

DISSERTATION

COMPUTATIONAL TOOLS TO IDENTIFY CORRELATES OF VACCINE-INDUCED
PROTECTION AGAINST TUBERCULOSIS

Submitted by

Amy Fox

Department of Microbiology, Immunology and Pathology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2021

Doctoral Committee:

Advisor: Marcela Henao-Tamayo

Co-Advisor: Brooke Anderson

Zaid Abdo

Bailey Fosdick

Copyright by Amy Fox 2021

All rights reserved

ABSTRACT

COMPUTATIONAL TOOLS TO IDENTIFY CORRELATES OF VACCINE-INDUCED PROTECTION AGAINST TUBERCULOSIS

Tuberculosis is a significant threat to human health. While the BCG vaccine exists to protect children from disseminated forms of tuberculosis, it fails to protect against pulmonary tuberculosis. Thus, a better vaccine is needed. However, the immune system in response to tuberculosis and the BCG vaccine is incompletely understood. We sought to develop novel analysis methods to help understand the immune system. This dissertation describes an analysis tool, cyto-feature engineering, that rapidly identifies flow cytometry immune cell populations utilizing experimental controls. The tool was corroborated through testing the pipeline on different types of flow cytometry datasets. Cyto-feature engineering was then utilized to understand the immune response to two immunomodulatory drugs—losartan and propranolol—when used in conjunction with the BCG vaccine. This study identified an increase in T cell responses due to drug administration, but ultimately failed to decrease bacterial burden in the lung and spleen. Other studies employed a new method for identifying immune cells correlated with various metabolites in the context of tuberculosis. The method can be utilized to generate hypotheses from secondary data sources and gain new biological insight. Using this method, we identified a potential correlation between CD45RA and arachidonic acid metabolism which could serve as a potential target for future vaccination studies. The research outlined in this dissertation will hopefully lead to better immunological analyses of data and the development of a better tuberculosis vaccine.

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my advisors, Marcela Henao-Tamayo and Brooke Anderson. They have taught me invaluable skills, provided guidance and feedback, and pushed me to be a better researcher and scientist. I would also like to thank the other members of my committee, Zaid Abdo and Bailey Fosdick, for their generous time and thoughtful feedback and suggestions every time we met.

I would also like to acknowledge the contributions of various collaborators that played a role in my dissertation. I would like to thank Laylaa Ramos for collecting the minipig data used in Chapter 4 and Nurul Islam for helping with the metabolomics pre-processing and answering all my questions about LC-MS. I am also grateful to Bailey Fosdick for guidance in the linear model analysis for Chapter 4. Additionally, I would like to express my appreciation to Andres Obregon-Henao, Taru Dutt, and Burton Karger for helping with all the large mouse studies and providing support both inside and outside of the lab.

To all my friends and peers, thank you for your understanding and encouragement throughout the last few years. I'd like to thank members of my cohort: Lyndsey Gray, Bekah McMinn, Jaz Donkoh, Joanie Ryan, and Laura St. Clair for providing a support system while we simultaneously took on the journey of a PhD program. I'd also like to thank my Friday beer crew: Charlotte Avanzi, Camron Pearce, James Terry, and McKenzie Fletcher for lending a sympathetic ear and making the last few years memorable.

I'd also like to thank my parents and family for always encouraging me to find the fun in science and persevere when it gets tough. From first-grade science fair projects to today, you have inspired me to always be curious and taught me that hard work leads to success.

Finally, I would like to thank my husband, Brian, who has patiently waited for me for years.

Thank you for putting up with my incessant science talk, the stress and complaints, and sticking with me through a long-distance relationship while I worked towards this degree. Without your love and unconditional support, I would not be where I am today.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iii
CHAPTER 1 — REVIEW OF THE LITERATURE.....	1
Background.....	1
Tuberculosis: Epidemiology, Disease Outcomes, Diagnosis, Treatments.....	3
MTB Bacteria and Immune Response.....	6
Vaccines.....	15
Animal Models.....	21
Techniques and Tools for Studying Mycobacterium tuberculosis.....	23
Research Rationale and Research Aims.....	27
References.....	29
CHAPTER 2 — CYTO-FEATURE ENGINEERING: A PIPELINE FOR FLOW CYTOMETRY ANALYSIS TO UNCOVER IMMUNE POPULATIONS AND ASSOCIATIONS WITH DISEASE.....	40
Introduction.....	40
Materials/Methods.....	42
Results.....	46
Discussion.....	62
References.....	66
CHAPTER 3 — IMMUNOMODULATORY DRUGS AS VACCINES AGAINST MYCOBACTERIUM TUBERCULOSIS.....	69
Introduction.....	69
Materials/Methods.....	72
Results.....	75
Discussion.....	95
References.....	100
CHAPTER 4 — NOVEL TECHNIQUE TO STUDY IMMUNOMETABOLISM DURING MYCOBACTERIUM TUBERCULOSIS INFECTION.....	105
Introduction.....	105
Materials/Methods.....	108
Results.....	113
Discussion.....	133
References.....	136
CHAPTER 5 — CONCLUDING REMARKS.....	139

APPENDIX 1 — SUPPLEMENTAL MATERIALS FOR CYTO-FEATURE ENGINEERING: A PIPELINE FOR FLOW CYTOMETRY ANALYSIS TO UNCOVER IMMUNE POPULATIONS AND ASSOCIATIONS WITH DISEASE CYTO-FEATURE ENGINEERING: A PIPELINE FOR FLOW CYTOMETRY ANALYSIS TO UNCOVER IMMUNE POPULATIONS AND ASSOCIATIONS WITH DISEASE.....	143
APPENDIX 2 — SUPPLEMENTAL MATERIALS FOR IMMUNOMODULATORY DRUGS AS VACCINES AGAINST MYCOBACTERIUM TUBERCULOSIS.....	157
APPENDIX 3 — SUPPLEMENTAL MATERIALS FOR NOVEL TECHNIQUE TO STUDY IMMUNOMETABOLISM DURING MYCOBACTERIUM TUBERCULOSIS INFECTION.....	165
APPENDIX 4— ACQUISITION OF HIGH-QUALITY SPECTRAL FLOW CYTOMETRY DATA.....	167
Significance Statement.....	167
Abstract.....	167
Introduction.....	168
Strategic Planning.....	169
Basic Protocol 1: Preparation of single-cell suspension for flow cytometry.....	169
Support Protocol 1: Lung preparation.....	173
Support Protocol 2: Counting Cells on Flow Cytometer.....	175
Basic Protocol 2: Surface and Intracellular Flow Cytometry Staining.....	179
Support Protocol 3: Single-Color Bead Controls.....	183
Reagents and Solutions.....	185
Commentary.....	193
References.....	204
LIST OF ABBREVIATIONS.....	205

CHAPTER 1 — REVIEW OF THE LITERATURE

Background

Prior to the recent Covid-19 pandemic in 2020-2021, tuberculosis remained the world's leading cause of death from an infectious agent.^{1,2} In 2019 alone, 1.4 million people died from the disease, which is caused by the bacteria *Mycobacterium tuberculosis* (MTB).² While tuberculosis is not commonly reported in the United States, in high-burden countries, such as India, about 10 million people fall ill with TB each year.² With the emergence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) MTB in the last few decades, MTB is a persistent threat to global health. However, there are still many unknowns about MTB and the immune response. It is imperative to develop a better understanding of the immune system in response to MTB infection so that we can prevent millions of deaths each year.

First administered in 1921, the bacille Calmette-Guérin (BCG) vaccine is commonly administered to prevent active MTB infection in countries with a high prevalence of tuberculosis. While the vaccine protects children from tuberculosis meningitis and miliary disease, it only confers variable protection against pulmonary tuberculosis in adults.^{3,4} It is not fully understood why this variability exists, but some reasons point to host genetic differences, divergent BCG strains, and the existence of non-tuberculosis mycobacteria (NTM) in the environment.⁵⁻⁷ Further studies have noted that vaccines administered farther away from the equator confer better protection.⁸ Despite the ongoing research, it is unclear how these factors and many others impact the inadequate BCG vaccine protection.

As the current BCG vaccine is substandard, many groups have been working to develop better tuberculosis vaccines or vaccination strategies. Initial vaccine testing often involves *in vitro* studies followed by experimental infections with animal models such as mice. In these studies, the colony forming units (CFUs) are used to assess the bacterial burden and determine if the animals are able to limit bacterial replication better. Flow cytometry and metabolomics are other techniques that can be used to investigate, respectively, the immune cells and small molecules elicited due to either vaccines or infection. However, current methods to characterize immune cells are labor intensive and highly variable. Additionally, few computational tools exist that are able to combine all of these experimental outputs for a holistic view of the study.

There are many different unknowns such as how MTB interacts with and evades the immune response and why BCG does not adequately protect. To answer questions about how the complex immune system responds to MTB infection and vaccines, there are a number of questions one must ask themselves before beginning an experiment. Due to the \$1.1 billion tuberculosis funding deficit, do I have the funding to perform an experiment?^{9,10} How will the MTB and BCG strain affect the results as both can elicit different immune responses? Moreover, how will vaccination route and timing affect protection? As all animal models are an imperfect representation of humans, which animal model should I use? As all techniques and analyses have explicit biases that can affect readouts, what measurements, techniques, and analyses should I use to answer my question? Finally, do I have the patience and time to wait 4-6 weeks for MTB to grow on agar plates? Some relevant aspects about tuberculosis and these factors are described here in this literature review.

Tuberculosis: Epidemiology, Disease Outcomes, Diagnosis, Treatments

Epidemiology

There are three primary factors that affect epidemiology of a disease: pathogen factors, environmental factors, and host factors. Pathogen factors, such as virulence and drug resistance, can greatly affect the distribution and incidence of MTB. For example, certain MTB strains such as HN878 are more virulent than others (discussed in more detail in the `MTB Bacteria and Immune Response` section). Additionally, multidrug-resistant (MDR) and extensively drug-resistant (XDR) tuberculosis strains are increasingly becoming common.¹¹ These strains are typically resistant to the drugs isoniazid and rifampin which can prolong treatment times and provide longer contagious periods for MTB to infect others.¹¹ Environmental factors that can increase the risk of developing active TB include air pollution and overcrowded living conditions.¹² For every 10µg/m³ increase in the air pollutant sulfur dioxide, there is a 4.6% increase in pulmonary tuberculosis incidence.¹³ Additionally, host factors such as co-infections and comorbidities can also increase the risk that an infected person will advance to active disease. High-risk individuals include those who are immunocompromised, taking medications for organ transplants, or undergoing chemotherapy. For example, patients with Human Immunodeficiency Virus (HIV) are 18 times more likely to develop active tuberculosis infection.¹⁴ Other common risk factors include malnutrition, Vitamin D deficiency and diabetes.^{12,15,16} People with Vitamin D deficiency are 5 times more likely to develop tuberculosis, and those with diabetes mellitus are 3 times more likely to develop tuberculosis.^{15,16} These comorbidities can shift the probability that a person develops active infection and contribute to the complexity of understanding the immune response.

Disease outcomes

There are two outcomes of MTB infection: latent infection and active infection. People with latent tuberculosis are infected with MTB, but asymptomatic, whereas active tuberculosis infection is associated with many non-specific symptoms such as fever, cough, weight loss, and night sweats.¹⁷ While only 5-10% of people infected with MTB will develop active disease, it is still not known why some people develop active disease and others maintain latent tuberculosis.¹⁸

Once infected, tuberculosis disease can manifest in different forms. The most common form is pulmonary tuberculosis which affects the lungs.¹⁹ Other disseminated forms such as miliary tuberculosis where bacteria enter the bloodstream and travels to other organs and meningeal tuberculosis which causes inflammation of the membranes around the brain and spinal cord also exist.^{19,20} These disseminated forms of tuberculosis are particularly difficult to diagnose due to the limited tools available.²¹

Diagnosis

Because of the nonspecific tuberculosis symptoms, it can be difficult to quickly diagnose patients. Diagnostic methods include sputum microscopy and culturing, tuberculin skin tests, and molecular PCR tests. Some of these methods can take weeks to diagnose or require follow-ups in the hospital which can be challenging in resource-limited areas. Historically, sputum microscopy and culturing have been used in low- and middle-income countries.²² Sputum microscopy involves coughing up sputum and visualizing bacteria with a microscope. This microscopy method only has a 34-80% sensitivity, and when there are less than 10,000 bacilli per milliliter, it

is highly unlikely that MTB will be diagnosed.^{23,24} For confirmatory microscopy results, the clinic must then culture the bacteria for up to 6 weeks to confirm that the bacteria is in fact MTB.²⁵ The Mantoux tuberculin skin test is an alternative diagnostic that involves injecting purified protein derivative (PPD) from MTB into the arm and examining reactivity.²⁶ If a person has been exposed to MTB, their body will respond to the PPD antigens and begin swelling at the site of PPD administration.²⁶ Test positivity is measured by inspecting this diameter of firm swelling.²⁶ However, it often gives false positive results in patients who have been vaccinated with BCG.²⁷ This removes the utility of the test in many areas because countries with high tuberculosis-burden often have high BCG-vaccination rates.²⁸ This test also requires a follow-up visit to a clinic within 48-72 hours, making it troublesome for patients who live far away.²⁶ Importantly, recent advances in technology have offered a new, rapid diagnostic. The Xpert MTB/RIF PCR test is now recommended by the World Health Organization, if available.²⁹ The Xpert MTB/RIF PCR test can diagnose MTB and drug resistance to rifampin in 2 hours.³⁰ With a sensitivity of 85% and specificity of 98%, the Xpert provides a better alternative to conventional diagnostic methods.³¹ Despite these recent advances, diagnosis is only the first step in treating TB.

Treatment

Although tuberculosis is generally considered a treatable disease, treatment can be a lengthy, brutal process. Patients are treated with multiple drugs for a minimum of 6 months. A typical regimen includes taking four drugs—isoniazid, rifampin, pyrazinamide, and ethambutol—in combination for two months, followed by isoniazid and rifampin for four additional months.³² This rigorous drug regimen can be toxic, causing liver and kidney injury, and has upsetting side

effects such as nausea, vomiting, and diarrhea, among others.³³⁻³⁵ Further, the use of these drugs must be monitored to prevent the emergence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) tuberculosis.

In addition to the complications with MDR and XDR emergence, the treatment of tuberculosis is exceedingly expensive. In the United States, it costs approximately \$17,000 to treat a single tuberculosis patient.³⁶ The cost, however, skyrockets when an MDR or XDR case is identified, costing \$134,000 and \$430,000 per patient, respectively.³⁶ Therefore, a vaccine that can prevent disease and this rigorous treatment would be a powerful tool in the fight against MTB.

MTB Bacteria and Immune Response

While research has been ongoing for decades, there are still many unknowns regarding specific defense mechanisms of MTB and how the immune system can be modulated to fight MTB infection. It is known that certain immune cells play important roles in infection, but conflicting research has shown both protective and non-protective qualities of most cell types and it is not known which combination of cell responses are needed to elicit superior immunity.

Bacteria

As MTB has been co-evolving with humans for at least 15,000 years, it has developed numerous mechanisms for evading the host immune response.³⁷ *Mycobacterium tuberculosis* is unique in that it is a Gram-positive bacteria, but it also retains properties of Gram-negative bacteria.³⁸ This means that the exceptionally thick layer of peptidoglycan typical of a Gram-positive bacteria and the outer lipid membrane typical of a Gram-negative bacteria provide additional defense against

antibiotics.³⁸ MTB uses these layers and a variety of mechanisms and pathways to evade the immune response. One example is the MTB secretory systems which contribute to virulence. Secretory systems ESX 1-5 release proteins that help MTB establish infection through breaking the phagosomal membrane, disrupting phagosome maturation, and curtailing MHC II antigen presentation.^{39,40} All of these components and many others help MTB multiply and play a role in pathogenicity.

Different strains of MTB also have varying levels of pathogenicity and virulence that can affect laboratory results based on the specific strain used. Much of the previous work studying MTB has used two laboratory strains—H37RV and Erdman.⁴¹ H37RV was isolated from a patient in 1905, and Erdman was isolated in 1945.⁴¹ In the decades since they have been isolated, the lab strains have been passaged many times. Each of these passages has the opportunity for genetic mutations to occur. Thus, these strains may now be more suited to a laboratory setting and non-representative of the genetic heterogeneity present in the world.⁴¹ To better represent the relevant MTB strains currently circulating in the world, more recent clinical strains, such as those from the Beijing family, are being used to evaluate the immunopathogenesis of this disease. These Beijing strains are characterized by high virulence and drug resistance.⁴² It is also thought that these strains may be more resistant to the BCG vaccine or that BCG vaccination may even facilitate spread of Beijing strains.^{43,44} A more recently isolated Beijing strain, HN878 is increasingly being used now in research. HN878 is an MTB strain that is characterized by high virulence; it caused 60 tuberculosis cases in Houston, TX from 1995-1998.^{41,45} Because the HN878 strain has not been passaged for over 100 years like H37RV, it more closely resembles

the MTB that is circulating and involved in active infections. Therefore, HN878 may be an advantageous MTB strain to use in research.

Innate Immune Response

The innate immune response is generally considered the first line of defense against an invading pathogen. MTB is transmitted through airborne particles when an infected person coughs or sneezes and is considered highly transmissible as it only takes 1-10 bacilli to become infected.⁴⁶

Once inhaled by an uninfected individual, the bacteria travel from the mouth or nose to the alveoli in the lungs. At this point, immune cells including neutrophils, alveolar macrophages, and dendritic cells use pattern recognition receptors to recognize specific pathogen-associated molecular patterns (PAMPs) on foreign bacteria. Once recognized, the bacteria are then phagocytosed. From here, the innate and adaptive immune response are intimately linked through a series of complex interactions.

Neutrophils are one of the first lines of defense against foreign particles such as bacteria. When neutrophils phagocytose bacteria, they can kill invading intracellular pathogens through releasing bactericidal proteins, reactive oxygen species, and nitric oxide.^{47,48} Neutrophils can also trap and kill extracellular bacteria through releasing neutrophil extracellular traps (NETs).⁴⁹ While these structures are efficient killers, they can also cause tissue damage in the surrounding tissue.⁴⁹

There is conflicting information on the role of neutrophils in MTB infection. Some studies show that neutrophils may be unable to effectively kill MTB.^{50,51} Researchers consider this ability for MTB to reside within neutrophils but not be killed as a potential mechanism for MTB to hide from other phagocytic cells.⁵¹ However, other studies show that neutrophil levels are associated

with tuberculosis protection.⁵² This points to an interactive role that neutrophils may play with other protective cells.⁵² This interactive role can be shown through the fact that neutrophils present antigens to T lymphocytes and promote dendritic cell maturation by binding to dendritic cells.⁵² Neutrophils also attract dendritic cells, monocytes, and lymphocytes by releasing cytokines and chemokines; these cells can, in turn, offer protection.⁵² In these ways, neutrophils play an important role in initiating the immune response, though more research is still needed to further elucidate neutrophilic MTB concealment and how this contributes to MTB clearance.

Macrophages recognize pathogen-associated molecular patterns (PAMPs) on MTB via Toll-like receptors (TLR)—primarily TLR2, TLR4, and TLR9.⁵³ TLR2 recognizes lipids such as lipoarabinomannan and mannosylated phosphatidylinositol in the cell wall of MTB and lipoproteins such as LpqH and LprG found in MTB membrane vesicles.⁵³ On the other hand, TLR4 binds with lipopolysaccharide found in Gram-negative bacteria, as well as lipomannan and mycobacterial proteins.⁵³ Finally, TLR9 recognizes undermethylated CpG motifs in MTB DNA.⁵³ The activation of these receptors causes downstream activation of NF- κ B which initiates the production of pro-inflammatory cytokines such as TNF, IL-1 β , and IL-12.⁵³ These cytokines trigger apoptosis of the macrophages and recruit other immune cells such as neutrophils and dendritic cells to the lungs. Additionally, IL-12 secreted by the macrophages promotes the production of IFN- γ by T cells and NK cells.⁵³ Another role that macrophages play is in granuloma formation. Infected macrophages release cytokines that recruit other macrophages, neutrophils, B cells and T cells.⁵⁴ The cells aggregate around the infected macrophages, and this mass of cells becomes fibrotic, walling itself off to infection.⁵³ While these granulomas can prevent MTB from dispersing to other areas of the lung, they also provide a location inaccessible

to drugs where the bacteria can grow.^{55,56} The primary type of granuloma in TB is the caseous granuloma, though others such as necrotizing neutrophilic, non-necrotizing, and fibrotic also exist.⁵⁷ Caseous granulomas are formed around a necrotic area with a layer of macrophages and an outermost layer of T cells and B cells.⁵⁷ Non-necrotizing granulomas, however, are primarily composed of macrophages.⁵⁷ These granuloma types can either be protective or can promote transmission of MTB. For example, granulomas with high numbers of neutrophils after granuloma necrosis can lead to transmission of MTB.⁵⁸ However, granulomas with fibrotic tissue can sequester MTB, controlling growth.⁵⁸ Thus, macrophages can exhibit both protective qualities through release of pro-inflammatory cytokines and MTB sequestration and non-protective qualities by providing a niche for MTB to grow and disseminate.

When dendritic cells uptake bacteria, the cells mature and migrate to the draining lymph node where they present the bacterial antigens to naive T cells.⁵⁹ Depending on the costimulatory molecules and Major Histocompatibility Complex (MHC), dendritic cells stimulate the activation of CD8 T cells (through MHCI) and differentiation of CD4 T cells (through MHCII) into Th1, Th2, or regulatory T cells which then migrate to the lungs.⁵⁹ Once back at the site of infection, these cells produce cytokines that are either able to control infection or assist it. Some studies show that MTB-infected dendritic cells secrete IL-12 and IFN- α , which induces the production of IFN- γ .⁶⁰ This IFN- γ has been shown to inhibit MTB replication and induce apoptosis of macrophages with high MTB bacterial loads.^{60,61} However, other studies show that MTB-infected dendritic cells have impaired maturation and ability to induce antigen-specific T cells, thereby hindering the immune response to MTB.^{62,63} Consequently, while we know that

dendritic cells are crucial for protection, it is not fully understood how MTB modulates dendritic cell ability to fight infection.

Adaptive Immune Response

The pathogen-specific adaptive immune response develops after encountering a pathogen for the first time. This adaptive immune response is generally composed of T cells and B cells, described below.

T cells develop in the thymus and are considered naive until they first encounter antigens presented by dendritic cells in the lymphoid organs.⁶⁴ These T cells then undergo clonal expansion and can differentiate into various T cell subsets specified by the cytokines present during antigen exposure.⁶⁴ Following activation, the T cells enter the bloodstream and travel to peripheral tissues to fight infection.⁶⁴

As evidenced through depletion studies, CD4 T cells are critical for controlling MTB infection.⁶⁵ CD4 T cells are capable of producing both proinflammatory and anti-inflammatory cytokines. In the presence of IL-12 and IFN- γ , CD4 T cells differentiate into T Helper 1 (Th1) cells.⁶⁶ Th1 cells are generally associated with combating intracellular pathogens.⁶⁶ In the context of tuberculosis, Th1 cells produce cytokines which are important in controlling MTB infection.⁶⁷ However, the cytokines IFN- γ and TNF- α alone are not enough to protect against MTB. Another cell type produced in response to infection are T Helper 2 (Th2) cells. Th2 cells generally combat extracellular pathogens and are involved in allergic inflammation; they are unable to protect against MTB infection.⁶⁸ Th2 cells express IL-4, IL-5, and IL-13 and are induced by

presence of IL-4.⁶⁶ GATA3, a Th2 regulator, suppresses differentiation of cells into Th1, and T-bet, a Th1 regulator, suppresses differentiation of cells into Th2.^{66,69} Thus, a delicate equilibrium is formed between Th1 and Th2 cells. An additional cell type important during infection are Regulatory T cells (Treg). Tregs produce IL-10, an anti-inflammatory cytokine that regulates the inflammatory response.⁷⁰ The timing and number of these Tregs is crucial. If too many Tregs infiltrate at the beginning of infection, they can downregulate inflammation and the protective immune response too much and allow pathogens to establish chronic infection.⁷¹ However, if there are too few Tregs, especially late in infection, inflammation caused by immune cells can lead to tissue damage.⁷¹ Similarly, a balance of Tregs is needed to control infection and prevent tissue damage. While we know that a balance of Th1, Th2, and Tregs is essential, there is a need to understand the optimal number of Th1, Th2, and Tregs necessary for protection.

CD8 T cells have two main effector functions to kill infected cells. The first function is to release cytotoxic granules such as perforin, a protein that binds to the plasma membrane of target cells and forms pores, and granzymes.⁶⁴ Cytotoxic granzymes can then initiate the caspase pathway and programmed cell death.⁶⁴ The role of granzymes and perforin in MTB infection is contested. Some research shows that these granules are capable of both killing extracellular MTB and reducing the number of viable intracellular MTB.⁷² Additional research shows that CD8 T cells utilize perforin to lyse MTB infected macrophages.⁷³ However, conflicting evidence has shown that the loss of perforin in knockout mice does not affect the course of MTB infection.⁷⁴ This may be explained by a perforin-independent method by human MTB-specific CD8 T cells for killing MTB.⁷⁵ The second mechanism for killing infected cells involves producing cytokines, such as IFN- γ , TNF- α , and IL-10. While IFN- γ can initiate apoptosis in infected macrophages,

TNF- α can initiate either apoptosis or necrosis of target cells.^{61,76} Whereas apoptosis results in decreasing the viability of MTB, necrosis is actually a mechanism by which MTB can escape macrophages and disseminate.⁷⁷ Thus, the expression of TNF- α can have opposing effects on the pathogenicity of MTB. Additionally, although IL-10 is typically considered an anti-inflammatory cytokine, recent human studies have identified antigen-specific CD8 T cells that produce Th2 cytokines such as IL-10; these cells have been shown to inhibit MTB growth and activate B cells.⁷⁸ Thus CD8 T cells play an important role in conjunction with other immune cells in combating MTB infection, but more research is needed to understand the optimal MTB killing mechanisms.

Memory T cells are a subset of long-lived antigen-specific CD4+ or CD8+ T cells that can provide protection upon reexposure.⁷⁹ While it is not fully understood how memory T cells are generated, there is evidence to support the theory that memory T cells are derived from effector cells.^{80,81} There are four main types of memory T cells—effector memory, central memory, and tissue-resident memory, and stem-cell like memory—which vary in their function and homing abilities. Effector memory T cells are typically found in circulation and peripheral tissues such as the lungs, skin, and gut.^{82,83} As these cells are found at initial sites of infection, they provide a first line of defense against an invading pathogen.⁸² It has been shown that BCG can elicit the production of these effector memory T cells which are crucial for protection.⁸⁴ Tissue-resident memory T cells, on the other hand, are maintained in specific tissues and do not recirculate.^{85,86} In this way, they can also act as the first line of defense against pathogens. Increases in these tissue-resident memory T cells in the lungs have been associated with better protection from MTB.^{87,88} If the initial site of infection is not contained, central memory cells can fight off

pathogens in lymphoid organs such as the lymph nodes, spleen, and bone marrow.^{82,83} It has been shown that CD4⁺ KLRG- IL-2⁺ central memory T cells can control MTB growth even when the population is boosted late in infection.⁸⁹ However, it is hypothesized that BCG does not adequately establish central memory T cells populations following vaccination.⁸⁴ Finally, stem cell-like memory T cells are capable of proliferating and differentiating into multiple different memory phenotypes.⁹⁰ During MTB infection, these stem cell-like memory T cells are induced and produce Th1 cytokines, which may offer protection.⁹¹ Interestingly, all of these cell types seem to have certain plasticity with the ability for central memory and effector T cells to differentiate into tissue-resident memory T cells.⁸⁵ While these cell types have all been shown to have some protective capabilities, there is a need for more research to elucidate the specific levels of each cell type required for protection.

While the utility of B cells in fighting MTB infection has only recently been recognized, there is increasing evidence to support the important role of B cells in fighting MTB infection through production of antibodies, secretion of cytokines, and presentation of antigens. During development, B cells undergo VDJ recombination in which a set of unique, diverse antigen receptors are produced.⁹² This allows the B cells to recognize a variety of different molecules. B cells are then activated when their B cell receptor binds to antigens; the B cells class switch, allowing them to produce either IgA, IgG, or IgE.⁹² These antibodies bind to foreign material and can neutralize pathogens, target cells for opsonization or lyse bacterial cells through activating the complement system.⁹³ In MTB infection, antibodies produced by B cells play a role in defense by targeting extracellular MTB for opsonization.^{94,95} B cells also secrete a variety of pro-inflammatory and anti-inflammatory cytokines that can modulate the immune response.⁹⁶ Two

examples of cytokines produced by B cells are IL-6 and IL-21. The production of IL-6 by B cells promotes polarization of Th1 cells, and IL-21 recruits T cells to the lungs and promotes CD8+ T cell priming.^{93,96,97} These T cells can then proliferate and kill MTB at the site of infection in the lungs. Finally, activated B cells can promote dendritic cell maturation and present MTB antigens to CD4+ T cells.^{98,99} Due to the neglect for MTB B cell research for many years, there is a need for more concerted efforts to understand the full capacity of B cells to fight MTB infection.

While extensive research has been performed to better understand the immune response to MTB, there are still many questions about how the immune system responds and protects against MTB infection. Thus, it is challenging to develop a vaccine or boosting strategy because a protective profile of cells does not yet exist.

Vaccines

Vaccines are crucial for eliminating diseases. In the development of a vaccine, there are many factors that can affect efficacy including the type of vaccine, the vaccination route, and the timing of vaccine administration. All of these factors also play a role in the lengthy timeline for developing a vaccine. Specific to the BCG, differing efficacy of vaccine strains also affect research and protection against MTB across the world. These factors are described briefly below.

Types of Vaccines

There are three main types of vaccines typically tested to combat MTB disease: 1) live, attenuated vaccines, 2) inactivated/killed vaccines, and 3) subunit/recombinant vaccines.

Attenuated live vaccines are composed of live bacteria or viruses that have been weakened to the point that they are incapable of causing disease.¹⁰⁰ As these bacteria and viruses can replicate and similarly mimic natural infection, they typically produce a strong, long-lasting immune response.¹⁰⁰ There are multiple ways to develop an attenuated vaccine. Traditionally, bacteria were cultured and passaged many times until they were less virulent or unable to cause disease in humans.¹⁰¹ This is how BCG was initially produced by Albert Calmette and Camille Guérin. They cultured *Mycobacterium bovis*, a *Mycobacteria* similar to MTB that causes disease in cattle, in a medium of ox bile, glycerin, and potato until the bacteria no longer caused tuberculosis in animal models.¹⁰² Now with the use of CRISPR/Cas9, genes relating to pathogenicity can be deleted to create a safer version of a replicating virus or bacteria while preserving many of the epitopes.^{103,104} An example of a live, attenuated tuberculosis vaccine developed in this way is the ESX5 KO. This vaccine utilizes an MTB mutant with the ESX5 secretion system deleted.¹⁰⁵ This attenuated vaccine has shown to be protective against MTB when used in conjunction with BCG in both mice and guinea pigs.¹⁰⁵ One disadvantage of live attenuated vaccines is that they have the ability to mutate and revert back to a pathogen that can cause disease.¹⁰⁰ Another disadvantage is that these vaccines must maintain cold-chain, and they may not be suitable for immunocompromised individuals.¹⁰⁰

Inactivated or killed vaccines typically use chemicals or heat to eradicate a pathogen's ability to replicate.¹⁰⁶ Because the pathogen cannot replicate or cause disease, they have better safety profiles for immunocompromised individuals than live vaccines.¹⁰⁶ However, because their protection is not as long lasting, a booster may be needed.¹⁰⁶ A heat-killed *Mycobacterium vaccae* vaccine for tuberculosis was tested in the 1990s, and while it was capable of stimulating

production of CD8+ T cells *in vitro*, it did not offer significant protection in human clinical studies.^{107,108}

Subunit or recombinant vaccines utilize specific parts of a pathogen, such as proteins and antigens, to generate an immune response.¹⁰⁹ They typically create a strong immune response to that specific antigen, however, because they only contain some parts of a pathogen, they are not as immunogenic as inactivated or live vaccines.¹¹⁰ Thus, these subunit vaccines can be given to immunocompromised individuals because the subunits of a pathogen alone cannot cause disease.¹⁰⁹ Recombinant vaccines, specifically, insert the DNA of an antigen into bacteria, yeast, or cells to stimulate the immune response.¹¹¹ MVA85A is an example of a recombinant vaccine tested against MTB. It is composed of a modified Vaccinia Ankara virus that expresses the protein 85A found in *Mycobacterium tuberculosis*.¹¹² While the vaccine showed promise in protecting against MTB in animals, it proved ineffective in human trials.¹¹²

Vaccination route

The route of administration can greatly affect the efficacy of vaccines. There are six main vaccination routes: intradermal, subcutaneous, intramuscular, oral, intranasal, and intravenous. While intradermal vaccines are injected into the dermis, or top layer of skin, subcutaneous vaccines are delivered to the subcutaneous tissue which lies between the skin and the muscles (Figure 1.1). Intramuscular vaccines are injected into the muscle. Vaccines with adjuvants that cause strong local inflammation are typically given intramuscularly to avoid skin irritation.¹¹³ Oral and intranasal vaccines are advantageous in that there is relatively no pain involved in these vaccinations which are delivered to the mouth and nose, respectively. Fairly uncommon for

vaccines, intravenous injections are injected directly into the veins to rapidly enter the bloodstream.

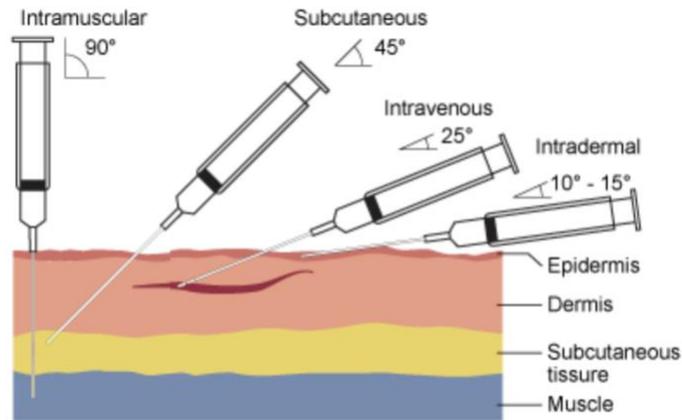


Figure 1.1: Vaccination route figure from “Clinical Procedures for Safer Patient Care” by Glynda Rees Doyle and Jodie Anita McCutcheon, used under a [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/).¹¹⁴ Vaccination routes illustrating intramuscular, subcutaneous, intravenous, and intradermal injections.

Typically, BCG is given to neonates via an intradermal injection.¹¹³ In a recent paper by Darrah et al., researchers compared MTB protection from intradermal, aerosolized, and intravenous BCG immunization in non-human primates.¹¹⁵ Intravenous vaccination induced more antigen-specific T cells, and six out of the ten intravenously vaccinated animals had no detectable MTB or signs of infection following infection.¹¹⁵ However, intravenous delivery of BCG to neonates has many challenges including ease of delivery.¹¹⁶ Even an intradermal injection is supposedly difficult to administer to babies.¹¹³ If an oral or intranasal vaccination route exhibited protective qualities, they may provide a better alternative to an intravenous or intradermal injection as they would be easier to administer to babies.

Vaccination Timing

Another factor to consider in developing vaccine studies is timing. Typically, the BCG vaccine is given at birth, but studies have shown that pushing back the time from birth to 10 weeks can increase the numbers of BCG-specific T cells that express IFN- γ , TNF- α , and IL-2.¹¹⁷ When considering a BCG boosting strategy the complexity of identifying the optimal time to give both the BCG and booster vaccine greatly increases.

Vaccine Development

The average time for vaccine development is 10-15 years.¹¹⁸ This vaccine development pipeline has six main stages (Figure 1.2). Traditionally, in the discovery research phase, basic laboratory research is performed to identify target antigens of interest, which may take 2-4 years.¹¹⁹ In the preclinical stage, experiments are performed *in vitro* in cells and then, if promising, in animal models.¹²⁰ Research can then move into vaccine trials in humans after approval by the FDA to begin testing. Phase I trials determine the safety of the vaccine in fewer than 100 individuals.¹²⁰ Phase II trials increase the number of participants to the hundreds and test the dosing, immunogenicity, and safety.¹²⁰ Finally, in Phase III trials, vaccines are tested to determine efficacy, side effects, and safety in a larger group of thousands of patients.¹²⁰ These vaccine trials can be exceedingly expensive and on average there is only a 6% percent chance of success.¹²¹ The cost of moving a vaccine from preclinical trials through the end of Phase 2a trials can cost on average \$31-68 million.¹²¹

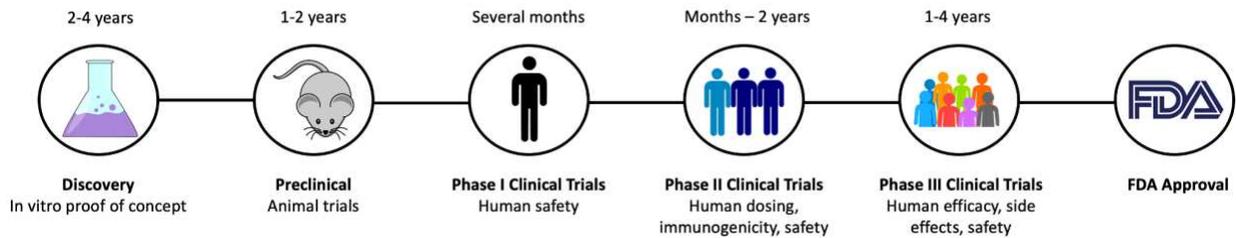


Figure 1.2: Vaccine Development and Timeline. It can take between 10-15 years for vaccines to move from the discovery phase through FDA approval.

In August 2020, there were 14 tuberculosis vaccine candidates in clinical trials.¹²² The most promising candidate is M72/AS01E which has shown about a 50% efficacy in preventing active TB infection during Phase II trials.² However, further studies in Phase III need to be conducted to confirm these results.

BCG Strain Efficacy

When Calmette and Guérin successfully created the BCG vaccine by serially passaging *Mycobacterium bovis*, the vaccine was disseminated to various laboratories across the world.¹²³ These strains were further passaged and genetically divergent daughter strains were produced.¹²³ The different strains produce distinct immune responses, different rates of adverse events and susceptibility to anti-tuberculosis drugs, and varying efficacy against MTB.⁷ Despite the widespread use of BCG, there is insufficient research to determine which BCG strain has the best protection.⁷ This paucity of data is due to many complexities. For example, there is surprisingly little documentation on the strains that countries utilize, and some countries use multiple strains of BCG to vaccinate individuals.⁷ Therefore, in designing vaccine studies, the specific BCG strain used can significantly impact results.

Animal Models

Although animal models cannot provide direct comparisons to human infection, it is not possible to test every potential vaccine or treatment on humans first. While humans are the only natural hosts of MTB, animal models can nevertheless provide valuable information about general pathogenesis and immunity.¹²⁴ Four animal models used in MTB research are described briefly below—mouse, guinea pig, minipig, and non-human primate.

Mice are commonly used in MTB studies for a multitude of reasons. They are inexpensive, easy to house, multiply quickly and can be genetically manipulated to study specific immune functions. There are similarly extensive immunological reagents for use in the murine model. Different strains of mice have varying susceptibility to MTB. C57BL/6 are more resistant to MTB than BALB/c mice which are more resistant than C3Heb/FeJ mice.^{125,126} However, one advantage of C3Heb/FeJ mice is that they form granuloma types that are more similar to humans such as necrotic tubercles.¹²⁷ Each of these mouse strains upregulates different genes and elicits varying levels of immune cells and can be used to study different aspects of disease and better understand susceptibility to disease.¹²⁸ While mice are a cost effective and simple model to work with, they do not exactly replicate human immunology and pathology. For example, mice do not develop latent infection.¹²⁹ Further, many mouse models used in studies are inbred. These inbred models offer advantages in that they are genetically similar, so the output measures have lower variability, and the results are more likely to be reproducible.¹³⁰ However, results with these mice do not account for genetic diversity exhibited in the real world. Recent innovations in the field of animal research have led to the development of a “Collaborative Cross” mouse population. The Collaborative Cross (CC) consists of recombinant inbred mouse lines derived

from eight inbred founder strains.¹³¹ This has led to high genetic variation within the CC strain which more closely mimics human genetic diversity.¹³¹ Nonetheless, using these mice in research requires larger sample sizes due to greater result variability.

Guinea pigs are another animal model used to study MTB. Guinea pigs can become infected with MTB after very low doses (20-50 bacilli), making them highly susceptible to MTB; almost all low-dose infections result in terminal endpoints 30 weeks after infection.¹³² Guinea pigs also exhibit pathology more similar to humans than mice. Lesions in guinea pigs can progress to necrosis similar to humans.¹³³ This is an important factor in MTB research as human necrotic tissue can shelter live MTB bacteria and offer protection from drugs.¹³⁴ Guinea pigs are also capable of producing CD1-restricted T cells.¹³⁵ In humans, group 1 CD1 molecules play an important role in MTB protection by presenting mycobacterium lipids and antigens to T cells.¹³⁵ Mice do not have these group1 CD1 molecules, providing another advantage to the guinea pig model.¹³⁵ However, there are few immunological reagents commercially available for guinea pigs.

Minipigs are another MTB animal model that have recently undergone evaluation.¹³⁶⁻¹³⁸ They are capable of coughing and sneezing, and natural transmission of MTB from infected to uninfected animals co-housed together has also been shown.¹³⁶ Physiologically, minipigs also have a highly similar lung structure to humans, and pathology in infected minipigs shows human-like heterogeneity of pulmonary lesions.¹³⁶ Pigs also produce double positive (CD4+ CD8+) T cells like those found in humans.^{139,140} In non-human primates, these double positive cells have been shown to exhibit cytolytic markers and Th1 cytokines and may play an important role in

controlling MTB infection.¹⁴¹ However, there are few immunological reagents available for minipigs. Another limitation with the minipig model is that they must be housed in large animal facilities as the adult minipigs can weigh between 45 and 220 pounds (depending on the breed).¹⁴² Moreover, due to funding scarcity, few studies have been conducted with MTB and minipigs to solidify the minipig as a satisfactory animal model for tuberculosis.

Non-human primates infected with MTB most closely resemble human disease as non-human primates are genetically, physiologically, and immunologically similar to humans.¹²⁴ Typically, vaccines and therapies are tested in other animal models before proceeding to use in non-human primates. Unlike other animal models, macaques exhibit both latent and chronic progressive infection.¹²⁴ They also develop a variety of lesions highly similar to human disease.⁵⁵ However, these animals are expensive and there are relatively few facilities in the US that can house these animals and perform these experiments.^{55,143}

Techniques and Tools for Studying *Mycobacterium tuberculosis*

In MTB research, there are a variety of tools used to characterize the immune response, study disease state, and identify small molecules. The two main techniques—flow cytometry and metabolomics—utilized in this dissertation are described here.

Flow cytometry

Flow cytometry is an example of a technique used to analyze the properties of cells. Cells are stained with fluorescently conjugated antibodies and directed single file through a flow cytometer. A laser light source excites the fluorescent antibodies causing them to emit light that

can then be measured. This gives information on the type and function of cells present in a sample.

Traditional analysis of flow cytometry data occurs on two-dimensional plots. Manual selection of cells on these plots, often referred to as “gating”, is how cell populations are identified. This has typically been done using controls called Fluorescence Minus One (FMOs).¹⁴⁴ These samples contain all of the markers within a panel except for one. This allows the user to know the amount of spillover from other fluorescent markers into each channel. A gate is placed around this negative population in FMOs and can be subsequently copied onto all samples (Figure 1.3). While this method has traditionally been used to gate cells, it is both subjective and time consuming.

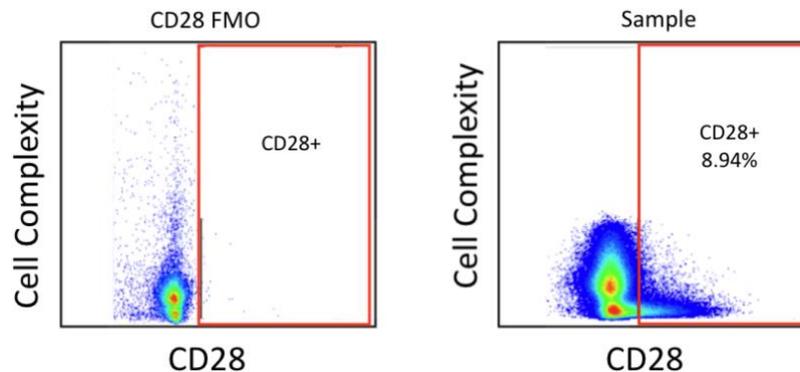


Figure 1.3: Using FMOs to gate a sample. The CD28 FMO contains all markers within the panel except for CD28 (left). Therefore, a threshold gate can be placed to show that any signal greater than what is shown in the FMO plot should be positive. This gate can then be copied onto the sample to identify the percentage of cells that express CD28 (right).

Efforts have been made to develop tools to analyze flow cytometry data more efficiently. Some of these tools include t-Distributed Stochastic Neighbor Embedding (t-SNE) and flow Self Organizing Maps (FlowSOM), among others.

t-SNE is a machine learning algorithm that transforms high-dimensional data into two-dimensional or three-dimensional plots.¹⁴⁵ It utilizes a Gaussian probability distribution to calculate similarities between data points.¹⁴⁵ While it is able to retain local similarities within the data, global similarities are often lost.¹⁴⁶ Further, this method is typically computationally intensive; it requires downsampling, or random sampling, to select only some of the cells from each sample.¹⁴⁷ While this shortens the amount of time to run the algorithm, important information can be lost.¹⁴⁷ This method also has difficulties identifying rare cells as many of the cells are removed in downsampling. Recent optimization in the code, referred to as opt-SNE, has reduced the computation time and improved data resolution, but datasets must still be downsampled and rare populations may still be lost.¹⁴⁷

FlowSOM is another machine learning algorithm developed in 2015.¹⁴⁸ It clusters cells into nodes based on similar marker expression. It then builds a minimal spanning tree to connect nodes that are most similar. While this algorithm provides great information on overall trends in the data, because each node is an average of all of the cells within the node, rare populations are often lost. Additionally, the expression of each marker within a node is not representative of all cells within the node.

While both of these analysis methods can provide information on the structure of the flow cytometry data, they are not able to incorporate important controls such as FMOs to determine positive and negative expression of markers. Further, these methods are not able to easily incorporate other data from the experiment such as colony forming units.

Metabolomics

Metabolomics is another measure that can be used to study disease in humans or animals.

Metabolites are small molecules formed during metabolism. Identification of metabolites during disease can help aid in diagnosis. It can also help understanding the underlying processes that are occurring during health and infection.

While genes give us valuable information on encoded capabilities, genes are subject to epigenetic changes.¹⁴⁹ Proteins, created from genes, give a closer look at processes taking place, but proteins are subject to post-translational changes.¹⁴⁹ Metabolites give us the most accurate representation of what is happening both inside and outside of cells. There are a variety of metabolites and metabolic pathways that play a role in MTB infection. MTB has also been shown to cause changes in cellular metabolic pathways that in turn affect cell function.¹⁵⁰

One machine used to study metabolomics is Liquid Chromatography-Mass Spectrometry (LC-MS). LC-MS utilizes the physical separation of liquid chromatography with mass analysis in mass spectrometry to detect compounds in a sample.¹⁵¹ Samples are injected into the liquid chromatography portion of the machine and analytes are separated based on interactions with the stationary phase such as polarity. The strength of these interactions determines how quickly the analytes flow through the column. The time it takes for these analytes to flow through the column is called retention time. The analytes are then vaporized and charged to an ionized state where magnetic fields are used to record the mass-to-charge ratio and abundance of each ion.

Analysis of metabolomics data requires peak detection to pick out the different components in the sample.^{152,153} Next, peaks are aligned across samples to correct for differing retention times.^{152,153} After alignment, hundreds or thousands of t-tests are used to calculate the differences between experimental groups.^{152,154} The analytes with differences can then be putatively identified or a pathway analysis can be run to determine the putative pathway differences between groups. One limitation of this method is that it is a univariate method for identifying interesting metabolites. In many experiments, there are covariates of interest that cannot be accounted for with this method such as time point and vaccination status. Furthermore, these methods cannot be combined with other dependent variables obtained during the experiment such as abundances of immune cells or bacterial burden.

Research Rationale and Research Aims

There is a deep need to develop a better tuberculosis vaccine. In order to do this, we must better understand the relationship between the immune response, MTB, and candidate vaccines. These relationships can be examined using novel tools to help us gain additional knowledge. The research outlined in this dissertation has three main goals:

Specific Aim 1: To develop a flow cytometry analysis pipeline that aids in identification of immune cell populations. *Hypothesis:* Researchers using this pipeline will be able to analyze data more efficiently and with less bias. In addition, results will be comparable to traditional gating analysis methods.

Specific Aim 2: To test a novel BCG vaccination boosting strategy utilizing two drugs—losartan and propranolol. *Hypothesis:* The use of losartan and propranolol will induce better protection against *Mycobacterium tuberculosis* infection by reducing inflammation, enhancing antigen presentation, and promoting T cell polarization to Th1 cells.

Specific Aim 3: To use a novel integrative metabolomics approach to identify correlations between immune cells and metabolites during *Mycobacterium tuberculosis* infection.

Hypothesis: There are metabolic profiles correlated with immune cells that can be identified via a novel flow cytometry and metabolomics analysis. Further, metabolites associated with glycolysis, the pentose phosphate pathways, and amino acid synthesis pathways will have a positive correlation with activated T cells.

The research in this dissertation aims to provide tools for analyzing large preclinical datasets and understanding how the immune system responds to both MTB and vaccines. These tools can hopefully provide researchers with new insights that can lead to better tuberculosis vaccines.

REFERENCES

1. World Health Organization. (2021). The true death toll of Covid-19: Estimating global excess mortality. <https://www.who.int/data/stories/the-true-death-toll-of-covid-19-estimating-global-excess-mortality>.
2. World Health Organization. (October 2020). Global Tuberculosis Report 2020.
3. Horvath, C.N., and Xing, Z. (2013). Immunization Strategies Against Pulmonary Tuberculosis: Considerations of T Cell Geography. *New Paradigm of Immunity to Tuberculosis* 783, 267-278. 10.1007/978-1-4614-6111-1_14.
4. Prevention, C.f.D.C.a. (May 4, 2016). BCG Vaccine.
5. Smith, C.M., Proulx, M.K., Olive, A.J., Laddy, D., Mishra, B.B., Moss, C., Gutierrez, N.M., Bellerose, M.M., Barreira-Silva, P., Phuah, J.Y., et al. (2016). Tuberculosis Susceptibility and Vaccine Protection Are Independently Controlled by Host Genotype. *Mbio* 7, e01516-16. 10.1128/mBio.01516-16.
6. Poyntz, H.C., Stylianou, E., Griffiths, K.L., Marsay, L., Checkley, A.M., and McShane, H. (2014). Non-tuberculous mycobacteria have diverse effects on BCG efficacy against *Mycobacterium tuberculosis*. *Tuberculosis* 94, 226-237. 10.1016/j.tube.2013.12.006.
7. Ritz, N., and Curtis, N. (2009). Mapping the global use of different BCG vaccine strains. *Tuberculosis* 89, 248-251. 10.1016/j.tube.2009.03.002.
8. Wilson, M.E., Fineberg, H.V., and Colditz, G.A. (1995). Geographic Latitude and the Efficacy of Bacillus Calmette Guerin. *Clinical Infectious Diseases* 20, 982-991. 10.1093/clinids/20.4.982.
9. United Nations and StopTB Partnership. (September 26, 2018). UN High-level Meeting on TB Key Targets & Commitments For 2022.
10. Group, T.A. (December 7, 2020). Tuberculosis Research Funding Trends 2005 – 2019.
11. Lange, C., Chesov, D., Heyckendorf, J., Leung, C.C., Udwadia, Z., and Dheda, K. (2018). Drug-resistant tuberculosis: An update on disease burden, diagnosis and treatment. *Respirology* 23, 656-673. 10.1111/resp.13304.
12. Schmidt, C.W. (2008). Linking TB and the Environment: An overlooked Mitigation Strategy. *Environmental Health Perspectives* 116, A478-A485. 10.1289/ehp.116-a478.
13. Liu, F.Q., Zhang, Z.X., Chen, H.Y., and Nie, S.F. (2020). Associations of ambient air pollutants with regional pulmonary tuberculosis incidence in the central Chinese province of Hubei: a Bayesian spatial-temporal analysis. *Environmental Health* 19, 51. 10.1186/s12940-020-00604-y.
14. World Health Organization. (October, 14, 2020). Tuberculosis.
15. Talat, N., Perry, S., Parsonnet, J., Dawood, G., and Hussain, R. (2010). Vitamin D Deficiency and Tuberculosis Progression. *Emerging Infectious Diseases* 16, 853-855. 10.3201/eid1605.091693.
16. Sola, E., Rivera, C., Mangual, M., Martinez, J., Rivera, K., and Fernandez, R. (2016). Diabetes mellitus: an important risk factor for reactivation of tuberculosis. *Endocrinology Diabetes and Metabolism Case Reports*, 160035. 10.1530/edm-16-0035.
17. Centers for Disease Control and Prevention. Tuberculosis (TB) Signs & Symptoms.
18. Centers for Disease Control and Prevention. (October 20, 2014). The Difference Between Latent TB Infection and TB Disease.

