than those with latent disease [70]. Neutralization of TNF-α results in disseminated Mtb infection in non-human primates [71]. Further, adoptive transfer of T cells from TNF +/+ mice are able to delay but not prevent overwhelming Mtb infection in RAG -/-, TNF -/- mice. [72]. Finally, TNF-α is essential to maintain granuloma integrity and TNF-inhibition is associated with reactivation of latent Mtb infection [73, 74].

While Th1 responses have canonically been thought to underlie protective immunity, there is accumulating evidence that Type 17 inflammation also plays a critical role in Mtb infection and response to Mtb vaccination [75, 76]. Th17 cells are capable of providing IFN-γ independent control of *M. bovis* infection, as adoptive transfer of IFN-γ -/- Th17 cells prolonged survival relative to naïve IFN-γ -/- T cells [77]. In the absence of IL-12 signaling, IL-23-mediated responses are sufficient to induce Mtb-protective Th1 and Th17 cells [78]. Further, group 3 innate lymphoid cells (ILC3s) rapidly accumulate in the lung following Mtb infection, and early expansion of ILC3s is

associated with immune protection in mice [79]. Taken together, with their role in other pulmonary infections, Th17 cells are likely a critical component of a protective, Mtb-specific immune response.

Other Infectious Granulomas

Cutaneous infection with *Mycobacterium leprae* (*M. leprae*), the causative agent of human leprosy infection, results in formation of dermal granulomas. In leprosy, bacteria infect dermal macrophages leading to recruitment of T cells. Importantly, leprosy infection underscores the importance of T cells in determining the outcome of infection. Canonically, there are 2 primary clinical forms of leprosy that can be distinguished by T cell phenotypes: (1) Tuberculoid leprosy is characterized by expansion of Th1-polarized T cells and improved bacterial control and (2) lepromatous leprosy is marked by Th2-polarization and poor control of bacteria [80]. Beyond the Th1-Th2 paradigm, the expression of granulysin by cytotoxic T cell is important for bacterial control [81], and Th17 cells have recently been implicated in control of leprosy infection [82, 83].

Granulomas that form around *Schistosoma mansoni (S. mansoni)* eggs display a strongly Th2-polarized immune response [84]. In contrast, Th1 dominant granulomatous inflammation is frequently observed in the setting of visceral leishmaniasis infection [85]. In granulomas that form in Salmonella infection, M2-like macrophages are less restrictive of Salmonella, while exposure to TNF limits replication of Salmonella [86]. While rarely observed clinically in the age of penicillin, secondary and tertiary syphilis infection can present as disseminated granulomatous disease [87]. Here, dermal syphilitic granulomas are composed of numerous plasma cells and histiocytes [88].

Sterile Granulomas

Granulomas can arise in response to numerous non-infectious stimuli including drug reactions and foreign material (e.g. surgical suture or biologic implants) [39]. Here, a broader understanding of granulomatous inflammation is instructive for a fuller exploration of the Mtb

granuloma. In the skin, foreign-body granuloma reactions can arise from dermal macrophage activation in response to surgical suture, tattoo ink, or polymer microparticles [89-91]. Granulomas also arise in setting of autoimmune inflammation in sarcoidosis [92, 93], Crohn's disease [94], large and medium vessel vasculitidies [95, 96], and atherosclerosis [97].

In sarcoidosis, granulomas are observed primarily in the lungs, lymph nodes and skin [98]. In contrast to tuberculous granulomas, sarcoid lesions less frequently exhibit extensive necrosis despite being driven by signaling pathways [99]. While Th1 cells are thought to predominate in sarcoid lesions, accumulation of IFN-γ Th17 cells has been observed in sarcoid granulomas [100, 101]. In large-vessel vasculitis (e.g. giant cell arteritis and Takayasu arteritis), it is thought that endothelial self-peptides lead to immune activation that culminates in the formation of vaso-occlusive granulomas [102]. Finally, atherosclerotic plaques share many of the primary features seen in Mtb granulomas including the accumulation of activated, foamy macrophages, formation of a fibrotic cap, and the presence of diverse effector T cells. Further, inflammatory macrophage activation [103] and the induction of lesional Th17 cells are associated with progression of atherosclerotic plaques [104, 105].

Gene Expression Profiling in Mtb Infection

Numerous studies have performed transcriptional profiling in blood and granulomas in *M. tuberculosis* infection in non-human primates and humans[106-111]. Expression of interferon-regulated gene pathways in peripheral blood prior to infection was associated with development of active disease in non-human primates [108]. Further, bulk transcriptional profiling of granulomas from rhesus macaques revealed elevation of inflammatory cytokines and chemokines among early granulomas, and a gradual loss of inflammation among older lesions [112].

Signatures from peripheral blood in humans have revealed a pattern of interferon expression that correlates with disease activity [109, 110]. Specifically, elevation of a neutrophil-driven pattern of Type I interferon responsive genes is observed in peripheral blood; although, it

is difficult to distinguish between changes in peripheral cell-type composition and intrinsic changes in gene expression [109]. Deconvolution of peripheral blood expression signatures along with Cytof profiling has revealed significant expansion of natural killer (NK) cell populations in successfully treated individuals [113]. Further, increased abundance of NK cells is associated with latency, while diminished pulmonary NK cells are associated with Mtb reactivation [113].

1.7 The non-human primate model of Mtb Infection

Mtb infection in the cynomolgus macaque most closely recapitulates the primary features of human tuberculosis [114, 115]. In this model, low-dose, bronchoscopic infection results in a spectrum of disease across animals in which 40% of animals develop active disease, 60% develop latent disease [116]. By 4-weeks post-infection multiple lesions can be detected through PET-CT imaging [116]. Necropsy of animals at 4-weeks post-infection has revealed that this time-point represents peak bacteremia. After 4 weeks, granuloma-level bacterial burdens begin to decline, presumably following activity of cytotoxic T cells and the onset of adaptive immunity (**Figure 1.6A-B**). Studies performed at 10 weeks post-infection reveal that animals harbor both sterile and non-sterile lesions irrespective of overall disease status (i.e. active or latent) [117].

While animals with active disease have higher cumulative bacterial burden than latent animals, granulomas within animals with active or latent disease display a similar range of bacterial burdens [117, 118]. However, animals with clinically latent disease have a higher percentage of sterilizing lesions compared to those with active disease (**Figure 1.6C**).

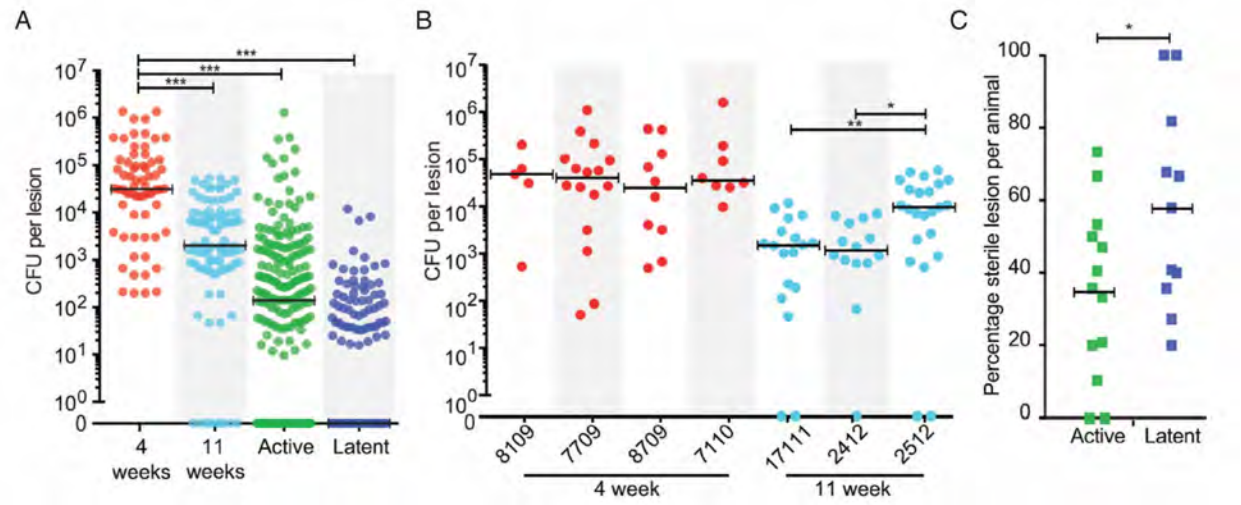


Figure 1.6 | Spectrum of Bacterial Burden in NHP Granulomas. Figure adapted from Lin*, Ford* et al.,[117]. A. CFU in lesions from monkeys at 4 weeks (4 animals, 68 lesions) is significantly higher than at 11 weeks (3 animals, 98 lesions), in active disease (13 animals, 222 lesions) and in clinically latent infection (11 animals, 145 lesions) ($p < 0.001$). B. Individual monkeys necropsied at 4 weeks and 11 weeks are shown. By 11 weeks there are significant differences between animals ($P < 0.05$). C. The percentage of sterile lesions in monkeys with latent infection (n=11) is significantly higher than in monkeys with active disease (n=13, $P < 0.05$).

Interestingly, non-human primates with established primary Mtb infection are resistant to acquisition of tuberculosis upon secondary challenge [119]. Here, non-human primates infected with orthogonally barcoded, but otherwise identical, strains of Mtb reveal significant resistance to secondary infection. In this case, granulomas form upon secondary challenge, but the majority of these lesions are either sterilizing or have low bacterial burdens. Temporal evolution is an essential feature of immune responses, but the exact nature of dynamic changes that underlie sterilization in secondary granulomas is not fully understood. Understanding alterations in the immunologic environment that supports granuloma formation is a key question that will be addressed in this dissertation.

1.8 Overview of Dissertation

This dissertation focuses on a central theme: how does cellular composition of MTB granulomas influence their ability to control bacterial replication? Here, we develop, optimize and

apply novel techniques for high-throughput single-cell mRNA sequencing to systematically address these questions.

In Chapters 2 and 3, we develop, optimize and apply Seq-Well [36, 120], a portable, low-cost platform for high-throughput single-cell mRNA sequencing. In chapter 2, we report the development of Seq-Well [36] and a series of validation experiments to demonstrate the effectiveness of the platform. Validation experiments include species mixing controls using mouse (NIH-3T3) and human (HEK293) cell lines, primary human cells (PBMCs), and the ability to work in challenging environments (e.g. BSL3 facilities). In Chapter 3, we develop a protocol that significantly improves the per-cell information content of scRNA-Seq data generated with Seq-Well [120]. Here, we developed a protocol (Seq-Well S^3) that improves recovery of cDNA molecules through the use of randomly-primed second strand synthesis. The improved sensitivity of Seq-Well S^3 results in enhanced detection of biologically informative molecules (e.g. transcription factors, cytokines, and cytokine receptors) that were previously under-sampled. Finally, we go on to construct an atlas of multiple inflammatory skin conditions using this improved technique.

In Chapter 4, we use Seq-Well to understand the relationship between the cellular and molecular features of Mtb granulomas and bacterial control in a non-human primate model of Mtb infection. Here, we performed Seq-Well on a total of 26 non-human primate granulomas from 4 cynomolgus macaques and obtained single-cell profiles for a total of 109,584 cells. We observe T cells to have the strongest relationship with reduced bacterial burden across animals, while we observe expansion of mast and plasma cells in high-burden lesions. Further, we observe a population of T1-T17 cells that are more frequent in low-burden lesions. Using longitudinal PET-CT imaging to track the development of granulomas throughout the course of infection, we observe a striking relationship between the timing of granuloma formation, bacterial burden and cellular composition. Collectively, our data reveal a nuanced relationship between cellular composition,

Finally, in Chapter 5, we provide a summary of previous work and provide an overview of ongoing and future work to control Mtb infection. Specifically, we discuss recent work aimed at understanding mechanisms of Bacillus Calmette-Guerin (BCG)-mediated protection in a multi-route vaccination in rhesus macaques [121]. Here, it was observed that intravenous administration of the BCG vaccine results in robust protection against subsequent Mtb challenge. Intravenous BCG vaccination proved more effective than aerosol and intradermal (high and low dose) vaccination. Notably, IV-vaccination was observed to induce a marked increase in the proportion of T cells in bronchoalveolar lavage (BAL). In collaboration with the NIH-VRC, we characterized cells from BAL at 13 and 25 weeks post-vaccination to understand the phenotypic state of immune cells induced across BCG vaccine routes.

References

1. Prakadan, S.M., A.K. Shalek, and D.A. Weitz, *Scaling by shrinking: empowering single-cell'omics' with microfluidic devices.* Nature Reviews Genetics, 2017. **18**(6): p. 345.

2. Tirosh, I., et al., *Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq.* Science, 2016. **352**(6282): p. 189-196.

3. Trombetta, J.J., et al., *Preparation of single-cell RNA-seq libraries for next generation sequencing.* Current protocols in molecular biology, 2014. **107**(1): p. 4.22. 1-4.22. 17.

4. Picelli, S., et al., *Smart-seq2 for sensitive full-length transcriptome profiling in single cells.* Nature methods, 2013. **10**(11): p. 1096-1098.

5. Picelli, S., et al., *Full-length RNA-seq from single cells using Smart-seq2.* Nature protocols, 2014. **9**(1): p. 171.

6. Shalek, A.K., et al., *Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells.* Nature, 2013. **498**(7453): p. 236-240.

7. Shalek, A.K., et al., *Single-cell RNA-seq reveals dynamic paracrine control of cellular variation.* Nature, 2014. **510**(7505): p. 363-369.

8. Rosenberg, A.B., et al., *Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding.* Science, 2018. **360**(6385): p. 176-182.

9. Macosko, E.Z., et al., *Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets.* Cell, 2015. **161**(5): p. 1202-1214.

10. Klein, A.M., et al., *Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.* Cell, 2015. **161**(5): p. 1187-1201.

11. Krutzik, P.O., et al., *Phospho flow cytometry methods for the analysis of kinase signaling in cell lines and primary human blood samples*, in *Flow cytometry protocols*. 2011, Springer. p. 179-202.

12. van Galen, P., et al., *Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity.* Cell, 2019. **176**(6): p. 1265-1281. e24.

13. Nam, A.S., et al., *Somatic mutations and cell identity linked by Genotyping of Transcriptomes.* Nature, 2019. **571**(7765): p. 355-360.

14. Smallwood, S.A., et al., *Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity.* Nature methods, 2014. **11**(8): p. 817-820.

15. Patel, A.P., et al., *Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma.* Science, 2014. **344**(6190): p. 1396-1401.

16. Gaublomme, J.T., et al., *Single-cell genomics unveils critical regulators of Th17 cell pathogenicity.* Cell, 2015. **163**(6): p. 1400-1412.

17. Braga, F.A.V., et al., *A cellular census of human lungs identifies novel cell states in health and in asthma.* Nature medicine, 2019. **25**(7): p. 1153-1163.

18. Vento-Tormo, R., et al., *Single-cell reconstruction of the early maternal–fetal interface in humans.* Nature, 2018. **563**(7731): p. 347-353.

19. Ordovas-Montanes, J., et al., *Allergic inflammatory memory in human respiratory epithelial progenitor cells.* Nature, 2018. **560**(7720): p. 649-654.

20. Smillie, C.S., et al., *Intra-and inter-cellular rewiring of the human colon during ulcerative colitis.* Cell, 2019. **178**(3): p. 714-730. e22.

21. Consortium, T.M., *Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris.* Nature, 2018. **562**(7727): p. 367.

22. Han, X., et al., *Mapping the mouse cell atlas by microwell-seq.* Cell, 2018. **172**(5): p. 1091-1107. e17.

23. Plasschaert, L.W., et al., *A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte.* Nature, 2018. **560**(7718): p. 377-381.

24. Montoro, D.T., et al., *A revised airway epithelial hierarchy includes CFTR-expressing ionocytes.* Nature, 2018. **560**(7718): p. 319-324.

25. La Manno, G., et al., *Molecular diversity of midbrain development in mouse, human, and stem cells.* Cell, 2016. **167**(2): p. 566-580. e19.

26. Regev, A., et al., *The human cell atlas white paper.* arXiv preprint arXiv:1810.05192, 2018.

27. Stewart, B.J., et al., *Spatiotemporal immune zonation of the human kidney.* Science, 2019. **365**(6460): p. 1461-1466.

28. Zilionis, R., et al., *Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species.* Immunity, 2019. **50**(5): p. 1317-1334. e10.

29. Hashimshony, T., et al., *CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq.* Genome biology, 2016. **17**(1): p. 77.

30. Hashimshony, T., et al., *CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification.* Cell reports, 2012. **2**(3): p. 666-673.

31. Jaitin, D.A., et al., *Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types.* Science, 2014. **343**(6172): p. 776-779.

32. Islam, S., et al., *Quantitative single-cell RNA-seq with unique molecular identifiers.* Nature methods, 2014. **11**(2): p. 163.

33. Sasagawa, Y., et al., *Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity.* Genome biology, 2013. **14**(4): p. 3097.

34. Shishkin, A.A., et al., *Simultaneous generation of many RNA-seq libraries in a single reaction.* Nature methods, 2015. **12**(4): p. 323-325.

35. Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells.* Nature communications, 2017. **8**(1): p. 1-12.

36. Gierahn, T.M., et al., *Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput.* Nature methods, 2017. **14**(4): p. 395-398.

37. Organization, W.H., *Global tuberculosis report 2019.* 2019.

38. Nunes-Alves, C., et al., *In search of a new paradigm for protective immunity to TB.* Nature Reviews Microbiology, 2014. **12**(4): p. 289-299.

39. Pagán, A.J. and L. Ramakrishnan, *The formation and function of granulomas.* Annual review of immunology, 2018. **36**: p. 639-665.

40. Cronan, M.R., et al., *Macrophage epithelial reprogramming underlies mycobacterial granuloma formation and promotes infection.* Immunity, 2016. **45**(4): p. 861-876.

41. Marakalala, M.J., et al., *Inflammatory signaling in human tuberculosis granulomas is spatially organized.* Nature medicine, 2016. **22**(5): p. 531.

42. Carow, B., et al., *Spatial and temporal localization of immune transcripts defines hallmarks and diversity in the tuberculosis granuloma.* Nature communications, 2019. **10**(1): p. 1-15.

43. Russell, D.G., et al., *Foamy macrophages and the progression of the human tuberculosis granuloma.* Nature immunology, 2009. **10**(9): p. 943-948.

44. Rohde, K.H., R.B. Abramovitch, and D.G. Russell, *Mycobacterium tuberculosis invasion of macrophages: linking bacterial gene expression to environmental cues.* Cell host & microbe, 2007. **2**(5): p. 352-364.

45. Martinez, F.O., L. Helming, and S. Gordon, *Alternative activation of macrophages: an immunologic functional perspective.* Annual review of immunology, 2009. **27**: p. 451-483.

46. Murray, P.J., et al., *Macrophage activation and polarization: nomenclature and experimental guidelines.* Immunity, 2014. **41**(1): p. 14-20.

47. Brodbeck, W.G. and J.M. Anderson, *Giant cell formation and function.* Current opinion in hematology, 2009. **16**(1): p. 53.

48. McNally, A.K. and J.M. Anderson, *Macrophage fusion and multinucleated giant cells of inflammation*, in *Cell Fusion in Health and Disease*. 2011, Springer. p. 97-111.

49. Herrtwich, L., et al., *DNA damage signaling instructs polyploid macrophage fate in granulomas.* Cell, 2016. **167**(5): p. 1264-1280. e18.

50. Cheng, S.-C., et al., *mTOR-and HIF-1α–mediated aerobic glycolysis as metabolic basis for trained immunity.* Science, 2014. **345**(6204): p. 1250684.

51. Arts, R.J., et al., *BCG vaccination protects against experimental viral infection in humans through the induction of cytokines associated with trained immunity.* Cell host & microbe, 2018. **23**(1): p. 89-100. e5.

52. Kaufmann, E., et al., *BCG educates hematopoietic stem cells to generate protective innate immunity against tuberculosis.* Cell, 2018. **172**(1-2): p. 176-190. e19.

53. Yao, Y., et al., *Induction of autonomous memory alveolar macrophages requires T cell help and is critical to trained immunity.* Cell, 2018. **175**(6): p. 1634-1650. e17.

54. Huang, L., et al., *Growth of Mycobacterium tuberculosis in vivo segregates with host macrophage metabolism and ontogeny.* Journal of Experimental Medicine, 2018. **215**(4): p. 1135-1152.

55. Braverman, J., et al., *HIF-1α is an essential mediator of IFN-γ–dependent immunity to Mycobacterium tuberculosis.* The Journal of Immunology, 2016. **197**(4): p. 1287-1297.

56. Knight, M., et al., *Lipid droplet formation in Mycobacterium tuberculosis infected macrophages requires IFN-γ/HIF-1α signaling and supports host defense.* PLoS pathogens, 2018. **14**(1): p. e1006874.

57. Lin, P.L., et al., *CD4 T cell depletion exacerbates acute Mycobacterium tuberculosis while reactivation of latent infection is dependent on severity of tissue depletion in cynomolgus macaques.* AIDS research and human retroviruses, 2012. **28**(12): p. 1693-1702.

58. Gideon, H.P., et al., *Variability in tuberculosis granuloma T cell responses exists, but a balance of pro-and anti-inflammatory cytokines is associated with sterilization.* PLoS pathogens, 2015. **11**(1).

59. Flynn, J.L., et al., *An essential role for interferon gamma in resistance to Mycobacterium tuberculosis infection.* The Journal of experimental medicine, 1993. **178**(6): p. 2249-2254.

60. Harris, J., et al., *T helper 2 cytokines inhibit autophagic control of intracellular Mycobacterium tuberculosis.* Immunity, 2007. **27**(3): p. 505-517.

61. Scott-Browne, J.P., et al., *Expansion and function of Foxp3-expressing T regulatory cells during tuberculosis.* The Journal of experimental medicine, 2007. **204**(9): p. 2159-2169.

62. Shafiani, S., et al., *Pathogen-specific regulatory T cells delay the arrival of effector T cells in the lung during early tuberculosis.* Journal of Experimental Medicine, 2010. **207**(7): p. 1409-1420.

63. Cooper, A.M., et al., *Disseminated tuberculosis in interferon gamma gene-disrupted mice.* The Journal of experimental medicine, 1993. **178**(6): p. 2243-2247.

64. Gutierrez, M.G., et al., *Autophagy is a defense mechanism inhibiting BCG and Mycobacterium tuberculosis survival in infected macrophages.* Cell, 2004. **119**(6): p. 753-766.

65. Redford, P.S., et al., *Enhanced protection to Mycobacterium tuberculosis infection in IL-10-deficient mice is accompanied by early and enhanced Th1 responses in the lung.* European journal of immunology, 2010. **40**(8): p. 2200-2210.

66. Sallin, M.A., et al., *Th1 differentiation drives the accumulation of intravascular, non-protective CD4 T cells during tuberculosis.* Cell reports, 2017. **18**(13): p. 3091-3104.

67. Egen, J.G., et al., *Macrophage and T cell dynamics during the development and disintegration of mycobacterial granulomas.* Immunity, 2008. **28**(2): p. 271-284.

68. Egen, J.G., et al., *Intravital imaging reveals limited antigen presentation and T cell effector function in mycobacterial granulomas.* Immunity, 2011. **34**(5): p. 807-819.

69. Beattie, L., et al., *Dynamic imaging of experimental Leishmania donovani-induced hepatic granulomas detects Kupffer cell-restricted antigen presentation to antigen-specific CD8+ T cells.* PLoS pathogens, 2010. **6**(3).

70. Harari, A., et al., *Dominant TNF-α+ Mycobacterium tuberculosis–specific CD4+ T cell responses discriminate between latent infection and active disease.* Nature medicine, 2011. **17**(3): p. 372.

71. Lin, P.L., et al., *Tumor necrosis factor neutralization results in disseminated disease in acute and latent Mycobacterium tuberculosis infection with normal granuloma structure in a cynomolgus macaque model.* Arthritis & Rheumatism: Official Journal of the American College of Rheumatology, 2010. **62**(2): p. 340-350.

72. Saunders, B., H. Briscoe, and W. Britton, *T cell-derived tumour necrosis factor is essential, but not sufficient, for protection against Mycobacterium tuberculosis infection.* Clinical & Experimental Immunology, 2004. **137**(2): p. 279-287.

73. Kindler, V., et al., *The inducing role of tumor necrosis factor in the development of bactericidal granulomas during BCG infection.* Cell, 1989. **56**(5): p. 731-740.

74. Gardam, M.A., et al., *Anti-tumour necrosis factor agents and tuberculosis risk: mechanisms of action and clinical management.* The Lancet infectious diseases, 2003. **3**(3): p. 148-155.

75. Gopal, R., et al., *Unexpected role for IL-17 in protective immunity against hypervirulent Mycobacterium tuberculosis HN878 infection.* PLoS pathogens, 2014. **10**(5).

76. Khader, S.A., et al., *IL-23 and IL-17 in the establishment of protective pulmonary CD4+ T cell responses after vaccination and during Mycobacterium tuberculosis challenge.* Nature immunology, 2007. **8**(4): p. 369-377.

77. Wozniak, T.M., et al., *Mycobacterium bovis BCG-specific Th17 cells confer partial protection against Mycobacterium tuberculosis infection in the absence of gamma interferon.* Infection and immunity, 2010. **78**(10): p. 4187-4194.

78. Khader, S.A., et al., *IL-23 compensates for the absence of IL-12p70 and is essential for the IL-17 response during tuberculosis but is dispensable for protection and antigen-*

specific IFN-γ responses if IL-12p70 is available. The Journal of Immunology, 2005. **175**(2): p. 788-795.

79. Ardain, A., et al., *Group 3 innate lymphoid cells mediate early protective immunity against tuberculosis.* Nature, 2019. **570**(7762): p. 528-532.

80. Modlin, R.L., *Th1-Th2 paradigm: insights from leprosy.* Journal of Investigative Dermatology, 1994. **102**(6): p. 828-832.

81. Ochoa, M.-T., et al., *T-cell release of granulysin contributes to host defense in leprosy.* Nature medicine, 2001. **7**(2): p. 174-179.

82. Saini, C., V. Ramesh, and I. Nath, *CD4+ Th17 cells discriminate clinical types and constitute a third subset of non Th1, Non Th2 T cells in human leprosy.* PLoS neglected tropical diseases, 2013. **7**(7).

83. Saini, C., et al., *Leprosy reactions show increased Th17 cell activity and reduced FOXP3+ Tregs with concomitant decrease in TGF-β and increase in IL-6.* PLoS neglected tropical diseases, 2016. **10**(4).

84. Chiaramonte, M.G., et al., *IL-13 is a key regulatory cytokine for Th2 cell-mediated pulmonary granuloma formation and IgE responses induced by Schistosoma mansoni eggs.* The Journal of Immunology, 1999. **162**(2): p. 920-930.

85. Moore, J., et al., *Functional complexity of the Leishmania granuloma and the potential of in silico modeling.* Frontiers in immunology, 2013. **4**: p. 35.

86. Pham, T.H., et al., *Salmonella-driven polarization of granuloma macrophages antagonizes TNF-mediated pathogen restriction during persistent infection.* Cell Host & Microbe, 2020. **27**(1): p. 54-67. e5.

87. Wu, S.J., et al., *Nodular tertiary syphilis mimicking granuloma annulare.* Journal of the American Academy of Dermatology, 2000. **42**(2): p. 378-380.

88. Hoang, M.P., W.A. High, and K.H. Molberg, *Secondary syphilis: a histologic and immunohistochemical evaluation.* Journal of cutaneous pathology, 2004. **31**(9): p. 595-599.

89. Molina-Ruiz, A.M. and L. Requena, *Foreign body granulomas.* Dermatologic clinics, 2015. **33**(3): p. 497-523.

90. Rudolph, C.M., et al., *Foreign body granulomas due to injectable aesthetic microimplants.* The American journal of surgical pathology, 1999. **23**(1): p. 113-117.

91. Veiseh, O., et al., *Size-and shape-dependent foreign body immune response to materials implanted in rodents and non-human primates.* Nature materials, 2015. **14**(6): p. 643-651.

92. Agostini, C., F. Adami, and G. Semenzato, *New pathogenetic insights into the sarcoid granuloma.* Current opinion in rheumatology, 2000. **12**(1): p. 71-76.

93. Linke, M., et al., *Chronic signaling via the metabolic checkpoint kinase mTORC1 induces macrophage granuloma formation and marks sarcoidosis progression.* Nature immunology, 2017. **18**(3): p. 293-302.

94. Wakefield, A.J., et al., *Granulomatous vasculitis in Crohn's disease.* Gastroenterology, 1991. **100**(5): p. 1279-1287.

95. Gioffredi, A., et al., *Eosinophilic granulomatosis with polyangiitis: an overview.* Frontiers in immunology, 2014. **5**: p. 549.

96. Krupa, W.M., et al., *Trapping of misdirected dendritic cells in the granulomatous lesions of giant cell arteritis.* The American journal of pathology, 2002. **161**(5): p. 1815-1823.

97. Thirunavukkarasu, S. and S.A. Khader, *Advances in cardiovascular disease lipid research can provide novel insights into Mycobacterial Pathogenesis.* Frontiers in cellular and infection microbiology, 2019. **9**: p. 116.

98. Sakthivel, P. and D. Bruder, *Mechanism of granuloma formation in sarcoidosis.* Current opinion in hematology, 2017. **24**(1): p. 59-65.

99. Bergeron, A., et al., *Cytokine patterns in tuberculous and sarcoid granulomas: correlations with histopathologic features of the granulomatous response.* The Journal of Immunology, 1997. **159**(6): p. 3034-3043.

100. ten Berge, B., et al., *Increased IL-17A expression in granulomas and in circulating memory T cells in sarcoidosis.* Rheumatology, 2012. **51**(1): p. 37-46.

101. Ramstein, J., et al., *IFN-γ–producing T-helper 17.1 cells are increased in sarcoidosis and are more prevalent than T-helper type 1 cells.* American journal of respiratory and critical care medicine, 2016. **193**(11): p. 1281-1291.

102. Villon, M.L.Z., J.A.L. de la Rocha, and L.R. Espinoza, *Takayasu arteritis: recent developments.* Current rheumatology reports, 2019. **21**(9): p. 45.

103. Peled, M. and E.A. Fisher, *Dynamic aspects of macrophage polarization during atherosclerosis progression and regression.* Frontiers in immunology, 2014. **5**: p. 579.

104. Nus, M., et al., *Endothelial Jag1-RBPJ signalling promotes inflammatory leucocyte recruitment and atherosclerosis.* Cardiovascular research, 2016. **112**(2): p. 568-580.

105. Erbel, C., et al., *Inhibition of IL-17A attenuates atherosclerotic lesion development in apoE-deficient mice.* The Journal of Immunology, 2009. **183**(12): p. 8167-8175.

106. Mehra, S., et al., *Transcriptional reprogramming in nonhuman primate (rhesus macaque) tuberculosis granulomas.* PloS one, 2010. **5**(8).

107. Zak, D.E., et al., *A blood RNA signature for tuberculosis disease risk: a prospective cohort study.* The Lancet, 2016. **387**(10035): p. 2312-2322.

108. Gideon, H.P., et al., *Early whole blood transcriptional signatures are associated with severity of lung inflammation in cynomolgus macaques with Mycobacterium tuberculosis infection.* The Journal of Immunology, 2016. **197**(12): p. 4817-4828.

109. Berry, M.P., et al., *An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis.* Nature, 2010. **466**(7309): p. 973-977.

110. Ottenhoff, T.H., et al., *Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis.* PloS one, 2012. **7**(9).

111. Maertzdorf, J., et al., *Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis.* PloS one, 2011. **6**(10).

112. Mehra, S., et al., *Transcriptional reprogramming in nonhuman primate (rhesus macaque) tuberculosis granulomas.* PloS one, 2010. **5**(8): p. e12266.

113. Chowdhury, R.R., et al., *A multi-cohort study of the immune factors associated with M. tuberculosis infection outcomes.* Nature, 2018. **560**(7720): p. 644-648.

114. Capuano, S.V., et al., *Experimental Mycobacterium tuberculosis infection of cynomolgus macaques closely resembles the various manifestations of human M. tuberculosis infection.* Infection and immunity, 2003. **71**(10): p. 5831-5844.

115. Flynn, J.L., et al., *Immunology studies in non-human primate models of tuberculosis.* Immunological reviews, 2015. **264**(1): p. 60-73.

116. Lin, P.L., et al., *Early events in Mycobacterium tuberculosis infection in cynomolgus macaques.* Infection and immunity, 2006. **74**(7): p. 3790-3803.

117. Lin, P.L., et al., *Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing.* 2014. **20**(1): p. 75.

118. Lin, P.L., et al., *Quantitative comparison of active and latent tuberculosis in the cynomolgus macaque model.* Infection and immunity, 2009. **77**(10): p. 4631-4642.

119. Cadena, A.M., et al., *Concurrent infection with Mycobacterium tuberculosis confers robust protection against secondary infection in macaques.* PLoS pathogens, 2018. **14**(10): p. e1007305.

120. Hughes, T.K., et al., *Highly efficient, massively-parallel single-cell RNA-seq reveals cellular states and molecular features of human skin pathology.* bioRxiv, 2019: p. 689273.

121. Darrah, P.A., et al., *Prevention of tuberculosis in macaques after intravenous BCG immunization.* Nature, 2020. **577**(7788): p. 95-102.

Chapter 2: Seq-Well: A Portable, Low-cost Platform for Single-Cell RNA-Seq of Low-Input Samples

Modified from: Seq-Wel: portable, low-cost RNA sequencing of single cells at high-throughput

Todd M. Gierahn*, Marc H. Wadsworth II* Travis K. Hughes*, Bryan D. Bryson, Andrew Butler, Rahul Satija, Sarah Fortune, J. Christopher Love, and Alex K. Shalek

*These authors contributed equally.

Authorship Attributions

TMG, MHW, TKH, JCL, and AKS developed the concepts and designed the study. TG, MHW, TKH, and BDB performed the experiments. All authors analyzed and interpreted the data. TMG, MHW, TKH, JCL, and AKS wrote the manuscript with feedback from all authors.

Introduction

The emergence of single-cell genomics has enabled new strategies for identifying the cellular and molecular drivers of biological phenomena [1-19]. Patterns in genome-wide mRNA expression measured by single-cell RNA-seq (scRNA-seq) can be leveraged to uncover distinct cell types, states and circuits within cell populations and tissues [1-5, 9-13]. To inform our understanding of healthy and diseased behaviors and eventually guide precision diagnostics and therapeutics, we need broadly applicable scRNA-seq methods that are easy to use and enable high-throughput studies of cellular phenotypes, particularly for low-input (≤104 cells) samples such as clinical specimens.

Typically, scRNA-seq involves isolating and lysing individual cells, then independently reverse transcribing and amplifying their mRNAs before generating barcoded libraries that are pooled for sequencing. Although manual picking [2, 5, 8], FACS sorting [1, 3, 4] or integrated microfluidic circuits [7, 9, 10] can isolate single cells, these approaches are constrained in scale by cost, time and labor. Recently developed massively parallel methods assign unique bar-codes to each cell's mRNAs during reverse transcription, enabling ensemble processing while retaining single-cell resolution. These techniques typically yield single-cell libraries of lower complexity, but higher throughput reduces the impact of the technical and intrinsic noise associated with each cell in analyses [11, 12]. The most commonly used approach relies on microfluidic devices to generate reverse-emulsion droplets that couple single cells with uniquely barcoded mRNA capture beads [11, 12]. Droplet-based techniques, however, can have inefficiencies in

encapsulation, introduce technical noise through differences in cell lysis time and require specialized equipment—limiting where, when and at what scale scRNA-seq can be performed.

One alternative to droplets is to use arrays of subnanoliter wells loaded by gravity, which reduces the need for peripheral equipment, decreases dead volumes and facilitates parallelization. As a proof of principle, cells and beads have been co-confined in nanowell arrays to perform targeted single-cell transcriptional profiling [13], yet the absence of a seal significantly impairs capture efficiency and increases cross-contamination (**Figure A1.1**). Nanowells have also been combined with microfluidic channels that facilitate oil-based single-cell isolation via fluid exchange [14]. Nevertheless, this design limits buffer exchange and necessitates integrated temperature and pressure controllers, impacting ease of use and portability [15]. Semiporous-membrane-covered nanowells have been used to link pairs of specific transcripts from single cells [16]; however, this approach used many beads per well, precluding the creation of unique single-cell libraries, and transcript capture and sealing efficiency were not addressed.

Results

To overcome these challenges, we developed Seq-Well, a simple, portable platform for massively parallel scRNA-seq (**Figure A1.2**). Seq-Well confines single cells and bar-coded poly(dT) mRNA capture beads in a PDMS array of ~86,000 subnanoliter wells. Wells accommodate only one bead, enabling single-bead loading efficiencies of ~95% (**Figure 2.1 *and* Figure A1.3A**). A simplified cell-loading scheme in turn permits capture efficiencies of around 80% (**Methods, Figure A1.3C**), with a dual occupancy rate that can be tuned by adjusting the number of cells loaded and visualized before processing (**Figure A1.3C**).

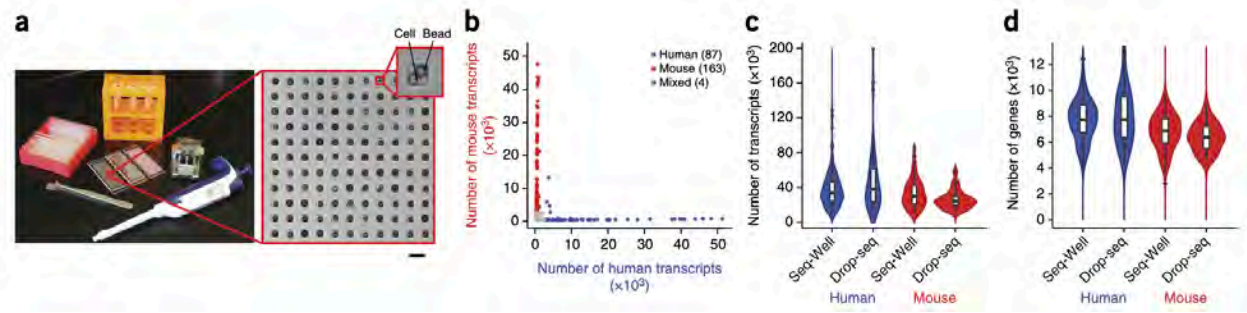


Figure 2.1 | Seq-Well: a portable, low-cost platform for high-throughput single-cell RNA-seq of low-input samples. (a) Equipment and array used to capture and lyse cells, respectively. Scale bar, 100 μm. **(b)** Sequencing a mix of human (HEK293) and mouse (NIH/3T3) cells reveals distinct transcript mapping and single-cell resolution. Cells with >2,000 human and <1,000 mouse transcripts are labeled as human, and cells with >2,000 mouse and <1,000 human transcripts are labeled as mouse. Of the 254 cells identified, 4 (1.6%) had a mixed phenotype. (c,d) Number of transcripts **(c)** and genes **(d)** detected in single-cell libraries generated by Seq-Well or Drop-seq (ref. 12; center-line: median; limits, first and third quartile; whiskers, ±1.5 IQR; points; values, >1.5 IQR). Using Seq-Well (Drop-seq), an average of 37,878 (48,543) transcripts or 6,927 (7,175) genes were detected among human HEK cells (n = 159 for Seq-Well; n = 48 for Drop-seq); and an average of 33,586 (26,700) transcripts or 6,113 (5,753) genes were detected among mouse 3T3 cells (n = 172 for Seq-Well; n = 27 for Drop-seq) at an average depth of 164,238 (797,915) reads per HEK cell and 152,488 (345,117) reads per 3T3 cell.

A key unique feature of Seq-Well is the use of selective chemical functionalization to facilitate reversible attachment of a semipermeable polycarbonate membrane (10-nm pore size) in physiologic buffers. This enables rapid solution exchange for efficient cell lysis but traps biological macromolecules for improved transcript capture and reduced cross-contamination (**Figure A1.4A**). The array's three-layer surface functionalization comprises an aminosilane base [20] crosslinked to a bifunctional poly(glutamate)–chitosan top via a p-phenylene diisothiocyante intermediate (**Figure A1.4; Methods**). In the outer layer, poly(glutamate) at the inner nanowell surfaces prevents nonspecific binding of mRNAs, while chitosan on the array's top sur-face encourages efficient sealing to the membrane (**Figure A1.4B-C; Methods**). To test sealing and buffer exchange, we monitored the fluorescence of dye-labeled, cell-bound antibodies before and after adding a guanidinium-based lysis buffer. We observed rapid diffusion of the antibodies throughout the wells within 5 minutes of buffer addition and—unlike in unsealed or previously described, membrane-covered BSA-blocked arrays [16]— we observed little change in

fluorescent signal over 30 min, suggesting robust retention of biological macromolecules despite the use of a strong chaotrope (**Figure A1.5; Methods**).

After lysis, cellular mRNAs are captured by bead-bound poly(dT) oligonucleotides that also contain a universal primer sequence, a cell barcode and a unique molecular identifier (UMI) (**Methods**). Next, the membrane is peeled off, and the beads are removed for subsequent bulk reverse transcription, amplification, library preparation and paired-end sequencing, as previously described [12] (**Methods**). Critically, beyond a disposable array and membrane, Seq-Well only requires a pipette, a manual clamp, an oven and a tube rotator to achieve stable, barcoded single-cell cDNAs (**Figure 2.1a**), so it can be performed almost anywhere.

To assess transcript capture efficiency and single-cell resolution, we profiled a mixture of 5 × 103 human (HEK293) and 5 × 103 mouse (3T3) cells using Seq-Well. The average fraction of reads mapping to exonic regions was 77.5% (**Figure A1.6**), demonstrating high-quality libraries. Shallow sequencing from a fraction of an array revealed highly organism-specific libraries, suggesting single-cell resolution and minimal cross-contamination (**Figure 2.1B** and **Figure A1.7A-C**). In the absence of membrane sealing, by comparison, we obtained poor transcript and gene detection as well as substantial cross-contamination (**Figure A1.1**). From deeper sequencing of a fraction of a second array, we detected an average of 37,878 mRNA transcripts from 6,927 genes in HEK cells and 33,586 mRNA transcripts from 6,113 genes in 3T3 cells, comparable to a droplet-based approach using the same mRNA capture beads (Drop-seq) [12] (**Figures 2.1C-D, A1.7** and **A1.8**). Upon downsampling to match read depths, we also observed levels of transcript and gene detection consistent with those of other massively parallel bead-based scRNA-seq methods (**Figure A1.7D-G; Methods**). Moreover, bulk RNA-seq data were strongly correlated with populations constructed in silico from individual HEK cells ($R = 0.751 \pm 0.073$ to $R = 0.983 \pm 0.0001$ for populations of 1–1,000 single cells, respectively), suggesting representative cell and transcript sampling (**Figure A1.9; Methods**). To examine Seq-Well's ability to resolve populations of cells in complex primary samples, we loaded human peripheral

blood mononuclear cells (PBMCs) into arrays in triplicate before beads, allowing us to perform on-array multicolor imaging cytometry (**Figure 2.2A-B**; **Methods**).

Sequencing one-third of the beads recovered from each array yielded 3,694 high-quality single-cell libraries (**Methods**). Unsupervised graph-based clustering revealed unique subpopulations corresponding to major PBMC types (**Figures 2.2B**, **A1.10, A1.11** and **A1.12; Methods**). Each array yielded similar subpopulation frequencies (**Figure 2.2C**), with detection efficiencies comparable to those of other massively parallel technologies (**Figure A1.13**). The proportion of each subpopulation determined by sequencing also matched on-array immunophenotyping results (**Figure 2.2A-B**).

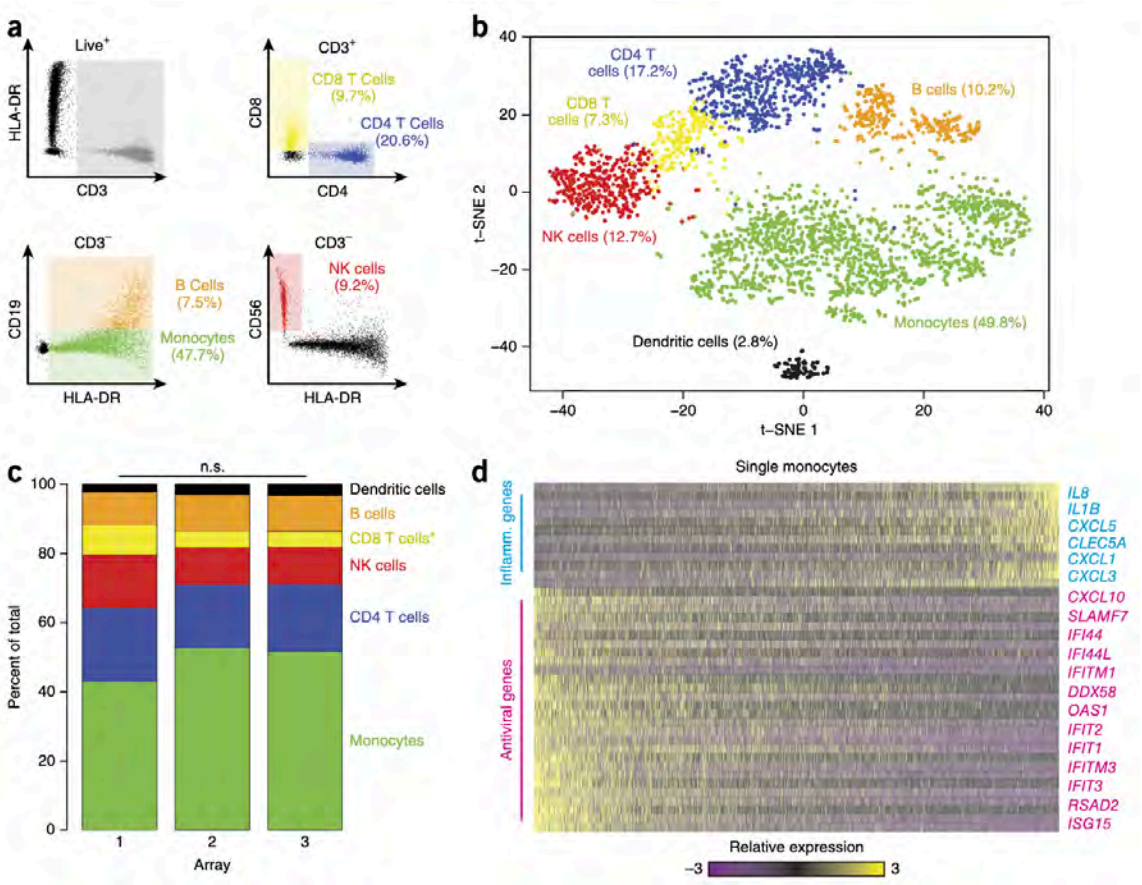


Figure 2.2 | Combined image cytometry and scRNA-seq of human PBMCs. (a) Hierarchical gating scheme used to analyze PBMCs labeled with a panel of fluorescent antibodies, loaded onto three replicate arrays and imaged before bead loading (see **Online Methods**). Myeloid cells (green) were identified as the population of hCD3(−) HLA-DR(+) CD19(−) cells; B cells (orange) as the subset of hCD3(−) HLA-DR(+) CD19 (+) cells; CD4 T cells (blue) as the subset of CD3(+) CD4(+) cells; CD8 T cells (yellow) as the CD3(+) CD8(+) subset of cells; and NK cells (red) as the subset of CD3(−) HLA-DR(−) CD56(+) CD16(+) cells. **(b)**

t-SNE visualization of clusters identified among 3,694 human Seq-Well PBMCs single-cell transcriptomes recovered from the imaged array and the two additional arrays (see **Online Methods**). Clusters (subpopulations) are labeled based on annotated marker gene (supplementary fig. 0). **(c)** Distribution of transcriptomes captured on each of the biological replicate arrays, run on separate fractions of the same set of PBMCs. No shifts are statistically significant (n.s. = not significant; see **Online Methods**) except for a slightly elevated fraction of CD8 T cells in array 1 (*, P = 1.0 × 10−11; Chi-square test, Bonferroni corrected). **(d)** Relative expression level of a set of inflammatory and antiviral genes among cells identified as monocytes. Inflamm., inflammatory.

Critically, sequencing provided additional information; in addition to resolving dendritic cells from monocytes (**Figure 2.2B**), we found significant variation among the monocytes (captured in PC3) due to differential expression of inflammatory and antiviral gene programs (**Figure 2.2D**) [1, 3]. Our results show that characterizing a sample in two ways using a single platform increases the amount of information that can be extracted from a precious specimen, while allowing analysis of one measurement to be interpreted in the context of the other.

Finally, to test the portability of Seq-Well, we profiled primary human macrophages exposed to Mycobacterium tuberculosis (H37Rv) in a BSL3 facility (**Methods**). In total, we recovered 14,218 macrophages (from a total of 40,000 loaded across experiments) with greater than 1,000 mapped transcripts from an M. tuberculosis-exposed and an unexposed array. Unsupervised analysis of 4,638 cells with greater than 5,000 transcripts per cell revealed five distinct clusters (**Figures 2.3A-B** and **A1.14A-B**). Two clusters had lower transcript capture and high mitochondrial gene expression (suggestive of low-quality libraries)[17] and were removed; the remaining three (2,560 cells) were identified in both the exposed and unexposed samples (**Figures 2.3A, A1.14C-D** and **A1.15**), and they likely represent distinct sub-phenotypes present in the initial culture.

We next examined common and cluster-specific gene enrichments (**Methods**). Although clusters 1 and 3 did not present strong stimulation-independent enrichments, cluster 2 uniquely expressed several genes associated with metabolism. Intriguingly, within each cluster we observed pronounced shifts in gene expression in response to M. tuberculosis (**Figure 2.3C**; **Methods**), with common enrichments for gene sets previously observed in response to

intracellular infection, LPS stimulation and activation of TLR7/8 receptor. Cluster 1 uniquely displayed stimulation-induced shift in several genes associated with cell growth, cluster 3 in transcripts associated with hypoxia, and cluster 2 (again) in genes linked to metabolism. Overall, these data suggest that basal cellular heterogeneity may influence ensemble M. tuberculosis responses. Equally important, they demonstrate Seq-Well's ability to acquire large numbers of single-cell transcriptomes in challenging experimental environments.

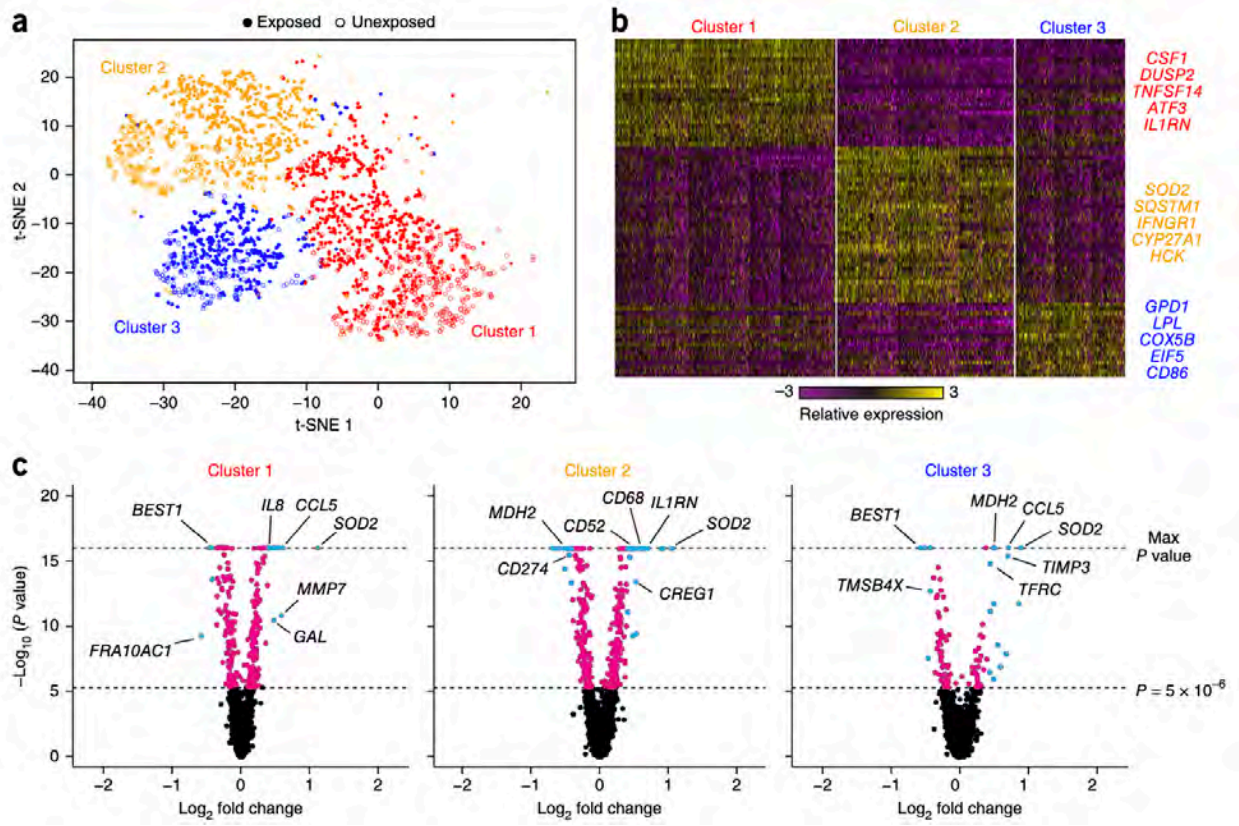


Figure 2.3 | Sequencing of TB-exposed macrophages in a BSL3 facility using Seq-Well. (a) t-SNE visualization of single-cell clusters identified among 2,560 macrophages (1,686 exposed, solid circles; 874 unexposed, open circles) generated using five principal components across 377 variable genes (see Online Methods). **(b)** Marker genes for the three phenotypic clusters of macrophages highlighted in a. **(c)** Differential expression between exposed and unexposed macrophages within each cluster showing genes enriched in cells exposed to M. tuberculosis. Cyan, genes with P values less than 5.0 × 10−6 (threshold for statistical significance, determined by a likelihood ratio test) and absolute log2 fold changes greater than 0.4 (threshold used for differential expression). Magenta, genes with P values less than 5.0 × 10−6 but absolute log2 fold changes less than 0.4. Black, remaining genes.

Conclusions

In conclusion, Seq-Well is a robust platform for scalable, single-cell transcriptomics applicable to almost any cellular suspension for which a reference genome or transcriptome exists. The technique is inexpensive, user friendly, portable, and efficient; it enables scRNA-seq to accelerate scientific and clinical discovery even when working with limited samples. Furthermore, the ability to measure protein secretion and cell-surface expression on the same platform [18, 19] foreshadows multi-omic single-cell measurements at scale.

Methods and Materials

PDMS Surface Functionalization

The PDMS arrays were functionalized as follows (**Figure A1.2**):

1. They were treated with air plasma (Harrick Plasma, PDC-32G) at maximum power for 5 minutes under mild vacuum (~ 10 mTorr).
2. They were then submerged for 10 minutes in 0.05% (3-Aminopropyl)triethoxysilane (APTES) in 95% ethanol for silanization.
3. The silanized arrays were spun dry at 500 rpm for 1 minute and baked at 90°C for 30 minutes.
4. Once dry, the arrays were brought to room temperature and submerged in acetone for 5 minutes to solvate the microwells.
5. The arrays were then moved to a solution of 0.2% *p*-Phenylene diisothiocyante (PDITC) in a 10% Pyridine and 90% Dimethylformamide (DMF) mixture and rocked at room temperature for 2 hours to activate the surface with amine-reactive isothiocyanate groups.
6. The arrays were briefly washed twice with DMF to remove residual PDITC, and twice with acetone.
7. The arrays were rocked for 20 minutes in second acetone wash at room temperature, and baked at 90°C for 2 hours.

8. After baking, the arrays were removed from the oven, brought to room temperature, submerged in a solution of 0.2% chitosan in 100 mM sodium acetate solution (pH 6.5) and incubated at 37°C for 2 hours to facilitate reaction of chitosan with the PDITC top surface.
9. The arrays were then washed four times with distilled water to remove excess chitosan.
10. The arrays were next submerged in 10 ug/mL poly-glutamic acid in a 100 mM sodium carbonate buffer (pH 10.0), placed in a vacuum chamber under mild (house) vacuum, and rocked overnight to facilitate solvation of the wells and reaction of the poly(glutamic) acid to the PDITC on the inner well surfaces.
11. The following morning, arrays were quenched through the addition of 0.01 volumes of 1M Tris pH 8.5 followed by rocking for 3 hours at room temperature.
12. Afterwards, the arrays were incubated at 4°C for 24 hours before use.

Bead synthesis

Barcoded oligo-dT beads (as described in Macosko *et al.* [12]) were purchased from Chemgenes (Wilmington, Massachusetts, USA; cat. no. MACOSKO-2011-10) at 10 umol scale (~100 arrays). Bead functionalization and reverse phosphoramidite synthesis was performed by Chemgenes Corporation using Toyopearl HW-65S resin (30 micron mean particle diameter) obtained from Tosoh Biosciences (cat. no. 19815). Surface hydroxyls were reacted with a PEG derivative to obtain an 18-carbon linker to serve as a support for oligo synthesis. Reverse-direction phosphoramidite synthesis was performed using an Expedite 8909 DNA/RNA synthesizer at 10 micromole scale with a coupling time of 3 min. Initially, a conserved PCR handle was synthesized followed by 12 rounds of split and pool synthesis to generate 16,777,216 unique barcode sequences. Addition of an 8-mer random sequence was performed to generate unique molecular identifiers (UMIs) on each capture oligo. Finally, a 30-mer poly-dT capture sequence was synthesized to enable capture of polyadenylated mRNA species.

Imaging differential surface functionalization

Differential labeling of the top and inner well surfaces was visualized by substituting 1 ug/mL PE–streptavidin for chitosan and 1 ug/mL AlexaFluor488-Streptavidin for the polyglutamate in the standard functionalization protocol (**Figure A1.3**). Carboxylation of the inner well surfaces was visualized by treating the functionalized array with 100 µg/mL EDC/10 µg/mL NHS MES (pH 6.0) solution for 10 min, washing twice with MES buffer, once with sodium borate buffer (pH 8.5), and incubating overnight with 1 µg/mL Alexa-Fluor 568-labeled antibody. Arrays were washed three times with phosphate-buffered saline (PBS) and imaged using Alexa Fluor 568 channel.

Visualizing lysate retention (imaging)

PBMCs were labeled with αCD45–AF647 (BioLegend 304020, diluted 1:20). Cells were washed and loaded onto two arrays previously blocked with 1% BSA solution for 30 min and one array functionalized with chitosan as described above. A polycarbonate membrane was attached to the chitosan-functionalized array as described above. The array was submerged in PBS and imaged for AF647 fluorescence to identify wells containing cells. The BSA-blocked arrays were imaged before membrane attachment because the membrane would detach when submerged in media. After imaging, a plasma-treated polycarbonate membrane was attached to one of the BSA-blocked arrays as described [16]. Briefly, the membrane was placed on the array with forceps, and all excess media were aspirated from the array. The open BSA-blocked array and the chitosan array were submerged in 5 mL of 5 M GCTN lysis buffer. 500 µL of lysis buffer was placed on the top of membrane attached to the BSA-blocked array as described [16]. 5 and 30 minutes later, 100 block positions were imaged on each array, encompassing 12,100 individual wells. Automated image analysis software was used to background subtract each image, identify cell and well locations and extract AF647 signal intensity of the cells and the well volumes (Figure S4).

Membrane Attachment

Once the arrays were loaded, the membrane is attached to seal the microwells. All excess media was aspirated from the array surface. Quickly, the glass slide underneath the membrane was used to lift the membrane out of the PBS. Excess PBS was blotted off and the glass slide was flipped over with the membrane sticking to the bottom of the glass slide. Forceps were used to slide the membrane so a small section was hanging off the glass slide. This part of the membrane was then touched to the array surface. A second glass slide was used to hold the membrane in place while the first slide was slid across the array surface, depositing the membrane onto the array. After the whole membrane was deposited, a glass slide was placed on top of the membrane and the array sandwich was placed in a manual hybridization clamp (Agilent). The clamp was incubated at 37C for 30 min to complete membrane binding. After 30 minutes, the array was transferred to a 4-well dish and submerged in 5 mL 5 M guanidine thiocyanate (GCTN). The top slide released in 1-2 minutes, leaving the open array sealed with the permeable membrane.

Cell Lysis And mRNA Hybridization

After cover slide removal, the arrays were transferred to 5 mL of lysis buffer (5M GTCN, 1% 2-mercaptoethanol, 1mM EDTA, and 0.1% Sarkosyl in 1X PBS, pH 6.0) and allowed to rock at room temperature for 20 minutes. The lysis buffer was then aspirated and arrays were washed once with 5 mL of hybridization buffer (2M NaCl, 1x PBS, 0.5% Tween20 (pH 7.5)). The hybridization buffer was aspirated and replaced with another 5 mL, followed by rocking for 40 minutes at room temperature.

Removal Of Beads

Following hybridization of mRNA transcripts to the beads, the arrays were washed once in Wash 1 solution (2 M NaCl, 3 mM MgCl2, 20 mM Tris, 0.5% Tween20 (pH 8.0), RNAse

Inhibitor). This solution was aspirated and replaced with another 5 mL of wash 1 buffer. The membrane was then peeled from the surface of the array using forceps. Beads were removed from microwell arrays by either centrifugation in a 50mL conical tube containing Wash 1 (3000 x g, 5 min), or by scraping the array with a microscope slide.

Reverse Transcription

Recovered beads were transferred to a 1.5 mL tube, pelleted and washed with 250 uL of 5X Maxima buffer (Life Technologies). After centrifugation and aspiration of the supernatant, the beads were resuspended in 200 uL of RT mix consisting of 80 uL of RNAse-free water, 40 uL of Maxima 5X RT Buffer, 40 uL of 20% Ficoll PM-400, 20 uL of 10mM dNTPs, 5 uL of RNase Inhibitor, 5 ul of 100 uM Template Switch Oligo, and 10 uL of 200U/uL Maxima H-RT. The beads were then incubated at room temperature for 30 minutes with end-over-end rotation, followed by 90 minutes at 50°C with rotation.

ExoI Digestion

Following reverse transcription, the beads were washed once with 1 mL of Tris-EDTA (TE) buffer (pH 8.0) + 0.5% SDS (Sigma) and twice with TE (pH 8.0) + 0.01% Tween20 (Sigma). The beads were washed with 1 mL of 10mM Tris-HCl (pH 8.0) and resuspended in 200 uL of ExoI solution, which contained 170 uL water, 20 uL of 10X ExoI buffer, and 10 uL of 20,000 U/mL ExoI (NEB Cat no. M0293). Reaction was incubated at 37°C for 45 minutes with end-over-end rotation.

PCR Amplification

The beads were washed once with TE (pH 8.0) + 0.5% SDS, twice with TE (pH 8.0) + 0.01% Tween20, once with 1 mL of water, and finally resuspended in 500 uL of water. A 10 □L aliquot of beads was resuspended in 20% PEG/NaCl solution and counted using a hemocytometer. For each PCR reaction, 1,500-2,000 beads were combined with 25 uL of KAPA

HiFi PCR Mastermix, 0.4 uL of 100uM ISPCR Primer, and 24.6 ul of water. PCR amplification of single-cell libraries was performed as follows: 95°C for 3 minutes; then 4 cycles of: 98°C for 20 seconds, 65°C for 45 seconds, 72C for 3 minutes; 9 cycles of: 98C for 20 seconds 67 C for 20 seconds and 72°C for 3 minutes; then a final extension at 72°C for 5 minutes. Following amplification, 5 uL from each PCR product were combined. The pooled PCR library was purified using Agencourt AMPure XP beads (Beckman Coulter Cat no. A63881) at a 0.6X volumetric ratio and quantified using Qubit Fluorometric Quantitation (Thermo Fisher).

Library Preparation And Sequencing

Tagmentation was performed on a total of 800 pg of pooled cDNA library combined with 5 uL of Amplicon Tagment enzyme and 10 uL of Nextera Tagmentation DNA (TD) Buffer and incubated at 55C for 5 minutes. Then, 5 uL of Neutralization Buffer NT were added, mixed by pipetting, and incubated at room temperature for 5 minutes. Amplification of tagmented fragments was performed by adding 15 uL of Nextera PCR Mix (Buffer NPM), 8 uL of H2O, 1 uL of 10 uM P5-SMART PCR Hybrid Oligo, and 1 uL of 10 uM Nextera N7XX oligo, where N7XX is a standard Nextera N700 adapter sequence capable of adding i7/p7 sequences. Amplification was then performed as follows: 95°C for 30 seconds and 12 cycles of: 95°C for 10 seconds, 55°C for 30 seconds, 72°C for 30 seconds; and a final extension at 72°C for 5 minutes. Tagmented and amplified sequences were purified using Agencourt AMPure XP beads at a 0.6x volumetric ratio and quantified using Qubit Fluorometric Quantitation. The size distribution of purified sequencing libraries was determined using the Agilent D1000 Screen Tape System (Agilent Genomics). The average fragment size of sequenced libraries was between 400 and 700 base pairs in length. Seq-Well libraries were all sequenced on an Illumina NextSeq500 at a final concentration of 2.4 pM. 20 bases were allocated (12 bp cell barcode and 8 bp UMI) for Read 1, which was primed using Custom Read 1 Primer [21].

Calculating bead loading efficiency

Bead loading efficiencies were determined by loading two functionalized arrays with beads as outlined above (**Figure A1.1**). Arrays were imaged in transmitted light and AF488 channel to capture bead autofluorescence. Automated image analysis was used to identify well locations and extract the 75^{th} percentile fluorescence intensity in each well. Histogram analysis of fluorescence intensities was used to identify empty wells and wells containing beads. Finally, manual review of 50 randomly selected image positions, each containing 121 nanowells, of a total of 690 positions was used to calculate the frequency of wells containing two beads.

Calculating cell loading efficiency

To calculate cell loading efficiencies and well occupancy distributions (**Figure A1.3**), HEK293 and 3T3 cells were labeled with Calcein AM (Life Technologies) and Calcein Violet (Life Technologies), respectively, per the manufacturer's recommendations. 200 µL of serial dilutions of a 1:1 mix of the cells at an estimated concentration of 1,000, 10,000 and 100,000 cells/mL were loaded in functionalized arrays in triplicate using the standard protocol. To determine the distribution of cells present in 200 µL of these solutions, the same volume of each solution was added to 12 wells of a 96-well plate. 690 array positions on each array were imaged in the transmitted light, AF488 and AF405 spectral channels. Overlapping images of each well of the 96-well plate were acquired in the same channels. Automated image analysis was used to identify well and cell locations in the array images. The overlapping images of the 96-well plate were stitched together based on *x*–*y* location of each image and analyzed in a similar manner to identify cell locations. All three dilutions were used to determine the distribution of well occupancy as a function of the number of cells loaded. The 10,000 cells/mL dilution were used to calculate cell loading efficiency.

Species-mixing experiments

Murine NIH/3T3 cells (ATCC, CRL-1658) were cultured in Dulbelcco's modified Eagle's medium (DMEM) with glutamate and supplemented with 10% fetal bovine serum (FBS) at 37 °C and 5% CO_2. Human 293T cells (ATCC, CRL-11268) were cultured at 37 °C and 5% CO_2 in DMEM with glutamate supplemented with 10% FBS. The media were removed from the culture flasks, which were then rinsed with 5 mL of 1× PBS. Cells were detached from the surface of the culture flasks by applying 3.5 mL of Trypsin-LE (Life Technologies) and incubating at room temperature for 5 min. Once cells had de-adhered, 10 mL of complete media was added, and cells were pelleted by spinning at 500× g for 10 min. Cell pellets were resuspended in 1 mL of media, and a 10 µL aliquot was used to count cells. A total of 100,000 HEK and 3T3 cells were again pelleted and resuspended in 1 mL of media. For species-mixing experiments, a total of 200 µL of a single-cell suspension containing 5,000 HEK and 5,000 NIH/3T3 cells was applied to the surface of two nanowell devices loaded with beads. In the first experiment, of the 60,000 beads collected from the array, 9,600 beads were pooled for subsequent processing and sequencing, from which we identified 254 high-quality cells with greater than 2,000 transcripts. In the second experiment, of the 25,000 beads collected from the array, 15,000 beads were pooled for subsequent processing and sequencing, from which we identified 331 high-quality cells with greater than 10,000 transcripts, greater than 2,000 genes, and greater than 90% transcript purity (i.e. >90% of transcripts from the same species). Also, as in Drop-Seq, we attempted to validate capture efficiency using ERCC spike-ins; however, this required us to load ERCCs onto the nanowell array by pipetting, which proved inefficient to properly assess capture efficiency since we could not evenly distribute ERCCs to nanowells.

