DISSERTATION

TARGETED COMPUTATIONAL ANALYSIS OF THE C3HEB/FEJ MOUSE MODEL FOR DRUG EFFICACY TESTING

Submitted by

Bryce Clifford Asay

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

Spring 2020

Doctoral Committee:

Advisor: Anne J. Lenaerts

John Belisle
Brian Munsky
Michael Lyons

ABSTRACT

TARGETED COMPUTATIONAL ANALYSIS OF THE C3HEB/FEJ MOUSE MODEL FOR DRUG EFFICACY TESTING

Efforts to develop effective and safe drugs for the treatment of tuberculosis (TB) require preclinical evaluation in animal models. Alongside efficacy testing of novel therapies, effects on pulmonary pathology and disease progression are monitored by using histopathology images from these infected animals. To compare the severity of disease across treatment cohorts, pathologists have historically assigned a semi-quantitative histopathology score that may be subjective in terms of their training, experience, and personal bias. Manual histopathology, therefore, has limitations regarding reproducibility between studies and pathologists, potentially masking successful treatments. This report describes a pathologist-assistive software tool that reduces these user limitations while providing a rapid, quantitative scoring system for digital histopathology image analysis. The software, called 'Lesion Image Recognition and Analysis' (LIRA), employs convolutional neural networks to classify seven different pathology features, including three different lesion types from pulmonary tissues of the C3HeB/FeJ tuberculosis mouse model. LIRA was developed to improve the efficiency of histopathology analysis for mouse tuberculosis infection models. The model approach also has broader applications to other diseases and tissues. This also includes animals that are undergoing anti-mycobacterial treatment and host immune system modulation. A complimentary software package called 'Mycobacterial Image Analysis' (MIA) had also been developed that characterizes the varying bacilli characteristics such as density, aggregate/planktonic bacilli size, fluorescent intensity, and total counts. This further groups the bacilli characteristic data depending on the seven different classifications that are selected by the user. Using this approach allows for an even more targeted

analysis approach that can determine how therapy and microenvironments influence the Mtb response.

ACKNOWLEDGMENTS

I want to thank my advisor Dr. Anne Lenaerts for taking a chance on me by accepting me into her lab. Especially since my approaches to my Ph.D. and research have been outside the scope of the lab's expertise. I know that it had to take a lot of faith and patience while I integrated other disciplines with the lab's research interests. I have learned so much from her about research and how to be successful in life. I could not have asked for a better mentor, and I hope for her continued success.

I want to thank Dr. Gregory Robertson for his mentorship and entertaining my questions. He has been incredibly generous in both his time and helping me develop as a writer and scientist. He made my time much more enjoyable at CSU, and I am excited to see his future successes. I want to thank Dr. Michael Lyons for all the wonderful discussions, time commitment, and debates that we had. They were enjoyable and made me question everything I was doing in a different light. Even though I did not always listen to his advice, I always held it in high regard, and I always took it to heart. I also want to thank Dr. Mike McNeil for all the biochemistry we did together and for encouraging me to think about the moral consequences associated with my research. Lastly, I want to thank Dr. Gerry Andrews and Dr. Myrna Miller from the University of Wyoming for helping me start my journey into research.

The Lenaerts laboratory, members both past and present, really deserve my thanks for helping me with my projects when I needed it most. I could not have done this without them.

Lastly, I want to thank my family. I know that my wife has felt like a single parent on occasion, and I feel like she has earned this Ph.D. as much as I have. I love my wife and kids more than they know!

TABLE OF CONTENTS

ABSTRACT……..ii

ACKNOWLEDGMENTS……iv

CHAPTER 1: LITERATURE REVIEW……1

1.1 Background……..1

1.2 Human Tuberculosis Disease And Diagnosis……5

1.3 Animal Models Of Tuberculosis……8

1.3.1 Non-Human Primate……..…….…….9

1.3.2 Marmoset……..……10

1.3.3 Rabbit…..……10

1.3.4 Guinea pig…..……11

1.3.5 Mouse……..11

1.3.6 C3HeB/FeJ Mouse Model……..……13

1.4 Scoring Histology On Microscopy Slides……..17

1.4.1 Histology Scoring Of Tuberculosis-Infected Tissues On Microscopic Slides……..…19

1.5 Image Analysis…….……21

1.5.1 Introduction……..……..21

1.5.2 Techniques In Computer Vision And Image Processing……..…..22

1.5.3 Machine Learning……26

1.5.4 Human Bias In Research……...29

1.5.5 Computer Vision (CV) In Clinical And Biomedical Research……32

1.6 Rationale……33

CHAPTER 2: DEVELOP LESION IDENTIFICATION AND AREA CALCULATION SOFTWARE (LIRA) FOR THE C3HEB/FEJ MOUSE MODEL……36

2.1 Introduction……36

2.2 Materials And Methods……41

2.2.1 Sample Collection From Achieved Animal Studies……41

2.2.2 Processing Of Samples To Generate The Digital Image Dataset……42

2.2.3 Lesion And Histopathology Classification Scheme……43

2.2.4 The LIRA Pipeline……45

2.2.5 Neural Network Training……46

2.2.6 Validation Of Software……47

2.2.7 Evaluation Metrics For Performance And Accuracy Of The Neural Networks……49

2.2.8 Computer Code Availability……50

2.2.9 Acknowledgements……50

2.3 Results……50

2.3.1 Ability Of CNN1 To Identify Caseous Necrotic (Type I) Lesions……50

2.3.2 Accuracy For Lira In Comparison To An Experienced TB Researcher……52

2.3.3 Improved Agreement Of Histopathology Classifications Among Pathologists Using LIRA……54

2.3.4 Trained Research Technicians + Lira……60

2.4 Discussion……61

CHAPTER 3: IMPLEMENTATION OF LIRA TO ASSESS TREATMENT RESPONSES USING THE C3HEB/FEJ MOUSE EFFICACY MODEL……66

3.1 Introduction……66

3.2 Materials And Methods……72

3.2.1 Animals……72

3.2.2 Bacteria……72

3.2.3 Chemotherapies……73

3.2.4 Aerosol Infection And Sample Collection……73

3.2.5 Pathology, Slide Preparation, And Imaging……73

3.2.6 Stereology……74

3.2.7 Scoring Of Pathology For LIRA Comparison……74

3.2.8 LIRA Analysis……75

3.3 Results……75

3.3.1 Design And Past Results Of The C3heb/Fej Efficacy Study (Study Performed By Dr. Emily Driver) ……75

3.3.2 LIRA Histopathology Analysis……77

3.3.3 Cohort Total Score Agreement Between Lira And The Classical Pathology Approach……81

3.3.4 Individual Lesions (Type I, II, III) Score Agreement Between Lira And The Classical Pathology Approach……83

3.4 Discussion And Conclusions……84

CHAPTER 4: DEVELOPMENT OF MYCOBACTERIAL IMAGE ANALYSIS (MIA) SOFTWARE TO QUANTIFY LESION SPECIFIC POPULATIONS OF MYCOBACTERIUM TUBERCULOSIS……90

4.1 Introduction……90

4.2 Materials And Methods……94

4.2.1 Mouse Infection……94

4.2.2 Staining And Imaging……94

4.2.3 MIA Algorithm Pipeline……95

4.2.3.1 Number Of Defined Bacterial Areas per um^2……..…..….....96

4.2.3.2 Individual Bacterial Area Analysis..…….....…….....…….....…..97

4.2.3.3 Percent Of Image Composed Of Visual Bacteria..…….....……..97

4.2.4 MIA Optimization and Validation..…….....…….....…….....…….....98

4.3 Results..…….....…….....…….....…….....…….....…….....…….....98

4.3.1 Verification Of Software……..….98

4.3.2 Verification Of The MIA Software On C3HeB/FeJ Histology Tissues…...101

4.3.3 Analysis Of Drug-Treated Bacterial Populations *In Vitro*.……..103

4.3.4 Mycobacterial Islands……108

4.4 Discussion……109

Chapter 5: Final Discussion And Future Directions…….……112

5.1 Final Discussion……...112

5.2 Future Directions……..……117

References……..……...122

Appendix: Heterogeneity Of Pulmonary Lesion Types In C3HeB/FeJ Mice Is Modulated By Specific Characteristics Of The *M. Tuberculosis* Strain Used……....135

List Of Abbreviations……..148

CHAPTER 1: LITERATURE REVIEW

1.1 – Background

Tuberculosis is the leading cause of death by an infectious organism, with an estimated 1.3 million deaths in 2017 alone [1]. The World Health Organization (WHO) estimates that 23% of the world's population has latent TB, and 10 million new cases of tuberculosis occurred in 2017 [1]. The burden of new TB infections and TB-related deaths disproportionally affect individuals in low-income countries, reflecting the strong connection between this devastating disease and conditions associated with poverty (i.e., poor health care, inadequate nutrition, overcrowded living conditions) [2]. The current UN Sustainable Development Goals (SDGs) includes a 90% reduction in the absolute number of TB deaths and an 80% reduction in new cases by 2030 compared with 2015 levels. Progression towards this goal has been made with a 2% annual decrease of incidence but falls short of the 4-5% that is needed per year to reach the 2030 milestone [1], [3], [4]. To attain the TB SDGs, significant progress needs to be made in multiple areas including: 1) the development of new, more effective drug therapies that can shorten treatment duration, 2) improved establishment of more predictive animal models that can guide new regimen development, 3) more robust and predictive diagnostics and biomarkers that can help guide clinical trials and individualized patient therapy, and 4) more protective vaccines.

First-line therapy for uncomplicated TB is lengthy, consisting of two months of isoniazid (INH), rifampicin (RIF), ethambutol (EMB), and pyrazinamide (PZA) and an additional 4 to 7 months of INH and RIF [5]. The first-line regimen is effective, with a success rate exceeding 95% [6]. However, poor treatment compliance, resistance to key first-line drugs, and associated co-morbidities - such as diabetes, human immunodeficiency virus (HIV) infection, alcoholism, smoking, cancer, etc – can have a dramatic impact on treatment success and/or lead to development and expansion of drug resistance [7], [8]. An additional issue is the increased incidence of resistance to first-line TB therapies. Specifically, resistance to isoniazid and

rifampicin is occurring in about 4.1% of the newly infected individuals and in 19% of patients who have been previously treated for the disease [9]. Second-line therapy for MDR-TB patients involves longer treatment durations appreciably and is associated with increased risk of treatment-associated adverse events (i.e., side effects), and significantly higher treatment cost. Current second-line therapies may include combinations of newer FDA-approved drugs such as bedaquiline (BDQ), nitroimidazoles, including pretomanid and delamanid, or older drugs including oxazolidinones such as linezolid, fluoroquinolones, ethionamide, or injectable aminoglycosides such as kanamycin, amikacin, or capreomycin.

Further complicating the challenge of successful treatment of drug-resistant TB is the recent emergence of extensively drug-resistant TB (XDR TB) - which is resistant to INH, RIF, any fluoroquinolone and at least one injectable second-line drug and totally drug-resistant TB - which shows an even wider array of drug resistance phenotypes. The rise of antibiotic resistance underscores the urgent need for the development of new and more effective TB drug regimens, which has been the strategic goal for all current TB drug development efforts. Progress towards this goal has been made with the approval of three new FDA approved drugs for the treatment of TB [10]. These include bedaquiline (brand name Sirturo), a diarylquinoline antibiotic that blocks the ability of Mtb to make adenosine 5'-triphosphate and two nitroimidazoles, pretomanid and delamanid, both prodrugs, which are thought to exert bactericidal effect by inhibiting bacterial cell wall mycolic acid biosynthesis, among other possible targets [11]. All three drugs were approved for the treatment of MDR/XDR TB [5], [12]. These recent successes reflect the concerted efforts of key stakeholders and funders to entice pharmaceutical companies and their research partners back to TB infectious disease research and development. However, the current scheme for TB drug development continues to be hampered by a lack of essential tools to predict treatment outcomes or identify more effective drug candidates and combinatory drug regimens.

The causative agent of TB is the bacillus bacterium *Mycobacterium tuberculosis* (Mtb). Typically associated with various disease states of the lungs, Mtb may also spread to other parts

of the body, leading to a host of TB-related disease complications, which is the subject of a recent review [13] and will not be discussed in detail here. The pathogenesis of pulmonary TB disease has been studied for decades, and the key steps and processes are now well described (see section 1.2, below; [14], [15]). One distinct feature of human TB patients with active pulmonary disease is the appearance of a variety of different pulmonary lesion types (TB-associated pathology), whose appearance can have a dramatic impact on treatment effectiveness, disease outcome, and the development and expansion of drug-resistance. Georges Canetti first reported on the different pathologic states in human TB patients that ultimately succumbed to disease [16]. From these early studies to more recent work using modern tools and methodologies, it is now abundantly clear, although the exact details are often disputed, that the presence of diverse lung pathology has a profound effect on bacterial phenotype [17]–[19]. The key microenvironmental differences cited include varying levels of oxygen, or pH, shifts in nutrient availability and carbon sources, impact or lack host immune system responses, and altered bacterial location (i.e., intracellular versus extracellular in necrotic caseum). From a drug discovery perspective, increased knowledge regarding these varying environmental factors has proven valuable in evaluating potential TB therapeutics under similar *in vitro*/*ex vivo* conditions. However, conventional *in vitro* assays such as minimum inhibitory concentration (MIC) assay and minimum bactericidal concentration (MBC) assay, which is most commonly used in drug discovery of new chemical entities [20], fail to recapitulate the pathological and environmental complexity found within an infected host or model host *in vivo*. Additionally, even when new therapeutics are discovered, drug efficacy and drug-partitioning are often seen to vary due to lesion-specific environmental heterogeneity found in humans and more advanced TB animal efficacy models presenting with complex-TB-associated lung pathologies [21]. For example, BDQ distributes slowly, and to a lesser extent, into caseum of fibrotic, necrotic granulomas found in TB-infected C3HeB/FeJ mice compared to non-encapsulated lymphocyte dominated cellular lesions observed in TB-infected BALB/c mice [22]. Pyrazinamide, in contrast, readily distributes into both necrotic

and non-necrotic lesions in mice, but its activity against bacilli is markedly reduced, as a result of near-neutral pH of the caseum microenvironment [22], [23].

Animal models are a fundamental part of both vaccine development, testing novel drug therapies, and basic disease research. Common animal models employed in TB research include non-human primates, Guinea pigs, mice, and rabbits [24]. Although no one model recapitulates all aspects of the human TB disease spectrum [18], all such models aim to recapitulate certain aspects of the disease in humans. The most advanced animal models seek to replicate, as much as possible, the human disease state includes the pathophysiological lesion conditions described in humans with active or latent TB lung disease.

Although animal models may better represent human situation relative to common *in vitro* models, not all animal models present with the entire range of pathology features and disease states observed in human TB patients. The inability to recapitulate these important aspects can potentially limit the predictive nature of some animal models and, thus, limit the extrapolation of data from *in vivo* animal models to the human situation. Therefore, the choice of the animal model employed during the drug discovery process is key, as is understanding potential limitations of the output from the use of these models. Additionally, conducting such experiments and analysis of the resultant data can be time-consuming and require highly specialized training and tools. As an example, a single histopathological image analysis conducted by our laboratory of 132 samples, or about 660 individual lung lobes, took nearly four weeks to reduce to a quantifiable data set by a dedicated technician (not shown). Such extensive endeavors lead to user fatigue and is subject to variability in data analysis and interpretation when extended over multiple days, laboratories, or individual researchers. Therefore, one approach to accelerate and improve TB drug development is to employ more predictive animal models that better mimic human disease, and concomitantly, to develop and employ faster, more robust methods to analyze and correlate resultant data produced in preclinical animal efficacy trials.

Image analysis and artificial intelligence (AI) have seen tremendous research growth during the last several years. The capabilities of AI in digital medical image analysis hold the potential to not only improve the speed by which large, complex datasets can be analyzed, but lead to improved accuracy and can facilitate wholly new analyses that were not possible using conventional methods. For example, AI was found to match or outperform qualified health care professionals in visually diagnosing human breast cancer and certain skin disease states based upon digital pathology images [25], [26]. AI strengths include a reduced level of bias, increased reproducibility, faster analysis time, and the ability to collect quantifiable data. Conversely, the limiting factor of AI in biological research is the small representative data sets employed for initial AI training, and an inability to correctly classify uncommon events that were not present in the original training set [27], [28]. A proficient and accurate pathology classification pipeline will not rely entirely on either traditional methodology or AI alone, but integrate the two approaches to reduce the limitations of either approach while retaining their strengths.

1.2- Human Tuberculosis Disease And Diagnosis

TB infection begins after a susceptible individual inhales aerosolized droplets (5-10 microns) of Mtb which travel to the alveolar space within the lung. If the host immune system fails to eradicate the pathogen, the disease progresses into primary tuberculosis. During this phase, distribution of nearly 40% of pulmonary bacilli in either alveolar macrophages or interstitial macrophages occurs, with the majority of replication occurring in alveolar macrophages [29], [30]. Upon becoming infected, macrophages produce chemokines and cytokines to attract other phagocytic cells such as additional alveolar macrophages, monocytes, and neutrophils. If disease progression continues, cells will form a caseating granuloma due to lack of oxygen and blood flow into the expanding lesion, which is composed of central caseous necrosis surrounded by epithelioid macrophages and lymphocytes [31]. If the host is still unable to control the infection, enlargement of the tubercle is proceeded by the dissemination of Mtb into the lymph nodes. The

expansion of the pulmonary tubercle and the involvement of the lymph node followed by healing is called a Gohn's complex [32]. The Gohn's complex includes calcified granulomas in the pulmonary tissue and hilar lymph nodes [14]. The bacilli continue to proliferate uninterrupted until an effective T-cell mediated immune response develops, which typically happens 2 to 6 weeks post-infection.

The period after the initial infection, in which the host is in a state of a persistent immune response to stimulation by the bacilli without overt signs of disease, is called a latent TB infection (LTBI). Individuals with LTBI are not infectious and are sputum negative [33]. About five to ten percent of LTBI individuals with no underlying medical conditions will experience a reactivation of the disease in their lifetime [34]. While a small population of reactivation cases has been observed where it took decades, recent studies support the idea that median reactivation time is anywhere from a few months to two years [35].

The poles of TB are latenet and active TB, however a dynamic process occurs between the latent and active stages of the diseases. Post-primary tuberculosis, which is the period that occurs after a latent infection, is normally restricted to the upper lobes of the lungs with no lymph node or other organ involvement. 90% of individuals with reactivated tuberculosis will recover without therapeutic intervention [14], [15]. In individuals that proceed to active TB, the disease may develop a heterogeneity of different lesion types of varying degrees of pathological severity. Post-primary TB pulmonary lesions are more representative of caseous pneumonia in comparison to the host tubercle response initially seen in primary tuberculosis. The lipid pneumonia stage lesions initially form with the accumulation of foamy macrophages in the alveoli with little edema and no leukocytes or fibrin present [15], [36]. A distinguishing feature is the obstruction of bronchi by infected cells and debris that shows up as a tree-in-bud structure in radiological imaging [37], [38]. Mtb organisms are found extensively in the alveolar macrophages [39], and resolution of the disease, or healing, happens in 95% of individuals [14], [40]. Lesions that fail to heal will undergo necrosis by fibrin losing its fibrillary structure resulting in the consolidation of the entire alveoli

structure and restricting cellular flow. These lesions are termed caseous pneumonia and are homogenous in their presentation [41]. Regions of caseous pneumonia will eventually soften, followed by fragmentation and expulsion of the fragmented lung via coughing. These areas of dislodged tissue form a cavity that is followed by massive amounts of bleeding [42]. Eventually, a thin wall of fibrosis and necrotic material forms at the cavity site. Resolution of untreated disease at the cavitation phase is infrequent, unlike what is seen previously in the lipid and caseous pneumonia phase. During and after the formation of the cavity, bacilli are present in small numbers, and it is only after the maturation of the cavity that the numbers of bacteria substantially increase.

TB diagnosis based on sputum culturing of bacteria and microscopy was originally developed by Robert Koch over 130 years ago and is still the existing method of diagnosing TB. Sputum culturing is considered the gold standard for the diagnosis of active TB disease but is dependent on the skill of the technician. Sputum culture exhibits a high specificity (98%), but has poor sensitivity (45-80%), and takes on average 2 to 8 weeks to provide definitive diagnostic results [43]. In response to these limitations, imaging techniques such as plain chest radiography and serial positron emission tomography (PET) combined with computed tomography (CT) imaging is used in more affluent areas where cost is not as much of a limitation. On radiographs, patients with primary TB Ghon complex, adenopathy, and pleural effusion are visible. In patients with reactivated TB, focal patchy opacities, cavitation, fibrosis, nodal calcification, and flecks of caseous material can be visualized [43]. In patients with severe immune deficiencies, such as patients with HIV, this method of diagnosis is not reliable due to the retardation of lesion development [44]. PET/CT imaging provides a more holistic approach by including anatomic information of the disease. Active TB readily uptakes 2-[18F]fluoro-2-deoxyglucose (18F-FDG) PET/CT (F-FDG PET/CT), which is used to detect active granulomas, and monitor the dynamics of activity and the extent of the lesions. Because this technique is not TB specific, it is prone to false positives. Besides, PET/CT image analysis does not identify closed ‘cold’ lesions, which can

contain viable bacteria. Visualization of lesion dynamics with F-FDG PET/CT has shown that treatment affects different lesion types disproportionately [45]. Overall, lung involvement stays relatively stable, but individual lesions experience dramatic variations in size and FDG avidity during and after effective therapy [46]. F-FDG PET/CT, therefore, represents a studied as an alternative approach to assessing treatment efficacy [45].

In summary, TB is a dynamic disease that presents with multiple different lesion types. Disease progression is rarely static, and heterogeneity in pathology is common, often manifesting as multiple different lesion types within a given infected individual. Each lesion presents a unique microenvironment in which Mtb must survive and often propagates. Newer methodologies seek to understand better how and when these disease manifestations occur and their dynamics with treatment over time.

1.3- Animal Models Of Tuberculosis

Robert Koch initially observed that there was a spectrum of tuberculosis disease that can occur both in humans and in experimentally infected animals [47]. Such historical observations led to the wider practice of using various experimentally infected animals to study TB disease, pathogenesis, and the advent of antimicrobial therapy, drug efficacy, and therapeutic safety. While no animal model can perfectly replicate a human TB infection, researchers have capitalized on key models that can recapitulate specific aspects of human TB disease. Frequently employed animal models for TB research have included non-human primates, marmosets, rabbits, Guinea pigs, and mice [18], [24], [48]. These models have been employed to explore specific aspects of TB infection and disease progression, including immunity to disease, vaccine efficacy trials, bacterial pathogenesis, and drug efficacy. There are many challenges in working with animal models. These include the need for specialized ABLS-3 facilities and approval of key oversight committees and agencies (e.g., Institutional Animal Care and Use Committee, Office of Laboratory Animal Welfare, etc.). Additionally, because of the complex interactions between the

host and pathogen, care should be taken to optimize the experiment to gather relevant, clear, and translatable data and to treat the animals as humanely as possible. While many animals respond to TB infection similarly to humans from physiological, pathological, and therapeutic perspectives, there are also significant species-by-species differences. Overall, TB animal models have proven invaluable in aiding our understanding of TB biology, pathogenesis and immunology [24], [48]. Additionally, animal models, especially the laboratory mouse, have proven essential for the advancement of TB therapeutic approaches and the development of new drug regimens. In the following sections, I will briefly describe the most commonly used animal models in TB research with an emphasis on the C3HeB/FeJ murine animal model, which is the focus of this dissertation.

1.3.1 - Non-Human Primate

The non-human primate model (NHP) for TB develops a spectrum of pulmonary lesion pathology that appears to be the closest disease representation of human disease. Cynomolgus macaques (CM) with a low infectious dose develop a heterogeneity of lesions comparable to human disease. In particular, the model can develop both active and latent TB, which is of particular importance since there is no other suitable *in vivo* model of LTBI [49]. Each granuloma develops from a single bacterium [50] and can be detected 2 to 4 weeks post-infection in both pulmonary and lymph node tissue [51], [52]. Initially, the animals develop a smaller and denser cellular granuloma that is typical of human tuberculosis. Additionally, they may develop caseating granulomas that will eventually see a reduction of epithelioid macrophages, giant cells, and lymphocytes, followed by subsequent mineralization [18]. As seen in humans, PET-CT scan analysis revealed that these lesions are dynamic during infection. While CM appears to most closely model human disease with their development and heterogeneity of lesion types, it is also a weakness of the model. NHP models are expensive to purchase, house, and due to the diversity of disease presentations, large numbers of animals are needed to obtain relevant statistical information, especially for endpoint necropsy studies. The need for large numbers can be

somewhat mitigated by PET-CT technologies, which enables the tracking of lesions of the same animal across multiple time points, thus reducing the number of animals needed per study.

1.3.2 - Marmoset

Marmosets are a commonly used NHP which recapitulates many common aspects of human TB disease. After a low-dose aerosol infection, the disease progresses proportionally to the virulence of the strain of Mtb used. Common pathological features develop 6 to 8 weeks post-infection and are composed of cellular, caseating necrotic lesions and Mtb strain-dependent cavitation [53]. A pathological difference between marmosets and CM is that marmosets fail to develop the calcified lesions seen in human disease. They do offer an advantage over CMs owing to their small size, reducing the cost of housing, and the capability of increasing the number of animals that can be used in a single study. Another benefit is that marmosets often develop as identical twins, which are ideal controls in intervention studies. There is also a wide availability of immunological reagents that can be used in contrast to other animal models such as rabbits or Guinea pigs.

1.3.3 - Rabbit

The New Zealand white rabbit TB animal model develops an Mtb strain-dependent heterogeneity of pathology presentations reminiscent of the diversity of the human disease. Developed pathology includes cellular, caseating, and cavitating lesions, but rabbits do not develop the calcified lesion normally associated with healing in human TB patients. Rabbits infected with H37Rv generally clear the infection, while animals infected with Erdman develop a chronic infection with 53% of the infected animals developing caseous necrotic granulomas [31]. When infected with a highly virulent strain of Mtb, such as the clinical Bejing HN878, animals will develop granulomas reminiscent of human disease and eventual cavity formation [54]. Mtb located in cavitary lesions are associated with phenotypic antibiotic resistance [55]. Resistance in

cavitary disease can be somewhat mitigated by treatment with anti-vascular endothelial growth factors, which have shown to reduce hypoxia and increase granuloma vasculature with a corresponding increase in treatment efficacy [38]. Rabbits are considered excellent models for modeling drug penetration, distribution, and cellular accumulation in TB granulomas. Rabbits are not more widely used for other aspects of TB biology, vaccine studies, or drug discovery owing to the lack of immunological reagents, increase biocontainment requirements due to Mtb shedding in their urine, and their overall larger size relative to mice.

1.3.4 - Guinea Pigs

Guinea pigs are highly susceptible to Mtb infection, and similar to humans exhibit pulmonary caseous necrosis, lymphadenopathy, and calcification of resolving lung lesions but only infrequently develop lesion cavitation [56]. They are considered the gold standard for vaccine efficacy because of the presence of CD1b molecules that responds to glycolipid antigens, also present in humans but not in mice [57]. Also, Guinea pigs can be used to replicate latent/dormant infection commonly seen in humans or serve as a secondary animal model to measure the sterilizing activity of anti-tuberculosis regimens developed using other animal models [58]. Guinea pig use is limited due to a requirement to provide daily vitamin C supplements, a limited repertoire of immunological reagents, and a higher overall cost relative to that of mice.

1.3.5 - Mouse

Mice have many practical advantages in comparison to other models such as less space requirements, ease of handling, low costs, a wide selection of immunological tools, and diverse availability of inbred, outbred, and transgenic strains [59]. C57BL/6 and BALB/c mice have similar immune responses as humans after exposure to Mtb but lack the development of lesion heterogeneity. They will develop a cellular lesion containing primarily epithelioid macrophages, large numbers of lymphocytes, and small amounts of neutrophils. It is uncommon for lesions to

develop necrosis, calcification, or cavitation which is considered a limitation in using traditional mouse models for TB research. Bacilli reside primarily intracellularly, which is in contrast to other animal models and humans that exhibit necrosis where a significant proportion of bacilli reside extracellularly.

While traditional mouse strains do not develop necrosis, several other mouse strains present with caseous necrotic pulmonary lesions upon infection with Mtb. These include the following mouse strains: CBA/J, I/StSnEgYCit, IL-13^{tg}, and C3HeB/FeJ. The CBA/J mouse strain is susceptible to an Mtb infection and has no known immunodeficiency [60]. At approximately 4 to 6 weeks post aerosol TB infection with Mtb Erdman 50-100 colony forming units (CFU) mice will develop caseous necrotic granulomas that may contain fibrosis and bronchiolar intraluminal exudate [61]. Only mice that exhibit an outward sign of sickness tend to have necrotic lesions. Mortality is relatively low, with most mice surviving 32 to 40 weeks after a high-dose aerosol infection.

Mice from the I/StSnEgYCit (I/St) strain are highly susceptible to an Mtb infection caused by the recessive allele Tbc-1^{s} and exhibit disease progression similar to human disease including both chronic and reactivated tuberculosis[62], [63]. Lung lesions occur roughly 2 weeks post intravenous infection with an infectious dose of 5x10^{5} CFU. It begins with an influx of T lymphocytes and neutrophils, and after 3 to 5 weeks post-infection, the formation of necrotic granulomas occur rapidly with hypoxic regions [64]. Similar to other murine animal models, there are higher levels of bacterial burdens in the lungs relative to the other organs. Mortality in this strain is very high, which occurs 3 to 4 weeks after a high-dose infection, but mice generally survive 20 to 25 weeks after a low-dose infection.

The IL-13 -transgenic mouse model overexpresses IL-13 by a transgene containing a genomic fragment containing the IL-13 gene and the human CD2 locus control region [65], [66]. About 3 weeks post aerosol infection with a low dose of 100CFU Mtb H37Rv small granulomatous lesions form, similar to what has been previously observed in BALB/c and C57BL/6 TB-infected

mice. At 6 weeks post-infection, organized central necrotic granulomas develop, which by 10 weeks have matured with a fibrotic capsule with measured levels of hypoxia. At about 14 weeks post-infection, mice presenting with these necrotic granulomas begin to be moribund, and by 20 weeks, most mice have succumbed to the disease [65], [67].

1.3.6 - C3HeB/FeJ Mouse Model

In contrast to traditional mouse models, the C3HeB/FeJ mouse strain develops a heterogeneity of pulmonary lesion types upon an Mtb infection that contains pathology analogous to human disease, including caseous necrotic granulomas. The model is often referred to as the "Kramnik mouse model" for the pioneering work that Dr. Igor Kramnik did in describing and adapting the model to Mtb [68]. C3HeB/FeJ mice are highly susceptible to Mtb infection but have no overt immunodeficiencies. Previous work has identified the intracellular pathogen resistance 1 (IPR1) isoform of the interferon-inducible-75 (Ifi75) gene responsible for this increase in susceptibility [69]. Macrophages from IPR1-negative mice will undergo necrosis instead of apoptosis even with the major pathways of macrophage responsiveness still intact (such as nitric oxide production and gamma interferon upregulation)[70].

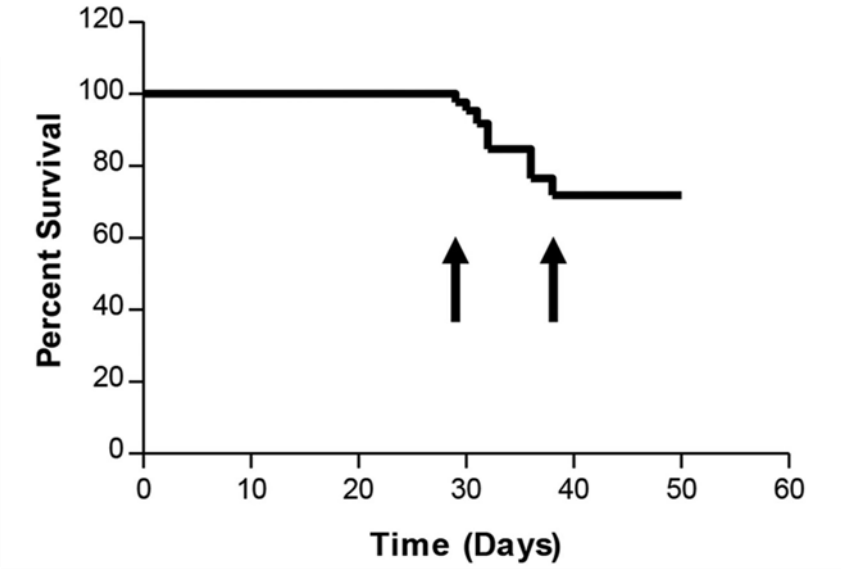


Figure 1.1. The experiment that demonstrates the survival curve and the window of mortality (between arrows) of TB infected C3HeB/FeJ mice (n=85) by aerosol infection.

After a low-dose aerosol infection (of approximately 55 CFU *Mtb* Erdman per mouse), an increase in bacillary load and disease progression is observed. By day 40, the bacterial pulmonary burden will be around 10^8 CFU [71], which is substantially higher than 10^6 CFU observed in BALB/c mice [72]. A significant mortality event, ranging from 10-40%, occurs between 28 to 45 days post-infection (Figure 1.1). The surviving mice will typically stay alive at least an additional 14 weeks with minimal mortality. Mortality can be predicted based upon the pre-infection weight and the development of severe cachexia, hunched posture, and unkempt appearance at about 4 weeks post-infection. It has also been observed that more virulent strains of Mtb, in general, have higher rates of mortality (<80%), while less virulent strains could have no mortality (0%). These early mortality rates are influenced by the culturing method for the infective strain used, the original bacterial source, route of infection, and the *M. tuberculosis* strain itself.

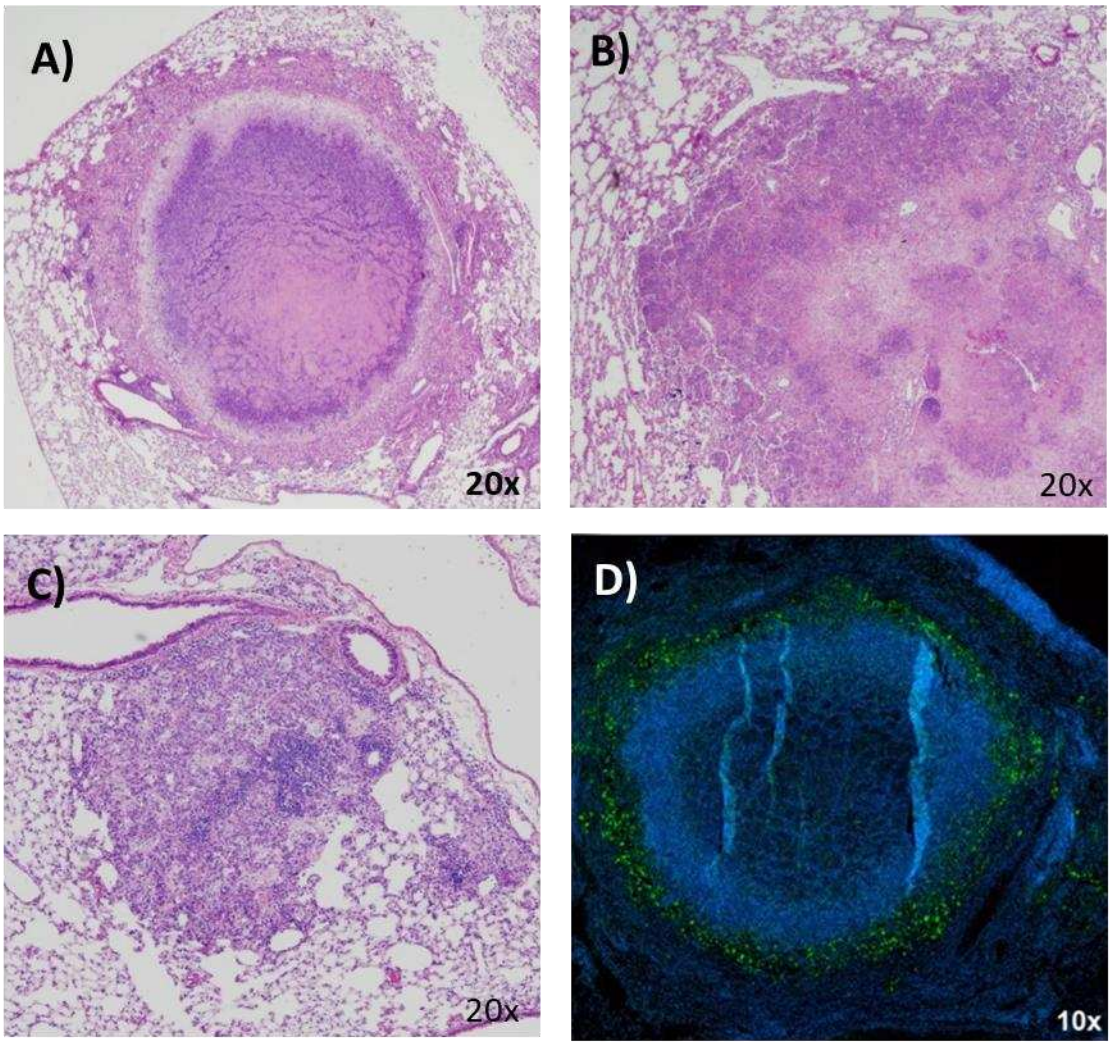


Figure 1.2. Heterogeneity of lesions that develop in the C3HeB/FeJ mice infected with M. tuberculosis Erdman. Type I lesions (A) are highly organized and composed of a caseous necrotic center, a band of neutrophils, and encapsulated in a fibrotic rim. Outside the fibrotic rim, the lesion is composed of epithelioid and activated macrophages, fibroblasts, and few lymphocytes. Type II lesions (B) are composed primarily by neutrophils but lack the fibrosis observed in Type I lesions. Type III (C) lesions are composed predominately of epithelioid and activated macrophages with a large number of lymphocytes throughout the lesion. (D) A Type I Lesion with bacteria stained using SYBR Gold (green) and background tissue stained with DAPI (blue) [73], [74].

After 8 weeks of a low-dose aerosol (LDA) infection with the Mtb Erdman strain, C3HeB/FeJ mice may develop three distinct lesion types, now identified as Type I, Type II, and Type III. Each lesion type develops independently of each other, and multiple lesion types can develop within a single animal. Type I lesions (Figure 1.2A) are fibrous encapsulated caseous necrotic lesions that closely resemble the human TB hallmark granuloma. Visible development

begins 4 to 5 weeks post-LDA infection with the central accumulation of foamy macrophages and neutrophils with margins composed of epithelioid macrophages and lymphocytes. By 7 to 10 weeks post-infection, the granuloma becomes highly organized in appearance. The environment within the core is hypoxic, filled with necrotic debris, and has a neutral pH [74], [75]. The inner ring surrounding the core is composed of a dense ring of neutrophils followed by foamy macrophages, all encapsulated within a fibrotic collagen rim. Outside the fibrotic rim, the lesion tissue is composed of fibroblasts, epithelioid and activated macrophages, and lymphocytes. As observed previously, C3HeB/FeJ mice aerosol infected with mouse passaged *Mtb* H37R or *Mtb* HN878, large Type I lesions may develop cavities in 47% of untreated mice 8 to 14 weeks post-infection, but the rate of incidence is influenced by infectious dose, and Mtb strain used [76]. Bacteria exist both intracellularly in foamy macrophages and extracellularly in the necrotic core (Figure 1.2D).

Type II lesions (Figure 1.2B) are consisting of a fulminant granulocytic pneumonia composed predominately of neutrophils, few lymphocytes, and unlike Type I lesions do not present with fibrosis. Histopathology analysis of early mortality (4 to 6 weeks) mice showed a considerable amount of lung consolidation composed almost entirely by Type II lesions. These lesions are rapidly forming and closely resemble polymorphonuclear alveolitis, sometimes seen in humans [41]. Visual inspection of Type II lesions shows a considerable number of bacteria that appear to be extracellular.

Type III lesions (Figure 1.2C) are indistinguishable from lesions seen in traditional murine models, including BALB/c or C57BL/6. These lesions are composed primarily of epithelioid and foamy macrophages with dispersed lymphocyte clusters. These immune controlled lesions are commonly associated with low mortality and few bacterial numbers. However, there are some exceptional conditions where it had been previously observed with a mouse passaged H37Rv strain that developed a considerable amount of lung consolidation composed of predominately Type III lesions with a corresponding high mortality rate (Unpublished Data).

For my thesis, I have opted to use the C3HeB/FeJ mouse model for a variety of reasons. The first is because it is a murine model we can house a larger number of animals per cohort than we could with other animal models such as Guinea pigs or NHPs. The larger numbers will allow us to have data that is more representative of the population. The second is that we have a large database of historical data, which includes pathology tissue blocks, images, bacterial burden data, and time points. The third is that the animal model develops complex pathology that mimics many aspects of human disease, especially the Type I granuloma. Machine learning requires large amounts of data, which is why the first two reasons were of such importance. Creating a predictive model would not have been feasible otherwise, especially with such complex data classifications. This complexity also allowed us to demonstrate that machine learning is a viable tool to perform targeted histopathology analysis, which currently is not available in the TB field. Lastly, the C3HeB/FeJ mouse model is becoming more widely used, especially in drug efficacy and determining Mtb virulence, which is why we wanted to develop tools to help research laboratories in their analysis.

1.4- Scoring Histology On Microscopy Slides

In clinical diagnosing and monitoring disease, a nominal classification may be sufficient in determining patient treatment options or disease progression. For example, cutaneous basal cell carcinoma does not require additional information because it is the clinical diagnosis. In other cases, having a simple classification may not provide enough information to determine the specific treatment option or disease prognosis. For example, breast carcinoma histology samples are assigned a histology grade based on cell morphology. The grade will determine the aggressiveness of cancer present and the corresponding treatment options [77]. For research purposes, histological scores are given as a way to perform quantifiable analysis across individuals, cohorts, and across independent studies.

A robust scoring system should have three defining characteristics, as suggested by Crissman and colleagues: (1) definable, (2) reproducible, and (3) produce meaningful results [78]. When performing analysis using a predetermined scoring system, several steps need to be taken to generate a useful score [79]. The first is to mask the samples of any identifiers to reduce any preformed biases that may result in the examination of the tissues. Failure to properly mask the samples may inadvertently create an observational bias that may distort the severity of disease or the effects of treatment [80]. The second is to perform a thorough examination of the samples and record all the information that may seem pertinent to the study. Important information may include severity, frequency, and type of lesions that may provide value in a larger scope than the specific study in which the data was collected [81]. The third is following the lesion parameters by defining the type of lesions that can be studied in the study. Defined parameters should encompass what applies to the specific use of the animal models, organ/tissue, and disease under the conditions of the experiment. The fourth is having clear scoring definitions and avoiding collection methods that collect ordinal data such as scoring language that includes terminology such as “mild” or “severe” [81], [82]. A more appropriate reporting methodology is the quantitative reporting of percent involvement, which enhances the reproducibility and sensitivity of an experiment. The fifth and last step is the reduction of diagnostic drift, which is the variation of values consistently during analysis [77]. Bias can happen if there are large numbers of samples, examination occurs over an extended period, category definitions are poorly defined, or multiple pathologists are examining the same cohort of samples. It is common practice in research to have one pathologist score all the microscopy slides in a practical timeframe to increase the consistency of results and limit the amount of variability that occurs between classifiers [77], [78].

When designed correctly, scoring systems should be reproducible and provide data that is impactful for the experiment. There are two commonly used approaches in the validation of a scoring system. The first is the *repeatability of the observer* in scoring consistently amongst a single experiment. A recent study by Begley and Ellis has demonstrated that when using current

standard methodology, it is difficult to achieve reproducibility. They repeated the experiments of 53 published high impact oncology studies to determine if they were reproducible, with only a shocking 11% (6/53) success rate [83]. While there has been an unspoken acknowledgment of issues with irreproducible data, the scientific community, for the most part, has taken the validity of published work as fact. With the rise of drug development costs, late-stage clinical trial failures, and lack of new therapeutics as discussed previously, this lack of reproducibility has begun receiving more attention [84], [85].