19. Loddenkemper, R., Lipman, M., and Zumla, A. (2016). Clinical Aspects of Adult Tuberculosis. *Cold Spring Harbor Perspectives in Medicine* 6, a017848. 10.1101/cshperspect.a017848.
20. Garg, R.K., Malhotra, H.S., and Gupta, R. (2015). Spinal cord involvement in tuberculous meningitis. *Spinal Cord* 53, 649-657. 10.1038/sc.2015.58.
21. Khan, F.Y. (2019). Review of literature on disseminated tuberculosis with emphasis on the focused diagnostic workup. *Journal of Family and Community Medicine* 26, 83-91. 10.4103/jfcm.JFCM_106_18.
22. Tuberculosis Coalition for Technical Assistance. (2006). International Standards for Tuberculosis Care.
23. Schluger, N.W., and Burzynski, J. (2006). PNEUMONIA | Mycobacterial. In *Encyclopedia of Respiratory Medicine*, (Academic Press), pp. 451-456.
24. Desikan, P. (2013). Sputum smear microscopy in tuberculosis: Is it still relevant? *Indian Journal of Medical Research* 137, 442-444.
25. Ryu, Y.J. (2015). Diagnosis of Pulmonary Tuberculosis: Recent Advances and Diagnostic Algorithms. *Tuberculosis and Respiratory Diseases* 78, 64-71. 10.4046/trd.2015.78.2.64.
26. Nayak, S., and Acharjya, B. (2012). Mantoux test and its interpretation. *Indian Dermatology Online Journal* 3, 2-6. 10.4103/2229-5178.93479.
27. Centers for Disease Control and Prevention. (April 15, 2016). Testing in BCG-Vaccinated Persons.
28. Zwerling, A., Behr, M.A., Verma, A., Brewer, T.F., Menzies, D., and Pai, M. (2011). The BCG World Atlas: A Database of Global BCG Vaccination Policies and Practices. *Plos Medicine* 8, e1001012. 10.1371/journal.pmed.1001012.
29. TB CARE. (2014). International Standards for Tuberculosis Care, 3rd Edition.
30. Lawn, S.D., and Nicol, M.P. (2011). Xpert (R) MTB/RIF assay: development, evaluation and implementation of a new rapid molecular diagnostic for tuberculosis and rifampicin resistance. *Future Microbiology* 6, 1067-1082. 10.2217/fmb.11.84.
31. Li, S.Y., Liu, B., Peng, M.L., Chen, M., Yin, W.W., Tang, H., Luo, Y.X., Hu, P., and Ren, H. (2017). Diagnostic accuracy of Xpert MTB/RIF for tuberculosis detection in different regions with different endemic burden: A systematic review and meta-analysis. *Plos One* 12, e0180725. 10.1371/journal.pone.0180725.
32. Centers for Disease Control and Prevention. Tuberculosis (TB) Highlights from the 2016 Treatment of Drug-Susceptible Tuberculosis Guidelines.
33. List, R. (2021). Rifater(Rifampin, Isoniazid and Pyrazinamide) side effects drug center.
34. Zhao, H., Wang, Y.B., Zhang, T., Wang, Q., and Xie, W. (2020). Drug-Induced Liver Injury from Anti-Tuberculosis Treatment: A Retrospective Cohort Study. *Medical Science Monitor* 26, e920350. 10.12659/msm.920350.
35. Chang, C.H., Chen, Y.F., Wu, V.C., Shu, C.C., Lee, C.H., Wang, J.Y., Lee, L.N., and Yu, C.J. (2014). Acute kidney injury due to anti-tuberculosis drugs: a five-year experience in an aging population. *Bmc Infectious Diseases* 14, 23. 10.1186/1471-2334-14-23.
36. Marks, S.M., Flood, J., Seaworth, B., Hirsch-Moverman, Y., Armstrong, L., Mase, S., Salcedo, K., Oh, P., Graviss, E.A., Colson, P.W., et al. (2014). Treatment Practices, Outcomes, and Costs of Multidrug-Resistant and Extensively Drug-Resistant Tuberculosis, United States, 2005-2007. *Emerging Infectious Diseases* 20, 812-820. 10.3201/eid2005.131037.

37. National Jewish Health. (2013). Tuberculosis: History.
38. Fu, L.M., and Fu-Liu, C.S. (2002). Is Mycobacterium tuberculosis a closer relative to Gram-positive or Gram-negative bacterial pathogens? *Tuberculosis* 82, 85-90. 10.1054/tube.2002.0328.
39. Roy, S., Ghatak, D., Das, P., and BoseDasgupta, S. (2020). ESX secretion system: The gatekeepers of mycobacterial survivability and pathogenesis. *European Journal of Microbiology and Immunology* 10, 202-209. 10.1556/1886.2020.00028.
40. Saini, N.K., Baena, A., Ng, T.W., Venkataswamy, M.M., Kennedy, S.C., Kunnath-Velayudhan, S., Carreno, L.J., Xu, J.Y., Chan, J., Larsen, M.H., et al. (2016). Suppression of autophagy and antigen presentation by Mycobacterium tuberculosis PE_PGRS47. *Nature Microbiology* 1, 16133. 10.1038/nmicrobiol.2016.133.
41. Borrell, S., Trauner, A., Brites, D., Rigouts, L., Loiseau, C., Coscolla, M., Niemann, S., De Jong, B., Yeboah-Manu, D., Kato-Maeda, M., et al. (2019). Reference set of Mycobacterium tuberculosis clinical strains: A tool for research and product development. *Plos One* 14, e0214088. 10.1371/journal.pone.0214088.
42. Abebe, F., and Bjune, G. (2006). The emergence of Beijing family genotypes of Mycobacterium tuberculosis and low-level protection by bacille Calmette-Guerin (BCG) vaccines: is there a link? *Clinical and Experimental Immunology* 145, 389-397. 10.1111/j.1365-2249.2006.03162.x.
43. Verrall, A.J., Chaidir, L., Ruesen, C., Apriani, L., Koesoemadinata, R.C., van Ingen, J., Sharples, K., van Crevel, R., Alisjahbana, B., Hill, P.C., and grp, I.S. (2020). Lower Bacillus Calmette-Guerin Protection against Mycobacterium tuberculosis Infection after Exposure to Beijing Strains. *American Journal of Respiratory and Critical Care Medicine* 201, 1152-1155. 10.1164/rccm.201912-2349LE.
44. Kremer, K., van der Werf, M.J., Au, B.K.Y., Anh, D.D., Kam, K.M., van Doorn, H.R., Borgdorff, M.W., and van Soolingen, D. (2009). Vaccine-induced Immunity Circumvented by Typical Mycobacterium tuberculosis Beijing Strains. *Emerging Infectious Diseases* 15, 335-339. 10.3201/eid1502.080795.
45. Manca, C., Tsenova, L., Bergtold, A., Freeman, S., Tovey, M., Musser, J.M., Barry, C.E., Freedman, V.H., and Kaplan, G. (2001). Virulence of a Mycobacterium tuberculosis clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-alpha/beta. *Proceedings of the National Academy of Sciences of the United States of America* 98, 5752-5757. 10.1073/pnas.091096998.
46. Dean, G.S., Rhodes, S.G., Coad, M., Whelan, A.O., Cockle, P.J., Clifford, D.J., Hewinson, R.G., and Vordermeier, H.M. (2005). Minimum infective dose of Mycobacterium bovis in cattle. *Infection and Immunity* 73, 6467-6471. 10.1128/iai.73.10.6467-6471.2005.
47. Dubey, M., Nagarkoti, S., Awasthi, D., Singh, A.K., Chandra, T., Kumaravelu, J., Barthwal, M.K., and Dikshit, M. (2016). Nitric oxide-mediated apoptosis of neutrophils through caspase-8 and caspase-3-dependent mechanism. *Cell Death & Disease* 7, e2348. 10.1038/cddis.2016.248.
48. Marra, M.N., Wilde, C.G., Griffith, J.E., Snable, J.L., and Scott, R.W. (1990). Bactericidal Permeability Increasing Protein has Endotoxin Neutralizing Activity. *Journal of Immunology* 144, 662-666.
49. Papayannopoulos, V. (2018). Neutrophil extracellular traps in immunity and disease. *Nature Reviews Immunology* 18, 134-147. 10.1038/nri.2017.105.

50. Corleis, B., Korbel, D., Wilson, R., Bylund, J., Chee, R., and Schaible, U.E. (2012). Escape of Mycobacterium tuberculosis from oxidative killing by neutrophils. *Cellular Microbiology* 14, 1109-1121. 10.1111/j.1462-5822.2012.01783.x.
51. Eruslanov, E.B., Lyadova, I.V., Kondratieva, T.K., Majorov, K.B., Scheglov, I.V., Orlova, M.O., and Apt, A.S. (2005). Neutrophil responses to Mycobacterium tuberculosis infection in genetically susceptible and resistant mice. *Infection and Immunity* 73, 1744-1753. 10.1128/iai.73.3.1744-1753.2005.
52. Lyadova, I.V. (2017). Neutrophils in Tuberculosis: Heterogeneity Shapes the Way? *Mediators of Inflammation* 2017, 8619307. 10.1155/2017/8619307.
53. Stamm, C.E., Collins, A.C., and Shiloh, M.U. (2015). Sensing of Mycobacterium tuberculosis and consequences to both host and bacillus. *Immunological Reviews* 264, 204-219. 10.1111/imr.12263.
54. Flynn, J.L., Chan, J., and Lin, P.L. (2011). Macrophages and control of granulomatous inflammation in tuberculosis. *Mucosal Immunology* 4, 271-278. 10.1038/mi.2011.14.
55. Flynn, J.L., Gideon, H.P., Mattila, J.T., and Lin, P. (2015). Immunology studies in non-human primate models of tuberculosis. *Immunological Reviews* 264, 60-73. 10.1111/imr.12258.
56. Dartois, V. (2014). The path of anti-tuberculosis drugs: from blood to lesions to mycobacterial cells. *Nature Reviews Microbiology* 12, 159-167. 10.1038/nrmicro3200.
57. Miranda, M.S., Breiman, A., Allain, S., Deknuydt, F., and Altare, F. (2012). The Tuberculous Granuloma: An Unsuccessful Host Defence Mechanism Providing a Safety Shelter for the Bacteria? *Clinical & Developmental Immunology*, 139127. 10.1155/2012/139127.
58. Ehlers, S., and Schaible, U.E. (2013). The granuloma in tuberculosis: dynamics of a host-pathogen collusion. *Frontiers in Immunology* 3, 411. 10.3389/fimmu.2012.00411.
59. Mihret, A. (2012). The role of dendritic cells in Mycobacterium tuberculosis infection. *Virulence* 3, 654-659. 10.4161/viru.22586.
60. Giacomini, E., Iona, E., Ferroni, L., Miettinen, M., Fattorini, L., Orefici, G., Julkunen, I., and Coccia, E.M. (2001). Infection of human macrophages and dendritic cells with mycobacterium tuberculosis induces a differential cytokine gene expression that modulates T cell response. *Journal of Immunology* 166, 7033-7041. 10.4049/jimmunol.166.12.7033.
61. Lee, J., and Kornfeld, H. (2010). Interferon- γ Regulates the Death of *M. tuberculosis*-Infected Macrophages. *J. Cell Death* 3, 1-11.
62. Hanekom, W.A., Mendillo, M., Manca, C., Haslett, P.A.J., Siddiqui, M.R., Barry, C., and Kaplan, G. (2003). Mycobacterium tuberculosis inhibits maturation of human monocyte-derived dendritic cells in vitro. *Journal of Infectious Diseases* 188, 257-266. 10.1086/376451.
63. Wolf, A.J., Linas, B., Trevejo-Nunez, G.J., Kincaid, E., Tamura, T., Takatsu, K., and Ernst, J.D. (2007). Mycobacterium tuberculosis infects dendritic cells with high frequency and impairs their function in vivo. *Journal of Immunology* 179, 2509-2519. 10.4049/jimmunol.179.4.2509.
64. Janeway, C.A.J., Travers, P., Walport, M., and Schlomchik, M.J. (2001). *Immunobiology: The Immune System in Health and Disease*, 5th edition (Garland Science).

65. Flory, C.M., Hubbard, R.D., and Collins, F.M. (1992). Effects of In Vivo T-Lymphocyte Subset Depletion On Mycobacterial Infections in Mice. *Journal of Leukocyte Biology* 51, 225-229.
66. Hohl, T.M. (2015). Chapter6: Cell Mediated Defense against Infection. In Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases, 8th Edition, (Elsevier), pp. 50-69.
67. Abebe, F., and Bjune, G. (2009). The protective role of antibody responses during Mycobacterium tuberculosis infection. *Clinical and Experimental Immunology* 157, 235-243. 10.1111/j.1365-2249.2009.03967.x.
68. Gallegos, A.M., van Heijst, J.W.J., Samstein, M., Su, X.D., Pamer, E.G., and Glickman, M.S. (2011). A Gamma Interferon Independent Mechanism of CD4 T Cell Mediated Control of M. tuberculosis Infection in vivo. *Plos Pathogens* 7, e1002052. 10.1371/journal.ppat.1002052.
69. Ylikoski, E., Lund, R., Kylanemi, M., Filen, S., Kilpelainen, M., Savolainen, J., and Lahesmaa, R. (2005). IL-12 up-regulates T-bet independently of IFN-gamma in human CD4(+) T cells. *European Journal of Immunology* 35, 3297-3306. 10.1002/eji.200526101.
70. Chaudhry, A., Samstein, R.M., Treuting, P., Liang, Y.Q., Pils, M.C., Heinrich, J.M., Jack, R.S., Wunderlich, F.T., Bruning, J.C., Muller, W., and Rudensky, A.Y. (2011). Interleukin-10 Signaling in Regulatory T Cells Is Required for Suppression of Th17 Cell-Mediated Inflammation. *Immunity* 34, 566-578. 10.1016/j.immuni.2011.03.018.
71. Smigielski, K.S., Srivastava, S., Stolley, J.M., and Campbell, D.J. (2014). Regulatory T-cell homeostasis: steady-state maintenance and modulation during inflammation. *Immunological Reviews* 259, 40-59. 10.1111/imr.12170.
72. Stenger, S., Hanson, D.A., Teitelbaum, R., Dewan, P., Niazi, K.R., Froelich, C.J., Ganz, T., Thoma-Uszynski, S., Melian, A., Bogdan, C., et al. (1998). An antimicrobial activity of cytolytic T cells mediated by granulysin. *Science* 282, 121-125. 10.1126/science.282.5386.121.
73. Lin, P.L., and Flynn, J.L. (2015). CD8 T cells and Mycobacterium tuberculosis infection. *Seminars in Immunopathology* 37, 239-249. 10.1007/s00281-015-0490-8.
74. Cooper, A.M., Dsouza, C., Frank, A.A., and Orme, I.M. (1997). The course of Mycobacterium tuberculosis infection in the lungs of mice lacking expression of either perforin- or granzyme-mediated cytolytic mechanisms. *Infection and Immunity* 65, 1317-1320. 10.1128/iai.65.4.1317-1320.1997.
75. Rozot, V., Vigano, S., Mazza-Stalder, J., Perreau, M., Goletti, D., Bart, P.A., Nicod, L., Pantaleo, G., and Harari, A. (2012). Mycobacterium tuberculosis-specific CD8 T cells are functionally and phenotypically different between latent infection and active disease. *Immunology* 137, 598-598.
76. Laster, S.M., Wood, J.G., and Gooding, L.R. (1988). Tumor Necrosis Factor Can Induce Both Apoptotic and Necrotic Forms of Cell Lysis. *Journal of Immunology* 141, 2629-2634.
77. Behar, S.M., Martin, C.J., Booty, M.G., Nishimura, T., Zhao, X., Gan, H., Divangahi, M., and Remold, H.G. (2011). Apoptosis is an innate defense function of macrophages against Mycobacterium tuberculosis. *Mucosal Immunology* 4, 279-287. 10.1038/mi.2011.3.

78. van Meijgaarden, K.E., Haks, M.C., Caccamo, N., Dieli, F., Ottenhoff, T.H.M., and Joosten, S.A. (2015). Human CD8(+) T-cells Recognizing Peptides from Mycobacterium tuberculosis (Mtb) Presented by HLA-E Have an Unorthodox Th2-like, Multifunctional, Mtb Inhibitory Phenotype and Represent a Novel Human T-cell Subset. *Plos Pathogens* *11*, e1004671. 10.1371/journal.ppat.1004671.
79. Burleson, G.R., Burleson, S.C.M., and Burleson, F.G. Chapter 30 - Pulmonary Immunology of Infectious Disease. In *Comparative Biology of the Normal Lung* (Second Edition), (Academic Press), pp. 581-600.
80. Youngblood, B., Hale, J.S., Kissick, H.T., Ahn, E., Xu, X.J., Wieland, A., Araki, K., West, E.E., Ghoneim, H.E., Fan, Y.P., et al. (2017). Effector CD8 T cells dedifferentiate into long-lived memory cells. *Nature* *552*, 404-+. 10.1038/nature25144.
81. Akondy, R.S., Fitch, M., Edupuganti, S., Yang, S., Kissick, H.T., Li, K.W., Youngblood, B.A., Abdelsamed, H.A., McGuire, D.J., Cohen, K.W., et al. (2017). Origin and differentiation of human memory CD8 T cells after vaccination. *Nature* *552*, 362-+. 10.1038/nature24633.
82. Orme, I.M., and Henao-Tamayo, M.I. (2018). Trying to See the Forest through the Trees: Deciphering the Nature of Memory immunity to Mycobacterium tuberculosis. *Frontiers in Immunology* *9*, 461. 10.3389/fimmu.2018.00461.
83. Gray, J.I., Westerhof, L.M., and MacLeod, M.K.L. (2018). The roles of resident, central and effector memory CD4 T-cells in protective immunity following infection or vaccination. *Immunology* *154*, 574-581. 10.1111/imm.12929.
84. Orme, I.M. (2010). The Achilles heel of BCG. *Tuberculosis* *90*, 329-332. 10.1016/j.tube.2010.06.002.
85. Mami-Chouaib, F., and Tartour, E. (2019). Editorial: Tissue Resident Memory T Cells. *Frontiers in Immunology* *10*, 1018. 10.3389/fimmu.2019.01018.
86. Ogongo, P., Porterfield, J.Z., and Leslie, A. (2019). Lung Tissue Resident Memory T-Cells in the Immune Response to Mycobacterium tuberculosis. *Frontiers in Immunology* *10*, 992. 10.3389/fimmu.2019.00992.
87. Copland, A., Diogo, G.R., Hart, P., Harris, S., Tran, A.C., Paul, M.J., Singh, M., Cutting, S.M., and Reljic, R. (2018). Mucosal Delivery of Fusion Proteins with Bacillus subtilis Spores Enhances Protection against Tuberculosis by Bacillus Calmette-Guerin. *Frontiers in Immunology* *9*, 346. 10.3389/fimmu.2018.00346.
88. Hart, P., Copland, A., Diogo, G.R., Harris, S., Spallek, R., Oehlmann, W., Singh, M., Basile, J., Rottenberg, M., Paul, M.J., and Reljic, R. (2018). Nanoparticle-Fusion Protein Complexes Protect against Mycobacterium tuberculosis Infection. *Molecular Therapy* *26*, 822-833. 10.1016/j.ymthe.2017.12.016.
89. Lindstrom, T., Knudsen, N.P.H., Agger, E.M., and Andersen, P. (2013). Control of Chronic Mycobacterium tuberculosis Infection by CD4 KLRG1(-) IL-2-Secreting Central Memory Cells. *Journal of Immunology* *190*, 6311-6319. 10.4049/jimmunol.1300248.
90. Ahmed, R., Roger, L., del Amo, P.C., Miners, K.L., Jones, R.E., Boelen, L., Fali, T., Elemans, M., Zhang, Y., Appay, V., et al. (2016). Human Stem Cell-like Memory T Cells Are Maintained in a State of Dynamic Flux. *Cell Reports* *17*, 2811-2818. 10.1016/j.celrep.2016.11.037.
91. Mpande, C.A.M., Dintwe, O.B., Musvosvi, M., Mabwe, S., Bilek, N., Hatherill, M., Nemes, E., Scriba, T.J., and Team, S.C.I. (2018). Functional, Antigen-Specific Stem Cell

- Memory (TSCM) CD4(+) T Cells Are Induced by Human Mycobacterium tuberculosis Infection. *Frontiers in Immunology* 9, 324. 10.3389/fimmu.2018.00324.
92. Alberts, B., Johnson, A., and Lewis, J. (2002). *Molecular Biology of the Cell*, 4th Edition (Garland Science).
 93. Rijnink, W.F., Ottenhoff, T.H.M., and Joosten, S.A. (2021). B-Cells and Antibodies as Contributors to Effector Immune Responses in Tuberculosis. *Frontiers in Immunology* 12, 640168. 10.3389/fimmu.2021.640168.
 94. Kumar, S.K., Singh, P., and Sinha, S. (2015). Naturally produced opsonizing antibodies restrict the survival of Mycobacterium tuberculosis in human macrophages by augmenting phagosome maturation. *Open Biology* 5, 150171. 10.1098/rsob.150171.
 95. Achkar, J.M., Chan, J., and Casadevall, A. (2015). Role of B Cells and Antibodies in Acquired Immunity against Mycobacterium tuberculosis. *Cold Spring Harbor Perspectives in Medicine* 5, a018432. 10.1101/cshperspect.a018432.
 96. du Plessis, W.J., Kleynhans, L., du Plessis, N., Stanley, K., Malherbe, S.T., Maasdorp, E., Ronacher, K., Chegou, N.N., Walzl, G., and Loxton, A.G. (2016). The Functional Response of B Cells to Antigenic Stimulation: A Preliminary Report of Latent Tuberculosis. *Plos One* 11, e0152710. 10.1371/journal.pone.0152710.
 97. Leal, I.S., Smedegard, B., Andersen, P., and Appelberg, R. (1999). Interleukin-6 and interleukin-12 participate in induction of a type 1 protective T-cell response during vaccination with a tuberculosis subunit vaccine. *Infection and Immunity* 67, 5747-5754. 10.1128/iai.67.11.5747-5754.1999.
 98. Maddur, M.S., and Bayry, J. (2015). B cells drive Th2 responses by instructing human dendritic cell maturation. *Oncoimmunology* 4, e1005508. 10.1080/2162402x.2015.1005508.
 99. Linge, I., Dyatlov, A., Kondratieva, E., Avdienko, V., Apt, A., and Kondratieva, T. (2017). B-lymphocytes forming follicle-like structures in the lung tissue of tuberculosis-infected mice: Dynamics, phenotypes and functional activity. *Tuberculosis* 102, 16-23. 10.1016/j.tube.2016.11.005.
 100. World Health Organization. *Vaccine Safety Basics - Module 2: Types of Vaccine and Adverse Reactions - Live Attenuated Vaccines*.
 101. Mak, T.W., and Saunders, M.E. (2005). Chapter 23 - Vaccines and Clinical Immunization. In *The Immune Response: Basic and Clinical Principles*, (Elsevier), pp. 695-750.
 102. Luca, S., and Mihaescu, T. (2013, March). History of BCG Vaccine. *Maedica* 8, 53-58.
 103. Doerflinger, M., Forsyth, W., Ebert, G., Pellegrini, M., and Herold, M.J. (2017). CRISPR/Cas9-The ultimate weapon to battle infectious diseases? *Cellular Microbiology* 19, e12693. 10.1111/cmi.12693.
 104. Tang, Y.D., Liu, J.T., Wang, T.Y., An, T.Q., Sun, M.X., Wang, S.J., Fang, Q.Q., Hou, L.L., Tian, Z.J., and Cai, X.H. (2016). Live attenuated pseudorabies virus developed using the CRISPR/Cas9 system. *Virus Research* 225, 33-39. 10.1016/j.virusres.2016.09.004.
 105. Tiwari, S., Dutt, T.S., Chen, B., Chen, M., Kim, J., Dai, A.Z., Lukose, R., Shanley, C., Fox, A., Karger, B.R., et al. (2020). BCG-Prime and boost with Esx-5 secretion system deletion mutant leads to better protection against clinical strains of Mycobacterium tuberculosis. *Vaccine* 38, 7156-7165. 10.1016/j.vaccine.2020.08.004.

106. World Health Organization. Vaccine Safety Basics - Module 2: Types of Vaccine and Adverse Reactions: Inactivated.
107. Skinner, M.A., Yuan, S., Prestidge, R., Chuk, D., Watson, J.D., and Tan, P.L.J. (1997). Immunization with heat-killed *Mycobacterium vaccae* stimulates CD8(+) cytotoxic T cells specific for macrophages infected with *Mycobacterium tuberculosis*. *Infection and Immunity* 65, 4525-4530. 10.1128/iai.65.11.4525-4530.1997.
108. Lahey, T., Arbeit, R.D., Bakari, M., Horsburgh, C.R., Matee, M., Waddell, R., Mtei, L., Vuola, J.M., Pallangyo, K., and von Reyn, C.F. (2010). Immunogenicity of a protective whole cell mycobacterial vaccine in HIV-infected adults: A phase III study in Tanzania. *Vaccine* 28, 7652-7658. 10.1016/j.vaccine.2010.09.041.
109. World Health Organization. Vaccine Safety Basics - Module 2: Types of Vaccine and Adverse Reactions: Subunit.
110. Vartak, A., and Sucheck, S.J. (2016). Recent Advances in Subunit Vaccine Carriers. *Vaccines* 4, 12. 10.3390/vaccines4020012.
111. Nascimento, I.P., and Leite, L.C.C. (2012). Recombinant vaccines and the development of new vaccine strategies. *Brazilian Journal of Medical and Biological Research* 45, 1102-1111. 10.1590/s0100-879x2012007500142.
112. Tameris, M.D., Hatherill, M., Landry, B.S., Scriba, T.J., Snowden, M.A., Lockhart, S., Shea, J.E., McClain, J.B., Hussey, G.D., Hanekom, W.A., et al. (2013). Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. *Lancet* 381, 1021-1028. 10.1016/s0140-6736(13)60177-4.
113. World Health Organization. Vaccine Safety Basics - Module 2: Route of Administration.
114. Doyle, G.R., and McCutcheon, J.A. (2015). Clinical Procedures for Safer Patient Care. BCcampus. <https://opentextbc.ca/clinicalskills/>
115. Darrah, P.A., Zeppa, J.J., Maiello, P., Hackney, J.A., Ii, M.H.W., Hughes, T.K., Pokkali, S., Swanson, P.A., Grant, N.L., Rodgers, M.A., et al. (2020). Prevention of tuberculosis in macaques after intravenous BCG immunization. *Nature* 577, 95-+. 10.1038/s41586-019-1817-8.
116. Sherwin, C.M.T., Medlicott, N.J., Reith, D.M., and Broadbent, R.S. (2014). Intravenous drug delivery in neonates: lessons learnt. *Archives of Disease in Childhood* 99, 590-594. 10.1136/archdischild-2013-304887.
117. Kagina, B.M.N., Abel, B., Bowmaker, M., Scriba, T.J., Gelderbloem, S., Smit, E., Erasmus, M., Nene, N., Walzl, G., Black, G., et al. (2009). Delaying BCG vaccination from birth to 10 weeks of age may result in an enhanced memory CD4 T cell response. *Vaccine* 27, 5488-5495. 10.1016/j.vaccine.2009.06.103.
118. Han, S. (2015). Clinical vaccine development. *Clinical and Experimental Vaccine Research* 4, 46-53. 10.7774/cevr.2015.4.1.46.
119. Philadelphia, C.o.P.o. (January 17, 2018). Vaccine Development, Testing, and Regulation.
120. U.S. Food and Drug Administration. (December, 14, 2020). Vaccine Development.
121. Gouglas, D., Le, T.T., Henderson, K., Kaloudis, A., Danielsen, T., Hammersland, N.C., Robinson, J.M., Heaton, P.M., and Rottingen, J.A. (2018). Estimating the cost of vaccine development against epidemic infectious diseases: a cost minimisation study. *Lancet Global Health* 6, E1386-E1396. 10.1016/s2214-109x(18)30346-2.

122. Mendez-Samperio, P. (2018). Development of tuberculosis vaccines in clinical trials: Current status. *Scandinavian Journal of Immunology* *88*, e12710. 10.1111/sji.12710.
123. Brosch, R., Gordon, S.V., Garnier, T., Eiglmeier, K., Frigui, W., Valenti, P., Dos Santos, S., Duthoy, S., Lacroix, C., Garcia-Pelayo, C., et al. (2007). Genome plasticity of BCG and impact on vaccine efficacy. *Proceedings of the National Academy of Sciences of the United States of America* *104*, 5596-5601. 10.1073/pnas.0700869104.
124. Kaushal, D., Mehra, S., Didier, P.J., and Lackner, A.A. (2012). The non-human primate model of tuberculosis. *Journal of Medical Primatology* *41*, 191-201. 10.1111/j.1600-0684.2012.00536.x.
125. Singh, A.K., and Gupta, U.D. (2018). Animal models of tuberculosis: Lesson learnt. *Indian Journal of Medical Research* *147*, 456-463. 10.4103/ijmr.IJMR_554_18.
126. Driver, E.R., Ryan, G.J., Hoff, D.R., Irwin, S.M., Basaraba, R.J., Kramnik, I., and Lenaerts, A.J. (2012). Evaluation of a Mouse Model of Necrotic Granuloma Formation Using C3HeB/FeJ Mice for Testing of Drugs against Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy* *56*, 3181-3195. 10.1128/aac.00217-12.
127. Henao-Tamayo, M., Obregon-Henao, A., Creissen, E., Shanley, C., Orme, I., and Ordway, D.J. (2015). Differential Mycobacterium bovis BCG Vaccine-Derived Efficacy in C3HeB/FeJ and C3H/HeOuJ Mice Exposed to a Clinical Strain of Mycobacterium tuberculosis. *Clinical and Vaccine Immunology* *22*, 91-98. 10.1128/cvi.00466-14.
128. Adam, L., Lopez-Gonzalez, M., Bjork, A., Palsson, S., Poux, C., Wahren-Herlenius, M., Fernandez, C., and Spetz, A.L. (2018). Early Resistance of Non-virulent Mycobacterial Infection in C57BL/6 Mice Is Associated With Rapid Up-Regulation of Antimicrobial Cathelicidin Camp. *Frontiers in Immunology* *9*, 1939. 10.3389/fimmu.2018.01939.
129. Vernon, A., and Bishai, W. (2020). Modeling Treatment of Latent Tuberculosis: Shortening the Leap of Faith? *American Journal of Respiratory and Critical Care Medicine* *201*, 405-406. 10.1164/rccm.201911-2160ED.
130. Brekke, T.D., Steele, K.A., and Mulley, J.F. (2018). Inbred or Outbred? Genetic Diversity in Laboratory Rodent Colonies. *G3-Genes Genomes Genetics* *8*, 679-686. 10.1534/g3.117.300495.
131. Threadgill, D.W., Miller, D.R., Churchill, G.A., and de Villena, F.P.M. (2011). The Collaborative Cross: A Recombinant Inbred Mouse Population for the Systems Genetic Era. *Ilar Journal* *52*, 24-31. 10.1093/ilar.52.1.24.
132. Clark, S., Hall, Y., and Williams, A. (2015). Animal Models of Tuberculosis: Guinea Pigs. *Cold Spring Harbor Perspectives in Medicine* *5*, a018572. 10.1101/cshperspect.a018572.
133. Turner, O.C., Basaraba, R.J., and Orme, I.M. (2003). Immunopathogenesis of pulmonary granulomas in the guinea pig after infection with Mycobacterium tuberculosis. *Infection and Immunity* *71*, 864-871. 10.1128/iai.71.2.864-871.2003.
134. Basaraba, R.J. (2008). Experimental tuberculosis: the role of comparative pathology in the discovery of improved tuberculosis treatment strategies. *Tuberculosis* *88*, S35-S47. 10.1016/s1472-9792(08)70035-0.
135. Hiromatsu, K., Dascher, C.C., LeClair, K.P., Sugita, M., Furlong, S.T., Brenner, M.B., and Porcelli, S.A. (2002). Induction of CD1-restricted immune responses in guinea pigs by immunization with mycobacterial lipid antigens. *Journal of Immunology* *169*, 330-339. 10.4049/jimmunol.169.1.330.

136. Ramos, L., Obregon-Henao, A., Henao-Tamayo, M., Bowen, R., Lunney, J.K., and Gonzalez-Juarrero, M. (2017). The minipig as an animal model to study Mycobacterium tuberculosis infection and natural transmission. *Tuberculosis* 106, 91-98. 10.1016/j.tube.2017.07.003.
137. Gil, O., Diaz, I., Vilaplana, C., Tapia, G., Diaz, J., Fort, M., Caceres, N., Pinto, S., Cayla, J., Corner, L., et al. (2010). Granuloma Encapsulation Is a Key Factor for Containing Tuberculosis Infection in Minipigs. *Plos One* 5, e10030. 10.1371/journal.pone.0010030.
138. Ramos, L., Obregon-Henao, A., Henao-Tamayo, M., Bowen, R., Izzo, A., Lunney, J.K., and Gonzalez-Juarrero, M. (2019). Minipigs as a neonatal animal model for tuberculosis vaccine efficacy testing. *Veterinary Immunology and Immunopathology* 215, 109884. 10.1016/j.vetimm.2019.109884.
139. Gerner, W., Talker, S.C., Koinig, H.C., Sedlak, C., Mair, K.H., and Saalmuller, A. (2015). Phenotypic and functional differentiation of porcine alpha beta T cells: Current knowledge and available tools. *Molecular Immunology* 66, 3-13. 10.1016/j.molimm.2014.10.025.
140. Clenet, M.L., Gagnon, F., Moratalla, A.C., Viel, E.C., and Arbour, N. (2017). Peripheral human CD4(+)CD8(+) T lymphocytes exhibit a memory phenotype and enhanced responses to IL-2, IL-7 and IL-15. *Scientific Reports* 7, 11612. 10.1038/s41598-017-11926-2.
141. Diedrich, C.R., Gideon, H.P., Rutledge, T., Baranowski, T.M., Maiello, P., Myers, A.J., and Lin, P.L. (2019). CD4CD8 Double Positive T cell responses during Mycobacterium tuberculosis infection in cynomolgus macaques. *Journal of Medical Primatology* 48, 82-89. 10.1111/jmp.12399.
142. Gutierrez, K., Dicks, N., Glanzner, W.G., Agellon, L.B., and Bordignon, V. (2015). Efficacy of the porcine species in biomedical research. *Frontiers in Genetics* 6, 293. 10.3389/fgene.2015.00293.
143. Hild, S.A., Chang, M.C.C., Murphy, S.J., and Grieder, F.B. (2021). Nonhuman primate models for SARS-CoV-2 research: Infrastructure needs for pandemic preparedness. *Lab Animal* 50, 140-141. 10.1038/s41684-021-00760-9.
144. Roederer, M. (2002). Compensation in flow cytometry. *Current protocols in cytometry Chapter 1*, Unit 1.14-Unit 11.14. 10.1002/0471142956.cy0114s22.
145. van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, 2579-2605.
146. Kobak, D., and Berens, P. (2019). The art of using t-SNE for single-cell transcriptomics. *Nature Communications* 10, 5416. 10.1038/s41467-019-13056-x.
147. Belkina, A.C., Ciccolella, C.O., Anno, R., Halpert, R., Spidlen, J., and Snyder-Cappione, J.E. (2019). Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. *Nature Communications* 10, 5415. 10.1038/s41467-019-13055-y.
148. Van Gassen, S., Callebaut, B., Van Helden, M.J., Lambrecht, B.N., Demeester, P., Dhaene, T., and Saeys, Y. (2015). FLOW-SOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry Part A* 87A, 636-645. 10.1002/cyto.a.22625.
149. Quiroz, E.N., Chavez-Estrada, V., Macias-Ochoa, K., Ayala-Navarro, M.F., Flores-Aguilar, A.S., Morales-Navarrete, F., Lopez, F.D., Escorcía, L.G., Musso, C.G., Martínez, G.A., et al. (2019). Epigenetic Mechanisms and Posttranslational Modifications

- in Systemic Lupus Erythematosus. *International Journal of Molecular Sciences* *20*, 5679. 10.3390/ijms20225679.
150. Vrieling, F., Kostidis, S., Spaink, H.P., Haks, M.C., Mayboroda, O.A., Ottenhoff, T.H.M., and Joosten, S.A. (2020). Analyzing the impact of *Mycobacterium tuberculosis* infection on primary human macrophages by combined exploratory and targeted metabolomics. *Scientific Reports* *10*, 7085. 10.1038/s41598-020-62911-1.
 151. Pitt, J.J. (2009). Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry. *The Clinical Biochemist Reviews* *30*, 19-34.
 152. Zhang, J.Q., Gonzalez, E., Hestilow, T., Haskins, W., and Huang, Y.F. (2009). Review of Peak Detection Algorithms in Liquid-Chromatography-Mass Spectrometry. *Current Genomics* *10*, 388-401. 10.2174/138920209789177638.
 153. Rainer, J. (2021). LCMS data preprocessing and analysis with xcms.
 154. Vinaixa, M., Samino, S., Saez, I., Duran, J., Guinovart, J.J., and Yanes, O. (2012). A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. *Metabolites* *2*, 775-795. 10.3390/metabo2040775.

CHAPTER 2— CYTO-FEATURE ENGINEERING: A PIPELINE FOR FLOW CYTOMETRY ANALYSIS TO UNCOVER IMMUNE POPULATIONS AND ASSOCIATIONS WITH DISEASE*

Introduction

Flow cytometers can now analyze up to 50 parameters (antigens, size, granularity, cytokines, transcription factors, etc.) per cell and millions of cells per sample.¹ Conventional flow cytometry data analysis uses manual gating of cells on 2D plots to distinguish populations 1-2 dimensions at a time; this makes it both subjective and time consuming (up to 15 hours per experiment).² Better methods are therefore critically needed to take full advantage of this powerful technology. Researchers have responded with open-source tools, including tools for automated gating to remove user input bias (e.g., OpenCyto) and tools to identify and cluster cell populations concurrently using all parameters (e.g., FlowSOM, t-SNE).³⁻⁵ While powerful advances, these new tools lack a straightforward way to integrate data from important technical controls or to compare resulting cell population with other experimental measurements. Work is ongoing across several research groups to extend existing open-source tools to address some of these gaps. CytoCompare and Cytofast, for example, focus on data analysis after clustering.^{6,7} However, few tools exist that allow users to incorporate the many flow cytometry controls

* Much of this work was published in Scientific Reports: Fox, A., Dutt, T.S., Karger, B. *et al.* Cyto-Feature Engineering: A Pipeline for Flow Cytometry Analysis to Uncover Immune Populations and Associations with Disease. *Sci Rep* 10, 7651 (2020). <https://doi.org/10.1038/s41598-020-64516-0>. It was published under a Creative Commons license; a copy of this license can be found here: <http://creativecommons.org/licenses/by/4.0/>. We have added an additional section “Testing available flow cytometry analysis techniques” which discusses other methods that we tested before ultimately developing the pipeline described in the paper to the supplemental materials.

required for good data acquisition and analysis, and the output from the clustering tools available is often difficult for immunologists to interpret.

We have developed an end-to-end method for analyzing flow cytometry data that aims to address these limitations. For flow cytometry data, a parameter often represents a biologically binary phenomenon—that a marker is present or missing on a cell. While variation exists in the flow cytometry measurements for that parameter, within cells in each binary group, that within-group variation is often uninformative noise. Our pipeline leverages this underlying biology—it uses feature engineering to create from the original data to create binary features for whether each cell has a positive or negative value for each marker. It does this using either external thresholds identified based on Fluorescence Minus One controls (FMOs) or the availability to separate the data based on clear population separation. The pipeline therefore identifies cell populations based on positive/negative combinations of each flow cytometry marker, a description that is readily interpretable by immunologists.

In four main steps, the pipeline: (1) cleans the data for live, single cells; (2) feature engineers the data based on FMO cutoffs or population separation; (3) analyzes the flow cytometry samples for all populations present in the sample and filters to populations above a population size threshold; (4) visualizes resulting populations through heatmaps of cell phenotypes and time series plots within experimental groups. Furthermore, it allows the use of statistical testing to identify cell populations associated with other experimental measurements (e.g., disease burden as measured through colony forming units) and novel populations induced by any experimental or clinical condition. All steps in the pipeline are modular, allowing each to be modified or replaced

depending on the research question and features of the experimental data. As a case study, we illustrate the pipeline on a study involving *Mycobacterium bovis* Bacille Calmette-Guérin (BCG)-vaccinated or control (Phosphate buffered saline (PBS)-injected) C57BL/6 mice infected with *Mycobacterium tuberculosis* (*M. tuberculosis*). We further validate the pipeline analyzing human whole blood for B and T cells.

Materials and Methods

Experimental Setup. The experiment was designed to compare T cell populations between BCG-vaccinated and control PBS-vaccinated mice following infection with *M. tuberculosis*. The populations could then be compared to the bacterial load (CFUs) to associate the immune cells with disease.

Animals. C57BL/6 mice were purchased from Jackson Laboratories (Bar Harbor, ME). The mice were retained in a BSL-3 facility at Colorado State University, and all experimental protocols were approved by the Institutional Animal Care and Use Committee at Colorado State University. All methods were carried out in accordance with relevant guidelines and regulations for care and use of laboratory animals.

Vaccinations. Mice were vaccinated subcutaneously with 1×10^6 CFU *M. bovis* BCG Pasteur or 100 μ L of phosphate buffered saline (PBS) (Corning, Corning, NY).

***Mycobacterium tuberculosis* Infection.** Twelve weeks after vaccination, mice were aerosol challenged with *M. tuberculosis* HN878 grown in 7H9 broth and stored at -80°C . The *M.*

tuberculosis was suspended in PBS and aerosolized using an aerosol chamber. Aerosolization delivered 164 CFU/animal confirmed via plating whole lung homogenates on the same day of infection.

Tissue Preparation. Mice were euthanized by CO₂ inhalation. The superior and middle lung lobes were removed and digested with DNase I- type IV Bovine (500 units/ml; Sigma-Aldrich, St. Louis, MO)/ Liberase (0.5 mg/mL; Sigma-Aldrich, St. Louis, MO) at 37°C for 30 minutes. The lungs were strained through a 70µm cell strainer and treated for 1 minute with Red Blood Cell Lysing Buffer (Sigma-Aldrich, St. Louis, MO) to lyse erythrocytes. Complete DMEM, composed of 500 mL of 1x DMEM (Corning, Corning, NY), 45 mL of fetal bovine serum (Atlas Biologicals, Fort Collins, CO), 4.5 mL MEM non-essential amino acids (Corning, Corning, NY), 4.5 mL of Penicillin Streptomycin (10,000 units/mL Penicillin, 10,000 µg/mL Streptomycin; ThermoFisher Scientific, Waltham, MA), and 4.5 mL of L-glutamine (Sigma-Aldrich, St. Louis, MO), is then added to neutralize the solution. Lung cells were resuspended in 500 µL of PBS (Corning, Corning, NY).

Flow cytometry and analysis. Following single-cell suspension of lung cells in PBS, cells were stimulated with 1x BD Golgi Stop, a protein transport inhibitor (BD Biosciences, Cat# 554724) in complete media for 3 hours in a 37 °C incubator. Cells were then washed with PBS and incubated with a 1:2,000 dilution of Zombie-NIR viability stain (BioLegend, San Diego, CA) for 15 minutes in the dark at room temperature. After a wash with FACS Staining buffer (PBS (Corning, Corning, NY) containing 2% Fetal Bovine Serum (Atlas Biologicals, Fort Collins, CO) and 0.05% sodium azide (Thermo Fisher Scientific, Waltham, MA)), cells were incubated for

30 minutes at 4 °C with a fluorescently-labeled surface antibody cocktail containing a 1:200 dilution of FC Block (See Appendix 1: Supplemental Table 2.4). Following a wash with FACS staining buffer, cells were incubated with 150 µL of 1x Permeabilization/Fixation buffer (Invitrogen, Carlsbad, CA) for 1 hour at room temperature. Cells were subsequently washed with 150 µL of 1x permeabilization buffer (Invitrogen, Carlsbad, CA) before being incubated with an intracellular cytokine antibody cocktail overnight at 4 °C in the dark. Cells were then washed and finally resuspended in 1x permeabilization buffer. Eighteen markers were used to analyze memory T cell expression: Sca-1, CD3, CD62L, CD122, CD28, PD-1, CD103, CD44, CD4, CD8, CTLA-4, CD27, CD153, KLRG-1, IL-17, IFN- γ , IL-10, and TNF- α (Appendix 1: Supplemental Table 2.4). Antibodies and reagents were purchased from BD Biosciences, BioLegend, or Thermo Fisher Scientific. 100,000 events were collected per sample on a Cytex Aurora Flow Cytometer (Cytex, Fremont, CA) and analyzed with FlowJo version 10.5.2 software and the cyto-feature engineering pipeline. Cell populations were identified by feature engineering of cells by phenotype and confirmed via traditional gating methods. Prior to the development of the pipeline, other flow cytometry analysis methods were also tested (Appendix 1: Supplemental Text T2.1).

Bacterial burden. One third lungs were placed in a Bullet Blender Blue (NextAdvance, Troy, NY) and homogenized at speed 8 for 4 minutes. Tissue homogenate was plated at a 1:5 dilution on 7H11 agar plates. The limit of detection is 15 CFUs for lungs.

Human clinical data. Human clinical data was obtained and processed as previously described.⁸⁻¹⁰ All experimental protocols involving human subjects were approved by the Ethics

Committee of the Institute of Medical Research of the Faculty of Medicine at the University of Antioquia and adhere to the ethical principles outlined in the Declaration of Helsinki. EDTA-anticoagulated (4 mL) whole cells were obtained from peripheral blood of healthy volunteers aged 20-30 years, after written informed consent to participate. Most of them worked in the Sede de Investigación Universitaria (SIU) at Universidad de Antioquia and signed a written informed consent approved by the Ethics Committee of the Institute of Medical Research of the Faculty of Medicine at the University of Antioquia. They declared that they were not taking any medication and that they had neither an autoimmune nor active infectious disease.

Thirty microliters of whole blood cells suspended in 100 μ L eBioscience™ Flow Cytometry Staining Buffer (Catalog number: 00-4222-26) and stained with fluorochrome-conjugated mouse anti-human 0,5 μ L CD45-PE-Cy7 (Clone HI30), 5 μ L CD3-PE (Clone: OKT3), 7 μ L CD19-Alexa Fluor® 488 (Clone: HIB19) and 5 μ L CD27-APC (Clone M-T271) monoclonal antibodies for twenty minutes at room temperature. Erythrocytes were lysed with 300 μ L of OptiLyse Buffer for 10 minutes and 300 μ L of sterile deionized water for an additional 10 minutes. The acquisition was performed in an LSR Fortessa II™ flow cytometer.

Statistical analysis and reproducible code. Statistical significance for CFUs was determined using unpaired t-tests ($p < 0.05$) using the *ggpubr* package, R version 3.6.2. The R and package versions can be found in Appendix 1: Supplemental Figure S2.6.

Data Availability. The datasets generated during and/or analyzed during the current study are available online. The flow cytometry data can be found at Flow Repository:

<https://flowrepository.org/id/RvFrMfWyY0wR3ZM3cvlsAsQPBmAOXmMUzURGm1D8V0ShqSNnH2UCrPdpttuqNS4>. All other data can be found at:

https://github.com/aef1004/cyto-feature_engineering in the “data” folder.

Code Availability. Code to analyze the data and produce the graphs can be found at:

https://github.com/aef1004/cyto-feature_engineering

Results

The workflow for the analysis pipeline includes reading flow cytometry data output, cleaning and feature engineering this data, performing data visualization, and performing hypothesis testing through integrating other experimental measurements (Figure 2.1). Each step is described in detail below, and a route map describing the method can be found in Appendix 1:

Supplemental Figure S2.1.

Cleaning data with gating input

Flow cytometers use a standardized file format for outputting data, the .fcs file, which includes cell measurements, metadata describing data collection, and the Median Fluorescent Intensities (MFIs) of the fluorescently-conjugated antibodies of fluorescent probes.¹¹ Multiple .fcs files generated from an experiment can be read into R and manipulated as a “flowSet” object.¹² Our pipeline begins by reading experimental data into a “flowSet” object, then cleaning the data using the *openCyto* package (Figure 2.1). This package provides infrastructure for the use of reproducible algorithms to gate cells based on marker density.³ However, it alone is unable to control for instances where clumps of cells pass through the flow cytometer lasers, producing

erroneous results and subsequently skewing the data. To address this phenomenon, the pipeline incorporates the “SingletGate” function from the *flowStats* package, removing doublet or larger cell clumps.¹³ The pipeline then funnels the data through the “mindensity” function, selecting for leukocytes via a threshold filter that distinguishes between populations based on cell density.³ Finally, a “mindensity” gate is used with a live/dead stain, to filter to data captured for live cells.³ The data is next converted from a “flowSet” object into a dataframe object that complies with the “tidy data” standards, allowing further pipeline steps to draw on the powerful suite of “tidyverse” tools in R.¹⁴

Feature engineering using FMOs

FMOs are often used in manual gating to control for spillover events, which are common during flow cytometry data collection.¹⁵ Take for example a panel consisting of 10 markers with different fluorophores. When excited, each of those 10 markers fluoresce at different intensities along the light spectrum. However, though they have different spectrums, tails of these spectrums can overlap. This overlap can lead to noise within a parameter’s measurements, and in extreme cases to the detection of false positives/negatives for presence or absence of a marker. FMOs are created experimentally; by running new samples where each sample has one marker removed, all cells are guaranteed to be truly negative on that marker. With FMOs, we can therefore identify a threshold for the maximum parameter values possible for true negative marker signal on cells to determine marker presence in fully stained samples.¹⁵ Incorporation of FMOs greatly reduces the subjectivity of manual gating and helps support unbiased analysis of flow cytometry data. Despite the importance of FMOs for accurate analysis, few flow cytometry computational tools exist that incorporate them into unsupervised analysis.¹⁶

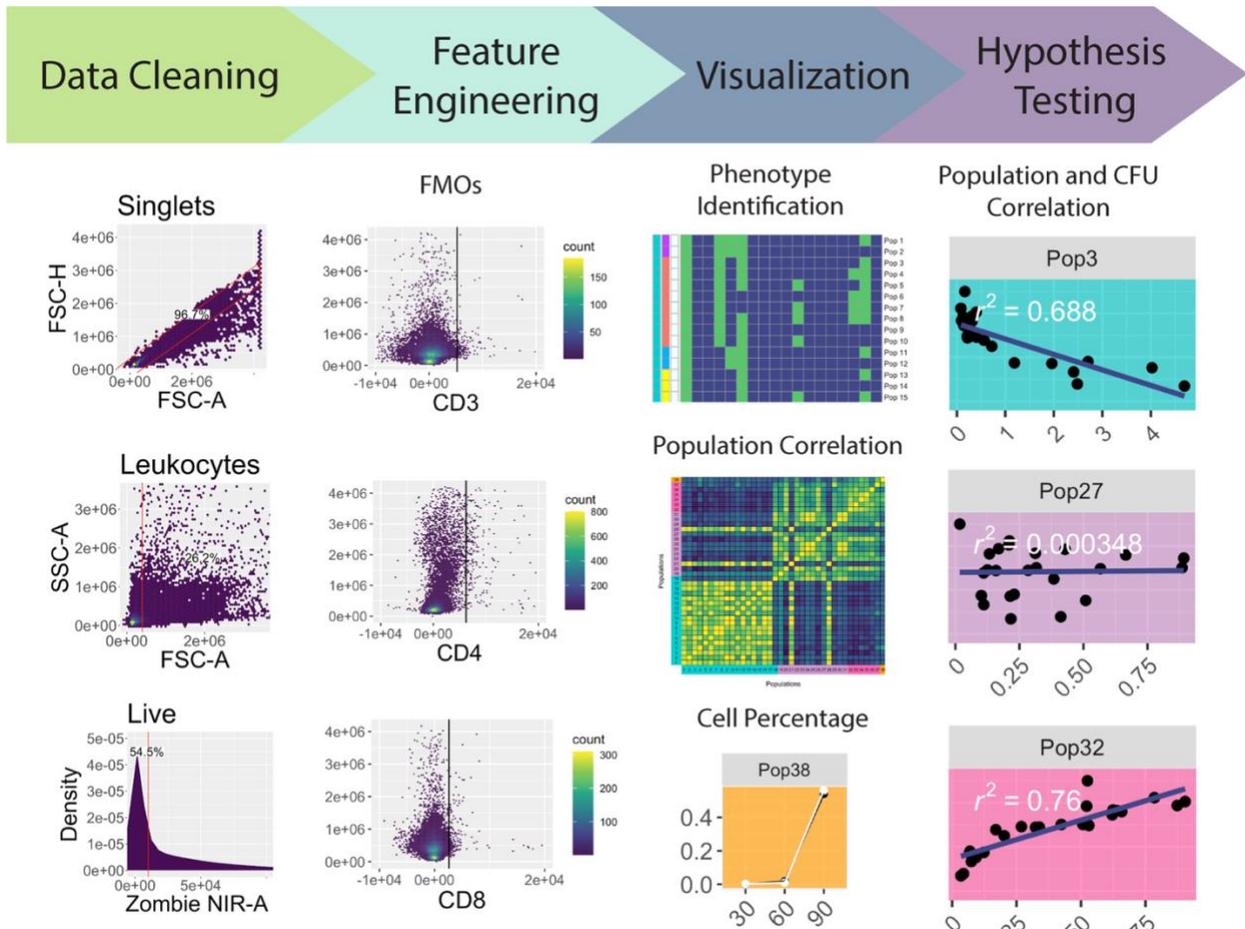


Figure 2.1: Pipeline workflow. Initial data cleaning is performed on all FMOs and samples. A singlet gate on the root population gates out doublets (top panel under “Data Cleaning”). Debris is removed from the samples with a “mindensity” gate from the R package *openCyto* (middle panel under “Data Cleaning”), and then the live cells (those negative for Zombie NIR) are gated, also using a “mindensity” gate (bottom panel under “Data Cleaning”). The data is then feature engineered into binary data based on FMOs. It is then filtered to a smaller number of populations that may help answer a research question, such as CD3+ cells. Finally, the data is visualized and statistically analyzed, for phenotype identification, population correlation, cell percentage, and population and CFU correlation.

Our pipeline processes the data from FMOs to include in further analysis. Traditionally, FMOs have been manually gated to identify the upper threshold of a parameter's value for negative cells. In our pipeline, we instead automate this analysis of the FMOs, measuring the threshold as the 99th percentile of the parameter values in each FMO (Figure 2.2). Noise can originate from very small particles or debris that pass through the flow cytometer. In an ideal world, a 100% threshold could be used, but in reality, the 99% threshold is used to account for this random noise. The user can assess this 99% threshold with the FMO plots in Figure 2.2 and adjust the thresholds if need be. The 99th percentile values are then saved and subsequently funneled into feature engineering of binary features (negative/positive) for each marker in the experimental data.

