Beyond Mutation: Epigenetic Drivers of Phenotypic Diversity and Survival in Mycobacteria

Beyond Mutation: Epigenetic Drivers of Phenotypic Diversity and Survival in Mycobacteria

A dissertation presented by

Alexandra Joelle Sakatos

to

The Committee on Biological Sciences in Public Health

In partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in the subject of
Biological Sciences in Public Health

Harvard University
Cambridge, Massachusetts

April 2016

Professor Sarah M. Fortune Alexandra Joelle Sakatos

Beyond Mutation: Epigenetic Drivers of Phenotypic Diversity and Survival in Mycobacteria

ABSTRACT

M. tuberculosis is the causative agent of a global health epidemic that kills ~1.5 million people each year. The outcomes of infection with Mtb are highly variable. Although many patients are able to control the infection in a quiescent state, others develop active disease. Furthermore, the progression of TB lesions has been shown to vary within a single individual. This phenotypic variability in the infecting population of Mtb may be responsible for the high rate of treatment failures, which can exceed 20% in some endemic areas (World Health Organization, 2015). Although genetic mutation can drive a portion of the observed phenotypic variability, mutation rates in mycobacteria are exceedingly low. Epigenetic factors are therefore likely to be responsible for the majority of observed diversity in infection and treatment outcomes.

Here, we investigated epigenetic drivers of phenotypic variability and survival in mycobacteria. We found evidence of high rates of phenotypic variability in response to drug treatment of *M. smegmatis*, a non-pathogenic, model for TB. Specifically, we found that two distinct subpopulations are able to grow in the presence of drug. These subpopulations exhibited heritability of their transcriptional profiles, growth properties, and ability to grow on drug across generations. We next found that hupB, a histone-like protein, is critical for the formation of these epigenetically regulated subpopulations. We also show that hupB regulates gene expression and is post-translationally modified, and that modification of hupB may drive the formation of one of these phenotypically drug-resistant subpopulations. These findings suggest that modification of a histone-like protein may drive epigenetic inheritance and phenotypic

variability in mycobacteria, which allows it to withstand antibiotic treatment. Finally, we also investigated alternative post-transcriptional mechanisms of hupB regulation.

TABLE OF CONTENTS

Title Page..... i
Copyright Page ii
Abstract iii
Table of Contents..... v
Dedication..... vii
Acknowledgements..... viii
List of Figures and Tables ix

Chapter One – Introduction..... 1

1.1 The Global Burden of tuberculosis..... 1

1.2 Treatment of TB 1

1.3 Phenotypic Diversity in TB infection..... 2

1.4 Phenotypic Diversity in response to antibiotics.....3

1.5 Molecular mechanisms driving phenotypic diversity.....4

1.6 Overview of chapters..... 6

Chapter Two – Identification of epigenetically regulated, drug resistant phenotypic variants in mycobacteria

2.1 Abstract..... 8

2.2 Introduction 9

2.3 Results 11

2.4 Discussion..... 21

2.5 Materials and Methods..... 24

Chapter Three – Investigating the role of HupB and HupB modification sites in the formation of phenotypic variants in mycobacteria
2.1 Abstract..... 31

2.2 Introduction 32

2.3 Results 34

2.4 Discussion..... 46

2.5 Materials and Methods..... 48

Chapter Four – Alternate Forms of HupB regulation

4.1 Abstract..... 51

4.2 Introduction 52

4.3 Results 55

4.4 Discussion..... 67

4.5 Materials and Methods..... 69

Chapter Five – Conclusion 71

Appendix 1 76

Appendix 2 83

References 86

DEDICATION

I dedicate this to my parents. Thank you for your infinite love and support.

ACKNOWLEDGEMENTS

I am grateful for the numerous individuals who have provided support, encouragement, and companionship throughout this process. I would first like to thank Dr. Sarah Fortune for her unending enthusiasm for science, which has served as a constant source of support and encouragement throughout this experience. To former and present members of the Fortune Lab, thank you for making the lab such a supportive, collaborative, and fun environment in which to work. To my fellow BPHers, I feel so lucky to have started this program with such a tremendous group of people. You have provided a tremendous source of support, fun, and laughter throughout graduate school. I couldn't imagine going through this experience without all of you. To all of my other friends who have provided valuable companionship and support through the years, thank you. To my family – thank you for your infinite love and support in all my endeavors. And to Peter, thank you for the love, support, and humor that you have never failed to provide.

LIST OF FIGURES AND TABLES

Chapter One
Figure 1.1Bactericidal activities of antibiotics within the first 14 days of treatment.
Figure 1.2The coexistence of sterile and active lesions within a single macaque

Chapter Two
Figure 2.1Identification of mycobacterial subpopulations that grow in the presence of drug.
Figure 2.2 Alternative models for small colony formations
Figure 2.3 Variants have distinct growth properties
Figure 2.4 Phenotypic drug resistance is heritable, semi-stable
Table 2.1 Whole-genome sequencing results
Figure 2.5 Variants have distinct transcriptional signatures

Chapter Three
Figure 3.1 Loss of HupB results in significant reduction of colony formation in the presence of INH and alters gene expression.
Figure 3.2 HupB is modified at lysine and arginine residues
Figure 3.3 Assessment of HupB protein levels and phenotype on INH following mutation of HupB modification sites
Figure 3.4 K86R mutation does not alter DNA binding affinity.
Figure 3.5 Growth curves
Figure 3.6 K86R mutation leads to specific loss of small colony variants
Figure 3.7 Differences in HupB K86 methylation were not detected between small and large colonies
Figure 3.8 Role of methylation in small colony formation

Chapter Four
Figure 4.1 HupB lysines are modified
Figure 4.2 Mutation of K3 to arginine or glutamine results in significantly reduced tolerance to A)isoniazid and B) rifampicin
Figure 4.3 Mutation of K3 lowers HupB protein levels.
Figure 4.4 Mutation of K3 lowers HupB transcript levels**.**
Figure 4.5 Opposing models for hupB regulation
Fiure 4.6 Protein levels of hupB K3R from the tet promoter are lowered relative to wild-type
Figure 4.7 Native HupB transcript is upregulated in the tet-K3R strain relative to wild-type.
Figure 4.8 5'end mapping reveals processing sites in the *hupB* transcript.
Figure 4.9 3'RACE reveals the presence of a truncated *hupB* transcript encoding only the N-terminal HU domain.
Figure 4.10 Expression of the 3'end of the *hupB* transcript is reduced by ~30% relative to the 5'end
Figure 4.11 Northern blot of hupB
Figure 4.12 Western blot of flag-tagged hupB

Appendix I

Figure 6.1. Expression of the 5'UTR and the ORF of narG, Rv1161
Figure 6.2. Ratio of 5'UTR expression relative to the downstream ORF for narG in normal vs. hypoxic conditions.

Appendix 2

Figure 7.1 Representative volcano plots for High vs Low iron and High vs Low pH.
Figure 7.2. Visualization of 5'ends by RNAseq aligns with 5' end mapping data

CHAPTER ONE: INTRODUCTION

1.1 The Global burden of tuberculosis

Tuberculosis is a chronic respiratory illness that kills approximately 1.5 million people per year, causing the most fatalities of any infectious disease worldwide. Infection with TB occurs when *M. tuberculosis*, the causative agent, is inhaled and taken up by alveolar macrophages. Immune cells aggregate around the infected macrophages, which forms a granuloma, a lesion which can block the airway and produce the primary symptom of TB, the chronic cough. The majority of those infected with TB are able to effectively control the infection, while about ten percent of them develop active disease. It has been estimated that nearly a third of the world's population is latently infected with Mtb. Immunocompromising factors such as HIV can accelerate the risk of reactivation of these latent infections.

1.2 Treatment of TB

Treatment of TB involves a four drug regimen that is taken over a period of up to nine months. The long treatment time is designed to reduce the risk of drug resistance and sterilize the infection. Still, incomplete treatment is a common source of treatment failure, and even in successfully treated cases, recurrent infections are common and may occur in up to ~30 percent of patients (Cox et al., 2006). Among recurrent infections, roughly half are due to relapse of the primary infection, rather than reinfection with a different strain (Bryant et al., 2013), Richa Gawande, unpublished data). Notably, relapse after completed treatment can occur in the absence of an underlying genetic mutation, implying that epigenetic factors may drive treatment failure in a portion of cases. Furthermore, biphasic killing of Mtb in patients in response to antibiotic treatment has been observed, suggesting that subpopulations of drug-tolerant bacteria may cause treatment failures (Figure 1.1, (Jindani, Doré, & Mitchison, 2003; Walter et al., 2015).

Understanding the factors that drive these variable infection and treatment outcomes is crucial to improving the control of this epidemic.

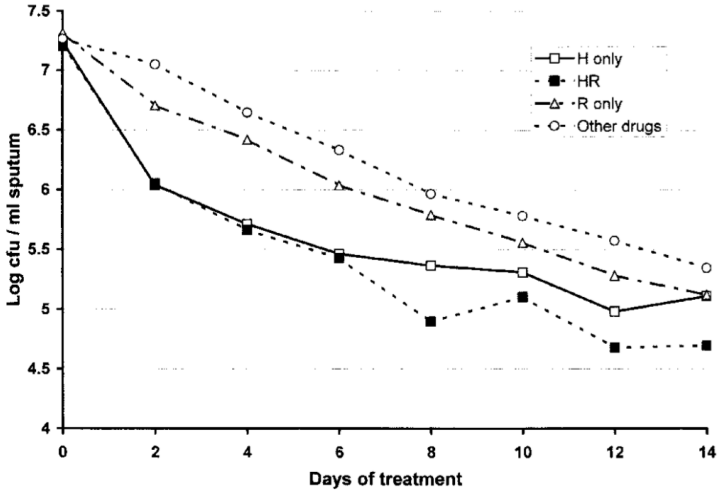


Figure 1.1. (Jindani et al., 2003) **Bactericidal activities of antibiotics within the first 14 days following treatment.** African patients with previously untreated, smear-positive, drug-sensitive TB were studied. Drugs were given daily and sputum was collected at indicated time points following the start of treatment. H= isoniazid only, R- rifampin only, HR – isoniazid and rifampin.

1.3 Phenotypic diversity in TB infection

Mtb has a limited capacity to create genetic diversity. Unlike many other bacteria, Mtb does not possess the machinery to undergo horizontal gene transfer. The mutation rate of Mtb is also exceedingly low, with estimates hovering around 2.4 x 10^{-10} mutations per base pair per generation. Still, the extent of phenotypic variability in Mtb has recently begun to be appreciated. A recent study elucidated the role of this variability in driving variable infection outcomes. Infection of a macaque with a clonal population of Mtb will result in both sterile and active granuloma lesions within a single animal (Figure 1.2, (Lin et al., 2014). These lesions can be founded by just one or a few bacteria, suggesting that diversity in the infecting bacterial population may drive the variable lesion outcomes within a person, and perhaps also the variable clinical outcomes observed between individuals and in response to antibiotic treatment.

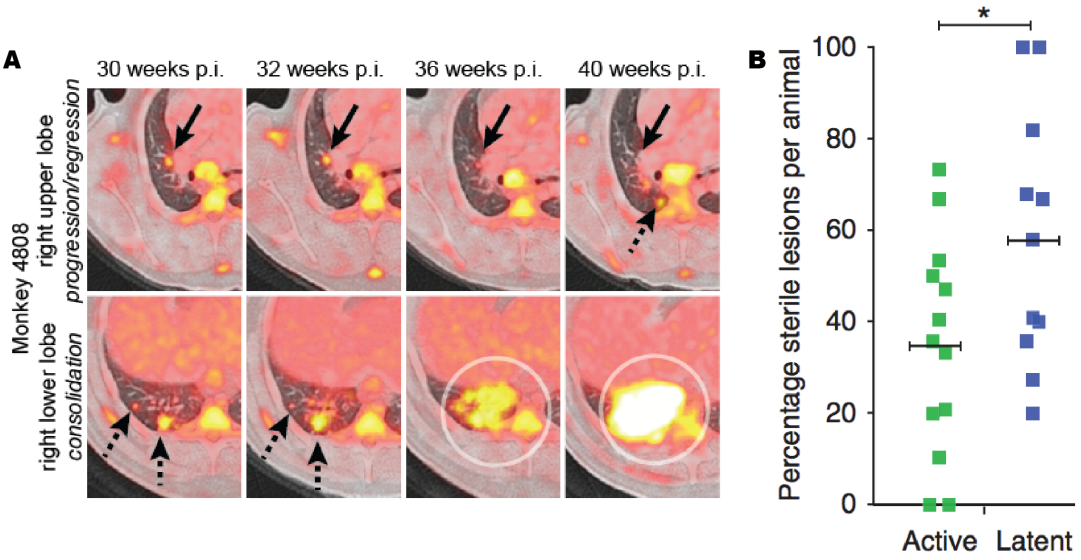


Figure 1.2. The coexistence of sterile and active lesions within a single macaque (Lin et al., 2014)**. A)** Tracking of 2 lesions over time within a single monkey. **B)** Percentage sterile lesions per animal – monkeys with both clinically active and latent disease exhibit lesions with a wide range of sterilizing activity.

1.4 Phenotypic diversity in response to antibiotics

Exposing bacterial populations to antibiotics *in vitro* frequently reveals the extent of phenotypic diversity in bacterial populations. Distinct subpopulations with phenotypes beneficial for survival can be selected for in the presence of drug. The first example of phenotypic diversity influencing the survival of bacteria in the presence of drug was described by Joseph Bigger in 1944. Bigger observed that upon exposure of a clonal population of bacteria to high concentrations of antibiotic, a fraction of cells would survive treatment. These cells, which he termed persisters, did not posses high-level genetic resistance, as evidenced by the fact that upon regrowth in media without drug and re-exposure, the progeny of the persisters exhibited the same susceptibility to drug as the starting population. Bigger hypothesized that that these cells were a subpopulation of non-dividing cells that could withstand antibiotic treatment due to their

halted growth. Sixty years later, Bigger's hypotheses were validated. Using high persistence forming strains of *E.coli*, individual persister cells were imaged before and after antibiotic treatment. Persisters were shown to be generated from both non-dividing and slowly growing cells that were pre-existing in the population (Balaban, Merrin, Chait, Kowalik, & Leibler, 2004).

Since the discovery of persisters, more recent studies have shown that a variety of phenotypically distinct subpopulations with differential susceptibilities to antibiotics can exist within a clonal bacterial population(Dhar & McKinney, 2007; Kester & Fortune, 2014). Although persisters generally represent an antibiotic-tolerant subpopulation with slow or nonexistent growth, a wide spectrum of growth rate heterogeneity in bacterial populations has since been demonstrated. In mycobacteria, using an rRNA reporter to report on the relative growth rates of individual cells, Manina et al showed that there is variability in growth rates across single cells under standard conditions and that this variability increases in the presence of drug and after host infection(Manina, Dhar, & McKinney, 2015). Actively growing persister cells were also identified in mycobacteria, suggesting that antibiotic tolerant subpopulations could be drawn from cells with a range of growth rates (Wakamoto et al., 2013).

1.5 Molecular mechanisms driving phenotypic diversity

Although phenotypic diversity in bacterial populations has been widely described, the known molecular mechanisms underpinning the formation and maintenance of these subpopulations are still scarce. The finding that mycobacteria grow and divide asymmetrically offered a significant advance to our understanding of how distinct cell types with differential antibiotic susceptibilities could arise from a clonal population (Aldridge et al., 2012). Asymmetric growth produced faster-growing cells, termed accelerators, which were more

susceptible to cell-wall acting drugs, and slower growing cells, which were more susceptible to rifampicin, a transcription targeting drug (Aldridge et al., 2012).