The second approach is to validate the *relationship between the relevant parameters of disease severity and the corresponding scores*. The relationship is, at times, calculated using a statistical correlation approach (e.g., Spearmans correlation for nonparametric data) to measure the strength and direction of the association between the two variables. The approach would provide a value between -1 and 1, and as the value approaches zero the lower the correlation. If the scoring system has a value close to 1, it is considered positively correlated, while a negative value indicates a negative correlation depending on the definition [86]. The context associated with the different aspects of the disease severity is important as individual factors may not contribute equally. An example is presented by using the C3HeB/FeJ mouse model development of the three lesion types. In this case, two mice display with 20% lung involvement, but one mouse presents with predominately Type II lesions associated with high mortality, whereas the other mouse has predominately Type III associated with low mortality, a score of 20% lung involvement would not be sufficient to classify these animals. Instead, each lesion type context needs to be taken into consideration with more weight for the Type II lesion classification and a less weighted value for the Type III lesions in the analysis.

1.4.1 - Histology Scoring Of Tuberculosis-Infected Tissues On Microscopic Slides

Histological scoring in tuberculosis research primarily focuses on pulmonary pathology and is highly dependent on the animal model in question. Different scoring methodologies are

being used in the field for the multiple TB animal models. In NHPs, the pathology focus is on pulmonary lesion characteristics such as the lesion architecture appearance, type, distribution patterns, and cellular composition. Based on the listed criteria, a score of 0-3 is applied for each characteristic, and an overall score is calculated [87]. In the rabbit model, the focus is primarily on pulmonary lesion characteristics but also includes the general condition of the animal. Parameters include the animal's weight, lesion type, lesion location (which lung lobe), pleural involvement, number of lesions, and granuloma presence on the mediastinal lymph node, kidneys, and appendix. Disease severity is calculated by the summation of individual pathology scores, with a maximum score of 50 that can be attained for the most severe cases [88]. In classical mouse models, such as BALB/c, each histological parameter of peribronchiolitis, perivasculitis, alveolitis, and granuloma formation is scored. The score ranges from 0-5, with 5 being the most severe form for that particular parameter. The lesion frequency and general lung involvement determine overall disease severity. A maximum score of 20 is what can be achieved using this system [89]. In the scoring of pathology of the C3HeB/FeJ mouse model in our laboratory by board-certified pathologist collaborators, each lesion type (I, II, III) is scored from 0-4. A score of 0 is defined as 0% lung involvement, a score of 1 is defined as 0-10% lung involvement, a score of 2 is defined as 10-25% involvement, a score of 3 means 25-50% involvement, and a score of 4 means greater than 50% involvement. Each lesion is also weighted differently with the Type I lesion score multiplied by 3, Type II lesion score multiplied by 4, and Type III multiplied by 2. Therefore, an animal can have a maximum score of 36, but most severe cases receive a score of around 20.

As demonstrated above, within the tuberculosis field, there is no standardized method to score histopathology in determining the efficacy of treatments or disease severity. It is subject to what has been previously done in the specific animal models, laboratories, and individual's preferences. This makes comparisons across different studies or laboratories difficult unless the same pathologist reviews the studies, following the same scoring criteria, with a short time period

between analyses. A more reproducible and informative methodology would be to provide exact measurements for each parameter that is of interest to the disease. An example using the TB C3HeB/FeJ mouse model would be to report the percent involvement for each specific lesion type. This will move the data from being nominal or ordinal to a ratio data type that can be used for statistical measurements and data analysis. The unbiased quantitive data decreases the user variability and increases the reproducibility of the results regardless of the time or individual performing the analysis. With the widespread availability of digital image analysis and machine learning scoring automation of analysis, not only the variability but the amount of time required by a human user has been significantly reduced.

1.5 - Image Analysis

1.5.1 - Introduction

Computer vision (CV) is an interdisciplinary field that aims to mimic human vision through the detection and/or classification of objects from visual media such as images, videos, or sensors. It has allowed digital optical analysis to be automated, to be faster, and more reliable than human operators [90]. Human visual analysis, while highly complex, is plagued with 'perceived' rather than 'factual' information when making decisions. This is because we have evolved to make quick decisions based upon visual stimuli that may be lacking key information or the ability to process large amounts of stimuli. After light hits an object, it takes about a tenth of a second for our brains to process this information, but during that tenth of a second, we are missing information that is still being provided and haven't yet been processed [91]. Our brains will fill in these blind spots with information from its surroundings and contextual clues, which is indistinguishable from the real information [92]. This approach is called 'top-down' visual processing, which is more computationally efficient than 'bottom-up' visual processing [93]. Our visual processing method impacts areas of intra and inter-observer variability, and the perception of dimension, color, and size [94], [95]. We can identify our visual shortcomings in the form of

optical illusions. While these neural shortcuts may have provided a competitive advantage for avoiding predators to our ancestors, in today's context, it delivers many unperceived disadvantages when performing manual visual tasks that require precision and reproducibility. In contrast, the biggest advantage humans currently have over computer vision is the flexibility, or the ease of transitioning between crystallized and fluid intelligence, when encountering deviating or previously unseen scenarios [96]. With a simplistic example, this means that if you had created a CV algorithm to classify images of cats and dogs and you gave it an image of a bird, it would try to classify it as either a dog or cat. Humans do not have this limitation because even if we had never seen a bird before, we would recognize it as neither a cat nor a dog and give it its own classification. With current technology limitations, there is still a need in research and clinical settings to not remove the human element in their analysis but complement the strengths of each approach by implementing a hybrid solution [97], [98].

1.5.2 - Techniques In Computer Vision And Image Processing

CV works by implementing an optical sensor (e.g., camera) and an algorithm to extract information about an image object. Most CV solutions utilize similar steps as listed: (1) digital image formation or capture using the optical sensor, (2) image preprocessing to improve or enhance desired qualities or features, (3) image segmentation in which objects of possible interest are separated from background noise, (4) quantification of object features, and (5) image interpretation [99] (Figure 1.3).

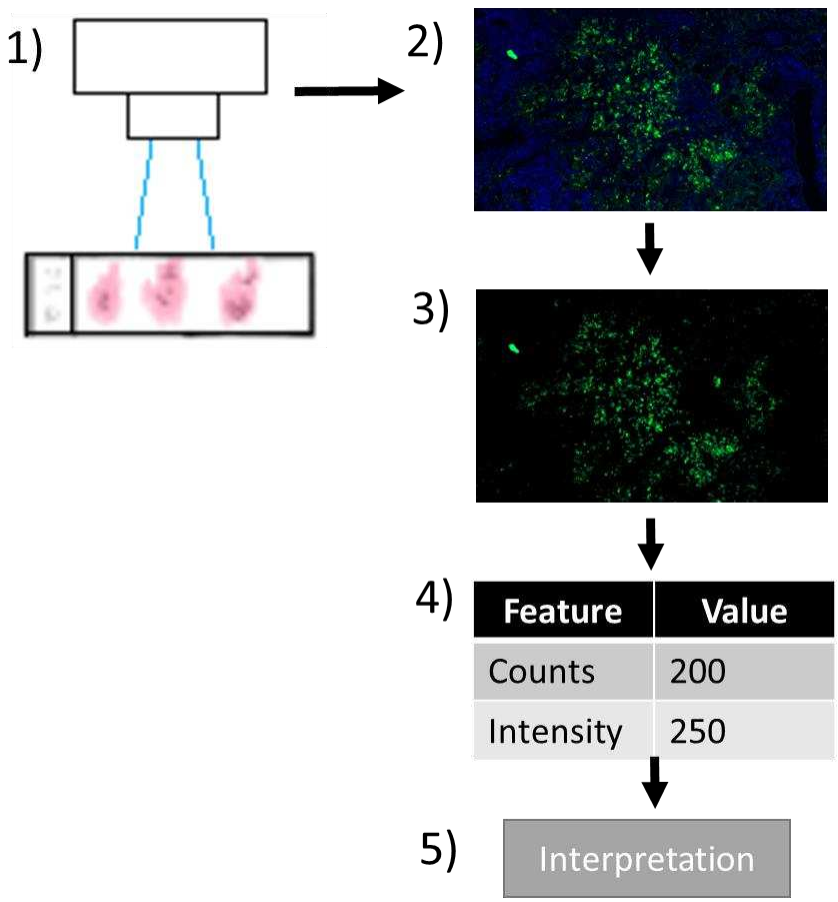


Figure 1.3. The steps used in image analysis: (1) digital image formation, (2) image preprocessing, (3) image segmentation, (4) image quantification, and (5) image interpretation

Image preprocessing aims to either correct the image quality or enhance for features that are of interest. Most images are subject to changes in illumination, atmospheric conditions, viewing angle, and/or instrument errors [100], [101]. In biomedical research, most digital image quality limitations are not the result of poor external conditions around the sensor, but of the sample itself. This is because the optical sensor (microscope, MRI, CT/PET, etc.) is in very controlled environments that have little deviation in angles, lighting, and exposure. Specifically, in microscopy, sample complexity and deviations in the preparation and staining of microscopic sections have the greatest influence on the degree of image pre-processing that may be required [102]. Common techniques used, include but are not limited to image blurring, sharpening,

normalization (color, size), distortion, erosion and dilation, morphological processing, and outlining [103] (Figure 1.4).

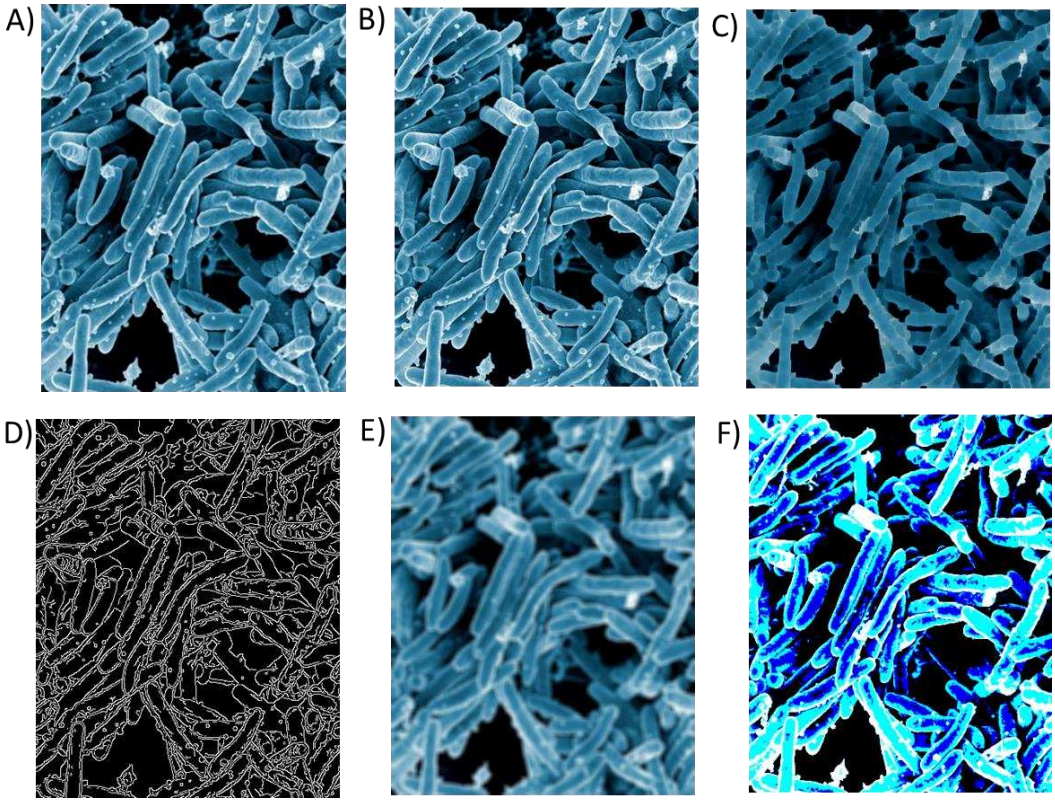


Figure 1.4. A comparison of the commonly used image analysis techniques: (A) Original image, (B) normalization, (C) erosion, (D) edge detection, (E) blurring, and (F) threshold segmentation. Image (A) taken from the NIAID website under the creative common license [104]

Image segmentation is the method of partitioning an image into separate, distinct objects. Each pixel within an image is assigned a label, and pixels that receive the same label are determined to be part of the same object. This is a common approach to focus on areas of significance, determine the location within the image as well as boundaries, and to reduce the amount of processing that may need to be performed for additional tasks. Image segmentation can be roughly divided into two different types; sematic and instance segmentation. Sematic segmentation assigns each pixel as either background or object. Instance segmentation assigns each pixel as a particular class; however, each class can also be considered a separate instance.

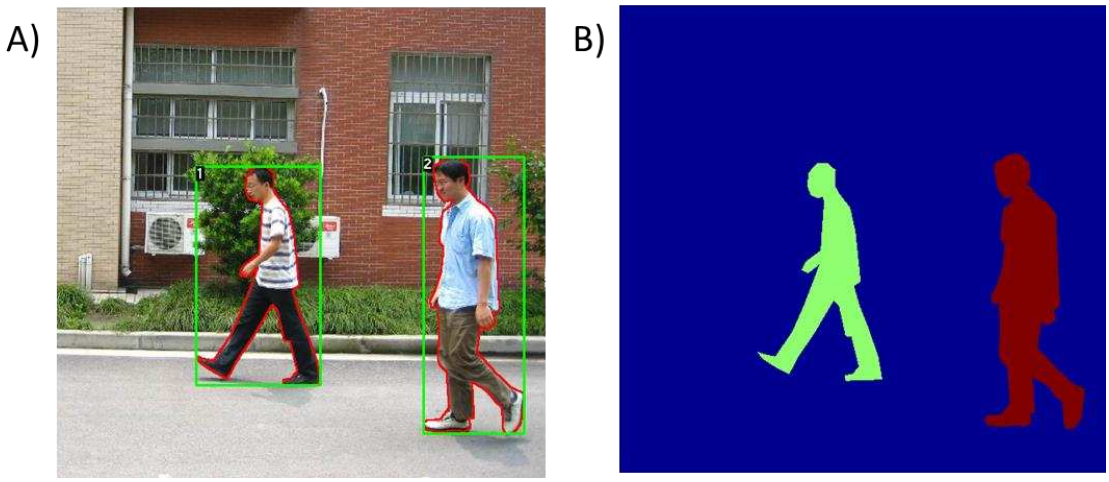


Figure 1.5. An example of cluster-based image segmentation for pedestrian detection. (A) An image is uploaded, clusters of pixels with similar characteristics are identified, and (B) objects of interest (green and red) are separated from the background (blue). Images were taken from the Penn-Fudan Database for Pedestrian Detection and Segmentation [105].

A common segmentation technique is “threshold segmentation” (Figure 1.4F), which partitions each area of an image-based upon pixel values. Calculations for this technique are simple, fast, and perform well with high contrast (e.g., fluorescence imaging). The method performs poorly if there is not a substantial amount of contrast or if there is an overlap in pixel values. Another approach is “edge detection segmentation,” which looks for discontinuous local features within an image. Also, in this case, this technique suffers if there is not a substantial amount of contrast between adjacent discontinuous pixels and if there is a high degree of complexity. The last approach for segmentation is clustering-based image segmentation. Clustering techniques divide pixel populations into a specified number of groups or ‘clusters.’ Pixels within a particular cluster have increased similarity of characteristics than those of other clusters. This approach is simple to implement but is computationally expensive to use, which limits it to only small datasets. A common example of image segmentation is detecting pedestrians. The algorithm determines which objects are of interest (the pedestrians) and subtracts the background (Figure 1.5). By implementing a segmentation approach, you can determine where the pedestrian is in the image.

Quantification of image objects measures specific characteristics of the segmented objects. This can include but is not limited to the number of objects, region of interest (ROI) intensities, ROI hues, ROI saturation, areas, shapes, and locations. While this information on its own is valuable, it may also be used in providing a classification (interpretation). For example, when providing grades for breast carcinoma, the information required includes each cell's shape, color, and the number of carcinoma cells relative to healthy cells. The interpretation step can either be automated to give a recommendation based upon the parameters or left up to a specialist who is trained to look for specific features from the quantification step. This approach is commonly used in microbiology settings to distinguish between the differently labeled cells using multiplex fluorescent imaging.

1.5.3 - Machine Learning

Arguably one of the biggest advances in CV is the implementation of machine learning (ML) to perform trivial and highly complicated CV tasks. ML can be defined as the implementation of computer algorithms and statistical models to perform specified tasks without having to program in explicit instructions. ML can be broken down into two main categories, 'supervised' and 'unsupervised' learning. Supervised learning uses a ground truth that 'learns' by using examples of the desired output. An example would be developing a classification algorithm that distinguishes between dogs and cats. Common supervised learning algorithms include methods such as logistic regression, naïve Bayes, support vector machines, and neural networks. Unsupervised learning does not have labeled outputs, so it is up to the algorithm to determine the output structure present within a given dataset. This approach is popular in data exploration when the researcher is in the process of discovering new information on data. Common unsupervised algorithms include k-means clustering, principal component analysis, autoencoders, and neural networks. Because of the flexibility of approaches developed for neural networks, they are utilized

in both supervised and unsupervised learning but are considered very computationally expensive [106], [107].

Supervised ML models are developed using representative data called 'training sets' to discover patterns through inference. Below is an illustrative example of the difference between traditional CV and ML in solving an image analysis problem. If a researcher wanted to enumerate fluorescently labeled bacteria within an image using traditional CV, they would calculate the hue, saturation, and lightness of representative pixel values of the bacteria versus the background. A range of values for each metric listed above will be collected, and objects of interest (bacteria) will be masked out. Masking is the process in which portions of the image are hidden, leaving only the specified pixels of interest. Segmentation will follow and further masking based upon cell size to eliminate objects that are considered false positives. The last step would be to implement a shape-specific contour algorithm, followed by enumerating the objects that are present. Traditional CV requires the user to program explicit instructions on how to detect the object in question, which can get complicated and can be non-specific depending on the circumstance. The benefit is that supervised ML only requires a small amount of training data and in some instances, can be fast to implement. A researcher can create a training set of representative images of bacteria and train an ML algorithm. There would be no need to perform preprocessing, determine pixel value ranges, or object sizes as the ML algorithm could infer these patterns on its own. The main advantage is that there is no prior information needed to train the algorithm, which is important when one is unsure about what the critical parameters are. The disadvantage is that depending on the complexity of the problem, the ML algorithm may require thousands if not millions of training set image data points to achieve a level of accuracy that would be acceptable. In the biological sciences field, acquiring a thousand separate images of multiple classifications may not be realistic, especially for digital histological data. This is why both approaches are still relevant in creating algorithms to analyze digital images.

In CV, there are a number of algorithms used in ML, which include convolutional neural networks (CNN), deep belief networks and deep Boltzmann machines, and stacked autoencoders. For the purpose of this thesis, we will focus on CNNs and their applications in object classification. The conception for CNN was inspired by the visual mechanisms and neural processing that takes place in biological organisms [108]. Just like a human brain works by the firing of highly connected neurons in response to a stimulus to create an output, CNNs follow a similar framework. An input value is passed to a neuron, which is either turned off or on. This continues for multiple connected neurons until an output is generated. Early implementations of neural connectivity, hierarchical organized transformations of images, and later additions of error gradient proved very successful [109]–[111]. This field is highly evolving, and over the last couple of years, more efficient algorithms and alternative CNN approaches have become available. In comparison to other image classification algorithms, little pre-processing is needed.

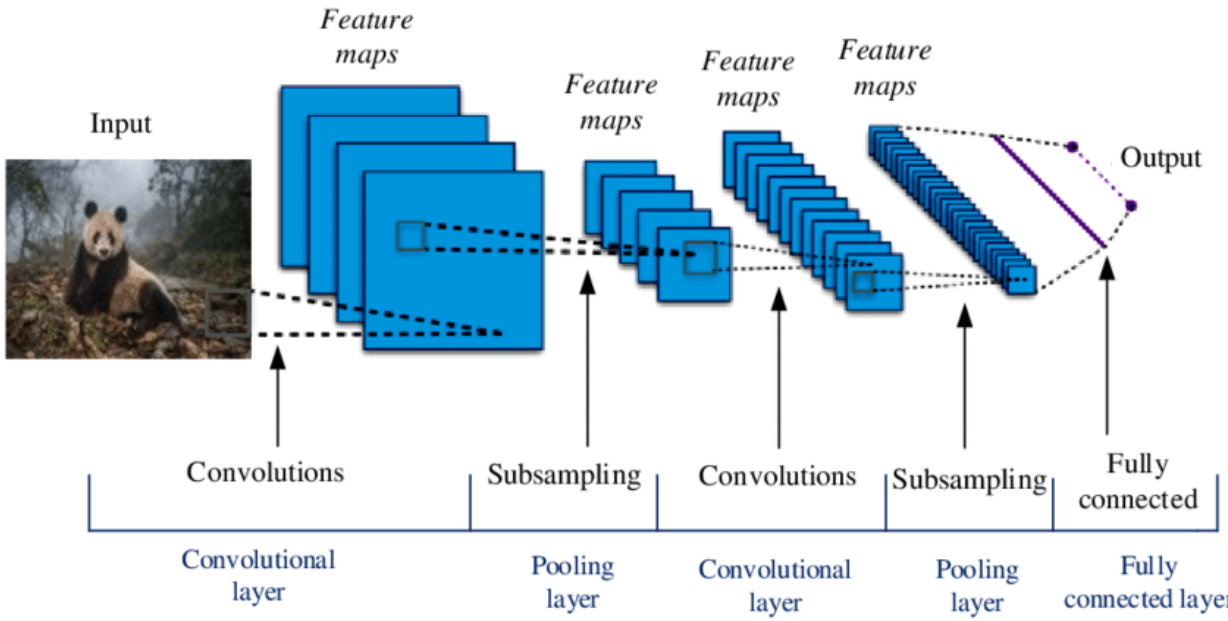


Figure 1.6. A visual representation of the architecture of a common CNN. This includes multiple convolutional and pooling layers before being processed by a fully connected layer. Adapted from Kaymak et al. [112].

At its core, a CNN is comprised of three main types of neural layers: (1) convolutional layers, (2) pooling layers, and (3) fully connected layers (Figure 1.6). Convolutional layers utilize various kernels to convolute an entire image and its intermediate feature maps into multiple feature maps. The purpose of a kernel is to take data as an input and transform it into a predetermined form. From a biology perspective, this step mimics human vision, where specific identifying features of an image such as color, shapes, and size are identified in the image. Pooling layers reduce the spatial dimensionality of the input image for the next convolutional layer. The reduction in size leads to a loss of information but is actually beneficial due to a decrease in computational cost and prevents overfitting of the network. Average pooling and max pooling are the two most common pooling algorithms used in this step [113]. The last step takes the feature maps that were created in the convolutional and pooling steps and passes the flattened image through a fully connected neural network. From a practical standpoint, this is when the learning takes place in distinguishing between the different images. Within a fully connected layer, there are multiple neurons, and each neuron is connected to the activation of the previous layer. Each neuron activation is computed by matrix multiplication and the bias offset. Eventually, the fully connected layers will convert the two dimension feature map into a one dimension feature vector. The derived vector will then be fed forward into the specified number of classification categories to produce the final classification result [114].

1.5.4 - Human Bias In Research

Historically, pathologists have been using a manual histopathology scoring system to analyze pathology images which are considered to be semi-quantitative at best. Manual scoring produces qualitative or semi-quantitative scores that are rich in detail about the disease but even with experienced pathologists can be subjective to user bias and interpretation [115]. Bias, or systemic error, could alter the results or diminish the validity of a study [80], [116]. As mentioned previously, with the rising cost of drug development and the difficulty for the science community

to generate reproducible work, this has increasingly become a larger concern. There are many potential sources of bias that may unintentionally arise when scoring pathology and these are often referred to as visual traps.

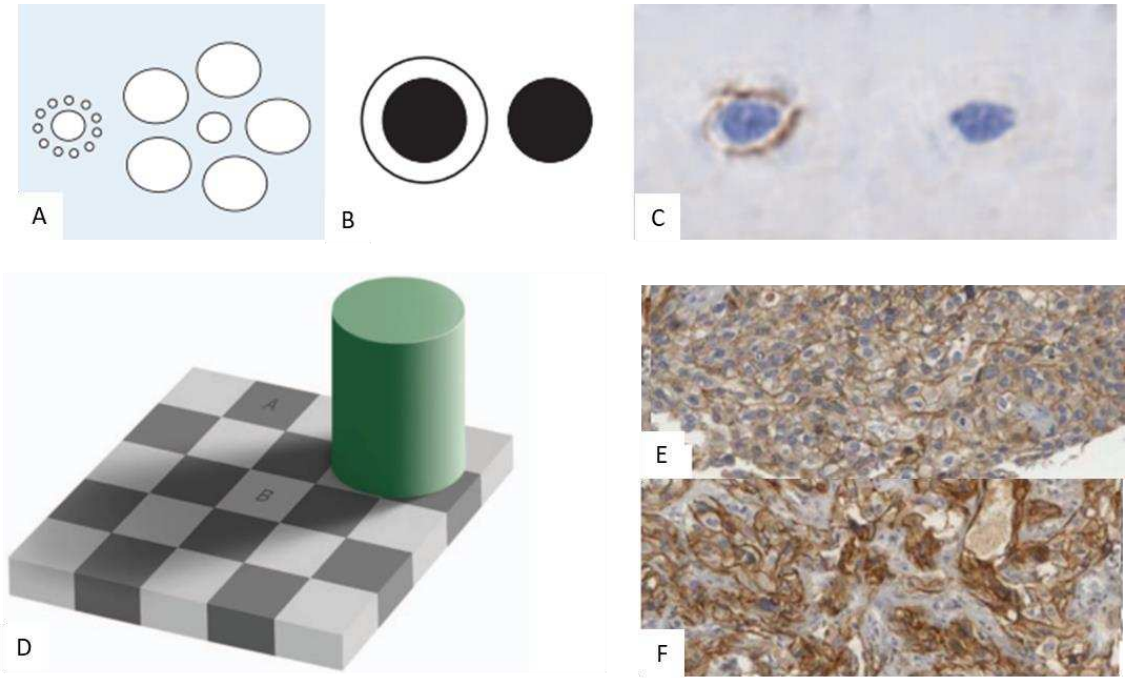


Figure 1.7. Examples of known visual biases, or optical illusions that can impact the ability of human researchers to make judgments and the corresponding pathology equivalent. Both (A) and (B) demonstrate how contextual clues influence our perception of size. This is also seen in pathology of a cell whose membrane is shown stained and unstained (C). The checkerboard illusion (D) demonstrates our inability to discern color gradients properly. Both checkers labeled A and B are of the exact same color. The pathology equivalent is shown where two images (E & F) have the same amount of staining (Brown) present. Adapted from Aeffner et al. [117].

Several potential visual trap scenarios have been described, such as the Ebbinghaus illusion of size, which demonstrates that our perception of size is influenced by the context in which it is presented [118], [119]. In Figure 1.7A, B, the inner circles are of the same size but contextually appear different due to the surrounding outer circles. Illusions of size is influenced by preconceived concepts and memory of the observer. This can be conceptualized in histopathology analysis of a perceived size difference for instance, by the addition of a labeled membrane to a nucleus (Figure 1.7C) [117]. The Checker Shadow Illusion (Figure 1.7D) is a visual

trap that demonstrates the faulty perception of brightness, contrast, and surrounding variation[120], [121]. Throughout our lives, we have visualized countless instances of objects that appear darker in shadows than in light. We have seen so many that our brains automatically assume that it is the same as we have previously experienced and make these compensations, whether they are real or not. Combined with our perception of ignoring soft edges of shadows and visualizing the sharp edges of the squares makes this illusion possible. Lastly, the distinction between hues and gradients of colors can also vary tremendously between individuals. It is estimated that the average human can distinguish between 30,000 to 10 million colors [122] that can be impacted by several factors such as the environment, age, health, and lifestyle choices. Color is perceived by not only the light refracted by a specific color but also by the light nearby, which is called color induction [123], [124]. For example, a dark yellow surrounded by red will appear as a greenish tint. In contrast, a 12-bit camera can distinguish 68 billion different shades of color, and a 14-bit camera can store up to 4 trillion shades [125]. Digital systems are far more superior to the human visual system in capturing hue and color gradient differences. A pathology example that illustrates the checker shadow illusion and color induction is demonstrated in Figure 1.7E, F. This example shows that two different levels of cytoplasmic staining in the context of adjacent membrane staining clearly influence the assigned scores. Using digital analysis, it is determined that both images would actually receive equivalent scores.

In addition to visual traps, human objectivity is also hampered by cognitive biases. This includes confirmation bias, which is defined as unknowingly interpreting new information as confirmation of previously held beliefs [126], [127]. An example would be the belief that left-handed people are more creative than right-handed people. The next time you run into a left-handed artist, regardless of reality, it will only confirm that bias. This is evident when an investigator will inadvertently seek and interpret data that support their given hypothesis. Confirmation bias is often reduced by blinding the data and implementing the scientific method when acquiring data. Another cognitive bias is diagnostic drift, which is the variation of score

values in a consistent manner throughout a study. Even experienced pathologists and clinicians are subject to diagnostic drift, especially during long studies with multiple time points [128], [129]. While this section is not an exhaustive review of potential human bias in research, it helps underscore the need for integration of an unbiased automated digital analysis in the pathology scoring system.

1.5.5 - Computer Vision (CV) In Clinical And Biomedical Research

Computer-aided analysis and diagnosis combine the flexibility and training of a human specialist with the reproducibility and accuracy of a machine. ML became feasible for biological laboratories with the development of graphical processing units (GPUs) in the early 2000s. GPUs are composed of hundreds of simple cores that perform thousands of concurrent hardware threads. Central processing units (CPU), in comparison, have a couple of complex cores with single-thread performance optimization. What this means is that GPUs can perform thousands of simple tasks simultaneously while a CPU can perform a few complex tasks at a time. The viability of CNN became apparent in 2012 when training with the ImageNet dataset had an error rate 10.8 percentage points lower than the next runner up which used more standard methodology [114]. Since 2019 it has been referenced over 40,000 times and is considered one of the most influence events in computer vision.

In the early 2000s, the threshold for entry in ML implementation was reduced thanks to advances in GPU in both cost and performance. In 2008, the utility of computational pathology had been clearly demonstrated by Thomas Fuchs with predictive analytical modeling of renal cell carcinoma in patients [130]. In 2012, the introduction of CNN in ML showed a significant improvement over the previous methodology for image analysis [131]. And in 2014, the development of generative adversarial nets have lowered the amount of training data required which has been a limitation for biological research [132].

AI has expanded into applications of biomedical diagnostic systems from previously other human domains such as facial recognition and self-driving cars. This includes oncology, pathology, radiology, and infectious diseases, which have seen rapid advancements in their respective fields [133]. Digital image analysis has allowed the high throughput of data at a relatively low cost while increasing the efficiency of the analyzing specialists. The TB field has seen a relatively slow expansion of ML compared to fields such as oncology; nonetheless, the TB field has begun utilizing these new methodologies to improve their analysis [134]–[136].

1.6 – Rationale

The objective of this dissertation was to develop computational analysis tools that take a targeted approach on acquiring data on specific host disease states and bacilli populations. These are metrics that are potentially important in evaluating the efficacy of novel and existing therapies for TB. Current targeted approaches can be time-consuming and require multiple steps, which introduce additional steps where errors can occur. They are also subject to multiple forms of user bias and rely on specialized knowledge to be able to accomplish accurately. These problems are further compounded when analyzing complex pathology, such as what is developed by the TB C3HeB/FeJ mouse model. To accurately analyze this murine model, the classifier needs to look at what pathology structures are formed and what the cellular makeup is for each lesion. With recent advancements in ML and CV, we were able to develop specialized analysis tools, which are called Lesion Image Recognition Analysis (LIRA) and Mycobacterial Image Analysis (MIA). Each tool required algorithm development, analysis verification using existing methodologies, and determining the limitations of employing these tools in future analysis. We propose that analysis done with these tools will result in faster analysis, an increased agreement among human classifiers, and more insightful data in determining the efficacy of drug therapies.

Specific Aim 1: To develop software utilizing deep convolutional neural networks that will assist in the identification and quantification of pulmonary lesions using digital microscopic images from the lungs of *M. tuberculosis*-infected C3HeB/FeJ mice. Hypothesis: Pathologists that are aided by a computational tool, such as LIRA, in their analysis will show a reduction in the amount of time needed to perform the analysis and more agreement between different pathologists than pathologists who use standard methodology.

Specific Aim 2: Compare pathologist results using LIRA or those made by a certified pathologist using actual tissue slides from a previous drug-efficacy trial. Analysis will include a review of the original study conclusions. Hypothesis: Histology, from a drug efficacy trial, when analyzed with the assistance of LIRA, will generate scores that are in agreement with a board-certified pathologist. In addition, using a more targeted analysis approach will expand upon the initial conclusions that were made.

Specific Aim 3: To generate software for the enumeration of fluorescently stained bacteria in pulmonary tissues of C3HeB/FeJ mice infected with *M. tuberculosis*. Hypothesis: Analysis of *Mycobacterium tuberculosis* located in infected C3HeB/FeJ mice tissue using MIA will identify differences in bacterial enumeration metrics (density, % involvement, size) among treatment cohorts and lesion types.

The development of computational tools to provide quantifiable data for analysis is an important tool for drug development. For histopathology analysis, these computational tools will allow the reduction in variability found between pathologist scoring while reducing the analysis time and scoring errors seen previously. It also allows researchers to collect metrics that previously were not available, which may lead to new discoveries. The metrics that are being generated using the novel software for histopathology include the percent lung involvement for

each lesion type, as well as metrics on bacterial density, shapes, and fluorescent intensities for bacterial populations found within each lesion microenvironment. In conjunction with fluorescent labeling for specific cells, this will allow researchers to quantify the effects of therapies on both the host and bacteria in targeted microenvironments within the host. This targeted data analysis approach will help expand our knowledge used in the development of new therapies and speed up the time required for new drug development.

CHAPTER 2: DEVELOP LESION IDENTIFICATION AND AREA CALCULATION SOFTWARE (LIRA) FOR THE C3HEB/FEJ MOUSE MODEL[1]

2.1- Introduction

Tuberculosis (TB) is a communicable disease that continues to affect modern society with a quarter of the world believed to have active or latent TB, primarily in low and middle-income countries [4], [137]. TB is transmitted by aerosol inhalation of the bacterium *Mycobacterium tuberculosis* (Mtb) from an infected individual. During the course of infection, a wide variety of pulmonary disease lesion presentations may concurrently present within the same host [16], [138], [139] . This pulmonary pathology includes, but is not limited to, inflammatory lesions, interstitial pneumonia, necrotic caseating granulomas encapsulated within a fibrotic rim, non-cavitary necrotic lesions, non-necrotic cellular lesions, and cavitary lesions [14], [41]. The heterogeneity of lesions within a single individual represent a myriad of microenvironments for Mtb that range from hypoxic regions, where bacteria are extracellular in caseum, to vascularized and more aerated regions, where Mtb is intracellular within various macrophage populations [140],[141]. Because of the complex environments that develop in the lung throughout infection, Mtb has developed adaptive strategies to alter its metabolism and replicative state to increase survival [22], [73], [142], [143]. The adaptation of Mtb to its environment results in bacterial populations exhibiting multiple distinct phenotypes [19], [140]. A critical component in understanding disease outcome and, importantly, treatment success is our understanding of

[1] Authors: Bryce C. Asay, Blake Blue Edwards, Jenna Andrews, Brendan K. Podell, Juan F. Muñoz Gutiérrez, Chad B. Frank, Forgivemore Magunda, Michael Lyons, Asa Ben-Hur and Anne J. Lenaerts. Bryce C. Asay conceptualized the project/software, created the training set, contributed to the coding, pipeline development, design of the validation testing, and performed the validation analysis. Paper under review in Scientific Reports.

these varying lesion microenvironments and bacterial phenotypes [144]–[146]. Two important clinical implications emerge from the lesion heterogeneity observed in the lungs of TB patients in terms of the effectiveness of drug treatment in the various lesion types and lesion compartments. Lesion characteristics differentially affect (1) drug penetration and distribution of TB drugs inside the lung lesions where the bacteria are located [147], and (2) the relative drug susceptibility of the bacilli contained within these lesions due to metabolic adaptation that could induce phenotypic drug tolerance and bacterial persistence [140], [148]. These combined factors result in marked inter- and intra-lesional variability of drug-mediated killing. We have reported previously on the C3HeB/FeJ mouse model, which after Mtb aerosol infection, develops a heterogeneity of pulmonary lesion types more reflective of the lung pathology seen in TB patients, and which has proven useful to further our understanding of disease progression as well as studying treatment responses for emerging therapies [22], [71], [74], [149], [150].

Murine models are employed in TB research due to their small size and low cost, and in addition, their physiology and genetics are well understood. Standard laboratory mouse strains such as C57BL/6 and BALB/c have been used most widely for TB research. However, these mouse strains show limitations by solely developing a single lesion type after Mtb infection, and they lack the lesion heterogeneity seen in TB patients. The C3HeB/FeJ TB infection model presents with a heterogeneity of lesion types more reflective of human disease [71], [74]. This mouse model was first described for tuberculosis by Igor Kramnik et al. (and is therefore also more commonly referred to as “the Kramnik mouse model”) [68]. These mice developed a *de novo* recessive allele, identified in a region at the 54.0-cM location of chromosome 1, termed the ‘super susceptibility to tuberculosis -1’ locus (sst1) [69]. The susceptible sst1 allele was reported to control the formation of caseous necrosis of pulmonary lesions. In C3HeB/FeJ mice after an Mtb infection with a low inoculum, the bacterial load increases to high bacterial numbers in lungs, and mice develop a chronic disease state with a dramatic progression of the lung pathology over time. After 8 weeks post-infection, three pulmonary lesion types can be identified which we

classified earlier [71], [74], as: (i) a highly organized encapsulated caseous necrotic granuloma defined as a Type I lesion, (ii) a neutrophil dominated lesion with cellular necrosis defined as a Type II lesion, and (iii) cellular lesions with distinct clusters of lymphocytes defined as a Type III lesion type. Both the Type I and II lesion types contain high bacterial numbers (approximately 10^{6-7} per lesion), whereas the smaller Type III lesions only show few bacteria (approximately 10^{2-3} per lesion). All three lesions types can be present at the same time within a single mouse, and even in a single lung lobe [74]. Occasionally cavities have been observed, although this has been an infrequent event in the C3HeB/FeJ mouse model infected with Mtb H37Rv or Erdman [38], [71].

The efficacy of tuberculosis vaccines [151], [152] or drugs [20], [153] is primarily measured by determining the reduction in bacterial load in target organs. The quantification of lung involvement in disease has proven to be informative as a secondary readout. A thorough pathology assessment can be informative to assess the effect of the therapeutic intervention on the disease itself, by studying improvement or worsening of inflammation or measures of tissue repair, such as fibrosis, as well as revealing potential immunotoxicity. Serial PET/CT scans of *M. tuberculosis*-infected animals, and TB patients have been used to monitor disease progression and response to treatment [53], [154], [155], allowing precise quantitation of the extent of disease and inflammatory response of the host. Parallel studies of CM and humans have shown comparable rates of radiological response to linezolid [156], and based on these initial studies, radiographic surrogate markers are being actively explored for use in human clinical trials of new agents and TB regimens [45], [157], [158]. For mouse models used in preclinical testing for TB, the treatment effect on pulmonary pathology is generally examined on microscopic images using a histological grading system by a veterinary pathologist specialized in TB [159]. The pathologist then assesses both macroscopic recognition of larger pathologic structures such as lesion types, as well as microscopic identification of predominant immune cell types. Because of the complex nature of the classification process, quantification of the lesions and their lung involvement for research purposes is not only a specialized task but highly time-consuming and variable upon the

individual performing the analysis. The person-to-person variability and subject nature of the method can lead to differing results, and this will, in addition, make inter-experiment comparisons of the histopathology results difficult. The time-consuming manual analysis can lead to increased user fatigue and reduced focus, which in itself can lead to the introduction of more inaccuracies.

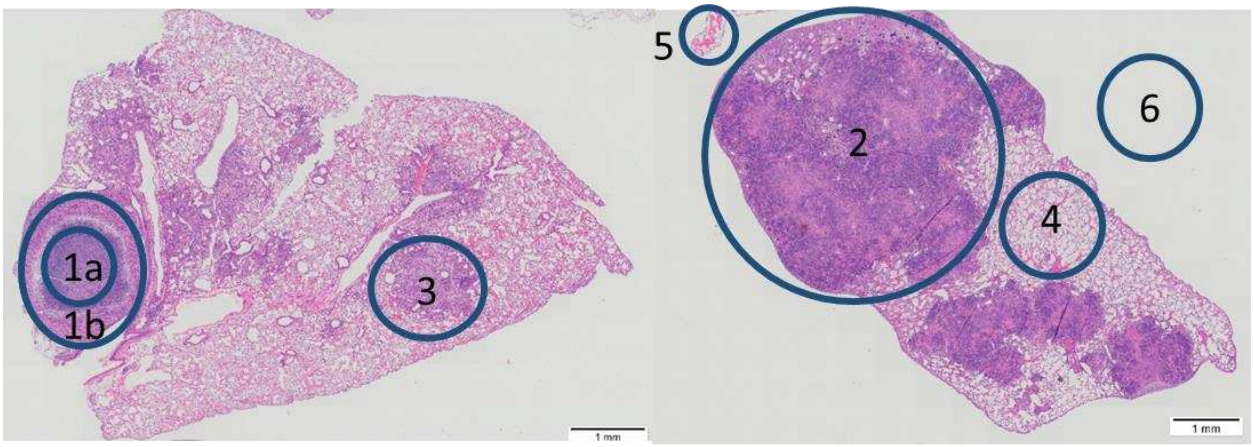


Figure 2.1

Classification	Number	Color	Prominent Features
Type I Core	1a	Red	Central necrosis surrounded by a dense neutrophilic debris layer, and a foamy macrophage layer which lies within the collagen rim.
Type I Rim	1b	Blue	Collagen rim surrounded by an outside layer composed of fibroblasts, epithelioid and activated macrophages, and clusters of lymphocytic cells.
Type II	2	Green	Large numbers of neutrophils at times visible as a honeycomb-like structure of lung parenchyma, peripheral DNA NETs. Along the periphery of the Type II lesion pockets of lymphocytes and compressed tissue can be observed. However, these are not included in our definition (see Type III and Healthy Tissue).
Type III	3	Yellow	Cellular inflammatory lesions composed primarily of epithelioid, foamy and activated macrophages, large numbers of lymphocytes, and small pockets of neutrophils.
Healthy	4	Pink	Healthy lung parenchyma, bronchiole structures excluding the airway, compressed tissue, blood arteries, veins and vessels, and any other pulmonary tissues that are commonly found in healthy individuals.
Miscellaneous (Misc.)	5	Purple	Classification encompasses a large repertoire of tissue pathology features that would not normally be found in healthy tissue, is not part of one of the specified TB lesions, or is not associated with the pulmonary tissue. The miscellaneous classification includes pleural pneumonia, adipose tissue, non-TB specific inflammation, as well as staining and processing artifacts.
Empty	6	None	Areas containing no sample, bronchial airways, or tissue tears as the result of histology preparation.

While histopathology analysis of TB lung lesions is complex, recent advances in machine learning can provide for certain aspects of such analyses to be automated. Areas such as digital image analysis in cancer research [25], [26], [160], magnetic resonance imaging / computed

tomography [161],[162], and clinical and anatomical pathology [163],[164] have seen an increase of analysis tools being developed using machine learning approaches. Many of the algorithms currently employed in the medical digital imaging field include CNNs [165]–[168], and support vector machines (SVM) [169],[170]. CNNs are neural networks designed specifically for images and address the issue of achieving classification invariance for object recognition using local feature extractors or filters. These feature extractors or filters are applied at increasing levels of granularity, thereby allowing the system to recognize features at increasing levels of abstraction similarly to how the brain processes visual stimuli [171], [172]. CNNs also have a demonstrated history of success in image classification, and a plethora of open-source tools and packages are widely available for use [27].