HEK population experiments

HEK293 cells were cultured in RPMI supplemented with 10% FBS. A total of 10,000 HEK293 cells were applied to a Seq-Well device and scRNA-seq libraries were generated from 24,000 beads and sequenced on a NextSeq 500. For the bulk RNA-seq sample, cellular lysate

from 40,000 HEK293 cells in 200 µL of lysis buffer (5 M GTCN, 1% 2-mercaptoethanol, 1 mM EDTA, and 0.1% Sarkosyl in 1× PBS, pH 6.0) was combined with 40,000 mRNA capture beads in a PCR tube and rotated end over end for 1 h. Afterward, the beads were washed, and a population sequencing library was constructed in an identical manner to that of the single-cell Seq-Well libraries but with reads from the different bead barcodes combined into one population. *In silico* populations were created by randomly sampling 1, 10, 100 or 1,000 HEK cells from a total of 1,453 cells with greater than 3,000 transcripts obtained from a Seq-Well array. Average Pearson correlation coefficients and their s.d. were calculated between 100 randomly generated *in silico* populations for each number of cells and the bead population (**Figure A1.9**).

Human PBMC experiments

Leukocytes isolated from a leukocyte reduction filter used during platelet aphoresis were purchased from Key Biologics (Memphis, Tennessee). The cells were shipped overnight at room temperature. PBMCs were isolated from the sample using a Ficoll–Hypaque (GE) gradient, washed two times with HBSS buffer, and frozen in 90% FBS/10%DMSO in aliquots of 10^7 cells. The day before the experiment, an aliquot was thawed and rested overnight in RPMI-1640 supplemented with 10% FBS, Pen–Strep, nonessential amino acids, sodium pyruvate, and HEPES buffer (RP10) at 10^6 cells/mL in a 50 mL conical tube. Cells were counted the next day, and 5 × 10^5 cells were pelleted, resuspended in 1 mL of CellCover solution, and processed as described above.

Array loading for imaging (PBMCs)

To quantify cell surface marker protein expression levels on array (**Figure 2.2A**), PBMCs were loaded first and imaged before bead addition to avoid potential detection issues associated with bead autofluorescence. Here, cells were resuspended in cold CellCover (Anacyte), an RNA stabilization reagent, and placed at 4 °C for 1 h. Cells were spun down and resuspended in a

cocktail containing αCD45-AF647 (BioLegend; HI30), αCD3-PerCP (BioLegend; UCHT1), αCD4-PECy5.5(eBioscience; SK3), αCD56-PECy5(BD Biosciences; B159), αCD8-APCCy7 (BioLegend; RPA-T8), αHLA-DR-PECy7 (BD Biosciences; L243), and αCD19-PE (BioLegend; HIB19) with all antibodies diluted 1:20 in RP10 media and were incubated at 4 °C for 30 min. Cells were washed twice with PBS and resuspended in CellCover10 buffer (CellCover supplemented with 10% FBS and 100 mM sodium carbonate (pH 10) buffer). Functionalized arrays were washed with 5 mL of CellCover10 buffer. 2.0×10^4 cells were loaded onto the array and washed twice with CellCover10 buffer, and finally the array was placed in 5 mL CellCover. Arrays were imaged with a Zeiss AxioVision microscope with Lumencor light source and EMCCD camera. Automated imaging software was used to identify cell locations within the images and extract signal intensities in each spectral channel. To generate spillover coefficients for each fluorophore, α−mouse beads (Bangs Labs) were stained individually with each antibody using the same protocol as the cells. Images of the singly stained beads were used to generate spillover coefficients for each fluorophore that were then used to calculate the amount of each fluorophore on each cell as previously described [22]. After imaging, arrays were washed with 5 mL CellCover10 media. Barcoded beads suspended in CellCover10 media were loaded into the array through gentle agitation. Arrays were washed 3x with CellCover10 without FBS and finally washed with CellCover. Arrays were then moved on to membrane attachment.

Human monocyte isolation

Primary human monocytes were isolated from deidentified human buffy coats obtained from the Massachusetts General Hospital Blood Bank using a standard Ficoll gradient and subsequent CD14 positive selection (Stemcell Technologies). Enriched monocytes were cultured in low-adherence flasks (Corning) for 9 d with RPMI media (Invitrogen) supplemented with 10% heat-inactivated FCS (Sigma-Aldrich).

Mycobacterium tuberculosis culture

Mycobacterium tuberculosis (Mtb) H37Rv expressing the E2-Crimson fluorescent protein was grown in Difco Middlebrook 7H9 media supplemented with 10% OADC, 0.2% glycerol, 0.05% Tween-80 and Hygromycin B (50 ug/mL).

Macrophage infection and flow cytometry

The Mtb culture was pelleted by centrifugation and washed once with RPMI + 10% FCS, sonicated briefly, and filtered through a 5 μm syringe filter. Monocyte-derived macrophages (MDM) were infected at an MOI of 10 for 4 h and then washed 3× with RPMI + 10% FCS. 24 h after infection, cells were washed briefly with 1× PBS. 10× Trypsin (Life Technologies) was added, and cells were incubated briefly at 37 °C to allow for cell detachment. Detached cells were spun down and resuspended in 1× PBS supplemented with 2% FCS and 1 mM EDTA and then passed through a mesh filter to eliminate clumps. Uninfected and infected cells were sorted by flow cytometry on an Aria IIu flow cytometer. Mtb-infected cells were identified by the presence of an E2-Crimson signal above the background autofluorescence of uninfected cells.

Transcriptome alignment and barcode collapsing

Read alignment was performed as in Macosko et al. [12]. Briefly, for each NextSeq sequencing run, raw sequencing data was converted to FASTQ files using bcl2fastq2 that were demultiplexed by Nextera N700 indices corresponding to individual samples. Reads were first aligned to both HgRC19 and mm10, and individual reads were tagged according to the 12-bp barcode sequence and the 8-bp UMI contained in read 1 of each fragment. Following alignment, reads were binned and collapsed onto 12-bp cell barcodes that corresponded to individual beads using Drop-seq tools (http://mccarrolllab.com/dropseq). Barcodes were collapsed with a single-base error tolerance (Hamming distance = 1), with additional provisions for single insertions or deletions. An identical collapsing scheme (Hamming distance = 1) was then applied to UMIs to

obtain quantitative counts of individual mRNA molecules. Quality metrics are presented in (**Figures A1.5** and **A1.8**).

Data normalization

Digital gene expression matrices were obtained by collapsing filtered and mapped reads for each gene by 8-bp UMI sequences within each cell barcode. For each cell, we performed library-size normalization. UMI-collapsed gene expression values for each cell barcode were scaled by the total number of transcripts and multiplied by 10,000. Scaled expression data were then natural-log transformed before analysis using Seurat [23].

Analyzing species-mixing experiments

In the first experiment, HEK cells were identified as those barcodes with greater than 2,000 human transcripts and less than 1,000 mouse transcripts, while barcodes with greater than 2,000 mouse transcripts and less than 1,000 human transcripts were identified as 3T3 cells. Cells with fewer than 2,000 total transcripts were considered indeterminate, while any cell with greater than 5,000 total transcripts and more than 1,000 nonmouse or nonhuman transcripts was considered a multiplet (**Figure 2.1D**). In the second experiment, HEK cells were identified as those barcodes with greater than 10,000 human transcripts, greater than 2,000 human genes, and greater than 90% human transcript alignment; while barcodes with greater than 10,000 mouse transcripts, greater than 2,000 mouse genes, and greater than 90% mouse transcript alignment were identified as 3T3 cells. Cells with fewer than 10,000 total transcripts were considered indeterminate, while any cells with greater than 10,000 total transcripts and more than 1,000 nonmouse or nonhuman transcripts were considered multiples (**Figure 2.1C** and **A1.8**).

PBMC analysis

We reduced the dimensionality of our data to 11 principle components that account for the majority of the variation (51.6% cumulative variance) among variable genes to achieve optimal discrimination of cell types identified through image cytometry. We identified seven distinct clusters of cells using the FindClusters function in Seurat with k.param = 50 (a measure of neighborhood size) and resolution = 0.75 (see below; **Figure A1.11**). Clusters corresponding to CD4+ T cells, CD8+ T cells, B cells, NK cells, monocytes, and dendritic cells were all identified on the basis of significant enrichment using an ROC test implemented in Seurat (**Figures A1.10** and **A1.11**). We removed 602 cells that comprised a distinct cluster enriched for expression of mitochondrial genes (**Figure A1.11**) and a lower mapping rate of new transcripts and genes per sequencing read (**Figure A1.12**), which likely represented single-cell libraries of low complexity. We then applied t-distributed stochastic neighbor embedding (t-SNE) using the cell loadings for the previously chosen 11 principle components to visualize the cells in two dimensions. Following sequence alignment, we analyzed a total of 4,296 cells in which at least 10,000 reads, 1,000 transcripts and 500 genes were detected with mRNA alignment rate greater than 65% (**Figure 2.2B-D**), which resulted in filtering of 1,670 cells with greater than 1,000 transcripts. We analyzed a total of 6,713 genes that were detected in at least 2.5% of filtered cells across six sequencing runs from three separate arrays. We identified 687 variable genes with log-mean expression values greater than 0.5 and dispersion (variance/mean) greater than 0.5. We observed optimal discrimination of cell types identified through image cytometry using 11 principal components that account for the majority of the variation (51.6% cumulative variance) among variable genes and visualized using the *t*-distributed stochastic neighbor embedding (t-SNE) algorithm. We performed 1,000 iterations of the Barnes–Hut implementation of the t-SNE algorithm using a 'perplexity' value of 40. We identified seven distinct clusters of cells using the FindClusters function in Seurat with k.param = 50 (a measure of neighborhood size) and resolution = 0.75 (see below; **Figure A1.11**). Clusters corresponding to CD4+ T cells, CD8+ T cells, B cells, NK cells,

monocytes and dendritic Cells were all identified on the basis of significant enrichment using an ROC test implemented in Seurat (**Figures A1.10** and **A1.11**). We removed 602 cells that comprised a distinct cluster enriched for expression of mitochondrial genes (**Figure A1.11**) and a lower mapping rate of new transcripts and genes per sequencing read (**Figure A1.12**), which likely represented single-cell libraries of low complexity. We examined proportions of various cell types across arrays and sequencing runs among 3,694 cells that passed the aforementioned filtering criteria. Statistical significance of differences in the proportion of clusters between separate arrays and sequencing runs was performed using a Chi-square test (**Figure 2.2C**). We further examined phenotypic variation within myeloid cells among identified principal components (**Figure 2.2D**) by ranking cells on the basis of their PC score among genes with highest loadings for each principal component.

Comparison of Seq-Well PBMCs to 10× genomics data

We performed comparisons of gene detection and transcript capture among PBMC cell types conserved between 3,590 PBMCs (excluding dendritic cells) obtained using Seq-Well and 2,700 PBMCs from the 10x Genomics platform (http://support.10xgenomics.com/single-cell/datasets/pbmc3k). To classify PBMC cell types within the 10x Genomics data, we first identified 446 variable genes with log-mean expression values greater than 0.5 and dispersion (variance/mean) greater than 0.5. We then performed graph-based clustering using 13 principal components, k.param of 50 and resolution of 0.75. Cell type identity of each cluster was established on the basis of gene enrichments. Comparisons of genes and transcripts were initially performed between B cells, CD4 T cells, CD8 T cells, monocytes and NK cells using raw data matrices. We refined these comparisons by separately downsampling genes and transcripts within each cell type in Seq-Well data to an average read depth of 69,000 reads per cell to match the reported sequencing depth using in publicly available 10x Genomics data.

Mycobacterium tuberculosis analysis

Following sequence alignment, we identified a total of 14,218 cells with greater than 1,000 mapped transcripts. Initially, we analyzed a subset of 4,638 macrophages with greater than 5,000 detected transcripts (**Figure A1.14A**) and a total of 9,381 genes expressed in at least 5% of filtered cells. Principal components analysis was performed among a set of 377 variables genes, defined by genes with log-mean expression greater than 0.5 and dispersion (variance/mean) greater than 0.5. We performed graph-based clustering, as described below, using the first five principal components since we observed that they captured the majority of the biological variation in our data set (63% cumulative variance), and that each additional principal component contributed less than 1% to the total variance. We performed 1,000 iterations of the t-SNE algorithm (Barnes–Hut implementation) using a 'perplexity' value of 30. We identified five distinct clusters of cells in the t-SNE plot using the FindClusters function in Seurat with k.param = 40 and resolution = 0.25 (**Figure A1.14**). We removed two clusters comprised of cells with reduced gene detection, transcript capture and enrichment for expression of mitochondrial genes. Following removal of low-quality cells, we analyzed three distinct clusters with total of 2,560 high-quality cells (**Figure 2.3A** and **A1.14**). Differential expression analysis was performed between clusters, and cells exposed and unexposed to TB within each t-SNE cluster using a likelihood ratio test in Seurat (**Figure 2.3**). We performed gene set enrichment analysis to examine association of expression differences observed between control macrophages exposed and unexposed to *M. tuberculosis* with previously published gene sets using GSEA. For each cluster, expression patterns between exposed and unexposed cells were made to complete GSEA databases.

Regressing out latent technical effects

Technical parameters governing sequencing data, such as the number of genes detected or the transcriptomics alignment rate, often vary significantly across single cells. We sought to conservatively remove these technical effects using a 'latent-variable' approach similar to that of

Buettner *et al.* [24]. Briefly, we fit a linear model to predict the expression value of each gene based on a set of technical metrics, as well as the total number of unique genes detected in that cell. In our analyses, we constructed models to adjust gene expression values for alignment rate of each cell. We considered the residual expression from this model as a 'corrected' gene expression value, and we used these values as input to the downstream clustering analyses.

Graph-based clustering of single-cell transcriptomes

For all single-cell clustering analyses, we used an approach similar to that of our recently proposed clustering strategy for Drop-seq data. Briefly, as in Macosko *et al.* [12] we first identified the set of genes that was most variable across our data set after controlling for the relationship in single-cell RNA-seq data that inherently exist between mean expression and variability by binning genes into 20 bins based on their average expression level and *z*-scoring dispersion (mean/variance) estimates within a bin. We excluded all genes which were detected in less than 2.5% of PBMCs (5% of monocytes for the Mtb experiments) and used a dispersion cutoff of 0.5 to select variable genes, resulting in the selection of 687 variable genes across 4,296 PBMCs and 377 variable genes across 4,638 macrophages.

We next reduced the dimensionality of our data set, using principal components analysis. As previously described in Macosko *et al*. [12] we ran PCA using the prcomp function in R. We then selected PCs for further downstream analysis (11 PCs in PBMC analysis and 5 PCs in TB analysis). As expected, markers for distinct cell types were highly represented among the genes with the largest scores along these PCs. We then applied *t*-distributed stochastic neighbor embedding (t-SNE) using cell loadings for the significant principal components as input to visualize the structure of our data in two dimensions.

Here we used graph-based clustering methods, similar to those that have been recently proposed for both single-cell RNA-seq and mass cytometry data [25, 26]. We first construct a Euclidean distance matrix on the loadings for the significant principal components as described

above and use this to construct a *K*-nearest neighbor graph (KNN, k = 50 in PBMC analysis; k = 40 in TB analysis). Our goal was to identify 'quasi-cliques' [26] or 'communities' [25] of cells that were highly interconnected across this graph. Therefore, we first converted the KNN graph into a weighted shared nearest neighbor (SNN) graph, where the weight between any two cells was represented by the percent overlap in their respective *K*-nearest neighborhoods (Jaccard distance), and we pruned low-quality edges with a Jaccard distance of <0.1 (less than 10% overlap in local neighborhoods). Finally, to group the cells into clusters, we used a recently developed method for modularity optimization, which aims to optimize a function describing the density of connections within a cluster versus connections between clusters, essentially to identify highly interconnected nodes within the SNN graph. Here, we applied the smart local moving algorithm, which is similar to the widely used 'Louvain' algorithm for community detection but implements a local moving heuristic that enables communities to be split up and iteratively reorganized in an attempt to improve the overall partition modularity. This grants the SLM algorithm additional freedom in identifying an optimal clustering solution, and we empirically observed increased sensitivity and consistency applying this approach to single-cell data.

References

1. Shalek, A.K., et al., *Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells.* Nature, 2013. **498**(7453): p. 236-240.

2. Lohr, J.G., et al., *Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer.* Nature biotechnology, 2014. **32**(5): p. 479.

3. Shalek, A.K., et al., *Single-cell RNA-seq reveals dynamic paracrine control of cellular variation.* Nature, 2014. **510**(7505): p. 363-369.

4. Tirosh, I., *Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq.* Science, 2016. **352**: p. 189-196.

5. Hashimshony, T., et al., *CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification.* Cell reports, 2012. **2**(3): p. 666-673.

6. Bendall, S.C., et al., *Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum.* Science, 2011. **332**(6030): p. 687-696.

7. Buenrostro, J.D., *Single-cell chromatin accessibility reveals principles of regulatory variation.* Nature, 2015. **523**: p. 486-490.

8. Smallwood, S.A., et al., *Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity.* Nature methods, 2014. **11**(8): p. 817-820.

9. Zeisel, A., et al., *Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq.* Science, 2015. **347**(6226): p. 1138-1142.

10. Treutlein, B., et al., *Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq.* Nature, 2014. **509**(7500): p. 371-375.

11. Klein, A.M., et al., *Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.* Cell, 2015. **161**(5): p. 1187-1201.

12. Macosko, E.Z., et al., *Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets.* Cell, 2015. **161**(5): p. 1202-1214.

13. Fan, H.C., G.K. Fu, and S.P. Fodor, *Combinatorial labeling of single cells for gene expression cytometry.* Science, 2015. **347**(6222): p. 1258367.

14. Bose, S., et al., *Scalable microfluidics for single-cell RNA printing and sequencing.* Genome Biology, 2015. **16**(1).

15. Yuan, J. and P.A. Sims, *An automated microwell platform for large-scale single cell RNA-seq.* Scientific reports, 2016. **6**(1): p. 1-10.

16. Dekosky, B.J., et al., *High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire.* Nature Biotechnology, 2013. **31**(2): p. 166-169.

17. Ilicic, T., et al., *Classification of low quality cells from single-cell RNA-seq data.* Genome biology, 2016. **17**(1): p. 29.

18. Yamanaka, Y.J., et al., *Single-cell analysis of the dynamics and functional outcomes of interactions between human natural killer cells and target cells.* Integrative Biology, 2012. **4**(10): p. 1175-1184.

19. Han, Q., et al., *Polyfunctional responses by human T cells result from sequential release of cytokines.* Proceedings of the National Academy of Sciences, 2012. **109**(5): p. 1607-1612.

20. Steinberg, G., et al., *Strategies for Covalent Attachment of DNA to Beads.* Biopolymers, 2004. **73**(5): p. 597-605.

21. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.* Cell, 2015. **161**(5): p. 1202-14.

22. Roederer, M., *Multiparameter FACS analysis.* Current protocols in immunology, 2002. **49**(1): p. 5.8. 1-5.8. 10.

23. Satija, R., et al., *Spatial reconstruction of single-cell gene expression data.* Nature biotechnology, 2015. **33**(5): p. 495-502.

24. Buettner, F., et al., *Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells.* Nature biotechnology, 2015. **33**(2): p. 155.

25. Levine, J.H., et al., *Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis.* Cell, 2015. **162**(1): p. 184-197.

26. Xu, C. and Z. Su, *Identification of cell types from single-cell transcriptomes using a novel clustering method.* Bioinformatics, 2015. **31**(12): p. 1974-1980.

Chapter 3: Highly Efficient, Massively-Parallel Single-Cell RNA-Seq Reveals Cellular States and Molecular Features of Human Skin Pathology

Modified from: Highly Efficient, Massively-Parallel Single-Cell RNA-Seq Reveals Cellular States and Molecular Features of Human Skin Pathology

Travis K Hughes*, Marc H Wadsworth II*, Todd M Gierahn*, Tran Do, David Weiss, Priscilla R. Andrade, Feiyang Ma, Bruno J. de Andrade Silva, Shuai Shao, Lam C Tsoi, Jose Ordovas-Montanes, Johann E Gudjonsson, Robert L Modlin, J Christopher Love, and Alex K Shalek. BioRxiv. 2019.

*These authors contributed equally.

Authorship Attributions

T.K.H., M.W.H., T.M.G., R.L.M., J.C.L., and A.K.S. designed the study. T.K.H., M.H.W., T.D., D.W., P.A., and B.A. collected skin samples and performed single-cell sequencing experiments. S.S., L.C.T., and J.E.G. performed immunofluorescent staining. T.K.H., M.H.W., T.M.G., and F.M. analyzed data under the guidance of J.O.M., R.L.M., J.C.L., and A.K.S. T.K.H., M.H.W., T.M.G., J.C.L., and A.K.S. wrote the manuscript with input from all authors.

Introduction

Although a nascent technology, single-cell RNA-sequencing (scRNA-Seq) has already helped define, at unprecedented resolution, the cellular composition of many healthy and diseased tissues [1-6]. The development of high-throughput methodologies has been crucial to this process, empowering the characterization of increasingly complex cellular samples. Unfortunately, current scRNA-Seq platforms typically demonstrate an inverse relationship between the number of cells that can be profiled at once and the amount of biological information that can be recovered from each cell. As a result, one must choose between quantity and quality – and thus comprehensiveness and fidelity – or alternatively employ two distinct approaches in parallel [7]. Indeed, inefficiencies in transcript capture among massively-parallel methods have limited our ability to resolve the distinct cell states that comprise broad cell types [8], as well as their essential molecular attributes and often lowly-expressed molecular features, such as transcription factors, affinity receptors, and signaling molecules.

Improving the fidelity of these methodologies is particularly important for resolving differences within heterogeneous populations of immune cells like lymphocytes and myeloid cells [9, 10]. Here, subtle differences in surface receptor, transcription factor and/or cytokine expression can profoundly impact cellular function, particularly in the setting of human pathology [11]. Enhancing data quality in high-throughput scRNA-Seq would facilitate a greater appreciation of the underlying molecular features that describe such cellular variation. Similarly, it would ease

integration with legacy datasets that often utilize lowly-expressed biomarkers, such as transcription factors, that are false-negative prone to discriminate subsets of cells.

Most high-throughput scRNA-Seq methods currently rely on early barcoding of cellular contents to achieve scale. Typically, these techniques recover single-cell transcriptomes for thousands of cells at once by leveraging reverse-emulsion droplets or microwells to isolate individual cells with uniquely barcoded poly-dT oligonucleotides which can then capture and tag cellular mRNAs during reverse transcription [12]. Afterward, an additional priming site is typically added to the 3' end of the synthesized cDNA to enable PCR-based amplification of all transcripts using a single primer (whole transcriptome amplification, WTA). A number of techniques have been described to add this second priming site [13, 14]. The most common uses the terminal transferase activity of certain reverse transcription enzymes to facilitate a "template-switch" from the original mRNA to a second defined oligonucleotide handle [15]. While simple to implement, this process has the potential to be highly inefficient, leading to the loss of molecules that have been captured and converted to cDNA but not successfully tagged with a secondary PCR priming site [16-18].

To overcome these limitations, we have developed a new, massively-parallel scRNA-Seq protocol we call Seq-Well S^3 (for "Second-Strand Synthesis"). Of the modifications incorporated into the Seq-Well S^3 workflow, one of the most significant is the use of a randomly-primed second-strand synthesis after reverse transcription to append the second PCR handle addition. Working with cell lines and peripheral blood mononuclear cells (PBMCs), we demonstrate that Seq-Well S^3 enables significant improvements in transcript and gene capture across sample types, facilitating studies of complex immune tissues at enhanced resolution.

To illustrate the utility of S^3, we apply it to generate a resource of single-cell transcriptional states spanning multiple inflammatory skin conditions. Skin represents the largest barrier tissue in the human body and is comprised of numerous specialized cell-types that help

maintain both immunological and physical boundaries between our inner and outer worlds [19]. The dermis and epidermis – the two primary compartments of human skin – play complementary roles in tissue structure and function [19] . The epidermis consists primarily of keratinized epithelial cells, which provide a physical barrier to the outside world; the dermis, meanwhile, provides structural support for the skin, with fibroblasts producing collagen and elastin fibrils, along with the other components of the extracellular matrix. Crucially, within the cellular ecosystem of human skin, there are numerous tissue-resident immune and parenchymal cells that are essential to homeostatic barrier function.

To date, single-cell analysis of human skin has revealed heterogeneity in immune and parenchymal cell types in health and disease [20-23]. Using Seq-Well S^3, we examine the cellular composition of normal skin and altered cellular phenotypes across multiple inflammatory skin conditions, specifically acne, alopecia areata, granuloma annulare, leprosy and psoriasis. With conditions that span autoimmune inflammation driven by adaptive immunity (alopecia), combined autoimmune and autoinflammatory (i.e. both adaptive and innate processes; psoriasis), reactive (acne), and granulomatous (granuloma annulare and leprosy) inflammation, we uncover a diverse spectrum of immune and parenchymal cellular phenotypes, as well as their molecular features, across multiple inflammatory skin conditions. Overall, our work presents an essential methodological advance, as well as a critical resource for understanding how diverse inflammatory responses can impact a single tissue, and the range of cellular phenotypes that are possible upon perturbation.

RESULTS

Second-Strand Synthesis (S^3) Leads to Improved Transcript Capture and Gene Detection

We hypothesized that use of “template-switching” to append a second PCR handle during reverse transcription might limit the overall recovery of unique transcripts and genes from individual cells in some massively-parallel scRNA-Seq methods such as Seq-Well and Drop-Seq

[2, 24]. Thus, we incorporated a randomly primed second-strand synthesis following first-strand cDNA construction (**Figures 3.1A** and **A2.1A**). Briefly, after reverse transcription, barcoded mRNA capture beads are washed with 0.1 molar sodium hydroxide to remove attached RNA template strands, and then a random second-strand synthesis is performed to generate double-stranded cDNA that is labeled on one end with the SMART sequence and reverse complement on the other (**Figure 3.1A** and **A2.1A; Methods**) [15, 25].

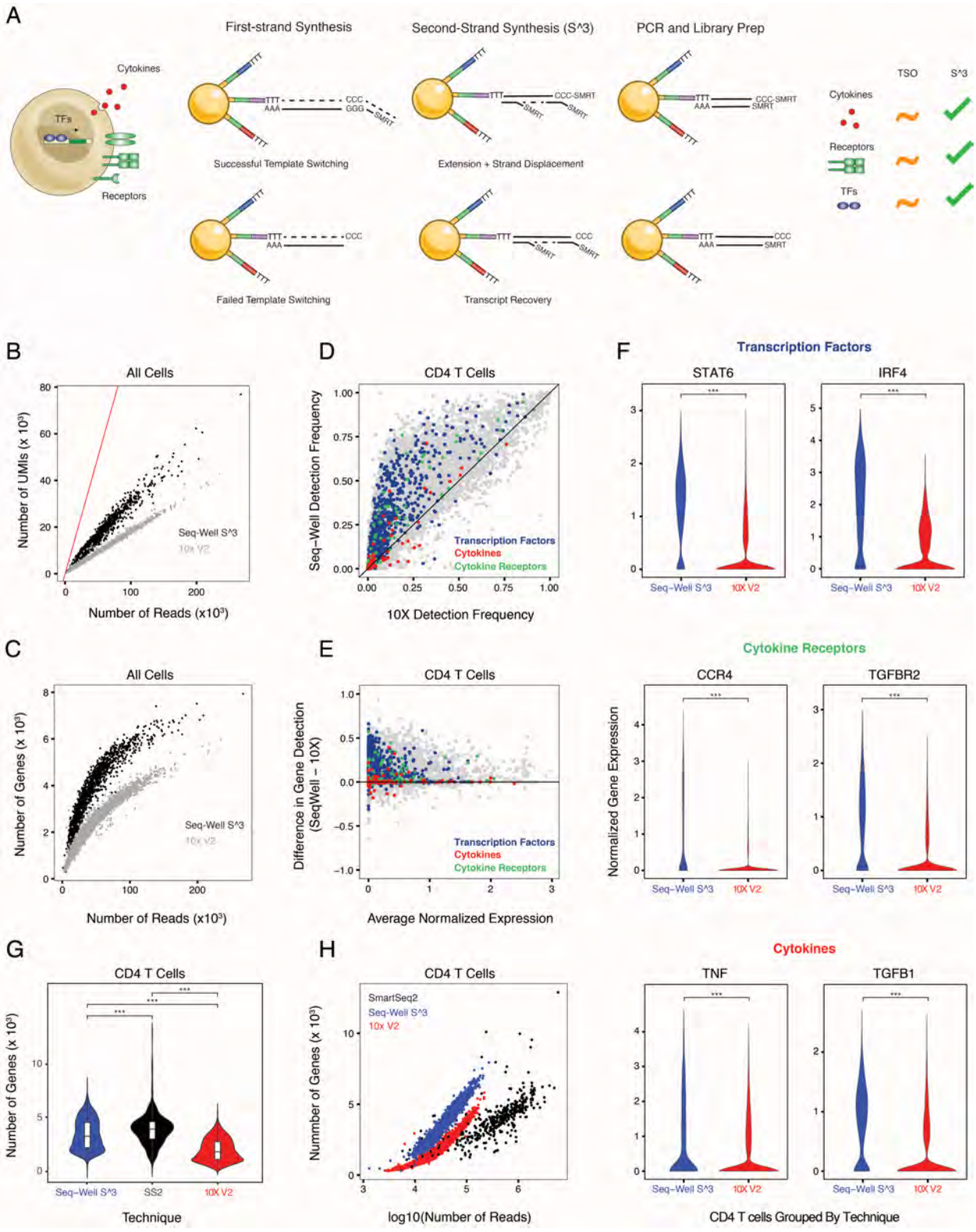


Figure 3.1 | Overview of Second Strand Synthesis (S^3). A. Conceptual illustration of the molecular features that define immune phenotypes – including transcription factors, cytokines and receptors – as well as the Seq-Well second-strand synthesis method (Seq-Well S^3) and how it improves detection of key

genes and transcripts. **B.** Scatterplot showing differences in per-cell transcript capture (y-axis) as a function of aligned reads per cell (x-axis) between 10x Genomics v2 (10x v2, grey) and Seq-Well S^3 (black). Red line indicates uniform line where transcripts per cell and aligned reads would be equivalent. **C.** Scatterplot shows the differences in per-cell gene detection (y-axis) as a function of aligned reads per cell (x-axis) between 10x v2 (grey) and Seq-Well S^3 (black). **D.** Scatterplot comparing gene detection rates in $CD4^+$ T cells between 10x v2 (x-axis) and Seq-Well S^3 (y-axis). Black line indicates point of equivalence in gene detection frequency between methods. Colors correspond to classes of genes including transcription factors (blue), cytokines (red), and receptors (green). **E.** Scatterplot comparing gene detection frequency (y-axis) between Seq-Well S^3 (positive values) and 10x v2 (negative values) as a function of the aggregate expression levels (log(scaled UMI + 1)) of an individual gene (x-axis). Black line indicates point of equivalence in gene detection frequency between methods. Colors correspond to classes of genes including transcription factors (blue), cytokines (red), and receptors (green). **F.** Violin plot (boxplots median +- quartiles) showing the distribution of per-cell transcript capture for Seq-well S^3 (blue; n = 1,485), 10x v2 (red; n = 2995), and Smart-Seq2 (black, n = 382). *** P-values < 1.0×10^{-10}. **G.** Violin plots showing the distribution of normalized expression values for select transcription factors, cytokines and cytokine receptors between Seq-Well S^3 and 10x v2. *** P-values < 1.0×10^{-10}. **H.** Scatterplot showing the relationship between aligned reads and genes detected per cell between Seq-Well S^3 (blue), 10x v2 (red) and Smart-Seq2 (black) in sorted PBMC $CD4^+$ T cells.

To examine the effectiveness of Seq-Well S^3 and optimize its performance, we first tested a number of conditions using cell lines (i.e., HEK293) and human PBMCs (**Figure A2.1B; Methods**). In these experiments, we observed that S^3 led to marked improvements in library complexity (i.e., the number of unique transcripts detected per aligned read; Seq-Well V1: 0.22 transcripts/aligned read; Seq-Well S^3: 0.68 transcripts/aligned read) at matched sequencing depth below saturation (**N.B.** higher library complexity implies that a greater amount of information remains to be detected through further sequencing). Seq-Well S^3 was further able to function in the absence of a template switching oligo (TSO); Seq-Well V1, meanwhile, failed to generate appreciable product without a TSO (**Figure A2.1B-E**). In species-mixing experiments using HEK293 (human) and NIH-3T3 (mouse) cell lines, the use of the S^3 protocol not only achieved significant increases in the numbers of unique transcripts captured and genes detected per cell compared to our original protocol for Seq-Well ($P < 0.05$, Mann-Whitney U Test; **Figure A2.1D; Methods**), but also comparable single-cell resolution (i.e., transcript purity) to Seq-Well V1 (**Figure A2.1F-G**).

To fully understand how S^3 would perform on more challenging primary cells, we applied it to human PBMCs (**Figure A2.1E** and **A2.2**; **Methods**), benchmarking against our original Seq-

Well protocol as well as commercial technology (10x genomics, v2 3' chemistry; hereafter 10x v2). For these comparisons, we down-sampled all resulting data to an average of 38,000 reads per cell to account for differences in sequencing depth across technologies. Critically, Seq-Well S^3 resulted in significant improvements in the complexity of our sequencing libraries compared to 10x v2 as determined by the number of transcripts and genes detected at matched read depth ($P < 0.05$, Mann-Whitney U Test & Linear Regression**; Figure 3.1B-C**; **Methods**). To confirm that these overall improvements were not driven by changes in the relative frequencies of different cell types captured by each technology, we also examined each subset independently (**Figure A2.2A-B**). For each cell type, we observed significant increases in the numbers of transcripts captured and genes detected using S^3 in pairwise comparisons between techniques ($P < 0.05$, Mann-Whitney U Test; $CD4^+$ T cells, Seq-Well V1: 1,044 $\pm$ 62.3 UMIs/cell; 10x v2: 7,671 $\pm$ 103.9 UMIs/cell; Seq-Well S^3: 13,390 $\pm$ 253.4 UMIs/cell; Mean $\pm$ Standard Error of the Median (SEM)); **Figure A2.2C, Methods**). Both Seq-Well S^3 and 10x v2 displayed increased sensitivity for transcripts and genes relative to Seq-Well v1 (Seq-Well S^3: 6-fold gene detection, 10-fold UMI detection), but Seq-Well S^3 showed the greatest efficiency (defined as genes recovered at matched read depth) to detect genes for each cell type (**Figure 3.1C**; **Figure A2.2**).

We sought to further understand whether these improvements resulted in enhanced detection of biologically relevant genes typically under-represented in high-throughput single-cell sequencing libraries [7]. Importantly, genes that were differentially detected (i.e., higher in S^3) within each cell type include numerous transcription factors, cytokines and cell-surface receptors (**Figure 3.1D-E, A2.2D-E**). For example, among $CD4^+$ T cells, we observe significantly increased detection of transcription factors (e.g., *STAT6,* and *IRF4*), cytokine receptors (e.g., *CCR4 and TGFBR2*) and cytokines (e.g., *TGFB1* and *TNF*; $P< 0.05$, Chi-Square Test, **Figure 3.1F** and **A2.2**).

We performed an additional comparison of enriched human $CD4^+$ T cells profiled using Seq-Well S^3, 10X v2, and Smart-Seq2, a commonly implemented microtiter plate-based approach (**Figure 3.1G-H; Methods**) [15]. Integrated analysis of aggregate gene detection revealed that Seq-Well S^3 detects more genes per cell than 10x v2 and nearly as many genes per cell as Smart-Seq2 in pairwise comparison of techniques (10x v2: 2,057 $\pm$ 18.7 genes/cell , Seq-Well S^3: 3,514 $\pm$ 36.2 genes/cell , SS2: 3,975 $\pm$ 74.0 genes/cell; mean $\pm$ SEM; $P < 0.05$, Mann-Whitney Test; **Figure 3.1G; Methods**). Further, comparing the frequency of gene detection between methods revealed crucial differences for transcription factors, cytokines and receptors/ligands (**Methods**). Surprisingly, we observe similar rates of gene detection between S^3 and Smart-Seq2 for a large number of biologically informative genes (**Figure A2.2F**). Critically, while comparable numbers of genes were detected across methods, Seq-Well S^3 detected more genes per aligned read than either 10x v2 or SS2 in pairwise comparisons ($P<0.05$, Mann-Whitney U Test; **Figure 3.1H; Methods**).

Finally, we performed additional comparisons between Seq-Well S^3 and the most recent version of the 10x genomics chemistry (10x v3 3' chemistry) (**Figure A2.3**). Here, we observe that Seq-Well S^3 detected similar levels of genes and transcripts as 10x V3 at comparable sequencing depth (**Figure A2.3 A-C).** For example, among CD4 T cells, we observe similar levels of gene detection and transcript capture per-cell between Seq-Well S^3 (**Figure A2.3C**). We note that we once again observe increased detection frequencies among biologically informative transcripts using Seq-Well S^3 (**Figure A2.3D-E**; **Methods).**

A Resource of Cellular States Across Healthy and Inflamed Skin

To demonstrate the utility of Seq-Well S^3 to comprehensively describe cellular states across human pathology at unprecedented resolution, we applied it to profile human skin samples spanning multiple, complex inflammatory skin conditions (**Figure 3.2**). We studied acne, alopecia

areata, granuloma annulare, leprosy, and psoriasis, as well as normal skin (**Figure 3.2A-C** and **A2.4A-C; Methods**). In total, we processed 19 skin biopsies (Acne, n=4; Alopecia, n = 1; granuloma annulare, n=2; leprosy, n=4; psoriasis, n=5; normal skin, n=3) by S^3 and, after data quality filtering, retained 38,274 high-quality single-cell transcriptomes (**Figure 3.2A-C; Methods**).

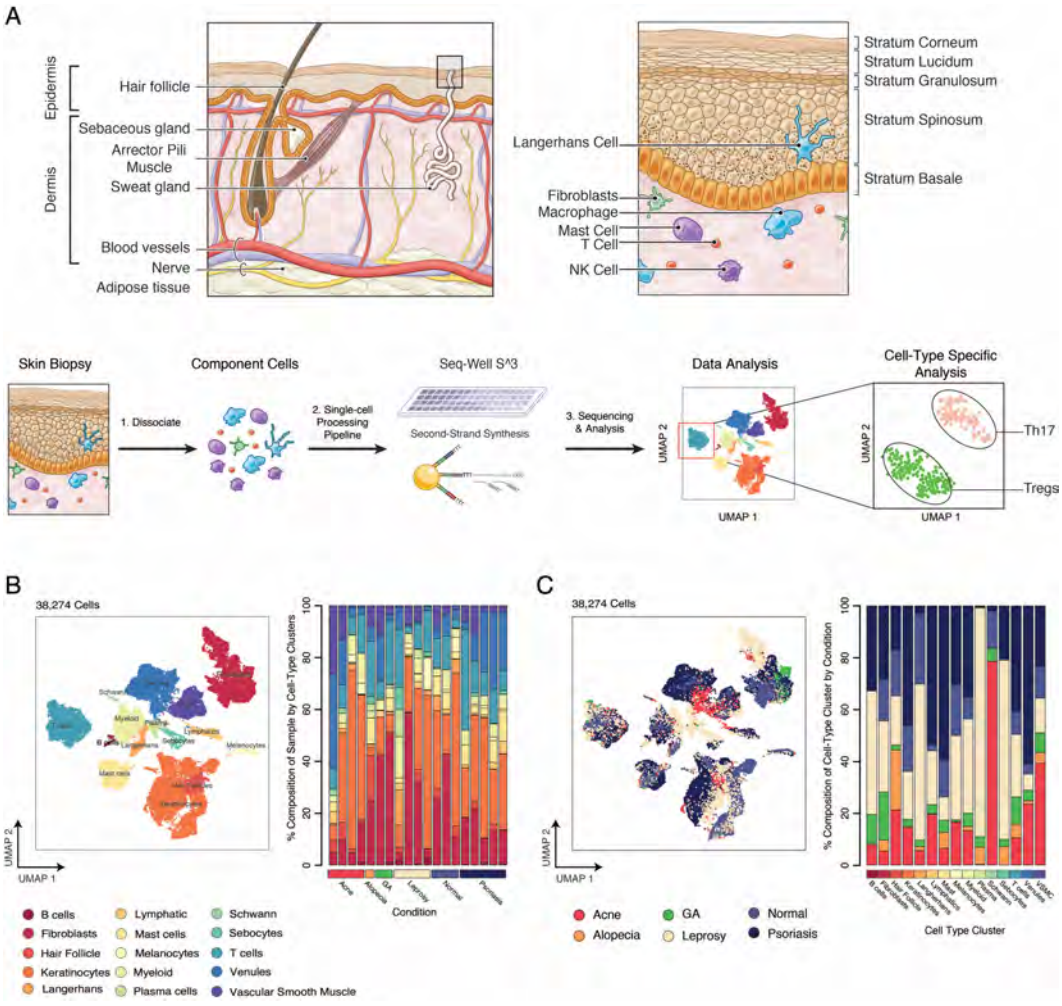


Figure 3.2 | Cell Types Recovered across Inflammatory Skin Conditions. **A.** (Top-Left) Diagram illustrating the anatomic organization and major features of human skin. (Top-Right) Cell-type composition of the epidermis and dermis. (Bottom) Sample processing pipeline used to generate a collection cellular states across skin inflammation. **B.** (Left) UMAP plot for 38,274 cells colored by cell-type cluster. (Right) Stacked barplot showing the cell-type composition for each of the 19 skin biopsies. **C.** (Left) UMAP plot for 38,274 cells colored by inflammatory skin condition. (Right) Stacked barplot showing the proportion of cells from each skin condition within phenotypic clusters.

To examine similarities and differences among these cells across the high-dimensional gene expression space, we selected variable genes, performed UMAP dimensionality reduction, and identified 35 clusters through Louvain clustering in Scanpy [26] (**Figure 3.2** and **A2.4A-C**; **Methods**). To collapse clusters to cell-types, we performed enrichment analyses to identify cluster-defining genes and then manually assigned cell-type identities based on the expression of known lineage markers (**Figure 3.2B** and **Figure A2.4A,D; Methods**). We further classified cells using SingleR [27] and observed close concordance between manually identified cell populations and automated classification where appropriate reference signatures existed (**Figure A2.4B; Methods**). To further support our annotations and groupings, we generated aggregate gene expression profiles and performed hierarchical clustering using a combined list of the top 50 cluster-defining genes for each cluster (**Figure A2.4C**; **Methods**). Ultimately, we recovered a total of 16 primary cell-types, within which there was considerable heterogeneity. The identified cell types include: B cells (marked by expression of *MS4A1* and *CD79A*), fibroblasts (*DCN* and *COL6A2*), hair follicles (*SOX9*), keratinocytes (*KRT5* and *KRT1*), Langerhans cells (*CD207*), lymphatic endothelial cells (*LYVE1*), mast cells (*CPA3* and *IL1RL1*), melanocytes (*MLANA*), myeloid cells (*CD68* and *CTSS*), plasma cells (*IGHG1*), Schwann cells (*SCN7A*), sebocytes (*DCD*), T cells (*CD3D* and *TRBC2*), venules cells (*SELE* and CD93), and vascular smooth muscle cells (VSMCs; *TAGLN*) (**Figure 3.2** and **A2.4A-E**). As a final quality measure, we examined the distribution of reads, transcripts and genes within each major cell population and observed consistent coverage across recovered cell types (**Figure A2.4F**). We next sought to define nuanced cell states within these immune, stromal and parenchymal populations – including T cells, myeloid cells, endothelial cells, dermal fibroblasts, and keratinocytes – across the spectrum of skin inflammation.

Seq-Well S^3 describes T cell states across inflammatory skin conditions

To determine the range of biological features that can be captured using Seq-Well S^3, we first focused on further characterizing T cells across the inflammatory skin conditions examined since each is known to significantly skew T cell phenotypes (**Figure 3.3**) [28, 29]. We performed dimensionality reduction and sub-clustering across T cells alone (**Figure 3.3A-B; Methods**). Our analysis revealed nine sub-clusters that closely correspond to NK cells and $CD8^+$ T cells, as well as several known $CD4^+$ T-helper cell (Th) subsets. As before, we used the enhanced sensitivity of S^3 for lineage defining transcripts to help annotate the identity of each sub-cluster; for example, in regulatory T cells and Th-17 cells we detected distinct expression of canonical transcription factors (e.g., *FOXP3* and *RORC,* respectively*)* and immune receptors (e.g. *TIGIT/CTLA4/IL2RA* and *CXCR6*, respectively) (**Figure 3.3C-D** and **A2.5**). Additionally, we cross-referenced each sub-cluster's marker genes against a series of curated signatures in the Savant database [30] to confirm our assignments, which highlighted similarity to previously characterized T cell and NK cell populations [31, 32] (**Figure A2.5A; Methods**).

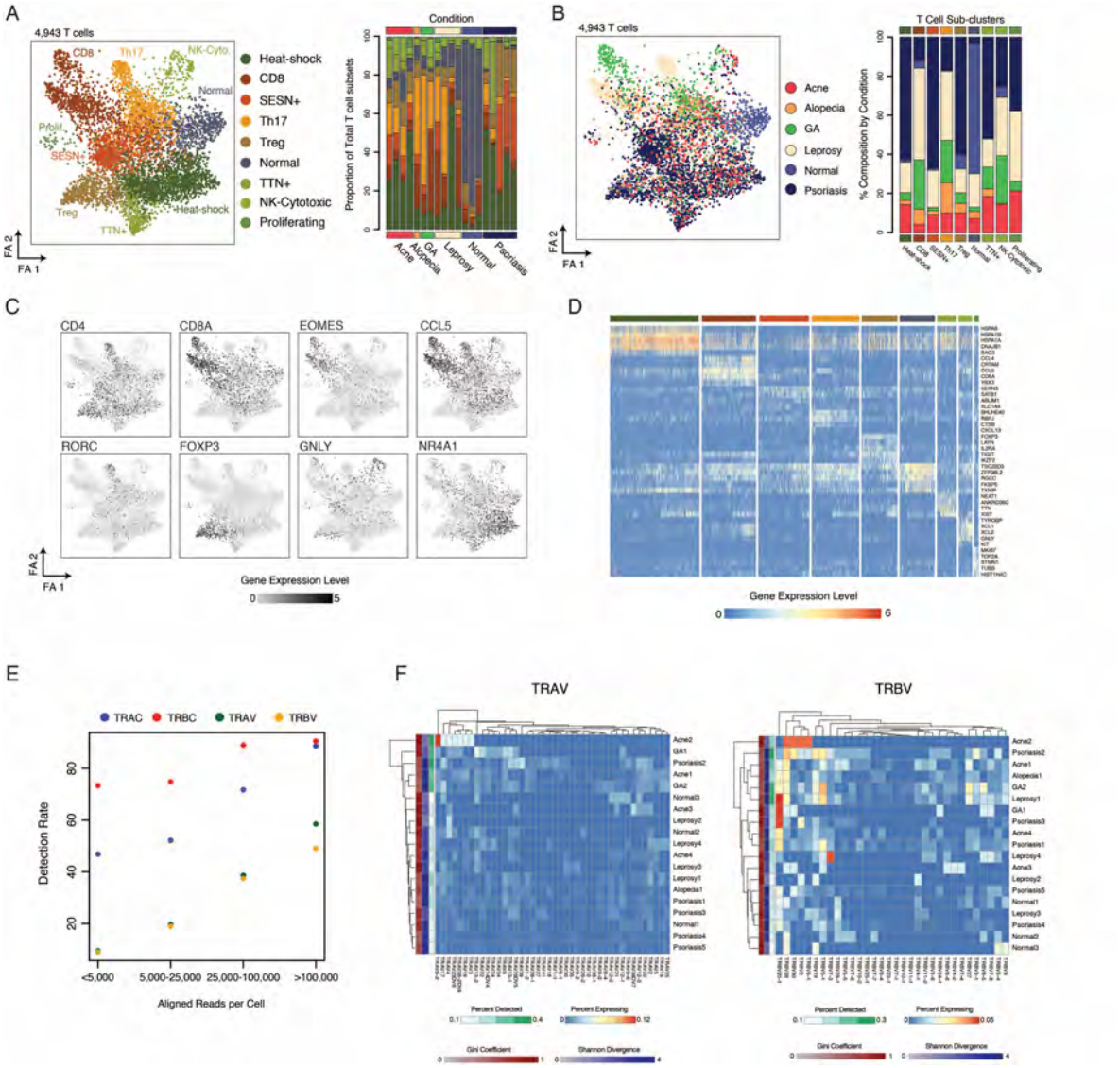


Figure 3.3 | Identification of Inflammatory T cell States using Seq-Well S^3. A. (Left) Force-directed graph of 4,943 T cells colored by the nine phenotypic sub-clusters identified by Louvain clustering. (Right) Stacked barplots showing the distribution of these T cell sub-clusters within each skin biopsy. **B.** (Left) Force-directed graph of 4,943 T cells colored by inflammatory skin condition. (Right) Stacked barplots showing the contribution of each inflammatory skin condition to the T cell sub-clusters. **C.** T cell force-directed graphs displaying normalized expression (log(scaled UMI + 1)) of a curated group of sub-cluster-defining gene. Higher expression values are shown in black. **D.** Heatmap showing normalized gene expression values (log(scaled UMI + 1)) for a curated list of sub-cluster-defining genes across nine T cell sub-clusters. **E.** Plot showing rates of detection of TCR genes from human skin T cells across a range of sequencing depths. **F.** Heatmaps showing the distribution of TRAV (left) and TRBV (right) gene expression among T cells within each sample. Within each sample (rows), the color represents the percent of T cells expressing a given TRAV or TRBV gene (columns). The sidebar shows (red) the gini coefficient (blue) the Shannon Divergence, and (green) the percent of T cells within each sample with non-zero expression of either TRAV or TRBV genes.

We next examined T cell phenotypes across inflammatory skin conditions to explore variability in T cell subset composition by skin pathology (**Figure 3.3A-B**). This analysis revealed potentially varied contributions to different classes of cutaneous inflammation. For example, we observe a population enriched for expression of canonical Th-17 genes including *RORC*, which encodes the Th-17 lineage-defining transcription factor RORγt [33] and is observed within 2 leprosy biopsies (**Figure 3.3A-B**), consistent with a role for Th-17 cells in controlling disease [34, 35].

We further found that a sub-cluster of T cells, which express *NR4A1*, a transcription factor that is a marker of dysfunctional T cells [36], and are enriched for genes involved in nuclear organization (*NEAT* and *ANKRD36*), was over-represented in psoriasis samples (**Figure 3.3B-C**). We also observe an expansion of regulatory T cells in 3 of 5 patients with psoriasis in addition to a population of T cells characterized by expression of *SESN3,* a marker of T cell senescence [37], *SATB1*, and *FURIN* (**Figure 3.3A-D**).

Directed analysis within $CD8^+$ T cells revealed a sub-grouping of activated $CD8^+$ T cells that express elevated levels of several inflammatory cytokines (*TNF*, *CCL4*, and *XCL1),* as well as specific affinity receptors (*FASLG* and *TNFRSF9*) and transcription factors (*KLF9* and *EGR2*); this phenotypic skewing was observed primarily in a patient with granuloma annulare (**Figure A2.5B - Top; Methods**). We also uncovered considerable variation within the cluster containing cytotoxic T cells and NK cells (Cytotoxic), where we found the highest degree of cytotoxic gene expression (*GNLY*, *GZMB*, and *PRF1*). Indeed, sub-clustering analysis of the cytotoxic cluster revealed 3 distinct sub-groups (**Figure A2.5B**): 1. a sub-group of NK cells (Cytotoxic-1) enriched for expression of *c-KIT*, *RANKL* (TNFSF11) and *GITR* (TNFSFR18); 2. sub-group of $CD16^+$ cells (Cytotoxic-2) expressing cytotoxic effector molecules (*GNLY*, *PRF1*, *GZMB*) and NK surface receptors, consistent with either NK cell or tri-cytotoxic CTL [38]; and, 3. a sub-group of $CD8^+$ T cells (Cytotoxic-3, *TNFSF8*, *SLAMF1*, *CLEC2D*, *CD5*) that express both TCR αβ and γδ constant

genes (**Figure A2.5B - Bottom**) [39]. Collectively, these data suggest that the cytotoxic cluster actually contains NK cells, γδ T cells, and activated cytotoxic T cells.

Profiling of T cell receptor expression is critical to understand T cell antigen specificity [40] (Zhang et al., 2018). Importantly, among $CD4^+$ T cells obtained from peripheral blood, we recovered TCR-V and TCR-J genes at a higher frequency on a per-gene basis using Seq-Well S^3 as compared to 10x v2 ($P < 0.05$, Chi-square Test; **Figure A2.5C; Methods**). Among $CD4^+$ T cells from peripheral blood, we observed paired detection of *TRAC* and *TRBC* in 1,293 of 1,485 $CD4^+$ T cells (87.1% Paired Detection Rate, **Figure A2.5C**). In the setting of skin inflammation, we explored TCR detection rates across a range of sequencing read depths. Overall, we detected *TRAC* in 53.5%, *TRBC* in 76.7% (**Figure 3.3E**), and paired detection in 45.1% of T cells. Among T cells with at least 25,000 aligned reads, we recovered paired α and β chains in 68.6%. Among cytotoxic cells, we observe expression of γ and δ constant genes (*TRGC* and *TRDC*), while the remaining T cell clusters exclusively express α and β TCR constant genes. These data further suggest that the cytotoxic cluster represents a diverse population of γδ, NK, and cytotoxic $CD8^+$ T cells that share common gene expression features and, potentially, roles in inflammation.

Finally, we examined the distribution of TCR V gene expression across inflammatory skin biopsies to identify clonally expanded T cells (**Figure 3.3F; Methods**). Here, we observe biased distributions of TRAV and TRBV genes (e.g. elevated Gini coefficients and low Shannon Divergence) within multiple biopsies including biopsies from leprosy and acne (Leprosy 2 and Acne 2, TRAV and TRBV Gini Coefficient > 0.85; **Figure 3.3F**).

Spectrum of Myeloid Cell States in Skin Inflammation

In the setting of cutaneous inflammation, myeloid cells play a key role in maintaining tissue homeostasis, wound healing and response to pathogens [41]. Using Seq-Well S^3, we were able

to identify numerous myeloid cell subpopulations defined by a combinations of surface markers, cytokines and lineage-defining transcription factors. Specifically, we independently analyzed 5,010 myeloid cells and identified 10 sub-clusters representing 4 primary myeloid cell types based on expression of canonical lineage markers and comparison to cell-type signatures in the Savant database: dendritic cells (*CLEC9A* and *CLEC10A*), Langerhans cells (*CD207* and *CD1A*), macrophages (*CD68* and *CD163*), and mast cells (*CPA3* and *TPSAB1*) (**Figure 3.4A** and **A2.5D-E**; **Methods**) [30].

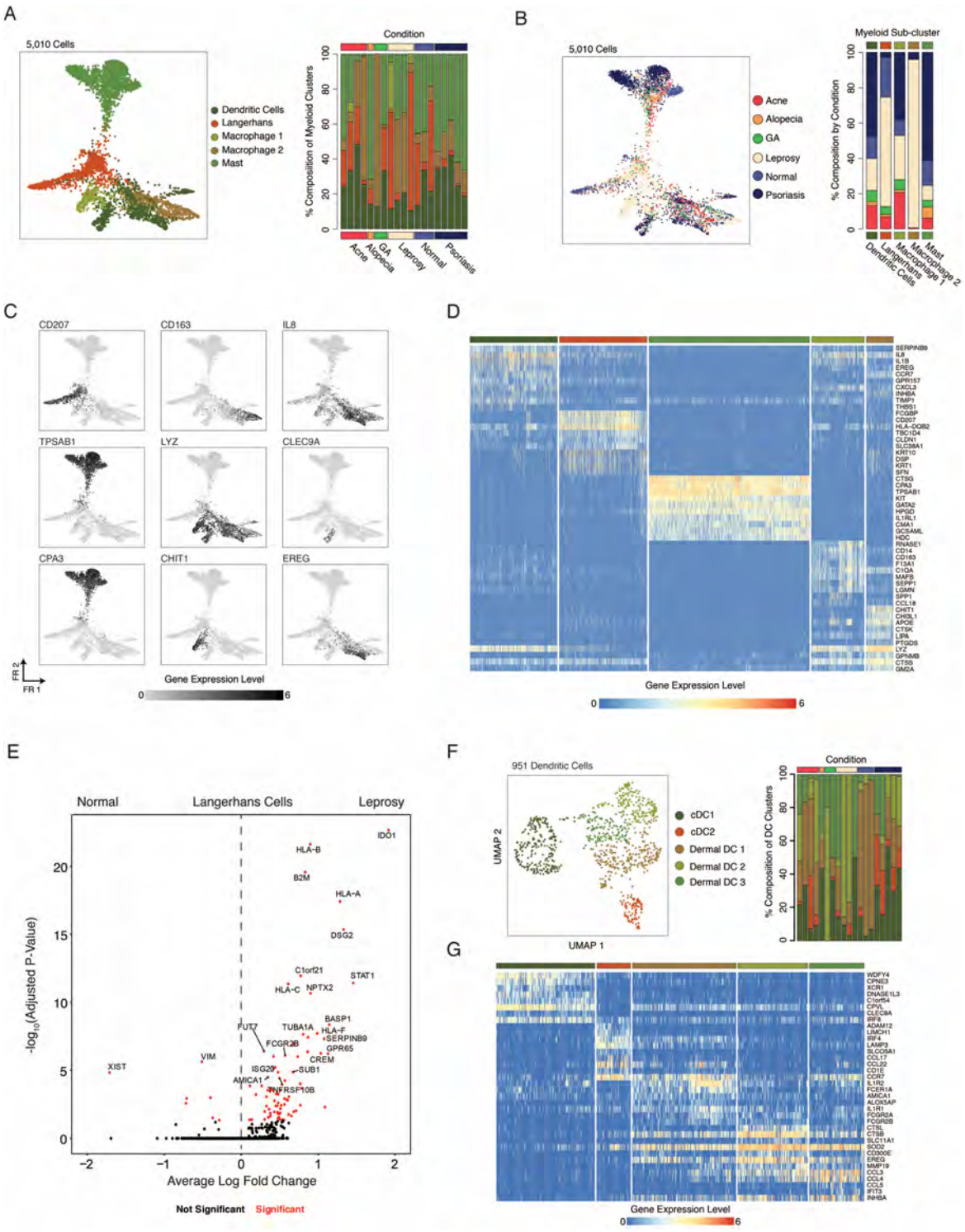


Figure 3.4 | Diverse Myeloid Cell States Uncovered using Seq-Well S^3. A. (Left) Force-directed graph of 5,010 myeloid cells colored by five phenotypic sub-clusters (NB, Langerhans cells were enriched from leprosy and normal skin). (Right) Stacked barplots showing the distribution of myeloid sub-clusters within each skin biopsy. **B.** (Left) Force-directed graph of 5,010 myeloid cells colored by inflammatory skin condition. (Right) Stacked barplots showing the contribution of each inflammatory skin condition to each myeloid sub-cluster. **C.** Force-directed graphs of 5,010 myeloid cells that highlighting expression of a

curated group of sub-cluster defining genes (log(scaled UMI + 1)). **D.** Heatmap showing the normalized expression (log(scaled UMI + 1)) of a curated list of myeloid cell-type cluster-defining genes. **E.** Volcano plot showing genes differentially expressed in Langerhans cells between leprosy (n_{cells} = 67) and normal skin (n_{cells} = 171). Log10-fold change values are shown on the x-axis and -log10 adjusted p-values are shown on the y-axis. **F.** (Left) UMAP plot for 951 dendritic cells from human skin colored by inflammatory skin condition. (Right) Stacked barplot showing the distribution dendritic cell sub-grouping within 19 skin biopsies. **G.** Heatmap showing the distribution of normalized gene expression levels (log(scaled UMI + 1)) for cluster-defining genes across dermal DC subpopulations.

Among the macrophages, our data reveal two distinct sub-clusters (**Figure 3.4A-B**). One macrophage sub-cluster spans normal skin as well as multiple types of skin inflammation and is characterized by elevated expression of previously characterized markers of dermal macrophages (*CD163*, *STAB1*, and *CEPP)* [42]. The second sub-cluster, meanwhile, is observed primarily in a single leprosy patient and is defined by genes involved in extracellular proteolysis (*LYZ*, *CHIT1*, and *CHI3L1*) [43].

Skin functions as both a physical and immunologic barrier, and is the primary site of exposure to environmental antigens. As such, multiple types of antigen-presenting cells (APCs) are distributed in both the dermis and epidermis. In the epidermis, there is a specialized population of antigen-presenting cells known as Langerhans cells. We initially identified Langerhans cells on the basis of expression of canonical markers (*CD207*, *CD1A*; **Figure 3.4C-D**) [44]. For two biopsies obtained from normal skin and leprosy (Normal skin 1 and Leprosy 1), we performed MACS enrichments from the epidermal section and loaded Langerhans cells as 5% of the total amount to increase recovery (**Methods**). When we directly compared Langerhans cells from leprosy and normal skin, we observed elevated expression of *IDO1*, *STAT1*, *HCAR3* and MHC class I molecules (*HLA-A*, *HLA-B* and *HLA-F*) in Langerhans cells in leprosy infection. (**Figure 3.4E**) [45, 46]. We further performed gene-ontology analysis among genes up-regulated in Langerhans cells from leprosy and observe enrichment of genes related to IFN-γ response.

Sub-analysis of the dendritic cell cluster revealed multiple sub-groups, including conventional and dermal dendritic cells (**Figure 3.4F**). Consistent with previous observations from peripheral blood, we saw a sub-group of dendritic cells that corresponds to cDC1 (*CLEC9A*, *IRF8*,

and *WDFY4*) [9] (P<0.05, Permutation Test, **Figure A2.5F**, **Methods**). In comparison to another set of DC signatures [10], we observe close correspondence to cDC1 cells (**Figure S5G, Methods**). We further report another DC sub-group representing cDC2 cells (*IRF4*, *SOCS2*, *SLCO5A1, CD1B, CD1E*) that is observed across all biopsies from patients with psoriasis (**Figure 3.4F, Figure A2.5H**; **Methods**) [47]. Importantly, we detect expression of *IL12B*, a subunit of the IL-23 cytokine, within the sub-group of $IRF4^+$ cDC2 cells (**Figure A2.5I**), which has previously been shown to promote mucosal type 17 inflammation via secretion of IL-23 [48]. Further, this sub-grouping of cDC2 cells expresses high levels of *CCL17* and *CCL22*, chemokines involved in T cell chemotaxis (**Figure 3.4G**) [49].

Among the dermal dendritic cells, we identified 3 subgroups that are broadly distinguished from the conventional dendritic cell clusters by expression of *CLEC10A* (**Figure A2.5I**), which has been shown to influence T cell cytokine responses in skin [50, 51]. Cells from dermal DC sub-group 1 show elevated expression of *IL1R1*, *IL1R2*, *CCR7* and Fc-receptor including *FCER1A*, *FCGR2A*, and *FCGR2B*, which are important for interfacing with humoral immunity (**Figure A2.5I**) [52]. We observe a second population of dermal DCs (Dermal DC sub-group 2), with elevated expression of cathepsins (*CTSL* and *CTSB*), and surface receptors (*CD300E* and *SLC11A1*), which collectively represent markers of DC activation (**Figure A2.5I**) [53]. Finally, a third sub-grouping of dermal DCs (Dermal DC sub-group 3) was distinguished elevated expression of pro-inflammatory chemokines up-regulated during DC maturation (*CCL3*, *CCL4,* and *CCL5*) [54] and soluble mediators (*EREG* and *INHBA*).

Detection of Endothelial Heterogeneity and Vascular Addressin Expression

Multiple types of endothelial cells exist within the dermis of the skin. As in most tissues, arterioles shuttle oxygenated blood to tissues terminating in a capillary bed that gives rise to post-capillary venules. Importantly, $DARC^+$ post-capillary venules are the primary site of egress of

immune cells from circulation into tissues [55]. Using the improved sensitivity of Seq-Well S^3, we sought to understand the spectrum of endothelial cell diversity and vascular addressin expression across multiple instances of skin inflammation [56]. We performed sub-clustering and dimensionality reduction across 8,571 endothelial cells (**Figure 3.5A-B** and **Methods**) and identified three primary sub-clusters of dermal endothelial cells defined by distinct expression patterns: vascular smooth muscle cells (VSMCs; *TAGLN*), endothelial cells (*CD93*) and lymphatic endothelial cells (*LYVE1*) (**Figure 3.5C**). Importantly, we found multiple sub-clusters of CD93^{+} endothelial across normal and inflamed skin biopsies (**Figure 3.5A-B**). For example, we observe a cluster of DARC^{-}, CD93^{+} endothelial cells (Venule sub-cluster 3) that displays elevated expression of *SLC9A3R2*, which is involved in endothelial homeostasis [57], and another that is proliferating (Venule sub-cluster 6) (**Figure 3.5D**).

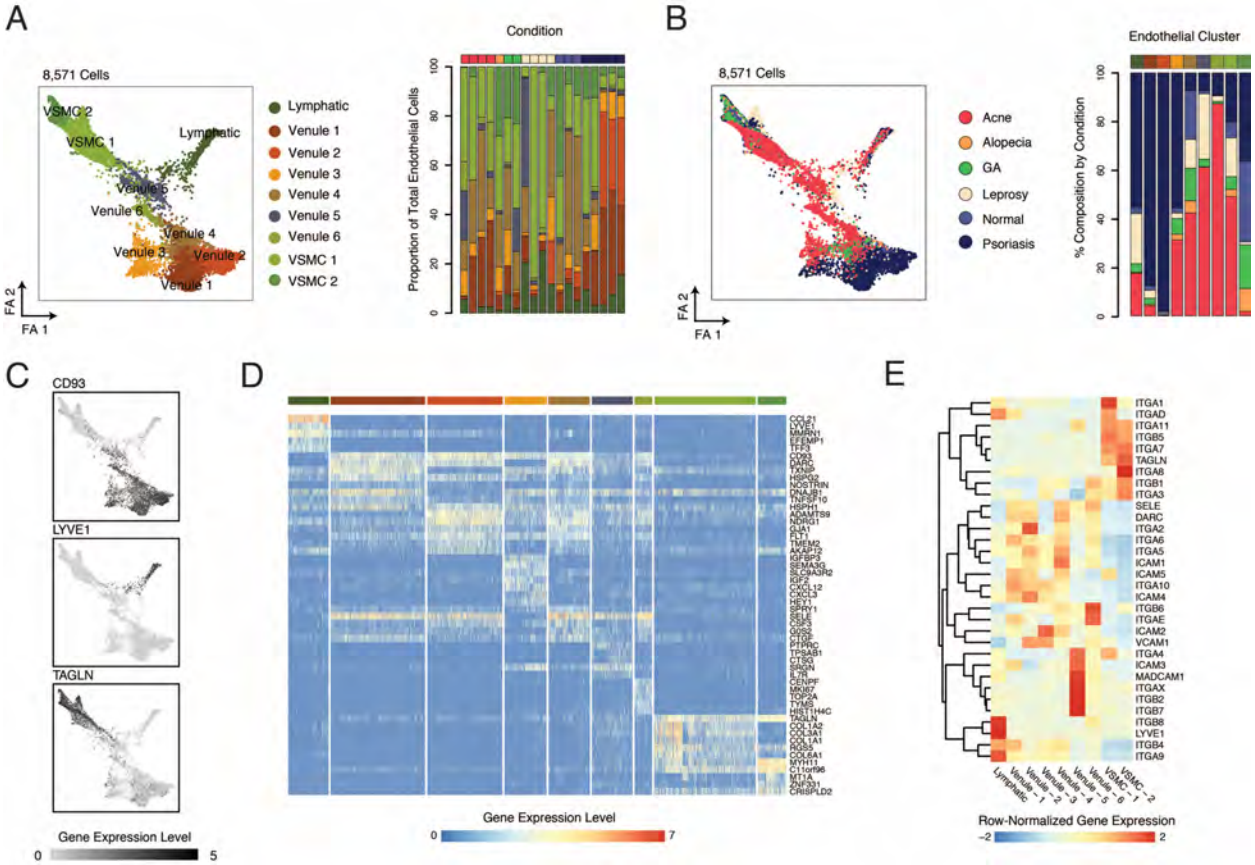


Figure 3.5 | A. Force-directed plots for 8,571 endothelial cells colored by phenotypic sub-cluster (left) and stacked barplot showing the distribution of endothelial phenotypic sub-clusters across samples (right). **B.** Force-directed plots for 8,571 endothelial colored by inflammatory skin condition (left) and stacked barplot showing the contribution of each inflammatory skin condition to endothelial phenotypic sub-clusters (right). **C.** Forced-directed plot colored by normalized expression level of genes that mark endothelial cell types: (Left) CD93, venules, (Middle) TAGLN, arterioles, (Right) LYVE1, lymphatics. **D.** Heatmap showing patterns of normalized gene expression levels (log(scaled UMI + 1)) across 9 clusters of endothelial cells. **E.** Heatmap showing row-normalized expression levels of vascular addressins across phenotypic sub-clusters of endothelial cells.