Feature engineering identifies all cell phenotypes present in the samples

Features are measurements in a dataset, such as the MFIs used in flow cytometry. Feature engineering is a machine learning technique that uses the original features in a dataset, possibly with the integration of external knowledge or data, to create new features that make the data easier to understand.^{17,18}

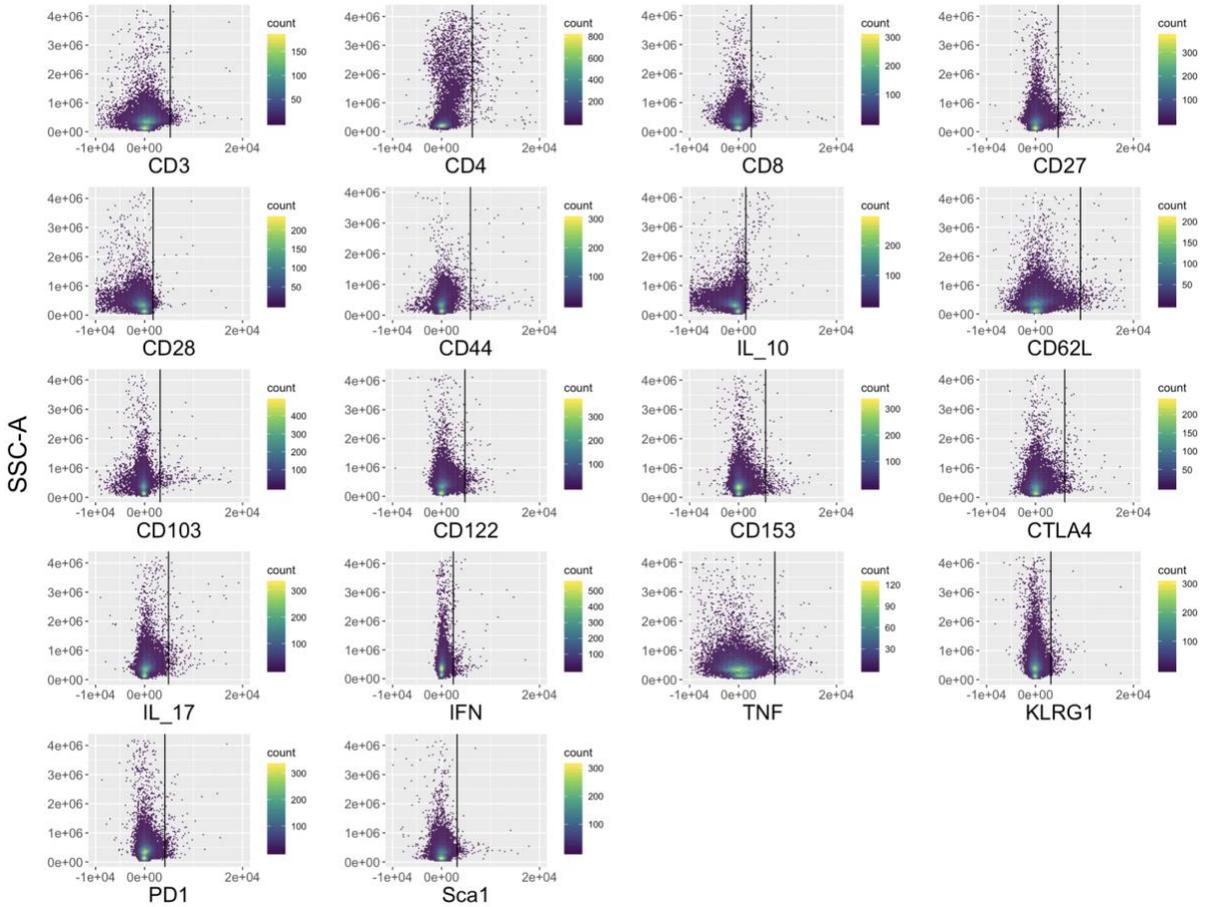


Figure 2.2: Numerical feature engineering of FMOs. The 18 FMOs are shown with the individual marker MFI expression on the x-axis and Side Scatter (SSC-A) on the y-axis. The black vertical line indicates the 99th percentile threshold for identifying positive versus negative cells (i.e., 99% of the data is located to the left of the line in each plot). These thresholds are used on the subsequent samples to feature engineer new parameters on whether a cell positively or negatively expresses each marker.

For flow cytometry, FMOs can add information about the possible range of expression measurements for cells that are truly negative for a marker. The threshold identified by FMOs can be used to create new, binary features that capture whether the expression of each marker is positive or negative for the cell, thus simplifying overly redundant, continuous MFI data with noise resulting from spillover. In the pipeline, we feature engineer each parameter using the thresholds identified from the FMOs, so that positive expression on cells (values above the FMO cutoff) equal 1, and negative cells equal 0.

For each cell in the experimental data, the cell phenotype is then identified based on the set of marker expressions (0's and 1's) of each population. Eighteen markers were used to elucidate memory T cell populations including markers for terminal differentiation and exhaustion in the *M. tuberculosis* case study. The pipeline identifies all cell populations (i.e., combinations of negative and positive marker expression values) for which at least one sample includes at least one cell. A total of 12,122 cell populations were identified in the samples for this study (Appendix 1: Supplemental Figure S2.2). As this number of populations is still very large, the data can be filtered to look at a smaller subset of the populations. In this case, we are specifically looking for CD3+ T cells that may mediate protection against *M. tuberculosis* infection. Immunologically, a protective population is unlikely to be present only in extremely small numbers. Therefore, in the filtering step of our pipeline we chose to filter to CD3+ T cell populations with population percentages greater than 0.5% in at least one sample (Figure 2.3a). This analysis filtered the cells to look specifically at larger populations, but an alternative filter could be used to look at rare populations that compose <0.1% of the sample, for example.

The pipeline then classifies specific lineages and subsets of cells according to marker expression (Appendix 1: Supplemental Tables 2.1 and 2.2).^{19,20} The pipeline identifies cell phenotypes that may be overlooked in manual analysis, due to combinations of markers that are uncommon (e.g., cells classified as “unknown” in this analysis) or combinations that have not yet been linked to a disease or condition of interest. In the case study, for example, there are six populations that express CD153; CD153 has only recently been shown to mediate protection against *M. tuberculosis*.²¹ These six double-negative populations likely would not have been gated for in a manual gating scheme, as thus far, the marker has only been shown to be present on CD4+ T helper cells.²¹

While we used T cell lineages and subsets in the case study analysis, the pipeline could easily be modified for different panels. If a panel aims to identify myeloid cells, they could be classified by lineage as macrophages, neutrophils, dendritic cells, and then subset further, for example, by alveolar macrophages and interstitial macrophages.

Correlation Between Identified Populations

The pipeline then visualizes a correlation matrix comparing the percentage of cells in each of the populations (Figure 2.3b). This allows users to explore associations between cell populations.

This plot can also allow for identifying unusual populations that do not behave like other populations within the same lineage. In the case study example, populations 21 and 28, which are T helper cell populations, are negatively correlated or uncorrelated with the other T helper cells (populations 19–31), and instead have patterns more similar to the identified double-negative cell populations. These are the only identified cell populations outside the

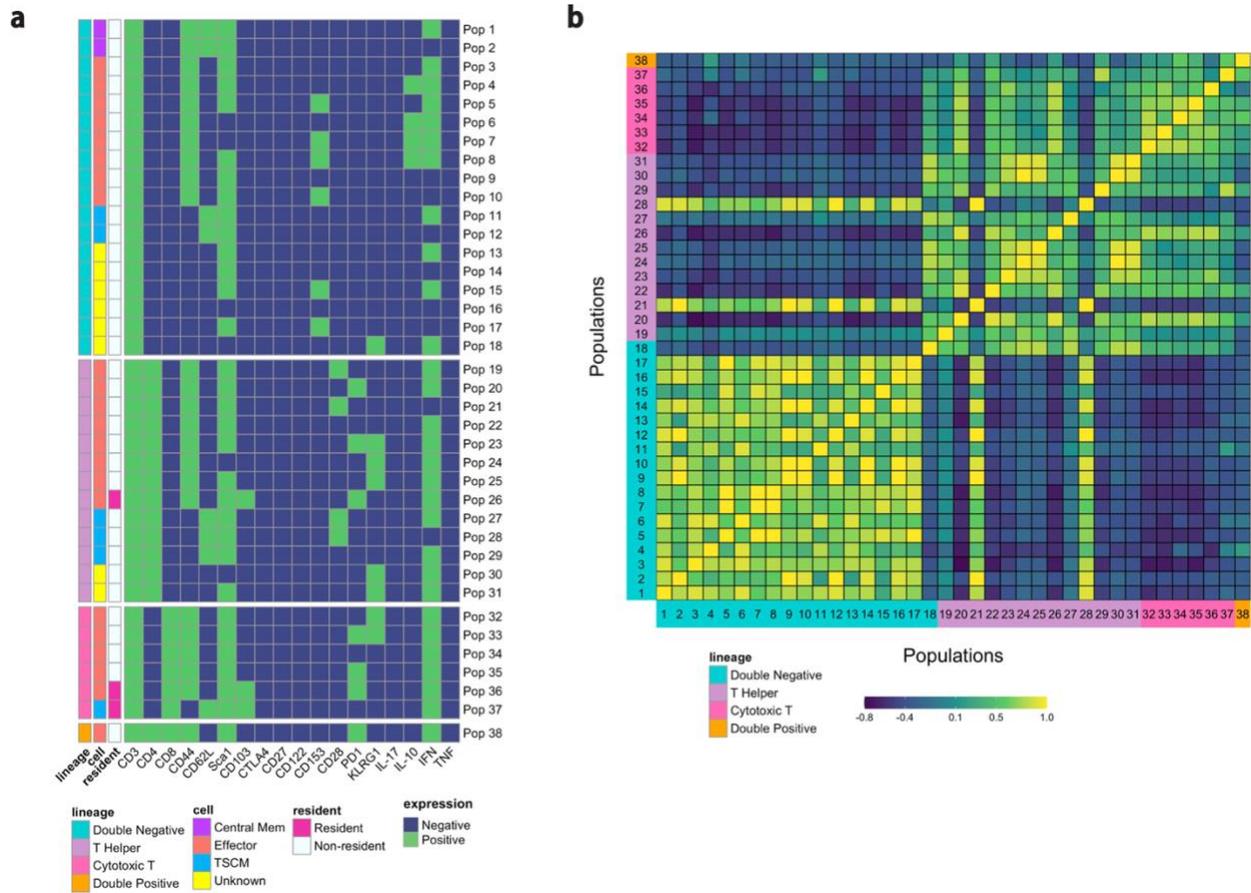


Figure 2.3: Feature engineering identifies CD3+ phenotypes and correlates identified populations. (a) The heatmaps show the CD3+ phenotypes that constitute greater than 0.5% of the live leukocytes in at least one sample. Green indicates positive expression, and blue indicates negative expression of all 18 markers used for analysis in the flow cytometry experiment on the x-axis. The plots are separated by four different CD3+ lineages based on CD4 and CD8 expression (double negative immune cells, T helper cells, cytotoxic T cells, and double positive T cells). The “cell” column classifies cells as either central memory, effector, T stem-cell like memory, or unknown cells that need to be explored further. The “resident” column indicates if the population is a resident cell, as determined by expression of CD103. **(b)** The correlation across study samples between the percentage of cells in each population can be used to see the similarities and differences between different populations. Yellow indicates high positive correlation and purple is high negative correlation. Populations are grouped by cell lineages, and each number on the x-axis and y-axis identifies a separate cell population, corresponding to the population numbers in Figure 2.3a.

double-negative cells that are negative for interferon-gamma (IFN- γ) expression. Importantly, IFN- γ is well known to be strongly associated with protection against *M. tuberculosis*.²² As these cell populations do not behave similarly to the other T helper cells, they may be a good candidate for further, more targeted exploration in later experiments.

Visualizing the percentage of cells in each population

Whereas less modular pipelines may provide more limited options for visualizations at later stages of analysis, our pipeline's modularity and its use of a common "tidy" data format provide the researcher wide flexibility to create visualizations suited to their research question and data characteristics through the *ggplot2* package.¹⁴ For this study, the percentage of cells in each population at each time point is plotted to compare the dynamic changes in populations over time and between groups (Figure 2.4). Some populations may steadily increase or decrease over time, while others behave in unexpected ways. Comparing the two groups, we can see that some cell populations (e.g., population 20) are similar over time regardless of vaccination status. It is also possible to elucidate differences based on cell lineages. For example, the largest differences in population percentages occurs at Day 30 post-infection, before BCG-vaccination protection begins to wane.²³ This difference primarily occurs in the double-negative CD3⁺ immune cell populations. This example visualization, and its accompanying code, is just one application of visualizations that could be created at this step of the workflow (see Appendix 1: Supplemental Figure S2.3 for two more examples).

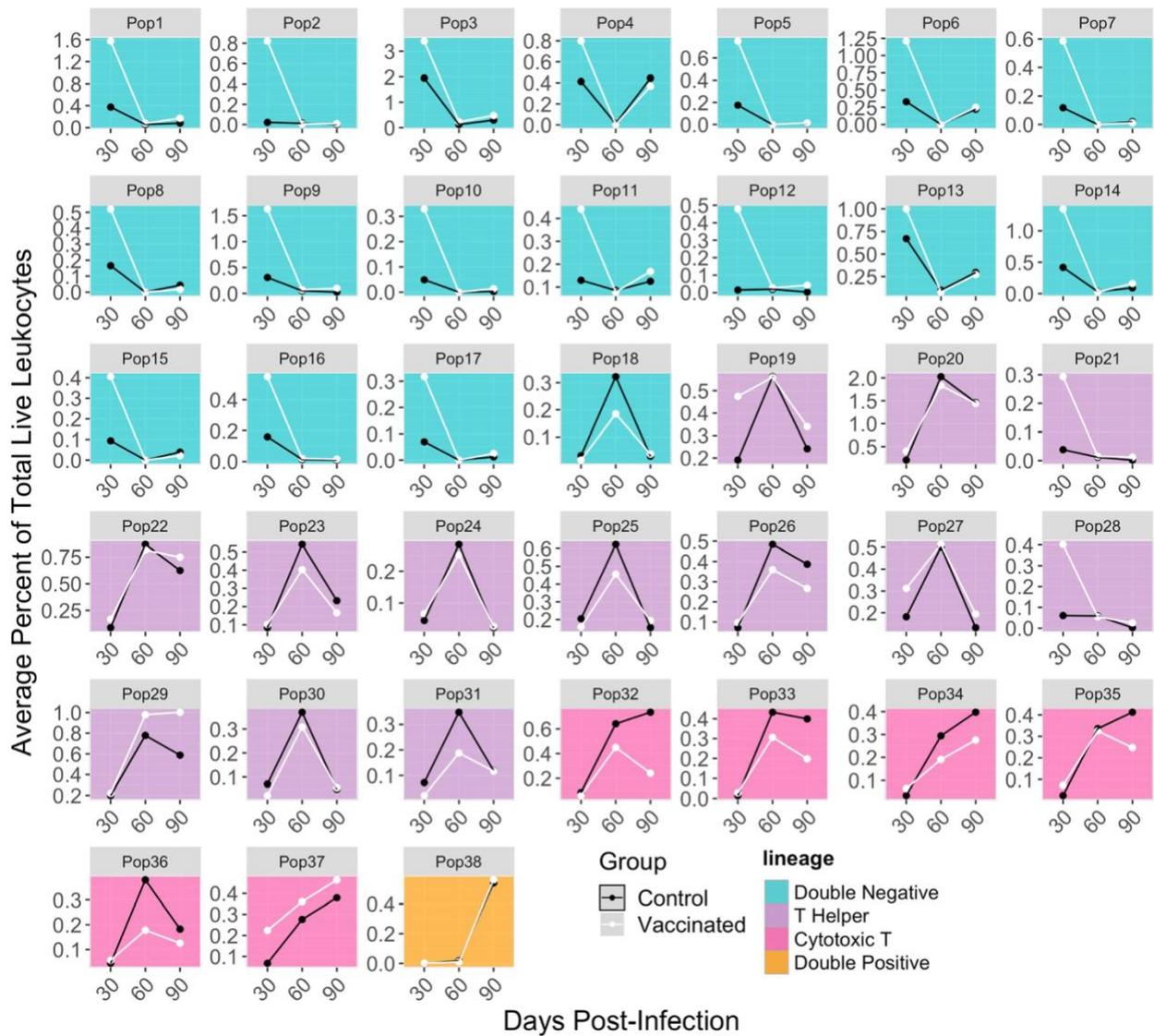


Figure 2.4: Time series of percentage of cells in each population. Each small plot shows the time series of a single cell population identified in the pipeline at the three measured time points post-infection. Separate lines are shown for vaccinated (“BCG”) versus control mice. Each point represents average cell populations across all mouse replicates (4–5 per time point and vaccination status). The small plots give cell populations in the same order as Figures 2.3 and 2.5, with population identification numbers corresponding across the three plots. The background color of each plot matches the cell lineages plotted in Figure 2.3a.

Integrating cell population measurements with other experimental measurements

At this stage, the pipeline allows the integration of cell population measurements with other data from the experiment, such as lesion scores or gene expression. In the *M. tuberculosis* study, bacterial burden (expressed as log₁₀-transformed Colony Forming Units (CFUs)) is a measurement of the number of bacteria found in the lung. These CFU measurements were found to vary between experimental groups in the case study data, with significantly higher bacterial burden at days 30, 60, and 90 post-infection in the control group compared to the vaccinated group (Figure 2.5a). It is of interest to investigate if certain cell populations identified through the pipeline, are associated with this measurement of bacterial burden, as this might help to identify cell populations possibly indicative of the host's response to infection with or without vaccination.

For this case study, the pipeline investigates associations between CFUs and cell populations using scatterplot visualizations and linear regression (Figure 2.5b). It tests the null hypothesis of a slope of zero for CFUs regressed on cell population size within a mouse's lung (Figure 2.5b and Appendix 1: Supplemental Table 2.3). Further, the coefficient of determination (r^2) was estimated between the CFUs and each cell population. This analysis identified that cells that co-express CD44, CD153, and IFN- γ (populations 5, 7, 8) could be candidates for future experimental studies. They are potentially protective against *M. tuberculosis* infection, as a decreased bacterial burden is associated with a higher percentage of these populations.

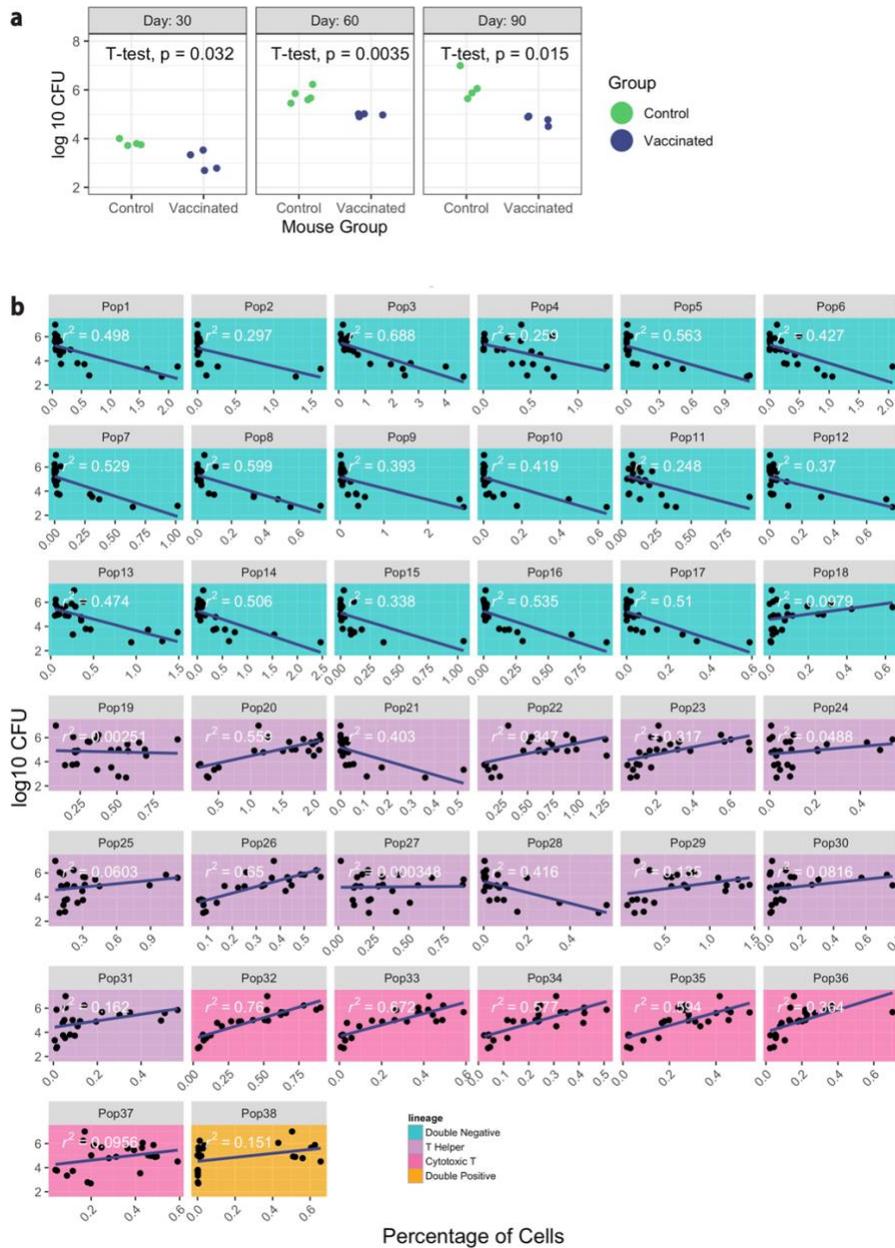


Figure 2.5: Populations associated with bacterial burden. (a) The log-10 *M. tuberculosis* CFUs in each mouse at each time point is shown, separated into vaccinated (“BCG”) and control groups. Unpaired t-tests were used to calculate statistical significance ($p < 0.05$). (b) Each small plot shows the association between a specific cell population and bacterial burden across all samples for the experiment. The small plots give cell populations in the same order as Figures 2.3 and 2.4, with population identification numbers and lineage colors corresponding across the three plots. The x-axis in each small plot gives the percentage of cells in a population, with each point providing the measurement from a single mouse. The y-axis gives the log-10 *M. tuberculosis* CFUs for that mouse. Estimates of how well the linear models fit the data between cell population sizes and log-10 CFUs are given on each plot (“ r^2 ”).

In this section of the pipeline, the user can modify the statistical model and visualization used in the case study for a wide range of alternatives. For example, the linear model fit in this step of the example pipeline is based on an assumption of linearity in the relationship between cell populations and CFUs, but exploratory analysis might identify that this assumption is incorrect. In this case, since the pipeline is modular, the code fitting the linear models can be replaced with R code to fit non-linear or non-parametric models. Further, the pipeline allows users to add analysis steps at this point. For example, when performing multiple statistical tests there is an increased possibility of identifying false statistically relevant comparisons.²⁴ The user might want to adjust the acceptable p-values for the multiple comparisons made in fitting linear models for each cell population. The pipeline could be extended with the Benjamin and Hochberg False Discovery Rate correction (Appendix 1: Supplemental Table 2.3) or other multiple testing corrections (e.g., Bonferroni or Benjamini & Yekutieli).²⁵⁻²⁷

Population validation via manual gating

As with the development of any new tool to analyze data, the pipeline must be tested against a traditional method, in this case, manual gating, to ensure that similar patterns and populations are captured in both analyses. To investigate if the estimated cell population sizes were similar between the automated and manual gating, a population for which we observed a relatively high percentage of cells (population 3) was manually gated (Appendix 1: Supplemental Figure S2.4). The percentage of cells in each of the populations at the different time points are very similar to the percentages calculated in this analysis pipeline with an average absolute difference of 0.26%. There is also a very high correlation coefficient ($\rho_{rs} = 0.99$, $p < 0.01$), indicating a strong positive association between the manual and pipeline gating (Appendix 1: Supplemental Figure S2.4).

Running time

Measurements for running time were made on a computer with 32GB RAM and a 4.2 GHz Intel Quad Core i7 processor. The initial data files contain a total of 7,299,424 cells (1.24GB); the FMO files contain 2,641,651 cells (0.57GB). Following gating, there were 468,754 cells from the FMOS and 1,023,402 cells from the sample data. These 1 million data points were input to the feature engineering algorithm. From the gating steps and feature engineering, through producing the plots in Figure 2.3a and Appendix 1: Supplemental Figure S2.2, the analysis script took 2.7 minutes running time. Half of this was spent on the gating steps (1.01 min). The running time was also evaluated based on the number of input cells to the feature engineering algorithm (Appendix 1: Supplemental Figure S2.5). It analyzes roughly 10,000 cells per second.

Testing pipeline with clinical human whole blood

Another dataset analyzing clinical human whole blood in a healthy individual was used to confirm the utility of this pipeline. This data was collected and compensated on a Fortessa II with a panel comprised of 5 markers: CD45, CD3, CD19, HLA-DR, and CD27. When used side-by-side, FMOs are a more robust depiction of the fluorescent marker composition in a sample as they account for spillover from other channels. However, it is not always possible to run FMOs. When there is good separation between the positive and negative populations within a marker, it is acceptable to base the MFI cutoffs on the sample data.¹⁵ The “mindensity” function from the *openCyto* package was used on the sample to determine thresholds for the feature engineering (Figure 2.6a). Following feature engineering, a total of 19 populations were identified in the sample (Figure 2.6b). The CD45+ populations, or leukocytes, were filtered and subset according to B cell and T cell lineages (Figure 2.6c). The percentage of cells in each of

the populations is shown in Figure 6d. Finally, the comparison of the percentage of cells between the manual gating and the automated pipeline indicated high similarity between the two methods (Figure 2.6e,f). The Spearman correlation coefficient (ρ_{rs}) between the two methods is 0.96 with a p-value < 0.001 . There is therefore slightly more discrepancy between automated and manual gating for this dataset compared to the murine lung cells, although the ranking agreement between the two methods is still very high. The use of this clinical dataset shows the utility and flexibility of the pipeline for conventional flow cytometry without the need for FMOs.

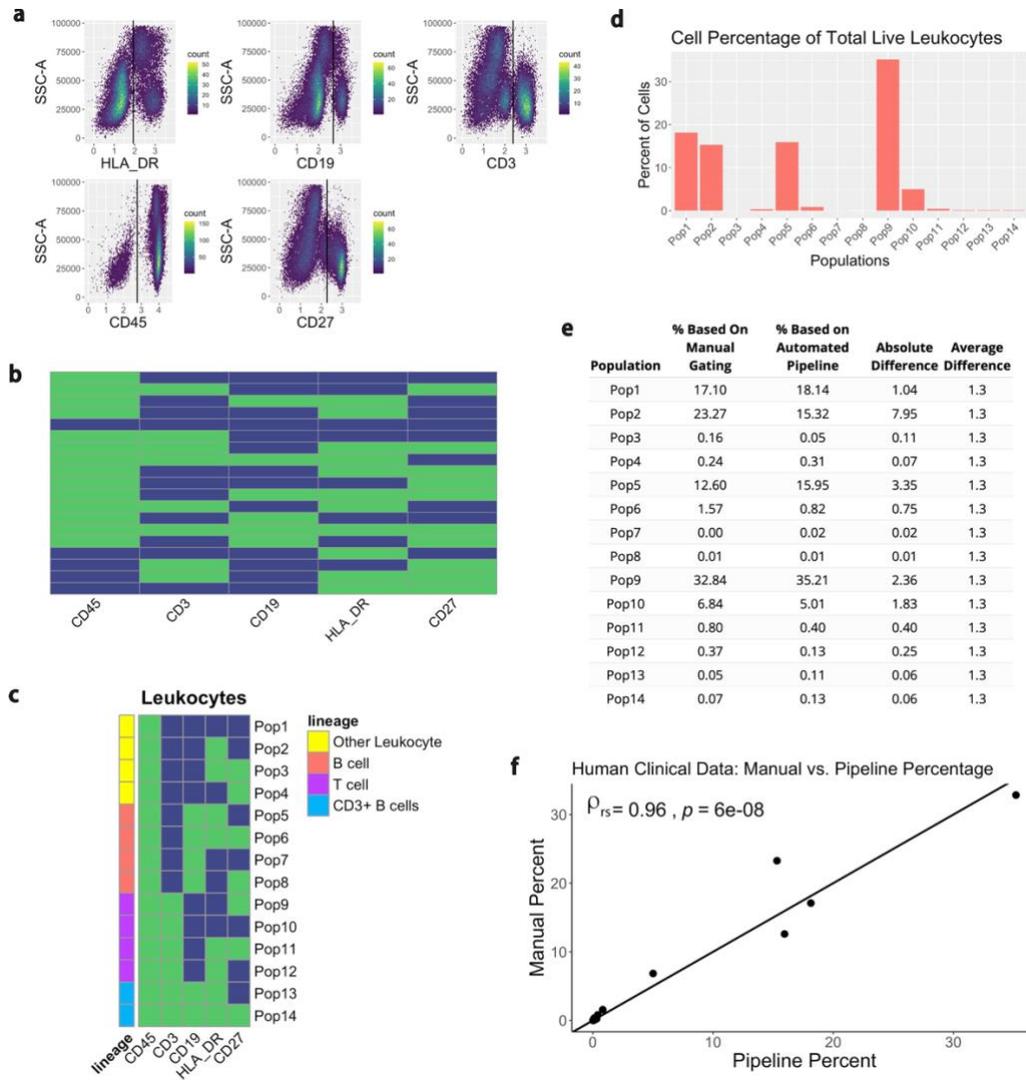


Figure 2.6: Testing the pipeline with human clinical samples. The data is first transformed so that it is viewable on a linear axis. (a) A “mindensity” gate from the R package *openCyto* is then used to determine the cutoff for positive and negative MFIs. (b) The data is feature engineered based on these cutoffs and identifies a total of 19 different populations. (c) The populations are then filtered to only view the leukocytes, or CD45+ populations. These populations are further subset by lineage. (d) The percentage of cells in each population is displayed. (e,f) Each population was manually gated in FlowJo and the difference in the percentage of cells from manual gating and the automated pipeline was compared. A table and plot compare the differences for each of the populations. Each point on the plot indicates a measurement for one of the 14 populations identified. The x-axis represents the value for the percentage of cells found via the automated pipeline, while the y-axis represents the percentage of cells found via manual gating. The $y = x$ line through the center of the plot indicates where the points would be located if both of the gating methods output identical values. The Spearman correlation coefficient (ρ_{rs}) comparing the methods and the p-value are displayed at the top left of the plot.

Discussion

The exploration of large and complex biological datasets requires steps of simplification and aggregation; scientific research is improved if these steps draw on biological knowledge and principles, rather than relying exclusively on subject-agnostic statistical techniques. Flow cytometers generate data with continuous measurements of an underlying binary phenomenon, with much of the added noise in measurements due to spillover from other channels. FMO controls can be used to informatively distinguish negative and positive cell populations for a marker in flow cytometry data and are used for this purpose commonly in traditional manual gating. However, they have not yet been integrated with the open-source tools being created to improve objectivity and efficiency of flow cytometry analysis. As with any simplification step, the pipeline presented here does lose some of the information inherent in the original measurements. In cells that are positive for a marker, for example, the measurement of expression density may have some biological meaning, which is lost in creating binary distinctions in the feature engineering. Future development of this pipeline could focus on extending the methods to help explore meaningful variation among the positive values for a marker. In the case of markers in which there is low, intermediate, and high expression, the 99th percentile FMO cutoff could be used to determine the negative population. From here, a function like “mindensity” used to analyze the clinical data could be used to determine the minimum density between the intermediate and high population, thus splitting the marker into three expression levels. If a more continuous expression level is of interest, instead of using the “mindensity” function to separate the positive markers into intermediate and high expression, the continuous marker expression could be retained.

This pipeline's modularity adheres to the Unix philosophy of combining small tools that each perform a discrete task to solve complex problems.²⁸ Excellent tools already exist in the R ecosystem, both specific to flow cytometry data (e.g., to perform initial gating on samples) and for more general data visualization and modeling. Rather than developing a "fixed" pipeline, where pre-existing tools are encapsulated within a framework that does not allow easy modifications or substitutions of steps in the pipeline, this pipeline is based on the principle of combining these existing tools from the R ecosystem. We designed this pipeline to use a common data format in later steps, which makes it useful to a variety of researchers, as it is easy to adapt with common R tools. For example, in the visualization and statistical testing stages, different statistical models and plots can easily be substituted for the linear regression that is used for the data described here. Another advantage of this modular pipeline is that it can be used on clinical and research cytometry samples from either conventional or spectral flow cytometers.

An added strength of the pipeline is that, since it is computationally simple, it does not require down-sampling of the raw flow cytometry data. For clustering algorithms that are computationally intensive (e.g., t-SNE), it is often necessary to down-sample data by taking a random sample of the data in order to improve computation time.^{29,30} This manipulation of data can lead to the masking of novel discoveries. Each sample has a different total live leukocyte cell count allowing us to consider percentage of cells rather than total cell count throughout the pipeline. As our goal is to identify immune cells that are biologically involved in the protection against disease, we filtered the cell populations to those comprising >0.5% of cells in at least one sample. In this case, even if only one animal in a group has a high percentage of cells in a certain population, it will be maintained to compare against all groups. In other contexts, for example, in cancer research where the goal is to identify very rare cell types, this percentage could be

decreased with a low-set filter (e.g., to maintain populations with ≤ 5 cells). The filtering and the feature engineering analyze the data in a way that makes it immediately interpretable to immunologists without the hassle and bias of manually looking for populations of interest or trying to understand complex clusters from other methods.

Another advantage of the pipeline is its ability to compile additional flow cytometry datasets. Other multiparametric approaches, such as UMAP and t-SNE, typically assign cells to unnamed clusters.^{5,31} To compile replicated data, these methods map additional data back to the original data. Unlike these approaches, cyto-feature engineering does not assign data to unnamed clusters, but rather keeps groupings explicitly tied to specific, named markers in the original data. Provided that replicated or additional data contains the same markers as the original data, it can be seamlessly added to the pipeline by adding the data to the data folder. Any additional cell populations identified will be added to the plots when the analysis is run. Further, the data with cyto-feature engineering does not need to be manually gated to determine statistical significance between populations. Statistical analysis is a built-in feature that can be calculated based on the identified populations.

Researchers have an inherent bias in the types of cells they gate when analyzing flow cytometry data. By identifying all phenotypes present in the dataset, this pipeline allows users to analyze a variety of populations in an unbiased manner. These cell types, such as the CD153+ cells in the murine lung data, may have gone undiscovered if not for the use of this fast and reproducible pipeline. Importantly, this analysis pipeline relies on high quality flow cytometry methodology, and/or FMO samples, as well as, strong panel design. Spillover from other channels can greatly impact the analysis, so researchers must ensure that the controls are prepared correctly.

The analysis pipeline described here, allows for the use of necessary and rigorous technical controls in flow cytometry. The pipeline is able to identify populations that may not be normally gated. Provided that the samples and controls remain constant, the automated pipeline analysis will consistently produce the same results, removing person-to-person bias. Further, this pipeline drastically reduces the amount of time typically required to analyze flow cytometry data. Overall, this strategy is envisioned to help identify the elusive nature of cellular phenotypes through fast and accurate analysis of flow cytometry data.

REFERENCES

1. BD Biosciences. BD FACSymphony™.
2. Verschoor, C.P., Lelic, A., Bramson, J.L., and Bowdish, D.M.E. (2015). An introduction to automated flow cytometry gating tools and their implementation. *Frontiers in Immunology* 6, 380. 10.3389/fimmu.2015.00380.
3. Finak, G., Frelinger, J., Jiang, W.X., Newell, E.W., Ramey, J., Davis, M.M., Kalams, S.A., De Rosa, S.C., and Gottardo, R. (2014). OpenCyto: An Open Source Infrastructure for Scalable, Robust, Reproducible, and Automated, End-to-End Flow Cytometry Data Analysis. *Plos Computational Biology* 10, e1003806. 10.1371/journal.pcbi.1003806.
4. Van Gassen, S., Callebaut, B., Van Helden, M.J., Lambrecht, B.N., Demeester, P., Dhaene, T., and Saeys, Y. (2015). FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry Part A* 87A, 636-645. 10.1002/cyto.a.22625.
5. van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, 2579-2605.
6. Platon, L., Pejowski, D., Gautreau, G., Targat, B., Le Grand, R., Beignon, A.S., and Tchitchek, N. (2018). A computational approach for phenotypic comparisons of cell populations in high-dimensional cytometry data. *Methods* 132, 66-75. 10.1016/j.ymeth.2017.09.005.
7. Beyrend, G., Stam, K., Hollt, T., Ossendorp, F., and Arens, R. (2018). Cytofast: A workflow for visual and quantitative analysis of flow and mass cytometry data to discover immune signatures and correlations. *Computational and Structural Biotechnology Journal* 16, 435-442. 10.1016/j.csbj.2018.10.004.
8. Arroyo, L., Rojas, M., Franken, K., Ottenhoff, T.H.M., and Barrera, L.F. (2016). Multifunctional T Cell Response to DosR and Rpf Antigens Is Associated with Protection in Long-Term Mycobacterium tuberculosis-Infected Individuals in Colombia. *Clinical and Vaccine Immunology* 23, 813-824. 10.1128/cvi.00217-16.
9. Marin, N.D., Paris, S.C., Rojas, M., and Garcia, L.F. (2013). Functional profile of CD4+ and CD8+ T cells in latently infected individuals and patients with active TB. *Tuberculosis* 93, 155-166. 10.1016/j.tube.2012.12.002.
10. Marin, N.D., Paris, S.C., Rojas, M., and Garcia, L.F. (2012). Reduced Frequency of Memory T Cells and Increased Th17 Responses in Patients with Active Tuberculosis. *Clinical and Vaccine Immunology* 19, 1667-1676. 10.1128/cvi.00390-12.
11. Spidlen, J., Moore, W., Parks, D., Goldberg, M., Bray, C., Bierre, P., Gorombey, P., Hyun, B., Hubbard, M., Lange, S., et al. (2010). Data File Standard for Flow Cytometry, Version FCS 3.1. *Cytometry Part A* 77A, 97-100. 10.1002/cyto.a.20825.
12. Hahne, F., LeMeur, N., Brinkman, R.R., Ellis, B., Haaland, P., Sarkar, D., Spidlen, J., Strain, E., and Gentleman, R. (2009). flowCore: a Bioconductor package for high throughput flow cytometry. *Bmc Bioinformatics* 10, 106. 10.1186/1471-2105-10-106.
13. Hahne, F., N, G., AH, K., C, W., and K, L. (2021). *flowStats: Statistical methods for the analysis of flow cytometry data*.
14. Wickham, H. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.

15. Roederer, M. (2002). Compensation in flow cytometry. *Current protocols in cytometry Chapter 1*, Unit 1.14-Unit 11.14. 10.1002/0471142956.cy0114s22.
16. Saeys, Y., Van Gassen, S., and Lambrecht, B.N. (2016). Computational flow cytometry: helping to make sense of high-dimensional immunology data. *Nature Reviews Immunology 16*, 449-462. 10.1038/nri.2016.56.
17. Zheng, A., and Casari, A. (2018). *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. O'Reilly.
18. Kuhn, M., and Johnson, K. (June 21, 2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models*. Taylor & Francis Group.
19. Henao-Tamayo, M.I., Ordway, D.J., Irwin, S.M., Shang, S.B., Shanley, C., and Orme, I.M. (2010). Phenotypic Definition of Effector and Memory T-Lymphocyte Subsets in Mice Chronically Infected with *Mycobacterium tuberculosis*. *Clinical and Vaccine Immunology 17*, 618-625. 10.1128/cvi.00368-09.
20. Wu, K., Wang, F., Guo, G., Li, Y.Q., Qiu, L.J., and Li, X. (2019). CD4(+) TSCMs in the Bone Marrow Assist in Maturation of Antibodies against Influenza in Mice. *Mediators of Inflammation*, 3231696. 10.1155/2019/3231696.
21. Sallin, M.A., Kauffman, K.D., Riou, C., Du Bruyn, E., Foreman, T.W., Sakai, S., Hoft, S.G., Myers, T.G., Gardina, P.J., Sher, A., et al. (2018). Host resistance to pulmonary *Mycobacterium tuberculosis* infection requires CD153 expression. *Nature Microbiology 3*, 1198-1205. 10.1038/s41564-018-0231-6.
22. Flynn, J.L., Chan, J., Triebold, K.J., Dalton, D.K., Stewart, T.A., and Bloom, B.R. (1993). An Essential Role for Interferon-Gamma in Resistance to *Mycobacterium-tuberculosis* Infection. *Journal of Experimental Medicine 178*, 2249-2254. 10.1084/jem.178.6.2249.
23. Ordway, D.J., Shang, S.B., Henao-Tamayo, M., Obregon-Henao, A., Nold, L., Caraway, M., Shanley, C.A., Basaraba, R.J., Duncan, C.G., and Orme, I.M. (2011). *Mycobacterium bovis* BCG-Mediated Protection against W-Beijing Strains of *Mycobacterium tuberculosis* Is Diminished Concomitant with the Emergence of Regulatory T Cells. *Clinical and Vaccine Immunology 18*, 1527-1535. 10.1128/cvi.05127-11.
24. Franceschi, P., Giordan, M., and Wehrens, R. (2013). Multiple comparisons in mass-spectrometry-based -omics technologies. *Trac-Trends in Analytical Chemistry 50*, 11-21. 10.1016/j.trac.2013.04.011.
25. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate – A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Statistical Methodology 57*, 289-300.
26. Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics 29*, 1165-1188.
27. Curtin, F., and Schulz, P. (1998). Multiple correlations and Bonferroni's correction. *Biological Psychiatry 44*, 775-777. 10.1016/s0006-3223(98)00043-2.
28. Gancarz, M. (1995). *The Unix Philosophy*.
29. Belkina, A.C., Ciccolella, C.O., Anno, R., Halpert, R., Spidlen, J., and Snyder-Cappione, J.E. (2019). Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. *Nature Communications 10*, 5415. 10.1038/s41467-019-13055-y.
30. Crawford, T.Q. *Dimensionality Reduction with the tSNE Algorithm*.

31. Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* 37, 38-+. [10.1038/nbt.4314](https://doi.org/10.1038/nbt.4314).

CHAPTER 3 — IMMUNOMODULATORY DRUGS AS VACCINES AGAINST MYCOBACTERIUM TUBERCULOSIS

Introduction

It is estimated that one quarter of the world's population is infected with the bacteria *Mycobacterium tuberculosis* (MTB), killing an estimated 1.4 million people each year.¹ The only licensed vaccine currently available against tuberculosis (TB) is the Bacille Calmette–Guérin (BCG) which is a live-attenuated strain of *Mycobacterium bovis*. While BCG vaccine confers protection against meningitis and disseminated TB in children, it fails to protect from pulmonary TB and the reactivation of latent infection.² Therefore, an improved vaccination boosting strategy is desperately needed.

Research efforts to develop a better vaccine regimen have focused on immunomodulation, or ways to modify the immune response. One approach is to combine the BCG vaccine with drugs that are already FDA-approved. This method of repurposing drugs is beneficial both to patients and the pharmaceutical industry as it bypasses the lengthy, expensive drug safety clinical trials. This brings safe treatments to patients faster and costs pharmaceutical companies less money.

An example of immunomodulators currently used in vaccines are adjuvants. Adjuvants are additional constituents added to vaccines that induce a stronger immune response. These adjuvants typically produce non-specific inflammation and recruit antigen presenting cells to the draining lymph nodes.³ In many cases, this inflammation is favored to achieve the desired immune response. However, recent work has shown that high doses of vaccine adjuvants can

cause migration of neutrophils and monocytes that suppress T cell responses and actually decrease the desired immune response.³ As a potential remedy, Mitchell et al. demonstrated that blocking monocyte recruitment to the lymph nodes can increase both the cellular and humoral immune responses to vaccination. This is thought to be true in the case of BCG. During BCG vaccination, it is hypothesized that the large influx of inflammatory cells leads to the rapid killing of BCG.⁴ This rapid killing decreases antigen presentation, and thus, T cell responses.⁴ Therefore, we believe if we reduce inflammation, macrophages will be exposed to antigen for longer which will help generate better T cell memory.

Losartan is an example of an angiotensin receptor blocker drug that has been shown to reduce inflammatory monocyte recruitment. While losartan is an FDA-approved drug for treating high blood pressure, recent studies have elucidated losartan's immunological effects. Cancer studies have shown losartan inhibits migration of CCL2-stimulated CCR2+ monocytes, cells that have been shown to promote metastasis when recruited to the lung.^{5,6} Other studies have shown that losartan's anti-inflammatory immunomodulatory effects impede the development of pro-inflammatory monocytes.⁷ This has been confirmed in studies where losartan reduces murine inflammatory macrophages and micro-metastases.⁵ Longitudinal studies have also demonstrated that the use of angiotensin-converting enzyme inhibitors (ACEi), inhibitors that similarly impede the renin angiotensin system, are associated with a decreased risk of active TB.⁸ In pneumonia, lower mortality associated with the use of these ACEis is due to the ACEi's ability to modulate the immune response and decrease levels of pro-inflammatory IL-6.⁹⁻¹¹ Thus, losartan provides a good candidate for reducing the number of inflammatory monocytes during vaccination.

Another drug, propranolol, could also be effective in modulating the immune response to vaccines by increasing T cell responses. Propranolol is a β 1- and β 2-adrenergic receptor antagonist used to lower blood pressure. In addition to being found on cells in the heart, kidney, and airway smooth muscles, β -2 adrenergic receptors are also found on immune cells such as Th0, Th1 and B cells.^{12,13} Typically, catecholamines such as epinephrine or norepinephrine, bind to and activate these β -adrenergic receptors, causing an increase in heart rate (Figure 3.1). This typical stimulation of receptors on antigen presenting cells and T lymphocytes promotes Th2 responses through inhibition of IL-12.¹⁴ However, propranolol competitively binds these receptors, effectively lowering heart rate and blood pressure. It has been shown that blocking the typical β -adrenergic receptor signaling pathway, such as with a drug like propranolol, leads to an increase in CD4 T cell proliferation and protective Th1 cytokines.^{15,16} The addition of β 2-adrenergic antagonists have also been shown to increase antigen presentation by dendritic cells.¹⁷ This increased antigen presentation can lead to a better T-cell mediated response. β 2 antagonists can increase CD8+ effector memory T cell expression of IFN- γ , TNF- α secretion, and cytolytic abilities, and propranolol treatment, specifically, has been shown to increase CD8+ T cell expression of T-bet, IFN- γ , and GzmB.^{18,19}

Our goal in this study was to determine if losartan and propranolol, when used with the BCG vaccine, could induce better protection against MTB infection. Specifically, we hypothesize that decreasing the number of inflammatory monocytes recruited to the lungs during vaccination will generate a better cell mediated and memory response. Losartan should reduce this influx of inflammatory macrophages, allowing dendritic cells to perform antigen presentation and thus

better T cell activation. We also hypothesized that propranolol would promote T cell polarization to Th1 cells and enhance cytolytic capabilities of CD8+ T cells.

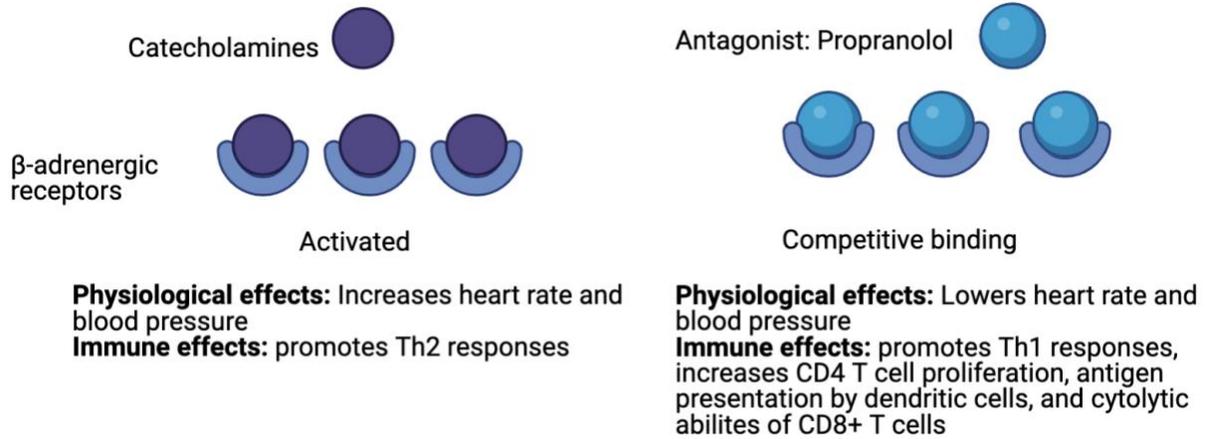


Figure 3.1: Effects of catecholamines and propranolol binding to β-adrenergic receptors.

Methods

Experimental Setup. The experiment was designed to determine if two drugs — propranolol and losartan — can act as immunomodulators in conjunction with the BCG vaccine to induce better protection against MTB. Drugs and vaccinations were administered as detailed in Figure 3.2. Mice received drugs either orally or via intraperitoneal injections two days prior to receiving the BCG vaccine. Mice in the injection groups received drugs for a total of eight days. Oral drugs were given weekly for four weeks. Lung and spleen tissue were taken at days -22, -1 before MTB infection and days 21 and 56 post infection. Bacterial burden and immune cell infiltration were quantified.

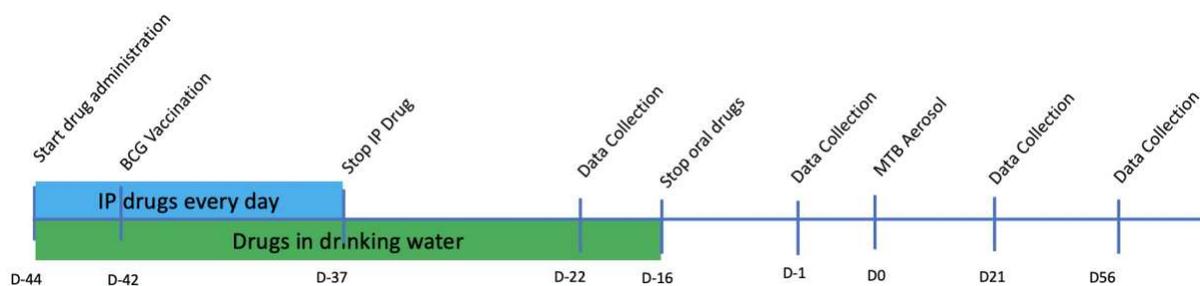


Figure 3.2: Experimental Design. A timeline of the administered drugs, vaccinations, and time points are detailed.

Animals. C57BL/6 mice were purchased from the Jackson Laboratory. The mice were maintained at a BSL-2 facility at Colorado State University until one week prior to infection with *Mycobacterium tuberculosis*. At which point they were transferred to a BSL-3 facility. Animals were acclimated for at least one week and fed water and chow ad libitum. All experimental protocols were approved by the Institutional Animal Care and Use Committee at Colorado State University.

BCG Vaccination. Mice were subcutaneously vaccinated with 5×10^4 colony forming units (CFU) Bacille Calmette-Guérin, Pasteur strain.

Drug Administration. Drugs were reconstituted in sterile biological grade water and sterile filtered with a $0.22 \mu\text{m}$ filter. Injections were reconstituted in phosphate buffer saline (PBS) and $100 \mu\text{L}$ was administered intraperitoneally. Losartan was injected daily at 60 mg/kg/day , while propranolol was given daily at 10 mg/kg/day .^{5,15} The systemic bioavailability of losartan after oral ingestion is about 33% causing us to increase the dose in the water.²⁰ For oral dosing, drugs were reconstituted in purified drinking water at the following concentrations: 200 mg/kg/day of losartan and/or 10 mg/kg/day of propranolol. The water was changed weekly.

Mycobacterium tuberculosis infection. Mice were infected via aerosol route with MTB HN878 using a Glas-Col apparatus (Terre Haute, IN) calibrated to deliver 50-100 bacilli per mouse. Mice received 50 CFU/mouse as confirmed via plating whole lung homogenates on 7H11 agar plates following the aerosol.

Animal euthanasia, necropsy, and tissue preparation. On the days of the experiment, mice were euthanized using CO₂ asphyxiation and lungs and spleens were harvested. Lung lobes were perfused with PBS prior to generating single-cell suspension. Lung and spleen tissues were prepared for single-cell suspension according to the protocol described by Fox et al.²¹

Flow cytometry and analysis. Multi parametric spectral flow cytometry was performed to analyze the immune cell phenotype changes during the course of drug treatment with and without infection. Since we want to evaluate a broad range of innate and adaptive cell types, we created two flow panels. The first panel contained 14 surface markers, showing an overview of the different cells present: CD11c, CD3, NK1.1, Ly6G, CD45, CD14, Ly6C, CD11b, CD4, CD8, CD19, CCR, CD64, and viability dye. The second panel contained a combination of 19 surface markers and transcription factors to analyze the presence of different types of T cells: Sca-1, CD3, CD62L, FoxP3, CD28, PD-1, ROR-γT, GATA3, CD103, CD44, CD69, CD4, CD25, CD8, T-bet, CD27, CD153, KLRG-1, and viability dye. Before staining, we counted the cells on an LSRII as described in Fox et al. so that we could calculate total numbers.²¹ To begin the staining process, cells were first stained with 1/2000 dilution of Zombie-NIR (fixable viability dye) and subsequently washed with fluorescence activated cell sorting staining buffer (FACS).

The surface markers and 1/200 dilution of FC receptor block were added and incubated for 30 minutes at 4°C and then washed with FACS. For the first panel that contained only surface markers, 4% paraformaldehyde (PFA) was added to the samples for 20 minutes, washed with FACS and then transferred to tubes. For the intracellular staining of transcription factors in the second panel, 1x FoxP3 transcription factor fix/perm buffer (ebiosciences) was added to the cells and incubated for 1 hour. The cells were subsequently washed with 1x FoxP3 Permeabilization buffer and the transcription factor antibodies were added and incubated overnight at 4°C. The following day, the cells were washed with FoxP3 Permeabilization buffer and transferred to flow cytometry tubes. 50,000 events for lungs and 100,000 events for spleen cells per sample were acquired on a 4-laser Cytex Aurora. Flow cytometry data was analyzed using the cyto-feature engineering pipeline.²²

BCG and MTB Bacterial burden. Fifty percent of the whole lung and 50% of the spleens were homogenized in PBS in a bullet blender (Next Advanced) for 4 minutes at speed 8. Homogenates were serially diluted in PBS at a dilution factor of 5 and plated on 7H11 agar plates. CFUs were counted after incubating plates for 5 weeks at 37°C.

Results

Use of losartan and propranolol did not reduce MTB bacterial burden

The goal of this experiment was to evaluate if the drugs losartan and propranolol could modulate the immune response to MTB, decrease MTB load in the lungs, and provide better protection than BCG alone. For this reason, we were able to use Dunnett's analysis to determine statistical significance from the BCG group. Comparisons were made for both the lung and spleen organs

and at each of the time points. The drugs did not affect the presence of live BCG present in the organs following vaccination (Figure 3.3). After infection, the only difference exhibited was in the PBS group which had significantly higher MTB CFUs in the lungs at day 21 than the BCG group. These results show that the vaccines in conjunction with the drugs did not control MTB infection better than the BCG vaccine alone.

Drugs affect the total numbers of cells that infiltrate the organs

To assess the general influx of cells into the organs, total numbers of live leukocytes were calculated using the counting bead protocol found in Fox et al. 2020.²¹ At day -22 the BCG group had a higher number of cells in the lung compared to all of the other groups (Figure 3.4). This suggests that losartan and propranolol reduced the influx of cells to the lungs.