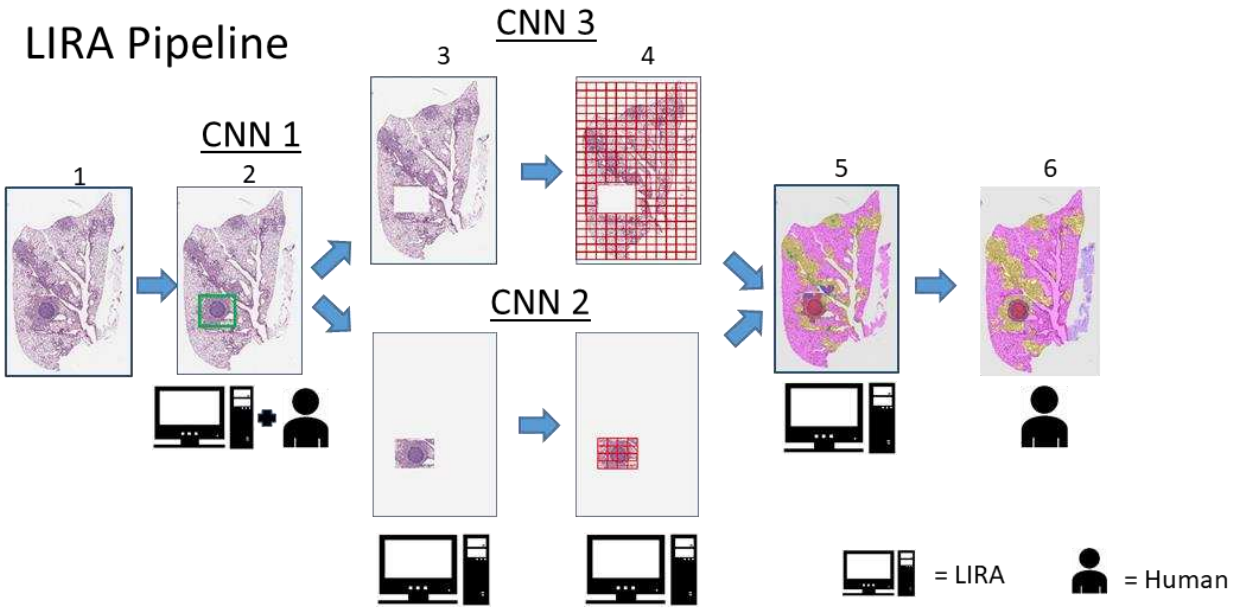


Figure 2.2. Graphical representation of the workflow required to analyze images. 1) Original image is uploaded , 2) CNN1 detects Type I lesions and is verified by a user, 3) Window location is cropped based upon classification, 4) Image broken up in 80 x 145 windows to generate our predictions, 5) Classification predictions generated using 1 of two CNNs, 6) User is allowed to edit the results one last time.

In this work, we propose a TB lesion machine learning classifier that follows established classification criteria developed in our laboratory, which distinguishes three different lesion types that develop in the C3HeB/FeJ mouse TB infection model [71], [74]. This digital image analysis

software package was developed as a modular neural network [173]–[176], consisting of three CNNs, each optimized for a specific sub-task, together with two human intervention checkpoints to limit the probability of misclassification (Figure 2.1, 2.2). This digital image analysis pipeline is named 'Lesion Image Recognition and Analysis', or LIRA. From its inception, LIRA was designed to work in conjunction with pathologists with the goal to improve the analysis of pulmonary pathology to be more efficient, accurate, and reproducible regardless of the individual analyzing the digital image slides.

2.2 - Materials And Methods

2.2.1 - Sample Collection From Archived Animal Studies

The digital images from pulmonary samples used for the development of the image analysis software package were derived from archived C3HeB/FeJ mouse studies conducted for the development of the mouse model [71], [74], for drug efficacy trials using existing and novel drugs for tuberculosis [22], [149], and studies focusing on virulence of various Mtb strains (unpublished results). Lung samples were collected from 14 previous independent mouse studies, conducted over several years (from 2012-2018), and samples were obtained from two different research laboratories at CSU. Mtb strains used for the mouse infection studies included *M. tuberculosis* Erdman (TMC 107, purchased from ATCC), HN878 (Clinical Isolate W210, CSU, Fort Collins, CO, available at BEI resources), or H37Rv (Trudeau Institute, Saranac Lake, New York). Bacteria for mouse infections were initially grown as a pellicle and further propagated in Proskauer-Beck (PB) medium containing 0.05% Tween 80 (Sigma Chemical Co., St. Louis, MO), never extending past mid-log phase [177]. Female C3HeB/FeJ mice, 6-8 weeks of age (Jackson Laboratories, Bar Harbor, ME) were housed in a bio-safety level III animal facility and exposed to an Mtb aerosol infection using a Glas-Col inhalation exposure system [178]. Mice were euthanized by CO_2 inhalation, between 4 to 10 weeks post aerosol exposure. Whole lungs were fixed by inflation with 4% paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA) in

phosphate-buffered saline solution (PBS) via cardiac perfusion before subsequently being transferred to a 70% ethanol solution 48 hours later.

2.2.2 - Processing Of Samples To Generate The Digital Image Dataset

Fixed lung samples were further processed at either Premier Laboratory (Boulder, CO), the Experimental Pathology Facility at CSU (Fort Collins, CO), or at the Microbiology, Immunology, and Pathology (MIP) Department. Pulmonary tissue samples were placed in a single cassette per mouse, then paraffin-embedded and sectioned. Tissue sections (5µm) with the largest surface area were used for further histopathology analysis and mounted on glass slides. Paraffin sections processed for Haemotoxylin and Eosin (H&E) staining underwent a Xylene bath for 10 min before undergoing gradient hydration. Samples were stained using a Leica ST5020 instrument with an initial staining step using Hematoxylin 560 (Leica Selectech, Buffalo Grove, IL) for 5 min and rinsed in tap water. Next, the slides were washed using a Define wash step (Leica Selectech, Buffalo Grove, IL) for 10 sec before being placed in a Blue Buffer Solution (Leica Selectech, Buffalo Grove, IL) for 1 min. The slides were then rinsed with Eosin for 1 min before being transferred to 96 Eosin Y 515 (Leica Selectech, Buffalo Grove, IL) for 3 min. The final step included a washing step with 96% EtOH for 6 min before being mounted using a LEICA CV5030 Mounting instrument (Leica Selectech, Buffalo Grove, IL), thereby using the Coverseal-X Xylene Mounting Media (Cancer Diagnostic Inc., Durham, NC). Slides were generated and stained at three facilities using similar, however not the exact same, protocols over a time span of several years (2012-2018). Digital image scans from microscopic slides containing C3HeB/FeJ lungs were generated from one of three microscopes. Image scans from Premier Laboratories (Longmont, CO) were generated on an Aperio Scanscope XT microscope (Nikon, Melville, NY) at 20x magnification. At the Experimental Pathology Core (CSU), the images were acquired using

an Olympus VS120 microscope (Olympus, Center Valley, PA) using the Olympus VS-ASW software (v.2.9; Olympus, Center Valley, PA) at 20x magnification. Images acquired within the MIP Department were generated on a Nikon Eclipse Ti microscope (Nikon, Melville, NY) at 20x magnification, using the Nikon NIS Elements AR software (v. 4.51.00; Nikon, Melville, NY). In total, for the training and subsequent validation of LIRA, lung sections from 176 mice were scanned digitally to generate 176 digital image scans (containing all five lung lobes for every mouse per microscopic slide and per digital image scan). Because of the heavy computational load that occurs from the large size of pathology scans, digital image files were generated separately for every individual lung lobe (resulting in five digital image files per mouse).

2.2.3 - Lesion And Histopathology Classification Scheme

The histopathology classifications used for the creation of the training, test, and validation sets included the original classifications of pulmonary lesion types in the C3HeB/FeJ mouse model, as previously described by Lenaerts et al. [71], [74]. Briefly, C3HeB/FeJ mice present with three distinct pulmonary lesion types after an Mtb aerosol infection. Type I lesions show well organized, caseous, necrotic lesions with a layer of foamy macrophages around a core composed of neutrophilic debris, which is surrounded by a collagen rim with interstitial macrophages admixed within the rim. Type I lesions contain high bacterial numbers, which are either extracellularly located in the lesion core, or intracellular in foamy macrophages. The Type II lesions are less organized with a massive recruitment of neutrophils, resulting in large areas of alveolar wall necrosis throughout the lung parenchyma. Type II lesions also contain high bacterial numbers, both intra- and extracellularly. Type III lesions develop as a result of the accumulation of lymphocytes, epithelioid, and foamy macrophages as well as small pockets of neutrophils. The latter is similar to the main lesion type described for other immunocompetent mouse strains, such as the BALB/c and C57BL/6 [72], [179], [180].

In addition to the lesion classifications, other histopathology categories were added for the machine learning approach to ensure the inclusion of healthy tissue, non-Mtb specific lung pathology, artifacts, and an empty slide feature. In total, seven classification categories were identified, which were as follows: Type I Rim, Type I Core, Type II, Type III, Healthy, Miscellaneous (Misc.), and Empty slide. Each classification is characterized by certain macroscopic pathology features, as well as microscopic features based on multiple different cell types. Each of the seven classifications is described in detail below and are presented in Figure 1.

The Type I (caseous necrotic) pulmonary lesion type is characterized by two distinct pathology features, which include a collagen rim and a caseous necrotic core. For this reason, the Type I lesion was categorized into two distinct classifications for the analysis, defined here as the 'Type I Rim' and 'Type I Core' categories. The Type I Core is composed of central necrosis surrounded by a dense neutrophilic debris layer, and a foamy macrophage layer which lies within the collagen rim. The Type I Rim composition includes the collagen rim and the outside layer composed of fibroblasts, epithelioid, and activated macrophages, and clusters of lymphocytic cells. The 'Type II' pulmonary lesion consists primarily of large numbers of neutrophils at times visible as a honeycomb-like structure of lung parenchyma. Along the periphery of the Type II lesion, the presence of aggregates of extracellular DNA, pockets of lymphocytes, and compressed tissue can be present. These pathology features are mostly found in close proximity and, therefore, also classified as Type II. The 'Type III' pulmonary lesions are cellular inflammatory lesions composed primarily of epithelioid, foamy and activated macrophages, large numbers of lymphocytes, and small pockets of neutrophils. The 'Healthy' classification is here reserved for tissue features, such as healthy lung parenchyma, bronchiole epithelium, blood vessels, compressed tissue, and any other pulmonary tissues that are present in healthy individuals. The 'miscellaneous' classification encompasses a broad array of tissue pathology features that are not found in healthy tissue, that are not part of the specified TB lesions, and are

not associated with healthy pulmonary tissue. The miscellaneous classification includes pleural pneumonia, adipose tissue, non-TB specific inflammation, as well as staining and processing artifacts. Lastly, the 'empty slide' classification was included as a separate category, although this classification was ultimately not included in the final performance and accuracy analysis of the results.

2.2.4 - The LIRA Pipeline

LIRA was developed as a modular neural network. This implementation of multiple neural networks working together follows the natural logic employed by trained researchers to classify lesions by intuitively first investigating on a more general, macroscopic level (for Type I lesion recognition), followed by a more detailed analysis on a microscopic level to identify other pathological features on a cellular level. In preliminary work, we used a single CNN to identify all seven classifications. However, due to similarities between multiple lesion types at a microscopic scale (such as the necrosis encountered in a Type I Core and a Type II lesion), multiple neural networks operating at different visual scales (macro and micro) were used to limit misclassification of lesion types. The final approach, illustrated in Figure 2.2, employs a modular architecture [173]–[176] with three CNNs and two human intervention checkpoints. After a digital image is uploaded, in a first step, the macro-classifier (CNN1) identifies the Type I physiological structure with its characteristic caseous necrotic center and collagen rim. Human feedback is required after this initial step to reduce the false positive rate of Type II lesions misclassified as Type I lesions (see also in result section 1). The macro-classifier CNN1 will then identify the individual image areas either positively or negatively for the Type I pathology features. Based upon the macro-classifier CNN1 prediction, the digital image areas are partitioned into two separate data sets, classified as 'Type I' or 'not Type I' lesion. Subsequently, both the positively or negatively identified image areas are further analyzed at a microscopic level based on cell types, by either the CNN2 or the CNN3 micro-classifiers. Both CNN2 and CNN3 require that the digital image sections are tiled

into individual image patches that are 80x145 pixels in size. In case a Type I lesion is detected after CNN1, then CNN2 analyzes these image patches further and classifies each image patch as either Type I Rim, Type I Core, Healthy Tissue, Miscellaneous, or Empty. The remaining image patches (not identified by CNN1) are in parallel further classified with CNN3 as either Type II, Type III, Healthy, Miscellaneous, or Empty.

The output of the implementation of the three CNNs uses raw image patch counts (RIPC) per classification category, which are the sum of the tiled image patches for each classification. RPIC are classified, classification locations recorded, and subsequently visualized as colored overlays on the digital image. Every classification category is represented by a different color for easy visualization, qualitative analysis and result verification (See Figure 2.1). The classifications with unique colors for the seven pathology classifications are then overlaid and smoothed with a node labeling algorithm [181], with each image patch classification being considered an individual node, onto the original H&E stained scanned digital image. The digital image of the lung lobe with colored overlay is then inspected and verified by the user at a second human intervention checkpoint, and at this point, the user can make changes before the final classification calculations are performed (Figure 2.2). Ultimately, the final output of LIRA using the two human intervention checkpoints includes the enumeration of the various lesion types and the calculation of area of lung involvement per classification category (presented as % lung involvement).

2.2.5 - Neural Network Training

All CNNs in the LIRA pipeline were trained using the Keras 2.0 machine learning framework in Python 3.5, on an NVIDIA Titan X GPU. Using a built-in human-in-the-loop classification tool for LIRA, trained research technicians assisted in creating the training set for CNN1 by collecting 791,000 individual image patches with a patch size of 128x128 pixels. The purpose of these patches was solely to identify the Type I lesion structure at a macroscopic level. Additionally, 1.2 and 1.3 million image patches with a patch size of 80x145 pixels were collected

in a similar manner for further classification at a microscopic level for both the Type I (CNN2) and Non-Type I lesion classification (CNN3), respectively.

Each CNN model was subsequently trained on the provided training set of image patches, with the following parameters as implemented in Keras 2.0. Values for training and validation were initially chosen using a Bayesian Optimization Algorithm, and values were further adjusted by hand. The composition of each of the neural networks is presented in Table 2.1.

Table 2.1. Tables explaining the specific details on the creation of our model. The top table is the makeup of the Type I detector (CNN1), and the bottom table contains the information for the other two CNN (CNN2&3).

Macroscopic CNN Model (Type 1 Detection Model)

Layer Type	Filter Size	Stride	Activation Function	Input Size	Output Size
Convolutional	7x7x16	1x1	ReLU	128x128x3	128x128x16
Max Pooling	2x2	2x2	N/A	128x128x16	64x64x16
Convolution	7x7x32	1x1	ReLU	64x64x16	64x64x32
Max Pooling	2x2	2x2	N/A	64x64x32	32x32x32
Convolution	3x3x32	1x1	ReLU	32x32x32	32x32x32
Convolution	3x3x64	1x1	ReLU	32x32x32	32x32x64
Max Pooling	2x2	2x2	N/A	32x32x64	16x16x64
Convolution	3x3x64	1x1	ReLU	16x16x64	16x16x64
Convolution	3x3x128	1x1	ReLU	16x16x64	16x16x128
Max Pooling	2x2	2x2	N/A	16x16x128	8x8x128
Convolution	3x3x64	1x1	ReLU	8x8x128	8x8x64
Avg Pooling	8x8	8x8	N/A	8x8x64	1x1x64
Flatten	N/A	N/A	N/A	1x1x64	64
Dense/Output	N/A	N/A	Softmax	64	2

Microscopic CNN Models (Both Lesion Classification Models)

Layer Type	Filter Size	Stride	Activation Function	Input Size	Output Size
Convolution	7x12x20	1x1	Sigmoid	80x145x3	74x134x20
Max Pooling	2x2	2x2	N/A	74x134x20	37x67x20
Convolution	6x10x40	1x1	Sigmoid	37x67x20	32x58x40
Max Pooling	2x2	2x2	N/A	32x58x40	16x29x40
Flatten	N/A	N/A	N/A	16x29x40	18560
Fully Connected	N/A	N/A	Sigmoid	18560	1024
Fully Connected	N/A	N/A	Sigmoid	1024	100
Fully Connected	N/A	N/A	Softmax	100	4

2.2.6 Validation Of Software

After training the CNNs, a validation step was included using a separate digital image set. Validation of the entire LIRA pipeline was performed by comparing results to a classical manual pathology readout. First, the performance of the LIRA software was assessed to recognize the various lesion types, i.e., calculation of the percent error per lesion type by LIRA when compared to the analysis by the experienced TB researcher (as baseline). Second, the percent agreement between four board-certified veterinary pathologists (American College of Veterinary Pathologists) was calculated by comparing the results after classifying image patches with and without LIRA assistance. Lastly, the histopathology readout was measured more robustly by calculating a reliability coefficient developed to measure the agreement among observers, using the Krippendorff's alpha coefficient [182]. The coefficient was calculated and based on data generated by four board-certified veterinary pathologists with and without the use of LIRA, as well as four research technicians with the use of LIRA.

To assess the agreement in the analysis by the pathologists with and without the use of LIRA, seven randomly selected scanned images of single lung lobes were selected for the software validation step. In a first step, the pathologists hand labeled each 80x145 pixel image patch using one out of seven classifications (same as described for LIRA). Data on every 80x145 pixel-sized image patch classification was collected as RIPC. The hand-labeled result by the pathologists and technicians generated a seven-color classification overlay for every digital image. The result also included the calculated RIPC for each of the seven classification categories for every digital image of a single lung lobe. In a second step, seven days later, the pathologists and research technicians were presented with the same seven digital images of lung lobes to review, this time with the assistance of LIRA predictions for both the macroscopic (CNN1) and microscopic classifications (CNN2, CNN3). Both the pathologists and technicians were requested to review the LIRA predictions, and to intervene and potentially modify the Type I lesion detection at the first intervention step (CNN1 LIRA prediction) and/or the individual image patch classifications at the second intervention step (CNN2, CNN3 LIRA prediction) if needed.

The analysis of the validation results included the calculation of the percent agreement in prediction results between the four pathologists[183]. Using percent agreement does not take into account that classifications may have been selected by chance; it is, however, still a commonly used metric and useful as a first measurement in the context of this work [184], [185]. The formula is $PA=N_{xA}N_{yA}+N_D\times100$ where PA is percent agreement, NA is the number of agreed upon ratings divided by the total sum of cases, and ND is the total cases of disagreements.

As mentioned before, the classical histopathology readout is only semi-quantitative and in some cases subjective, which makes comparisons of the LIRA prediction results to a gold standard unreliable. To study the variance in readouts between the different pathologists after manually reviewing the digital images, we opted for an inter-annotator comparison approach to calculate the agreement among the pathologists with and without LIRA. The Krippendorff's alpha coefficient calculation using nominal data and any number of observers was implemented here as an analysis tool to quantify the agreement between the prediction results of multiple observers with or without assistance with LIRA. The formula is provided below, where n_{uc} = the number of values c assigned to unit u. $n_{uc} \leq$ m observers, n_{uk} by analogy $n_{u.} = \Sigma_c n_{uc}$ = the number of values assigned to unit u, $n_{.c} = \Sigma_{u|nu.\geq 2} n_{uc}$ = the number of pairable values c occurring in the reliability data (omitting all units with lone or no values: $n_{u.} \leq 1$) $n.. = \Sigma_{u|nu.\geq 2} n_{u.}$ = the total number of all pairable values in the reliability data (omitting all units with lone or no values: $nu. \leq 1$); $n.. \leq mN$.

$$_{metric}\alpha = 1 - \frac{D_o}{D_e} = 1 - (n.. - 1)\frac{\sum_u \frac{1}{n_u. - 1}\sum_c \sum_{k>c} n_{uc} n_{uk\,metric}\delta^2_{ck}}{\sum_c \sum_{k>c} n_{.c} n_{.k\,metric}\delta^2_{ck}}$$

The Krippendorff's alpha approach takes into account chance classifications, can be used with nominal data, using more than two raters, with multiple categories, plus the approach can be used with large datasets [186]. Each image patch was considered as a separate observation, and the classifications from all four pathologists were used for calculations. The Krippendorff's alpha

approach provides a coefficient between 0 and 1, with α < 0.667 being rejected, α ≥ 0.667 being the lowest acceptable limit, and α≥ 0.800 being considered to have good agreement [187].

2.2.7 - Evaluation Metrics For Performance And Accuracy Of The Neural Networks

For CNN1 the sensitivity was calculated by $\frac{True\ Positive}{(True\ Positive+False\ Negative)} * 100$ and specificity was calculated using $\frac{True\ Negative}{(True\ Negative+False\ Positive)} * 100$. These numbers were acquired by counting the number of true positives, true negatives, false positives, and false negatives within 70 digital images of lung lobes that had been classified by the CNN1 detector. This approach did not stipulate that the entire lesion was detected, but any area of the lesion be correctly or incorrectly identified.

In order to determine the accuracy of CNN2 and CNN3, a standard or ground truth classification was established on 12 digital images by a researcher with substantial experience analyzing C3HeB/FeJ mouse pathology (author Bryce C. Asay). The researcher manually labeled each image patch using one of the seven different classifications described earlier. The researcher was also the main individual training LIRA, and therefore this analysis was mainly aimed to investigate whether LIRA performed adequately and met expectations. The 12 digital images containing a lung lobe consisted of three images primarily composed of Healthy Tissue, three images primarily composed of Type I lesions, three images composed of Type II images, and three images composed primarily of Type III lesions. Total image patch counts for only the classification of interest were then used to calculate the percent error for each individual classification for LIRA compared to the manual labeling result by the researcher.

2.2.8 - Computer Code Availability

At publication, the computer code will be made open source to the research community (Github), and we will include a statement in the Methods section, under the heading "Code

availability", indicating where and how the code can be accessed without any restrictions. This will be made publicly available through the Lenaerts lab Github repository.

2.2.9 - Acknowledgments

We would like to thank the TB Drug Accelerator (BMGF) and the Global Alliance for TB Drug Development for their assistance with earlier mouse studies, which provided the necessary histopathology samples for the training of the new software. We thank the Bill & Melinda Gates Foundation for funding this work, Grant number OPP 1126594.

2.3 - RESULTS

2.3.1 - Ability Of CNN1 To Identify Caseous Necrotic (Type I) Lesions

After training CNN1 with the digital image training set, a validation step was introduced to evaluate the sensitivity, specificity, and false-positive rate of classifying Type I lesion cores. After an image is uploaded, the Type I object detector (CNN1) is used to macroscopically detect the caseous necrotic center of Type I lesions (Figure 2.1, 2.2). LIRAs CNN1, the Type I object detector, showed to be consistent in identifying and delineating Type I lesion structures, with a sensitivity of 86.36%. CNN1, however, shows a high false-positive rate by misclassifying other lesion types for a Type I with the specificity of 55.11%. Most misclassifications occurred when large areas of cellular necrosis were visible in the center of Type II lesions, which on a single image patch shows a similar cellular composition as the Type I cores. The high recall rate is likely the result of the training set for the creation of CNN1, having an insufficient number of images needed for training. To decrease the false positive rate for CNN1, continued training with a larger training set might be an area of improvement for future iterations. Despite the high recall rate, we opted to retain CNN1 as it facilitated the rapid identification of Type I lesions for the user, and it provided an accurate demarcation of its margins from surrounding lesions or tissue. For instance, the validation results showed that the inclusion of CNN1 decreased the inconsistencies seen after

hand labeling among multiple pathologists in identifying and demarcation of the Type I Rim. Overall, the addition of CNN1 proved to be a substantial time saver as more time is allocated to accurately identify the margins of the Type I lesions than in the removal of potential misclassifications.

2.3.2 - The Accuracy For LIRA In Comparison To An Experienced TB Researcher

Table 2.2. LIRAs accuracy measured by percent error, which measures the precision of each of the predicted classifications.

	LIRA % Error
Healthy Tissue (n=3)	1.975
Type I - Caseum (n=3)	17.324
Type I - Rim (n=3)	8.324
Type II (n=3)	38.374
Type III (n=3)	2.829
Total (n=12)	12.107

After the initial positive or negative classification by the macroscopic object detector CNN1, the areas designated as Type I were further classified by CNN2, and the remaining image patches were classified in parallel by CNN3 (Figure 2.1, 2.2). To determine the accuracy of the overall pipeline (Figure 2.2), a ground truth was established on 12 digital images by a researcher with substantial experience analyzing C3HeB/FeJ mouse pathology. These 12 images consisted of lung lobes composed of three Type I dominated tissues, three Type II dominated tissues, three Type III dominated tissues, and three predominately Healthy tissues. Total image patch counts for only the classification of interest were used to calculate the percent error per individual classification. The results for LIRA without human intervention were then compared to the manual labeling results by the researcher. As seen in Table 2.2, the overall percent error for all seven

classifications was 12.11% across all 12 digital images. Healthy Tissue and Type III had the lowest percent error, while Type II had the highest percent error at 38.37%.

The high percent error observed for the Type II lesions for LIRA compared to the experienced researcher was the result of poor demarcation by LIRA of the surrounding lesion tissue, which is often composed of compressed tissue and in some cases, aggregates of extracellular DNA (potentially NETs). Compressed tissue is the result of lesions pressing into healthy lung parenchyma, which changes the morphological structure. The aggregates of extracellular DNA are considered part of a Type II lesion, whereas compressed tissue, as defined here, is considered “Healthy Tissue”. Another area contributing to the decrease in accuracy by LIRA was from processing artifacts to generate the microscope slides. This included areas of high red blood cell number, which were incorrectly classified by LIRA as Type III. Also, minor differences in the staining procedure of the microscope slide affected the intensity and colors of the tissue and its images, which resulted in slight differences in classification by LIRA. Lastly, sectioning artifacts such as microtome occasional knife chatter, lines, or fragment holes were also prone to result in misclassification by LIRA.

2.3.3 - Improved Agreement Of Histopathology Classifications Among Pathologists Using LIRA

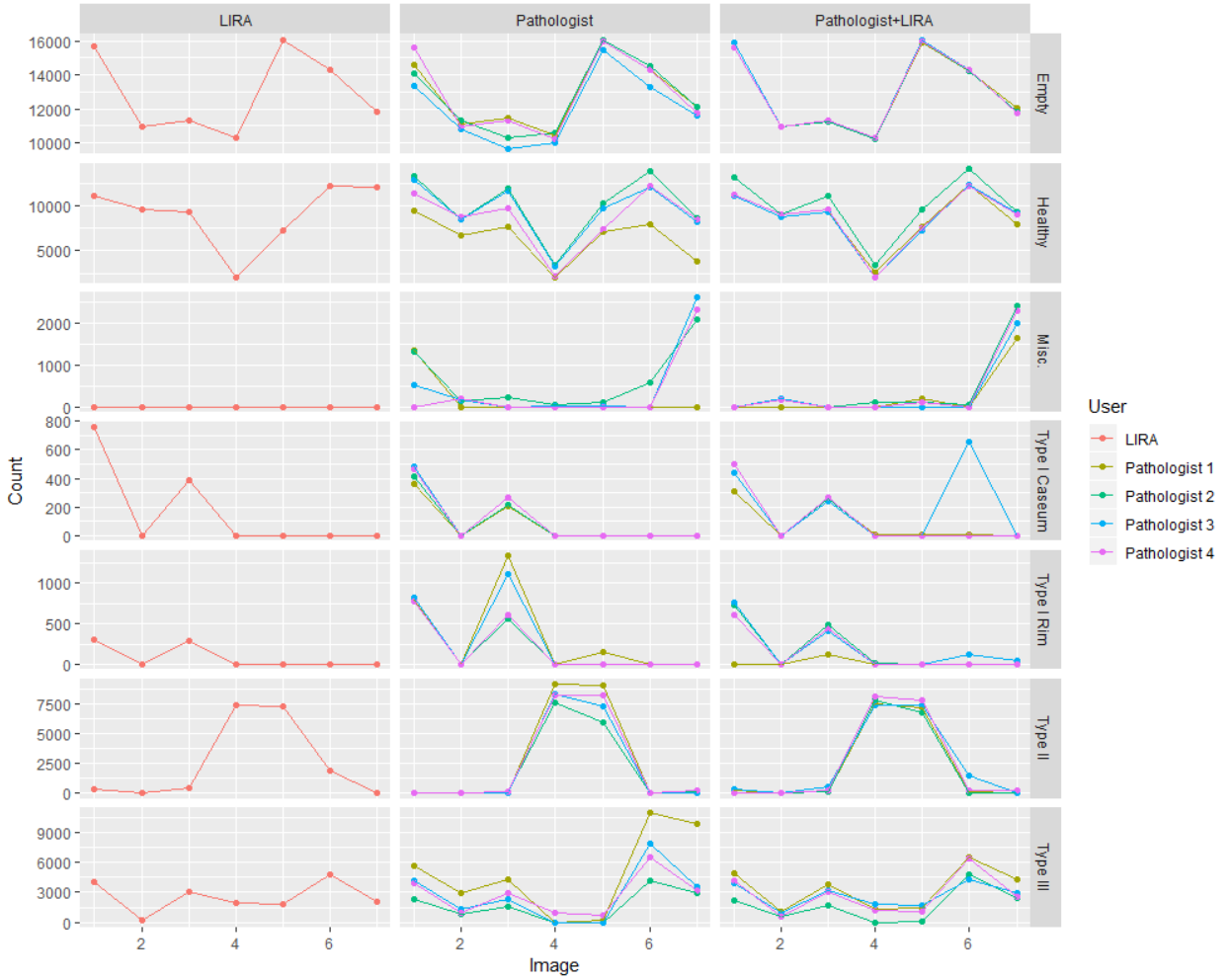


Figure 2.3. Line graph representation of each cohort: LIRA, Pathologist, Pathologist+LIRA. Left Y-axis is the Raw Image Counts generated during classifications, Right Y-axis is each of the classifications, and the bottom X-axis is the image number. Each colored line represents an individual pathologist or LIRA. As visualized, there is a smaller range of raw image patch counts when the pathologists are assisted by LIRA (Pathologist+LIRA) than when the pathologists are unassisted (Pathologists). Many of the values that the Pathologists+LIRA gave are only marginally different than the classifications that LIRA proposed.

To evaluate the performance of LIRA in terms of reproducibility and concordance between users, a second validation step was included using a previously unseen digital image set. Seven randomly-selected scanned images of single lung lobes were used for this validation step. The goal for this validation step was to compare the pathology classification results from the classical

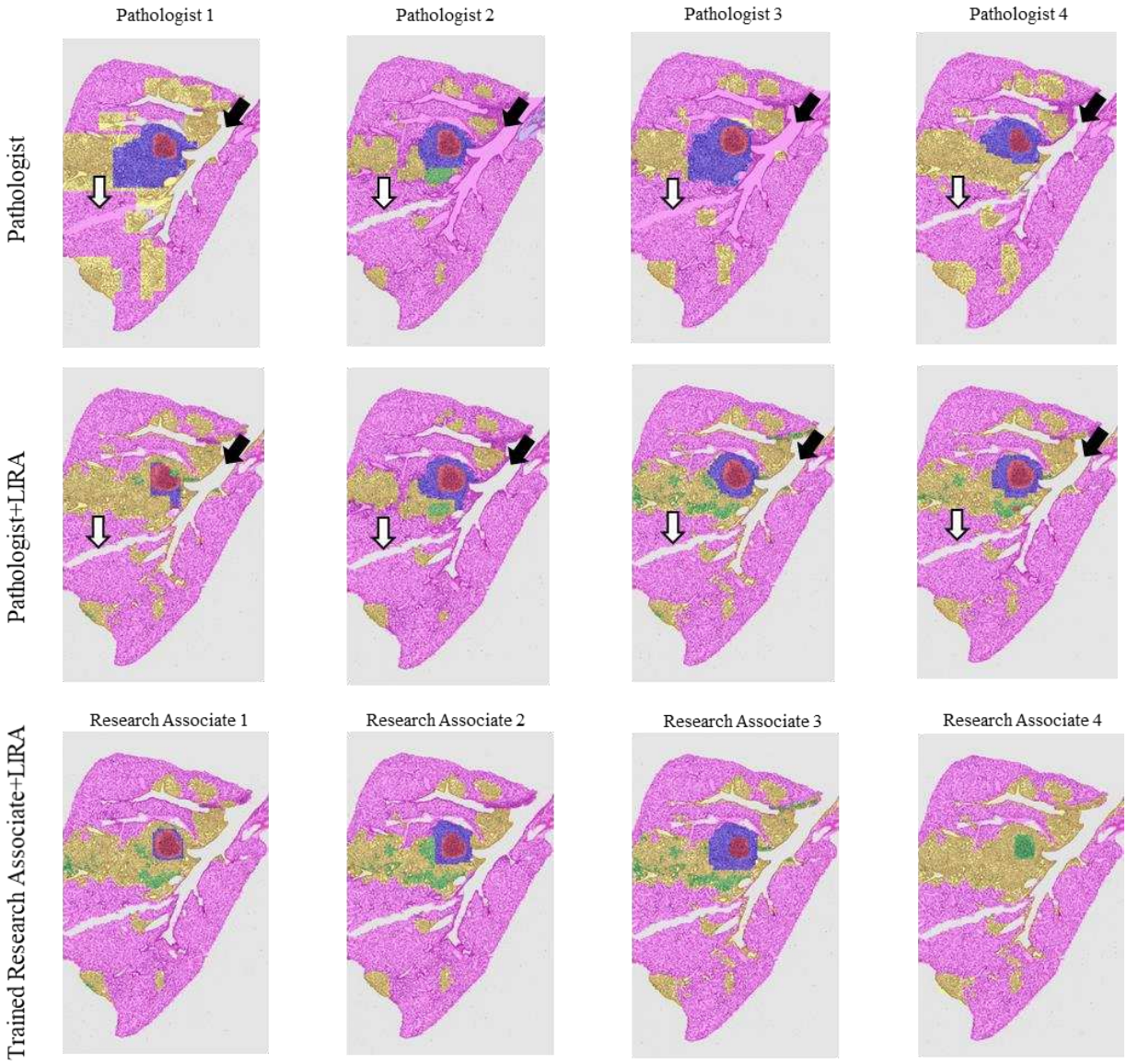


Figure 2.4. Image 3, which consists of Type I Core (Red), Type I Rim (Blue), Type II lesions (Green), Type III (Yellow), and Healthy Tissue (Pink). The top row consists of pathologists hand labeling the tissue, the middle row consists of pathologists with the assistance of LIRA, and the bottom row is the labeling by the research associates with the assistance of LIRA. Black arrows are pointing out the bronchial airways, and the white arrow is pointing out areas of artifact generation.

pathology analysis approach of four pathologists or four research technicians, to the prediction results of LIRA with pathologist assistance (Figure 2.3). The comparison focused on three criteria: 1) the accuracy of both approaches in identifying the correct pathology classification; 2) the accuracy of demarcation of lesion margins that affect the area of lung involvement, and 3) the

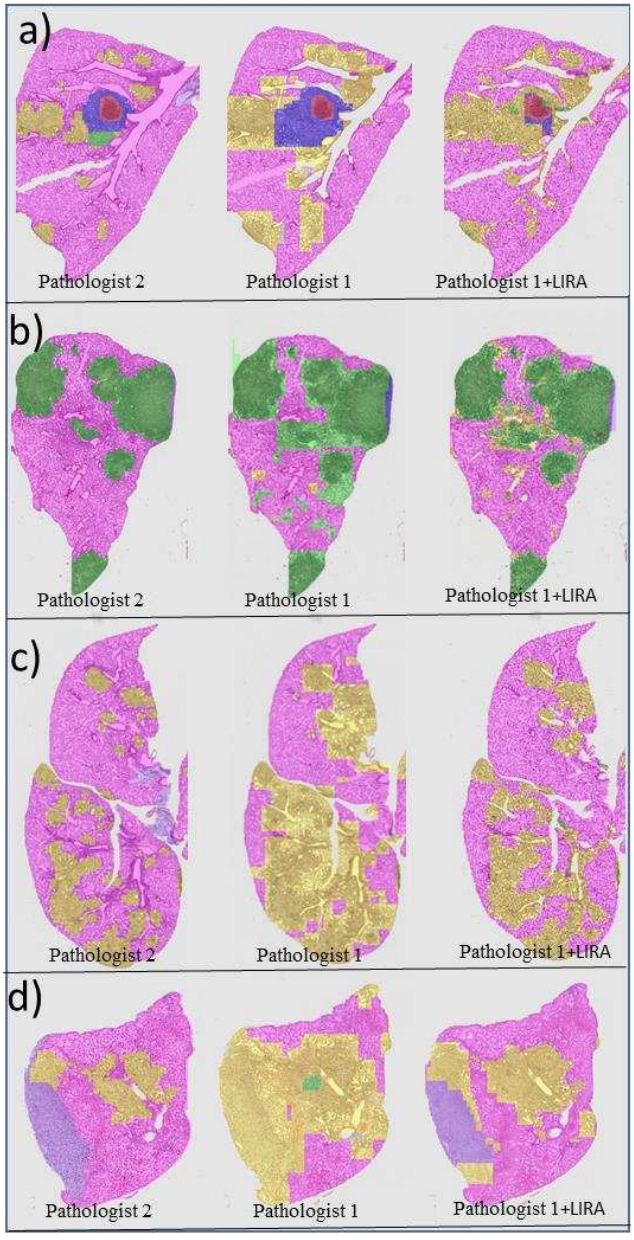


Figure 2.5. Color overlay results of 4 images from our validation set with (a) Image 3, (b) Image 5, (c) Image 6, and (d) Image 7. The far-left images are the best-labeled Pathologist image, the middle images are the poorly labeled Pathologist images, and the far-right are the poorly labeled Pathologist image but with LIRA assistance. When a pathologist who labels poorly has assistance

from LIRA, the final result is in more in accordance with the best-labeled pathologist example. This allows studies to be comparable across multiple pathologists, limiting the amount of error that might have been introduced.

variability between users. For this purpose, the pathologists were first given seven images to classify each image patch by hand. After one week, the pathologists used the same seven digital images in a random order to classify each image patch now with the assistance of LIRA. Using the classical pathology approach, the results of the four pathologists showed a consensus on the specific lesion classification in general, but not always on the extent of the percent lung involvement (Figure 2.4). The disease classification categories with the most variability in results were Type I Rim and Type II lesion types, which subsequently also impacted the Healthy tissue classification indirectly. The variation in results within these categories was mostly the result of inaccurate demarcation of lesion margins after manual labeling by the pathologists. Specific examples detailing the high variability for each of these classifications are presented in Figure 2.5 and Supplemental Figures 2.3 & 2.4.

As LIRA makes classification predictions per image patch at a microscopic level, it became clear in the validation step that both CNN2 and CNN3 can identify certain areas of a single lesion as one classification whereas another area of the same lesion could be identified as a different classification. An example of this can be observed in Figure 2.4 (Image 3), which consists of one Type I and multiple Type III lesions. In this example, for the pathologist-only readout, an observed disagreement was seen for the demarcation of the Type I Rim and the start of the Type III lesion that is adjacent to it. Only one of the four pathologists (pathologist 2) had accurately identified neutrophil-rich regions of the Type III lesion (in green), which is an unusual event for a Type III lesion. For the pathologist-assisted LIRA results (Figure 2.4), three out of four pathologists decided to reduce the lung area for the Type I lesion after being prompted by LIRA. In addition, all four pathologists recognized regions of neutrophil infiltration only after the assistance of LIRA (in green). Another disagreement between pathologists was observed regarding the classification

of bronchial airways and artifacts. For the pathologist-only readout, two out of four pathologists classified the bronchial airways as Healthy tissue, and the two remaining pathologists classified these regions as Empty slide. The same was observed for a tissue processing artifact visible on the image scan (a fragment hole artifact), which was classified by two pathologists as Healthy tissue and by the other two as Empty slide. With the assistance of LIRA, the results for all four pathologists were now in agreement, and both the bronchial airways and the artifact were classified correctly as empty slide. The change in the classification that occurred at the pathologists' checkpoints after CNN2 and CNN3 showed to have a significant impact on the output data in terms of reducing the variability and increasing reproducibility of the results.

Table 2.3. For each image, the percent agreement and Krippendorfs alpha were calculated to determine the agreement between pathologists' classifications without (Path) and with LIRA (Path+LIRA).

	Percent Agreement		**Krippendorfs Alpha**	
Image	*Path*	*Path+LIRA*	*Path*	*Path+LIRA*
1	84%	94%	0.761	0.904
2	92%	97%	0.864	0.951
3	84%	94%	0.757	0.902
4	92%	94%	0.860	0.903
5	90%	95%	0.849	0.920
6	84%	93%	0.757	0.891
7	82%	93%	0.725	0.896
Mean	**87%**	**94%**	**0.796**	**0.910**

Reproducibility of histopathology analysis regardless of the pathologist analyzing the data or the variation that can occur over time is important when comparing data across preclinical studies or laboratories. To measure whether the agreement between pathologists increased with the assistance of LIRA, multi-rater percent agreement, and the Nominal Krippendorff's alpha approach were used. Each image patch was considered a separate observation for the analysis,

and the percent agreement was calculated per image. The multi-rater average percent agreement without LIRA for the four pathologists across all images was 87%. Images showing the lowest agreement consisted of primarily Type I (Images 1&3) and the Misc. classification (Image 7) (Table 2.3 and Figure 2.5D). With the assistance of LIRA, the percent agreement for the four pathologists increased across all images, and there was a 53% decrease in disagreement overall to achieve an overall agreement of 94%. The most prominent increase in percent agreement was observed for the images containing Type I lesions and the Misc. classification.

Using a more refined additional metric of analysis, the Krippendorff's statistical analysis approach was used for calculating the nominal alpha coefficient. Again, each image patch classification by the four pathologists was seen as a separate observation. Manual labeling of seven pathology features for seven images by the four pathologists showed an average alpha of 0.796 (Table 2.3), which is interpreted as tentative. Four of the seven classified images were below the 0.80 threshold, which means the pathologists failed to achieve good agreement on the classification of those images. In contrast, when performing the Krippendorff's analysis on the data generated by the four pathologists with the assistance of LIRA, the calculated average alpha was 0.91 (Table 2.3). The Krippendorff's analysis results for all seven images were above the 0.80 threshold, and results for all seven images were above the 0.8 threshold, which is considered good. These results confirmed our observations using percent error, providing further evidence of the benefits of LIRA assistance.

The increase in agreement among pathologists with the assistance of LIRA was the result of two main beneficial factors. First, the use of the LIRA software increased the accuracy of the demarcation of lesion margins, which affected the percent lung involvement in various categories indirectly. In addition, LIRA prompted the pathologist to revisit and re-analyze specific areas in the lungs, often small in size, which were often overlooked using the classical histopathology approach. Examples are shown in Figure 3 (Pathologist 1, Pathologist 1 + LIRA) and Figure 4 (Images 6 & 7), which show the improvement of lesion margin demarcation with LIRA assistance.

With the use of LIRA, the classification of the pathological features not only improved the agreement between pathologists, but the results tended to be closer to the predictions made by the most experienced TB pathologist (Pathologist 4) (Figure 2.3). An example where LIRA prompted the user to re-analyze their initial prediction is shown in Figure 3D. Here, a large area of fibrotic tissue is observed, which should have been classified as miscellaneous as it does not fit any other definition. With the assistance of LIRA, the pathologist changed the classification to miscellaneous, which was in agreement with the other three pathologists. The accuracy of classifying image features in the miscellaneous category is mostly dependent on the background and level of expertise of the user, and therefore this category will always be the most difficult to standardize across multiple users. The assistance of LIRA in the miscellaneous category, however, did minimize the misclassifications in this category and was an added benefit.