Further, we sought to understand the distribution of vascular addressins expressed by DARC$^+$ endothelial sub-populations (**Figure 3.5E**) [58]. Notably, across sub-populations of CD93$^+$ endothelial cells (Venule sub-clusters 1-6), we observe variation in expression of vascular addressins (**Figure 3.6E**). Among post-capillary venules, we observe broadly elevated expression of *ITGA5*, *ITGA6*, *ICAM2*, and *ITGA2*, while VSMCs express higher levels of *ITGA7*, *ITGA8*, and *ITGB5*. Further, we observe the highest expression of *ITGB4*, *ITGB8,* and *ITGA9*, among lymphatic endothelial cells (**Figure 3.5E**).

Altered Dermal Fibroblast Identities in Skin Inflammation

Dermal fibroblasts provide structural support and are the primary source of extracellular matrix components within the skin. Previous studies have demonstrated significant variation among dermal fibroblasts based on their relationship to anatomic features of the skin [59, 60]. To deeply catalogue diverse fibroblast cell states across inflamed skin, we performed dimensionality reduction and sub-clustering within the 7,237 fibroblasts identified across all samples and conditions (**Figure 3.6A-B; Methods**). In comparison to inflamed biopsies, fibroblasts from normal skin display enrichments in *LTBP4*, *IGFBP5*, and *TCF4* (Fibroblast Clusters 2 and 8). Consistent with previous single-cell studies of dermal fibroblasts, we observe a sub-population of fibroblasts (Fibroblast Cluster 3) that express *COL11A1*, *DPEP1* and *RBP4,* where these cells were suggested to have a role in connective tissue differentiation (**Figure 3.6C**) [20].

In granuloma annulare (GA), we observe two distinct fibroblast populations. Specifically, fibroblasts from GA patient 1 (sub-cluster 0) display elevated expression of protease inhibitor 16 (*PI16*), which inhibits the function of MMP2 [61], and *ITIH5,* a protease inhibitor important for maintenance of dermal hyaluronic acid that is overexpressed in skin inflammation (**Figure 3.6C-D**) [62]. Fibroblasts from GA patient 2 (sub-cluster 7), meanwhile, express elevated levels of *SPOCK1* (Avg-Log FC: 0.99), *CRLF1 (*Avg-Log FC: 1.38), and *COMP* (Avg-Log FC: 1.35), a cartilage protein that is upregulated in matrix-producing fibroblasts following myocardial infarction [63] (**Figure 3.6C-D**).

We also observe distinct fibroblast phenotypes in leprosy infection. Specifically, we find a population of fibroblasts (Fibroblast Cluster 1) marked by combined expression of and *POSTN* (Periostin) and *MMP11*, a marker of fibroblasts in basal cell carcinoma [64] (**Figure 3.6D).** In another leprosy biopsy, we observe a population of pro-inflammatory fibroblasts (Fibroblast Cluster 5) that express elevated levels of *SFRP2*, *PRSS23* and *IL6* (**Figure 3.6D**). Finally, among all 5 psoriasis biopsies, we observe a population of pro-inflammatory fibroblasts (Fibroblast Cluster 4) marked by elevated expression of *CCL19*, *TNFSF13B* (BAFF), and *CXCL12* (**Figure 3.6C-D**). Notably, expression of *CCL19* and *BAFF* by synovial fibroblasts have been implicated in the progression of rheumatoid arthritis [65, 66].

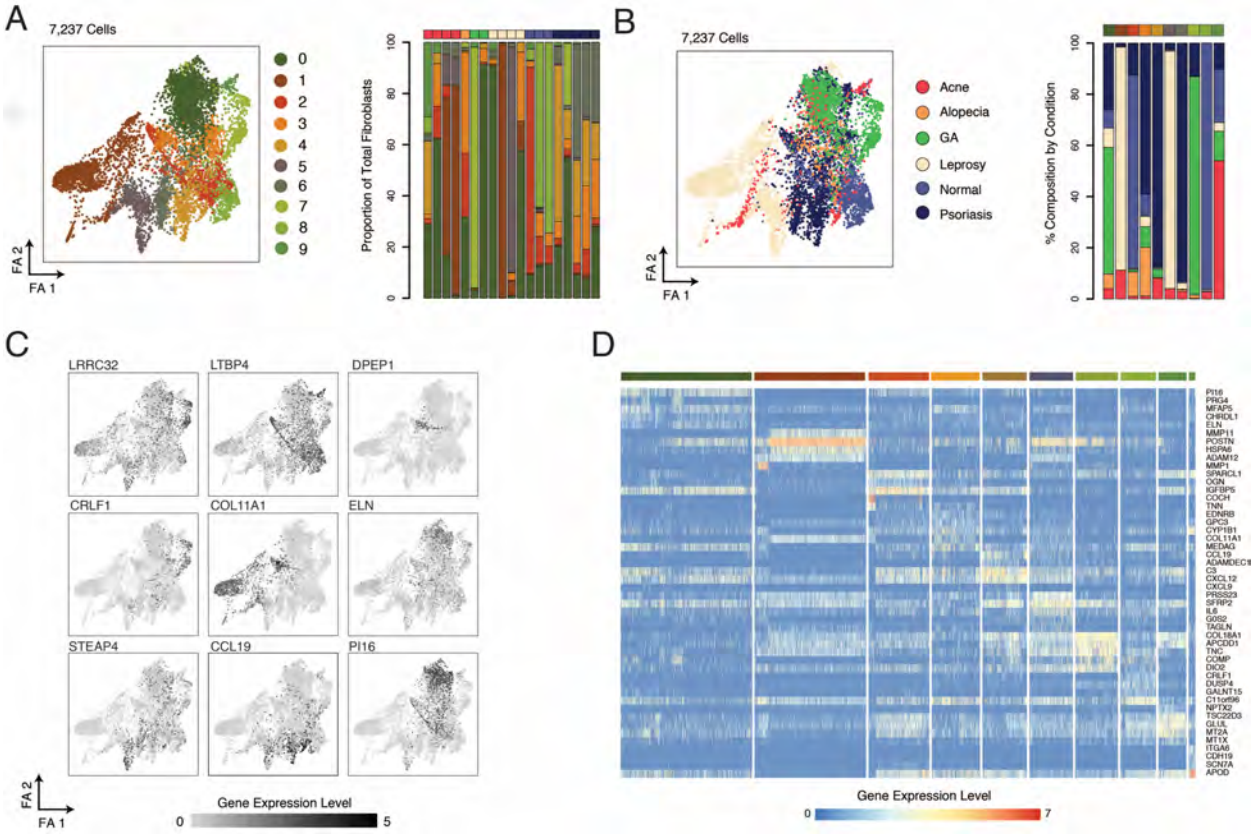


Figure 3.6 | A. Force-directed plots for 7,237 fibroblasts colored by phenotypic sub-cluster (left) and stacked barplot showing the distribution of fibroblast phenotypic sub-clusters across samples (right). **B.** Force-directed plots for 7,237 fibroblasts colored by inflammatory skin condition (left) and stacked barplot showing the contribution of each inflammatory skin condition to fibroblast phenotypic sub-clusters (right). **C.** Force-directed graphs highlighting fibroblast cluster defining genes. **D.** Heatmap showing the normalized gene expression levels (log(scaled UMI + 1)) of fibroblast cluster-defining genes.

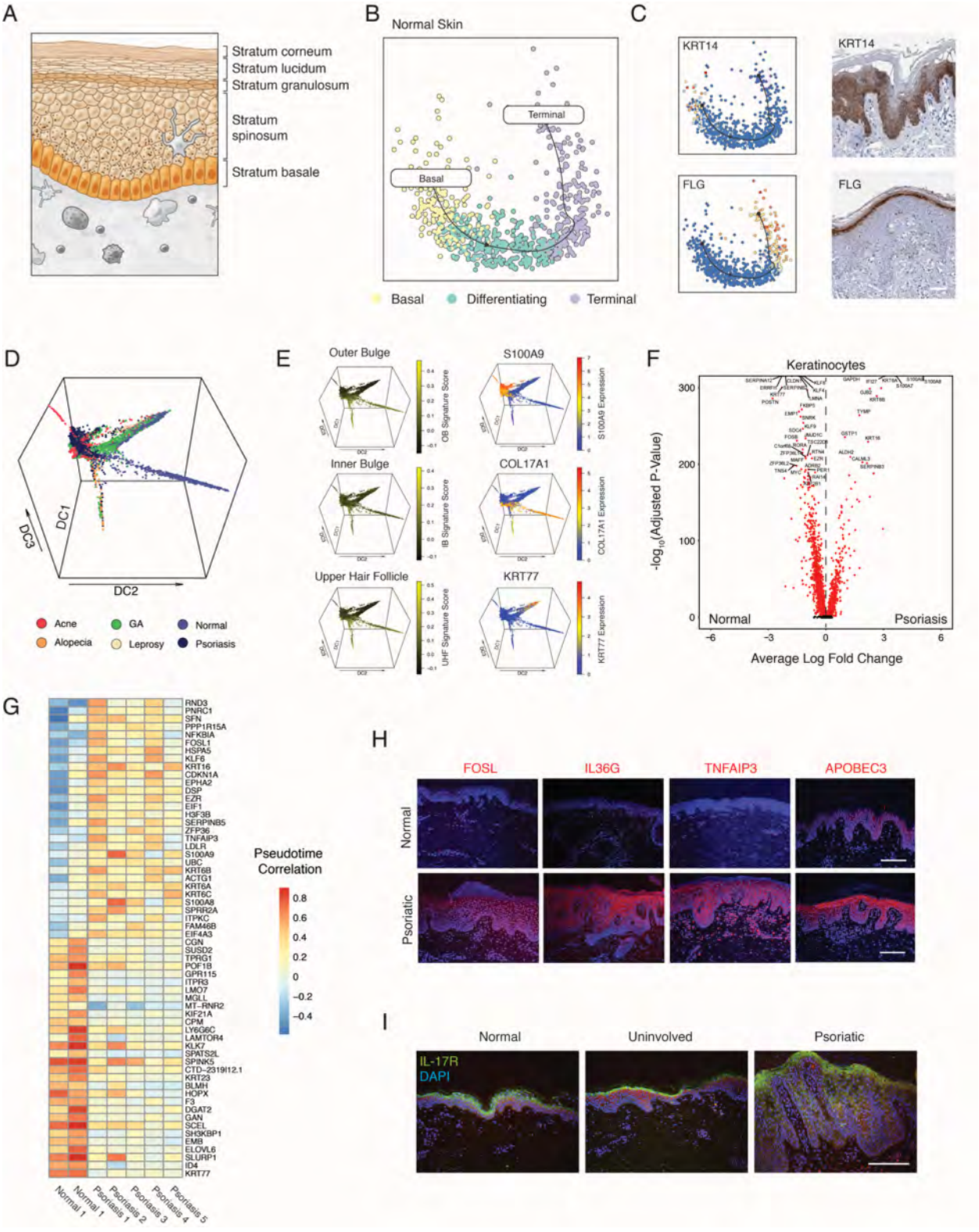
Keratinocyte Differentiation Trajectories

Within the epidermis, keratinocytes undergo a stereotyped differentiation process in which cells acquire altered morphology and phenotype as they mature (**Figure 3.7A**) [67]. Under physiologic conditions, basal keratinocytes continuously divide to give rise to the remaining cells of the epidermis [68]. Using keratinocytes from normal skin, we performed pseudo-temporal analysis to reconstruct the differentiation process of normal epidermal keratinocytes (**Figure 3.7B**; **Methods**) [69]. More specifically, in normal skin, we first identified a population of keratinocytes enriched for expression of *KRT14*, a marker of basal keratinocytes (**Figure 3.7C**) [70]. We then used known patterns of cytokeratin expression to infer localization of keratinocytes along a

supervised differentiation trajectory (**Figure 3.7C** and **A2.6A**) [4]. Our trajectory analysis revealed patterns of transcription factor and cytokeratin expression that closely correspond to previously established signatures of keratinocyte maturation in both normal skin samples for which we performed trajectory analysis (**Methods**; **Figure A2.6A-B**) [21]. Consistent with immunohistochemical staining from the Human Protein Atlas (**Figure 3.7C**) [71], we observed enriched expression of filaggrin (*FLG*), a protein in the outer layers of the epidermis [72], among keratinocytes that lie at the terminal points in the pseudo-temporal ordering (**Figure 3.7C** and **Figure A2.6B**).

Figure 3.7 | Keratinocyte Differentiation Trajectories. A. Diagram showing the layers of the epidermis and morphologic changes associated with keratinocyte differentiation. **B.** t-SNE plot showing differentiation trajectory of keratinocytes from normal skin from basal cells (yellow) through differentiating cells (aqua) and terminal keratinocytes (purple). **C.** (Top-left) tSNE plot of normal keratinocytes with normalized *KRT14* expression values overlayed. (Top-right) Immunohistochemistry staining showing the expression of *KRT14* from the human protein atlas (Uhlén et al., 2015). (Bottom-left) tSNE plot of normal keratinocytes with normalized *FLG* expression values overlayed. (Bottom-right) Immunohistochemistry staining of *FLG* from the human protein atlas(Uhlén et al., 2015). Scale bars = 50 microns. **D.** Diffusion map of 10,777 keratinocytes colored by inflammatory skin condition. Axes correspond to diffusion components 1,2, and 3. **E.** Diffusion map of keratinocytes colored by signatures of hair follicle-specific gene expression (Left: Outer Bulge, Inner Bulge and Upper Hair Follicle) and genes that distinguish basal (COL17A1), normal (KRT77) and inflamed (S100A9) keratinocytes. **F.** Volcano plot showing genes differentially expressed between psoriatic and normal keratinocytes. **G.** Heatmap showing gene-specific Pearson correlation values between diffusion pseudotime and gene expression for 2 normal skin biopsies and 5 psoriatic biopsies. **H.** (Top) Immunofluorescence staining in normal (above) and psoriatic (below) for FOSL, IL36G, TNFAIP3, and APOBEC3. All images stained for nuclei (DAPI) and gene of interest (Red Fluorescence). Scale bar = 100 microns. **I.** Immunofluorescence staining for IL-17R expression (green) in normal (left), uninvolved (middle), and psoriatic skin (right). Scale bar = 100 microns.

Continued


A
Stratum corneum
Stratum lucidum
Stratum granulosum
Stratum spinosum
Stratum basale
B
Normal Skin
Terminal
Basal
Basal
Differentiating
Terminal
C
KRT14
KRT14
FLG
FLG
D
DC1
DC3
DC2
Acne
Alopecia
GA
Leprosy
Normal
Psoriasis
E
Outer Bulge
S100A9
Inner Bulge
COL17A1
Upper Hair Follicle
KRT77
OB Signature Score
IB Signature Score
UHF Signature Score
S100A9 Expression
COL17A1 Expression
KRT77 Expression
F
Keratinocytes
-log10(Adjusted P-Value)
Normal
Psoriasis
Average Log Fold Change
G
Pseudotime Correlation
Normal 1
Normal 1
Psoriasis 1
Psoriasis 2
Psoriasis 3
Psoriasis 4
Psoriasis 5
H
FOSL
IL36G
TNFAIP3
APOBEC3
Normal
Psoriatic
I
Normal
Uninvolved
Psoriatic
IL-17R
DAPI

Having established a trajectory for normal keratinocyte differentiation, we next examined patterns of keratinocyte differentiation across pathologic conditions. To identify conserved and unique patterns across conditions, we constructed a combined diffusion map among the 10,777 keratinocytes of the epidermis and hair follicle recovered across all 19 skin biopsies (**Figure 3.7D-E**; **Methods**). While keratinocytes from most conditions closely align with normal differentiation, we observe marked deviation in the differentiation trajectory of psoriatic keratinocytes (**Figure 3.7D**). We further observe distinct trajectories for basal keratinocytes (*COL17A1*) and cells of the hair follicle, where we see enrichment of published hair follicle signatures (**Figure 3.7E**) [73]. Consistent with previous observations, differential expression analysis reveals significant up-regulation of antimicrobial peptides (*S100A7*, *S100A8*, *S100A9*) and pro-inflammatory cytokines (*IL36G*, *IL36RN*) in psoriatic keratinocytes (**Figure 3.7E-F**) [74].

Based on increased sensitivity of Seq-Well S^3 to detect transcription factors observed in peripheral lymphocytes, we hypothesized that our data might enable identification of novel transcriptional regulators of psoriatic keratinocytes. To identify potential drivers of the psoriatic disease process within the epidermis, we performed differential pseudo-time correlation analysis between psoriatic and normal keratinocytes (**Methods**). Specifically, we separately constructed pseudo-time trajectories for normal (n=2) and psoriatic keratinocytes (n=5), calculated correlation values between diffusion pseudo-time and gene expression levels, and examined the difference in correlation values between psoriatic and normal keratinocytes (**Figure 3.7G** and **A2.6A-B**). Notably, we observed positive correlation of *FOSL1*, an AP-1 transcription factor, with diffusion pseudo-time in psoriatic keratinocytes, implying that *FOSL1* is aberrantly expressed along the differentiation trajectory of psoriatic keratinocytes. To validate this observation, we performed immunofluorescence staining for FOSL1 protein, and measured increased levels of FOSL1 in psoriatic skin (**Figure 3.7H; Methods**). We validated the distribution of additional genes overexpressed or differentially correlated with diffusion pseudo-time in psoriatic keratinocytes

(including *TNFAIP3, IL36G*, and *APOBEC3*) at the protein level (**Figure 3.7H** and **A2.6A**; **Methods**). Finally, we examined the relationship between differential expression and difference in pseudotime correlation (**Figure A2.6C**). Here, we observe no overall relationship between differential expression and differential pseudotime, suggesting a more complicated picture of dysregulated gene expression in psoriatic keratinocytes.

To further define differences in gene expression patterns between normal and psoriatic keratinocytes, we scored the expression levels of known cytokine response signatures using a series of reference signatures gene lists derived from population RNA-Seq of cultured keratinocytes exposed to multiple cytokines including IL-17A, IL-4, IL-13, TNFα, INFα and IFNγ (**Figure A2.6D**, **Methods**) [75]. While IL-17 has been previously implicated in the pathogenesis of psoriasis, here we infer the identity of cells that dominate the IL-17 response, localizing the expression of IL-17 responsive genes to spinous keratinocytes [76]. To validate this observation, we performed immunofluorescent staining for IL-17R protein and measured the highest staining within spinous keratinocytes exclusively within psoriatic skin (**Figure 3.7I; Methods**). Collectively, these data provide novel insights into the localization IL-17 response in psoriatic keratinocytes.

Discussion

Here, we present an enhanced technique for high-throughput scRNA-Seq – Seq-Well S^3 – that affords improved sensitivity for transcript capture and gene detection. Through use of a templated second-strand synthesis, S^3 recovers information typically lost in bead-based high-throughput scRNA-Seq protocol such as Seq-Well or Drop-Seq. Specifically, S^3 reclaims mRNA molecules that are successfully captured and reverse transcribed but not labeled with a second primer sequence through template switching (**Figure 3.1 and A2.1**). Using Seq-Well S^3, we obtain a 5-10 fold increase in the number of unique molecules captured from cells at similar sequencing depth relative Seq-Well v1 (**Figures 3.1, A2.1** and **A2.2**) [24]. Beyond aggregate increases in the number of transcripts recovered per-cell, the improvements in sensitivity made

possible by Seq-Well S^3 enable enhanced detection, and thus deeper examination, of lineage-defining factors in immune and parenchymal cells – such as transcription factors, cytokines, and cytokine receptors among lymphocytes (**Figure 3.1 and A2.2**) – which are often transiently or lowly expressed [77]. Among $CD4^+$ T cells isolated from PBMCs, for example, we observed rates of gene detection similar to those observed in Smart-Seq2, a best-in-class microtiter plate-based method (**Figure 3.1F-G** and **A2.2F**).

Similarly, using Seq-Well S^3, we report improved paired detection of α and β TCR sequences from T cells in peripheral blood and tissue biopsies (**Figures 3.3G** and **A2.5C**). Among $CD4^+$ T cells from PBMCs, we recover paired TCR α and β constant genes in 87.1% of cells. Together with targeted enrichment, amplification and sequencing, we anticipate that Seq-Well S^3 will enable improvements in TCR reconstruction and deep characterizations of clonotype-phenotype relationships at scale [78]. Collectively, our validation experiments show that Seq-Well S^3 significantly augments the amount of information that can be recovered in massively-parallel scRNA-seq experiments, enabling high-resolution profiling of low-input biopsy samples at scale.

With this enhanced method, here, we move towards a draft atlas of human skin inflammation by creating a compendium of cell-types and states for the broader research community [79]. Through use of Seq-Well S^3, we survey, at unprecedented resolution, the diversity of cell-types and states – e.g., among tissue resident T cells and myeloid cells – present across multiple types of skin inflammation. For example, GA and leprosy are two granulomatous diseases characterized by aggregates of lymphocytes and macrophages within the dermis, which are thought to arise from a delayed-type hypersensitivity response to *M. leprae* infection (leprosy) or an unknown agent (GA) [80, 81]. Here, we find that both are characterized by the presence of Immature $CD8^+$ CTLs and a cluster of cytotoxic cells containing $CD8^+$ T-CTL, $\gamma\delta$ and NK cells; **Figure 3.3**). Although all conditions contain $CD163^+$ dermal macrophages and various DC subpopulations, we observe a unique macrophage population in leprosy defined by expression of

extracellular proteases [82, 83] (Fulco et al., 2014; Verreck et al., 2004). Moreover, GA uniquely contained specific populations of fibroblasts expressing *SPOCK1*, *CRLF1,* and *COMP* (**Figure 3.6**), which likely reflect remodeling of the dermis with mucin deposition and alternation of elastin fibers [84, 85].

Acne, meanwhile, is an inflammatory disease thought to arise in response to infection with *P. acnes*, resulting in the formation of lesions that resemble a wound following eruption of the hair follicle into the dermis [86]. Here, we observe 2 clusters of endothelial cells marked by expression of *SLC9A3R2*, a marker of endothelial homeostasis, and a signature of proliferation (Venule clusters 3 and 4, **Figure 3.5**) [57]. This increased angiogenesis and endothelial proliferation is most consistent with the proliferative phase of wound healing in acne [87].

In psoriasis, T cells are thought to be a primary driver of inflammation, with dendritic cells playing a central role in the recruitment and polarization of T cells that contribute to the hyperproliferation of keratinocytes in the disease [29]. Across 5 patients with psoriasis, we report a sub-cluster of DCs ($IRF4^+$ cDC2) that displays elevated expression of *CCL17*, *CCL22* and *IL12B* (**Figure 3.4G** and **Figure A2.5I**). Importantly, a similar population of dermal cDC2 cells has recently been shown to drive psoriatic inflammation in mice and humans through the recruitment of inflammatory T cells [88, 89]. Although we detected a diversity in T cell subtypes in psoriatic lesions, we note few Th-17-like cells [90]. We further report a conserved population of fibroblasts in psoriasis that express pro-inflammatory cytokines and chemokines including CCL19 and BAFF, which have been implicated in Rheumatoid Arthritis and systemic sclerosis [65, 91].

Leveraging the increased sensitivity of Seq-Well S^3, we performed pseudo-time correlation analysis to uncover an altered differentiation trajectory of keratinocytes compared to normal skin (**Figure 3.7** and **A2.6**). From our pseudo-time correlation analysis, we detected dysregulated expression of FOSL1 in biopsies from 5 patients with psoriasis as a putative transcription factor involved in psoriatic differentiation, a finding which we validated through

immunofluorescent staining of healthy and psoriatic skin (**Figure 3.7G**). Further, previous studies using *in vitro* keratinocyte based systems have suggested that more differentiated keratinocytes were the main responders to IL-17A, given larger effect sizes in differentiated compared to monolayer keratinocyte [76]. Using data generated with Seq-Well S^3 cross-analyzed against an IL-17 response signature in keratinocytes, we show that IL-17 responses are observed in keratinocytes from all layers of the epidermis, but that these responses are stronger in keratinocytes derived from more differentiated layers of the psoriatic epidermis (**Figure S7D**). This observation is corroborated by co-localization of the IL-17 receptor subunits (IL-17RA/IL-17RC) in the upper layers of psoriatic epidermis (**Figure 3.7I**).

Cost is a key factor in the selection of a scRNA-Seq method. Prior to sequencing, the cost of the Seq-Well S^3 protocol is significantly less than those of commercial scRNA-Seq platforms. Specifically, the pre-sequencing cost of a single Seq-Well array is $343, while the pre-sequencing costs for 10x genomics V2/V3 are $2,100 (**Methods**). Sequencing data for Seq-Well S^3 libraries are generated using Illumina NextSeq500/550 75-cycle kits ($1,570/kit), while Illumina NextSeq500/550 150-cycle kits are needed for 10x Genomics V2/V3 ($3,010 /kit). When comparing alignment rates between Seq-Well S^3 and 10x Genomics v3, we observe lower alignment rates for Seq-Well S^3 (52.3% alignment rate) as compared to 10x Genomics V3 (89.5% alignment rate), which may be due, in part, to differences in read length. However, when considering both differences in efficiency and the costs of the two sequencing kits, both methods require approximately the same amount of money to obtain 50k aligned reads ($0.30 per cell for Seq-Well S^3 versus $0.33 for 10x v3). Finally, in either technology, the cost of processing per sample can be further reduced through sample multiplexing [92, 93].

In conclusion, we describe a powerful massively-parallel scRNA-Seq protocol that enables improved transcript capture and gene detection from low-input clinical samples. Seq-Well S^3 provides a competitive alternative for high-throughput scRNA-Seq that is compatible with clinical

study designs due to flexible stopping points (post-reverse transcription), has a high degree of technical reproducibility, and is compatible with targeted enrichment of molecules of interest [78, 94]. Here, Seq-Well S^3 provides novel insights into putative mechanisms and the cellular localization of previously appreciated and unknown responses to specific inflammatory mediators in immunologic skin conditions in a fashion not previously achievable. Increases in the sensitivity of gene and transcript detection are increasingly important as single-cell atlasing efforts shift from detection of large differences between cell types within normal tissue to identification of subtle differences in cell state across cell types within diseased tissues. The increased sensitivity of gene detection and transcript capture afforded by S^3 enhances the strength of inferences that can be drawn from these types of single-cell data, as evidenced by the range of immune, stromal and parenchymal cell states uncovered in human skin inflammation. The S^3 protocol (**Supplementary Seq-Well S^3 Protocol - Appendix 3**) is easy to integrate into current bead-based RNA-Seq platforms, such as Drop-Seq, making it broadly useful for the single-cell community, particularly in the setting of human disease. Importantly, S^3's increases in library complexity and sequencing efficiency reduce costs relative to plate-based protocols, and providing researchers with a powerful and cost-effective alternative to commercial solutions in a format that can be deployed almost anywhere.

Materials and Methods

Single-Cell Processing Pipeline

We utilized Seq-Well, a massively-parallel, low-input scRNA-Seq platform for clinical samples, to capture the transcriptome of single cells. A complete, updated protocol for Seq-Well S^3 is included as a Supplementary Protocol (**Appendix 3**) and is hosted on the Shalek Lab website (www.shaleklab.com). Briefly, 10-15,000 cells were loaded onto a functionalized-polydimethylsiloxane (PDMS) array preloaded with uniquely-barcoded mRNA capture beads (Chemgenes; MACOSKO-2011-10). After cells had settled into wells, the array was then sealed with a hydroxylated polycarbonate membrane with pore sizes of 10 nm, facilitating buffer exchange while confining biological molecules within each well. Following membrane-sealing, subsequent buffer exchange permits cell lysis, mRNA transcript hybridization to beads, and bead removal before proceeding with reverse transcription. The obtained bead-bound cDNA product then underwent Exonuclease I treatment (New England Biolabs; M0293M) to remove excess primer before proceeding with second strand synthesis.

Templated Second Strand Synthesis

Following Exonuclease I treatment, beads were washed once with 500uL of a TE-SDS (0.5% SDS) solution, and twice in 500uL of a TE-Tween (0.01% Tween) solution. After the second TE-TW wash the beads were solvated with 500uL of 0.1M NaOH and mixed for 5 minutes at room temperature using an end-over-end rotator with intermittent agitation to denature the mRNA-cDNA hybrid product on the bead. Following denaturing, the NaOH was removed and beads were washed once with 1M TE, and then combined with a mastermix consisting of 40uL 5x maxima RT buffer, 80uL 30% PEG8000 solution, 20uL 10mM dNTPs, 2uL 1mM dn-SMART oligo, 5uL Klenow Exo-, and 53ul of DI ultrapure water. Second strand synthesis was carried out by incubating the beads for 1 hour at 37°C with end-over-end rotation and intermittent agitation. Following incubation, beads were sequentially washed twice with 0.5 mL of TE buffer with 0.01% Tween

20, and once with 0.5 mL of TE. Immediately prior to PCR amplification, beads were washed once with 0.5 mL of water and resuspended in 0.5 mL of water.

PCR Amplification

After second strand synthesis, PCR amplification was performed using KAPA HiFi PCR Mix (Kapa Biosystems KK2602). Specifically, a 40uL PCR Mastermix consisting of 25 uL of KAPA 5X Mastermix, 0.4 uL of 100 uM ISPCR oligo, and 14.6 uL of nuclease-free water was combined with 2,000 beads per reaction. For each sample, the total number of PCR reactions performed varied based on the number of beads recovered following second strand synthesis. PCR amplification was performed using the following cycling conditions: an initial denaturation at 95°C for 3 minutes, then 4 cycles of 98°C for 20 seconds, 65°C for 45 seconds, and 72°C for 3 minutes, followed by 9-12 cycles of 98°C or 20 seconds, 67°C or 20 second, and 72°C for 3 minutes, and then a final extension of 72°C for 5 minutes. Following PCR amplification, WTA products were isolated through two rounds of SPRI purification using Ampure Spri beads (Beckman Coulter, Inc.) at both 0.6X and 0.8x volumetric ratio and quantified using a Qubit.

Optimization of Second Strand Synthesis

We performed a series of experiments to validate the performance of the second-strand synthesis protocol relative other techniques. For the comparison of the Seq-Well protocol with and without second-strand synthesis, we performed species-mixing experiments and PBMC comparisons. For species-mixing experiments, we applied a mixture of 5,000 HEK293 and 5,000 NIH-3T3 cells to a loaded Seq-Well device, while for PBMC comparisons, we loaded a total of 10,000 PBMCs. In optimization experiments, PBMCs were thawed and immediately loaded directly onto Seq-Well devices without stimulation. Following bead removal, beads were split into separate reverse transcription reactions with and without the template-switching oligo. After reverse transcription and ExoI treatment, beads for each comparison were processed separately with and without the second-strand synthesis protocol.

Specifically, we performed a series of optimization experiments to validate the effectiveness of Seq-Well S^3. Specifically, we performed a series of control experiments using beads from a single Seq-Well array loaded with 10,000 PBMCs. For each array we split the beads into six equal fractions and performed the following controls: (1) we performed PCR amplification without the use of second-strand synthesis. (2) we performed random second-strand synthesis followed by PCR amplification. (3) we omitted the template switching oligo without the use of second-strand synthesis. (4) we omitted the template switching oligo but subsequently performed random second-strand synthesis. (5) we examined the effect of heat inactivation of the reverse transcription reagent without the use of second strand synthesis. (6) we examined the effect of heat inactivation of the reverse transcription reagent followed by random second strand synthesis (**Figure A2.1B-C**). Following PCR amplification, products were obtained from all conditions with the exception of Condition 3 (Seq-Well V1/ No TSO), which did not yield appreciable WTA product.

Comparison of 10X Genomics (V2 Chemistry) and Seq-Well S^3

Human PBMC were thawed and rested overnight. Cells were stimulated for 18 hours by adding aCD3 (UCHT1) and aCD28 (CD28.2) antibodies to the bulk PBMC culture at a concentration of 1mg/mL and 5 mg/mL, respectively, and $CD4^+$ T cells were enriched following stimulation using magnetic negative selection (Stemcell Technologies). Following isolation, T cells were stained with calcein violet live stain (Thermo), Sytox dead stain (Thermo), and aCD45-AF647 (HI30) antibody at 4C for 30 minutes. After two washes, aliquots of the cells were placed on ice and sorted directly into RLT buffer using a Sony SH800S Sorter for Smart-Seq2 processing and another unstained sample for 10X Chromium analysis. Once the cells were delivered, a third aliquot was loaded onto a Seq-Well array. Single-cell libraries were generated using the Smart-Seq 2, 10X V2, and Seq-Well S^3 protocols.

Sequencing Library Preparation

A total of 1ng of WTA product at a concentration of 0.2 ng/uL was combined with 10 uL of Buffer TD and 5 ul of Buffer ATM and incubated at 55°C for 5 minutes. Following tagmentation, 5 uL of Buffer NT was added and incubated at room temperature for 5 minutes to neutralize the reaction. A total of 8 uL of nuclease-free water, 15uL of buffer NPM, 1 uL of Custom P5 hybrid Oligo, and 1 uL of N700 Index oligo were combined and PCR amplification was performed using the following cycling conditions: an initial denaturation of 95°C for 30 seconds, then 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, followed by a final extension of 72°C for 5 minutes. PCR products were isolated through two rounds of SPRI purification (0.6x and 0.8x volumetric ratios) and quantified using a Qubit. Library size distributions were determined using an Agilent Bioanalyzer D1000 High Sensitivity Screen tape.

DATA ANALYSIS OF COMPARISON EXPERIMENTS

DNA Sequencing and Alignment of PBMC Optimization samples and Downsampling

PBMC optimization experiments were all sequenced on NextSeq500 75 cycle kits. Sequencing read alignment was performed using version 1 of the Drop-Seq pipeline (Macosko et al., 2015). NextSeq runs were loaded at a final concentration of 2.2pM using NextSeq 550 v2 sequencing kits at the Ragon Institute. Briefly, for each sequencing run, raw sequencing reads were converted from bcl files to FASTQs using bcl2fastq and demultiplexed using Nextera N700 indices that corresponded to individual samples. Demultiplexed FASTQs were then aligned using an implementation of DropSeqTools v1.0 maintained by the Broad Institute for data analysis, and aligned to the Hg19 genome using standard parameters. Individual reads were tagged with a 12-bp barcode and 8-bp unique molecular identifier (UMI) contained in Read 1 of each sequencing fragment. Following alignment, aligned read 2 sequences were grouped by the 12-bp cell barcodes and subsequently collapsed by the 8-bp UMI for digital gene expression (DGE) matrix extraction and generation.

PBMC Comparison Experiments

We generated data matrices for PBMC data from 10x genomics and Seq-Well S^3. Initially, we performed downsampling to an average sequencing depth of 42,000 reads per cell. Specifically, downsampling was performed on Seq-Well S^3 to match the sequencing depth of 10x Genomics v2. For each data set, we performed variable gene identification and selected variable genes for downstream analysis (Seq-Well S^3, 856 variable genes and 10x Genomics v2, 516 genes). We performed principal components analysis and selected the first 20 principal components to perform a t-SNE dimensionality reduction. We then performed cluster identification and discovered clusters representing $CD4^+$ T cells, $CD8^+$/NK cells, and B cells for each of the technology platforms (**Figure A2.2A**). We examined the proportion of cell types recovered between Seq-well S^3 and 10x Genomics v2 and performed a Chi-Square test to examine differences in the proportion of recovered cell types ($P = 0.971$).

Within each cell type identified between Seq-Well S^3 and 10x Genomics V2, we examined differences in aggregate gene detection and transcript capture (**Figure A2.2C**). We initially performed a Lilliefors test to assess normality of the distribution of genes and UMIs for each technique. Based on these results, we determined to use a Mann-Whitney U Test to determine difference in aggregate gene and transcript detection between techniques (**Figures 3.1, A2.1, and A2.2**). As a measure of library complexity, we examined the linear relationship between the number of UMIs captured and aligned sequencing reads. Specifically, across cell types for each technique, we plotted the number of UMIs against the number of aligned reads and calculated the slope of the regression line for each condition (**Figure 3.1B-C**). For comparisons of library complexity, we constructed a multivariable linear regression model in which the number of transcripts per cell was modeled follows: nUMI ~ Intercept + B1*nReads + B2*Technique + B3*nReads*Technique. From these models, we determined statistical significance of the difference in slope (i.e. library complexity) based on p-values for the interaction

term B3*nReads*Technique, the magnitude and significance of which correspond to a difference in slope (i.e. library complexity or the number of UMIs per aligned read) (**Figure 3.1B** and **A2.1B-C**). For example, in a library of low-complexity application of additional sequencing reads might result in detection of a new transcript in every 20th aligned read (i.e. slope = 0.05). Conversely, a library of high complexity, might result in detection of a new transcript with every 4 aligned reads (i.e. slope = 0.25). Critically, these comparisons should be performed on libraries that have been sequenced or down-sampled to similar depths as over-sequencing can augment the relative perception of differences in library complexity. Specifically, libraries that have been "over-sequenced" will appear to have lower complexity because unique molecular identifiers will eventually accumulate additional reads upon saturation.

Comparison of Gene Detection Rates

For each cell-type cluster, we calculated the rate of detection for each gene as the proportion of cells with a non-zero expression value. Gene detection rates were separately calculated across $CD4^+$ T cells, B cells, $CD8^+$/NK cells, and monocytes for both Seq-Well S^3 and 10x Genomics v2. We further examined differences in gene detection rates among transcription factors, cytokines and surface receptors. For comparisons of relationship between gene-detection rates and overall expression levels, we calculated the expression level of individual genes as the average normalized expression value within each cell type for all cells identified in both Seq-Well S^3 and 10x v2 data (**Figure 3.1 and A2.2**). To test the statistical significance of differences in gene detection frequencies, we performed a chi-square test using the number of cells in which a given gene had a non-zero expression values for each technique. We further performed comparisons between Seq-Well S^3 and 10x Genomics v3 (**Figure A2.3**).

PROFILING CELL STATES IN HUMAN SKIN INFLAMMATION

IRB Statement

Informed written consent was obtained from human subjects under a protocol approved by the institutional review boards of the University of Michigan and University of California Los Angeles (UCLA). This study was conducted according to the Declaration of Helsinki Principles.

Processing of Human Skin

Skin biopsies were obtained from a total of 9 patients at the University of California, Los Angeles and University of Southern California Hansen's Clinic. For each sample, a 4-mm punch biopsy was obtained following local anesthesia and was placed immediately into 10 mL of RPMI on ice. Initially, skin biopsies were incubated in 5mL of a 0.4% Dispase II solution (Roche Inc.) at 37°C for 1 hour with vigorous shaking. The dermis and epidermis were then carefully separated using forceps and transferred to separate tubes for additional processing. Epidermal samples were placed in 3mL of 0.25% Trypsin and 10U/mL DNAse for 30 minutes at 37°C. Trypsin was neutralized with 3mL of fetal calf serum (FCS), and the tissue was passed through a 70-micron nylon cell strainer which was washed with 5mL of RPMI. Epidermal cells were then pelleted at 300xg for 10 minutes and counted. Dermal samples were minced with a scalpel and incubated in a solution of 0.4% collagenase 2 and 10 U/mL DNAse for 2 hours at 37°C with agitation. The cell suspension was passed through a 70-micron cell strainer and washed with 5mL of RPMI. Cells were pelleted at 300xg for 10 minutes, resuspended in 1mL of RPMI and counted. MACS enrichment for $CD207^{+}$ cells was performed for epidermal fractions from biopsies from normal skin 1 and leprosy 1.

Sequencing and Alignment of Skin Samples

Sequencing read alignment was performed using version 2 of the Drop-seq pipeline previously described in Macosko et al. Briefly, for each Nova-Seq sequencing run, raw sequencing reads were converted from bcl files to FASTQs using bcl2fastq based on Nextera N700 indices that corresponded to individual samples. Demultiplexed FASTQs were then aligned to the Hg19 genome using STAR and the DropSeq Pipeline on a cloud-computing platform

maintained by the Broad Institute. Individual reads were tagged with a 12-bp barcode and 8-bp unique molecular identifier (UMI) contained in Read 1 of each sequencing fragment. Following alignment, reads were grouped by the 12-bp cell barcodes and subsequently collapsed by the 8-bp UMI for digital gene expression (DGE) matrix extraction and generation.

Tissue Immunofluorescence Staining

Formalin fixed, paraffin-embedded tissue slides obtained from psoriasis patients and normal controls were heated for 30 min at 60°C, rehydrated, and epitope retrieved with Tris-EDTA, pH 6. Slides were blocked, incubated with primary antibody APOBEC3 (LS-C98892-400; Lifespan bioscience), FOSL (A03927; Boster), IL-36G (sc-80056; Santa Cruz Biotechnology), TNFAIP3 (ab74037, Abcam), IL-17RC (LS-C400522, Lifespan bioscience), and IL-17RA (LS-C359381, Lifespan bioscience) overnight at 4 °C. Slides were then washed and incubated with Donkey anti-Rabbit IgG 594, Donkey anti-Mouse IgG 488, or Donkey anti-Rat IgG 594 (all from Invitrogen) for 1 h at room temperature. Slides were washed and prepared in mounting medium with 4',6-diamidino-2-phenylindole (DAPI) (VECTASHIELD Antifade Mounting Medium with DAPI, H-1200, VECTOR). Images were acquired using Zeiss Axioskop 2 microscope and analyzed by SPOT software 5.1. Images presented are representative of at least three experiments from separate donors.

DATA ANALYSIS OF SKIN SAMPLES

Cell Quality Filtering

Cells were initially filtered on the basis of gene detection (> 500 genes per cell) and transcript detection (> 700 umis per cell) for inclusion in downstream analysis. Further, cells with fractional representation of mitochondrial genes greater than 40% were excluded. To account for potential transcript spreading, we removed any duplicated or hamming=1 barcodes among

samples sequenced on the same Nova-Seq runs. For each sample, we performed variable gene identification and calculated 30 principal components. Within each sample, we performed jackstraw simulations to identify significant principal components that were then used to perform t-SNE dimensionality reduction and clustering for each sample using only significant principal components. Within each sample, clusters defined exclusively by mitochondrial gene expression, indicative of low-quality cells, were removed from downstream analysis.

Removal of Ambient RNA Contamination

Within each sample, we removed ambient RNA contamination using SoupX (Young and Behjati, 2018). Initially, we determined appropriate UMI thresholds to estimate background contamination using EmptyDrops (Lun et al., 2019). Specifically, we examined the distribution of P-values UMI thresholds between 30 and 100 and selected the UMI threshold in which the distribution most closely approximated a uniform distribution. For each sample, we calculated an array-specific 'soup' profile among barcodes below the UMI threshold. To calculate estimated per-cell contamination fractions, we manually selected genes observed to be bimodally expressed across cells, which suggest that these genes are predominantly expressed in a single-cell type, but are observed at low-levels in other cell types for which endogenous expression would not be expected. For each array, we removed individual transcripts most likely to be contamination from each single-cell based on the estimated contamination fraction. Specifically, individual transcripts were sequentially removed from each single-cell transcriptome until the probability of subsequent transcripts being soup-derived was less than 0.5 to generate a background-corrected UMI matrix for each Seq-Well S^3 array.

Doublet Removal

We performed doublet removal for each sample individually using DoubletFinder (McGinnis et al., 2018). For each sample, we calculated the expected doublet rate based on the cell loading density. For each sample, a total of 20,000 cells were loaded to a loaded Seq-Well

device containing 85,000 wells (lambda = 20,000). For each array, we calculated an expected doublet rate of 2.37%. For each array, we generated pseudo-doublets using the following parameter values in DoubletFinder: proportion.artificial = 0.25 and proportion.NN = 0.01. Cells were identified as doublets based on their rank order in the distribution of the proportion of artificial nearest neighbors (pANN). Specifically, we identified the pANN value for the cell at the expected doublet percentile and used the corresponding pANN value as a threshold to remove additional cells with pANN greater than or equal to this value.

Analysis of Combined Skin Dataset

After background and doublet correction, we performed integrated analysis on a combined dataset of 49,373 cells. We performed variable gene identification and dimensionality reduction to identify 38 cell type clusters using Louvain clustering (Resolution = 1.75). We identified genes enriched across clusters to identify generic cell types. We performed an initial round of dimensionality reduction and cluster identification among cell types used in subsequent analysis (i.e. T cells, myeloid cells, B and plasma cells, endothelial cells, fibroblasts, and keratinocytes). Based on the initial sub-clustering results for each cell type, we removed sub-clusters defined by residual contamination not corrected for by SoupX background correction and doublet filtering. In total, we filtered 11,099 cells from sub-clusters defined by residual contamination: 1,471 from the T cell sub-analysis, 497 from the myeloid sub-analysis, 2,444 from the endothelial sub-analysis, 2,512 from the fibroblast sub-analysis, and 4,175 from the keratinocyte sub-analysis.

After this stringent quality control filtering step, a total of 38,274 cells were included in downstream analysis of the atlas of skin inflammation. We first performed variable gene identification and identified 5,897 genes as variably expressed. We performed UMAP dimensionality reduction to generate a 2-dimensional representation of gene expression data, and we identified a total of 35 cell type clusters using Louvain clustering (Resolution = 1.5) in Scanpy (Wolf et al., 2018). To understand similarity of identified clusters, we performed

hierarchical clustering of identified cell type clusters (**Figure A2.2C**). Specifically, across clusters we generated a list composed of the top 25 cluster-defining genes from each cluster. Average gene expression values within each across the 511 unique cluster-defining genes was used to perform hierarchical clustering. A dendrogram was generated to display the similarity of clusters, and the observed relationships were used to inform rational combination of related cell type clusters for combined analysis (**Figure A2.2C**). Cell type assignments were assigned through a combination of literature-based assessment of expression signatures and manual curation. In total, we identified clusters representing 15 major cell types including B cells, dendritic cells, fibroblasts, hair follicle, keratinocytes, Langerhans cells, lymphatics, mast cells, macrophages, neurons, plasma cells, sweat gland, T cells, vascular smooth muscle cells (VSMCs) and venules.

Cell-type Identification Using SingleR

To validate cell-type identification based on the combination of literature and manual curation, we performed automated cell-type identification using SingleR (Aran et al., 2019). Here, we classified 38,274 using the blueprint encode reference dataset. Cell types assigned by SingleR were compared to the manually assigned cell-type classifications (**Figure A2.4B**). Where suitable reference populations exist in the blueprint encode dataset, we observe close correspondence between manual classification and automated classification in SingleR.

Identification of T cell Sub-Clusters

We performed sub-analysis for numerous cell-types to examine additional variation within major cell types. Among the 4,943 T cells identified in the total dataset, we identified 5,574 variable genes that were used to construct a force-directed graph. We further used this set of variable genes to perform Louvain clustering (Resolution = 0.8) and identified a total of nine T cell sub-clusters. Cell-type identities were established by examining the expression of known marker genes corresponding to $CD4^+$ T helper and $CD8^+$ T cell subsets. We performed comparison of identified T cell signatures to previously identified signatures in Savant (Lopez et al., 2017)

(**Figure A2.5A**). To further define variation among both CD8 T cells and NK-Cytotoxic cells, we performed additional sub-grouping analyses. In both cases, we performed sub-analyses in which UMAP dimensionality reduction was performed using a total of 5 principal components calculated across variable genes using Seurat. For $CD8^+$ T cells, we identified sub-groupings in Seurat using a resolution of 0.3, while a resolution of 0.6 was used for sub-grouping analysis for NK-Cytotoxic T cells (**Figure A2.5B**).

T cell Receptor Detection and Clonal Expansion

Initially, we examined the detection rates for TCR α and β (Constant, V and J genes) among $CD4^+$ T cells from experiments performed on PBMCs using the Seq-Well V1, Seq-Well S^3 protocol and 10x v2 (**Figure A2.5C**). Specifically, detection of constant genes was determined by non-zero values for either TRAC or TRBC2 genes for α and β constant genes, respectively. Similarly, detection of TRAV and TRBV sequences was determined on the basis of a non-zero value for any TRAV or TRBV gene, We further examined the rate of TCR detection across 2,908 T cells obtained from human skin biopsies. Specifically, we examine detection rates across multiple sequencing depths: <5,000, 5,000-25,000, 25-000-100,000, and > 100,000 aligned reads per cell. (**Figure 3.3E**).

We further examined the extent of conservation of V gene usage as a proxy for clonal expansion. Specifically, for skin sample, we examined the extent of conserved V gene expression (**Figure 3.3F**). We examined the distribution of TRAV and TRBV gene expression and identified the V-gene (TRAV or TRBV) with the highest expression level for each T cell. We examined the distribution of TRAV and TRBV genes within each sample to identify potential clonal expansions. To understand the divergence from uniform distribution, we calculated the gini coefficient and the Shannon divergence for TRAV and TRBV sequences within each sample (**Figure 3.3F**).

Identification of Myeloid Heterogeneity

We performed sub-analysis of myeloid populations observed in skin, which include Dendritic cells, Macrophages, Mast cells, and Langerhans cells identified in global analysis of 38,274 total cells. Using a combined dataset of 5,010 myeloid cells, we performed variable gene identification and dimensionality reduction in Scanpy. Specifically, we constructed a force-directed graph across 6,599 variable genes and performed Louvain clustering (resolution = 0.80), and we obtained 10 sub-clusters of myeloid cells (**Figure A2.5D**).

To understand differences in Langerhans cells in normal skin, we performed differential expression analysis within each cluster of Langerhans cells (**Figure A2.5D**). Results presented in **Figure 3.4E** represent differential expression results between Langerhans cells from Myeloid cluster 8 between normal and leprosy skin biopsies. We further performed gene-set enrichment analysis among genes upregulated in Langerhan's cells from leprosy samples by comparing to signatures contained in the MSigDb database.

We performed additional sub-grouping among 951 dendritic cells and performed UMAP dimensionality reduction and Louvain clustering (resolution =0.45). We identified 5 sub-groupings of dendritic cells, and identified genes enriched within each cluster by performing a wilcox test in Seurat. To understand how dendritic cells related to previous findings, we performed comparisons to published signatures of dendritic cell phenotypes [9, 10]. Specifically, we generated expression scores using the top 10 genes using the AddModuleScore function in Seurat and examined the distribution of signature scores (**Figure A2.5F-G**). We determined significance of cluster enrichment by performing 1,000 permutations in which cell and signature score identifiers were randomly re-assigned. Upon permutation testing, we observed significant enrichment among DC sub-group 4 (cDC1) for a signature corresponding to CLEC9A^{+} cDC1 cells ($P<0.05$, Permutation Test; **Figure A2.5F**).

Identification of Endothelial Heterogeneity

We performed variable gene identification across 8,571 endothelial cells identified in the global analysis across 19 skin biopsies. We generated a force-directed graph and performed Louvain clustering (resolution = 0.6) among 5,082 variable genes, which revealed 9 sub-clusters of endothelial cells used in downstream analysis. To identify genes enriched in each endothelial sub-cluster, we performed a wilcox test in Seurat. We further examined the distribution of addressin expression across endothelial subsets using the set of addressins detected in endothelial cells (**Figure 3.5E**).

Identification of Fibroblast Heterogeneity

We performed variable gene identification across 7,237 fibroblasts identified in global analysis. We performed dimensionality reduction across 4,825 variable genes and performed Louvain clustering (Resolution = 0.6), revealing 10 sub-clusters of fibroblasts. For each fibroblast sub-cluster, we examined the relative contribution of cells from each sample and condition. Further, we examined the distribution of fibroblast sub-clusters within each sample. For each sub-cluster, we performed enrichment analysis to identify cluster-defining genes.

Pseudo-temporal Reconstruction of Epidermal Keratinocytes

Initially, we performed diffusion analysis across all keratinocytes and hair follicle cells using the Diffmap function in Scanpy [26], which implements a method for diffusion pseudotime reconstruction (Haghverdi et al., 2015). Within normal and psoriatic keratinocytes, a population of basal keratinocytes was identified on the basis of expression of *KRT14* (**Figure 3.7C**). Initially, we performed pseudo-temporal analysis within normal keratinocyte separately, using the basal keratinocyte population as the origin of the pseudo-temporal ordering. After observing distinct developmental trajectories among psoriatic keratinocytes (**Figure 3.7D**), we performed differential expression analysis between normal and psoriatic keratinocytes in Seurat using a wilcox rank-sum test across total keratinocytes and among basal, differentiating, and terminal keratinocytes.

Differentiation trajectories for individual samples were constructed using dyno (https://rdrr.io/github/dynverse/dyno/man/dyno.html) [69]. Specifically, we used Scorpius to generate diffusion pseudo-time reconstructions of normal and psoriatic keratinocytes (**Figure A2.6**). We examined gene expression patterns correlated with pseudo-temporal order across normal keratinocyte populations. We performed differential pseudo-time correlation analysis between normal and psoriatic keratinocytes for 2 normal skin samples and 5 psoriasis samples. Here, for each pseudo-temporal reconstruction, we performed linear regression between pseudo-time values and gene expression values for normal and psoriatic keratinocytes. We then calculated the difference in pseudo-time correlation between psoriatic and normal keratinocytes to identify genes that are uniquely involved in the development of psoriatic keratinocytes (**Figure 3.7G**).

Keratinocyte Cytokine-Response Profiles

Among both normal and psoriatic keratinocytes, we generated cytokine response scores using a series of reference datasets. Specifically, bulk RNA-sequencing references were previously generated from in vitro experiments in which cultured keratinocytes were stimulated with cytokines, individually or in combination. We used expression signatures generated by exposing keratinocytes to IL-17A, IL17-A + TNF-α, TNF-α, IFN-α, IL-4, IL-13, and IFN-γ. Expression signature were generated relative an unstimulated control population of keratinocytes. For each cytokine condition, we used the top 100 differentially expressed genes to generate a cytokine response score across both psoriatic and normal keratinocytes. We then examined the extent of cytokine response across basal, differentiating and terminal keratinocytes between normal and psoriatic keratinocytes.

REFERENCES

1. Klein, A.M., et al., *Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.* Cell, 2015. **161**(5): p. 1187-1201.

2. Macosko, E.Z., et al., *Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets.* Cell, 2015. **161**(5): p. 1202-1214.

3. Montoro, D.T., et al., *A revised airway epithelial hierarchy includes CFTR-expressing ionocytes.* Nature, 2018. **560**(7718): p. 319-324.

4. Ordovas-Montanes, J., et al., *Allergic inflammatory memory in human respiratory epithelial progenitor cells.* Nature, 2018. **560**(7720): p. 649-654.

5. Smillie, C.S., et al., *Intra-and inter-cellular rewiring of the human colon during ulcerative colitis.* Cell, 2019. **178**(3): p. 714-730. e22.

6. Vento-Tormo, R., et al., *Single-cell reconstruction of the early maternal–fetal interface in humans.* Nature, 2018. **563**(7731): p. 347-353.

7. Consortium, T.M., *Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris.* Nature, 2018. **562**(7727): p. 367.

8. Braga, F.A.V., et al., *A cellular census of human lungs identifies novel cell states in health and in asthma.* Nature medicine, 2019. **25**(7): p. 1153-1163.

9. Villani, A.-C., et al., *Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors.* Science, 2017. **356**(6335): p. eaah4573.

10. Dutertre, C.-A., et al., *Single-cell analysis of human mononuclear phagocytes reveals subset-defining markers and identifies circulating inflammatory dendritic cells.* Immunity, 2019. **51**(3): p. 573-589. e8.

11. Puel, A., et al., *Defective IL7R expression in T-B+ NK+ severe combined immunodeficiency.* Nature genetics, 1998. **20**(4): p. 394-397.

12. Prakadan, S.M., A.K. Shalek, and D.A. Weitz, *Scaling by shrinking: empowering single-cell'omics' with microfluidic devices.* Nature Reviews Genetics, 2017. **18**(6): p. 345.

13. Sasagawa, Y., et al., *Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity.* Genome biology, 2013. **14**(4): p. 3097.

14. Shishkin, A.A., et al., *Simultaneous generation of many RNA-seq libraries in a single reaction.* Nature methods, 2015. **12**(4): p. 323-325.

15. Picelli, S., et al., *Smart-seq2 for sensitive full-length transcriptome profiling in single cells.* Nature methods, 2013. **10**(11): p. 1096-1098.

16. Islam, S., et al., *Highly multiplexed and strand-specific single-cell RNA 5' end sequencing.* Nature protocols, 2012. **7**(5): p. 813-828.

17. Kapteyn, J., et al., *Incorporation of non-natural nucleotides into template-switching oligonucleotides reduces background and improves cDNA synthesis from very small RNA samples.* BMC genomics, 2010. **11**(1): p. 413.

18. Zajac, P., et al., *Base preferences in non-templated nucleotide incorporation by MMLV-derived reverse transcriptases.* PloS one, 2013. **8**(12).

19. Kabashima, K., et al., *The immunological anatomy of the skin.* Nature Reviews Immunology, 2019. **19**(1): p. 19-30.

20. Tabib, T., et al., *SFRP2/DPP4 and FMO1/LSP1 define major fibroblast populations in human skin.* Journal of Investigative Dermatology, 2018. **138**(4): p. 802-810.

21. Cheng, J.B., et al., *Transcriptional programming of normal and inflamed human epidermis at single-cell resolution.* Cell reports, 2018. **25**(4): p. 871-883.

22. He, H., et al., *Single-cell transcriptome analysis of human skin identifies novel fibroblast subpopulation and enrichment of immune subsets in atopic dermatitis.* Journal of Allergy and Clinical Immunology, 2020.

23. Kim, D., et al., *Targeted therapy guided by single-cell transcriptomic analysis in drug-induced hypersensitivity syndrome: a case report.* Nature Medicine, 2020. **26**(2): p. 236-243.

24. Gierahn, T.M., et al., *Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput.* Nature methods, 2017. **14**(4): p. 395-398.

25. Picelli, S., et al., *Full-length RNA-seq from single cells using Smart-seq2.* Nature protocols, 2014. **9**(1): p. 171.

26. Wolf, F.A., P. Angerer, and F.J. Theis, *SCANPY: large-scale single-cell gene expression data analysis.* Genome biology, 2018. **19**(1): p. 15.

27. Aran, D., et al., *Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage.* Nature immunology, 2019. **20**(2): p. 163-172.

28. Diani, M., G. Altomare, and E. Reali, *T cell responses in psoriasis and psoriatic arthritis.* Autoimmunity reviews, 2015. **14**(4): p. 286-292.

29. Lowes, M.A., M. Suárez-Fariñas, and J.G. Krueger, *Immunology of psoriasis.* Annual review of immunology, 2014. **32**: p. 227-255.

30. Lopez, D., et al., *SaVanT: a web-based tool for the sample-level visualization of molecular signatures in gene expression profiles.* BMC genomics, 2017. **18**(1): p. 824.

31. Bezman, N.A., et al., *Molecular definition of the identity and activation of natural killer cells.* Nature immunology, 2012. **13**(10): p. 1000.

32. Best, J.A., et al., *Transcriptional insights into the CD8+ T cell response to infection and memory T cell formation.* Nature immunology, 2013. **14**(4): p. 404.

33. Ivanov, I.I., et al., *The orphan nuclear receptor RORγt directs the differentiation program of proinflammatory IL-17+ T helper cells.* Cell, 2006. **126**(6): p. 1121-1133.

34. Saini, C., V. Ramesh, and I. Nath, *CD4+ Th17 cells discriminate clinical types and constitute a third subset of non Th1, Non Th2 T cells in human leprosy.* PLoS neglected tropical diseases, 2013. **7**(7).

35. Saini, C., et al., *Leprosy reactions show increased Th17 cell activity and reduced FOXP3+ Tregs with concomitant decrease in TGF-β and increase in IL-6.* PLoS neglected tropical diseases, 2016. **10**(4).

36. Liu, X., et al., *Genome-wide analysis identifies NR4A1 as a key mediator of T cell dysfunction.* Nature, 2019. **567**(7749): p. 525-529.

37. Lanna, A., et al., *A sestrin-dependent Erk–Jnk–p38 MAPK activation complex inhibits immunity during aging.* Nature immunology, 2017. **18**(3): p. 354.

38. Balin, S.J., et al., *Human antimicrobial cytotoxic T lymphocytes, defined by NK receptors and antimicrobial proteins, kill intracellular bacteria.* Science immunology, 2018. **3**(26): p. eaat7668.

39. Söderström, K., et al., *Natural killer cells trigger osteoclastogenesis and bone destruction in arthritis.* Proceedings of the National Academy of Sciences, 2010. **107**(29): p. 13028-13033.

40. Zhang, L., et al., *Lineage tracking reveals dynamic relationships of T cells in colorectal cancer.* Nature, 2018. **564**(7735): p. 268-272.

41. Malissen, B., S. Tamoutounour, and S. Henri, *The origins and functions of dendritic cells and macrophages in the skin.* Nature Reviews Immunology, 2014. **14**(6): p. 417-428.

42. Fuentes-Duculan, J., et al., *A subpopulation of CD163-positive macrophages is classically activated in psoriasis.* Journal of Investigative Dermatology, 2010. **130**(10): p. 2412-2422.

43. Di Rosa, M., et al., *Evaluation of CHI3L-1 and CHIT-1 expression in differentiated and polarized macrophages.* Inflammation, 2013. **36**(2): p. 482-492.

44. Romani, N., et al., *Langerhans cells–dendritic cells of the epidermis.* Apmis, 2003. **111**(7-8): p. 725-740.

45. Hunger, R.E., et al., *Langerhans cells utilize CD1a and langerin to efficiently present nonpeptide antigens to T cells.* The Journal of clinical investigation, 2004. **113**(5): p. 701-708.

46. Pinheiro, R.O., et al., *Innate immune responses in leprosy.* Frontiers in immunology, 2018. **9**: p. 518.

47. Guilliams, M., et al., *Unsupervised high-dimensional analysis aligns dendritic cells across tissues and species.* Immunity, 2016. **45**(3): p. 669-684.

48. Schlitzer, A., et al., *IRF4 transcription factor-dependent CD11b+ dendritic cells in human and mouse control mucosal IL-17 cytokine responses.* Immunity, 2013. **38**(5): p. 970-983.

49. Stutte, S., et al., *Requirement of CCL17 for CCR7-and CXCR4-dependent migration of cutaneous dendritic cells.* Proceedings of the National Academy of Sciences, 2010. **107**(19): p. 8736-8741.

50. Kashem, S.W., et al., *Nociceptive sensory fibers drive interleukin-23 production from CD301b+ dermal dendritic cells and drive protective cutaneous immunity.* Immunity, 2015. **43**(3): p. 515-526.

51. Kumamoto, Y., et al., *CD301b+ dermal dendritic cells drive T helper 2 cell-mediated immunity.* Immunity, 2013. **39**(4): p. 733-743.

52. Guilliams, M., et al., *The function of Fcγ receptors in dendritic cells and macrophages.* Nature Reviews Immunology, 2014. **14**(2): p. 94-108.

53. Brckalo, T., et al., *Functional analysis of the CD300e receptor in human monocytes and myeloid dendritic cells.* European journal of immunology, 2010. **40**(3): p. 722-732.

54. Jin, P., et al., *Molecular signatures of maturing dendritic cells: implications for testing the quality of dendritic cell therapies.* Journal of translational medicine, 2010. **8**(1): p. 4.

55. Schön, M.P., T.M. Zollner, and W.-H. Boehncke, *The molecular basis of lymphocyte recruitment to the skin: clues for pathogenesis and selective therapies of inflammatory disorders.* Journal of Investigative Dermatology, 2003. **121**(5): p. 951-962.

56. von Andrian, U.H. and T.R. Mempel, *Homing and cellular traffic in lymph nodes.* Nature Reviews Immunology, 2003. **3**(11): p. 867-878.

57. Bhattacharya, R., et al., *NHERF-2 maintains endothelial homeostasis.* Blood, The Journal of the American Society of Hematology, 2012. **119**(20): p. 4798-4806.

58. Thiriot, A., et al., *Differential DARC/ACKR1 expression distinguishes venular from non-venular endothelial cells in murine tissues.* BMC biology, 2017. **15**(1): p. 45.

59. Driskell, R.R., et al., *Distinct fibroblast lineages determine dermal architecture in skin development and repair.* Nature, 2013. **504**(7479): p. 277-281.

60. Driskell, R.R. and F.M. Watt, *Understanding fibroblast heterogeneity in the skin.* Trends in cell biology, 2015. **25**(2): p. 92-99.

61. Hazell, G.G., et al., *PI16 is a shear stress and inflammation-regulated inhibitor of MMP2.* Scientific reports, 2016. **6**: p. 39553.

62. Huth, S., et al., *Inter-α-trypsin inhibitor heavy chain 5 (ITIH 5) is overexpressed in inflammatory skin diseases and affects epidermal morphology in constitutive knockout mice and murine 3D skin models.* Experimental dermatology, 2015. **24**(9): p. 663-668.

63. Fu, X., et al., *Specialized fibroblast differentiated states underlie scar formation in the infarcted mouse heart.* The Journal of clinical investigation, 2018. **128**(5): p. 2127-2143.

64. Micke, P., et al., *In situ identification of genes regulated specifically in fibroblasts of human basal cell carcinoma.* Journal of Investigative Dermatology, 2007. **127**(6): p. 1516-1523.

65. Pickens, S.R., et al., *Characterization of CCL19 and CCL21 in rheumatoid arthritis.* Arthritis & Rheumatism, 2011. **63**(4): p. 914-922.

66. Reyes, L.I., et al., *Dexamethasone inhibits BAFF expression in fibroblast-like synoviocytes from patients with rheumatoid arthritis.* Cytokine, 2008. **42**(2): p. 170-178.

67. Fuchs, E., *Epidermal differentiation: the bare essentials.* The Journal of cell biology, 1990. **111**(6 Pt 2): p. 2807-2814.

68. Fuchs, E. and S. Raghavan, *Getting under the skin of epidermal morphogenesis.* Nature Reviews Genetics, 2002. **3**(3): p. 199-209.

69. Saelens, W., et al., *A comparison of single-cell trajectory inference methods.* Nature biotechnology, 2019. **37**(5): p. 547-554.

70. Pellegrini, G., et al., *p63 identifies keratinocyte stem cells.* Proceedings of the national academy of sciences, 2001. **98**(6): p. 3156-3161.