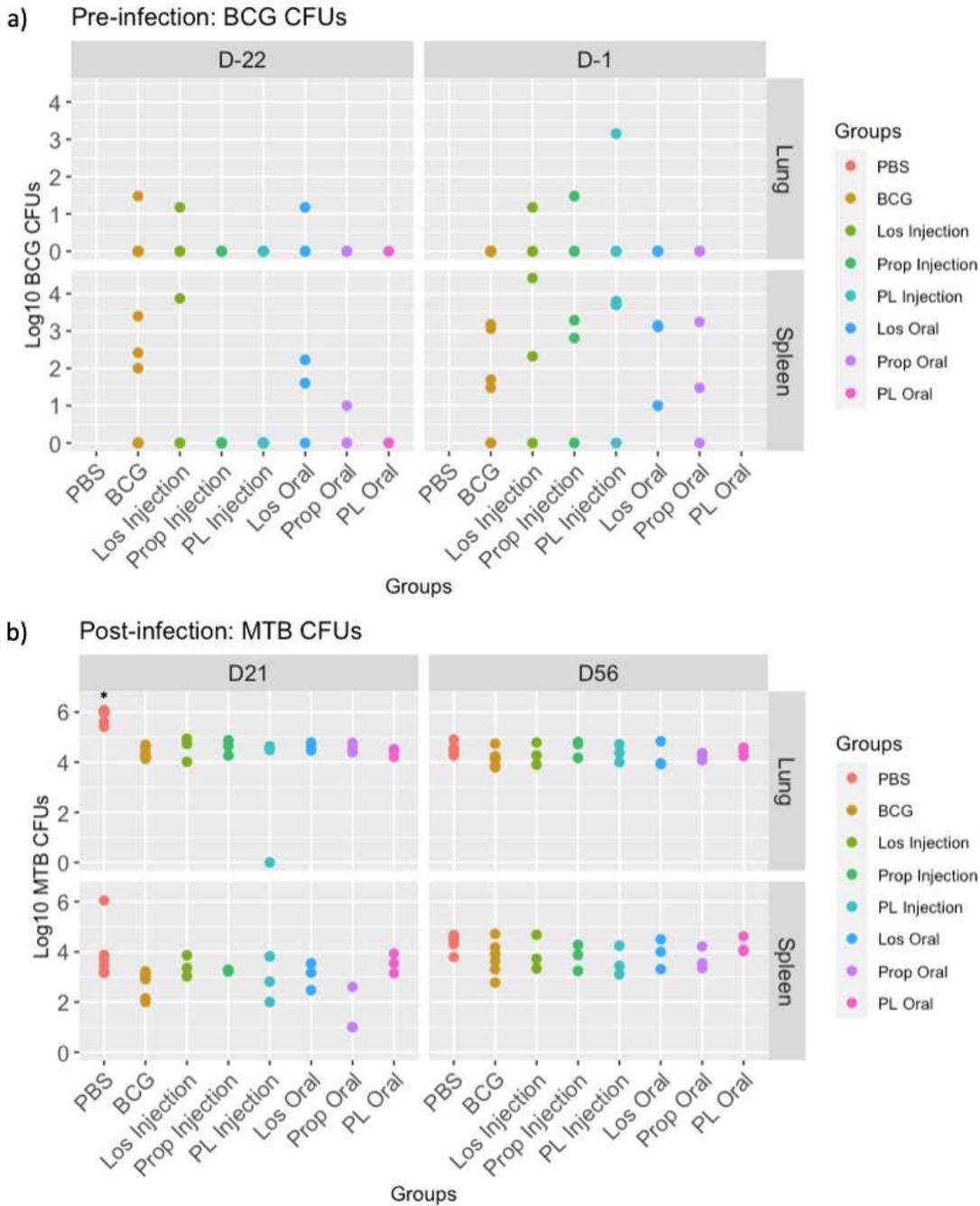


Figure 3.3: Higher CFUs in the PBS group compared to the BCG group at day 21 in the lung. The log₁₀ CFUs for either the (a) live BCG vaccine at days -22 and -1 or (b) MTB infection at days 21 and 56 in each mouse is shown. The CFUs in the lungs are shown on the top row and the spleen is shown on the bottom row. Dunnett's analysis was performed to all groups to the BCG group (*p<0.05). Los: Losartan, Prop: Propranolol, PL: Propranolol + Losartan

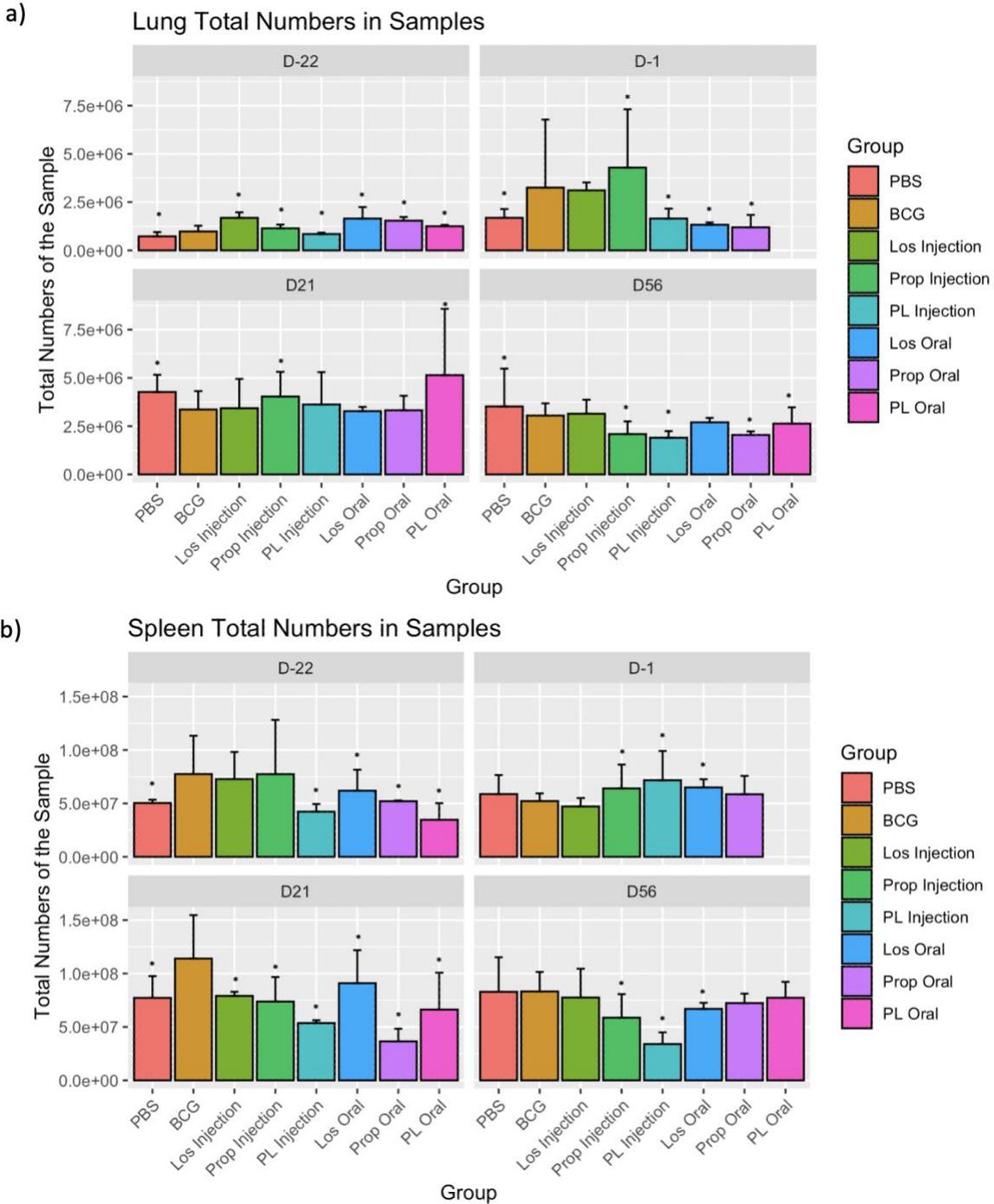


Figure 3.4: Total numbers of cells in the organs. The total numbers of cells in the (a) lungs and (b) spleens is displayed in the bar charts, faceted by the time points. Dunnett’s test was used to calculate statistically significant differences between the BCG group and all other groups (* $p < 0.05$).

Before infection, the lung PBS control group had lower total cell counts (day -22: $7.3 \times 10^5 \pm 2.4 \times 10^5$ cells, day -1: $1.7 \times 10^6 \pm 4.9 \times 10^5$ cells) compared to the BCG group (day -22: $9.8 \times 10^5 \pm 3.3 \times 10^5$ cells, day -1: $3.3 \times 10^6 \pm 3.8 \times 10^6$ cells), but after infection, the cell counts were significantly higher. This is consistent with what we would expect as the BCG vaccine was able to induce proliferation of cells in response to the vaccine stimulus. After the PBS group is aerosolized with MTB, the immune response ramps up to handle the infection. At day 56, while the lung PBS group had more cells compared to the BCG group, all of the groups except for the losartan injection had significantly fewer cells. At this time point, there were no differences in CFU counts. For this reason, it does not appear as if the sheer number of cells in the lungs had an effect on the bacterial burden.

The number of cells in the spleen maintained a more constant number compared to the lungs over the course of the study (Figure 3.4). Interestingly, at day 21 the number of cells in the BCG group is significantly higher ($1.1 \times 10^8 \pm 4.4 \times 10^7$ cells) than all of the other groups. However, by day 56 the number of cells in the BCG group decreases to a similar level as the PBS group.

In both the lung and spleen, the drug groups were unable to produce high levels of cells as shown in the BCG group. While these differences did not ultimately have a difference on the bacterial burden, we wanted to explore the composition of the cell populations to determine the types of cells produced.

Losartan increases the numbers of specific immune cell populations in the lungs compared to the BCG group

Both predefined and data-driven analyses were used to identify the flow cytometry cell populations. For the predefined analysis, cell phenotypes were determined according to predefined combinations of positive and negative expression of various markers (Appendix 2: Supplemental Table 3.1). This analysis identified populations that we anticipated seeing in the lungs.

Based on the predefined analysis, most differences between groups occurred at day -22, after the mice had received both the BCG vaccination and drugs (Figure 3.5, Appendix 2: Supplemental Table 3.2). The losartan groups tended to have the largest differences in comparison to the BCG group. In particular, in the lungs the losartan injection group had higher numbers of alveolar macrophages, monocytes, B cells, CD4 and CD8 T cells, T Helper 1, Regulatory T cells, and Effector T cells at day -22 compared to the BCG group. The oral losartan group similarly had higher numbers of naive T cells, T Helper 1 cells, and central memory T cells at this time point. Both of these groups had lower total numbers of cells in the lungs than the BCG group at day -22, showing that the increase in these cell types is not due to differences in the total number of cells in the organ. This suggests that losartan is able to mount a greater immune response early during, and after administration of the drug, and that losartan is acting as an immunomodulator that aids in the recruitment of beneficial immune cells in order to clear infection. We expected to see an increase in T cell activation in the losartan groups and a decrease in inflammatory macrophages. The losartan groups supported this theory at day -22 with an increase in CD4, CD8, effector T cells, Th1 cells. However, we did not see a decrease in inflammatory

macrophages in these groups; in fact, we saw an increase in alveolar macrophages in the losartan injection group at day -22. Further, these changes in cell numbers did not affect the final outcome as these cell types did not sustain high levels after infection at days 21 and 56. By day 56, there are no differences in total numbers of predefined phenotypes between BCG and any of the other groups.

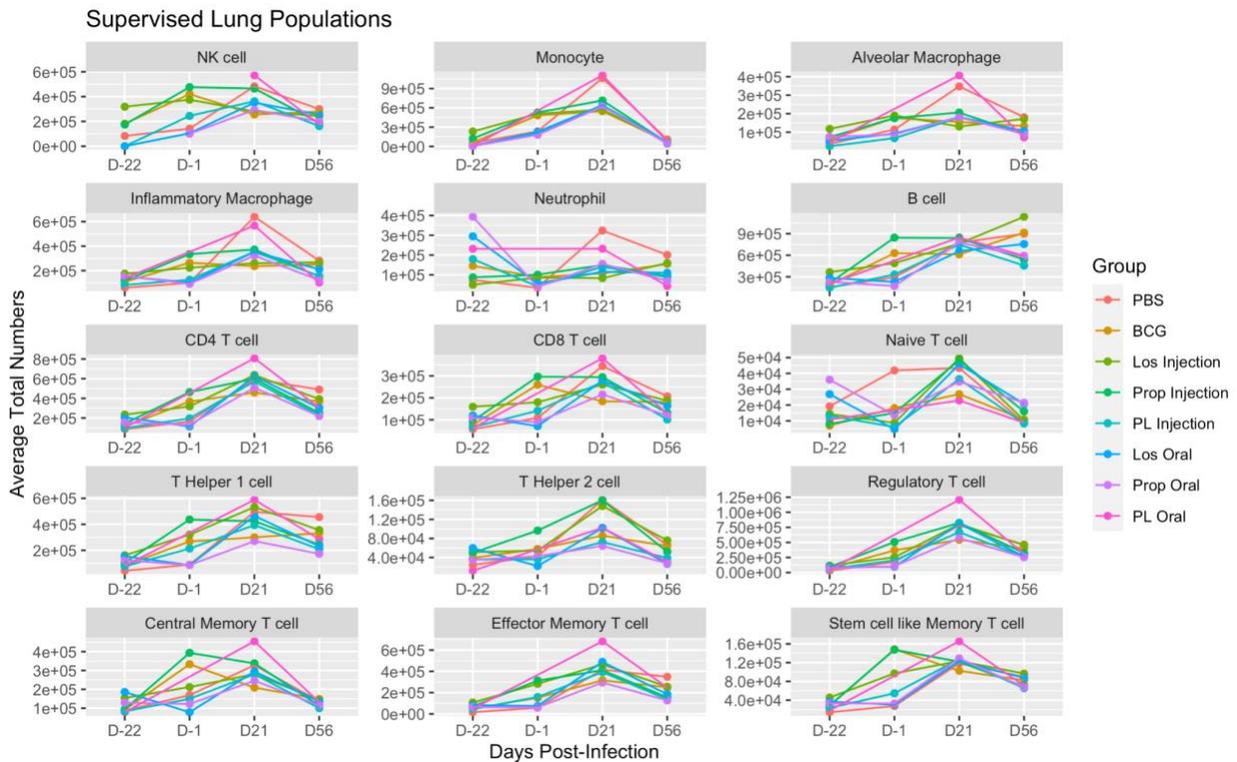


Figure 3.5: Predefined Lung Phenotypes Over Time. The lung phenotypes were assessed by identifying cells that we expected to see in our samples. The total number of cells in each of the groups over time is shown in the small plots where each small plot is a different cellular phenotype. Dunnett’s test was used to calculate statistically significant differences between the BCG group and all other groups. Due to the high number of statistically significant differences ($p < 0.05$), the differences in total numbers are shown in Appendix 2: Supplemental Table 3.2.

Immunomodulatory drugs affect the number of dendritic cells, monocytes, and CD8 T cells in the lungs as evidenced through data-driven analysis of general populations

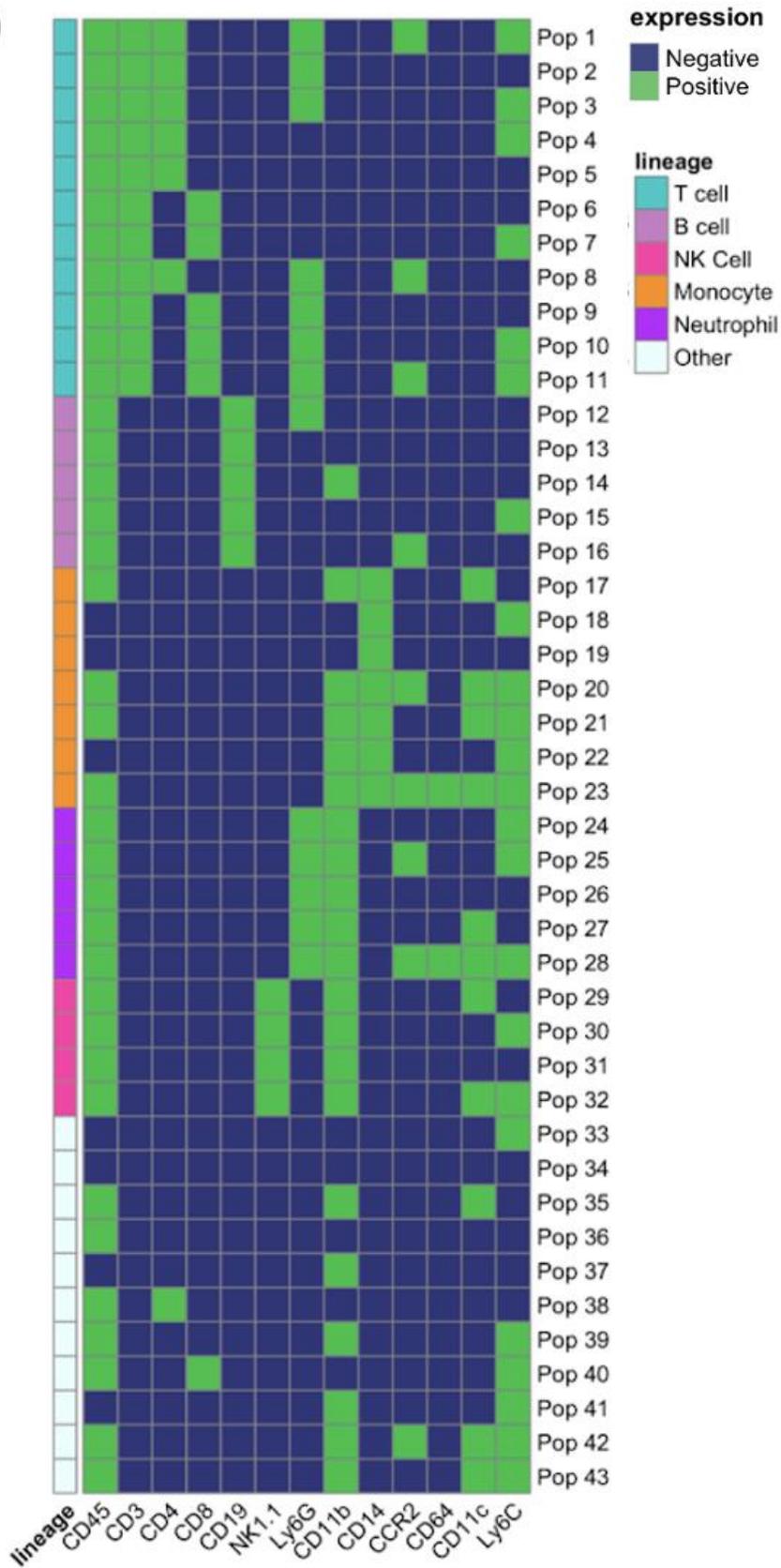
While looking at predefined flow cytometry phenotypes provides valuable information, we also wanted to use an approach to allow the data to tell us the unique phenotypes present. Data-driven analysis was performed using cyto-feature engineering which identifies all combinations of markers that form a phenotype in a sample.²² For the data-driven general lung analysis, 4,715 populations were initially identified in the sample. These populations were filtered to those for which at least one sample contains greater than 2% of the cells— reducing the number of populations to 43 (Figure 3.6a).

At day -22 in the general lung analysis, there were significantly higher numbers of CD11b+ myeloid-derived dendritic cell population (Population 43: CD45+ CD11b+ CD11c+ Ly6C+ and negative for all other markers in panel) in the losartan injection group compared to the BCG group (Figure 3.6b, Appendix 2: Supplemental Table 3.3). This same population was similarly increased at day 21 in both the PBS and oral propranolol-losartan (PL) groups compared to BCG. As described in the literature review, dendritic cells play a crucial role in linking the innate and adaptive immune response by presenting antigens to T cells and promoting differentiation of naive T cells into T Helper cells.²³ These T Helper cells in-turn create a positive feedback loop by releasing cytokines that promote the differentiation of more T Helper cells.²³ We similarly saw higher numbers of T Helper 1 cells (Th1) in the predefined lung analysis corroborating these results. These Th1 cells are capable of secreting IFN- γ which in turn activates macrophages that can phagocytose and kill MTB.²⁴ However, if differentiation of different T Helper cells such as Th17 are induced, the positive feedback loop can lead to inflammation and tissue damage.²⁴ Th17 cells, in particular, contribute to TB pathology.²⁴

Two other CD8+ Ly6G+ Ly6C+ populations (population 10: CD45+ CD3+ CD8+ Ly6G+ Ly6C+ and negative for all other markers in panel) and population 11: CD45+ CD3+ CD8+ Ly6G+ Ly6C+ CCR2+ and negative for all other markers in panel) were found to be increased in the oral groups (losartan oral and propranolol oral) at day -22, and in the PBS group at day 21 (Figure 3.6b). Research by Matsuzaki et al. has shown that Ly6G/Ly6C+ CD8+ memory T cells produce IFN- γ and exhibit cytotoxic abilities.²⁵ While these cells were initially elicited by the oral drugs at day -22, the populations decreased rapidly and were not induced to the same levels after infection.

Interestingly, a number of cell types that were higher in the losartan groups at day -22 were similarly higher in the PBS group at day 21 in comparison to the BCG control. For example, Populations 15, 20, 21, 23, and 42 were increased in the losartan injection group at day -22. These same populations were higher in the PBS groups compared to the BCG groups at day 21. Population 15 (CD45+ CD19+ Ly6C+ CD3- and negative for all other markers in panel) is likely a plasma cell capable of producing antibodies and anti-inflammatory cytokines such as IL-10 and IL-35.²⁶⁻²⁸ Populations 20, 21, 23, and 42 are monocytes/dendritic cells. While the losartan-BCG vaccinated mice had activated plasma cells, monocytes, and dendritic cells at day -22, because the PBS control group had not encountered a vaccine/infection before, we expected to see these populations only after MTB infection.

a)



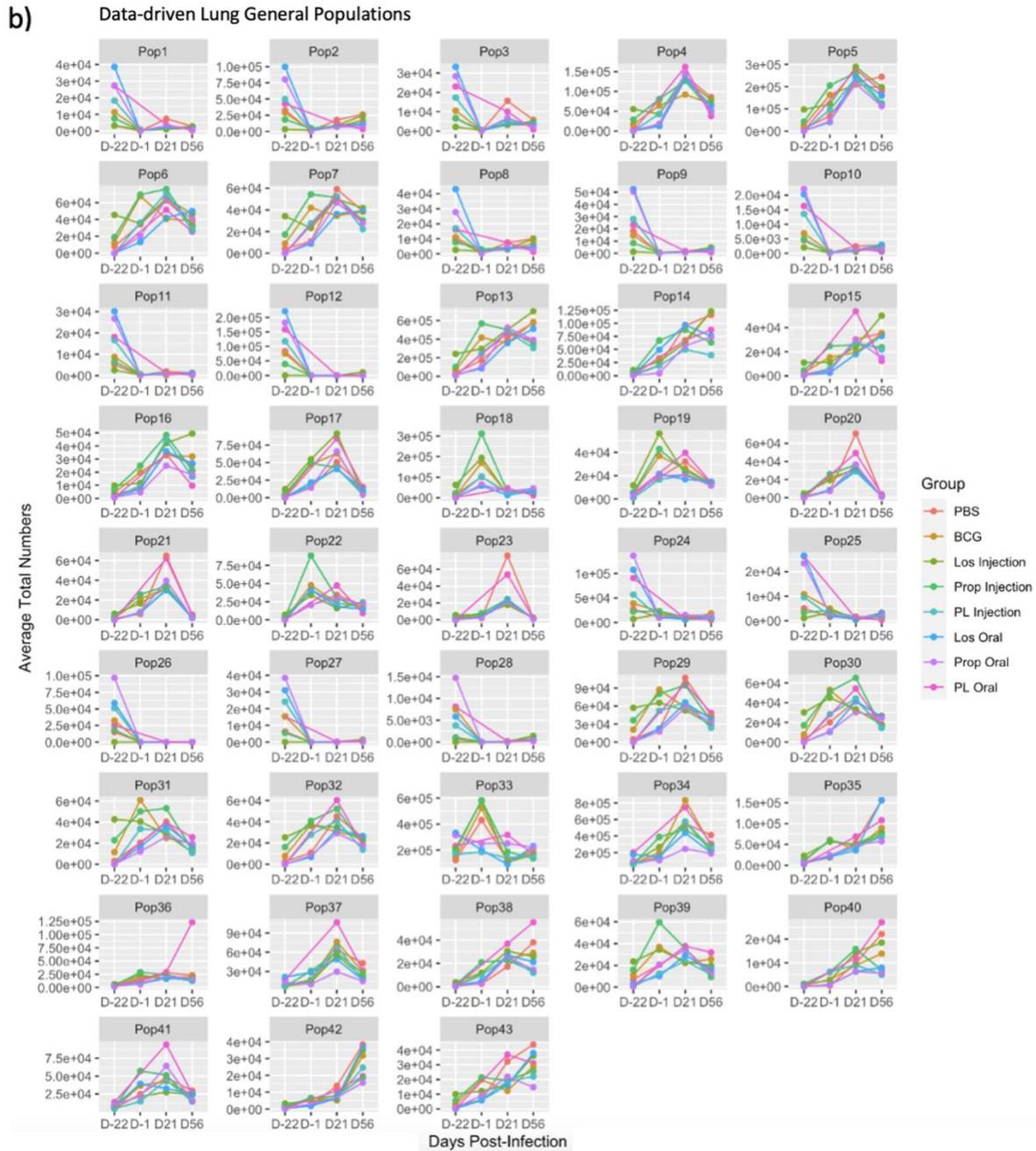


Figure 3.6: Data-driven Evaluation of General Populations in the Lungs (a) The heatmap shows the feature engineered flow cytometry lung phenotypes that make up at least 2% of a sample. Green indicates positive expression and blue indicates negative expression of each maker on the x-axis used in the analysis. (b) The total number of cells in each of the groups over time is shown in the small plots where each small plot is a different phenotype corresponding to the populations listed in (a). Dunnett’s test was used to calculate statistically significant differences between the BCG group and all other groups. Due to the high number of statistically significant differences ($p < 0.05$), the differences in total numbers are shown in Appendix 2: Supplemental Table 3.3.

Losartan and propranolol increase the number of CD8+ T cells in the lungs as shown through data-driven T cell analysis

While the general analysis allowed us to look at a variety of different cells in the lungs, we wanted to take a closer look at the specific types of T cells present. Cyto-feature engineering identified 23,159 unique populations in the T cell panel lung samples. The populations were filtered to cells that express CD3 and for which at least one sample contains said population greater than 1%. This resulted in 22 populations (Figure 3.7a). The most differences occurred at D-22 in the losartan groups (both losartan injection and losartan oral) (Figure 3.7b, Appendix 2: Supplemental Table 3.4), confirming the findings in the predefined analysis. A CD3+ CD8+ CD44+ KLRG1+ Tbet+ cell (Population 19) is particularly of interest as it was increased in both losartan groups compared to the BCG group. The presence of CD44, KLRG1, and Tbet points to the fact that this cell population is a terminally differentiated effector T cell with low memory potential.^{29,30} These results appear to show that losartan has increased the adaptive immune response to BCG. If there are terminally differentiated effector T cells present, it means that cells were activated first that then pushed the development of terminally differentiated cells. However, because this cell type was only present at elevated levels in the time point before infection, it likely did not play a role in combating infection. Population 14 (CD3+ CD8+ CD4- CD44- CD62L+ Sca1- CD103+ Tbet- and negative for all other markers in panel) is also of merit as it had higher numbers in oral groups (oral losartan and oral propranolol) at day -22, the PBS group at day -1, and the oral losartan group at day 56. Population 14 exhibits markers that are reminiscent of both naive T cells and tissue-resident cells.³¹ It is possible that we captured this population of cells as they differentiated from naive T cells to tissue-resident cells.³² These

tissue-resident memory cells serve as a first line of defense against an invading pathogen and are associated with MTB protection.^{33,34}

Propranolol decreases the numbers of B cells and T cells in the spleen

Compared to the lungs, there were not nearly as many differences in the predefined phenotypes between groups at the early time points in the spleen (Appendix 2: Supplemental Table 3.5). At day 21 post infection, however, there were more differences in cell types. Particularly in the propranolol oral group, there were decreased numbers of NK cells, monocytes, inflammatory macrophages/monocytes, B cells, CD4 T cells, CD8 T cells, T Helper 1, T Helper 2, Regulatory T cells, central memory, effector memory, and stem-cell like memory T cells in comparison to the BCG group (Figure 3.8). At this time point, there are similarly fewer total cells in the spleen in this group compared to the BCG group. However, all of the other groups that had fewer total cells did not show the same decreased cell type counts. While we expected the propranolol groups to show an increase in Th1 and CD4 T cells, these results show the opposite in the oral propranolol group. It appears as though oral propranolol dosing delays the response to MTB infection because by day 56, the Th1 and CD4 T cells had similar numbers to the BCG group. Interestingly at day 56, the groups that received propranolol injections (propranolol-losartan [PL] injection and propranolol injection) had fewer neutrophils in the spleen, potentially showing lower levels of inflammation. This is consistent with cancer literature that has shown that propranolol reduces the influx of neutrophils to primary tumors.³⁵ Similarly, at day 56, MTB has disseminated to the spleen. While it is true that we want a certain level of neutrophils to recruit other cells to the site of infection, depleting neutrophils late in infection does not affect MTB burden in the spleen.³⁶

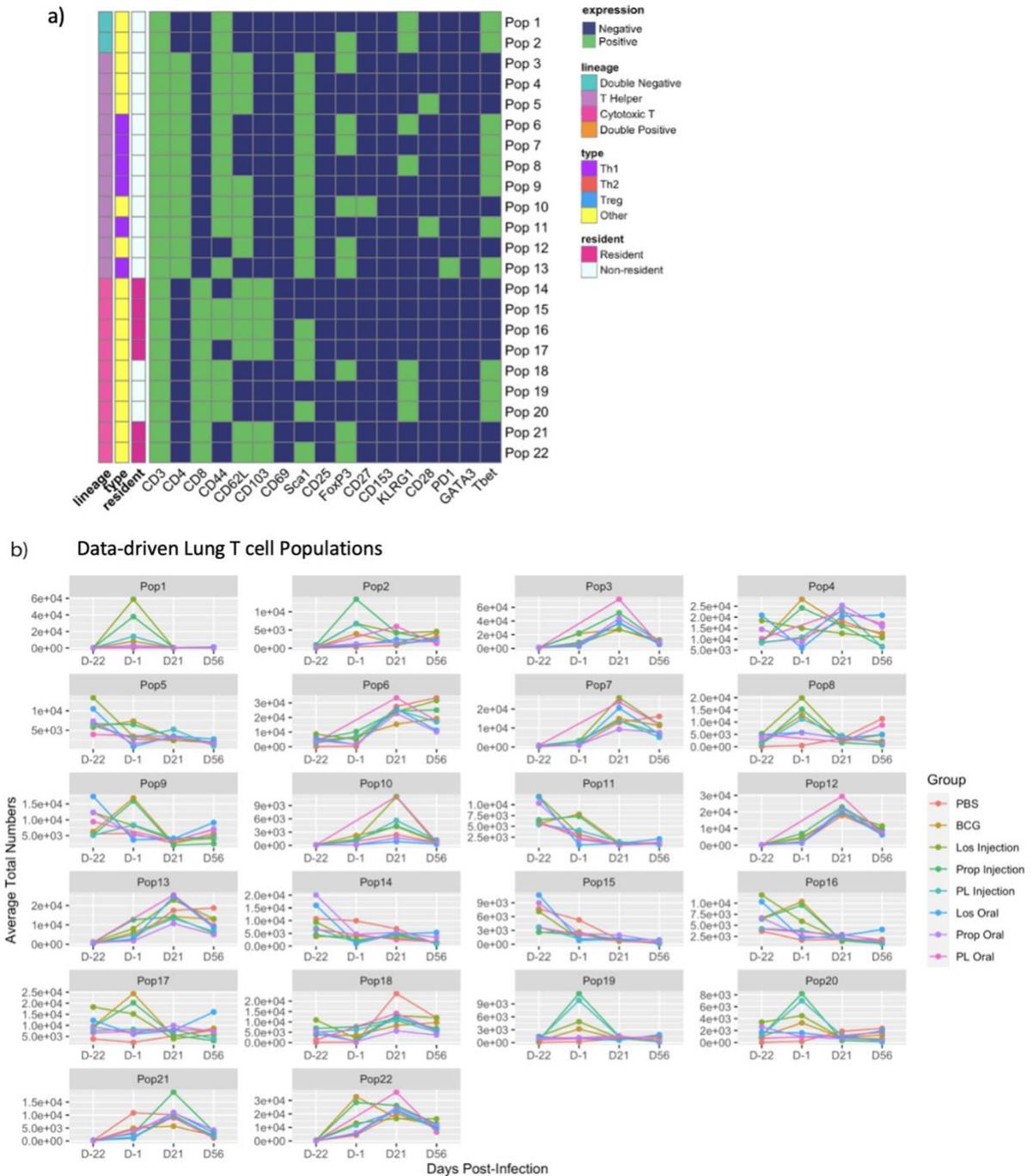


Figure 3.7: Data-driven Analysis of T cells in Lungs. (a) The heatmap shows the feature engineered flow cytometry lung phenotypes that are CD3+ and make up at least 1% of a sample. Green indicates positive expression and blue indicates negative expression of each maker on the x-axis used in the analysis. (b) The total number of cells in each of the groups over time is shown in the small plots where each small plot is a different phenotype corresponding to the populations listed in (a). Dunnett’s test was used to calculate statistically significant differences between the BCG group and all other groups. Due to the high number of statistically significant differences ($p < 0.05$), the differences in total numbers are shown in Appendix 2: Supplemental Table 3.4.

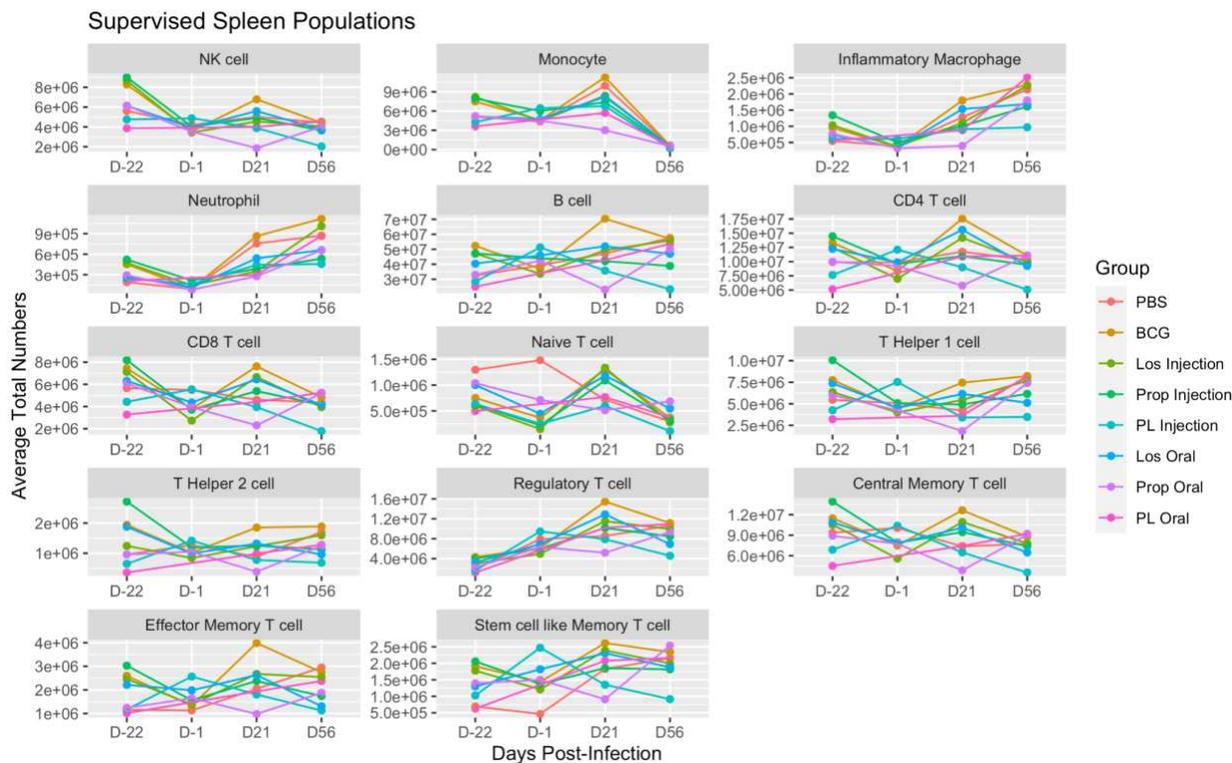


Figure 3.8: Predefined Spleen Phenotypes Over Time. The spleen phenotypes were assessed by identifying cells that we expected to see in our analysis. The total number of cells in each of the groups over time is shown in the small plots where each small plot is a different phenotype. Dunnett’s test was used to calculate statistically significant differences between the BCG group and all other groups. Due to the high number of statistically significant differences ($p < 0.05$), the differences in total numbers are shown in Appendix 2: Supplemental Table 3.5.

Losartan-Propranolol injection group exhibited lower numbers of B cells in the spleen in comparison to the BCG group as evidenced through the data-driven analysis of general populations

3,838 general spleen populations were identified via feature engineering and filtered to 23 populations that make up 1% of at least one sample (Figure 3.9a). The most differences in the spleen occurred at D21. There were significantly more CD45+ CD19+ CD11b+ cells (Population 12) in the BCG group compared to all of the drugged groups except for the oral losartan

regardless of the drug administration route at day 21 (Figure 3.9b, Appendix 2: Supplemental Table 3.6). Research has shown that the majority of B cells that express CD27 also express CD11b, making it likely that this is a memory B cell population.³⁷ Upon re-exposure to an antigen, memory B cells are able to swiftly differentiate into plasma cells that can produce antibodies.³⁸ The oral propranolol group also had 12 different populations at D21 with significantly fewer cells compared to the BCG group. By D56, there were only differences between the BCG group and the propranolol-losartan (PL) injection group with lower cell counts in various B cell populations including Population 8 which resembles activated B cells (CD45+ CD19+ CD3- and negative for all other markers in panel), Population 10, which are functional B cells (CD45+ CD19+ CD14+ CD3- and negative for all other markers in panel), Population 11, an interesting natural killer-like B cell (CD45+ CD19+ CD14+ NK1.1+ CD3- and negative for all other markers in panel), and Population 14, B cells expressing CCR2 (CD45+ CD19+ CCR2+ CD3- and negative for all other markers in panel) in the propranolol-losartan (PL) injection group. As described previously, B cells play a role in T cell differentiation through the secretion of polarizing cytokines.³⁹ B cells themselves can also present antigens to T cells; this ability has been shown to play a role in BCG immunity.⁴⁰ Another role for B cells is demonstrated through the increasing evidence that shows that MTB may be susceptible to antibodies.⁴¹ However, these differences in B cells did not impact protection as there were no differences in the bacterial burden in the mice.

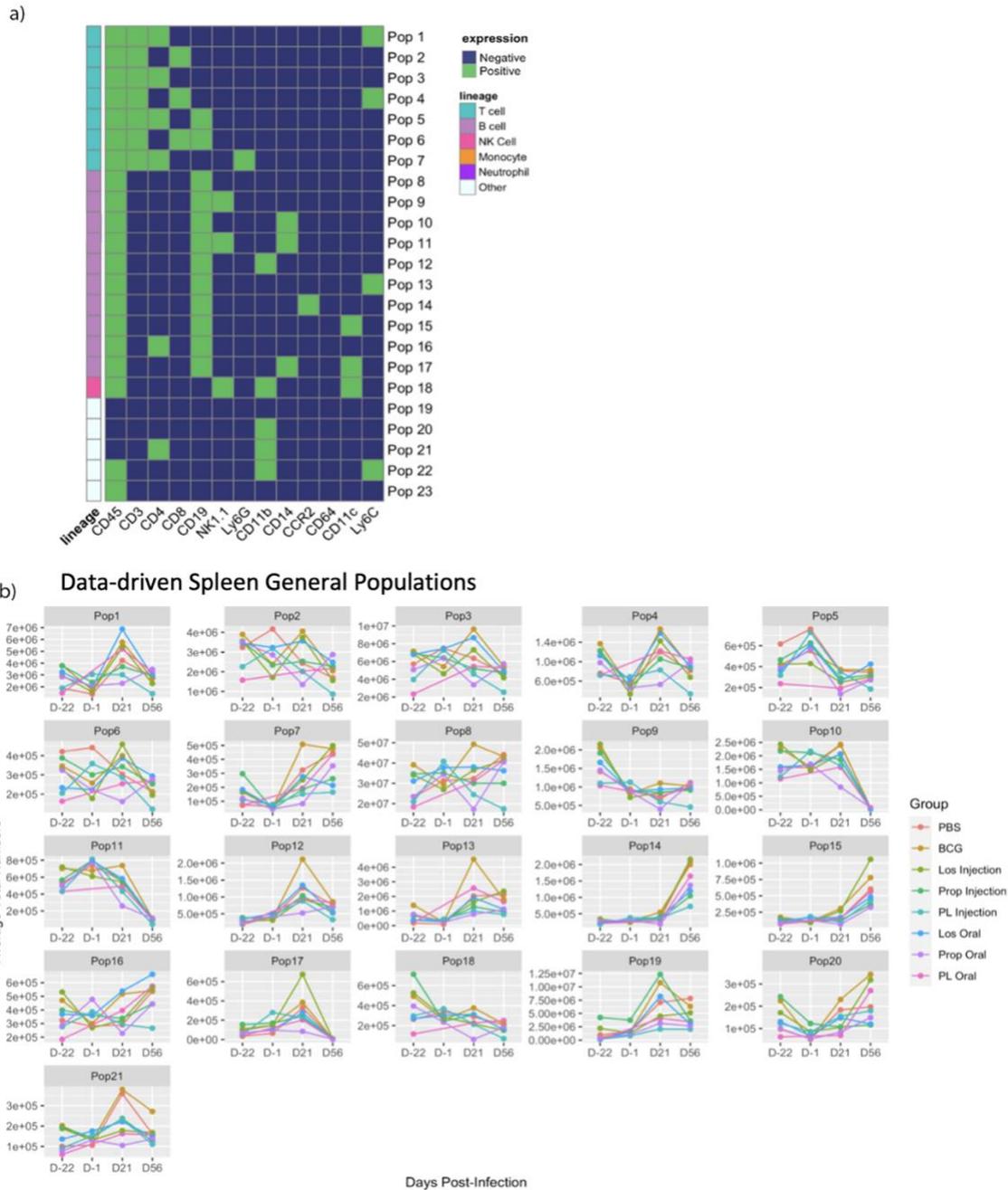


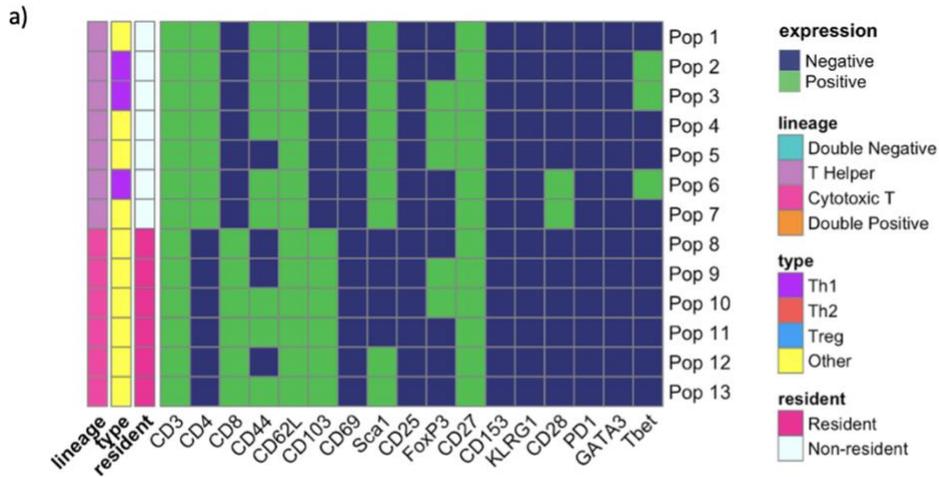
Figure 3.9: Data-driven Evaluation of General Populations in the Spleen (a) The heatmap shows the feature engineered flow cytometry lung phenotypes that make up at least 1% of a sample. Green indicates positive expression and blue indicates negative expression of each marker on the x-axis used in the analysis. (b) The total number of cells in each of the groups over time is shown in the small plots where each small plot is a different phenotype corresponding to the populations listed in (a). Dunnett’s test was used to calculate statistically significant differences between the BCG group and all other groups. Due to the high number of statistically significant differences ($p < 0.05$), the differences in total numbers are shown in Appendix 2: Supplemental Table 3.6.

Propranolol affects central memory and tissue-resident T cell numbers in the spleen as shown through data-driven evaluation of T cells

Taking a closer look at the T cells in the spleen, we filtered the 24,703 spleen T cell populations to 13 populations after removing populations that do not express CD3 or make up less than 1% of a sample (Figure 3.10a). Compared to BCG, both of the groups that received propranolol orally (oral propranolol and oral PL) had lower numbers of two central memory T cell populations characterized by CD3⁺ CD4⁺ CD44⁺ CD62L⁺ Sca1⁺ CD27⁺ CD28⁺ Tbet⁺ and negative for all other markers in panel (population 6) and CD3⁺ CD4⁺ CD44⁺ CD62L⁺ Sca1⁺ CD27⁺ CD28⁺ Tbet⁻ and negative for all other markers in panel (population 7) at day 21 (Figure 3.10, Appendix 2: Supplemental Table 3.7). Interestingly at day 56, the oral propranolol group had a higher number of a similar, but slightly different central memory T cell population (Population 2: CD3⁺ CD4⁺ CD44⁺ CD62L⁺ Sca1⁺ CD27⁺ Tbet⁺ and negative for all other markers in panel) that rapidly expanded 18 times from day 21 to day 56. Central memory T cells are often found in lymphoid tissues such as the spleen where they rapidly proliferate upon secondary exposure to antigens.⁴² Adoptive transfer of antigen-specific central memory T cells have also shown to be protective against tuberculosis disease in mice.⁴³ This suggests that the rapid expansion seen in the oral propranolol group could be protective, though no differences were shown in the CFUs at this time point.

Interestingly, at day 56, the oral propranolol group interestingly had significantly higher numbers of tissue-resident cells (Population 8: CD3⁺ CD8⁺ CD44⁻ CD62L⁺ CD103⁺ CD27⁺ and negative for all other markers in panel, Population 11: CD3⁺ CD8⁺ CD44⁺ CD62L⁺ CD103⁺ CD27⁺ and negative for all other markers in panel) in comparison to the BCG group. Similar to

Population 14 in the T cell lung data, Population 8 appears to be cells that are transitioning from a naive phenotype to a tissue-resident phenotype. Once more, Population 11 may be central memory cells that are actively differentiating into tissue-resident cells.⁴⁴ Tissue-resident cells are non-circulating cells that reside in tissue for prolonged periods of time so that they can respond upon secondary infection. Studies have shown that increases in tissue-resident memory T cells are associated with a decrease in MTB bacterial burden.^{33,45} While we did not see this in our study, it is possible had we continued the study to a later time point, we would have seen a decrease in bacterial burden in these groups.



Data-driven Spleen T cell Populations

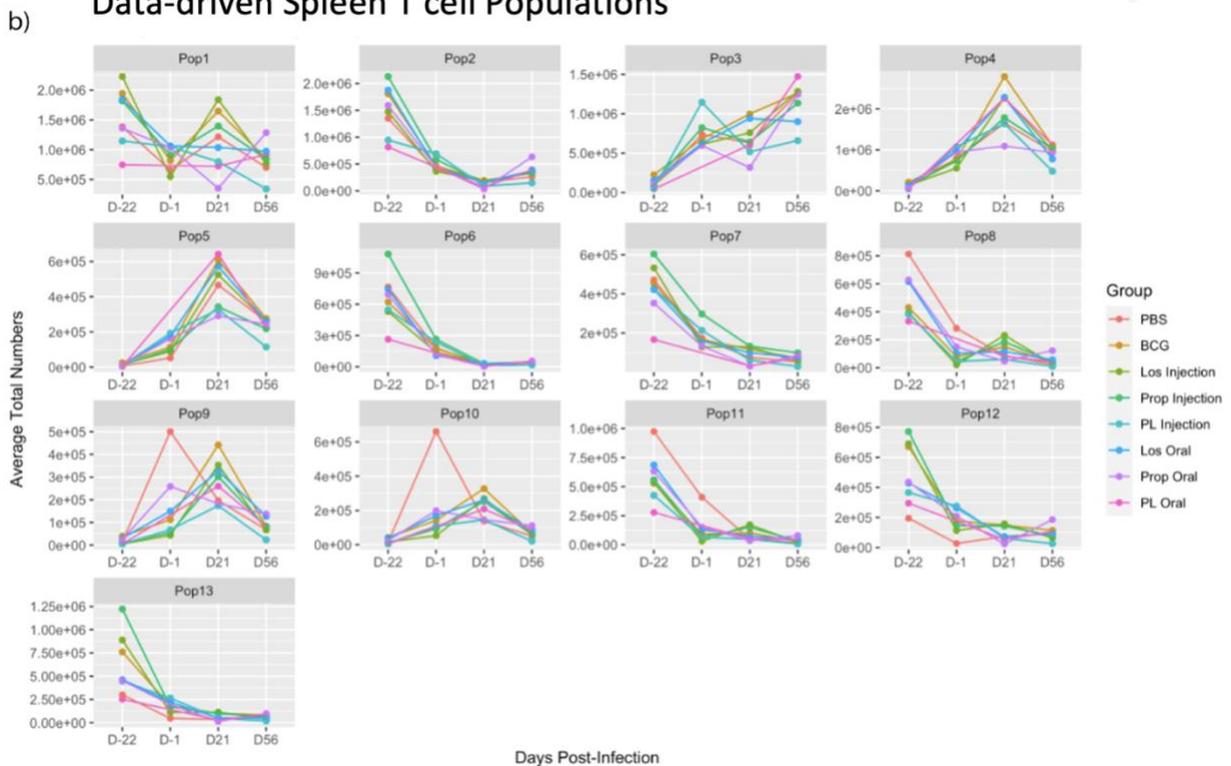


Figure 3.10: Data-driven Evaluation of T cells in the Spleen. (a) The heatmap shows the feature engineered flow cytometry lung phenotypes that are CD3+ and make up at least 1% of a sample. Green indicates positive expression and blue indicates negative expression of each marker on the x-axis used in the analysis. (b) The total number of cells in each of the groups over time is shown in the small plots where each small plot is a different phenotype corresponding to the populations listed in (a). Dunnett’s test was used to calculate statistically significant differences between the BCG group and all other groups. Due to the high number of statistically significant differences ($p < 0.05$), the differences in total numbers are shown in Appendix 2: Supplemental Table 3.7.

Discussion

As discussed previously, there is a need for a better tuberculosis vaccine. The purpose of this study was to modify the immune response to BCG—using the drugs losartan and propranolol—to induce better protection against *Mycobacterium tuberculosis*. In this study, we identified various cell types elicited in the lungs and spleens of BCG or PBS vaccinated mice treated with or without our immunomodulatory drugs and challenged with MTB.

In response to MTB infection, neutrophils and alveolar macrophages are the first line of defense.⁴⁶ These cells release cytokines to promote chemotaxis of other macrophages, dendritic cells, and neutrophils to the site of infection. Eight to twelve days after infection, dendritic cells migrate to the lymph nodes to present antigens to T cells.⁴⁶ Fourteen to seventeen days after infection, these activated T cells migrate to the lungs to fight the infection.⁴⁶ Therefore, we would expect to see high levels of T cells in the lungs at day -22, which is two–three weeks after BCG vaccination. We also observed some differences at day 21 after infection, primarily in the PBS group. This is due to the fact that this group had not been primed with a previous vaccination, and thus was generating an initial immune response to vaccination with an influx of monocytes, dendritic cells, and T cells. We did not see many differences in the immune cell populations in the lungs at day 56. This is aligned with the fact that BCG protection generally wanes by day 60.⁴⁷ In comparison to the lungs, the immune response in the spleen is generally delayed because MTB disseminates to the spleen about two weeks after infection.⁴⁸ Therefore, we saw many differences in immune cells at days 21 and 56 in the spleen soon after MTB appeared.

Looking at the cell populations that exhibited lasting differences at day 56, there were lower numbers of splenic neutrophils in the propranolol injection groups (propranolol-losartan injection and propranolol injection). This decrease in inflammation could be due to the anti-inflammatory characteristics that propranolol has shown to exhibit in human cells through quenching IL-13 and TNF- α cytokines.⁴⁹ We saw 11 differences in cell types in the spleen at day 56 between BCG and other groups across all of the analyses. Interestingly, all of these groups received propranolol (oral propranolol, injected propranolol, and injected propranolol-losartan). However, we only saw 2 differences in the lung at day 56, and both of these differences were in losartan groups (oral losartan and oral propranolol-losartan). Propranolol was able to induce more changes in the spleen while losartan had more effects in the lungs.

We expected the losartan groups to show a decrease in the influx of inflammatory macrophages which would lead to slower killing of BCG. This slower killing would allow more time for antigens to be presented to T cells, and thus an increase in T cell activation. While we saw an initial increase at day -22 in various T cell populations in the lung (CD4, CD8, effector, Th1 cells), these cells did not maintain high levels after infection. However, it is possible that these cells were initiating an activation cascade by releasing cytokines and chemokines, which we did not evaluate in our current study. This could potentially be assessed in future studies. Further, contrary to what Regan et al. demonstrated in micro-metastases studies, the use of losartan did not appear to reduce murine inflammatory macrophages/monocytes during MTB infection in this study. This could potentially be explained by the different animal models used in both studies. Regan et al. utilized BALB/c mice whereas we used C657BL/6 mice. BALB/c mice have been shown to induce a greater influx of inflammatory cells to the lungs during MTB infection than

C57BL/6 mice.⁵⁰ Therefore, it is possible that losartan was able to reduce this increased influx in BALB/c mice to a lower baseline level using losartan. We further hypothesized that propranolol would promote T cell polarization to Th1 cells and increase CD8+ T cell cytolytic abilities. We actually saw a decrease in the Th1 cells in the oral propranolol group at day 21, but by day 56, the numbers had increased to similar levels as the BCG group. Interestingly, we saw an increase in CD8+ Ly6G+ T cells in the oral propranolol group at day -22. Research by Matsuzaki et al. has shown that Ly6G+ CD8 T cells exhibit cytotoxic effects.²⁵ Therefore, it appears as though our hypothesis was partially correct, though this increase in cells was not present after infection and did not affect MTB burden.

While B cells have previously been overlooked in MTB research, there is increasing evidence that B cells play a role in protection. In this study we saw decreased levels in a variety of B cells in the PL injection group compared to the BCG control group after infection at day 56 in the spleen. Particularly of interest is the CCR2+ B cell population (Spleen-Population 14). It has been shown that CCR2 is present on immature B cells in the spleen; as the B cells mature, they lose CCR2.⁵¹ This expression of CCR2 reduces chemotaxis of immature B cells to other organs such as the lymph nodes.⁵¹ Perhaps these immature B cells would have matured at a later time point to help fight the MTB infection. Another fascinating B cell identified in this analysis was an NK-like B cell (Spleen-Population 11). NK-like B cells have recently been recognized as a subpopulation of B cells.⁵² These cells are primarily found in the spleen and mesenteric lymph nodes and produce proinflammatory IL-12 and IL-18.⁵² These secreted cytokines work together to enhance the Th1 response. While IL-12 plays a role in the differentiation of Th1 cells and can modulate the natural killer response, IL-18 promotes IFN- γ production by Th1 cells.^{53,54} Up until

day 56, the levels of NK-like B cells were relatively high. If we had been able to sustain higher numbers of NK-like B cells at day 56, maybe we would have seen increases in Th1 cell numbers and better protection from MTB.

