Investigating the Effects of Mycobacterium tuberculosis Bacterial Heterogeneity on Innate Immune Cell Tropism

Investigating the Effects of *Mycobacterium tuberculosis* Bacterial Heterogeneity on Innate Immune Cell Tropism

Wiktor Karpinski

A Thesis in the Field of Biology

for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

May 2024

Abstract

Tuberculosis (TB) is an infectious disease caused by the pathogen *Mycobacterium tuberculosis (*Mtb*)*. Tuberculosis is the leading cause of death by a single infectious agent worldwide despite the use of a live attenuated vaccine and treatment antibiotics. Mtb spreads from person to person by aerosolized particles primarily through inhalation, where it encounters cells of the innate immune system. Previous research has found that TB patients harbor a diverse population of Mtb, and further evidence suggests that the bacterial population measurably evolved in response to selective pressures driven by environments within hosts. Yet, there is limited research on how these heterogenous populations respond to different kinds of immune cells. The aim of this research is to investigate differences in cell tropism through phagocytic uptake – a ubiquitous function of many innate immune cells- when infected by heterogenic clinical isolates of *Mycobacterium tuberculosis.* Here, several innate immune cells (monocytes, neutrophils, macrophages, and dendritic cells) from healthy, TB-naïve human donors were infected for four hours with a unique molecularly barcoded library of Mtb clinical isolates from Vietnam. It was found that uptake by cells altered relative abundance of strains compared to background growth in media. Interestingly, there were several strains within sub-lineages and clades that may be driving phagocytic uptake within cell types, such as monocytes and neutrophils, suggesting cell tropism that is unique to individual strains. Together, these data suggest that bacteria heterogeneity plays an important role in phagocytic uptake in human innate immune cells.

Acknowledgments

This work would not have been possible without the many people supporting my goals. First I would like to thank Dr. Sarah Fortune for taking on the role of my research advisor and for this exciting opportunity to carry out this research in her lab. You created a supportive environment in which my work thrived. I would also like to especially thank Dr. Nicole Howard for your tireless efforts to guide me through countless hours of data analysis, execution of protocols, and for patiently answering all questions that I had pertaining to this work. I cannot express my sincerest gratitude enough for your time. I have grown so much as a scientist under your supervision. Special shout-out to Dr. Peter Culviner for guiding me through the wonderful world of coding, to Dr. Sydney Stanley who alongside Dr. Howard laid the foundations for this work, to Dr. Michael Chase for bridging the gap between sequencing and data analysis, and to my fellow research assistants at the Fortune Lab- Forrest Hopkins, Andrew Vickers, and Byron Roman as well as the rest of the members of the Fortune Lab and Rubin Lab for your comradery.

Thank you to Dr. James Morris who was very helpful in my thesis process and to all those on my thesis committee.

And lastly, I would like to give a special thanks to my wife, Maura, without your support this work would not have been possible.

Table of Contents

List of Figures

Chapter I

Introduction

This chapter will introduce the background of the experiment, the preliminary work that inspired this research, and why it is important in the field of tuberculosis. Several topics will be introduced, including tuberculosis incidence, the pathogen associated with the disease (*Mycobacterium tuberculosis*) and the importance of bacterial heterogeneity, its secretion systems, and how it interacts with innate immune cells in the human body.

Background of the Problem

Tuberculosis (TB) is an infectious disease caused by the pathogen *Mycobacterium tuberculosis* (Mtb). Tuberculosis is the leading infectious cause of death worldwide. In 2017, 10 million people were affected by TB, and 1.6 million people died of TB disease, according to the World Health Organization. In the last 200 years, TB has killed more than one billion people- more deaths than from influenza, malaria, smallpox, HIV/AIDS, plague, and cholera combined. Treatment for TB exists, primarily in developed countries, however, drug resistance is a continued threat (National Institute for Allergies and Infectious Diseases, 2020).

The emergence of increasingly drug-resistant strains of Mtb reveals the pathogen's ability to adapt to antibiotic pressures. Previous research has shown the influence of bottlenecks, population sub-division, and purifying selection on the genetic diversity of Mtb circulating within human hosts. In addition, it was found that TB

patients harbor a diverse population of Mtb, and further evidence suggests that the bacterial population measurably evolved in response to selective pressures driven by environments within hosts (O'Neill, Mortimer, & Pepperell, 2015).

The Fortune Lab at the Harvard T.H. Chan School of Public Health has engineered a library of digitally barcoded plasmids that were introduced into cultured sputum samples of Mtb-infected patients from Vietnam, referred to as Vietnam Clinical Isolates (VCI). This resource is a powerful tool that can be used to track behaviors of 159 Mtb strains from evolutionarily distinct lineages (L1, L2, and L4), including a laboratory reference L4 strain, Erdman. The "barcode" used to track these strains consists of a random 18-mer stably inserted into the L5 phage integration site of the bacterial genome. This results in a low chance of integrating the same barcode into two different strains (Stanley et al., 2023). This collective barcoded library was pooled when each individual stain reached an OD of 0.5, stored at -80C, and made available for downstream assays.

Preliminary work in the Fortune Lab (by Dr. Nicole Howard and Dr. Sydney Stanley) observed the behavior of the VCI pool in murine bone marrow-derived macrophages (BMDMs) from a panel of mice deficient in components of the immune system known to be important for Mtb infection and restriction. At defined time points, the intracellular bacteria were recovered, briefly expanded, and then the molecular barcodes were sequenced to determine the relative abundance of the strains. Interestingly, there was evidence that Mtb phylogenetic structure, rather than host cell genetic background, appeared to have the strongest effect on macrophage uptake (data not shown/unpublished data).

Mtb has been shown to inhabit several different cell types over the course of infection (Liu, Liu, & Ge, 2017) through dissemination which has been linked to specific Mtb genes. The mechanism by which Mtb infects these different cell types can trigger different types of immune responses, including increased survival of Mtb in specific cell types, dampening immune effects, tissue destruction, and host selective pressure, suggesting there may be a benefit in cell tropism. Due to the complicated nature of working with clinical isolates, such studies have not been able to clearly demonstrate the relationship between a large number of Mtb clinical isolates and innate immune cells. The Fortune Lab, however, does have the ability to investigate this relationship due to its expansive library of molecularly barcoded Mtb clinical isolates from a Vietnamese population and access to anonymous, healthy, TB-naïve blood donors from the greater Boston area. Following a similar method from the preliminary studies mentioned previously, innate immune cells may be isolated from whole blood and infected with the VCI library, where relative uptake using phagocytosis as a feature of interest may be investigated. Phagocytosis is a function of many innate immune cells, including bone marrow-derived cells of myeloid origin such as neutrophils, monocytes, and the mature form of monocytes (Allenspach & Torgenson, 2013). By isolating these cells from healthy human blood and infecting them with the VCI pool, a hypothesis forms where there will be differences in phagocytic uptake of strains between cell types due to bacterial preference on a lineage, clade, or individual strain basis. The relative bacterial abundances of each strain taken up will be measured by amplicon sequencing the molecular barcodes of strains that are biologically relevant. The relevance will be

determined by a number of quality control metrics throughout the experiment and in data analysis.

The findings of this study have important implications of whether uptake of Mtb is dependent upon lineage-to-lineage, clade-to-clade, or within individual strains that may be driving differences in uptake within certain myeloid cells. Such findings may help shed some light on the complicated relationships between host-pathogen interactions and bacterial heterogeneity in *Mycobacterium tuberculosis.* Further, investigating lineage-clade-strain-specific uptake differences/unique genes associated with uptake within myeloid cells may lead to interesting treatment options in future studies. There is a possibility that this hypothesis is incorrect since such a study has not been attempted before. If it is found that there are no significant differences in phagocytic uptake across cell types, then this study will be useful in future studies surrounding the relationship between Mtb and innate immune cells.

Tuberculosis

Tuberculosis (TB) is one of the oldest recorded infections in humans and is one of the deadliest diseases to this day despite the worldwide use of a live attenuated vaccine and treatment antibiotics (Smith, 2003). The causative agent of the disease-*Mycobacterium tuberculosis* (Mtb*)*– is an aerobic bacterium with an unusual, waxy coating on its cell surface primarily due to the presence of mycolic acid (Fu & Fu-Liu, 2002). Mtb is an extremely successful pathogen that adapts to survive within the host. It utilizes a range of effector proteins to evade destruction by the host immune system and shapes its environment to reside in granulomas (found mostly in the lungs), which are sophisticated and organized structures made up of immune cells that are established by

the host in response to persistent infection (Chai et al., 2018). This allows the pathogen to remain latent in a host for many years. A host infected with a latent form of TB has no clinical symptoms, radiological abnormalities, or microbiological evidence (Lee 2016).

Certain cases of untreated latent TB can progress to active TB. Approximately 5% to 10% of untreated latent TB cases will progress to active TB, with about half of these cases developing within two years of infection. Individuals with compromised immune systems, especially those afflicted by untreated human immunodeficiency virus (HIV), have a higher risk of developing TB disease. When Mtb overcomes host immune system defenses and begins to multiply, the disease progresses from latent to active at which point an individual becomes infectious, where it can spread via aerosol droplets through coughing, talking, and singing (World Health Organization, 2023.) (Furin et al., 2019).

In 2012, there were an estimated 8.6 million incident cases of TB globally. In high-income countries, such as Western Europe, Canada, Australia, the United States of America, and New Zealand, the incidence of TB has been declining over several decades due to healthcare availability and treatment options (Glaziou et al., 2015). However, the Centers for Disease Control and Prevention reported a small increase in TB cases in 2021 (from 2.2 to 2.4 cases per 100,000 population) most likely due to the COVID-19 pandemic (CDC, 2022).

Countries with a high TB burden include India, China, and South Africa. Countries in the top 10 worldwide in terms of incidence rates are mostly in Africa. In South Africa and Swaziland, it is estimated that at least 1 out of 100 people develop TB each year (Glaziou et al., 2015). This has led to genetically distinct populations of Mtb, which are classified into 7-9 lineages.

Bacterial Heterogeneity of *Mycobacterium tuberculosis*

The lineages of Mtb found globally are collectively referred to as the *Mycobacterium tuberculosis* complex (MTBC). This consists of *Mycobacterium tuberculosis* sensu stricto (Mtb) (lineages 1, 2, 3, 4, and 7) and Mtb var. *africanum* (lineages 5 and 6, also referred to as *M. africanum*). Recently, two new Mtb lineages have been proposed (lineages 8 and 9) from the African Great Lakes region. The MTBC lineages vary in their geographic distribution and spread. They are endemic in different regions around the world; thus, it is believed that the strain types are specifically adapted to different human populations. Lineage 2 (L2) is found to be mobile, due to evidence of recent spread from Asia to Europe and Africa. Lineage 4 (L4) is common in Europe and southern Africa, with regions of high TB incidence and HIV co-infection (Napier, Campino, & Merid et al., 2020). Lineage 1 (L1) is believed to have migrated from East Africa to South America during the 16^{th}-18^{th} century slave trade (Conceição et al., 2019). Evidence suggests that MTBC lineages can determine the transmission, control, and clinical outcome of pulmonary and extra-pulmonary tuberculosis. More specifically, variational phenotypes include differences in drug resistance, host response, virulence, transmissibility, disease site, and severity. These phenotypes may offer advantages for those MTBC lineages and lead to an increased likelihood of disease spread and a poorer prognosis for patients. The relationship between increased virulence and poorer prognosis is unclear, with some studies suggesting increased mortality risk with less virulent strains. Interestingly, studies have shown that inter-strain variation exists within lineages. For example, it was found that two different Beijing sub-lineage strains, “ancient” (atypical) and “modern” (typical), showed differences in drug resistance, geographical distribution,

and virulence patterns (Napier, Campino, & Merid et al., 2020. Forrellad et al., 2013). More specifically, patients infected with L1 are generally older and have a higher case fatality rate even though it is less resistant than L2, while L2 is known to have high virulence, faster disease progression, and association with antibiotic resistance. L4 is the most widespread lineage on a global scale and is associated with cavitations and treatment failures (Rutaiwha et al., 2019. Stucki et al., 2016. Ordaz-Vasquez et al., 2023).