71. Uhlén, M., et al., *Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.

72. Sandilands, A., et al., *Filaggrin in the frontline: role in skin barrier function and disease.* Journal of cell science, 2009. **122**(9): p. 1285-1294.

73. Joost, S., et al., *Single-cell transcriptomics reveals that differentiation and spatial signatures shape epidermal and hair follicle heterogeneity.* Cell systems, 2016. **3**(3): p. 221-237. e9.

74. Li, B., et al., *Transcriptome analysis of psoriasis in a large case–control sample: RNA-seq provides insights into disease mechanisms.* Journal of Investigative Dermatology, 2014. **134**(7): p. 1828-1838.

75. Tsoi, L.C., et al., *Atopic Dermatitis Is an IL-13–Dominant Disease with Greater Molecular Heterogeneity Compared to Psoriasis.* Journal of Investigative Dermatology, 2019. **139**(7): p. 1480-1489.

76. Chiricozzi, A., et al., *IL-17 induces an expanded range of downstream genes in reconstituted human epidermis model.* PloS one, 2014. **9**(2).

77. Zhu, J., H. Yamane, and W.E. Paul, *Differentiation of effector CD4 T cell populations.* Annual review of immunology, 2009. **28**: p. 445-489.

78. Tu, A.A., et al., *TCR sequencing paired with massively parallel 3' RNA-seq reveals clonotypic T cell signatures.* Nature immunology, 2019. **20**(12): p. 1692-1699.

79. Angelo, M., et al., *Multiplexed ion beam imaging of human breast tumors.* Nature medicine, 2014. **20**(4): p. 436.

80. Modlin, R.L., et al., *Immunopathologic Demonstration of T Lymphocyte Subpopulations and Interleukin 2 in Graneloma Annulare.* Pediatric dermatology, 1984. **2**(1): p. 26-32.

81. Beretta-Piccoli, B.T., et al., *Cutaneous granulomatosis: a comprehensive review.* Clinical reviews in allergy & immunology, 2018. **54**(1): p. 131-146.

82. de Oliveira Fulco, T., et al., *Effect of apoptotic cell recognition on macrophage polarization and mycobacterial persistence.* Infection and immunity, 2014. **82**(9): p. 3968-3978.

83. Verreck, F.A., et al., *Human IL-23-producing type 1 macrophages promote but IL-10-producing type 2 macrophages subvert immunity to (myco) bacteria.* Proceedings of the National Academy of Sciences, 2004. **101**(13): p. 4560-4565.

84. Piette, E.W. and M. Rosenbach, *Granuloma annulare: clinical and histologic variants, epidemiology, and genetics.* Journal of the American Academy of Dermatology, 2016. **75**(3): p. 457-465.

85. Yun, J.H., et al., *Clinical and pathological features of generalized granuloma annulare with their correlation: a retrospective multicenter study in Korea.* Annals of dermatology, 2009. **21**(2): p. 113-119.

86. Beylot, C., et al., *Propionibacterium acnes: an update on its role in the pathogenesis of acne.* Journal of the European Academy of Dermatology and Venereology, 2014. **28**(3): p. 271-278.

87. Holland, D., et al., *Inflammation in acne scarring: a comparison of the responses in lesions from patients prone and not prone to scar.* British Journal of Dermatology, 2004. **150**(1): p. 72-81.

88. Kim, T.-G., et al., *Skin-Specific CD301b+ Dermal Dendritic Cells Drive IL-17− Mediated Psoriasis-Like Immune Response in Mice.* Journal of Investigative Dermatology, 2018. **138**(4): p. 844-853.

89. Zaba, L.C., et al., *Identification of TNF-related apoptosis-inducing ligand and other molecules that distinguish inflammatory from resident dendritic cells in patients with psoriasis.* Journal of Allergy and Clinical Immunology, 2010. **125**(6): p. 1261-1268. e9.

90. Hawkes, J.E., et al., *Discovery of the IL-23/IL-17 signaling pathway and the treatment of psoriasis.* The Journal of Immunology, 2018. **201**(6): p. 1605-1613.

91. Mathes, A.L., et al., *Global chemokine expression in systemic sclerosis (SSc): CCL19 expression correlates with vascular inflammation in SSc skin.* Annals of the rheumatic diseases, 2014. **73**(10): p. 1864-1872.

92. Stoeckius, M., et al., *Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics.* Genome biology, 2018. **19**(1): p. 1-12.

93. McGinnis, C.S., et al., *MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices.* Nature methods, 2019. **16**(7): p. 619.

94. van Galen, P., et al., *Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity.* Cell, 2019. **176**(6): p. 1265-1281. e24.

Chapter 4: Cellular correlates of M. tuberculosis control in pulmonary granulomas revealed through single-cell mRNA sequencing

Travis K. Hughes*, Hannah Gideon*, Marc Wadsworth, Andy Tu, Nicole Grant, J. Chris Love, JoAnne Flynn[#], Sarah M. Fortune[#], and Alex K. Shalek[#]

*These authors contributed equally to this work.

Authorship Attribution

TKH, HPG, JLF, SMF, and AKS designed the study. TKH, HPG and MHW performed Seq-Well experiments. TKH and AT performed TCR enrichment experiments and analysis. NG performed mast cell imaging experiments. TH and HPG performed data analysis. TKH, HPG, JLF, SMF and AKS wrote the manuscript with input from all authors.

Introduction

Mycobacterium tuberculosis (Mtb) remains a major threat to global health. The latest World Health Organization analysis of the global burden of tuberculosis (TB) estimates 10 million new cases and 1.5 million deaths per year in 2018 [1]. While ~90% of those infected individuals remain asymptomatic, the features of protective immunity are still largely unknown.

Mtb infection is characterized by the formation of granulomas, primarily in the lungs and thoracic lymph nodes [2, 3]. The TB granuloma is an organized structure of immune, parenchymal and stromal cells that forms in response to persistent Mtb infection, and consists of macrophages, neutrophils, and lymphocytes [4-7]. Granulomas function both as the niche in which the bacillus can grow or persist and an immunological microenvironment in which host cells interact to control or kill Mtb and prevent dissemination. The mere presence of a granuloma is insufficient to control infection; instead, proper functioning of all granulomas in a host determines the ultimate outcome of infection [5-9]. T cells, macrophages and neutrophils in granulomas are generally considered critical to the control of initial and persistent Mtb infection, mediating the inflammatory balance [10]. Recent studies in non-human primates (NHPs) have shown that granulomas have independent trajectories within a host with respect to inflammation (measured as FDG avidity by PET-CT imaging), size, number and types of cells, bacterial burden, dissemination risk and ability to kill Mtb [3, 8, 11-14].

Bacterial burden in individual granulomas is highest at 4-6 weeks post-infection and decreases over time even in animals with active disease, while the number of Mtb chromosomal

equivalents (CEQ), a measure of cumulative burden, is similar in granulomas at different times [13, 15]. In a study of 38 animals, the onset of a significant decrease in the bacterial burden occurs at 10 weeks post infection, where animals have both restrictive and permissive granulomas. In the present study, we chose selected this time point (10-weeks post-infection) to interrogate granuloma-level immune profiles that contribute to bacterial control.

Comprehensive study of the cellular composition of Mtb granulomas using standard methods (e.g. flow cytometry or mass cytometry) has been limited by the small number of cells that can be obtained from individual lesions. Single-cell sequencing (scRNA-Seq) enables unbiased profiling of immune, parenchymal and stromal responses in complex tissues. In the present study, we performed high-throughput scRNA-Seq using the Seq-Well platform [16] to understand the cellular and molecular correlates of Mtb control at the level of individual granulomas.

Results

Study population

Four cynomolgus macaques were infected with a low dose of Mtb (Erdman strain) for 10 weeks and included in this study (**Figure 4.1A**). Progression of Mtb infection and granuloma formation was monitored with PET-CT scans at 4, 8, and 10 weeks following Mtb infection. At necropsy, granulomas were harvested and bacterial burden was interrogated using (1) serial CFU assays to obtain an estimate of viable bacterial load and (2) sequencing to determine the number of Mtb chromosomal equivalents as a measure of cumulative bacterial burden (**Figure 4.1A**). For each animal, we recorded total lung flurodeoxyglucose (FDG) activity by PET-CT and observed that overall level of lung inflammation ranged from low to high FDG activity at 4 weeks post-infection (**Figure 4.1B**). Over the 10-week course of infection, each animal followed a different trajectory in total lung FDG (**Figure 4.1B**). We also monitored granuloma-level dynamics throughout the course of infection, measuring PET intensity (i.e. FDG uptake) (**Figure 4.1C**).

At necropsy, individual granulomas were excised from the lung of the infected animals. We randomly selected a total of 26 granulomas from 4 animals that were identified by pre-necropsy PET-CT scans for scRNA-seq analysis. Granulomas were isolated during necropsy and mechanically and enzymatically dissociated to generate a single-cell suspension (**Methods**). Approximately 15,000-20,000 cells from each granuloma were applied to Seq-Well devices loaded with barcoded mRNA capture beads. A portion of the single-cell suspension was further used for enumeration of bacterial burden (CFU) and cumulative bacterial load (CEQ) for each granuloma (**Methods**).

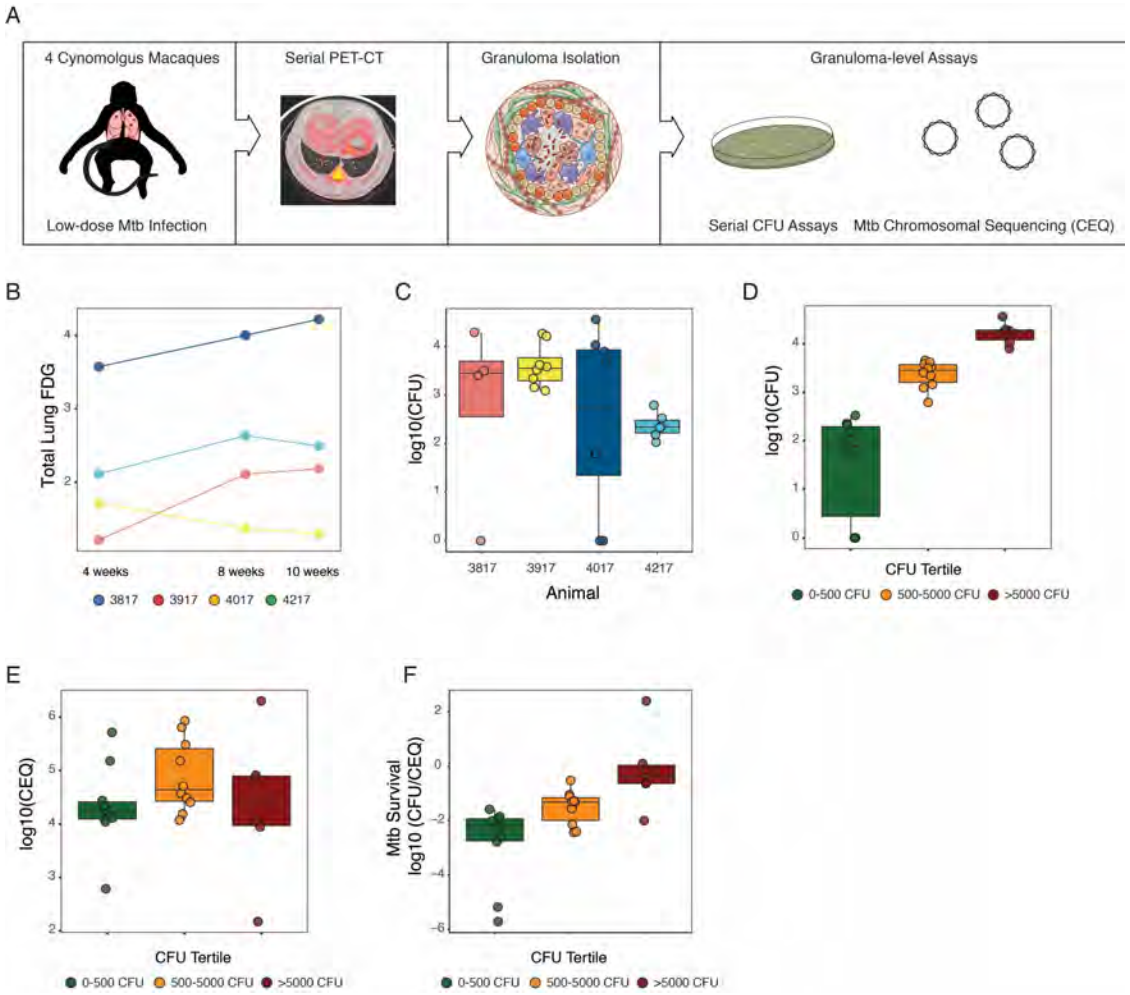


Figure 4.1 | Dynamics of *M. tuberculosis* infection in non-human primates. (A) Illustration of the study design. 4 cynomolgus macaques were bronchoscopically infected with a low-dose inoculum of Mtb (Erdman strain). Serial PET-CT was performed at 4, 8, and 10-weeks post-infection. 26 granulomas were randomly selected at necropsy. End-point bacterial burden was determined through serial CFU assays, while cumulative bacterial burden was determined by sequencing of Mtb chromosomal equivalents (Methods).

(B) Scatterplot showing total lung FDG (log10 – FDG) for each animal at 4, 8 and 10-week PET-CT scans. **(C)** Boxlplots showing the distribution of granuloma-level CFU in 4 cynomolgus macaques. **(D)** Boxplots showing granuloma-level CFU organized by tertiles. Colors correspond to CFU tertiles (Green: 0-500 CFU, Yellow: 500-5000 CFU, and Red: >5000 CFU). **(E)** Boxplots showing granuloma-level CEQ organized by CFU tertiles. Colors correspond to CFU tertiles (Green: 0-500 CFU, Yellow: 500-5000 CFU, and Red: >5000 CFU). **(F)** Boxplots showing the ratio between CFU (viable bacterial at end-point) and CEQ (cumulative bacterial burden). Higher log10(CFU/CEQ) ratios correspond to increased Mtb survival.

Among these 26 granulomas, we observe a broad range of granuloma-level bacterial burden with CFU values ranging from no bacterial growth to 4.56 $\log_{10}$ CFU/granuloma, which enables interrogation of the relationship between granuloma composition and bacterial control (**Figure 4.1C-D**). We evaluated the bacterial burden by tertiles (low: 0-500 CFU, mid: 500-5,000 CFU, and high: >5,000 CFU) to identify trends in cellular composition across the range of bacterial burden (**Figure 4.1D**). Importantly, we observed a significant difference in bacterial burden between granulomas with low CFU (median: 1.78 $\log_{10}$ CFU/granuloma) and high CFU (median: 4 $\log_{10}$ CFU/granuloma) (**Figure 4.1D**; $P < 0.0001$).

We next sought to understand whether low CFU values merely reflect absence of bacterial growth or the presence of immune control. Here, we initially examined the CEQ values across CFU tertiles and observed similar cumulative bacterial loads between granulomas with low (median 4.31 $\log_{10}$ CEQ/granuloma), mid (4.88 $\log_{10}$ CEQ/granuloma) or high CFU values (4.37 $\log_{10}$ CEQ/granuloma) (**Figure 4.1E**), suggesting that all granulomas supported similar cumulative growth of Mtb over the course of infection. In order to quantify the extent of killing, we calculated a ratio between CFU (i.e. end-point bacterial burden) and CEQ (i.e. cumulative bacterial burden) (**Figure 4.1F**; **Methods**). We found that that granulomas with the lowest bacterial burdens had significantly higher killing (2.3 $-\log_{10}$ CFU/CEQ per granuloma) than lesions with the highest bacterial burden (0.62 $-\log_{10}$ CFU/CEQ per granuloma, $P=0.0006$, Kruskal-Wallis test; **Figure 4.1F**), indicating that the immune composition of these granulomas could play a greater role in Mtb-killing and restricting bacterial growth.

Single-cell Sequencing of Mtb Pulmonary Granulomas

In order to perform an unbiased evaluation of the cellular composition of TB lung granulomas, we performed high-throughput scRNA-seq on 26 granulomas from 4 animals using the Seq-Well platform (**Figure 4.2A**) [16]. A single-cell suspension of individual granulomas was applied to the Seq-Well device and processed as previously described (**Methods**) [16]. After robust technical correction, we performed downstream analysis on 109,584 high-quality single-cell transcriptomes (**Methods**). Single-cell transcriptional profiles separated into 24 gross clusters which were assigned a generic cell type identity using a combination of computational classification and manual curation (**Methods**).

In total, we identified 13 generic cell types, with numerous sub-clusters, particularly among T cells and macrophages. Specifically, we found B cells (*CD79A* and *BANK1*), conventional dendritic cells (cDCs, *CLEC9A*), plasmacytoid dendritic cells (pDCs, *LILRA4*), endothelial cells (*CD93*), erythrocytes (*HBB*), fibroblasts (*COL1A1*), macrophages (*LYZ*), mast cells (*CPA3*), neutrophils (*CSF3R* and *PLEK*), plasma cells (*JCHAIN*), T cells (*CD3D* and *IL7R*), Type 1 Pneumocytes (*AGER*) and Type 2 Pneumocytes (*SFTPB* and *SFTPC*) (**Figure 4.2B-C**). For each generic cell type, we observed exclusive enrichment of a specific cell-type gene expression signature for each cluster. We furthered performed extensive validation of cell-type assignments through comparison to reference expression signatures from multiple sources including the Tabula muris [17], Mouse Cell Atlas [18], and Savant database (**Methods**) [19].

Figure 2.

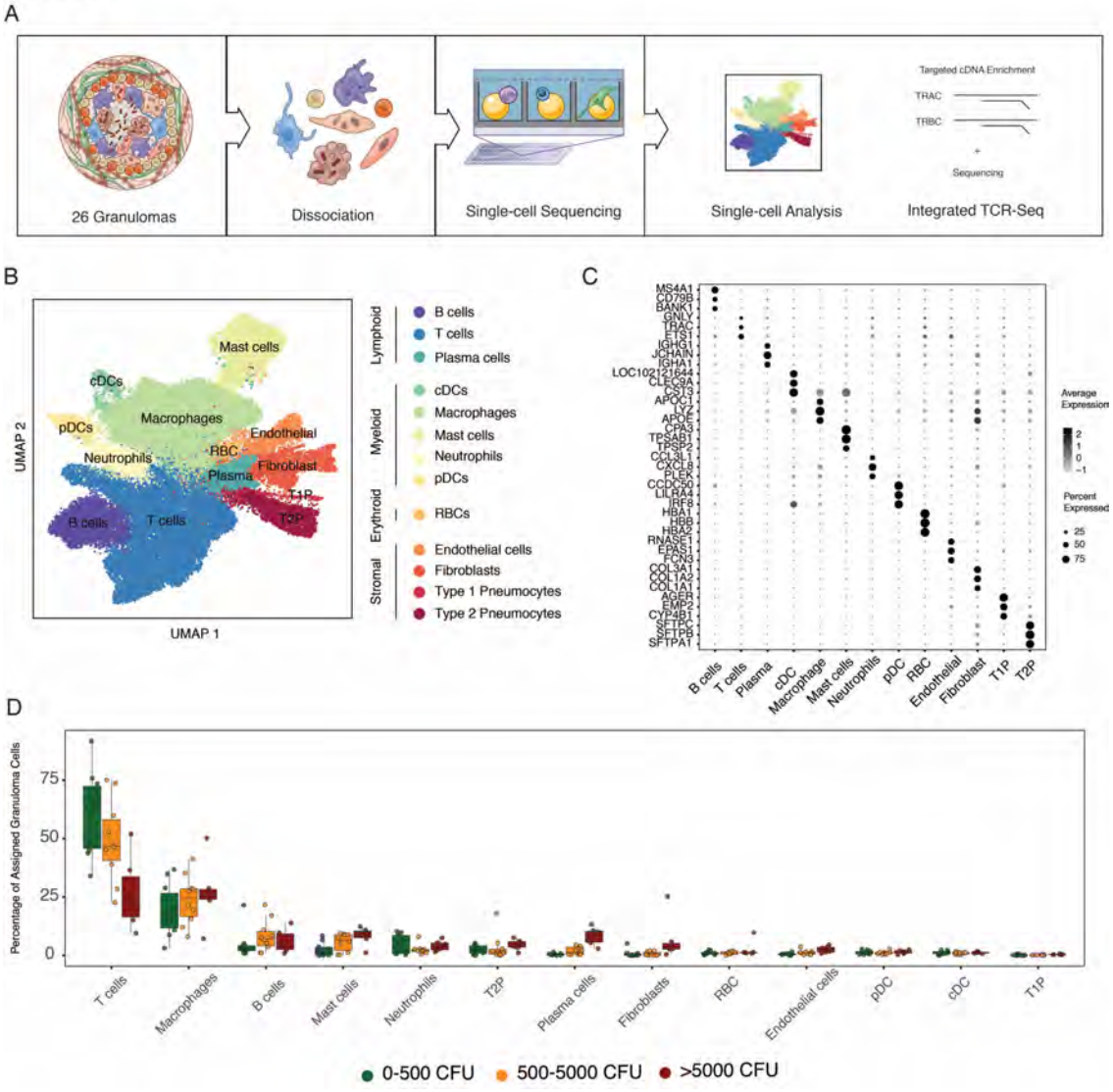


Figure 4.2 | Single-cell sequencing of MTB granulomas. (A) Overview of single-cell workflow and analytic pipeline. 26 granulomas were collected at necropsy and dissociated. We performed scRNA-Seq on 15,000 – 20,000 cells using the Seq-Well platform including targeted enrichment and sequencing of αβ TCR sequences. **(B)** UMAP plot of 109,584 cells from 26 granulomas colored by 13 generic cell types. **(C)** Dot-plot showing expression levels of genes enriched across 13 generic cell types. Color intensity corresponds to the level of gene expression, while the size of dots represents the percent of cells with non-zero expression in each cluster. **(D)** Boxplots showing the relationship between proportional composition of generic cell types and CFU tertiles. Significance of associations between cell-type composition and CFU are shown below.

Relationship of Cell-Type Composition to Bacterial Burden

To establish the cellular composition of each granuloma, we calculated the fractional composition for each generic cell type as the percentage of assigned cells within each granuloma (**Methods**). Across granulomas, we observed T cells (Median +/- S.D. : 45.1%, +/- 21.2.) and

macrophages (23.4% +/- 11.4) to be the most abundant cell types, followed by mast cells (5.74% +/- 4.1), B cells (4.1% +/- 6.1) and neutrophils (2.48% +/- 3.1).

To understand how the cellular composition of each granuloma associated with granuloma-level bacterial control, we examined potential links between bacterial burden and the frequency of each cell type across granulomas. For each granuloma, we examined the relationship between proportional composition of identified cell types (n of cells/ Total cells in each granuloma) and granuloma-level bacterial burden. Overall, we observed a negative correlation with bacterial burden and T cell proportion (Spearman's rho: -0.5156, p = 0.007), suggesting that T cells may play a major role in bacterial control (**Figure 4.2D**). Interestingly, our analysis revealed a strong, positive correlation with bacterial burden for plasma cells (Spearman's Rho = 0.7391, p < 0.0001), endothelial cells (Spearman's rho = 0.584, p= 0.0013), and mast cells (Spearman's rho = 0.5482, p = 0.0037). Further, we observed an association between fibroblasts and Type 1 pneumocytes and increased bacterial burden.

T cell population Diversity

To identify granular correlates of immune protection in Mtb granulomas, we performed sub-clustering analysis among T cells, which were most strongly associated with bacterial clearance. Initially, we identified and removed a single cluster of 3,144 T and NK cells marked by residual expression of genes derived from macrophages and mast cells (**Methods**). Following removal of this cluster, we performed a separate dimensionality reduction among 41,622 T cells and identified 12 sub-clusters of T and NK cells across 26 granulomas (**Figure 4.3A**; **Methods**). For each T cell sub-population, we contextualized gene expression patterns through by comparison to literature-derived expression signatures (bulk and single-RNA sequencing) and manual curation (**Methods**) [19, 20].

Figure 3.

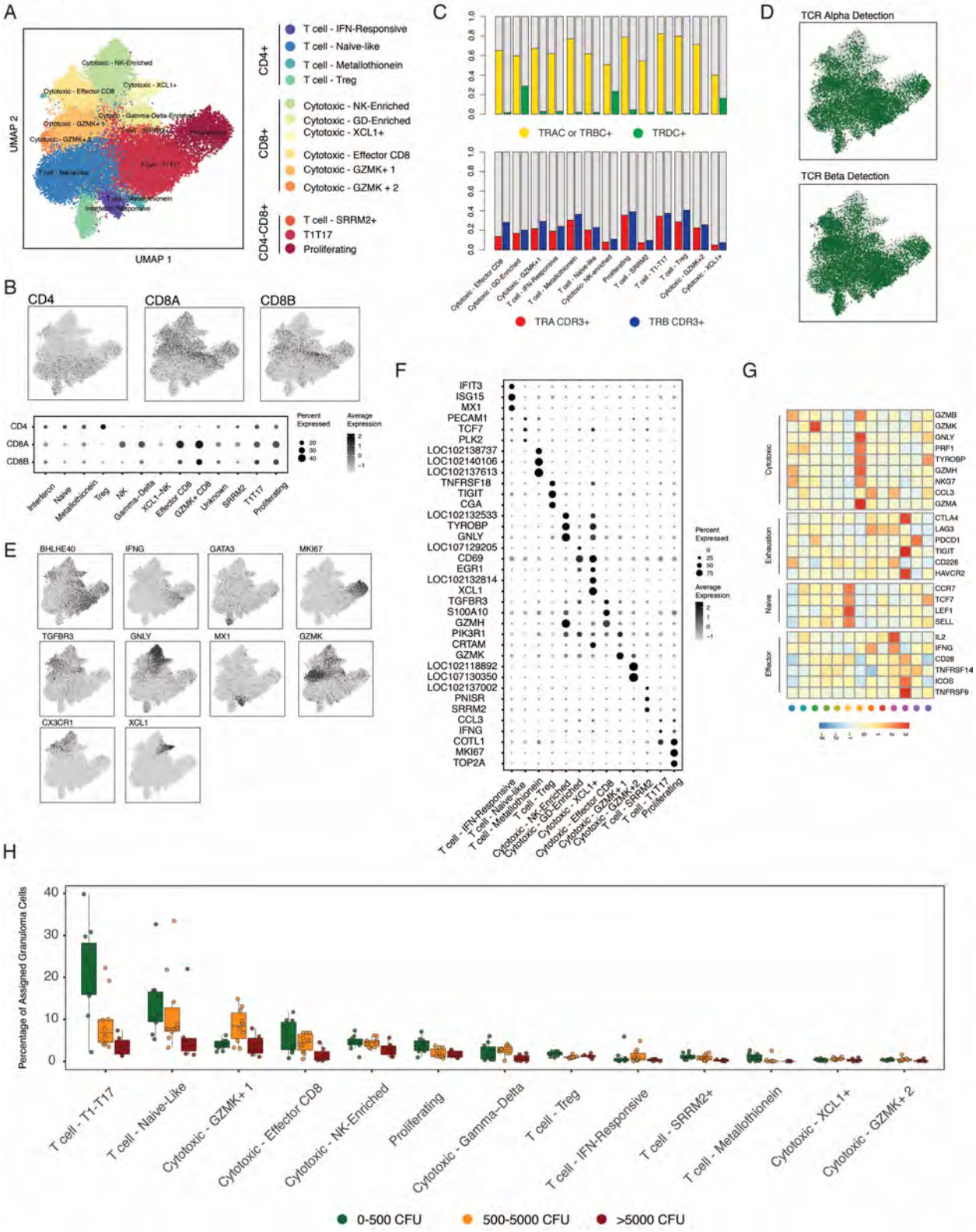


Figure. 4.3 | Relationship of T cell phenotype to granuloma-level bacterial burden (A) UMAP plot of 41,222 T cells colored by T cell sub-type. **(B)** (Top) UMAP plots of 41,222 T cells colored by the normalized

gene expression for CD4, CD8A, and CD8B. (Bottom) Dotplot showing expression of CD4, CD8A and CD8B across 12 T cell sub-clusters. Color intensity corresponds to the level of gene expression, while the size of dots represents the percent of cells with non-zero expression in each cluster. **(C)** (Top) Stacked barplot showing the rates of detection of TRAC, TRBC1 and TRBC2 (TRDC?) within 12 T cell clusters. (Bottom) Stacked barplot showing the rates of alpha and beta TCR recovery across 12 T cell clusters. **(D)** (Top) UMAP plot of 41,222 T cells colored by recovery of TCR α CDR3 sequences. (Bottom) UMAP plot of 41,222 T cells colored by recovery of TCR β CDR3 sequences. **(E)** UMAP plots colored by normalized expression levels for select T cell cluster-defining genes. **(F)** Dotplot showing the expression of T cell cluster-defining genes. Color intensity corresponds to the level of gene expression, while the size of dots represents the percent of cells with non-zero expression in each cluster. **(G)** Heatmap showing the distribution of T cell functional programs across T cell sub-clusters. Values shown in heatmap are row-normalized expression values. **(H)** Boxplots showing the relationship between T cell sub-clusters and granuloma-level CFU. Cell-types are ordered by overall abundance and colors correspond to CFU Tertiles (Green: 0-500 CFU, Yellow: 500-5000 CFU, and Red: >5000 CFU).

To assign functional identities to T cell clusters, we initially examined the expression of *CD4*, *CD8A* and *CD8B* and the recovery of $\alpha\beta$ and $\gamma\delta$ TCR sequences across clusters **(Figure 4.3B-D**). We initially identified 5 populations with elevated expression of either CD8A and CD8B and elevated expression of cytotoxic gene expression (e.g. granulysin, granzymes and NK-receptors) (**Figure 4.3E**). Among cytotoxic cells, we discovered 2 populations of innate cytotoxic cells: one population that is consistent with NK and cytotoxic T cells, and is distinguished by the highest expression of cytotoxic effector molecules (*GZMA*, *GZMB*, *GNLY*, and *PRF1*) and lowest= rates of TCR recovery (TRA Recovery: 8.2%, TRB Recovery: 11.1%) (**Figure 4.3D**). We further identified a population expressing gamma-delta T cell receptors (*TRGC* and *TRDC)*, surface receptor CD69 and cytokines (XCL1). Additional sub-clustering of the gamma-delta expressing population revealed 2 distinct sub-populations: a subpopulation of $\gamma\delta$-T cells and another of XCL1+ NK cells, which have recently been shown to potentiate anti-tumor immune responses [21] (**Figure 4.3F; Methods**).

We further identified 3 subpopulations of adaptive cytotoxic cells that express both CD8A and CD8B along with increased recovery of $\alpha\beta$ TCR sequences (**Figure 4.3D**). Among adaptive cytotoxic cells, we found a population of effector CD8 T cells (*CX3CR1*, *GZMB*, and *ZEB2*) and a population of GZMK+ CD8 T cells (*GZMK*, *CCL5*, and *CXCR4*). We also observed a third, low-abundance population of cytotoxic cells defined by relative expression of *GZMK* (**Figure 4.3F**).

We identified 3 populations of CD4+ T cells including a population of T cells expressing genes associated with a naïve phenotype (*CCR7*, *LEF1*, and *SELL*), regulatory T cells (*FOXP3*, *IKZF2*, and *IL1RL1*), and interferon-responsive T cells (*OAS2*, *MX1*, and *ISG15*) (**Figure 4.3E-F**). Consistent with published annotations, the population of naïve-like T cells is characterized by elevated expression of *CCR7*, *LEF1*, *SELL*, and *TCF7* (**Figure 4.3E-G**). Consistent with tissue-Tregs, we observe elevated expression of the IL-33 receptor ST2 (*IL1RL1*) in addition to canonical Treg markers (*FOXP3* and *TIGIT*) (**Figure 4.3F**) [22].

We further identified 2 populations of T cells that consist of both CD4+ and CD8+ T cells: one population of proliferating T cells (*MKI67* and *TOP2A*) and another population of T1-T17 cells defined by a unique pattern of transcription factors (*RORA*, *RORC*, *BHLHE40*, *HIF1A*, and *RBPJ*), surface receptors (*CCR6, CXCR3* and *IL23R*), and cytokines (*IFNG* and *TNF*) (**Figure 4.3E-F**). Finally, we observe two low-abundance populations characterized by expression of genes encoding multiple metallothioneins and another cluster distinguished by expression of SRRM2, an RNA splicing factor [23].

T cell phenotypic diversity is associated with granuloma-level bacterial control

Based on the observation that T cells most strongly correlate with protection in aggregate, we sought to understand how the phenotypic composition of T cells across granulomas relates to bacterial burden. Specifically, we observe that the abundance of naïve T cells ($P = 0.03$), effector CD8 T cells ($P = 0.02$), proliferating T cells ($P=0.032$) and T1-T17 cells ($P = 0.0013$) significantly correlate with reduced granuloma-level bacterial burden (**Figure 4.3H**).

Notably, we observe the most pronounced expansion of T1-T17 cells in granulomas with the lowest bacterial burdens. Th1 and Th17 cells have previously been observed to correlate with protective MTB vaccine responses and in clearance of pulmonary infection [24-27]. Among T1-T17 cells from granulomas, we observe enrichment of genes associated with neuro-pathogenic Th17 function (*RORC*, *BHLHE40*, *SATB1* and *RBPJ*) [28-32] in addition to Th1 effector function

including *IFNG* and *TNF* (**Figure 4.4A**). The population of T1T17 cells is further enriched for expression of *CCR6* and *CXCR3*, markers of Th1* cells that have been implicated in TB-protective immune responses [33]. Notably, this population is reminiscent of a recently described population of lung-resident 'ex-Th17' cells, which have been observed to express *IL-23R*, *RBPJ* and *BHLE40* following exposure to killed Klebsiella [34].

Based on the expression of both CD4 and CD8 T cells within this cluster, we performed additional sub-clustering analysis within the population of T1-T17 cells and identified 4 sub-groups (**Figure 4.4B**). Specifically, T1-T17 cluster 1 is characterized by expression of *CD4*, *IL7R*, and *TXNIP*; sub-cluster 2 is characterized by increased relative expression of cytotoxic effector molecules including *GZMA*, *GZMB*, *GZMH*, *GZMK, GNLY*, and *PRF1*; sub-cluster 3 displays elevated expression of *IFNG*, *TNF*, and *LTA,* while sub-cluster 4 expresses increased levels of heat-shock and DNA-damage associated genes (**Figure 4.4C-D**). We observe a significant association with T1T17 sub-populations and reduced bacterial burden with the exception of sub-cluster 3, which is defined by expression of *IFNG* and *TNF* and did not reach statistical significance (**Figure 4.4E**).

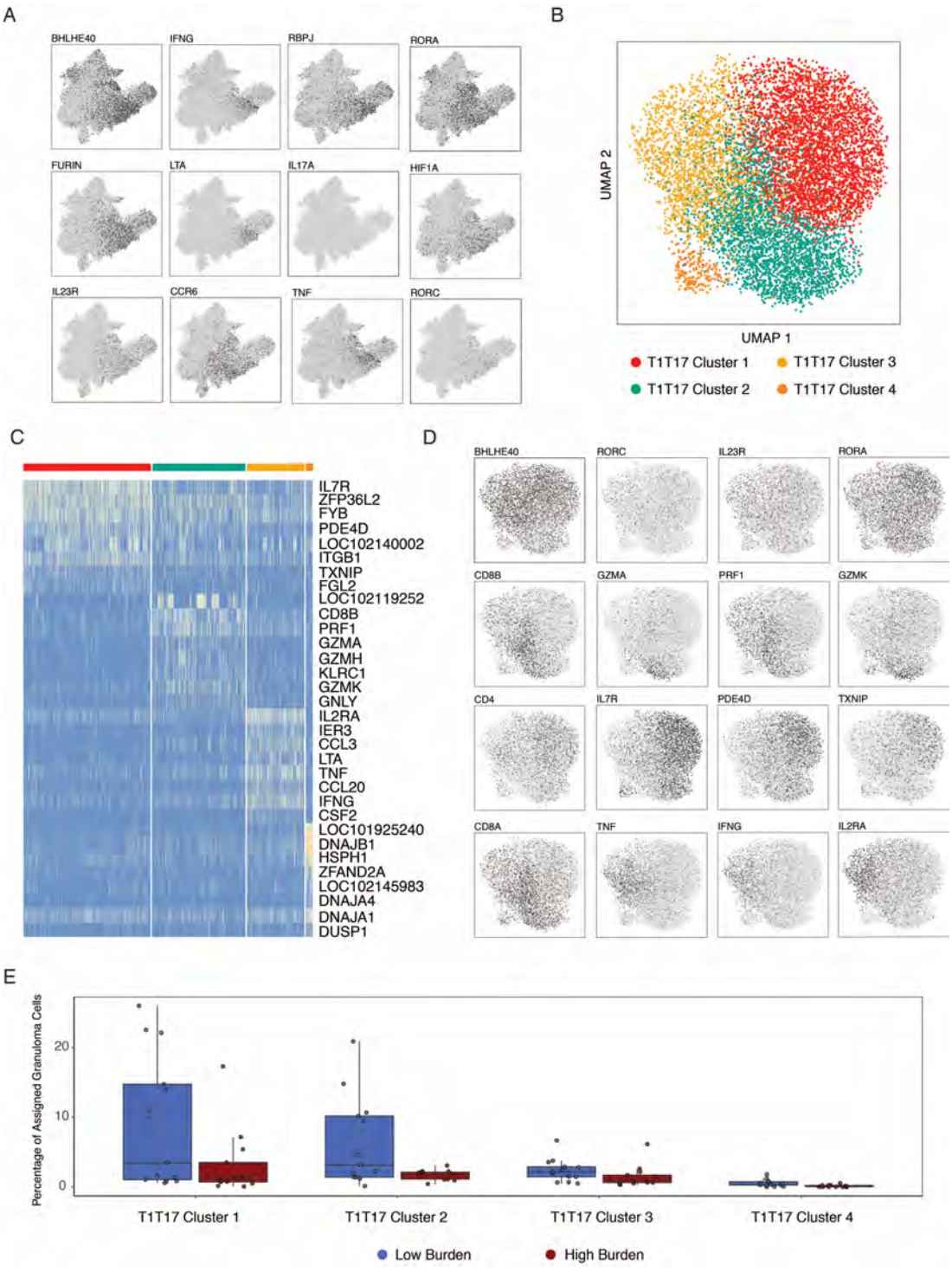


Figure 4.4 | Phenotypic Diversity in T1-T17 cells. (A) UMAP plot of 41,222 T cells colored by normalized expression values for T1-T17 cluster-defining genes. **(B)** UMAP plot of 9,234 T1-T17 cells colored by T1-T17 phenotypic sub-cluster. **(C)** Heatmap showing the distribution of T1-T17 phenotype-defining gene expression. **(D)** UMAP plots of 9,324 T1-T17 cells colored by normalized gene expression values for selected sub-population defining genes. **(E)** Boxplots showing the relationship between the proportional abundance of T1-T17 sub-clustering and granuloma-level bacterial burden.

Relationship of T cell population to enhanced Mtb killing

We further examined differences in cell-type composition between lesions with the highest levels of bacterial killing (low CFU/CEQ, viable/cumulative Mtb ratio) and those with the lowest killing (**Methods**). In this analysis, T cells (P=0.002) are the most highly associated with elevated bacterial killing, while mast cells (P=0.001), endothelial cells (P = 0.016), and macrophages (P=0.038) are associated with reduced bacterial killing (**Figure 4.5A**). Among T cell sub-population, we observe the strongest relationships with increased bacterial killing among effector CD8 cells (P=0.008), naïve-like T cells (P=0.008), and T1-T17 cells (P=0.018).

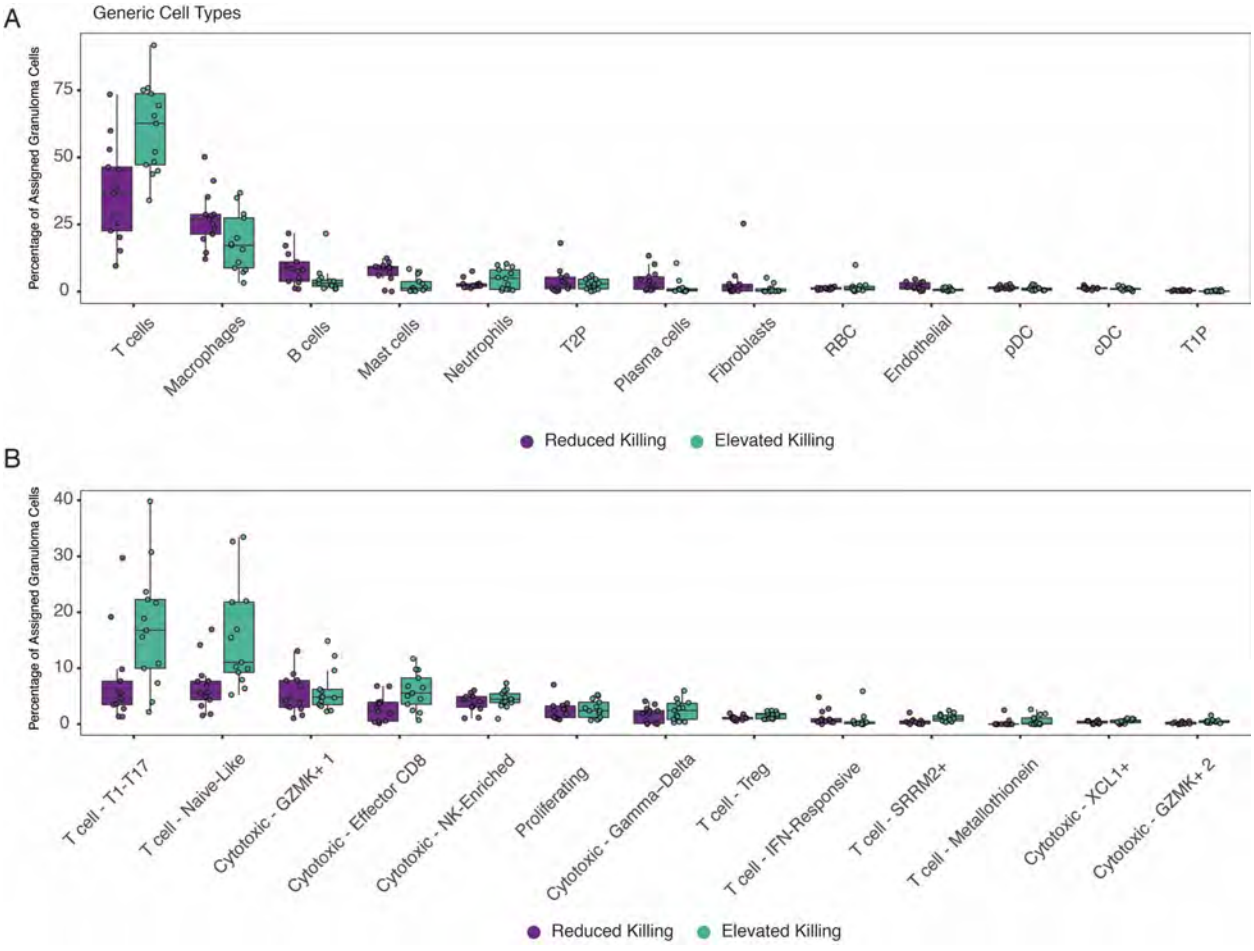


Figure 4.5 | Relationship between Cell-type Composition and Mtb Killing. (A) Boxplots showing the relationship between the level of bacterial killing and cell-type composition among 13 generic cell types. Lesions with reduced Mtb killing are shown in purple while lesions with elevated Mtb killing are in teal. **(B)** Boxplots showing the relationship between the level of bacterial killing and the relative abundance of T cell sub-populations.

Clonotype-phenotype relationships in MTB granulomas

We hypothesized that differences in antigen specificity might contribute to differences in bacterial burden between high and low burden lesions. To understand the extent of T cell clonality and its relationship to Mtb-restrictive T cell phenotypes, we performed enrichment of $\alpha\beta$ TCR sequences from whole-transcriptome amplification libraries and recovered CDR3 sequences from granuloma T cells (TCR-α: 22.5%, TCR-β: 27.3%, paired-$\alpha\beta$: 11.0%) through targeted sequencing (**Methods**) [35]. Further, we examined the extent of clonal expansion among both TCR-α and TCR-β sequences across lesions (**Figure 4.6A**). Here, we observed extensive sharing of expanded clones (i.e. CDR3 alpha or beta sequences with n >= 2) between granulomas within each animal, but we fail to observe public clones between animals. Based on this observation, we focused our analysis of clonotype sharing between high and low burden lesions on Animal 4017, the animal with the broadest distribution of bacterial burdens, and observed significant sharing of expanded TCRβ CDR3 sequences between high and low burden granulomas (**Figure 4.6B**).

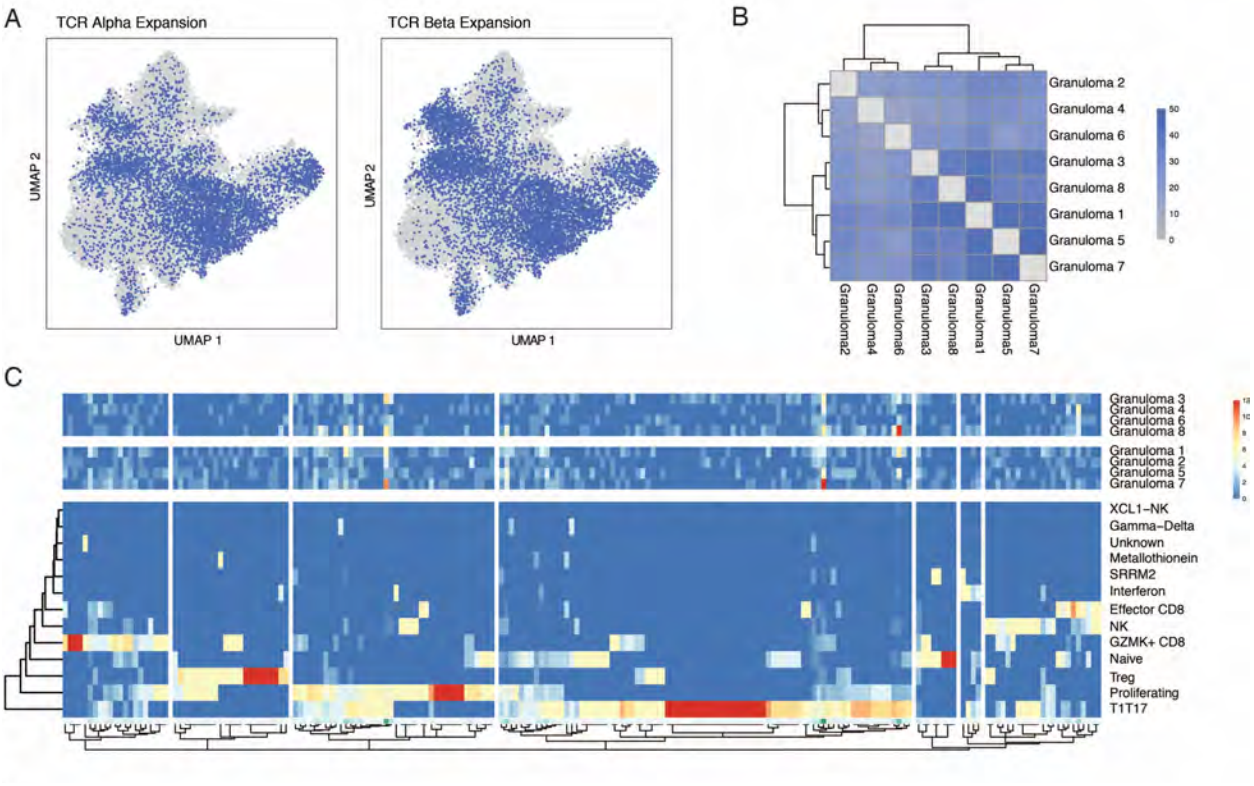


Figure 4.6 | Relationship of T cell clonality to restrictive T cell phenotypes. (A) UMAP plots of 41,222 T cells showing expanded CDR3 alpha (left) and CDR3 beta sequences (right). Cells colored in blue have CDR3 sequences that occur at least twice. **(B)** Heatmap displaying the number of shared CDR3 sequences between granulomas in animal 4017. **(C)** Top: Heatmaps showing the distribution of expanded clones between high (n=4) and low burden (n=4) lesions in animal 4017. Bottom: Heatmap showing the distribution of T cell phenotypic clusters (row) within each expanded clones (columns) in animal 4017. Heatmap colors within each column show the proportion of each phenotype within a clone, while the size of each clone is shown in green below. Column clustering represents similarity in the distribution of phenotypes within clones, while row clustering represents similarity of T cell sub-clusters.

We next sought to understand the relationship between TCR clonality and T cell phenotype. Within animal 4017, we observed the highest extent of clonal expansion among GZMK+ CD8 T cells, effector CD8 T cells, Th1-Th17 cells and proliferating T cells (**Figure 4.6A)**. We examined the phenotypic distribution of individual TCR clones and observed that individual clones are largely restricted to a single T cell phenotype (**Figure 4.6C**). We further examined the extent to which expanded clones adopt different phenotypes between high and low burden granulomas. Here, we observed similar distributions of T cell phenotypes between high and low-burden lesions within expanded TCR clones (**Figure 4.6C**).

Donor-unrestricted T cell represent a heterogeneous class of invariant T cells that recognize non-canonical T cell antigens [36]. We used recovered TCR sequences to identify rare populations of donor-unrestricted T cells including MAIT cells, iNKTs and GEMs (**Methods**). Among DURT cells, we observe the highest frequency of T cells with the *TRAV1-2*/*TRAJ33* MAIT-associated TCR combination (240/41,222; 0.6%). We further observe a population of iNKT cells (*TRAV10-1*/*TRAJ18*) TCR sequences (20/41,222, 0.05%) and GEM cells (*TRAV1-2*/*TRAJ9*). While we are able to identify these populations, their low frequency precludes our ability to accurately assess their relationship to granuloma-level bacterial burden.

Macrophage Population Diversity

We performed sub-clustering analysis among 27,670 macrophages and identified 10 sub-clusters of macrophages (**Figure 4.7A-C; Methods**). To classify granuloma macrophage populations, we initially performed comparisons to literature-derived myeloid signatures [37]. Specifically, we observe multiple populations of macrophages defined by expression of *FUCA1* and *LGMN* (**Figure 4.7D**). We further observe a population of macrophages enriched for expression of lipid-response genes including *FABP4*, *PPARG,* and *MRC1* and another population of inflammatory macrophages enriched for expression of *CXCL9*, *CXCL10*, and *CXCL11*. We identify 2 primary populations of monocytes: (1) a population of inflammatory monocytes (*VCAN*, *S100A6*, and *THBS1*) and (2) a population of CD16+ monocytes (*FCGR3A* and *CX3CR1*). Finally, we observed a population of macrophages that was distinguished by persistent expression of soup-defining and non-macrophage, lineage-defining gene expression (**Methods**), which possibly represents an efferocytotic macrophage population.

For each macrophage sub-population, we examined the relationship with granuloma-level bacterial burden. Here, we observe a significant relationship between PPARG-high, alveolar macrophages and elevated bacterial burden (P=0.0023) (**Figure 4.7E**). Gene expression patterns

in this cluster are consistent with alveolar macrophages that also express a number of genes associated with alternative activation including *MRC1* and *MCEMP1*.

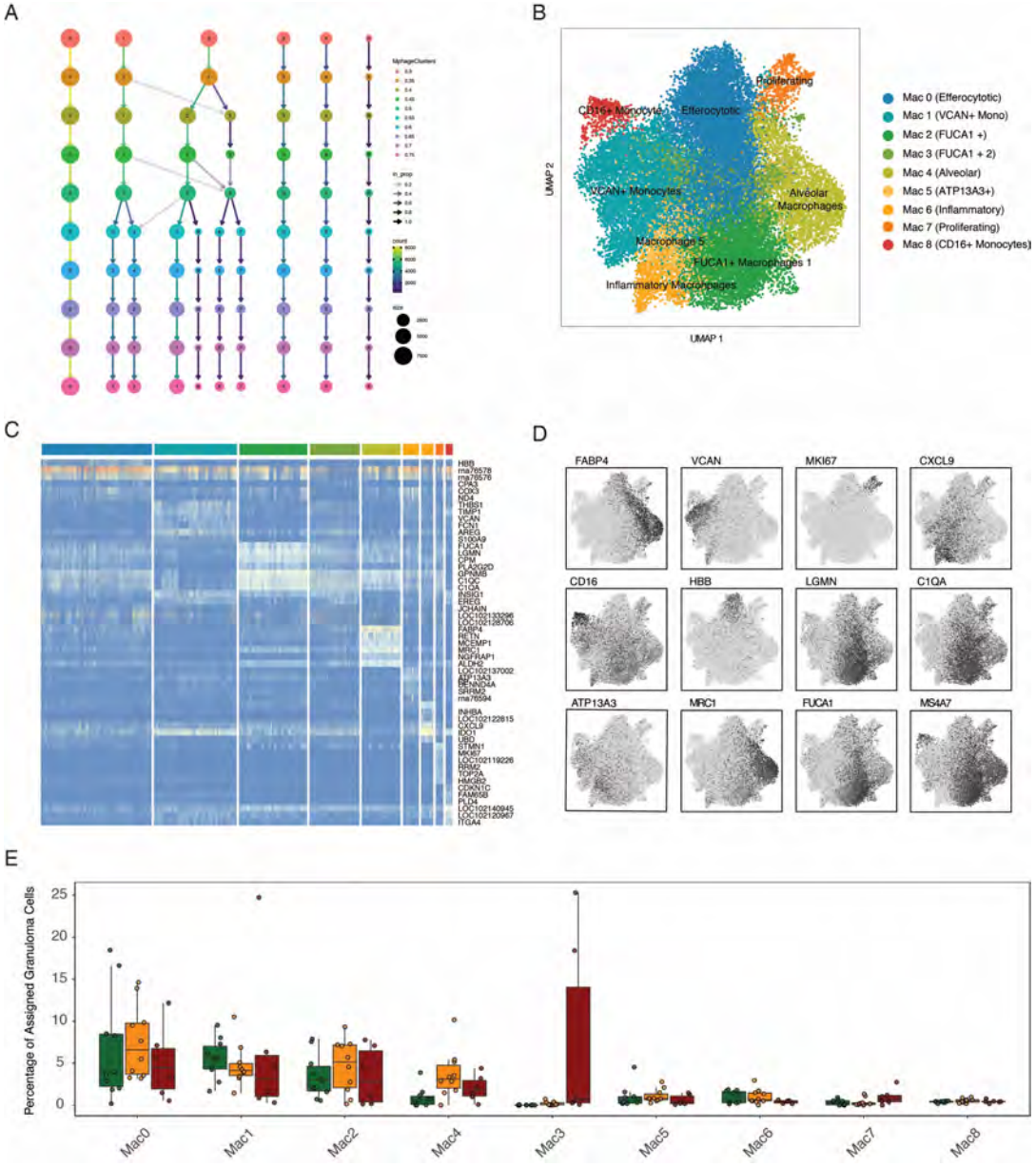


Figure 4.7 | Identification of Macrophage sub-clusters. (A) Waterfall plot showing stability of macrophage clustering to multiple clustering resolutions. **(B)** UMAP plot of 27,670 macrophages from 26 granulomas colored by macrophage sub-clusters. **(C)** Heatmap showing the distribution of cluster-defining gene expression across macrophage subsets. **(D)** UMAP plots of 27,670 macrophages colored by log-normalized expression values for select cluster-defining genes. **(E)** Boxplots showing relationship between the proportion of macrophage sub-populations and granuloma-level bacteria burden (Green: 0-500 CFU), Orange: 500-5000 CFU, Red: > 5000 CFU).

Relationship between timing of granuloma formation and granuloma composition

We hypothesized that timing of granuloma formation might influence the relationship between cell type composition and bacterial control. Examination of granuloma trajectories during the course of infection reveals a subset of lesions (n= 11) only observed on pre-necropsy (10-week) PET-CT scans (**Figure 4.8A**). Notably, we observe a striking relationship between the timing of granuloma formation and granuloma-level bacterial burden, where "late-blooming" lesions have significantly lower CFU values (Original Lesions: Avg-CFU= 9,112 CFU, Late-blooming lesions: Avg-CFU = 159 CFU, P= 0.0035) (**Figure 4.8A**). We next examined differences in cumulative bacterial load to understand whether low CFU among late-blooming lesions was attributable to reduced bacterial growth. Here, we observed similar levels of cumulative growth between original and late-blooming lesions (**Figure 4.8C**; Avg-log10CEQ Original: 4.73, Avg-CEQ Late-blooming: 4.29, P = 0.204), suggesting that low bacterial-burden in late-blooming lesions is likely the result of improved immunologic control. Further, we also observed that late-blooming lesions developed to a median size similar to that of the original granulomas (Original: 2.7 mm, Late-blooming: 2.1 mm) (**Figure 4.8D**).

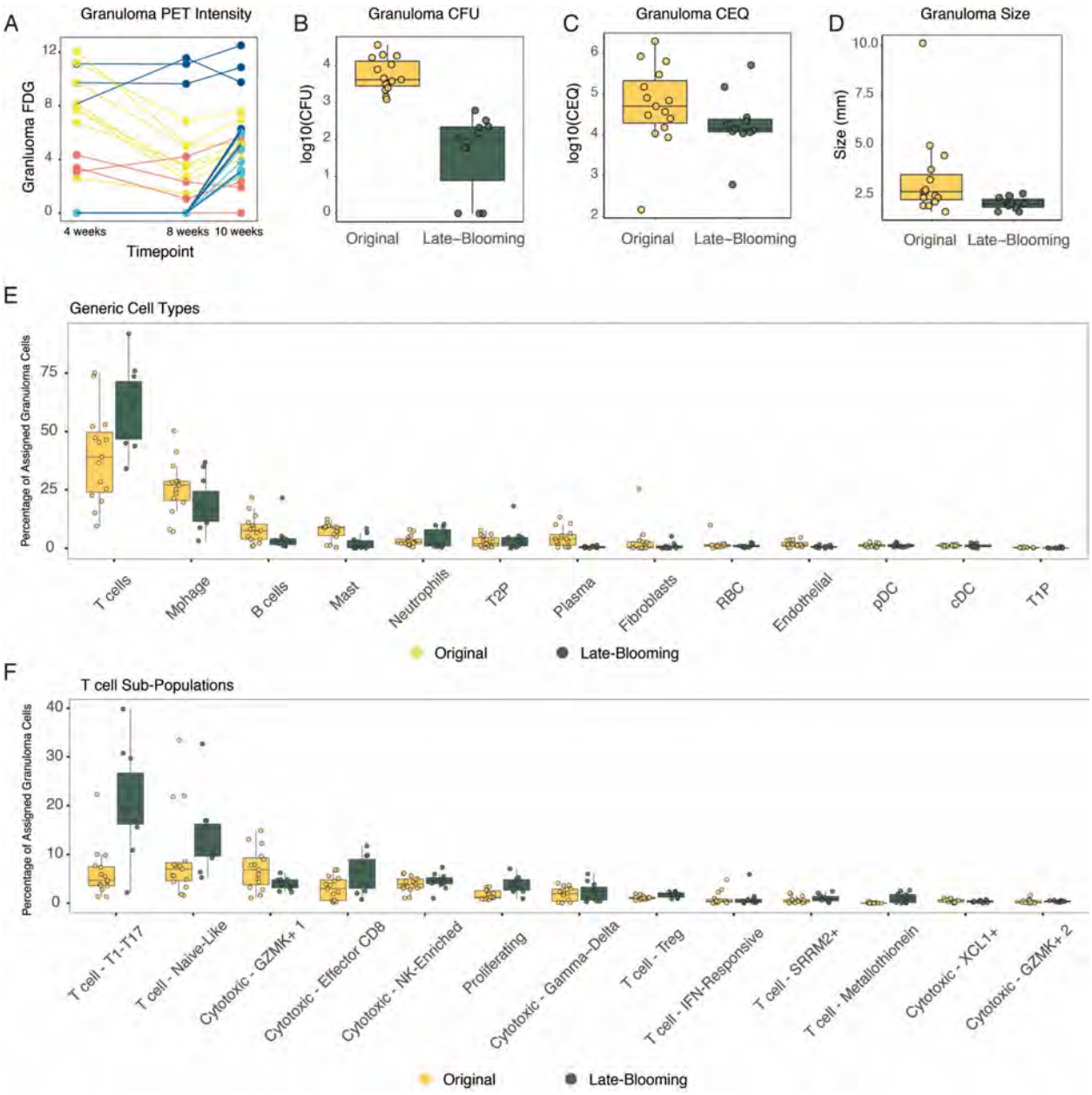


Figure 4.8 | Granuloma Composition Associated with Timing of Granuloma Formation. (A) Scatterplot showing granuloma-level FDG values for 26 granulomas at 4, 8, 10 weeks post-infection. **(B)** Boxplots showing log10-CFU values for granulomas grouped by time of initial observation on PET-CT imaging. **(C)** Boxplots showing log10-CEQ values for granulomas grouped by time of initial observation on PET-CT imaging. **(D)** Boxplots showing granuloma size (mm) for granulomas grouped by time of initial obsesrvation on PET-CT imaging. **(E)** Boxplots showing the relationship between timing of granuloma formation and abundance of generic cell types. **(F)** Boxplots showing the relationship between timing of granuloma formation and abundance of T cell sub-populations.

We examined differences in granuloma composition between original lesions (i.e. those observed on all PET-CT scans) and late-forming lesions (i.e. those observed only at pre-necropsy scans). Here, the associations observed in these comparisons strongly mirror those observed

between high and low-burden comparisons. Specifically, we observe strong associations between mast cells and plasma cells with earlier granuloma formation, while we observe increased proportion of naïve, effector CD8, and T1T17 T cells in late-blooming granulomas (**Figure 4.8F**).

To further investigate the relationship between granuloma cell-type composition and bacterial burden while adjusting for lesional age, we performed comparisons between original lesions with the highest CFU (n= 6, avg. CFU = 18,375) and lowest CFU (n = 6, avg. CFU =2,305; **Figure 4.9A**). Here, we observe previously unappreciated associations between lower bacterial burden and innate lymphoid populations, including the cluster enriched for NK cells (P= 0.036) and gamma-delta T cells (0.0002). Notably, the association of innate cytotoxic populations with lower bacterial burden in early-forming lesions reiterates a role for these populations in early control of MTB infection, while the lower overall proportion of T1-T17 cells and a lack of significant association (P = 0.11) with bacterial control in original lesions suggests that these cells accumulate primarily in late-blooming lesions following the onset of adaptive immunity (**Figure 4.9B**).

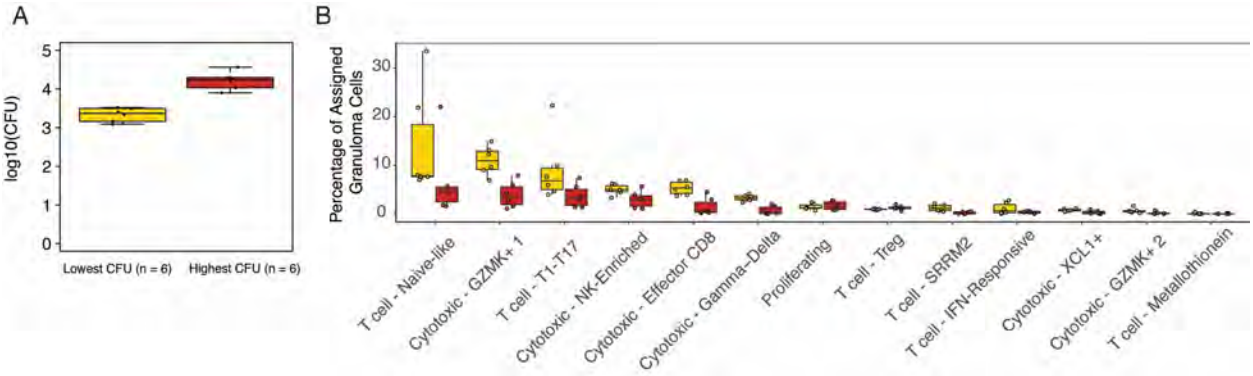


Figure 4.9 | Cellular correlates of bacterial control in original lesions. (A) Boxplot showing differences in granuloma-level bacterial burden between original granulomas with the lowest bacterial burden and original granulomas with the highest bacterial burden. **(B)** Boxplots showing differences in cell-type composition between original granulomas with the highest (n=6) and lowest (n=6) bacterial burdens.

In situ validation of mast cell association with granuloma-level bacterial burden

To validate the observed relationship between mast cells and granuloma-level bacterial burden, we performed staining on a distinct set of 10-week NHP granulomas (n=10) (**Figure 4.10**).

Here, we used performed immunohistochemistry and staining for tryptase to validate the presence of mast cells in NHP granulomas (**Figure 4.10A**). Moreover, examination of the relationship between the number of tryptase-positive cell counts per granulomas reveals a positive relationship between the frequency of mast cells and granuloma-level bacterial burden (**Figure 4.10B**). Upon enumeration of cKIT+ (CD117) cells, we observe an attenuated relationship between burden (**Figure 4.10B**); however, cKIT (CD117) is a less specific marker of mast cells than tryptase.

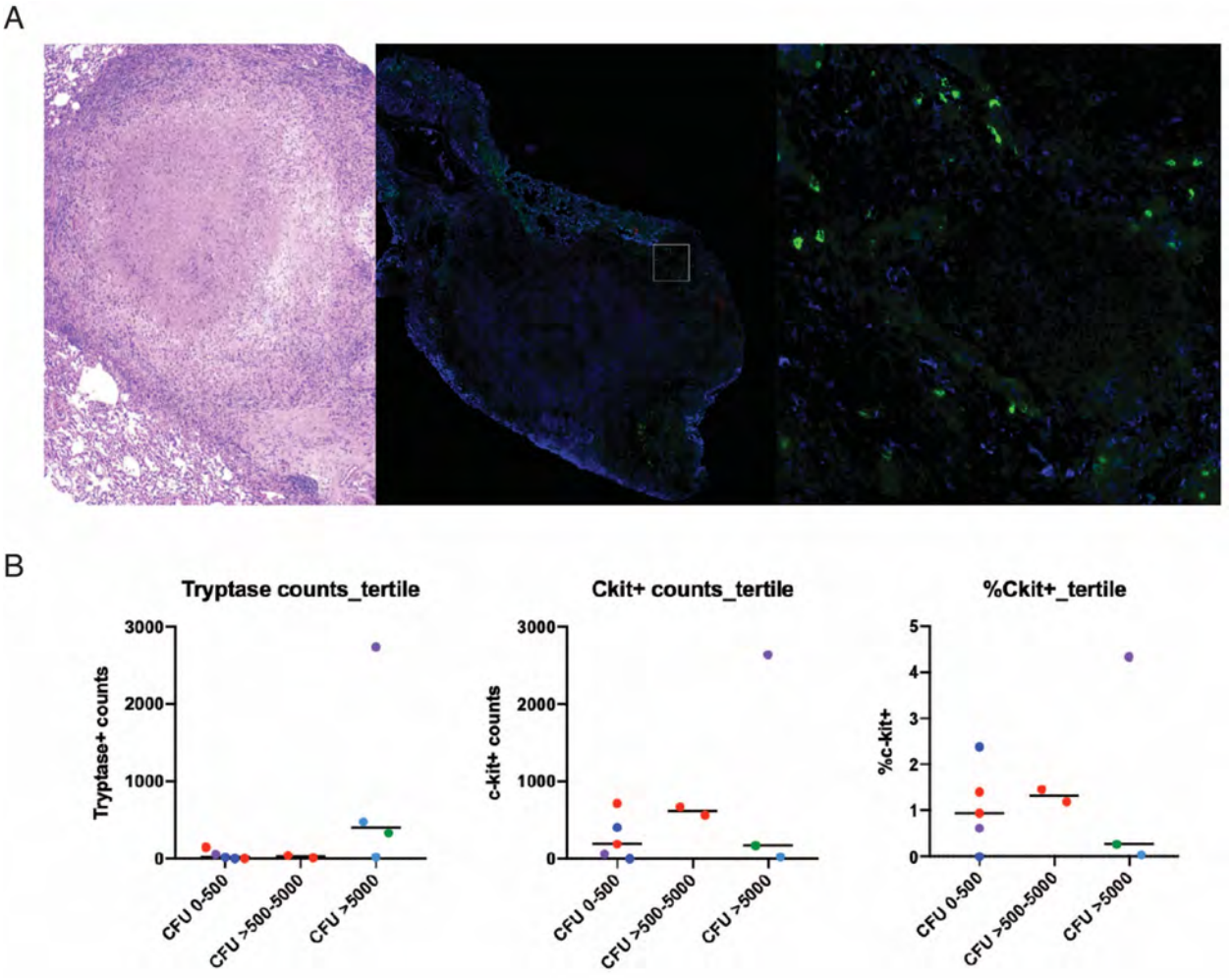


Figure 4.10 | In-situ validation of Granuloma Mast Cells. (A) Left: H&E staining of a necrotic NHP granulomas at 10-weeks post-infection. Right: Fluorescent image showing Tryptase express within the lymphocytic region. **(B)** Quantification of the number of tryptase + cells (left), the number of cKIT+ cells (middle), and the percentage of cKIT+ cells from a representative sample of 10-week lesions with varying bacterial burden.