Throughout the data-driven flow analysis, we identified cell populations that exhibited cell markers of multiple known T cell types (e.g., Population 8 in the spleen that expressed both naive T cell markers and tissue-resident markers). These cells are typically difficult to identify using traditional manual gating techniques. It is possible that at our experimental time points we took a snapshot of the cell populations as they differentiated from one cell type to another.

One limitation of the study was that we had a very small sample size per group ($n=3$). This significantly reduced the power of all of our statistical analysis. Further, we could have extended the timeline of the study to include a time point around day 75. In our analysis, we saw a rapid increase in a central memory population in the oral propranolol group from day 21 to day 56. In the same group, we saw cells that were potentially differentiating from naive T cells into tissue-resident cells. Both of these cell types may play a role in protection, but we may have captured the data too early. Another limitation of our study is that we did not include groups that received Propranolol and Losartan without BCG vaccination or without infection to understand the effect of the individual drugs on the immune response. However, adding these groups would have greatly increased the number of mice beyond the feasibility of this study.

Despite the differences that we saw in activated T cells, macrophages, and neutrophils before infection in groups that received the drugs, there were no differences in the bacterial burden of

the animals after day 21. Therefore, it does not appear as if losartan and propranolol produced a lasting effect on the immune system. This could be due to dosage, administration of drugs, or timing. Additional studies would need to be performed to determine if the use of these drugs can aid in tuberculosis protection.

REFERENCES

1. World Health Organization. (October 2020). Global Tuberculosis Report 2020.
2. World Health Organization. BCG Vaccine.
3. Mitchell, L.A., Henderson, A.J., and Dow, S.W. (2012). Suppression of Vaccine Immunity by Inflammatory Monocytes. *Journal of Immunology* *189*, 5612-5621. 10.4049/jimmunol.1202151.
4. Moliva, J.I., Turner, J., and Torrelles, J.B. (2017). Immune Responses to Bacillus Calmette-Guerin Vaccination: Why Do They Fail to Protect against Mycobacterium tuberculosis? *Frontiers in immunology* *8*, 407-407. 10.3389/fimmu.2017.00407.
5. Regan, D.P., Coy, J.W., Chahal, K.K., Chow, L., Kurihara, J.N., Guth, A.M., Kufareva, I., and Dow, S.W. (2019). The Angiotensin Receptor Blocker Losartan Suppresses Growth of Pulmonary Metastases via AT1R-Independent Inhibition of CCR2 Signaling and Monocyte Recruitment. *Journal of Immunology* *202*, 3087-3102. 10.4049/jimmunol.1800619.
6. Qian, B.Z., Li, J.F., Zhang, H., Kitamura, T., Zhang, J.H., Campion, L.R., Kaiser, E.A., Snyder, L.A., and Pollard, J.W. (2011). CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* *475*, 222-U129. 10.1038/nature10138.
7. Merino, A., Alvarez-Lara, M.A., Ramirez, R., Carracedo, J., Martin-Malo, A., and Aljama, P. (2012). Losartan prevents the development of the pro-inflammatory monocytes CD14(+)CD16(+) in haemodialysis patients. *Nephrology Dialysis Transplantation* *27*, 2907-2912. 10.1093/ndt/gfr767.
8. Wu, J.Y., Lee, M.T.G., Lee, S.H., Tsai, Y.W., Hsu, S.C., Chang, S.S., and Lee, C.C. (2016). Angiotensin-Converting Enzyme Inhibitors and Active Tuberculosis: A Population-Based Study. *Medicine* *95*, e3579. 10.1097/md.0000000000003579.
9. Mortensen, E.M., Nakashima, B., Cornell, J., Copeland, L.A., Pugh, M.J., Anzueto, A., Good, C., Restrepo, M.I., Downs, J.R., Frei, C.R., and Fine, M.J. (2012). Population-Based Study of Statins, Angiotensin II Receptor Blockers, and Angiotensin-Converting Enzyme Inhibitors on Pneumonia-Related Outcomes. *Clinical Infectious Diseases* *55*, 1466-1473. 10.1093/cid/cis733.
10. Gullestad, L., Aukrust, P., Ueland, T., Espevik, T., Yee, G., Vagelos, R., Froland, S.S., and Fowler, M. (1999). Effect of high- versus low-dose angiotensin converting enzyme inhibition on cytokine levels in chronic heart failure. *Journal of the American College of Cardiology* *34*, 2061-2067. 10.1016/s0735-1097(99)00495-7.
11. Antunes, G., Evans, S.A., Lordan, J.L., and Frew, A.J. (2002). Systemic cytokine levels in community-acquired pneumonia and their association with disease severity. *European Respiratory Journal* *20*, 990-995. 10.1183/09031936.02.00295102.
12. Scanzano, A., and Cosentino, M. (2015). Adrenergic regulation of innate immunity: a review. *Frontiers in Pharmacology* *6*, 171. 10.3389/fphar.2015.00171.
13. Emeny, R.T., Gao, D., and Lawrence, D.A. (2007). beta 1-adrenergic receptors on immune cells impair innate defenses against Listeria. *Journal of Immunology* *178*, 4876-4884. 10.4049/jimmunol.178.8.4876.
14. PaninaBordignon, P., Mazzeo, D., DiLucia, P., Dambrosio, D., Lang, R., Fabbri, L., Self, C., and Sinigaglia, F. (1997). beta(2)-agonists prevent Th1 development by selective

- inhibition of interleukin 12. *Journal of Clinical Investigation* *100*, 1513-1519. 10.1172/jci119674.
15. Mohammadpour, H., MacDonald, C.R., Qiao, G., Chen, M., Dong, B., Hylander, B.L., McCarthy, P.L., Abrams, S.I., and Repasky, E.A. (2020). B2 adrenergic receptor-mediated signaling regulates the immunosuppressive potential of myeloid-derived suppressor cells. *Cancer Research* *80*. 10.1158/1538-7445.sabcs19-p3-01-02.
 16. Ashrafi, S., Shapouri, R., Shirkhani, A., and Mahdavi, M. (2018). Anti-tumor effects of propranolol: Adjuvant activity on a transplanted murine breast cancer model. *Biomedicine & Pharmacotherapy* *104*, 45-51. 10.1016/j.biopha.2018.05.002.
 17. Seiffert, K., Hosoi, J., Torii, H., Ozawa, H., Ding, W.H., Campton, K., Wagner, J.A., and Granstein, R.D. (2002). Catecholamines inhibit the antigen-presenting capability of epidermal Langerhans cells. *Journal of Immunology* *168*, 6128-6135. 10.4049/jimmunol.168.12.6128.
 18. Estrada, L.D., Agac, D., and Farrar, J.D. (2016). Sympathetic neural signaling via the beta 2-adrenergic receptor suppresses T-cell receptor-mediated human and mouse CD8(+) T-cell effector function. *European Journal of Immunology* *46*, 1948-1958. 10.1002/eji.201646395.
 19. Bucsek, M.J., Qiao, G.X., MacDonald, C.R., Giridharan, T., Evans, L., Niedzwecki, B., Liu, H.C., Kokolus, K.M., Eng, J.W.L., Messmer, M.N., et al. (2017). beta-Adrenergic Signaling in Mice Housed at Standard Temperatures Suppresses an Effector Phenotype in CD8(+) T Cells and Undermines Checkpoint Inhibitor Therapy. *Cancer Research* *77*, 5639-5651. 10.1158/0008-5472.can-17-0546.
 20. Ruddy, M.C., and Kostis, J.B. (2005). Angiotensin II Receptor Antagonists.
 21. Fox, A., Dutt, T.S., Karger, B., Obregon-Henao, A., Anderson, G.B., and Henao-Tamayo, M. (2020). Acquisition of High-Quality Spectral Flow Cytometry Data. *Current protocols in cytometry* *93*, e74-e74. 10.1002/cpcy.74.
 22. Fox, A., Dutt, T.S., Karger, B., Rojas, M., Obregon-Henao, A., Anderson, G.B., and Henao-Tamayo, M. (2020). Cyto-Feature Engineering: A Pipeline for Flow Cytometry Analysis to Uncover Immune Populations and Associations with Disease. *Scientific Reports* *10*. 10.1038/s41598-020-64516-0.
 23. Kim, B., and Kim, T.H. (2018). Fundamental role of dendritic cells in inducing Th2 responses. *Korean Journal of Internal Medicine* *33*, 483-489. 10.3904/kjim.2016.227.
 24. Lyadova, I.V., and Panteleev, A.V. (2015). Th1 and Th17 Cells in Tuberculosis: Protection, Pathology, and Biomarkers. *Mediators of Inflammation* *2015*, 854507. 10.1155/2015/854507.
 25. Matsuzaki, J., Tsuji, T., Chamoto, K., Takeshima, T., Sendo, F., and Nishimura, T. (2003). Successful elimination of memory-type CD8(+) T cell subsets by the administration of anti-Gr-1 monoclonal antibody in vivo. *Cellular Immunology* *224*, 98-105. 10.1016/j.cellimm.2003.08.009.
 26. Khodadadi, L., Cheng, Q., Radbruch, A., and Hiepe, F. (2019). The Maintenance of Memory Plasma Cells. *Frontiers in Immunology* *10*, 721. 10.3389/fimmu.2019.00721.
 27. Rao, M., Valentini, D., Poiret, T., Doodoo, E., Parida, S., Zumla, A., Brighenti, S., and Maeurer, M. (2015). B in TB: B Cells as Mediators of Clinically Relevant Immune Responses in Tuberculosis. *Clinical Infectious Diseases* *61*, S225-S234. 10.1093/cid/civ614.

28. Dang, V.D., Hilgenberg, E., Ries, S., Shen, P., and Fillatreau, S. (2014). From the regulatory functions of B cells to the identification of cytokine-producing plasma cell subsets. *Current Opinion in Immunology* 28, 77-83. 10.1016/j.coi.2014.02.009.
29. Lazarevic, V., Glimcher, L.H., and Lord, G.M. (2013). T-bet: a bridge between innate and adaptive immunity. *Nature Reviews Immunology* 13, 777-789. 10.1038/nri3536.
30. Samji, T., and Khanna, K.M. (2017). Understanding memory CD8(+) T cells. *Immunology Letters* 185, 32-39. 10.1016/j.imlet.2017.02.012.
31. Corgnac, S., Boutet, M., Kfoury, M., Naltet, C., and Mami-Chouaib, F. (2018). The Emerging Role of CD8(+) Tissue Resident Memory T (T-RM) Cells in Antitumor Immunity: A Unique Functional Contribution of the CD103 Integrin. *Frontiers in Immunology* 9, 1904. 10.3389/fimmu.2018.01904.
32. Mackay, L.K., Rahimpour, A., Ma, J.Z., Collins, N., Stock, A.T., Hafon, M.L., Vega-Ramos, J., Lauzurica, P., Mueller, S.N., Stefanovic, T., et al. (2013). The developmental pathway for CD103(+)CD8(+) tissue-resident memory T cells of skin. *Nature Immunology* 14, 1294-+. 10.1038/ni.2744.
33. Copland, A., Diogo, G.R., Hart, P., Harris, S., Tran, A.C., Paul, M.J., Singh, M., Cutting, S.M., and Reljic, R. (2018). Mucosal Delivery of Fusion Proteins with *Bacillus subtilis* Spores Enhances Protection against Tuberculosis by *Bacillus Calmette-Guerin*. *Frontiers in Immunology* 9, 346. 10.3389/fimmu.2018.00346.
34. Hart, P., Copland, A., Diogo, G.R., Harris, S., Spallek, R., Oehlmann, W., Singh, M., Basile, J., Rottenberg, M., Paul, M.J., and Reljic, R. (2018). Nanoparticle-Fusion Protein Complexes Protect against *Mycobacterium tuberculosis* Infection. *Molecular Therapy* 26, 822-833. 10.1016/j.ymthe.2017.12.016.
35. Wrobel, L.J., Bod, L., Lengagne, R., Kato, M., Prevost-Blondel, A., and Le Gal, F.A. (2016). Propranolol induces a favourable shift of anti-tumor immunity in a murine spontaneous model of melanoma. *Oncotarget* 7, 77825-77837. 10.18632/oncotarget.12833.
36. Pedrosa, J., Saunders, B.M., Appelberg, R., Orme, I.M., Silva, M.T., and Cooper, A.M. (2000). Neutrophils play a protective nonphagocytic role in systemic *Mycobacterium tuberculosis* infection of mice. *Infection and Immunity* 68, 577-583. 10.1128/iai.68.2.577-583.2000.
37. Kawai, K., Tsuno, N.H., Matsushashi, M., Kitayama, J., Osada, T., Yamada, J., Tsuchiya, T., Yoneyama, S., Watanabe, T., Takahashi, K., and Nagawa, H. (2005). CD11b-mediated migratory property of peripheral blood B cells. *Journal of Allergy and Clinical Immunology* 116, 192-197. 10.1016/j.jaci.2005.03.021.
38. Lyashchenko, K.P., Vordermeier, H.M., and Waters, W.R. (2020). Memory B cells and tuberculosis. *Veterinary Immunology and Immunopathology* 221, 110016. 10.1016/j.vetimm.2020.110016.
39. Harris, D.P., Haynes, L., Sayles, P.C., Duso, D.K., Eaton, S.M., Lepak, N.M., Johnson, L.L., Swain, S.L., and Lund, F.E. (2000). Reciprocal regulation of polarized cytokine production by effector B and T cells. *Nature Immunology* 1, 475-482. 10.1038/82717.
40. Rijnink, W.F., Ottenhoff, T.H.M., and Joosten, S.A. (2021). B-Cells and Antibodies as Contributors to Effector Immune Responses in Tuberculosis. *Frontiers in Immunology* 12, 640168. 10.3389/fimmu.2021.640168.

41. Achkar, J.M., Chan, J., and Casadevall, A. (2015). Role of B Cells and Antibodies in Acquired Immunity against Mycobacterium tuberculosis. *Cold Spring Harbor Perspectives in Medicine* 5, a018432. 10.1101/cshperspect.a018432.
42. Orme, I.M., and Henao-Tamayo, M.I. (2018). Trying to See the Forest through the Trees: Deciphering the Nature of Memory immunity to Mycobacterium tuberculosis. *Frontiers in Immunology* 9, 461. 10.3389/fimmu.2018.00461.
43. Vogelzang, A., Perdomo, C., Zedler, U., Kuhlmann, S., Hurwitz, R., Gengenbacher, M., and Kaufmann, S.H.E. (2014). Central Memory CD4(+) T Cells Are Responsible for the Recombinant Bacillus Calmette-Gu, rin Delta ureC::hly Vaccine's Superior Protection Against Tuberculosis. *Journal of Infectious Diseases* 210, 1928-1937. 10.1093/infdis/jiu347.
44. Enamorado, M., Khouili, S.C., Iborra, S., and Sancho, D. (2018). Genealogy, Dendritic Cell Priming, and Differentiation of Tissue-resident Memory CD8(+) T Cells. *Frontiers in Immunology* 9, 1751. 10.3389/fimmu.2018.01751.
45. Perdomo, C., Zedler, U., Kuehl, A.A., Lozza, L., Saikali, P., Sander, L.E., Vogelzang, A., Kaufmann, S.H.E., and Kupz, A. (2016). Mucosal BCG Vaccination Induces Protective Lung-Resident Memory T Cell Populations against Tuberculosis. *Mbio* 7, e01686-16. 10.1128/mBio.01686-16.
46. O'Garra, A., Redford, P.S., McNab, F.W., Bloom, C.I., Wilkinson, R.J., and Berry, M.P.R. (2013). The Immune Response in Tuberculosis. In *Annual Review of Immunology*, Vol 31, D.R. Littman, and W.M. Yokoyama, eds. (Annual Reviews), pp. 475-527. 10.1146/annurev-immunol-032712-095939.
47. Ordway, D.J., Shang, S.B., Henao-Tamayo, M., Obregon-Henao, A., Nold, L., Caraway, M., Shanley, C.A., Basaraba, R.J., Duncan, C.G., and Orme, I.M. (2011). Mycobacterium bovis BCG-Mediated Protection against W-Beijing Strains of Mycobacterium tuberculosis Is Diminished Concomitant with the Emergence of Regulatory T Cells. *Clinical and Vaccine Immunology* 18, 1527-1535. 10.1128/cvi.05127-11.
48. Chackerian, A.A., Alt, J.M., Perera, T.V., Dascher, C.C., and Behar, S.M. (2002). Dissemination of Mycobacterium tuberculosis is influenced by host factors and precedes the initiation of T-cell immunity. *Infection and Immunity* 70, 4501-4509. 10.1128/iai.70.8.4501-4509.2002.
49. Hajighasemi, F., and Mirshafiey, A. (2016). Regulation of inflammatory cytokines in human immunocompetent cells by propranolol in vitro. *European Respiratory Journal* 48. 10.1183/13993003.congress-2016.PA1083.
50. Adam, L., Lopez-Gonzalez, M., Bjork, A., Palsson, S., Poux, C., Wahren-Herlenius, M., Fernandez, C., and Spetz, A.L. (2018). Early Resistance of Non-virulent Mycobacterial Infection in C57BL/6 Mice Is Associated With Rapid Up-Regulation of Antimicrobial Cathelicidin Camp. *Frontiers in Immunology* 9, 1939. 10.3389/fimmu.2018.01939.
51. Flaishon, L., Becker-Herman, S., Hart, G., Levo, Y., Kuziel, W.A., and Shachar, I. (2004). Expression of the chemokine receptor CCR2 on immature B cells negatively regulates their cytoskeletal rearrangement and migration. *Blood* 104, 933-941. 10.1182/blood-2003-11-4013.
52. Wang, S., Xia, P.Y., Chen, Y., Huang, G.L., Xiong, Z., Liu, J., Li, C., Ye, B.Q., Du, Y., and Fan, Z.S. (2016). Natural Killer-like B Cells Prime Innate Lymphocytes against Microbial Infection. *Immunity* 45, 131-144. 10.1016/j.immuni.2016.06.019.

53. Trinchieri, G. (2003). Interleukin-12 and the regulation of innate resistance and adaptive immunity. *Nature Reviews Immunology* 3, 133-146. 10.1038/nri1001.
54. Nakanishi, K. (2018). Unique Action of Interleukin-18 on T Cells and Other immune Cells. *Frontiers in Immunology* 9, 763. 10.3389/fimmu.2018.00763.

CHAPTER 4 — NOVEL TECHNIQUE TO STUDY IMMUNOMETABOLISM DURING MYCOBACTERIUM TUBERCULOSIS INFECTION

Introduction

Despite over 100 years of research, there is still a mystery surrounding the key immune cell players in tuberculosis infection and vaccination.^{1,2} Different research points to both protective and non-protective qualities of most cells. Thus, it is not understood which immune cells need to be elicited to develop a better vaccine. One approach to a more complete understanding of the immune system is through immunometabolism. Immunometabolism is a relatively new field of research, emerging in the last two decades. The term immunometabolism refers to the study of metabolic pathways that play a role in immune cell populations. Various immune cells require metabolites to grow, proliferate, and activate.³ As such, immune cells and metabolites are intimately involved.

One way to study immunometabolism and the immune system is through comparing different data observed over the course of an experiment. However, while these experiments are critical to understanding the science, they can be both expensive and time consuming. Another pathway to gain scientific knowledge is through making better use of data that has already been collected. Large databases containing terabytes of experimental data exist.⁴ Much of this data is publicly available, making it accessible and affordable for hypothesis generation. There has also been a push from the National Institutes of Health to gain new immunological insights from existing datasets.⁵ Thus, it is important to extract more knowledge and ensure better use of difficult-to-collect and costly data.

Two tools used to study immunometabolism are flow cytometry and liquid chromatography-mass spectrometry (LC-MS). Flow cytometry is a tool used to quantify immune cells present in the organs, tissues, or bodily fluids of animals. Traditionally, flow cytometry data has been analyzed with manual gating on biaxial plots, taking considerable time and becoming increasingly difficult with higher numbers of recorded marker parameters. With the invention of unsupervised methods such as FlowSOM and cyto-feature engineering, extensive work has been performed on developing tools to efficiently analyze large flow cytometry datasets.^{6,7} On the other hand, LC-MS and LC-MS/MS are used to study immunometabolism and identify molecular compounds and metabolites within samples. These tools can be used to explore chemical compounds, drug and toxin metabolites, and metabolic pathways. The traditional way to analyze LC-MS data is to pre-process the data using packages such as 'xcms' to identify peaks, and then perform a series of t-tests to compare the fold change of metabolic density between two experimental groups such as control and treated.⁸⁻¹⁰ While this method is beneficial in identifying key molecular differences between groups, it does not account for measurable dependent variables and confounding factors. While there has been extensive research on analyzing flow cytometry and metabolomics data separately, little research has been done integrating the two and comparing flow cytometry immune cell populations with metabolomics.

Flow cytometry and metabolomics techniques can be used to understand potential relationships between immune cells and metabolic pathways. Though we may not identify exact metabolites of interest, metabolites and biochemicals serve as tags for these metabolic pathways and can capture information on whether these pathways are upregulated or downregulated. There is clear

evidence that metabolic pathways produce metabolites and amino acids (precursors to metabolites) that are required for T cell activation and proliferation. Specifically, research has shown that activated T cells utilize glucose and glutamine through a pathway called glycolysis, as well as NADPH generated through the pentose phosphate pathway, to proliferate.¹¹ Further, amino acid synthesis pathways utilize small molecules produced in both glycolysis and the pentose phosphate pathway to produce amino acids necessary for activated T cells to function.^{12,13} For example, CD8+ T cell responses diminish when there is a low abundance of the amino acid methionine.¹⁴ Another amino acid, alanine, is also required to activate naive and memory T cells.¹³ To measure the activation of these cells, SLA-DQ, a class II swine leukocyte antigen found on T cells and antigen presenting cells, can be used.¹⁵⁻¹⁸ By identifying relationships between immune cells and metabolites, these cells and metabolites can be targeted for drugs and vaccines.

In this paper, we introduce a novel technique for correlating immune cell populations with metabolomics data in the context of a tuberculosis vaccine study from 2015. The technique exhibits the utility of re-analyzing old datasets, though it can also be used on new data sources that utilize LC-MS or LC-MS/MS data and have additional numerical dependent variables such as flow cytometry or gene expression data. We developed two hypotheses: 1) we would identify metabolic profiles correlated with immune cells with this novel methodology 2) we would see an association between activated T cells expressing SLA-DQ and glycolysis, the pentose phosphate pathways, and amino acid synthesis pathways. We expected to see these differences in the vaccinated group and after infection because of the increased T cell activation due to the

vaccination and infection. Regardless of the initial study outcome, the technique can be used to gain knowledge and generate hypotheses using dependent variables.

Methods

Experimental Setup. Neonatal minipigs were either left unvaccinated or vaccinated with BCG; they were subsequently infected with *Mycobacterium tuberculosis* (Figure 4.1). Blood was collected at two time points, pre-infection and post-infection, for subsequent flow cytometry and metabolomics analysis.

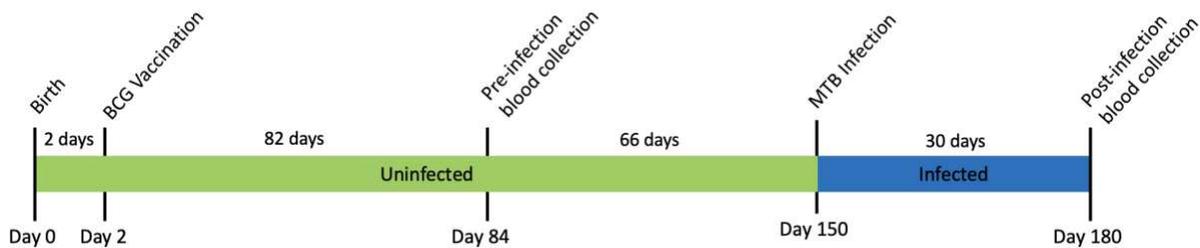


Figure 4.1: Experimental timeline. Minipigs were either vaccinated with BCG (Bacille Calmette-Guérin) or left unvaccinated two days after birth. Blood was collected approximately 12 weeks later at a pre-infection time point. The minipigs were subsequently infected with *Mycobacterium tuberculosis* (MTB). Finally, blood from a post-infection time point was collected 30 days after infection.

Animals. Two pregnant Sinclair minipigs were purchased from Sinclair Bio-Resources 2-3 weeks before expected farrowing dates. Ten healthy neonatal piglets were used for the study as described in Ramos et al 2019.¹⁷ All experimental protocols were approved by the Institutional Animal Care and Use Committee at Colorado State University.

Vaccinations. Five minipigs randomly distributed across piglets from the two mothers were intradermally vaccinated at the base of the tail 48-hours after birth with 0.05ml Bacille Calmette-

Guérin (Statens Serum Institute, Copenhagen, Denmark). Five piglets remained unvaccinated to serve as controls. This early vaccination mimics human neonatal vaccination timelines.

Mycobacterium tuberculosis infection. Five months post-birth, the minipigs were inoculated with an aerosolized target dose of 25 CFU of the virulent W-Beijing *Mycobacterium tuberculosis* HN878 as described in Ramos et al. 2017.¹⁹

Blood collection. Blood was collected from minipigs at two time points, once before infection at 84 days, and again post-infection at 180 days. The pre-infection time point at day 84, was chosen for two reasons. First, at this age minipigs are considered adolescents and thus the immune response is more mature. Further, it occurs about 12 weeks after BCG vaccination which is around the time when BCG-specific T cell levels peak in humans (around 10 weeks post-vaccination).²⁰ The post-infection time point at day 30 was chosen because BCG protection is generally observed 30 days following infection in mice, though this protection wanes over time.²¹ Blood samples were taken so that the animals did not need to be sacrificed at each time point. The blood samples were separated into peripheral blood mononuclear cells (PBMCs) for flow cytometry analysis and serum for metabolomics analysis.

PBMC Preparation. Approximately 2ml of blood was collected into heparin tubes from each animal. Separation of peripheral blood mononuclear cells (PBMC) was performed using Lympholyte[®] Cell-Separation medium (Cedarlane) following manufacturer's recommendations. PBMC were then placed in RPMI-1640 medium with supplements and 10% FBS (cRPMI) in 96

well plates at 10^6 cells/ml. The PBMC were incubated for 18 hours at 37°C in 5% CO_2 in cRPMI before use in flow cytometry analysis.

Flow Cytometry and Analysis. Flow cytometry was used to assess the specific immune cells in the PBMCs. The panel included 8 surface markers which can be used to characterize the specific immune cells: CD3, CD4, CD8, CCR7, SLADQ, CD172, CD45RA, live-dead stain (Fixable Viability Stain 510). Staining was performed as described in Ramos et al. 2019. Two time points were assessed in this analysis: 12 weeks post-birth and 30 days post infection. Cells were first gated on single live leukocytes and then analyzed with a cyto-feature engineering pipeline.⁷ CCR7 was removed from analysis because it was not difficult to determine the cutoff between the positive and negative populations from the data. The pipeline analysis resulted in a data frame where each row is a different live cell from a sample characterized by the combination of positive and negative expression of each marker denoted as either a 1 (positive) or 0 (negative).

Metabolomics and Analysis.

Samples for metabolomics were prepared following previous publications with minor modification.²²⁻²⁴ Briefly, serum samples collected from minipigs were thawed on ice and metabolites were extracted from an aliquot of serum using LC-MS grade methanol (final concentration 75%, v/v) containing internal standards (phenylalanine d5 175 ng/mL, 1-methyl tryptophan 37 ng/mL and arachidonoyl amide 30 ng/mL) for protein precipitation. Samples were incubated at -20°C for one hour, brought back to room temperature, and vortexed briefly. The samples were centrifuged at 14,000 rpm at room temperature for 30 minutes. Supernatant was transferred to new Eppendorf tubes and dried under vacuum. The dried metabolite residue was

suspended in 50% methanol and centrifuged at 14,000 rpm for 30 minutes. The supernatant was transferred to vials and 6 μ L of sample was subjected to LC-MS analysis. Quality control (QCs) samples were prepared by pooling an aliquot of serum and processed along with experimental samples. The QC samples were analyzed at the beginning of the experiment and following every 5 samples.

The Agilent 1290 series LC system was used to carry out analysis of extracted metabolites with gradient elution mode using solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in acetonitrile). Samples were injected at a composition of mobile phase 98% solvent A and 2% solvent B, and held for 2.5 minutes. Solvent B was then raised to 40% at the 15-minute time mark and to 98% at the 25-minute time mark. This condition was maintained until the 33-minute time point before returning to the initial solvent composition at 34 minutes. Post-run was set to 8 minutes for this analysis. Metabolites were separated chromatographically on an Atlantis T3 column (2.1 x 150mm, 3 μ m, Waters) with a mobile phase flow rate of 0.300 mL/min. Column oven temperature and autosampler temperature were 35°C and 4°C, respectively.

Eluent from the LC system was directed to the time-of-flight mass spectrometer (MS-TOF, model 6230, Agilent Technologies) equipped with an electrospray ionization system. The following instrumental parameters were set for MS analysis: capillary voltage = 4000; fragmentor voltage = 120; drying gas temperature = 330°C; drying gas flow = 11 L/min; nebulizer pressure = 45 psig. Data was acquired in positive ion mode at a scan range of 75 -1700 m/z and scan rate of 1.5 spectra/sec. The mass spectrometer was tuned to verify the performance

of the instrument before beginning the sample analysis. Instrument control, data acquisition and analysis were ascertained by MassHunter software (Agilent Technologies).

With this LC-MS method we would expect to identify a variety of metabolites including amino acids, polar lipids, and simple carbohydrates. However, some of the settings can bias our results. For example, it can be difficult to analyze hydrophobic lipids due to the extraction protocol and LC-MS method. Further, with the use of a positive ion mode, metabolites that typically ionize in a negative mode are not recorded.²⁵ Therefore, some of our metabolites of interest could be outside of our viewing window.

The LC-MS raw (Agilent .d file) data files were converted to mzML files using open source Proteowizard software.²⁶ These data files were then processed by xcms software in the R programming environment.⁹ The *centWave* algorithm in xcms was used to detect chromatographic peaks. The molecular features were grouped across the samples, and the features were further considered only if they were present in more than 40% of samples in at least one sample class (i.e., pre-infection group, post-infection group, vaccinated group, unvaccinated group). Molecular features can vary in their retention time across LC runs; this retention time variability was corrected using the *obiwarp* algorithm.²⁷ A table with integrated values for each aligned peak, which provides measures of the intensity of each measured metabolite feature in each sample was saved as a .csv file for further statistical analysis.

The metabolomics were further analyzed using a pipeline whose development is described in the Results; further tools used in the developed pipeline were the Human Metabolome DataBase Version 4.0, MetaboAnalyst 5.0, and R version 4.0.2.

Results

Broad overview of analysis pipeline

To understand relationships between immune cells and metabolites in the context of a tuberculosis vaccine study, we developed a novel pipeline to identify associations between immune cell populations and metabolite pathways in data collected from BCG-vaccinated versus unvaccinated minipigs before and after infection with *Mycobacterium tuberculosis* (MTB). The pipeline begins with separate pre-processing of flow cytometry and metabolomics data to prepare the two types of data for integrated analysis (Figure 4.2). The two data types are then integrated by correlating the flow cytometry immune cell populations with the metabolomics molecular features: after reducing the dimensions of the metabolomic data with a principal components analysis, we fit multiple linear regression models, with controls included for sampling time point, for every combination of metabolite principal component and immune cell population. This allows us to identify principal components of the metabolomics data that are associated with variation in immune cell populations in a sample. Once we have identified principal components of the metabolite measurements that are associated with immune cell populations, we utilize the variable loadings to identify the specific molecular features that are drivers of that principal component.²⁸ These molecular features can then be used to either identify putative metabolomic pathways (i.e., pathways that, if linked with the immune cell population, would be consistent with the data observed through metabolomics) or perform a more targeted analysis of specific metabolite features with LC-MS/MS. As an additional step, to clarify that associations identified

as interesting are not spurious results of the multiple comparisons applied through testing, we create global p-value histograms for each immune cell population.

Flow cytometry pre-processing and data exploration with cyto-feature engineering

The first step in the flow cytometry analysis workflow was to identify immune cells and markers of interest. These can be utilized to form correlations with the metabolomics data to gain insight on important biological relationships. Flow cytometry results for the minipig data were analyzed via three methods: 1) singular marker expression 2) predefined cellular phenotypes 3) data-driven cellular phenotypes. Analysis of Covariance (ANCOVA) tests were used to test differences both between vaccinated and unvaccinated minipigs and pre- and post- infection. Singular marker expression looked at the overall expression of the 6 flow markers used to analyze the cells. While there were no differences in the singular marker expression between the unvaccinated and vaccinated minipigs pre- and post- infection, there were differences across all animals, unvaccinated and vaccinated from pre-infection to post-infection, with significantly higher levels of CD172, CD45RA, and SLA-DQ after infection (Figure 4.3). There were also significantly lower levels of CD4 after infection.

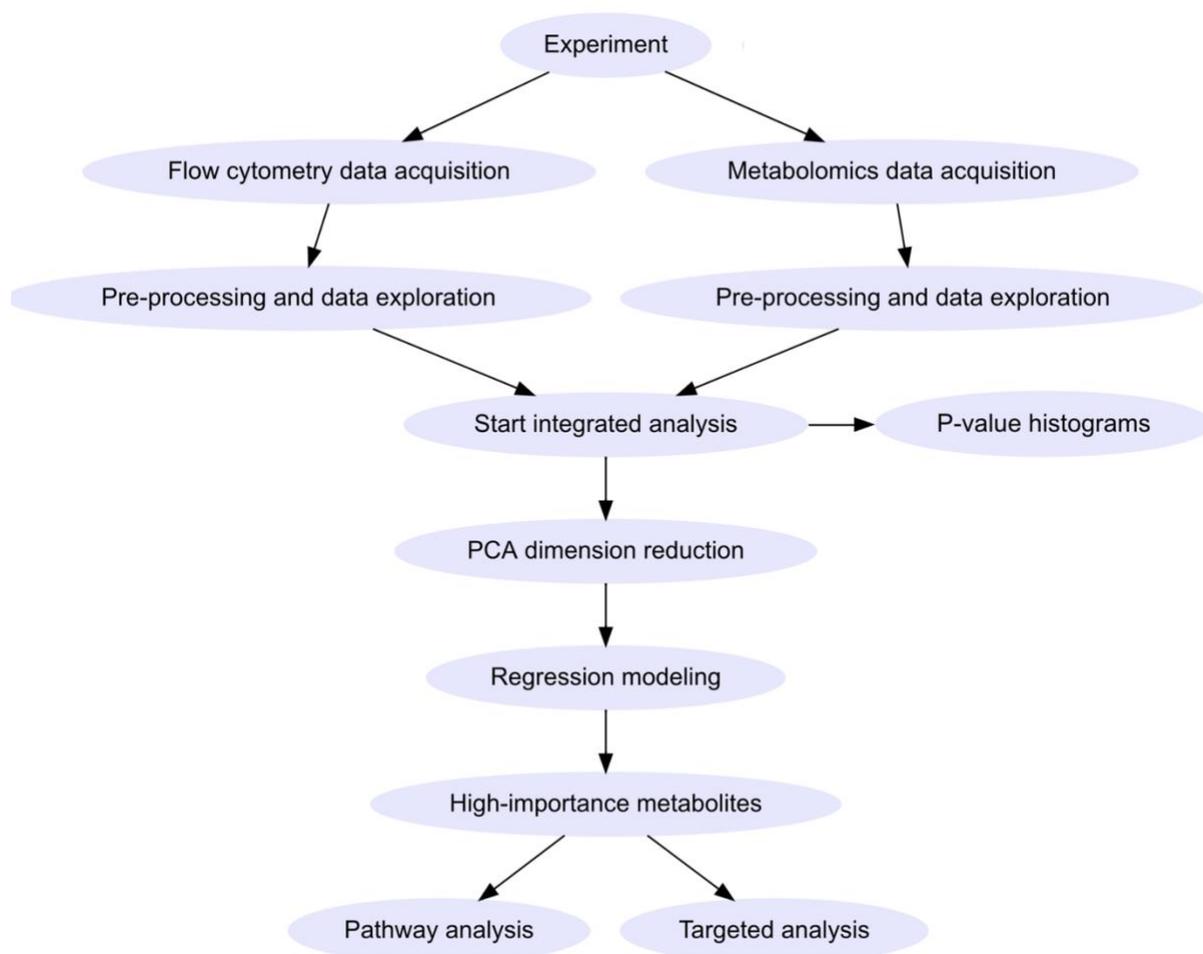


Figure 4.2: Analysis pipeline workflow. Both flow cytometry and metabolomics data are recovered from an experiment. Flow cytometry data is analyzed with a cyto-feature engineering pipeline that identifies all immune cell populations present in the data. Concurrently, metabolomics data is processed to identify peaks and then molecular features in the data. Metabolomics data is then analyzed with a principal components analysis (PCA) to reduce the number of dimensions. Regression models are created to identify correlations between each principal component and immune cell population. The high-importance metabolites associated with the principal components for the significant linear models are determined based on loading cutoffs. Finally, a pathway analysis is performed with the high-importance metabolites to identify the specific metabolic pathways correlated with the cell populations. Alternatively, a targeted analysis for specific metabolites could be performed if the metabolomics data is from an LC-MS/MS experiment. As an additional check, global p-value histogram analyses for each immune cell population are used to confirm that there are compelling correlations between the immune cells and molecular features within the dataset.

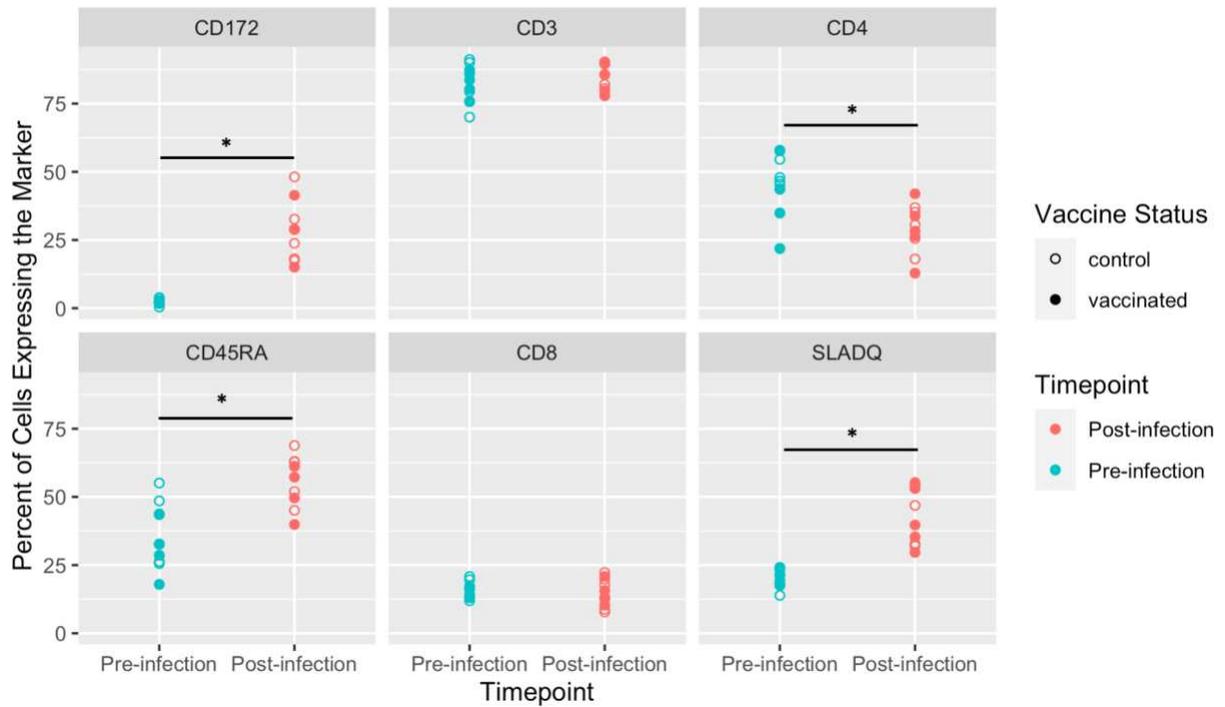


Figure 4.3: Singular marker expression of PBMCs pre- and post-infection via flow cytometry. The percentage of live leukocytes expressing each marker is displayed on the y-axis with the two time points on the x-axis. Asterisks indicate statistically significant differences between groups (* $p < 0.05$).

We then analyzed predefined phenotypes, which are classified as a combination of positive and negative expression of two markers (CD4 and CD8). These predefined phenotypes can shed light on general populations that we expect to see in the minipigs, mainly CD4 T cells (CD4+ CD8-), CD8 T cells (CD4- CD8+), double negative T cells (CD4-CD8-) and double positive T cells (CD4+ CD8+). Again, there were similarly no differences in the predefined phenotypes between vaccinated and unvaccinated minipigs, but there were differences across all animals when comparing pre- and post-infection, with higher levels of double negative T cells and lower levels of CD4 T cells after infection (Figure 4.4).

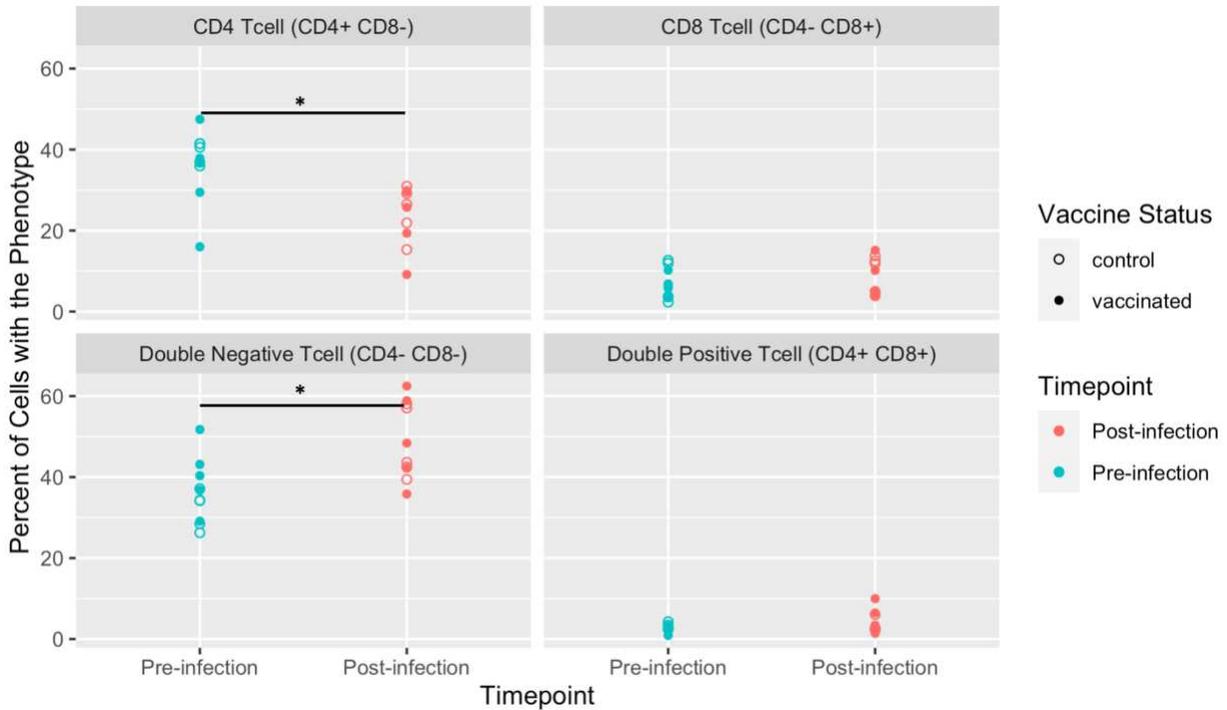


Figure 4.4: Predefined PBMC cellular phenotypes pre- and post-infection via flow cytometry. The percentage cells in each population is displayed on the y-axis with the two time points on the x-axis. Asterisks indicate statistically significant differences between groups (* p < 0.05).

Finally, we used an approach to allow data-driven identification of cellular phenotypes, also known as populations, present in the minipigs. This allows us to identify populations that we may not expect to see in the data. Using the cyto-feature engineering method described in Fox et al., 63 immune cell phenotypes were identified.⁷ These phenotypes are described as all possible negative and positive expression combinations of markers within a sample. These phenotypes were then filtered down to only cells that were CD3+ and constituted at least 1% of one sample, which resulted in 25 phenotypes (Figure 4.5a). These cellular populations are numbered to keep track of them (e.g., Pop1, Pop2), but the interpretation of these populations is based on the combination of positive and negative markers (i.e., Pop1 is a CD3+CD4-CD8-CD45RA-SLADQ-CD172- cell). While there was only one difference between the vaccination groups (in

Pop3: CD3+CD45RA+SLADQ+CD4-CD8-CD172-), there were statistically significant differences between the two time points in 22 of the 25 populations (Figure 4.5b).

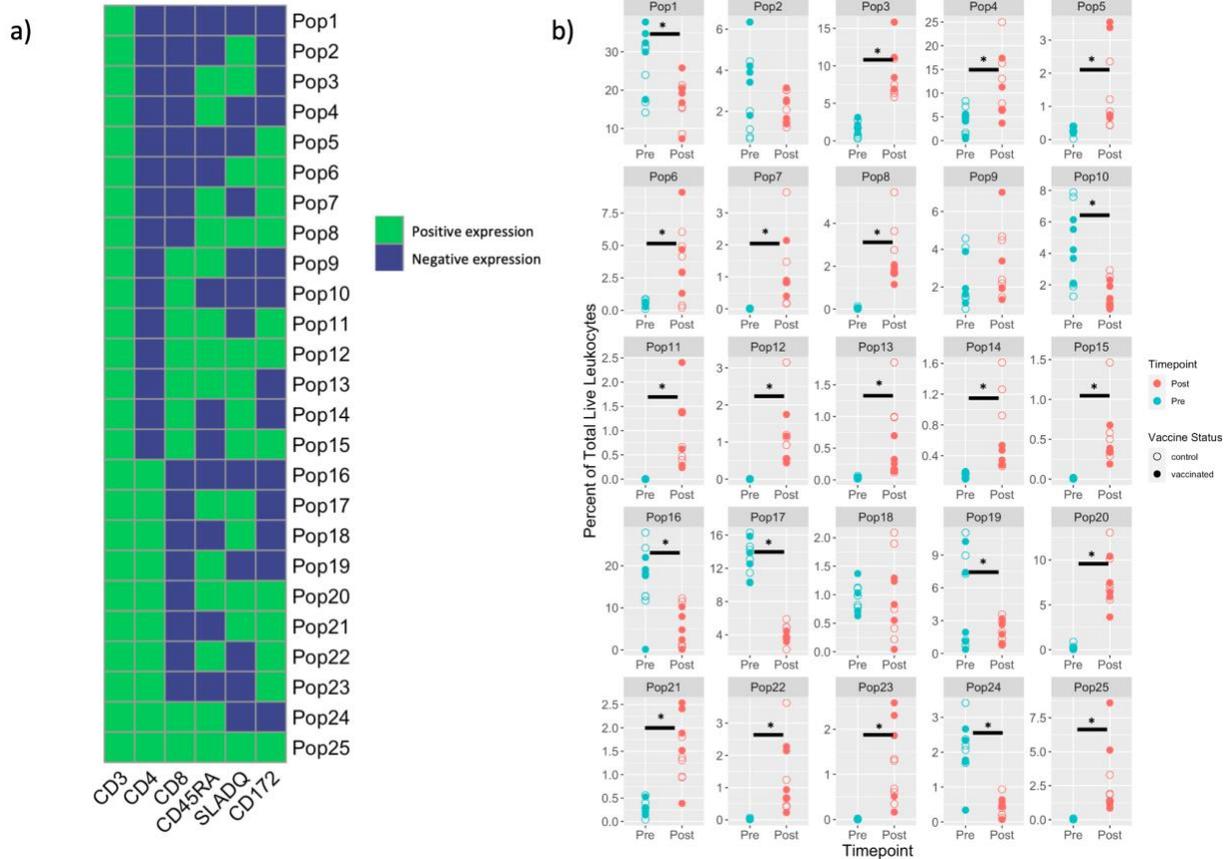


Figure 4.5: Data-driven PBMC cell populations pre- and post-infection via flow cytometry. (a) Cyto-feature engineering was used to identify 25 unique populations that express CD3 and constitute greater than 1% of the live leukocytes in at least one sample. Green indicates positive expression of a marker on the x-axis and blue indicates negative expression of a marker. (b) Each small plot corresponds to a cell population described in a. The percentage of cells in each population is displayed on the y-axis with the two time points on the x-axis. Asterisks indicate statistically significant differences between groups (* $p < 0.05$).

Thus, in total, there were 35 flow cytometry immune cell populations of interest based on this analysis: 6 singular marker expression populations, 4 predefined cellular phenotypes, and 25

data-driven cellular phenotypes. These 35 flow cytometry immune cell populations will be referenced further in these analyses.

Metabolomics pre-processing, data exploration, and PCA dimension reduction

Metabolomics LC-MS data was initially pre-processed with the R package ‘xcms’ to identify molecular features present in the minipigs. A total of 4,504 molecular features were identified. Because this number is too high to try to identify each of the individual metabolites, we tested a variety of methods for filtering to the most relevant molecular features.

Principal component analysis (PCA) was used to visualize the high-dimensional metabolomics data and reduce the data to 19 principal components (Appendix 3: Supplemental Figure S4.1). Representative PCA plots are shown in Figure 4.6. About 50% of the variance in the data can be explained by principal components 1 and 2. There was a lot of overlap in similarities when comparing the vaccinated to unvaccinated minipigs, but there was separation between the pre-infection and post-infection time points along PC1 and PC2. This is very similar to what we saw in the flow cytometry when comparing vaccination status and time point.

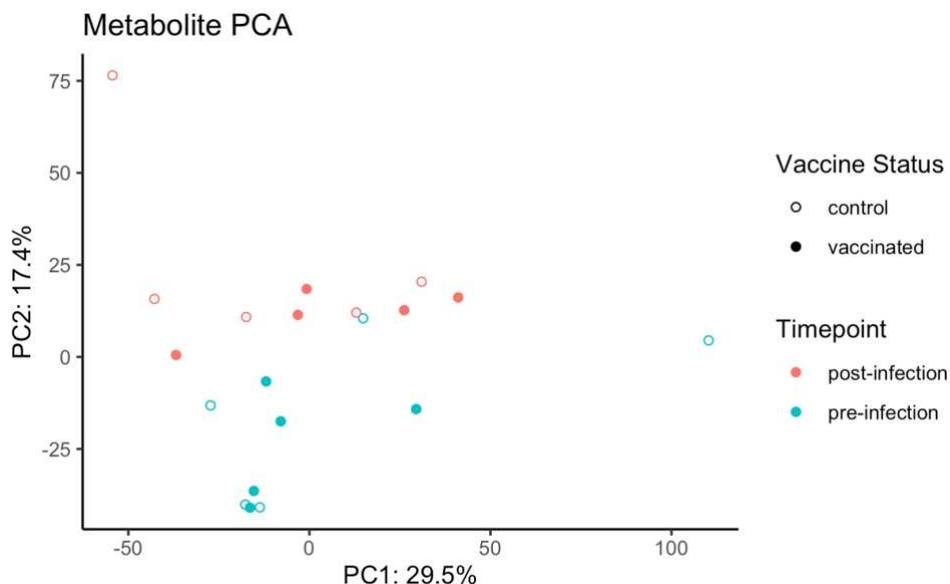


Figure 4.6: Metabolomics PCA Plot. Differences between the pre-infection and post-infection time points were recognized along PC2 while separation on the PC plots due to vaccination status was not.

Regression modelling

After a brief exploration of the immune cell population and metabolomics data, we moved into the data integration stage. When preparing for regression modelling, we could use simple linear models to correlate each molecular feature with each cell population, but we would encounter the multiple comparisons problem.²⁹ When multiple comparisons are made (such as in this case where 4,504 comparisons are made for each of the 35 immune cell populations), there is a high probability that the number of statistically significant features is inflated due to chance. To account for this multiple comparison issue, we reduced the number of correlations and inferences made by utilizing dimensionality reduction methods. Dimensionality reduction is a common technique that transforms high-dimensional data to low-dimensional data while maintaining key information.³⁰ Principal components analysis is an example of dimensionality reduction which reduces the number of features in data but retains information on variance in the data. Therefore,

we can reduce the dimensions of the metabolite data by using the 19 principal components rather than the 4,504 individual molecular features in our regression modelling.

Regression modelling frameworks are incredibly flexible and allow us to control for many different factors. Because we did not see any differences between the vaccinated and unvaccinated animals in our preliminary analysis, we did not include a control for vaccination status in our regression modelling. However, we saw differences between the two time points. The goal of this analysis was to identify changes in immune cell populations that can be explained by changes in various metabolites; this relationship should remain the same regardless of time point. We used an interaction linear model first to weed out the pairs where the relationship between immune cells and molecular features was different for the two time points. We then moved to a less complex additive linear regression model to determine which markers and metabolites were correlated. These regression modeling steps are described below.

We performed linear regressions on each of the 19 principal components with the 35 immune cell populations to see if the flow cytometry results were correlated with each principal component, adjusting for time point interactions. This interaction linear model (Eq. 1) was used to determine if the interaction between the principal components (PC) value and the immune cell percentage was different based on the timepoints; this interaction model allowed there to be a different relationship at the two time points.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \quad (\text{Eq. 1})$$

Where Y : Immune cell percentage; X_1 : PC value, X_2 : Time point

If a molecular feature is truly correlated with an immune cell population, we would expect to see the same or at least very similar correlation regardless of confounding factors such as time points. Thus, we filtered to correlations where the p-value for the interaction between PC value and time point was not significant ($p > 0.05$) meaning that the relationship between the PC value and the immune cell percentages was not consistent across time points. This removed six correlations that were not of interest. From here, we could use a less complex additive linear model after removing the six correlations where there was an interaction between the time point and PC value (Eq. 2).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad (\text{Eq. 2})$$

Where Y : Immune cell percentage; X_1 : PC value, X_2 : Time point

After filtering our additive linear model to those where the p-value < 0.01 for the correlation between immune cell percentage and PC value, there were 13 immune cell-principal component pairs (Table 4.1).

Table 4.1: Statistically significant additive linear model pairs. The table displays the immune cell population-principal component pairs in which the p-value for the slope is significant ($p < 0.01$) for an additive linear model.