Whole Genome Sequencing of *Mycobacterium tuberculosis*

Single nucleotide polymorphisms (SNPs) can be identified using whole genome sequencing (WGS) that distinguish Mtb isolates in a genotype-matched cluster. SNPs can be mapped on a phylogenetic tree that demonstrates the evolutionary direction of SNP accumulation within isolates relative to a most recent common ancestor, a hypothetical reference point of which all collective isolates are descended from. The genetic distance between isolates can be estimated by the number of SNPs between them. This provides a measure of how closely related they are genetically. Mtb acquires mutations more slowly compared to other pathogens, but the accumulation of SNPs can still be observed over time (Nelson et al., 2022).

Human Innate Immune System

Humans rely on the innate immune system during the first critical hours and days of exposure to a pathogen. Innate immunity can be categorized into four defensive barriers: anatomic, physiologic, endocytic/ phagocytic, and inflammatory. In vertebrates and invertebrates alike, the recognition of a microbial invader is usually followed by engulfment by a phagocytic cell. Phagocytic cells can be organized into two categories:

professional and non-professional phagocytes. Professional phagocytes consist primarily of monocytes, polymorphonuclear neutrophils, monocyte-derived macrophages, and tissue-resident macrophages- cells whose main function is to phagocytose. Non-professional phagocytes include all other cell types that can perform phagocytosis, such as dendritic cells (DCs) (Lim et al., 2017).

In vertebrates, macrophages are found in tissues throughout the body and are especially abundant in infection-prone areas, such as the lung and the gut. These long-lived cells are among the first cells to encounter a pathogen. Neutrophils are the second major family of phagocytes in vertebrates, which are primarily found in blood but not in normal, healthy tissues. They are recruited to sites of infection by activated macrophages and other signaling peptides released by the pathogen itself (Alberts, Johnson, & Lewis et al., 2002).

Monocytes are highly plastic and heterogeneous, and environmental stimulation can cause them to change their functional phenotype. When a tissue is damaged or infected, monocytes are recruited to the site, where they can differentiate into tissue macrophages and dendritic cells (Yang et al., 2014). M-CSF and GM-CSF are two important cytokines that monocytes utilize to differentiate into different flavors of macrophages. M-CSF (also referred to as M2) shows a mostly homeostatic expression pattern, whereas GM-CSF (also referred to as M1) is a product of cells activated during inflammatory or pathologic conditions. In other words, M-CSF is associated with tissue homeostasis and repair while GM-CSF is associated with pro-inflammatory properties (Ushach & Zlotnik, 2016). Dendritic cells, meanwhile, can be differentiated from monocytes with cytokines GM-CSF and IL-4 (Yokota et al., 2009).

Mycobacterium tuberculosis spreads from person to person by aerosolized particles primarily through inhalation. Once in the lower respiratory tract, Mtb first comes into contact with alveolar macrophages and eventually is exposed to monocytes, neutrophils, macrophages, and dendritic cells. In addition, studies that have examined sputum cultures of patients afflicted by active TB identified neutrophils as the predominant phagocytic cell infected with Mtb. Further, it has been found that the laboratory strain H37Rv was found to be more efficient compared to clinical strains in inducing late apoptosis/necrosis in neutrophils (Hilda, Narasimhan, & Das, 2015). Once Mtb colonizes the host, an immune response triggers the formation of granulomas which are cell aggregates constituted by macrophages, granulocytes, and lymphocytes, among others. Mtb can infect several cell types, including neutrophils, macrophages, and endothelial cells through interactions of mycobacterial glycolipids with receptors in host cells to promote bacilli internalization (Jacobo-Delgado et al., 2023). Once internalized by the cell, Mtb can reside in different cellular compartments such as phagosomes and autophagosomes. The host utilizes several cellular and immunological strategies to control the infection that competes with a broad range of Mtb evasion and virulence mechanisms. Successful evasions of host control mechanisms allow for bacterial survival and replication.

In addition to being the first cells to come in contact with aerosolized Mtb, macrophages are also the preferred cellular host and site of replication. During early infection, macrophages are polarized into an M1 (GM-CSF) state. As chronic infection progresses, these macrophages may shift to an M2 (M-CSF) type that regulates excessive inflammation and promotes tissue healing. The loss of immune pressure by the M1-

derived molecules allows Mtb to thrive in M2 environments. The importance of M2 polarization for Mtb survival in host macrophages is reflected by the stronger polarization bias driven by virulent strains. The ESAT-6 molecule, for example, produced by virulent strains of Mtb induces a stronger M2 polarization compared to attenuated strains that lack ESAT-6. Virulent Mtb is known to induce less apoptosis and more necrosis than attenuated Mtb. Cell death caused by necrosis, in addition to an M2 anti-inflammatory environment, provides an opportunity for Mtb to proliferate more effectively. M1-modulated environments and apoptosis promote stronger T-cell responses, which are less favorable for Mtb growth (Thiriot et al., 2020).

Mtb infection is not limited to the lungs, as extrapulmonary infection may occur in any organ. The most common site of these infections is lymph nodes (LN), yet little is known about how Mtb influences LN structure and function (Ganchua et al., 2018). The transportation of Mtb from the lung to LN is usually facilitated by dendritic cells. Previous studies have shown that the activation and cell trafficking ability of Mtb-infected dendritic cells was dependent upon the type of Mtb strain. Further, suppressing the migration of human monocyte-derived dendritic cells and modulating its cell trafficking ability may be a mechanism used by *Mycobacterium tuberculosis* to paralyze the early innate immune response of its host (Mihret, 2012).

Secretion Systems of *Mycobacterium tuberculosis*

Protein secretion is an important feature in bacteria for adaptation and survival in their natural surrounding environments. Further, these systems are crucial for interacting with host cells by exporting toxins/signal proteins that allow bacteria to cause pathology. These secretion systems have evolved over time to exhibit specific functions, such as

those involved in pathogenicity. Several gene clusters encoding proteins that are secreted by mycobacteria have been identified and may contribute to mycobacterial pathogenesis. One such system in *Mycobacterium tuberculosis*, named ESX-1, is one of the most well-studied secretion systems due to its importance in virulence and, more specifically, the ability of one of its secreted effector proteins, EsxA, to induce phagosomal rupture in host phagocytes (Groshel et al., 2016). ESX-1 is conserved in many other mycobacteria, such as *M. smegmatis* and *M. leprae* (although in *M. smegmatis* ESX-1 is associated with DNA transfer, not virulence). Mtb utilizes the ESX-1 secretion system to deliver virulence proteins during infection of host cells (Coros et al., 2008).

PE/PPE Proteins

The PE/PPE families have been found exclusively in mycobacteria and are comprised of two large unrelated families of acidic and glycine-rich proteins whose genes are clustered. PE and PPE genes make up about 10% if the coding capacity of the Mtb H37Rv genome (H37Rv is the most widely used Mtb strain, and its genome is globally used as the Mtb reference sequence (Poonam et al., 2022). A majority of these genes can be found upstream or within the ESX operons. Most of the PE/PPE proteins are localized on the cell surface and/or are secreted, and they induce a strong immune response in the host. However, the specific function of the PE/PPE family of proteins is still unknown. They have been connected to virulence of mycobacteria (e.g., the disruption of *ppe38* observed in L2 strains has been associated with a hypervirulent phenotype) as a source for antigenic biodiversity and as modulators of the host immune response, yet more investigation is necessary in order to better understand the exact role these proteins play in pathogenesis (Gómez-González et al., 2023. Forrellad et al., 2013).

Barcoded Mtb Clinical Strain Library

A molecular barcode can be integrated into the DNA of Mtb to identify it during sequencing. The use of molecular barcoded Mtb bacteria can provide insights into strain abundance before and after exposure to a cell. This particular “barcode”, engineered at the Fortune Lab, is transformed with an integrating plasmid library carrying randomized 18bp DNA sequence tag that inserts at the L5 phage integration site of the Mtb genome. With this approach, which is shown in Figure 1, a library of 159 strains was generated. The barcode is amplified from isolated genomic DNA in two rounds of nested PCR steps to prepare it for sequencing. In the first round, a pool of four primers with a degenerate variable-length spacer anneals a random 9-mer “molecular counter” that allows for the enumeration of PCR templates rather than amplicons. The degenerate variable-length spacer, or “phasing region”, introduces sequence variability required for Illumina-based sequencing of relatively low-complexity libraries. The second round of PCR adds sequencing adapters and multiplexing indices (Stanley et al., 2023).

A custom Python pipeline, BARTI, is used to identify barcodes from Illumina sequencing. It applies cutoff to read samples by analyzing those that are greater than 10,000, as well as identifying barcode sequences surrounded by constant motifs. It removes barcodes that are within one base pair from another more abundant barcode to avoid false readings. True reads are further separated from false reads by using second derivatives (Martin et al., 2017).

Research Aims, Goals, and Hypothesis

The primary research goal of this thesis was to investigate phagocytic differences in cellular tropism of several innate immune cell types when infected with *Mycobacterium tuberculosis* strains from various lineages and to observe unique variants that may be driving these changes. Studies have established that different myeloid cells play a distinct role in Mtb infection, thus it is important to investigate the relationship between Mtb uptake and cell types. Based on previous literature and evidence seen in preliminary mouse models, it was hypothesized that there would be observable Mtb lineage differences that were phagocytosed between cell types. These differences may have downstream consequences in disease outcomes. Two specific aims have been crafted to assess this hypothesis, and the details of the study design can be found in Figure 1.

Primary Goal: Assess relative uptake of Mtb clinical strains in an *in vitro* model of cellular phagocytosis in several cell types (neutrophils, monocytes, macrophages (GM-CSF & M-CSF), and dendritic cells) and observe unique variants that may be responsible for differences in uptake.

Specific Aim 1: Investigate lineage-level, clade-level, and strain-level differences in relative abundance through phagocytic uptake in several human innate immune cells.

- Methods: Isolate and differentiate immune cells of interest from fresh whole human blood, infect cells with appropriate MOI of VCI Mtb, extract bacterial genomic DNA from lysed immune

cells, attach molecular barcodes to each condition via PCR, sequence barcodes, run on BARTI pipeline.

- Expected Results: There will be differences in bacterial relative abundance that is dependent upon lineages, clades, and individual strains.

Specific Aim 2: Identify genes of interest that may be responsible for differences in phagocytic uptake of Mtb based on phylogenetic structure and genetic pairwise comparison.

- Methods: A previously constructed whole-genome phylogenetic tree of the VCI pool will be visualized in iTOL, Strains will be rank ordered based on uptake features and variants will be identified using the breseq software package mapping to the H37Rv genome and run through Pyseer grouping all variants in a given coding region.
- Expected Results: There will be unique variants present in outlier strains within clades that may be responsible for uptake.