Cell-Cell Interactions Correlate with Granuloma-level Bacterial Burden

To understand which cell populations co-occur in Mtb granulomas, we calculated pairwise Pearson correlation values between cell types across 26 granulomas (**Figure 4.11A**). Here, we defined 3 primary groups of cells whose abundance is correlated across granulomas using hierarchical clustering (**Figure 4.11A**). We next sought to understand how the cumulative abundance of correlated cell-type groupings varied with granuloma-level bacterial burden. Here, we found that group 1 (green), which consists primarily of T cell sub-populations, is expanded in low-burden lesions, while it is diminished in high-burden lesions. Further, we observed that group 3 (red), which includes mast cells, plasma cells and stromal populations, is broadly expanded in high-burden lesions (**Figure 4.11B**). While these findings largely re-iterate associations of individual cell populations and bacterial burden, correlation of among cell populations suggests that these populations might reflect increased communication between these populations.

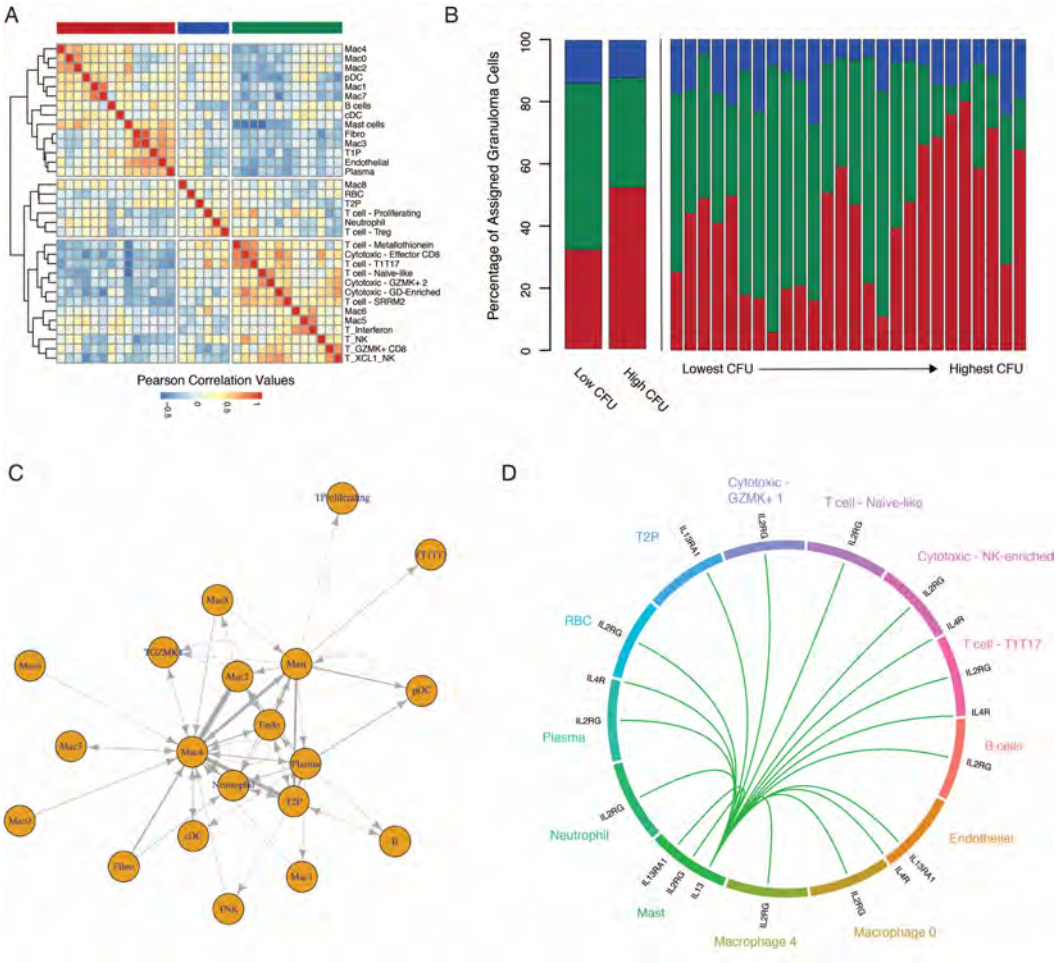


Figure. 4.11 | Cell-Cell Interactions. (A) Heatmap showing pairwise Pearson correlation values among specific cell-type proportions across 26 granulomas. **(B)** Stacked barplots showing the relationship between the distribution of correlated cell-types between high and low burden lesions (left), and across all lesions ordered from lowest burden (left) to highest burden (right). **(C)** Network graph showing cell-cell interaction map of receptor-ligand edge weights over-represented in high burden lesions. **(D)** Circos plot showing receptor-ligand interactions involving mast cell IL-13 signaling increased in high burden lesions.

To explore potential differences in cell-cell signaling between high and low burden lesions, we constructed edge weights for receptor-ligand pairs within each granuloma. Specifically, for each potential interacting cell-type pair we adjusted receptor-ligand edge weights to account for differences in the abundance of the sender cell type, relative receptor expression, and the percent of receptor positive cells (**Methods**). We performed comparisons within cell-cell interaction networks to identify receptor-ligand edge weights over-represented in high or low burden lesions. Initially, we examined interacting cell partners over-represented in high burden lesions by summarizing receptor-ligand edge weights (**Figure 4.11C**). In high burden lesions, we broadly observe increased communication between stromal and immune populations including increased interactions between Type 2 pneumocytes, endothelial cells, and alveolar macrophages (**Figure 4.11C**). Further, in high-burden lesions we observe significant increases in receptor-ligand edge weights involving mast cell IL-13 signaling. Specifically, we observe that IL-13 broadly acts on T cells and macrophage subsets (**Figure 4.11D**), suggesting that mast cells might serve to alter the balance of inflammation within high burden lesions.

Discussion

We have performed high-throughput single-cell profiling of 26 granulomas that span a wide range of bacterial burdens in a non-human primate model of Mtb infection that most closely recapitulates the essential features of human tuberculosis. Critically, by collecting lesions at 10 weeks post-infection, we are able to profile lesions where differences in instantaneous bacterial burden are most likely to result from immune activity. This study design affords a unique window into immune factors that mediate resistance to Mtb infection at the level of individual granulomas.

While we observe broad similarity in the spectrum of immune cell states across high and low burden granulomas, the composition of immune populations varies significantly.

We observe an inverse relationship between the abundance of T cells and the level of granuloma-level bacterial burden. Further, we identify multiple sub-populations of T and NK cells that show varied association with bacterial burden. Most notably, we observe significant increase in the proportion of T1-T17 cells in low-burden granulomas (**Figure 4.3H**). Emergence of a Th1-Th17 population in the lung plays a central role in protective vaccination and natural immunity against numerous pathogens including *B. pertussis* [38-42], *C. albicans* [43, 44], *H. influenzae* [45], *P. aeruginosa* [46, 47], *S. pneumoniae* [48-50], Leishmaniasis [51], and C. neoformans [52-54]. Further, Th17 cells play an important role in *M. tuberculosis* infection and vaccine response [24-27, 55-57].

We observe a significant relationship between when granulomas are first observed on PET-CT and bacterial burden (**Figure 4.8B**). Here, granulomas that are observed only at 10-week scans have consistently lower bacterial burdens, despite having similar cumulative bacterial burden. Granulomas that form at the later stages of infection do so in the presence of a robust adaptive immune response, whereas tissue-level and innate immune responses appear to dominate the early response to Mtb infection. During the early stages of infection, dendritic cells migrate to draining lymph nodes, where they preferentially induce differentiation of effector T cells. The emergence of clonally expanded Th1-Th17 cells shapes the ability of newly formed granulomas to control bacteria by altering ecologic balance from immune tolerance to bacterial control.

Collectively, our data suggest a model of granuloma development in which tissue-mediated cytokine signaling likely leads to early expansion of ST2+ immune populations (e.g. mast cells and tissue Tregs) that generate a tolerogenic immune environment, primarily directed at mitigation of tissue damage and maintenance of tissue function [58]. Notably, we observe

proportional expansion of mast cells, elevated mast cell expression of IL-13, and increased IL-13 receptor-ligand edge weights in high burden granulomas.

In future experiments, we will more fully explore the temporal evolution of immune responses in Mtb granulomas by analyzing a set of granulomas at 4-weeks post-infection. Further, we will examine the effects of depletion of CD8 T cells on bacterial control. Secondary challenge of animals with existing Mtb infection results in formation of granulomas with remarkable capacity to control bacteria. We will further investigate the relationship between cell-type composition, T cell clonality and bacterial control in the setting of re-infection to determine if a similar mechanism of T1-T17 immunity operates in this context. Finally, we will explore whether Mtb-protective immune responses elicited following intravenous BCG vaccination are similar to those observed in the present study.

Materials and Methods

Ethics Statement

All experimental manipulations, protocols, and care of the animals were approved by the University of Pittsburgh School of Medicine Institutional Animal Care and Use Committee (IACUC). The protocol assurance number for our IACUC is D16-00118. Our specific protocol approval numbers for this project are 18124275 and IM-18124275-1. The IACUC adheres to national guidelines established in the Animal Welfare Act (7 U.S.C. Sections 2131 - 2159) and the Guide for the Care and Use of Laboratory Animals (8^{th} Edition) as mandated by the U.S. Public Health Service Policy.

All macaques used in this study were housed at the University of Pittsburgh in rooms with autonomously controlled temperature, humidity, and lighting. Animals were singly housed in caging at least 2 square meters apart that allowed visual and tactile contact with neighboring conspecifics. The macaques were fed twice daily with biscuits formulated for nonhuman primates, supplemented at least 4 days/week with large pieces of fresh fruits or vegetables. Animals had

access to water *ad libitem*. Because our macaques were singly housed due to the infectious nature of these studies, an enhanced enrichment plan was designed and overseen by our nonhuman primate enrichment specialist. This plan has three components. First, species-specific behaviors are encouraged. All animals have access to toys and other manipulata, some of which will be filled with food treats (e.g. frozen fruit, peanut butter, etc.). These are rotated on a regular basis. Puzzle feeders foraging boards, and cardboard tubes containing small food items also are placed in the cage to stimulate foraging behaviors. Adjustable mirrors accessible to the animals stimulate interaction between animals. Second, routine interaction between humans and macaques are encouraged. These interactions occur daily and consist mainly of small food objects offered as enrichment and adhere to established safety protocols. Animal caretakers are encouraged to interact with the animals (by talking or with facial expressions) while performing tasks in the housing area. Routine procedures (e.g. feeding, cage cleaning, etc) are done on a strict schedule to allow the animals to acclimate to a routine daily schedule. Third, all macaques are provided with a variety of visual and auditory stimulation. Housing areas contain either radios or TV/video equipment that play cartoons or other formats designed for children for at least 3 hours each day. The videos and radios are rotated between animal rooms so that the same enrichment is not played repetitively for the same group of animals.

All animals are checked at least twice daily to assess appetite, attitude, activity level, hydration status, etc. Following *M. tuberculosis* infection, the animals are monitored closely for evidence of disease (e.g., anorexia, weight loss, tachypnea, dyspnea, coughing). Physical exams, including weights, are performed on a regular basis. Animals are sedated prior to all veterinary procedures (e.g. blood draws, etc.) using ketamine or other approved drugs. Regular PET/CT imaging is conducted on most of our macaques following infection and has proved very useful for monitoring disease progression. Our veterinary technicians monitor animals especially closely for any signs of pain or distress. If any are noted, appropriate supportive care (e.g. dietary supplementation, rehydration) and clinical treatments (analgesics) are given. Any animal

considered to have advanced disease or intractable pain or distress from any cause is sedated with ketamine and then humanely euthanatized using sodium pentobarbital.

Research Animals

Four Cynomolgus macaques (*Macaca fascicularis*), >4 years of age, (Valley Biosystems, Sacramento, CA) were housed within a Biosafety Level 3 (BSL-3) primate facility as previously described and as above. Animals were infected with low dose *M tuberculosis* (Erdman strain) via bronchoscopic instillation of 7-12 colony-forming units (CFUs)/ monkey to the lower lung lobe. Animals were infected for a period of 10 weeks and Infection was confirmed by tuberculin skin test conversion. Serial clinical, microbiologic, immunologic, and radiographic examinations were performed, as previously described.

Serial PET-CT Imaging

Animals underwent PET-CT scans after Mtb infection at 4weeks, 8 weeks and pre necropsy (i.e. 10 weeks post-infection). Animals were sedated, intubated and imaged by 2-deoxy-2-^{18}F-D-deoxyglucose (FDG) PET imaging (microPET Focus 220 preclinical PET scanner, Seimens Molecular Solutions) and CT scanner (Neurologica Corp) within our biosafety level 3 facility. The total lung FDG avidity was analyzed using Osirix viewer, an open-source PACS workstation and DICOM viewer. The whole lung was segmented on CT by using the Growing region algorithm on the Osirix viewer to create a ROI of normal lung (Hounsfield units < 200). The closing tool was used to include individual nodules and other pulmonary disease. The ROI was transferred to the co-registered PET scan and manually edited to ensure all pulmonary disease was included. Voxels outside the ROI were set to zero and voxels with an SUV greater than or equal to normal lung (SUV > 2.3) were isolated. Finally, the “Expert ROIs” plug-in was then used to export the data from these isolated ROIs to a spreadsheet where the total SUV per voxel were summed to represent the total lung FDG avidity.

Necropsy

Necropsy was performed as previously described [ref]. Briefly, an ^{18}F-FDG PET-CT scan was performed on every animal 1-3 days prior to necropsy to measure disease progression and identify individual granulomas. At necropsy, monkeys were maximally bled and humanely sacrificed using pentobarbital and phenytoin (Beuthanasia; Schering-Plough, Kenilworth, NJ). Individual lesions previously identified by PET-CT and those that were not seen on imaging from lung and mediastinal lymph nodes were obtained for histological analysis, bacterial burden, and immunological studies [ref]. A veterinary pathologist described gross pathologic findings. To quantify gross pathologic disease (disease burden), a necropsy score was developed in which points were given for TB disease: number, size, and pattern of granulomas distributed in each lung lobe and mediastinal lymph node and in other organs each lung lobe, lymph node, and visceral organ were included and enumerated, and used to establish an overall pathology overall score. The size of each granuloma was measured at necropsy and by pre necropsy scans and representative sections of each tissue were homogenized into single-cell suspensions for immunologic studies, flow cytometric analysis, and bacterial burden.

Bacterial burden

200ul of each granuloma homogenate were plated in serial dilutions onto 7H11 medium, and the CFU of *M. tuberculosis* growth were enumerated 21 days later to determine the number of bacilli in each granuloma [ref]. As a quantitative measure of overall bacterial burden, a CFU score was derived from the summation of the log-transformed CFU/gram of each sample at the time of necropsy.

Single-cell mRNA Sequencing

We performed high-throughput single-cell mRNA sequencing (scRNA-Seq) using the Seq-Well platform [16]. Initially, a single-cell suspension was obtained from granulomas isolated at necropsy. Intact granulomas were isolated via manual dissection and carefully separated from surrounding parenchyma. Isolated granulomas were cut into smaller pieces using a razor blade, before being digested with DNAse and Collagenase I and mechanically dissociated using a GentleMacs. For each granuloma, a single-cell suspension was obtained and applied to the surface of a loaded Seq-Well device. Following cell loading, Seq-Well devices were reversibly sealed with a polycarbonate membrane and incubated at 37C for 30 minutes. After membrane sealing, Seq-Well devices were submerged in lysis buffer (5M guanidine thiocyanate, 10mM EDTA, 0.1% Beta-mercaptoethanol, 0.1% Sarkosyl) and rocked for 20 minutes. Following cell lysis, arrays were rocked for 40 minutes in 2M NaCl to promote hybrdization of mRNA to bead-bound capture oligos.

Reverse Transcription and PCR Amplification

Beads were removed from arrays by centrifugation and reverse transcription was performed at 52C for 2 hours. Following reverse transcription arrays were washed with TE-SDS (TE Buffer + 0.1% SDS) and twice with TE-Tween (TE Buffer + 0.01% Tween20). Following ExoI digestion, PCR amplification was performed to generate whole-transcriptome amplification (WTA) libraries. Specifically, a total of 2,000 beads were amplified in each PCR reaction using 16 cycles as previously described [10]. Following PCR amplification, SPRI purification was performed at 0.6x and 0.8x volumetric ratios and eluted samples were quantified using a Qubit. Sequencing libraries were prepared by tagmentation of 800 pg of cDNA input using Illumina Nextera XT reagents. Tagmented libraries were purified using 0.6x and 0.8x volumetric SPRI ratios and final library concentrations were determined using a Qubit. Library size distributions were established using an Agilent TapeStation with D1000 High Sensitivity ScreenTapes (Agilent, Inc., USA).

Sequencing and Alignment

Libraries for each sample were sequenced on a NextSeq550 75 Cycle High Output sequencing kit (Illumina Inc., Sunnyvale, CA, USA). For each library, 20 bases were sequenced in read 1, which contains information for cell barcode (12 bp) and unique molecular identifier (UMI, 8bp), while 50 bases were obtained for each read 2 sequence. Cell barcode and UMI tagging of transcript reads was performed using DropSeqTools v1.12 [59]. Barcode and UMI-tagged sequencing reads were aligned to the Macaca fascicularis v5 genome (https://useast.ensembl.org/Macaca_fascicularis/Info/Index) using the STAR aligner. Aligned reads were then collapsed by barcode and UMI sequences to generate digital gene expression matrices with 10,000 barcodes for each array.

Single-cell TCR Reconstruction

We performed single-cell TCR reconstruction across all 28 granulomas. Initially, we performed targeted enrichment of TCR constant genes using biotinylated probes as in Tu et al. (Nature Immunology 2019). In short, DNA was denatured with a 15 minute incubation at 95C, 80-mer hybridization probes (5' biotin) specific to alpha and beta TCR constant chain were added along with a blocking oligo (SMART universal PCR primer) to prevent re-annealing of denatured cDNA molecules. Hybrdization was accomplished through a 60 minute incubation at 65C. Hybridized products were enriched using streptavidin coated magentic beads (M-270 Dynabeads). Magnetic beads were sequentially washed with buffers contained in the IDT xGen® hybridization kit (e.g. Wash Buffer 1, Stringent Wash Buffer, Wash Buffer 2, and Wash Buffer 3; Cat # 1080584 IDT DNA, Inc.). After washing beads, PCR was performed to amplify enriched cDNA molecules. Amplified TCR products were purified using custom SeraPure Beads at a 0.64x ratio. Following an initial primer extension, additional PCR amplification was performed using a series of V-specific primer sequences. Sequencing was performed by combining alpha and beta TCR libraries at equimolar ratios.

Data Processing and Quality Control

Initially, we generated a combined dataset of 202,386 barcodes after applying thresholds of 500 genes and 750 transcripts (UMIs), after examining a range of cell inclusion thresholds. For each array, we initially visualized cells from each array using t-SNE and performed Louvain clustering in Seurat. For many arrays, we observed large clusters of cell barcodes that were not marked by distinct cell-type defining gene expression. Instead, these cells were marked by distributed, low-level expression of genes presumed to originate from other cell types (e.g. *HBB* from erythrocytes, *JCHAIN* from plasma cells, and *CPA3* from mast cells). To understand the identity of these barcodes more fully, we initially examined sequencing quality metrics. Notably, these non-descript clusters did not significantly differ in the total number of aligned reads, detected genes, UMIs/cell, or mitochondrial percentage.

To more fully understand the identity of these clusters, we attempted multiple modeling approaches:

1. We attempted to model these clusters as array-specific doublets. Here, we constructed models in which pseudo-doublets/multiplets (n=2, 5, 10, 15, or 20 cells) were created from random sampling of the remaining cell-type clusters. However, in these models, we failed to observe significant overlap between generated pseudo-multiplets and the clusters with non-distinct gene expression patterns.
2. We attempted to create random cells by binomial sampling a pseudo-population average expression vector generated by summation across well-defined cell populations. In these models, we failed to observe direct overlap between the simulated mixed population and those clusters with non-distinct gene expression patterns.
3. Finally, we examined whether these clusters might represent deep sampling of ambient contamination or cellular debris by generating a “contamination” scoring scheme. First, to

identify the clusters within each array, we calculated 30 principal components (this was observed to consistently capture the majority of variation in each array), and performed Louvain clustering (resolution = 1.25) using all significant principal components (JackStraw Empirical P-value < 0.05). Next, within each array, we generated cluster-specific "contamination" scores that consisted of 3 components:

a. **A measure of array-specific background contamination by cluster ("soup expression").** For each array, we generated a background expression profile based on low-UMI barcodes (See ***Correction for Residual Background Contamination*** below for full details). We identified a set of "soup"-defining genes at a range of thresholds for soup-defining gene expression (0.01, 0.005, 0.001, and 0.0005; based on the proportional contribution of each gene to the cumulative soup expression profile for each array). We then generated a series of array-specific, background-contamination scores for the set of soup-defining transcripts using the AddModuleScore function in Seurat. We observed that clusters with ambiguous/overlapping expression of lineage-defining gene expression signatures (Erythrocytes: HBB, Plasma cells: JCHAIN, Mast cells: CPA3, etc.) were significantly enriched for soup-defining gene expression. Finally, to calculate "contamination' scores, we used expression scores for soup genes at a threshold of 0.001, and calculated the average soup-profile score both unadjusted and scaled values across average soup-signature scores for each cluster within each array.
b. **An estimate of biological signal ("biological signal")**. Here, we examined the average log-fold change for the top 5 genes enriched within each cluster. For clusters dominated by ambient RNA, we broadly observed lower fold changes for their biological signature genes compared to clusters characterized by expression of canonical cluster-defining genes. In cases where the highest average log-fold change

values within a cluster were below the "return threshold" in Seurat, we set the value to the default return threshold of 0.2.

c. **A measure of co-expression of lineage-defining genes ("soup linage coexpression")**. We manually selected 5 genes that were recurrently over-represented in clusters suspected to arise from ambient contamination and cellular debris. Specifically, we selected the following genes: HBB (An erythrocyte-defining gene), JCHAIN (A plasma cell defining gene), COL3A1 (A fibroblast defining gene), SFTPC (A type 2 pneumocyte defining genes), and CPA3 (A mast cell defining gene). For each cell barcode, we calculated the number of these five genes with non-zero expression. Within each cluster, we then calculated the average co-expression of these genes and subtracted one from this average to allow for endogenous expression of 1 lineage-defining gene. This parameter was specifically added to avoid exclusion of *bona fide* cell clusters with high-background contamination (presumably due to low endogenous RNA content) and low biological signal (e.g., naïve T cells). Here, we specifically observed that cell populations which scored high for markers of a single lineage yet had higher soup-expression scores presented with lower rates of co-expression of these soup and lineage defining transcripts relative to clusters which did not, likely representing ambient RNA and debris.

Using these three values, we then calculated cluster-specific background "contamination" scores for each array in 2 ways:

$$Contamination\ Score\ 1 = \frac{(Soup\ Expression)\ x\ (Soup\ Lineage\ Coexpression)}{Biological\ Signal}$$

$$Contamination\ Score\ 2 = \frac{(Scaled - Soup\ Expression)\ x\ (Soup\ Lineage\ Coexpression)}{Biological\ Signal}$$

These two "contamination" scores quantify both the (1) absolute and (2) relative soup-profile contamination in downstream cluster classification.

Next, for each array, we performed clustering to identify clusters with array-specific ambient contamination and debris. More specifically, we performed hierarchical clustering using a total of 7 variables: the 2 contamination scores (shown above),three scaled soup scores (soup gene thresholds: 0.01, 0.05 and 0.001), the average log-fold change for the top 5 cluster genes, and soup/lineage gene co-expression to identify clusters defined by ambient contamination. For each array, we cut the hierarchical clustering tree at the first branch point to identify clusters with a signature of ambient contamination. In total, we identified 41 array-specific clusters comprising 56,590 barcodes from 21 out of 32 total arrays, as characterized by ambient RNA contamination and cellular debris and removed them in all subsequent analyses.

Correction for Residual Background Contamination

After removal of cell barcodes that were derived from background contamination and extracellular debris, we performed additional correction for ambient RNA contamination among remaining cell barcodes on an array-by-array basis. Among filtered cell barcodes, we observed array-specific, ambient RNA contamination marked by ectopic expression of cell-type defining genes (e.g. widespread expression of *JCHAIN*, *HBB*, and *CPA3* etc.). Specifically, we observed this contamination to vary in relation to the overall distribution of cell-types recovered from each array. To correct for residual ambient contamination within each array, we used SoupX [60] to: (1) generate array-specific profiles of background contamination and (2) estimate per-cell contamination fractions, and (3) generate corrected background-corrected UMI counts matrices. To generate background expression profiles, we first generated counts matrices containing up to

50,000 barcodes to assemble a collection of low-UMI cell barcodes that presumably represent extracellular mRNA. For each array, a UMI threshold for background expression was determined using EmptyDrops [61] to estimate the likelihood distribution of low-UMI barcodes having optimal cellular identity. Using an array-specific UMI-threshold (Range: 20-100 UMIs), we separately generated composite background profiles for each array. To estimate per-cell contamination fraction, we first identified a set of lineage-defining genes with bimodal expression patterns across cells (i.e., lineage defining genes with leaky expression). Finally, the composite soup profile was subtracted from each the transcriptional profile of each cell based on the estimated contamination fraction. For each array, we removed individual transcripts most likely to be contamination from each single-cell based on the estimated contamination fraction. Specifically, individual transcripts were sequentially removed from each single-cell transcriptome until the probability of subsequent transcripts being soup-derived was less than 0.5 to generate a background-corrected counts matrix for each array.

Separation of Doublets

Within each array, we also performed doublet identification and separation using DoubletFinder. To account for differences in cell loading densities and expected cell doublet frequencies, we generated array-specific estimates of the expected number of doublets. For example, a total of 20,000 cells applied to a Seq-Well device containing 85,000 wells (lambda = 20,000), we would calculate an expected doublet rate of >2.37% (since not all of the array's surface area contains wells). For each array, we generated pseudo-doublets using DoubletFinder [62]. Here, we optimized the pK parameter estimates for each array separately by performing a parameter sweep in which we selected the pK value with the maximum bimodality coefficient, while we maintained pN = 0.25 across all arrays based on published recommendations [62]. Cells were identified as doublets based on their rank order in the distribution of the proportion of artificial nearest neighbors (pANN). Specifically, we identified the pANN value for the cell at the expected

doublet percentile and used the corresponding pANN value as a threshold to remove additional cells in the event of ties. In total, we excluded 3,852 cells as doublets.

Integrated Cell Type Classification

Following the aforementioned quality filtering, we obtained a combined dataset of 109,584 cells. We performed an initial dimensionality reduction on these cells by selecting variable genes, principal component analysis (PCA), UMAP dimensionality reduction and Louvain clustering using Scanpy [63]. To identify broad cell-types, we examined cluster assignments at multiple levels of clustering resolution (Resolutions: 0.5 to 2.25). We selected a cluster resolution of 1.00 because this was the resolution beyond which branching did not result in discovery of clusters that represent distinct cell lineages (e.g., division of Type 1 and Type 2 pneumocytes). To define these major cell-populations, we first performed extensive comparisons to existing signatures of lung parenchyma and immune cell populations in the Tabula Muris [17] and Mouse Cell Atlas [18] studies. Specifically, we collected lung single-cell sequencing data from both studies and calculated enriched gene expression signatures for each lung cell-type cluster using a Wilcox rank-sum test. For each cluster, we selected the top 20 genes as a cluster-specific expression signature and then used them to score all cells in the granuloma dataset. We calculated the average signature score within each cluster and examined the distribution of signature score within each granuloma cell-type and determined significance via permutation testing.

We further performed extensive literature-based curation of gene expression signatures in support of identified cell type classifications. Through these approaches, we identified multiple uniform populations that were consistently detected irrespective of cluster resolution. These populations include conventional DCs, plasmacytoid DCs, pneumocytes (Type 1 and 2), endothelial cells, fibroblasts, erythrocytes, regulatory T cells, and proliferating cells. At this point in our analysis, we had already observed significant sub-variation among T cell and macrophage clusters; however, we generated broad grouping of T cells and macrophages and then proceeed

to sub-cluster those separately using variable genes specific to each (See ***Sub-clustering of Granuloma T cells and Macrophages***).

Cell Type Assignment of Proliferating Cells

Among our top-level clusters was one defined by markers of cellular proliferation (*MKI67*, *TOP2A*, and *CDK1*). To identify the underlying cell type identity for these cells, we executed a separate dimensionality reduction and clustering among 3,123 cells defined by this proliferation signature. We performed UMAP dimensionality reduction and Louvain clustering at multiple clustering resolutions (0.4-0.8), and selected a resolution of 0.70. For each of the major cell-type identified in the global clustering analysis, we generated a gene signature using the top 20 enriched genes and scored the proliferating cells clusters using the AddModuleScore function in Seurat. We then examined the distribution of cell-type signature scores across each of the sub-clusters of proliferating cells and re-assigned clusters based on enrichment of lineage-specific gene expression. Here, we assessed the significance of the cluster scores using a permutation test. More specifically, we performed 1,000 permutations in which we first down-sampled the proliferating clusters to have the same number of cells. We then randomized the cluster assignments of the cells and calculated the average generic cell-type signature score for each randomized cluster. The significance of a cell-type score for each proliferating cluster was determined by comparing the observed average signature score to the random null distribution). Through this approach, we identified and re-assigned distinct clusters of proliferating B cells, macrophages, neutrophils, plasma cells, and T cells to their respective cell types.

Filtering of Soup-Defining Transcripts

To avoid artifacts from ambient RNA contamination and cellular debris in sub-clustering of T cells and macrophages, we excluded genes that were observed to be soup-defining for any array. Specifically, we identified a set of 210 soup defining genes that comprised 0.001 of total

soup expression in any array. The threshold of 0.001 was selected to maximize the cumulative fraction of soup expression with the least number of genes to avoid removing underlying biology. In a further effort to avoid removing cell-type specific biology, we retained any genes with average log-fold changes greater than 1.00 in T cells and macrophages compared to all other generic cell types. In total, we removed 204 and 180 genes prior to sub-clustering analysis of T cells and macrophages, respectively.

Sub-clustering of Granuloma T cells

Across the complete set of 44,766 T and NK cells, we initially performed Louvain clustering at a range of resolution of values (0.30 – 0.75) and examined the relationships between cluster membership. In this analysis, we consistently observed a cluster defined by persistent expression of contaminating transcripts derived from macrophage and mast cells (Cluster 4 - Louvain Resolution 0.60). To confirm that these cells did not represent persistent doublets, we scored all T cells by expression of the top 20 cluster defining T cells and observed similar signature scores between the contaminated cell population. To understand whether residual contamination obscured additional T cell biology, we performed sub-clustering within the “contaminated” T cell cluster. In this analysis, we failed to observe T cell clusters not identified among the remaining non-contaminated populations. Since this contamination cluster was not observed to obscure a novel T cell phenotype, we decided to exclude this population from downstream analysis. Following removal of the cluster of T cells defined by residual contamination, we performed dimensionality reduction and clustering at multiple clustering resolutions (Louvain resolution: 0.25 – 0.75). In this final analysis, we identified 12 T cell populations at a clustering resolution of 0.75.

Classification of T cells Populations

We then classified T cell populations using a combination of manual curation and comparison to literature-derived sequences. We compared granuloma T cell populations to

publicly available T cell population and single-cell mRNA-Seq signatures. Specifically, we performed comparisons in the following ways:

1. For each T cell cluster, we compared its cluster-defining gene to publicly available databases of immune signatures, including MSigDb [64] and Savant [19]. Specifically, we performed gene-set enrichment for the top 20 cluster-defining genes in Piano using a hypergeometric test with correction for multiple hypothesis testing using the Benjamini-Hochberg method for false discovery.
2. We compared each T cell cluster to literature-derived signatures of T cells from another scRNA-Seq study. Here, we generated cell signature scores in Seurat using the AddModuleScore function using gene expression signatures obtained from human lung cancer [20]. To determine the significance of these score, we performed 1,000 permutations in which T cell cluster identity was randomly re-assigned to generate a null distribution of module scores.
3. Finally, we performed extensive manual curation based on literature evidence. For each cell population, we carefully examined patterns of enrichment and performed extensive literature search to support classification of T cell sub-populations. For example, we identified regulatory T cells on the basis of expression of known regulatory T cell markers (FOXP3, IKZF1, and TNFSF18/GITR). However, in many cases, surface markers used to define canonical T cell populations do not directly correspond to single-cell mRNA sequencing data.

Initially, we distinguished T cells from NK cells based on express recovery of $\alpha\beta$-TCR sequences. We identified NK/cytotoxic cells based on reduced recovery of $\alpha\beta$-TCR sequences and similarity to NK gene signatures (GZMH, GZMB, GNLY, and PRF1). We identified an additional population of T cells defined by expression of $\gamma\delta$ TCR constant genes. Upon sub-

clustering this population of cells (***See Sub-clustering of Gamma-Delta T cell Cluster***), we identified a second NK population characterized by elevated expression of XCL1.

Next, we defined T cell clusters based on the expression of CD4 and CD8A and CD8B. Specifically, we identified 3 T cell clusters with elevated expression of CD4, 2 T cell clusters with elevated expression of CD8A/B, and 2 clusters with mixed expression. We identified multiple CD4+ T cell populations, including naïve T cells (*LEF1*, *TCF7*, *SELL*, and *CCR7*), regulatory T cells (*FOXP3*, *IKZF1*, and *IL1R1*), and a population of interferon-responsive CD4+ T cells (*MX1* and *ISG15*). Among CD8+ T cells, we identified a population of CD8 T cells characterized by expression of GZMK and a population of effector CD8 T cells (*GZMB*, *CX3CR1*, and *TGFBR3*) (Fig. SX and Table SX). Finally, we observed 2 populations of T cells expressing both CD4 and CD8 including proliferating T cells (*MKI67* and *TOP2A*) and a population of T1/T17 cells (*RORA*, *RBPJ*, and *BHLHE40*). We performed additional sub-clustering and observed multiple clusters that segregate closely with expression of CD4 and CD8 (***See Sub-clustering of T1-T17 Cells***).

Sub-clustering of Gamma-Delta and Cytotoxic Cluster

We performed additional sub-clustering with within the population of 2,377 $\gamma\delta$ and cytotoxic T cells. Specifically, we performed dimensionality reduction and clustering at multiple resolutions (0.30 – 0.75). Here, we observed 2 primary populations of cells: (1) a population of $\gamma\delta$ T cells and (2) a population of XCL1+ NK cells. We performed differential expression analysis to determine differences in gene expression between these clusters upon which we based the classification of these cells.

Sub-clustering of T1-T17 Cells

Based on the observation that both CD4+ and CD8+ T cells comprise the Th1/Th17 cluster, we performed sub-clustering analysis within this population. We performed variable gene

identification, dimensionality reduction and Louvain clustering (Resolution = 0.55) and identified 4 distinct sub-population. 3 of 4 sub-clusters within the T1T17 population correspond closely to CD4 (Cluster 1) and CD8 T cells (Cluster 2 and 3). Within the T1T17 population, we observed a fourth population defined by expression of heat-shock and DNA-damage response proteins.

T cell Functional Analysis

To identify functional gene expression signatures of bacterial control, we performed differential expression within each T cell subset between T cells recovered from high and low burden lesions using a Wilcox test. Within each T cell subset, we further performed correction for multiple testing using false-discovery rate correction (Benjamini-Hochberg). We specifically performed directed analysis to examine signatures of T cell exhaustion and cytotoxic effector function among CD8 T cell sub-populations. In each case, we examined the expression of individual genes in addition to combined metrics of gene expression across multiple genes. To calculate gene expression signature scores from multiple genes, we used the AddModuleScore function in Seurat. We examined the distribution of functional gene programs within and between T cell sub-populations.

Analysis of TCR Data

We initially examined the rates of TCR recovery for both alpha and beta CDR3 sequences across T cell clusters. Specifically, we recovered αCDR3 sequences in 22.5%, β-CDR3 in 27.3% and paired $\alpha\beta$-CDR3 sequences in 11% of T cells. We examined the extent of clonal expansion among CDR3 sequences based on sharing of alpha and beta CDR3 sequences. Here, expanded clones were identified as CDR3 alpha or beta sequences observed at least twice. We examined the distribution of alpha and beta CDR3 sequences across granulomas.

Having observed few public clones, we sought to understand the relationship between clonality and phenotype among expanded clones within individual animals. For each clone, we calculated the distribution of T cell phenotypes with expanded clones for both α and β CDR3 sequences. In particular, we primarily focused our analysis to animal 4017, where we sampled lesions with a clear bifurcation in bacterial burden, For each expanded clone, we examined the distribution of phenotypes between high and low-burden lesions.

Sub-clustering of Granuloma Macrophages

Across 27,670 macrophages, we initially performed Louvain clustering and dimensionality reduction at multiple clustering resolutions. In our initial clustering, we identified a cluster defined by contaminating transcripts derived from other cell types including mast cells (*KIT* and *CLU*), T cells (CD3D and), and plasma cells (*JCHAIN*) and soup-defining gene expression. We compared the distribution of macrophage-defining gene expression in this cluster to other clusters and observed this cluster to have enriched signature scores relative other clusters. We then performed a sub-clustering analysis in which we used the top 20 cluster-defining genes from non-contaminated macrophage clusters as variable genes.

We examined the enrichment of macrophage expression signatures and determined that the population of macrophages that have a core macrophage expression program. While we observe a population of macrophages that is primarily soup-defining gene expression, we did not exclude this cluster due to the possibility that this represents an efferocytotic macrophage population.

Classification of Macrophage Populations

We established the identities of the macrophage clusters through a combination of manual curation and comparison to published gene expression signatures from both population and single-cell mRNA sequencing studies. More specifically:

1. For each macrophage cluster, we performed similar comparison to databases of immune signatures including MSigDb and Savant (See ***Identification of T cell Populations***).
2. We generated gene expression signatures from published scRNA-Seq studies of macrophage states. For example, we used a recently published atlas of myeloid states in lung [37] to score granuloma macrophages. Further, we generated a list of myeloid expression signatures using lung myeloid cells from the mouse cell atlas. For each study, we generated signatures for the top 20 cluster-defining genes to generate gene expression signatures. Signature scores were generated for each cell using the AddModuleScore function in Seurat.
3. Finally, in cases where we failed to discover existing description of a macrophage population, we performed extensive literature searches to contextualize possible identity of macrophage populations.

Comparison to signatures from Zillonis et al., reveals a high degree of similarity to multiple clusters. Specifically, we observe multiple populations of macrophages defined by expression of APOE and APOC1. We identified a consistent population characterized by expression of FABP4, MRC1, and PPARG. We observe another population of inflammatory macrophages enriched for expression of CXCL9, CXCL10, and CXCL11. We identify 2 primary populations of monocytes: (1) a population of inflammatory monocytes (*VCAN*, *S100A6*, and *THBS1*) and (2) a population of CD16+ monocytes (*FCGR3A* and *CX3CR1*).

Granuloma-level Correlates of Bacterial Burden

Initially, we calculated the proportion of cell type clusters observed in each granuloma with the numerator as the number of cells of a specific phenotype and the denominator as the total number of cells in a given granuloma.

We then examined cellular and molecular correlates of granuloma-level bacterial burden in two ways:

1. We performed correlation analysis of cell-type proportions and end-point CFU level across all granulomas using a Spearman-rank correlation test. All Spearman rank correlation tests were adjusted for animal-specific effects and corrected for multiple testing using Benjamini-Hochberg False-Discovery correction.
2. We performed a T-test for each cell-type between high and low-burden lesions, correcting for multiple testing using Benjamini-Hochberg False-discovery correction.

Relationship between bacterial killing and cell type composition

We performed targeted sequencing (CEQ Values) from granuloma lysates to determine the number of chromosomal equivalents from each lesion. This metric provides a measure of cumulative bacterial burden. As a measure of granuloma-level bacteria killing, we used the ratio of CEQ (cumulative bacterial burden) and CFU (CFU: viable bacterial load). For the CEQ:CFU ratio, higher values correspond to elevated bacterial killing and lower values correspond to lower bacterial killing. To understand the relationship between cell-type composition and CEQ:CFU ratio, we examined the continuous relationship between CFU:CEQ ratios and proportional cell type composition using a spearman rank-correlation test. We also divided lesions into 2 groups based on CEQ:CFU ratio and performed a T-test, correcting for multiple testing using Benjamini-Hochberg False-discovery correction.

Relationship between cell-type composition and timing of granuloma formation

Based on serial PET-CT imaging, we identified lesions that formed at different points in infection. Specifically, there was a group of lesions that were observed starting at 4-weeks post-infection that we classified as ‘original’ lesions and a distinct group of lesions that were observed only at 10-week scans (i.e. pre-necropsy scans) that we classified as ‘late-blooming’.

We examined the relationship between the timing of granuloma formation and granuloma-level cell-type composition. Specifically, we divided lesions into 2 groups based on the scan on which they were first observed. Lesions observed on PET-CT scans at 4, 8 and 10 week scans were classified as “original” lesions (n = 15), while those lesions observed only at 10-week scans were classified as “late-blooming” lesions. We calculated cell-type proportions within each granuloma as the number of cells of each type divided by the total of number cells in each granuloma. We examined differences in the abundance of cell-types between original and late-blooming lesions using a T-test and significance of association was determined through Benjamini-Hochberg FDR correction.

Since we observe a strong association between bacterial burden and timing of granuloma formation, we performed additional analysis restricted to original lesions. Here, we selected original lesions with the highest (n=6, Avg. CFU = 18,375) and lowest (n=6 Avg. CFU = 2,305). Here, we performed a T-test to examine differences in the cell type composition between original lesions with the highest bacterial burden and original lesions with the lowest bacterial burden, correcting for multiple testing using Benjamini-Hochberg False-discovery correction.

Co-variation in Granuloma Composition

We calculated correlations in cell-type proportions to identify underlying structure in the co-occurrence of cell types across all granulomas. Specifically, we calculated Pearson correlation coefficients for all pair-wise cell-type combinations. We also examined the robustness of granuloma composition analysis to spearman correlations. Further, we performed hierarchical clustering to identify clusters of correlated cell-types across granulomas, and calculated the proportional composition of correlated cell-type clusters within each lesion. To understand the relationship between identified cell-type clusters and granuloma-level bacterial burden, we examined the abundance of correlated cell types by grouping lesions by bacterial burden (CFU ranges: 0-500, 500-5,000, and >5,000).

Cell-Communication Analysis

To examine cell-cell interactions, we first generated a curated list of receptor-ligand pairs through a combination of publicly-available databases and literature review [65]. Within each granuloma, we generated edge weights between cell types for a given receptor ligand pair by multiplying the average receptor expression in Cell Type 1 by the average ligand expression in Cell Type 2. Edge weights were constructed for all receptor-ligand pairs and pairwise-cell type combinations within granulomas individually. Within each granuloma, we performed a total of 1000 permutations for each receptor-ligand pair in which cell-type identifiers were randomly resorted and the resulting edge weight was recorded. For each receptor-ligand pair, the significance of the observed value was calculated from a z-score comparison of the observed value relative the permuted values.

We further performed adjustment of receptor-ligand edge weights at multiple levels. (1) We examined differences in unweighted receptor ligand edge weights Comparisons of receptor-ligand edge weights on the basis of bacterial burden (granuloma-level CFU) and timing of granuloma formation (time of first PET-CT observation). (2) To account for differences in the relative abundance of 'sender' cell types, we multiplied receptor-ligand edge weights by the proportion of 'sender' cell types. In effect, this generates a pool of 'sender' cell derived ligand that is available to act upon cell types bearing appropriate receptors. (3) To account for variability in the expression of receptor expression across cell population, we weighted receptor-ligand edge-weights by the proportion of total receptor expression within the receiving cell type cluster relative the aggregate receptor expression. In this scheme, receptors with more uniform expression across cell type clusters will be down-weighted to reflect non-autonomous sinks of extracellular ligands, while receptors predominantly expressed by a single cell type will be up-weighted. (4) Finally, we adjusted receptor-ligand edge weights to account for the percent of cells within the receiver population expressing a given receptor.

To identify axes of intercellular communication with differential weights across granulomas, we performed t-tests of receptor-ligand edge weights between (1) high-burden and low-burden lesions, (2) original and late-blooming lesions. We filtered results based on the following criteria: (1) average permutation p-values within high or low-burden lesions < 0.05, (2) p-value from student's t-test < 0.05, (3) fold-change > 0, and (4) receptor-ligand edge weights were ranked by receptor-ligand edge weights. We created a Shiny app to facilitate exploration and interpretation of cell-cell interaction data (https://tkhughes2.shinyapps.io/Shiny/).

References

1. Organization, W.H., *Global tuberculosis report 2019.* 2019.

2. Russell, D.G., C.E. Barry, and J.L. Flynn, *Tuberculosis: what we don't know can, and does, hurt us.* Science, 2010. **328**(5980): p. 852-856.

3. Lin, P.L., et al., *Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing.* Nature medicine, 2014. **20**(1): p. 75.

4. Flynn, J.L., et al., *Non-human primates: a model for tuberculosis research.* Tuberculosis (Edinb), 2003. **83**(1-3): p. 116-8.

5. Phuah, J.Y., et al., *Activated B cells in the granulomas of nonhuman primates infected with Mycobacterium tuberculosis.* Am J Pathol, 2012. **181**(2): p. 508-14.

6. Ulrichs, T. and S.H. Kaufmann, *New insights into the function of granulomas in human tuberculosis.* J Pathol, 2006. **208**(2): p. 261-9.

7. Mattila, J.T., et al., *Microenvironments in tuberculous granulomas are delineated by distinct populations of macrophage subsets and expression of nitric oxide synthase and arginase isoforms.* J Immunol, 2013. **191**(2): p. 773-84.

8. Gideon, H.P., et al., *Variability in tuberculosis granuloma T cell responses exists, but a balance of pro- and anti-inflammatory cytokines is associated with sterilization.* PLoS Pathog, 2015. **11**(1): p. e1004603.

9. Flynn, J.L., et al., *Immunology studies in non-human primate models of tuberculosis.* Immunol Rev, 2015. **264**(1): p. 60-73.

10. O'Garra, A., et al., *The immune response in tuberculosis.* Annu Rev Immunol, 2013. **31**: p. 475-527.

11. Lin, P.L., et al., *Radiologic Responses in Cynomolgus Macaques for Assessing Tuberculosis Chemotherapy Regimens.* Antimicrob Agents Chemother, 2013. **57**(9): p. 4237-4244.

12. Coleman, M.T., et al., *PET/CT imaging reveals a therapeutic response to oxazolidinones in macaques and humans with tuberculosis.* Sci Transl Med, 2014. **6**(265): p. 265ra167.

13. Cadena, A., et al., *Early in vivo infection dynamics in non-human primates infected with Mycobacterium tuberculosis (MPF3P. 813)*. 2014, Am Assoc Immnol.

14. Martin, C.J., et al., *Digitally barcoding Mycobacterium tuberculosis reveals in vivo infection dynamics in the macaque model of tuberculosis.* MBio, 2017. **8**(3): p. e00312-17.

15. Cadena, A.M., et al., *Concurrent infection with Mycobacterium tuberculosis confers robust protection against secondary infection in macaques.* PLoS pathogens, 2018. **14**(10): p. e1007305.

16. Gierahn, T.M., et al., *Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput.* Nature methods, 2017. **14**(4): p. 395-398.

17. Consortium, T.M., *Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris.* Nature, 2018. **562**(7727): p. 367.

18. Han, X., et al., *Mapping the mouse cell atlas by microwell-seq.* Cell, 2018. **172**(5): p. 1091-1107. e17.

19. Lopez, D., et al., *SaVanT: a web-based tool for the sample-level visualization of molecular signatures in gene expression profiles.* BMC genomics, 2017. **18**(1): p. 824.

20. Guo, X., et al., *Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing.* Nature medicine, 2018. **24**(7): p. 978-985.

21. Böttcher, J.P., et al., *NK cells stimulate recruitment of cDC1 into the tumor microenvironment promoting cancer immune control.* Cell, 2018. **172**(5): p. 1022-1037. e14.

22. Panduro, M., C. Benoist, and D. Mathis, *Tissue tregs.* Annual review of immunology, 2016. **34**: p. 609-633.

23. Zanini, I.M., et al., *Human cactin interacts with DHX8 and SRRM2 to assure efficient pre-mRNA splicing and sister chromatid cohesion.* J Cell Sci, 2017. **130**(4): p. 767-778.

24. Scriba, T.J., et al., *Distinct, specific IL-17-and IL-22-producing CD4+ T cell subsets contribute to the human anti-mycobacterial immune response.* The Journal of Immunology, 2008. **180**(3): p. 1962-1970.

25. Van Dis, E., et al., *STING-activating adjuvants elicit a Th17 immune response and protect against Mycobacterium tuberculosis infection.* Cell reports, 2018. **23**(5): p. 1435-1447.

26. Khader, S.A., et al., *IL-23 and IL-17 in the establishment of protective pulmonary CD4+ T cell responses after vaccination and during Mycobacterium tuberculosis challenge.* Nature immunology, 2007. **8**(4): p. 369-377.

27. Gopal, R., et al., *Unexpected role for IL-17 in protective immunity against hypervirulent Mycobacterium tuberculosis HN878 infection.* PLoS pathogens, 2014. **10**(5).

28. Yasuda, K., et al., *Satb1 regulates the effector program of encephalitogenic tissue Th17 cells in chronic inflammation.* Nature communications, 2019. **10**(1): p. 1-14.

29. Lin, C.-C., et al., *Bhlhe40 controls cytokine production by T cells and is essential for pathogenicity in autoimmune neuroinflammation.* Nature communications, 2014. **5**(1): p. 1-13.

30. Lin, C.-C., et al., *IL-1–induced Bhlhe40 identifies pathogenic T helper cells in a model of autoimmune neuroinflammation.* Journal of Experimental Medicine, 2016. **213**(2): p. 251-271.

31. Yu, F., et al., *The transcription factor Bhlhe40 is a switch of inflammatory versus antiinflammatory Th1 cell fate determination.* Journal of Experimental Medicine, 2018. **215**(7): p. 1813-1821.

32. zu Horste, G.M., et al., *RBPJ controls development of pathogenic Th17 cells by regulating IL-23 receptor expression.* Cell reports, 2016. **16**(2): p. 392-404.

33. Acosta-Rodriguez, E.V., et al., *Surface phenotype and antigenic specificity of human interleukin 17–producing T helper memory cells.* Nature immunology, 2007. **8**(6): p. 639-646.

34. Vesely, M.C.A., et al., *Effector TH17 Cells Give Rise to Long-Lived TRM Cells that Are Essential for an Immediate Response against Bacterial Infection.* Cell, 2019. **178**(5): p. 1176-1188. e15.

35. Tu, A.A., et al., *TCR sequencing paired with massively parallel 3' RNA-seq reveals clonotypic T cell signatures.* Nature immunology, 2019. **20**(12): p. 1692-1699.

36. Van Rhijn, I. and D.B. Moody, *Donor unrestricted T cells: a shared human T cell response.* The Journal of Immunology, 2015. **195**(5): p. 1927-1932.

37. Zilionis, R., et al., *Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species.* Immunity, 2019. **50**(5): p. 1317-1334. e10.

38. Warfel, J. and T. Merkel, *Bordetella pertussis infection induces a mucosal IL-17 response and long-lived Th17 and Th1 immune memory cells in nonhuman primates.* Mucosal immunology, 2013. **6**(4): p. 787-796.

39. Ross, P.J., et al., *Relative contribution of Th1 and Th17 cells in adaptive immunity to Bordetella pertussis: towards the rational design of an improved acellular pertussis vaccine.* PLoS pathogens, 2013. **9**(4).

40. Fedele, G., et al., *Attenuated Bordetella pertussis vaccine candidate BPZE1 promotes human dendritic cell CCL21-induced migration and drives a Th1/Th17 response.* The Journal of Immunology, 2011. **186**(9): p. 5388-5396.

41. Schnoeller, C., et al., *Attenuated Bordetella pertussis vaccine protects against respiratory syncytial virus disease via an IL-17–dependent mechanism.* American journal of respiratory and critical care medicine, 2014. **189**(2): p. 194-202.

42. Higgins, S.C., et al., *TLR4 mediates vaccine-induced protective cellular immunity to Bordetella pertussis: role of IL-17-producing T cells.* The Journal of Immunology, 2006. **177**(11): p. 7980-7989.

43. Pandiyan, P., et al., *CD4+ CD25+ Foxp3+ regulatory T cells promote Th17 cells in vitro and enhance host resistance in mouse Candida albicans Th17 cell infection model.* Immunity, 2011. **34**(3): p. 422-434.

44. Lin, L., et al., *Th1-Th17 cells mediate protective adaptive immunity against Staphylococcus aureus and Candida albicans infection in mice.* PLoS pathogens, 2009. **5**(12).

45. Li, W., et al., *Recognition of conserved antigens by Th17 cells provides broad protection against pulmonary Haemophilus influenzae infection.* Proceedings of the National Academy of Sciences, 2018. **115**(30): p. E7149-E7157.

46. Wu, W., et al., *Th17-stimulating protein vaccines confer protection against Pseudomonas aeruginosa pneumonia.* American journal of respiratory and critical care medicine, 2012. **186**(5): p. 420-427.

47. Liu, J., et al., *Early production of IL-17 protects against acute pulmonary Pseudomonas aeruginosa infection in mice.* FEMS Immunology & Medical Microbiology, 2011. **61**(2): p. 179-188.

48. Wright, A.K., et al., *Experimental human pneumococcal carriage augments IL-17A-dependent T-cell defence of the lung.* PLoS pathogens, 2013. **9**(3).

49. Moffitt, K.L., et al., *TH17-based vaccine design for prevention of Streptococcus pneumoniae colonization.* Cell host & microbe, 2011. **9**(2): p. 158-165.

50. Moffitt, K.L., R. Malley, and Y.-J. Lu, *Identification of protective pneumococcal TH17 antigens from the soluble fraction of a killed whole cell vaccine.* PloS one, 2012. **7**(8).

51. Wu, W., L. Huang, and S. Mendez, *A live Leishmania major vaccine containing CpG motifs induces the de novo generation of Th17 cells in C57BL/6 mice.* European journal of immunology, 2010. **40**(9): p. 2517-2527.

52. Zhang, Y., et al., *Robust Th1 and Th17 immunity supports pulmonary clearance but cannot prevent systemic dissemination of highly virulent Cryptococcus neoformans H99.* The American journal of pathology, 2009. **175**(6): p. 2489-2500.

53. Müller, U., et al., *IL-13 induces disease-promoting type 2 cytokines, alternatively activated macrophages and allergic inflammation during pulmonary infection of mice with Cryptococcus neoformans.* The Journal of Immunology, 2007. **179**(8): p. 5367-5377.

54. Murdock, B.J., et al., *Early or late IL-10 blockade enhances Th1 and Th17 effector responses and promotes fungal clearance in mice with cryptococcal lung infection.* The Journal of Immunology, 2014. **193**(8): p. 4107-4116.

55. Khader, S.A., et al., *IL-23 compensates for the absence of IL-12p70 and is essential for the IL-17 response during tuberculosis but is dispensable for protection and antigen-specific IFN-γ responses if IL-12p70 is available.* The Journal of Immunology, 2005. **175**(2): p. 788-795.

56. Wozniak, T.M., et al., *Mycobacterium bovis BCG-specific Th17 cells confer partial protection against Mycobacterium tuberculosis infection in the absence of gamma interferon.* Infection and immunity, 2010. **78**(10): p. 4187-4194.

57. da Costa, A.C., et al., *A new recombinant BCG vaccine induces specific Th17 and Th1 effector cells with higher protective efficacy against tuberculosis.* PLoS One, 2014. **9**(11).

58. Liu, Q., et al., *IL-33–mediated IL-13 secretion by ST2+ Tregs controls inflammation after lung injury.* JCI insight, 2019. **4**(6).

59. Macosko, E.Z., et al., *Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets.* Cell, 2015. **161**(5): p. 1202-1214.

60. Young, M.D. and S. Behjati, *SoupX removes ambient RNA contamination from droplet based single cell RNA sequencing data.* BioRxiv, 2018: p. 303727.

61. Lun, A.T., et al., *EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data.* Genome biology, 2019. **20**(1): p. 63.

62. McGinnis, C.S., L.M. Murrow, and Z.J. Gartner, *DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors.* Cell systems, 2019. **8**(4): p. 329-337. e4.

63. Wolf, F.A., P. Angerer, and F.J. Theis, *SCANPY: large-scale single-cell gene expression data analysis.* Genome biology, 2018. **19**(1): p. 15.

64. Liberzon, A., et al., *Molecular signatures database (MSigDB) 3.0.* Bioinformatics, 2011. **27**(12): p. 1739-1740.

65. Ramilowski, J.A., et al., *A draft network of ligand–receptor-mediated multicellular signalling in human.* Nature communications, 2015. **6**(1): p. 1-12.

Chapter 5: Conclusions and Future Directions

Authorship Attributions

Parts of this chapter are based on work described in:

Prevention of Mycobacterium tuberculosis infection and disease in nonhuman primates following intravenous BCG immunization.

Patricia A. Darrah, Joseph J. Zeppa, Joshua A. Hackney, Marc H. Wadsworth II, Travis K. Hughes, Supriya Pokkali, Phillip A. Swanson, Nicole L. Grant, Mark A. Rodgers, Megha Kamath, Charles A. Scanga, Chelsea M. Causgrove, Dominick J. Laddy, Aurelio Bonavia, Danilo Casimiro, Philiana Ling Lin, Edwin Klein, Alexander G. White, Pauline Maiello, Alex K. Shalek, Mario Roederer, JoAnne Flynn, Robert A. Seder. (2020) *Nature*.

5.1 Summary and Discussion of Main Findings

We have worked to develop and deploy a novel technology platform to understand the relationship between cellular composition and bacterial burden in Mtb granulomas. Development of the Seq-Well platform has uniquely enabled high-throughput single-cell mRNA sequencing to study *M. tuberculosis* granulomas in BSL3 facilities (**Chapter 2**). The use of a portable, low-cost device enables the technique to be performed in virtually any lab environment around the world. Further, improvements in the sensitivity of transcript capture have enabled increased per-cell information content (**Chapter 3**). These improvements have enabled increased capture efficiency of biologically informative lowly expressed transcripts including cytokines, receptors, and transcription factors. Finally, using Seq-Well, we have performed high-throughput scRNA-seq to explore the cellular and molecular features of bacterial control in MTB granulomas (**Chapter 4**).

In Chapter 2, we report the development of Seq-Well, a portable, low-cost platform for high-throughput single-cell mRNA sequencing. Here, uniquely barcoded mRNA capture beads are confined in a portable PDMS device that contains approximately 85,000 individual microwells. Following loading of beads, cells are loaded and microwell arrays are reversibly sealed using a semipermeable membrane. Following lysis, mRNA is captured and reverse transcription is

performed to generate cDNA. Following PCR amplification and Illumina library preparation, high-throughput sequencing is performed to generate single-cell sequencing data (**Figure 2.1** and **2.2**). Critically, membrane sealing of Seq-Well devices enables increased capture efficiency of cellular mRNAs while minimizing transcript loss and cross-talk between adjacent wells. To validate the performance of Seq-Well in challenging environments and demonstrate its compatibility with select pathogen research, we performed single-cell sequencing on macrophages exposed to Mtb in a BSL3 facility. Here, we observed multiple macrophage populations that display altered expression patterns following exposure to Mtb (**Figure 2.3**).

In Chapter 3, we developed an improved version of the Seq-Well protocol (Seq-Well S^3) that results in significant increases in per-cell gene detection and transcript capture. Specifically, Seq-Well S^3 increases per-cell information content by random second strand synthesis that more efficiently appends a second primer sequence to cDNA molecules following reverse transcription (**Figure 3.1**). In the original version of the Seq-Well protocol, the use of template switching results in loss of molecules that were successfully captured and reversed transcribed, but without successful template switching, are not efficiently amplified during PCR amplification. The second-strand synthesis technique enables recovery of those molecules that are otherwise lost in transcription by random hybridization to first strand cDNA molecules on mRNA capture beads. We validated the performance of the Seq-Well S^3 technique in cell lines and primary cells and show that the performance of the technique is on par with best in class commercial alternatives (e.g. 10x Genomics v2 and v3) (**Figure A2.1-3**). We further use the improved Seq-Well S^3 to generate an atlas of inflammatory skin conditions. Specifically, we sequenced a total of 38,274 cells from 19 skin biopsies across multiple inflammatory skin conditions including acne, alopecia areata, granuloma annulare, leprosy, psoriasis and normal human skin (**Figures 3.2-3.6**). In this work we simultaneously demonstrate the utility of Seq-Well S^3 to uncover disease relevant biology by identifying multiple disease-specific immune and stromal phenotypes.

In Chapter 4, we used Seq-Well to explore the relationship between Mtb granuloma cell-type composition and granuloma-level bacterial burden. Here, cynomolgous macaques (n=4) were bronchoscopically infected with a low-dose inoculum of Mtb (7-11 bacilli) and serial PET-CT was performed to track the development of granulomas at 4, 8 and 10 weeks post-infection. In total, we performed Seq-Well on 26 granulomas from 4 animals and obtained 109,584 single cells. Across lesions, we examined the relationship between cell-type composition and granuloma-level bacterial burden. Here, we discovered the strongest link between abundance of T cells and reduced bacterial burden, while we found an unexpected link between mast and plasma cell with increased bacterial burden. Further, we observed elevated proportions of specific T cell subsets with reduced bacterial burden including Naïve T cells, effector CD8 T cells, and a population of T1T17 cells (**Figure 4.3**). We further performed high-throughput immune receptor profiling for granuloma T cells, and observe close relationship between T cell clonality and phenotype (**Figure 4.6**). We further report a nuanced relationship between timing of granuloma formation, cell-type composition, and bacterial burden (**Figure 4.8**). Specifically, in early-forming lesions we observe expansion of mast and plasma cells, while late-blooming, low-burden lesions are enriched for a population of T1-T17 cells following the onset of adaptive immunity. Finally, we examined expression of ligand-receptor pairs across lesions and observed increased mast cell expression of IL-13 in high burden lesions acting on multiple T cell and macrophage populations (**Figure 4.11**).

5.2 Strengths and Weaknesses of Approaches

Strengths

Single-cell mRNA sequencing enables unbiased characterization of complex populations of immune and non-immune cells the comprise TB granuloma and other complex tissue-immune ecosystems. Previous methods for single-cell profiling (e.g. flow cytometry or mass cytometry) are critically limited in the number of cellular features that can be simultaneously analyzed, which

limits the potential for novel discovery. Further, these methods generally require higher numbers of cells than can be obtained from most granulomas. Among methods for single-cell mRNA sequencing, Seq-Well [1] uniquely enables single-cell analysis of individual Mtb granulomas. Through single-cell sequencing, we have been able to identify the presence of previously unappreciated cells populations in Mtb granulomas like mast cells. Using recently developed methods, we have also been able to reconstruct TCR immune repertoires from TB granulomas [2]. Importantly, our data reveal an important role for T1-T17 cells in the control of Mtb infection within granulomas.

Impact

The impact of the techniques that I have worked to develop, optimize and apply has extended far beyond my own work. Since the publication of Seq-Well, I have worked to train over 100 scientists from around the world to perform the technique, and I have personally performed experiments on every continent minus Antarctica. Through numerous collaborations, I have worked to optimize experiments in the setting of Ebola infection (Collaboration with Pardis Sabeti and NIH-IRF), Malaria Infection (Mahidol Tropical Research Institute, Bangkok, Thailand), Zika infection (Gehrke lab – MIT), human Mtb infection (Leslie – AHRI, Durban, South Africa), leprosy infection (Modlin Lab, UCLA and Sarno Lab, Fiocruz), environmental enteric dysfunction (Gartner Lab, UCSF and Kelley Lab, Zambia), multi-route BCG vaccination (Seder and Roederer Labs, NIAID-VRC), variant detection in acute myeloid leukemia [3] (Berstein Lab, MGH), and many others. In addition to academic collaborations, this work has resulted in multiple patent applications which have led to the founding of Honeycomb, a company working to develop a commercial version of the Seq-Well platform.

Weaknesses

While Seq-Well enables high-throughput characterization of complex tissue samples and immune responses, the need for tissue dissociation prior to cell isolation eradicates the spatial architecture of MTB granulomas. Further, as currently implemented, accurate and comprehensive discovery of infection status of macrophages is currently not possible. In the initial study of MTB granulomas described in Chapter 4, single-cell sequencing was performed using the original version of the Seq-Well protocol, which has reduced gene detection and transcript recovery relative the Seq-Well S^3 protocol (Chapter 3).

In the experiments outlined in chapter 4, there are a limited number of animals (n=4) used for infection and sequencing. While we have made efforts to ensure that associations are not driven by animal-specific features of disease, additional animal would provide increased confidence in the strength of our findings. To this end, we have performed infections in an additional 4 cynomolgus macaques (42 granulomas). Granulomas were obtained at 10-weeks post-infection; however, this set of granulomas is composed exclusively of “original” lesions (i.e. those observed on PET-CT at 4, 8, and 10-week scans), which has led to a significant reduction in the range of granuloma-level bacterial burdens observed. Despite the reduced range of bacterial burden, we are able to observe elevated proportions of mast cells, consistent with original, high-burden lesions to the set of 26 granulomas described in Chapter 4.

In work described in this dissertation, we have not performed interventional studies in mouse or NHP to validate findings associated with improved bacterial control. To this end, we have since begun to perform studies in which antibody depletion of CD8 T cells is performed prior to infection. We have also performed scRNA-Seq in the context of Mtb bre-infection, where lesions that form upon secondary challenge show remarkable ability to contain bacteria [4]. Finally, in an attempt to understand early events in granuloma formation, we have performed scRNA-Seq in lesions at 4-weeks post-infection.

5.3 Alternative approaches:

Spatially-resolved Profiling of Mtb Granulomas

Mtb granulomas have a stereotyped spatial architecture that undoubtedly influences cell-cell interactions and their influence on the structure and function of granulomas. Unfortunately, mechanical and enzymatic dissociation of granulomas prior to application to Seq-Well devices eradicate valuable information about the spatial relationships among granuloma cell-types.

Fortunately, multiple methods for high-throughput spatial profiling are available. Protein-based methods for highly multiplexed imaging enable spatial profiling of primary samples. Multiplexed ion-beam imaging (MIBI) utilizes antibodies labeled with isotopically pure heavy metals and can simultaneously profile up to 100 targets [5]. Notably, there is ongoing work to explore phenotypic findings generated in the context of Mtb granulomas using MIBI in a Gates collaboration with Mike Angelo at Stanford.