Immune Cells	Principal Components	P value
CD45RA	PC4	5.8e-04
CD172	PC3	2.8e-03
Pop7: CD3+CD45RA+CD172+CD4-CD8-SLADQ-	PC3	3.1e-04
Pop8: CD3+CD45RA+SLADQ+CD172+CD4-CD8-	PC3	1.3e-05
Pop10: CD3+CD8+CD4-CD45RA-SLADQ-CD172-	PC4	2.9e-04
Pop11: CD3+CD8+CD45RA+CD172+CD4-SLADQ-	PC6	2.4e-03
Pop12: CD3+CD8+CD45RA+SLADQ+CD172+CD4-	PC3	1.1e-04
Pop15: CD3+CD8+SLADQ+CD172+CD4-CD45RA-	PC3	4.4e-06
Pop16: CD3+CD4+CD8-CD45RA-SLADQ-CD172-	PC4	7.8e-03
Pop19: CD3+CD4+CD45RA+CD8-SLADQ-CD172-	PC4	1.2e-03
Pop20: CD3+CD4+CD45RA+SLADQ+CD172+CD8-	PC3	5.7e-04
Pop22: CD3+CD4+CD45RA+CD172+CD8-SLADQ-	PC3	9.9e-04
Pop23: CD3+CD4+CD172+CD8-CD45RA-SLADQ-	PC6	5.7e-03

When visualizing the 13 correlation plots, we observed that some of the pairs identified in this analysis had very low or non-existent immune cell populations pre-infection. Because the goal of this analysis was to identify correlations where the relationship remained regardless of time point, we then filtered these populations to those where both of the time points exhibited variability in both the immune cell and metabolomics observations. This resulted in 4 immune cell-principal component pairs (Figure 4.7).

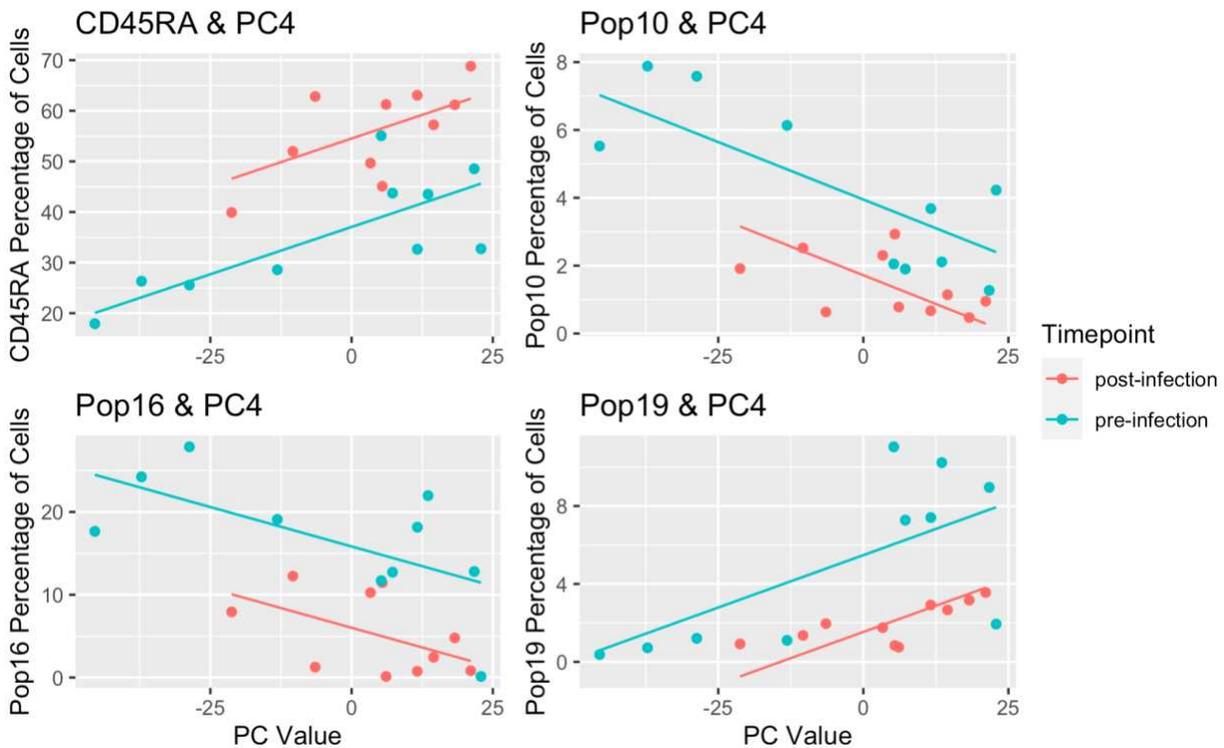


Figure 4.7: Immune cell-principal component pairs for which there was a correlation between the observations, and variability in both observations. The x-axis in each small plot gives the principal component value for PC4, and the y-axis gives the percentage of cells in each of the respective flow cytometry immune cell populations. Additive linear models are shown for the two time points: pre-infection and post-infection, where each point represents a minipig. Pop10: CD3+CD8+CD4-CD45RA-SLADQ-CD172-; Pop16: CD3+CD4+CD8-CD45RA-SLADQ-CD172-; Pop19: CD3+CD4+CD45RA+CD8-SLADQ-CD172-

Gathering high-importance metabolites associated with principal components

All of the significant correlations occurred between immune cell populations and PC4. PC4 is a measure of the molecular features, so we could then go back into this principal component and see which of the specific molecular features were driving this component. We could do this by visualizing the loading that each molecular feature has on the principal component. If the loading for a molecular feature is near 0, then it has little effect on the principal component, but if the loading is highly positive or highly negative, it has a greater effect. We then compiled the molecular features that had a large impact on PC4, mainly those with loading ≥ 0.03 and ≤ -0.03

(Figure 4.8). For PC4, there were 140 molecular features with loading ≥ 0.03 and 171 molecular features with loading ≤ -0.03 .

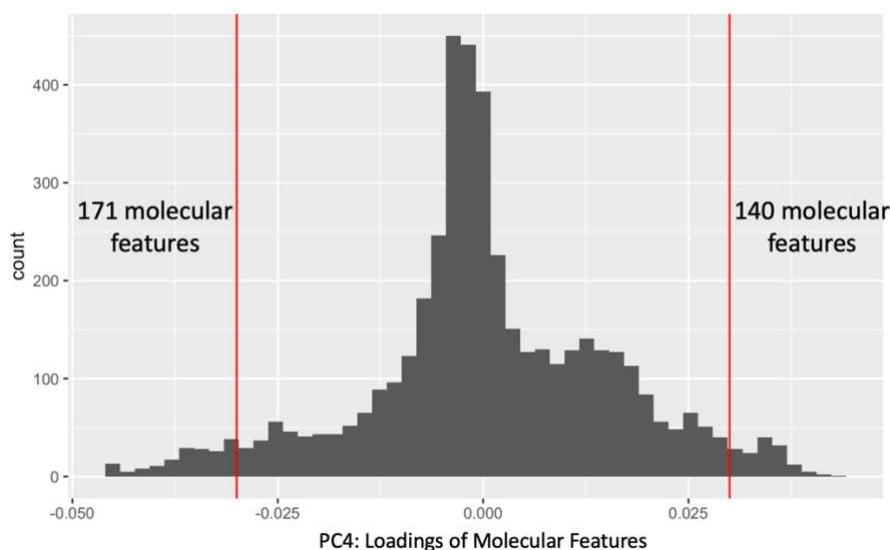


Figure 4.8: Histogram of the loadings of the molecular features (MF) that contribute to PC4. The x-axis denotes the loadings of the molecular features, and the y-axis denotes the number of molecular features with each loading. Red lines indicate the cutoffs for molecular features considered “important” and used in further analyses.

Pathway analysis

While it is not possible to definitely identify metabolites using LC-MS data, a pathway analysis can be used to identify putative pathways that are plausible and consistent with our data. To perform the putative pathway analysis, we utilized the list of molecular features that had large impacts on PC4. The m/z masses were queried on the Human Metabolome Database to identify all possible metabolites for each m/z value. MetaboAnalyst’s pathway analysis was then used to identify the putative pathways. The list of features for the highly positive and highly negative loadings for PC4 were sent through MetaboAnalyst separately. The Homo sapiens (KEGG) reference genome was used because a minipig reference genome does not exist. Hypergeometric tests were used to calculate p-values for the probability that each of the pathways was identified.

The list of putative pathways was then filtered to those in which in the p-value ≤ 0.01 (Table 4.2).

Table 4.2: Significant putative pathways correlated with flow cytometry immune cell populations. The “Metabolites in Pathway” denotes the total number of metabolites in the pathway on the MetaboAnalyst database where the “identified metabolites” gives the number of metabolites putatively identified in the data.

Putative Metabolic Pathways Correlated with Immune Cell Populations

Correlated PC	Putative Pathway	Metabolites In Pathway	Identified Metabolites	P value
PC4	Arachidonic acid metabolism	36	29	6.9e-21
PC4	Linoleic acid metabolism	5	5	3.3e-05
PC4	Arginine and proline metabolism	38	8	4.3e-03

This pathway analysis identified three metabolic pathways—arachidonic acid metabolism, linoleic acid metabolism, arginine and proline metabolism—that were consistent with the data. Arachidonic acid metabolism had the lowest p-value out of these three pathways. For arachidonic acid metabolism, there were 29 metabolites that were consistent with patterns in our molecular feature data. This is not to say that all 29 of these metabolites were found in our data, but it is plausible that some of these metabolites were present.

Biological Significance

Interestingly, we potentially identified 29 out of 36 metabolites present in the MetaboAnalyst database associated with arachidonic acid metabolism (Figure 4.9, Appendix 3: Supplemental Table 4.1). When we drilled back down to the original molecular features that were associated with arachidonic acid metabolism, we saw a negative correlation between the relative abundance of the metabolites and the percentage of cells that express CD45RA (Figure 4.10).

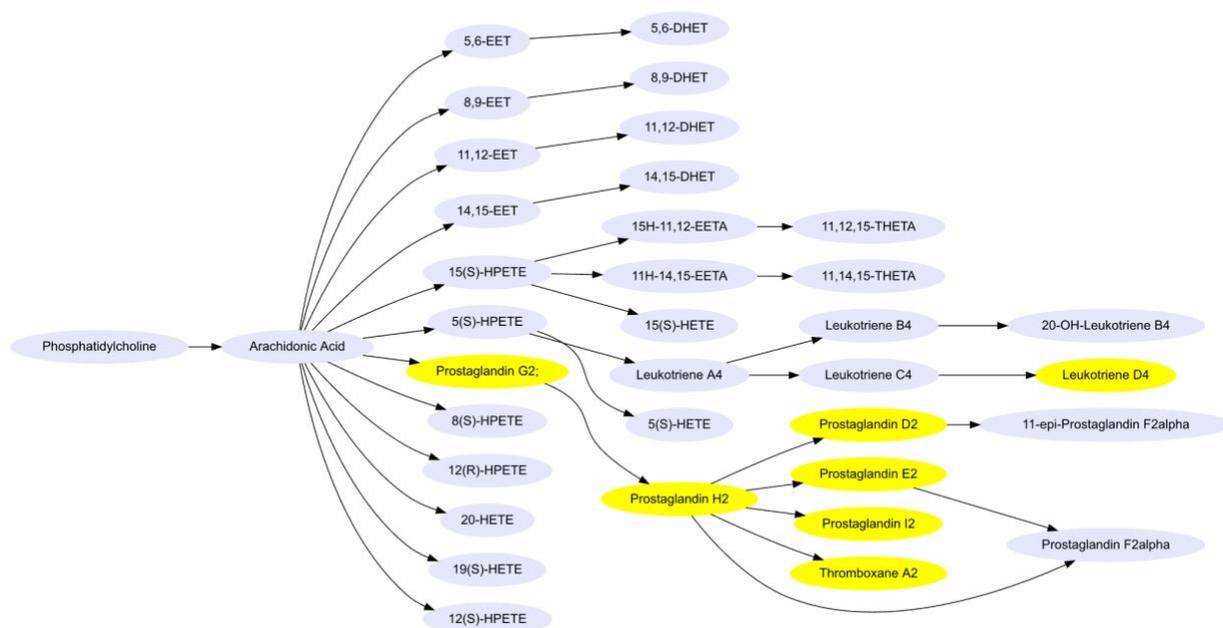


Figure 4.9: Simplified arachidonic acid metabolism plot adapted from MetaboAnalyst and KEGG. The nodes in lavender are the metabolites that were putatively identified in our analysis while the metabolites highlighted in yellow were not. KEGG IDs for each of the metabolites can be found in Appendix 3: Supplemental Table 4.1.

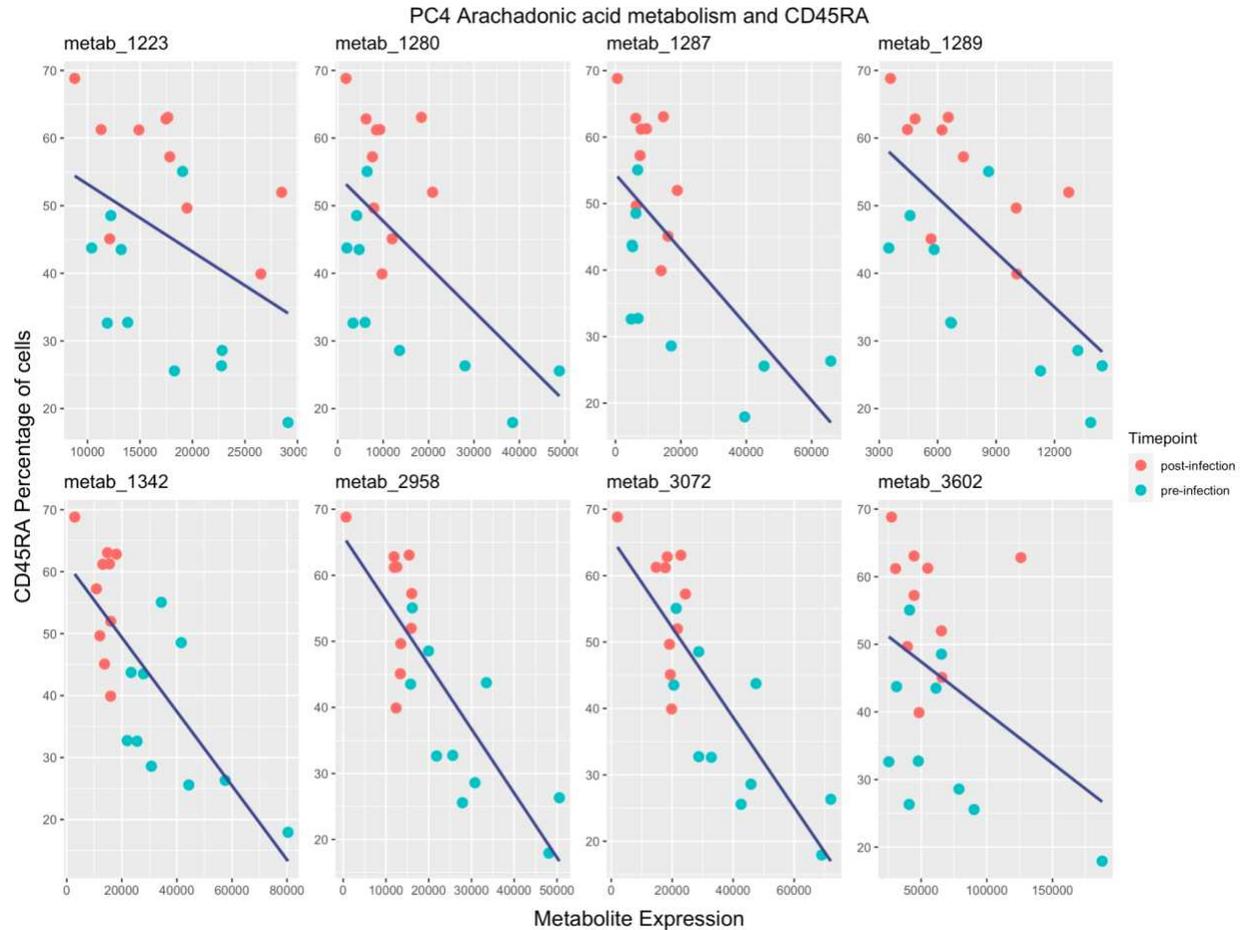


Figure 4.10: Molecular features potentially associated with arachidonic acid metabolism have a negative relationship with CD45RA. Each small plot shows the association between a different molecular feature and CD45RA. The x-axis in each small plot gives the relative metabolite expression, and the y-axis gives the percentage of cells that express CD45RA.

CD45RA is a marker that is expressed both on naive T cells (CD45RA⁺ CCR7⁺), but also a subset of effector memory T cells that re-express CD45RA upon stimulation (CD45RA⁺ CCR7⁻).³¹ Similarly, we can see that generally, post-infection, there is a higher percentage of CD45RA⁺ cells. This could potentially be explained by the fact that there is stimulation due to infection in the minipig that promotes the differentiation of naive T cells to effector memory T cells.

Arachidonic acid metabolism, on the other hand, produces a variety of small molecules such as prostaglandins, eicosanoids, and leukotrienes that can greatly affect immune cell function. For

example, leukotriene B4 has been shown to increase chemotaxis of effector T cells, but not naive or central memory T cells.^{32,33} With the chemotaxis of effector T cells to the tissue, we may expect to see a lower level of CD45RA cells present in the blood as these cells have already migrated to the infected tissue. While it is not possible to know if any of the metabolites identified in this analysis are definitely Leukotriene B4, this could explain why we see a lower CD45RA expression when arachidonic acid metabolism, and therefore, leukotriene B4, is increased.

P-value histograms and family-wise error rate

As a supplemental check, we wanted to confirm that there is compelling evidence to support the theory that the metabolites were correlated with various immune cell populations in our data. Because there were 4,504 molecular features, there are many opportunities to detect correlations that are just due to chance. We used standard statistical methods that focus on family-wise error rates to determine if the number of strong correlations are expected. For each of the 35 immune cell populations, we performed linear regressions with each of the 4,504 molecular features and created a p-value histogram that shows the p-values for each of the 4,504 test comparisons (Figure 4.11). By doing this, we could visualize the family of tests for each immune cell population under the null hypothesis that there are no correlations between the immune cells and molecular features.

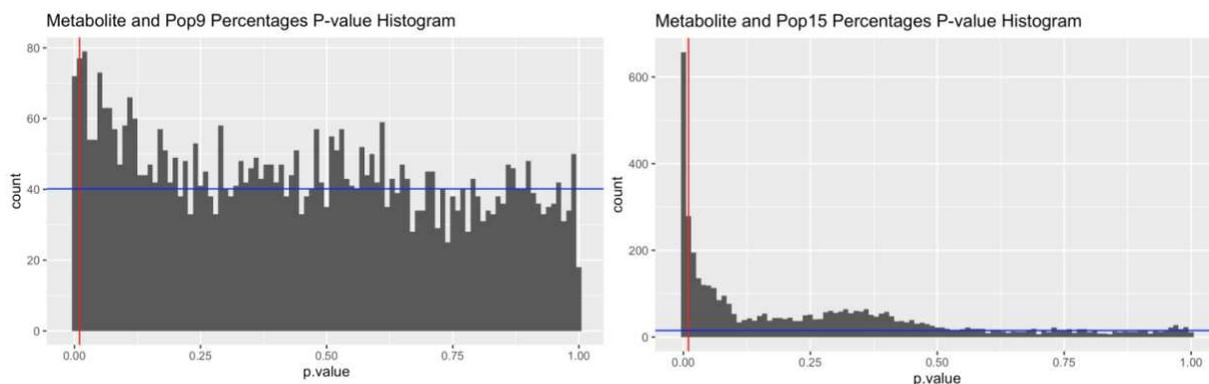


Figure 4.11: Histograms of p-values associated with correlation plots. Each histogram plot shows all of the p-values for the correlation of an immune cell population with each molecular feature. The red line represents our statistically significant p-value cut off ($\alpha = 0.01$) and the blue line represents the false discovery rate.

Some tests are going to be positive just as a result of chance—in that case we would see something like the left plot on Figure 4.11 where there is a fairly uniform distribution of p-values from 0-1. With high statistical power and statistically significant results, we would expect our results to have a peak below our statistically significant p-value cut off—showing that there is likely a correlation between the immune cells and the metabolites (Figure 4.11, right). This does not necessarily mean that every single test below our p-value cutoff is true, but we can calculate the false discovery rate (FDR) (Eq. 3) to estimate the fraction of false rejections.

$$\text{False Discovery Rate} = \frac{2 * (\# \text{ pvalues} > 0.5) * \alpha}{\# \text{ pvalues} \leq \alpha} \quad (\text{Eq. 3})$$

This false discovery rate cannot show which of the tests below our statistically significant p-value cutoff are true and which are due to chance, but we can estimate the percentage that are true. There was a large difference in the range of FDR values (1.8%-36.2%) based on each immune cell populations (Table 4.3). Based on this data, we can see that there are certain

immune cell populations such as Pop15 (CD3+CD8+SLADQ+CD172+CD4-CD45RA-) where there are clear correlations between the immune cell populations and molecular features because of the peak below our p-value cutoff and the low FDR. Therefore, we can confirm that there are interesting correlations occurring.

When we compare the results of the four correlations identified through our regression modelling to the corresponding FDRs, all of the immune cells and metabolites that we interrogated have a FDR less than 15%. This cannot guarantee that the correlations we identified are true but provides more evidence to support our theory that the immune cells and metabolic pathways that we identified are correlated.

Table 4.3: False Discovery Rate for each immune cell population. Table showing each of the immune cell populations and the number of molecular features that are correlated with the immune cells (column: Number p-values ≤ 0.01), the number of p-values calculated to be false within this number (column: False Discovery Number), and the calculated False Discovery Rate. The table is arranged by the population type and then ascending false discovery rate.

Summary of P-value Histogram False Discovery Rate

Population Type	Immune Cell Populations	Number P-values ≤ 0.01	Number Falsely Significant	False Discovery Rate (%)
Singular Marker Expression	CD172	671	30	4.4
Singular Marker Expression	SLADQ	561	33	5.9
Singular Marker Expression	CD3	301	23	7.8
Singular Marker Expression	CD45RA	503	44	8.8
Singular Marker Expression	CD4	149	40	26.5
Singular Marker Expression	CD8	94	47	50.4
Predefined Phenotype	CD8_Tcell	226	39	17.2
Predefined Phenotype	DN_Tcell	117	30	25.5
Predefined Phenotype	CD4_Tcell	164	45	27.2
Predefined Phenotype	DP_Tcell	50	48	96.9
Unsupervised Phenotype	Pop15	822	15	1.8
Unsupervised Phenotype	Pop12	770	20	2.6
Unsupervised Phenotype	Pop8	739	22	3.0
Unsupervised Phenotype	Pop24	728	22	3.0
Unsupervised Phenotype	Pop20	743	29	4.0
Unsupervised Phenotype	Pop22	499	24	4.8
Unsupervised Phenotype	Pop7	587	31	5.3
Unsupervised Phenotype	Pop17	621	37	6.0
Unsupervised Phenotype	Pop19	371	26	7.0
Unsupervised Phenotype	Pop21	392	35	9.0
Unsupervised Phenotype	Pop3	431	44	10.2
Unsupervised Phenotype	Pop14	325	33	10.3
Unsupervised Phenotype	Pop11	313	37	11.8
Unsupervised Phenotype	Pop1	325	39	12.0
Unsupervised Phenotype	Pop6	229	28	12.4
Unsupervised Phenotype	Pop16	373	47	12.5
Unsupervised Phenotype	Pop10	375	48	12.7
Unsupervised Phenotype	Pop23	278	39	14.1
Unsupervised Phenotype	Pop4	264	48	18.1
Unsupervised Phenotype	Pop13	175	37	21.3
Unsupervised Phenotype	Pop25	160	34	21.3
Unsupervised Phenotype	Pop5	135	40	30.0
Unsupervised Phenotype	Pop9	111	40	36.2
Unsupervised Phenotype	Pop2	82	43	52.8
Unsupervised Phenotype	Pop18	23	65	284.4

Discussion

Although the purpose of the primary experiment was to identify immunological and metabolic differences between vaccinated and unvaccinated minipigs in response to *Mycobacterium tuberculosis* infection, we did not see differences between the vaccinated and unvaccinated minipigs at either of the time points in either the flow cytometry or metabolomics data. This could potentially be due to the low BCG vaccine dose administered. As expected, we did see differences in both the flow cytometry and metabolomics data before and after infection.

Although the primary experiment was unsuccessful in terms of seeing differences between vaccinated and unvaccinated animals, we were able to re-analyze the data with new techniques to gain new knowledge. We originally hypothesized that we could identify metabolic profiles correlated with immune cells through this novel flow cytometry and metabolite analysis pathway. We similarly hypothesized that we would see correlations between T cells expressing SLADQ, and glycolysis, the pentose phosphate pathway, and amino acid synthesis pathways. While we were not able to identify metabolites relating to these pathways, the analysis described here offered a few promising leads. We discovered correlations of CD45RA+ cells, Pop10 (CD3+CD8+CD4-CD45RA-SLADQ-CD172-), Pop16 (CD3+CD4+CD8-CD45RA-SLADQ-CD172-), and Pop19 (CD3+CD4+CD45RA+CD8-SLADQ-CD172-) with arginine and proline metabolism, arachidonic acid metabolism, and linoleic acid metabolism. In particular, we found the negative relationship between CD45RA and arachidonic acid metabolism. Because we were testing so many comparisons, it is possible that this correlation was spurious, but we could use this as a driver for future experiments to specifically test this relationship. If we can establish that this relationship is true, and that the CD45RA+ cells shown here are primarily effector memory

T cells, increasing arachidonic acid metabolism during vaccination could improve vaccine efficacy by promoting chemotaxis of these cells to the lungs. To corroborate these results, future research could utilize the marker CCR7 in addition to CD45RA to monitor chemotaxis and identify levels of naive T cells (CD45+ CCR7+), central memory T cells (CD45RA- CCR7+), and effector memory T cells (CD45+ CCR7-) both in the blood and in the lungs. These cells could then be cell-sorted and LC-MS/MS could be run to confirm that leukotriene B4 or other specific arachidonic acid metabolites are correlated with CD45RA expression.

This novel technique utilizes two measurable dependent variables to identify interesting trends in the data. As long as there is variation in the observations, we are able to utilize this method and identify valuable information from the data. Another advantage of this pipeline is that it considers multiple comparison problems in two ways. First, it uses global p-value histograms to confirm that there are indeed correlations of interest between observations. It also utilizes principal components to represent molecular features which greatly reduces the dimensions of the data. This innovative approach to selecting molecular features, can be used to putatively identify pathways as described in this paper or to identify molecular features to run through further LC-MS/MS analysis. Unfortunately, in this experiment, we were unable to perform LC-MS/MS due to low sample volume. However, this technique and pathway analysis can still be used for generating a hypothesis for future experiments.

The analysis could have been improved if we had a larger sample size. There was noise in the data, meaning that there may be some cases where the slopes in the regression modelling were not considered significant due to this noise. Samples sizes are generally determined to power an

initial hypothesis. However, when utilizing secondary data, the power may decrease when testing different hypotheses. Another limitation is that we utilized the Human Metabolome Database instead of a minipig metabolome database. While there are many similar metabolites in both humans and minipigs, we could have missed some minipig-specific metabolites using the human database.

Leveraging secondary datasets to understand immunometabolism can help in the fight against infectious diseases, and to develop better vaccines and treatments, as highlighted by the recent call from the NIH to take full advantage of secondary datasets.⁵ The technique described here allows for the use of measurable dependent variables to identify correlations in data and considers multiple comparison issues. Overall, this technique can help advance our immunometabolism knowledge using secondary datasets.

REFERENCES

1. Centers for Disease Control and Prevention. (2021). World TB Day 2021.
2. Moliva, J.I., Turner, J., and Torrelles, J.B. (2017). Immune Responses to Bacillus Calmette-Guerin Vaccination: Why Do They Fail to Protect against Mycobacterium tuberculosis? *Frontiers in immunology* 8, 407-407. 10.3389/fimmu.2017.00407.
3. O'Neill, L.A.J., Kishton, R.J., and Rathmell, J. (2016). A guide to immunometabolism for immunologists. *Nature Reviews Immunology* 16, 553-565. 10.1038/nri.2016.70.
4. Nature. Scientific Data: Data Repository Guide. <https://www.nature.com/sdata/policies/repositories>.
5. Health, N.I.o.A.a.I.D.a.N.I.o.M. (February 12, 2021). Notice of Special Interest (NOSI): Secondary Analysis of Existing Datasets for Advancing Immune-mediated and Infectious Disease Research.
6. Van Gassen, S., Callebaut, B., Van Helden, M.J., Lambrecht, B.N., Demeester, P., Dhaene, T., and Saeys, Y. (2015). FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry Part A* 87A, 636-645. 10.1002/cyto.a.22625.
7. Fox, A., Dutt, T.S., Karger, B., Rojas, M., Obregon-Henao, A., Anderson, G.B., and Henao-Tamayo, M. (2020). Cyto-Feature Engineering: A Pipeline for Flow Cytometry Analysis to Uncover Immune Populations and Associations with Disease. *Scientific Reports* 10. 10.1038/s41598-020-64516-0.
8. Rainer, J. (2021). LCMS data preprocessing and analysis with xcms.
9. Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification. *Analytical Chemistry* 78, 779-787. 10.1021/ac051437y.
10. Vinaixa, M., Samino, S., Saez, I., Duran, J., Guinovart, J.J., and Yanes, O. (2012). A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data. *Metabolites* 2, 775-795. 10.3390/metabo2040775.
11. van der Windt, G.J.W., and Pearce, E.L. (2012). Metabolic switching and fuel choice during T-cell differentiation and memory development. *Immunological Reviews* 249, 27-42. 10.1111/j.1600-065X.2012.01150.x.
12. Berg, J.M., Tymoczko, J.L., and Stryer, L. (2002). *Biochemistry - 5th Edition*. W. H. Freeman.
13. Castellano, F., and Molinier-Frenkel, V. (2020). Control of T-Cell Activation and Signaling by Amino-Acid Catabolizing Enzymes. *Frontiers in Cell and Developmental Biology* 8, 613416. 10.3389/fcell.2020.613416.
14. Bian, Y.J., Li, W., Kremer, D.M., Sajjakulnukit, P., Li, S.S., Crespo, J., Nwosu, Z.C., Zhang, L., Czerwonka, A., Pawlowska, A., et al. (2020). Cancer SLC43A2 alters T cell methionine metabolism and histone methylation. *Nature* 585, 277-+. 10.1038/s41586-020-2682-1.
15. Lunney, J.K., Ho, C.S., Wysocki, M., and Smith, D.M. (2009). Molecular genetics of the swine major histocompatibility complex, the SLA complex. *Developmental and Comparative Immunology* 33, 362-374. 10.1016/j.dci.2008.07.002.

16. Chen, F.X., Xie, J., Li, N.L., Zhou, Y., Xin, L.J., and Chou, K.Y. (2005). Novel SLA-DQ alleles and their recombinant molecules in xenogeneic stimulation of human T cells. *Transplant Immunology* *14*, 83-89. 10.1016/j.trim.2005.02.002.
17. Ramos, L., Obregon-Henao, A., Henao-Tamayo, M., Bowen, R., Izzo, A., Lunney, J.K., and Gonzalez-Juarrero, M. (2019). Minipigs as a neonatal animal model for tuberculosis vaccine efficacy testing. *Veterinary Immunology and Immunopathology* *215*, 109884. 10.1016/j.vetimm.2019.109884.
18. Aguilar, G.I.S., Beshah, E., Vengroski, K.G., Zarlenga, D., Jauregui, L., Cosio, M., Douglass, L.W., Dubey, J.P., and Lunney, J.K. (2001). Cytokine and lymphocyte profiles in miniature swine after oral infection with *Toxoplasma gondii* oocysts. *International Journal for Parasitology* *31*, 187-195. 10.1016/s0020-7519(00)00159-4.
19. Ramos, L., Obregon-Henao, A., Henao-Tamayo, M., Bowen, R., Lunney, J.K., and Gonzalez-Juarrero, M. (2017). The minipig as an animal model to study *Mycobacterium tuberculosis* infection and natural transmission. *Tuberculosis* *106*, 91-98. 10.1016/j.tube.2017.07.003.
20. Hanekom, W.A. (2005). The immune response to BCG vaccination of newborns. *Human Immunology: Patient-Based Research* *1062*, 69-78. 10.1196/annals.1358.010.
21. Ordway, D.J., Shang, S.B., Henao-Tamayo, M., Obregon-Henao, A., Nold, L., Caraway, M., Shanley, C.A., Basaraba, R.J., Duncan, C.G., and Orme, I.M. (2011). *Mycobacterium bovis* BCG-Mediated Protection against W-Beijing Strains of *Mycobacterium tuberculosis* Is Diminished Concomitant with the Emergence of Regulatory T Cells. *Clinical and Vaccine Immunology* *18*, 1527-1535. 10.1128/cvi.05127-11.
22. Molins, C.R., Ashton, L.V., Wormser, G.P., Andre, B.G., Hess, A.M., Delorey, M.J., Pilgard, M.A., Johnson, B.J., Webb, K., Islam, M.N., et al. (2017). Metabolic differentiation of early Lyme disease from southern tick-associated rash illness (STARI). *Science Translational Medicine* *9*, eaa12717. 10.1126/scitranslmed.aal2717.
23. Fitzgerald, B.L., Molins, C.R., Islam, M.N., Graham, B., Hove, P.R., Wormser, G.P., Hu, L.D., Ashton, L.V., and Belisle, J.T. (2020). Host Metabolic Response in Early Lyme Disease. *Journal of Proteome Research* *19*, 610-623. 10.1021/acs.jproteome.9b00470.
24. Fitzgerald, B., Graham, B., Delorey, M., Pegalajar-Jurado, A., Islam, M., Wormser, G., Aucott, J., Rebman, A., Soloski, M., Belisle, J., and Molins, C. (September 25, 2020). Metabolic response in patients with post-treatment Lyme disease symptoms/syndrome. *Clinical Infectious Diseases*.
25. Nordstrom, A., Want, E., Northen, T., Lehtio, J., and Siuzdak, G. (2008). Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics. *Analytical Chemistry* *80*, 421-429. 10.1021/ac701982e.
26. Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* *30*, 918-920. 10.1038/nbt.2377.
27. Prince, J.T., and Marcotte, E.M. (2006). Chromatographic alignment of ESI-LC-MS proteomics data sets by ordered bijective interpolated warping. *Analytical Chemistry* *78*, 6140-6152. 10.1021/ac0605344.
28. Zhang, Z.H., and Castello, A. (2017). Principal components analysis in clinical studies. *Annals of Translational Medicine* *5*, 351. 10.21037/atm.2017.07.12.

29. Franceschi, P., Giordan, M., and Wehrens, R. (2013). Multiple comparisons in mass-spectrometry-based -omics technologies. *Trac-Trends in Analytical Chemistry* 50, 11-21. 10.1016/j.trac.2013.04.011.
30. Holmes, S., and Huber, W. (2019). Chapter 7: Multivariate Analysis. In *Modern Statistics for Modern Biology*, (Cambridge University Press).
31. Tian, Y., Babor, M., Lane, J., Schulten, V., Patil, V.S., Seumois, G., Rosales, S.L., Fu, Z., Picarda, G., Burel, J., et al. (2017). Unique phenotypes and clonal expansions of human CD4 effector memory T cells re-expressing CD45RA. *Nature Communications* 8, 1473. 10.1038/s41467-017-01728-5.
32. Goodarzi, K., Goodarzi, M., Tager, A.M., Luster, A.D., and von Andrian, U.H. (2003). Leukotriene B-4 and BLT1 control cytotoxic effector T cell recruitment to inflamed tissues. *Nature Immunology* 4, 965-973. 10.1038/ni972.
33. Ott, V.L., Cambier, J.C., Kappler, J., Marrack, P., and Swanson, B.J. (2003). Mast cell-dependent migration of effector CD8(+) T cells through production of leukotriene B-4. *Nature Immunology* 4, 974-981. 10.1038/ni971.

CHAPTER 5 — CONCLUDING REMARKS

Tuberculosis is still a disease that kills more than 1.4 million people every year. The current vaccine does not adequately protect against pulmonary tuberculosis, and it is not fully understood why the BCG vaccine fails to offer sufficient protection. Ongoing research focuses on which immune cells or correlates of protection are necessary to improve protection. Thus, the goals of our studies were 1) to develop a flow cytometry analysis pipeline that utilizes feature engineering and aids in identification of immune cell population 2) to test a novel BCG vaccination boosting strategy utilizing two drugs—losartan and propranolol and 3) to use a novel integrative metabolomics approach to identify correlations between immune cells and metabolites during *Mycobacterium tuberculosis* infection.

Novel flow cytometry computational tools can aid in identification of immune cell populations

Flow cytometers can now analyze up to 50 parameters per cell and millions of cells per sample; however, conventional methods to analyze data are subjective and time-consuming. To address these issues, we have developed a novel flow cytometry analysis pipeline to identify a plethora of cell populations efficiently. Coupled with feature engineering and immunological context, researchers can immediately extrapolate novel discoveries through easy-to-understand plots. The R-based pipeline uses Fluorescence Minus One (FMO) controls or distinct population differences to develop thresholds for positive/negative marker expression. The continuous data is transformed into binary data, capturing a positive/negative biological dichotomy often of interest in characterizing cells. Next, a filtering step refines the data from all identified cell phenotypes to

populations of interest. The data can be partitioned by immune lineages and statistically correlated to other experimental measurements. The pipeline's modularity allows customization of statistical testing, adoption of alternative initial gating steps, and incorporation of other datasets. Validation of this pipeline through manual gating of two datasets (murine splenocytes and human whole blood) confirmed its accuracy in identifying even rare subsets. Lastly, this pipeline can be applied in all disciplines utilizing flow cytometry regardless of cytometer or panel design. The code is available at https://github.com/aef1004/cyto-feature_engineering. Though work is ongoing, future studies could focus on further developing the pipeline into a comprehensive R package that is submitted to Bioconductor.

BCG boosting with immunomodulatory drugs, losartan and propranolol, does not induce better protection against *Mycobacterium tuberculosis* infection

A better boosting strategy for the BCG vaccine is desperately needed. The purpose of this study was to modify the immune response to BCG—using the drugs losartan and propranolol—to induce better protection against *Mycobacterium tuberculosis*. This boosting strategy was tested in C57BL/6 mice that were subsequently infected with *Mycobacterium tuberculosis*. We hypothesized that losartan would reduce the number of inflammatory monocytes to the lungs and allow better antigen presentation and thus better T cell activation. We further hypothesized that propranolol would promote T cell polarization to Th1 cells and enhance cytolytic capabilities of CD8⁺ T cells. We conceived that these immunomodulatory effects would offer better protection against MTB and reduce the bacterial burden. While this boosting strategy induced higher levels of T cells (Th1, CD8⁺, CD4⁺) in the losartan groups and the propranolol oral groups after administration of the drugs, we did not see a reduction in inflammatory monocytes. Further,

these high T cell numbers did not sustain after infection, and the boosting strategy did not reduce the bacterial burden in the mice that received the immunomodulatory drugs. Future studies could further explore the dosage and timing of immunomodulatory drug administration.

Novel analysis method identifies potential correlations between immune cell populations and metabolic pathways.

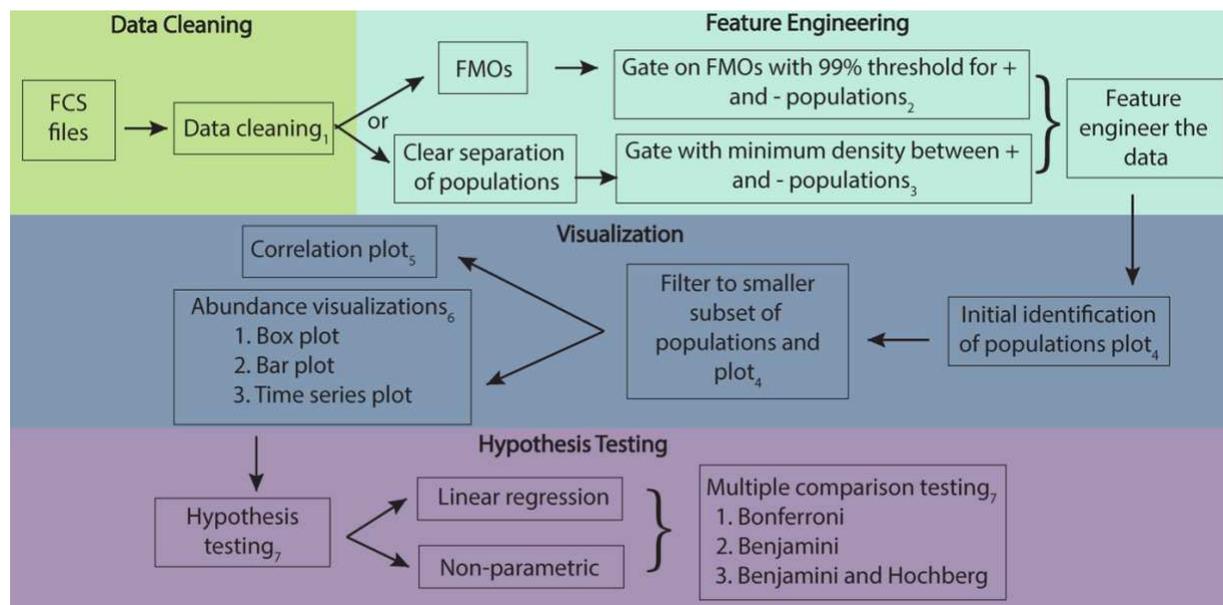
Immunometabolism is an important field to understand the immune response to infection. We developed a new method for identifying correlations between immune cells and metabolites. The new method utilizes global p-value histogram analyses and PCA clustering to account for multiple comparison issues. It further utilizes linear models with the principal components from a PCA analysis to identify correlations. The high-importance metabolites from these principal components can be identified and correlated with specific immune cells. We hypothesized that we could utilize this novel method to identify correlations between metabolic profiles and immune cells. We further hypothesized that metabolites associated with glycolysis, the pentose phosphate pathways, and amino acid synthesis pathways would have a positive correlation with activated T cells. While we did not identify this specific correlation, we putatively identified four immune cell populations (CD45RA⁺ cells, CD3⁺ CD4⁺ CD8⁻ CD45RA⁻ SLADQ⁻ CD172⁻ cells, CD3⁺ CD8⁺ CD4⁻ CD45RA⁻ SLADQ⁻ CD172⁻ cells, and CD3⁺ CD4⁺ CD45RA⁺ CD8⁻ SLADQ⁻ CD172⁻ cells) correlated with arginine and proline metabolism, arachidonic acid metabolism, and linoleic acid metabolism. Future studies could conclude if CD45RA and memory T cells are negatively correlated with arachidonic acid metabolism through the use of

cell sorting and LC-MS/MS. If this is corroborated, the link between CD45RA and arachidonic acid metabolism could be used as a potential target for future vaccination studies.

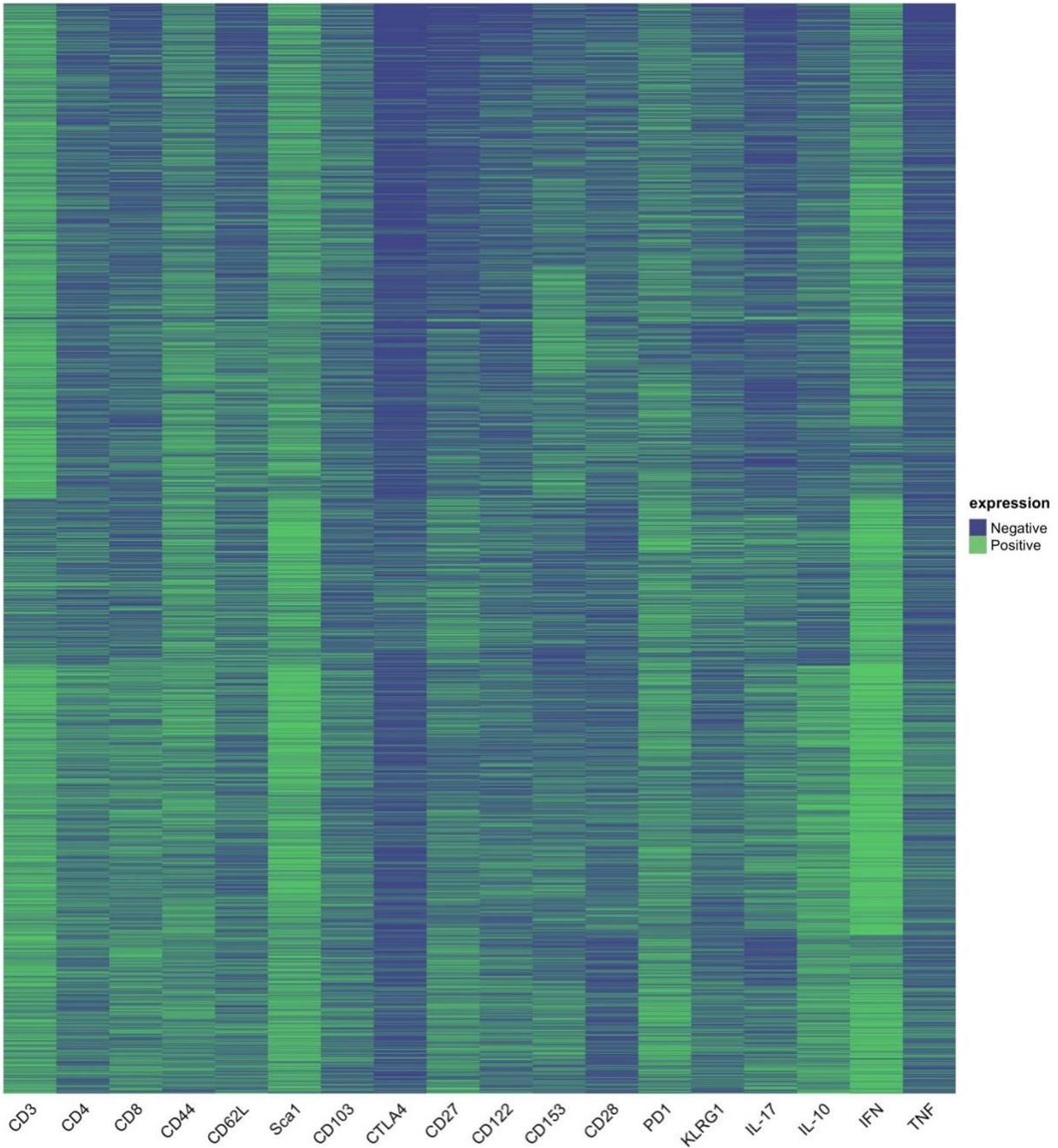
Concluding remarks

The road ahead is a long one, but collectively these studies have increased both our knowledge and workflows for tuberculosis research. The cyto-feature engineering pipeline can be used to reduce bias in analyses and time spent analyzing data. The new method for correlating metabolites with immune cells can also be used to gain insight from secondary datasets and generate hypotheses. Although the immunomodulatory vaccine study did not offer increased protection against MTB, studies like these are important to understand the immune response. Further, other scientists can learn from this study and either adjust their study accordingly or avoid repeating it. The fight for a better vaccine has been a long one, but cumulatively, these new analysis tools and immunological knowledge will help in the fight against MTB and other infectious diseases.

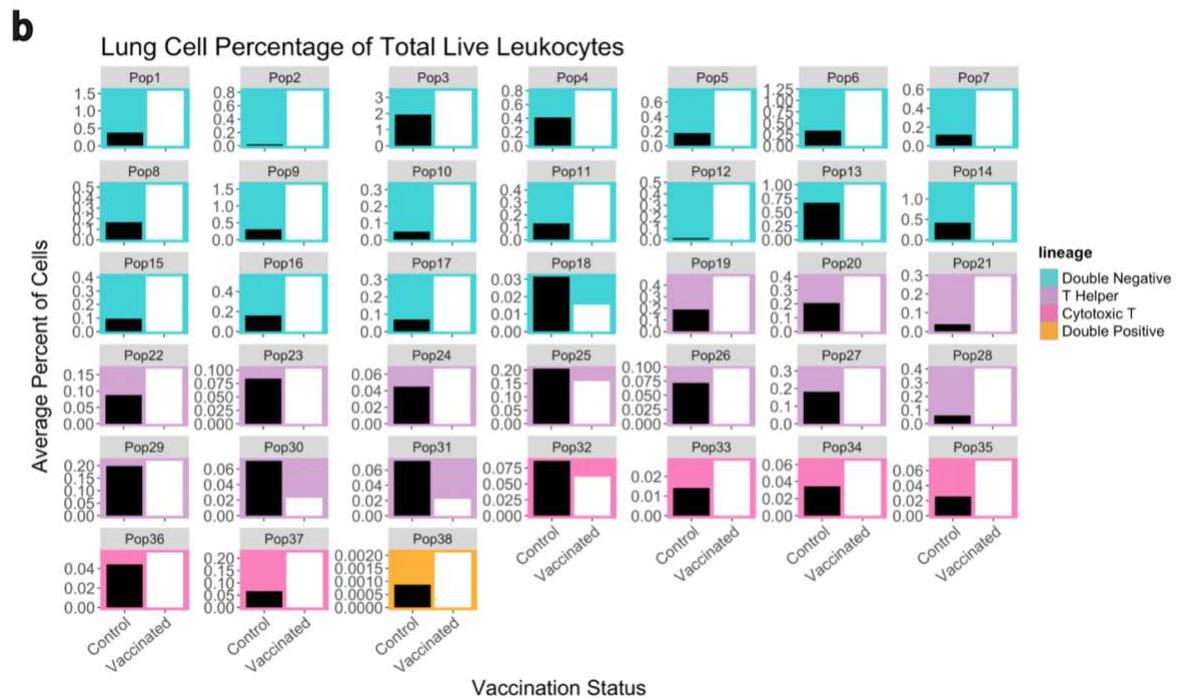
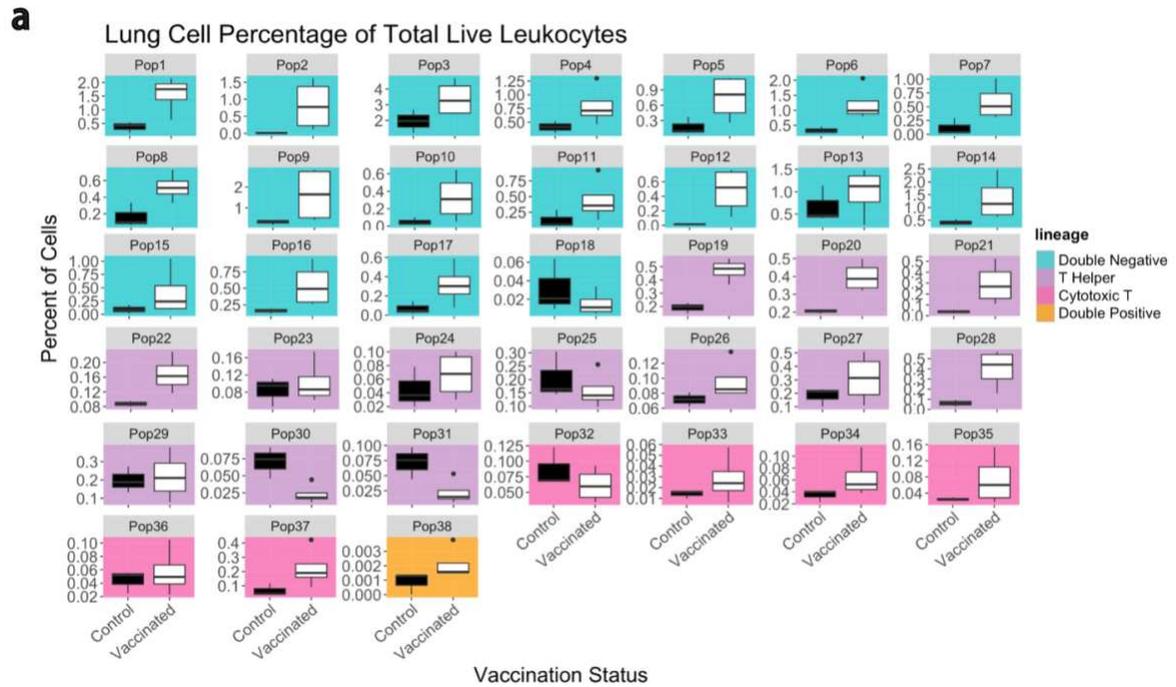
APPENDIX 1 — SUPPLEMENTAL MATERIALS FOR CYTO-FEATURE ENGINEERING:
 A PIPELINE FOR FLOW CYTOMETRY ANALYSIS TO UNCOVER IMMUNE
 POPULATIONS AND ASSOCIATIONS WITH DISEASE
 ENGINEERING: A PIPELINE FOR FLOW CYTOMETRY ANALYSIS TO UNCOVER
 IMMUNE POPULATIONS AND ASSOCIATIONS WITH DISEASE



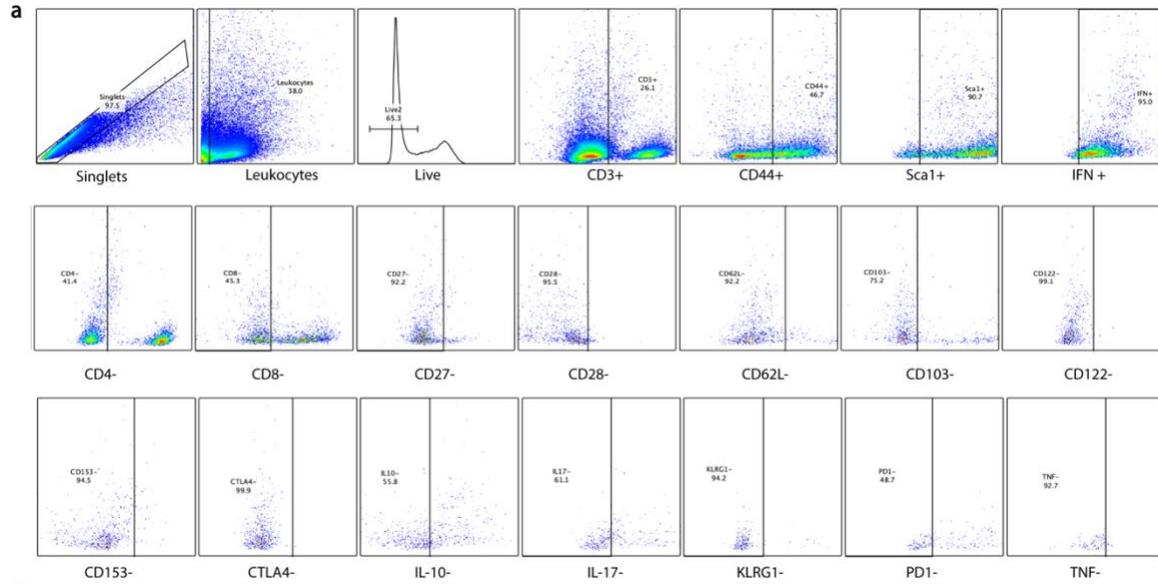
Supplemental Figure S2.1: Pipeline route map. The route map details the workflow of the pipeline based on the samples acquired and the visualiations and hypothesis testing of interest. The packages used at each of the stages are noted as follows: (1) *ncdfFlow*: to read in the flow cytometry data to R, *openCyto*: to facilitate automated gating for data cleaning, *ggcyto*: to visualize the initial gating strategy, *tibble*: to convert the S4 flow data object to a data frame, *flowCore*: to transform the data if needed, *dplyr*: to tidy the data (2) *quantile*: to perform the 99% FMO threshold cutoff (3) *openCyto*: to calculate the minimum density between the positive and negative cell populations (4) *pheatmap*: to plot the identified cell populations (5) *superheat*: to find correlations between different cell populations (6) *ggplot2*: to plot cell population abundances and changes in abundances over time (7) *stats*: to perform linear regressions, calculate p-values, and execute multiple comparison testing, *ggplot2*: to plot the cell abundances against other data from the experiment.