Chapter II

Materials and Methods

The materials and methods section describes the isolation of innate immune cells from whole blood, FACS protocols, VCI library outgrowth and infection, gDNA extraction, sequencing, and downstream analysis. Additionally, this section describes the phylogenetic structure for the VCI pool, as well as identifying unique variants in individual strains.

Innate Immune Cell Isolation from Whole Blood

Whole blood from healthy, anonymous, deidentified donors was obtained from Research Blood Components (RBC) in Boston, MA. Following a protocol approved by the Harvard IRB, purified populations of monocytes and neutrophils/PMNs were isolated using Lymphoprep (Stemcell), Leucosep tubes (VWR), and HBSS/EDTA (Thermofisher) by centrifugation. Neutrophils/PMNs were isolated from red blood cells at 18C-22C using red blood lysis buffer. Monocytes were isolated from PBMCs and platelets using lysis buffer, centrifugation, and a CD14+ isolation kit (Stemcell) at 18C-22C. Monocytes were further differentiated into MCSF/GMCSF macrophages and dendritic cells using appropriate cytokines (50ng/mL of hMCSF (Biolegend) for M0/M2 macrophages, 50ng/mL of hGMCSF (Biolegend) for M1 macrophages, 50ng/mL of hGMCSF+ hIL4 for dendritic cells (Biolegend) diluted in RP10 (Roswell Park Memorial Institute (RPMI-1640) medium + 10% fetal bovine serum (FBS), 1% HEPES). These cells were incubated at 37C, 5% CO2 for 7 days to allow for appropriate differentiation. All cells were plated at 1e6 cells/mL in tissue culture treated 24 well plates (Portevin et al., 2011).

FACS Analysis

Cells (at a concentration of 1e5- 1e6) were pre-incubated in FcBlock (Biolegend) for 20 minutes at 4C in flow staining media (Hanks' Balanced Salt Solution (HBSS)+ 10% FBS). Cells were incubated with antibodies, all at a 1:200 dilution, against CD14-GFP (clone M5E2) and CD66b-PB (clone G10F5) (Biolegend) for 30 minutes at 4C, washed, and fixed with 4% PFA for 10 minutes in the dark. Flow cytometry data was obtained on an iQue3 (Sartorius) and analyzed using FlowJo (version 10.8.2). Based on established literature, monocytes were defined as CD14+CD66b- and neutrophils as CD66b+CD14-. Flow data confirmed monocytes expressing high signaling of CD14 and neutrophils expressing high signaling of CD66b (Figure 2), whereas previous studies have shown that macrophages express high CD16 and low CD14, while dendritic cells express high levels of DC-SIGN (data not shown) (Bournazos 2016). Additionally, the purity of monocytes and neutrophils isolated directly from whole blood was assessed with the same protocol (Figure 3).

VCI Pool Infection and 7H9 Outgrowth

The barcoded VCI library was grown at 37C with shaking for 48 hours in 7H9 media (7H9 salts (Middlebrook), 1% glycerol, 10% OADC (Middlebrook), 0.05% Tween-80) + 20ug/mL kanamycin (kan) to mid-log phase prior to in vitro infections. The culture was washed with PBS and resuspended in R10 media. Immune cells were infected with the prepared VCI library at various multiplicity of infections (MOIs) (1, 5, 10) for four hours at 37C, 5% CO2. After infection, cells were washed 3x with warm sterile PBS and lysed with cold deionized distilled water for 20 minutes at room temperature. A small aliquot of the cell lysate was used to serially dilute and plate for intracellular CFU on

7H11+kan agar plates. Plates were incubated for 2-3 weeks at 37C and colonies were counted to determine optimal MOI infection for downstream analysis (data not included), to which an MOI of 1 was selected. The remainder of the cell lysates were inoculated into 7H9 + kan media and grown for 7-10 days. The infectious inoculum was also used to inoculate 7H9+ kan cultures grown in triplicate, to mimic a "cell-free" control condition. Cultures were kept in a matched growth phase until subsequent genomic DNA (gDNA) extraction.

Mycobacterial Genomic DNA Extraction

Mycobacterial cultures were spun down and resuspended in TE-NaCL for gDNA extraction. Samples were transferred to bead beater tubes (MPbio), mixed with phenol:chlorophorm:isoamyl alcohol (PCIA), and mechanically pulsed in a Precelly machine twice for thirty seconds. Samples were centrifuged, supernatant was removed, combined with 25ug/mL RNAse (Invitrogen), and incubated for 1 hour at 37C. PCIA was added again, and samples were taken from biosafety level 3 (BSL3). Samples were then centrifuged in phase lock tubes (5prime) to remove residual phenol and the supernatant was mixed with isopropanol/sodium acetate and incubated overnight at -20C. Samples were finally washed with 70% EtOH and resuspended in nuclease-free water (Fisher Scientific).

Next Generation Sequencing of Mycobacterium

gDNA samples were quantified using a Qubit 4 Fluorometer and diluted to 25ng/uL. Two rounds of PCR were performed to attach sequence-specific primers and Ilumina adapter primers to DNA products for barcode identification. Bands of ~367 bp

were confirmed by running a DNA gel (data not shown). DNA product was purified using AMPure XP beads and quantified using Qubit 4 Fluorometer. The total library was quantified by ViiA7 qPCR, then diluted to 4nM and loaded into an Ilumina V3 150bp kit. Samples were sequenced on MiSeq (Ilumina) and barcodes were identified using a custom Python pipeline, BARTI, as detailed in Martin et al.

Data Analysis

Using a similar method outlined in Stanley et al., barcode abundances were compared in the technical replicates by calculating the barcode normalized to read counts to control for differences in sample read depth and barcode abundance in the input library: $\frac{\frac{barcode\ X\ read\ count\ in\ condition\ A}{total\ barcode\ read\ counts\ in\ condition\ A}}{\frac{average\ barcode\ X\ read\ count\ in\ input}{average\ total\ barcode\ read\ counts\ in\ input}}$. These values were then input into R studio (version 2023.12.01+402) to create a PCA plot to determine whether the VCI library grew differently in 7H9 outgrowth versus cell types as well as confirming separation between donors or cell types (Figure 4). To more closely capture the 7-10 day time period in which bacteria-containing cell lysates were grown, the barcode counts of cell conditions were normalized to the "cell-free" control condition or 7H9 outgrowth as follows: $\frac{\frac{barcode\ X\ read\ count\ in\ condition\ A}{total\ barcode\ read\ counts\ in\ condition\ A}}{\frac{average\ barcode\ X\ read\ count\ in\ 7H9\ outgrowth}{average\ total\ barcode\ read\ counts\ in\ 7H9\ outgrowth}}$. Log2FC values were calculated in each subsequent condition to be able to visualize results more easily. These strains were then organized by lineage and graphed in Prism (version 10.1.1).

Construction and Visualization of Phylogenetic Tree

A phylogenetic tree was constructed according to previously published work (Stanley et al., 2023). A joint variant call file was created with alignment, variant calling, and mutation annotations where reads were aligned to the inferred ancestral genome of the *Mycobacterium tuberculosis* complex most common ancestor. This whole-genome phylogenetic tree was constructed based on 33,566 sites from the joint VCF using fastTree as detailed in the extended methods by Stanley et al. and visualized in iTOL (version 6.8.2).

Identifying Unique Variants in Strains

Log_2FC values were separated by lineage and rank ordered to determine quantifiable regions of unique uptake (data not shown). Variants were identified using the breseq software package mapping to the H37Rv genome (accession: NC_000962.3). Variant types were annotated using the SnpEff software package and the effects of missense mutations were predicted using SIFT 4G. The pyseer package using a linear mixed model was used to associate the <PHENOTYPE> with predicted functional variants (INDELs or missense mutations < 0.05 SIFT score) in each coding region; pyseer was run in burden mode grouping all variants in a given coding region. Identification of variants unique to particular strains were identified by visualization of the phylogenetic tree in ITOL followed by use of custom scripts built on pysam (Berrick Lab, breseq, Pyseer 1.3.10 Documentation, Hegar).

Chapter III

Results

This section highlights the results from the experiments listed in specific aims 1 and 2. The growth consistency of bacterial strains, cell tropism of innate immune cells, and genetic pairwise comparisons of phylogenetically structured strains are presented here. Accompanying relevant figures are found in the 'Figures' section.

Growth Consistency of Bacterial Strains

In order to compare barcode abundance accounted for differences in sample read depth, the VCI inoculum pool was assessed within two separate experiments (stock expt. 1 and stock expt. 2) and across two different rounds (round 1 and round 2). There was correlation of the inoculum pool across several rounds of experiments (Figure 5) with some experiments having more drift than others, especially as barcode abundances became sparser within strains (data not shown). These correlations show overall that the input library retained consistent barcode abundance across different experimental dates. Further, the VCI inoculum pool (also referred to as VCI library) grown in 7H9+ kan media for several days had consistent barcode abundance across four separate infections (Figure 6) and within replicates (Figure 6), despite the potential for competition for resources when growing in nutrient-containing media, in addition to accounting for differences in doubling time which vary across the isolates. It was for these reasons that it must be assumed that outgrowth in 7H9 will be inequivalent for every strain. Additionally, barcode relative abundance highly correlated within biological replicates of donors **(**Figure 7**)**. Therefore, these replicates within donors were averaged for

downstream analyses, while donors were analyzed separately to account for host heterogeneity. Altogether, the VCI outgrowth separated from the cell types quite clearly (Figure 4), suggesting that strains underwent some environmental pressure that was separate from outgrowth in media alone. Up to this point, analyses were performed at the barcode level; within the pool, each strain was represented by one, two, or three clones (66 strains were represented by three clones, 64 strains by two clones, and 29 strains by a single clone (Stanley et al., 2023)) to create biological replicates of the strains. The Log_2FC values for each barcode across each condition were compared using a Pearson R correlation coefficient analysis, and biological replicates that fell below the 1.5* interquartile range cutoff were considered outliers and removed (according to the distribution of all Pearson correlation coefficients across the individual barcodes (n= 260) in the pool) (Figure 8 and Figure 9). Additionally, samples with low sequencing depth (read counts <20) were also removed.

Cell Tropism Across Innate Immune Cells

The resulting data showed that uptake by cells altered relative abundance of strains compared to outgrowth in 7H9 alone (Figure 4). Given prior literature, the association between cell type and lineage was assessed by lineage-specific and clade-specific patterns of uptake (Sarkar et al., 2012, Reiling et al., 2013). When viewing the relative abundance of VCI strains across innate immune cells as a whole, it was found that strains were taken up most similarly by monocytes and neutrophils, monocytes and M-CSF macrophages, neutrophils and M-CSF macrophages, and GM-CSF macrophages and dendritic cells (Figure 11). Interestingly, within monocytes and neutrophils, but not in other cell types, L1 strains were relatively less abundant than L2 and L4 strains (Figure

12), which may suggest that there are lineage-specific differences that drive uptake within those cell types, yet follow-up studies must be done to confirm this, with a particular focus on sub-clades within L1 that may be contributing to these differences. Prior literature has established that innate immune cells behave differently when exposed to Mtb, due to their function, phagocytic receptors, certain cell types taking up mycobacterium more freely, etc. (B. Shoshana Zha, et al., 2022, Sarkar et al., 2012, Reiling et al., 2013) Together data suggest there are differential interactions between the Mtb strains of the VCI pool and the panel of innate immune cells. However, it is still suspected that certain lineage outliers are driving these changes and contributing to cell tropism effects seen.