In summary, the results of the validation tests showed a clear benefit of using LIRA assistance in the pathology readings by board-certified pathologists, and this by improving the accuracy in the demarcation of lesion margins as well as prompting the user to re-analyze certain areas thereby increasing conformity and reproducibility of an otherwise subjective readout. Additionally, compared to previous manual pathology analysis methods employed by the four pathologists taking about 15-30 min manual effort per lung lobe, an average decrease in manual analysis time of 82% was observed, now reducing the histopathology manual analysis to 3-10 min semi-automated effort per lung lobe (with 5 minutes on average per lung lobe).

2.3.4 - Trained Research Technicians + LIRA

In a final validation assessment, we investigated whether LIRA could be easily employed by research technicians, and how their prediction results would compare to those made by specialized pathologists. For this purpose, four research technicians were first trained in the identification of lesion types of the C3HeB/FeJ animal model (Trained Research Technician + LIRA). The technicians were then asked to analyze the same seven digital images of a single

lung lobe, in that same manner as performed by the pathologists for the validation step. The results showed that for the research technicians with the assistance of LIRA, similar classification predictions were obtained as described earlier for the pathologists. Most notable was again the accurate demarcation of the lesions with LIRA assistance (Figure 2.4). The results by research technicians assisted by LIRA exhibited a slightly higher variability when compared to the results of the pathologists assisted by LIRA. One area where the research technicians using LIRA showed less accurate results was in the identification of the Miscellaneous Tissue category. For example, in Image 7, a large area of fibrotic inflammation was not identified by the research technicians, whereas this was correctly classified by the pathologists using LIRA (Figure 2.4, Misc., Image 7). Two out of four research technicians correctly identified fibrotic inflammation (Figure 2.4, Image 5), which was also observed by the pathologists.

In summary, the output of the research technicians with LIRA was remarkably similar to that of the experienced pathologists using LIRA, except when uncommon or infrequent tissue types were present. Taken together, the data of the validation study indicates that even with the implementation of AI, a certain level of pathology expertise is still preferred to quantify certain unique lesion pathology accurately.

2.4 - Discussion

Traditional histopathological analysis generally involves an experienced board-certified pathologist, often using a semi-quantitative scoring system to identify and quantify unique disease-specific pathology features [79], [188]. Specific scoring criteria are developed which can distinguish between various disease states, thereby informing about disease progression or on effects of treatment intervention. These scoring criteria are then applied in double-blind histopathology analysis to reduce bias and variability. However, issues such as reproducibility between studies and variability in readouts by different pathologists still exist, and this continues to be an area of active research [116], [189]. In non-clinical TB animal studies used in drug and

vaccine development, studies are often long term, require multiple time points, and generally occur in more than one laboratory to confirm that novel treatments or interventions are reproducibly efficacious. To evaluate the progression of pulmonary lesions in TB animal models, a histological grading system was developed years ago, based on grading granulomatous lesions for inflammatory cell numbers and their infiltrative distribution pattern [159]. The pathologist would then calculate a mean total lesion score for individual tissues; as normal (score of 1), mild (score: 1–3), moderate (score: 3–6), severe (score: 6–8) to reflect the disease state per animal. The variability in histopathology analysis between studies, time points, and individual pathologists, owing to the subjective nature of this method, was often substantial thereby making it difficult to interpret results. In addition, studies would be time-consuming, and results over time will be affected by user fatigue. In this report, we describe the development of novel software for a rapid automated and unbiased digital image analysis using a machine learning approach based on convolutional neural networks to evaluate the histopathology of mouse models for tuberculosis in a quantitative manner. The goal was to make the software intuitive for its user, as well as facilitate and accelerate histopathology analysis for the non-expert.

We report here on the training and validation of a model for the accurate classification of seven pathology features using histopathology images. The model uses human classifications to learn robust features from a large number of H&E stained image patches. The model is currently suitable for quantitative lesion identification of the C3HeB/FeJ mouse model for tuberculosis. Initial attempts tried to implement a single convolutional neural network to identify all seven histopathology classifications. However, significant misclassification was observed caused by similar cell compositions in more than one lesion type. Therefore, LIRA was created using three individual neural networks to create a single modular network to generate more accurate predictions based on both macroscopic- and microscopic histopathology events. Currently, the accuracy of the model is not sufficient to rely solely on the implementation of the LIRA software without any human intervention. Two human intervention checkpoints for the user are still

required: first, to confirm and potentially modify the Type I lesion prediction (CNN1) and second, to confirm the individual image patch classifications made by LIRA (CNN2, CNN3). Moreover, human intervention can help prevent the misclassification of infrequent pathology events and artifacts on which the specific CNN had not previously been trained. Of note, certain misclassifications by LIRA may ultimately be beneficial to the process as they can bring a particular pathology feature to the attention of the user, and can prompt the researcher for closer analysis. Future iterations will focus on improving the model by including additional datasets for training when more studies become available to increase LIRA's performance under all circumstances.

There were multiple advantages to using assistive software in identifying and quantifying disease states with digital image analysis. First, we observed a significant reduction in analysis time and user-fatigue, whereby an 82% decrease in time was achieved for quantitative analysis in comparison to the standard histopathology scoring methodology. Second, with the assistance of the software, the variability of the predicted classifications among different pathologists was greatly reduced. Of importance, with the assistance of LIRA, a more accurate pathology comparison will be possible across multiple animal studies, different longitudinally time points, and studies performed at multiple laboratories. Where LIRA outperforms the individual pathologist is the consistent and accurate demarcation of lesion margins, as well as the detection of small regions on a microscopic level, which both substantially affects the quantitative analysis of the lesion area involvement. Pathologists and LIRA, in most instances, both correctly classify the lesion type, but the variability in the results from classical pathologist readouts is derived from the imprecise or inconsistent demarcation of lesion margins. Where pathologist expertise and intervention is still needed and preferred is for the identification of infrequent pathology events or unusual artifacts. As also described for other applications in medical image analysis, diagnostic confidence never reaches 100%, and combining machine learning and physician or pathologist experience will enhance system performance [190].

During the development of the LIRA software, it became apparent that the quality of tissue sample preparation can impact the accuracy of the results of the software. LIRA is affected by features such as the presence of high red blood cell numbers and processing artifacts, which led to the software incorrectly classifying these regions as pulmonary lesions instead of healthy tissue. Deviation from the sample preparation protocol presented in our methods section, including using frozen sections or changes in section thickness, will also result in a decrease in accuracy. When sample preparation varies, it is recommended that a new model be trained on image patches generated with these new protocols. In addition, the current iteration of LIRA is designed to identify lesion types for C3HeB/FeJ pulmonary tissue collected between 6 to 10 weeks post-infection. We have tested LIRA on more commonly used BALB/c or C57BL/6 pulmonary tissues infected with Mtb, and results showed these models are suitable for LIRA use.

A future direction to improve the current LIRA pipeline are to generate a new classification for the surrounding Type II tissue and to increase the sensitivity of CNN1. LIRA predictions for Type II lesions could be significantly improved by creating a separate class for the surrounding tissue, primarily composed of aggregates of extracellular DNA and compressed tissue. And an increase of the Type I detector (CNN1) sensitivity might allow omission of the first human intervention checkpoint. The current false-positive rate for the Type I Core is the result of the limited sample numbers that were available at the time of training CNN1. With an increase in the data available for training and the utilization of generative adversarial networks (GANs) [132] and pre-trained neural networks [191] represent possible future solutions.

A potential adaptation of LIRA's architecture for use in TB preclinical models could be an enhancement beyond just classification and quantification of infection in tissues. Researchers can combine lesion classification features with additional quantifiable measurements. Of particular interest in TB drug and vaccine development is to include the mycobacterial quantification metrics of each individual lesion. The goal would be to integrate precise measurements on bacterial numbers after fluorescent staining [74], [192], as well as bacterial aggregation sizes, level of

fluorescence, and the average number of bacteria as a metric of area. With the availability of matrix-assisted laser desorption ionization (MALDI) imaging to assess drug levels across pulmonary lesions [18], [193], one can envision the development of a model that would integrate data sets from digital lesion pathology with bacterial metrics as well as drug exposure.

In conclusion, LIRA does not replace the histopathology analysis by pathologists but instead intends to improve the accuracy, speed, and reproducibility of the analysis. While there are certain limitations with the current model, these can easily be corrected and adjusted. The ability to more quickly and more accurately assess the effect of treatment interventions on histopathology of target organs will improve the evaluation of the efficacy of novel antibiotics and vaccines. Our demonstrated approach is not limited to TB and can be modified to include additional diseases and animal models, by creating new training sets and modifying the original architecture of LIRA to meet the needs of the researcher.

CHAPTER 3: IMPLEMENTATION OF LIRA TO ASSESS TREATMENT RESPONSES USING THE C3HEB/FEJ MOUSE EFFICACY MODEL.

3.1 - Introduction

Not everyone who becomes infected with Mtb will develop active tuberculosis. In 90% of cases, the infected individual will exhibit latent disease, which is when a person is infected with Mtb but does not have overt signs of an active TB infection. They are not infectious, and the only sign of infection is a positive reaction to the tuberculin skin test or TB blood test. The risk of developing active disease is higher for individuals with an impaired immune system, especially those with a co-HIV infection or diabetes. Even though symptoms are benign in latent TB, treatment is essential to reduce the risk that the infection will progress to active TB disease. Depending on the state of the disease (active or latent), age of the patient, overall health, if antibiotic resistance is confirmed, and location of the infection within the host, the duration and specific drug regimen will vary. For latent disease, the treatment consists of Isoniazid (INH) and Rifapentine (RPT) for 3 months [194]. For active disease, there are treatment regimen guidelines for drug-susceptible TB, drug-resistant TB, multidrug-resistant TB (MDR TB), and the uncommon extensively resistant tuberculosis (XDR TB) [195]. There are two different phases in treatment regimens in patients for drug-susceptible TB disease, using first-line therapies, called the intensive phase and the continuation stage. The intensive phase includes a combination of INH, RIF, PZA, and EMB taken 7 days/week for 8 weeks. This is followed by the continuation phase which consists of additional treatment of INH and RIF for 7 days/week for an additional 4-7 months [5]. For the purposes of the aim analysis, I will only be providing a condensed background synopsis on the antibiotic INH. INH is bactericidal to actively dividing extracellular and intracellular bacilli, and bacteriostatic to slow-growing bacilli. It is a prodrug that must be activated by the Mtb catalase-peroxidase enzyme KatG which later forms a covalent adduct with the NAD cofactor. The INH-NAD adduct acts as a binding competitive inhibitor for InhA in the FAS II fatty acid

biosynthesis pathway[196]. In the presence of INH, actively growing Mtb will undergo cell lysis, but after removal of the drug the surviving bacilli will resume rapid growth and division[197]. Drug resistant TB relies on second-line therapies which entail longer treatment times of 18+ months, increased negative side effects, and consist of a combination of thioamides, diarylquinoline, nitroimidazole, fluoroquinolones, and aminoglycosides. XDR TB treatment is completely dependent on the drugs that are still effective and currently relies heavily on the newly developed drugs such as BDQ or delamanid [198].

Because treatment is at a minimum 6 months in duration, non-adherence to TB treatment is of major concern. In a recent study, it was reported that 33% of patients missed either one or two doses within a 2 week period [199]. Failure to properly adhere to treatment schedules can result in an increased risk of the spread of TB and may lead to the development of drug resistance. While resistance is considered an uncommon event, it is increasing to an annual rate of more than 20% [200]. This is a disturbing statistic since, according to the European Centre for Disease Prevention and Control (ECDC) 2015 report, MDR TB in Europe had a success rate of 46% and only 23.2% for XDR TB [201]. The WHO observed slightly higher rates with 56% of patients with MDR TB having a successful outcome [202]. Drug resistance is an increasing problem, but with the discovery of novel therapies, such as delamanid, pretomanid, and BDQ, we can increase the success rate of therapy.

When a potential new drug candidate is identified, it goes through iterative phases of testing. This process is costly, with only a 13.8% probability of success that the proposed compound will be approved [203]. The percentage of success is important in prioritizing research funds and time. Having models that more accurately represent human outcomes is critical to help mitigate the risks involved. Compounds are initially tested *in vitro*, commonly with cell assays, and if positive results are seen, compounds are further tested *in vivo* using animal models. Animals are commonly used in TB pharmaceutical research because they offer the complexities of a host environment not available *in vitro* and help predict the efficacy, safety, and the optimal regimen

for human patients [204], [205]. There are many metrics measured in animal models used to determine the efficacy of TB drugs such as animal morbidity (weight, posture, and physical appearance), mortality, pathology, and bacterial burden (culturable and visual). Some aspects, such as monitoring the animal's weight, are quantitative with little room for user interpretation, while other metrics such as pathology currently rely on a mixture of qualitative and quantitative results to accurately describe the disease. Different types of data types collected during an experiment provide various benefits and limitations that need to be considered in their relation to proving or disproving the hypothesis.

In our laboratory, the C3HeB/FeJ mouse model has been developed to be used as a tool in preclinical development of novel drugs and regimens. Throughout a drug efficacy trial one of the most important metrics that is measured is the severity of pulmonary lesion development. Untreated C3HeB/FeJ mice aerosol infected with a moderately virulent strain of Mtb consistently develop three independent lesions types termed Type I, Type II, and Type III (Figure 1.2) [74], [206], [207]. Type I lesion can be broken down into two separate classifications, the Type I Core and the Type I Rim. The Type I Core is everything from the caseous necrotic center to the fibrotic rim, and the Type I Rim is from the rim outward comprised mostly of fibroblasts, epithelioid, and activated macrophages, and occasional lymphocytes. Type II lesions are composed of a rapidly expanding caseous necrotic center comprised primarily of neutrophils but lacking the encapsulation of Type I lesions. Type III lesions are cellular, inflammatory lesions containing macrophages, lymphocytes, and small pockets of activated neutrophils. 5 weeks post aerosol infection lesions become distinguishable [74], and historically our laboratory has collected pathology within the 8 to 10 weeks window after aerosol. This is to allow for sufficient advancement of the pathology of Type I lesions, as well as the enrichment of mice that have Type I lesions. 10 weeks past the aerosol infection, the Type I lesions continue to grow in size which makes histology processing, especially sectioning, difficult to perform. Drug treatment in this mouse model and others have altered lung pathology development, generally by improving lung

pathology over time. This is demonstrated by previous studies that during treatment, lesion progression will either begin to resolve or halt lesion progression [208], [209]. However, it has been observed that efficacious antimicrobial drugs can also result in a substantial increase in the inflammatory response and lesion development. This immunopathologic event caused by treatment can result in a spike in morbidity and increased lung consolidation. In our experience, we found treatment with a new experimental drug being tested in our laboratory resulted in Type I lesions, which progressed to cavitary disease (personal communication, Dr. A. Lenaerts). In scoring pulmonary pathology, signs such as healing of lung pathology and decrease in lesion involvement are usually implying a drug is efficacious, but it is important to take into account that drug treatment could initially worsen lung pathology

Treatment efficacy is also influenced by multiple factors, including the lesion-specific properties that influence drug penetration and access [210], [211]. Humans and the C3HeB/FeJ TB animal model both display a heterogeneity of lesion types that have varying microenvironments and physical lesion structures. Another similarity is that multiple different lesion types can exist within a single host independent of each other which can hinder attempts for successful treatment [22], [149]. The fibrotic rim within the C3HeB/FeJ mouse model Type I lesions is densely vascularized meaning the pharmacotherapy is often successful in these regions. In contrast, the Type I caseous center has had most of its vascular architecture destroyed causing a failure to properly mount an adequate immune response and little drug availability from the blood. This means the drug needs to diffuse across the cellular rim to the necrotic center without the aid of active or facilitated transport mechanisms. If an adjunct treatment strategy were available that would improve drug delivery into the encapsulated lesions (by, for instance, decreasing the integrity of the Type I in C3HeB/FeJ mice), it might shorten the length of treatment. Potential adjunct therapies include inhibition of the eicosanoid pathway [212], statins [213], [214], and vitamin D[215]. One adjunct therapy that had been explored in our laboratory was the use of losartan (LK) in an attempt to inhibit the formation of the fibrotic capsule in the Type I lesion [206].

The hypothesis was that C3HeB/FeJ TB infected mice treated with LK will have inhibited progression of the collagen rim in the Type I lesion allowing for greater penetration of drugs resulting in reduced bacterial burdens. Adjunct therapies, such as LK, will need to be combined with current clinical therapies to see an increase of efficacy or a decrease in therapy duration.

LK is an antihypertension drug that is commonly used to treat high blood pressure and provide protection to kidneys from damage resulting from diabetes. It works by inhibiting the vasoconstriction of blood vessels and increase blood flow[216]. This inhibition by LK specifically results in the reduction of fibrinogen, collagen I, collagen II, fibronectin, and TGF-β1[217]. There has been a reported secondary effect of LK in the reduction of fibrosis in multiple diseases such as skeletal muscle injuries [218], ovarian cancer[86], and bleomycin-induced lung fibrosis[219]. An earlier study in our laboratory evaluated the addition of LK to existing drug treatments in the TB C3HeB/FeJ animal model with the aim of using this anti-fibrolitic chemotherapy to manipulate the collagen rim of the Type I granuloma (study by Dr. Emily Driver, a Ph.D. student at the time)[206]. The goal of the previous study was primarily aimed at determining the effect of adjunct combination regimen on bacterial loads in the lungs and used a manual scoring system for the lung histopathology analysis.

Current histopathology scoring metrics utilized a semi-quantitative methodology that is non-standardized and highly susceptible to user bias [78], [220]. To reduce bias and errors, common pathology parameters and guidelines are often used, such as standardized diagnosis metrics, grading, staging, and biomarker scoring. In the case of TB, no set scoring standards have been developed, which requires the researchers to develop their own. This makes comparing results across laboratories, studies, or even time points difficult. Even if the scoring systems are similar, they can result in varying interpretations. It also can mask results because the scoring parameters are not sensitive enough to distinguish differences. This is exemplified in TB scoring, where an ordinal score is given based upon arbitrary boundaries of the disease percent involvement. For instance, if a particular treatment cohort averages 12% lung involvement but

another cohort averages 23% lung involvement, with the current scoring system, they can both receive the same pathology score, even though biologically, the differences might be significant. With the recent advancements in ML and CV, a more targeted analysis with less human bias can be accomplished. This allows the reporting of quantitative assessments that are reproducible, less prone to human bias, and can be standardized across multiple laboratories and time points.

As described in Chapter 2, we demonstrated that LIRA increases the user agreement across multiple pathologists scoring the same microscopy slides. This analysis was performed on mice following their natural disease progression without any form of treatment or host manipulation. Next, LIRA's performance needed to be tested on lung lesions from the C3HeB/FeJ mouse model that had been treated, therapies included immunotherapy and anti-mycobacterial drug treatments. ML classification models can be sensitive to changes in the image input, and the introduction of drug treatment can be a potential source of noise that may impact the proficiency of the model. It is important that the model is verified using standard pathology scoring methodology as a control, using C3HeB/FeJ histopathology images that have additional image noise from drug treatment, which may influence overall performance. We are hereby using a past C3HeB/FeJ study with a combination treatment of LK with classical TB drug treatment as an example of how to use the LIRA software in a TB drug treatment study. With the improved approach for histopathology analysis using LIRA, we hope in addition to investigate: 1) the effects of standard TB drugs on lungs pathology, 2) assess the potential of LK as adjunct therapy for TB, and 3) assess how using a targeted computational analysis can alter the interpretation of the final results.

Of most interest to my thesis was investigating how implementing a targeted data analysis approach can change the final interpretation of the data. By comparing historical data analysis interpretations using techniques that take a less targeted approach with data generated using LIRA, we have shown that there is a difference in the study outcomes. Masking of potentially efficacious drugs or major biological events not only makes drug efficacy more costly but also

impedes the development of other potential new therapies. With cost, reproducibility, and time becoming major hindrances in the development of new therapies, it is imperative that tools are developed that overcome these obstacles. LIRA represents a new tool for TB drug development that addresses many of the obstacles associated with human pathology scoring while also providing additional quantifiable data that was not available previously. LIRA is not only applicable to normal disease progression analysis as demonstrated in Aim 1 (chapter 2) but also can determine the severity of disease compared to a board-certified pathologist while quantifying additional data metrics for further analysis.

3.2- Materials and Methods

All methodologies for the animals, bacteria, chemotherapies, aerosol infection, and stereology were performed previously within the Lenaerts laboratory. Inclusion of this information for my dissertation is for convenience purposes but more detailed information can be found in *Driver et al.* [206]. The histology preparation, staining, microscopy, pathology scoring, and analysis were planned and executed by Bryce Asay for the purpose of this dissertation.

3.2.1 Animals

Female C3HeB/FeJ mice aged 6 to 8 weeks were purchased from Jackson Laboratories, Bar Harbor ME. Animals were housed in a bio-safety level III animal facility and were routinely monitored. Mice were specific pathogen-free which was verified by testing sentinel mice housed within each colony.

3.2.2 Bacteria

M. tuberculosis Erdman strain (TMCC 107) was used for aerosol infections and was prepared as previously described [221], [222]. In summary, Mtb bacilli were grown as a pellicle to generate seed lots. Working stocks were cultured by growing individual seed lots to mid-log phase

in PB medium containing 0.05% Tween 80 (Sigma Chemical Co., St. Louis, MO) in three passages. Before storage at -70℃ titers will be calculated by colony counting on 7H11 agar plates and divided into 1.5 ml aliquots.

3.2.3 Chemotherapies

3 weeks post aerosol infection mice were treated with one of 3 drug combinations: INH (25mg/kg), LK(50 mg/kg), and INH+LK (25+50 mg/kg) for 6 weeks. Due to the high acute mortality, the dose for LK was reduced to 20 mg/kg at week 2 of treatment for both the single and combination therapy.

3.2.4 Aerosol Infection and Sample Collection

C3HeB/FeJ mice were inoculated with a low-dose aerosol infection of *Mtb Erdman* strain in a Glas-Col inhalation exposure system resulting in an average of 15 bacteria in the lungs. To determine the infectious dose, five mice were sacrificed the next day[222]. At 3 and 6 weeks of treatment, mice were euthanized by carbon dioxide inhalation, the right caudal lung was collected for histology, and whole lungs were homogenized, and the bacterial burden was calculated by serial dilution on 7H11 agar plates.

3.2.5 Pathology, Slide Preparation, and Imaging

At the time of necropsy, the right caudal lobe is collected and fixed in 10% neutral buffered formalin in phosphate-buffered saline. Samples were embedded in paraffin wax and sectioned to a thickness of 5µm using a microtome before being mounted on a slide. Before further processing, the paraffin is dissolved using either Histoclear (Electron Microscopy Sciences, Hartfield, PA) or xylene and gradually rehydrated using multiple baths with a systematic decrease in alcoholic concentration. Each slide is then stained with Hematoxylin and Eosin and imaged using an

Olympus BX41 with Olympus DP70 camera controlled by Olympus DP Software (Olympus, Melville, NY).

3.2.6 Stereology

Lung, lesion, and collagen deposition was originally analyzed using Stereo Investigator 8.2 (MBF Biosciences, Williston, VT) [223]. The workstation was composed of a Nikon 80i Research Microscope equipped with a bright field CFU objective (2x/0.01 Plan Apo, 4x/0.2 Plan Apo, 100x/1.4 Plan Apo), 2-z-axis computer-controlled stepping stage with linear grid encoders, Z-axis motorized specimen stage, color digital camera, and virtual slice Zoomify software. Lung and lesion areas were determined by using the area fraction fractionator module at the 40x objective. The data was presented as the mean ratio of lesion area to lung area from randomly selected image patches. Investigators were blinded to the cohorts before analysis to reduce bias.

3.2.7 Scoring Of Pathology for LIRA Comparison

Standard pathologist methodology used lesion area approximations with its corresponding scores as follows: <1% lung receives a score of 0, 1-10% receives a score of 1, 10-25% receives a score of 2, 25-50% receives a score of 3, and >50% receives a score of 4. To further characterize the severity of disease, scores would be weighted: Type I lesions being multiplied by three, Type II lesions being multiplied by four, and Type III lesions multiplied by two. The weighted scores would be summed and the mean calculated for each cohort with a minimum score of 0 for no visible disease involvement and a high of 28 for a mix of both Type II and Type III lesions. This scoring methodologies major drawback is that a lung tissue that has 100% Type II lung involvement would receive a score of 16 while a mixture of 50% Type II and 50%Type III would have a high of 28. This makes scores above 16 unreliable because Type II lesions are more severe than Type III.

Analysis with LIRA provides results that are a quantitative score that is less prone to user bias. Percent involvement is calculated and provides exact values that can be used in statistical analysis. In order to compare the two methodologies the scores generated by LIRA needed to be converted to the same scoring methodologies employed by the pathologist. This was accomplished by summing each lung lobe total image patch counts for each individual lesion, calculating the percent involvement for each, and translating it using the criteria outlined by the pathologist.

3.2.8 LIRA Analysis

Data was generated using LIRA by a graduate research assistant (GRA) highly experienced in C3HeB/FeJ pathology analysis and a research associate (RA) with training in the analysis but as much experience. The agreement of scores generated by the pathologist, GRA, and RA was measured using Krippendorff's alpha as previously described in Aim 1 (See Chapter 2 for more detailed explanation).

3.3- Results

3.3.1. Design And Past Results Of The C3HeB/FeJ Efficacy Study (Study Performed By Dr. Emily Driver, [206])

The main advantages of using a prioir study for MIAs verification is that we can reuse preiouvsly collected samples, and the original analysis had been done with methodologies commonly used in histology analysis to help reduce bias in the analysis. The original aim of the study performed previously was to use an anti-fibrolitic chemotherapy to manipulate the collagen rim of the Type I granuloma in an attempt to increase drug penetration [206]. The study consisted of TB infected animals either being treated with INH, LK, or INH+LK with pathology collected at 3 and 6 weeks post-treatment. Samples were collected at both 3 and 6 weeks from the initial time of treatment. Samples were either fixed in PFA before H&E stained or homogenized and serially

diluted to determine bacterial burdens. The CFU, percent lung involvement, lesion classification, and percent collagen involvement were measured. It was predicted that the combination of INH and LK would have reduced bacterial burden and reduced disease lung involvement. The original conclusion from the study stated that there was no difference in the percentage of lesion involvement and collagen deposition between any of the cohorts at both time points [206] (Figure 3.1). Despite this finding, the researchers noted that animals treated with INH and INH+LK appeared to have more Type I lesion involvement but did not have any quantifiable data to support their claims. Additionally, there was no other significant observation about the pathology severity of the LK and control cohorts except that the LK cohort had less collagen deposition. Analysis tools during the original study were limited in the scope and granularity that could be applied to the samples. Most notably, the stereology technique used randomized sections of tissue to extrapolate percent lesion involvement. LIRA, in comparison, achieves absolute percent involvement of the slide section while also providing the percent involvement of each classification. It is prudent to determine if having a more targeted approach, such as LIRA, would altar the original final conclusions that had been made.

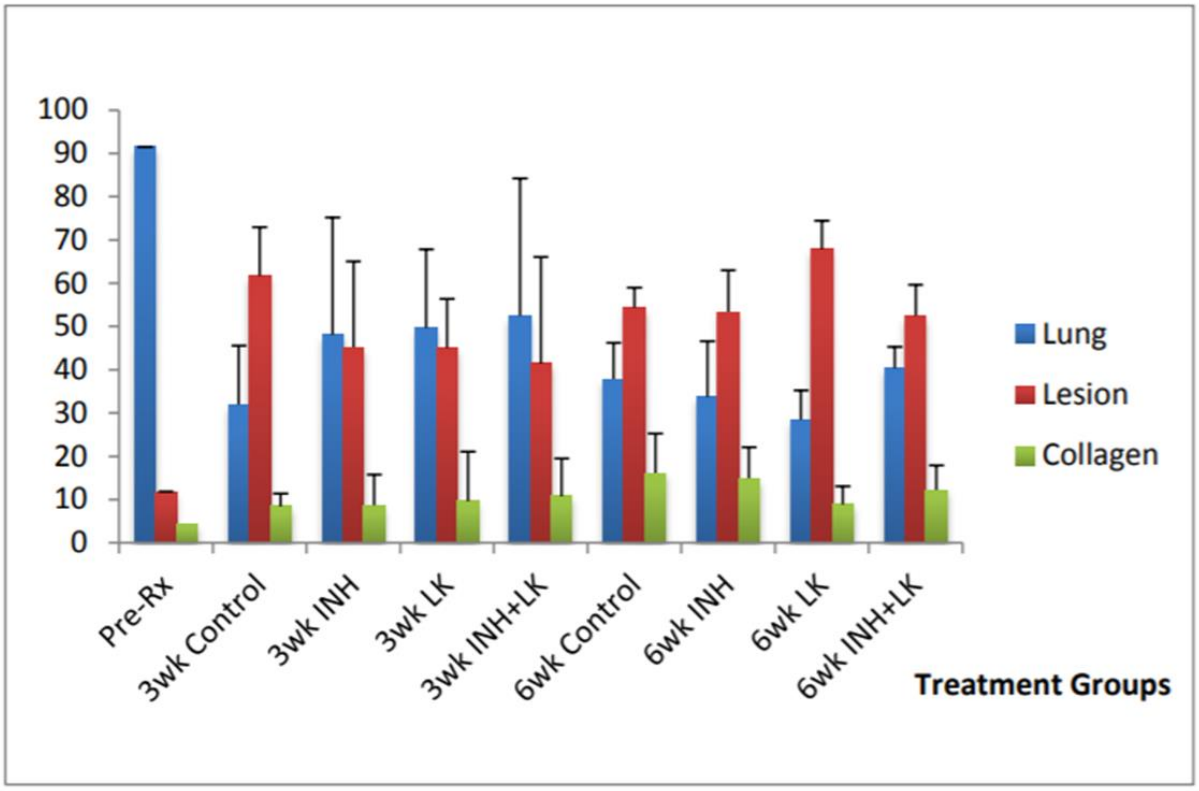


Figure 3.1. It was determined in the original LK study that there was no substantial difference in lesion development between the different cohorts over 6 weeks of treatment. Tissue measurements were performed using stereology. Adapted from Driver dissertation [206].

3.3.2 LIRA Histopathology Analysis

There were two aims in reanalyzing the histopathology of the original drug treatment study. The first aim was to discern if using a targeted analysis tool, such as LIRA, can give additional information in drug treatment studies. The second aim was to assess whether the values that were derived from LIRA were in agreement with the pathology scores given by a board-certified pathologist. Again, this was the first time LIRA was used on a TB drug treatment study using C3HeB/FeJ mice. For both aims, H&E stained slides of lungs from each of the treatment cohorts were given to a board-certified pathologist who was blinded of any identifier information. Using standardized methodology for scoring lung pathology [88], [224], each sample was given a pathology score based upon each lesion classification's percent lung involvement from 0-4, with

0 being no disease present and 4 being the most severe for each lesion type. Single scores were weighted based upon the severity of the lesion type, and all the scores across all the lesion types were summed to provide an end score used for comparison. These scores were compared to the converted values produced by a graduate research assistant (GRA) with extensive experience analyzing pathology from the TB C3HeB/FeJ mouse model and a research associate (RA) with a moderate amount of experience.

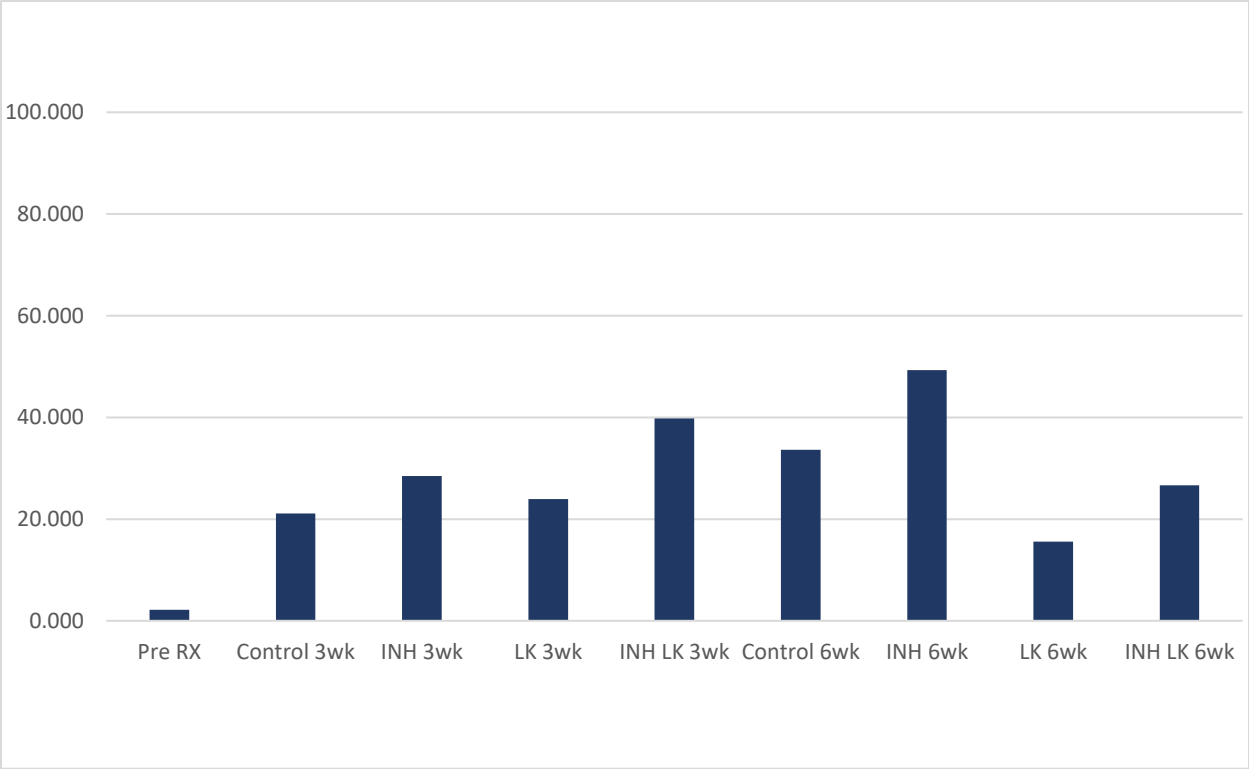


Figure 3.2. Using LIRA there was a substantial difference in overall disease severity between the INH 6wk and LK 6wk. This was in contrast to the original study that had been performed which had determined there was no difference between the different cohorts (Figure 3.1).

For the prior studies in this section, the analysis performed using the stereology had determined that there was no significant difference between the different cohorts and time points. With the ability of LIRA to quantify percent lung involvement for each lesion type, trends for each cohort were observed that were not discovered previously. All groups showed similar pathology at 3 weeks post-treatment for all lesion classifications, but by 6 weeks, a significant difference in

lesion severity and type was observed between the INH and LK cohorts (Figure 3.2). The INH cohort showed a substantial increase in lesion severity, particularly for Type I lung involvement, while the LK cohort showed a substantial decrease in lesion severity with scarce amounts of Type I development.

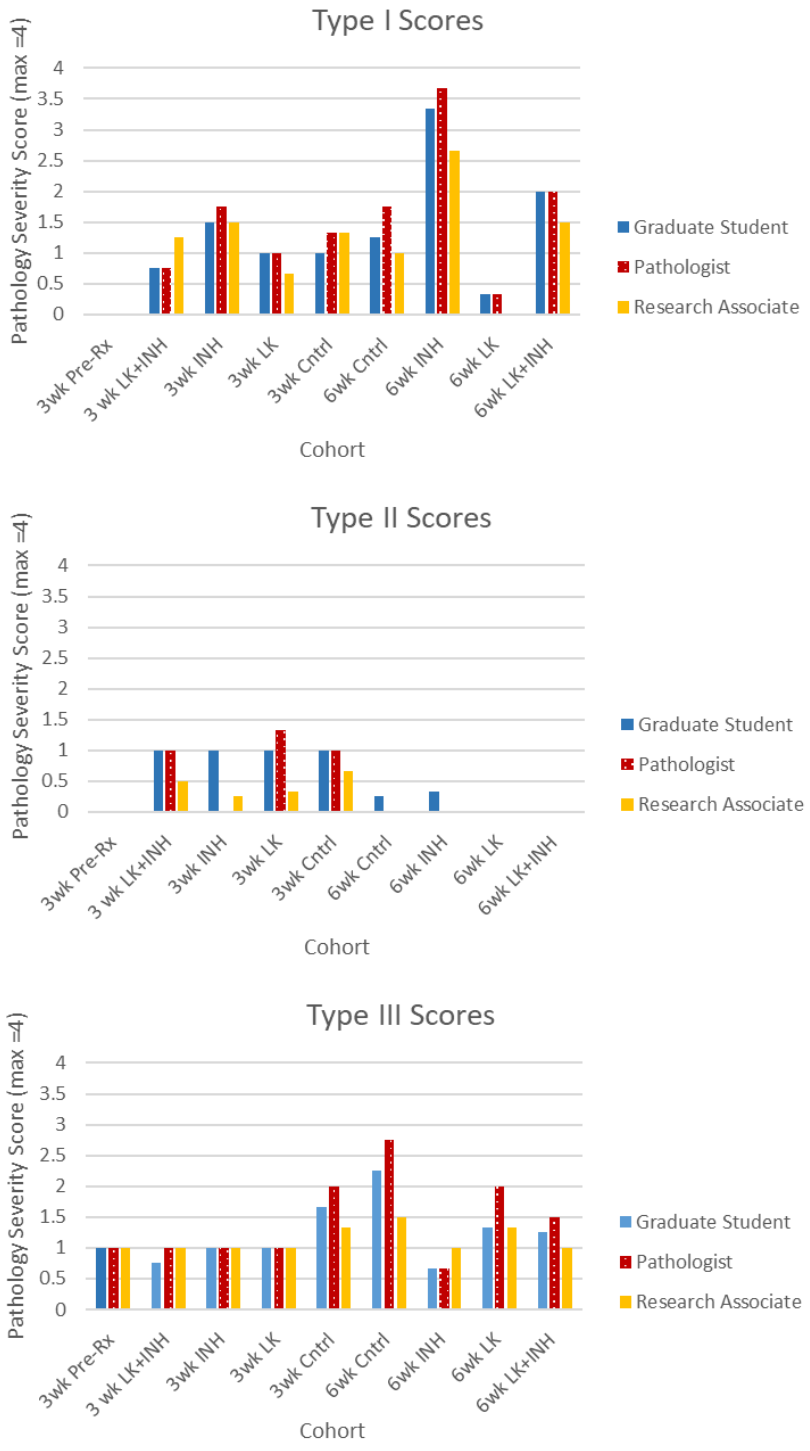


Figure 3.3. Histology analysis performed using LIRA by quantifying the three different lesion types separately. Unlike previously, differences were seen between three and 6 week time points of treatment and between the different cohorts.

The untreated control developed moderate amounts of both Type I and III and a slight increase in Type III lesion involvement as expected. The INH treatment cohort after 3 weeks of treatment showed a slightly increased Type I lesion involvement (score=1.6), which was somewhat increased compared to the untreated control group (score=1.2). This becomes even more pronounced after 6 weeks of treatment with INH, where lung involvement of Type I lesions is considerably higher (score = 3.2) in comparison to the control (score = 1.3) (Figure 3.3). There were fewer Type II lesions (score = 0.1) and Type III (score = 0.8) involvement at 6 weeks with INH treatment. The LK cohort at 6 week treatment showed almost no Type I (score=0.2) or Type II (score=0.0) lesion involvement and some Type III (score = 1.5). Dual treatment (LK+INH) showed mixed results with Type I lesions having the most lung involvement (score=1.8), followed by Type III (score=1.25), and Type II (score=0) lesion involvement. The dramatic increase in Type I pathology with the INH cohort was unexpected since previous drug efficacy trials demonstrated reduced lesion involvement or halted the severity of lesions. Yet it had a reduction in culturable pulmonary CFU and the lowest mortality out of the other cohorts. In contrast, the LK cohort had the lowest level of lesion severity and a reduction in culturable CFU comparable to INH but had the highest mortality out of all the cohorts. The results demonstrate that by utilizing a more targeted approach, such as LIRA, the interpretation of the collected data provides additional insights not identified earlier.

3.3.3 Cohort Overall Score Agreement Between LIRA And The Classical Pathology Approach

When comparing the overall severity and classification generated during the analysis by the board-certified pathologist versus the RA, there was no substantial difference between the users. The agreement of pathology scoring was calculated using Krippendorff's alpha with a score of 0.843, which can be interpreted as a good agreement between the GRA, RA, and pathologist. The only exception was for the INH treatment group after 3 weeks of treatment, whereby the 'expert' scored a 0, and the GRA had a score of 1 for Type II lesions. Post analysis revealed it

stemmed from the disagreement of atypical lesions that were still in the process of formation. The atypical lesion was the outcome of the early collection time point of the lung tissue when the Type I lesions are still being formed with a partial fibrotic rim on one side and high numbers of neutrophils on the other side of the immature Type I lesion (Figure 3.4). The early-stage Type I lesions are difficult to identify, and classification is largely based on expertise and information given prior to the analysis. This particular example highlights the need for both AI and human classifiers working together to achieve optimal results. In summary, there was a good agreement in the summary scores given for overall lesion severity at the cohort level when measuring the agreement between the board-certified pathologist and trained researchers.

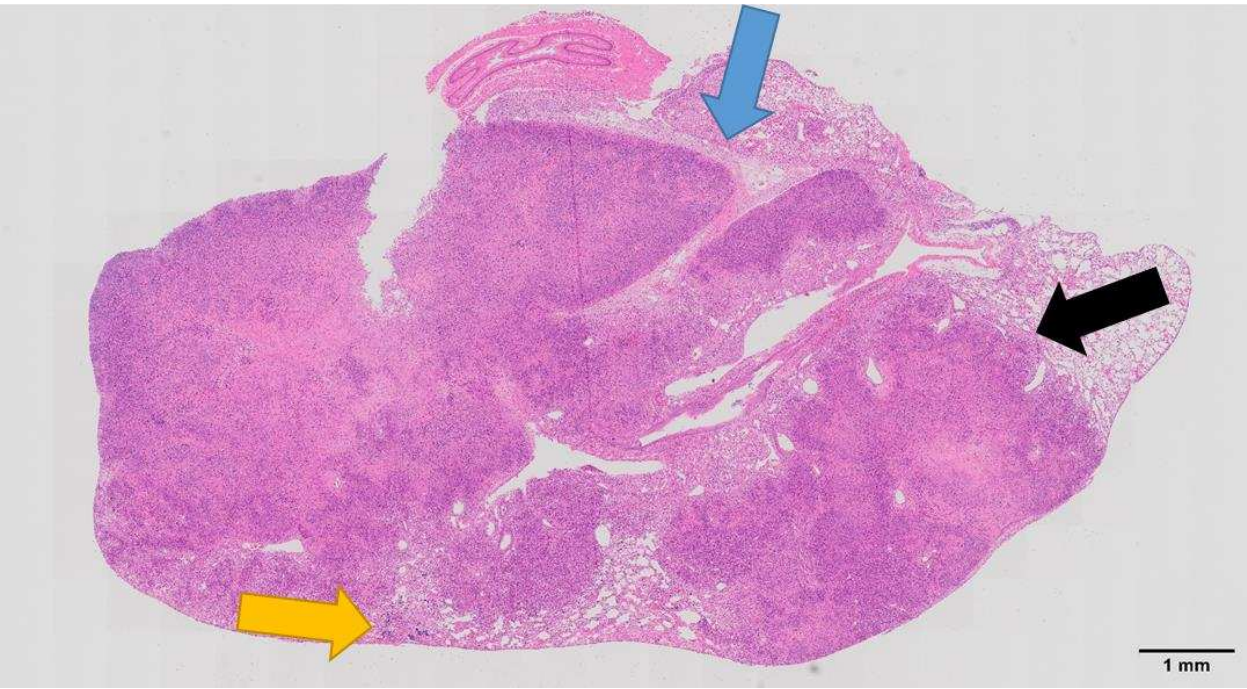


Figure 3.4. Example of a partially formed Type I lesion. Portions of the lesion have the early stages of fibrosis and collagen development (blue arrow), while other portions have no fibrosis black arrow) and DNA pooling characteristic of Type II lesions (yellow arrow).