Supplemental Figure S2.2: Identification of all phenotypes in the lung samples. After feature engineering the data, 12,122 total populations were identified. Each row represents a unique cell phenotype, where green indicates positive expression and blue indicates negative expression of each marker.

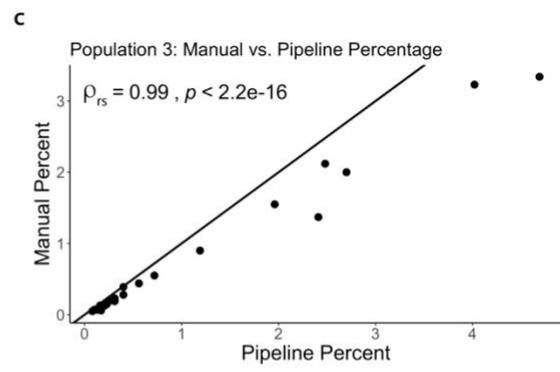


Supplemental Figure S2.3: Alternative visualizations for cell percentages. All of the plot backgrounds denote the cell lineages as described in Figure 2.3a. The day 30 data was used for all of the plots in this figure. a) Box plots show the distribution of the percentage of cells in each population. b) Bar plots show the average percentage of cells in the two experimental groups in each population.

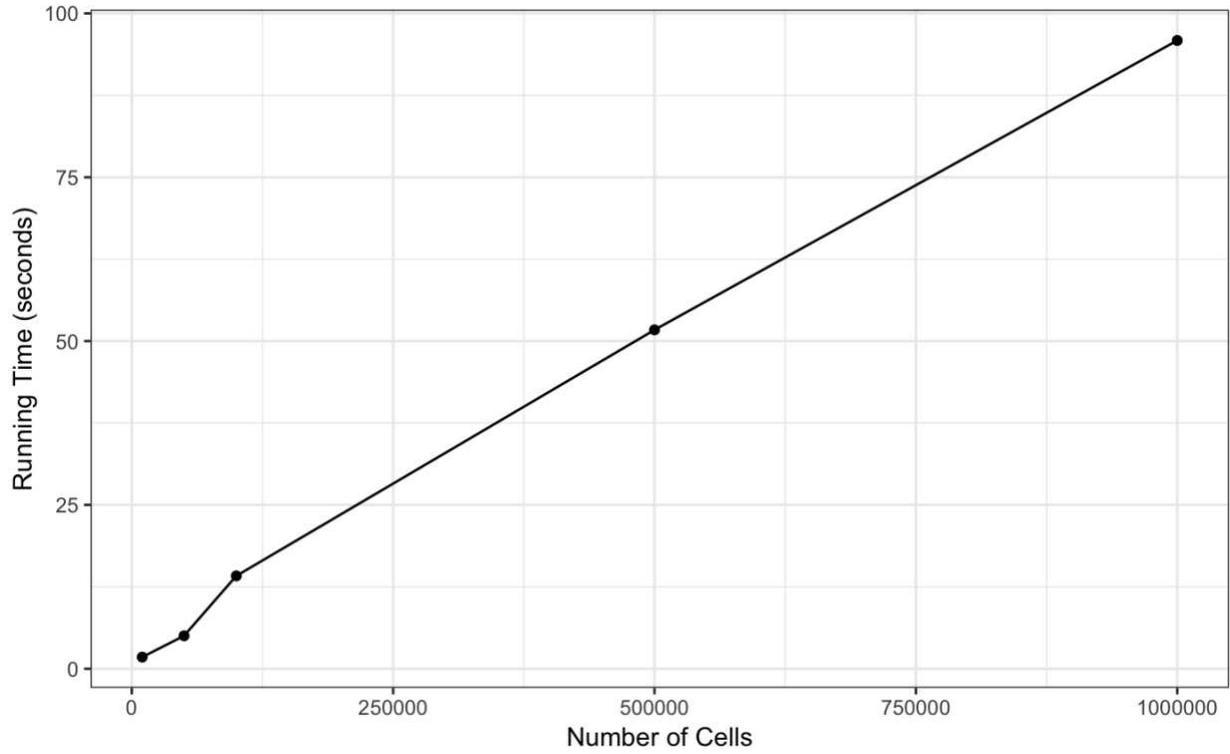


b

Timepoint	Group	Number	% Based On Manual Gating	% Based on Automated Pipeline	Absolute Difference	Average Difference
30	Control	A	2.00	2.70	0.70	0.26
30	Control	B	1.55	1.96	0.41	0.26
30	Control	C	0.90	1.19	0.29	0.26
30	Vaccinated	A	2.12	2.48	0.36	0.26
30	Vaccinated	B	3.23	4.02	0.79	0.26
30	Vaccinated	C	1.37	2.41	1.04	0.26
30	Vaccinated	D	3.34	4.69	1.36	0.26
60	Control	A	0.09	0.14	0.06	0.26
60	Control	B	0.07	0.10	0.02	0.26
60	Control	C	0.05	0.08	0.03	0.26
60	Control	D	0.07	0.13	0.06	0.26
60	Control	E	0.06	0.17	0.11	0.26
60	Vaccinated	A	0.19	0.25	0.06	0.26
60	Vaccinated	B	0.13	0.21	0.09	0.26
60	Vaccinated	C	0.15	0.23	0.08	0.26
60	Vaccinated	D	0.16	0.23	0.07	0.26
90	Control	A	0.19	0.31	0.12	0.26
90	Control	B	0.13	0.16	0.03	0.26
90	Control	C	0.23	0.31	0.09	0.26
90	Control	D	0.28	0.40	0.12	0.26
90	Vaccinated	A	0.55	0.72	0.17	0.26
90	Vaccinated	B	0.44	0.56	0.12	0.26
90	Vaccinated	C	0.39	0.40	0.01	0.26
90	Vaccinated	D	0.16	0.22	0.05	0.26



Supplemental Figure S2.4: Confirmation of populations via manual gating. (a) Population 3, defined by CD3+ CD44+ Sca1+ IFN- γ + CD8- CD27- CD28- CD62L- CD69- CD103- CD122- CD153- CTLA4- IL10- IL17- KLRG1- PD1- TNF- α , was manually gated in FlowJo using the FMOs. (b) The difference in the percentage of cells in population 3 for each mouse at each time point was calculated, as well as the absolute average difference. (c) A comparison of manual and pipeline gating results for this population. Each point represents the measurements of population 3 in a single mouse. The points' position on the x-axis gives the population 3 measurement based on the pipeline while the position on the y-axis gives the population based on manual gating. The diagonal line provides a reference of $x = y$ (i.e., where points would fall if results from manual and pipeline gating were identical). The Spearman correlation coefficient (ρ_{rs}) and p-value are displayed on the plot.



Supplemental Figure S2.5: Linear running time for feature engineering algorithm. The x-axis displays the number of cells that are put into the feature engineering algorithm and the y-axis displays the amount of time in seconds that it takes to compute the feature engineering using the input number of cells.

```

R version 3.6.2 (2019-12-12)
Platform: x86_64-apple-darwin15.6.0 (64-bit)
Running under: macOS Catalina 10.15.3

Matrix products: default
BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions/A/libBLAS.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] grid      parallel  stats      graphics  grDevices  utils      datasets  methods   base

other attached packages:
 [1] Rgraphviz_2.30.0      graph_1.64.0      BiocGenerics_0.32.0  jpeg_0.1-8.1
 [5] broom_0.5.5          ggpubr_0.2.5      magrittr_1.5         kableExtra_1.1.0
 [9] superheat_0.1.0     viridis_0.5.1    viridisLite_0.3.0   scales_1.1.0
[13] forcats_0.5.0       dplyr_0.8.5      purrr_0.3.3         readr_1.3.1
[17] tidyr_1.0.2         tibble_2.1.3     tidyverse_1.3.0     ggcorrplot_0.1.3
[21] gridExtra_2.3       readxl_1.3.1     stringr_1.4.0       pheatmap_1.0.12
[25] ggcyto_1.14.1       ncdfFlow_2.32.0  BH_1.72.0-3         RcppArmadillo_0.9.850.1.0
[29] ggplot2_3.3.0       data.table_1.12.8 openCyto_1.24.0     flowWorkspace_3.34.1
[33] flowCore_1.52.1

loaded via a namespace (and not attached):
 [1] colorspace_1.4-1    ggsignif_0.6.0    ellipsis_0.3.0    rprojroot_1.3-2   mclust_5.4.5
 [6] corpcor_1.6.9       fs_1.3.2          clue_0.3-57       rstudioapi_0.11   farver_2.0.3
[11] hexbin_1.28.1      remotestats_2.1.1 IDPmisc_1.1.20    fansi_0.4.1       mvtnorm_1.1-0
[16] lubridate_1.7.4    xml2_1.2.5        splines_3.6.2     R.methodsS3_1.8.0 mnormt_1.5-6
[21] robustbase_0.93-5 knitr_1.28        pkgload_1.0.2     jsonlite_1.6.1    cluster_2.1.0
[26] dbplyr_1.4.2       png_0.1-7         R.oo_1.23.0       BiocManager_1.30.10 rrcov_1.5-2
[31] compiler_3.6.2     httr_1.4.1        backports_1.1.5   assertthat_0.2.1 Matrix_1.2-18
[36] lazyeval_0.2.2     cli_2.0.2         prettyunits_1.1.1 htmltools_0.4.0   tools_3.6.2
[41] gtable_0.3.0       glue_1.3.2        Rcpp_1.0.3        Biobase_2.46.0    cellranger_1.1.0
[46] vctrs_0.2.4        nlme_3.1-145     xfun_0.12         ps_1.3.2          testthat_2.3.2
[51] rvest_0.3.5        lifecycle_0.2.0  devtools_2.2.2   gtools_3.8.1     DEoptimR_1.0-8
[56] zlibbioc_1.32.0    MASS_7.3-51.5    hms_0.5.3         RBGL_1.62.1       RColorBrewer_1.1-2
[61] yaml_2.2.1         memoise_1.1.0    latticeExtra_0.6-29 stringi_1.4.6     desc_1.2.0
[66] pcaPP_1.9-73       flowClust_3.24.0 pkgbuild_1.0.6    flowViz_1.50.0   rlang_0.4.5
[71] pkgconfig_2.0.3    matrixStats_0.56.0 evaluate_0.14     fda_2.4.8.1      lattice_0.20-40
[76] labeling_0.3       ks_1.11.7        processx_3.4.2    tidyselect_1.0.0 plyr_1.8.6
[81] R6_2.4.1           generics_0.0.2   DBI_1.1.0         pillar_1.4.3     haven_2.2.0
[86] withr_2.1.2        modelr_0.1.6     crayon_1.3.4     KernSmooth_2.23-16 ellipse_0.4.1
[91] rmarkdown_2.1      usethis_1.5.1    callr_3.4.2       reprex_0.3.0     digest_0.6.25
[96] webshot_0.5.2     R.utils_2.9.2    flowStats_3.44.0 RcppParallel_5.0.0 stats4_3.6.2
[101] munsell_0.5.0     sessioninfo_1.1.1

```

Supplemental Figure S2.6: Cyto-feature engineering pipeline session info. The R and package versions used for this manuscript are listed.

Supplemental Table 2.1: Definition of T cell lineages. The markers used to classify the T cell lineages are denoted in the “Associated Markers” column.

Cell Lineage	Associated Markers
Double Negative	CD3+ CD4- CD8-
T helper	CD3+ CD4+ CD8-
Cytotoxic T	CD3+ CD4- CD8+
Double Positive	CD3+ CD4+ CD8+

Supplemental Table 2.2: Definition of T cell types. The markers used to classify the T cell types are denoted in the “Associated Markers” column.

Cell Type	Associated Markers
Naive	CD44- CD62L+ Sca1-
Effector	CD44+ CD62L-
Central Memory	CD44+ CD62L+
Stem-cell Like Memory	CD44- CD62L+ Sca1+
Resident	CD103+

Supplemental Table 2.3: Adjusted p-values for populations associated with bacterial burden. The r squared and p-values for the linear regression lines in Figure 2.5b are displayed for each population. The Benjamini and Hochberg False Discovery Rate correction is used to adjust the p-values based on the multiple tests that are performed. The significance column depicts if the adjusted p-value is less than 0.05.

population	r squared	p-value	Adjusted p-value	Significance
Pop1	0.4979925	0.0001170	0.0003176	TRUE
Pop2	0.2974035	0.0058494	0.0085491	TRUE
Pop3	0.6881335	0.0000005	0.0000103	TRUE
Pop4	0.2587603	0.0111376	0.0156751	TRUE
Pop5	0.5633154	0.0000239	0.0001124	TRUE
Pop6	0.4265429	0.0005402	0.0012829	TRUE
Pop7	0.5294946	0.0000559	0.0001930	TRUE
Pop8	0.5988273	0.0000092	0.0000657	TRUE
Pop9	0.3932679	0.0010393	0.0019748	TRUE
Pop10	0.4185797	0.0006337	0.0014124	TRUE
Pop11	0.2480886	0.0132496	0.0179816	TRUE
Pop12	0.3704975	0.0015978	0.0028912	TRUE
Pop13	0.4738976	0.0002002	0.0005071	TRUE
Pop14	0.5063059	0.0000967	0.0002826	TRUE
Pop15	0.3377243	0.0029005	0.0045925	TRUE
Pop16	0.5346706	0.0000493	0.0001872	TRUE
Pop17	0.5098544	0.0000890	0.0002820	TRUE
Pop18	0.0979274	0.1365125	0.1621086	FALSE
Pop19	0.0025116	0.8161077	0.8381647	FALSE
Pop20	0.5592121	0.0000266	0.0001124	TRUE
Pop21	0.4033427	0.0008554	0.0017109	TRUE
Pop22	0.3465869	0.0024747	0.0040886	TRUE
Pop23	0.3172453	0.0041589	0.0063216	TRUE
Pop24	0.0487987	0.2995875	0.3162312	FALSE
Pop25	0.0602502	0.2476404	0.2688667	FALSE
Pop26	0.6502334	0.0000020	0.0000186	TRUE
Pop27	0.0003482	0.9310329	0.9310329	FALSE
Pop28	0.4158473	0.0006690	0.0014124	TRUE
Pop29	0.1353798	0.0769004	0.0942650	FALSE
Pop30	0.0816482	0.1758853	0.1965777	FALSE
Pop31	0.1622194	0.0510141	0.0668461	FALSE
Pop32	0.7600534	0.0000000	0.0000011	TRUE
Pop33	0.6721796	0.0000009	0.0000120	TRUE
Pop34	0.5765940	0.0000169	0.0000916	TRUE
Pop35	0.5944313	0.0000104	0.0000657	TRUE
Pop36	0.3640287	0.0018010	0.0031108	TRUE
Pop37	0.0955722	0.1415766	0.1630276	FALSE
Pop38	0.1514692	0.0601433	0.0761815	FALSE

Supplemental Table 2.4: Flow cytometry antibodies. The T cell Surface panel antibody cocktail is prepared in FACS staining buffer with a 1:10 dilution of Brilliant Violet Buffer (BD). The T cell Intracellular panel antibody cocktail is prepared in permeabilization buffer.

T cell Panel – Surface

Fluor	Marker	Dilution	Catalog	Clone	RRID
BB515	Sca-1	1:1000	BD: 8127577	D7	RRID: AB_2739218
Alexa 532	CD3	1:50	Invitrogen: 58-0032-82	17A2	RRID: AB_11217479
PE Dazzle 594	CD62L	1:500	BioLegend: 104448	MEL-14	RRID: AB_2566163
PE Cy5	CD122	1:100	BioLegend: 123220	TM- β 1	RRID: AB_2715962
PerCP Cy5.5	CD28	1:50	BioLegend: 102114	37.51	RRID: AB_2073850
PerCP e710	PD-1	1:100	Invitrogen: 46-9981-82	RMP1-30	RRID: AB_11151142
APC R700	CD103	1:200	BD: 565529	M290	RRID: AB_2739282
APC Fire750	CD44	1:1000	BioLegend: 103062	IM7	RRID: AB_2616727
BV480	CD4	1:100	BD: 565634	RM4-5	RRID: AB_2739312
BV570	CD8	1:100	BioLegend: 100739	53-6.7	RRID: AB_10897645
BV605	CTLA-4	1:50	BioLegend: 106323	UC10-4B9	RRID: AB_2566467
BV650	CD27	1:100	BioLegend: 124233	LG.3A10	RRID: AB_2687192
BV711	CD153	1:50	BD: 740751	RM153	RRID: AB_2740419
BV785	KLRG-1	1:100	BD: 565477	2F1	RRID: AB_2739256

T cell Panel – Intracellular

Fluor	Marker	Dilution	Catalog	Clone	RRID
PE	IL-17	1:100	BioLegend: 506904	TC11-18H10.1	RRID: AB_315464
PE Cy7	IFN- γ	1:100	Invitrogen: 25-7311-82	XMG1.2	RRID: AB_469680
BV421	IL-10	1:100	BioLegend: 505021	JESS-16E3	RRID: AB_10900417
Pacific Blue/e450	TNF- α	1:100	BioLegend: 506318	MP6-XT22	RRID: AB_893639

Supplemental Text T2.1

Testing available flow cytometry analysis techniques

Prior to the development of this pipeline, we tested a variety of methods to analyze flow cytometry data more efficiently than with manual gating. We primarily tested two publicly available methods—t-SNE and FlowSOM.

t-SNE, as previously described in the literature review, stands for t-Distributed Stochastic Neighbor Embedding. It is capable of transforming high-dimensional data to low dimensional plots.¹ In the context of flow cytometry, it does this by identifying clusters of cells based on similarity of marker expression. While this method is useful in exploring data and identifying hidden patterns, it does not allow for automated population identification or automated sample comparisons.² Thus, it is primarily a data exploration and visualization technique. This technique is also computationally intensive; samples must first be “downsampled,” or randomly sampled, to reduce the number of data points.³ This can cause the loss of rare, but valuable populations of interest. Further, analysis and separation of data is highly dependent on the specific markers used. While t-SNE is good at separating cells in flow cytometry panels where there are not many overlapping markers (e.g., a panel to differentiate between T cells, B cells, and NK cells), it has more difficulties stratifying cells that may co-express similar markers (e.g., a panel that contains many different activation markers) (Figure S2.7).⁴ Additionally, when visualizing populations like those in Figure S2.7, the populations must be manually gated and then applied to the t-SNE. While t-SNE has utility in initial data exploration, we wanted a method that did not have to utilize manual gating, would retain our data without downsampling, and could utilize controls.

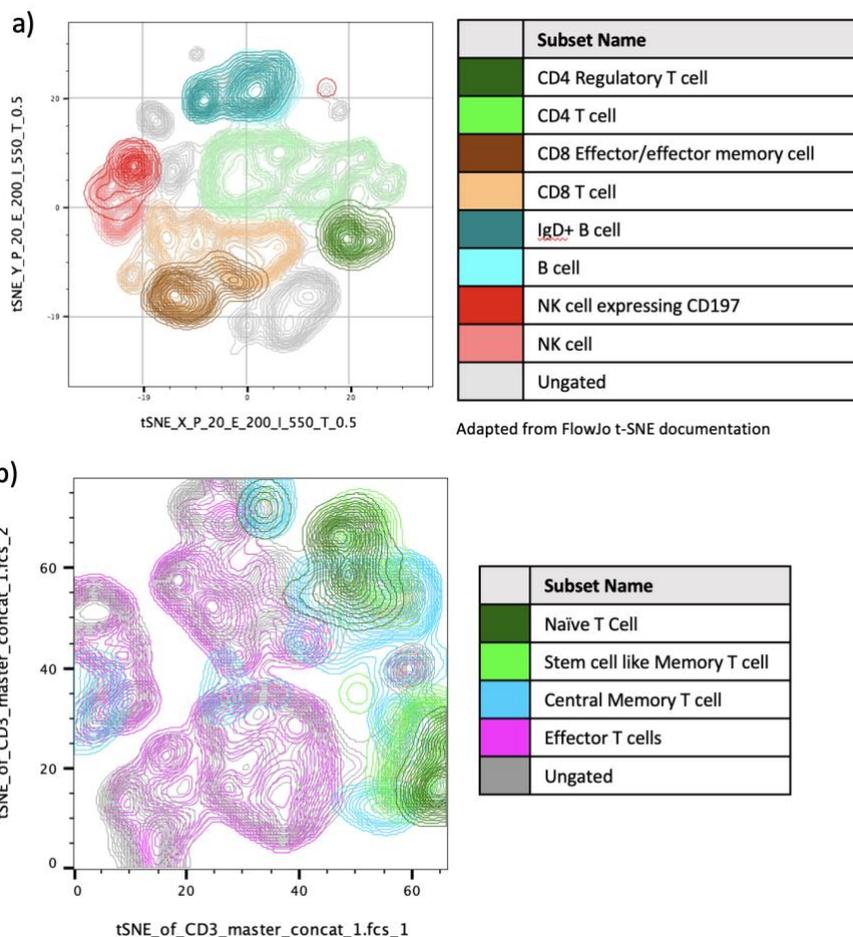


Figure S2.7: T-SNE comparison (a) Adapted from FlowJo t-SNE documentation website.⁵ t-SNE showing good separation of different immune cell populations (b) t-SNE example of bad separation of immune cell populations due to overlap of many markers.

FlowSOM (Flow Self-Organizing Map), is another technique that we tested for flow cytometry analysis. FlowSOM is a visualization aid for clustering and dimensionality reduction.⁶ It operates by training a self-organizing map (SOM) (Figure S2.8a). It starts by initializing a set number of nodes, or clusters, by pulling several random cells from the data.⁶ It then compares every subsequent cell to each of the initialized nodes.⁶ The interrogated cell will join the “nearest neighbor” node, or the node to which it is most similar, and the node phenotype will become an average of the cells within it.⁶ After the SOM has been developed, the nodes are connected in a

minimal spanning tree according to similarity between nodes.⁶ The nodes can further be clustered into “metaclusters” that exhibit similar marker expression.⁶ Because the resulting nodes from the FlowSOM are a heterogeneous mixture of cells which are described as an average of all the cells within the node, it can be difficult to interrogate specific cell types of interest or rare cells. Another issue with this method is that some markers generally have low expression on cells and others have high expression levels. This can cause specific markers to appear as though there is no expression on the cells, such as the case of PDL1 in Figure S2.8b. We tested normalization methods to scale the data so that the lowly expressed and highly expressed markers were not misrepresented. One normalization method involved scaling the data by subtracting the average expression of a particular marker and dividing by the standard deviation (Figure S2.8c). While this conveyed more information about what is considered high expression and low expression for a particular marker, the visualization still lacked important data from controls such as FMOs. We also developed an additional analysis process that collected the number of cells in each group and calculated fold changes between groups (Figure S2.8d). However, we could not perform statistical analyses on specific populations of interest. While FlowSOM offered some advantages over t-SNE in that we did not need to perform any manual gating, it still did not allow for the incorporation of FMO controls.

Both t-SNE and FlowSOM have their advantages and disadvantages. Ultimately, we wanted a method that could utilize our FMO controls to identify positive and negative expression on cells and could utilize statistical analyses to quantify differences between groups and allow for the incorporation of external data. Thus, we developed a pipeline that could fulfill these requirements.

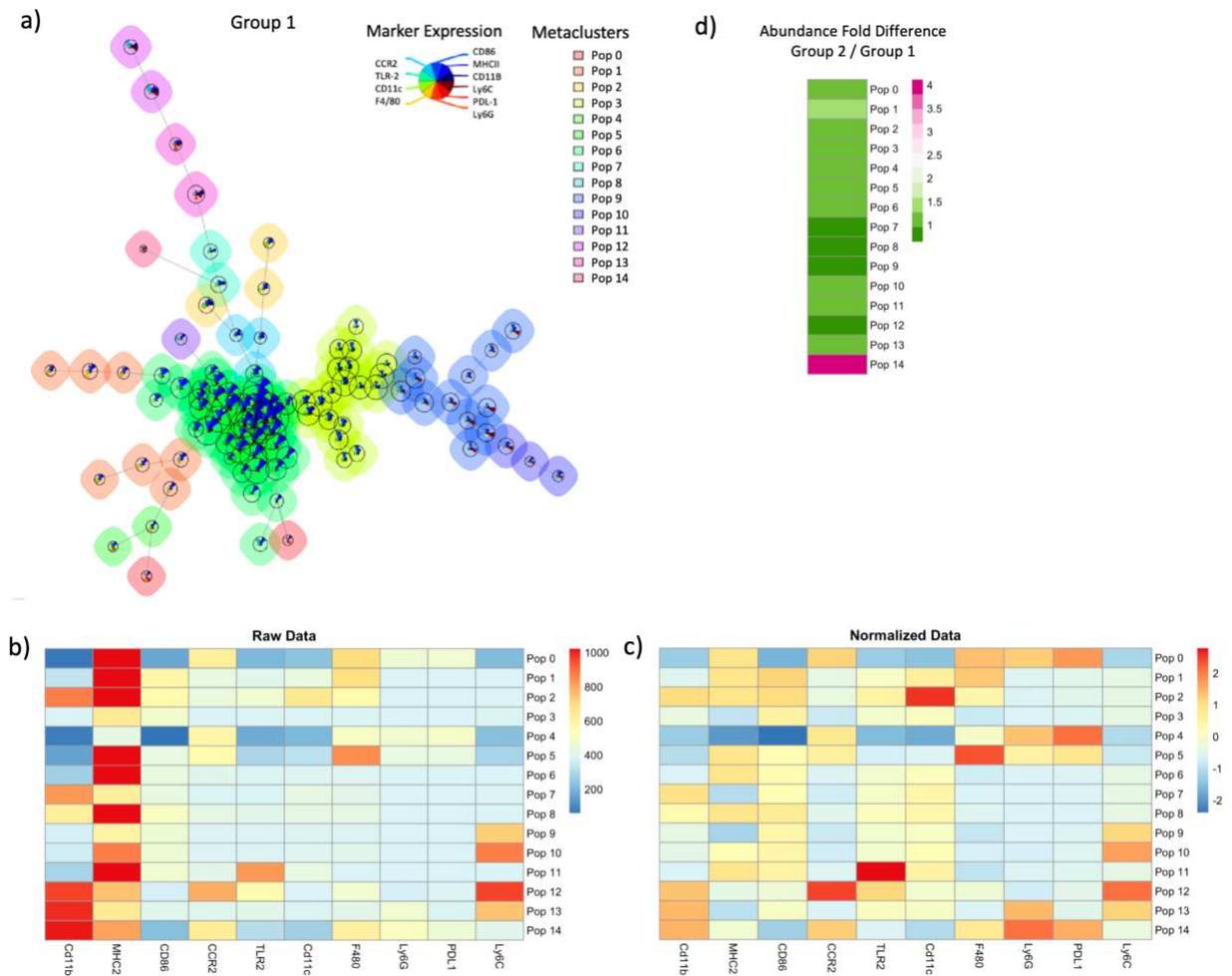
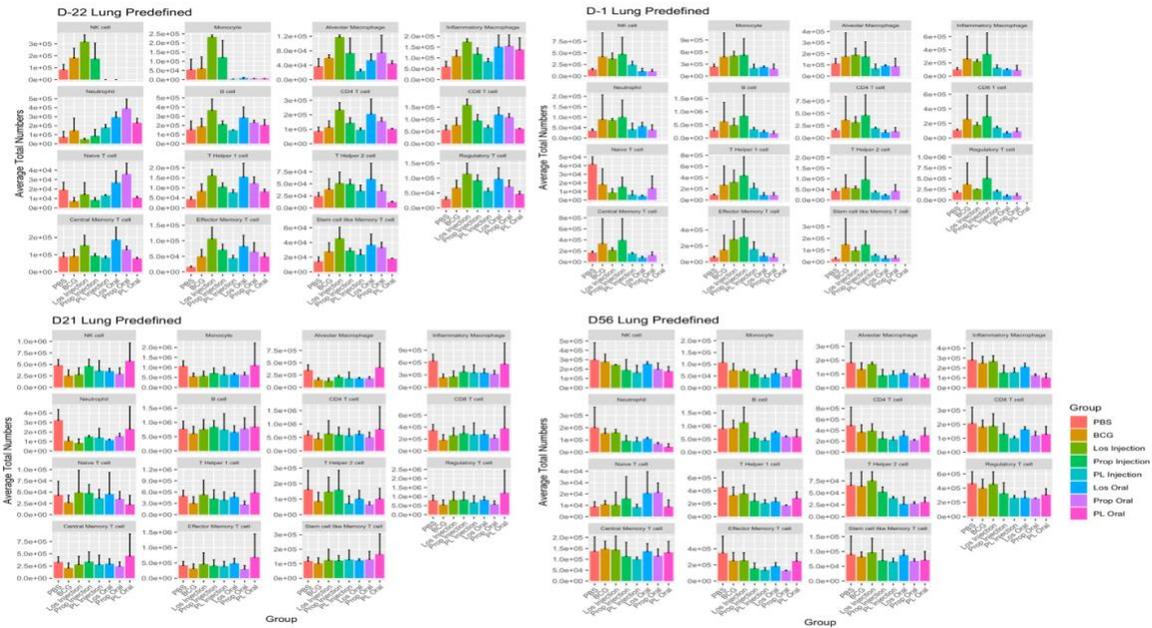


Figure S2.8: FlowSOM Testing. (a) FlowSOM map. The background color for the nodes displays the assigned metaclusters 0 through 14. Within each node, pie charts show the average marker expression of cells within the node. The size of the nodes indicates the number of cells within each cluster. (b) Heatmap of the average marker expression level in each metacluster. The x-axis displays the ten flow cytometry markers in the panel, and the y-axis shows the different metaclusters that match with those shown in Figure S2.8a. Red indicates high expression level of a particular marker and blue indicates low expression. (c) Heatmap of the normalized data for the average marker expression level in each metacluster. The data is normalized by subtracting the average expression of a particular marker and dividing by the standard deviation. (d) Heatmap of the fold change in the abundance of cells in each metacluster between two experimental groups. Pink indicates large fold differences between the groups and dark green indicates little or no differences in the number of cells between groups.

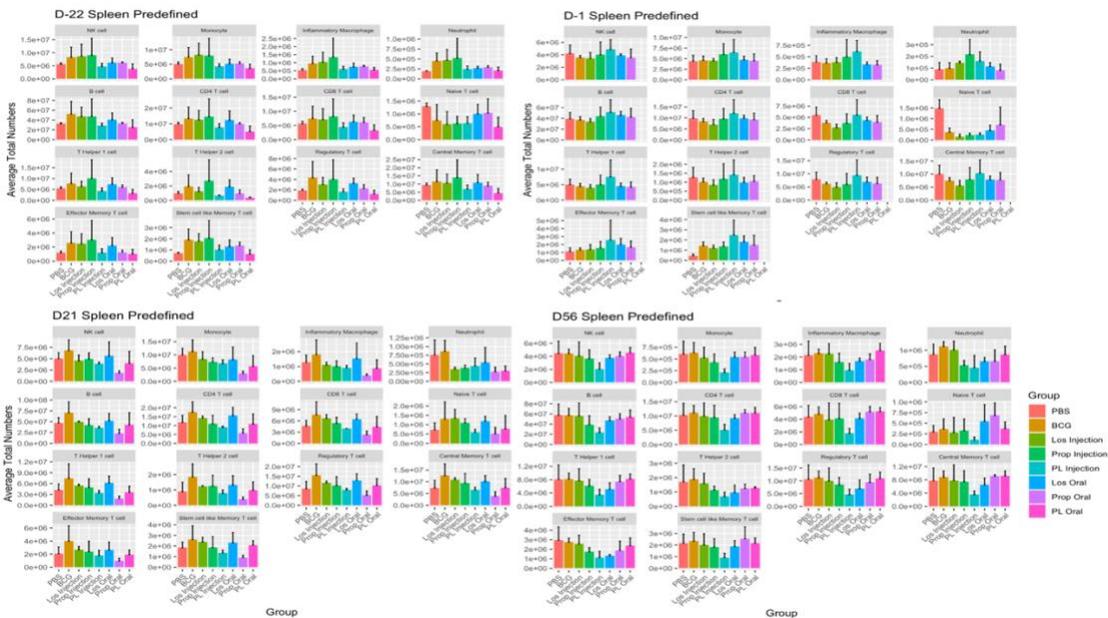
REFERENCES

1. van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, 2579-2605.
2. Mair, F., Hartmann, F.J., Mrdjen, D., Tosevski, V., Krieg, C., and Becher, B. (2016). The end of gating? An introduction to automated analysis of high dimensional cytometry data. *European Journal of Immunology* 46, 34-43. 10.1002/eji.201545774.
3. Belkina, A.C., Ciccolella, C.O., Anno, R., Halpert, R., Spidlen, J., and Snyder-Cappione, J.E. (2019). Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. *Nature Communications* 10, 5415. 10.1038/s41467-019-13055-y.
4. Lucchesi, S., Nolfi, E., Pettini, E., Pastore, G., Fiorino, F., Pozzi, G., Medaglini, D., and Ciabattini, A. (2020). Computational Analysis of Multiparametric Flow Cytometric Data to Dissect B Cell Subsets in Vaccine Studies. *Cytometry Part A* 97, 259-267. 10.1002/cyto.a.23922.
5. Crawford, T.Q. Dimensionality Reduction with the tSNE Algorithm.
6. Van Gassen, S., Callebaut, B., Van Helden, M.J., Lambrecht, B.N., Demeester, P., Dhaene, T., and Saey, Y. (2015). FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry Part A* 87A, 636-645. 10.1002/cyto.a.22625.

APPENDIX 2 — SUPPLEMENTAL MATERIALS FOR IMMUNOMODULATORY DRUGS
AS VACCINES AGAINST MYCOBACTERIUM TUBERCULOSIS



Supplemental Figure S3.1: Lung Predefined Phenotypes



Supplemental Figure S3.2: Spleen Predefined Phenotypes

Supplemental Table 3.1: Markers used to determine cell phenotypes for the predefined cells

Cell Lineage	Associated Markers
NK Cell	CD45+ NK1.1+
Monocyte	CD45+ CD14+
Alveolar Macrophage	CD45+ CD11c+ CD64+
Inflammatory Macrophage/Monocyte	CD45+ Ly6C+ CCR2+
Neutrophil	CD45+ Ly6G+ CD11b+
B cell	CD45+ CD19+
CD4 T cell	CD3+, CD4+
CD8 T cell	CD3+, CD8+
Naive T cell	CD3+ CD44- CD62L+ Sca1-
T Helper 1 cell	CD3+ Tbet+
T Helper 2 cell	CD3+ GATA3+
Regulatory T cell	CD3+ FoxP3+
Effector T cell	CD3+ CD44+ CD62L-
Central Memory T cell	CD3+ CD44+ CD62L+
Stem-cell like memory T cell	CD3+ CD44- CD62L+ Sca1+

Supplemental Table 3.2: Statistically significant differences in predefined lung predefined populations in the lungs. There were 19 statistically significant differences between the groups.

timepoint	phenotype	contrast	estimate	p.value
D-22	Alveolar Macrophage	Los Injection - BCG	57676.08	0.0139197
D-22	B cell	Los Injection - BCG	176101.65	0.0381166
D-22	CD4 T cell	Los Injection - BCG	120439.24	0.0080551
D-22	CD8 T cell	Los Injection - BCG	82398.62	0.0003900
D-22	Central Memory T cell	Los Oral - BCG	93660.81	0.0119237
D-22	Effector Memory T cell	Los Injection - BCG	56672.49	0.0099534
D-22	Monocyte	Los Injection - BCG	171748.74	0.0005774
D-22	Naive T cell	Los Oral - BCG	19893.53	0.0103080
D-22	Naive T cell	Prop Oral - BCG	29025.09	0.0002240
D-22	Neutrophil	Prop Oral - BCG	248219.81	0.0037057
D-22	NK cell	Los Oral - BCG	-182010.62	0.0308221
D-22	NK cell	PL Injection - BCG	-182022.30	0.0308108
D-22	Regulatory T cell	Los Injection - BCG	47360.48	0.0416117
D-22	Regulatory T cell	PBS - BCG	-37786.52	0.0480029
D-22	T Helper 1 cell	Los Injection - BCG	78111.20	0.0135328
D-22	T Helper 1 cell	Los Oral - BCG	71469.53	0.0260895
D-1	Naive T cell	PBS - BCG	23757.90	0.0126236
D21	Inflammatory Macrophage	PBS - BCG	403653.81	0.0115962
D21	Neutrophil	PBS - BCG	217456.93	0.0051968

Supplemental Table 3.3: Statistically significant differences in data-driven populations in the general lung analysis. There were 55 statistically significant differences between the groups.

timepoint	population	contrast	estimate	p.value	timepoint	population	contrast	estimate	p.value
D-22	Pop1	Los Oral - BCG	27218.097	0.0179385	D-22	Pop30	Los Injection - BCG	22486.5316	0.0008573
D-22	Pop2	Los Oral - BCG	70694.676	0.0143542	D-22	Pop31	Los Injection - BCG	31063.8705	0.0018847
D-22	Pop3	Los Oral - BCG	22676.803	0.0214438	D-22	Pop32	Los Injection - BCG	17544.5425	0.0056068
D-22	Pop4	Los Injection - BCG	39075.472	0.0016153	D-22	Pop33	Los Oral - BCG	190463.9022	0.0491581
D-22	Pop5	Los Injection - BCG	74106.153	0.0007596	D-22	Pop34	Los Oral - BCG	109924.6893	0.0120716
D-22	Pop6	Los Injection - BCG	33011.385	0.0058165	D-22	Pop34	PL Oral - BCG	131337.2627	0.0091418
D-22	Pop7	Los Injection - BCG	25233.977	0.0008825	D-22	Pop35	Los Injection - BCG	13150.1277	0.0108177
D-22	Pop8	Los Oral - BCG	32061.242	0.0036501	D-22	Pop36	Los Injection - BCG	3286.8979	0.0033848
D-22	Pop9	Los Oral - BCG	37855.079	0.0089232	D-22	Pop37	Los Oral - BCG	12748.9924	0.0116731
D-22	Pop9	Prop Oral - BCG	35818.125	0.0139413	D-22	Pop38	Los Injection - BCG	2769.4353	0.0229390
D-22	Pop10	Los Oral - BCG	13474.027	0.0050872	D-22	Pop40	Los Injection - BCG	719.9728	0.0400406
D-22	Pop10	Prop Oral - BCG	15249.260	0.0015383	D-22	Pop41	PL Oral - BCG	8509.3708	0.0097676
D-22	Pop11	Los Oral - BCG	21344.587	0.0014780	D-22	Pop42	Los Injection - BCG	2067.9605	0.0122264
D-22	Pop11	Prop Oral - BCG	17905.580	0.0077521	D-22	Pop43	Los Injection - BCG	6980.3427	0.0031814
D-22	Pop13	Los Injection - BCG	173638.888	0.0006099	D-1	Pop12	Prop Injection - BCG	109.8529	0.0220983
D-22	Pop14	Los Injection - BCG	7094.273	0.0148053	D21	Pop1	PBS - BCG	6282.3646	0.0000100
D-22	Pop15	Los Injection - BCG	8243.750	0.0000919	D21	Pop3	PBS - BCG	12387.8095	0.0001353
D-22	Pop17	Los Injection - BCG	9112.841	0.0001914	D21	Pop10	PBS - BCG	1823.8298	0.0311425
D-22	Pop18	Los Injection - BCG	51147.678	0.0002720	D21	Pop11	PBS - BCG	1718.6541	0.0051794
D-22	Pop20	Los Injection - BCG	3146.334	0.0084856	D21	Pop15	PL Oral - BCG	34191.2902	0.0352691
D-22	Pop21	Los Injection - BCG	4575.520	0.0021931	D21	Pop20	PBS - BCG	40699.4934	0.0013535
D-22	Pop22	Los Injection - BCG	6103.662	0.0004999	D21	Pop21	PBS - BCG	32664.4880	0.0153534
D-22	Pop23	Los Injection - BCG	4048.956	0.0060839	D21	Pop23	PBS - BCG	57676.7875	0.0073604
D-22	Pop24	Los Oral - BCG	68342.228	0.0249686	D21	Pop33	PL Oral - BCG	196071.9712	0.0024963
D-22	Pop24	Prop Oral - BCG	97222.051	0.0011248	D21	Pop42	PBS - BCG	8476.9136	0.0009431
D-22	Pop26	Prop Oral - BCG	64388.206	0.0074616	D21	Pop43	PBS - BCG	19864.9420	0.0036708
D-22	Pop29	Los Injection - BCG	36531.405	0.0355178	D21	Pop43	PL Oral - BCG	24526.1798	0.0033992
					D56	Pop36	PL Oral - BCG	103444.6628	0.0005359

Supplemental Table 3.4: Statistically significant differences in data-driven populations in the T cell lung panel. There were 21 statistically significant differences between the groups.

timepoint	population	contrast	estimate	p.value
D-22	Pop2	Prop Injection - BCG	641.2841	0.0048832
D-22	Pop4	Los Oral - BCG	12385.3731	0.0247858
D-22	Pop5	Los Injection - BCG	7599.9963	0.0436810
D-22	Pop8	Los Injection - BCG	3463.3650	0.0387341
D-22	Pop9	Los Oral - BCG	11118.6345	0.0145350
D-22	Pop14	Los Oral - BCG	12217.5936	0.0100453
D-22	Pop14	Prop Oral - BCG	16308.7041	0.0006093
D-22	Pop15	Los Oral - BCG	8098.3230	0.0065058
D-22	Pop15	PBS - BCG	5053.6900	0.0473656
D-22	Pop15	Prop Oral - BCG	6370.9111	0.0395621
D-22	Pop17	Los Injection - BCG	8760.8728	0.0261881
D-22	Pop18	Los Injection - BCG	7196.5113	0.0001915
D-22	Pop18	PBS - BCG	-3654.6950	0.0197907
D-22	Pop19	Los Injection - BCG	1093.1503	0.0213512
D-22	Pop19	Los Oral - BCG	1034.0393	0.0314360
D-22	Pop20	Los Injection - BCG	2406.0327	0.0209235
D-22	Pop22	PBS - BCG	-371.8116	0.0149271
D-1	Pop14	PBS - BCG	6007.0099	0.0030786
D-1	Pop21	PBS - BCG	5908.9944	0.0297298
D21	Pop2	PL Oral - BCG	4546.5744	0.0181400
D56	Pop14	Los Oral - BCG	3794.9672	0.0317220

Supplemental Table 3.5: Statistically significant differences in predefined populations in the spleen. There were 17 statistically significant differences between the groups.

timepoint	phenotype	contrast	estimate	p.value
D-1	Naive T cell	PBS - BCG	1117130.3	0.0001679
D21	NK cell	Prop Oral - BCG	-4939185.2	0.0077998
D21	Monocyte	Prop Oral - BCG	-8247643.8	0.0108950
D21	Inflammatory Macrophage	Prop Oral - BCG	-1403474.8	0.0356598
D21	B cell	Prop Oral - BCG	-47363651.6	0.0061469
D21	CD4 T cell	Prop Oral - BCG	-11772957.8	0.0057311
D21	CD8 T cell	Prop Oral - BCG	-5321238.4	0.0132010
D21	T Helper 1 cell	Prop Oral - BCG	-5615763.4	0.0175036
D21	T Helper 2 cell	Prop Oral - BCG	-1465315.3	0.0139054
D21	Regulatory T cell	Prop Oral - BCG	-10270485.3	0.0093691
D21	Central Memory T cell	Prop Oral - BCG	-8775394.9	0.0104118
D21	Effector Memory T cell	Prop Oral - BCG	-3007964.1	0.0392468
D21	Stem cell like Memory T cell	Prop Oral - BCG	-1704607.5	0.0452400
D56	Monocyte	PL Injection - BCG	-418662.0	0.0430695
D56	Neutrophil	PL Injection - BCG	-661325.7	0.0106824
D56	Neutrophil	Prop Injection - BCG	-584598.1	0.0271947
D56	B cell	PL Injection - BCG	-33723980.4	0.0484040

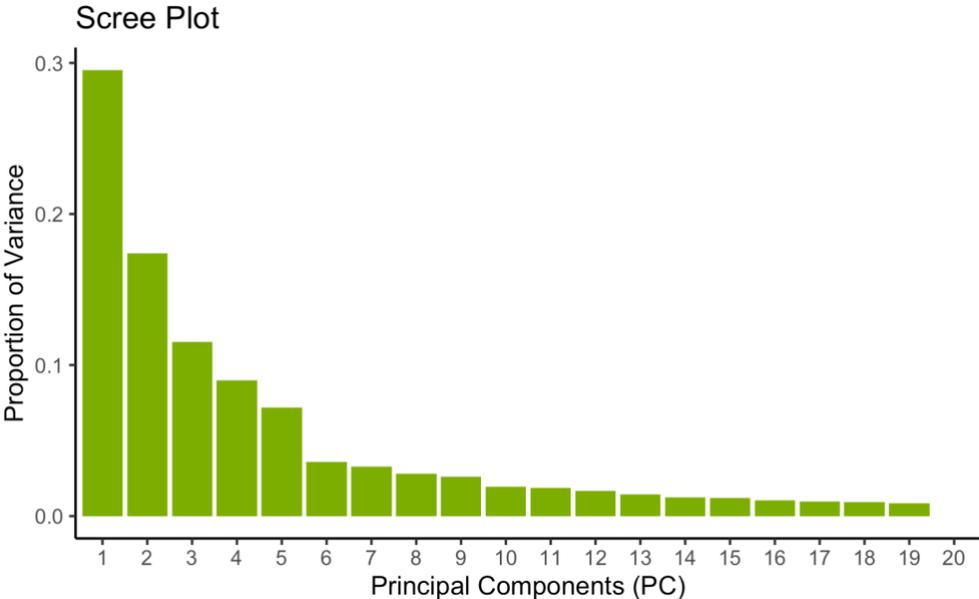
Supplemental Table 3.6: Statistically significant differences in data-driven populations in the general spleen analysis. There were 21 statistically significant differences between the groups.

timepoint	population	contrast	estimate	p.value
D-1	Pop2	PBS - BCG	1784015.65	0.0490991
D21	Pop3	Prop Oral - BCG	-6289107.84	0.0134258
D21	Pop5	Prop Oral - BCG	-234120.33	0.0074349
D21	Pop6	Prop Oral - BCG	-238578.85	0.0323409
D21	Pop8	Prop Oral - BCG	-31994971.09	0.0080520
D21	Pop9	Prop Oral - BCG	-701525.93	0.0352826
D21	Pop10	Prop Oral - BCG	-1579609.42	0.0285788
D21	Pop11	Prop Oral - BCG	-478681.09	0.0157345
D21	Pop12	Los Injection - BCG	-1188798.76	0.0267256
D21	Pop12	PL Injection - BCG	-1232945.47	0.0206326
D21	Pop12	PL Oral - BCG	-1178175.13	0.0284275
D21	Pop12	Prop Injection - BCG	-1080288.15	0.0496460
D21	Pop12	Prop Oral - BCG	-1591622.38	0.0023164
D21	Pop14	Prop Oral - BCG	-362504.39	0.0272533
D21	Pop15	Prop Oral - BCG	-239451.15	0.0440247
D21	Pop18	Prop Oral - BCG	-314847.20	0.0085535
D21	Pop21	Prop Oral - BCG	-274634.76	0.0299055
D56	Pop8	PL Injection - BCG	-25765844.50	0.0459158
D56	Pop10	PL Injection - BCG	-49293.20	0.0266366
D56	Pop11	PL Injection - BCG	-70962.78	0.0160837
D56	Pop14	PL Injection - BCG	-1274825.70	0.0217121

Supplemental Table 3.7: Statistically significant differences in data-driven populations in the T cell spleen panel. There were 17 statistically significant differences between the groups.

timepoint	population	contrast	estimate	p.value
D-22	Pop5	PBS - BCG	-18331.99	0.0238227
D-22	Pop12	PBS - BCG	-475739.08	0.0471049
D-1	Pop8	PBS - BCG	201919.48	0.0006439
D-1	Pop9	PBS - BCG	388749.48	0.0009834
D-1	Pop10	PBS - BCG	518532.07	0.0000902
D-1	Pop11	PBS - BCG	319541.47	0.0001655
D-1	Pop12	PBS - BCG	-148989.06	0.0350230
D21	Pop2	Prop Oral - BCG	-160011.74	0.0491167
D21	Pop3	Prop Oral - BCG	-679930.62	0.0162524
D21	Pop4	Prop Oral - BCG	-1701509.90	0.0261104
D21	Pop6	PL Oral - BCG	-27874.56	0.0069552
D21	Pop6	Prop Oral - BCG	-26881.60	0.0094542
D21	Pop7	PL Oral - BCG	-85066.32	0.0166553
D21	Pop7	Prop Oral - BCG	-86659.06	0.0144519
D56	Pop2	Prop Oral - BCG	300059.48	0.0456795
D56	Pop8	Prop Oral - BCG	84138.24	0.0100535
D56	Pop11	Prop Oral - BCG	51844.93	0.0229724

APPENDIX 3 — SUPPLEMENTAL MATERIALS FOR NOVEL TECHNIQUE TO STUDY IMMUNOMETABOLISM DURING MYCOBACTERIUM TUBERCULOSIS INFECTION



Supplemental Figure S4.1: Scree plot displaying the proportion of variance accounted for by each of the 19 principal components in the principal components analysis.

Supplemental Table 4.1: Summary of Arachidonic acid metabolites and associated KEGG IDs.

Summary of Arachidonic Acid Metabolites and Associated KEGG ID

Name	KEGG ID
11-epi-Prostaglandin F2alpha	C05959
11,12-DHET	C14774
11,12-EET	C14770
11,12,15-THETA	C14782
11,14,15-THETA	C14814
11H-14,15-EETA	C14813
12(R)-HPETE	C14812
12(S)-HPETE	C05965
14,15-DHET	C14775
14,15-EET	C14771
15(S)-HETE	C04742
15(S)-HPETE	C05966
15H-11,12-EETA	C14781
19(S)-HETE	C14749
20-HETE	C14748
20-OH-Leukotriene B4	C04853
5,6-DHET	C14772
5,6-EET	C14768
5(S)-HETE	C04805
5(S)-HPETE	C05356
8,9-DHET	C14773
8,9-EET	C14769
8(S)-HPETE	C14823
Arachidonic Acid	C00219
Leukotriene A4	C00909
Leukotriene B4	C02165
Leukotriene C4	C02166
Leukotriene D4	C05951
Phosphatidylcholine	C00157
Prostaglandin D2	C00696
Prostaglandin E2	C00584
Prostaglandin F2alpha	C00639
Prostaglandin G2	C05956
Prostaglandin H2	C00427
Prostaglandin I2	C01312
Thromboxane A2	C02198

APPENDIX 4 — ACQUISITION OF HIGH-QUALITY SPECTRAL FLOW CYTOMETRY DATA[#]

Significance Statement

With new high-throughput flow cytometry, data analysis has become highly complex. Using open-source software, it is now possible to explore these large datasets, simplifying the seemingly complex data. However, to perform these analyses, sample preparation, staining procedure, and use of controls must follow rigorous protocols. In this Current Protocols article, we describe the best practices for preparation and acquisition of spectral flow cytometry samples. Following this protocol will lead to clean results that can be used with the cyto-feature engineering data analysis pipeline described previously.

Abstract

Flow cytometry allows for the visualization of physical, functional, and/or biological properties of cells including antigens, cytokines, size, and complexity. With increasingly large flow cytometry panels able to analyze up to 50 parameters, there is a need to standardize flow cytometry protocols to achieve high quality data that can be input to analysis algorithms. Without this clean data, algorithms may incorrectly categorize the cell populations present in the samples. In this protocol, we outline a comprehensive methodology to prepare samples for polychromatic flow cytometry. The use of multiple washing steps and rigorous controls creates high-quality data with good separation between cell populations. Experimental data acquired

[#] This appendix has been published in Current Protocols of Cytometry: Fox, A., Dutt, T. S., Karger, B., Obregón-Henao, A., Anderson, G. B., & Henao-Tamayo, M. (2020). Acquisition of high-quality spectral flow cytometry data. Current Protocols in Cytometry, 93, e74. doi: 10.1002/cpcy.74.

using this protocol can be analyzed via computational algorithms that perform end-to-end analysis.

Basic Protocol 1: Preparation of single-cell suspension for flow cytometry

Support Protocol 1: Lung preparation

Support Protocol 2: Counting cells on a flow cytometer

Basic Protocol 2: Surface and intracellular flow cytometry staining

Support Protocol 3: Single-color bead controls

Introduction

High-dimensional flow cytometry data, containing excess of 15 parameters, is difficult to analyze using conventional analysis methods such as manual gating of cells on 2-dimensional plots. In the past decade, researchers have worked to develop data analysis tools for flow cytometry. However, data input must be reliable for these tools to accurately analyze it. By executing this current protocol, it will be possible to acquire clean flow cytometry data that can be input into data analysis pipelines. With increasingly complex flow cytometry development, it is important to acquire data following a very strict and reproducible flow cytometry staining procedure to ensure high quality data prior to analysis. By the end of this protocol, the high-quality flow cytometry data and controls are ready to be entered into and analyzed using a cyto-feature engineering pipeline.¹

Basic Protocol 1 describes how to prepare single-cell suspension for flow cytometry after harvesting either lung or spleen from an animal. This protocol can further be modified to harvest

cells from other organs. Basic Protocol 2 describes how to discriminate between live and dead cells using a viability dye and stain cells with surface and intracellular markers.

Strategic planning

Prior to employing this protocol for an experiment, the flow cytometry panel should be optimized for the user's flow cytometer to ensure there is no overlap between different fluorophores. Users should familiarize themselves with their flow cytometer configuration and the possible combination of fluorophores that can be used with that cytometer. Panel design will be based on this configuration and possible fluorophores. Excellent resources for panel design include the website FluoroFinder and a recent publication in Current Protocols.^{2,3}

BASIC PROTOCOL 1: Preparation of single-cell suspension for flow cytometry

Introductory paragraph

In this protocol, tissues are dissociated into single-cell suspension in preparation for flow cytometry. Whole tissue is macerated through a cell strainer after which red blood cells are lysed. Note this protocol should occur in a biosafety cabinet to prevent sample contamination and ensure technician safety.