In combination with existing data, techniques for spatially-resolved sequencing would provide a complimentary view of granulomatous inflammation. Methods for spatial transcriptomics are rapidly maturing, and offer exciting opportunities for biological discovery. First, there are a number of techniques that provide sub-cellular resolution using either multiplexed variations of fluorescent *in situ* hybridization (FISH) or *in situ* amplification and sequencing [6-9]. These methods, however, have significant technical and analytic requirements. Another set of techniques capture mRNA from frozen tissue sections using barcoded oligos covalently linked to a glass slide or spatially patterned arrays of mRNA capture beads. For example, Slide-Seq enables spatial sequencing with ~10 micron resolution and is compatible with frozen tissue sections [10]. These techniques provide the benefit of unbiased transcriptional profiling while retaining spatial information and offer exciting possibilities to study granuloma biology.

Targeted Host-Pathogen Profiling

While current forms of mRNA capture beads rely on capture of poly-adenylated mRNA species, terminal modification of mRNA capture beads with capture sequences complimentary to

mRNA target sequences would enable directed hybridization and capture of non-poly-adenylated RNA species. Specifically, modification of bead synthesis strategies to incorporate a universal adapter sequence would enable stoichiometric addition of customized capture sequences using a single hybridization and extension prior to a given experiment [11]. In this framework, target sequences could be complimentary to user-defined genetic constructs (e.g. CRISPR screens), pathogen transcripts, or host mRNAs (e.g. Immune receptors or recurrently mutated proteins). In combination with methods for targeted enrichment and sequential library construction that retain cell barcode and UMI, bead modification could result in significant improvements to single-cell pathogen detection, mutation calling [3, 12] and immune receptor sequencing [2, 11].

Multi-omic profiling of Mtb Granulomas

The field of single-cell genomics has made great progress over the course of my PhD training. Improvements in single-cell sequencing technologies have enabled new levels of biological inquiry and can now be performed in virtually any lab around the world. While methods for generation of high-throughput single-cell mRNA sequencing libraries have largely come of age, new methods for single-cell multi-omics offer exciting possibilities. Numerous methods have been developed that incorporate additional axes of single-cellular information including CRISPR perturbation [13], variant detection [3, 12], immune receptor reconstruction [2], chromatin accessibility and lineage tracing.

Recently developed methods enable simultaneously interrogation of gene expression and chromatin accessibility at single-cell resolution [14]. In particular, these methods would enable interrogation of epigenetic changes in granuloma immune populations that correlate with bacterial burden and the temporal evolution of immune states during infection. In granuloma macrophage populations, this would enable understanding of epigenetic reprogramming of macrophages in sterilizing granulomas. Here, experiments could be designed to understand the epigenetic and

transcriptional basis of trained innate immunity in the context of re-infection and IV-BCG vaccination.

In vitro models of granuloma formation

Finally, there are multiple in vitro models of Mtb granuloma formation that could be used to validate our findings and screen candidate molecules for host-directed therapy. In one of the earliest *in vitro* models of granuloma formation (c. 1925), Mtb can be combined with blood in hanging droplets, which results in the formation of epithelioid macrophages and multi-nucleated giant cells [15]. More recently, Mtb granuloma formation *in vitro* has been achieved through co-culture of immune cells and Mtb in alginate microspheres [16]. Importantly, this model would enable formation of hundreds or thousands of granulomas using immune cells with user-defined input composition. In combination with live-dead reporter strains, this could provide a scalable system to track the effects of cellular composition, host-directed and anti-bacterial therapies on granuloma-level bacterial burden. Finally, lung-on-chip models offer the possibility of studying early dynamics of Mtb infection at the level of the lung epithelium [17].

5.4 Leveraging Single-Cell sequencing results for MTB vaccine design

Rational design of an effective MTB vaccine remains an elusive goal in the prevention of tuberculosis infection. In addition to understanding protective immune responses in the context of natural infection (**Chapter 4**), we have explored the landscape of mucosal immune cells in the setting of BCG (Bacillus Calmette-Guerin) vaccination. Currently, BCG, the only available TB vaccine, is administered intradermally and prevents disseminated TB infection in infants but has variable efficacy in adolescents and adults [18, 19]. Here, we collaborated with Drs. Robert Seder, Mario Roederer and Patricia Darrah at the NIH-NIAID Vaccine Research Institute to understand immune responses associated with multi-route BCG administration.

The central aim of this study was to assess how the route and dose of BCG vaccination influence systemic and tissue-resident T cell immunity, and protection after Mtb challenge [20]. In this study, Rhesus macaques were vaccinated with 5×10^7 colony-forming units (CFUs) of BCG by intradermal (high dose and low dose), aerosol or IV routes, or with a combination of both AE (5×10^7 CFUs) and ID (5×10^5 CFUs; AE/ID). Importantly, IV administration of BCG resulted in significant protection where nine out of ten IV-BCG-immunized NHPs had no lung FDG activity throughout the challenge phase (Fisher's exact test, $P < 0.001$) with six out of ten NHPs having no granulomas throughout the course of infection.

IV BCG vaccination resulted in significant changes in BAL cell numbers with a 5-10- fold increase in total cells made up primarily of conventional T cells. Notably, the altered cellular composition of BAL was sustained up to six months after IV BCG vaccination. To understand the phenotypic identity of BAL T cell responses, we performed single-cell mRNA sequencing with Seq-Well to comprehensively assess transcriptional states among T cells that might underlie protective vaccine responses (**Figure 5.1**).

Single-cell mRNA sequencing was performed in a total of 15 animals across 5 vaccine routes with 3 animals in each group which include unvaccinated controls, low-dose intradermal, high-dose intradermal, aerosol, and intravenous routes of administration. Sequencing was performed on cell obtained from BAL at 2 time points (13 weeks and 25 weeks post-vaccination). In total, we recovered 160,000 cells that include B cells, eosinophils, epithelial cells, macrophages, mast cells, neutrophils, proliferating cells, and T cells (**Figure 5.1A**). Consistent with observations from flow cytometry, we observe increases in the proportion of BAL T cells in the IV-BCG vaccine group (**Figure 5.1B**). Further, we observed persistence of this proportional increase in T cells from week 13 to week 25 post-immunization among animals receiving intravenous vaccine.

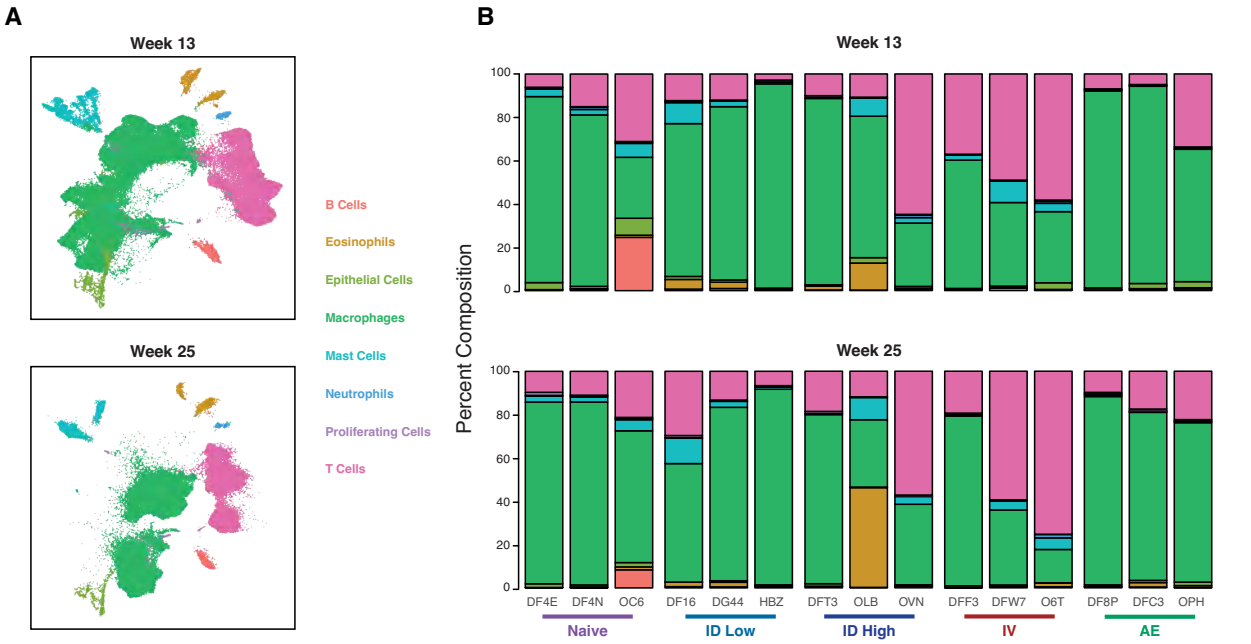


Figure 5.1 | Seq-Well of ~160,000 cells from a multi-route NHP BCG vaccination study. Adapted from Darrah et al. *Nature* 2020. (**A**) UMAPs of cells recovered from Week 13 (top) and Week 25 (bottom)**.** (**B**) Proportion of cell types (colored as in **A**) recovered from BALs by route and by animal at Week 13 (top) and Week 25 (bottom).

We examined correlated patterns of gene expression within unstimulated and PPD-stimulated T cells from BAL to identify groups of genes whose coordinated activity differed by vaccine route (**Figure 5.2A-C**). A total of seven significant T cell modules were identified among in vitro-stimulated T cells 13 weeks after immunization and used to generate expression scores across all T cells at weeks 13 and 25. Among these, we identified a stimulation-inducible module of gene expression, module 2, enriched for memory T cell functionality (**Figure 5.2C**), primarily expressed in a population of BAL CD4 T cells from IV-BCG-immunized NHPs at week 13, and maintained until week 25 (**Figure 5.2B**). Differential gene expression analysis, comparing T cells positive and negative for module 2, showed enrichment of genes previously associated with protection against TB including *IFNG*, *TBX21*, *RORC*, *TNFSF825* and *IL21R* (**Figure 5.2C**) [21].

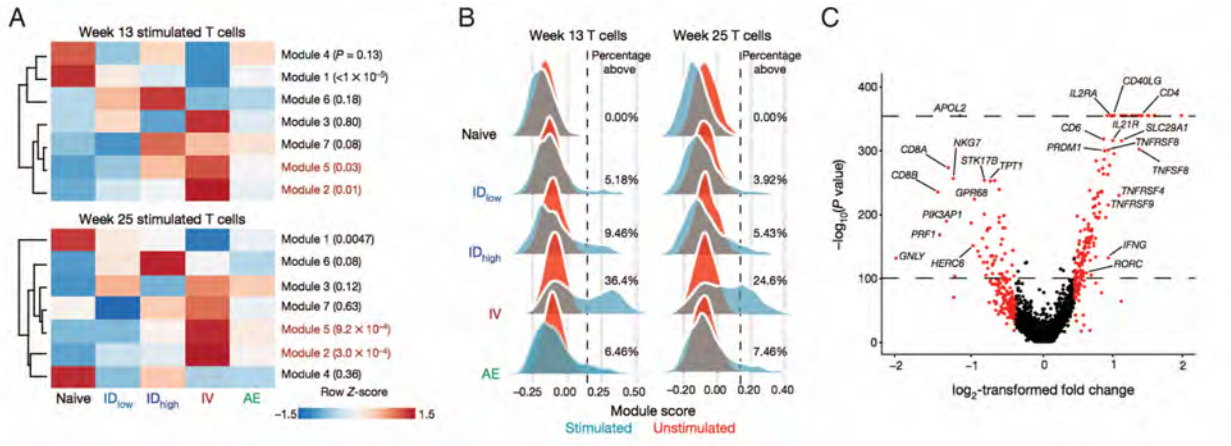


Figure 5.2 | A non-canonical Th1/Th17 hybrid T cell is enriched in IV-BCG NHPs. (**A**) Z-scored heat maps of the average cellular score for modules identified in week 13 PPD-stimulated T cells at weeks 13 and 25 after BCG vaccination. Red P values indicate modules uniquely elevated in the IV BCG group (one-way ANOVA) (**B**) Distributions of module 2 expression in unstimulated and stimulated T cells at weeks 13 and 25 for each group. Percentage module 2-positive is shown; positivity (dashed line) defined as 2 s.d. above the mean score of the unvaccinated (Naive) NHPs (**C**) Volcano plot showing differentially expressed genes between T cells positive and negative for module 2 at week 13 (P values calculated using the likelihood ratio test with Bonferroni correction).

Collectively, findings from natural infection and protective vaccination provide an emerging picture of Mtb-protective immune responses. In the setting of Mtb granulomas, we observe bacterial control to most strongly associate with a population of T1-T17 cells. In the setting of intravenous BCG vaccination, we observe a unique module of correlated gene expression that is enriched for genes associated with Th1 and Th17 effector function. In each case, it appears that in sufficient numbers, T cells that with some combination of Type 1 and Type 17 effector function are capable of providing protection against subsequent mycobacterial challenge. In natural infection, this is best illustrated by observations that late-blooming lesions have significantly lower bacterial burdens presumably since they arise following the onset of adaptive immunity. In the following section, we will discuss the role of Type 1 and Type 17 inflammation in Mtb immunity.

5.5 – The role of Type 1 and Type 17 Inflammation in Mtb Protective Immune Responses

The role of T cells in immune control of Mtb infection has long been appreciated, but the identity and balance of T cell phenotypes that lead to control remains poorly understood. Based

on evidence in the literature and presented in this dissertation, it is most likely that Mtb-protective T cells reside along a spectrum of Type 1 and Type 17 inflammation. While there is a notable lack of consistent nomenclature regarding these hybrid T cell responses in the literature, where descriptions range from Th1, Th17, Th1* [22], non-classical Th1 [23], Th1-Th17, Tc17 [24, 25], ex-Th17 [26], ILC3s [27], or T1-T17 cells as in Chapter 4, these all share many consistent features [28].

For example, a population of Mtb-specific Th1 cells (Th1*) characterized by expression of CCR6 and CXCR3 has been described to co-express IFN-γ and TBET in addition RORγt [22]. Notably, Th1* cells do not produce IL-17 and are increased in patients with latent Mtb infection [29, 30]. Th1-Th17 cells have further been described to result in protective immunity in the setting of both natural infection and vaccination. Emergence of a Th1-Th17 population in the lung plays a central role in protective vaccination and natural immunity against numerous pathogens including *B. pertussis* [31-35], *C. albicans* [36, 37], *H. influenzae* [38], *P. aeruginosa* [39, 40], *S. pneumoniae* [41-43], Leishmaniasis [44], and C. neoformans [45-47]. Further, Th17 cells have been previously shown to play an important role in *M. tuberculosis* infection and vaccine response [48-54].

Recent evidence suggests that a population of ex-Th17 cells are induced following exposure to killed *Klebsiella* and are capable of inducing protective immune responses upon infectious challenge [26]. Here, ex-Th17 cells were observed to lose expression of Ror-γT and IL-17 and upregulate expression of *RBPJ*, *BHLHE40*, *IL-23R* and *IL-7R*. Notably, this pattern bears striking resemblance to that observed among the protective T1-T17 population identified in MTB granulomas described in Chapter 4 (**Figure 4.3 and 4.4**). Collectively, the evidence described in this dissertation and throughout the literature suggests that a combination of Th1 and Th17 effector function consistent with Th1* or ex-Th17 cells is an essential feature of TB-protective immunity.

5.6 – A Balancing Act: Tolerance and Immunity in MTB granulomas

Tissues have evolved tolerance mechanisms to maintain homeostatic function in the face of noxious environmental stimuli, commensal microbiota, and tissue damage [55-58]. These mechanisms serve to limit inflammation and tissue damage and re-direct the host immune system to support organ function and repair. In the context of infection, the evolutionary impulse for organs to safeguard homeostatic function creates an inertia that must be overcome to incite a sterilizing immune response. At the same time, the innate immune system has evolved to display a broad number of pattern recognition receptors specific to components of pathogens (e.g. Toll-like receptors and nucleic acid sensors). Compared to other organisms that cause pulmonary infections, MTB is relatively devoid of molecular patterns that trigger a robust innate immune response (e.g. LPS) and the initial infectious dose of Mtb is generally very low. Moreover, virulent strains of MTB are able to directly suppress TLR signaling [59]. In the view, MTB is the most successful pathogen due in part to its initial immunologic insignificance.

Sequestration of persistent antigens and limiting surrounding tissue damage are central functions of granuloma formation. In this view, granulomas represent an evolutionary détente in which persistent antigens are tolerated but sequestered to prevent end-organ failure due to an overexuberant host immune response. For example, Type 2 polarized granulomatous inflammation in *S. mansoni* infection highlights a key evolutionary function to limit inflammation, sequester parasite eggs and ultimately preserve tissue function [60]. Notably, with impaired Type 2 immunity, unchecked immunity activation leads to increased immunopathology and mortality. Specifically, in *S. mansoni* infection, deletion of IL-13RA2 results in an overall reduction in the number of viable pathogen eggs, but improved pathogen control is accompanied by increased immunopathology and reduced overall survival [61].

In the earliest stages of infection (0-4 weeks post-infection), I propose that MTB “exploits” tissue-level signals (i.e. the “keep-breathing” instinct of the lung) to initiate granulomatous environments that are tolerant to persistent bacterial replication. In the absence of over-riding

alarm bells that accompany canonical pulmonary pathogens, Mtb is able to infect macrophages and initiate granuloma formation. However, in the absence of strong over-riding, PAMP signaling, tissue-level signals designed to maintain tissue function predominate, and early polarization towards wound-healing and pro-fibrotic signaling enable the generation of immune-privileged environments amenable to long-term pathogen viability. Largely secluded from the "outside world", these granulomas allow the slowly growing pathogen a safe haven to outlast the robust immune response that ensues, enabling resurgence of infection upon eventual changes in host fitness.

During later stages of infection (4-8 weeks post-infection), granulomas form in a distinct immune environment. While it is possible that alveolar macrophages and bone-marrow derived monocytes have adopted a more responsive phenotype following exposure to acute products of inflammation. The largest difference, however, is the emergence of an adaptive immune response. In the single-cell sequencing data from natural infection (Chapter 4), there are hints that such a mechanism might explain differences in granuloma-level bacterial control. In late-blooming, low-burden lesions, we observe increases in the number of relative composition of T cells (**Figure 4.8**). Specifically, we observe expansion of a population of T1-T17 cells that is highly associated with bacterial control. Further, in early-forming lesions alone, we observe lower numbers of this population and do not detect a significant relationship between bacterial burden (**Figure 4.9**).

5.7 Future Investigation

Designing Effective Host-directed Therapies in Established MTB infection

Based on observation from studies of natural infection (Chapter 4) and protective vaccination (Chapter 5), robust T-cell responses are capable of providing sterilizing immunity against Mtb. Intravenous BCG vaccination results in widespread dissemination of BCG throughout lymphoid (e.g. lymph nodes and spleen) and non-lymphoid tissues leading. Studies in

non-human primates reveal that local immune control is possible at the level of granulomas, even in the context of global loss of control [62]. For example, a single animal reveals robust elimination of bacteria among late-blooming lesions despite global loss of bacterial control, as evidence by elevated total lung FDG (**Figure 4.1**). While these data hold promise for vaccine development to prevent the acquisition of Mtb infection in unexposed individuals, the larger challenge is immunologic reprogramming within established granulomas that support persistent bacterial replication using host-directed therapy [63].

Unless eliminated, bacteria within persistent lesions can lie dormant for a period of months, years, or decades until changes in host fitness enable reactivation of latent infection. It is likely that protective, antigen-specific T cells are present throughout the lung following the onset of adaptive immunity, but are not able to eliminate Mtb from persistently high-burden lesions, which may critically limit the efficacy of T-cell targeted host-directed therapy. From an ecologic perspective, there is likely a carrying capacity for the number of T cells that can reside in granulomas, where the accumulation of immunologically ineffective T cells likely prevents the subsequent recruitment of protective antigen-specific T cells in early-forming lesions [64, 65]. Further, fibrosis represents a physical barrier that might limit access of protective T cells to the granuloma microenvironment among early forming lesions.

Importantly, IL-13 represents a key cytokine associated with fibrotic changes in multiple tissues [66-69]. Blockade of IL-13 is associated with reduction in hepatic fibrosis in the context of *S. mansoni* infection [70] and reduction in pulmonary fibrosis [71]. Further, lung injury results in IL-33 dependent expansion of ST2+ Tregs that in turn produce IL-13 leading to alternative activation of macrophages and tissue repair [72]. IL-13 also signaling represents a promising target for host-directed therapy based on scRNA-seq data from MTB granulomas. In the context of 10-week NHP granulomas, we observe increased mast cell IL-13 expression in high burden lesions with increased receptor-ligand potentials acting on multiple T cell populations in high-burden, early-forming lesions (**Figure 4.11D**). Moreover, IL-13 is a type 2 cytokine that limits the

ability of macrophages to control intracellular Mtb infection [73]. Importantly, IL-13 targeting antibodies are already used clinically to treat a range of Type 2 inflammatory conditions including atopic dermatitis and chronic rhinosinusitis [74-76]. Collectively, these observations make IL-13 a promising target for host-directed therapy in established Mtb infection.

Understanding the role of tissue-level sensitization in determining MTB infection outcome

Individuals exposed to MTB have highly divergent responses: some individuals develop fulminant disease while others are able to control and even eliminate infection [77]. While we have begun to define the features of protective immune responses at the level of individual granulomas, we still lack a comprehensive understanding of the factors that predict host-level control of Mtb infection between individuals.

Epidemiologic studies of Mtb infection have uncovered groups of individuals who are able to persistently resist infection despite prolonged exposure to Mtb [78, 79]. The first evidence that some individuals are able to avoid infection despite prolonged exposure emerged from the study of nurses at Boston City Hospital [80] and a study of 1,175 sailors aboard a ship in close quarters with an index TB patient [81]. More recent cohort studies include household contacts of Mtb index cases in Kampala, Uganda and a cohort of South African gold miners with estimated annual risk of Mtb infection of 20% [78]. These studies have reveled subsets of individuals termed “resisters” who remain persistently IGRA-negative, meaning they lack peripheral T cells that produce IFN-γ in response to Mtb antigens, despite high levels of Mtb exposure. Recent evidence suggests that resisters (IGRA-negative, exposed individuals) have distinct patterns of adaptive immune responses characterized by elevated T cell expression of CD40LG, increased antibody avidity and unique patterns of Fc glycosylation [82].

Genome-wide association studies (GWAS) have been performed to identify genetic risk factors associated with the risk of Mtb acquisition [83-87]. While these studies have identified multiple genetic susceptibility loci including 18q11.2, JAG1, CCL17, and ASAP1, there has often

been limited replication across multi-ethnic cohorts [88]. The lack of coherent genetic predisposition highlights the importance of understanding how variation in host-levels factors influences response to Mtb infection. Here, key unanswered questions include how allergic sensitization, prior/co-incident infection, and nutritional status influence subsequent host-response to Mtb infection.

Allergy is associated with Type 2 polarized immune responses and can lead to tissue remodeling within respiratory epithelium [89, 90] that might alter the balance of Mtb tolerance and immunity in individuals with pre-existing allergic conditions. Epidemiologic studies have shown an inverse relationship between tuberculosis infection and development of asthma, where individuals with positive PPDs develop asthma less frequently [91, 92]. Further, some studies report that TB infection protects against atopic rhinitis, while others observe no association [93, 94]. Adding to this complexity, childhood BCG vaccination represents a potential confounding variable as it may reduce the likelihood of atopy [95]. However, a recent study of Mtb and allergy in children, revealed a lack of association between BCG vaccination and reduction of allergy along with a strong, positive association between allergic symptoms (eczema and asthma) and Mtb infection [96]. Epidemiologic studies require careful interpretation, particularly cross-sectional studies, as the temporal relationship between allergy and infection is difficult to establish. Ultimately, interventional studies are needed to address the causal nature of allergy in host response to Mtb. Interestingly, evidence from mice suggests that exposure to dust mite allergen prior to infection impairs host immunity to Mtb [97], but further studies are necessary to more fully elucidate the role of allergic sensitization in Mtb risk.

Helminth infection is associated with increased expression of type 2 cytokines including IL-4, IL-5 IL-9, IL-10, and IL-13 and mucosal expansion of mast cells, eosinophils and basophils [98]. There is widespread helminth infection throughout TB-endemic regions of the world, and it is possible that pre-existing helminth infection could modulate host-response to Mtb infection. In mice, prior infection with *N. brasiliensis* results in impaired control of Mtb infection in an IL-4

dependent manner [99]. Further, infection with *S. mansoni* attenuates protection from BCG vaccination in mice [100]. In humans, helminth co-infection modulates immune control of Mtb infection and response to anti-mycobacterial therapy [101, 102].

Finally, impaired host nutrition and low body mass index (LBMI) are risk factors for acquisition and reactivation of Mtb infection [103]. Protein calorie restriction results in rapid decline of Mtb-infected mice [104]. Interestingly, malnourished individuals with latent Mtb infection (LTBI) show elevated levels of IL-5 and IL-13 and reductions in numerous pro-inflammatory cytokines including IFN-gamma, TNF-alpha, IL-1beta, and IL-22 relative individuals with LTBI and normal BMIs [105]. Collectively, an improved understanding of potentially modifiable host factors that influence response to Mtb infection will be a critical factor in determining the effectiveness of prophylactic and therapeutic interventions to prevent and treat Mtb infection in humans.

References

1. Gierahn, T.M., et al., *Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput.* Nature methods, 2017. **14**(4): p. 395-398.

2. Tu, A.A., et al., *TCR sequencing paired with massively parallel 3′ RNA-seq reveals clonotypic T cell signatures.* Nature immunology, 2019. **20**(12): p. 1692-1699.

3. van Galen, P., et al., *Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity.* Cell, 2019. **176**(6): p. 1265-1281. e24.

4. Cadena, A.M., et al., *Concurrent infection with Mycobacterium tuberculosis confers robust protection against secondary infection in macaques.* PLoS pathogens, 2018. **14**(10): p. e1007305.

5. Angelo, M., et al., *Multiplexed ion beam imaging of human breast tumors.* Nature medicine, 2014. **20**(4): p. 436.

6. Lee, J.H., et al., *Highly multiplexed subcellular RNA sequencing in situ.* Science, 2014. **343**(6177): p. 1360-1363.

7. Lee, J.H., et al., *Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues.* Nature protocols, 2015. **10**(3): p. 442.

8. Chen, K.H., et al., *Spatially resolved, highly multiplexed RNA profiling in single cells.* Science, 2015. **348**(6233): p. aaa6090.

9. Wang, X., et al., *Three-dimensional intact-tissue sequencing of single-cell transcriptional states.* Science, 2018. **361**(6400): p. eaat5691.

10. Rodriques, S.G., et al., *Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution.* Science, 2019. **363**(6434): p. 1463-1467.

11. Hanson, W.M., et al., *Reversible oligonucleotide chain blocking enables bead capture and amplification of T-cell receptor α and β chain mRNAs.* Journal of the American Chemical Society, 2016. **138**(35): p. 11073-11076.

12. Nam, A.S., et al., *Somatic mutations and cell identity linked by Genotyping of Transcriptomes.* Nature, 2019. **571**(7765): p. 355-360.

13. Dixit, A., et al., *Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens.* Cell, 2016. **167**(7): p. 1853-1866. e17.

14. Cao, J., et al., *Joint profiling of chromatin accessibility and gene expression in thousands of single cells.* Science, 2018. **361**(6409): p. 1380-1385.

15. Lewis, M.R., *The formation of macrophages, epithelioid cells and giant cells from leucocytes in incubated blood.* The American journal of pathology, 1925. **1**(1): p. 91.

16. Tezera, L.B., et al., *Dissection of the host-pathogen interaction in human tuberculosis using a bioengineered 3-dimensional model.* Elife, 2017. **6**: p. e21283.

17. Thacker, V.V., et al., *A lung-on-chip infection model reveals protective and permissive roles of alveolar epithelial cells in tuberculosis.* bioRxiv, 2020.

18. Mangtani, P., et al., *Protection by BCG vaccine against tuberculosis: a systematic review of randomized controlled trials.* Clinical infectious diseases, 2014. **58**(4): p. 470-480.

19. Harris, R.C., et al., *Systematic review of mathematical models exploring the epidemiological impact of future TB vaccines.* Human vaccines & immunotherapeutics, 2016. **12**(11): p. 2813-2832.

20. Darrah, P.A., et al., *Prevention of tuberculosis in macaques after intravenous BCG immunization.* Nature, 2020. **577**(7788): p. 95-102.

21. Booty, M.G., et al., *IL-21 signaling is essential for optimal host resistance against Mycobacterium tuberculosis infection.* Scientific reports, 2016. **6**(1): p. 1-13.

22. Acosta-Rodriguez, E.V., et al., *Surface phenotype and antigenic specificity of human interleukin 17–producing T helper memory cells.* Nature immunology, 2007. **8**(6): p. 639-646.

23. Basdeo, S.A., et al., *Ex-Th17 (nonclassical Th1) cells are functionally distinct from classical Th1 and Th17 cells and are not constrained by regulatory T cells.* The Journal of Immunology, 2017. **198**(6): p. 2249-2259.

24. Nanjappa, S.G., et al., *Tc17 cells mediate vaccine immunity against lethal fungal pneumonia in immune deficient hosts lacking CD4+ T cells.* PLoS pathogens, 2012. **8**(7).

25. Hamada, H., et al., *Tc17, a unique subset of CD8 T cells that can protect against lethal influenza challenge.* The Journal of Immunology, 2009. **182**(6): p. 3469-3481.

26. Vesely, M.C.A., et al., *Effector TH17 Cells Give Rise to Long-Lived TRM Cells that Are Essential for an Immediate Response against Bacterial Infection.* Cell, 2019. **178**(5): p. 1176-1188. e15.

27. Ardain, A., et al., *Group 3 innate lymphoid cells mediate early protective immunity against tuberculosis.* Nature, 2019. **570**(7762): p. 528-532.

28. Sallusto, F., *Heterogeneity of human CD4+ T cells against microbes.* Annual review of immunology, 2016. **34**: p. 317-334.

29. Arlehamn, C.S.L., et al., *Memory T cells in latent Mycobacterium tuberculosis infection are directed against three antigenic islands and largely contained in a CXCR3+ CCR6+ Th1 subset.* PLoS pathogens, 2013. **9**(1).

30. Arlehamn, C.L., et al., *Transcriptional profile of tuberculosis antigen–specific T cells reveals novel multifunctional features.* The Journal of Immunology, 2014. **193**(6): p. 2931-2940.

31. Warfel, J. and T. Merkel, *Bordetella pertussis infection induces a mucosal IL-17 response and long-lived Th17 and Th1 immune memory cells in nonhuman primates.* Mucosal immunology, 2013. **6**(4): p. 787-796.

32. Ross, P.J., et al., *Relative contribution of Th1 and Th17 cells in adaptive immunity to Bordetella pertussis: towards the rational design of an improved acellular pertussis vaccine.* PLoS pathogens, 2013. **9**(4).

33. Fedele, G., et al., *Attenuated Bordetella pertussis vaccine candidate BPZE1 promotes human dendritic cell CCL21-induced migration and drives a Th1/Th17 response.* The Journal of Immunology, 2011. **186**(9): p. 5388-5396.

34. Schnoeller, C., et al., *Attenuated Bordetella pertussis vaccine protects against respiratory syncytial virus disease via an IL-17–dependent mechanism.* American journal of respiratory and critical care medicine, 2014. **189**(2): p. 194-202.

35. Higgins, S.C., et al., *TLR4 mediates vaccine-induced protective cellular immunity to Bordetella pertussis: role of IL-17-producing T cells.* The Journal of Immunology, 2006. **177**(11): p. 7980-7989.

36. Pandiyan, P., et al., *CD4+ CD25+ Foxp3+ regulatory T cells promote Th17 cells in vitro and enhance host resistance in mouse Candida albicans Th17 cell infection model.* Immunity, 2011. **34**(3): p. 422-434.

37. Lin, L., et al., *Th1-Th17 cells mediate protective adaptive immunity against Staphylococcus aureus and Candida albicans infection in mice.* PLoS pathogens, 2009. **5**(12).

38. Li, W., et al., *Recognition of conserved antigens by Th17 cells provides broad protection against pulmonary Haemophilus influenzae infection.* Proceedings of the National Academy of Sciences, 2018. **115**(30): p. E7149-E7157.

39. Wu, W., et al., *Th17-stimulating protein vaccines confer protection against Pseudomonas aeruginosa pneumonia.* American journal of respiratory and critical care medicine, 2012. **186**(5): p. 420-427.

40. Liu, J., et al., *Early production of IL-17 protects against acute pulmonary Pseudomonas aeruginosa infection in mice.* FEMS Immunology & Medical Microbiology, 2011. **61**(2): p. 179-188.

41. Wright, A.K., et al., *Experimental human pneumococcal carriage augments IL-17A-dependent T-cell defence of the lung.* PLoS pathogens, 2013. **9**(3).

42. Moffitt, K.L., et al., *TH17-based vaccine design for prevention of Streptococcus pneumoniae colonization.* Cell host & microbe, 2011. **9**(2): p. 158-165.

43. Moffitt, K.L., R. Malley, and Y.-J. Lu, *Identification of protective pneumococcal TH17 antigens from the soluble fraction of a killed whole cell vaccine.* PloS one, 2012. **7**(8).

44. Wu, W., L. Huang, and S. Mendez, *A live Leishmania major vaccine containing CpG motifs induces the de novo generation of Th17 cells in C57BL/6 mice.* European journal of immunology, 2010. **40**(9): p. 2517-2527.

45. Zhang, Y., et al., *Robust Th1 and Th17 immunity supports pulmonary clearance but cannot prevent systemic dissemination of highly virulent Cryptococcus neoformans H99.* The American journal of pathology, 2009. **175**(6): p. 2489-2500.

46. Müller, U., et al., *IL-13 induces disease-promoting type 2 cytokines, alternatively activated macrophages and allergic inflammation during pulmonary infection of mice with Cryptococcus neoformans.* The Journal of Immunology, 2007. **179**(8): p. 5367-5377.

47. Murdock, B.J., et al., *Early or late IL-10 blockade enhances Th1 and Th17 effector responses and promotes fungal clearance in mice with cryptococcal lung infection.* The Journal of Immunology, 2014. **193**(8): p. 4107-4116.

48. Scriba, T.J., et al., *Distinct, specific IL-17-and IL-22-producing CD4+ T cell subsets contribute to the human anti-mycobacterial immune response.* The Journal of Immunology, 2008. **180**(3): p. 1962-1970.

49. Van Dis, E., et al., *STING-activating adjuvants elicit a Th17 immune response and protect against Mycobacterium tuberculosis infection.* Cell reports, 2018. **23**(5): p. 1435-1447.

50. Khader, S.A., et al., *IL-23 and IL-17 in the establishment of protective pulmonary CD4+ T cell responses after vaccination and during Mycobacterium tuberculosis challenge.* Nature immunology, 2007. **8**(4): p. 369-377.

51. Khader, S.A., et al., *IL-23 compensates for the absence of IL-12p70 and is essential for the IL-17 response during tuberculosis but is dispensable for protection and antigen-specific IFN-γ responses if IL-12p70 is available.* The Journal of Immunology, 2005. **175**(2): p. 788-795.

52. Gopal, R., et al., *Unexpected role for IL-17 in protective immunity against hypervirulent Mycobacterium tuberculosis HN878 infection.* PLoS pathogens, 2014. **10**(5).

53. Wozniak, T.M., et al., *Mycobacterium bovis BCG-specific Th17 cells confer partial protection against Mycobacterium tuberculosis infection in the absence of gamma interferon.* Infection and immunity, 2010. **78**(10): p. 4187-4194.

54. da Costa, A.C., et al., *A new recombinant BCG vaccine induces specific Th17 and Th1 effector cells with higher protective efficacy against tuberculosis.* PLoS One, 2014. **9**(11).

55. Okabe, Y. and R. Medzhitov, *Tissue biology perspective on macrophages.* Nature immunology, 2016. **17**(1): p. 9.

56. Ayres, J.S. and D.S. Schneider, *Tolerance of infections.* Annual review of immunology, 2012. **30**: p. 271-294.

57. Schneider, D.S. and J.S. Ayres, *Two ways to survive infection: what resistance and tolerance can teach us about treating infectious diseases.* Nature Reviews Immunology, 2008. **8**(11): p. 889-895.

58. Medzhitov, R., D.S. Schneider, and M.P. Soares, *Disease tolerance as a defense strategy.* Science, 2012. **335**(6071): p. 936-941.

59. Pathak, S.K., et al., *Direct extracellular interaction between the early secreted antigen ESAT-6 of Mycobacterium tuberculosis and TLR2 inhibits TLR signaling in macrophages.* Nature immunology, 2007. **8**(6): p. 610-618.

60. Chiaramonte, M.G., et al., *IL-13 is a key regulatory cytokine for Th2 cell-mediated pulmonary granuloma formation and IgE responses induced by Schistosoma mansoni eggs.* The Journal of Immunology, 1999. **162**(2): p. 920-930.

61. Mentink-Kane, M.M., et al., *IL-13 receptor α 2 down-modulates granulomatous inflammation and prolongs host survival in schistosomiasis.* Proceedings of the National Academy of Sciences, 2004. **101**(2): p. 586-590.

62. Lin, P.L., et al., *Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing.* Nature medicine, 2014. **20**(1): p. 75.

63. Wallis, R.S. and R. Hafner, *Advancing host-directed therapy for tuberculosis.* Nature Reviews Immunology, 2015. **15**(4): p. 255-263.

64. Egen, J.G., et al., *Intravital imaging reveals limited antigen presentation and T cell effector function in mycobacterial granulomas.* Immunity, 2011. **34**(5): p. 807-819.

65. Egen, J.G., et al., *Macrophage and T cell dynamics during the development and disintegration of mycobacterial granulomas.* Immunity, 2008. **28**(2): p. 271-284.

66. Fichtner-Feigl, S., et al., *IL-13 signaling through the IL-13α 2 receptor is involved in induction of TGF-β 1 production and fibrosis.* Nature medicine, 2006. **12**(1): p. 99-106.

67. Kaviratne, M., et al., *IL-13 activates a mechanism of tissue fibrosis that is completely TGF-β independent.* The Journal of Immunology, 2004. **173**(6): p. 4020-4029.

68. Reiman, R.M., et al., *Interleukin-5 (IL-5) augments the progression of liver fibrosis by regulating IL-13 activity.* Infection and immunity, 2006. **74**(3): p. 1471-1479.

69. Gieseck III, R.L., et al., *Interleukin-13 activates distinct cellular pathways leading to ductular reaction, steatosis, and fibrosis.* Immunity, 2016. **45**(1): p. 145-158.

70. Chiaramonte, M.G., et al., *An IL-13 inhibitor blocks the development of hepatic fibrosis during a T-helper type 2–dominated inflammatory response.* The Journal of clinical investigation, 1999. **104**(6): p. 777-785.

71. Jakubzick, C., et al., *Therapeutic attenuation of pulmonary fibrosis via targeting of IL-4- and IL-13-responsive cells.* The Journal of Immunology, 2003. **171**(5): p. 2684-2693.

72. Liu, Q., et al., *IL-33–mediated IL-13 secretion by ST2+ Tregs controls inflammation after lung injury.* JCI insight, 2019. **4**(6).

73. Harris, J., et al., *T helper 2 cytokines inhibit autophagic control of intracellular Mycobacterium tuberculosis.* Immunity, 2007. **27**(3): p. 505-517.

74. Yang, G., et al., *Anti-IL-13 monoclonal antibody inhibits airway hyperresponsiveness, inflammation and airway remodeling.* Cytokine, 2004. **28**(6): p. 224-232.

75. Simpson, E.L., et al., *Two phase 3 trials of dupilumab versus placebo in atopic dermatitis.* New England Journal of Medicine, 2016. **375**(24): p. 2335-2348.

76. Bachert, C., et al., *Effect of subcutaneous dupilumab on nasal polyp burden in patients with chronic sinusitis and nasal polyposis: a randomized clinical trial.* Jama, 2016. **315**(5): p. 469-479.

77. Simmons, J.D., et al., *Immunological mechanisms of human resistance to persistent Mycobacterium tuberculosis infection.* Nature Reviews Immunology, 2018. **18**(9): p. 575-589.

78. Wallis, R.S., *Mathematical models of tuberculosis reactivation and relapse.* Frontiers in microbiology, 2016. **7**: p. 669.

79. Whalen, C.C., et al., *Secondary attack rate of tuberculosis in urban households in Kampala, Uganda.* PloS one, 2011. **6**(2).

80. Badger, T.L. and W.W. Spink, *First-infection type of tuberculosis in adults: a five-year study of student nurses at the Boston city hospital.* New England Journal of Medicine, 1937. **217**(11): p. 424-431.

81. Houk, V., et al., *The epidemiology of tuberculosis infection in a closed environment.* Archives of Environmental Health: An International Journal, 1968. **16**(1): p. 26-35.

82. Lu, L.L., et al., *IFN-γ-independent immune markers of Mycobacterium tuberculosis exposure.* Nature medicine, 2019. **25**(6): p. 977-987.

83. Mahasirimongkol, S., et al., *Genome-wide association studies of tuberculosis in Asians identify distinct at-risk locus for young tuberculosis.* Journal of human genetics, 2012. **57**(6): p. 363-367.

84. Uren, C., et al., *A post-GWAS analysis of predicted regulatory variants and tuberculosis susceptibility.* PloS one, 2017. **12**(4).

85. Thye, T., et al., *Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11. 2.* Nature genetics, 2010. **42**(9): p. 739-741.

86. Png, E., et al., *A genome wide association study of pulmonary tuberculosis susceptibility in Indonesians.* BMC medical genetics, 2012. **13**(1): p. 5.

87. Curtis, J., et al., *Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration.* Nature genetics, 2015. **47**(5): p. 523-527.

88. Miao, R., et al., *Genetic variants at 18q11. 2 and 8q24 identified by genome-wide association studies were not associated with pulmonary tuberculosis risk in Chinese population.* Infection, Genetics and Evolution, 2016. **40**: p. 214-218.

89. Ordovas-Montanes, J., et al., *Allergic inflammatory memory in human respiratory epithelial progenitor cells.* Nature, 2018. **560**(7720): p. 649-654.

90. Van Crombruggen, K., et al., *Pathogenesis of chronic rhinosinusitis: inflammation.* Journal of Allergy and Clinical Immunology, 2011. **128**(4): p. 728-732.

91. von Hertzen, L., et al., *Mycobacterium tuberculosis infection and the subsequent development of asthma and allergic conditions.* Journal of allergy and clinical immunology, 1999. **104**(6): p. 1211-1214.

92. Wong, G., et al., *Asthma, atopy and tuberculin responses in Chinese schoolchildren in Hong Kong.* Thorax, 2001. **56**(10): p. 770-773.

93. Soysal, A., et al., *Lack of an inverse association between tuberculosis infection and atopy: By T-cell-based immune assay (RD1–ELISpot).* Pediatric allergy and immunology, 2008. **19**(8): p. 709-715.

94. Obihara, C.C., et al., *Inverse association between Mycobacterium tuberculosis infection and atopic rhinitis in children.* Allergy, 2005. **60**(9): p. 1121-1125.

95. Shirakawa, T., et al., *The inverse association between tuberculin responses and atopic disorder.* Science, 1997. **275**(5296): p. 77-79.

96. Flohr, C., et al., *Tuberculosis, bacillus Calmette–Guerin vaccination, and allergic disease: findings from the International Study of Asthma and Allergies in Childhood Phase Two.* Pediatric allergy and immunology, 2012. **23**(4): p. 324-331.

97. Piñeros, A., et al., *M2 macrophages or IL-33 treatment attenuate ongoing Mycobacterium tuberculosis infection.* Scientific reports, 2017. **7**(1): p. 1-12.

98. Allen, J.E. and R.M. Maizels, *Diversity and dialogue in immunity to helminths.* Nature Reviews Immunology, 2011. **11**(6): p. 375-388.

99. Potian, J.A., et al., *Preexisting helminth infection induces inhibition of innate pulmonary anti-tuberculosis defense by engaging the IL-4 receptor pathway.* Journal of Experimental Medicine, 2011. **208**(9): p. 1863-1874.

100. Elias, D., et al., *Schistosoma mansoni infection reduces the protective efficacy of BCG vaccination against virulent Mycobacterium tuberculosis.* Vaccine, 2005. **23**(11): p. 1326-1334.

101. George, P.J., et al., *Helminth infections coincident with active pulmonary tuberculosis inhibit mono-and multifunctional CD4+ and CD8+ T cell responses in a process dependent on IL-10.* PLoS pathogens, 2014. **10**(9).

102. Babu, S. and T.B. Nutman, *Helminth-tuberculosis co-infection: an immunologic perspective.* Trends in immunology, 2016. **37**(9): p. 597-607.

103. McMurray, D.N., et al., *Micronutrient Status and Immune Function in Tuberculosis a.* Annals of the New York Academy of Sciences, 1990. **587**(1): p. 59-69.

104. Chan, J., et al., *Effects of protein calorie malnutrition on tuberculosis in mice.* Proceedings of the National Academy of Sciences, 1996. **93**(25): p. 14857-14861.

105. Anuradha, R., et al., *Coexistent malnutrition is associated with perturbations in systemic and antigen-specific cytokine responses in latent tuberculosis infection.* Clin. Vaccine Immunol., 2016. **23**(4): p. 339-345.

Appendix 1

Corresponds to Chapter 2.

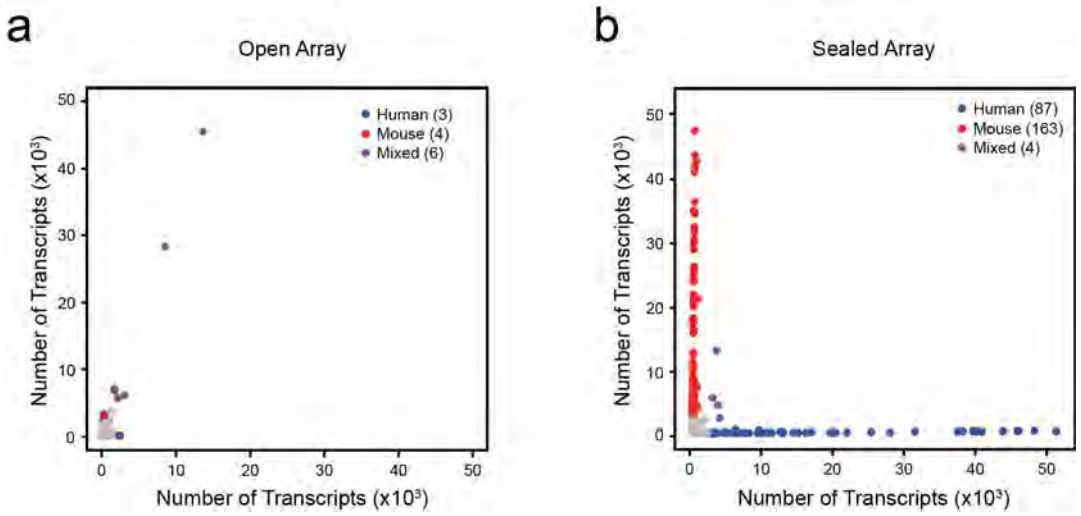


Figure A1.1. (A) An open array format results in decreased gene and transcript capture, and increased cross-contamination, relative to the membrane sealing implemented in Seq-Well. **(B)** Species mixing experiments with reversible membrane sealing using Seq-Well provides increased gene/transcript capture and improved single-cell resolution.

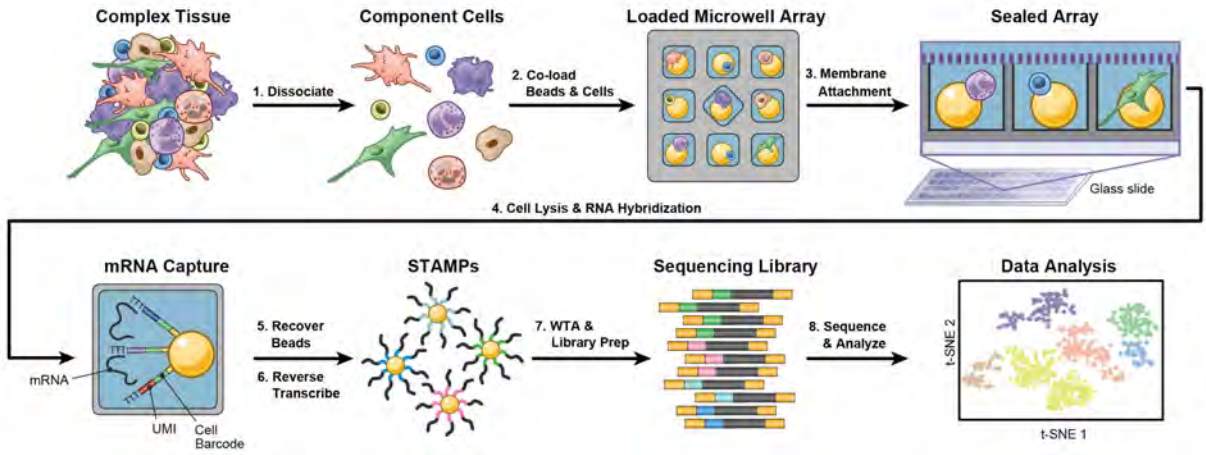
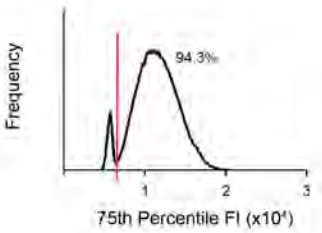


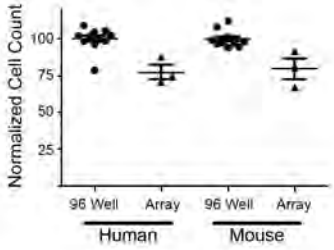
Figure A1.2. Cells are obtained from complex tissues or clinical biopsies, and digested to form a single-cell suspension. Barcoded mRNA capture beads are added to the surface of the microwell device, settling into wells by gravity, and then a single-cell suspension is applied**.** The device is sealed using a semi-permeable membrane that, upon addition of a chemical lysis buffer, confines cellular mRNAs within wells while allowing efficient buffer exchange. Liberated cellular transcripts hybridize to the bead-bound barcoded poly(dT) primers that contain a cell barcode (shared by all probes on the same bead but different between beads) and a unique molecular identifier (UMI) for each transcript molecule. After hybridization, the beads are removed from the array and bulk reverse transcription is performed to generate single-cell cDNAs attached to beads. Libraries are then made by a combination of PCR and tagmentation, and sequenced. After, Single-cell transcriptomes are assembled *in silico* using cell barcodes and UMIs.

a



Beads/well	0	1	>1
Array 1	5.0%	94.5%	0.5%
Array 2	5.7%	93.9%	0.4%

b



c

Loading Density	Empty	1 Human	1 Mouse	2+ Human	2+ Mouse	Human & Mouse Multiplet	Total Multiplet	Human & Mouse Multiplet Freq. (%)	Total Multiplet Freq. (%)
200	83,200 ± 293	134 ± 23	177 ± 53	0.33 ± 0.6	10.3 ± 17.9	0.7 ± 0.6	11.3 ± 19	0.1 ± 0.2	2.7 ± 4.4
2,000	81,100 ± 320	1,310 ± 150	1,080 ± 170	9.7 ± 3.2	23.3 ± 13	10.3 ± 3.1	43.3 ± 11	0.4 ± 0.1	1.8 ± 0.5
20,000	60,700 ± 560	13,100 ± 340	9,190 ± 250	976 ± 100	598 ± 23	1,100 ± 160	2,603 ± 250	4.8 ± 0.6	11.4 ± 0.9

Figure A1.3. (A) Two arrays were loaded with barcoded beads through intermittent rocking. After washing, arrays were imaged in transmitted light and AF488 channel to capture bead autofluorescence. A plot of the frequency of the 75th percentile AF488 well intensity across the array (Panel 1) and the frequency of wells containing zero, one and multiple beads is displayed (Panel 2). **(B)** 200 □L of a 1:1 mix of fluorescently labeled human (HEK 293) and mouse (3T3) cell solution was loaded into 3 arrays and 12 wells of a 96 well plate. The number of cells loaded into each array and well as enumerated by fluorescent imaging is plotted, normalized to the average number of cells/well in the 96 well plate. Mean and standard error are denoted by line and error bars respectively. **(C)** 2x10^2, 2x10^3, and 2x10^4 total cells of a 1:1 mix of fluorescently labeled HEK 293T and 3T3 cells were loaded onto three functionalized arrays each. All arrays were fluorescently imaged to enumerate the number of each cell line in each array microwell. The mean ± standard deviation of the number of empty, single and multiple occupancy wells across the three replicate arrays for each loading density is displayed along with the mean ± standard deviation of the percentage of occupied wells containing a cell from each species.

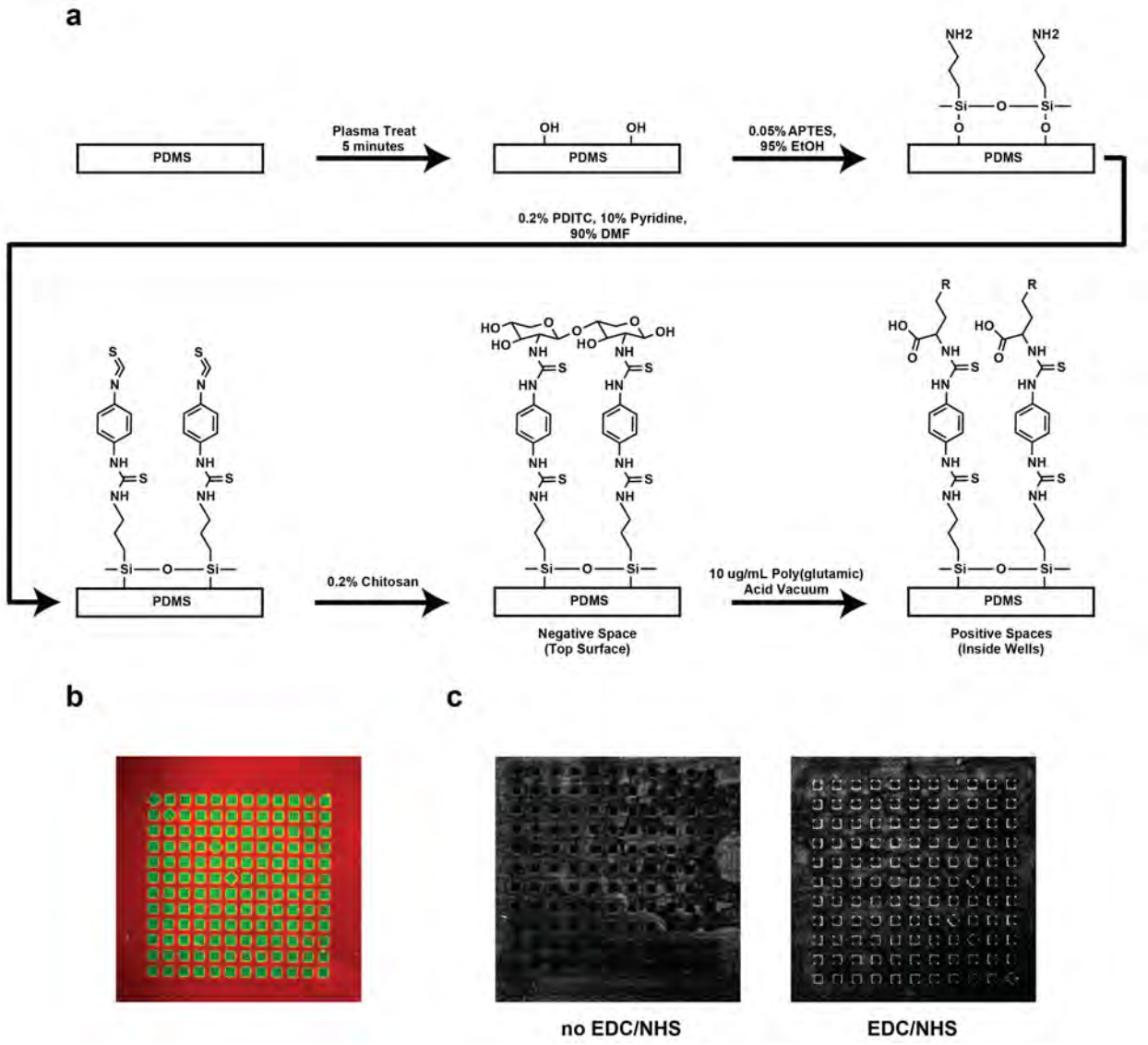


Figure A1.4. (A) The surface of the PDMS device is initially treated with an air plasma under mild vacuum, terminating the surface in hydroxyls. This PDMS surface is aminated using (3-Aminopropyl)triethoxysilane (APTES). The amine surface is then activated with PDITC to create an isothiocyanate surface. The isothiocyanate on the top surface of the array (negative space) is covalently linked to chitosan polymers through their amine group. The hydrophobicity of the isothiocyanate surface prevents solvation of the microwells with the aqueous chitosan solution, preventing chitosan from reacting with the inner well surfaces (positive space). These surfaces are subsequently reacted with the free amine of poly(glutamic) acid polymers under vacuum to drive the solvation of the wells. **(B)** The top surface of a PDITC-activated array was coated with streptavidin-PE (red) and the inner well surfaces were coated with streptavidin-AF488 (green) using same method used to functionalize with chitosan and poly(glutamate). **(C)** Two chitosan/poly(glutamate) bi-functionalized arrays were submerged in MES buffer without (Panel 1) or with (Panel 2) 100 µg/mL EDC and 10 µg/mL NHS for 10 minutes. The arrays were washed and then submerged in PBS solution containing 1 µg/mL AF568-labeled antibody overnight. After washing, arrays were imaged for AF568 fluorescence.

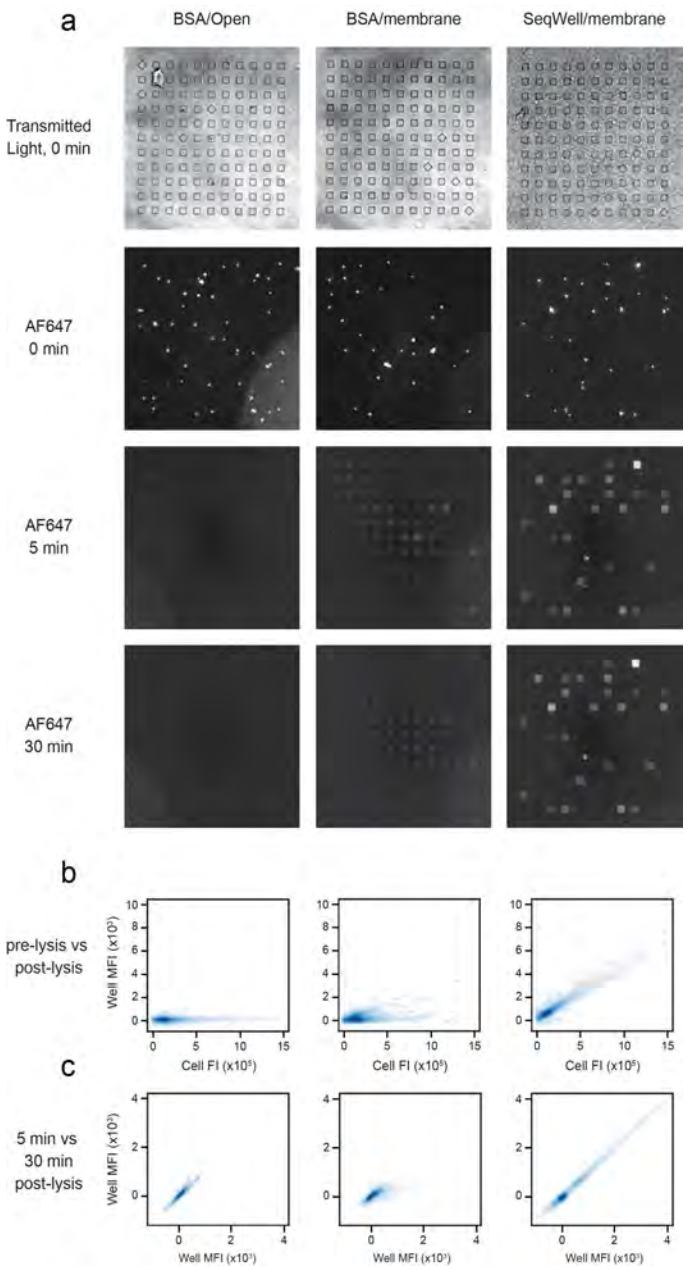
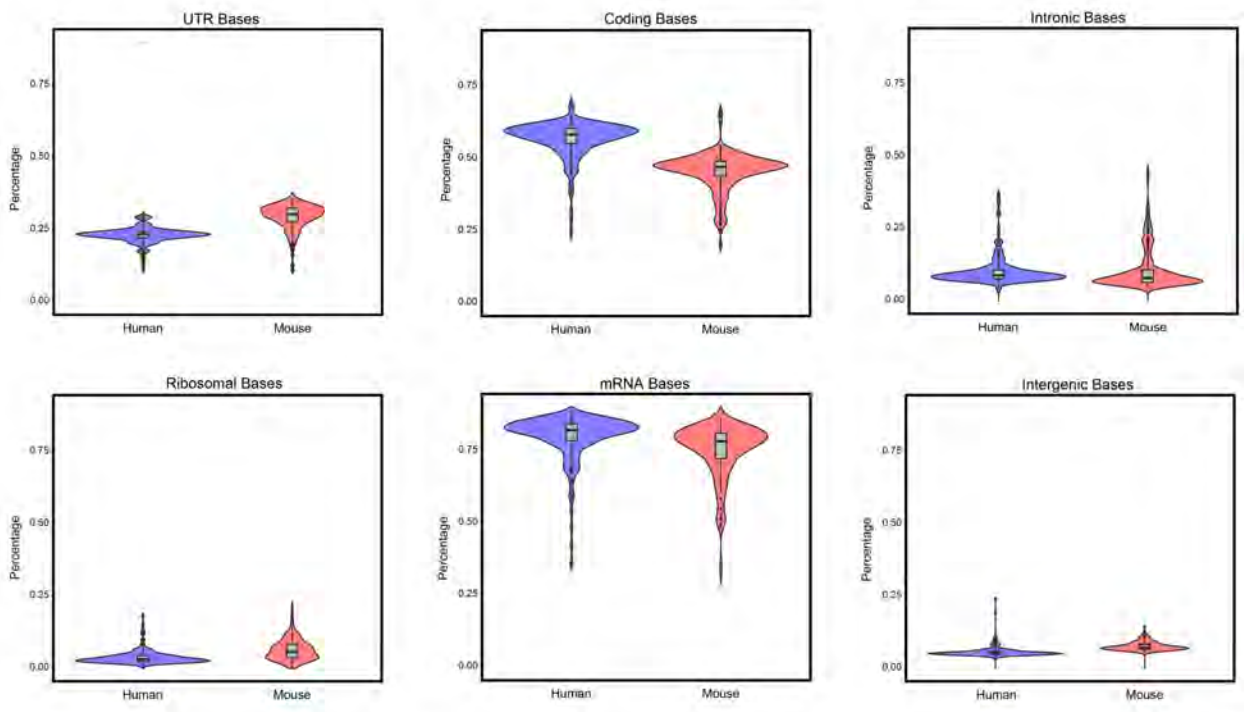


Figure A1.5. PBMCs labeled with CD45-AF647 were loaded into two BSA-blocked arrays and one array functionalized with chitosan and poly(glutamate). A semipermeable membrane was attached to one of the BSA-blocked arrays and the chitosan:polyglutamate functionalized array prior to addition of lysis buffer. **(A)** Example images of transmitted light and AF647 fluorescence of the arrays before and 5 and 30 minutes after addition of lysis buffer are displayed for each array. **(B)** The total fluorescence intensity (FI) of all pixels associated with cells within a well is plotted against the median fluorescent intensity (MFI) of the volume of the same well 5 minutes after lysis for 12,100 wells from each array. **(C)** The MFI of the well volume 5 minutes after lysis is plotted against the MFI of the volume of the same well 30 minutes after lysis for the same 12,100 wells from each array.



Species	Ribosomal Bases	mRNA Bases	Intergenic Bases	Intronic Bases	Coding Bases	UTR Bases
Human	3.88%	80.0%	5.79%	10.4%	56.9%	23.1%
Mouse	6.32%	75.2%	7.68%	10.9%	45.6%	29.6%

Figure A1.6. Read mapping quality matrices were generated for each sample for human (blue) and mouse (red) cells, aligned to hg19 and mm10, respectively. High quality samples had relatively higher percentages of annotated genomic (genic) and exonic transcripts and low percentages of annotated intergenic and ribosomal transcripts (Center-line: Median; Limits: 1st and 3rd Quartile; Whiskers: +/- 1.5 IQR; Points: Values > 1.5 IQR).

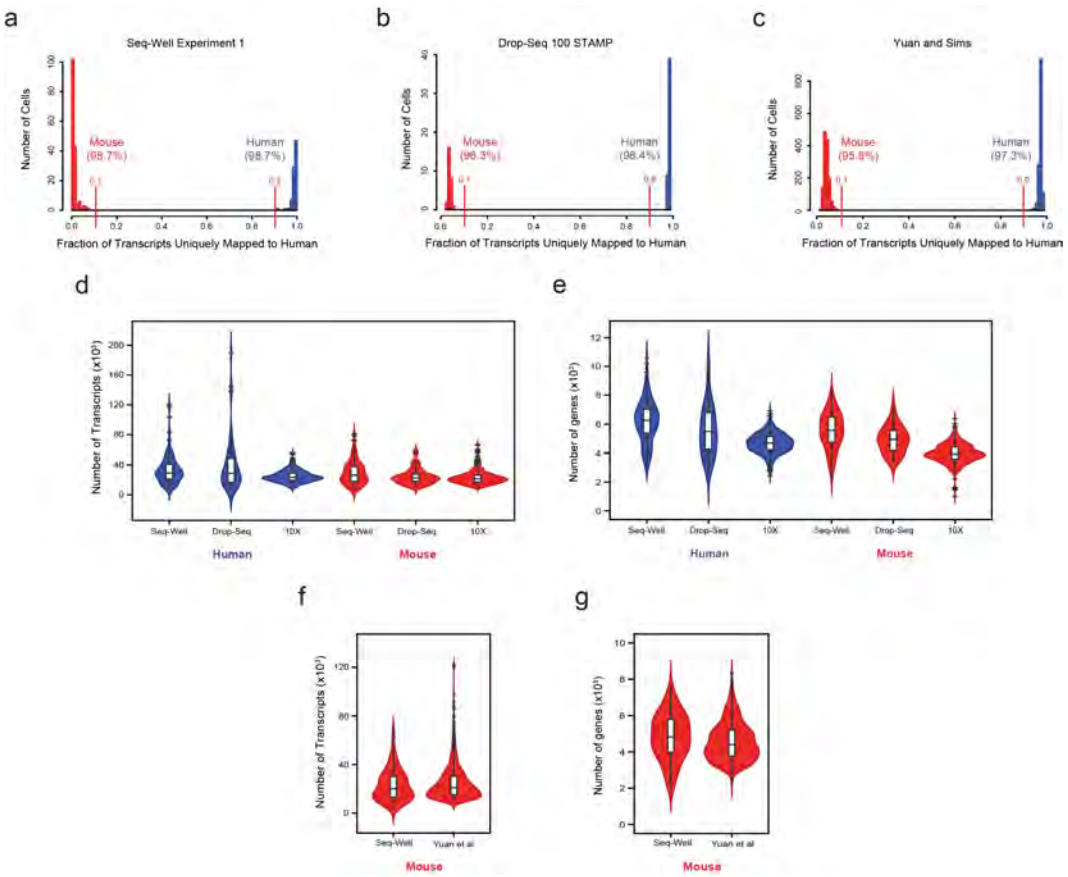


Figure A1.7 Histograms of the percent cross-species contamination in **(A)** Seq-Well, **(B)** Drop-Seq (Ref. 12), and **(C)** Yuan and Sims. In each plot, cells with greater than 90% of human transcripts are displayed in blue and cells with less than 10% human transcripts are displayed in red. **(D)** Transcript capture in human (blue) and mouse (red) cell lines across three massively-parallel, bead-based single-cell sequencing platforms (Seq-Well, Drop-Seq, and 10X Genomics, with downsampling to an average read-depth of 80,000 reads per cell, consistent with 10X genomics data (Center-line: Median; Limits: 1st and 3rd Quartile; Whiskers: +/- 1.5 IQR; Points: Values > 1.5 IQR). We detect an average of 32,841 human transcripts and 29,806 mouse transcripts using Seq-Well compared to an average of 39,400 human transcripts and 24,384 mouse transcripts using Drop-Seq, an average of 24,751 human transcripts and 22,971 mouse transcripts using 10X Genomics (available from). **(E)** Gene detection across human and mouse cell lines across the same three single-cell sequencing platforms with down-sampling to the average read-depth of 80,000 reads per cell, consistent with 10X genomics (Center-line: Median; Limits: 1st and 3rd Quartile; Whiskers: +/- 1.5 IQR; Points: Values > 1.5 IQR). We detect an average of 6,174 human genes and 5,528 mouse genes using Seq-Well, an average of 5,561 human genes and 4,903 mouse genes using Drop-Seq and an average of 4,655 human genes and 3,950 mouse genes using 10X Genomics. **(F)** Downsampling to an average of 42,000 reads per cell consistent with data published in Yuan and Sims 2016, results in average detection of 23,061 mouse transcripts using Seq-Well compared to an average of 24,761 mouse transcripts using the Yuan and Sims platform (Center-line: Median; Limits: 1st and 3rd Quartile; Whiskers: +/- 1.5 IQR; Points: Values > 1.5 IQR). **(G)** Downsampling to an average of 42,000 reads per cell results in average detection of 4,827 mouse genes using Seq-Well compared to an average of 4,569 mouse genes using the Yuan and Sims platform (Center-line: Median; Limits: 1st and 3rd Quartile; Whiskers: +/- 1.5 IQR; Points: Values > 1.5 IQR).