Genetic Pairwise Comparison of Phylogenetically Structured Strains

Strains of the VCI pool were organized by their phylogenetic structure and a tree was created (Figure 10). Log_2FC for all cell types separated by donor were projected on the tree, and patterns of uptake were observed. A Pyseer script was run on the phenotype of interest, in this case, the average Log_2FC, across and within all cell types and no significant hits were found (data not shown). Next, conditions were separated by donors and cell types, and the closest hit to statistical significance was a VapB antitoxin in monocytes from a single donor ($p=0.053$, data not shown). Instead, SNPs specific to most divergent outlier strains were identified since uptake is likely impacted by many different genes thus, a GWAS approach may not be the most appropriate method to utilize. Divergent uptake patterns in comparison to nearest neighbor strains were noted in strains 178, 456, 179, and 187. Additionally, they were found to be either the most or least abundant based on rank order of Log_2FC of strains within each lineage (data not shown).

In order to determine genetic factors that might contribute to these diverging phenotypes in nearest-neighbor strains, a pairwise genetic comparison was used to identify unique single nucleotide polymorphism (SNP) variants. Strain 456 had several unique variants not found anywhere else on the tree (Figure 13), including several related to cell wall and cell processes (such as Rv1424c and Rv3695) and genes previously linked to virulence, such as Rv3812 and the ESX-1 associated protein Rv3866. Strain 178 also carries unique variants in cell wall and cell processes (Rv0064 and Rv2281) (Figure 14). In the L2 subclade containing strains 179*, 187*, 18, 13, 44, 314, 252, and 251 it was found that both strains 179 and 187 had higher Log_2FC (especially in monocytes and neutrophils) than neighboring strains within the same subclade. Unique variants to strains of interest were then identified that may play an important role in these uptake differences (Figure 10). For strains 179 and 187, 0 unique variants were shared by 179 and 187 and not shared by the outgroup. Thus, it is suspected that: there is a single mutation that 179 and 187 share that explains the data, but this type of analysis was unable to capture it; 179 and 187 each have unique mutations that cause this effect (requiring two mutations); the 179-13 shares a mutation that causes increased Log_2FC, and the 18-13 clade shares a mutation that reverts this effect (requiring two mutations). More testing will need to be done to investigate this observation. There were several strains within sub-lineages, clades, and individual strains that may be driving phagocytic uptake within cell types, suggesting cell tropism that is unique to individual strains in the VCI pool.

Chapter IV

Discussion

The results from the previous chapter are discussed here. The significance of cell tropism and unique variants of individual strains that may be of interest (456 & 178) are expanded upon. Future studies are speculated upon, and how they may benefit the field of tuberculosis.

Significance of Results

Together this work utilizes a barcoded Mtb collection of clinical strains to comprehensively assess how bacterial genetic heterogeneity impacts phagocytic uptake by human innate immune cells, in order to identify patterns of cell tropism that may vary by lineage, clade, or even individual strains. To better understand how diverse strains of Mtb may differentially interact with critical immune cell populations to influence rates of uptake, genetic pairwise comparisons were performed on several strains that were quantitively selected and unique variants were observed. The pipeline for creating and analyzing barcodes in the Vietnamese Mtb clinical isolates was pioneered by the Fortune Lab, which provides this research with an opportunity to investigate complex relationships between unique Mtb strains and their propensities to be phagocytosed by various human innate immune cells such as monocytes, neutrophils, M-CSF/GM-CSF macrophages, and dendritic cells.

Although there was little evidence to definitively suggest that Mtb uptake was driven in a lineage-specific manner, the data produced in this study implies uptake dynamics may be affected by clades and individual strains rather than entire lineages.

Similar studies in the field, which usually rely on much fewer strains to explore lineage-level differences, may not accurately capture the complexity of the genetic diversity that exists even within Mtb lineages. Strains from L1 appeared to have restricted access into monocytes and neutrophils, which may be attributed to mechanisms of uptake such as functionality or phagocytic receptors when it comes to pathogenic infections. It is unclear whether it is something unique about L1 (such as surface proteins), shared receptors of monocytes and neutrophils, or a combination of the two that may be driving this. More studies must be performed to further untangle the relationship between L1 and monocytes and neutrophils. Such follow-up studies may have an important impact on disease outcomes in patients infected by L1 strains of Mtb.

When looking at individual strains and their unique variants, a few interesting features were found in strains belonging to L1. Strain 456, which was taken up at a higher Log_2FC than its neighbors, and its neighbor 178 had several unique variants that should be more closely investigated, especially in those that are important in cell wall and cell processes. Interactions between mycobacterial cell walls and receptors on host cells are important in phagocytic uptake, thus emphasizing the importance of further investigating these variants. When looking for unique variants across cell type and donor, only one variant (VapB) from a single donor of monocytes was close to statistical significance ($p=0.053$). The other two monocyte donors had a poorer beta score but moved in the same direction as the first donor (data not shown), so this experiment may have not been able to fully capture this variant in all three donors. Further, host genetic background may be influencing these results, and follow-up studies with more donors may strengthen this finding.

Follow-up studies on these strains should be performed to identify important genetic differences that may be responsible for this observation. When strains were grouped by cell type and overlaid donors, similarities and differences were found across cell types. It is suspected that these changes are being driven by individual strains as opposed to cell tropism. Strong differences were not observed in the relative abundance of different Mtb strains across the cell types assessed in this study. Frequently, strains were either strongly or poorly taken up by all cell types, suggesting that Mtb may use similar mechanisms to trigger phagocytosis, but the ability varies strain to strain. The innate immune cells included here share a number of known receptors known to mediate uptake of Mtb, including complement receptors and mannose receptors. In addition to the follow-up studies proposed earlier surrounding unique variants in single strains, it would be important to shed light upon mechanisms of uptake that are highly shared across Mtb strains. Further investigation is necessary to explain the complicated relationship between *Mycobacterium tuberculosis* and host cells, which may have downstream consequences on disease outcomes.

Research Limitations

Discerning differences between biological and technical significance proved to be a challenge in this analysis. The methodology of obtaining barcode reads from gDNA extracted from infected cells isolated from fresh blood is a lengthy process. While the aim was to analyze barcode reads from three donors for this experiment, there was a processing limit of one to two donors per run. Variation between batches could conceivably be due to technical variation or donor-driven effects. In order to correct for batch-to-batch variation, technical controls for normalization were included.

Additionally, the *in vitro* experiments (1e6 cells/mL per cell condition) were not reflective of the cell composition in the human body. *In vitro* experiments that accurately capture a human immune environment are extremely difficult to execute, thus it may be of interest to perform similar experiments in a mouse model.

The blood samples are obtained from healthy donors from the Boston area. This unfortunately will not be an accurate sample of a large population. More specifically, it would be interesting to obtain healthy blood samples from a population in Vietnam to observe how these Vietnam strains behave in our experiments. The blood samples that were obtained were anonymous and any identifying markers were destroyed by RBC. Thus, the role of demographics in cell tropism, such as age, race, and biological sex, could not be investigated.

Many strains in the VCI pool were dropped either due to poor reproducibility or other failures to meet an acceptable standard. It would be ideal to somehow rescue the lost strains in order to paint a clearer picture of what is collectively happening in the VCI library.

Altogether, there were several limiting factors in this research. Some may be corrected, such as technical failures of the strains or testing more donors. Others may be more difficult to achieve, such as mimicking a true human immune environment. It would be interesting to see if correcting these limitations would significantly impact the results obtained here.

Figures

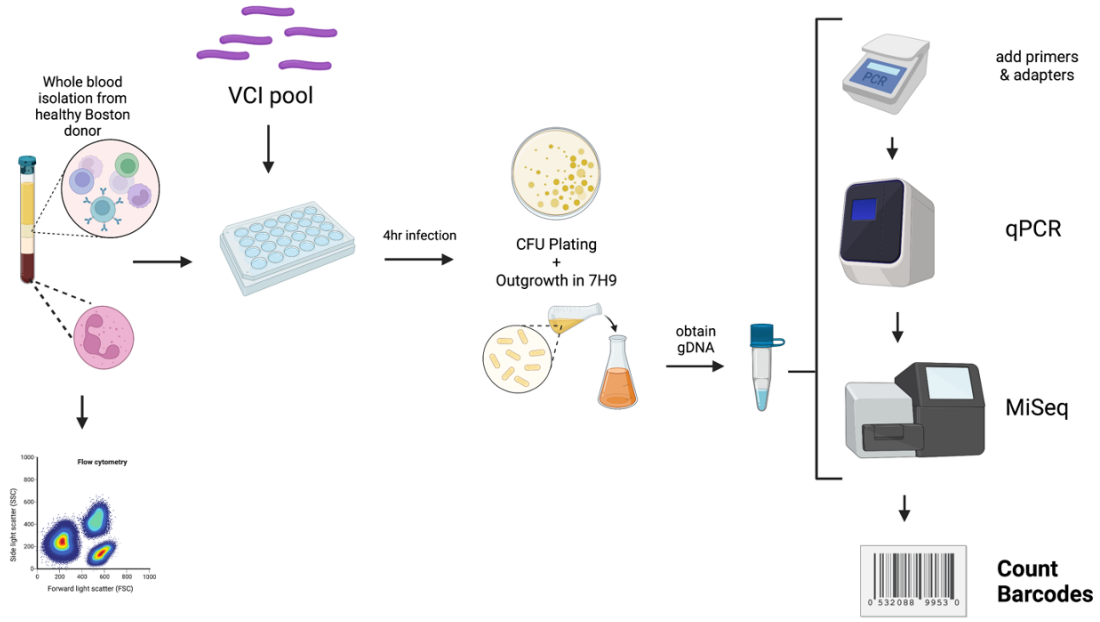


Figure 1. Experimental overview.

Whole blood was isolated from healthy, TB-naïve donors from the Boston area and cell populations were confirmed via FACS analysis. Cells were infected for four hours with VCI pool (VCI library) and intracellular bacteria were grown in 7H9 media for 7-10 days and plated for CFU as a confirmation of appropriate MOI. gDNA was obtained and sequenced on MiSeq, and bacteria relative abundance was quantified using the Fortune Lab's Python BARTI script. Figure generated using BioRender.

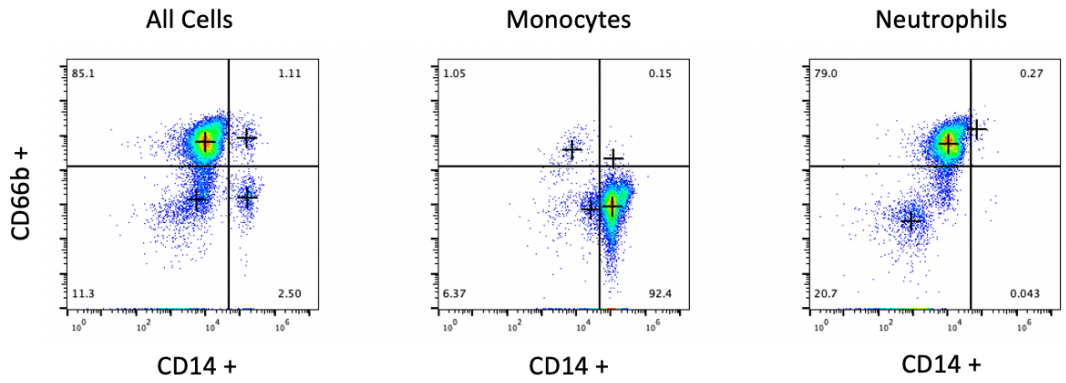


Figure 2. FACS analysis of innate immune cells isolated from whole blood.