Differential gene expression is a common strategy that can also create phenotypic variants with different drug susceptibilities. HipA, a toxin that alters bacterial growth, was identified as a crucial mediator of persistence in a screen for high-persistence mutants. Later studies showed that variable expression of hipA in subsets of cells may underlie their persister phenotype. Binding of the hipA toxin to its cognate antitoxin can also vary across individual cells. The nonlinear binding kinetics of toxin/antitoxin interactions could result in threshold based amplification of variability in the toxin phenotypes across cells (Balaban, 2011). Variable expression of other genes, including virulence factors and efflux pumps, has also been linked to altered drug susceptibilities in subsets of cells(Adams et al., 2011; Arnoldini et al., 2014; Sánchez-Romero & Casadesús, 2014). In mycobacteria, stochastic pulsing of katG correlated with survival of single cells in the presence of INH (Wakamoto et al., 2013), raising the intriguing possibility that gene expression variability at relatively short time scales could drive differences in drug susceptibilities.

Bistable expression of genes in a bacterial population can be established and maintained by a combination of noise and positive feedback loops. Noisy gene expression creates cells with different expression levels of a certain gene. When gene expression passes a certain threshold in a given cell, a positive feedback loop can be initiated, where the gene amplifies its own expression. This positive feedback results in at least two subpopulations of cells, with high and low expression of a given gene. Positive feedback loops can be maintained across generations, resulting in a heritable, bistable phenotype. Classic examples of bistability regulated by positive feedback loops are competence development in Bacillus subtilis and induction of the lac operon

in *E. coli*. More recently, expression of relA, the mycobacterial stringent response regulator, was also shown to be bimodally distributed and governed by noise and a positive feedback loop(Sureka et al., 2008). Growth rate feedback onto gene expression can also promote and maintain distinct subpopulations of cells (Kiviet et al., 2014). Finally, DNA methylation can drive differential gene expression in subsets of cells, as in the *pap* switch in *E. coli.*

1.6 Overview of chapters

In this dissertation we investigated epigenetic drivers of phenotypic variation and survival in mycobacteria. In **Chapter 2**, we imaged *M.smegmatis* colonies growing in the presence of drug and find that two distinct colony types appear in the presence of drug. These colony types consisted of two distinct subpopulations of cells with different transcriptional signatures and growth properties that were heritable across generations. Furthermore, these cells were heritably drug resistant, such that they retained their increased ability to grow on drug even after expansion in media without drug, suggesting that they comprise a fundamentally distinct population from the canonically described persister cells. **In Chapter 3,** we found that a nucleoid associated protein, hupB, may drive the formation of both populations of epigenetically regulated drug resistant variants. We also investigate the function of posttranslational modification of hupB and find that mutation of a hupB methylation site specifically abrogates the formation of one of these subpopulations. This finding suggests that bacteria may be able to heritably alter their nucleoid structure to affect gene expression in subpopulations cells and thereby create cells with distinct phenoytpes, akin to the use of histone modification in eukaryotes. Lastly, in **Chapter 4**, we investigate alternate mechanisms by which hupB function might be regulated in response to fluctuating environments.

CHAPTER TWO

Identification of epigenetically regulated, drug resistant phenotypic variants in mycobacteria

Authors: Alexandra Sakatos[1], Sarah Fortune[1]

[1] Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA

Author Contributions A.S. and **S.F.** designed the experiments. **A.S.** conducted the experiments. **A.S.** and **S.F**. drafted the manuscript.

2.1 Abstract

Phenotypic variants of bacterial cells within a clonal population typically persist upon exposure to antibiotics despite the sensitivity of their genetically identical brethren. The persistence of these subpopulations may be responsible for the slow eradication of bacterial infections such as Mtb in response to antibiotic treatment. A hallmark of these non-genetically determined survivors of drug exposure has historically been the lack of stable inheritance of the drug resistance phenotype across multiple generations. Here, using high-throughput quantitative plating and dynamic imaging, we identified novel mycobacterial subpopulations that exhibit heritable, epigenetically determined drug resistance which allows them to grow in the presence of antibiotic. We show that at least two distinct epigenetically regulated lineages of bacteria thrive in the presence of drug. These two subpopulations have distinct growth rates and transcriptional signatures that are heritable across generations.

2.2 Introduction

Phenotypic diversity in bacterial populations is a key barrier in the design of effective antibiotics. Antibiotic treatment failure has historically been attributed to a subpopulation of nongrowing cells, termed persisters. However, more recent studies have identified subsets of cells that continue to divide in the presence of antibiotic, and these actively growing subsets of cells may in fact be responsible for the slow eradication of bacterial infections after antibiotic therapy ((Adams et al., 2011; Claudi et al., 2014; Wakamoto et al., 2013))(Walter et al., 2015). A hallmark of these phenotypic variants that withstand drug treatment has historically been the lack of stable inheritance of the drug resistance trait. Upon re-growth in media without drug, the resulting progeny are killed with the same kinetics as the starting population. Nonetheless, dynamic persistence appeared to track between sister cells, suggesting that epigenetic inheritance may play a role in the survival of these actively growing subpopulations(Wakamoto et al., 2013). This growing body of evidence suggests that the subpopulations of cells that survive in the presence of drug may be more diverse and dynamic than originally thought. Understanding the subpopulations of cells that can withstand antibiotic treatment is crucial to improving the design of future therapies.

Non-growing persisters have been challenging to enrich and characterize due to their rarity in the populations and their inability to divide in the presence of drug. In contrast, actively growing, phenotypically drug resistant subpopulations of cells could theoretically be selected for and enriched in the presence of drug. We employed an adaptation of a previously established high-throughput quantitave dynamic imaging technique to identify and characterize mycobacterial subpopulations that grow in the presence of drug (Levin-Reisman et al., 2010). Using this technique, we monitored growth of *M. smegmatis* in the presence of a range of

isoniazid (INH) concentrations. We found that the numbers of colonies formed in the presence of drug dropped sharply at the beginning of the concentration range and then fell more slowly at increasing concentrations, consistent with the selection of drug resistant subpopulations. We characterized these drug resistant subpopulations and found that at least two distinct subpopulations grow in the presence of drug. These two subpopulations, termed large and small variants, have distinct transcriptional signatures and growth properties. Strikingly, these subpopulations of cells exhibited stable epigenetic inheritance of their growth properties and drug resistance across generations.

2.3 Results

Identification of mycobacterial subpopulations that grow in the presence of drug

To identify subpopulations of mycobacteria that grow in the presence of drug, we used an adaptation of the ScanLag technique, established by Nathalie Balaban, which combines plating with high-throughput dynamic imaging to identify rare bacterial growth variants(Levin-Reisman et al., 2010). We grew an isogenic population *M. Smegmatis* in rich media without drug, and then plated aliquots of the cultured cells on agar plates containing an increasing gradient of drug concentrations (Fig 1A). We reasoned that if, in the presence of drug, the entire population were growing slowly, all the cells plated would form colonies slowly at the same time. Alternatively, if privileged subpopulations of cells were growing more rapidly, a subset of colonies would come up earlier. We observed the latter- as drug concentration increased, a subset of the cells plated formed colonies, implying that the inhibition of growth observed at the population level at MIC levels of drug is actually due to a mixture of bacterial cells with different growth properties and drug susceptibilities, consistent with what others have found at higher drug concentrations (Fig.1B) (Wakamoto et al, 2013). Surprisingly, we observed that at ~MIC levels, the colonies came up at two distinct time points, suggesting that two distinct cell lineages are preferentially thriving in the presence of drug (Fig1C,D). The bimodal distribution of colony size observed at the final time point further supported this notion (Fig1C,E). The high rate at which these colonies arose (5-10% of the populations) argues against a mutational mechanism for their formation.

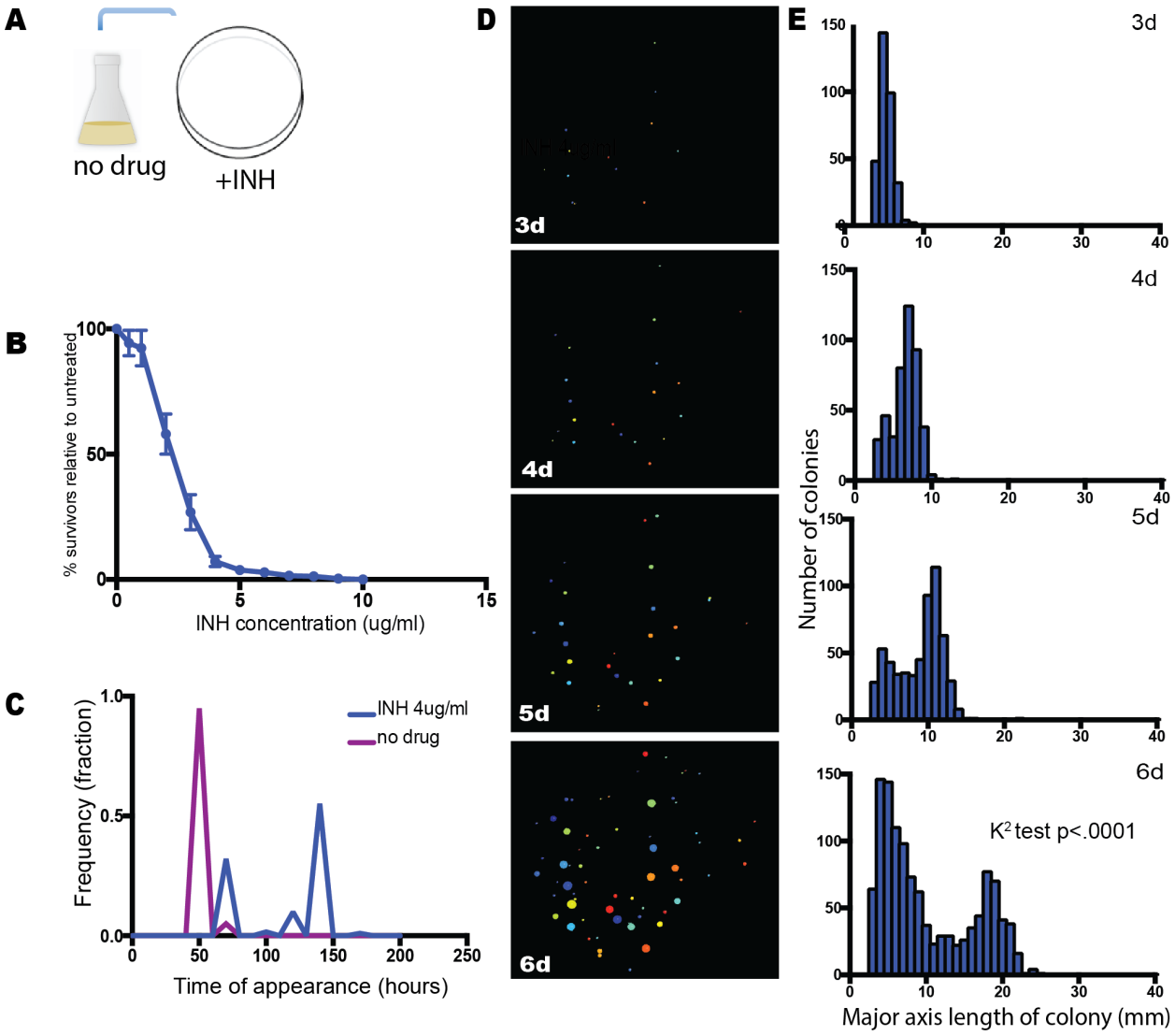


Figure 2.1 Identification of mycobacterial subpopulations that grow in the presence of drug. A) Schematic of experiment **B)** Number of colonies formed after plating on indicated concentrations of INH **C)** Time of initial detection of colonies in the presence and absence of 4 ug/ml INH. **D)** Example series of one plate tracked over time by cell profiler (4ug/ml INH). Plates were run through custom program (Supp fig 1) and colored circles indicate the actual size of the detected colonies. **E)** Histogram of major axis length of colonies for all colonies plated, over time. (4ug/ml INH)

Variants have distinct growth properties

The observation that the small and large colony variants come up at different times and grow to different sizes suggests that either, 1) the founding cells of the different colony types have different lag times, or 2) heritable differences in growth kinetics between the cells making up the colonies exist in the presence of drug (figure 2.2). To distinguish between these hypotheses, we picked individual large and small colonies and replated them on plates without drug. We found that upon initial replating, cells from small colonies produced more small colonies than cells from large colonies, implying that the different colony phenotypes are comprised of distinct subpopulations of cells. To further test this hypothesis directly at a single cell level, we grew cells from large and small colony variants in a microfluidic device with 4ug/ml and monitored their growth over ~48 hours. We found that cells from large colonies were larger and had shorter division times than cells from small colonies.

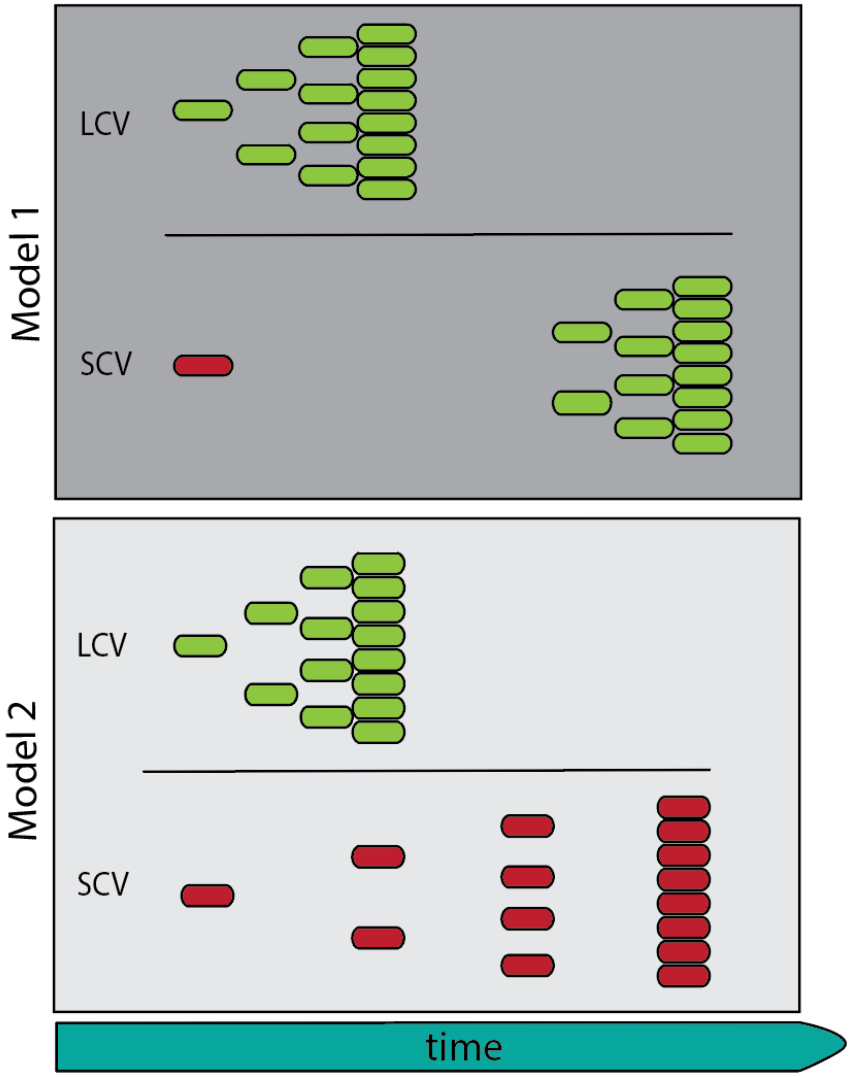


Figure 2.2 Alternative models for small colony formation. In the first model, the founding cells of the SCVs have a longer lag time than those of the LCVs, once they start dividing, the resulting progeny is identical to the LCV. In the second model, the SCVs are comprised of cells that heritably grow slower than the cells in the LCVs.