3.3.4 Individual Lesions (Type I, II, & III) Score Agreement Between LIRA And The Classical Pathology Approach

For Aim 2, the objective was to determine if the specific lesion scores (e.g., Type I, II, or III) that were derived from LIRA were in agreement with the pathology scores given by a board-certified pathologist. This was measured using Krippendorff's alpha, which measures the agreement between the individual user's score. Type I lesion scores were mostly in agreement with the Krippendorff's alpha value of 0.879 which would be considered 'good agreement'. The 6 week LK group had a noteworthy observation showing no Type I lesion classifications by the RA. The reason for this failure of recognizing the Type I lesions was the atypical structure which required previous experience to correctly discern these forms of abnormal events, as demonstrated by the pathologist and GRA (Figure 3.5). Type II lesion scores showed to have a 'tentative agreement' with a Krippendorff's alpha value of 0.690. The tentative agreement resulted from the improved recognition of LIRA at a microscopic level of small pockets of neutrophils. The small clusters of neutrophils are generally not noticed by manual scoring of lesions, as this would be far too time-consuming and are easily overlooked. As a result, the GRA and RA reported very low levels of Type II involvement, and the pathologist indicated no involvement. Type III lesion scores were on the limit between 'good and tentative agreement' with a Krippendorff's alpha score of 0.796. There were no significant outliers in the scoring agreement between the pathologists with scores showing a consensus among the users. It is important to note that when each classification agreement is measured separately, all alpha scores are above the minimum 0.667 which is the lowest conceivable limit. In summary, agreement for each lesion classification was all above the acceptable limit, with Type I scores demonstrating the highest agreement for severity and Type II having the lowest agreement.

3.4 Discussion And Conclusions

In this chapter, the objective was to assess LIRA's performance using a past drug trial in comparison to scores generated by a pathologist and evaluate whether we would be able to gain more information with reduced bias and is quantitative for our histopathology analysis. For this purpose, we were able to use a previous C3HeB/FeJ mouse study with animals treated with both LK and INH to investigate how a more target analysis using computational tools, such as LIRA, can enhance the end result interpretation.

One major limitation of traditional pathology analysis is bias variation that occurs between researchers and laboratories, as previously discussed. To reduce the impact of these limitations, LIRA has been developed as a software tool to assist in digital image analysis for the TB C3HeB/FeJ animal model, with a specific emphasis on TB drug development purposes. In Aim 1, we had previously demonstrated that assistive analysis with LIRA on lung pathology images of infected but untreated C3HeB/FeJ mice showed a more rapid and quantitative readout compared to traditional methodologies, which are only semi-quantitative. One of the limitations of current CNN methodologies is that most models only classify what the model has been trained for. A simple example would be a classifier trained in classifying cats and dogs when given an image of a bird would try to classify the image as a cat or dog, not a bird. Therefore, it was important to investigate whether any changes in the experimental design of the mouse study, such as antimicrobial treatment or host immune response manipulation, would alter the histopathology to the extent that LIRA's performance would decline.

To perform the experiment, LIRA's scores were converted using the score rubric provided by the pathologist, which has been previously outlined. The analysis between the different users (Pathologist, GRA, RA) in which disease involvement for each lesion type was given a score from 0-4 and weighted to reflect each lesion's disease severity. The overall disease severity had almost unanimous agreement except for one cohort, the 3 week INH. The reason for this was the result of the misclassification of an immature Type I lesion (Figure 3.4). These partially-formed lesions

take on the appearance of atypical Type II lesions with small isolated pockets of fibrosis. For future analysis, it has become apparent that disclosure of time points is essential to prevent any misclassification that could potentially occur.

Another important observation in the study was the disagreement between scores for the 6 week LK cohort for the individual Type I scores. Both the pathologist and GRA were able to identify poorly formed Type I lesions (Figure 3.5) due to the fibrotic inhibitory effects of the LK treatment. The RA was not able to identify these lesions, most likely due to LIRA not being trained on atypical Type I lesions. This highlights that LIRA is not a replacement for expertise but merely augments the current classification ability of the user. This limitation may be overcome by expanding the training dataset used in training the CNN to predict atypical Type I lesions more accurately or limiting the usage of LIRA to users experienced in the C3HeB/FeJ pathology.

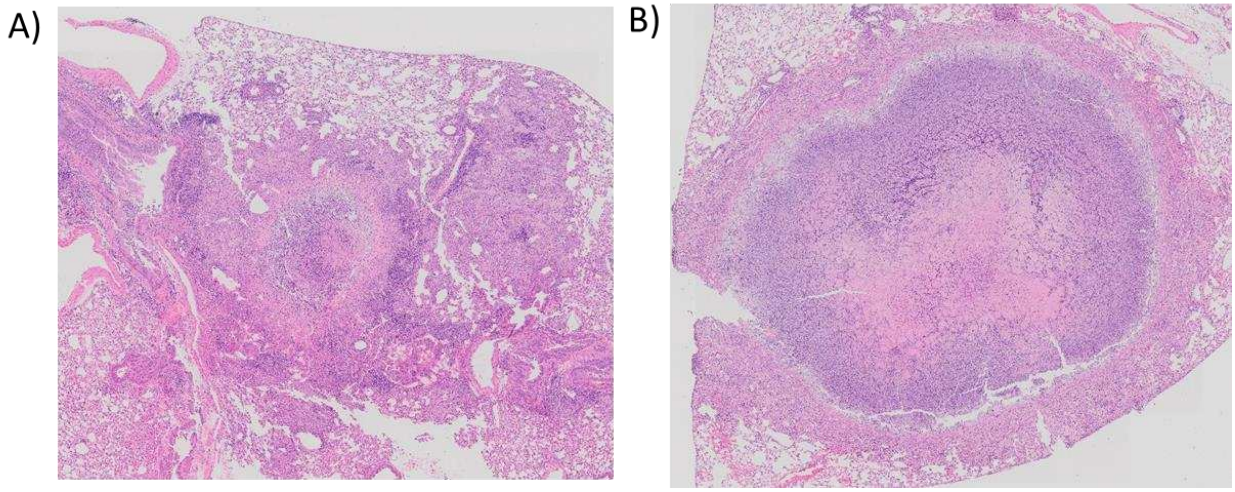


Figure 3.5. Visual comparison of the development of Type I lesions with treatment with either LK (A) or INH (B). Type I lesions from the LK cohort was not as developed as Type I lesions seen within the Control and INH treated mice.

Individual Type III lesion scores showed little variation between the different groups, but the individual Type II lesion scores showed several discrepancies between the classifiers. Most notably, when Type II lesions were the predominant lesion type, the three classifiers were generally in agreement. The discrepancy occurs when predominantly Type I and Type III lesions

have small pockets of neutrophils that are identified by LIRA. When the pathologist is making their classification, they do so at a macroscopic level. In contrast, LIRA performs the majority of its analysis at a microscopic level identifying each individual image patch as one of seven classifications. By following the traditional scoring rubric used even if the pulmonary tissue has only 2% involvement, which would have been deemed inconsequential, it will receive a score of 1. This highlights one of the major limitations with converting ratio data values into a form of ordinal data, which is that there is a substantial loss of information that was previously present in the prior data format. As highlighted earlier, a score of 12%, 24%, and 25% would receive the scores 2, 2, and 3 respectively masking potentially significant differences or similarities. Converting the pathologist scores would be the ideal option but is not possible since no direct measurements are done but are subjective to the pathologist's interpretation of lesion involvement.

The second aim was to determine how using a targeted analysis using computational tools may enhance the result interpretation compared to using traditional methodology. The original study relied on stereological techniques that required a randomized sample selection of different lung areas to represent the population. Based upon this approach, it was concluded that there was no significant difference in the lesion involvement between the different cohorts, except for a possible increase in Type I involvement for the INH treated cohorts [206].

Table 3.1. Mouse mortality separated by treatment cohorts. Adapted from Driver dissertation [206].

Treatment Group	# animals	% of group
Control	10	29
INH	0	0
LK	12	50
INH+LK	7	29

Using LIRA, we have observed substantial differences between the different cohorts that warrant attention. The first is that there was substantially more total lesion involvement for the INH group than previously reported. This consisted primarily of an increase of Type I lesion involvement and received an average score of 3 out of 4, implying considerable lung consolidation. The second observation was that the LK cohort showed considerably less lung consolidation when compared to any other treatment group, consisting of Type I lesions with an average score of 1.4, implying a low degree of lung consolidation. It has been observed previously within the laboratory that lesion severity will, at times, get initially worse with successful treatment due to increased inflammation by dead bacteria. This study marks one of the first times that this effect has been quantified. Both the results from the LK and INH are counterintuitive of the paradigm that effective treatment results in the decrease in lung involvement.

When comparing mortality, the LK group within the study showed significant mortality (50%), followed by the untreated controls and INH+LK cohort (29%), and no mortality for the INH group (0%) (Table 3.1). The mortality in the LK treatment group was of particular interest because the LK group showed a considerable reduction in bacterial burden from 8.57 $\pm$ 0.58 log10 CFU at 3 weeks to 6.01 $\pm$ 0.45 log10 CFU at 6 weeks. A similar reduction in bacterial load was also seen in the INH group, which at 3 weeks had 8.01 $\pm$ 0.38 log 10 and was reduced down to 6.29 $\pm$ 0.34 log10 at 6 weeks. However, INH treatment as expected resulted in increased survival.

This indicates that bacterial burden and lesion severity alone cannot be used as an indication of increased survivability in mice, but additional quantifiable metrics are needed.

Future studies will be using time points that extend past the 10 week post aerosol infection for which LIRA had been originally trained. While we have successfully demonstrated here that animals treated and images collected 6 to 10 weeks post aerosol infection will achieve comparable scores given by a board-certified pathologist, no analysis has been done for cohorts analyzed past this time point. A future direction would be as soon as data becomes available to perform a similar experiment using treated animals at later time points (10 to 20 weeks post aerosol infection) to find the limit of LIRAs accuracy. If the model fails to classify the pulmonary lesions correctly, there can be two approaches to correct these inconsistencies. The first would be to create a new CNN model that specializes in animals 13 to 20 weeks post aerosol infection. This means that based on the time of collection of the tissue samples, they will have to go to one of two different models that either is trained for samples 6 to 10 weeks or 10 to 20 weeks. The second approach would be to have one classification algorithm that can handle any time point. This will require continuing the training of the original CNN classification algorithm with additional data that had been collected past the 13 week time point. The second approach is not recommended since it would require substantially more data than the first approach. Collecting additional data is difficult and is the limiting factor with biology-based ML. In summary, users using LIRA generate scores comparable to a board-certified pathologist and provide additional data that aids in coming to a correct study conclusion.

In summation, LIRA achieves comparable results to a pathologist, is quantitative, and can thereby provide additional insights that may influence the interpretation of the experiment outcome. LIRA's major advantage is that the calculated disease severity includes the targeted percent involvement for each lesion type, how many Type I lesions are present, and does so in a timely and reproducible manner. As data acquisition and analysis becomes more complex, the development of computational tools will be required to more accurately represent what is actually

occurring in the animal models during infection. As shown without these tools, conclusions that do not accurately depict the experiments results may unintentionally overlook successful treatments or important observations.

CHAPTER 4: DEVELOPMENT OF MYCOBACTERIAL IMAGE ANALYSIS (MIA) SOFTWARE TO QUANTIFY LESION SPECIFIC POPULATIONS *OF MYCOBACTERIUM TUBERCULOSIS*

4.1 - Introduction

Human TB is a dynamic disease that develops a variety of pathological and immunological features in the lung, thereby creating a complex niche of pulmonary microenvironments for Mtb to occupy. For that reason, phenotypic analysis needs to be done in the context of the TB lesion. Lesion microenvironments in the lung that develop as the result of infection include varying levels of iron (Fe) [225], vascularization [226], hypoxia [227], pH [18], [19], carbon and energy availability [228], and immunological pressures. Mtb is present either intra or extracellularly in the lung. Over the course of months and even years, TB lesions are constantly fluctuating in both size and metabolic activity, as demonstrated using PET/CT imaging in human patients [46], [158]. Mtb has multiple mechanisms to successfully adapt to the myriad of environments and stresses that it will encounter. For example, within macrophages, Mtb will increase its lipid metabolism through isocitrate (icl), siderophore production to scavenge iron, and sigma factor production[229]. Within the necrotic granuloma, Mtb has been shown to elevate host lipid and cholesterol production[230], increase rates of replication [139], and induction of the dosR and devR transcription regulators[231]. Understanding exactly how TB responds to its environments and the different heterogenic bacterial populations that compose each specific lesion type is essential in designing targeted therapies to increase the chances for a successful treatment.

TB treatment is highly effective in the first weeks of treatment but is generally slowed by the occurrence of drug tolerance. Two types of drug tolerance have been described: the first is stochastic and seen as persistence after drug treatment of actively replicating bacilli (Class I), and the second is caused by the specific environmental locations where residing Mtb bacteria are non-replicating (Class II), [232]. TB drug development tries to overcome both types of drug tolerance to cure the infection and shorten treatment.

The first type of drug tolerance, or Class I, addresses the observation that multiple phenotypic populations can exist concurrently within a single lesion after drug treatment. The surviving Mtb populations are reversibly tolerant to one or more standard therapies but not always to therapy combinations[232]. Previous dogma had assumed that there were only two distinct populations of Mtb: dormant and active [233], [234]. Active bacteria were defined as replicating and metabolically active. This population of bacteria are susceptible to both antimicrobial treatment and host immune responses. Dormancy was defined as non-replicating bacteria that are in a quiescent metabolic state. Bacilli in this dormant state were considered to be resistant to both antimicrobial therapies and host immune responses [235]. This had previously explained how difficult it was to achieve sterilization during treatment and why some hosts seemed to stay in the latent phase of the TB infection. It has been observed that the transcriptional profile of drug-resistant Mtb is in a slow-growing, metabolically, and synthetically downregulated state [236]. Recent studies have expanded upon this hypothesis by showing that within a single uniform environment, multiple phenotypic populations can exist concurrently [237], [238]. *McKinney et al.* demonstrated *in vitro* that the surviving bacterial population under treatment with isoniazid showed three distinct populations; the first was a subpopulation that quickly resumed growing and remained metabolically active, the second was non-replicating but remained metabolically active, and the third population was both non-replicating and was in a quiescent metabolic state [140]. *In vivo,* they are represented by a biphasic reduction in the Mtb burden that continues before failing to reduce the burden below a certain level of detection. More importantly, Class I phenotypic resistance has been observed in human patients that are receiving the standard drug therapies[239] and within murine [71] and Guinea pig animal models [240].

The second type of drug tolerance is reflected by a bacterial population located within specific lesion compartments that do not behave uniformly across the various lesion types (Class II) [232]. Each compartment microenvironment can present with different challenges that need to be addressed to design therapies that target a variety of phenotypic bacterial populations,

successful delivery, and environmental conditions (e.g., pH) to clear an infection. Stressful environmental conditions can induce phenotypic drug tolerance that can halt net Mtb replication. Non-replication can be the result of transcriptional responses to the environment or an altered cell wall [241]. Host microenvironments are further complicated by the fact that within a human host, multiple different lung lesion types can exist simultaneously during infection [14]. Lesions are also highly dynamic and are consistently changing throughout the lifetime of the patient. Similar to pulmonary TB in patients, the C3HeB/FeJ animal model also develops a variety of different lesion types in the lungs displaying various levels of hypoxia, intra and extracellular bacteria, diversity of carbon and energy sources, and immunological pressures [74]. By utilizing the C3Heb/FeJ TB mouse model, we can better understand phenotypic antibiotic resistance that can arise as the result of environmental stimuli.

Utilizing an animal model that develops a heterogeneity of lesions types allows for the targeted visual analysis of specified bacterial populations within specific lesion compartments. In the C3HeB/FeJ mouse model, there are five main pulmonary environmental niches where Mtb can be visualized (also detailed in Chapter 2). The first and second environmental niches are the Type I lesion that can be further divided into the Type I Core and the Type I Rim. The Type I Core is hereby defined as the caseous necrotic center enclosed within the fibrotic rim where the bacteria are extracellular, in a lipid-rich environment, neutral pH [75], and are under hypoxic conditions due to a lack of vascularization [74]. The Type I Rim consists of the collagen rim as well as the consolidated region just outside the fibrotic rim. It is neovascularization, and bacilli are generally intracellular within the surrounding macrophages. Type II lesions are not highly vascularized, have large areas of cellular necrosis, karyorrhectic debris, and bacilli are intracellular within neutrophils. Type III lesions are the result of a host cellular immune response, and a limited number of bacteria are intracellular within interstitial macrophages. The last environment is the uninvolved lung parenchyma, where very few to no bacteria are located, here defined as Healthy tissue. The optimal animal model will attempt to mimic the complexity

observed in human TB disease since a treatment that would be effective in one particular lesion may not be efficacious in another lesion.

By microscopic analysis, the heterogeneity in bacterial populations has been reported on by acid-fast staining of either *in vitro* cultures or in tissues [73], [192]. The stress of the host immune response can affect the bacterial transcription thereby modulating the composition of the mycobacterial cell wall. Mimicking intracellular conditions *in vitro*, using low oxygen, nutrient deprivation, and acidic pH, the bacilli can lose its acid-fastness and accumulate triacylglycerol and wax ester [242], [243]. In macrophages, a subgroup of Mtb mycolic acids are substantially altered [244]. Bacilli in standard laboratory mouse strains have also been documented to lose their ability to retain the Ziehl Nielsen acid-fast stain over time [245]. Even though these studies highlight the ability of Mtb to alter their cell wall *in vitro*, this was also demonstrated previously *in vivo* in our laboratory. *Ryan et al*. visualized the cell wall heterogeneity of Mtb lesion populations in Guinea pigs and mice by using two separate staining techniques: Auramine-Rhodamine (AR) and fluorescent *in situ* hybridization (IF). These stains identified three separate bacterial subpopulations in equal proportions that either stained with AR, IF, or a combination of both stains and whose populations were homogenous throughout the tissue. This included tissues that were hypoxic and/or necrotic, which always had at least three distinct populations [73]. Both *Ryan et al.* and *McKinney et al.* establish that antibiotic targeting of a single Mtb population will not be efficacious in successful treatment and prevention of relapse.

Aim 3 of my dissertation proposes the development of novel quantitative software that will enumerate bacteria at the various locations in the lungs of the C3HeB/FeJ mouse model. This software will measure the number of visual CFU in a specified area (density), percentage of the visual field composed of bacteria (% bacteria), and individual bacterial CFU area. The software will be referred to as the **M**ycobacterial **I**mage **A**nalysis or MIA for the remainder of this paper. The goal for this initial study was to test the MIA software and to use MIA for the first time on a drug efficacy trial by quantifying the different bacterial metrics for each lesion type found within

drug-treated C3HeB/FeJ mice. For this purpose, microscopic slides from an earlier mouse study were used to optimize and validate the MIA software. The C3HeB/FeJ mouse study was described in this dissertation in Aim 2 (see Chapter 3), and detailed the treatment with either isoniazid (INH) with or without losartan (LK). This mouse study was chosen as comprehensive data was already available, by having both lung CFU data from plating on 7H11 plates, pathology data after manual histopathology, as well as analysis by LIRA.

For this Aim, the bacterial enumeration for every lung lesion compartment of the C3HeB/FeJ mice was quantified for every treatment group. The same pathology classifications for the diverse lung lesion compartments were used as detailed in Aim 1 (Chapter 2): Type I Core, Type I Rim, Type II, Type III, Healthy, and Miscellaneous. The same classifications were again used for the quantification with MIA, except for the addition of the Type I Core Outside. Tissue samples collected previously were stained using SYBR Gold, which is a DNA/RNA nucleic acid stain that is selective for Mtb [192]. SYBR Gold has a >1000 fold fluorescent enhancement when binding to both double or single-stranded DNA or RNA and has low levels of fluorescent background noise. This makes it ideal for image analysis by providing a stark contrast between the object of interest and the background noise. Using the SYBR Gold staining method, we optimized the MIA software to this stain and showed the utility of MIA in a preliminary study to accurately identify bacterial clusters and distinguish changes that occur between time points and treatment cohorts.

4.2 – Methods

4.2.1 Mouse Infection

C3HeB/FeJ mice infected with Mtb Erdman strain via aerosol infection (See Chapter 3 Methods)

4.2.2 Staining and Imaging

Histology slide samples were stained with a diluted 1:1000 SYBR dye (Invitrogen, Eugene, Oregon) that was diluted with a solution composed of phenol crystals (8g), glycerin(60mL), isopropanol (14mL), and distilled water (26mL). The slide was flooded with SYBR Gold solution on a heating block at 65° C for 5 min and allowed to cool for 5min at room temperature. Slides were washed in an acid alcohol (0.5% HCL, 70% isopropanol) bath 3 times for 3 min each followed by a water rinse. Samples were counterstained with Hematoxylin Quick stain and mounted using Prolong Gold antifade mounting medium [192]. Samples were imaged at 20x using both DAPI (Tissue) and FITC (SYBR+ bacteria) channels (See Chapter 2 imaging methods). The histology slides that had been processed for the INH+LK cohort were damaged, and no bacterial metrics were acquired using MIA as a result.

4.2.3 MIA Algorithm Pipeline

MIA was developed in the Python (version 3.6.7) language using Tkinter, PIL, Numpy, Pandas, and OpenCV libraries to perform the analysis. Along each major step of the image processing, representative images are created to help the user understand what objects were being gated and collected for each region of interest (ROI) (Figure 4.1). The ROI is the specific tissue section that is selected by the user. Available microenvironment classifications to subdivide the populations include Type I Core, Type I Core Outside, Type I Rim, Type II, Type III, Healthy, and Unknown. The Type I Core is defined as the area in the center of the caseous necrotic lesion. The Type I Core Outside is the outer cellular edge of the core that is primarily composed of the foamy macrophages and neutrophils which are encapsulated within the fibrotic rim. The Type I Rim is the lesion tissue that starts on the outside of the collagen rim and the surrounding involved tissue. The Type II, Type III, Healthy, and Unknown were defined previously in Chapter 2.

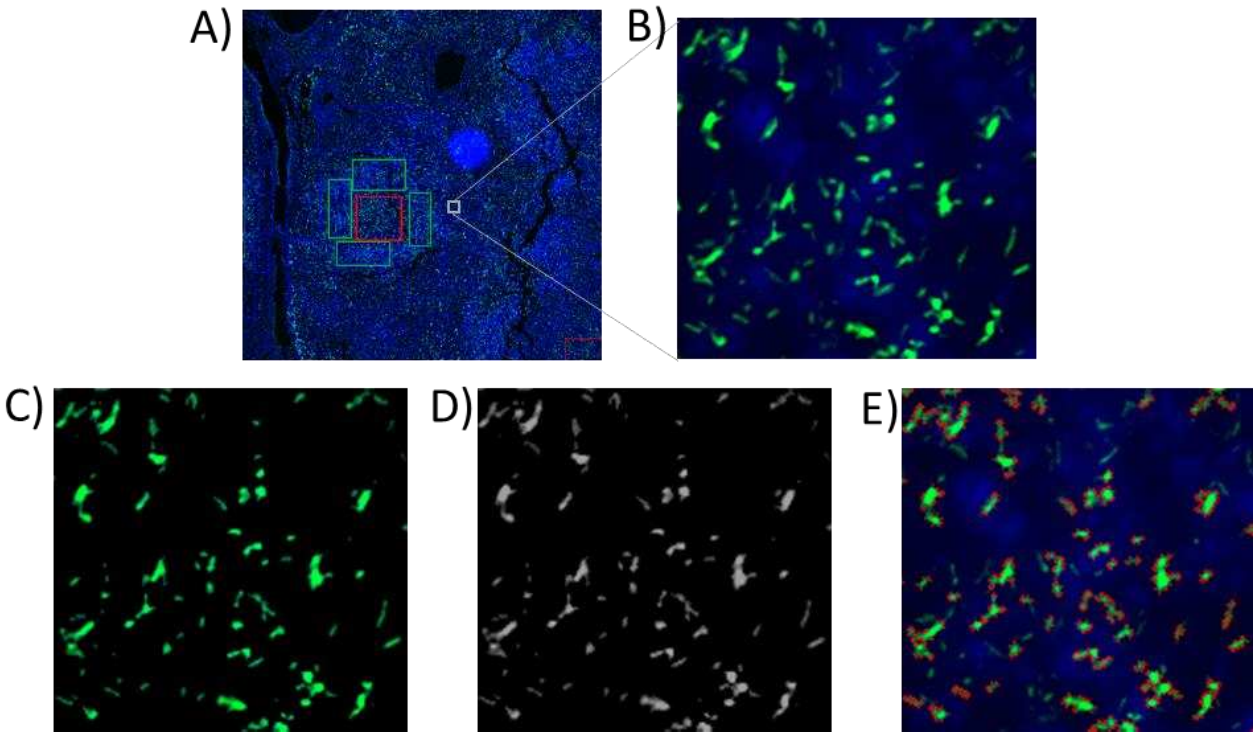


Figure 4.1. Visual representation of each of the steps performed in MIA's pipeline analysis. The image is uploaded into MIA, and the user classifies the lesion area (A), the ROI is cropped (B), converted to an HSV image where the H channel is isolated and a mask is generated based upon the desired stain color channel (C), the image is further converted to a grayscale image (D), and the remaining image object edges are detected and outlined in red (E).

4.2.3.1 Number Of Defined Bacterial Areas per μm^2

The various steps in the MIA pipeline are discussed below. In the initial step, the user identifies the classification for a particular ROI (Figure 4.1A), and a separate temporary cropped image is created for the analysis (Figure 4.1B). The temporary cropped RGB image is converted to an HSV, and the color is masked for the specified staining parameters on the H channel (Figure 4.1C). For instance, the SYBR Gold will only accept pixel values between the minimum =[50,1,1] and maximum =[85,255,255]. The image which only contains the specified color of interest will be converted to a grayscale image (Figure 4.1D). Using OpenCV's contour function the edges of detected objects (e.g., bacilli) are found. (Figure 4.1E). The area of each object is calculated, and

objects, or aggregates, that are too small or too large are removed from the detected object list. The remaining objects are highlighted in red as verification that the correct objects of interest had been detected. The number of remaining objects in the list is the number of objects (bacilli) that are found within the ROI. The total number of bacilli is divided by the total converted image area to provide the density of defined bacterial areas per unit^2. The default unit setting is μm^2.

4.2.3.2 Individual Bacterial Area Analysis

Each identified object of interest contoured edge coordinates are stored. Based upon each coordinate pairs, the sum total of the number of pixels located within the coordinates is the pixel area. Pixels are subjective, meaning that they need to be converted to standardized units to be able to make meaningful comparisons. The image size and individual objects of interest size in pixels are converted using the below formula:

$$Image\ Pixel\ Size = \frac{(Camera\ Pixel\ Size)(Binning\ Size)}{(Objective\ Lens)(Lens\ Magnification)(C\ Mount)}$$

Each converted object of interest area is stored as a separate within the .csv, along with the identifying metadata for large data analysis.

4.2.3.3 Percent of Image Composed of Bacteria

The summed area of each object of interest is divided by the total image area to calculate the percentage of the image that is composed of the objects of interest.

$$(\%\ Bacteria\ in\ Image) = \frac{\sum(Individual\ Bacterial\ Area)}{Total\ Image\ Area}$$

Additional metrics not used in this paper are also calculated and provided in a separate .csv file, such as the fluorescent intensity of the mean objects of interest.

4.2.4 MIA Optimization and Validation

Each of the three bacterial metrics (density, area, and percentage) are based on values incorrectly identifying the objects of interest and correctly identifying the edge coordinates. Object identification and edge detection were validated using Flow Cytometry Beads (Invitrogen CountBright beads lot# 1971889, cat_num: C36950) with a manufacturer value of 7.08 ± 0.478µm as the diameter. Beads were placed on a slide using a Shandon Cytospin 6 and mounted on microscopic slides using mm24 mounting medium. Slides were imaged at both 10x and 20x objectives and saved as .png files to prevent file compression and possible distortion of image objects. The accuracy of MIA object enumeration was calculated using values determined by manual observation and those processed by MIA. Pixel areas calculated by MIA were converted to µm, and the mean of all identified objects was used to compare to the bead values reported by the manufacturer. Accuracy of the analysis can vary based upon the objective of the microscope used to acquire the images. For example, an image taken at 100x will more accurately demonstrate the bead size than an image taken at 10x. The acceptable threshold of the calculated MIA value for the mean needed to be within the bead standard deviation of the mean provided by the manufacturer.

4.3 - Results

4.3.1 - Verification of the Software on Control Samples

MIA was designed to measure three important bacterial metrics: 1) the number of bacteria per tissue area, which are defined further as “density”, 2) the individual bacterial area, and 3) the percentage of the image that is made up of the bacteria. All three metrics rely on two core algorithms that either enumerate the number of fluorescently labeled objects of interest (e.g., bacilli) or calculate the individual object of interest area. To validate the algorithms and the corresponding conversions, the fluorescent flow cytometry beads were placed on a microscopic

slide and scanned to create a digital image. Visual representations of each of the steps used in the analysis were collected for both visual and quantifiable verification. The essential steps included placing the beads on a slide, scanning to create a digital image, and passing the image through MIA's core function for quantification. The function includes the creation of a mask to isolate the fluorescently labeled flow cytometry beads (Figure 4.2B) and calculates the object of interest edges by contouring (Figure 4.2C). MIAs object detection achieved 99.04% accuracy by identifying 208 out of 210 possible beads. The discrepancy was the result of two beads that had co-localized with another bead and appeared as one single object (Figure 4.2C, blue arrow). This means that it correctly identified all isolated objects, but cannot consistently discern objects that have co-localized together.

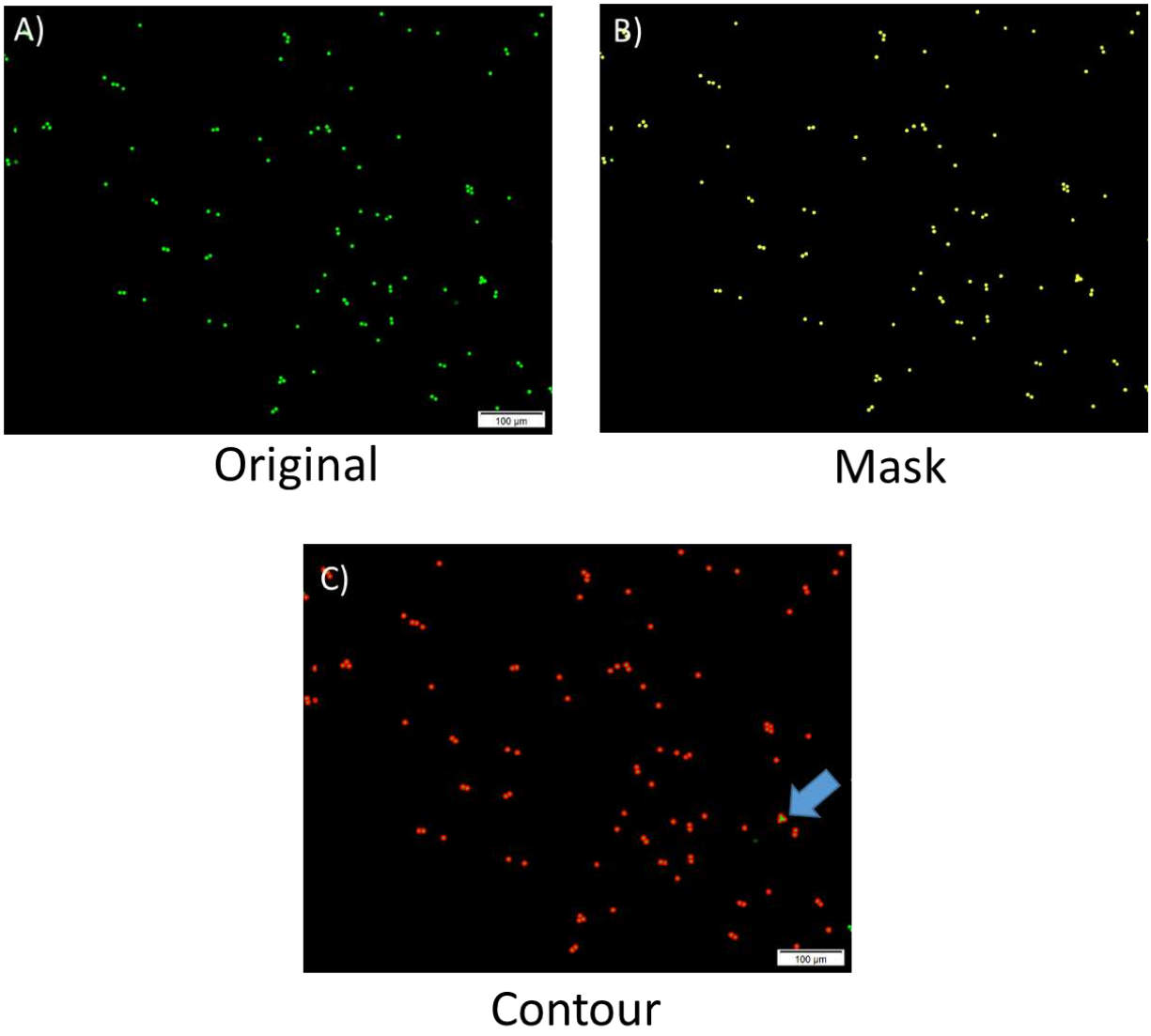


Figure 4.2. Visual representation of MIA's ability to identify objects of interest, quantify the number of objects, and determine their area. Blue arrow indicates bead cluster that was identified as 1 object but was actually three separate objects that had colocalized together.

Next, we needed to evaluate MIA's ability to correctly calculate the areas of objects of interest within an ROI. It was critical to have the results in a format that is easily interpreted and compared to images generated from other microscopes, cameras, and objectives. After the software has identified individual objects of interest and determined their location within the image, the number of pixels within each object are calculated. By knowing several variables about

the microscope used to generate the image (such as Pixel length, objective, binning), the pixels can be converted to metric units such as μm^2. According to the manufacturer's documentation, the flow cytometry beads have a diameter of 7.08 ± 0.478 µm and a mean bead area of 38.4845 μm^2. Measuring 200 beads using both 10x and 20x objectives, MIA calculated the mean bead diameter to be 6.66 ± 3.16 µm and the mean area to be 34.8563 μm^2. Manufacturers will perform a more robust analysis than what was performed here, but the purpose here was to demonstrate that MIA's calculated values were within an acceptable range to the values provided by the manufacturer. If accuracy is still a concern, images can be scanned at a higher magnification, or a higher resolution camera can be employed. In summary, the calculated values for both object of interest enumeration and object of interest area calculations were within acceptable ranges to be used for our analysis.

4.3.2 – Verification Of The MIA Software On C3HeB/FeJ Histology Tissues

MIA had demonstrated accuracy in correctly identifying flow cytometry beads within an image and calculating the area of each bead. We next wanted to evaluate LIRA's ability to identify Mtb within a histology sample correctly. The first piece of data that was needed was to determine the minimum size of bacilli that could be visualized. The average length of a bacilli is 2.71 ± 1.05 µm in length, and the average diameter of the cell was 0.345 ± 0.029 µm [246]. This means the minimum visual area of an Mtb cell that is cut perpendicular would be 0.16 μm^2 and is the minimum value used for MIA. As demonstrated in Figure 4.3B, MIA was able to identify all possible individual and aggregated bacteria within an image, excluding those that are touching the edges of the image. In addition, the user can select specified sizes as demonstrated in Figure 4.3C. Correctly identifying the edges of the Mtb bacilli is the primary algorithm required of MIA, and as

demonstrated, it correctly identifies the coordinates for each bacillus and has the option to select area size as selected by the user.

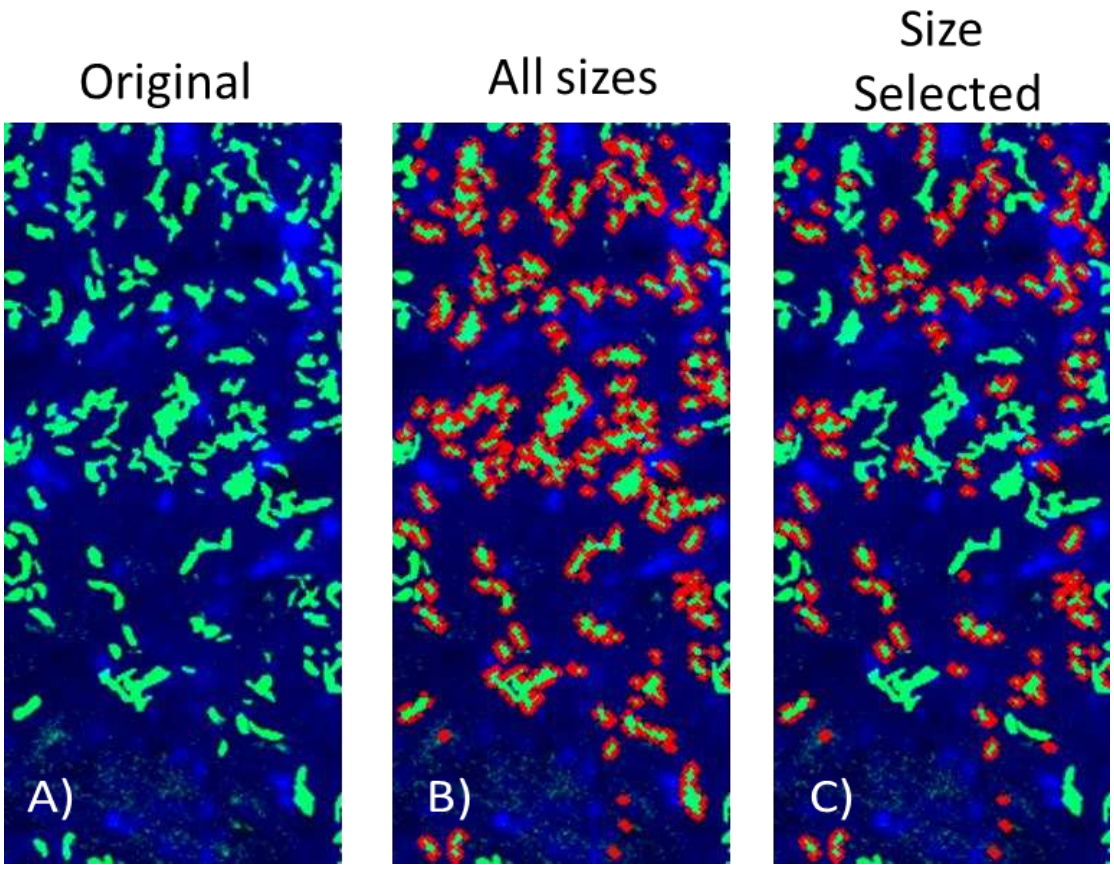


Figure 4.3. Visual representation of MIA's ability to identify SYBR Gold stained bacilli within a histology tissue sample. Bacilli (Green) would be outlined in red if they were identified by MIA. A) the original image, B) all objects not touching identified by MIA, and C) bacilli selected based on size.

4.3.3 – Analysis Of Drug-Treated Bacterial Populations *In Vivo*

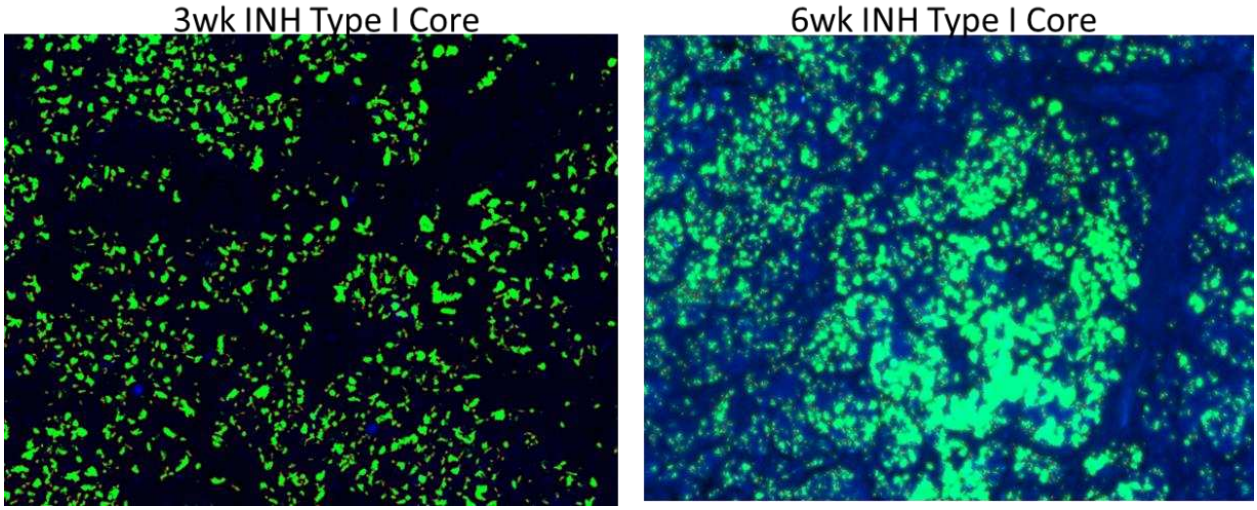


Figure 4.5. Difference in bacterial aggregation within the INH cohort Type I Core at 3 week and 6 week treatment. A similar effect also observed in the Control Cohort.

Within the prior study performed by a previous Ph.D. student in our laboratory (Emily Driver), it was concluded that there was no substantial visual bacterial burden difference between time points and cohorts [206]. When analyzing these same slides with MIA, there were differences in bacterial numbers and clustering observed between the various treatment groups. For instance, within the INH and Control mouse groups for the core of the caseous necrotic lesions (Type I Core), MIA showed a significant aggregation occurring between 3 and 6 weeks (Figure 4.5).

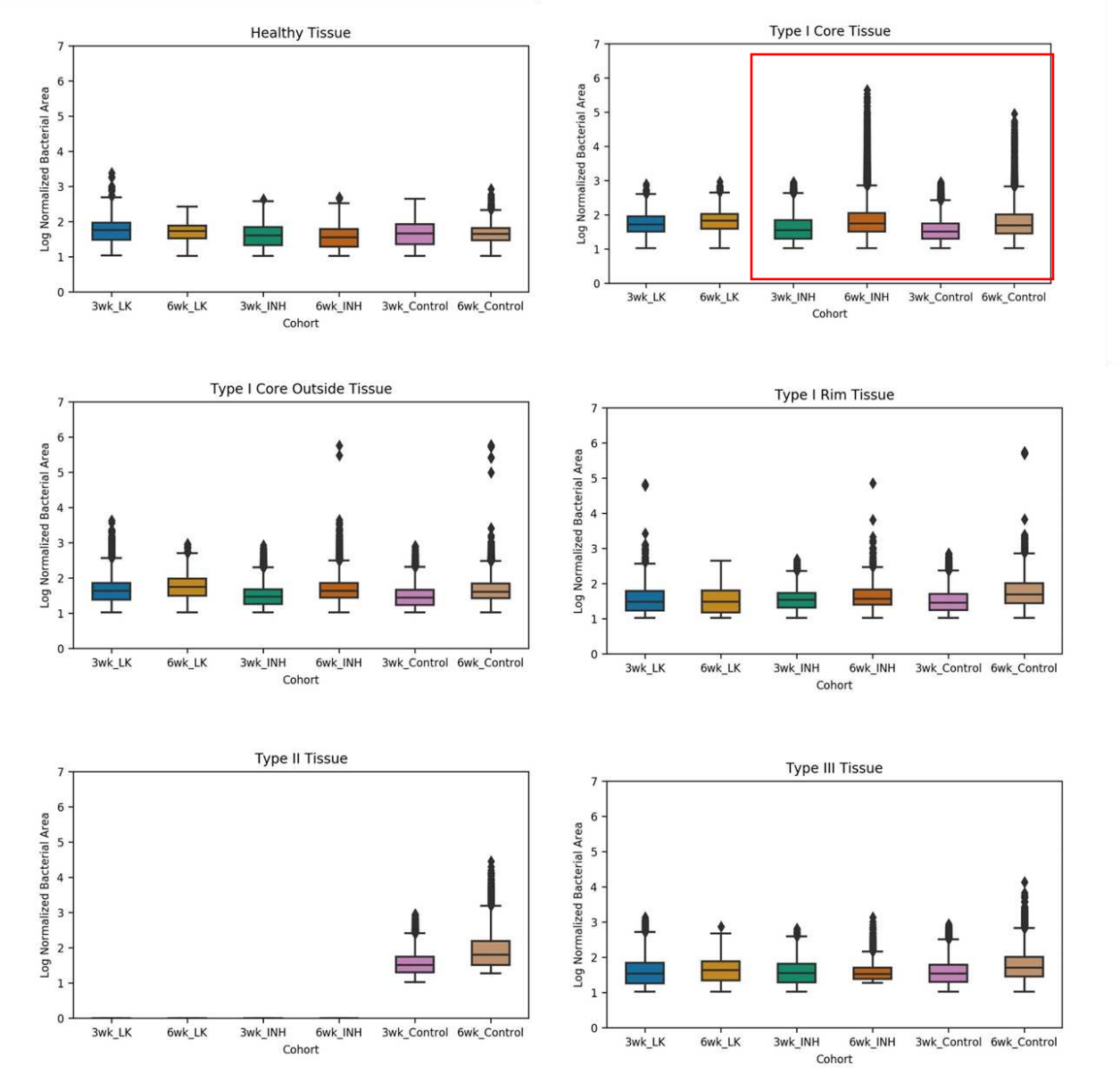


Figure 4.6. Boxplots of $\log_{10}$ normalized individual bacilli and aggregate areas separated by lesion type and cohort. Red box indicates the specific cohort of interest for the analysis.