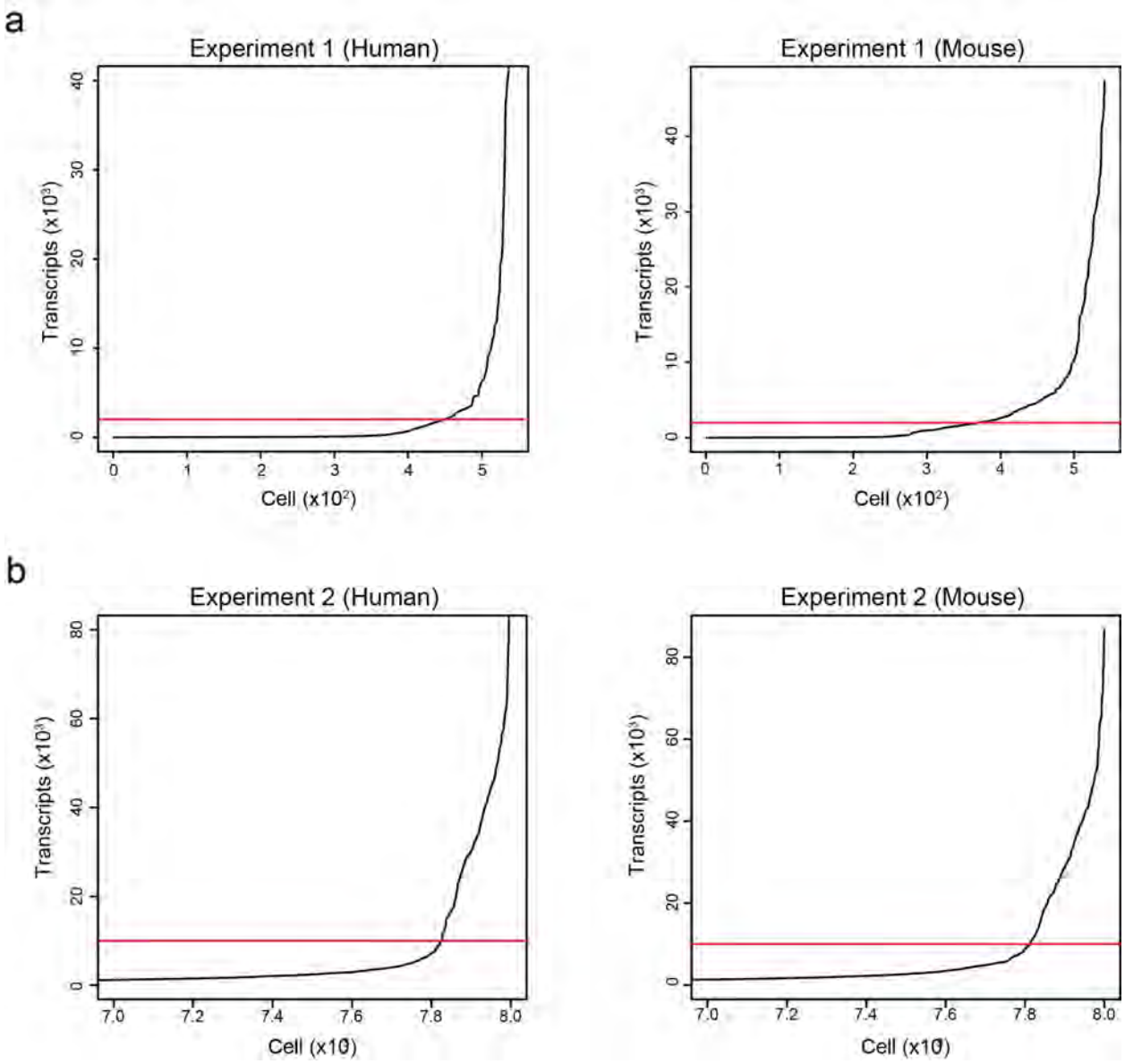


Figure A1.8. We sequenced two arrays (**A-B**) to confirm single-cell resolution and minimal cross-contamination between mouse and human cells. We called cells by plotting the cumulative distribution of transcripts and making a cutoff at the elbow in the curve. In the first experiment **(A)**, which was used to validate our single-cell resolution, we shallowly sequenced the array and made the cutoff at 2,000 transcripts. In the second experiment **(B)**, where we sequenced the array deeply to allow a competitive comparison to Drop-Seq, we made our cutoff at 10,000 transcripts.

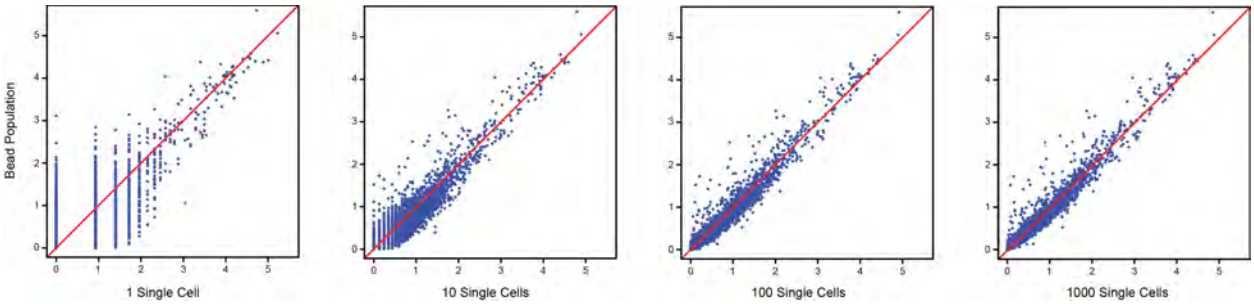


Figure A1.9. Scatterplots showing the correlation between gene expression estimates from bulk populations (40,000 HEK cells and 40,000 mRNA capture) and populations generated in-silico from 1, 10, 100, and 1,000 randomly-sampled single HEK293 cells (1 Cell: R = 0.751 ± 0.0726; 10 Cells: R = 0.952 ± 0.008; 100 Cells: R = 0.980 ± 0.0006; 1000 Cells: R = 0.983 ± 0.0001)**.**

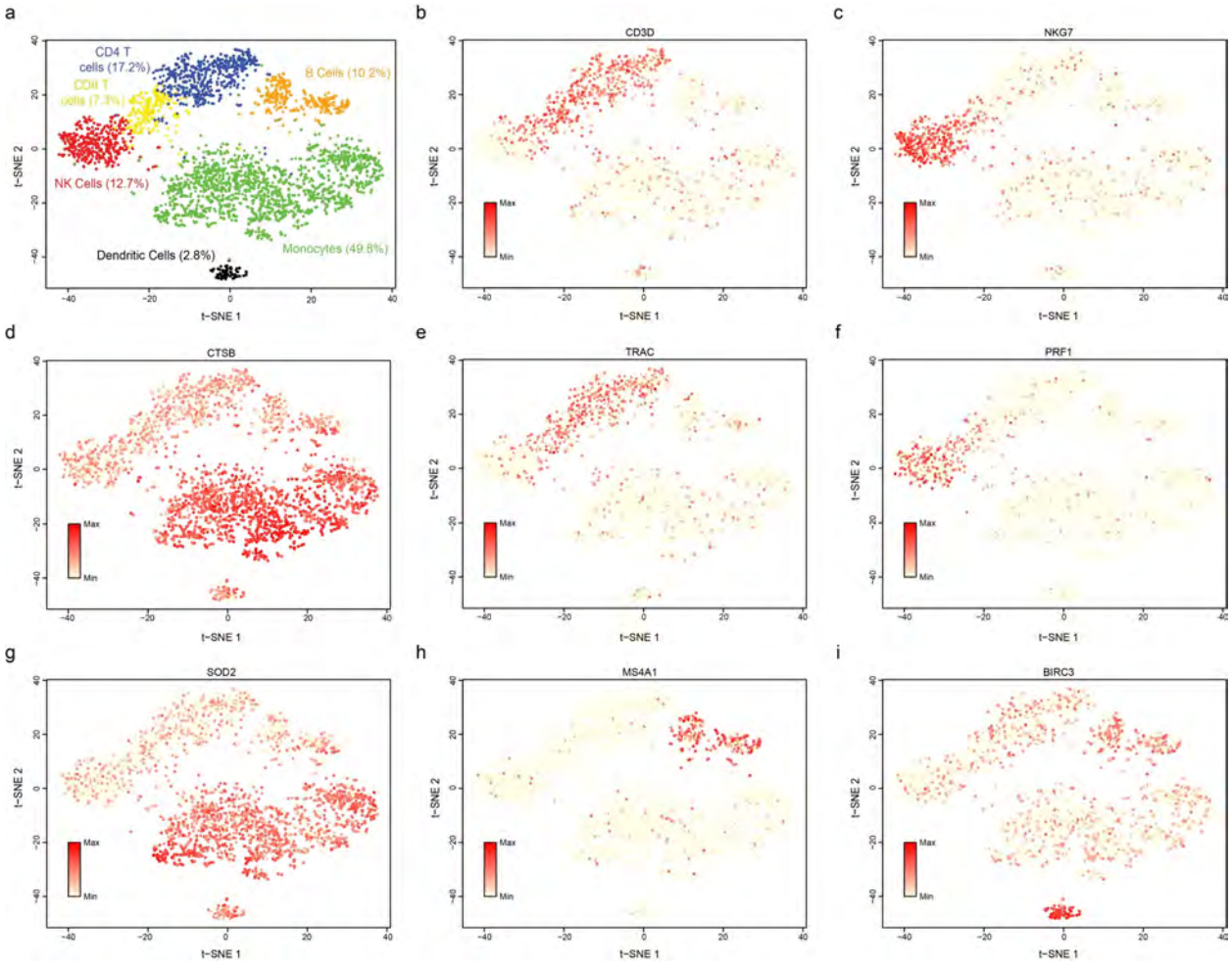


Figure A1.10. (A) Clusters identified through graph-based clustering (**Methods**) correspond to major immune cell populations. **(B,E)** CD4 T cells are characterized by expression of CD3D and T-cell receptor expression without pronounced expression of cytoxic genes NKG7 and PRF1. **(C,F)** CD8 T cells are defined by expression of NKG7 and PRF1. **(D,G)** Monocytes are defined by expression of cathepsin B (CTSB) and SOD2. **(E)** Natural killer cells are characterized by expression of cytotoxic genes in the absence of T cell receptor expression. **(H)** B cells are marked by elevated expression of MS4A1 (CD20) transcripts. **(I)** Dendritic cells are enriched for expression of BIRC3.

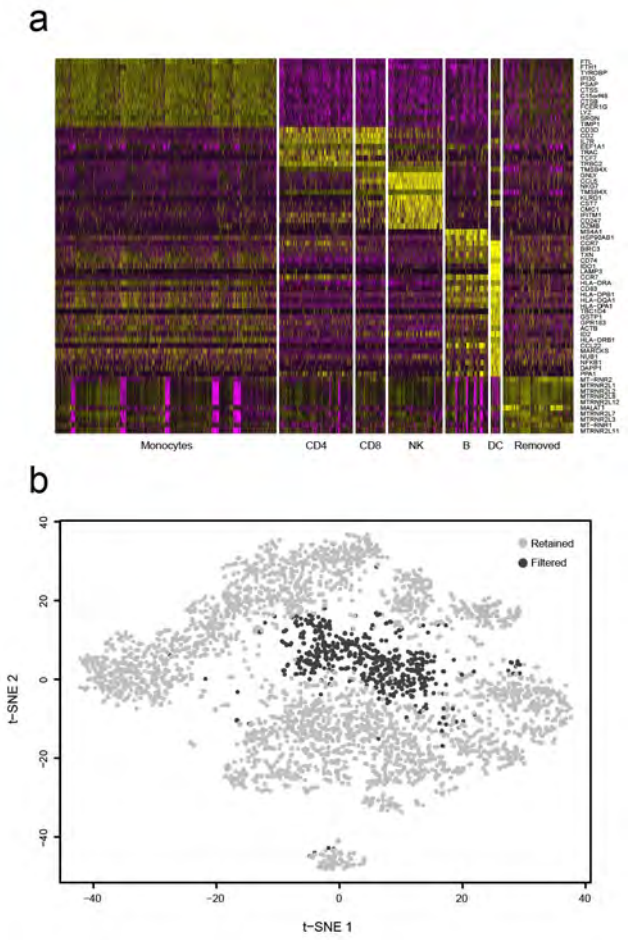


Figure A1.11. (A) Genes enriched in each cluster were identified using an "ROC" test in Seurat, comparing cells assigned to each cluster to all other cells. A heatmap was constructed using enriched genes found to define each cluster. One cluster of 602 cells that demonstrated exclusive enrichment of mitochondrial genes was removed as these likely represent low-quality or dying cells. **(B)** We generated a t-SNE projection of 4,296 cells with greater than 10,000 reads, 1,000 transcripts, 500 genes, and 65% transcript mapping. We removed a total of 602 cells from the final analysis found to be strongly enriched for expression of mitochondrial genes. The remaining 3,694 cells form distinct clusters enriched for lineage-defining that distinguish cells types from one another.

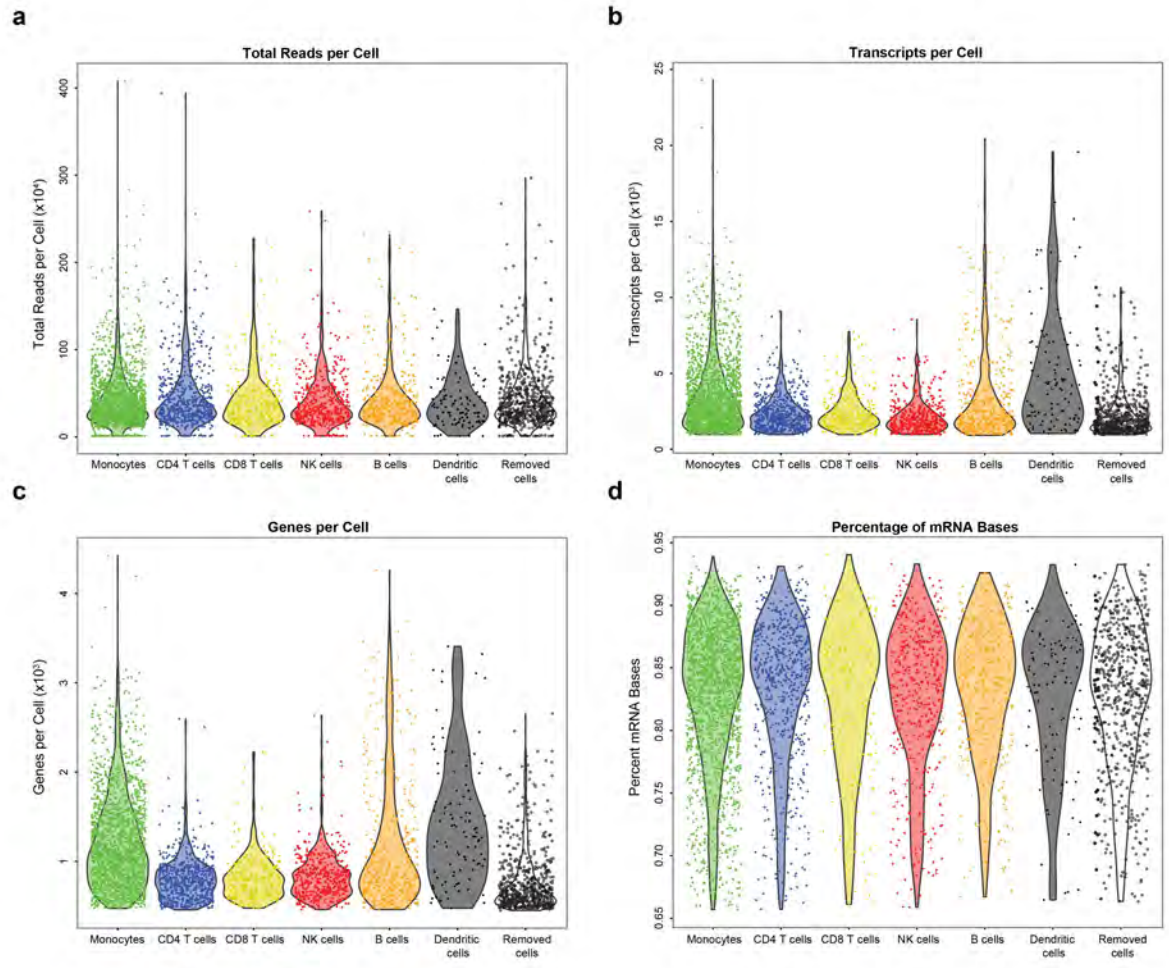


Figure A1.13. (A) Comparison of transcript capture (top) and gene detection (bottom) between Seq-Well and 10X Genomics within PBMC cell types prior to downsampling (colored as in Figure 2; Center-line: Median; Limits: 1st and 3rd Quartile; Whiskers: +/- 1.5 IQR; Points: Values > 1.5 IQR). Among B cells (orange), an average of 1,315 genes and 3,632 transcripts were detected using Seq-Well and an average of 710 genes and 1,910 transcripts were detected in 10X Genomics data. Among CD4 T cells (blue), an average of 861 genes and 2,444 transcripts were detected using Seq-Well and an average of 815 genes and 2,370 transcripts were detected in 10X Genomics data. Among CD8 T cells (yellow), an average of 885 genes and 2,574 transcripts were detected using Seq-Well and an average of 809 genes and 2,029 transcripts were detected in 10X Genomics data. Among Monocytes (green), an average of 1,288 genes and 3,568 transcripts were detected using Seq-Well and an average of 974 genes and 2,835 transcripts were detected in 10X Genomics data. Among NK cells (red), an average of 902 genes and 2,338 transcripts were detected using Seq-Well and an average of 907 genes and 1,943 transcripts were detected in 10X Genomics data. **(B)** Transcript capture (top) and gene detection (bottom) upon downsampling of Seq-Well data to an average read depth 69,000 reads per cell (Center-line: Median; Limits: 1st and 3rd Quartile; Whiskers: +/- 1.5 IQR; Points: Values > 1.5 IQR). Upon downsampling, in Seq-Well, an average of 1,048 genes and 3103 transcripts were detected among B cells, 735 genes and 2,221 transcripts among CD4 T cells, 763 genes and 2,353 transcripts among CD8 T cells, 1,052 genes and 3,105 transcripts among monocytes, and 789 genes and 2,041 transcripts among NK cells.

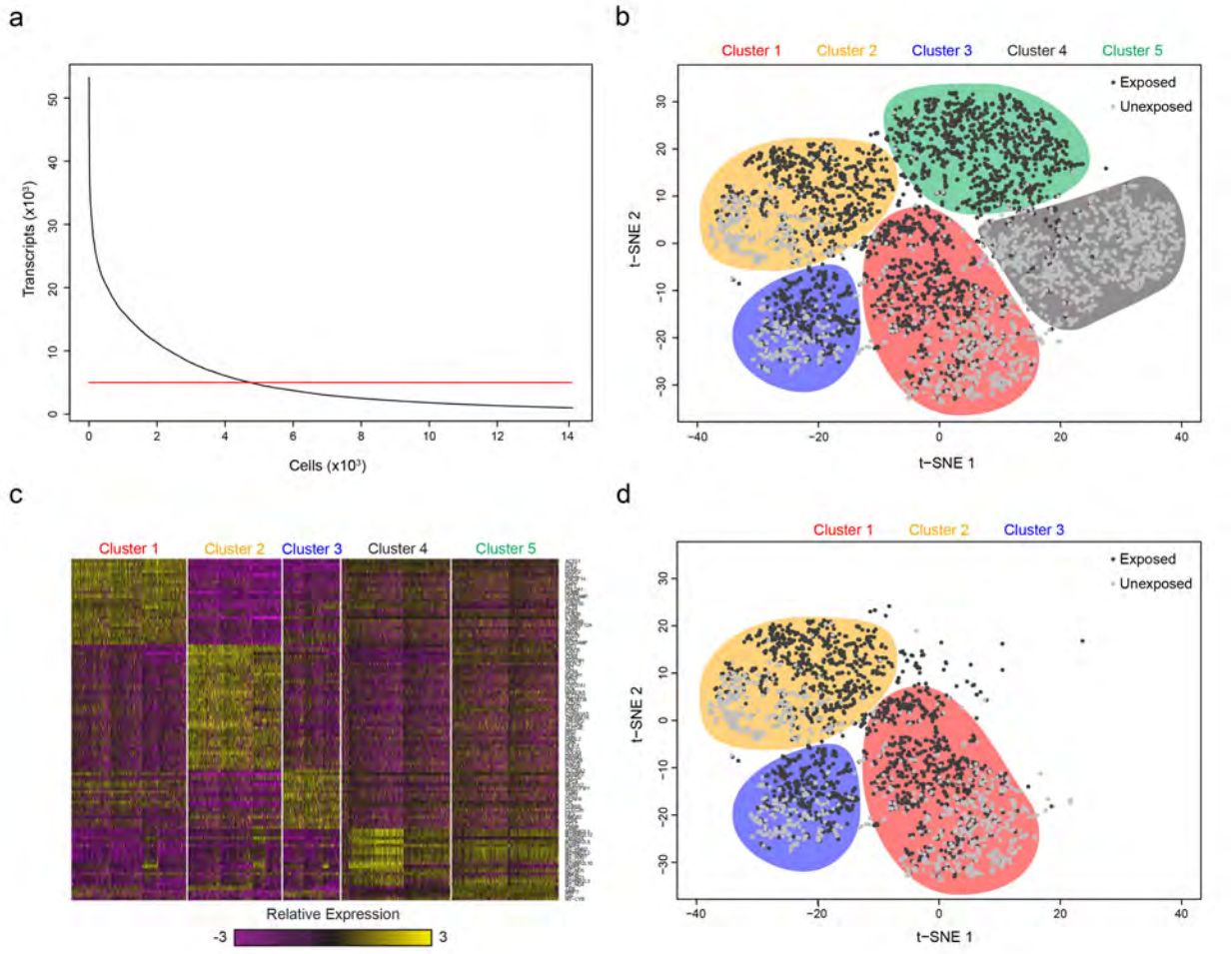


Figure A1.14. (A) Using a threshold of 5,000 detected transcripts, we identified 4,638 macrophages. **(B)** Among these 4,638 cells, we identified 5 distinct clusters of macrophages by performing graph-based clustering over 5 principal components (377 variable genes). **(C)** Clusters 1-3 are defined by unique gene expression signatures, while Clusters 4 and 5 are defined by expression of mitochondrial genes, suggesting low-quality cells. **(D)** Following removal of cells within Clusters 4 and 5, there remain a total of 2,560 cells in Clusters 1-3.

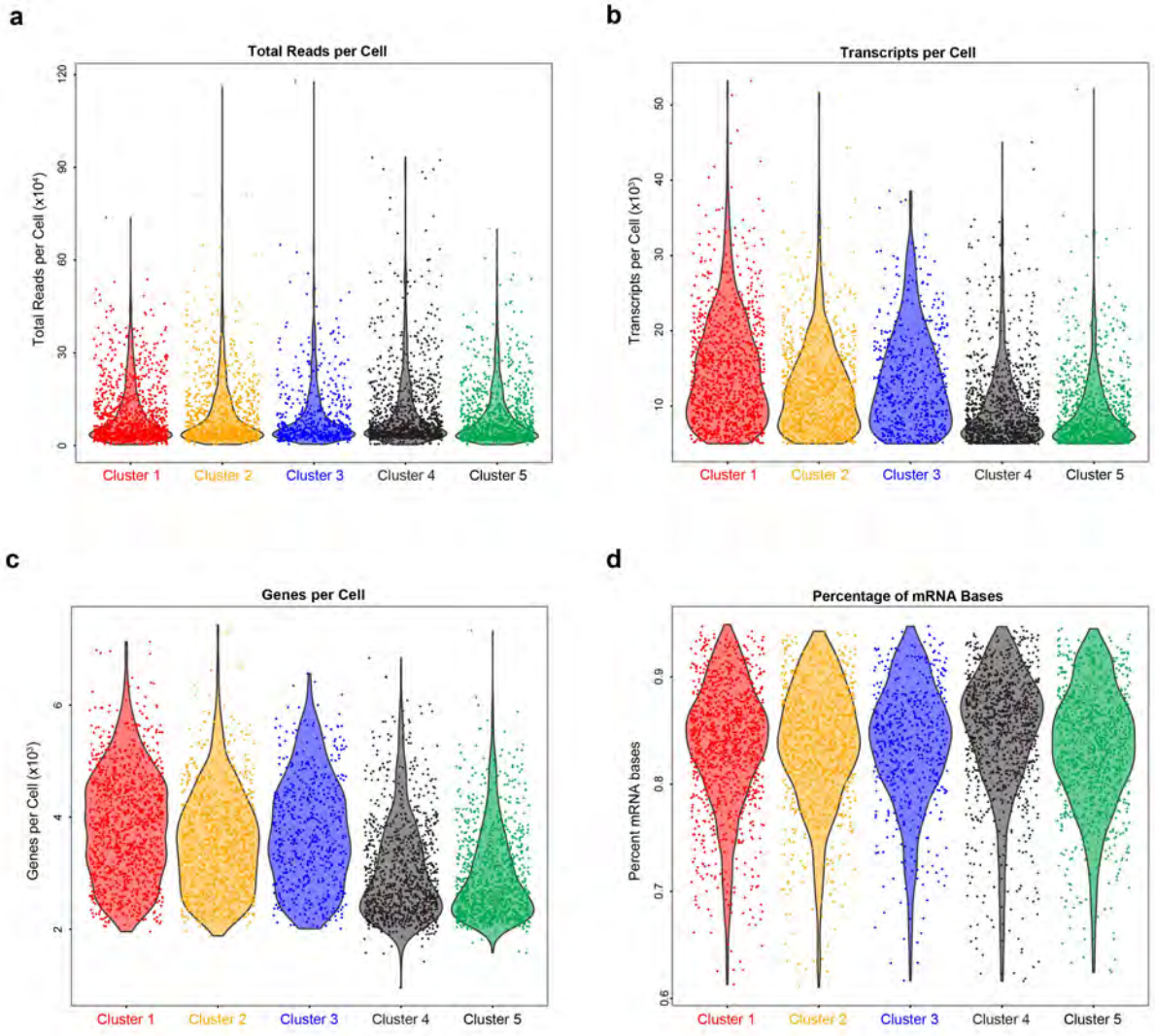


Figure A1.15. **(A-C)** Violin plots depicting reads **(A)**, transcripts **(B)**, and genes **(C)** per cell, separated by cluster. **(D)** Percent mRNA bases per cell, separated by cluster.

Appendix 2

Corresponds to Chapter 3.

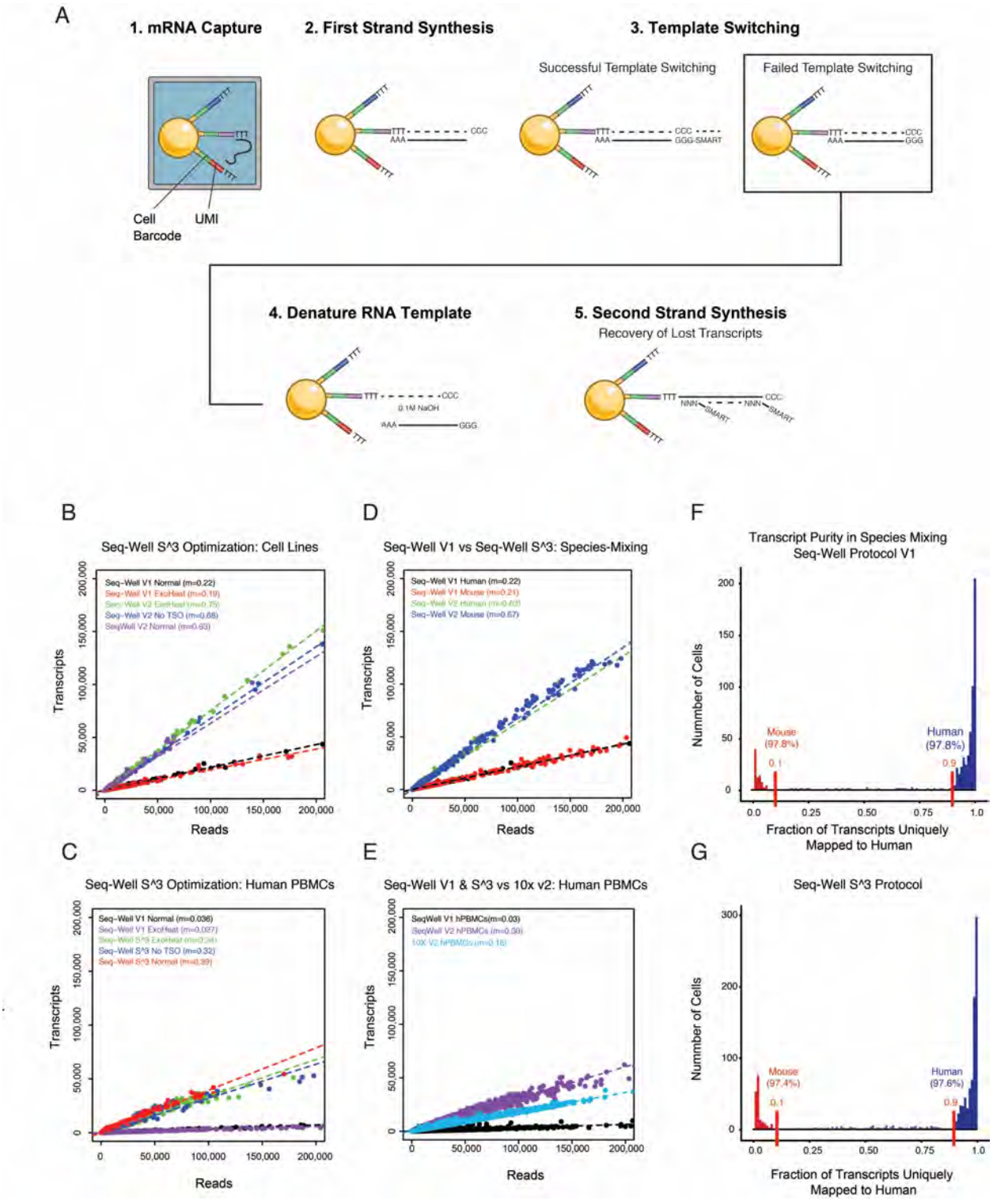


Figure A2.1. (A) Illustration of the second strand synthesis procedure: (1) mRNA is captured via poly-T priming of poly-adenylated mRNA; (2) First strand synthesis is performed to generate single-stranded cDNA template on bead-bound sequences; (3) Successful template switching: The use of enzymes with terminal transferase activity generates a 3' overhang of 3 cytosines. Template switching utilizes this overhang to append the SMART sequence to both ends of the

cDNA molecule during first strand synthesis. Failed Template Switching: If template switching fails, this results in loss of previously primed and reverse transcribed mRNA molecules; (4) mRNA template is chemically denatured using 0.1M NaOH; (5) Second strand synthesis is performed using a random-octamer with the SMART sequence in the 5' orientation; and, (6) Following second strand synthesis, PCR amplification, library preparation and sequencing are performed to generate data. (**B)** Scatterplots show the relationship between transcript detection (y-axis) and number of aligned reads per cell (x-axis) for a series of optimization experiments using HEK293 and NIH-3T3 cell lines. (**C)** Scatterplots show the relationship between transcript detection (y-axis) and the number of aligned reads per cell (x-axis) for a series of optimization experiments using PBMCs. (**D)** Scatterplots that illustrate the relationship between number of transcripts detected (y-axis) and number of aligned reads per cell (x-axis) between Seq-Well V1 and Seq-Well S^3 in species mixing experiments using HEK293 and NIH-3T3 cells. **(E)** Scatterplots that illustrate the relationship between number of transcripts detected (y-axis) and number of aligned reads per cell (x-axis) between Seq-Well V1 and Seq-Well S^3 in a series of optimization experiment using human PBMCs. **(F)** Histograms that show the fraction of transcripts uniquely mapped to the human genome for each cell for Seq-Well V1. Colors indicate species classification for cells with at least 90% purity of human (blue) or mouse (red) mapping. **(G)** Histograms that show the fraction of transcripts uniquely mapped to the human genome for each cell for Seq-Well S^3. Colors indicate species classification for cells with at least 90% purity of human (blue) or mouse (red) mapping.

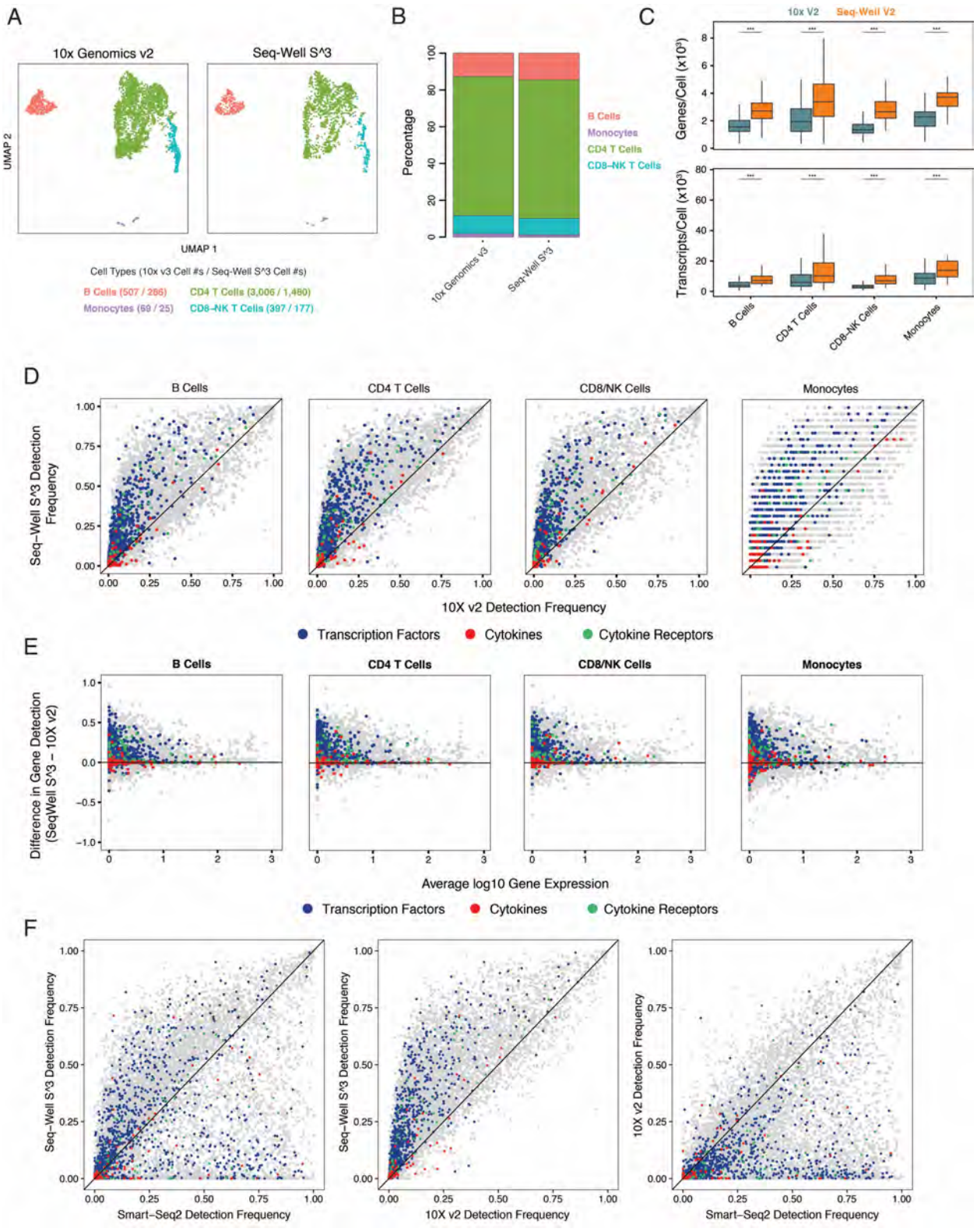


Figure A2.2. (A) UMAP plots showing detected cell-types among PBMCs using 10X v2 (left) and Seq-Well S^3 (right). **(B)** Stacked barplots show the proportion of cell types recovered using Seq-Well S^3 (left) and 10X v2 (right). **(C)** Top: Boxplots (median +- quartiles) showing the distribution

of per cell gene detection from 10X v2 (left) and Seq-Well S^3 (right). Bottom: Boxplots (median +/- quartiles) showing the distribution of per cell- transcript capture from 10X v2 (left) and Seq-Well S^3 (right). **(D)** Scatterplots showing a comparison of gene detection frequencies between Seq-Well S^3 (y-axis) and 10x v2 (x-axis) for each cell type. **(E)** Scatterplots showing the difference in gene detection between Seq-Well S^3 and 10X v2 (y-axis) as a function of average normalized expression (log(scaled UMI + 1)) (x-axis). **(F)** Scatterplots showing a comparison of gene detection frequencies among sorted $CD4^+$ T cells between **(Left)** Seq-well S^3 (y-axis) and 10x v2 (x-axis), (**Middle)** Seq-Well S^3 (y-axis) and Smart-Seq2 (x-axis), and **(Right)** 10x v2 (y-axis) and Smart-Seq2 (x-axis).

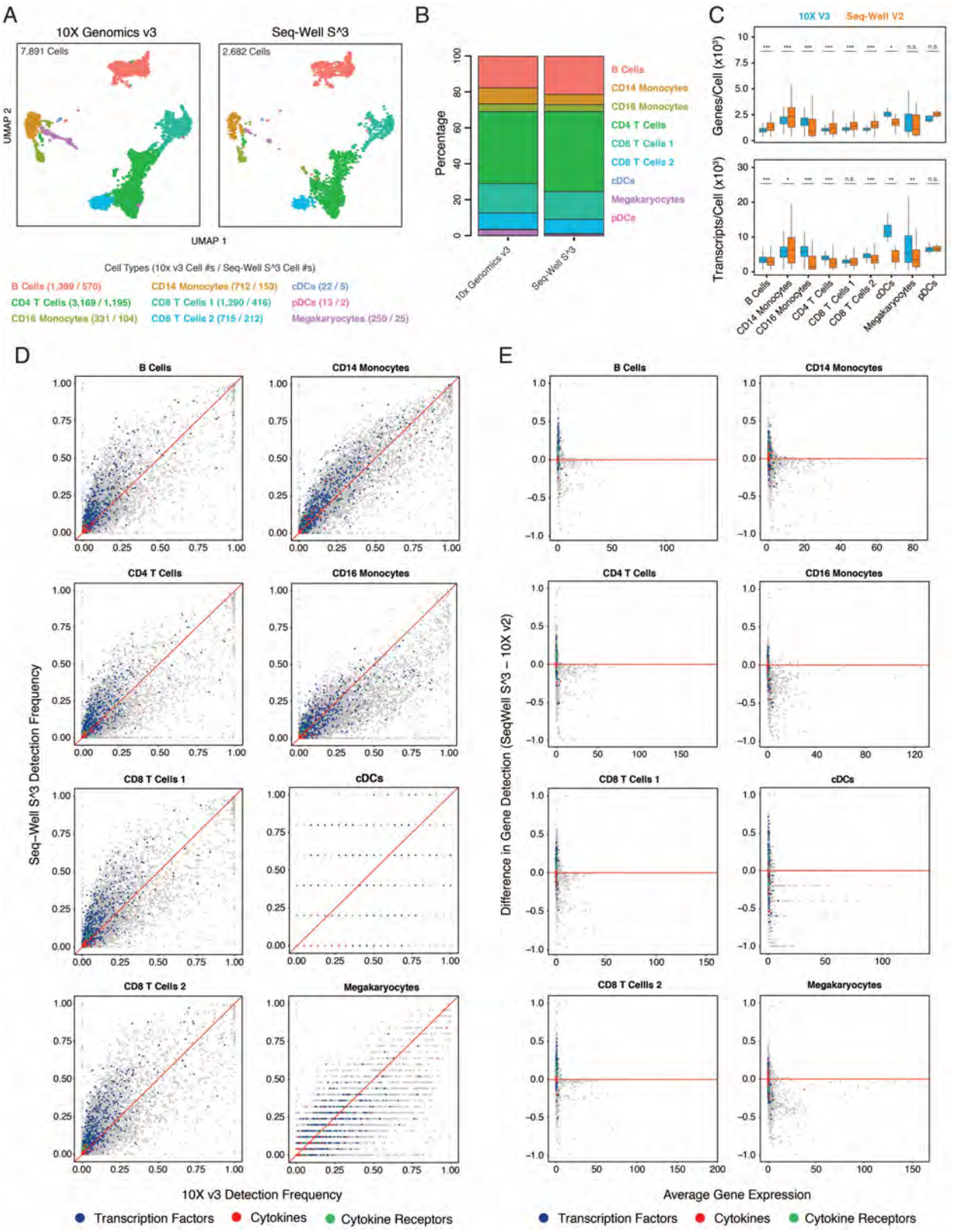


Figure A2.3. (A) UMAP plots for 7,891 PBMCs from 10x Genomics V3 (left) and 2,682 PBMCs from Seq-Well S^3 (right). **(B)** Stacked barplots showing the distribution of recovered cell types

between 10x Genomics V3 and Seq-Well S^3. **(C)** Top: Boxplots (median +- quartiles) showing the distribution of per cell gene detection from 10X v2 (left) and Seq-Well S^3 (right). Bottom: Boxplots (median +/- quartiles) showing the distribution of per cell-gene detection from 10X v2 (left) and Seq-Well S^3 (right). **(D)** Scatterplots showing a comparison of gene detection frequencies between Seq-Well S^3 (y-axis) and 10x v3 (x-axis) for each cell type. **(E)** Scatterplots showing the difference in gene detection between Seq-Well S^3 and 10X v3 (y-axis) as a function of average normalized expression (log(scaled UMI + 1)) (x-axis).

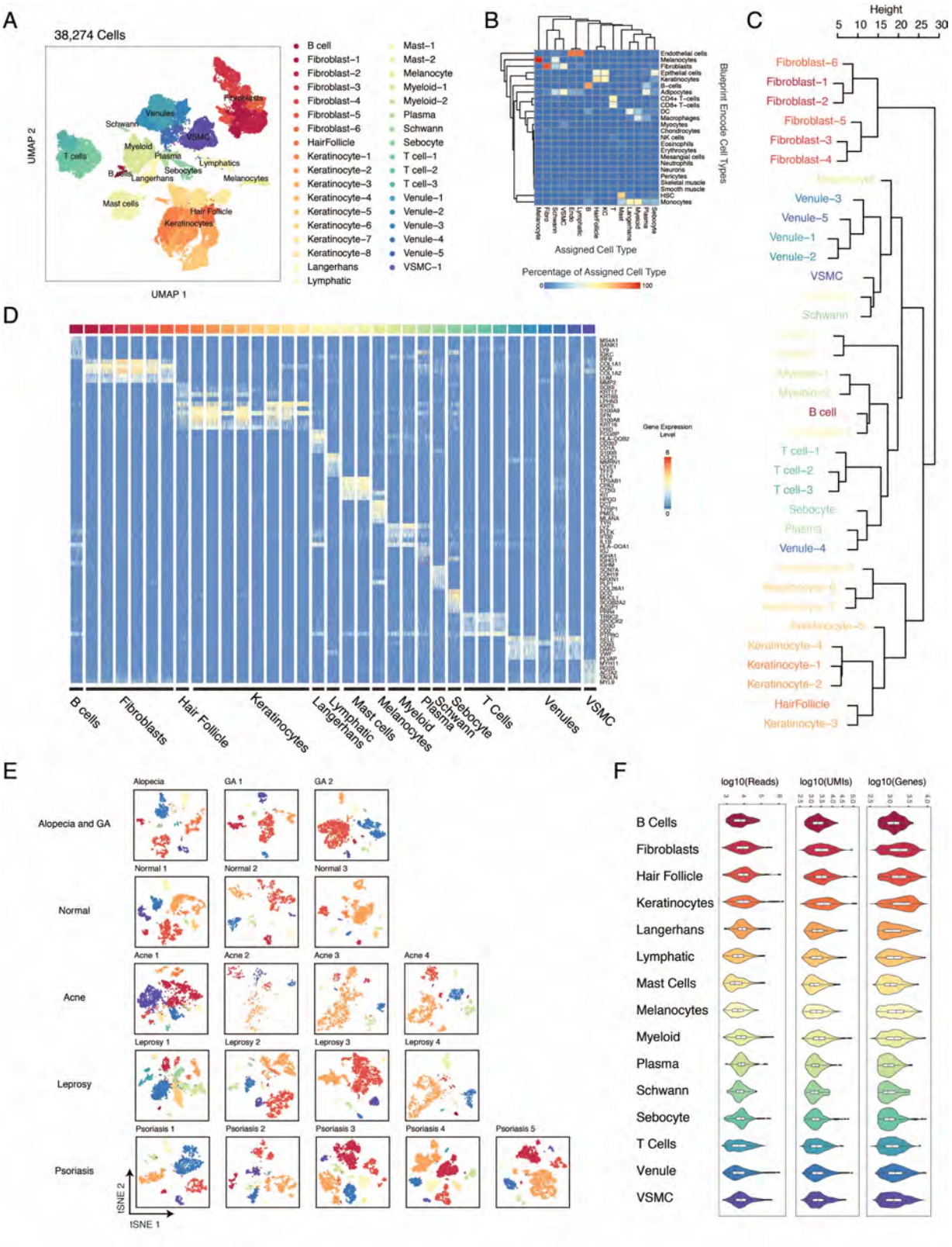

A
38,274 Cells
UMAP 2
UMAP 1
Fibroblasts
Venules
Schwann
VSMC
Myeloid
Plasma
T cells
Lymphatics
B cells
Sebocytes
Langerhans
Melanocytes
Mast cells
Hair Follicle
Keratinocytes
B cell
Fibroblast−1
Fibroblast−2
Fibroblast−3
Fibroblast−4
Fibroblast−5
Fibroblast−6
HairFollicle
Keratinocyte−1
Keratinocyte−2
Keratinocyte−3
Keratinocyte−4
Keratinocyte−5
Keratinocyte−6
Keratinocyte−7
Keratinocyte−8
Langerhans
Lymphatic
Mast−1
Mast−2
Melanocyte
Myeloid−1
Myeloid−2
Plasma
Schwann
Sebocyte
T cell−1
T cell−2
T cell−3
Venule−1
Venule−2
Venule−3
Venule−4
Venule−5
VSMC−1
B
Assigned Cell Type
Blueprint Encode Cell Types
Percentage of Assigned Cell Type
C
Height
5 10 15 20 25 30
Fibroblast−6
Fibroblast−1
Fibroblast−2
Fibroblast−5
Fibroblast−3
Fibroblast−4
Venule−3
Venule−5
Venule−1
Venule−2
VSMC
Schwann
B cell
T cell−1
T cell−2
T cell−3
Sebocyte
Plasma
Venule−4
Keratinocyte−4
Keratinocyte−1
Keratinocyte−2
HairFollicle
Keratinocyte−3
D
Gene Expression Level
B cells
Fibroblasts
Hair Follicle
Keratinocytes
Langerhans
Lymphatic
Mast cells
Melanocytes
Myeloid
Plasma
Schwann
Sebocyte
T Cells
Venules
VSMC
E
Alopecia
GA 1
GA 2
Alopecia and GA
Normal 1
Normal 2
Normal 3
Normal
Acne 1
Acne 2
Acne 3
Acne 4
Acne
Leprosy 1
Leprosy 2
Leprosy 3
Leprosy 4
Leprosy
Psoriasis 1
Psoriasis 2
Psoriasis 3
Psoriasis 4
Psoriasis 5
Psoriasis
tSNE 2
tSNE 1
F
log10(Reads)
log10(UMIs)
log10(Genes)
B Cells
Fibroblasts
Hair Follicle
Keratinocytes
Langerhans
Lymphatic
Mast Cells
Melanocytes
Myeloid
Plasma
Schwann
Sebocyte
T Cells
Venule
VSMC

Figure A2.4. (A) UMAP plot for 38,274 cells colored by 35 cell type clusters. **(B)** Comparison of cell-type classification to results obtained using SingleR. Color corresponds to the percentage of manually classified cells types (columns) assigned to cell-type references contained in the Blueprint Encode dataset (rows). **(C)** Dendrogram of hierarchical clustering shows similarity of cell type clusters among top 25 cluster-defining genes. **(D)** Heatmap showing the relative expression of cell-type defining gene signatures across 38,274 cells (**Table S4**). **(E)** t-SNE plots for each of the nine skin biopsies colored by generic cell type. **(F)** Violin plots show the distribution of per-cell quality metrics displayed in UMAP embedding of 38,274 cells colored by colored generic cell-type classification (**Figure 3.2B**).

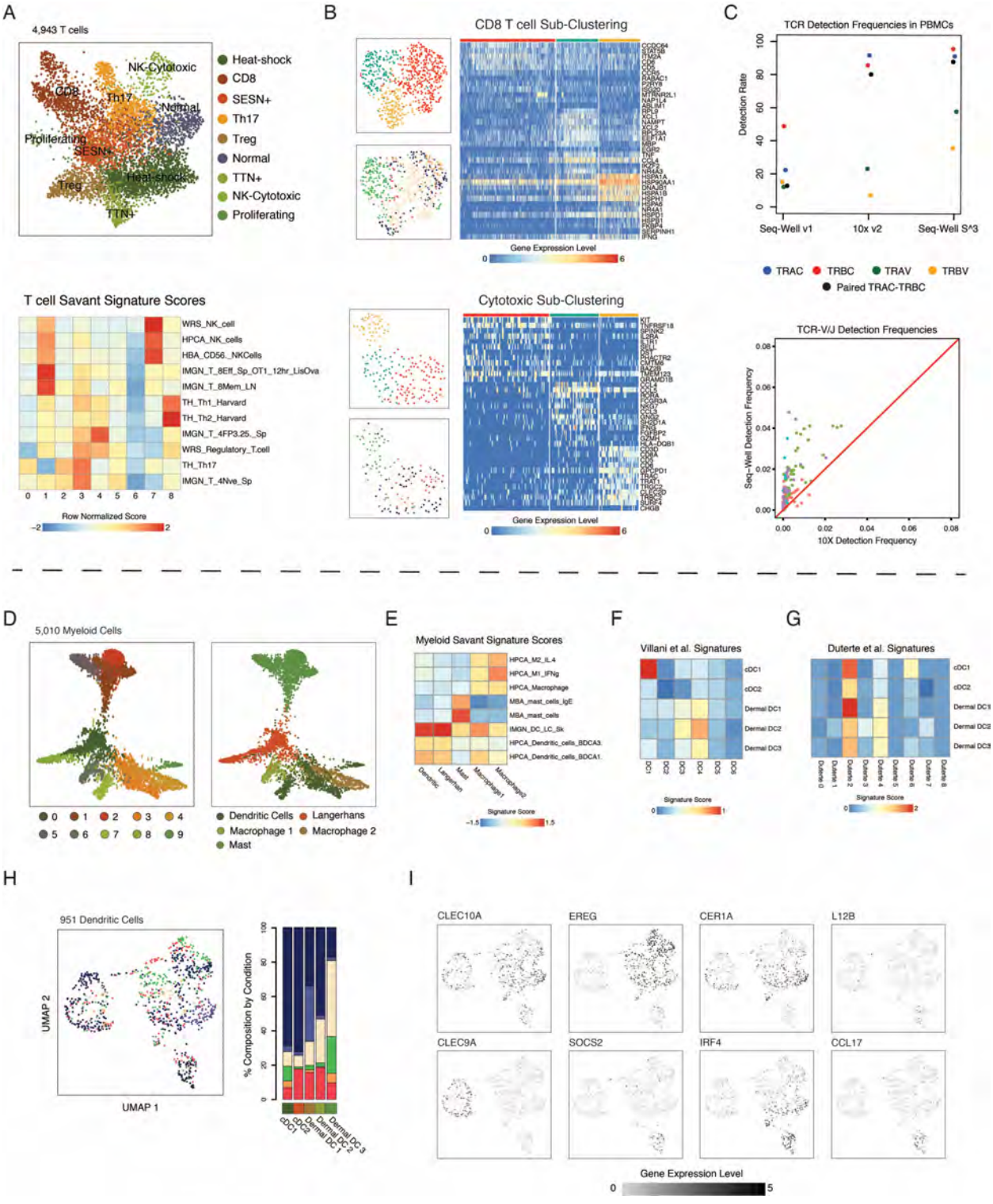


Figure A2.5. (A) (Top) Force-directed graph of 4,943 T cells colored by T cell sub-cluster. (Bottom) Heatmap of gene-set enrichment scores based on comparison of T cell phenotypic sub-clusters to a curated list of reference signatures in the Savant database. **(B)** Sub-grouping results

for CD8 T cells (top) (**Table S7**) and cytotoxic cells (bottom) (**Table S8**). For each analysis, t-SNE plots colored by inflammatory skin condition (top-left) and sub-cluster (bottom-left) are shown. For each clusters, heatmaps show gene expression patterns across T and NK cells sub-types (right). **(C)** (Top) Detection rates for TCR genes for PBMCs in Seq-Well v1, 10x v2. and Seq-Well S^3. (Bottm) Detection frequency of TCR V-J (e.g. TRAV/J and TRBV/J) genes in $CD4^+$ T cells from peripheral blood between Seq-Well S^3 (y-axis) and 10x v2 (x-axis). Colors correspond to TRAJ (red), TRAV (green), TRBJ (blue), and TRBV (purple) genes. **(D)** (Left) Force-directed graph of 5,010 myeloid cells colored by myeloid sub-clusters (Louvain resolution = 0.6). (Right) Force-directed graph of 5,010 myeloid cells colored by myeloid phenotypes. **(E)** Heatmap of gene-set enrichment scores based on comparison of myeloid phenotypic sub-clusters to a curated list of reference signatures in the Savant database. **(F)** Heatmap showing average signature score across 5 dermal DC populations based on dendritic cell signatures from *Villani et al. Science 2017*. **(G)** Heatmap showing average signature score across 5 dermal DC populations based on signatures from et al. Immunity 2019. **(H)** (Left) UMAP plot for 951 dendritic cells from human skin colored by inflammatory skin condition. (Right) Stacked barplot showing composition of dendritic cells within each of nine skin biopsies by DC sub-cluster. **(I)** UMAP plots colored by normalized expression levels for DC sub-grouping-defining genes.

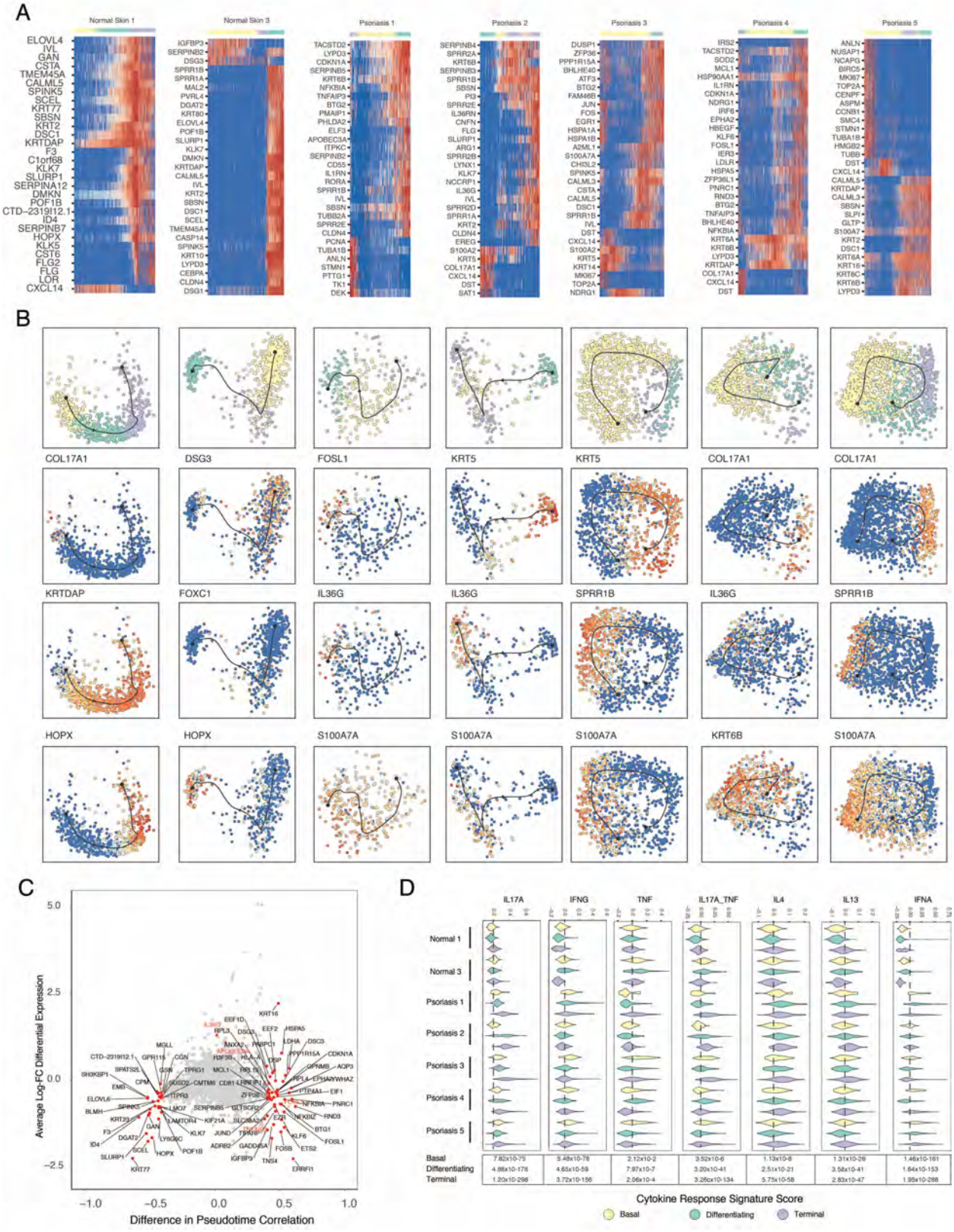

A
Normal Skin 1
Normal Skin 3
Psoriasis 1
Psoriasis 2
Psoriasis 3
Psoriasis 4
Psoriasis 5
B
COL17A1
DSG3
FOSL1
KRT5
KRT5
COL17A1
COL17A1
KRTDAP
FOXC1
IL36G
IL36G
SPRR1B
IL36G
SPRR1B
HOPX
HOPX
S100A7A
S100A7A
S100A7A
KRT6B
S100A7A
C
Average Log-FC Differential Expression
Difference in Pseudotime Correlation
D
IL17A
IFNG
TNF
IL17A_TNF
IL4
IL13
IFNA
Normal 1
Normal 3
Psoriasis 1
Psoriasis 2
Psoriasis 3
Psoriasis 4
Psoriasis 5
Basal
Differentiating
Terminal
Cytokine Response Signature Score
Basal
Differentiating
Terminal

Figure A2.6. (A) Heatmaps showing enrichment of genes along pseudo-temporal trajectories for keratinocytes from 2 normal skin biopsies and 5 psoriatic keratinocyte biopsies. **(B)** Plots showing the expression of trajectory-defining genes for 2 normal and 5 psoriatic biopsies. **(C)** Scatterplot showing the relationship between differential expression between psoriatic and normal keratinocyte (y-axis) and differential pseudotime correlation (x-axis). Genes highlighted in red have differential correlation values greater than 0.4 or less than -0.4. **(D)** Violin plots showing localization of cytokine response signatures in basal, differentiating and terminal keratinocytes across 2 normal and 5 psoriatic biopsies. P-values from T-tests showing differences in cytokine signature scores for basal, differentiating and terminal keratinocytes between normal and psoriatic keratinocytes.

Appendix 3

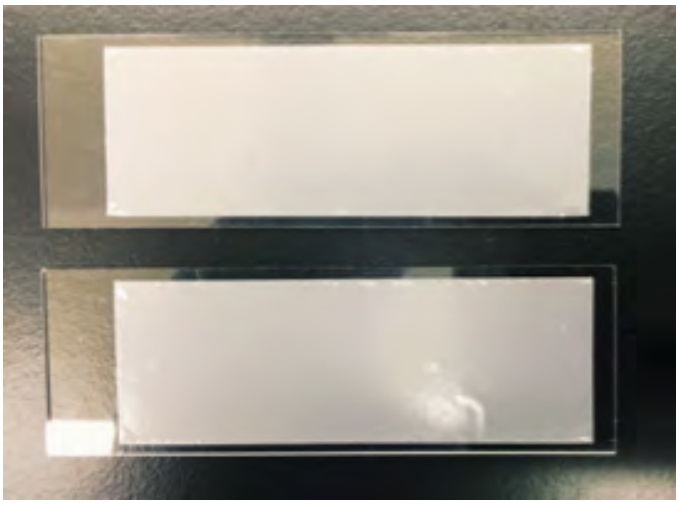
Seq-Well Protocol

Membrane Preparation

1. Carefully place a pre-cut (22 x 66 mm) polycarbonate membrane onto a glass slide using a gloved finger and tweezers to separate the membrane and paper.

   Note 1: Make certain the shiny side of the polycarbonate membrane is facing up to be in contact with the oxygen plasma and eventually the surface of the array.

   Note 2: Discard any membranes that have creases or other large-scale imperfections.



2. Place membranes onto a shelf in the plasma cleaner.

   Note 1: Shelves are not provided, but any piece of glass will do.

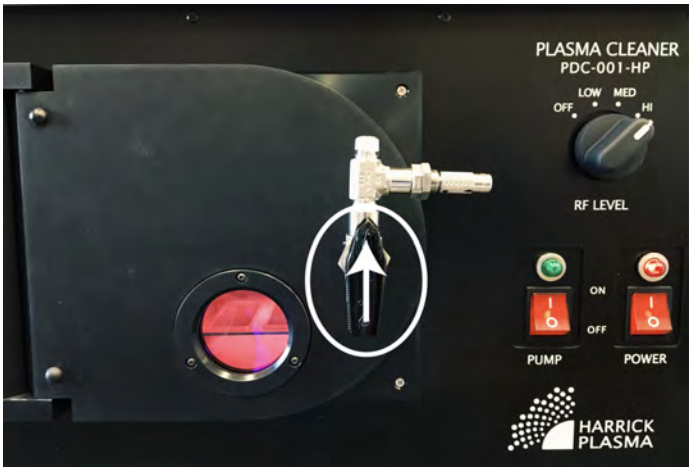
   Note 2 (optional): If you have two shelves, place membranes on the bottom shelf to reduce risk of them flying after vacuum is removed.

3. Close the plasma cleaner door, then turn on the main power and pump switch. To

form a vacuum, ensure that the 3-way valve lever is at the 9:00 position as shown below and that the door is completely shut.

4. Allow vacuum to form for 2-3 minutes. Once the vacuum has formed, simultaneously turn the valve to 12:00 while turning the power to the Hi setting (shown below).

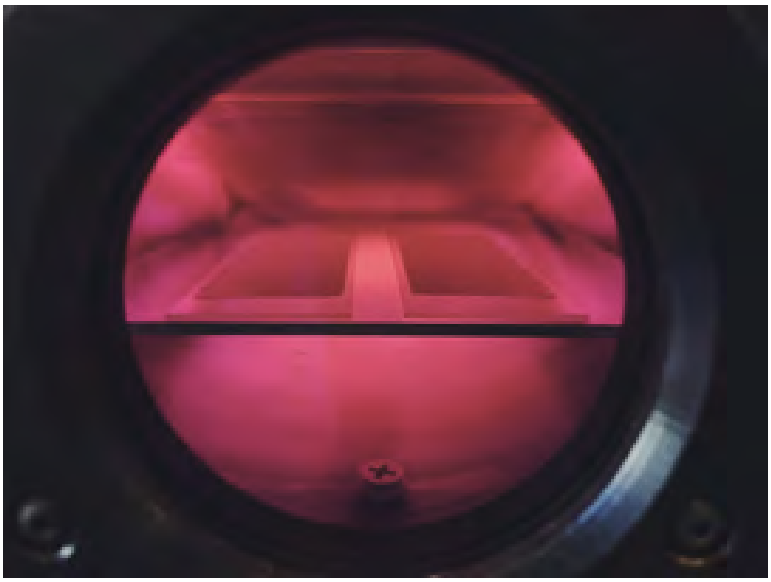
   Note: The plasma should be a bright pink. If not, adjust the air valve to increase or decrease the amount of oxygen entering the chamber.



5. Treat membranes with plasma for 5-7 minutes.

   Note: We treat membranes for 7 minutes, but treatment times can vary.

Experiment Notes



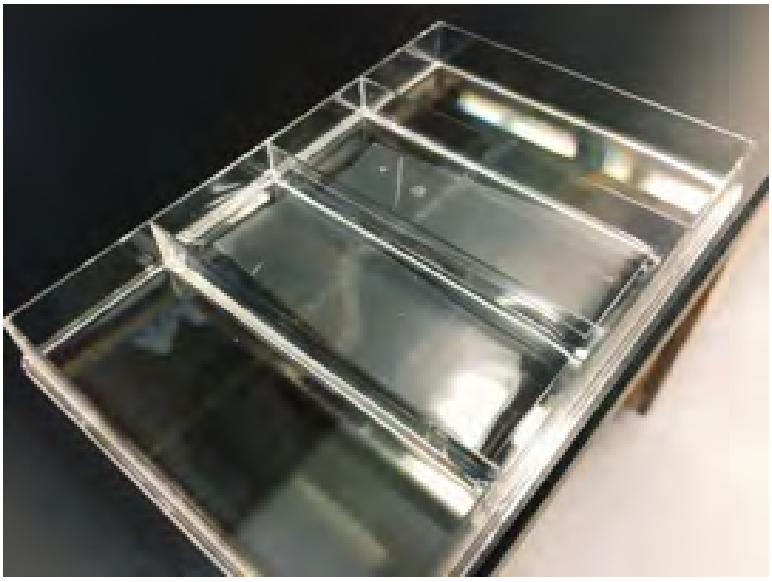
6. **Critical** – After treatment, in the following order: **(1)** turn the RF level valve from HIGH to OFF, **(2)** turn the air valve from the 12:00 position to the 9:00 position, and **(3)** then turn off the power followed by turning off the vacuum. Then *slowly open* the valve until air can be heard entering the chamber (approximate valve position shown below). Leave until door opens (~5 min).