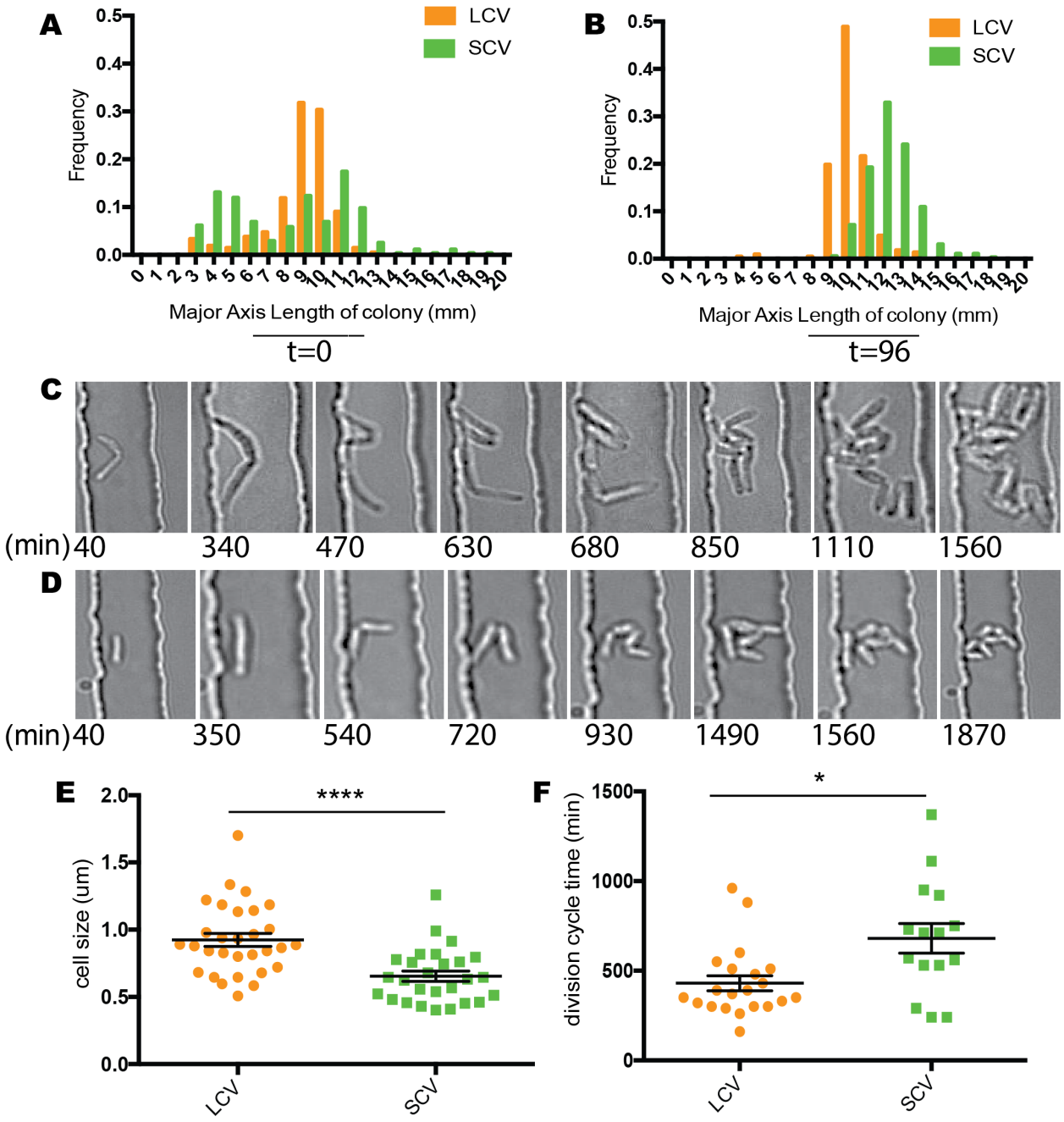


Figure 2.3. Variants have distinct growth properties. A and B) LCVs and SCVs were picked from INH plates, resuspended in 7h9 media and replated onto plates without drug. The histograms of the resulting colony sizes from each colony type after immediate replating **(A)** and after 96 hours in culture **(B)** are shown. **C,D,E, and F)** Large and small colonies were picked from INH plates and cells were loaded into a microfluidic device and imaged for ~48 hours. **C)** Images of one LCV cell over time, **D)** Images of one SCV cell over time **E)** Quantification of cell size at birth **F)** Quantification of time from birth to division for individual cells.

Phenotypic drug resistance is heritable, semi-stable

The rate at which colonies appear on the drug plate is significantly higher (~5-10% of the population) than would be expected given a mutational mechanism. However, the observation of colony-level phenotypes suggests that there is a stable, heritable determinant of drug resistance and growth variants. To determine the stability of the drug resistant phenotype, we picked small and large variants and re-plated them on drug immediately (t=0) and at staggered time points following outgrowth in rich media without drug. We found that phenotypic drug resistance was semi-heritable. Upon immediate re-exposure to antibiotics, the progeny exhibited significantly increased resistance compared to the original population. This increased resistance gradually decreased after regrowth in rich media without drug, implying a semi-heritable mechanism of drug resistance in the presence of sublethal antibiotic (Figure 2.4, *A*). One explanation for the observed stability of the phenotypic drug resistance could be that the cells failed to grow and divide upon reculturing in rich media, thereby allowing the cells to retain their previously acquired adaptation to drug. We show that in fact upon reculturing of the large and small colonies in rich media, they begin growing and dividing within the first 24 hours (Figure 2.4, *B*). We also measured growth at a single cell level after culturing in media without drug found that all cells went through 3-6 division cycles in the first 24 hours, further confirming that lack of growth does not explain the persistence of the drug resistance phenotype (data not shown).

The high rate (5-10% of the population) at which these colonies form in the presence of drug argues against a mutational mechanism for their formation. To determine whether genetic mutations may underlie the colony phenotypes, we picked individual large and small colonies from a single plate and cultured them in liquid media without drug, and then did whole-genome

sequencing on the genomic extracts. We found that one of the three small colony variants had a mutation in a beta-lactamase gene, and two of the three large colony variants had a mutation in a tetR transcriptional regulator. These results suggest that although some of the colonies we see may be genetic mutants, genetic mutation is not a requirement *per se* for the establishment of the small and large colony phenotypes. We also sequenced pools of small and large colonies in triplicate and did not find any mutations that uniquely associated with either small or large colonies.

It is also a possibility that the two large colonies with the same mutation were sister cells. To control for the possibility of isolating sister cells in future experiments, we can plate independent cultures on INH plates and sequence colonies formed from the individual cultures.

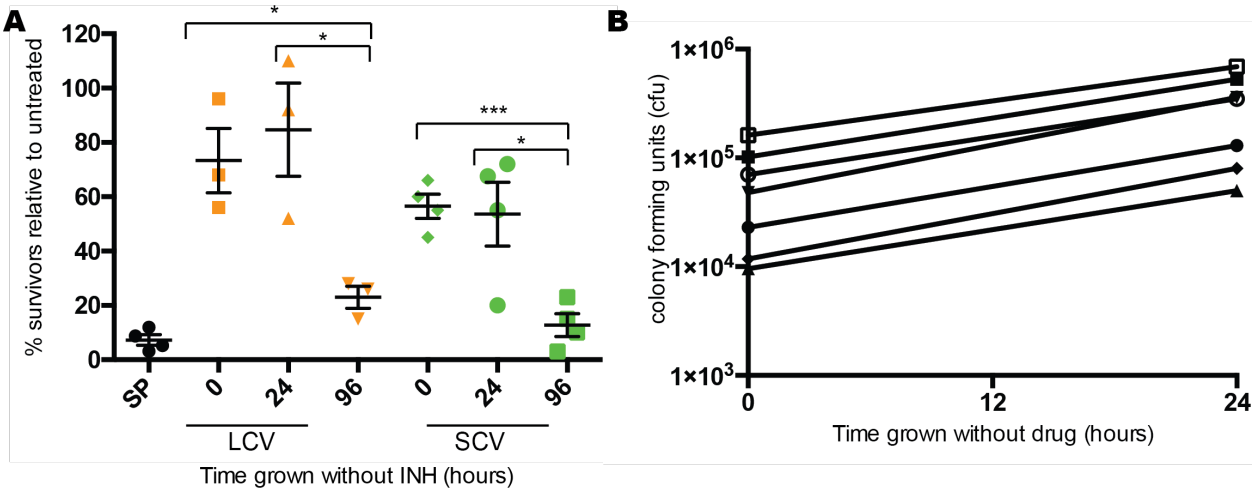


Figure 2.4. Phenotypic drug resistance is heritable, semi-stable. A) Small colonies were picked, resuspended in liquid media and plated immediately (t=0) and then passaged through liquid media without drug and plated at indicated time points. Ratio of colonies on INH plates / plates without INH were measured. SP, starting population **B)** Stable drug resistance after 24 hours is not due to stagnant growth after culturing of small colonies: Cfu counts after reculturing large and small colonies in liquid media.

Colony	SNP	Base pair change	GENE
Large-1	None	na	na
Large-2	801144	G>C	MSMEG_0712
Large-3	801144	G>C	MSMEG_0712
Small-1	2522298	C>T	MSMEG_2439
Small -2	None	na	na
Small -3	None	na	na

Table 2.1. Whole genome sequencing results. Individual small and large colonies were picked in triplicate from INH plates and grown overnight in liquid media. Whole-genome sequencing was performed on genomic extracts.

Variants have distinct transcriptional signatures.

We hypothesized that transcriptional differences may underlie the distinct growth properties of the cells forming large and small colonies. To test this hypothesis, we pooled collections of large and small colonies in duplicate and did RNA-seq to compare transcriptional profiles of the two cell types. We found evidence of extensive transcriptional differences between the large and small colony variants (Figure 2.5). The data suggest that global downregulation of gene expression may drive the slow growth of the small colonies. 533 genes were found to be significantly downregulated ($p<.01$, foldchange>2) in the small colonies, including ~20 genes that were downregulated more than 10 fold. We performed pathway enrichment analysis of the downregulated genes using DAVIDtools, and found that genes involved in transcription regulation were significantly enriched among the downregulated genes, further implicating a general reduction of transcriptional activity with the formation of the small colony variant phenotype (Figure S1).

The existence of widespread transcriptional differences between the small and large colonies suggests that perhaps structural remodeling of the nucleoid is occurring in the distinct variant forms. If structural changes were driving the observed differences in gene expression, we would expect that the chromosomal position of a gene would affect its transcriptional state. In fact, the distribution of differentially expressed genes across the genome appears nonrandom, such that neighboring genes are significantly more likely to be similarly regulated, even after excluding gene pairs that exhibit operonic structure (data not shown). Nucleoid remodeling has been shown to regulate transcription in other bacteria (Kar, Edgar, & Adhya, 2005; Weng & Xiao, 2014). In eukaryotic cells, chromatin states can be inherited across generations, and can drive the formation and maintenance of phenotypically diverse subpopulations. The ability to inherit a particular nucleoid structure across generations would offer a novel mechanism by which bacteria could similarly form and maintain phenotypically distinct subpopulations of cells.

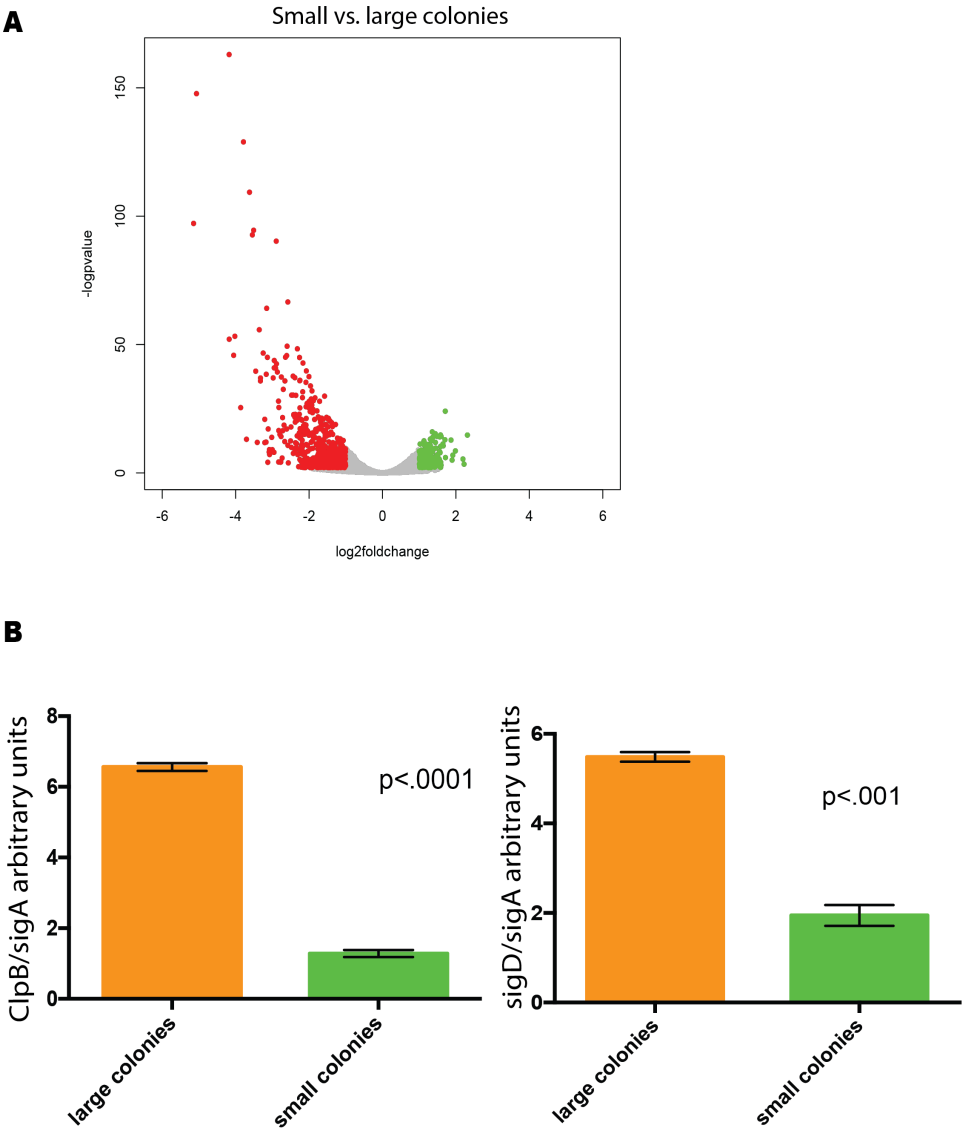


Figure 2.5. Variants have distinct transcriptional signatures. Pools of large and small colonies were collected in duplicate, RNA was extracted from samples and RNA-seq libraries were made and sequenced. **A)** Volcano plot of RNAseq results. **B)** qPCR validation of genes differentially expressed in small and large colonies.

2.4 Discussion

Chronic bacterial infections are notoriously difficult to clear, and epigenetic factors are likely to be responsible for many treatment failures(Bryant et al., 2013). In Mtb, a biphasic response to antibiotic treatment has been observed in patients over time, where roughly 90% of bacteria are killed within the first two days post-treatment, while the last 10% percent of the population exhibits a significantly slower rate of kill (Figure 1.1)((Jindani et al., 2003; Walter et al., 2015). The recalcitrance of this subpopulation to antibiotics is thought to be partially responsible for the extended treatment period for Mtb, which can last from 6-9 months(Jindani et al., 2003). The makeup of this slowly eliminated drug tolerant subpopulation is largely unknown, though transcriptional adaptation has been hypothesized to play a role(Jindani et al., 2003). Here we show that treating *M. smegmatis* with INH results in a biphasic kill curve with similar kinetics. We find that this biphasic killing is due to the selection of two distinct, epigenetically regulated lineages of bacteria, that continue to grow in the presence of drug. Selection of these actively growing lineages *in vivo* could delay the eradication of Mtb infection and increase the likelihood of developing drug resistance.