For the measurement by MIA of bacterial areas for both the INH and Control cohort, the number of pixels was calculated for each object of interest by the enumeration algorithm, and each pixel was converted to μm^2. Populations were log base 10 normalized and then visualized using boxplots. For the different treatment cohorts and lesion compartments, there were no substantive differences observed in the mean and interquartile range. What was of interest were

some of the outliers found in each cohort. For both the INH and Control groups, MIA showed that in the caseous necrotic lesions (Type I Lesion) a dramatic increase in large bacterial areas could be observed after 3 to 6 weeks treatment (Figure 4.6).

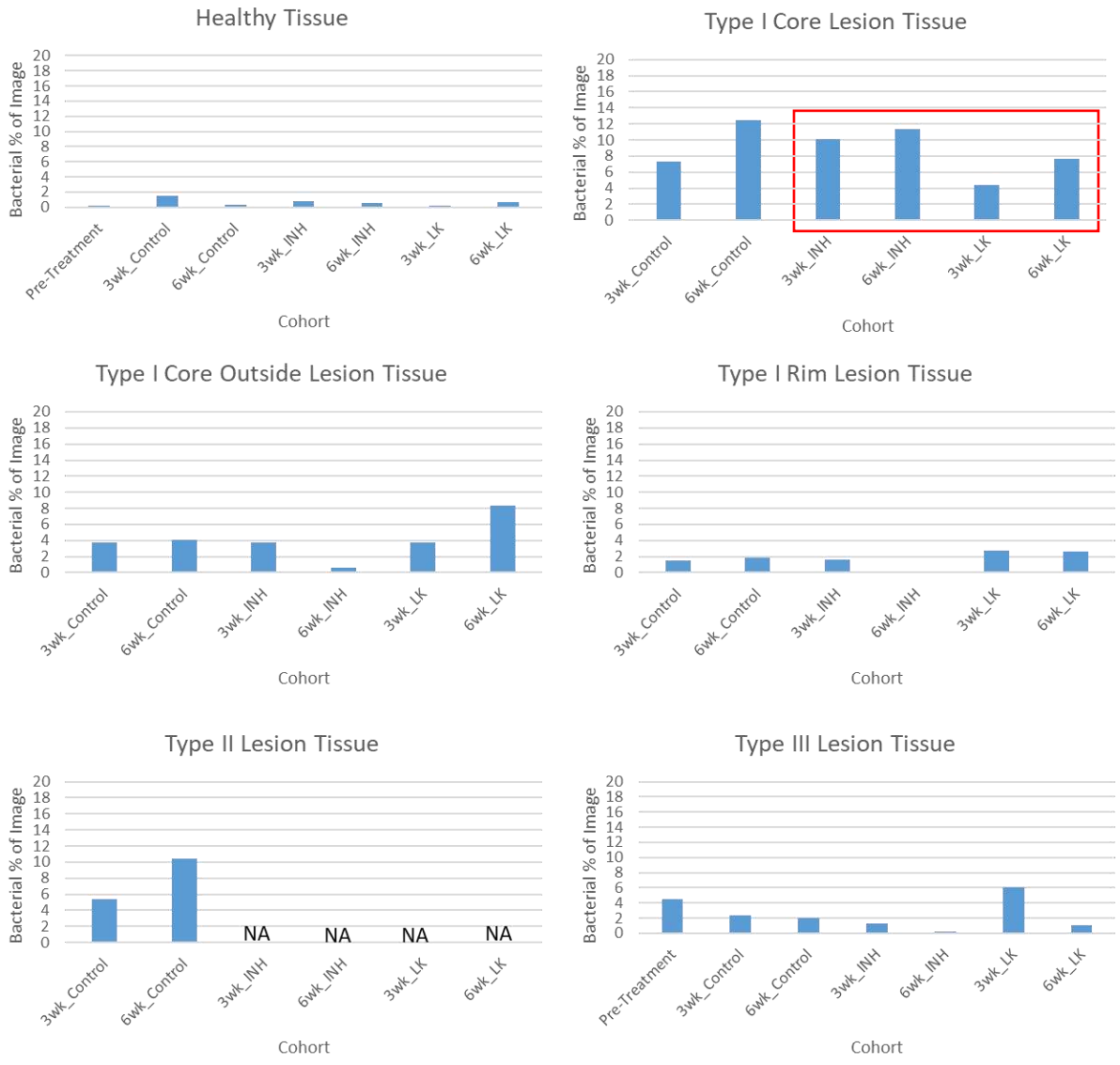


Figure 4.7. Percentage of the image that is composed of visible bacteria. Red box indicates the specific cohort of interest for the analysis. NA indicates that no sample was measured for that particular data point.

Another measurement by MIA calculates the percentage of the image composed of visible bacteria (Figure 4.7 Type I Core Lesion Tissue). This measurement showed an increase in the percent bacteria present for both Control and INH Cohort. Results were verified using a separate calculation by summing all green pixels within the image, which also resulted in a percentage of the image occupied by bacteria (results not shown). While this is a less accurate measurement, as it does not discriminate object size, it did achieve similar values thereby confirming the results obtained by MIA.

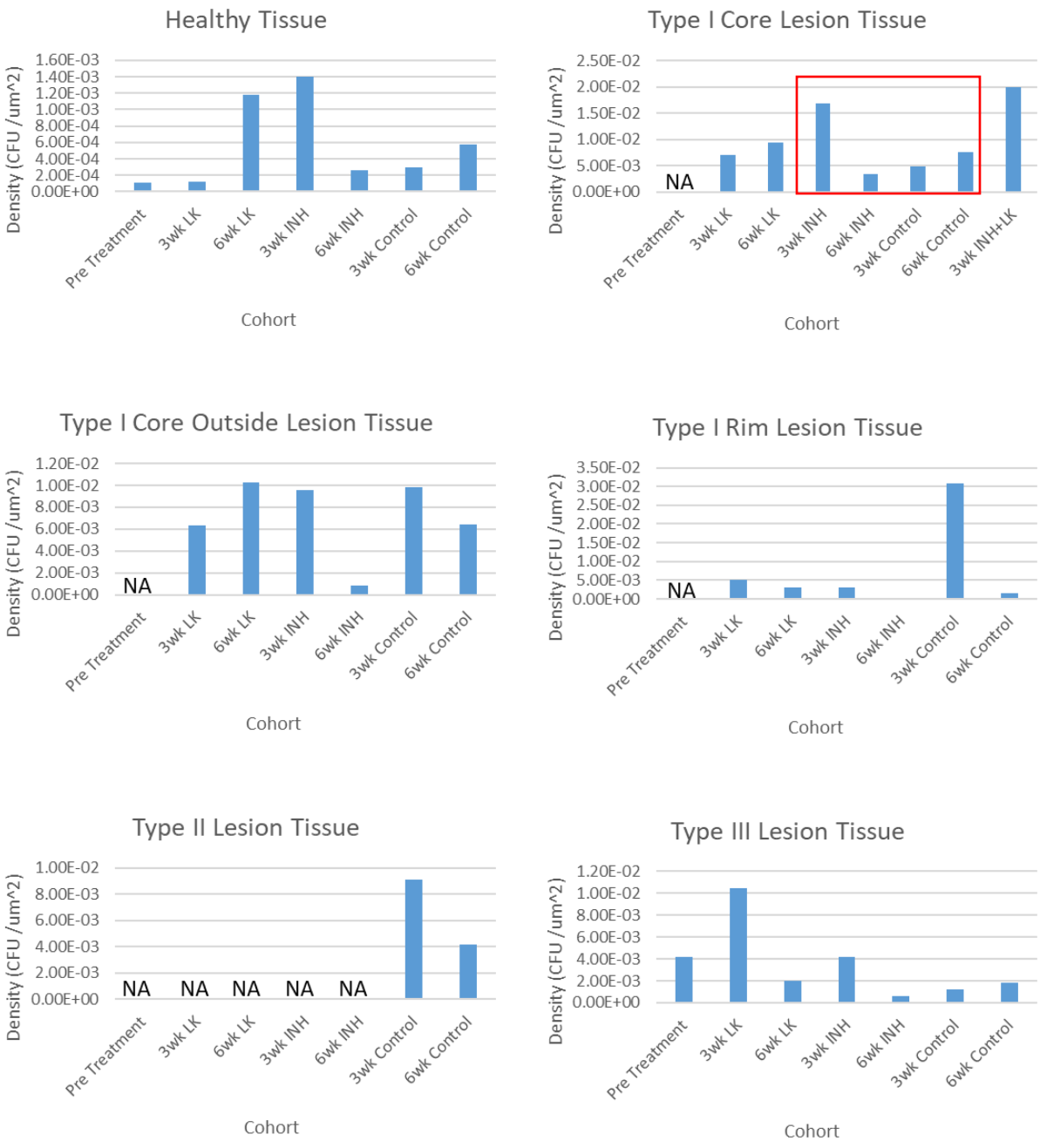


Figure 4.8. Bacterial densities of each lesion type calculated in terms of (defined bacterial areas) / μm^2. The red box indicates the specific cohort of interest for the analysis. NA indicates that no sample was collected for that data point.

The next measurement by MIA was obtained by calculating bacterial density. By applying this on the same mouse study, we saw a decrease in bacterial density from 3 to 6 weeks for the INH cohort, but a weak increase in the Control cohort (Figure 4.8). The INH decrease was as

expected whereas the slight increase in density for the Control cohort was not. After closer visual inspection of the Control cohort, we noticed less bacterial burden at 3 weeks, which could account for the slight increase in defined bacterial areas density (Figure 4.8). Because the other two metrics behaved as predicted, the results suggest the achieved values for the bacterial density for the Control cohort are real.

Of note, these measurements by MIA described above were performed on this one pilot study only, and therefore should be seen as preliminary results. The reason for the quantification of the various measurements by MIA was to 1) test the MIA software for the first time on a C3HeB/FeJ mouse efficacy study to assess flaws in the software, and 2) see if there were any advantages of using MIA, such as finding discernible differences among treatment groups that were not seen in the past after manual histopathology analysis. These results showed that quantification by MIA was successful and that it was able to find differences in bacterial numbers and aggregation in the various treatment groups.

4.3.4 - Mycobacterial Islands

One observation that was made during our analysis and verified quantitatively using MIA was the development of consolidated Mtb super aggregates in the center of fully developed Type I lesions (Figure 4.9), hereby named mycobacterial islands in this Chapter. These mycobacterial islands developed 6 weeks post-treatment for both the INH and Control cohorts. In contrast, the LK cohort did not develop any islands and had an evenly distributed Mtb burden across the Type I Core. The mechanism of the formation of these super aggregates is unknown and will require more targeted analysis to discern. Even though these islands have been seen in previous studies within our laboratory, future studies will have to show the mycobacterial island phenomenon is reproducible and not caused by technical aspects such as differential staining or artifact generation.

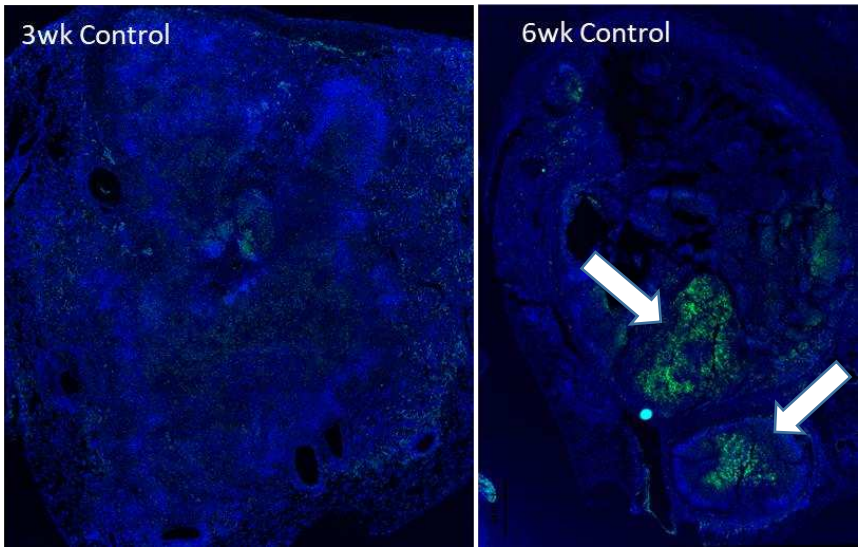


Figure 4.9. Each drug treatment imaged with DAPi and SYBR GOLD at 20x for both 3 week (3wk) and 6 week (6wk) treatment time points. For the Control and INH cohorts, there is the development of mycobacterial islands (white arrows).

4.4- Discussion

TB disease is highly complex due to the lack of uniformity of responses for both host lesions and bacterial populations as the disease progresses. A TB-infected individual can develop a variety of different lesions with vastly different levels of carbon and energy availability, immune system response, environmental stresses, and physical barriers. LIRA was developed to quantify the complexities of the host pathology by quantifying various pulmonary lesions that develop throughout a TB infection, specifically in the C3HeB/FeJ mouse animal model. Within these host lesion compartments, recent studies have shown that multiple different bacterial phenotypes, such as variations in cell wall composition, metabolism, and replication properties, are present within a single lesion compartment. Previous limitations in collecting quantitative data on bacterial populations included the proprietary image format used during scanning, the memory requirements to store and process the images, bacterial enumeration in tissue, and the different lesion compartments. MIA was developed as a means to provide quantifiable data of bacterial

populations in different compartments by lowering the barrier for quantitative analysis for researchers not familiar with coding. In the C3HeB/FeJ animal model, MIA can more accurately measure the density of bacilli in the necrotic core (Type I Core compartment) between a treated and non-treated animal.

One of the limitations of generating this value is that it cannot distinguish single bacteria from aggregated bacteria. This is not a new issue [247], and this could be addressed in future iterations of the software. A predicted number of bacilli can be determined within an aggregate by the user by calculating a frequency distribution of individual planktonic bacteria areas and applying it to each bacterial aggregate to predict a range of bacteria that are present. Calculating the mean distribution is important because the mean Mtb length is 2.71 ± 1.05 μm and the mean diameter is 0.345 ± 0.029 μm that could have even greater variability based upon strain used, location within the lesion, and the specific stress it is encountering. This would only provide an estimate of the bacterial number. The comparison of bacterial burden between cohorts or compartments is more accurately represented by the bacterial percentage of the image that is calculated by MIA.

The use of MIA on the drug efficacy study performed by Emily Driver [206] was used as a preliminary test case to determine the validity of MIA as a new tool to quantify bacterial populations. In these preliminary observations with MIA, either anti-mycobacterial treatment with INH or no modulation of the immune system (Control cohort) resulted in the formation of mycobacterial islands within the necrotic lesion cores (Type I Core Lesion Tissue compartment). No mycobacterial island formations were formed within the treated LK cohort, even though there were some caseous necrotic lesions present. The opposite occurred where there was either no change or a substantial decrease in measuring bacterial aggregation. This test case showed that 1) the MIA software was easy to use on a C3HeB/FeJ mouse efficacy study, and 2) was able to discern differences among treatment groups that were not seen in the past after manual

histopathology analysis. The next step will be to analyze more *in vivo* studies to assess the effect of treatment response on the bacterial populations in the various locations of the lung.

Now that MIA has been optimized and validated on both control and *in vivo* samples, its capabilities can be expanded using additional staining methods to characterize and potentially identify different bacterial populations. This can include using immunofluorescence with antibodies against specific mycobacterial targets [248], stains that bind to mycobacterial cell wall components such as Auramine Rhodamine [249], and *in situ* hybridization probes [250]. The applications are numerous, and for drug development, these additional metrics can broaden the current understanding of how various drug treatments impact specific populations of Mtb within various tissue/lesion compartments. With the availability of large data and image analysis methodologies, it is possible not just to measure subsets of populations but to analyze on a single cell basis.

In summary, MIA can accurately enumerate bacteria to calculate a bacterial density in an area on a microscopic slide, calculate the area of each object of interest, and determine how much of an ROI is composed of Mtb. By combining MIA, which quantifies bacterial properties, with LIRA, which quantifies host pathology response, the depth of information that can be gleaned from each experiment is greatly enhanced from previous manual methodologies.

CHAPTER 5 – FINAL DISCUSSION AND FUTURE DIRECTIONS

5.1 - Final Discussion

Throughout a TB infection, the human host shows progressive pulmonary pathology characterized by lesion heterogeneity, which can present itself from inflammatory up to destructive cavitary lesion types. A single patient can concurrently develop multiple different lesions with diverse microenvironments such as varying levels of pH, energy and carbon availability, vascularization, and immune cell interactions. The best method for research purposes to recapitulate the complexities of human TB disease are animal models such as Guinea Pigs, non-human primates, and certain mouse species. Animal modeling is essential in the development of new and more effective therapies but to maximize their potential, new tools are needed to quantify the disease presentations and bacterial locations. Artificial intelligence and computer vision represent a marked improvement in current manual measurement methodologies in terms of reproducibility and efficiency. These techniques also give additional insights and more detailed data that were not previously available and are helping to expand our knowledge on therapy efficacy and host-pathogen interactions. The results presented in this dissertation are the first steps in the development of digital tools to quantify the lesion involvement and bacterial locations in the C3HeB/FeJ TB animal model. This includes novel software tools using new code for the quantification of the different pulmonary lesion tissues in C3HeB/FeJ mice using Lesion Image Recognition Analysis (LIRA) and the quantification of bacterial metrics based upon the lesion compartment in which the population is located using Mycobacterial Image Analysis (MIA).

Aim 1 (Chapter 2) details the process for writing and developing *de novo* software utilizing deep convolutional neural networks for the identification and quantification of pulmonary lesions using digital microscopic images from the lungs of *M. tuberculosis* infected C3HeB/FeJ mice. Using the Lenaerts laboratories catalog of histopathology image scans from previous C3HeB/FeJ mouse studies, three neural networks were successfully trained to classify and quantify

pulmonary lesion tissues from mice between 6 to 10 weeks post aerosol infection. The first neural network identifies the structural features that are found within the caseous necrotic Type I lesion, the second neural network identifies subcomponents of the Type I lesion, and the third neural network classifies neutrophilic Type II and cellular Type III lesions. Multiple advantages were discovered by using the LIRA software in identifying and quantifying pathology. The first advantage using LIRA was a remarkable reduction in analysis time by 82% when compared to the manual histopathology method used previously. The time-saving aspect is significant since analysis in the past used to take several weeks to months, depending on the size of the study. Analysis that takes an excessive amount of time to complete can lead the researcher to experience user fatigue. User fatigue increases the amount of bias or mistakes thus causing unintentional skewing or incorrect reporting of the data. While unintentional, it may lead to mistaken results or assumptions about the data that may impact final conclusions. The second advantage of LIRA over manual histopathology analysis became apparent by showing greater reproducibility between users. For this purpose, a single study was analyzed by different pathologists either by manual histopathology analysis or by using LIRA. Manual histopathology analysis showed a percent agreement among users of 87%, but with the assistance of LIRA, this increased to 94%. This is an important finding as reproducibility and unbiased analysis will enable us to compare multiple studies, time points, and analysis done by different individuals or laboratories. This will increase the accuracy of the data, which otherwise might be lost when higher variability is present. The third advantage of LIRA is the precision on a microscopic level that LIRA classifies the various lesions types. With the use of LIRA, there was a more accurate demarcation of lesion margins, as well as the detection of small isolated regions that may not be identifiable at a macroscopic level used by manual analysis. Both the pathologist and LIRA were able to classify the various lesion types correctly, but the added precision that LIRA provides increases the confidence of the final generated quantitative results. The purpose of LIRA was never to replace a human classifier but to work in tandem with each other. LIRA's strength is its

performance for repetitive analysis, but LIRA performs poorly when given new information that was not included in its initial training set. Humans struggle to perform repetitive tasks consistently but are highly skilled in fluid intelligence problems that require new information or solutions. When paired together both LIRA and the human classifier complement each other to provide the most optimal and accurate analysis.

Aim 2 (Chapter 3) detailed the use and performance of the LIRA software on histopathology images from a drug-treated test case mouse study. For this purpose, we analyzed the results of a previously conducted drug efficacy C3HeB/FeJ trial in comparison to the standard manual histopathology analysis performed by a certified pathologist. LIRA had originally been trained using tissue samples from untreated animals 6 to 10 weeks post initial aerosol infection. Machine learning algorithms can be very sensitive to changes in the input data, such as if the new input data differs significantly from the training data set that was used to generate the model. Because the TB C3HeB/FeJ animal model is often used in drug efficacy trials, it was important to determine if drug treatment would impact LIRA's ability to classify pathology accurately. A previously conducted drug study using isoniazid (INH) and losartan (LK) was chosen to perform the validation of the LIRA software. It was especially advantageous because the study had been previously analyzed using standard methodology, allowing for a reduction in analysis bias when comparing old and new methodologies, and the study also included both antibiotic and immunotherapy cohorts to allow for more robust testing. In this chapter, for the validation of LIRA we had three goals; 1) we compared the results from the scores generated by a pathologist using standard methodology, 2) we analyzed the LIRA results in terms of the outcome of the therapies on pulmonary pathology, and 3) we determined if using a more targeted analysis approach such as LIRA would change or add to the initial interpretation of these results. Using Krippendorff's alpha, a statistical method that is able to measure the agreement among different classifiers, we measured the agreement of classification scores of non-pathologist researchers to that of a board-certified pathologist. The alpha score can range from 0 to 1 with 0 being "perfect disagreement",

1 being “perfect agreement”, 0.99 to 0.80 being “good agreement”, and 0.79 to 0.667 being “tentative” (lowest acceptable limit). The Krippendorff’s alpha value for each lesion type were all above the lowest acceptable values for agreement, and the overall score was 0.879 which is considered “good agreement”. This means that the scores generated by the non-pathologists were in agreement with the scores generated by the pathologist. Initial histopathology analysis of the drug efficacy trial had come to the conclusion that there were no significant changes in pathology between the different treatment groups; however, with LIRA substantial changes were observed. The first result was the striking increase in Type I lesion involvement for the INH cohort and the decrease in involvement for all the lesion types in the LK cohort. These LIRA findings were somewhat unexpected when they are compared to earlier CFU results and mortality data from similar mouse studies. Specifically, both treatment cohorts had comparable levels of bacterial reduction, but LK had the highest mortality (50%), while the INH cohort had the lowest mortality (0%). Taking the study results together, the LK cohort exhibited minimal lesion development, a reduction in the bacterial burden comparable to INH, and yet had a incredibly high mortality. Even though more investigation into the results is required, this implies that a reduction in both pathology and bacterial burdens may not always be a reliable indicator of drug efficacy.

In Aim 3 (Chapter 4), we generated software for the enumeration of fluorescently stained bacteria in pulmonary tissues of C3HeB/FeJ mice infected with *M. tuberculosis*. Tuberculosis disease is the result of the interaction of the host immune system and the Mtb pathogen. LIRA is a quantitative tool that measures the response of the host; however, another quantitative tool was needed to measure the bacillary response, which is why MIA was developed. MIA relies on more traditional computer vision techniques to be applied to specific staining procedures to quantify bacterial metrics. The analysis creates a mask based upon the selected fluorophore stain color and acquires the x,y coordinates of the objects of interest that are present in the image. The edge coordinates of the objects within a selected region of interest (ROI) can be used to calculate the

object area (μm^2), the object density within the ROI (defined bacterial areas / μm^2), and the percentage of the image that is composed of the objects of interest (bacterial % of the image). The software was initially verified using flow cytometry beads as a control to verify the correct measurement of an object area and number of objects present. The mean of the calculated diameter of a flow cytometry beads was 6.66µm, which was well within the standard deviation of the mean provided by the manufacturer of 7.08 ± 0.478µm. The performance of the MIA software with the control beads showed a 99.04% accuracy in identifying the desired individual objects of interest that were present in a given image. We next tested the MIA software on an actual efficacy study using histology slides collected from the C3HeB/FeJ mice. The previous analysis, like in Aim 2, had determined that there were no significant differences between the visualized stained bacterial numbers. The lung sections from the same mouse study were used as described in Aim 2 (Chapter 3) to measure the bacterial metrics within each lesion compartment. What these preliminary results showed was first of all that the MIA software was easy to use and highly time-efficient. A few interesting preliminary observations showed some apparent changes in bacterial aggregation and density-dependent on the treatment groups. In both the Control and INH treatment groups, a general trend of increased Mtb aggregation was observed for most tissue compartments, especially in the cores of the necrotic Type I lesion tissue. In contrast, the LK group had a demonstrated decrease in the aggregation of bacteria within the tissue compartments, with an increase in bacterial density. Lastly, for the bacterial percentage (or overall visual burden of bacteria within an image) the untreated Control cohort remained the same, the INH treatment group had a reduction in most areas of the lung except for the lesion core, and the LK treatment cohort saw a general increase in all tissue compartments except in the Type I lesions.

These initial results are preliminary at best and more studies will have to be analyzed; however, this test case using the losartan study in C3HeB/FeJ mice proved that the MIA software was easy to use and useful in determining differences between the various treatment groups

which could not be distinguished using previous manual methods. For future work, one could hypothesize that environmental stresses such as the immune system (Control) or anti-mycobacterial therapy (INH) may promote the formation of mycobacterial aggregates and islands within the caseous necrotic cores in lungs. The formation of these aggregates may then negatively impact either the host or chemotherapy ability to eliminate the pathogen. Future work is required to show these observations are reproducible and to prove this hypothesis further.

5.2 – Future Directions

The research presented in this dissertation is only the initial step for a larger image analysis pipeline, the purpose of which is to increase the efficiency of data collection during an Mtb infection. This section will serve as a road map for my successors on future directions and further refinements of this pipeline.

Chapter 2 demonstrated that the LIRA software saves significant amounts of time and improves considerably on reproducibility on histopathology analysis of digital images while using this initial version of the LIRA software. However, several refinements could be implemented to improve the current workflow. First, LIRA's pipeline analysis is currently a mixture of both command line and the general user interface (GUI). Because the majority of users will be non-computer scientists, a large proportion of functions within the script will need to be called using graphical user interface (GUI) widgets. GUI widgets are the buttons you press in a piece of software that perform a specific task. Second, LIRA can only currently handle image file types that are non-proprietary (e.g., .jpg, .png, .tif). For this purpose, we wrote a script to convert the digital image scan files into a format that can be recognized by LIRA and the script still needs to be integrated into LIRA's current workflow. Third, several functions (e.g., iterative looping) in LIRA could be further optimized using a faster programming language such as C or Cython. The new library of optimized code will need to be created to be callable in Python, and LIRA's code will need to be updated. Lastly, it would be beneficial to implement an object detector to reduce the

amount of processing on empty slide window patches. This can be accomplished by performing the following stepwise approach: by uploading the original image, followed by creating a reduced image size copy (e.g. 10% of the original size), then determining the coordinates of the pulmonary tissue compartments in the image by creating a simple object detector, and only running CNN1, CNN2, and CNN3 on these specific window patches that are located within these coordinates. This will greatly reduce the large datasets that need to be processed and accelerate overall analysis time.

TB is a complex disease with heterogeneity in lesion types present in the lungs. Lesion classification requires the identification of both the pathology features at a macroscopic level and cell type composition at a microscopic level. Therefore the LIRA software was specially designed to classify lesions using both macroscopic and microscopic classifications to address the complicated pulmonary pathology. While using the LIRA software on several C3heB/FeJ mouse studies over the last year, a realization occurred that two additional classifications could improve the current model. The first additional classification would include the detection of cavity formation in the Type I lesions at a macroscopic level. The second additional classification relates to the Type II lesions which often have pockets of DNA along the margin. These pockets of DNA resemble neutrophil extracellular traps (NETs), which seem to be consistently mislabeled by LIRA as Type III lesions. Creating a new classification, called NETs, and retraining the network should improve the overall model accuracy. The model architecture that the current LIRA software uses to make predictions on C3HeB/FeJ mouse models could, in the future, also be applied to additional TB animal models and disease types not currently integrated. New iterations of LIRA could include digital pathology analysis on the Guinea Pig, the rabbit model, on NHP, zebrafish models, and increased optimization for the BALB/c mice. These iterations will require the creation of new digital image data sets to create additional classifiers specific for these models.

Chapter 3 demonstrated that LIRA is effective in correctly classifying data generated from a test case drug study using C3HeB/FeJ mice from 6 to 10 weeks post aerosol infection. At the

time of LIRA's creation, the only available data for the training dataset was a larger number of tissue samples all collected within the six to ten week timeframe. Current studies are generating new data that extends past that timeframe. LIRA's current iteration has noticeably poorer performance on more advanced pulmonary pathology past 12 weeks post aerosol infection, and therefore the model training dataset (CNN2 and CNN3) would benefit from updates with the current 12+ week data to improve accuracy. Chapter 4 demonstrated that the MIA software could effectively quantify bacterial singlets and aggregates, and separate bacterial populations based upon their lesion compartment location. Preliminary data showed that *in vivo* bacteria under stress might form larger aggregates or mycobacterial islands in cores of caseous lesions. To further investigate this hypothesis, we propose to implement both MIA and LIRA to more broadly study the effect on bacteria in more mouse efficacy studies, thereby using a wider range of current TB therapies, enhancing host immunity, or control agents designed to effect bacterial replication. An effective study will include therapies that target the bacteria and host therapies that either enhance or inhibit the host immune system. Treatments should include first-line drugs such as rifampin, ethambutol, and pyrazinamide as well as newly approved drugs for TB such as BDQ and linezolid. As a positive control and to assess the effect of the host on bacterial numbers and aggregation, mice could be vaccinated before initial infection to elicit a stronger immune response. A good candidate would be an intramuscular DNA vaccine containing *Hsp65* and IL-12 genes that provided a 40% improvement of survival in Mtb infected primates [251]. As a negative control, additional therapies that change the integrity of granuloma formation can be used. This could include Enbrel [252], [253] which is a TNFα inhibitor, or Bevacizumab [254], [255] which neutralizes vascular endothelial growth factor. Bevacizumab would be preferred, in the context of studying environmental stresses, since previous studies have shown that adjunct therapy of Bevacisumab increased the efficacy of first-line therapies by restoring the vascular architecture and consequently reducing hypoxia [21], [254]. Based on our initial results, we suspect that we will not observe the formation of the mycobacterial islands in the immune impaired cohorts. In

contrast, the cohorts that are receiving some form of therapy (e.g., antibiotics or vaccine) will develop moderate amounts of mycobacterial island formation. The Lenaerts laboratory is also investigating an experimental proprietary compound that appears to promote cavity formation within caseous necrotic lesions. The use of LIRA, in combination with MIA, will greatly enhance these studies by quantitatively monitoring both lesion development and the corresponding bacterial phenotype responses from initial infection to terminal cavity formation.

The analysis of MIA might become even more impactful with the utilization of additional staining approaches that target other bacterial phenotypes or host cells. The MIA software in this dissertation was developed using the SYBR Gold staining method developed in our laboratory. However, the software was designed to gate for both Auromine Rhodamine (AR) and Ziehl Neelsen acid-fast staining. With minimal effort, the use of additional fluorophore channels can be integrated into the software by creating additional min and max HSV values specific to the wavelength color of the objects of interest. By using a wider diversity in fluorescent probes such as for detection of ribonucleic acid (RNA) by FISH, antibodies for use in immunohistochemistry, and other dyes will further the bacterial as well as host cell phenotypes and compositions. Of particular interest to our laboratory would be the further evaluation of the heterogeneity of bacterial populations within each lung compartment. We especially are interested in determining what proportion of bacilli are viable, metabolically active, are actively replicating, or show an altered cell wall. MIA can quantify an entire visual bacterial population within each lesion compartment, and in combination with the plethora of probes and staining approaches available more data and information can be generated. Of particular interest to me would be to study the number of Mtb phenotypes existing within each lesion compartment, and whether there is a predisposition for specific phenotypes for different environments. The end goal would be to be able to determine the effect of treatments on various bacterial phenotypes and in the different lung compartments.

LIRA and MIA were designed as two separate software components which are intended in the future to be incorporated together as a single software. There are several steps that need

to happen before both software programs can be integrated to work in tandem. The first step will be the improvement of the current staining methodology in tissue samples. Because MIA relies on masking, the isolation of specific pixel values, the fluorescent signal needs to be pristine without background or false positive signals. Autofluorescence occurs with the use of aldehyde fixation (formaldehyde or glutaraldehyde), staining with H&E, and in tissues in the presence of high numbers of red blood cells. Attempts have been to reduce autofluorescence using counterstains such as potassium permanganate and Sudan Black with only limited success. What has proven successful was the implementation of a new perfusion technique reducing red blood cells that are naturally autofluorescent. The second area of improvement would be the image capturing devices. Current available microscopes on CSU campus and their software are not well suited for high image resolution throughput of large tissue datasets. This equipment shortcoming is probably the most limiting step in utilizing a high throughput analysis pipeline. Optimally, one would need a digital slide scanner to capture bright-field images of the H&E stained tissue, as well as taking corresponding scans of multiple fluorescent channels of a single area at the same time. At the moment, with the current equipment available, this process would be highly time-consuming and is the reason that LIRA and MIA are still two separate entities to perform this type of analysis. The future vision is to overcome these obstacles and have a seamless analysis process that quantifies every relevant piece of data so that no important event is missed. Such a system is not feasible without utilizing the strengths of both AI and human analysis and understanding how they can be synergized together in a biological context. This approach will broaden our understanding of the interplay between the host and Mtb pathogen in the context of drug efficacy testing. With the continued optimization of analysis with new computational tools, new therapies can be developed to help aid in the elimination of tuberculosis.

REFERENCES

[1] "WHO | Global tuberculosis report 2018," *WHO*. [Online]. Available: http://www.who.int/tb/publications/global_report/en/. [Accessed: 20-May-2019].

[2] "WHO | 3. Tuberculosis control," *WHO*. [Online]. Available: https://www.who.int/trade/distance_learning/gpgh/gpgh3/en/index7.html. [Accessed: 17-Oct-2019].

[3] CDCMMWR, "World TB Day — March 24, 2019," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 68, 2019.

[4] "Tuberculosis (TB)," *World Health Organization*. [Online]. Available: http://www.who.int/news-room/fact-sheets/detail/tuberculosis. [Accessed: 28-Jun-2018].

[5] "Treatment for TB Disease | Treatment | TB | CDC," 09-May-2019. [Online]. Available: https://www.cdc.gov/tb/topic/treatment/tbdisease.htm. [Accessed: 21-May-2019].

[6] G. Elzinga, M. C. Raviglione, and D. Maher, "Scale up: meeting targets in global tuberculosis control," *The Lancet*, vol. 363, no. 9411, pp. 814–819, Mar. 2004.

[7] K. E. Dooley and R. E. Chaisson, "Tuberculosis and diabetes mellitus: convergence of two epidemics," *Lancet Infect. Dis.*, vol. 9, no. 12, pp. 737–746, Dec. 2009.

[8] S.-H. Shieh, J. C. Probst, F.-C. Sung, W.-C. Tsai, Y.-S. Li, and C.-Y. Chen, "Decreased survival among lung cancer patients with co-morbid tuberculosis and diabetes," *BMC Cancer*, vol. 12, p. 174, May 2012.

[9] "WHO | Surveillance of drug-resistant TB," *WHO*. [Online]. Available: http://www.who.int/tb/areas-of-work/monitoring-evaluation/mdr-tb-surveillance/en/. [Accessed: 21-May-2019].

[10] O. of the Commissioner, "FDA approves new drug for treatment-resistant forms of tuberculosis that affects the lungs," *FDA*, 11-Sep-2019. [Online]. Available: http://www.fda.gov/news-events/press-announcements/fda-approves-new-drug-treatment-resistant-forms-tuberculosis-affects-lungs. [Accessed: 15-Nov-2019].

[11] U. Manjunatha, H. I. Boshoff, and C. E. Barry, "The mechanism of action of PA-824," *Commun. Integr. Biol.*, vol. 2, no. 3, pp. 215–218, 2009.

[12] A. M. Upton *et al.*, "In Vitro and In Vivo Activities of the Nitroimidazole TBA-354 against Mycobacterium tuberculosis," *Antimicrob. Agents Chemother.*, vol. 59, no. 1, pp. 136–144, Jan. 2015.

[13] O. M. P. Jolobe, "Disseminated tuberculosis," *QJM Int. J. Med.*, vol. 110, no. 5, pp. 331–331, May 2017.

[14] R. L. Hunter, "PATHOLOGY of POST PRIMARY TUBERCULOSIS of the LUNG: AN ILLUSTRATED CRITICAL REVIEW," *Tuberc. Edinb. Scotl.*, vol. 91, no. 6, pp. 497–509, Nov. 2011.

[15] R. J. Basaraba and R. L. Hunter, *Pathology of Tuberculosis: How the Pathology of Human Tuberculosis Informs and Directs Animal Models*, vol. 5. 2017.

[16] W. C. Tobie, "The Tubercle Bacillus, in the Pulmonary Lesion of Man. Histobacteriology and Its Bearing on the Therapy of Pulmonary Tuberculosis. George Canetti," *Q. Rev. Biol.*, vol. 32, no. 2, pp. 201–201, Jun. 1957.

[17] G. Kaplan *et al.*, "Mycobacterium tuberculosis growth at the cavity surface: a microenvironment with failed immunity," *Infect. Immun.*, vol. 71, no. 12, pp. 7099–7108, Dec. 2003.

[18] A. Lenaerts, C. E. Barry, and V. Dartois, "Heterogeneity in tuberculosis pathology, microenvironments and therapeutic responses," *Immunol. Rev.*, vol. 264, no. 1, pp. 288–307, Mar. 2015.

[19] N. Dhar, J. McKinney, and G. Manina, "Phenotypic Heterogeneity in Mycobacterium tuberculosis," *Microbiol. Spectr.*, vol. 4, no. 6, Nov. 2016.

[20] "Comprehensive analysis of methods used for the evaluation of compounds against Mycobacterium tuberculosis | Elsevier Enhanced Reader." [Online]. Available: https://reader.elsevier.com/reader/sd/pii/S1472979212001400?token=449F410384EEC3560E3B8291E5A3EB9859582A3666B34059575EBF118B982C6FB4E01C9D17DEC645A21FE158D19C3F88. [Accessed: 05-Jun-2019].
[21] A. Kolloli and S. Subbian, "Host-Directed Therapeutic Strategies for Tuberculosis," *Front. Med.*, vol. 4, Oct. 2017.
[22] S. M. Irwin *et al.*, "Bedaquiline and Pyrazinamide Treatment Responses Are Affected by Pulmonary Lesion Heterogeneity in Mycobacterium tuberculosis Infected C3HeB/FeJ Mice," *ACS Infect. Dis.*, vol. 2, no. 4, pp. 251–267, Apr. 2016.
[23] N. Strydom *et al.*, "Tuberculosis drugs' distribution and emergence of resistance in patient's lung lesions: A mechanistic model and tool for regimen and dose optimization," *PLOS Med.*, vol. 16, no. 4, p. e1002773, Apr. 2019.
[24] A. Williams and I. M. Orme, "Animal Models of Tuberculosis: An Overview," *Microbiol. Spectr.*, vol. 4, no. 4, 2016.
[25] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.
[26] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, vol. 8150, K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 411–418.
[27] G. Litjens *et al.*, "A Survey on Deep Learning in Medical Image Analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[28] J.-G. Lee *et al.*, "Deep Learning in Medical Imaging: General Overview," *Korean J Radiol*, vol. 18, no. 4, pp. 570–584, Aug. 2017.
[29] R. E. Butler *et al.*, "The Balance of Apoptotic and Necrotic Cell Death in Mycobacterium tuberculosis Infected Macrophages Is Not Dependent on Bacterial Virulence," *PLOS ONE*, vol. 7, no. 10, p. e47573, Oct. 2012.
[30] L. Huang, E. V. Nazarova, S. Tan, Y. Liu, and D. G. Russell, "Growth of Mycobacterium tuberculosis in vivo segregates with host macrophage metabolism and ontogeny," *J. Exp. Med.*, vol. 215, no. 4, pp. 1135–1152, Apr. 2018.
[31] Y. C. Manabe *et al.*, "Different strains of Mycobacterium tuberculosis cause various spectrums of disease in the rabbit model of tuberculosis," *Infect. Immun.*, vol. 71, no. 10, pp. 6004–6011, Oct. 2003.
[32] K. Sakamoto, "The Pathology of Mycobacterium tuberculosis Infection," *Vet. Pathol.*, vol. 49, no. 3, pp. 423–439, Jan. 2012.
[33] I. G. Sia and M. L. Wieland, "Current concepts in the management of tuberculosis," *Mayo Clin. Proc.*, vol. 86, no. 4, pp. 348–361, Apr. 2011.
[34] G. W. Comstock, "Epidemiology of Tuberculosis," *Am. Rev. Respir. Dis.*, vol. 125, no. 3P2, pp. 8–15, Mar. 1982.
[35] M. A. Behr, P. H. Edelstein, and L. Ramakrishnan, "Revisiting the timetable of tuberculosis," *BMJ*, vol. 362, 2018.
[36] V. Cornil and L. Ranvier, "Pathological histology of the respiratory apparatus," in *Manual of pathological histology*, Philadelphia: Henry C Lea, 1880, pp. 407–445.
[37] J. Andreu, J. Cáceres, E. Pallisa, and M. Martinez-Rodriguez, "Radiological manifestations of pulmonary tuberculosis," *Pulm. Infect.*, vol. 51, no. 2, pp. 139–149, Aug. 2004.