Reagents, Solutions, and Test Animals

- Lung of mouse
- Spleen of mouse
- DMEM 1x with 4.5g/L glucose without L-glutamine, sodium pyruvate (Corning, cat# 15-017-CV)

- Red Blood Cell (RBC) Lysing Buffer (Sigma, cat# R7757)
- Complete media (see recipe)
- Phosphate Buffered Saline (VWR, cat. No. 45000-446)

Hardware and Instruments

- Biosafety cabinet (e.g., Labconco Type A2)
- 70 μ m Cell Strainers (Corning, cat# 352350)
- 60 x 15 mm Petri Dishes (Thermo Fisher Scientific, cat# AS4051)
- 3 mL LuerLok Syringes (BD Biosciences, cat# 309657)
- 5 mL serological pipets (Thermo Fisher Scientific, cat# 170355)
- Motorized Serological Pipette Filler (SCIOLOGEX, cat# 740200029999)
- 15 mL Conical Tubes (Thermo Fisher Scientific, cat# 12565269)
- Tabletop Centrifuge (e.g., Beckman Coulter Allegra 6)
- 1 mL Pipettor (Sigma Aldrich, cat# EP3124000121)
- 1 mL Pipet Tips (VWR, cat# 83007-382)
- 10 mL serological pipets (Thermo Fisher Scientific, cat# 170356)
- Flow cytometer (e.g., BD LSR-II)
- 200 μ L pipettor (Sigma Aldrich, cat# EP3124000083)
- 200 μ L pipet tips (VWR, cat# 53508-810)
- 96-well plates with V-bottom (Sigma Aldrich, cat# M9686)
- Paper towels (Supply Works, cat# SCAHB9201)

Protocol steps

1. Harvest spleen and lung into 1 mL of DMEM media in a 15 mL conical tube.

Organs may be required for multiple experimental tests. In this case, we take ½ spleen and two of the lung lobes for flow cytometry. The remaining lung lobes and spleen are utilized for enumerating colony forming units or other downstream applications. However, this could be modified according to the experimental design and goal.

For lungs, follow Supplemental Protocol 1 prior to step 2 below.

2. Add organ and liquid into 70 µm cell strainer placed inside a petri dish.
3. Macerate organs using 3 mL plunger in order to pass cells through the strainer mesh.

Remove the 3 mL plunger from the 3 mL syringe and use only the non-rubber side of the plunger.

4. Flush cell strainer mesh with 5 mL of DMEM media and harvest cells from the Petri Dish after resuspending. Transfer cells back into the 15 ml conical tube.
5. Centrifuge cells at 380 G-force for 10 min at 4°C.

The RPM (revolutions per minute) for specific centrifuges can be calculated with the following equation where g is the g-force and Radius is the radius of the rotor in centimeters.

$$g = (1.118 \times 10^{-5}) \times \text{Radius} \times \text{RPM}^2$$

6. Discard supernatant and resuspend cell pellet in 1 mL RBC lysing buffer – incubate for 1 minute at room temperature.

During the incubation time, cells should be mixed by running the 15 mL conical tubes against the grate of the biosafety cabinet, disrupting the pellet.

7. Add 6 mL of complete media to dilute RBC lysing buffer.
8. Centrifuge cells at 380 G-force for 10 min at 4°C.
9. Discard supernatant by gently flipping tube upside down.
10. Resuspend cells in PBS and keep at 4°C while counting.

Resuspend spleens in 800 µl of PBS and lungs in 400 µL of PBS.

11. Count cells in each tube using a flow cytometer (See Supplemental Protocol 2).

Alternatively, cells could be counted with a hemocytometer or cell counter, but this method is more time consuming.

12. Adjust cell suspension to 2×10^7 cells/mL for spleen and 5×10^6 cells/mL for lung.
13. Add 200 µL of lung cells (1×10^6 total cells) or 100 µl of spleen cells (2×10^6 total cells) to each well of a 96-well plate (Figure A.1).

There should be a well for every stained sample, a well for every Fluorescence Minus One control (FMO), a single well that will not be stained (Unstained), and a single well that will only be stained with live-dead (Zombie-NIR) stain. Single-color controls will also need to be stained, but beads can be used for these controls (see support protocol 3). Due to different levels of autofluorescence and viability, each organ requires the following controls: FMOs, unstained cells, and live-dead staining. To run these controls for each organ, cells can be pooled from different experimental groups or biological replicates within a group.

14. Centrifuge plates at 380 G-force for 10 min at 4°C.

At this point, it should be possible to see a pellet of cells at the bottom of each well.

15. Remove supernatant by gently tapping the liquid out of the wells onto a paper towel.

16. Wash cells by adding 100 μ L of PBS to each well and centrifuge at 380 G-force for 10 minutes at 4°C.

This wash minimizes any residual protein in the sample, which can quench the Zombie NIR viability stain.

17. Remove supernatant by gently tapping the liquid out of the wells onto a paper towel.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Lung Sample											
B												
C	FMO Ly-6A	FMO CD3	FMO CD62L	FMO CD122	FMO CD28	FMO PD-1	FMO CD103	FMO CD44	FMO CD4	FMO CD8	FMO CD152	FMO CD27
D	FMO CD153	FMO KLRG-1	FMO IL-17A	FMO IFN- γ	FMO IL-10	FMO TNF- α						
E												
F												
G												
H										Unstained		Zombie-NIR

Figure A.1: Cell plate layout with controls. Example of the required cell samples for flow cytometry. Note that there will be a second plate with beads used for single-color controls

SUPPORT PROTOCOL 1: Lung Preparation

Introductory paragraph

To maximize cell yield and viability, lungs must first be enzymatically digested before dissociation and cell suspension. Collagen in the lungs' extracellular matrix will be digested to improve cell dissociation when macerating through a 70 μ m cell strainer. Inevitably, some cells

lyse and release DNA during this procedure. DNase is included in the dissociation buffer to digest released DNA and minimize the formation of a viscous material trapping the cells. If this protocol is conducted properly, single-cell suspensions with a cell viability >70-80% will be obtained from the lungs.

Reagents, Solutions, and Test Animals

- Lung sample
- 70% ethanol
- DMEM 1x with 4.5g/L glucose without L-glutamine, sodium pyruvate (Corning, cat# 15-017-CV)
- 2x DNase/Liberase (see recipe)

Hardware and Instruments

- Biosafety cabinet (e.g., Labconco Type A2)
- 60 x 15 mm Drosophila Supplies Small Petri Dishes (Thermo Fisher Scientific, cat# AS4051)
- Razor blades (VWR, cat # 55411-050)
- 37°C incubator

Protocol steps

1. Make superficial cuts in the lung in a petri dish using 2 razors (rinse razors with 70% ethanol and then DMEM).

Use one razor to hold tissue in place and the other to chop. Move tissue and chop in a crosshatch pattern. Be careful not to chop tissue too deep—the tissue needs to be kept whole.

2. Transfer lungs and liquid back into the 15 mL conical tubes.
3. Add 1 mL of 2x DNase/Liberase to each tube and incubate at 37°C for 30 minutes.

The final concentration per sample of Liberase is 0.25 mg/mL and the concentration per sample of DNase is 0.125 mg/mL.

SUPPORT PROTOCOL 2: Counting Cells on Flow Cytometer

Introductory paragraph

To calculate the total number of cells obtained from an organ, the cells must be counted. Unless cell death is a readout, dead cells are usually excluded from this analysis as dead cells tend to be more autofluorescent and bind antibodies nonspecifically. Dead cells are stained with a viability dye such as 7-AAD, which binds to DNA in dead cells with compromised cell membranes.

Therefore, live cells are those in which there is no 7-AAD present. Fluorescent counting beads of known concentration are added to the samples to determine the ratio of beads to live cells. For convenience, two different viability dyes (7-AAD and Zombie-NIR) were used in this protocol: 7-AAD to count cells on the cytometer and Zombie-NIR to analyze samples after surface/intracellular staining. This staining approach will be required when counting cells using equipment that does not have a laser to detect Zombie-NIR. One viability dye could, however, be used for both procedures upon further optimization. One advantage of using 7-AAD for the counting procedure is that it allows for quicker counting analysis as it can directly be added to

flow tubes without washing steps. If this protocol is conducted properly, the concentration of live leukocytes in each sample can be calculated.

Reagents, Solutions, and Test Animals

- FACS Staining Buffer (see recipe)
- 7-AAD Viability Staining (Thermo Fisher, cat# 00-6993-50)
- CountBright Absolute Counting beads (Thermo Fisher, cat# C36950)

Hardware and Instruments

- Biosafety cabinet (e.g., Labconco Type A2)
- 5 mL Polypropylene round bottom flow tubes (VWR, cat# 60819-794)
- Flow cytometer (e.g., LSR II)

Protocol steps

1. For each sample, add 200 μ L of FACS Staining Buffer with 0.025 μ g of 7-AAD to a flow tube.
2. Add 25 μ L of counting beads and 25 μ L of cells to each tube from step 1.
3. Collect 50,000 events on the flow cytometer.

Ensure gains (voltages) for FSC and SSC allow for both the beads and cells to be seen on the plots (See Figure A.2). This can be visualized by setting the SSC-A axis to log-scale.

4. Place a gate around the beads and count the number of beads as shown in Figure A.2.
5. Gate the cells based on live leukocytes as shown in Figure A.2.

6. Calculate the volume of PBS to add to each sample based on the desired cell concentration using the following calculations. An example is provided in Table A.1.
- a. **Sample volume:** Volume in which the cells are suspended (e.g., 400 μL of PBS for lungs or 800 μL of PBS for spleen)
 - b. **Bead concentration/ μL :** Concentration of beads per μL (this information is written on the vial of beads. Note that bead concentration may vary between different lots and should be modified accordingly.)
 - c. **# Beads:** Number of beads based on gating (see Figure A.2)
 - d. **# Live Leukocytes:** Number of live leukocytes based on gating (see Figure A.2)
 - e. **Acquired Volume of Beads (μL):** Calculated volume based on dividing the # Beads by the Bead Concentration/ μL
 - f. **Total # Live Leukocytes:** Calculated by multiplying the sample volume by the # Live Leukocytes and dividing by acquired volume of beads. (As beads and cells were added at a 1:1 ratio (step 2), the volume of acquired beads and acquired cells should be the same)
 - g. **Desired # cells/mL:** Desired concentration of cells per mL for the experiment
 - h. **Volume to Add (mL):** Calculated by dividing the Total # Live Leukocytes/mL by desired # of cells/mL and subtracting the sample volume

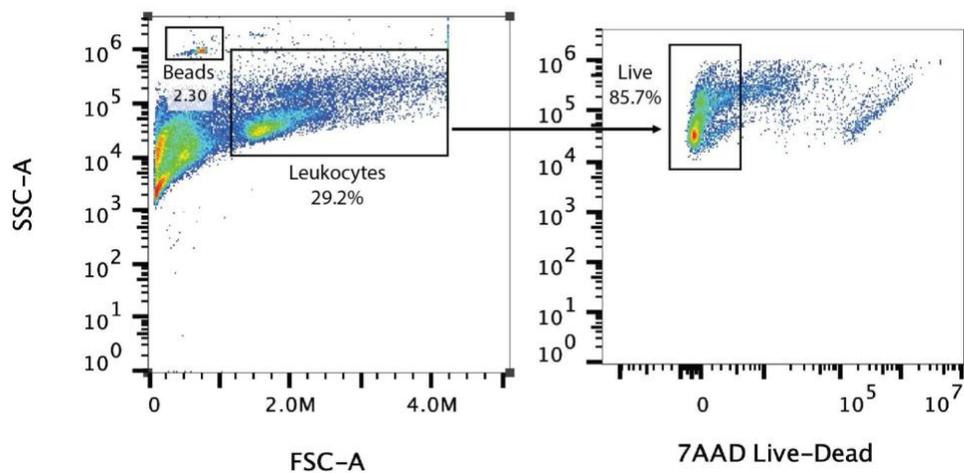


Figure A.2: Cell counting gating strategy. This figure shows placement of the bead gate and the leukocyte gate following cell counting on a flow cytometer. The leukocytes are further gated to limit cells to the live leukocytes.

Table A.1: Example of Calculations for determining the number of live cells in a sample.

Sample ID	Sample volume (μL)	Bead concentration/ μL	# Beads	# Live Leukocytes	Acquired Volume of Beads	Total # Live Leukocytes	Desired # Cells/mL	Volume to Add (mL)
Spleen	800	1040	590	28,933	0.57	4.08E7	2E7	1.24
Lung	400	1040	2,519	21,942	2.42	3.62E6	5E6	0.32

BASIC PROTOCOL 2: Surface and Intracellular Flow Cytometry Staining

Introductory paragraph

In this protocol, an initial 6-hour incubation with a Protein Transport Inhibitor prevents intracellular proteins from being secreted, causing the accumulation of such proteins in cells. This allows for the intracellular proteins to be stained. Following this incubation, anti-mouse CD16/32 is added to block FC-receptors on leukocytes, which prevents non-specific antibody binding. Dead cells are then stained using the viability dye, Zombie-NIR, and surface antibodies are added. To stain for intracellular markers, fixation/permeabilization buffer is added to allow antibodies to pass through the plasma membrane. If this protocol is conducted properly, cells stained with the fluorophore-marker pairs can be visualized on a flow cytometer.

If the panel only contains surface markers, the 6-hour incubation with a Protein Transport Inhibitor and steps with Permeabilization/Fixation can be skipped. If so, cells are simply blocked with anti-CD16/32, stained with Zombie-NIR, and then stained with surface antibodies. Thereafter, cells are washed with FACS Staining Buffer and fixed by incubating with 4% PFA for 20 minutes.

Reagents, Solutions, and Test Animals

- 1000X Protein Transport Inhibitor [BD GolgiStop] (BD Biosciences, cat# 554724)
- Complete Media (see recipe)
- Phosphate Buffered Saline (VWR, cat# 45000-446)
- Zombie-NIR Fixable Viability Kit (VWR, cat# 10761-492)
- FACS Staining Buffer (see recipe)

- Anti-mouse CD16/32 (Biolegend, cat# 101330)
- T cell surface antibody cocktail (see recipe)
- FMOs (see recipe)
- Perm/Fix buffer (see recipe)
- 1X Permeabilization buffer (see recipe)
- T cell intracellular antibody cocktail (see recipe)

Hardware and Instruments

- Biosafety cabinet (e.g., Labconco Type A2)
- 10 μ L pipettor (Sigma Aldrich, cat# EP3124000024)
- 10 μ L pipet tips (Thermo Fisher Scientific, cat# 2707454)
- Motorized Serological Pipette Filler (SCIOLOGEX, cat# 740200029999)
- 5 mL serological pipets (Thermo Fisher Scientific, cat# 170355)
- 37°C CO₂ incubator (e.g., VWR water-jacketed CO₂ incubator)
- Tabletop Centrifuge (e.g., Beckman Coulter Allegra 6)
- Paper towels (Supply Works, cat# SCAHB9201)
- 200 μ L pipettor (Sigma Aldrich, cat# EP3124000083)
- 200 μ L pipet tips (VWR, cat# 53508-810)
- 5 mL Polypropylene round bottom flow tubes (VWR, cat# 60819-794) [note that polystyrene tubes may need to be used depending on the cytometer]

Protocol steps

1. Prepare a 1X solution of Protein Transport Inhibitor (GolgiStop) in Complete Media.

2. Add 100 μ l of above prepared media with the Protein Transport Inhibitor to each well on the 96-well plate containing pelleted cells.
3. Incubate plate in a CO₂ incubator set at 37°C for 6 hours.
4. Remove plate from incubator and centrifuge at 380 G-force for 10 minutes at 4°C.

All centrifugations will be at 380 G-force for 10 min at 4°C.

5. Remove supernatant by gently tapping the liquid out of the wells onto a paper towel.
6. Wash cells by adding 100 μ L of PBS to each well and centrifuging at 380 G-force for 10 minutes at 4°C.
7. Remove supernatant by gently tapping all of the liquid out of the wells onto a paper towel, being careful not to disrupt the pellet.
8. Add 100 μ l of Zombie-NIR live/dead stain (1:2000) dilution in PBS to each well except for the “Unstained Sample well” and incubate for 15 minutes at room temperature in the dark.

Zombie-NIR must be diluted in PBS rather than FACS Staining Buffer because the FACS Staining Buffer contains Fetal Bovine Serum that quenches the Zombie signal. Note that all further steps should take place in the dark so that the fluorescently labeled antibodies are not photobleached.

9. Centrifuge cells at 380 G-force for 10 minutes at 4°C and remove the supernatant.
10. Wash cells by adding 100 μ L of FACS Staining Buffer and centrifuge at 380 G-force for 10 minutes at 4°C.
11. Remove supernatant.
12. Wash cells again by adding 100 μ L of FACS Staining Buffer and centrifuge at 380 G-force for 10 minutes at 4°C.

13. Remove supernatant.
14. Incubate cells with FACS Staining Buffer containing 2.5 µg/mL of anti-mouse CD16/32 for 20 minutes at 4°C.

This antibody blocks non-specific FC receptor binding.

15. Centrifuge cells at 380 G-force for 10 minutes at 4°C and remove the supernatant.
16. Add the T cell surface marker antibody panel to all the sample wells. Also add the appropriate surface marker FMOs to the wells.

Do not add intracellular antibodies at this point.

17. Incubate cells at 4°C for 30 minutes in the dark.
18. Centrifuge cells at 380 G-force for 10 minutes at 4°C and remove the supernatant.
19. Wash cells by adding 150 µL of FACS Staining Buffer and centrifuge at 380 G-force for 10 minutes at 4°C.
20. Remove supernatant.
21. Add 150 µl of 1X Perm/Fix buffer and incubate for 1 hour at room temperature.
22. Centrifuge the cells at 380 G-force for 10 minutes at 4°C and wash the cells with 150 µl of 1X Permeabilization Buffer.
23. Remove supernatant.
24. Centrifuge cells at 380 G-force for 10 minutes at 4°C and wash cells with 150 µl of Permeabilization Buffer again.
25. Remove supernatant.
26. Add 100 µl of the T cell intracellular antibody cocktail and intracellular FMO antibodies cocktail to respective wells.
27. Incubate plate overnight at 4°C in dark.

The Next Day

1. Centrifuge cells at 380 G-force for 10 minutes at 4°C.
2. Wash cells by adding 150 μ l of Permeabilization Buffer and centrifuge at 380 G-force for 10 minutes at 4°C.
3. Remove supernatant.
4. Suspend cells in 100 μ L of Permeabilization Buffer.
5. Transfer cells to flow cytometry tubes that contain an additional 200 μ L of Permeabilization Buffer.
6. Read samples on a flow cytometer.

SUPPORT PROTOCOL 3: Single-Color Bead Controls

Introductory paragraph

Single-color bead controls, also known as reference controls, are used to visualize the spectral signature for each fluorophore on the flow cytometer. Because spectral unmixing is dependent on clear separation of positive and negative populations and an exact spectra match, beads are often used. Each marker in the flow cytometry panel must have a single-color control made. If this protocol is conducted properly, there should be a positive and negative population with clear separation on the flow cytometer. Further, the spectral signature for each single-color control should be unique. This will allow the flow cytometer to perform spectral unmixing on the samples.

Reagents, Solutions, and Test Animals

- Individual stains (see supplemental for list of antibodies)
- UltraComp eBeads (Fisher, cat# 01-2222-42)
- 96-well plates with V-bottom (Sigma Aldrich, cat# M9686)
- FACS Staining Buffer (see recipe)

Hardware and Instruments

- 10 μ L pipettor (Sigma Aldrich, cat# EP3124000024)
- 10 μ L pipet tips (Thermo Fisher Scientific, cat# 2707454)
- 200 μ L pipettor (Sigma Aldrich, cat# EP3124000083)
- 200 μ L pipet tips (VWR, cat# 53508-810)
- Tabletop Centrifuge (e.g., Beckman Coulter Allegra 6)
- 5 mL Polypropylene round bottom flow tubes (VWR, cat# 60819-794)
- Cytex Aurora Flow Cytometer

Protocol steps

1. Prepare individual stains by adding the appropriate antibody dilution to 100 μ L of FACS Staining Buffer.
2. Add 1 drop of UltraComp beads for each single-color control to wells on a 96-well plate.
3. Centrifuge beads at 380 G-force for 10 minutes at 4°C.
4. Remove supernatant by gently tapping the liquid out of the wells onto a paper towel.
5. Add 100 μ L of the individual stains to each well.
6. Incubate for 10 minutes at room temperature.

7. Centrifuge beads at 380 G-force for 10 minutes at 4°C.
8. Remove supernatant.
9. Wash beads by adding 100 µL of FACS Staining Buffer and centrifuge again at 380 G-force for 10 minutes at 4°C.
10. Remove supernatant by gently tapping the liquid out of the wells onto a paper towel.
11. Wash beads again by adding 100 µL of FACS Staining Buffer and centrifuge again at 380 G-force for 10 minutes at 4°C.
12. Remove supernatant.
13. Add 100 µL FACS Staining buffer to each well and transfer the samples to flow tubes that contain 200 µL FACS Staining buffer.

Reagents and Solutions

2x DNase/Liberase

Materials

- Liberase (5.2 Wunch units/mg) (Sigma, cat# 5401127001)
- DMEM 1x with 4.5g/L glucose without L-glutamine, sodium pyruvate (Corning, cat# 15-017-CV)
- DNase I [type IV Bovine] (3,000 units/mg) (Sigma, cat# D5025-150KU)
- 50 mL Conical Tubes (Thermo Fisher Scientific, cat# 12565270)

Protocol steps

1. Mix 100 mg of Liberase (520 units) and 50 mg of DNase (150,000 units) in DMEM and adjust the total volume to equal 200 mL.

2. Mix the solution well and aliquot into conical tubes based on experimental needs taking into consideration that each lung sample will require 1 mL of this solution.
3. Store at -80°C.

Complete Media

Materials

- DMEM 1x with 4.5g/L glucose without L-glutamine, sodium pyruvate (Corning, cat# 15-017-CV)
- MEM Nonessential amino acids 100X (Corning, cat# 25-025-CI)
- Penicillin streptomycin (Thermo Fisher Scientific, cat# 15140-122)
- L-glutamine (Thermo Fisher Scientific, cat# 25030081)
- Heat-inactivated Fetal Bovine Serum (Thermo Fisher Scientific, cat# MT35011CV)

Protocol steps

1. Add 4.5 mL MEM amino acids, 4.5 mL Penicillin streptomycin, 4.5 mL L-glutamine, and 45 mL Fetal Bovine Serum to the bottle of 500 mL 1x DMEM in a sterile hood.
2. Store at 4°C.

FACS Staining Buffer

Materials

- Heat-inactivated Fetal Bovine Serum (Thermo Fisher Scientific, cat# MT35011CV)
- Phosphate Buffered Saline (VWR, cat. No. 45000-446)
- Sodium Azide (Fisher Chemical, cat# S2271-25)

Protocol steps

1. Add 10 mL of Fetal Bovine Serum to a 500mL bottle of PBS.
2. Weight 0.25 g of Sodium Azide and add to the PBS/FBS solution.
3. Store at 4°C.

FMOs

Materials

- Same materials as those listed in T cell surface antibody cocktail and T cell intracellular antibody cocktail

Protocol steps

1. Create one FMO for every marker in the flow cytometry panel.
2. Add a 1:10 dilution of Brilliant Stain buffer and the appropriate dilutions of all the markers except for one to FACS Staining Buffer to make up a 100 µL volume.

The surface FMOs will be prepared in FACS Staining Buffer and the intracellular FMOs will be prepared in Permeabilization Buffer.

1X Perm/Fix Buffer

Materials

- Fixation/Permeabilization Concentrate (Thermo Fisher Scientific, cat# 501129082)
- Fixation/Permeabilization Diluent (Thermo Fisher Scientific, cat# 501129081)

Protocol steps

1. Add 1-part Fixation/Permeabilization Concentrate to 3-parts Fixation/Permeabilization Diluent.

Example: 12.5 mL concentrate + 37.5 mL diluent

2. Store at 4°C

1X Permeabilization Buffer

Materials

- Permeabilization Buffer 10X (Invitrogen, cat# 00-8333-56)
- Sterile Deionized Water

Protocol steps

1. Add 1-part Permeabilization Concentrate to 9-parts Sterile Deionized Water.

Example: 5 mL Perm buffer + 45 mL water

2. Store at 4°C.

T cell intracellular antibody cocktail

Materials (see Table A.2)

- Permeabilization Buffer (see recipe)
- Anti-mouse IL-17A, PE (BioLegend Cat# 506904, RRID: AB_315464)
- Anti-mouse IFN- γ , PE Cyanine 7 (Thermo Fisher Scientific Cat# 25-7311-82, RRID: AB_469680)
- Anti-mouse IL-10, Brilliant Violet 421 (BioLegend Cat# 505021, RRID: AB_10900417)
- Anti-mouse TNF- α , Pacific Blue (BioLegend Cat# 506318, RRID: AB_893639)
- Anti-mouse CD16/32 (Biolegend, cat# 101330)

Table A.2: Antibodies and concentrations for T cell intracellular antibody cocktail

Fluorophore	Marker	Clone	Effective Concentration
PE	IL-17A	TC11-18H10.1	2 µg/mL
PE Cy7	IFN- γ	XMG1.2	2 µg/mL
BV421	IL-10	JES5-16E3	1 µg/mL
Pacific Blue/e450	TNF- α	MP6-XT22	5 µg/mL
-	CD16/32	93	2.5 µg/mL

Protocol steps

1. Prepare intracellular antibody cocktail in permeabilization buffer.

Brilliant Violet Buffer does not need to be added to this cocktail because there is only 1 Brilliant Violet stain present.

2. Add the antibodies at the dilution listed above.
3. Store at 4°C in the dark.

Protect antibody cocktail from the light.

T cell surface antibody cocktail

Materials (See table A.3)

- FACS Staining Buffer (see recipe)
- Brilliant Stain Buffer (BD Biosciences, cat# 566349)
- Rat Anti Ly-6A/E, BB515 (BD Biosciences Cat# 565397, RRID: AB_2739218)
- Anti-mouse CD3, Alexa Fluor 532 (Thermo Fisher Scientific Cat# 58-0032-82, RRID: AB_11217479)
- Anti-mouse CD62L, PE/Dazzle 594 (BioLegend Cat# 104448, RRID:AB_2566163)

- Anti-mouse CD122, PE/Cy5 (BioLegend Cat# 123220, RRID: AB_2715962)
- Anti-mouse CD28, PerCP/Cyanine 5.5 (BioLegend Cat# 102114, RRID: AB_2073850)
- Anti-mouse PD-1, PerCP-eFluor 710 (Thermo Fisher Scientific Cat# 46-9981-82, RRID: AB_11151142)
- Anti-mouse CD103, APC-R700 (BD Biosciences Cat# 565529, RRID: AB_2739282)
- Anti-mouse CD44, APC/Fire 750 (BioLegend Cat# 103062, RRID: AB_2616727)
- Anti-mouse CD4, Brilliant Violet 480 (BD Biosciences Cat# 565634, RRID: AB_2739312)
- Anti-mouse CD8, Brilliant Violet 570 (BioLegend Cat# 100739, RRID: AB_10897645)
- Anti-mouse CD152, Brilliant Violet 605 (BioLegend Cat# 106323, RRID: AB_2566467)
- Anti-mouse CD27, Brilliant Violet 650 (BioLegend Cat# 124233, RRID: AB_2687192)
- Anti-mouse CD153, Brilliant Violet 711 (BD Biosciences Cat# 740751, RRID: AB_2740419)
- Anti-mouse KLRG-1, Brilliant Violet 786 (BD Biosciences Cat# 565477, RRID: AB_2739256)
- Anti-mouse CD16/32 (Biolegend, cat# 101330)

Protocol steps

1. Prepare surface antibody cocktail in FACS Staining Buffer with a 1:10 dilution of Brilliant Violet Buffer.

If two or more brilliant violet dyes are used together, it is important to add

Brilliant

Violet Buffer and FACS Staining Buffer together before adding antibodies to avoid aggregation of the antibodies labeled with brilliant violet dyes.

2. Add the antibodies at the dilution listed above.
3. Store at 4°C in the dark.

The antibody cocktail should be protected from light, for instance by wrapping in tin foil.

Table A.3: Antibodies and concentrations for T cell surface antibody cocktail

Fluorophore	Marker	Clone	Effective Concentration
BB515	Ly-6A/E (Sca-1)	D7	0.2 µg/mL
Alexa Fluor 532	CD3	17A2	4 µg/mL
PE Dazzle 594	CD62L	MEL-14	0.4 µg/mL
PE Cy5	CD122	TM-β1	2 µg/mL
PerCP Cy5.5	CD28	37.51	4 µg/mL
PerCP e710	PD-1	RMP1-30	2 µg/mL
APC R700	CD103	M290	1 µg/mL
APC Fire750	CD44	IM7	0.2 µg/mL
BV480	CD4	RM4-5	2 µg/mL
BV570	CD8	53-6.7	1 µg/mL
BV605	CD152 (CTLA-4)	UC10-4B9	4 µg/mL
BV650	CD27	LG.3A10	2 µg/mL
BV711	CD153	RM153	2 µg/mL
BV786	KLRG-1	2F1	2 µg/mL
-	CD16/32	93	2.5 µg/mL

Commentary

Background

Flow cytometry is a technique used to analyze the physical, functional and/or biological properties of cells including antigens, cytokines, size, and granularity. Cells are stained with fluorophore-conjugated antibodies. The cells are then sent single file through the flow cytometer, where a laser light source excites the fluorescently tagged antibodies, which can then emit light that is measured.⁴

Traditionally flow cytometry has only been able to examine a handful of parameters at a time. However, with the advent of spectral flow cytometry, there has been a large increase in the number of possible parameters that can be measured. Traditional analysis methods for flow cytometry include manual selection (gating) of cells on two-dimensional plots, often using expensive software such as FlowJo or FCS Express.⁵ While these programs are user-friendly, they lack the ability to analyze high-dimensional data quickly and efficiently. Researchers and computational biologists have been working to develop analysis tools that utilize feature engineering, clustering, and dimensionality reduction algorithms to address this complex data. However, to ensure accurate results with these tools, data needs to be acquired following very strict flow cytometry staining procedures, leading to clean and accurate data before analysis. Further, it is crucial that the correct controls are collected, and the flow cytometer is calibrated to a set standard.

With all types of flow cytometry, single-color controls are required for data acquisition. In conventional flow cytometry, these controls are used to build the compensation matrix. In

spectral flow cytometry, these controls are used to perform spectral unmixing, which separates the expression of similar fluorophore emissions based on the entire emission spectra of each individual fluorophore. This allows for multiple fluorophore-marker pairs to be used that would be indistinguishable with the same number of lasers on a traditional flow cytometer.⁶

Fluorescent Minus One controls, or FMOs, are an example of a “best practice” control used in both conventional flow cytometry and spectral flow cytometry.⁷ FMOs contain all the markers in a flow cytometry panel except for one. For example, if a panel contains 4 markers (e.g., CD45, CD3, CD4, CD8), then a CD45 FMO would contain CD3, CD4, and CD8, but not CD45. These controls allow for the user to account for any spillover from other fluorophores into the specific marker channel that is of interest.⁷ These FMOs allow users to place informed “gates” around populations that are either negative or positive for a marker.

Critical Parameters

There are multiple factors that influence the success of these protocols.

Protocol 1: Preparation of single-cell suspension for flow cytometry

While using razors to perform crosshatch cuts in the lungs, careful movements should be used to ensure lungs are not cut into pieces. There should only be imprints of the razors into the lungs rather than full cuts through the lungs. If the lungs are digested into small pieces following incubation with DNase/liberase, it is difficult to macerate them through the 70 µm filter.

Special attention should also be paid to the controls used in these experiments. These controls, which include FMOs and single-color beads, are critical in both spectral unmixing on the cytometer and determination of positive and negative populations during data analysis. There should be 1 FMO and 1 single-color control for every fluorophore-marker pair in an experiment.

Protocol 2: Surface and intracellular flow cytometry staining

Panel design is a highly important step prior to flow cytometry staining. Note that the same antibody clone and fluorochrome should be used for subsequent sampling to standardize the fluorescent signal. Ferrer-Font et al. and Mahnke & Roederer et al. describe in detail how to design and optimize a flow cytometry panel.^{3,8}

Performing a pilot optimization study is highly recommended to ensure that the fluorophore-marker pairs do not overlap in emission wavelength on a conventional cytometer or spectral signatures on a spectral cytometer. If the signatures are too similar, there will be a problem with spectral unmixing or signal overlap. In this case, change one of the problematic fluorophores to another that is not in use. Otherwise, the panel's complexity must be reduced.

Low or no signal in certain markers can be due to low marker expression on cells. In this case, the cells can be stimulated with PMA-ionomycin. Another check for the low signal would be to perform antibodies dilution series (example dilutions: 1:50, 1:100, 1:200, 1:400, 1:800), to rule out insufficient antibody concentration. If there is still low expression, the fluorophore could be too dim for the marker. It may be best to try a brighter fluorophore-marker pair.⁸

Particularly when using fixation/permeabilization steps, there can be an increase in cell autofluorescence. This autofluorescence can mask signal of other markers, particularly dim ones, if the proper controls are not used. The unstained sample should be used to assess autofluorescence, meaning that all steps (except for the addition of fluorescent antibodies) should be performed on these cells, including the fixation/permeabilization, anti-mouse CD16/32, and washes.

Tandem dyes, such as PE/Cy5 or APC/Fire750, are fluorescent molecules comprised of two covalently bound fluorophores, in which the energy emitted by one fluorophore excites the second one. While these dyes have largely increased the number of fluorophores available for use in flow cytometry, special attention should be paid to their properties. It is not uncommon for tandem dyes to degrade or decouple. This can occur when the fluorophores are exposed to light, long fixation and permeabilization steps, or if the tandem dyes are not properly stored at 4°C.⁹

During the staining procedure, extra precaution should be used in the washing steps. If a washing step is omitted, there may not be good separation between populations. Particularly in the dimension of Zombie-NIR, there cannot be any protein present in the diluent. Proteins, such as those from fetal bovine serum (FBS), Bovine Serum Albumin (BSA), or those released during red blood cell lysis can bind to Zombie-NIR, therefore lowering its effective concentration. Hence, washing cells with PBS prior to adding Zombie-NIR is a crucial step.

Troubleshooting

Table A.4: Description of possible problems, causes, and solutions for the flow cytometry cell preparation and acquisition.

Problem	Potential Cause	Potential Solution
No cells on flow cytometer	Cells were knocked out of the plate when performing the wash steps Cytometer is clogged	When decanting supernatant, tap more gently Run 10% bleach or 33% Contrad through the cytometer until the clog is removed
Fluorescent signal in the negative population of the single-color bead control	Non-specific binding of antibody to negative beads	Wash the beads with FACS Staining Buffer again to remove non-specific binding on the negative beads
Spectral signature of single-color control does not match expected signature for the fluorophore	Contamination of control with another fluorescent antibody Tandem dye degraded	Re-prepare control Replace tandem dye with new vial
High proportion of cells are dead	Cells left in DNase/liberase for too long Cells left in RBC lysing buffer for too long Centrifuged too long and too hard Cells incubated with Protein Transport Inhibitor for too long or too high of a concentration	Reduce the amount of time in the DNase/liberase Reduce the amount of time in the RBC lysing buffer Reduce the centrifuge speed and/or time Reduce the amount of time or concentration of Protein Transport Inhibitor
Fluorophore spill-over into other channels	Spectra of two markers is indistinguishable	Substitute a different fluorophore-marker pair

<p>Low signal in certain markers</p>	<p>Marker expression might be too low or inexistent</p> <p>Experimental design (e.g., time point or vaccine stimulation) does not elicit certain markers</p> <p>Markers chosen were too dim</p> <p>Tandem dyes degraded or decoupled</p>	<p>Check that the marker is supposed to be expressed in the particular cells or animal model</p> <p>Choose a time point closer to vaccination or stimulate the cells with PMA-ionomycin</p> <p>Choose a brighter fluorophore for the specific marker or optimize dilution used</p> <p>Use a new vial of the tandem dye</p>
<p>Not a clear separation in the live/dead sample stain</p>	<p>Residual protein left in the media is binding to and quenching Zombie-NIR</p> <p>Single-color control does not contain both positive and negative populations</p>	<p>Add additional washing step with PBS prior to staining with Zombie-NIR</p> <p>Use ArC Amine Reactive Compensation Bead Kit (Thermo Fisher, cat# A10346) to run single-color live-dead control</p>
<p>Bad separation between negative and positive populations</p>	<p>Antibody concentration is either too high or too low</p> <p>Unbound antibodies were not adequately washed from the samples</p>	<p>Optimize dilution of antibodies</p> <p>Add additional centrifugation and PBS wash</p>

Statistical Analysis

The flow cytometry data acquired following this protocol can be used with almost any type of statistical analysis. The single-color controls are used either for spectral unmixing in the case of spectral flow cytometry, or the development of a compensation matrix in conventional flow cytometry. The FMOs can be utilized either in manual gating strategies in FlowJo or FCSExpress, or in the novel analysis pipeline, cyto-feature engineering.¹ The cyto-feature engineering is an end-to-end analysis pipeline that can further form correlations between flow cytometry data and additional readouts. The data can also be analyzed with other tools such as OpenCyto, which implements automated gating, FlowSOM, which creates self-organizing maps, or t-SNE which performs dimensionality reduction.¹⁰⁻¹² To compare differences in marker expression or cell populations (defined by combinations of markers) between experimental groups, statistical tests such as Analysis of Variance (ANOVA) and Tukey Honest Significant Difference (HSD) can be used.

Understanding Results

After gating each sample and FMO on single cells, leukocytes, and live cells, the gates on the FMOs can be placed. An example of a good FMO (Figure A.3) is shown with CD4. Note that the cells are concentrated in a cluster and are not very spread out. When this FMO gate is placed on a sample, there is also clear separation between the two populations (Figure A.3).

An example of spillover is shown in Figure A.4. This problem can be diagnosed by plotting the marker of question against the other markers. A hallmark sign of spillover is signal expression along the line $Y=X$. In this case, a different fluorophore must be substituted.

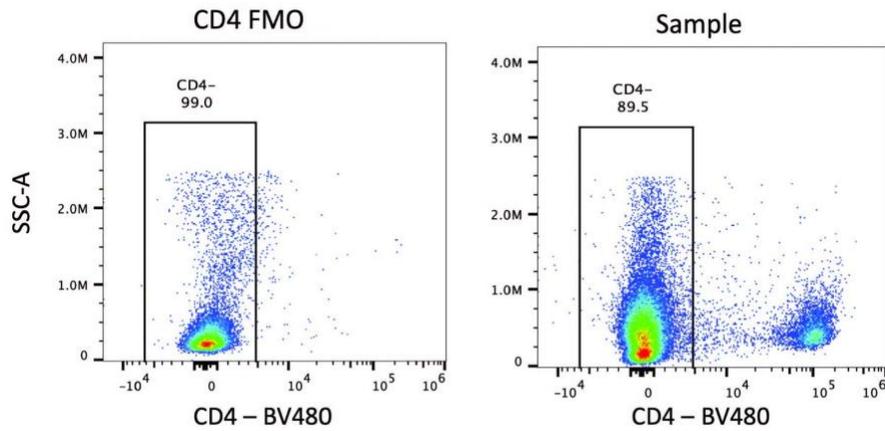


Figure A.3: Good fluorescence minus one (FMO) sample and gating strategy. An example of a good FMO is shown (left), in which the cells are tightly clustered and there is no expression of CD4 (as expected). The gate is placed on the negative CD4 population in the CD4 FMO and copied onto subsequent samples. There is further good separation between the positive and negative populations in the sample (right).

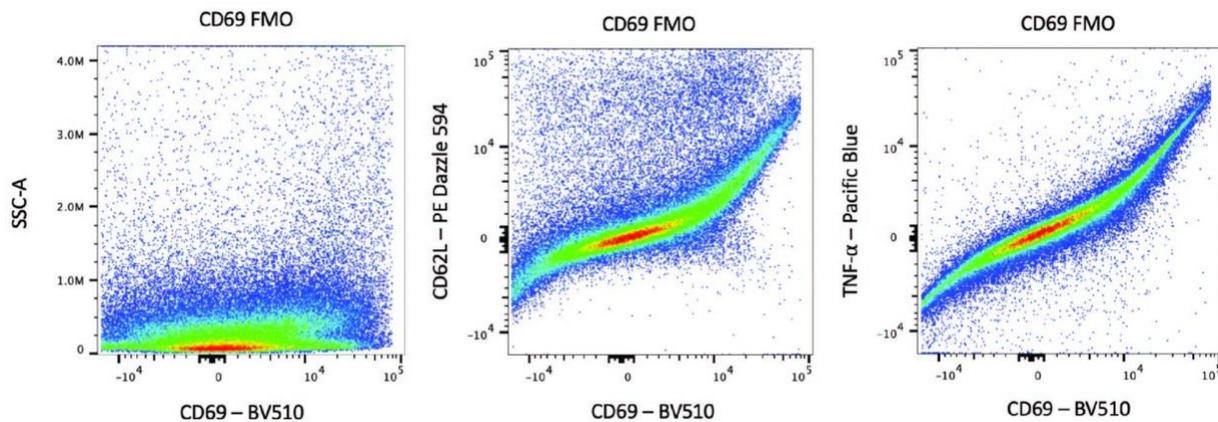


Figure A.4: Bad fluorescence minus one (FMO) sample due to spillover. An example of a bad FMO is shown, in which the population is spread across the x-axis (left). The source of the problem can be diagnosed by visualizing the marker vs. other panel markers (middle and right). The curved shape of the CD69 in the BV510 channel vs. PE/Dazzle 594 channel and the Pacific Blue channel shows that there is spillover from the other channels.

The single-color reference controls should display low signal in the negative population and a positive spectral signature identical to the fluorophore of interest (Figure A.5). If there is signal in the negative population, then there may be non-specific binding of the antibody to the negative beads (Figure A.6). Further, if the spectral signature does not match the expected signature exactly, there may be contamination from other fluorophores (Figure A.7). In both of these cases, the single-color controls should be re-run.

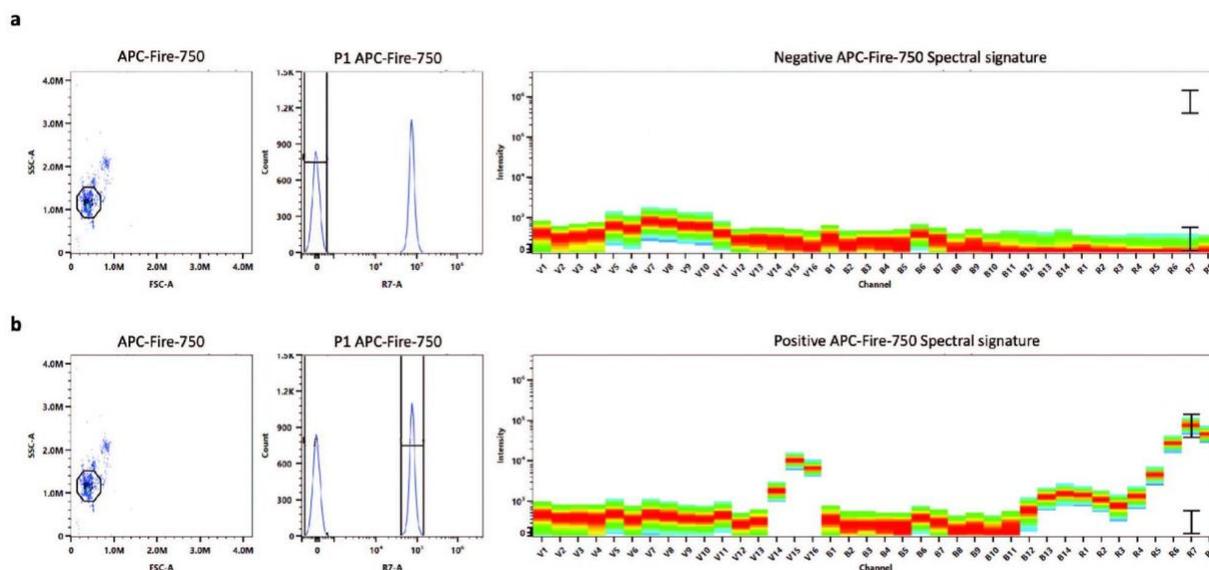


Figure A.5: Good single-color control with beads. A good APC Fire 750 single-color control with beads is shown. (a) The negative spectral signature should not show any signal. (b) The positive spectral signature should match the expected fluorophore spectral signature exactly.

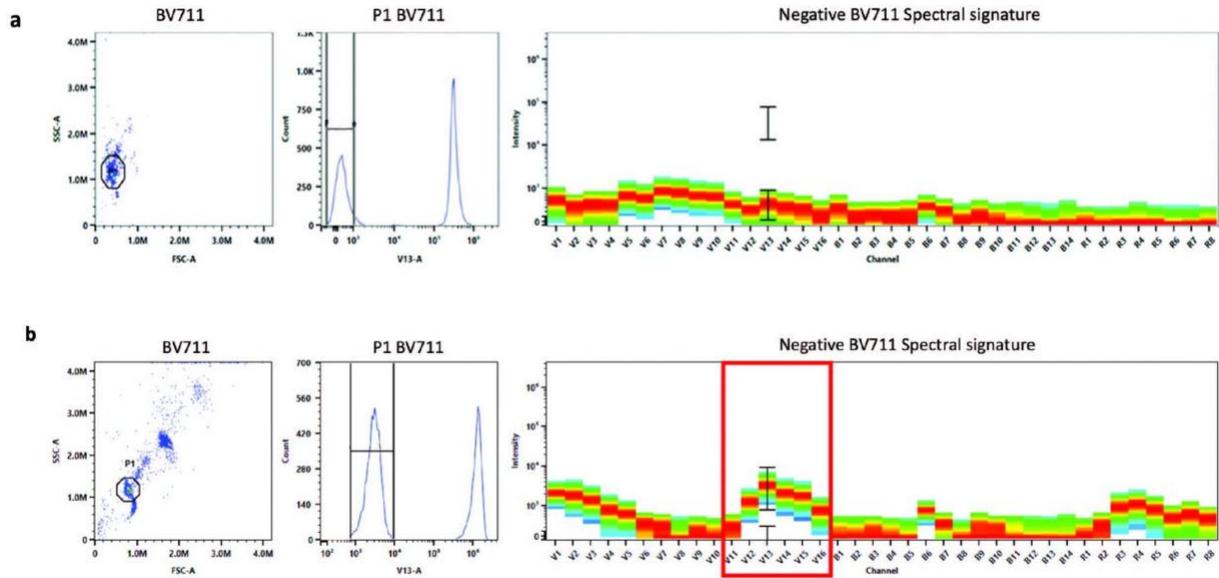


Figure A.6: Bad single-color control due to fluorescent signal in negative population. (a) An example is shown where the fluorescent signal in the negative population is low indicating that the single-color control has been run properly. (b) There is fluorescent signal (indicated in the red box) in the negative population indicating that there may be non-specific binding of the fluorophore. The BV711 single-color control should be prepared again and rerun.

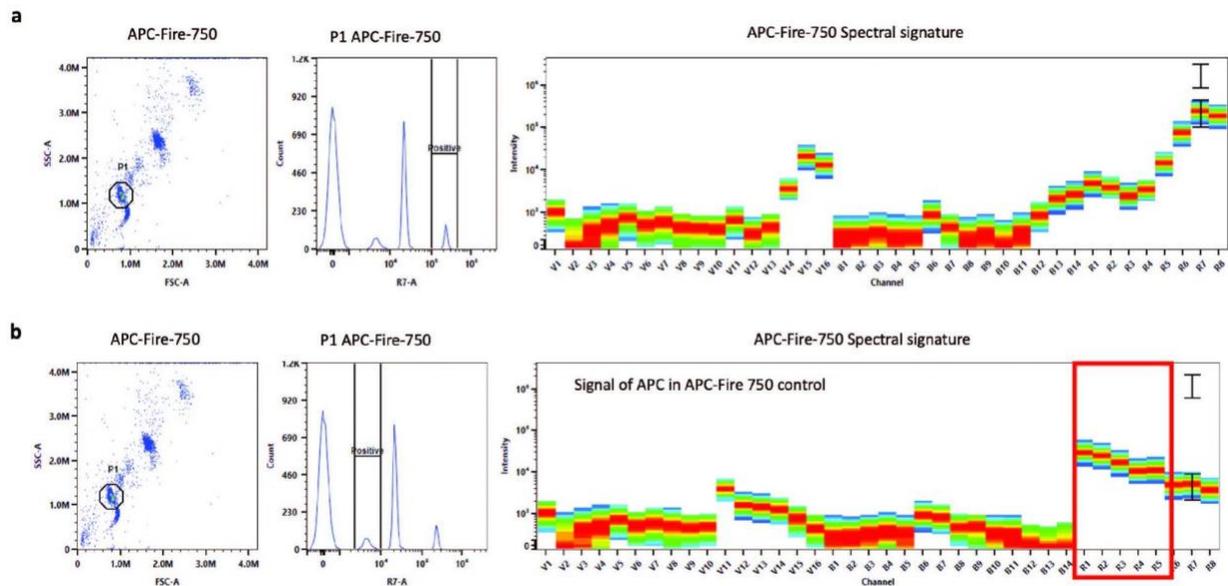


Figure A.7: Bad single-color control due to incorrect spectral signature. (a) The expected APC Fire 750 spectral signature is shown. (b) In the sample, the spectral signatures do not match. This is an example of APC signal in the APC Fire-750 single-color control. The control should be prepared again.

Complete analysis of the data acquired using this protocol can be found in the manuscript currently under review at “Cyto-feature engineering: a pipeline for flow cytometry analysis to uncover immune populations and associations with disease.” The code to analyze the data can be found here: https://github.com/aef1004/cyto-feature_engineering.

Time Considerations

The time required for these protocols depends on the number and type of samples. In protocol 1, it takes about 1.5 hours to prepare one sample from tissue harvesting to adding cells to a plate for staining. However, if it is the lung tissue being stained, 30 minutes should be added to this time due to the incubation with DNase/liberase.

To perform intracellular cytokine staining, an initial incubation step of 6 hours must occur. Following this incubation, it takes approximately 4.5–5 hours to stain. The plate must then sit overnight with the intracellular antibodies. The next day washing steps take approximately 30 minutes. Finally, it takes approximately 1-2 minutes to read one sample with 100,000 cells (with an initial concentration of 500,000 cells per tube).

Internet Sources

This website contains the code to analyze flow cytometry data using the Cyto-feature engineering analysis method. https://github.com/aef1004/cyto-feature_engineering

REFERENCES

1. Fox, A., Dutt, T.S., Karger, B., Rojas, M., Obregon-Henao, A., Anderson, G.B., and Henao-Tamayo, M. (2020). Cyto-Feature Engineering: A Pipeline for Flow Cytometry Analysis to Uncover Immune Populations and Associations with Disease. *Scientific Reports* 10. 10.1038/s41598-020-64516-0.
2. FluoroFinder (2020). FluoroFinder Flow Cytometry Panel Builder. <https://fluorofinder.com/>.
3. Ferrer-Font, L., Pellefigues, C., Mayer, J.U., Small, S.J., Jaimes, M.C., and Price, K.M. (2020). Panel Design and Optimization for High-Dimensional Immunophenotyping Assays Using Spectral Flow Cytometry. *Current protocols in cytometry* 92, e70-e70. 10.1002/cpcy.70.
4. Becton, D.a.C. (2012). Introduction to Flow Cytometry: A Learning Guide. In.
5. Verschoor, C.P., Lelic, A., Bramson, J.L., and Bowdish, D.M.E. (2015). An introduction to automated flow cytometry gating tools and their implementation. *Frontiers in Immunology* 6, 380. 10.3389/fimmu.2015.00380.
6. Biosciences, C. (2019). Spectral Analysis Meets Flow Cytometry.
7. Roederer, M. (2002). Compensation in flow cytometry. *Current protocols in cytometry Chapter 1*, Unit 1.14-Unit 11.14. 10.1002/0471142956.cy0114s22.
8. Mahnke, Y.D., and Roederer, M. (2007). Optimizing a multicolor immunophenotyping assay. *Clinics in Laboratory Medicine* 27, 469-+. 10.1016/j.cll.2007.05.002.
9. Biologicals, N. Tandem Dye Conjugated Antibodies
10. Finak, G., Frelinger, J., Jiang, W.X., Newell, E.W., Ramey, J., Davis, M.M., Kalams, S.A., De Rosa, S.C., and Gottardo, R. (2014). OpenCyto: An Open Source Infrastructure for Scalable, Robust, Reproducible, and Automated, End-to-End Flow Cytometry Data Analysis. *Plos Computational Biology* 10, e1003806. 10.1371/journal.pcbi.1003806.
11. Van Gassen, S., Callebaut, B., Van Helden, M.J., Lambrecht, B.N., Demeester, P., Dhaene, T., and Saeys, Y. (2015). FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry Part A* 87A, 636-645. 10.1002/cyto.a.22625.
12. Eshghi, S.T., Au-Yeung, A., Takahashi, C., Bolen, C.R., Nyachienga, M.N., Lear, S.P., Green, C., Mathews, W.R., and O'Gorman, W.E. (2019). Quantitative Comparison of Conventional and t-SNE-guided Gating Analyses. *Frontiers in Immunology* 10, 1194. 10.3389/fimmu.2019.01194.

LIST OF ABBREVIATIONS

ACEi – Angiotensin-Converting Enzyme Inhibitors

ANCOVA – Analysis of Covariance

ANOVA – Analysis of Variance

BCG – Bacille Calmette-Guérin

BSA – Bovine Serum Albumin

CC – Collaborative Cross

CFU – Colony Forming Units

FACS – Fluorescence Activated Cell Sorting Staining Buffer

FBS – Fetal Bovine Serum

FDR – False Discovery Rate

FlowSOM – Flow Self-Organizing Map

FMO – Fluorescence Minus One

HIV – Human immunodeficiency virus

HSD – Honest Significant Difference

LC-MS – Liquid Chromatography-Mass Spectrometry

MDR – Multidrug-Resistant

MF – Molecular Features

MFI – Median Fluorescent Intensities

MHC – Major Histocompatibility Complex

MS-TOF – Time-of-Flight Mass Spectrometer

MTB – Mycobacterium tuberculosis

NET – Neutrophil Extracellular traps

NTM – Non-Tuberculosis Mycobacteria

PAMP – Pathogen-associated molecular pattern

PBMC – Peripheral Blood Mononuclear Cells

PBS – Phosphate Buffered Saline

PC – Principal Component

PCA – Principal Components Analysis

PFA – Paraformaldehyde

PL – Propranolol-Losartan

PPD – Purified Protein Derivative

QC – Quality Control

RBC – Red Blood Cell

RPM – Revolutions Per Minute

SOM – Self-Organizing Map

TB – Tuberculosis

Th1 – T Helper 1 cells

Th2 – T Helper 2 cells

Treg – Regulator T cells

t-SNE – t-Distributed Stochastic Neighbor Embedding

XDR – Extensively Drug-Resistant