Variable colony phenotypes have previously been observed in response to stressors in other organisms. Small colony variants (SCVs) in *Staphylococcus aureus* were first discovered in the early 1900s. Interest in them increased in the last couple decades due to growing evidence that SCVs are associated with chronic, persistent infections which are notoriously difficult to treat(Kahl, 2014; Proctor et al., 2006). SCVs have since been identified in other bacteria, including *Pseudomonas aeruginosa*. Still, the term SCV is descriptive, and the mechanisms underlying their formation are diverse. A variety of *in vitro* and *in vivo* stress conditions have been shown to create the small colony phenotype, ranging from intracellular infection to

exposure to cold temperatures and low pH(Leimer et al., 2016; Mirani, Aziz, & Khan, 2015; Tuchscherr et al., 2011). The SCVs that have been identified under these array of conditions have fallen into two broad classes : 1) stable SCVs that do not revert to the normal colony phenotype upon subculture, and are caused by an underlying genetic mutation, and 2) nonstable SCVs that rapidly revert to the normal colony phenotype upon subculturing. Stresses such as antibiotics and starvation have been shown to increase lag times of individual cells (Fridman, Goldberg, Ronin, Shoresh, & Balaban, 2014). Therefore, it has remained unclear whether these nonstable SCVs are formed by a founding cell population with increased lag times, or if the SCVs are comprised of a lineage of cells that heritably grow differently in the presence of the applied stress (Figure 2.2). These two alternative models imply starkly different underlying mechanisms and biological consequences, and thus discriminating between them is essential to improve our understanding of bacterial survival strategies.

In our work, we used a combination of high-throughput dynamic and quantitative imaging to identify and define mycobacterial subpopulations that are able to grow in the presence of sublethal antibiotics. We found that two colony phenotypes appeared in the presence of drug, which we termed large and small colony variants. Using a combination of live-cell imaging, rna-seq, and whole-genome sequencing, we determined that these subpopulations have distinct transcriptional profiles and growth properties that are heritable across generations and cannot be explained by underlying genetic mutations. Remarkably, these phenotypic variants exhibited a heritable ability to grow on drug. Following regrowth in media without drug, the progeny of these variants showed increased resistance to drug relative to the original population even after 24 hours of subculture, after which drug resistance gradually returned to the level of the original population. Variant types could also be distinguished from each other after initial

subculturing, further indicating that the variant forms are heritable across generations. Live-cell imaging confirmed that heritable differences in growth dynamics between the variant types could be also be detected at a single-cell level. These findings show that phenotypically drug resistant, epigenetically regulated lineages of mycobacteria are selected for in the presence of sublethal antibiotics.

Our experiments do not formally determine whether these subpopulations are pre-existing in the population or whether they arise *de novo* in the face of drug treatment. The high rate at which we detect these resistant populations in the presence of drug (up to 10% of the population) suggests that drug is at least increasing the capacity of the bacteria to form these variants, however future experiments must be done to discriminate between these possibilities.

Although here we used *M. smegmatis* and sublethal antibiotics to isolate these phenotypic variants, colony size variation has been observed in mycobacteria in a variety of conditions. Notably, small and large colonies have been observed after plating *M. tuberculosis* from macrophages and mice (data not shown). Accordingly, single cell variation has been observed in *M. tuberculosis* in response antibiotic and immune stress(Manina et al., 2015). One study used an rRNA fluorescent reporter in *M. tuberculosis* as a measure of the growth rates of single mycobacterial cells, and found that variation in reporter levels across individual cells increased in response to drug and immune stress (Manina et al., 2015). Reporter levels also correlated between mother and daughter cells, suggesting a possible role epigenetic inheritance. These findings together with our work suggest that perhaps small and large colony variants may be privileged, epigenetically regulated subpopulations that are selected for under a variety of conditions and contribute to the relcalcitrance of Mtb to immune control and antibiotic treatment.

2.5 Materials and Methods

Plating assay

M. smegmatis cultures were started from individual colonies in triplicate and grown to stationary phase (approx. 2 days). Cultures were then normalized, back diluted and grown to an OD ~1.0. Cultures were then diluted from between 10e-3 to 10e-5 and 100ul aliquots were plated on 7h10 plates supplemented with OADC (Middlebrook) and 20%tween80 (2.5%). Isoniazid was diluted in water at 5mg/ml and added to plates at indicated concentrations. Plates were incubated at 37 degrees Celsius for 4-6 days. Plates were imaged at defined time points following plating, using an Android camera phone. Plate images were analyzed using a custom cellprofiler pipeline to count and size colonies (sup fig).

Live cell imaging

Small and large colonies were picked from 7h10 plates with 4ug/ml INH, and resuspended in 1ml 7h9 supplemented with OADC (Middlebrook) and sonicated. Cells were loaded onto a microfluidic device and images were acquired every 10 min for 48 hours.

Reversion assays

Small and large colonies were picked from 7h10 plates with 4ug/ml INH, resuspended in 1 ml 7h9 supplemented with OADC (Middlebrook), sonicated, diluted from 10e-1 to 10e-5 and 100 ul aliquots were plated on 7h10 plates with and without INH. Resuspended cells were grown in 5 ml 7h9, incubated at 37 degrees with shaking, and replated using above methods at defined time points.

RNAseq experiments

Cells were harvested and resuspended in 1 ml Trizol, and bead beat 2x on a RNA was extracted using Directzol columns (Zymo Research) and ribosomal RNA was removed using the magnetic

RiboZero-Bacteria kit (Illumina). Libraries were prepared with the KAPA mRNA-seq kit, and sequenced on the MiSeq (Illumina).

Supplementary Figure 2.1. Enrichment analysis results of genes downregulated in small colonies. Genes downregulated more than 2 fold with a p value <.01 were entered into DAVIDtools functional enrichment analysis program (533 genes total). Count: number of genes annotated in given category. Bonferoni: statistical significance of enrichment of functional category.

Category	Term	Count	Bonferroni
SP_PIR_KEYWORDS	transcription regulation	89	8.77E-19
SP_PIR_KEYWORDS	Transcription	89	1.93E-18
GOTERM_MF_FAT	GO:0003700~transcription factor activity	93	1.15E-18
SP_PIR_KEYWORDS	dna-binding	94	2.67E-17
GOTERM_MF_FAT	GO:0030528~transcription regulator activity	95	1.26E-16
GOTERM_BP_FAT	GO:0051252~regulation of RNA metabolic process	99	3.07E-16
GOTERM_BP_FAT	GO:0045449~regulation of transcription	99	6.40E-16
GOTERM_BP_FAT	GO:0006355~regulation of transcription, DNA-dependent	99	2.64E-16
GOTERM_BP_FAT	GO:0006350~transcription	88	2.64E-15
GOTERM_MF_FAT	GO:0003677~DNA binding	112	1.93E-13
COG_ONTOLOGY	Transcription	58	3.31E-10
INTERPRO	IPR012287:Homeodomain-related	38	7.98E-07
INTERPRO	IPR001647:Transcriptional regulator, TetR-like, DNA-binding, bacterial/archaeal	37	5.27E-07
INTERPRO	IPR011991:Winged helix repressor DNA-binding	40	1.01E-04
SMART	SM00418:HTH_ARSR	7	0.007871707
SMART	SM00345:HTH_GNTR	15	0.004132243
INTERPRO	IPR011711:GntR, C-terminal	13	0.033169098
INTERPRO	IPR000524:Bacterial regulatory protein GntR, HTH	15	0.057320882
INTERPRO	IPR001845:Bacterial regulatory protein, ArsR	7	0.100706863
INTERPRO	IPR002529:Fumarylacetoacetase, C-terminal-like	3	1
INTERPRO	IPR004843:Metallophosphoesterase	3	1
INTERPRO	IPR011234:Fumarylacetoacetase, C-terminal-related	3	1
INTERPRO	IPR002397:Cytochrome P450, B-class	6	1
INTERPRO	IPR001128:Cytochrome P450	7	1

Supplementary Figure 2.1. Enrichment analysis results of genes downregulated in small colonies. (cont'd)

KEGG_PATHWAY	msm00071:Fatty acid metabolism	11	0.124032908
SP_PIR_KEYWORDS	oxidoreductase	58	0.493972901
SMART	SM00346:HTH_ICLR	6	0.410058047
INTERPRO	IPR005119:LysR, substrate-binding	9	0.778456976
INTERPRO	IPR000847:Bacterial regulatory protein, LysR	9	0.91964819
SP_PIR_KEYWORDS	Sigma factor	6	0.795827854
KEGG_PATHWAY	msm00980:Metabolism of xenobiotics by cytochrome P450	4	0.807097479
KEGG_PATHWAY	msm00640:Propanoate metabolism	8	0.967697747
KEGG_PATHWAY	msm00380:Tryptophan metabolism	8	0.795833622
KEGG_PATHWAY	msm00650:Butanoate metabolism	9	0.962599747
SMART	SM00342:HTH_ARAC	5	0.81180621
KEGG_PATHWAY	msm00280:Valine, leucine and isoleucine degradation	9	0.724285512
KEGG_PATHWAY	msm00350:Tyrosine metabolism	6	0.589355221
KEGG_PATHWAY	msm00632:Benzoate degradation via CoA ligation	8	0.961412606
KEGG_PATHWAY	msm00620:Pyruvate metabolism	6	0.944080355
KEGG_PATHWAY	msm00072:Synthesis and degradation of ketone bodies	3	0.993250487
KEGG_PATHWAY	msm00643:Styrene degradation	3	0.993250487
KEGG_PATHWAY	msm00790:Folate biosynthesis	3	0.996970971
COG_ONTOLOGY	Secondary metabolites biosynthesis, transport, and catabolism	16	0.688734637
SP_PIR_KEYWORDS	iron	15	0.975797855

Supplementary Figure 2.1. Enrichment analysis results of genes downregulated in small colonies. (cont'd)

SP_PIR_KEYWORDS	2Fe-2S	7	0.993679508
GOTERM_MF_FAT	GO:0009055~electron carrier activity	34	0.953114845
INTERPRO	IPR018062:Helix-turn-helix, AraC type, subdomain 2	4	0.999867172
INTERPRO	IPR005471:Transcriptional regulator IclR, N-terminal	6	0.999401129
INTERPRO	IPR014757:Transcriptional regulator IclR, C-terminal	6	0.999401129
INTERPRO	IPR005829:Sugar transporter, conserved site	9	0.999831435
INTERPRO	IPR013249:RNA polymerase sigma factor 70, region 4 type 2	6	0.999976201
INTERPRO	IPR007627:RNA polymerase sigma-70 region 2	6	0.999996469
INTERPRO	IPR000838:RNA polymerase sigma factor 70, ECF, conserved site	4	0.999994519
INTERPRO	IPR015893:Tetracycline transcriptional regulator, TetR-like, C-terminal	4	0.999994519
INTERPRO	IPR002577:Helix-turn-helix, HxlR type	4	0.999994519
INTERPRO	IPR000005:Helix-turn-helix, AraC type	5	0.999999605
GOTERM_MF_FAT	GO:0016564~transcription repressor activity	6	0.998077312
INTERPRO	IPR014284:RNA polymerase sigma-70	6	0.999999959
INTERPRO	IPR017927:Ferredoxin reductase-type FAD-binding domain	5	0.999999999
INTERPRO	IPR002198:Short-chain dehydrogenase/reductase SDR	20	0.999999999
INTERPRO	IPR010852:Protein of unknown function DUF1470	3	1
INTERPRO	IPR001663:Aromatic-ring-hydroxylating dioxygenase, alpha subunit	3	1
INTERPRO	IPR012307:Xylose isomerase-type TIM barrel	4	0.999999998
INTERPRO	IPR001242:Condensation domain	4	0.999999998
INTERPRO	IPR013022:Xylose isomerase-like, TIM barrel	4	1
INTERPRO	IPR018060:Helix-turn-helix, AraC domain	5	1

Supplementary Figure 2.1. Enrichment analysis results of genes downregulated in small colonies. (cont'd)

INTERPRO	IPR002529:Fumarylacetoacetase, C-terminal-like	3	1
INTERPRO	IPR004843:Metallophosphoesterase	3	1
INTERPRO	IPR011234:Fumarylacetoacetase, C-terminal-related	3	1
INTERPRO	IPR002397:Cytochrome P450, B-class	6	1
INTERPRO	IPR001128:Cytochrome P450	7	1
GOTERM_MF_FAT	GO:0016645~oxidoreductase activity, acting on the CH-NH group of donors	4	0.999997484
GOTERM_MF_FAT	GO:0016987~sigma factor activity	6	0.999999781
GOTERM_MF_FAT	GO:0016566~specific transcriptional repressor activity	4	0.999999996
GOTERM_BP_FAT	GO:0006352~transcription initiation	6	0.999999838
PIR_SUPERFAMILY	PIRSF002815:acetate operon repressor	3	0.99953442
PIR_SUPERFAMILY	PIRSF000429:Ac-CoA_Ac_transf	4	0.998947125

CHAPTER THREE

Investigating the role of HupB and HupB modification sites in the formation of phenotypic variants in mycobacteria

Authors: Alexandra Sakatos[1], John Leszyk[2], Alejandra Garces[1], Michael Chase,[1] Sarah Fortune[1]

[1] Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA

[2] University of Massachussetts Medical School, Worcester, MA

Author Contributions A.S. and **S.F.** designed the experiments. **A.G** did whole-cell proteomic screen **M.C**. analyzed proteomic data. **J.L.** did the gel band mass spectrometry and parallel reaction monitoring. **A.S.** conducted all other experiments. **A.S.** and **S.F**. drafted the manuscript.

3.1 Abstract

Phenotypic heterogeneity is a common bet-hedging strategy used by bacteria to maximize the likelihood of survival in a variety of conditions. In chapter 2, we demonstrated a form of phenotypic diversification in *M. smegmatis* that creates at least two subpopulations of bacteria that grow in the presence of drug. These subpopulations have distinct transcriptional profiles and growth properties that are heritable across generations in the absence of an underlying genetic mutation. In this chapter, we describe a novel epigenetic mechanism that may drive the formation of these mycobacterial subpopulations. We show that the mycobacterial histone-like protein, hupB, is required for the formation of both subpopulations to grow in the presence of drug. We found that hupB is modified in *M. smegmatis* and *M. tuberculosis* in a whole cell proteomics screen, and we validated the modification sites with purified hupB. To study the function of these modification sites, we made a series of hupB mutants to abrogate modification at specific sites. We found that one of these mutations, K86R, resulted in the specific loss of the small colony variant subpopulation. This mutation also resulted in altered gene expression in late stationary phase, a stress condition. These results suggest a model in which modification of hupB heritably alters gene expression in a subpopulation of cells that form small colony variants in the presence of drug.

3.2 Introduction

Phenotypic heterogeneity has been observed in a wide variety of bacteria. In *M. tuberculosis*, phenotypic heterogeneity likely contributes to the failure of infection control by either the immune system or antibiotic treatment. Although the importance of phenotypic heterogeneity has been demonstrated, our knowledge of the molecular mechanisms driving phenotypic diversity in bacterial populations is still scarce.

In chapter 2, we demonstrated a novel form of phenotypic heterogeneity in *M. smegmatis* that is heritable and epigenetically regulated. A limited number of examples of heritable, epigenetic mechanisms that have the potential to drive this form of heterogeneity have been described in bacteria. DNA methylation is the most well-characterized mechanism by which bacterial cells can heritably alter gene expression in a subpopulations of cells, and a recent study showed that DNA methylation regulates gene expression in *M. tuberculosis*(Shell et al., 2013). Another established mechanism involves positive feedback loops, as in the lac operon in *E. coli*. In eukaryotes, histone modifications are widely used to heritably alter gene expression and chromatin structures in subpopulations of cells. Currently, no analogous form of epigenetic regulation has been described in bacteria. However, bacteria, including mycobacteria, do possess DNA binding proteins that perform functions similar to those of eukaryotic histones. These nucleoid-associated proteins are largely responsible for the maintenance and plasticity of the bacterial chromosome(Kar et al., 2005; Luijsterburg, Noom, Wuite, & Dame, 2006; Swinger & Rice, 2004; Wang, Li, Chen, Xie, & Zhuang, 2011). In *E. coli*, mutation of the nucleoid-associated protein HU results in altered nucleoid organization and altered expression of stress-responsive gene HU mutations also affect other DNA-dependent processes such as DNA replication, repair and chromosome segregation.