From a mixed starting population of cells from a single donor (left), isolation yielded pure populations of monocytes (center, 92.4% purity) and neutrophils (right, 79% purity).

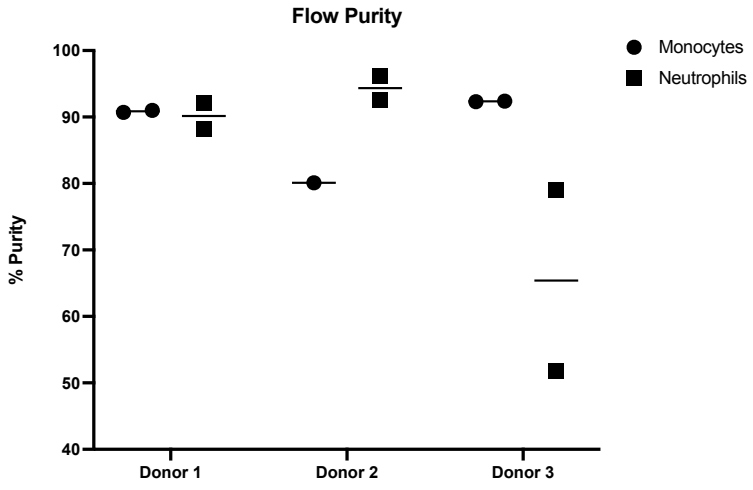


Figure 3. Purity of isolated cells from whole blood.

Assessing purity of cell isolation from three different donors for monocytes (%CD14+, CD66b-) and neutrophils (%CD66b+, CD14-) was assessed via FACS.

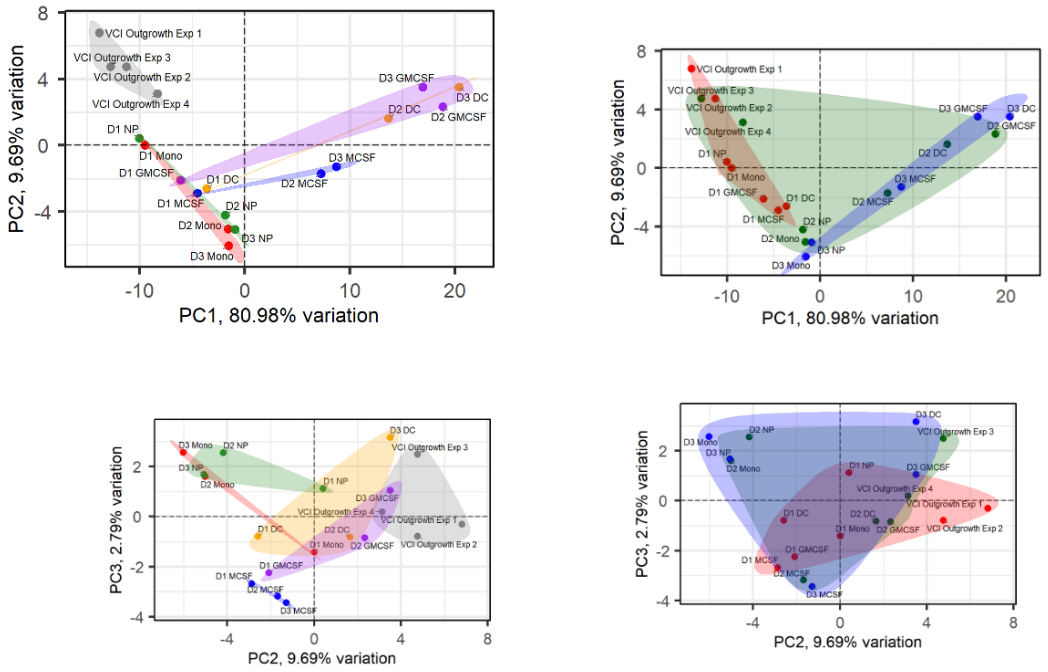
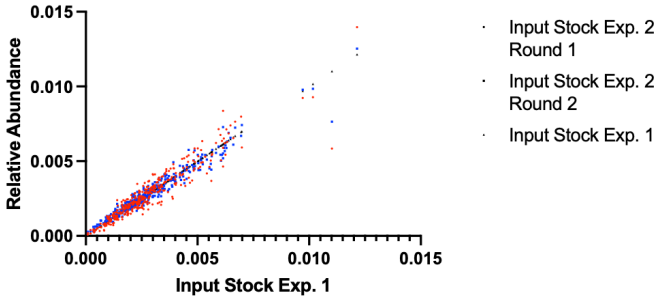


Figure 4. PCA plots of cell condition and outgrowth normalized to stock.

Principal component analysis (PCA) of barcode relative abundance across all conditions. Plots are colored by condition (left) or donor (right).



Input Stock Exp. 1 vs. Input Stock Exp. 2 Round 1	Input Stock Exp. 1 vs. Input Stock Exp. 2 Round 2	Input Stock Exp. 2 Round 1 vs. Input Stock Exp. 2 Round 2
0.94	0.98	0.97

Figure 5. Relative abundance growth consistency of VCI inoculum pool.

Correlation of the normalized barcode read counts of the VCI inoculum pool across experiments. Pearson correlation coefficients were calculated.

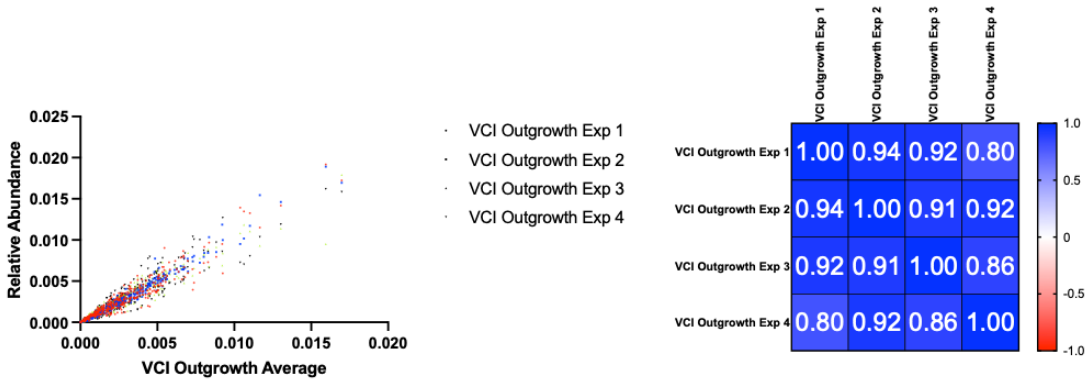


Figure 6. Growth consistency of VCI inoculum in 7H9 +Kan media.

VCI inoculum pool was grown for 7-10 days in 7H9 +kan, concurrently with innate cell samples, for each experiment. Barcode read counts were normalized to total reads for each sample. Correlation of the normalized barcode read counts of the 7H9 outgrowth condition across experiments is shown. Data analyzed using Pearson R correlation (right).

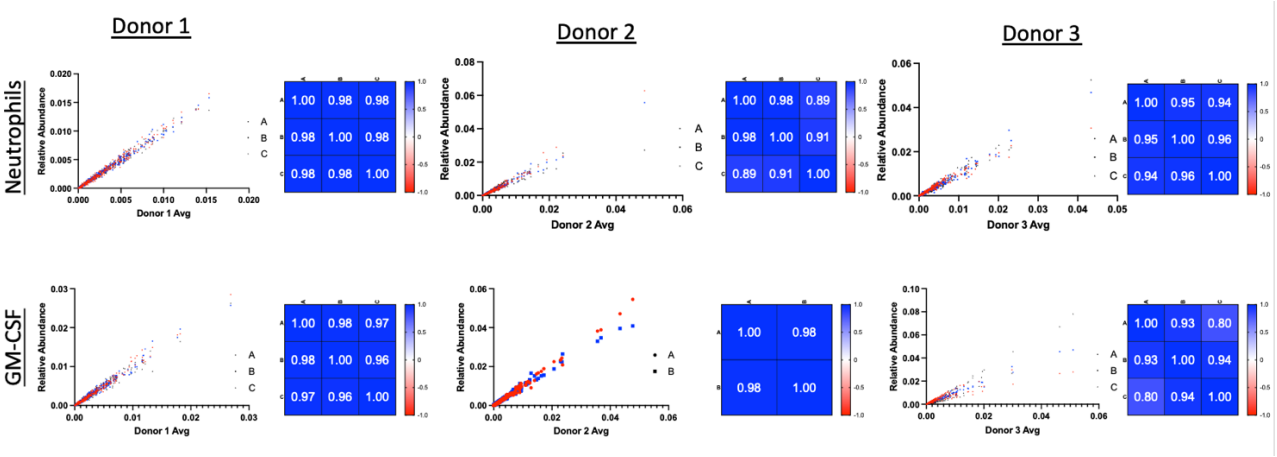


Figure 7. Biological replicate correlation of donors across cell types.

Relative barcode abundance of three (3) biological replicates within each donor in neutrophils and GM-CSF macrophages (GMCSF). A Pearson R correlation coefficient matrix is shown.

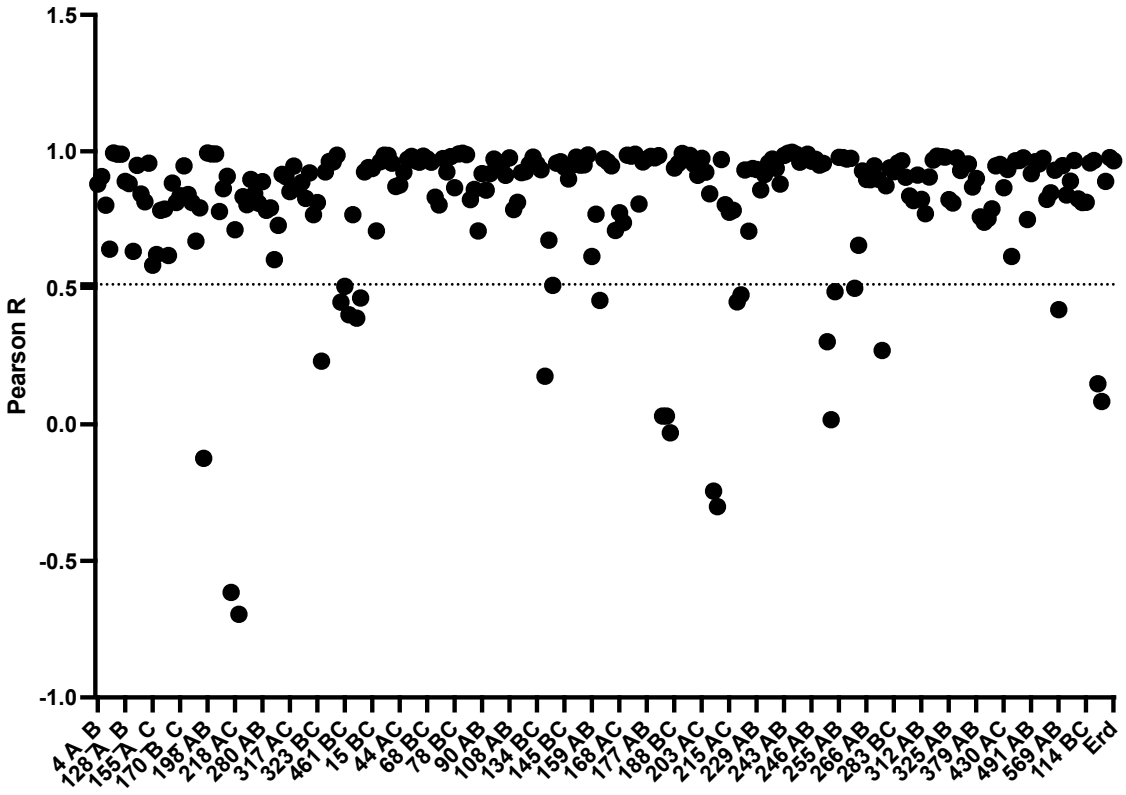


Figure 8. Pearson R correlation of VCI strain replicates.