[38] A. A. Ordonez *et al.*, "Radioiodinated DPA-713 Imaging Correlates with Bactericidal Activity of Tuberculosis Treatments in Mice," *Antimicrob. Agents Chemother.*, vol. 59, no. 1, pp. 642–649, Jan. 2015.
[39] R. L. Hunter, J. K. Actor, S.-A. Hwang, V. Karev, and C. Jagannath, "Pathogenesis of Post Primary Tuberculosis: Immunity and Hypersensitivity in the Development of Cavities," *Ann. Clin. Lab. Sci.*, vol. 44, no. 4, pp. 365–387, Sep. 2014.
[40] R. L. Hunter, "Tuberculosis as a three-act play: A new paradigm for the pathogenesis of pulmonary tuberculosis," *Tuberc. Edinb. Scotl.*, vol. 97, pp. 8–17, Mar. 2016.
[41] G. Canetti, "Biology of the mycobacterioses. Pathogenesis of tuberculosis in man," *Ann. N. Y. Acad. Sci.*, vol. 154, no. 1, pp. 13–18, Sep. 1968.
[42] K. Chatterjee, B. Colaco, C. Colaco, M. Hellman, and N. Meena, "Rasmussen's aneurysm: A forgotten scourge," *Respir. Med. Case Rep.*, vol. 16, pp. 74–76, Jan. 2015.
[43] A. O. Ankrah, T. S. van der Werf, E. F. J. de Vries, R. A. J. O. Dierckx, M. M. Sathekge, and A. W. J. M. Glaudemans, "PET/CT imaging of Mycobacterium tuberculosis infection," *Clin. Transl. Imaging*, vol. 4, pp. 131–144, 2016.
[44] D. C. Perlman *et al.*, "Variation of Chest Radiographic Patterns in Pulmonary Tuberculosis by Degree of Human Immunodeficiency Virus-Related Immunosuppression," *Clin. Infect. Dis.*, vol. 25, no. 2, pp. 242–246, Aug. 1997.
[45] R. Y. Chen *et al.*, "PET/CT imaging correlates with treatment outcome in patients with multidrug-resistant tuberculosis," *Sci. Transl. Med.*, vol. 6, no. 265, pp. 265ra166-265ra166, Dec. 2014.
[46] M. T. Coleman *et al.*, "PET/CT imaging reveals a therapeutic response to oxazolidinones in macaques and humans with tuberculosis," *Sci. Transl. Med.*, vol. 6, no. 265, pp. 265ra167-265ra167, Dec. 2014.
[47] "Historical Perspectives Centennial: Koch's Discovery of the Tubercle Bacillus." [Online]. Available: https://www.cdc.gov/mmwr/preview/mmwrhtml/00000222.htm. [Accessed: 16-Nov-2019].
[48] A. K. Singh and U. D. Gupta, "Animal models of tuberculosis: Lesson learnt," *Indian J. Med. Res.*, vol. 147, no. 5, pp. 456–463, May 2018.
[49] K. Patel, S. S. Jhamb, and P. P. Singh, "Models of Latent Tuberculosis: Their Salient Features, Limitations, and Development," *J. Lab. Physicians*, vol. 3, no. 2, pp. 75–79, 2011.
[50] P. L. Lin *et al.*, "Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing," *Nat. Med.*, vol. 20, no. 1, pp. 75–79, Jan. 2014.
[51] P. L. Lin *et al.*, "Early events in Mycobacterium tuberculosis infection in cynomolgus macaques," *Infect. Immun.*, vol. 74, no. 7, pp. 3790–3803, Jul. 2006.
[52] E. L. Rayner *et al.*, "Early lesions following aerosol infection of rhesus macaques (Macaca mulatta) with Mycobacterium tuberculosis strain H37RV," *J. Comp. Pathol.*, vol. 149, no. 4, pp. 475–485, Nov. 2013.
[53] L. E. Via *et al.*, "Differential Virulence and Disease Progression following Mycobacterium tuberculosis Complex Infection of the Common Marmoset (Callithrix jacchus)," *Infect. Immun.*, vol. 81, no. 8, pp. 2909–2919, Aug. 2013.
[54] G. Kaplan and L. Tsenova, "Pulmonary Tuberculosis in the Rabbit," in *A color atlas of comparative pathology of pulmonary tuberculosis*, 2010, pp. 107–130.
[55] M. E. Visser *et al.*, "Baseline Predictors of Sputum Culture Conversion in Pulmonary Tuberculosis: Importance of Cavities, Smoking, Time to Detection and W-Beijing Genotype," *PLoS ONE*, vol. 7, no. 1, Jan. 2012.
[56] S. Clark, Y. Hall, and A. Williams, "Animal models of tuberculosis: Guinea pigs," *Cold Spring Harb. Perspect. Med.*, vol. 5, no. 5, pp. a018572–a018572, Dec. 2014.
[57] C. C. Dascher *et al.*, "Immunization with a mycobacterial lipid vaccine improves pulmonary pathology in the guinea pig model of tuberculosis," *Int. Immunol.*, vol. 15, no. 8, pp. 915–925, Aug. 2003.

[58] S. S. Kashino, D. R. Napolitano, Z. Skobe, and A. Campos-Neto, “Guinea pig model of Mycobacterium tuberculosis latent/dormant infection,” *Microbes Infect.*, vol. 10, no. 14, pp. 1469–1476, Nov. 2008.
[59] T. F. Vandamme, “Use of rodents as models of human diseases,” *J. Pharm. Bioallied Sci.*, vol. 6, no. 1, pp. 2–9, Jan. 2014.
[60] S. Major, J. Turner, and G. Beamer, “Tuberculosis in CBA/J Mice,” *Vet. Pathol.*, vol. 50, no. 6, pp. 1016–1021, Mar. 2013.
[61] G. L. Beamer *et al.*, “Interleukin-10 promotes Mycobacterium tuberculosis disease progression in CBA/J mice,” *J. Immunol. Baltim. Md 1950*, vol. 181, no. 8, pp. 5545–5550, Oct. 2008.
[62] T. A. Skvortsov, D. V. Ignatov, K. B. Majorov, A. S. Apt, and T. L. Azhikina, “Mycobacterium tuberculosis Transcriptome Profiling in Mice with Genetically Different Susceptibility to Tuberculosis,” *Acta Naturae*, vol. 5, no. 2, pp. 62–69, 2013.
[63] T. K. Kondratieva *et al.*, “A new model for chronic and reactivation tuberculosis: Infection with genetically attenuated Mycobacterium tuberculosis in mice with polar susceptibility,” *Tuberculosis*, vol. 113, pp. 130–138, Dec. 2018.
[64] F. Sánchez *et al.*, “Multigenic Control of Disease Severity after Virulent Mycobacterium tuberculosis Infection in Mice,” *Infect. Immun.*, vol. 71, no. 1, p. 126, Jan. 2003.
[65] L. Heitmann *et al.*, “The IL-13/IL-4Rα axis is involved in tuberculosis-associated pathology,” *J. Pathol.*, vol. 234, no. 3, pp. 338–350, Nov. 2014.
[66] C. L. Emson, S. E. Bell, A. Jones, W. Wisden, and A. N. J. McKenzie, “Interleukin (IL)-4–independent Induction of Immunoglobulin (Ig)E, and Perturbation of T Cell Development in Transgenic Mice Expressing IL-13,” *J. Exp. Med.*, vol. 188, no. 2, pp. 399–404, Jul. 1998.
[67] Y.-J. Jung, R. LaCourse, L. Ryan, and R. J. North, “Evidence inconsistent with a negative influence of T helper 2 cells on protection afforded by a dominant T helper 1 response against Mycobacterium tuberculosis lung infection in mice,” *Infect. Immun.*, vol. 70, no. 11, pp. 6436–6443, Nov. 2002.
[68] I. Kramnik, W. F. Dietrich, P. Demant, and B. R. Bloom, “Genetic control of resistance to experimental infection with virulent Mycobacterium tuberculosis,” *Proc. Natl. Acad. Sci.*, vol. 97, no. 15, pp. 8560–8565, Jul. 2000.
[69] H. Pan *et al.*, “Ipr1 gene mediates innate immunity to tuberculosis,” *Nature*, vol. 434, no. 7034, pp. 767–772, Apr. 2005.
[70] B.-S. Yan *et al.*, “Progression of Pulmonary Tuberculosis and Efficiency of Bacillus Calmette-Guérin Vaccination Are Genetically Controlled via a Common sst1-Mediated Mechanism of Innate Immunity,” *J. Immunol.*, vol. 179, no. 10, p. 6919, Nov. 2007.
[71] E. R. Driver *et al.*, “Evaluation of a Mouse Model of Necrotic Granuloma Formation Using C3HeB/FeJ Mice for Testing of Drugs against Mycobacterium tuberculosis,” *Antimicrob. Agents Chemother.*, vol. 56, no. 6, pp. 3181–3195, Jun. 2012.
[72] S. D. Chakravarty, J. Xu, B. Lu, C. Gerard, J. Flynn, and J. Chan, “The Chemokine Receptor CXCR3 Attenuates the Control of Chronic Mycobacterium tuberculosis Infection in BALB/c Mice,” *J. Immunol.*, vol. 178, no. 3, pp. 1723–1735, Feb. 2007.
[73] G. J. Ryan *et al.*, “Multiple M. tuberculosis Phenotypes in Mouse and Guinea Pig Lung Tissue Revealed by a Dual-Staining Approach,” *PLOS ONE*, vol. 5, no. 6, p. e11108, Jun. 2010.
[74] S. M. Irwin *et al.*, “Presence of multiple lesion types with vastly different microenvironments in C3HeB/FeJ mice following aerosol infection with Mycobacterium tuberculosis,” *Dis. Model. Mech.*, vol. 8, no. 6, pp. 591–602, Jun. 2015.
[75] J.-P. Lanoix *et al.*, “Selective Inactivity of Pyrazinamide against Tuberculosis in C3HeB/FeJ Mice Is Best Explained by Neutral pH of Caseum,” *Antimicrob. Agents Chemother.*, vol. 60, no. 2, pp. 735–743, Feb. 2016.
[76] A. A. Ordonez *et al.*, “Mouse model of pulmonary cavitary tuberculosis and expression of matrix metalloproteinase-9,” *Dis. Model. Mech.*, vol. 9, no. 7, pp. 779–788, 01 2016.

[77] S. S. Cross, “Grading and scoring in histopathology,” *Histopathology*, vol. 33, no. 2, pp. 99–106, 1998.
[78] J. W. Crissman *et al.*, “Best Practices Guideline: Toxicologic Histopathology,” *Toxicol. Pathol.*, vol. 32, no. 1, pp. 126–131, Jan. 2004.
[79] K. N. Gibson-Corley, A. K. Olivier, and D. K. Meyerholz, “Principles for valid histopathologic scoring in research,” *Vet. Pathol.*, vol. 50, no. 6, Nov. 2013.
[80] G. T. Sica, “Bias in Research Studies,” *Radiology*, vol. 238, no. 3, pp. 780–789, Mar. 2006.
[81] “Qualitative and Quantitative Analysis of Nonneoplastic Lesions in Toxicology Studies - Cynthia Shackelford, Gerald Long, Jeffrey Wolf, Carlin Okerberg, Ronald Herbert, 2002.” [Online]. Available: https://journals.sagepub.com/doi/abs/10.1080/01926230252824761. [Accessed: 30-Jul-2019].
[82] B. Thoolen *et al.*, “Proliferative and Nonproliferative Lesions of the Rat and Mouse Hepatobiliary System,” *Toxicol. Pathol.*, vol. 38, no. 7_suppl, pp. 5S-81S, Dec. 2010.
[83] C. G. Begley and L. M. Ellis, “Drug development: Raise standards for preclinical cancer research,” *Nature*, vol. 483, pp. 531–533, Mar. 2012.
[84] “Checklists work to improve science,” *Nature*, vol. 556, p. 273, Apr. 2018.
[85] P. B. Stark, “Before reproducibility must come preproducibility,” *Nature*, vol. 557, p. 613, May 2018.
[86] J. L. Hecht, J. Kotsopoulos, M. A. Gates, S. E. Hankinson, and S. S. Tworoger, “Validation of Tissue Microarray Technology in Ovarian Cancer: Results from the Nurses’ Health Study,” *Cancer Epidemiol. Prev. Biomark.*, vol. 17, no. 11, pp. 3043–3050, Nov. 2008.
[87] P. L. Lin *et al.*, “Quantitative Comparison of Active and Latent Tuberculosis in the Cynomolgus Macaque Model,” *Infect. Immun.*, vol. 77, no. 10, pp. 4631–4642, Oct. 2009.
[88] M. S. Jassal, G. G. Nedeltchev, J. Osborne, and W. R. Bishai, “A modified scoring system to describe gross pathology in the rabbit model of tuberculosis,” *BMC Microbiol.*, vol. 11, no. 1, p. 49, Mar. 2011.
[89] J. DORMANS *et al.*, “Correlation of virulence, lung pathology, bacterial load and delayed type hypersensitivity responses after infection with different Mycobacterium tuberculosis genotypes in a BALB/c mouse model,” *Clin. Exp. Immunol.*, vol. 137, no. 3, pp. 460–468, Sep. 2004.
[90] P. W. Snook, “Comparison between Digital and Manual Interpretation of High Altitude Aerial,” p. 4.
[91] J. McFadyen, M. Mermillod, J. B. Mattingley, V. Halász, and M. I. Garrido, “A Rapid Subcortical Amygdala Route for Faces Irrespective of Spatial Frequency and Emotion,” *J. Neurosci.*, vol. 37, no. 14, p. 3864, Apr. 2017.
[92] B. V. Ehinger, K. Häusser, J. P. Ossandón, and P. König, “Humans treat unreliable filled-in percepts as more real than veridical ones,” *eLife*, vol. 6, p. e21761, May 2017.
[93] “Visual perception,” 11-Oct-2017. [Online]. Available: https://qbi.uq.edu.au/brain/brain-functions/visual-perception. [Accessed: 02-Aug-2019].
[94] E. Samei and E. A. Krupinski, *The Handbook of Medical Image Perception and Techniques*. Cambridge University Press, 2018.
[95] D. Memmert, “The effects of eye movements, age, and expertise on inattentional blindness,” *Conscious. Cogn.*, vol. 15, no. 3, pp. 620–627, Sep. 2006.
[96] A. K. Barbey, “Network Neuroscience Theory of Human Intelligence,” *Trends Cogn. Sci.*, vol. 22, no. 1, pp. 8–20, Jan. 2018.
[97] “IBM pitched Watson as a revolution in cancer care. It’s nowhere close,” *STAT*, 05-Sep-2017. .
[98] H. L. Semigran, D. M. Levine, S. Nundy, and A. Mehrotra, “Comparison of Physician and Computer Diagnostic AccuracyComparison of Physician and Computer Diagnostic AccuracyLetters,” *JAMA Intern. Med.*, vol. 176, no. 12, pp. 1860–1861, Dec. 2016.
[99] D. Mery, F. Pedreschi, and A. Soto, “Automated Design of a Computer Vision System for Visual Food Quality Evaluation,” *Food Bioprocess Technol.*, vol. 6, no. 8, pp. 2093–2108, Aug. 2013.

[100] M. Sonka, V. Hlavac, and R. Boyle, “Image pre-processing,” in *Image Processing, Analysis and Machine Vision*, M. Sonka, V. Hlavac, and R. Boyle, Eds. Boston, MA: Springer US, 1993, pp. 56–111.
[101] J. G. Moik, *Digital processing of remotely sensed images*. 1980.
[102] Y.-R. Van Eycke, J. Allard, I. Salmon, O. Debeir, and C. Decaestecker, “Image processing in digital pathology: an opportunity to solve inter-batch variability of immunohistochemical staining,” *Sci. Rep.*, vol. 7, no. 1, pp. 1–15, Feb. 2017.
[103] Y. Alginahi, “Preprocessing Techniques in Character Recognition,” in *Character Recognition*, M. Mori, Ed. Sciyo, 2010.
[104] NIAID, *Mycobacterium tuberculosis Bacteria*. 2015.
[105] “Pedestrian Detection Database.” [Online]. Available: https://www.cis.upenn.edu/~jshi/ped_html/. [Accessed: 15-Oct-2019].
[106] S. S. Haykin and S. S. Haykin, *Neural networks and learning machines*, 3rd ed. New York: Prentice Hall, 2009.
[107] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, “Exploring Strategies for Training Deep Neural Networks,” *J Mach Learn Res*, vol. 10, pp. 1–40, Jun. 2009.
[108] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” *J. Physiol.*, vol. 160, no. 1, pp. 106-154.2, Jan. 1962.
[109] J. Bruna, S. Chintala, Y. LeCun, S. Piantino, A. Szlam, and M. Tygert, “A mathematical motivation for complex-valued convolutional networks,” *ArXiv150303438 Cs Stat*, Mar. 2015.
[110] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[111] K. Fukushima and S. Miyake, “Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition,” in *Competition and Cooperation in Neural Nets*, 1982, pp. 267–285.
[112] Ç. Kaymak and A. Uçar, *A Brief Survey and an Application of Semantic Image Segmentation for Autonomous Driving*. 2018.
[113] D. Hutchison *et al.*, “Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition,” in *Artificial Neural Networks – ICANN 2010*, vol. 6354, K. Diamantaras, W. Duch, and L. S. Iliadis, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 92–101.
[114] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.
[115] P. Whiting, A. W. S. Rutjes, J. B. Reitsma, A. S. Glas, P. M. M. Bossuyt, and J. Kleijnen, “Sources of Variation and Bias in Studies of Diagnostic Accuracy: A Systematic Review,” *Ann. Intern. Med.*, vol. 140, no. 3, p. 189, Feb. 2004.
[116] M.-C. Rousselet *et al.*, “Sources of variability in histological scoring of chronic viral hepatitis,” *Hepatology*, vol. 41, no. 2, pp. 257–264, Feb. 2005.
[117] F. Aeffner *et al.*, “The Gold Standard Paradox in Digital Image Analysis: Manual Versus Automated Scoring as Ground Truth,” *Arch. Pathol. Lab. Med.*, vol. 141, no. 9, pp. 1267–1275, Sep. 2017.
[118] A. E. Rey, G. T. Vallet, B. Riou, M. Lesourd, and R. Versace, “Memory plays tricks on me: perceptual bias induced by memory reactivated size in Ebbinghaus illusion,” *Acta Psychol. (Amst.)*, vol. 161, pp. 104–109, Oct. 2015.
[119] S. Coren and J. T. Enns, “Size contrast as a function of conceptual similarity between test and inducers,” *Percept. Psychophys.*, vol. 54, no. 5, pp. 579–588, Nov. 1993.
[120] M. K. Albert, “Occlusion, transparency, and lightness,” *Vision Res.*, vol. 47, no. 24, pp. 3061–3069, Nov. 2007.

[121] “Adelson E. Lightness perception and lightness illusions. In: Gazzaniga M, ed. The New Cognitive Neurosciences. 2nd ed. Cambridge, MA: MIT Press; 2001: 339–351 - Google Search.” [Online]. Available: https://www.google.com/search?q=Adelson+E.+Lightness+perception+and+lightness+illusions.+In%3A+Gazzaniga+M%2C+ed.+The+New+Cognitive+Neurosciences.+2nd+ed.+Cambridge%2C+MA%3A+MIT+Press%3B+2001%3A+339%E2%80%93351&rlz=1C1GCEA_enUS787US787&oq=Adelson+E.+Lightness+perception+and+lightness+illusions.+In%3A+Gazzaniga+M%2C++ed.+The+New+Cognitive+Neurosciences.+2nd+ed.+Cambridge%2C+MA%3A+MIT+Press%3B+2001%3A++339%E2%80%93351&aqs=chrome..69i57&sourceid=chrome&ie=UTF-8. [Accessed: 04-Aug-2019].
[122] “OSA | Number of discernible object colors is a conundrum.” [Online]. Available: https://www.osapublishing.org/josaa/abstract.cfm?uri=josaa-30-2-264. [Accessed: 04-Aug-2019].
[123] A. Valberg, “Color induction: Dependence on luminance, purity, and dominant or complementary wavelength of inducing stimuli,” *JOSA*, vol. 64, no. 11, pp. 1531–1540, Nov. 1974.
[124] R. B. Lotto and D. Purves, “An empirical explanation of color contrast,” *Proc. Natl. Acad. Sci.*, vol. 97, no. 23, pp. 12834–12839, Nov. 2000.
[125] “14-bit vs 12-bit RAW - Can You Tell The Difference?” [Online]. Available: https://photographylife.com/14-bit-vs-12-bit-raw. [Accessed: 04-Aug-2019].
[126] R. S. Nickerson, “Confirmation Bias: A Ubiquitous Phenomenon in Many Guises,” p. 46.
[127] H. Ditrich, “Cognitive fallacies and criminal investigations,” *Sci. Justice J. Forensic Sci. Soc.*, vol. 55, no. 2, pp. 155–159, Mar. 2015.
[128] I. B. Anderson, T. I. Sørensen, and A. Prener, “Increase in incidence of disease due to diagnostic drift: primary liver cancer in Denmark, 1943-85,” *BMJ*, vol. 302, no. 6774, pp. 437–440, Feb. 1991.
[129] E. F. McInnes and C. L. Scudamore, “Review of approaches to the recording of background lesions in toxicologic pathology studies in rats,” *Toxicol. Lett.*, vol. 229, no. 1, pp. 134–143, Aug. 2014.
[130] T. J. Fuchs, P. J. Wild, H. Moch, and J. M. Buhmann, “Computational Pathology Analysis of Tissue Microarrays Predicts Survival of Renal Clear Cell Carcinoma Patients,” in *Proceedings of the international conference on Medical Image Computing and Computer-Assisted Intervention MICCAI*, Berlin, Heidelberg, 2008, vol. 5242, pp. 1–8.
[131] D. Cireşan, U. Meier, and J. Schmidhuber, “Multi-column Deep Neural Networks for Image Classification,” *ArXiv12022745 Cs*, Feb. 2012.
[132] I. J. Goodfellow *et al.*, “Generative Adversarial Networks,” *ArXiv14062661 Cs Stat*, Jun. 2014.
[133] “Artificial intelligence in healthcare | Nature Biomedical Engineering.” [Online]. Available: https://www.nature.com/articles/s41551-018-0305-z. [Accessed: 05-Aug-2019].
[134] S. Kouchaki *et al.*, “Application of machine learning techniques to tuberculosis drug resistance analysis,” *Bioinforma. Oxf. Engl.*, Nov. 2018.
[135] “Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization | Scientific Reports.” [Online]. Available: https://www.nature.com/articles/s41598-019-42557-4. [Accessed: 05-Aug-2019].
[136] S. Lopez-Garnier, P. Sheen, and M. Zimic, “Automatic diagnostics of tuberculosis using convolutional neural networks analysis of MODS digital images,” *PLOS ONE*, vol. 14, no. 2, p. e0212094, Feb. 2019.
[137] “The top 10 causes of death,” *World Health Organization*. [Online]. Available: http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. [Accessed: 28-Jun-2018].
[138] David S. Barnes, *The making of a social disease tuberculosis in nineteenth-century France*. Berkeley, Place of publication not identified, Berkeley, Calif.: University of California Press, 1995.

[139] J. L. Cook, “Pathogenesis of Human Pulmonary Tuberculosis: Insights from the Rabbit Model By Arthur M. Dannenberg, Jr. Washington, D.C.: ASM Press, 2006. 468 pp., illustrated. $149.95 (cloth),” *Clin. Infect. Dis.*, vol. 44, no. 9, pp. 1257–1257, May 2007.
[140] G. Manina, N. Dhar, and J. D. McKinney, “Stress and Host Immunity Amplify Mycobacterium tuberculosis Phenotypic Heterogeneity and Induce Nongrowing Metabolically Active Forms,” *Cell Host Microbe*, vol. 17, no. 1, pp. 32–46, Jan. 2015.
[141] S. Gagneux, “Host–pathogen coevolution in human tuberculosis,” *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 367, no. 1590, pp. 850–859, Mar. 2012.
[142] D. F. Warner and V. Mizrahi, “Tuberculosis Chemotherapy: the Influence of Bacillary Stress and Damage Response Pathways on Drug Efficacy,” *Clin. Microbiol. Rev.*, vol. 19, no. 3, pp. 558–570, Jul. 2006.
[143] R. S. Wallis, M. Palaci, and K. Eisenach, “Persistence, Not Resistance, Is the Cause of Loss of Isoniazid Effect,” *J. Infect. Dis.*, vol. 195, no. 12, pp. 1870–1871, Jun. 2007.
[144] R. R. Kempker, M. Kipiani, V. Mirtskhulava, N. Tukvadze, M. J. Magee, and H. M. Blumberg, “Acquired Drug Resistance in Mycobacterium tuberculosis and Poor Outcomes among Patients with Multidrug-Resistant Tuberculosis,” *Emerg. Infect. Dis.*, vol. 21, no. 6, pp. 992–1001, Jun. 2015.
[145] C. D. Hamilton *et al.*, “The value of end-of-treatment chest radiograph in predicting pulmonary tuberculosis relapse,” Sep-2008. [Online]. Available: https://www.ingentaconnect.com/content/iuatld/ijtld/2008/00000012/00000009/art00014. [Accessed: 01-Aug-2018].
[146] H.-R. Kim *et al.*, “Impact of Extensive Drug Resistance on Treatment Outcomes in Non-HIV-Infected Patients with Multidrug-Resistant Tuberculosis,” *Clin. Infect. Dis.*, vol. 45, no. 10, pp. 1290–1295, Nov. 2007.
[147] B. Prideaux *et al.*, “The association between sterilizing activity and drug distribution into tuberculosis lesions,” *Nat. Med.*, vol. 21, no. 10, pp. 1223–1227, Oct. 2015.
[148] E. J. Muñoz-Elías, J. Timm, T. Botha, W.-T. Chan, J. E. Gomez, and J. D. McKinney, “Replication Dynamics of Mycobacterium tuberculosis in Chronically Infected Mice,” *Infect. Immun.*, vol. 73, no. 1, pp. 546–551, Jan. 2005.
[149] S. M. Irwin *et al.*, “Limited Activity of Clofazimine as a Single Drug in a Mouse Model of Tuberculosis Exhibiting Caseous Necrotic Granulomas,” *Antimicrob. Agents Chemother.*, vol. 58, no. 7, pp. 4026–4034, Jul. 2014.
[150] A. A. Ordonez *et al.*, “Mouse model of pulmonary cavitary tuberculosis and expression of matrix metalloproteinase-9,” *Dis. Model. Mech.*, vol. 9, no. 7, pp. 779–788, 01 2016.
[151] I. M. Orme, “Tuberculosis Vaccine Types and Timings,” *Clin Vaccine Immunol*, vol. 22, no. 3, pp. 249–257, Mar. 2015.
[152] V. K. Sambandamurthy *et al.*, “Long-Term Protection against Tuberculosis following Vaccination with a Severely Attenuated Double Lysine and Pantothenate Auxotroph of Mycobacterium tuberculosis,” *Infect. Immun.*, vol. 73, no. 2, pp. 1196–1203, Feb. 2005.
[153] L. E. Via *et al.*, “A Sterilizing Tuberculosis Treatment Regimen Is Associated with Faster Clearance of Bacteria in Cavitary Lesions in Marmosets,” *Antimicrob. Agents Chemother.*, vol. 59, no. 7, pp. 4181–4189, Jul. 2015.
[154] A. G. White *et al.*, “Analysis of 18FDG PET/CT Imaging as a Tool for Studying Mycobacterium tuberculosis Infection and Treatment in Non-human Primates,” *JoVE J. Vis. Exp.*, no. 127, pp. e56375–e56375, Sep. 2017.
[155] L. Via *et al.*, “Infection Dynamics and Response to Chemotherapy in a Rabbit Model of Tuberculosis using [F-18]2-Fluoro-Deoxy-D-Glucose Positron Emission Tomography and Computed Tomography,” *Antimicrob. Agents Chemother.*, vol. 56, pp. 4391–402, Jun. 2012.

[156] M. Lee *et al.*, “Linezolid for Treatment of Chronic Extensively Drug-Resistant Tuberculosis,” *N. Engl. J. Med.*, vol. 367, no. 16, pp. 1508–1518, Oct. 2012.
[157] S. T. Malherbe *et al.*, “Persisting PET-CT lesion activity and M. tuberculosis mRNA after pulmonary tuberculosis cure,” *Nat. Med.*, vol. 22, no. 10, pp. 1094–1100, Oct. 2016.
[158] P. L. Lin *et al.*, “Radiologic Responses in Cynomolgus Macaques for Assessing Tuberculosis Chemotherapy Regimens,” *Antimicrob. Agents Chemother.*, vol. 57, no. 9, pp. 4237–4244, Sep. 2013.
[159] R. J. Basaraba *et al.*, “Lymphadenitis as a major element of disease in the guinea pig model of tuberculosis,” *Tuberculosis*, vol. 86, no. 5, pp. 386–394, Sep. 2006.
[160] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, “Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model,” *Sci. Rep.*, vol. 7, no. 1, p. 4172, Jun. 2017.
[161] Z. Yan *et al.*, “Multi-Instance Deep Learning: Discover Discriminative Local Anatomies for Bodypart Recognition,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1332–1343, May 2016.
[162] G. Wu, M. Kim, Q. Wang, Y. Gao, S. Liao, and D. Shen, “Unsupervised Deep Feature Learning for Deformable Registration of MR Brain Images,” in *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2013, vol. 16, pp. 649–56.
[163] Y. Xu *et al.*, “Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features,” *BMC Bioinformatics*, vol. 18, May 2017.
[164] “Machine Learning Methods for Histopathological Image Analysis.” [Online]. Available: https://reader.elsevier.com/reader/sd/24A2E9B94F1FBA96FAF83C82BEAB7F6D62B6160BEB3D2BAE2E8D14DB52DA2639EA3E0031E6FEB1853F99A9A97720D6AD. [Accessed: 28-Jun-2018].
[165] B. Kieffer, M. Babaie, S. Kalra, and H. R. Tizhoosh, “Convolutional Neural Networks for Histopathology Image Classification: Training vs. Using Pre-Trained Networks,” *ArXiv171005726 Cs*, Oct. 2017.
[166] L. Hou *et al.*, “Automatic histopathology image analysis with CNNs,” in *2016 New York Scientific Data Summit (NYSDS)*, 2016, pp. 1–6.
[167] N. Hatipoglu and G. Bilgin, “Classification of histopathological images using convolutional neural network,” in *2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2014, pp. 1–6.
[168] P. Mobadersany *et al.*, “Predicting cancer outcomes from histology and genomics using convolutional networks,” *Proc. Natl. Acad. Sci.*, p. 201717139, Mar. 2018.
[169] M. K. S. Varma, N. K. K. Rao, K. K. Raju, and G. P. S. Varma, “Pixel-Based Classification Using Support Vector Machine Classifier,” in *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 2016, pp. 51–55.
[170] A. Tzotsos, *A Support Vector Machine Approach for Object Based Image Analysis*. .
[171] Y. Lecun and Y. Bengio, “Convolutional networks for images, speech, and time-series,” *Handb. Brain Theory Neural Netw.*, 1995.
[172] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[173] Y. Ding, Q. Feng, T. Wang, and X. Fu, “A modular neural network architecture with concept,” *Neurocomputing*, vol. 125, pp. 3–6, Feb. 2014.
[174] B. L. M. Happel and J. M. J. Murre, “Design and evolution of modular neural network architectures,” *Neural Netw.*, vol. 7, no. 6, pp. 985–1004, Jan. 1994.
[175] “Deep and Modular Neural Networks | SpringerLink.” [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-662-43505-2_28. [Accessed: 18-Jul-2018].

[176] C. S. Pattichis, F. Schnorrenberg, C. N. Schizas, M. S. Pattichis, and K. Kyriacou, “A Modular Neural Network System for the Analysis of Nuclei in Histopathological Sections,” in *Computational Intelligence Processing in Medical Diagnosis*, Physica, Heidelberg, 2002, pp. 291–322.
[177] P. L. Dunn and R. J. North, “Virulence ranking of some Mycobacterium tuberculosis and Mycobacterium bovis strains according to their ability to multiply in the lungs, induce lung pathology, and cause mortality in mice,” *Infect. Immun.*, vol. 63, no. 9, pp. 3428–3437, Sep. 1995.
[178] B. P. Kelly, S. K. Furney, M. T. Jessen, and I. M. Orme, “Low-dose aerosol infection model for testing drugs for efficacy against Mycobacterium tuberculosis.,” *Antimicrob. Agents Chemother.*, vol. 40, no. 12, pp. 2809–2812, Dec. 1996.
[179] M. A. De Groote *et al.*, “Comparative Studies Evaluating Mouse Models Used for Efficacy Testing of Experimental Drugs against Mycobacterium tuberculosis,” *Antimicrob. Agents Chemother.*, vol. 55, no. 3, pp. 1237–1247, Mar. 2011.
[180] J. C. Cyktor, B. Carruthers, R. A. Kominsky, G. L. Beamer, P. Stromberg, and J. Turner, “IL-10 Inhibits Mature Fibrotic Granuloma Formation during Mycobacterium tuberculosis Infection,” *J. Immunol.*, p. 1202722, Feb. 2013.
[181] “Labeling Nodes Using Three Degrees of Propagation.” [Online]. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0051947. [Accessed: 19-Jul-2018].
[182] M. Gamer, J. Lemon, and I. F. P. Singh, *irr: Various Coefficients of Interrater Reliability and Agreement*. 2019.
[183] J. L. Fleiss, “Measuring nominal scale agreement among many raters.,” *Psychol. Bull.*, vol. 76, no. 5, pp. 378–382, 1971.
[184] K. A. Hallgren, “Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial,” *Tutor. Quant. Methods Psychol.*, vol. 8, no. 1, pp. 23–34, 2012.
[185] M. L. McHugh, “Interrater reliability: the kappa statistic,” *Biochem. Medica*, vol. 22, no. 3, pp. 276–282, Oct. 2012.
[186] K. Krippendorff, “Computing Krippendorff’s Alpha-Reliability,” p. 12.
[187] K. Krippendorff, *Content analysis an introduction to its methodology*, Fourth Edition. Los Angeles: SAGE, 2019.
[188] R. Klopfleisch, “Multiparametric and semiquantitative scoring systems for the evaluation of mouse model histopathology - a systematic review,” *BMC Vet. Res.*, vol. 9, no. 1, p. 123, Jun. 2013.
[189] M. A. Gavrielides, B. D. Gallas, P. Lenz, A. Badano, and S. M. Hewitt, “Observer Variability in the Interpretation of HER2/neu Immunohistochemical Expression With Unaided and Computer-Aided Digital Microscopy,” *Arch. Pathol. Lab. Med.*, vol. 135, no. 2, pp. 233–242, Feb. 2011.
[190] D. D. Miller and E. W. Brown, “Artificial Intelligence in Medical Practice: The Question to the Answer?,” *Am. J. Med.*, vol. 131, no. 2, pp. 129–133, Feb. 2018.
[191] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?,” in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 3320–3328.
[192] G. J Ryan, H. Shapiro, and A. J Lenaerts, *Improving acid-fast fluorescent staining for the detection of mycobacteria using a new nucleic acid staining approach*, vol. 94. 2014.
[193] L. Blanc, A. Lenaerts, V. Dartois, and B. Prideaux, “Visualization of Mycobacterial Biomarkers and Tuberculosis Drugs in Infected Tissue by MALDI-MS Imaging,” *Anal. Chem.*, vol. 90, no. 10, pp. 6275–6282, May 2018.
[194] “Treatment Regimens for Latent TB Infection | Treatment | TB | CDC,” 18-Jul-2019. [Online]. Available: https://www.cdc.gov/tb/topic/treatment/ltbi.htm. [Accessed: 24-Oct-2019].
[195] World Health Organization and Stop TB Initiative (World Health Organization), Eds., *Treatment of tuberculosis: guidelines*, 4th ed. Geneva: World Health Organization, 2010.

[196] G. S. Timmins and V. Deretic, “Mechanisms of action of isoniazid,” *Mol. Microbiol.*, vol. 62, no. 5, pp. 1220–1227, Dec. 2006.
[197] “Dynamic Persistence of Antibiotic-Stressed Mycobacteria | Science.” [Online]. Available: https://science.sciencemag.org/content/339/6115/91. [Accessed: 29-Aug-2019].
[198] “Tuberculosis Drugs and Mechanisms of Action | NIH: National Institute of Allergy and Infectious Diseases.” [Online]. Available: https://www.niaid.nih.gov/diseases-conditions/tbdrugs. [Accessed: 24-Oct-2019].
[199] Y. Tang *et al.*, “Non-adherence to anti-tuberculosis treatment among internal migrants with pulmonary tuberculosis in Shenzhen, China: a cross-sectional study,” *BMC Public Health*, vol. 15, May 2015.
[200] C. Lange, D. Chesov, J. Heyckendorf, C. C. Leung, Z. Udwadia, and K. Dheda, “Drug-resistant tuberculosis: An update on disease burden, diagnosis and treatment,” *Respirol. Carlton Vic*, vol. 23, no. 7, pp. 656–673, Jul. 2018.
[201] “Tuberculosis surveillance and monitoring in Europe 2015,” *European Centre for Disease Prevention and Control*, 17-Mar-2015. [Online]. Available: https://www.ecdc.europa.eu/en/publications-data/tuberculosis-surveillance-and-monitoring-europe-2015. [Accessed: 24-Oct-2019].
[202] “WHO | Global tuberculosis report 2019,” *WHO*. [Online]. Available: http://www.who.int/tb/publications/global_report/en/. [Accessed: 24-Oct-2019].
[203] C. H. Wong, K. W. Siah, and A. W. Lo, “Estimation of clinical trial success rates and related parameters,” *Biostatistics*, vol. 20, no. 2, pp. 273–286, Apr. 2019.
[204] O. of the Commissioner, “Step 1: Discovery and Development,” *FDA*, Apr. 2019.
[205] J. L. Davis, “Chapter 2 - Pharmacologic Principles,” in *Equine Internal Medicine (Fourth Edition)*, S. M. Reed, W. M. Bayly, and D. C. Sellon, Eds. W.B. Saunders, 2018, pp. 79–137.
[206] E. R. Driver, “C3HeB/FeJ mice as a novel preclinical mouse model for Mycobacterium tuberculosis : an analysis of the host pathogenesis and the in vivo environment of the necrotic granuloma, The,” Thesis, Colorado State University. Libraries, 2015.
[207] J. Harper *et al.*, “Mouse Model of Necrotic Tuberculosis Granulomas Develops Hypoxic Lesions,” *J. Infect. Dis.*, vol. 205, no. 4, pp. 595–602, Feb. 2012.
[208] A. J. Lenaerts *et al.*, “Location of Persisting Mycobacteria in a Guinea Pig Model of Tuberculosis Revealed by R207910,” *Antimicrob. Agents Chemother.*, vol. 51, no. 9, pp. 3338–3345, Sep. 2007.
[209] S. J. Sasindran and J. B. Torrelles, “Mycobacterium Tuberculosis Infection and Inflammation: what is Beneficial for the Host and for the Bacterium?,” *Front. Microbiol.*, vol. 2, Jan. 2011.
[210] V. Dartois and C. E. Barry, “Clinical pharmacology and lesion penetrating properties of second- and third-line antituberculous agents used in the management of multidrug-resistant (MDR) and extensively-drug resistant (XDR) tuberculosis.,” *Curr. Clin. Pharmacol.*, vol. 5, no. 2, pp. 96–114, May 2010.
[211] J. P. Sarathy *et al.*, “Prediction of Drug Penetration in Tuberculosis Lesions,” *ACS Infect. Dis.*, vol. 2, no. 8, pp. 552–563, Aug. 2016.
[212] G. V. Rayasam and T. S. Balganesh, “Exploring the potential of adjunct therapy in tuberculosis,” *Trends Pharmacol. Sci.*, vol. 36, no. 8, pp. 506–513, Aug. 2015.
[213] E. Hennessy, C. Adams, F. J. Reen, and F. O’Gara, “Is There Potential for Repurposing Statins as Novel Antimicrobials?,” *Antimicrob. Agents Chemother.*, vol. 60, no. 9, pp. 5111–5121, Sep. 2016.
[214] J.-W. C. Alffenaar, O. W. Akkerman, and R. van Hest, “Statin Adjunctive Therapy for Tuberculosis Treatment,” *Antimicrob. Agents Chemother.*, vol. 60, no. 11, pp. 7004–7004, Nov. 2016.
[215] D. A. Jolliffe *et al.*, “Adjunctive vitamin D in tuberculosis treatment: meta-analysis of individual participant data,” *Eur. Respir. J.*, vol. 53, no. 3, Mar. 2019.

[216] "Losartan: MedlinePlus Drug Information." [Online]. Available: https://medlineplus.gov/druginfo/meds/a695008.html. [Accessed: 29-Aug-2019].
[217] R. D. Cohn *et al.*, "Angiotensin II type 1 receptor blockade attenuates TGF-beta-induced failure of muscle regeneration in multiple myopathic states," *Nat. Med.*, vol. 13, no. 2, pp. 204–210, Feb. 2007.
[218] K. Garg, B. T. Corona, and T. J. Walters, "Losartan administration reduces fibrosis but hinders functional recovery after volumetric muscle loss injury," *J. Appl. Physiol.*, vol. 117, no. 10, pp. 1120–1131, Sep. 2014.
[219] M. Molina-Molina *et al.*, "Losartan attenuates bleomycin induced lung fibrosis by increasing prostaglandin E_2 synthesis," *Thorax*, vol. 61, no. 7, p. 604, Jul. 2006.
[220] M. Robinson *et al.*, "Quality assurance guidance for scoring and reporting for pathologists and laboratories undertaking clinical trial work," *J. Pathol. Clin. Res.*, vol. 5, no. 2, pp. 91–99, 2019.
[221] A. J. Lenaerts *et al.*, "Preclinical Testing of the Nitroimidazopyran PA-824 for Activity against Mycobacterium tuberculosis in a Series of In Vitro and In Vivo Models," *Antimicrob. Agents Chemother.*, vol. 49, no. 6, pp. 2294–2301, Jun. 2005.
[222] I. M. Orme, "Immune Responses in Animal Models," in *Tuberculosis*, T. M. Shinnick, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1996, pp. 181–196.
[223] D. J. Ordway *et al.*, "Evaluation of Standard Chemotherapy in the Guinea Pig Model of Tuberculosis," *Antimicrob. Agents Chemother.*, vol. 54, no. 5, pp. 1820–1833, May 2010.
[224] Y. Zhang, M. M. Wade, A. Scorpio, H. Zhang, and Z. Sun, "Mode of action of pyrazinamide: disruption of Mycobacterium tuberculosis membrane transport and energetics by pyrazinoic acid," *J. Antimicrob. Chemother.*, vol. 52, no. 5, pp. 790–795, Nov. 2003.
[225] K. Kurthkoti *et al.*, "The Capacity of Mycobacterium tuberculosis To Survive Iron Starvation Might Enable It To Persist in Iron-Deprived Microenvironments of Human Granulomas," *mBio*, vol. 8, no. 4, pp. e01092-17, Sep. 2017.
[226] H. Polena *et al.*, "Mycobacterium tuberculosis exploits the formation of new blood vessels for its dissemination," *Sci. Rep.*, vol. 6, Sep. 2016.
[227] M. C. Tsai *et al.*, "Characterization of the tuberculous granuloma in murine and human lungs: cellular composition and relative tissue oxygen tension," *Cell. Microbiol.*, vol. 8, no. 2, pp. 218–232, 2006.
[228] M.-J. Kim *et al.*, "Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism," *EMBO Mol. Med.*, vol. 2, no. 7, pp. 258–274, Jul. 2010.
[229] S. K. Ward, B. Abomoelak, S. A. Marcus, and A. M. Talaat, "Transcriptional Profiling of Mycobacterium Tuberculosis During Infection: Lessons Learned," *Front. Microbiol.*, vol. 1, Nov. 2010.
[230] W. Lee, B. C. VanderVen, R. J. Fahey, and D. G. Russell, "Intracellular Mycobacterium tuberculosis Exploits Host-derived Fatty Acids to Limit Metabolic Stress," *J. Biol. Chem.*, vol. 288, no. 10, pp. 6788–6800, Mar. 2013.
[231] "Mycobacterial survival strategies in the phagosome: defence against host stresses. - PubMed - NCBI." [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/19438516. [Accessed: 31-Oct-2019].
[232] "Targeting phenotypically tolerant Mycobacterium tuberculosis." [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5367488/. [Accessed: 31-Oct-2019].
[233] L. G. Wayne, "Synchronized replication of Mycobacterium tuberculosis.," *Infect. Immun.*, vol. 17, no. 3, p. 528, Sep. 1977.
[234] L. G. Wayne, "Dormancy of Mycobacterium tuberculosis and latency of disease," *Eur. J. Clin. Microbiol. Infect. Dis. Off. Publ. Eur. Soc. Clin. Microbiol.*, vol. 13, no. 11, pp. 908–914, Nov. 1994.