(Chodavarapu, Felczak, Yaniv, & Kaguni, 2008; Huisman et al., 1989; Jaffé, Vinella, & D'Ari, 1997; Kamashev & Rouvière-Yaniv, 2000; Ogawa, Wada, Kano, Imamoto, & Okazaki, 1989)

The mycobacterial histone-like protein HupB is a nucleoid-associated protein with a unique two-domain structure. It is homologous to the prokaryotic HU family in its N-terminus and the eukaryotic histone H1 in its C-terminus(Mukherjee, Bhattacharyya, & Grove, 2008). Histone H1 is specifically involved in the formation of higher order chromosome structures and DNA condensation, indicating that HupB's C-terminus may give it the capacity to influence nucleoid architecture. Studies have shown that HupB binds DNA with high affinity, with a preference for AT-rich DNA that is characteristic of the regulatory regions of ORFs(S. Kumar, Sardesai, Basu, Muniyappa, & Hasnain, 2010; Sharadamma et al., 2011). HupB is an essential gene in *M. tuberculosis* and essential under certain conditions of stringent growth in *M. smegmatis*.

We and others have shown that the mycobacterial hupB is modified at lysine and arginine residues. A recent study found that acetylation of hupB alters its ability to bind and compact DNA *in vitro*(Ghosh, Padmanabhan, Anand, & Nagaraja, 2016). These findings raise the intriguing possibility that modification of hupB could drive altered DNA compaction, and thereby alter gene expression, in subpopulations of cells. In this chapter we investigate this possibility.

3.3. Results

HupB is required for the formation of drug resistant colonies in the presence of INH

In chapter 2, we demonstrated that at least two privileged subpopulations, large and small colony variants are able to grow in the presence of MIC-levels of INH. These subpopulations exhibited a heritable ability to grow in the presence of drug in the absence of genetic mutation. The heritability of the drug resistance phenotype in the absence of an underlying genetic mutation implies that epigenetic factors are likely to govern the formation of these drug resistant subpopulations. HupB is a DNA binding protein that has previously been implicated in growth regulation and survival after INH exposure((Lewin et al., 2008; Niki et al., 2012; Whiteford, Klingelhoets, Bambenek, & Dahl, 2011). We tested a hupB deletion for its ability to grow in the presence of INH. We found that the hupB deletion resulted in a significant reduction in the numbers of small and large colonies formed, despite exhibiting normal growth in media without drug (Figure 3.1, 3.6). We wondered whether hupB might drive the formation of these drug-resistant subpopulations by regulating gene expression. We did RNA-seq in the WT vs hupB deletion strain and found that the lack of hupB resulted in depression of 77 genes. We then did gene set enrichment analysis (GSEA) and found that the genes that were derepressed in the hupB deletion strain significantly overlapped with those that were downregulated in the small colony variants, suggesting hupB represses gene expression to drive small colony formation (Subramanian et al., 2005). We also did chip-seq and found that hupB binds at ~800 sites around the genome, suggesting that it may regulate gene expression via global nucleoid remodeling as opposed to binding to specific promoters of the regulated genes.

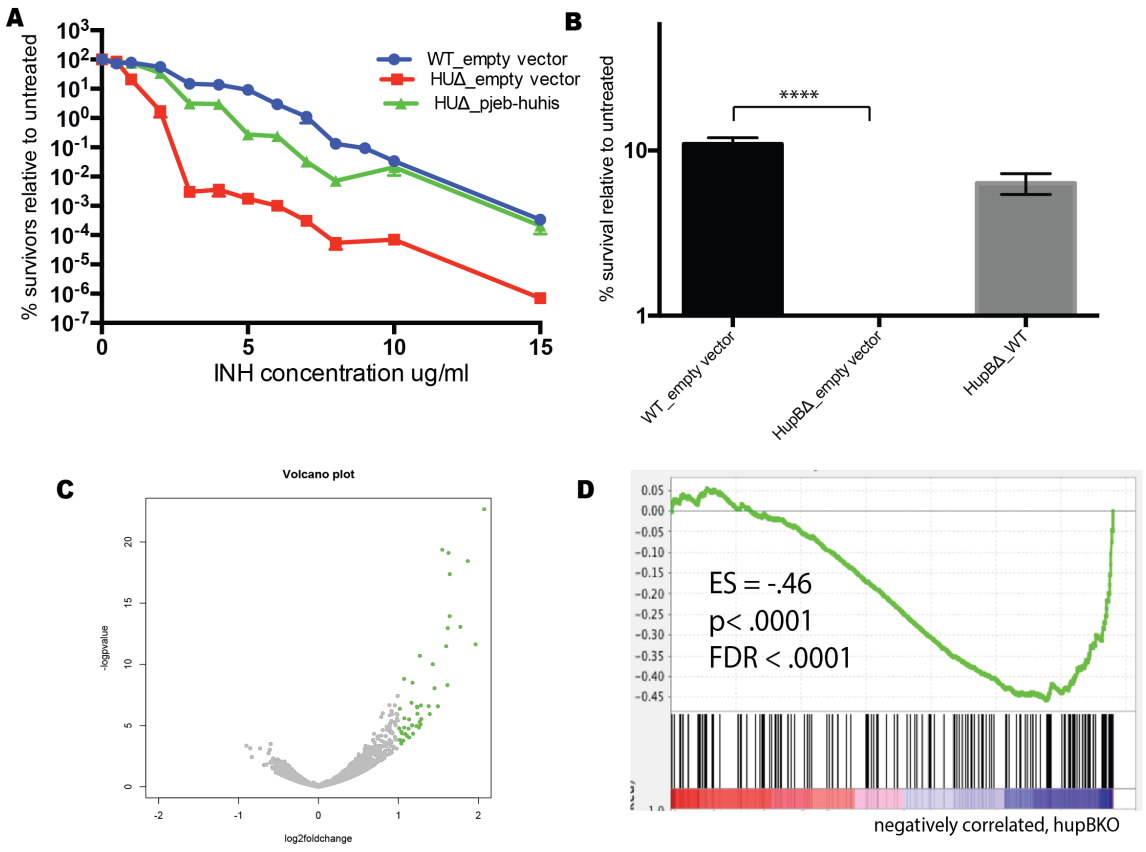
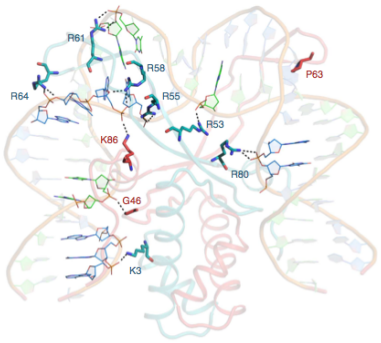


Figure 3.1 Loss of HupB results in significant reduction of colony formation in the presence of INH and alters gene expression A) Wt, HupB deletion, and complemented HupB strains were plated on 7h10 plates supplemented with increasing concentrations of INH, and colonies were counted after 3-6 days. **B**) Fraction of colonies on INH 4 ugh/ml plates/ colonies on 7h10 plates without drug **C)** Volcano plot of RNAseq data of HupB deletion vs WT cultures in late stationary phase. **D)** Gene set enrichment analysis (GSEA) of genes differentially expressed in the hupBKO and genes downregulated >4 fold with a p<.01 in the small colony variants. There is a significant, negative correlation between these two sets of genes, suggesting that hupB represses at least a subset of the genes that are downregulated in the small colonies.

HupB is modified at lysine and arginine residues.

The mycobacterial HupB has a C-terminal extension that is unique among eubacterial histone-like proteins, in that it contains a series of 'PAKKA' repeats that are homologous to the eukaryotic histone H1 sequence. Several of the lysines in the C-terminus have been demonstrated to be methylated ((Pethe et al., 2002)). A recent study intriguingly showed that hupB is also acetylated at lysine residues, and that acetylation of hupB *in vitro* alters it's ability to bind and compact DNA (Ghosh et al., 2016). The same group also showed that Eis acetylates hupB and that overexpression of Eis alters nucleoid morphology in *M. smegmatis.* In parallel to these studies, we found that hupB is modified at multiple lysine and arginine residues in a whole cell proteomics screen (Figure 3.2). Several of these modification sites were conserved between *M. smegmatis* and *M. tuberculosis* (Figure 3.2). In order to test the functions of these modification sites, we complemented the hupB deletion strain with a series of hupB mutant constructs. We then assessed the levels of protein in each of these mutant strains. We found that several of these mutations resulted in significantly reduced hupB protein levels, making it difficult to specifically asses the role of modification at those residues (Figure 3.3). We continued to study the mutants that did not display any reduction in protein levels, including the K86R mutant and the R55K mutant (Figure 3.3, 3.4). We found a phenotype for only the K86R mutant (Fig. 3.7, R55K data not shown) and therefore proceeded to further characterize the effect of the K86R mutation on hupB function. We found that the K86R mutation did not alter the DNA binding affinity of hupB *in vitro* (Figure 3.4).

Residue	Modification state
K3	Acetyl, Methyl
R55	Methyl
K70 or K72	Acetyl
K86	Methyl
K94	Acetyl
K103	Methyl



T. Bhomick et al, 2014

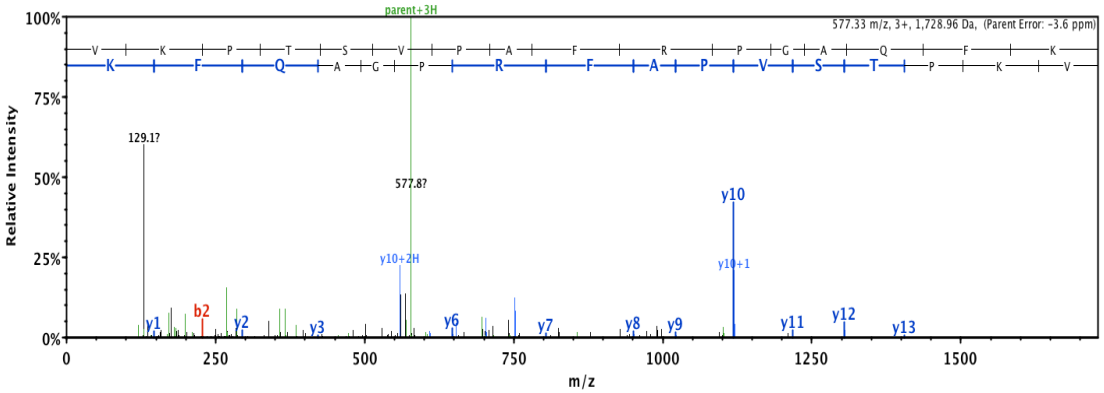
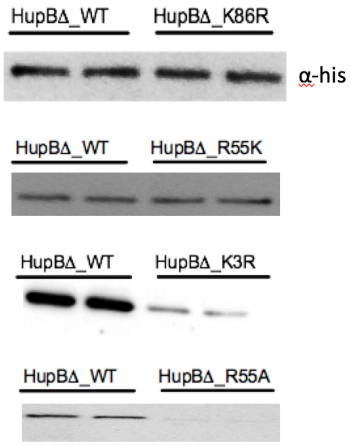


Figure 3.2. **HupB is modified at lysine and arginine residues. A)** A whole cell proteomics screen identified modified lysine and arginine residues in HupB. Several of these sites overlapped with putative DNA binding regions predicted by the crystal structure of HupB, which was solved by T.Bhomick et al. **B)** We purified his-tagged HupB from M. smegmatis and validated the modification sites. Shown here is the spectra of the K86 methylated peptide.

A.



B.

Mutant	Phenotype on INH	HupB protein levels
K3R	**yes**	**reduced**
K3Q	**yes**	**reduced**
K72R	**no**	**Not tested**
K94R	**no**	**Not tested**
K86R	**yes**	**normal**
K86A	**yes**	**Slightly reduced**
K86L	**yes**	**Slightly reduced**
R55A	**Not tested**	**reduced**
R55K	**no**	**no**

Figure 3.3 Assessment of HupB protein levels and phenotype on INH following mutation of HupB modification sites. A) We assessed protein levels of HupB point mutants by tagging mutants with a his tag and then probing for their levels by western blot. Several mutations resulted in a reduction of HupB levels. A subset of the mutations are shown here. **B)** Chart of the mutants made, whether they had a phenotype on INH, and their protein levels relative to wild-type as measured by western blot.

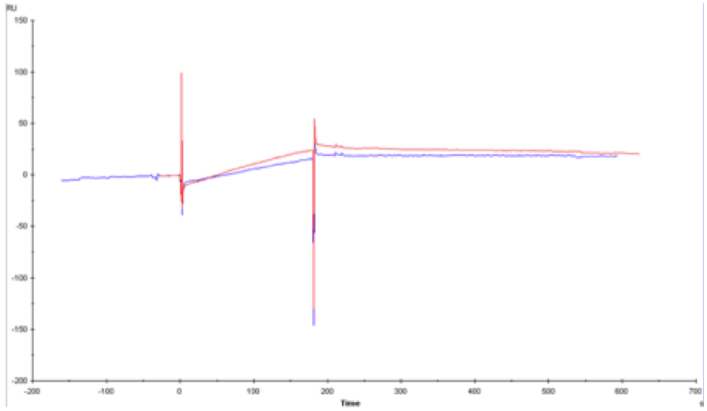


Figure 3.4 K86R mutation does not alter DNA binding affinity. WT and K86R forms of hupB were purified from *E.coli*. Surface plasmon resonance (SPR) was performed measuring the affinity of the protein to a 30 nucleotide oligo. Displayed binding is normalized to background binding to the chip

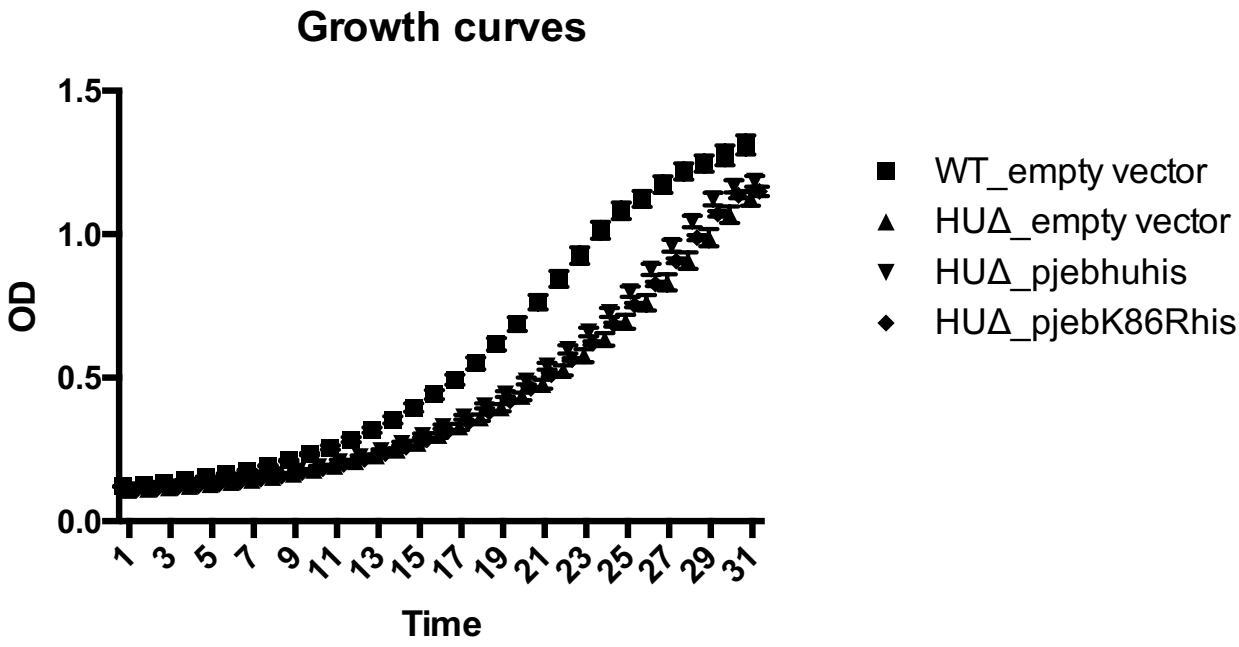


Figure 3.5. Growth curves. All strains were grown in triplicate and ODs were read on the Via II7 plate reader. Although all strains with the HupB deletion background grew slower than the WT strain, there were no significant differences in growth between the WT and mutant complemented strains.