*Pearson R correlations of all barcode replicates in the pool. Outlier cut-off (Q3-[1.5 * IQR]) shown as a dotted line. 17 strains fall below the outlier cutoff. (L1: 197, 218, 328, 461, 505 and 525; L2: 138, 159, 181, 188, 203, 224, 254, 259, 276, and 569; L4: 326).*

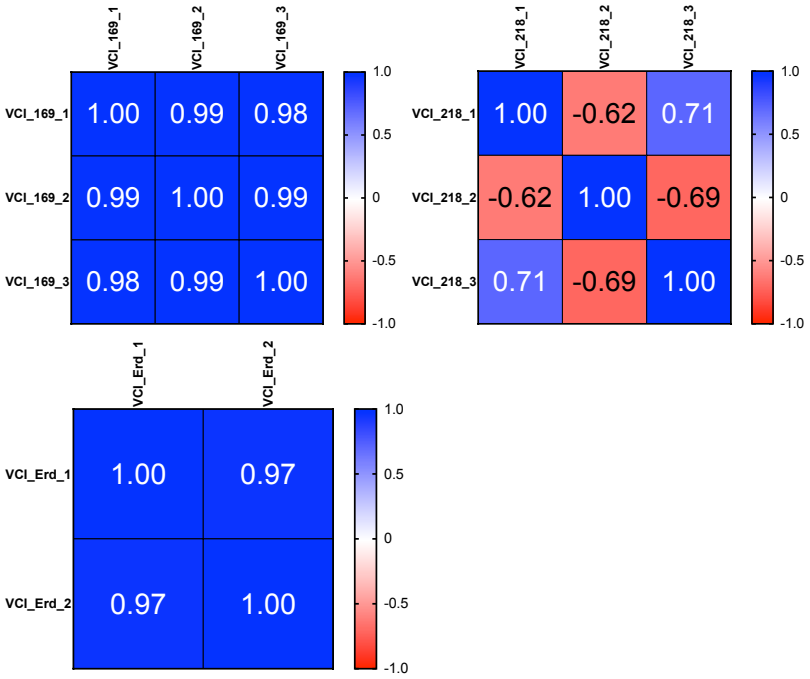


Figure 9. Representative Pearson R correlation heat maps of VCI strain replicates.

Representative Pearson R heat maps from three strains with replicates (169, top left; 218, top right; and Erdman, bottom) used to determine outliers.

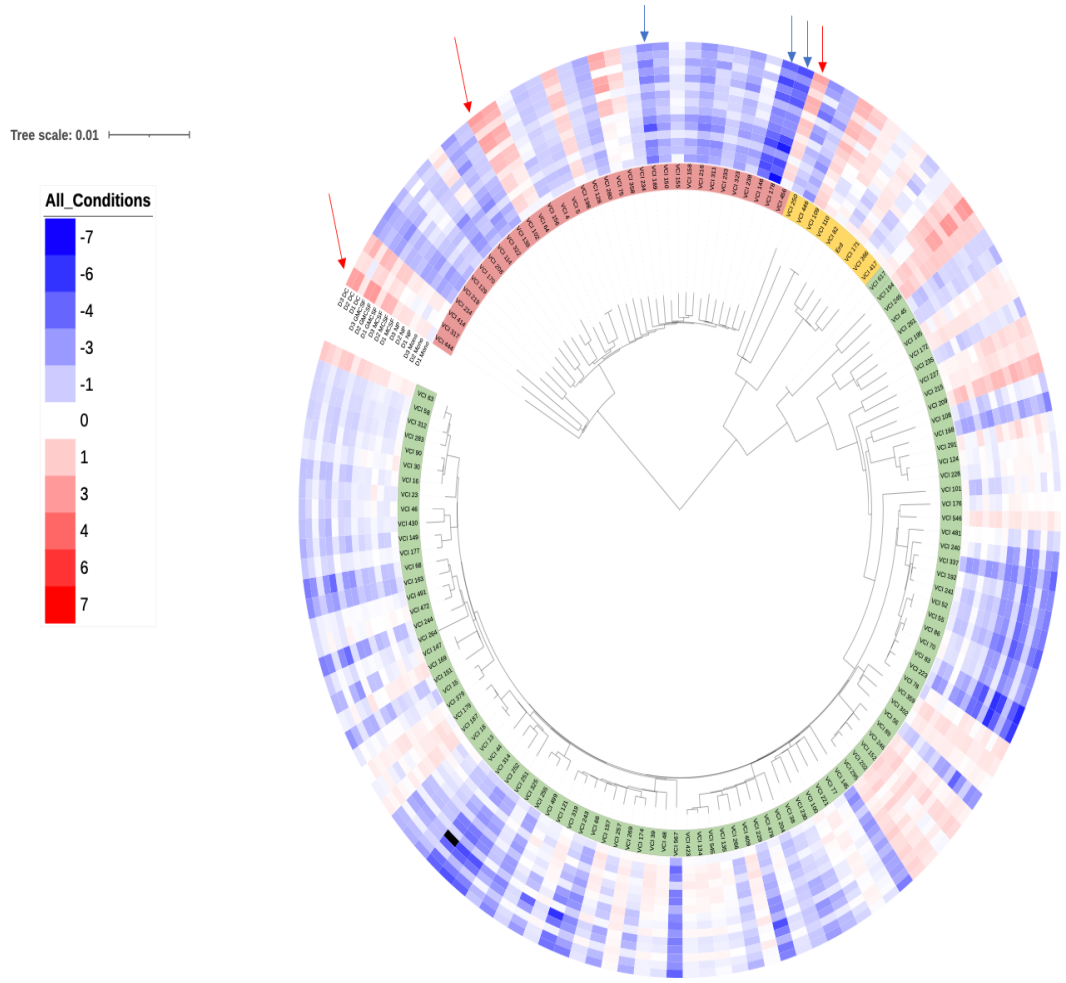


Figure 10. Log_2FC values of myeloid cells mapped to phylogenetic tree of VCI pool.

Log_2FC values of relative abundance of VCI pool that was internalized by a particular cell type from three donors mapped to the VCI phylogenetic tree (Blue = low relative abundance. Red = high relative abundance). Strains on internal track color color-coded by lineage (red= L1, yellow= L4, green =L2). Strains that were dropped by Pearson R correlation or failure to meet acceptable QC were removed from the tree. A black bar indicates no data in that particular strain within lineage. Red (high) and blue (low) arrows point to top 3 and bottom 3 strains by rank order. Figure made in iTOL.

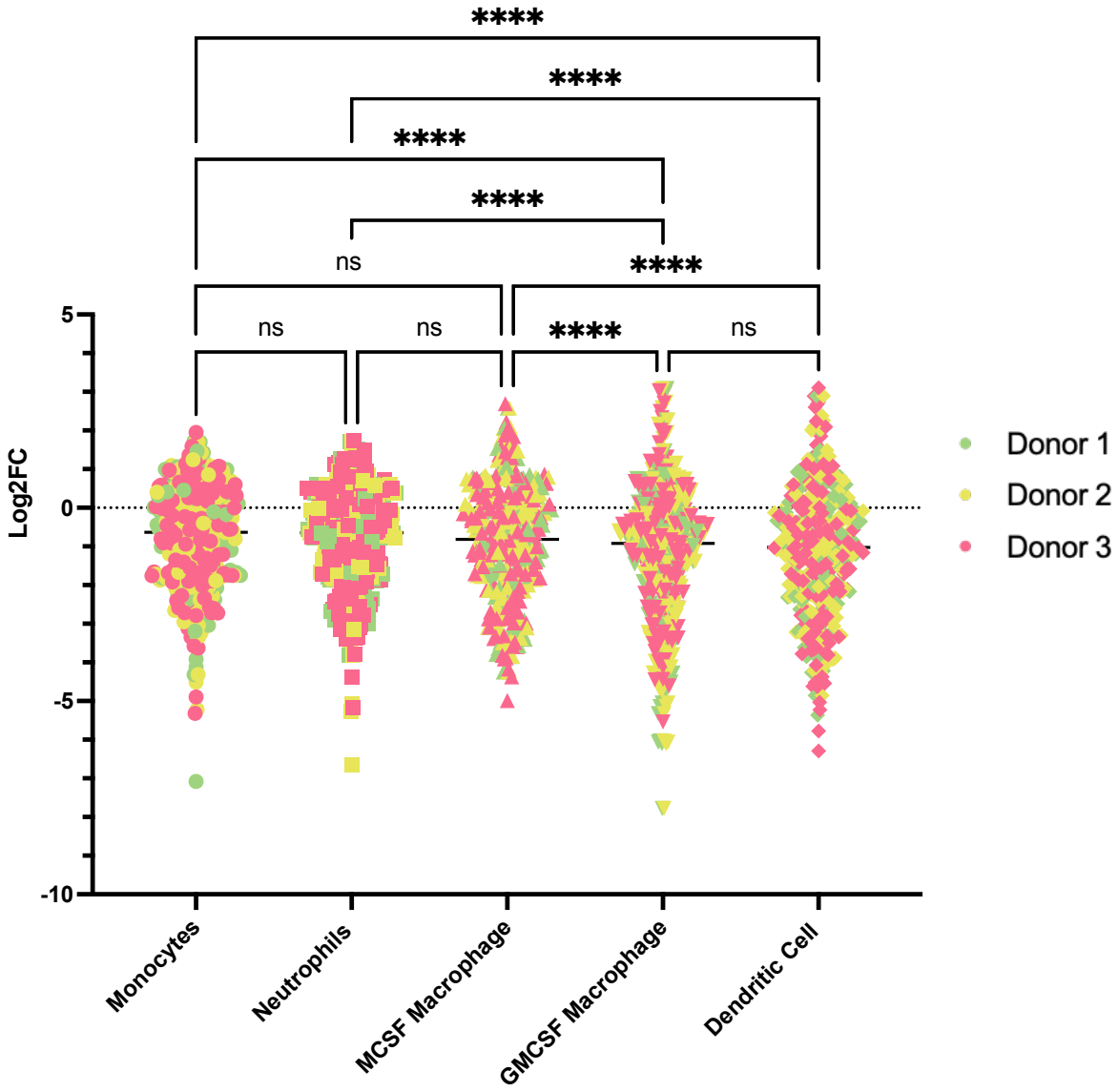


Figure 11. Log_2FC of internalized Mtb grouped by cell type.

Log2FC values of relative strain abundance of VCI pool that was internalized by a particular cell type from three donors. One-way ANOVA done in Prism (version 10.1.1).