[235] M. Gengenbacher and S. H. E. Kaufmann, “Mycobacterium tuberculosis: Success through dormancy,” *Fems Microbiol. Rev.*, vol. 36, no. 3, pp. 514–532, May 2012.
[236] N. D. Walter *et al.*, “Transcriptional Adaptation of Drug-tolerant Mycobacterium tuberculosis During Treatment of Human Tuberculosis,” *J. Infect. Dis.*, vol. 212, no. 6, pp. 990–998, Sep. 2015.
[237] G. L. Hobby, K. Meyer, and E. Chaffee, “Observations on the Mechanism of Action of Penicillin.,” *Proc. Soc. Exp. Biol. Med.*, vol. 50, no. 2, pp. 281–285, Jun. 1942.
[238] “Gene expression profile of Mycobacterium tuberculosis in a non-replicating state. - PubMed - NCBI.” [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/15207493. [Accessed: 05-Nov-2019].
[239] H. L. Torrey, I. Keren, L. E. Via, J. S. Lee, and K. Lewis, “High Persister Mutants in Mycobacterium tuberculosis,” *PLoS ONE*, vol. 11, no. 5, May 2016.
[240] Z. Ahmad *et al.*, “Biphasic Kill Curve of Isoniazid Reveals the Presence of Drug-Tolerant, Not Drug-Resistant, Mycobacterium tuberculosis in the Guinea Pig,” *J. Infect. Dis.*, vol. 200, no. 7, pp. 1136–1143, Oct. 2009.
[241] E. C. Hett and E. J. Rubin, “Bacterial Growth and Cell Division: a Mycobacterial Perspective,” *Microbiol. Mol. Biol. Rev. MMBR*, vol. 72, no. 1, pp. 126–156, Mar. 2008.
[242] C. Deb *et al.*, “A novel in vitro multiple-stress dormancy model for Mycobacterium tuberculosis generates a lipid-loaded, drug-tolerant, dormant pathogen,” *PloS One*, vol. 4, no. 6, p. e6077, Jun. 2009.
[243] A. Bhatt *et al.*, “Deletion of kasB in Mycobacterium tuberculosis causes loss of acid-fastness and subclinical latent tuberculosis in immunocompetent mice,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 12, pp. 5157–5162, Mar. 2007.
[244] Y. Yuan, Y. Zhu, D. D. Crane, and C. E. Barry, “The effect of oxygenated mycolic acid composition on cell wall function and macrophage growth in Mycobacterium tuberculosis,” *Mol. Microbiol.*, vol. 29, no. 6, pp. 1449–1458, Sep. 1998.
[245] P. Seiler *et al.*, “Cell-Wall Alterations as an Attribute of Mycobacterium tuberculosis in Latent Infection,” *J. Infect. Dis.*, vol. 188, no. 9, pp. 1326–1331, Nov. 2003.
[246] H. Yamada, M. Yamaguchi, K. Chikamatsu, A. Aono, and S. Mitarai, “Structome Analysis of Virulent Mycobacterium tuberculosis, Which Survives with Only 700 Ribosomes per 0.1 fl of Cytoplasm,” *PLoS ONE*, vol. 10, no. 1, Jan. 2015.
[247] R. Hazan, Y.-A. Que, D. Maura, and L. G. Rahme, “A method for high throughput determination of viable bacteria cell counts in 96-well plates,” *BMC Microbiol.*, vol. 12, p. 259, Nov. 2012.
[248] D. Qin, X. He, K. Wang, X. J. Zhao, W. Tan, and J. Chen, “Fluorescent Nanoparticle-Based Indirect Immunofluorescence Microscopy for Detection of Mycobacterium tuberculosis,” *J. Biomed. Biotechnol.*, vol. 2007, 2007.
[249] G. Tarhan, L. Ordulu, F. Gümüşlü, I. Ceyhan, and S. Cesur, “[Comparison of auramine-rhodamine and Erlich-Ziehl-Neelsen staining methods for the diagnosis of tuberculosis],” *Mikrobiyol. Bul.*, vol. 37, no. 2–3, pp. 131–136, Jun. 2003.
[250] J. Shah *et al.*, “Dual color fluorescence in situ hybridization (FISH) assays for detecting Mycobacterium tuberculosis and Mycobacterium avium complexes and related pathogens in cultures,” *PLoS ONE*, vol. 12, no. 4, Apr. 2017.
[251] M. Okada *et al.*, “Novel therapeutic vaccine: Granulysin and new DNA vaccine against Tuberculosis,” *Hum. Vaccin.*, vol. 7, no. sup1, pp. 60–67, Jan. 2011.
[252] S. Mezouar, I. Diarra, J. Roudier, B. Desnues, and J.-L. Mege, “Tumor Necrosis Factor-Alpha Antagonist Interferes With the Formation of Granulomatous Multinucleated Giant Cells: New Insights Into Mycobacterium tuberculosis Infection,” *Front. Immunol.*, vol. 10, 2019.

[253] S. D. Chakravarty *et al.*, “Tumor necrosis factor blockade in chronic murine tuberculosis enhances granulomatous inflammation and disorganizes granulomas in the lungs,” *Infect. Immun.*, vol. 76, no. 3, pp. 916–926, Mar. 2008.
[254] M. Datta *et al.*, “Anti-vascular endothelial growth factor treatment normalizes tuberculosis granuloma vasculature and improves small molecule delivery,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 6, pp. 1827–1832, Feb. 2015.
[255] N. P. Kumar, V. V. Banurekha, D. Nair, and S. Babu, “Circulating Angiogenic Factors as Biomarkers of Disease Severity and Bacterial Burden in Pulmonary Tuberculosis,” *PloS One*, vol. 11, no. 1, p. e0146318, 2016.
[256] G. Middlebrook, R. J. Dubos, and C. Pierce, “VIRULENCE AND MORPHOLOGICAL CHARACTERISTICS OF MAMMALIAN TUBERCLE BACILLI,” *J. Exp. Med.*, vol. 86, no. 2, pp. 175–184, Jul. 1947.
[257] H. Bloch, “Studies on the virulence of tubercle bacilli; isolation and biological properties of a constituent of virulent organisms,” *J. Exp. Med.*, vol. 91, no. 2, pp. 197–218, pl, Feb. 1950.

APPENDIX

HETEROGENEITY OF PULMONARY LESION TYPES IN C3HEB/FEJ MICE IS MODULATED BY SPECIFIC CHARACTERISTICS OF THE *M. TUBERCULOSIS* STRAIN USED

A1. Introduction

C3HeB/FeJ mice are a TB animal model that develops caseous necrotic pulmonary granulomas resembling the human granuloma that develops during a TB infection. These granulomas form into three morphologically distinct types that have been classified as encapsulated granulomas with central necrosis (Type I), fulminant neutrophilic alveolitis (Type II), and cellular inflammatory lesions (Type III). It is not fully understood how external factors, such as the bacterial strain used during infection, will influence the host response of determining the lesion type that will be developed. In our laboratory, a culture batch of *Mycobacterium tuberculosis Erdman* unexpectedly exhibited heightened levels of virulence, which included a substantially increased mortality rate, high levels of bacterial cording, and substantial levels of pulmonary lesion involvement. This event was not isolated to our laboratory but was also observed in a collaborator's culture that found similar events. Taking advantage of these hypervirulent cultures, we investigated how variables such as the virulence of the TB strain, method of propagation, infectious dose, time within storage, and the level of bacterial aggregation influence lesion development after an aerosol infection.

This is especially relevant for the C3HeB/FeJ animal model because it has been reported among laboratories that use this animal model that they observe different results than what our laboratory has been reporting. It is especially perplexing since our laboratory can consistently achieve reproducible outcomes of desired lesion development, particularly for the development and selection of the Type I lesion. By investigating the variables that can impact lesion

development, it is our ultimate goal to eventually standardize the model to develop consistent pathology regardless of the laboratory that is performing the study.

A.2 – Materials And Methods

A.2.1 - Bacterial Strains And Culture Methods

There were two different culturing methods that were employed throughout the study. The first is the general culturing methodology used (See Chapter 2 for more information) for all strains except for the CSU hypervirulent Mtb culture. The second is the slight deviation that occurred from the normal standard culturing protocol. The steps are: 1) loop full of Lenaerts seed stock, was streaked on a 7H11/ ADC agar plate to create a bacterial lawn. 2) 22 days later, 1 loop full of the Mtb lawn is used to inoculate 20mL of 7H9/ADC liquid culture for 14 days. 3) 5mL of the 7H9/ADC culture is then transferred into 45mL of PB liquid culture with 0.05% Tween 80 for 7 days. 4) 5mL of the culture is transferred to a fresh 45mL of PB liquid culture with Tween 80 0.05% Tween 80 for 7 days. 5) The culture is upscaled to 500mL of PB media with 0.05% Tween 80 for 6 days. 6) An additional 0.02% Tween is added before being harvested the next day.

A.2.2 - Mouse Infection

C3HeB/FeJ mice were aerosol infected using the strain specified in the study. More detailed information can be found in Chapter 2.

A.2.3 - Staining And Imaging

Bacteria cultures were heat-fixed for 5 minutes on a hot plate and additionally with 4%PFA for 48 hours. After the fixation step, the bacteria were stained using the standard SYBR Gold protocol found within Chapter 2. Tissue preparation, staining, and all microscopy imaging were performed as outlined in Chapter 2.

A.2.4 – LC/MS

10mL of bacterial stock was centrifuged at 2000g to obtain a pellet. The supernatant was discarded and 4mL of 2:1 chloroform:methanol solution was added. The pellet was re-suspended and allowed to sit overnight. The suspended solution was centrifuged again at 2000g to obtain a pellet, and the organic layer is removed. The organic layer was treated with 100µl of 0.2 M NaOH in methanol for 30 minutes at 37C°. After the allotted time, the solution is neutralized with glacial acetic acid until the pH reaches about 7. 200µl of chloroform-methanol 2:1 and 100µl of H2O were added, and the organic phase was removed. The samples were analyzed on an Agilent 6224 LCMS-TOF using the Sartain method in negative mode. Data was analyzed using Mass Hunter, XCMS, and Excel software. Relative abundance for molecular features of interest were normalized using global normalization.

A.3 – Results

A.3.1 – Initial Observations Of The Hypervirulent Strain

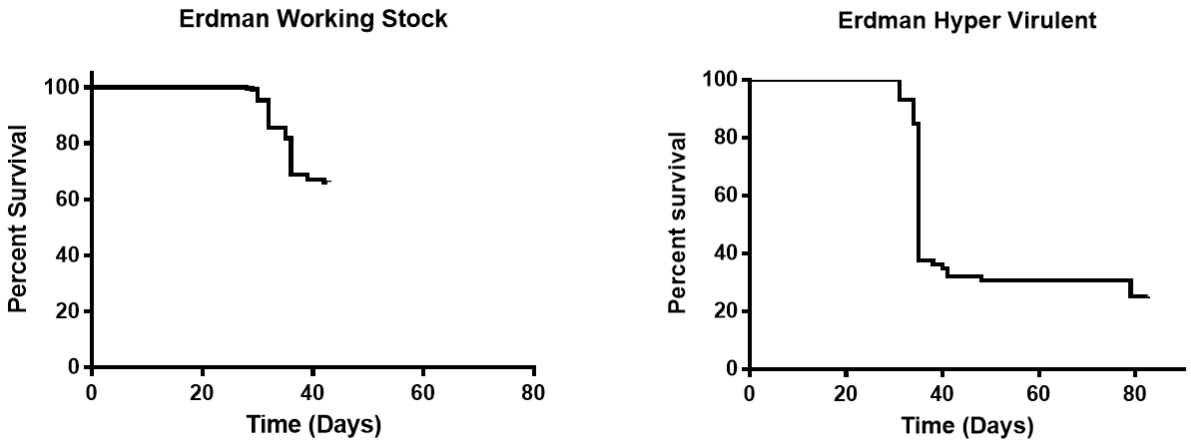


Figure A.1. Graphs of both the standard working stock (Erdman Working Stock) and the culture batch that exhibited hypervirulent characteristics (Erdman Hyper Virulent).

During a routine Mtb aerosol infection with the C3HeB/FeJ mouse model, a new batch of Mtb (hypervirulent Mtb) exhibited an unprecedented amount of mortality (75%) in comparison to the previous culture batch (26%) (working stock Mtb) (Figure A.1). What was more surprising was

that this was also observed in the more Mtb resistant BALB/c strain (100%) (data not shown). This increase in virulence is not explainable by differing levels of infectious doses because the hypervirulent Mtb had an infectious dose of 70 CFU, and the working stock Mtb had an infectious dose of 69 CFU. These results were verified in three independent studies confirming that between batches, our working stock had a substantial increase in virulence.

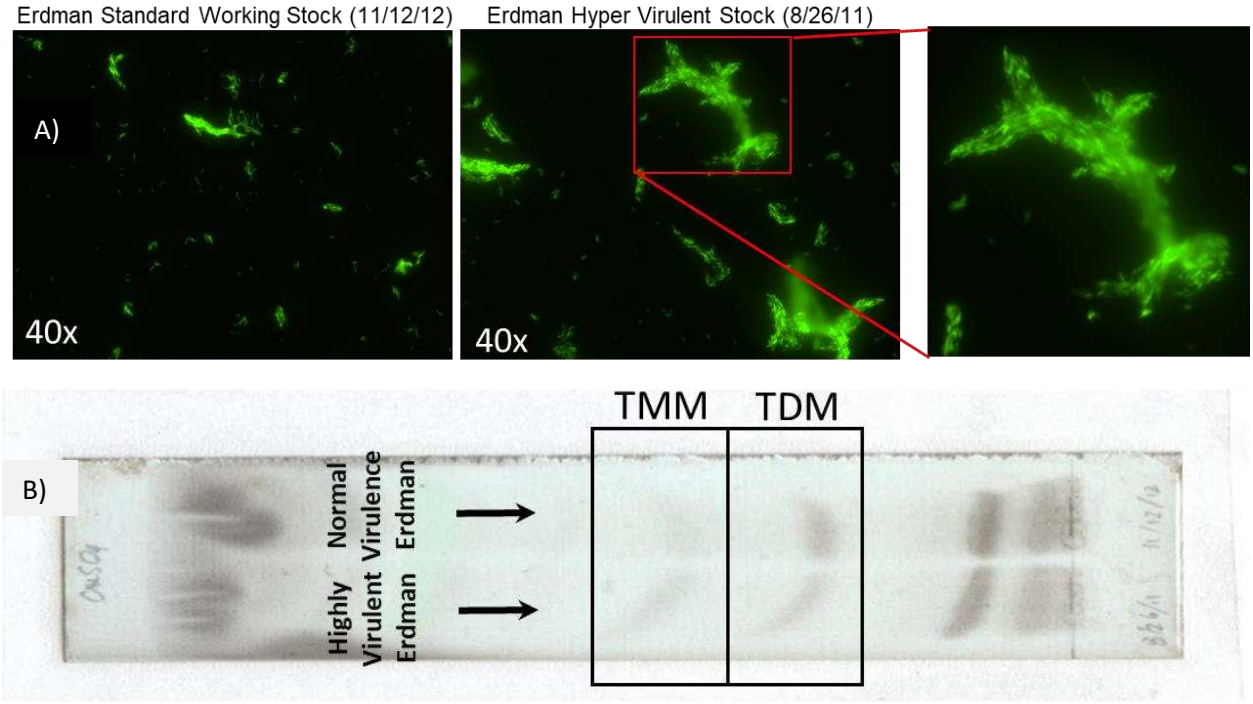


Figure A.2. Microscopy (40x) images of the *Mtb Erdman* culture batches of both the Standard Working Stock (Top Panel Left), the Hyper Virulent Stock (Top Panel Middle), and a digitally enhanced Mtb aggregate (Top Panel Right). The Hyper Virulent Stock showed an increase in bacterial aggregation when compared to the Standard Working Stock. In the bottom panel is the TLC of both the TMM and TDM for both culture batches, which showed no significant differences.

To investigate the cause of virulence further, a batch of our working stock Mtb and our hypervirulent Mtb were stained using SYBR Gold and digitally imaged. The hypervirulent Mtb culture had a substantial increase in visual cording than the working stock (Figure A.2A). Mtb cording is the aggregation of bacteria in a bundle like structure via a parallel arrangement. It is one of many indicators used to predict the virulence of Mtb strains. The degree of cording is determined by the cell wall levels of trehalose-6,6′-dimycolate (TDM), trehalose-6,6′-

monomycolate (TMM), and its modifications [256], [257]. Due to the contrasting levels of cording that was present between the two culture batches, the levels of TDM and TMM were measured by both thin layer chromatography (TLC) and liquid chromatography-mass spectrometry (LC-MS). It was initially predicted that because of the increased level of cording present in the hypervirulent Mtb culture that there would be increased levels of both TDM and TMM. The TLC had no significant differences in either TMM or TDM between the two culture batches (Figure A.2B). This was an unexpected finding that was verified using LC-MS, but the LC-MS did identify a substantial increase in an unidentified molecular feature that was present (Figure A.3).

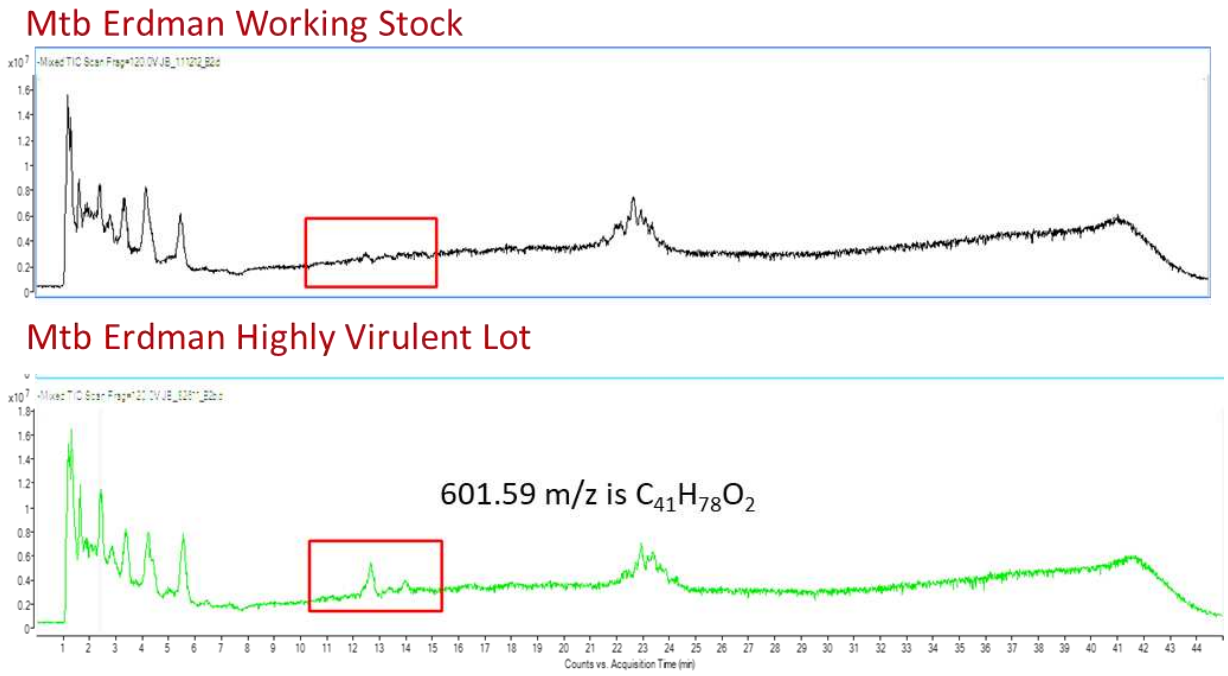
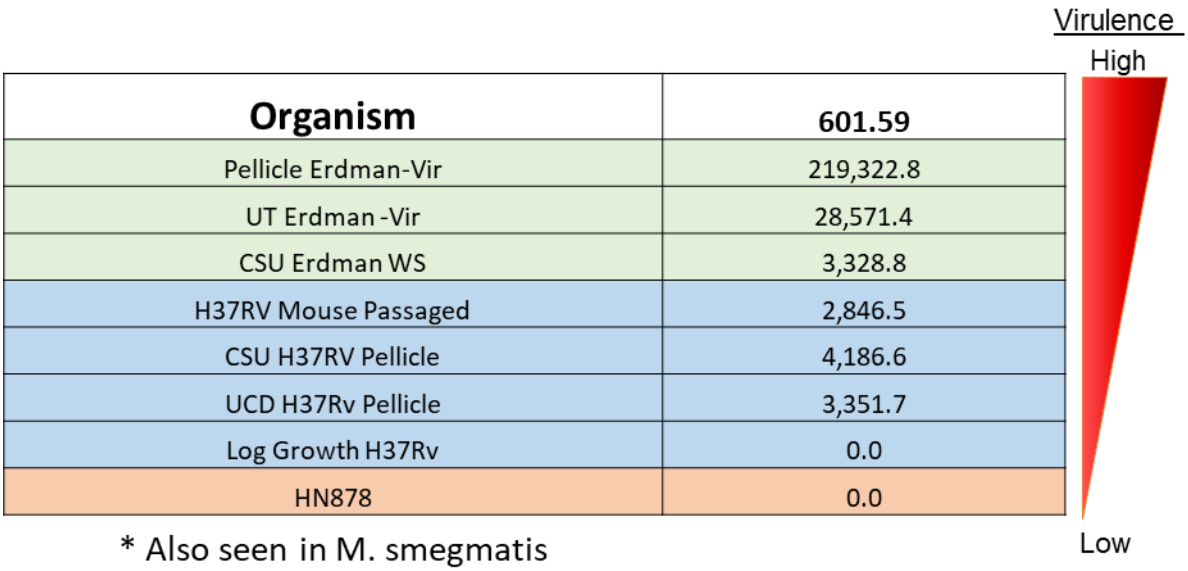


Figure A.3. Graph of the FFA molecular features obtained by the LC/MS. An uncharacterized class of molecular features (red box) is present in the Highly Virulent (Bottom) culture batch but is greatly reduced in the standard Working Stock (Top). The 601.59 m/z is of particular interest, and the proposed molecular formula is listed above.

When performing a global analysis of free fatty acids (FFA) in the negative mode, there was a group of uncharacterized molecular features that had elevated levels of relative abundance in the virulent Mtb culture (Figure A.3). The uncharacterized late series molecular features had

an m/z values of 657.6528, 643.6371, 629.6214, 615.6057, 601.59, 587.5743, 573.5586, 657.6528, and 629.6214. The 615.6057 m/z and 601.59 m/z peaks stood out among these features because they had a fold increase of 12.1 and 9.6 respectively between the two samples. Because of the lack of significant difference in TDM / TMM levels, and there was such a substantial increase in these uncharacterized FFA, it was prudent to investigate these findings further. While the chemical structure of these uncharacterized molecules is unknown, it is suspected that they are part of the mycolic acid biosynthesis pathway. Using GC/MS, it was determined that there is at least one double bond present, and it is highly likely that it contains two.

A.3.2 – Observation Of Increased Virulence In Other Mtb Strains



Organism	601.59
Pellicle Erdman-Vir	219,322.8
UT Erdman -Vir	28,571.4
CSU Erdman WS	3,328.8
H37RV Mouse Passaged	2,846.5
CSU H37RV Pellicle	4,186.6
UCD H37Rv Pellicle	3,351.7
Log Growth H37Rv	0.0
HN878	0.0

* Also seen in M. smegmatis

Figure A.4. Mtb strains (left column) are organized by the level of virulence observed *in vivo* with the corresponding normalized relative abundance of the 601.59 m/z molecular feature (right column). Strains highlighted in green are the *Mtb Erdman* strain, blue are the Mtb H37Rv strain, and pink is Mtb clinical isolate.

Building upon our previous findings, we expanded our analysis to include additional laboratory strains with varying degrees of tested virulence within the C3HeB/FeJ mouse model.

Using LC/MS to obtain the relative abundance values, the strains were rank-ordered from the most virulent strain to the least virulent strain (Figure A.4). A general trend emerged of the most virulent strains exhibiting the highest levels of relative abundance of the 601.59 m/z peak except for the clinical HN878 strain. Under a typical infection, the HN878 strain induces high levels of mortality, lesion development, and bacterial burden. However, a successive mouse experiment that used the specific culture batch of HN878 had exhibited low levels of lesion development, bacterial burden, and almost no mortality. This suggests that there is a correlation between the presence of the 601.59 m/z peak and the level of virulence that is detected.

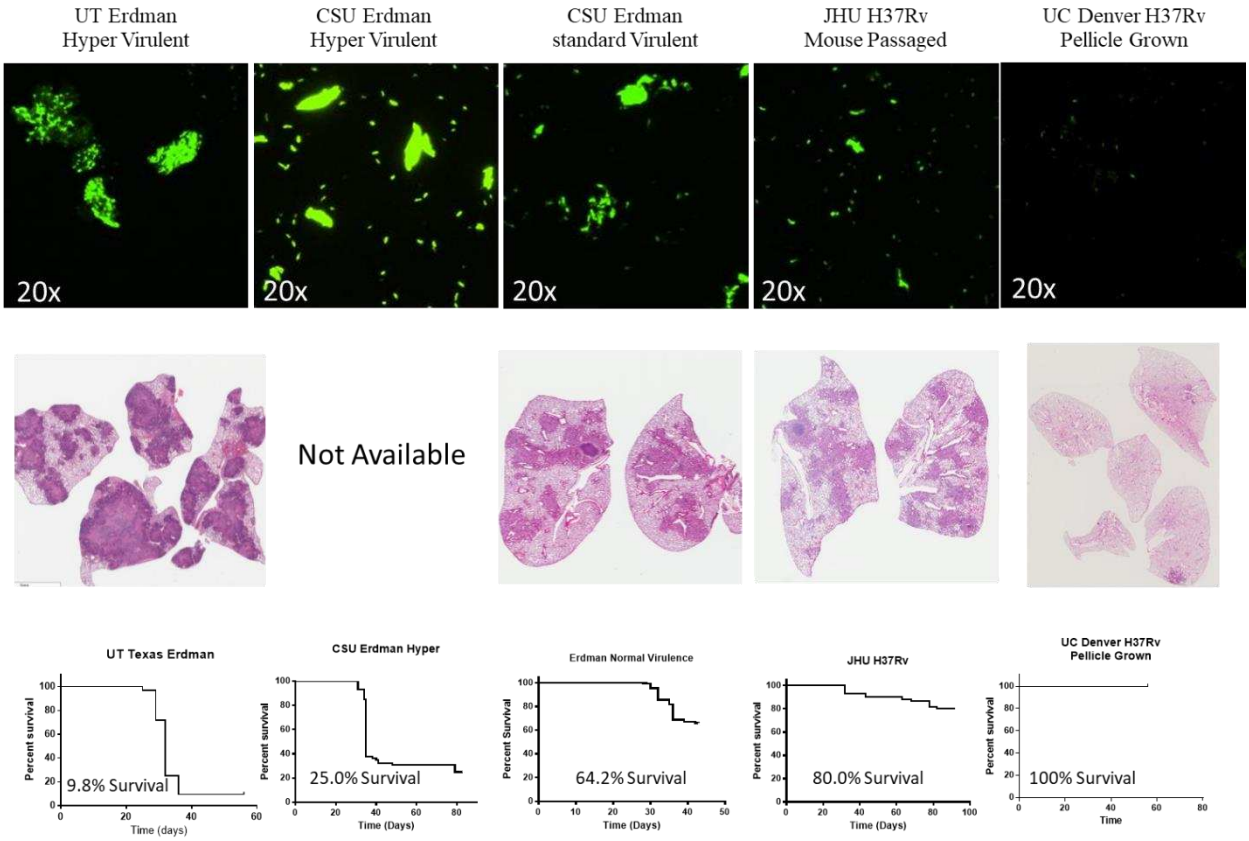


Figure A.5. SYBR Gold stained Mtb strains digital image (1st row), corresponding lesion H&E histology (2nd row), and the mortality over time (3rd row). The most virulent strain of Mtb is placed on the 1st column, while the least virulent strain is on the last column.

We next investigated if there was a correlation between virulence and the level of Mtb aggregation observed at the tie of infection. 6 to 8 week female C3HeB/FeJ mice were infected

with either a batch of hypervirulent *Mtb Erdman* from the University of Texas (Shenan Strain), a CSU working *Mtb Erdman* culture, an *Mtb H37Rv* mouse passaged culture from John Hopkins University, and an *Mtb H37Rv* strain that was grown as a pellicle from University of Colorado. Data for the CSU hypervirulent culture was used from the previous infections. When comparing the pathology and mortality of the infected mice, a trend emerges that more aggregated bacteria exhibit higher levels of lesion involvement and mortality (Figure A.5).

A.3.3 Loss Of Virulence In Long Term Storage

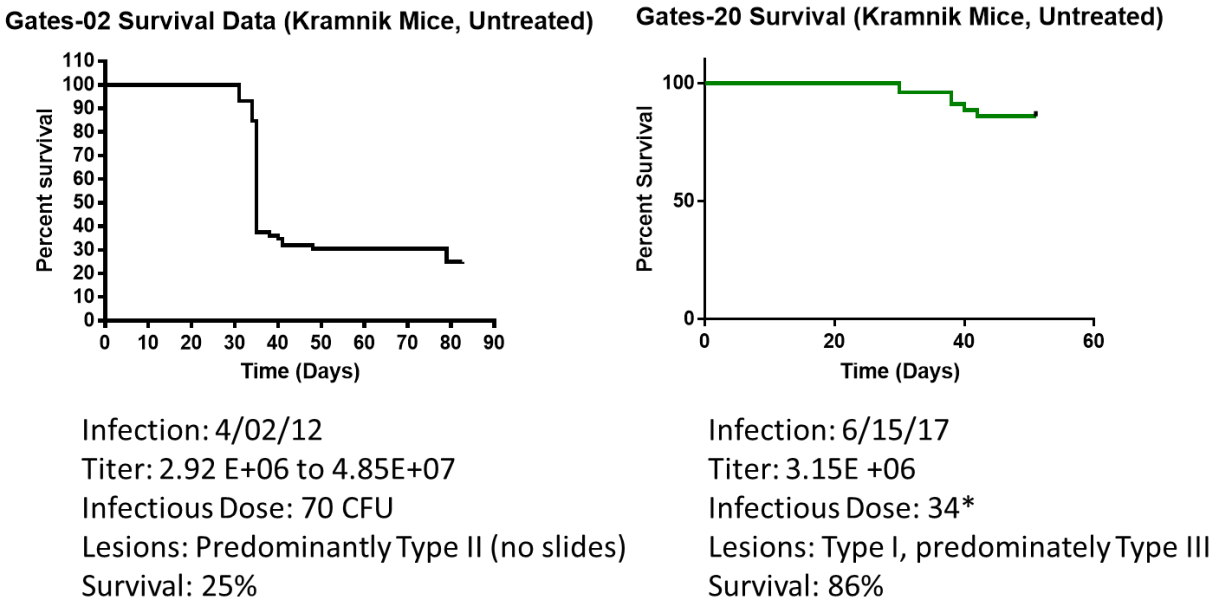


Figure A.6. The same batch of Hyper Virulent Erdman Mtb with decreased levels of virulence over time. Mortality graphs (top row) and the infection details (bottom row) show a demonstrated loss of virulence over a 5 year period within a -80℃ freezer.

After the identification of trends in both bacterial aggregation and the presence of the 601.59 m/z molecular feature being correlated with various levels of virulence, an additional mouse infection was performed. Despite previous results from three different studies, using the exact strain, the hypervirulent Mtb culture appeared to lose its virulence (Figure A.6). No studies

have been published on this effect before, but it has been previously observed in our laboratory that Mtb cultures stored long term in cold storage (-80℃) will lose their virulence. Taking this information into account, we investigated both the level of aggregation and the presence of the 601.59 m/z peak. No visual phenotypic changes were observed in the levels of aggregation between the two Mtb cultures, but there was a unique decrease in the relative abundance of the 601.59 m/z peak over time. This is further confirmation of a possible correlation between using the 601.59 m/z as an indicator of virulence but ultimately meant that there were no more samples in which to continue our testing.

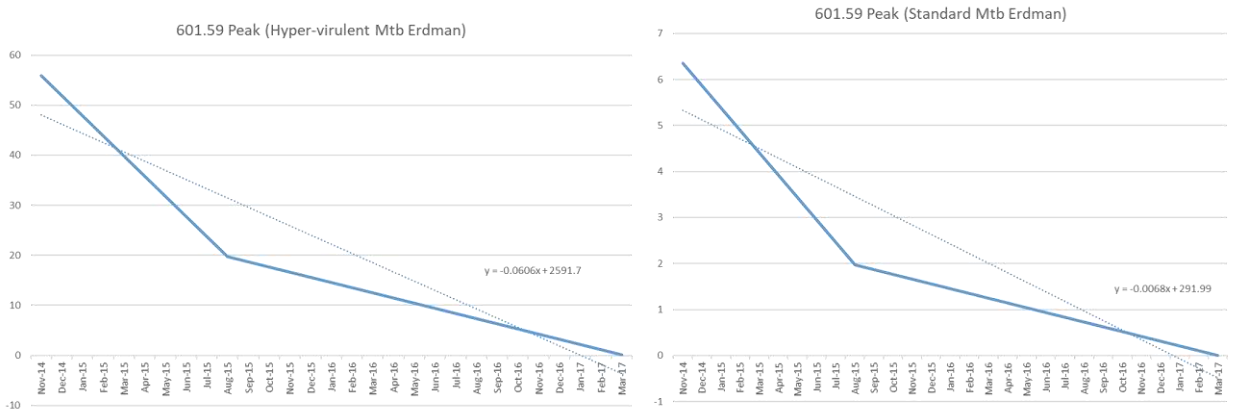


Figure A.7. The globally normalized 601.59 m/z molecular with relative abundance on the y-axis and the time in months over the x-axis. A decline in relative abundance is observed over a three year period.

A.3.4 – Attempt To Induce Hyper Virulence

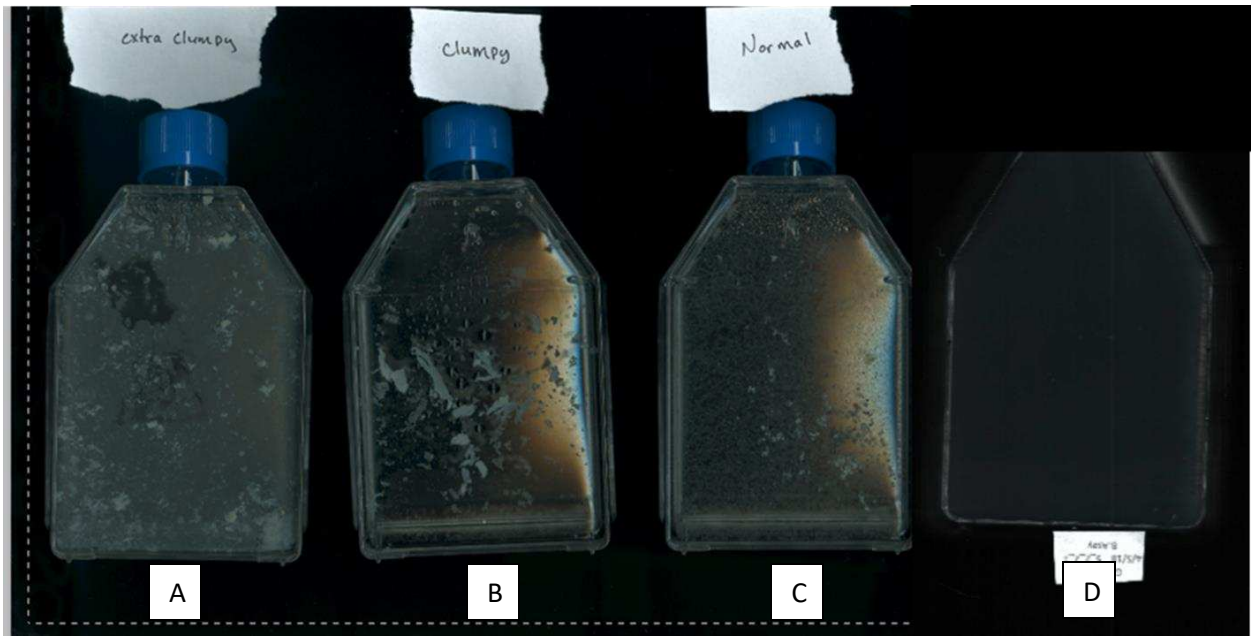


Figure A.8. Phenotypic variations in the biological replicates were observed during sample culturing. Phenotypes include (A) a highly clumped pellicle, (B) corded like pellicles, (C) standard pellicle formation, (D), and cells that were entirely planktonic with no visible aggregation. After furthering passaging, all phenotypes converged to the standard pellicle formation.

Due to the loss of virulence during long term storage to continue our research, we needed to induce hypervirulence again in our working stock of Mtb. Several deviations from the normal culturing protocol had been identified that might have contributed to the increase in virulence from the working stock. This included the initial streaking of bacilli on a 7H11/ADC plate before being transferred to a liquid 7H9/ADC culture tube and a substantial increase in Tween 80 at the final step. It was hypothesized that Mycobacterium tuberculosis switched between different media types will display higher levels of mortality and Type II lesion involvement than bacteria kept in PB media. During culturing, we had further unexpected results that required further investigation. To help capture the increase in virulence, three biological replicates were kept at each stage of culture propagation. Instead of a single phenotype that was comparable to the other replicates, there was instead four different phenotypes that were observed (Figure A.8). The first was a

densely clumped culture that had a thick film on the surface (originally labeled 'extra clumpy'). The second was the development of large filamentous rafts that had comparable macroscopic features similar to what is observed microscopically with cording (originally labeled 'clumpy'). The third was the development of individual rafts that were of moderate size and is what is expected during normal culture propagation (originally labeled 'normal'). The fourth was a completely planktonic culture that exhibited a refractive sheen similar to what is observed in an oil spill (originally labeled 'planktonic'). By the third passage within the PB culture flask, all three cultures had converted over to the third phenotype (normal).

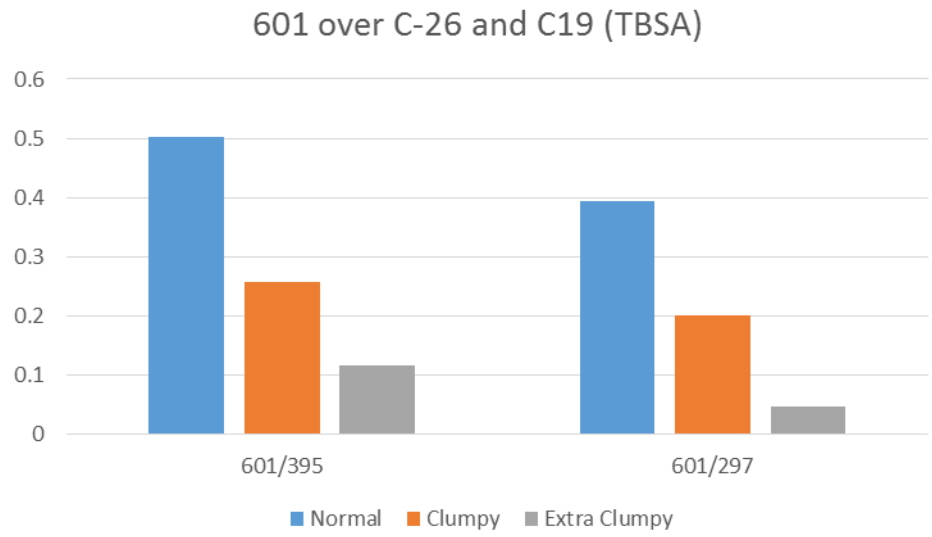


Figure A.9. Globally normalized relative abundance of the 601.59 m/z molecular feature in three of the different cultured phenotypes. In contrast to what was initially observed with the hypervirulent strain, it does not seem that the molecular features of interest are correlated with Mtb aggregation or cording.

LC/MS analysis of three observed phenotypes indicates that there are observable differences in the relative abundance of the 601.59 m/z peak, but it is the opposite of what had been previously observed (Figure A.9). More research is needed to investigate the correlation of virulence to bacterial aggregation and the class of uncharacterized late series FFA.

A.4 Discussion

Our laboratory can generate consistent disease presentation within the C3HeB/FeJ mouse model. Because the mouse strain develops lesions similar to human pathology, it is increasingly being used as a means to score the virulence of various Mtb strains. This may be an issue because, through personal communications with other laboratories, they are observing results that vary from what is generated in our laboratory. The aim of this project was to determine what variables influence the virulence of the Mtb strain used and the host response during an infection. Even though we consistently get reproducible results with our *Mtb Erdman* strain, we did have a cultured batch that became extremely virulent, not only the C3HeB/FeJ mouse model but also the more resistant BALB/C mice. This provided us an opportunity to help standardize the TB C3HeB/FeJ murine animal modal for use in other laboratories.

Under initial investigation, it was observed that mycobacterial strains that exhibited higher levels of virulence had increased quantities of cording/aggregation and levels of the molecular feature 601.59 m/z. This virulence phenotype is unstable and can be modulated by varying the propagation methods, route of infection, and as discovered recently, even the amount of time in being in cold storage. With the loss of the hypervirulent Mtb culture, further work is needed to move forward. The first project will be to more stringently investigate how propagation methods impact the level of virulence within a specific strain. A proposed experiment would be to use three strains of Mtb (Erdman, H37Rv, HN878) and measure how culturing bacteria through mouse passaging, pellicle passaging, and under stress impacts the virulence within the C3HeB/FeJ mice model. The second would be to determine the molecular structure of the 601.59m/z molecular feature and determine if the correlation is causative, indirect, or coincidental. The third would be to investigate further the four phenotypes that were observed during our attempt to recreate the virulent strain of Mtb that was created previously. This includes performing an aerosol infection to observe how even under identical culturing methodologies, how phenotypic variations that emerge can influence the host response. In conclusion, knowing about how even under identical

culturing methods multiple different Mtb phenotypes can form, virulence decreases with time, and how a possible new class of FFA can be used as an indicator for virulence are just the first stepping stones in standardizing the TB C3HeB/FeJ murine animal model.

LIST OF ABBREVIATIONS

TB - Tuberculosis

LIRA – Lesion Image Recognition and Analysis

MIA – Mycobacterial Image Analysis

CFU - Colony Forming Units

WHO - World Health Organization

SDG - Sustainable Development Goals

HIV - Human Immunodeficiency Virus

INH - Isoniazid

RIF- Rifampicin

EMB - Ethambutol

PZA – Pyrazinamide

BDQ – Bedaquilline

Mtb – *Mycobacterium tuberculosis*

XDR TB – Extensively Drug-Resistant Tuberculosis

MIC - Minimum Inhibitory Concentration

MBC - Minimum Bactericidal Concentration

AI – Artificial Intelligence

LTBI - Latent Tuberculosis Infection

18F-FDG - 2-[18F]fluoro-2-deoxyglucose

NHP - Non-Human Primate Model

CM - Cynomolgus macaques

I/St - I/StSnEgYCit Mouse Strain

IPR1 - Intracellular Pathogen Resistance 1

Ifi75 - Interferon-Inducible-75

LDA - Low-Dose Aerosol

CV – Computer Vision

ROI – Region of Interest

ML – Machine Learning

CNN – Convolutional Neural Networks

GPU - Graphical Processing Units

CPU - Central Processing Units

SST1 - Super Susceptibility To Tuberculosis -1 Locus

PET - Positron Emission Tomography

CT - Computed Tomography

SVM - Support Vector Machines

PB - Proskauer-Beck Medium

PBS - Phosphate Buffered Saline Solution

MIP - Microbiology, Immunology, and Pathology

H&E - Haemotoxylin and Eosin

Misc. – Miscellaneous

CNN1 – LIRAs Type I Macro-Classifier

RIPC - Raw Image Patch Counts

CNN2 – LIRAs Type I Micro-Classifier

CNN3 – LIRAs Non-Type I Micro-Classifier

GANs - Generative Adversarial Networks

MALDI - Matrix Assisted Laser Desorption Ionization

RPT - Rifapentine

MDR TB - Multidrug-Resistant TB

ECDC - European Centre for Disease Prevention and Control

LK – Losartan

GRA - Graduate Research Assistant

RA – Research Associate

Fe – Iron

AR - Auramine-Rhodamine

IF - in situ Hybridization

GUI - General User Interface

NETs - Neutrophil Extracellular Traps

RNA - Ribonucleic Acid

TDM - trehalose-6,6′-dimycolate

TMM - trehalose-6,6′-monomycolate

TLC - Thin Layer Chromatography

LC-MS - Liquid Chromatography–Mass Spectrometry

FFA - Free Fatty Acids