HupB K86R mutation results in specific loss of small colony variants

Approximately ten percent of the K86 peptide that we characterized showed evidence of a modification. This could indicate that either ten percent of hupB in every cell is modified, or that different subpopulations of cells have different hupB modification profiles. In the latter possibility, hupB modification could theoretically alter dna binding and nucleoid compaction in a subset of cells, thereby diversifying the population into two distinct physiological states. We therefore wondered whether modification of HupB could drive the formation of the epigenetically regulated drug resistant variants that we discovered in chapter 2. We grew the WT and K86R mutant strains in media without drug and then plated on 7h10 plates supplemented with 4 ug/ml INH. Intriguingly, we found that although the K86R mutant strain formed large colony variants at the same rate as the WT, the K86R mutant resulted in the specific loss of small colony variants, while forming large colonies at the same rate as WT (Figure 3.6).. This phenotype was not due to a general growth defect of the mutant strain, since the K86R mutant strain grew at the same rate as the WT strain in 7H9 media without drug (Figure 3.5). We also did RNA-seq in the WT and K86R mutant strain and again found a general depression of gene expression in the mutant (Figure 3.6). We did GSEA to determine if the differentially expressed genes were enriched for genes that were downregulated in the small colonies. Although there was no a significant enrichment, the data again trended toward a negative correlation, suggesting that K86 modification may downregulate genes in the small colonies (Figure 3.6)

The specific loss of small colony variants in the K86R mutant strain suggests that perhaps modification of the K86 residue is driving the formation of the small colony variant subpopulation. If this is true, we might expect to see an enrichment of K86 methylation in the

small colonies. We extracted whole cell lysates from collections of small and large colonies in duplicate and measured K86 methylation using parallel reaction monitoring (PRM), a data-dependent form of mass spectrometry quantification. Using this method, we did not see any significant differences in K86 methylation between the large and small colonies (Figure 3.7). However, this method is likely not sensitive enough to detect quantitative differences in peptides that are of relatively low abundance in whole cell lysates, as the K86 peptide is. It is also possible that K86 methylation is altered at specific sites in the genome in small vs large colonies, making the difference difficult to detect using a bulk assay.

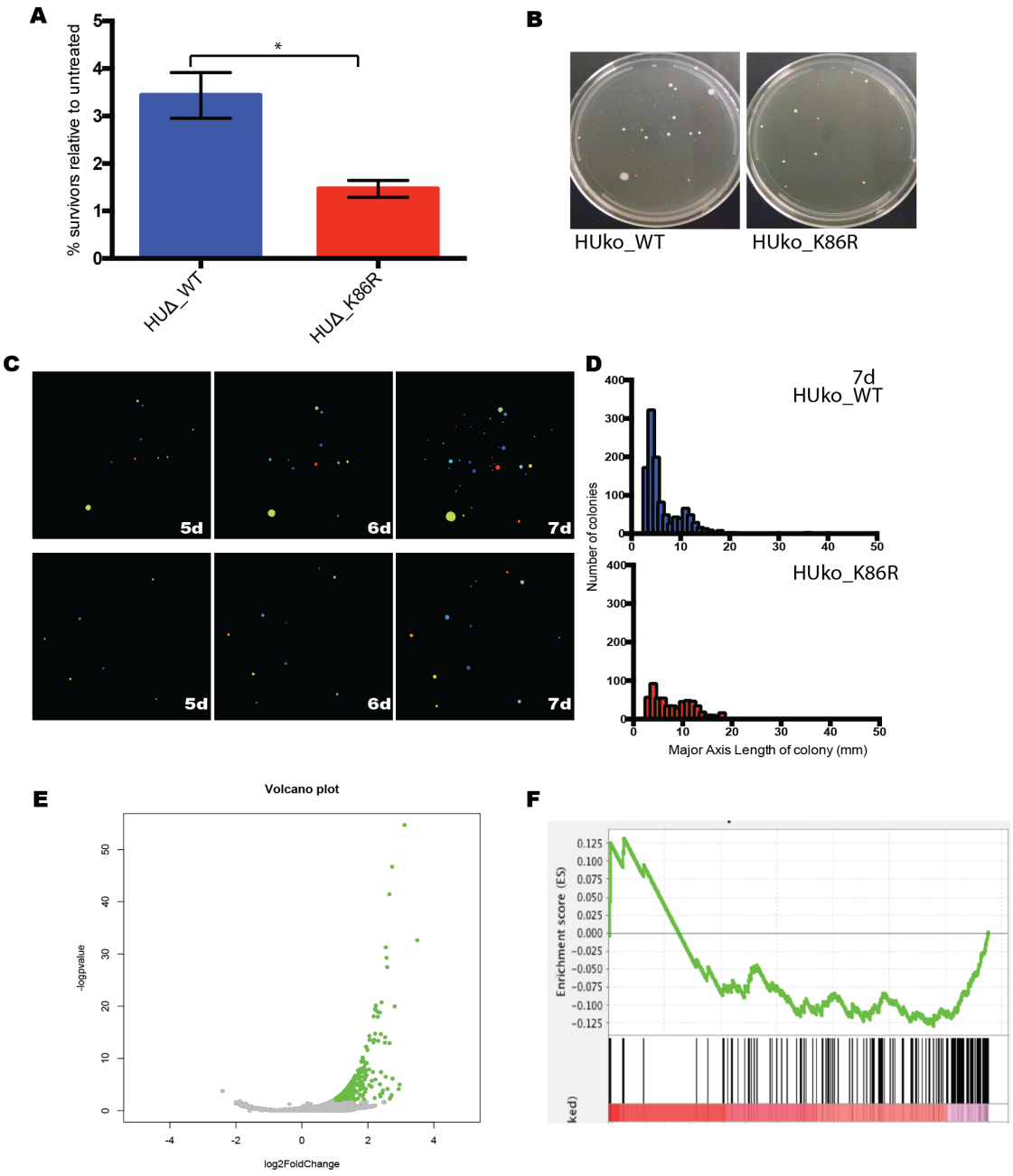
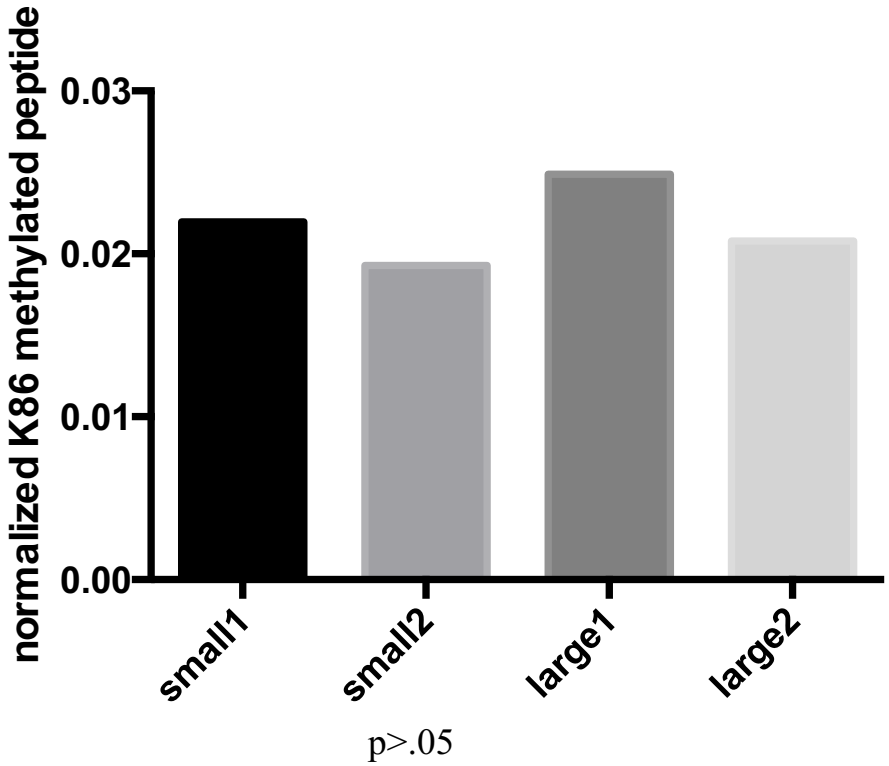


Figure 3.6.K86R mutation results in the specific loss of small colonies **A)** Total colony formation in the presence of INH in complemented HU deletion strain complemented with WT and K86R mutant. **B)** Representative plate images. **C)** Colony tracking over time – top panel, WT, bottom panel, K86R mutant. **D)** Histograms at day 7 for WT and K86R strains**. E)** Volcano plot of RNA-seq data of K86R mutant vs WT, RNA collected during late stationary phase. **F)** GSEA results of genes differentially regulated in the K86R mutant against genes downregulated more than 4 fold in the small colony variants. The data trend towards a negative correlation of between the two gene sets, suggesting that hupB may downregulate genes in the small colonies via K86.



p>.05

Figure 3.7 Differences in HupB K86 methylation were not detected between small and large colonies. The peptide containing K86 was targeting in whole cell lysates from small and large colonies using parallel reaction monitoring.

Role of methylation in small colony formation.

In order to ask whether methylation of hupB drives small colony formation, we attempted to block methylation in two ways. First, we found that three methyltransferases were upregulated in the small colonies in our RNAseq data. We knocked out one of these, a SAM-dependent methyltransferase (MSMEG_6483), and asked whether the loss of this methyltransferase blocked small colony formation. Although the distribution of colony sizes in the knockout looked different from that in the WT strain (Figure 3.8A) it was not clear whether small colonies were specifically eliminated. We also used an indirect methylation inhibitor, 3-Deazaneplanocin A (DZnep) to attempt to block methylation in *M. smegmatis*. This inhibitor has been shown to block methylation in eukaryotic cells via inhibiton of S-adenosylmethionine. In a preliminary experiment, adding this inhibitor to plates with INH appeared to result in a specific loss of small colonies (Figure 3.8B), which is consistent with our hypothesis. Future experiments must to done to confirm this result.

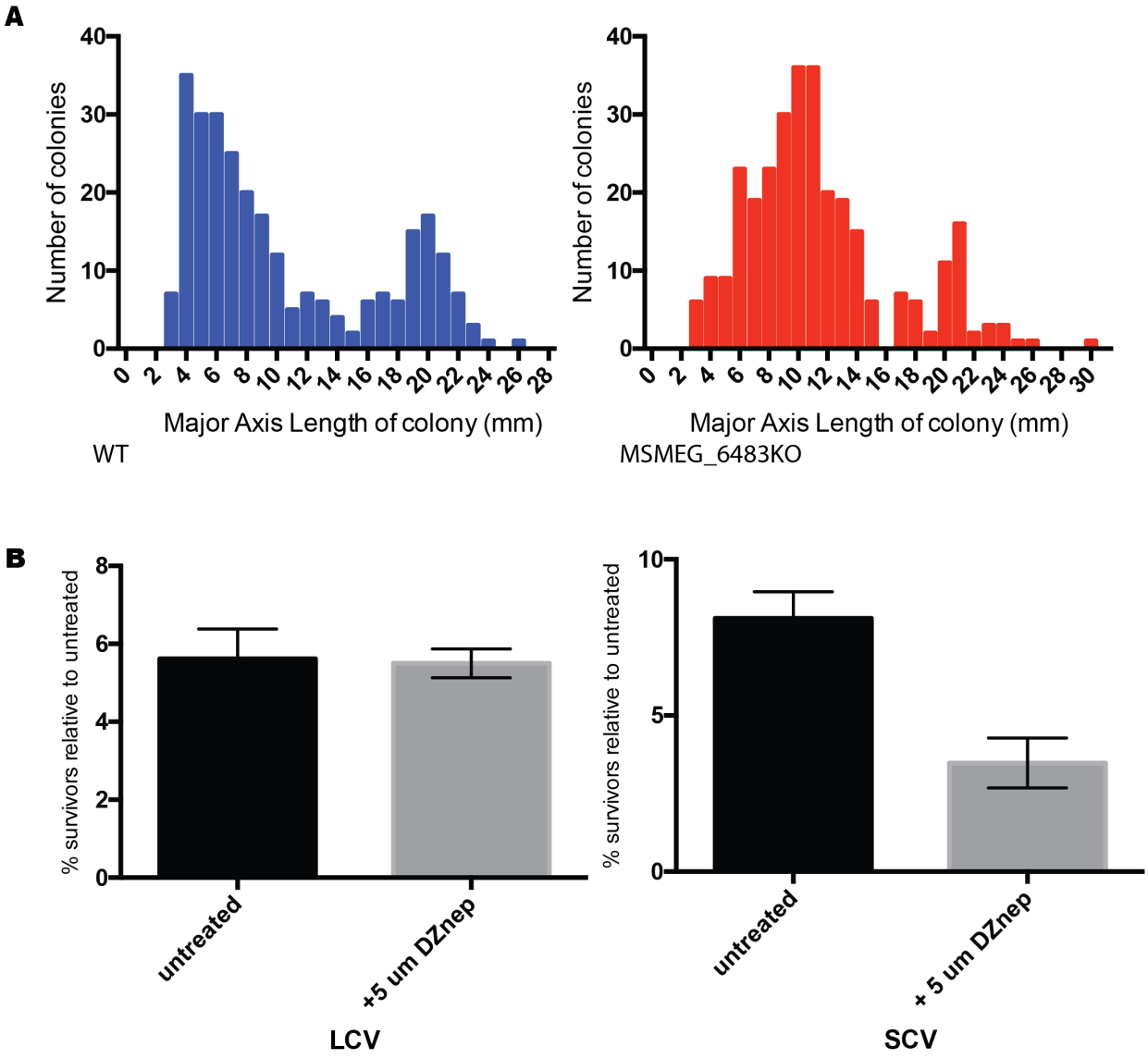


Figure 3.8 Role of methylation in small colony formation A) Distribution of colony sizes in the presence of inh after 5 days in the WT and MSMEG_6483 KO strains. **B)** Quantification of numbers of LCVs and SCVs in the presence of DZnep, a S-adenosylmethionine inhibitor.

3.4. Discussion

Epigenetic inheritance allows the forces of evolution to act on phenotypic variation, resulting in the selection of privileged subpopulations that are uniquely adapted to any given environment. Previously established mechanisms of epigenetic inheritance in bacteria, such as DNA methylation and positive feedback loops, typically involve regulation of only one or a few genes. Our findings suggest the existence of a fundamentally distinct form of epigenetic regulation in bacteria, whereby modification of a histone-like protein has widespread effects on gene expression that are inherited across generations and bifurcates the populations into two distinct cell types. A recent study showed that acetylation alters the ability of hupB to bind and compact DNA, suggesting a plausible mechanism by which these changes in gene expression could be implemented(Ghosh et al., 2016). This form of heritable epigenetic regulation mirrors the role of histone modifications in eukaryotes, and is the first indication that an analogous mechanism may exist in bacteria. Intriguingly, histone modifications were recently shown to drive the induction of a reversibly drug tolerant state in subpopulations of cancer cells, suggesting unexpected parallels of drug survival strategies across these distant cell types(Sharma et al., 2010).