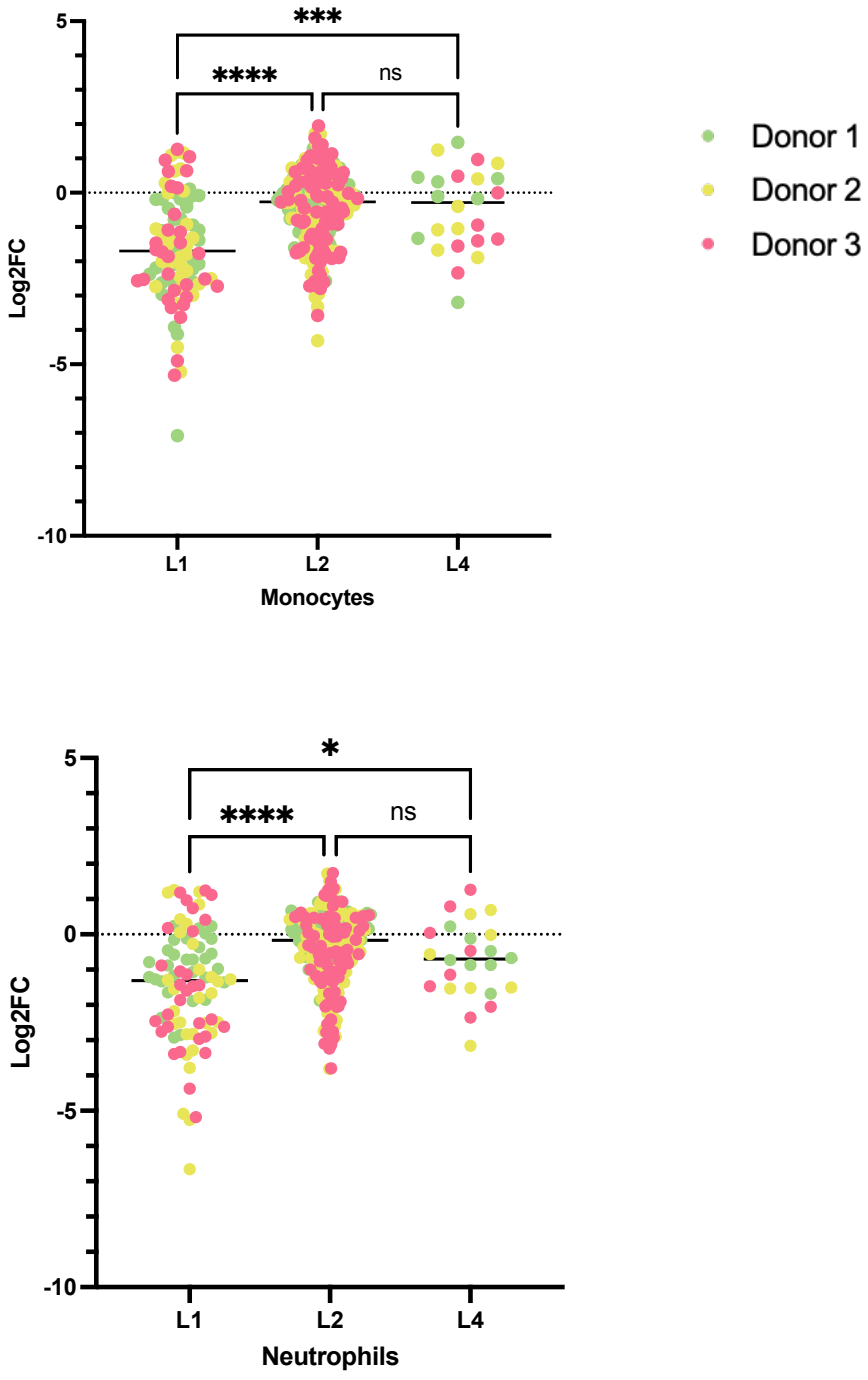
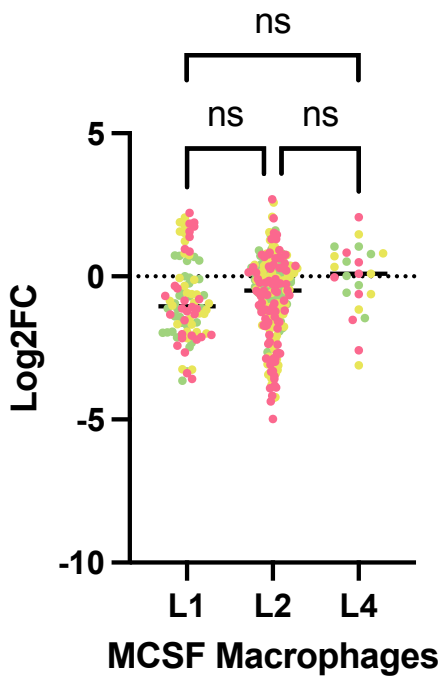


Figure 12. Log_2FC of internalized Mtb grouped by lineage.



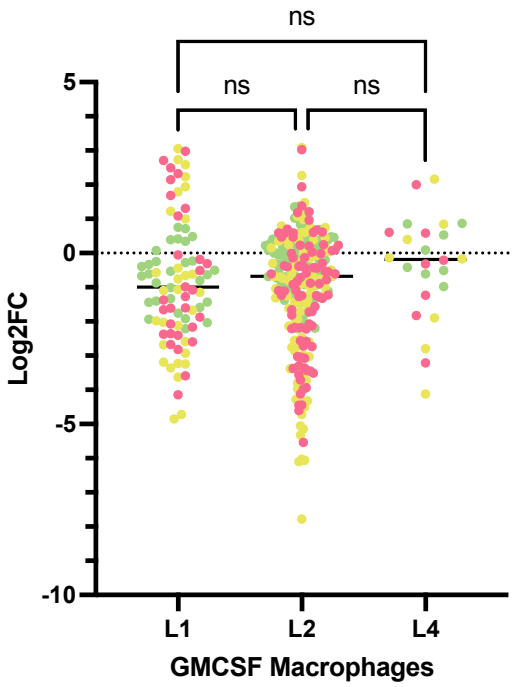


Figure 12 (cont'd). Log_2FC of internalized Mtb grouped by lineage.

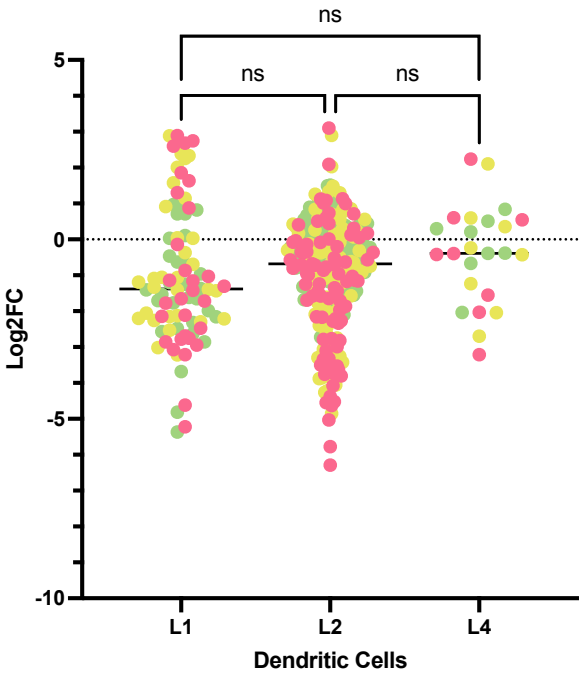


Figure 12 (cont'd). Log_2FC of internalized Mtb grouped by lineage.

Log2FC values of relative abundance of VCI pool that was internalized by a particular cell type from three donors separated by lineage. One-way ANOVA done in Prism (version 10.1.1).

Variants unique to 456	Functions, processes, and essentiality
Rv0204c	not enough data.
Rv0928	involved in active transport of inorganic phosphate across the membrane (import), this is one of the proteins required for binding-protein-mediated phosphate transport, cell wall and cell processes, essential.
Rv1257c	not enough data.
Rv1424c	function unknown, cell wall and cell processes, non-essential.
Rv2439c	not enough data.
Rv3109	involved in molybdenum cofactor biosynthesis; involved in the biosynthesis of molybdopterin precursor Z from guanosine, intermediary metabolism and respiration, non-essential.
Rv3262	required for coenzyme F420 production: involved in the conversion of FO into F420, intermediary metabolism and respiration, non-essential.
Rv3695	function unknown, cell wall and cell processes, non-essential.
Rv3708c	involved at the second step in the common biosynthetic pathway leading from asp to the cell wall precursor MESO-diaminopimelate, to LYS, to met, to ILE and to THR [catalytic activity: L-aspartate-semialdehyde + orthophosphate + NADP(+) = L-aspartyl phosphate + NADPH]., intermediary metabolism and respiration, essential.
Rv3812	function unknown, thought to be involved in virulence, PE/PPE, non-essential.
Rv3853	binds to RNASE E and inhibits RNASE E endonucleolytic cleavages, regulatory proteins, non-essential.
Rv3866	function unknown, but product is ESX-1 secretion-associated protein EspG1, cell wall and cell processes, non-essential.

Figure 13. Unique variants to strain 456.

SNP variants unique to strain 456 were determined from the WGS data. The 12 variants predicted to affect protein function (INDELs or missense with a SIFT score of <0.05) are shown here. Functions were annotated using Mycobrowser.

Variants unique to 178	Functions, processes, and essentiality
Rv0064	function unknown, cell wall and cell processes, non-essential.
Rv0080	function unknown, non-essential.
Rv0205	function unknown, non-essential.
Rv0414c	involved in thiamine biosynthesis, intermediary metabolism and respiration, essential.
Rv1033c	sensor part of the two-component regulatory system TRCS/TRCR, involved in transcriptional autoactivation: TRCR activates its own expression by interacting with the at-rich sequence of the TRCR promoter, essential.
Rv2281	involved in phosphate transport, cell wall and cell processes, non-essential.
Rv2684	thought to be involved in active transport of arsenical compounds across the membrane (export): arsenic resistance by an export mechanism, responsible for the translocation of the substrate across the membrane, cell wall and cell processes, non-essential.
Rv2880c, Rv2879c	not enough data.
Rv2921c	probably involved in the reception and insertion of a subset of proteins at the membrane: possibly membrane receptor for FFH|Rv2916c, cell wall and cell processes.
Rv3054c	function unknown, non-essential.
Rv3371	may be involved in synthesis of triacylglycerol, lipid metabolism, non-essential.
Rv3500c	not enough data.
Rv3507	not enough data.
Rv3551	function unknown, intermediary metabolism and respiration, non-essential.
Rv3755c	function unknown.

Figure 14. Unique variants to strain 178.

SNP variants unique to strain 178 were determined from the WGS data. The 12 variants predicted to affect protein function (INDELs or missense with a SIFT score of <0.05) are shown here. Functions were annotated using Mycobrowser.

References

Alberts, Bruce, et al. “Molecular Biology of the Cell.” *Nih.gov*, Garland Science, 2002, www.ncbi.nlm.nih.gov/books/NBK21054/.

B. Shoshana Zha, et al. “Bacterial Strain–Dependent Dissociation of Cell Recruitment and Cell-To-Cell Spread in Early M. Tuberculosis Infection.” *MBio*, vol. 13, no. 3, American Society for Microbiology, June 2022, https://doi.org/10.1128/mbio.01332-22. Accessed 5 Mar. 2024.

“Barrick Lab :: Breseq.” *Barricklab.org*, barricklab.org/twiki/bin/view/Lab/ToolsBacterialGenomeResequencing.

CDCTB. “Reported TB in the U.S., 2020.” *Centers for Disease Control and Prevention*, 15 Nov. 2023, www.cdc.gov/tb/statistics/reports/2022/default.htm.