7. Remove slides (with membranes) from the oven and transfer to a 4-well dish.

Note 1: If membranes have slightly folded over, slowly flip the membrane back using needle nose tweezers.

Note 2: If membranes have blown off the slide entirely, repeat above procedure to ensure you know which side was exposed to plasma.

8. Using a P1000 pipette, gently hydrate one end of the membrane with a single drop of 1xPBS so that it adheres to the slide before dispensing the entire volume. Once the membrane is hydrated, continuing add 1xPBS until you reach 5 mL (use either a serological pipette or P1000 pipette to complete hydration).



9. Remove any air bubbles underneath the membrane using wafer forceps or a pipette tip.
10. Membranes are now functionalized and ready for use.

    Note 1: Membranes solvated with 1xPBS should be used within **48 hours.**

    Note 2: If transporting solvated membranes (e.g. between buildings), remove all but ~1 mL of 1xPBS to prevent membranes from flipping within the dish.

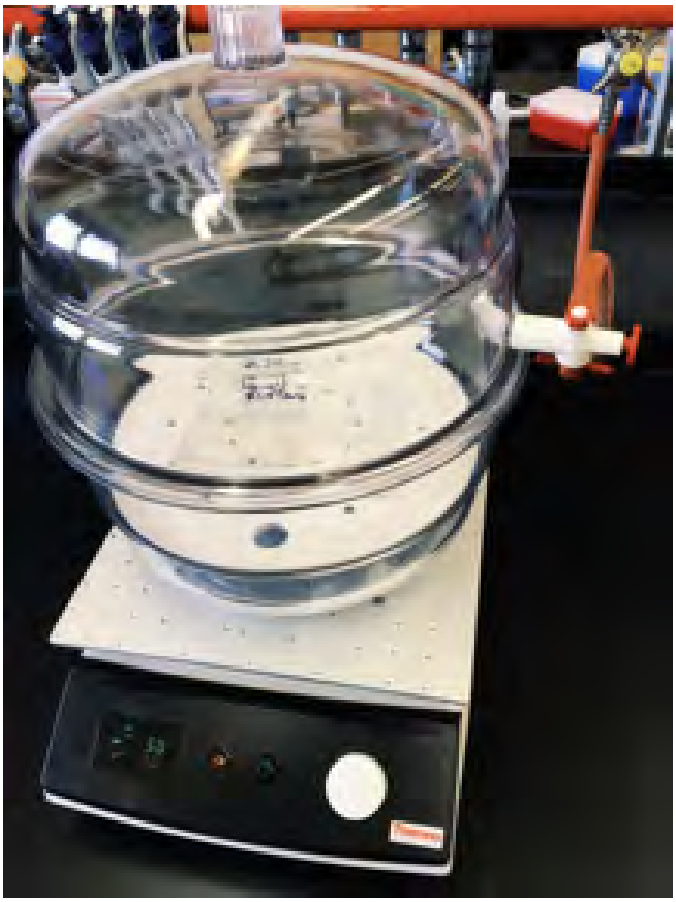
    Note 3: Alternatively, membranes initially solvated in 1xPBS can be dried and stored for 4 weeks at room temperature. To dry them out, carefully remove membranes, keeping them on their glass slides, from the 1xPBS solution, transfer the membranes to the benchtop, cover them with a tip box, and let them dry for 15-20 minutes. As the membranes dry they'll become opaque which is normal.

    Note 4: Before use the membranes should be rehydrated with 5 mL of 1xPBS. Drying out membranes is helpful when traveling or when running seq-well in a laboratory without access to a plasma cleaner.

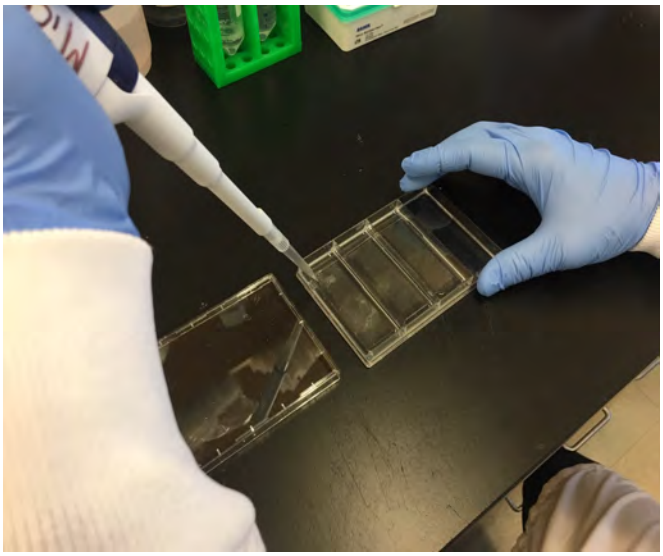
Experiment Notes

Bead Loading

1. Aspirate storage solution and solvate each array with 5 mL of bead loading buffer (BLB; **See Sub-Appendix D: Buffers Guide**).
2. Place array(s) under vacuum with rotation (50 RPM) for 10 minutes to remove air bubbles in wells. **Note:** Rotation is optional



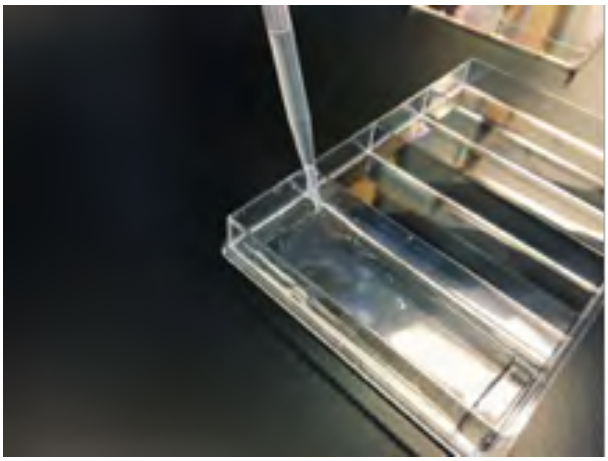
3. Aliquot ~110,000 beads from stock into a 1.5 mL tube and spin on a tabletop centrifuge for 15 seconds to form a pellet.
4. Aspirate storage buffer and wash beads twice in 500 uL of BLB.
5. Pellet beads, aspirate BLB, and resuspend beads in 200 uL of BLB.
   Note: For each array, it's recommended to load ~110,000 beads.
6. Before loading beads, thoroughly aspirate BLB from the dish containing the array(s), being careful not to aspirate or dry the PDMS surface of the array(s).
7. Using a P200 pipette, apply 200 uL containing 110,000 beads, in a drop-wise fashion, to the surface of each array (see image below and Bead Loading Diagram on page 10).

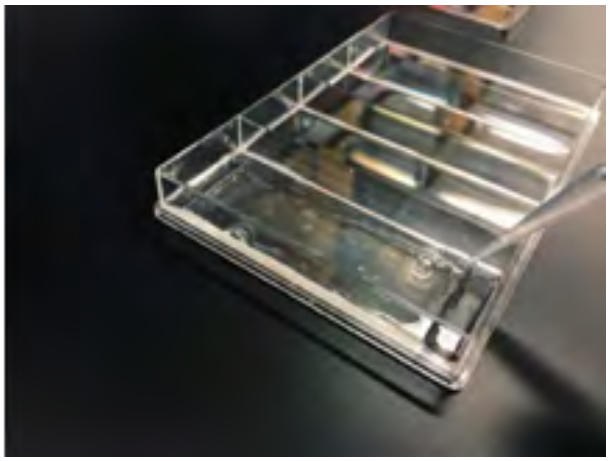


8. Allow the arrays to sit for 5 minutes, rocking them intermittently in the x & y direction.

   Pro-Tip: This step can be extended to 10 minutes to allow the beads more time to settle. However, make sure to monitor the surface of the array so that it doesn't dry out.

9. Thoroughly wash array(s) to remove excess beads from the surface. For each wash:
   1. Position each array so that it sits in the center of the 4-well dish.
   2. Dispense 500 uL of BLB in the upper right corner of each array and 500 uL in the bottom right corner of each array. Be careful not to directly pipette onto the microwells, as it can dislodge beads.
   3. Using wafer forceps or a pipette tip, push each array against the left side of the 4-well dish to create a capillary flow; this will help remove beads from the surface.
   4. Aspirate the liquid, reposition each array, and repeat on the other side.





10. Repeat step 9 as necessary. Periodically examine the array(s) under microscope to confirm that no loose beads are present on the surface, as this will interfere with

membrane attachment. Usually it takes 4 washes/side to thoroughly remove excess beads (this depends on your original loading density).

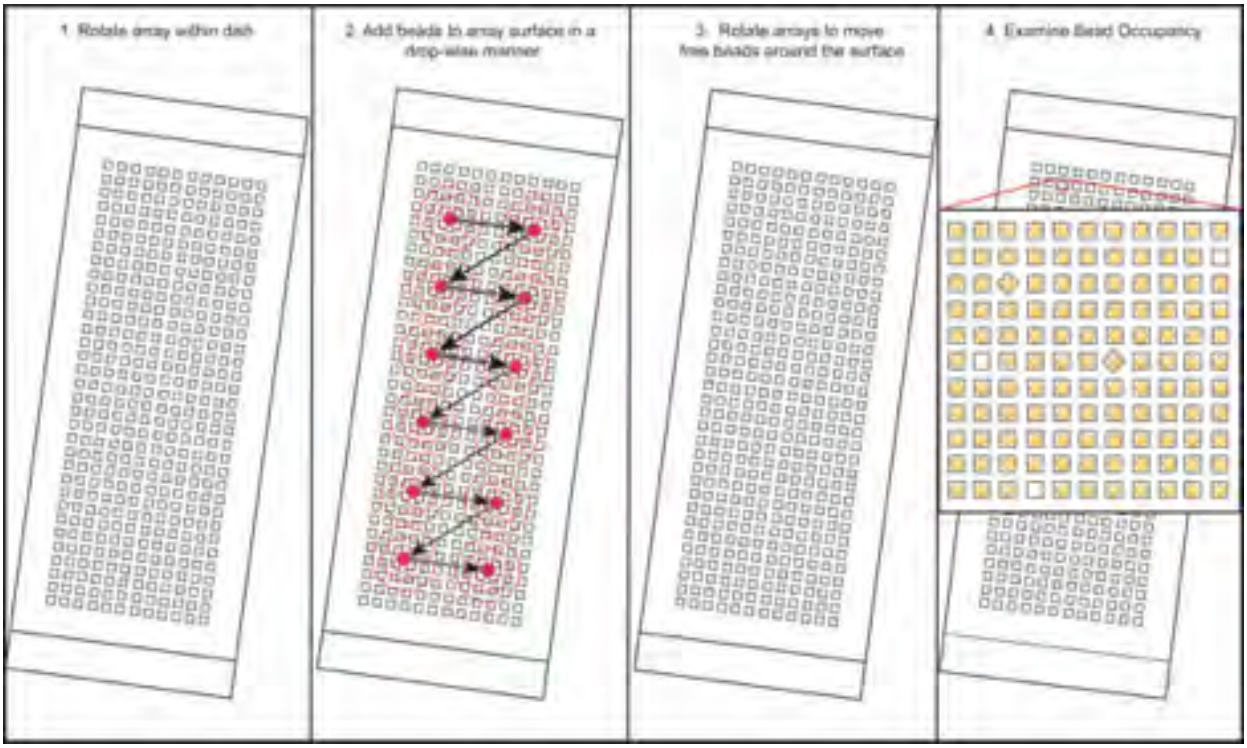
11. Once excess beads have been removed from the surface, solvate each array with 5 mL of BLB and proceed to cell loading.

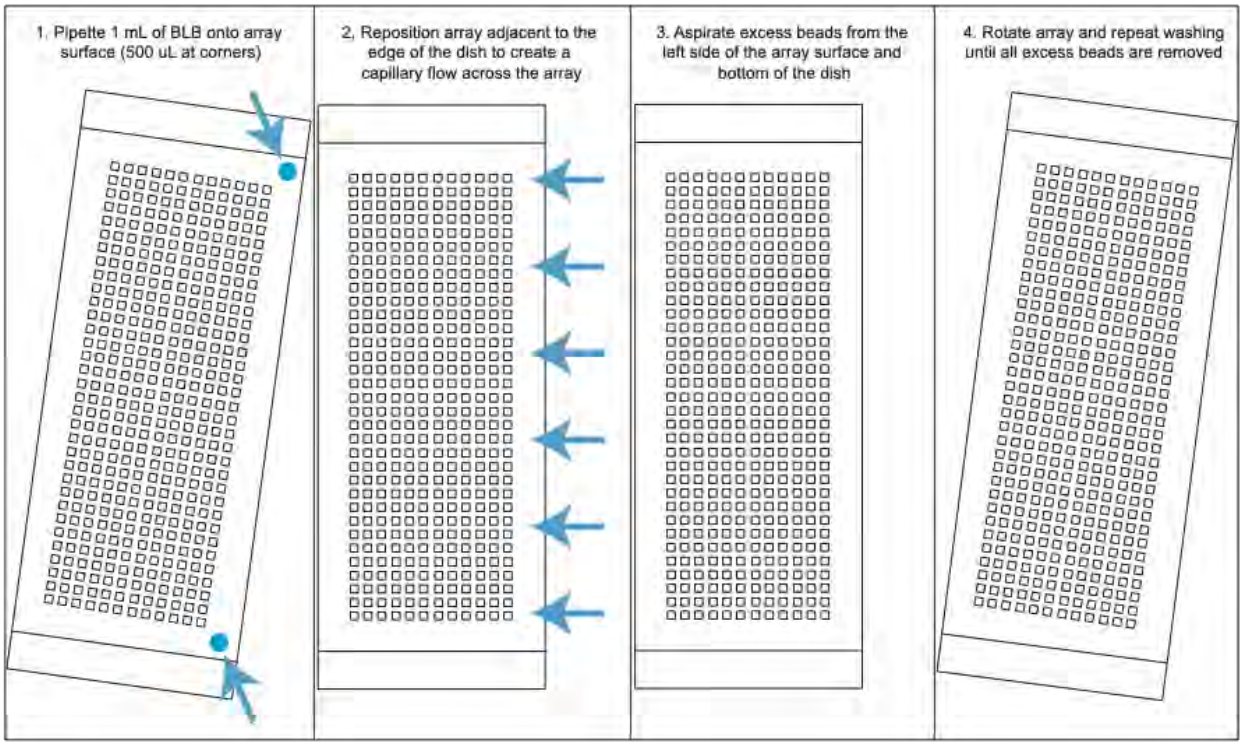
    Notes:

    1. If continuing to cell loading immediately (i.e., within 6 hours), loaded arrays should be stored in 5 mL of BLB.
    2. If you are not going to use the arrays on the day they're loaded, remove the BLB buffer, rinse the arrays once with 5 mL of 1xPBS, and then solvate the arrays with 5 mL of quenching buffer. Arrays can be stored in quenching buffer for 10 days (**See Sub-Appendix D: Buffers Guide**).

Experiment Notes

Bead Loading Diagrams





Cell Loading

1. At this point, your array should be loaded with beads and sitting in 5 mL of BLB.
2. Obtain the cell or tissue sample and prepare a single cell suspension using an optimized protocol for tissue dissociation.
3. While preparing your single-cell suspension, aspirate the BLB from each array (or quenching buffer) and rinse the array twice in 5 mL of 1xPBS to bring the solution in the four-well dish to physiological pH.
4. After the second wash, aspirate the 1xPBS and soak the loaded array in 5 mL of RPMI +10% (RP-10) FBS for 5 minutes.

   Note 1: This step is performed to mitigate non-specific adhesion of cells to primary amines on the top surface of the array.

   Note 2: Any supplemented media can be used in place of RP-10.
5. After obtaining a single-cell suspension, count cells using a hemocytometer and make a new solution of 10,000-15,000 cells in 200 uL of RP-10.

   Note 1: You can use your preferred media for prepping the cell loading solution.

   Note 2: Be sure to not use automated cell counters, particularly following tissue dissociation. This can provide an inaccurate cell count, compromising the experiment.
6. Thoroughly aspirate the RP-10/supplemented media (to ensure the array will not move during cell loading).
7. Center your array in the well and then apply the cell loading solution onto the surface in a dropwise fashion (similar to how beads were applied in the previous section).
8. Allow cells to settle for 10 minutes, intermittently rock the array in the x & y direction.
9. Wash array 4x with 5 mL of 1xPBS to remove the serum. For each wash, gently rock the array in the x & y direction, and then aspirate the 1xPBS. Once you have aspirated the 1xPBS out of the dish, gently tilt the 4-well dish toward you and aspirate directly off the bottom border of the array; this will help to completely remove the excess serum on the surface of the array

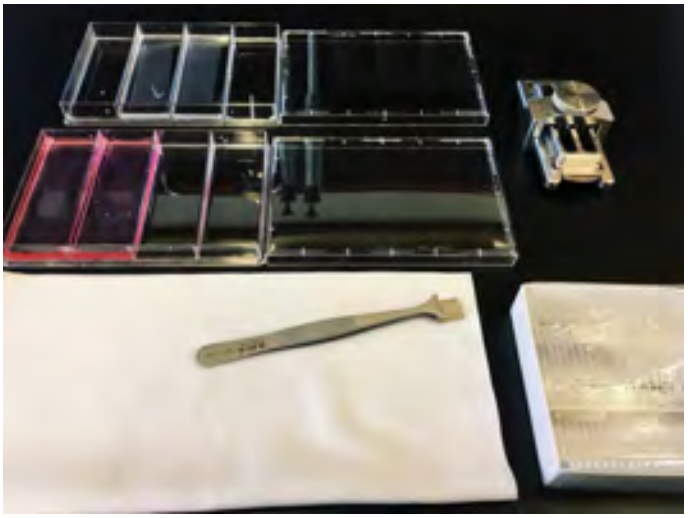
   Note: These washes are critical to remove excess serum which can interfere with successful membrane attachment.
10. Aspirate the final 1xPBS wash and replace with 5 mL of RPMI media ***without*** FBS.

    Note: You can use any media here as long as it **does not** contain serum.

Experiment Notes

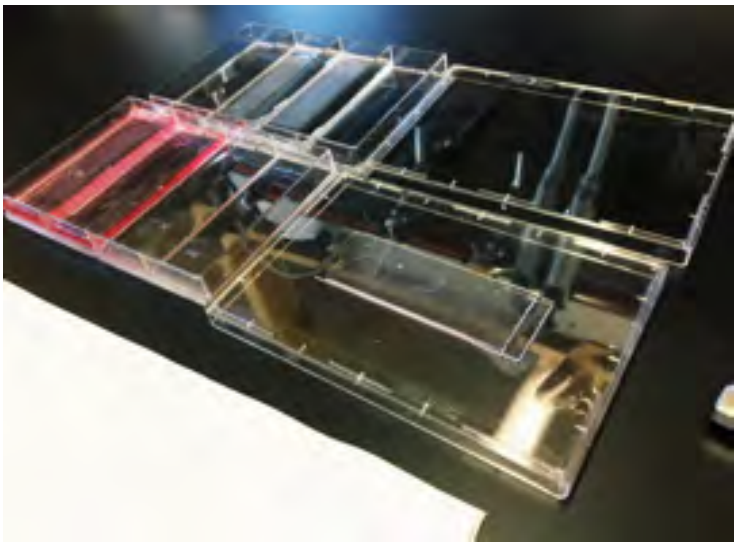
Membrane Sealing

1. Gather the following materials before sealing the array(s):

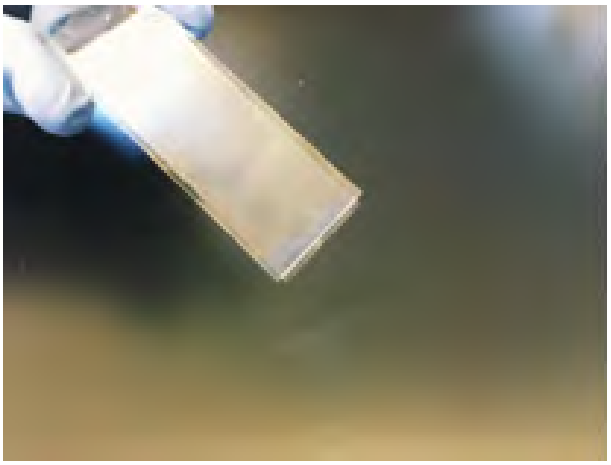
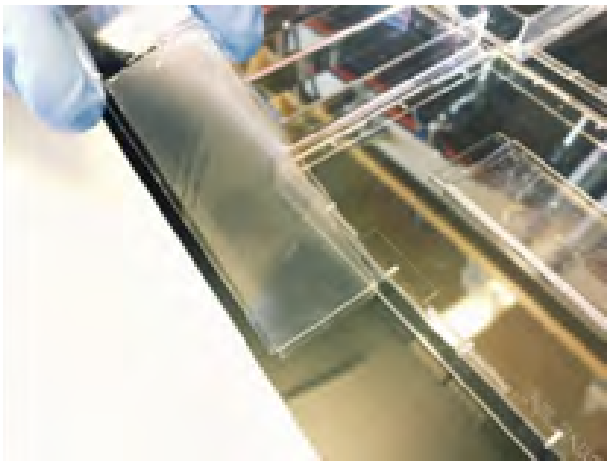


- Array loaded with beads and cells (See Bead/Cell Loading)
- Pre-treated membrane (See Membrane Preparation)
- Wafer forceps (or P1000 pipette tip)
- Paper towels
- Agilent clamp
- Clean microscope slides

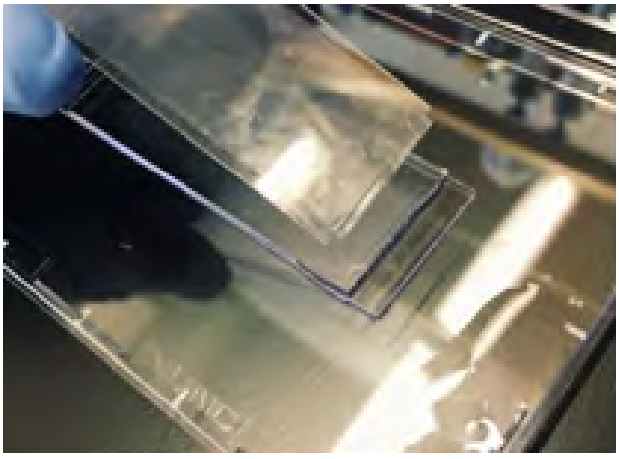
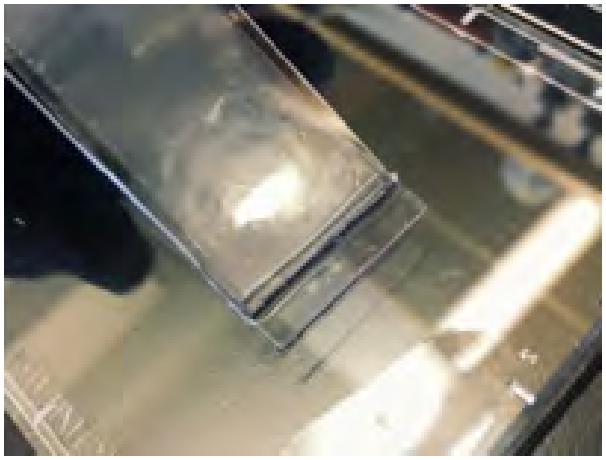
2. Use the wafer forceps to transfer the array from media to the lid of a 4-well dish, being careful to ensure that the array is not tilted.



3. Once the array is positioned on the lid of a 4-well dish, carefully aspirate excess liquid from around the edge of the array and the exposed surface of the glass slide. (**Note**: Be careful not to aspirate directly from the PDMS surface).
4. Using wafer forceps or a pipette tip, remove a pre-treated membrane from the 4-well dish.
5. Gently dab away moisture from the glass slide on the paper towel until the membrane does not spontaneously change position on the glass slide.
6. Carefully position the membrane on the center of the microscope slide, leaving a small membrane overhang (2-3 mm) beyond the edge of slide.

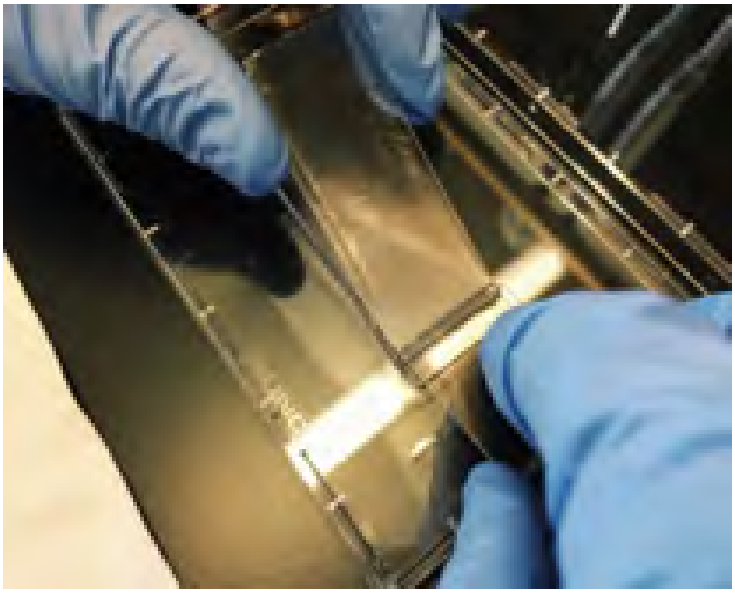



7. Holding the membrane in your left hand, invert the microscope slide so that the treated surface of the membrane is facing down.

Experiment Notes

8. Place the overhang of the membrane in contact with the PDMS surface of the array just beyond the boundary of the microwells.



9. Using a clean slide held in your right hand, firmly hold down the overhang of the membrane against the PDMS surface of the array.
10. **Critical Step:** While maintaining pressure with your right hand to hold the membrane in place, gently apply the membrane.

    Note 1: For optimal results, use only the weight of the slide to apply the membrane with the left hand.

    Note 2: Attempts to manually seal the microwell device using excess pressure result in a 'squeegee' effect, effectively removing moisture from the membrane while fixing membrane creases in place.

    Note 3: As you apply the membrane you should see a fluid interface form and expand as direct, uniform contact between the slide and the array will naturally remove some of the media as the membrane is applied.

    Note 4: You can use either your left or right hand for membrane-sealing (most people use their dominant hand to apply the membrane). Please practice this step before the actual experiment to figure out which hand you're most comfortable with.
11. After applying the membrane, carefully pry the array and membrane from the surface of the lid and transfer to an Agilent clamp.
12. After transferring the sealed array to the clamp, place a glass slide on top of the sealed array.

13. Close the clamp and tighten to the point of resistance, then place it in a 37C incubator for 30-40 minutes.

    Note: This time is flexible and depends on the incubator. If you want to decrease this incubation time, please optimize on cell lines before proceeding with precious samples.

14. Repeat membrane-sealing protocol procedure if running multiple arrays.

Experiment Notes

Cell Lysis & Hybridization

1. Remove the clamp from the incubator, and then remove the array from the Agilent clamp. (Note: At this point, the glass slide will be attached to the array and membrane).
2. Submerge the array, with top slide still attached, in 5 mL of complete lysis buffer (**See Sub-Appendix D: Buffers Guide**).
3. Gently rock the array in lysis buffer until the top glass slide spontaneously detaches.

   Note 1: Do not pry the top slide off as this can reverse membrane sealing. The time necessary for detachment of the top slide varies (10 seconds – 10 minutes).

   Note 2: If the top slide does not release after 10 minutes, gently pry the top slide off using wafer forceps or a pipette tip. Just be careful.
4. Once the top slide has detached, place the arrays on a horizontal rotator for 20 minutes at 50-60 rpm.
5. After 20 minutes, remove the lysis buffer and wash each array with 5 mL of hybridization Buffer (**See Sub-Appendix C: Buffers Guide**).

   Note 1: Use a separate waste container for lysis buffer because guanidine thiocyanate can react with bleach in TC traps to create cyanide gas.

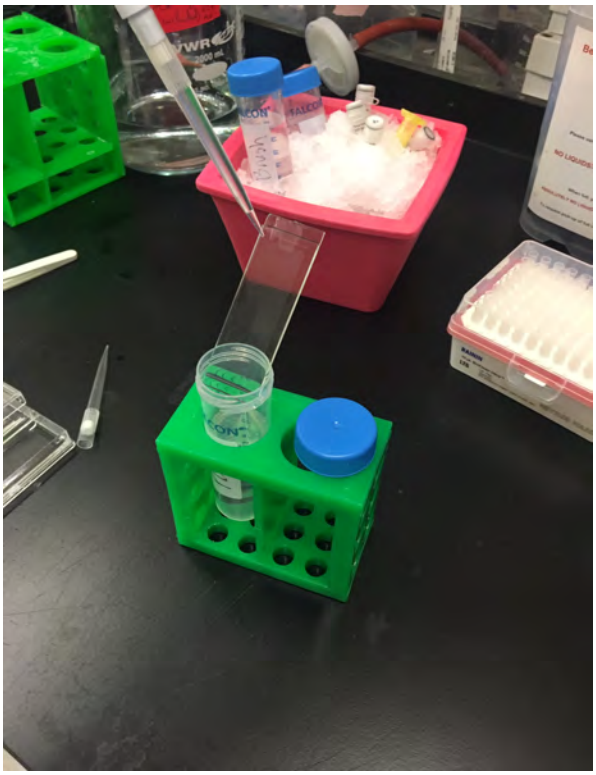
   Note 2: The hybridization buffer used to wash the array post-lysis may contain trace amounts of guanidine thiocyanate and should, therefore, be disposed of in the lysis buffer waste container.
6. Aspirate hybridization buffer and add another 5 mL of hybridization buffer to each array and rotate for 40 minutes at 50-60 rpm.
7. While the arrays are rocking in hybridization buffer, prepare RT master mix. (See **Reverse Transcription & Exonuclease Digestion**)

Experiment Notes

Bead Removal Method 1

1. After the arrays have rocked in hybridization buffer for 40 minutes, carefully peel back each membrane using fine-tipped tweezers.
2. Place array into a 50 mL conical containing 30-40 mL of Wash 1 solution.
3. Holding the array above the 50mL conical (shown below), repeatedly dispense approximately 1 mL of Wash 1 solution from the conical across the surface of the array to dislodge beads (**See Sub-Appendix D: Buffers Guide**).

   Note: Vigorously dispense Wash 1 buffer to remove beads.
4. Repeat these 10 times, periodically checking to see if beads are dislodging.



5. After repeatedly rinsing the array from top to bottom, use a clean glass slide to ***gently*** scrape the array to remove any beads that remain in the array.

   Note: At this point it is possible to visually inspect the array to assess bead removal.
6. Once you are satisfied with bead removal, place the empty array back in the 4-well disk, cap the 50 mL conical, and pellet beads for 5 minutes at 1000xg.

   Note 1: You can visually inspect the success of your bead removal by looking at the arrays under a light microscope. *(continues on the next page)*

   Note 2: Where possible, use a swinging bucket centrifuge to collect beads. The use of a fixed-rotor centrifuge can lead to the formation of a bead pellet on the elbow rather than the bottom of the conical tube, which can lead to inefficient recovery.

7. After centrifugation, aspirate all but ~1 mL of excess Wash Buffer, collect the beads using a P1000 pipette, and transfer beads suspended in wash buffer to a separate 1.5 mL eppendorf tube for each array.

Reverse Transcription & Exonuclease Digestion

Reverse Transcription (RT)

1. Prepare the following RT mastermix during the hybridization step:

40 uL	H_2O
40 uL	Maxima 5X RT Buffer
80 uL	30% PEG8K
20 uL	10 mM dNTPs (Clontech)
5 uL	RNase Inhibitor (Lucigen)
5 uL	100 uM Template Switch Oligo
10 uL	Maxima H-RT

   Note: Add the Maxima H-RT enzyme to the mastermix immediately before adding to beads.

2. Centrifuge eppendorf tubes containing collected beads for 1 minute at 1000xg.
3. Remove supernatant and resuspend in 250 uL of 1X Maxima RT Buffer and centrifuge beads for 1 minute at 1000xg.
4. Aspirate 1X Maxima RT Buffer and resuspend beads in 200 uL of the RT mastermix.
5. Incubate at room temperature for 30 minutes with end-over-end rotation. After 30 minutes, incubate at 52C for 90 minutes with end-over-end rotation.

   Note: The reverse transcription reaction can proceed overnight, if necessary.

7. Following the RT reaction, wash beads once with 500 uL of TE-SDS, and twice with 500 uL of TE-Tween (TE-TW). **Following Reverse Transcription, beads can be stored at 4C in TE-TW.**

Exonuclease I Treatment

1. Prepare the following Exonuclease I Mix:

20 uL	10x ExoI Buffer
170 uL	H_2O
10 uL	ExoI

2. Centrifuge beads for 1 minute at 1000xg and aspirate the TE-TW solution.

3. Resuspend in 500 uL of 10 mM Tris-HCl pH 8.0.
4. Centrifuge beads again, remove supernatant and resuspend beads in 200 uL of exonuclease I mix.
5. Incubate at 37C for 50 minutes with end-over-end rotation.
6. Wash the beads once with 500 uL of TE-SDS, twice with 500 uL TE-TW.

Beads can be stored at 4C in TE-TW.

Second Strand Synthesis & PCR

Second Strand Synthesis

(*Beginning after 2nd wash of TE-TW after Exo treatment)*

1. Prepare the following 2nd strand synthesis mix:

40 uL	Maxima 5X RT Buffer
80 uL	30% PEG8000
20 uL	10 mM dNTPs (Clontech)
2 uL	1 mM dN-SMRT oligo
5 uL	Klenow Enzyme
53 uL	H_2O

Note: Add the Klenow enzyme immediately before adding to beads.

2. After aspiration of 2nd TE-TW wash, resuspend beads in 500 uL 0.1 M NaOH.

Note: Make the 0.1 M NaOH solution fresh for each experiment.

3. Rotate tube for 5 min at room temp, then spin (800xg for 1 minute) and aspirate supernatant.
4. Wash once with 500 uL of TE-TW, and once with 500 uL 1xTE
5. Resuspend beads in 200 uL 2nd strand synthesis reaction and rotate end-over-end at 37C for 1 hr.
6. Wash beads twice with 500 uL TE-Tween and once with 500 uL TE
7. Proceed directly with the PCR protocol.

PCR (Whole Transcriptome Amplification (WTA))

1. Prepare the following PCR mastermix:

25 uL	2X KAPA HiFi Hotstart Readymix
14.6 uL	H_2O
0.4 uL	100 uM SMART PCR Primer
40 uL	per reaction

2. Wash beads once with 500 uL of water, pellet beads, remove supernatant and resuspend in 500 uL of water.

   Note 1: If you do not want to count the beads then after the 500 uL water wash in step 2, resuspend the beads in 240 uL of water and proceed to step 6.

   Note 2: If you choose this path, prepare mastermix for 24 PCR reactions for each array being processed.

3. Mix well (do not vortex) to evenly resuspend beads and transfer 20 uL of beads to a separate 1.5 mL tube to count the beads.

   Note: Don't vortex beads as this can result in bead fragmentation.

4. Pellet the small aliquot of beads, aspirate the supernatant, and resuspend in 20 uL of bead counting solution (10% PEG, 2.5 M NaCl).

   Note: The bead counting solution aids in even dispersion of beads across a hemocytometer.

5. Count the beads using a hemocytometer.
6. Add 40 uL of PCR mastermix per reaction to 96-well plate.
7. Add 1,500 – 2,000 beads per reaction in 10 uL of water for a total volume of 50 uL per PCR reaction, making certain to PCR the entire array.
8. Use the following cycling conditions to perform whole-transcriptome amplification:

   Start:

95C	3 minutes

   4 Cycles:

98C	20 seconds
65C	45 seconds
72C	3 minutes

   9-12 Cycles:

98C	20 seconds
67C	20 seconds
72C	3 minutes

   Final Extension:

72C	5 minutes
4 C	Infinite hold

   Note: The total number of PCR cycles necessary for amplification depends on the cell type used.

   - 13 cycles are optimal for cell lines or larger cells (e.g. macrophages)

- 16 cycles are optimal for primary cells**Purification of PCR products and analysis on the BioAnalyzer or Agilent TapeStation**

1. Pool PCR products from between 6 and 8 PCR reactions in a 1.5 mL microcentrifuge tube so that you have 10-12,000 beads/1.5 mL microcentrifuge tube.
2. Purify PCR products using Ampure SPRI beads and the following protocol:
   Note: Please refer to the Ampure SPRI bead official protocol for more details.
   A. Spri at 0.6x volumetric ratio.
   B. Allow the tubes to sit on the tube-rack off the magnet for 5 minutes, and then place the rack on the magnet for 5 minutes.
   C. Perform 3 washes with 80% ethanol (**Note:** At each wash step rotate each tube 180 degrees 6 times to allow beads)
   to pass through the ethanol solution to the opposite side of the tube.
   D. After the third wash, remove the 80% ethanol wash solution. Further, use a P200 with fresh tips to remove any residual ethanol and allow beads to dry for 10-15 minutes. (**Note**: Beads will have a cracked appearance once dry). Remove the rack from the magnet, elute dried beads in 100 uL, place the rack on the magnet and then transfer the 100 uL supernatant which contains eluted DNA to a new 1.5 mL microcentrifuge tube or 96-well plate.
   E. Spri the 100 uL at 1.0x volumetric ratio and repeat steps b and c
   F. After the third wash, allow the beads to dry for 15 minutes, remove the rack from the magnetic, elute the beads in 15 uL, place the rack back on the magnet and then transfer the 15 uL to a new 1.5 mL microcentrifuge tube or 96-well plate.
3. Run a BioAnalyzer High Sensitivity Chip or Agilent D5000 High Sensitivity Screentape according to the manufacturer's instructions. Use 2 uL of the purified cDNA sample as input (**Note:** Your WTA library should be fairly smooth, with an average bp size of 0.7-2 kbps).
4. Proceed to library preparation or store the WTA product at 4C (short-term) or -20C (long-term).

Library Preparation

Tagmentation of cDNA with Nextera XT

1. Ensure your thermocyclers are setup for Tagmentation (step 5) & PCR (step 9).
2. For each sample, combine 1000 pg of purified cDNA with water in a total volume of 5 uL. It's ideal to dilute your PCR product in a separate tube/plate so that you can add 5 uL of that for tagmentation.

   Example: For 1000 pg reactions, dilute PCR product, in a new plate, to 200 pg/uL, then you can add 5 uL of this to a reaction tube for a 1000 pg reaction.

   Note 1: We typically perform Nextera reactions in duplicate for WTA product from each pool of 6-8 PCR reactions. For example, if you recover 3 pools/array, you would run a total of 6 nextera reactions.

   Note 2: These volumes can be reduced by half to reduce reagent costs, if desired.
3. To each tube, add 11 uL of Nextera TD buffer, then 4 uL of ATM buffer (the total volume of the reaction is now 20 uL).
4. Mix by pipetting ~5 times. Centrifuge plate at 1000x g for 10-15 seconds.
5. Incubate at 55C for 5 minutes.
6. Add 5 uL of Neutralization Buffer. Mix by pipetting ~5 times. **Note:** Bubbles are normal.
7. Incubate at room temperature for 5 minutes.
8. Add to each PCR tube:

15 uL	Nextera PCR mix
8 uL	H_2O
1 uL	10 uM New-P5-SMART PCR hybrid oligo
1 uL	10uM Nextera N7XX oligo

(continues on the next page)

9. After sealing and centrifuging (1 minute at 1000xg) the PCR plate, run the following PCR program:

Start:

72C	3 minutes
95C	30 seconds

12 cycles:

95C	10 seconds
55C	30 seconds
72C	30 seconds

Final Extension:

72C	5 minutes
4C	Infinite hold

Purification of PCR products and analysis on the BioAnalyzer or Agilent TapeStation

1. If you performed Nextera reactions in duplicates, please pool duplicates before proceeding with step 2. If you ran a single Nextera reaction for each pooled WTA, proceed directly to step 2.
2. Purify PCR products using Ampure SPRI beads and the following protocol:

   Note: Please refer to the Ampure SPRI bead official protocol for more details.

   A. Spri at 0.6x volumetric ratio.

   B. Allow the tubes to sit on the tube-rack off the magnet for 5 minutes, and then place the rack on the magnet for 5 minutes.

   C. Perform 3 washes with 80% ethanol (**Note:** At each wash step rotate each tube 180 degrees 6 times to allow beads).
   to pass through the ethanol solution to the opposite side of the tube.

   D. After the third wash, remove the 80% ethanol wash solution. Further, use a P200 with fresh tips to remove any residual ethanol and allow beads to dry for 10-15 minutes. (**Note**: Beads will have a cracked appearance once dry).
   (continues on the next page)
   Remove the rack from the magnet, elute dried beads in 100 uL, place the rack on the magnet and then transfer the 100 uL supernatant which contains eluted DNA to a new 1.5 mL microcentrifuge tube or 96-well plate.

   E. Spri the 100 uL at 1.0x volumetric ratio and repeat steps b and c

F. After the third wash, allow the beads to dry for 15 minutes, remove the rack from the magnetic, elute the beads in 15 uL, place the rack back on the magnet and then transfer the 15 uL to a new 1.5 mL microcentrifuge tube or 96-well plate.

3. Run a BioAnalyzer High Sensitivity Chip or Agilent D1000 High Sensitivity Screentape according to the manufacturer's instructions.
   - Use 1 uL of the purified cDNA sample as input.
   - Your tagmented library should be fairly smooth, with an average bp size of 400-800 bp.
   - Smaller-sized libraries might have more polyA reads
   - Larger libraries may have lower sequence cluster density and cluster quality.

   Note: We have successfully sequenced libraries from 400-800bp.
5. Proceed to sequencing.

Sequencing

Once your sequencing library has passed the proper quality controls, you're ready to proceed to sequencing. For a detailed loading protocol, please consult the Illumina website for a step-by-step manual. (https://support.illumina.com/downloads.html)

NextSeq500 – Shalek Lab protocol

1. Make a 5 uL library pool at 4 nM as input for denaturation.
2. To this 5 uL library, add 5 uL of 0.2 N NaOH (make this solution fresh).
3. Flick to mix, then spin down and let tube sit for 5 minutes at room temperature.
4. After 5 minutes, add 5 uL of 0.2 M Tris-HCl pH 7.5.
5. Add 985 uL of HT1 Buffer to make a 1 mL, 20 pM library (solution 1).
6. In a new tube (solution 2), add 165 uL of solution 1 and dilute to 1.5 mL with HT1 buffer to make a 2.2 pM solution – this is the recommended loading concentration.

   Note: Optimal loading concentration is 1.8-2.5 pM
7. Follow Illumina's guide for loading a NextSeq500 Kit

Sequencing specifications for the MiSeq or NextSeq:

Read 1: 20 bp *

Read 2: 50 bp

Read 1 Index: 8 bp ← *only necessary if you are multiplexing samples*

Custom Read 1 primer

Sequencing specifications for the Nova-Seq:

Read 1: 20 bp *

Read 2: 50-80 bp

Read 1 Index: 8 bp

Read 2 Index: 8 bp (optional, but recommended)

Custom Read 1 primer

Note 1: If you're loading on a Nova-Seq you'll want to use dual-indexing to mitigate index switching.

Note 2: Read 1 can sometimes be 21 base pairs; this depends on the company and bead lot you are ordering from. Please consult with your bead provider to determine which read length to use.

NextSeq 500:

(http://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/nextseq/nextseq-custom-primers-guide-15057456-01.pdf)

(Follow Illumina's guide for custom primers)

MiSeq:

(http://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-system-custom-primers-guide-15041638-01.pdf)

Array Synthesis

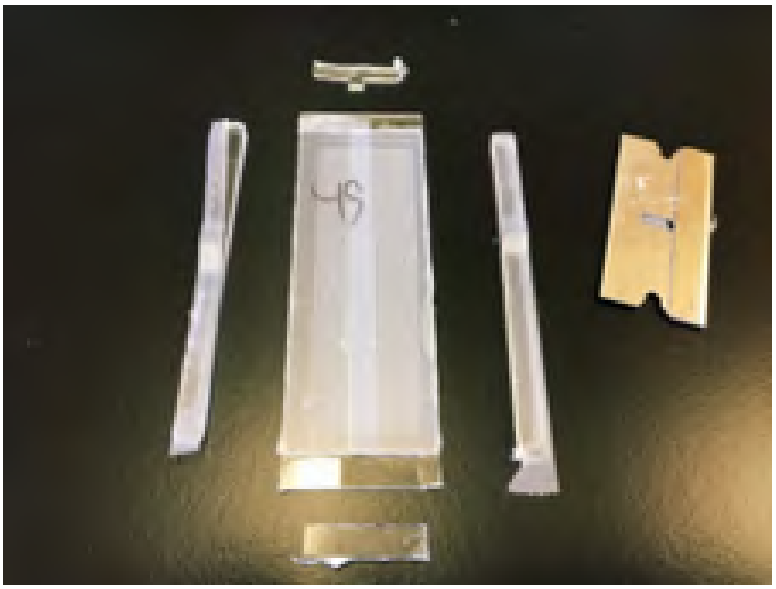
Synthesis Day 0: Pouring PDMS Arrays

1. Combine Sylgard crosslinker with Sylgard base at a 1:10 ratio and mix vigorously for 5 minutes to create a PDMS master mix.
2. Once mixing is complete, put your PDMS master mix under vacuum for 20 minutes to remove any air bubbles.
3. Use a 10 mL syringe to inject 6-10 mL of PDMS master mix into molds with mounted PDMS masters.
4. Incubate at 70C for 2.5 hours.

Synthesis Day 1: Array Functionalization Part 1

Note: For this section, make all solutions fresh!

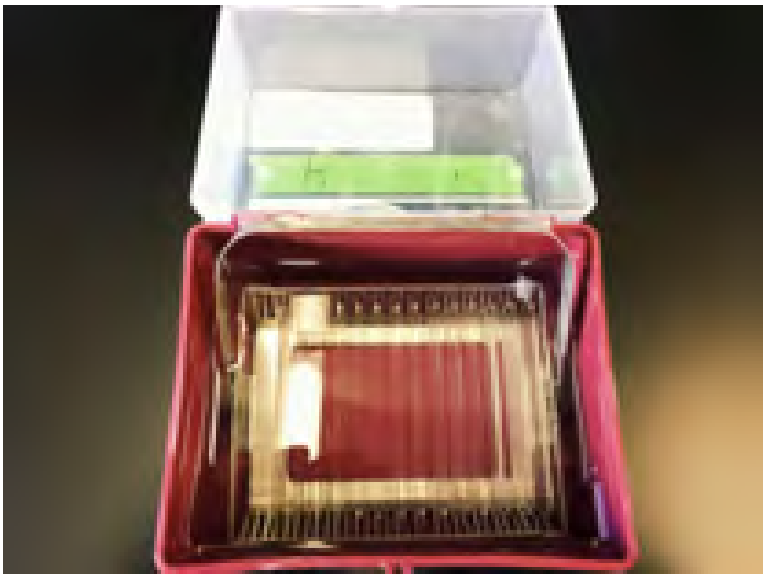
1. Remove excess PDMS from edges of the glass slide.



2. Use scotch tape to remove excess PDMS from the surface of the array and the glass slide.
3. Place clean arrays into a metal slide basket

(continues on the next page)

4. Rinse arrays in 100% ethanol for 5 minutes, then let dry at room temperature (RT) for 15 minutes.



5. Plasma treat arrays on high for 5-7 minutes.

   Note 1: Adjust the air valve so that the plasma is pink.

6. Following plasma treatment, immediately submerge arrays in 350 mL of 0.05% APTES in 95% ethanol for 15 minutes.
7. Spin dry arrays **(500 RPM** for 1 minute).

   Note: Our rotor model is TX-10000 75003017 (Thermo) with a rotor radius of 209 mm. 500 RPM on this instrument is ~ 60xg.

8. Incubate at room temperature for 10 minutes.
9. Submerge in 300 mL of acetone and rock until all bubbles are out of the wells; this

typically takes approximately 5 minutes.

10. Place in 350 mL of 0.2% PDITC/10% pyridine/90% DMF solution in a glass chamber (or polypropylene tip box) for 2 hours at room temperature.

    Note: While this is rocking, prepare your chitosan solution (See **Sub-Appendix D**)

11. After the PDITC soak, wash arrays briefly in two boxes of 300 mL DMF.

    Note: For each brief wash, simply dunk the arrays in the solution 5-10 times and then transfer to the new solution.

12. Dunk and wash the arrays in 300 mL of acetone.
13. Move to a fresh 350 mL of acetone and rock for 20 minutes.
14. Spin dry arrays (**500 RPM** for 1 minute).
15. Place arrays at 70C for 2 hours.
16. Remove from oven and let sit at room temperature for 20 minutes.
17. Submerge arrays in 350 mL of 0.2% chitosan solution (pH 6.0-6.1; See **Sub-Appendix D**) and incubate at 37C for 1.5 hours.
18. Wash arrays 4x in separate 300 mL distilled water baths.
19. Submerge in 350 mL of 20 ug/mL aspartic acid, 2 M NaCl, and 100 mM sodium carbonate solution (pH 10.0).
20. Place in vacuum chamber and apply house vacuum.

Note: You should see bubbles form indicating the solvation of wells.

21. Place vacuum chamber (still connected to house vacuum) on a rocker and rock (50-70 RPM) overnight at room temperature.

Synthesis Day 2: Array Functionalization – Part 2

1. The following morning, remove arrays from vacuum and rotate at 50-60 RPM for 3 hours at room temperature.
2. Place arrays at 4C and soak 24 hours before use.
   Note: Arrays can be stored in the aspartic acid solution for 3 months at 4C.

Master Mounting Protocol

1. Mix and degas PDMS in normal 1:10 ratio
2. While PDMS degases, use sandpaper to gently score back of silicon master and base plate to improve adhesion. Careful – silicon masters are brittle.
3. Wash back of master and base plate with 95% ethanol until no more dust is removed when wiping surface clean with paper towel.
4. Use gloved finger to spread vacuum grease on bottom of BasePlate2 around square holes where the nanowell arrays will be cast. You want a relatively thick layer, even on skinny parts between array holes, to make sure there is a seal between master and plate.
5. Carefully lower BasePlate2 onto the array side of the master making sure to not touch the array area with any of the greased surface. 4 array masters should fit into the 4 square holes. Gently slide plate against master to center the arrays.
6. Place Base Plate 1 on paper towels to catch PDMS running off plate.
7. Pour ~30 mL of mixed PDMS in center of Base Plate 1.
8. Place master/BasePlate2 sandwich on top of the PDMS.
9. Gently apply pressure in the center of the master while making circular motions to push PDMS out from between layers. You want to see PDMS coming out of all sides to ensure a complete coat.
10. Screw 6/32 screws into respective holes on base plate very gently. Too much pressure too fast may crack master. Do not fully tighten. Do your best to make screws even – look at width of crack between base plates on all sides and make equal.
11. Place both top plates on top.

12. Screw 10/24 screws into their holes just enough such that they catch. Again, do not fully tighten.
13. Place in 90C oven for 3 hours.
14. May need to do one dummy round of arrays to remove any PDMS or grease that got onto the nanowell features.

Buffers Guide

CellCover10

Reagents

- CellCover (Anacyte Art. No. 800-125)
- FBS (Thermo Fisher Scientific Cat. No. 10437028)
- Sodium Carbonate (Sigma Cat. No. 223530-500G)

Working Concentrations

- 10% FBS
- 100 mM Sodium Carbonate

Bead Loading Buffer

Reagents

- Sodium Carbonate (Sigma Cat No. 223530-500G)
- BSA (Sigma Cat No. A9418-100G)
- Water (Thermo Fisher Scientific Cat No. 10977023)

Quick Preparation Guide (50 mL)

1. 2.5 mL 2 M Sodium Carbonate
2. 42.5 mL H_2O
3. Add 5 mL BSA (100 mg/mL)
4. Titrate with glacial acetic acid to achieve a pH of 10.0

Working Concentrations

- 100 mM Sodium Carbonate
- 10% BSA

Complete Lysis Buffer

Reagents

- Pre-lysis buffer
- 10% Sarkosyl (Sigma Cat No. L7414)
- 100% 2-Mercaptoethanol (Sigma Cat No. M3148-25ML)

(continues on the next page)

Quick Preparation Guide (50 mL)

1. 47.25 mL Pre-Lysis Buffer
2. 250 uL 10% Sarkosyl
3. 500 uL BME

Working Concentrations

- 5 M Guanidine Thiocyanate
- 1 mM EDTA
- 0.50% Sarkosyl
- 1.0% BME

Hybridization Buffer

Reagents

- 5 M NaCl (Thermo Fisher Scientific Cat No. 24740011)
- 1x PBS (Thermo Fisher Scientific Cat No. 10010023)
- 8% (v/v) PEG8000 (Sigma Cat No. 83271-500ML-F)

Quick Preparation Guide (50 mL)

1. 20 mL 5 M NaCl
2. 26 mL of PBS
3. 4 mL PEG8000

Working Concentrations

- 2 M NaCl

Wash Buffer

Reagents

- 5 M NaCl (Thermo Fisher Scientific Cat No. 24740011)
- 1 M $MgCl_2$ (Sigma Cat No.63069-100ML)
- 1 M Tris-HCl pH 8.0 (Thermo Fisher Scientific Cat No. 15568025)
- Water (Thermo Fisher Scientific Cat No. 10977023)
- 8% (v/v) PEG8000 (Sigma Cat No. 83271-500ML-F)

(continues on the next page)

Quick Preparation Guide (50 mL)

1. 20 mL 5 M NaCl
2. 150 uL 1 M $MgCl_2$
3. 1 mL 1 M Tris-HCl pH 8.0
4. 24.85 mL H_2O
5. 4 mL PEG8000

Working Concentrations

- 2 M NaCl
- 3 mM $MgCl_2$
- 20 mM Tris-HCl pH 8.0

Array Quenching Buffers

Reagents

- Sodium Carbonate (Sigma Cat No. 223530-500G)
- 1 M Tris-HCl pH 8.0 (Thermo Fisher Scientific Cat No. 15568025)
- Water (Thermo Fisher Scientific Cat No. 10977023)

Quick Preparation Guide (50 mL)

1. 2.5 mL 2 M Sodium Carbonate
2. 500 uL 1 M Tris-HCl pH 8.0
3. 47 mL H_2O

Working Concentrations

- 100 mM Sodium Carbonate
- 10 mM Tris-HCl pH 8.0

0.2% Chitosan Solution

Reagents

- Chitosan (Sigma Cat No. C3646-100G)
- Water (Thermo Fisher Scientific Cat No. 10977023)

(continues on the next page)

Quick Preparation Guide

1. Add 1 gram of chitosan to 500 mL of DI water
2. Autoclave solution (40 minutes sterilization, 20 minutes dry)
3. Allow chitosan solution to come to room temperature, and then add 2-3 mL of glacial acetic acid.

Note: The chitosan will not start dissolving until the pH is acidic, and even then it will not fully dissolve. This is ok.

4. Add 50 mL 5 M NaCl, then titrate the chitosan solution with NaOH to bring the pH to 6.2.

TE - Tween Storage Solution

- 10 mM Tris pH 8.0 + 1 mM EDTA
- 0.01% Tween-20

Quick Preparation Guide (50 mL)

1. 49.95 mL H_2O
2. 5 uL Tween-20

TE - SDS Solution

- 10 mM Tris pH 8.0 + 1 mM EDTA
- 0.5% SDS

Quick Preparation Guide (50 mL)

1. 49.75 mL H_2O
2. 250 uL SDS

Bead Removal Method 2 ("Spin-Out")

1. Remove membrane and place array into an empty 50 mL conical tube.
2. Ensure that the array is angled within the tube as shown below.
   Note: The array might move around at this point, which isn't something to worry about.
3. Add 48-50 mL of Wash 1 solution (See Buffers Guide)
4. Place the insert so the array is secured angled as shown in the image below.
5. Secure the lid and seal with parafilm, if necessary.
6. Put the sealed conical in a centrifuge, making certain the PDMS surface of the array is facing away from the rotor arm (See Diagram Below).
7. Centrifuge at 2000 x g for 5 minutes to remove the beads.
8. At this point you should see a small, but visible, pellet of beads at the bottom of the tube.
9. Aspirate 5 - 10 mL of Wash 1 solution to enable easier removal of the array.
10. Remove the array and carefully position it over the top of the 50 mL tube.
11. Repeatedly wash any remaining beads from the surface of the array over the surface of the 50 mL falcon tube using 1 mL of Wash 1 remaining in the tube.
12. Spin again at 2000 x g for 5 minutes to pellet beads.
13. Aspirate all wash 1 solution except for ~ 1mL.
    Note: Be careful to not disturb the pellet of beads.
14. Transfer beads to a 1.5 mL centrifuge tube and proceed to reverse transcription.

Shopping List

Device Manufacturing

- Dow Corning Sylgard 184 Silicone Encapsulant Clear 0.5 kg kit (Part No. 184 SIL ELAST KIT 0.5 PG)
- Protolabs Custom Array Molding Plates (Please refer to www.shaleklab.com/seq-well
    - Make out of aluminum and make sure to tap holes only on base plateBasePlate1 v3.1 (Bottom plate you mount the wafer to)
- BasePlate2 v3.1 (Divider for arrays)
- TopPlate1 v3.1 (Plate that holds the glass slides)
- TopPlate2 v3.1 (Top plate)
- 45 micron Silicon Master Wafer Size (Please refer to www.shaleklab.com/seq-well)
- Master, pre-silanized – (FlowJem, Inc. Toronto, Canada)
- Corning 72x25 Microscope Slides (Corning Life Sciences Cat. No. 2947)
- 6/32 ¼" Hex Screws
- 5/8" Hex 10/24 Screws
- Hex Screwdriver
- Vacuum grease
- 80 grit sandpaper
- 95% ethanol in spray bottle

Array Functionalization

Equipment

- Plasma Oven (Harrick Plasma PDC-001-HP)
- 2x 30-slide rack slotted (VWR Cat No. 25461-014)
- 16x20 cm staining dish (VWR Cat No. 25461-018)
- Vacuum Desiccator (VWR Cat No. 24988-164)
- Sterile 4-well dishes (Thermo Fisher Scientific Cat No. 267061)

(continues on the next page)

Reagents

- 200 proof ethanol (VWR Cat No. 89125-188)
- (3-Aminopropyl)triethoxysilane (Sigma Cat No. A3648)
- Acetone (Avantor Product No. 2440-10)
- p-Phenylene Diisothiocyanate (PDITC) (Sigma Cat No. 258555-5G)
- Pyridine (Sigma 270970-1L)
- Dimethylformamide (DMF) (Sigma Cat No. 227056-1L)
- Chitosan (Sigma Cat No. C3646-100G)
- Poly(L-glutamic) acid sodium solution (Sigma Cat No. P4761-100MG)
- 5M NaCl (Sigma Cat No. S6546-1L)
- Sodium Carbonate (Sigma Cat No. S2127-500G)

Buffer Reagents

Bead Loading Buffer

- Sodium Carbonate (Sigma Cat No. 223530-500G)
- BSA (Sigma Cat No. A9418-100G)
- Water (Thermo Fisher Scientific Cat No. 10977023)

Complete Lysis

- Guanidine Thiocyanate, (Sigma Cat No. AM9422)
- 0.5 M EDTA (Thermo Fisher Scientific Cat No. 15575020)
- Water (Thermo Fisher Scientific Cat No. 10977023)
- 10% Sarkosyl (Sigma Cat No. L7414)
- 100% 2-Mercaptoethanol (Sigma Cat No. M3148-25ML)

Hybridization Buffer

- 5 M NaCl (Thermo Fisher Scientific Cat No. 24740011)
- 1x PBS (Thermo Fisher Scientific Cat No. 10010023)
- PEG-8K (50%) (Fisher Scientific Cat No. BP337-100ML)

(continues on the next page)

Wash Buffer

- 5 M NaCl (Thermo Fisher Scientific Cat No. 24740011)
- 1 M $MgCl_2$ (Sigma Cat No.63069-100ML)
- 1 M Tris-HCl pH 8.0 (Thermo Fisher Scientific Cat No. 15568025)
- Water (Thermo Fisher Scientific Cat No. 10977023)

Array Quenching Buffer

- Sodium Carbonate (Sigma Cat No. 223530-500G)
- 1 M Tris-HCl pH 8.0 (Thermo Fisher Scientific Cat No. 15568025)
- Water (Thermo Fisher Scientific Cat No. 10977023)

RT Reagents

- UltraPure Distilled Water (Thermo Fisher Scientific Cat No. 10977023)
- Maxima 5x RT Buffer/Maxima H-RT (Thermo Fisher Scientific Cat No. EPO0753)
- 20% Ficoll PM-400 (Sigma Cat No. F5415-50mL)
- 10 mM dNTPs (New England BioLabs Cat No. N0447L)
- RNAse Inhibitor (Thermo Fisher Scientific Cat No. AM2696)
- Template Switching Oligo (Order from IDT)

Exonuclease Reagents

- Exonuclease I (*E. coli*) (New England Biolabs Cat No. M0293S)

Second Strand Synthesis Reagents

- Maxima 5x RT Buffer/Maxima H-RT (Thermo Fisher Scientific Cat No. EPO0753)
- 10 mM dNTPs (New England BioLabs Cat No. N0447L)
- dN-SMART Oligo (Order from IDT)
- UltraPure Distilled Water (Thermo Fisher Scientific Cat No. 10977023)
- Klenow Exo- (New England BioLabs Cat No. M0212S)
- 30% PEG8000 (Sigma-Aldrich 89510-1KG-F)

PCR Reagents

- IS PCR Primer (Order from IDT)
- KAPA HiFi Hotstart Readymix PCR Kit (Kapa Biosystems Cat No. KK-2602)

Nextera Reagents

- Nextera XT DNA Library Preparation Kit (96 samples) (Illumina FC-131-1096)
- New-P5-SMART PCR Hybrid Oligo (Order from IDT)
- Nextera N70X Oligo (Order from Illumina)

Operating Equipment

- Polycarbonate (PCTE) 0.01 micron 62x22 mm precut membranes, 100 count (Sterlitech Custom Order)
- mRNA Capture Beads (Chemgenes Cat No. MACOSKO-2011-10)
- Lifter Slips, 25x60mm (Electron Microscopy Science Cat No. 72186-60)
- Agilent Clamps (Agilent Technologies Cat No. G2534A)

Sequences

Barcoded Bead SeqB:

5'–Bead–Linker--TTTTTTTAAGCAGTGGTATCAACGCAGAGTAC-
JJJJJJJJJJJJNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTT--3'

Template Switching Oligo (TSO):
AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG

dN-Smart Randomer (dN-SMRT):
AAGCAGTGGTATCAACGCAGAGTGANNNGGNNNB

Smart PCR Primer (TSO_PCR):
AAGCAGTGGTATCAACGCAGAGT

New-P5-SMART PCR Hybrid Oligo (P5-TSO_Hybrid):

AATGATACGGCGACCACCGAGATCTACACGCCTGTC-
CGCGGAAGCAGTGGTATCAACGCAGAGT*A*C

Custom Read 1 Primer (Read_1_Custom_SeqB):
GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC

Appendix 4

Complete Bibliography

Publications

Wadsworth MH*, **Hughes TK***, Shalek AK. Marrying microfluidics and microwells for parallel, high-throughput single cell genomics. (2015) *Genome Biology.* 16:129. *These authors contributed equally to this work.

Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, Fallahi-Sichani M, Dutton-Regester K, Lin J, Cohen O, Shah P, Lu D, Genshaft AS, **Hughes TK**, Ziegler CGK, Kazer SW, Gaillard A, Kolb KE, Villani A, Johannessen CM, Andreev AY, Van Allen EM, Bertagnolli M, Sorger PK, Sullivan RJ, Flaherty KT, Frederick DT, Jané-Valbuena J, Yoon CH, Rozenblatt-Rosen O, Shalek AK, Regev A, Garraway LA. Dissecting the multicellular ecosystem of metastatic melanoma by single cell RNA-seq. (2016) *Science*. 352 (6282),189-196.

Gierahn, TM*, Wadsworth, MH II*, **Hughes TK***, Bryson BD, Butler A, Satija R, Fortune S, Love JC, Shalek AK. Seq-Well: portable, low-cost RNA sequencing of single cells at high-throughput. *Nature Methods* (2017). *These authors contributed equally to this work.

Mead BE, Ordovas-Montanes J, Braun AP, Levy LE, Bhargava P, Szucs MJ, Ammendolia DA, McMullan MA, Yin X, **Hughes TK**, Wadsworth MHW, Ahmad R, Rakoff-Nahoum S, Carr SA, Langer R, Collins JJ, Shalek AK, Karp JM. Harnessing single-cell genomics to improve the physiological fidelity of organoid derived cell types. (2018) *BMC Biology*. 16(1):62.

Ordovas-Montanes J, Dwyer DF, Nyquist SK, Buchheit KM, Vukovic M, Deb C, Wadsworth MHW, **Hughes TK**, Kazer SW, Yoshimoto E, Cahill KN, Bhattacharyya N, Katz HR, Berger B, Laidlaw TM, Boyce JA, Barrett NA, Shalek AK. Allergic inflammatory memory in human respiratory epithelial progenitor cells. (2018) *Nature*. 560: 649-654.

Jiarui Ding, Xian Adiconis, Sean K Simmons, Monika S Kowalczyk, Cynthia C Hession, Nemanja D Marjanovic, **Travis K Hughes**, Marc H Wadsworth, Tyler Burks, Lan T Nguyen, John YH Kwon, Boaz Barak, William Ge, Amanda J Kedaigle, Shaina Carroll, Shuqiang Li, Nir Hacohen, Orit Rozenblatt-Rosen, Alex K Shalek, Alexandra-Chloé Villani, Aviv Regev, Joshua Z Levin. Systematic comparative analysis of single cell RNA-sequencing methods. In Press. *Nature Biotechnology*. (2019).

Peter van Galen, Volker Hovestadt, Marc H Wadsworth II, **Travis K Hughes**, Gabriel K Griffin, Sofia Battaglia, Julia A Verga, Jason Stephansky, Timothy J Pastika, Jennifer Lombardi Story, Geraldine S Pinkus, Olga Pozdnyakova, Ilene Galinsky, Richard M Stone, Timothy A Graubert, Alex K Shalek, Jon C Aster, Andrew A Lane, Bradley E Bernstein. Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity. (2019) *Cell*. 176(6): 1265-1281.

Toby P Aicher, Shaina Carroll, Gianmarco Raddi, Todd Gierahn, Marc H Wadsworth, **Travis K Hughes**, Chris Love, Alex K Shalek. Seq-Well: A sample-efficient, portable picowell platform for massively parallel single-cell RNA sequencing. *Single Cell Methods* (2019).

Travis K Hughes*, Marc H Wadsworth*, Todd M Gierahn*, Tran Do, David Weiss, Priscilla R Andrade, Feiyang Ma, Bruno J de Andrade Silva, Shuai Shao, Lam C Tsoi, Jose Ordovas-Montanes, Johann E Gudjonsson, Robert L Modlin, J Christopher Love, Alex K Shalek. Highly Efficient, Massively-Parallel Single-Cell RNA-Seq Reveals Cellular States and Molecular Features of Human Skin Pathology. (2019) *BiorXiv*. *These authors contributed equally to this work.

Patricia A. Darrah, Joseph J. Zeppa, Joshua A. Hackney, Marc H. Wadsworth II, **Travis K. Hughes**, Supriya Pokkali, Phillip A. Swanson, Nicole L. Grant, Mark A. Rodgers, Megha Kamath, Charles A. Scanga, Chelsea M. Causgrove, Dominick J. Laddy, Aurelio Bonavia, Danilo Casimiro, Philiana Ling Lin, Edwin Klein, Alexander G. White, Pauline Maiello, Alex K. Shalek, Mario Roederer, JoAnne Flynn, Robert A. Seder. *Prevention of Mycobacterium tuberculosis infection and disease in nonhuman primates following intravenous BCG immunization.* (2020) *Nature*.