Here we show that while HupB is required for formation of both small and large colonies, the K86 residue is solely required for the formation of the small colonies. Though the large colonies also appear to be an antibiotic resistant subpopulation that is dependent on the presence of HupB, the precise signals that drive their formation are unknown. Since the K86R mutation is not sufficient to drive large colony formation, the lack of methylation at this residue is unlikely to be the sole driver of large colony formation. In future work it would be interesting to

determine whether the posttranslational modifications we observed at additional residues in HupB are playing are role in the formation of the large colony subpopulation. Additionally, we have not ruled out that DNA methylation plays a role in the formation and maintenance of these colony phenotypes. In eukaryotes, histone modifications have been shown to work in concert with DNA methylation implement and maintain gene expression changes across generations. It would be interesting in future work to determine whether there is cross-talk between hupB modification and DNA methylation in mycobacteria.

The mechanism by which modification of hupB could alter gene expression of a subset of genes in a subpopulation of cells is as yet unknown. In eukaryotes, it is well established that histones can package chromosomal DNA into domains of varying compaction and transcriptional activity. Similar structural and functional organization of the bacterial nucleiod has more recently begun to be appreciated(Dame, 2005; Dillon & Dorman, 2010). Domains of varying superhelicites and levels of gene expression have been described in bacteria, and these domains are highly plastic in response to changing environments and growth phases. In *E. coli*, mutation of the nucleoid-associated protein HU results in altered nucleoid organization and altered expression of stress-responsive genes(Luijsterburg et al., 2006; Oberto, Nabti, Jooste, Mignot, & Rouviere-Yaniv, 2009). Whether modification of hupB can specifically alter nucleoid organization around specific domains, thereby affecting gene expression of a subset of genes, is yet to be explored.

3.5 Materials and Methods

Construction of HupB deletion and mutant strains. HupB deletion strain was constructed using previously published protocols to make unmarked deletion mutants. The complement constructs were made in pjEB402. The hupB ORF and 200 bp upstream of the ORF were amplified with primers and cut with XbaI and HindII, and ligated into a pjeB402 vector cut with the same enzymes. Mutations were made by doing site-directed mutagenesis PCR using pfu Ultra II HS polymerase.

INH plating assays. Cultures were grown overnight from single colonies in triplicate in 5ml 7h9, shaking at 37 degrees. The next day cultures were all normalized and grown overnight again to an OD ~1.0. Cells were diluted as indicated and 100 ul aliquots were plated on 7h10 plates supplemented with OADC (Middlebrook) with or without INH at indicated concentrations.

HupB purification. His-tagged hupB was expressed in *M. smegmatis* in pjeb under its native promoter and the recombinant strain was grown in 1 L of 7H9 with 25ug/ml of kanamycin to an OD of ~1.0. Cells were pelleted for 10 min at 4000 rpm, and pellets were resuspended in lysis buffer (200 mM NaCL, 50 mM Tris-HCL, ph 7.9, .1 mM EDTA, .1 mM DTT, 5% glycerol). Cells were lysed using a french press, then centrifuged for 10 min at 10,000. Supernatant was added to cobalt beads (TALON) that were washed 5x in wash buffer (50 mM Tris-HCL, ph 7.9, 0.1 mM EDTA, 500mM NaCL, 10 mM imidazole) and rocked overnight at 4 degrees. Beads were then spun down at low speed and liquid was removed, then beads were washed 3x for 10 min in wash buffer. HupB was eluted with 1 mL of elution buffer (50 mM Tris-HCL, pH 7.9, 0,1 mM EDTA, 400 mM NaCl, 500 mM imidazole). Eluate was mixed with 2x tricine SDS-

PAGE buffer (Life technologies) and run on a 10-20% tricine gel (Life technologies). Band was cut out and gel band mass spectrometry was performed.

Parallel Reaction monitoring (PRM) Collections of small and large colonies were picked from 7h10 plates containing 4 ug/ml INH in duplicate. Cells were resuspended in lysis buffer (200 mM NaCL, 50 mM Tris-HCl, pH 7.9, 0.1 mM EDTA, 0.1 mM DTT, 5% glycerol) added to glass beads and bead beat 6x at max speed for 30 sec each. Samples were placed on ice in between each bead beating cycle for 2 min. Beads were then spun down at 4000 rpm for 5 min. Supernatant was taken off the beads and mixed with an equal volume of 2x SDS page sample buffer and boiled for 2 min at 98 degrees. 25 ul of samples were run on 10-20% tricine gel (Life technologies) Gel was run for ~15 min at 105 V. Samples were cut out of gel and sent for PRM processing.

Chip-Seq *M. smegmatis* expressing hupB-his at an integrated promoter was grown to log phase. Chip-seq was performed as described in (Jaini et al., 2014). The monoclonal his antibody from Novagen was used for the pulldown. Libraries were constructed with the Illumina Tru-seq kit.

CHAPTER FOUR

Alternate forms of HupB regulation

Authors: Alexandra Sakatos[1], Scarlet Shell[1], Sarah Fortune[1]

[1] Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA

Author Contributions A.S., S.S. and **S.F.** designed the experiments. **S.S** did 5'end mapping, and protein tagging screen was done jointly by **A.S**. and **S.S**. **A.S.** and **S.F**. drafted the manuscript.

3.1 Abstract

M. tuberculosis is able to effectively adapt to a wide variety of harsh environmental conditions. This ability allows Mtb to establish chronic infection and also resist antibiotic treatment. The molecular mechanisms by which Mtb can withstand these conditions are not well understood, and developing a greater understanding of these mechanisms can inform the development of improved anti-tuberculosis treatments. Bacterial nucleoid-associated proteins have the ability to drive global transcriptional changes in response to the environment, and are therefore prime candidates for facilitating adaptive cellular programs during infection. In this chapter we investigate the function and post-transcriptional regulation of one nucleoid-associated protein, hupB, in mycobacteria. We show that hupB is important for survival of *M. smegmatis* in response to both isoniazid and rifampin treatment. We then investigate two potential mechanisms by which hupB function may be regulated in response to a changing environment. We show that hupB is modified, and test the hypothesis that modification of hupB drives autoregulation of the hupB promoter. We then tested the hypothesis that processing of the *hupB* transcript regulates hupB function by creating two protein isoforms of hupB.

3.2 Introduction

The remarkable ability of *M. tuberculosis* to tolerate stress contributes significantly to its global disease burden. *M. tuberculosis* is phagocytosed by alveolar macrophages, where it is able to reside for extended periods by evading host immunity. In the majority of patients, collections of immune cells will form around the infected macrophages, forming a granuloma lesion. *M. tuberculosis* can survive within these lesions for many years, and these latent infections can reactivate to cause disease[9]. Within the granuloma, *M. tuberculosis* is exposed to a variety of environmental stressors including hypoxia, nutrient deprivation, and low pH(Browning, Grainger, & Busby, 2010). Furthermore, treatment of tuberculosis can take six to nine months, after which not all infections are fully cleared. This extended treatment time may be attributed in part to the emergence of bacteria that can tolerate drug treatment(Adams et al., 2011; Chao & Rubin, 2010). *M.tuberculosis'* ability to adapt during chronic infection and antibiotic treatment is often attributed to metabolic and transcriptional reprogramming in response to stress, but we have a limited understanding of the molecular mechanisms that underlie these responses.

Bacterial nucleoid-associated proteins have the capacity to implement global cellular changes in response to the environment. In eukaryotes, it is well established that histones can package chromosomal DNA into domains of varying compaction and transcriptional activity. Similar structural and functional organization of the bacterial nucleiod has more recently been demonstrated(Dame, 2005; Dillon & Dorman, 2010; Kar et al., 2005; Luijsterburg et al., 2006; Wang et al., 2011). Domains of varying superhelicites and levels of gene expression have been described in bacteria, and these domains are highly plastic in response to changing environments and growth phases. In addition to regulation of these distinct domains, non-specific, global changes in chromatin

compaction in response to stress have also been shown in bacteria. Imaging of the mycobacterial nucleoid has shown strong growth phase and nutrient-dependent changes in chromosomal organization(Ghatak, Karmakar, Kasetty, & Chatterji, 2011). Nucleoid-associated proteins are largely responsible for the maintenance and plasticity of the bacterial chromosome(Berger et al., 2010; Swinger & Rice, 2004). In *E. coli*, mutation of the nucleoid-associated protein HU results in altered nucleoid organization and altered expression of stress-responsive genes(Luijsterburg et al., 2006; Oberto et al., 2009). HU mutations also affect other DNA-dependent processes such as DNA replication, repair and chromosome segregation(Chodavarapu et al., 2008; Huisman et al., 1989; Jaffé et al., 1997; Kamashev & Rouvière-Yaniv, 2000). The diverse effect of these proteins on cellular processes makes them uniquely suited to rapidly institute global responses to changing environments.

The mycobacterial histone-like protein HupB is a nucleoid-associated protein with a unique two-domain structure. HupB is an essential gene in *M. tuberculosis* and essential under certain conditions of stringent growth in *M. smegmatis*. A number of studies point to a specific role for HupB in mycobacterial stress responses. HupB is upregulated after exposure to host infection, cold shock, dormancy, and during stationary phase(M. Kumar et al., 2011; Ogawa et al., 1989; Shires & Steyn, 2001). Deletion of HupB in *M. smegmatis* results in increased susceptibility to antibiotic treatment and UV exposure(Whiteford et al., 2011). HupB may also play a role in regulation of the viable but non-culturable state (VBNC). The VBNC is a state in which the bacteria are nonreplicating and not culturable on standard nutrients, but able to be resuscitated by the application of "resuscitation factors". Although its significance is controversial, VBNC may be a model for the state of Mtb in the host during latent infection(Chao & Rubin, 2010). *M. smegmatis* HupB deletion strains enter the VBNC state significantly later than wild-type strains, and then become irreversibly non-culturable. Overall, these findings suggest that HupB plays a critical role in the mycobacterial response to stress.

Posttranscriptional regulation of HupB would allow for rapid responses to environmental changes. Although HupB has been shown to be important for survival under different conditions, the mechanisms by which its activity may be regulated in response to the environment are not known. This proposal focuses on two potential mechanisms of hupB regulation, posttranslational modification and RNA processing.

In eukaryotes, it is well established that histone modifications, such as methylation and acetylation, promote changes in gene expression in response to environmental signals. However, an analogous mechanism in prokaryotes is unknown. Mass-spectrometry data suggests that HupB is modified at multiple lysine residues (Figure1), raising the question of whether these modifications could give mycobacteria the capacity to similarly regulate cell state. Aim 1 investigates the functional significance of these modifications.

RNA processing is a similarly novel form of prokaryotic gene regulation. Alternative splicing in eukaryotes frequently creates different protein isoforms from a single gene. In bacteria, it is generally assumed that each gene encodes one protein. Surprisingly, genome-wide mapping of transcript 5' ends in *M.tuberculosis* revealed abundant RNA processing events both within coding sequences and in intergenic regions (Scarlet Shell, Fortune lab). Though previous work has identified the function of at least one intergenic processing event in mycobacteria, the significance of processing events within genes has not been investigated(Sala, Forti, Magnoni, & Ghisotti, 2008). Intriguingly, the global 5' end mapping suggested that the *hupB* transcript is cleaved in between the coding sequence for the N- and C-terminal domains. This observation raised the exciting possibility that RNA processing of the *hupB* transcript could create two functionally distinct protein isoforms. In the second part of this chapter, we investigats this possibility.

‘

3.3 Results

HupB is posttranslationally modified at lysine residues

Mass spectrometry data suggested that HupB is posttranslationally modified at lysine residues, with either methylation or acetylation (Figure 4.1). Since these modifications are largely conserved between *M. tuberculosis* and its non-pathogenic relative, *M. smegmatis*, we investigated the function of these modified sites in *M. smegmatis*. I mutated the indicated lysines to arginine, which prevents acetylation of the residue while conserving its positive charge. Since it was previously demonstrated that deletion of HupB reduced survival in the presence of isoniazid (ref), I plated the mutants on subinhibitory concentrations of isoniazid. I found a survival defect for the K3 mutant only (Figure 4.2). I also mutated the K3 site to glutamine, which is purpoted to mimic acetylation, and saw the same survival defect (Figure 4.2). This defect was also apparent after plating on rifampicin, a distinct class of antibiotic, suggesting that the phenotype is a result of a defect in a general stress response pathway.

Lysine	Smegmatis Whole cell mass spec	Pull down 1	Pull down 2	Pull down 3,4	TB whole cell mass spec
K3	Acetyl	Methyl	none	none	Acetyl
K70 or K72	Acetyl	Acetyl	Methyl	none	Acetyl
K94	Acetyl	Acetyl	Methyl	none	N/A- residue not conserved
K103	none	Acetyl	Methyl	none	Methyl

with Mike Chase and Alejandra Garces

Figure 4.1. Lysines, including K3, of HupB are modified. A) Proteomics identified modified lysine residues in HupB. Whole cell experiments were done with global protein extracts from log phase cultures. Pull down experiments were done with purified his-tagged HupB from *M. smegmatis*

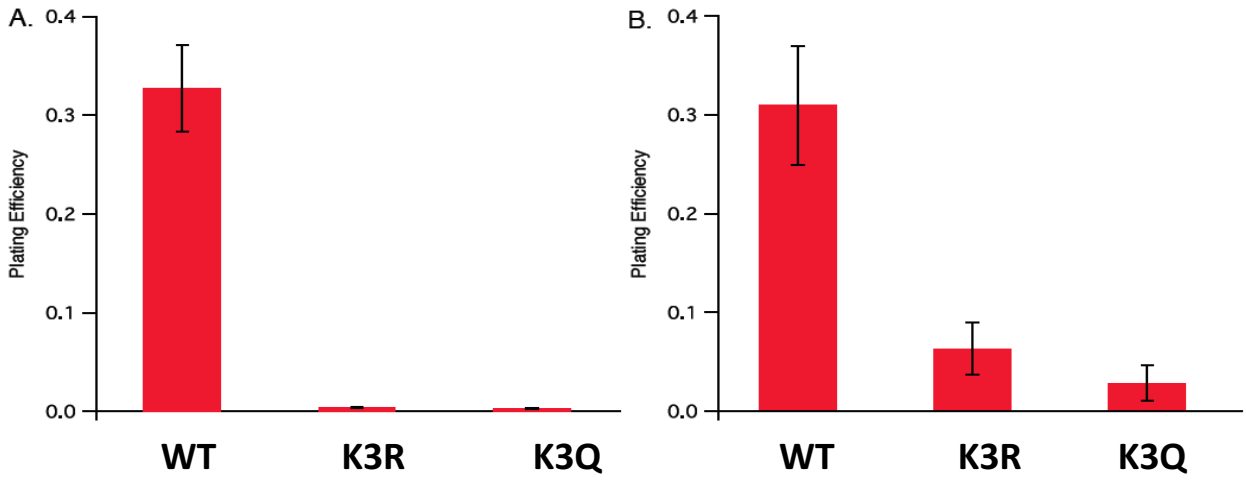


Figure 4.2. Mutation of K3 to arginine or glutamine results in significantly reduced tolerance to A)isoniazid and B) rifampicin. Cells were grown in 7h9+kan, normalized to an OD of 1.0 then diluted by 1x10-7 and 1x10-4 and plated on either 7H10 or 7h10 + drug, respectively. Colonies were counted after 4 days, and plating efficiency was calculated by taking the ratio of colonies grown on drug/number without drug.

The K3R mutation results in reduced hupB protein and hupB transcript levels

The K3 mutant appeared to phenocopy the hupB deletion (Figure 4.2). We therefore decided to look at the expression level of the mutant protein to determine if the reason for the defect was lowered hupB levels. We tagged the wild-type and mutant strains with a his tag and probed for protein abundance by western blot. We found that in fact, both the K3R and K3Q mutant proteins had significantly reduced protein levels relative to wt (Figure 4.3). To test whether the lowered protein levels in the K3 mutants was due to reduced protein stability or differences in transcript abundance or stability, we did qPCR of HupB in the wild-type and mutant strains.