Chai, Qiyao, et al. “Mycobacterium Tuberculosis: An Adaptable Pathogen Associated with Multiple Human Diseases.” *Frontiers in Cellular and Infection Microbiology*, vol. 8, no. 158, May 2018, https://doi.org/10.3389/fcimb.2018.00158.

Chitale, Poonam, et al. “A Comprehensive Update to the Mycobacterium Tuberculosis H37Rv Reference Genome.” *Nature Communications*, vol. 13, no. 1, Nov. 2022, p. 7068, https://doi.org/10.1038/s41467-022-34853-x.

Clinical Applications of Immunogenetics. Elsevier, 2022, https://doi.org/10.1016/c2020-0-01767-6. Accessed 8 Mar. 2023.

Conceição, Emilyn Costa, et al. “Mycobacterium Tuberculosis Lineage 1 Genetic Diversity in Pará, Brazil, Suggests Common Ancestry with East-African Isolates Potentially Linked to Historical Slave Trade.” *Infection, Genetics and Evolution*, vol. 73, Sept. 2019, pp. 337–41, https://doi.org/10.1016/j.meegid.2019.06.001. Accessed 2 Mar. 2024.

Coros, Abbie, et al. “The Specialized Secretory Apparatus ESX-1 Is Essential for DNA Transfer in *Mycobacterium Smegmatis*.” *Molecular Microbiology*, vol. 69, no. 4, July 2008, pp. 794–808, https://doi.org/10.1111/j.1365-2958.2008.06299.

Forrellad, Marina A., et al. “Virulence Factors of The Mycobacterium Tuberculosis complex.” *Virulence*, vol. 4, no. 1, Jan. 2013, pp. 3–66, https://doi.org/10.4161/viru.22329.

Fu, L. M., and C. S. Fu-Liu. “Is Mycobacterium Tuberculosis a Closer Relative to Gram-Positive or Gram–Negative Bacterial Pathogens?” *Tuberculosis*, vol. 82, no. 2-3, June 2002, pp. 85–90, https://doi.org/10.1054/tube.2002.0328.

Furin, Jennifer, et al. “Tuberculosis.” *The Lancet*, vol. 393, no. 10181, Apr. 2019, pp. 1642–56, https://doi.org/10.1016/s0140-6736(19)30308-3.

Ganchua, Sharie Keanne C., et al. “Lymph Nodes Are Sites of Prolonged Bacterial Persistence during Mycobacterium Tuberculosis Infection in Macaques.” *PLOS Pathogens*, edited by Marcel A. Behr, vol. 14, no. 11, Nov. 2018, p. e1007337, https://doi.org/10.1371/journal.ppat.1007337. Accessed 3 Mar. 2021.

Glaziou, P., et al. “Global Epidemiology of Tuberculosis.” *Cold Spring Harbor Perspectives in Medicine*, vol. 5, no. 2, Oct. 2014, pp. a017798–98, https://doi.org/10.1101/cshperspect.a017798.

Gröschel, Matthias I., et al. “ESX Secretion Systems: Mycobacterial Evolution to Counter Host Immunity.” *Nature Reviews Microbiology*, vol. 14, no. 11, Sept. 2016, pp. 677–91, https://doi.org/10.1038/nrmicro.2016.131.

Gómez-González, Paula Josefina, et al. “Functional Genetic Variation in Pe/Ppe Genes Contributes to Diversity in Mycobacterium Tuberculosis Lineages and Potential Interactions with the Human Host.” *Frontiers in Microbiology*, vol. 14, Frontiers Media, Oct. 2023, https://doi.org/10.3389/fmicb.2023.1244319. Accessed 2 Jan. 2024.

Heger, Andreas. “Pysam: Package for Reading, Manipulating, and Writing Genomic Data.” *PyPI*, pypi.org/project/pysam/. Accessed 2 Mar. 2024.

Hilda, J. Nancy, et al. “Mycobacterium Tuberculosis Strains Modify Granular Enzyme Secretion and Apoptosis of Human Neutrophils.” *Molecular Immunology*, vol. 68, no. 2, Dec. 2015, pp. 325–32, https://doi.org/10.1016/j.molimm.2015.09.019. Accessed 30 Apr. 2022.

Jacobo-Delgado, Yolanda M., et al. *Mycobacterium Tuberculosis Cell-Wall and Antimicrobial Peptides: A Mission Impossible?* May 2023, https://doi.org/10.3389/fimmu.2023.1194923.

Lee, Seung Heon. “Tuberculosis Infection and Latent Tuberculosis.” *Tuberculosis and Respiratory Diseases*, vol. 79, no. 4, 2016, p. 201, https://doi.org/10.4046/trd.2016.79.4.201. Accessed 14 Jan. 2020.

Lim, Justin J., et al. “Diversity and Versatility of Phagocytosis: Roles in Innate Immunity, Tissue Remodeling, and Homeostasis.” *Frontiers in Cellular and Infection Microbiology*, vol. 7, May 2017, https://doi.org/10.3389/fcimb.2017.00191.

Liu, Cui Hua, et al. “Innate Immunity in Tuberculosis: Host Defense vs Pathogen Evasion.” *Cellular & Molecular Immunology*, vol. 14, no. 12, Sept. 2017, pp. 963–75, https://doi.org/10.1038/cmi.2017.88.

Martin, Constance J., et al. “Digitally Barcoding Mycobacterium Tuberculosis Reveals in Vivo Infection Dynamics in the Macaque Model of Tuberculosis.” *MBio*, edited by Christina L. Stallings, vol. 8, no. 3, May 2017, https://doi.org/10.1128/mbio.00312-17. Accessed 4 Apr. 2020.

Mihret, Adane. “The Role of Dendritic Cells in Mycobacterium Tuberculosis Infection.” *Virulence*, vol. 3, no. 7, Nov. 2012, pp. 654–59, https://doi.org/10.4161/viru.22586.

Napier, Gary, et al. “Robust Barcoding and Identification of Mycobacterium Tuberculosis Lineages for Epidemiological and Clinical Studies.” *Genome Medicine*, vol. 12, no. 1, Dec. 2020, https://doi.org/10.1186/s13073-020-00817-3. Accessed 3 Apr. 2022.

Nelson, Kristin N., et al. “Mutation of Mycobacterium Tuberculosis and Implications for Using Whole-Genome Sequencing for Investigating Recent Tuberculosis Transmission.” *Frontiers in Public Health*, vol. 9, Jan. 2022, https://doi.org/10.3389/fpubh.2021.790544.

NIAID. “Tuberculosis | NIH: National Institute of Allergy and Infectious Diseases.” *Nih.gov*, 25 Oct. 2019, www.niaid.nih.gov/diseases-conditions/tuberculosis.

O’Neill, Mary B., et al. “Diversity of Mycobacterium Tuberculosis across Evolutionary Scales.” *PLOS Pathogens*, edited by Sarah M Fortune, vol. 11, no. 11, Nov. 2015, p. e1005257, https://doi.org/10.1371/journal.ppat.1005257. Accessed 2 Oct. 2021.

Ordaz-Vázquez, Anabel, et al. “Mycobacterium Tuberculosis Lineage 4 Associated with Cavitations and Treatment Failure.” *BMC Infectious Diseases*, vol. 23, no. 1, Mar. 2023, p. 154, https://doi.org/10.1186/s12879-023-08055-9. Accessed 2 Mar. 2024.

Portevin, Damien, et al. “Human Macrophage Responses to Clinical Isolates from the Mycobacterium Tuberculosis Complex Discriminate between Ancient and Modern Lineages.” *PLoS Pathogens*, edited by Debra E. Bessen, vol. 7, no. 3, Mar. 2011, p. e1001307, https://doi.org/10.1371/journal.ppat.1001307. Accessed 4 Sept. 2020.

Reiling, Norbert, et al. “Clade-Specific Virulence Patterns of Mycobacterium Tuberculosis Complex Strains in Human Primary Macrophages and Aerogenically Infected Mice.” *MBio*, edited by Eric J. Rubin, vol. 4, no. 4, Aug. 2013, https://doi.org/10.1128/mbio.00250-13. Accessed 12 Mar. 2022.

Rutaihwa, Liliana K., et al. “Insights into the Genetic Diversity of Mycobacterium Tuberculosis in Tanzania.” *PLOS ONE*, edited by Riccardo Manganelli, vol. 14, no. 4, Apr. 2019, p. e0206334, https://doi.org/10.1371/journal.pone.0206334. Accessed 9 Nov. 2022.

Sarkar, Rajesh, et al. “Modern Lineages of Mycobacterium Tuberculosis Exhibit Lineage-Specific Patterns of Growth and Cytokine Induction in Human Monocyte-Derived Macrophages.” *PLoS ONE*, edited by Olivier Neyrolles, vol. 7, no. 8, Aug. 2012, p. e43170, https://doi.org/10.1371/journal.pone.0043170. Accessed 31 Jan. 2021.

Smith, Issar. “Mycobacterium Tuberculosis Pathogenesis and Molecular Determinants of Virulence.” *Clinical Microbiology Reviews*, vol. 16, no. 3, July 2003, pp. 463–96, https://doi.org/10.1128/cmr.16.3.463-496.2003.

Stanley, Sydney, et al. “High-Throughput Phenogenotyping *of Mycobacteria Tuberculosis* Clinical Strains Reveals Bacterial Determinants of Treatment Outcomes.” *BioRxiv (Cold Spring Harbor Laboratory)*, Cold Spring Harbor Laboratory, Apr. 2023, https://doi.org/10.1101/2023.04.09.536166. Accessed 2 Mar. 2024.

Stucki, David, et al. “Mycobacterium Tuberculosis Lineage 4 Comprises Globally Distributed and Geographically Restricted Sublineages.” *Nature Genetics*, vol. 48, no. 12, Oct. 2016, pp. 1535–43, https://doi.org/10.1038/ng.3704. Accessed 29 May 2020.

Thiriot, Joseph D., et al. “Hacking the Host: Exploitation of Macrophage Polarization by Intracellular Bacterial Pathogens.” *Pathogens and Disease*, vol. 78, no. 1, Feb. 2020, https://doi.org/10.1093/femspd/ftaa009. Accessed 25 Feb. 2023.

“Usage — Pyseer 1.3.10 Documentation.” *Pyseer.readthedocs.io*, pyseer.readthedocs.io/en/master/usage.html#phenotype-and-covariates. Accessed 2 Mar. 2024.

Ushach, Irina, and Albert Zlotnik. “Biological Role of Granulocyte Macrophage Colony-Stimulating Factor (GM-CSF) and Macrophage Colony-Stimulating Factor (M-CSF) on Cells of the Myeloid Lineage.” *Journal of Leukocyte Biology*, vol. 100, no. 3, Sept. 2016, pp. 481–89, https://doi.org/10.1189/jlb.3RU0316-144R.

WHO. “Tuberculosis.” *World Health Organization*, World Health Organization, 7 Nov. 2023, www.who.int/news-room/fact-sheets/detail/tuberculosis.

Yang, Jiyeon, et al. “Monocyte and Macrophage Differentiation: Circulation Inflammatory Monocyte as Biomarker for Inflammatory Diseases.” *Biomarker Research*, vol. 2, no. 1, 2014, p. 1, https://doi.org/10.1186/2050-7771-2-1.

Yokota, Aya, et al. *GM-CSF and IL-4 Synergistically Trigger Dendritic Cells to Acquire Retinoic Acid-Producing Capacity*. no. 4, Apr. 2009, pp. 361–77, https://doi.org/10.1093/intimm/dxp003. Accessed 16 May 2023.