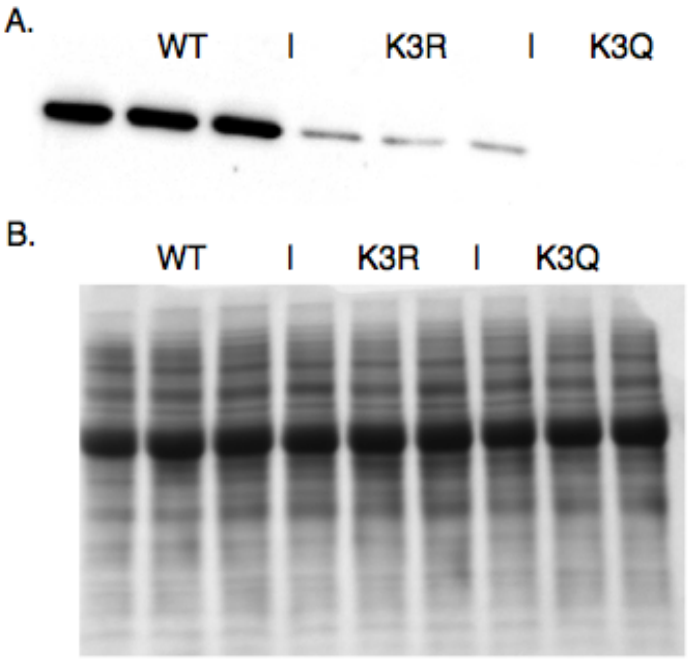


Figure 4.3. Mutation of K3 lowers HupB protein levels. A) Western blot indicates that K3R and K3Q mutations result in lower abundance of his-tagged HupB. B) Coomassie staining demonstrates that these mutations do not affect global protein levels.

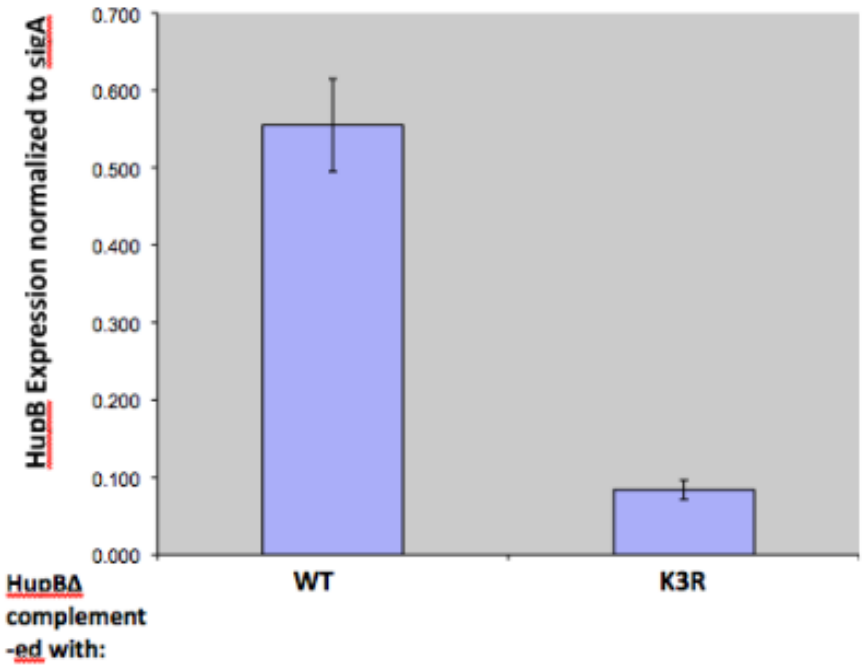


Figure 4.4. Mutation of K3 lowers HupB transcript levels. Total RNA was extracted from WT and K3R log phase cultures, then cdna was made and probed with SYBR green for qPCR.

HupB does not regulate its own promoter via K3

Surprisingly, qPCR revealed that the K3R mutant transcript was also lowered relative to the wild-type transcript. Since we had expressed both the wt and mutant transcript from the native hupB promoter, this result raised the intriguing possibility that the K3 mutant was negatively autoregulating its own promoter. However, the alternative possibility was that the K3 mutant transcript had reduced stability relative to wild-type. To distinguish between these possibilities, we episomally expressed his-tagged wt and K3R HupB mutants under the Tet promoter, and then probed for the native hupB transcript with qPCR. We reasoned that if the K3R mutation autoregulates its own promoter, reducing transcript levels, then protein levels of the his-tagged hupB expressed under the tet promoter should not differ between the WT and K3R mutant. Additionally, the native HupB transcript levels in the strain overexpressing the

K3R hupB mutant should be reduced. Alternatively, if the hupB k3R mutant transcript is directly destabilized, then the protein levels of the hupB K3R mutant expressed from the tet promoter should be lowered relative to wild-type.

The episomally expressed K3R mutant protein levels were reduced relative to episomally expressed wild-type HupB. Furthermore, when we probed for the native transcript in these strains, I found that the strain expressing the K3R mutant from the tet promoter had increased native transcript levels relative to the wild-type strain. A caveat to this experiment is that overexpression of the wt but not the K3R mutant inhibits bacterial growth – and this is likely why the transcription in the wt strain is reduced relative to the mutant. Regardless, the results do not support the autoregulation hypothesis. Therefore, the K3R mutation likely reduces hupB transcript stability and does not cause autoregulation of its own promoter.

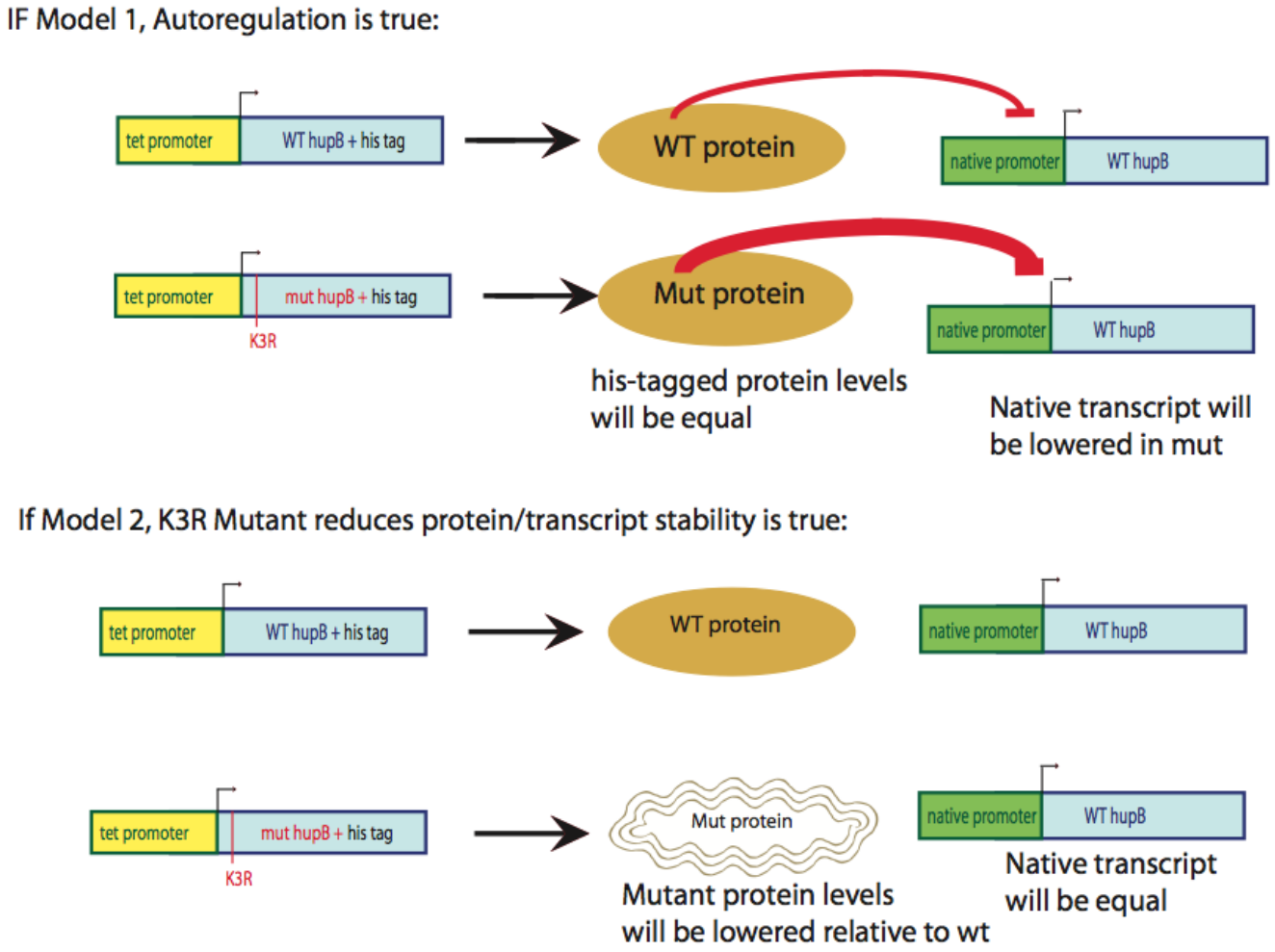


Figure 4.5. Opposing models for hupB regulation. In the first model, hupB autoregulates is own promoter via K3. If that is true, then expression of the K3 mutant protein from an episomal tet promoter will result in lowered levels of the native hupB transcript. In the second model, the K3 mutant protein is destabilized. In this model, the mutant protein levels expressed from the episomal promoter will be lowered relative to WT.

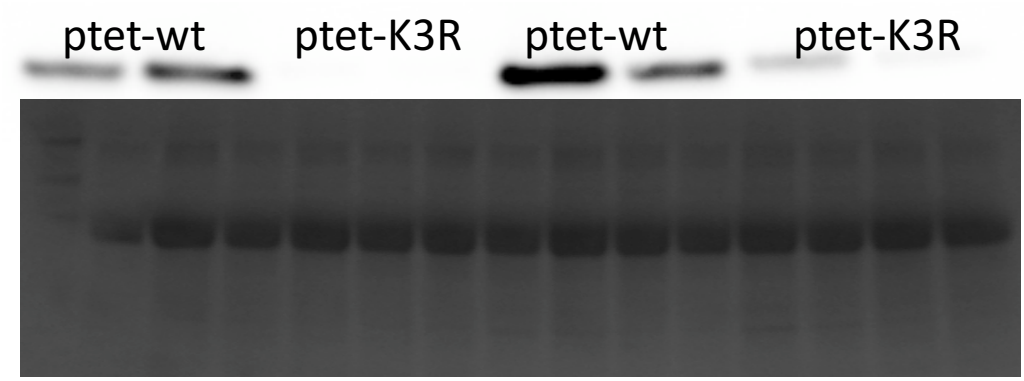


Figure 4.6. Protein levels of hupB K3R from the tet promoter are lowered relative to wild-type. Cultures were induced with 50 ng/ml tet and protein was extracted 60 and 90 minutes post- induction. Western blots were probed with an anti-his antibody. Biological duplicates are shown. A) Western blot indicates that protein levels of the K3R mutant are lowered relative to wild type. B) Coomassie loading control

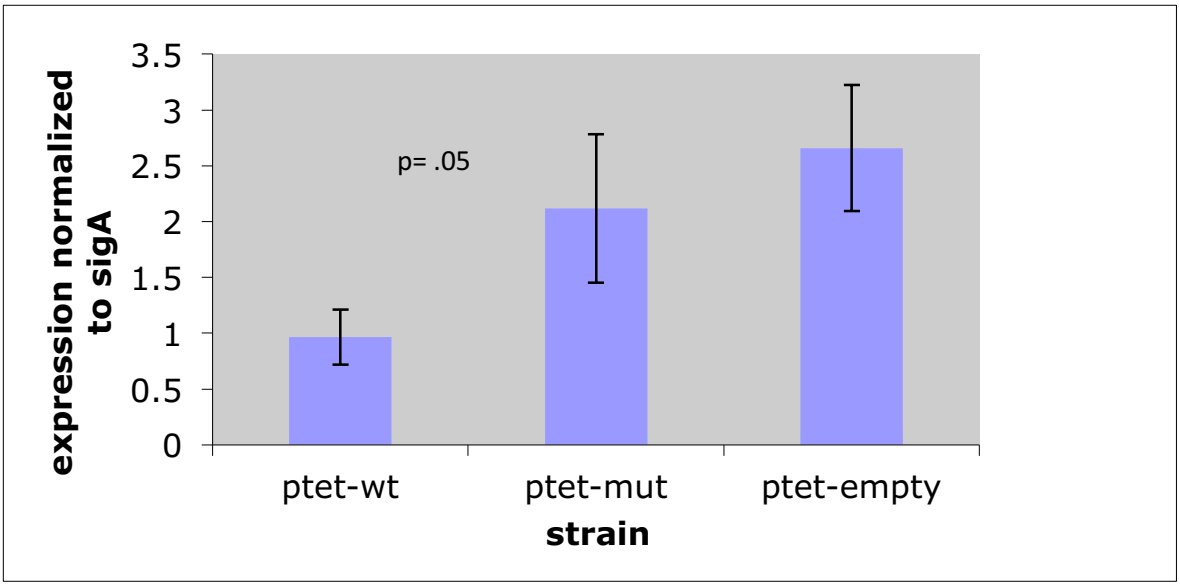
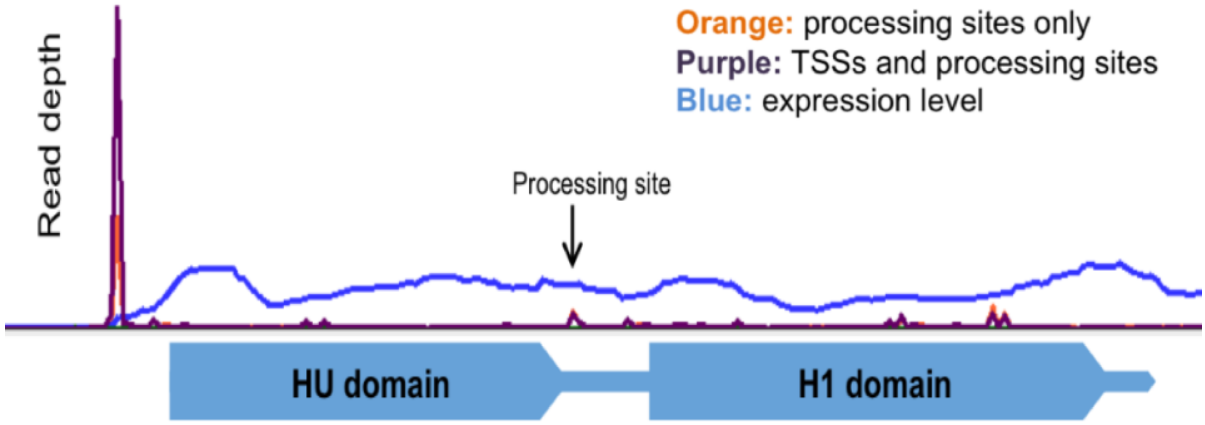


Figure 4.7. Native HupB transcript is upregulated in the tet-K3R strain relative to wild-type.
Cultures were induced with ATC, and RNA was extracted 3 hours post-induction. Relative expression of native hupB was quantified with qPCR .

The *hupB* transcript is processed.

Mycobacteria, unlike all other bacteria, have a sole, unique homolog to the prokaryotic HU protein, with a C-terminal extension that is homologous to the eukaryotic histone H1. This C-terminal extension may allow mycobacteria to tolerate extreme environmental stress(Salerno et al., 2009). Thus, regulation targeting the C-terminal domain might give mycobacteria the flexibility to tune their physiology to different metabolic states. In light of its unique domain arrangement, RNA processing presents a potentially exciting mechanism of HupB regulation. Intriguingly, global 5' end mapping in *M. tuberculosis* revealed RNA processing sites that were clustered in the 3'end of the *hupB* transcript (Figure 4.8). Using 3'RACE and qPCR, we confirmed that a truncated *hupB* transcript encoding only the N-terminal HU domain exists in the cell (Figure 4.9, 4.10). The qPCR data suggests that the 3' end of the transcript is degraded after the initial cut. We also did a northern blot that seemed to further suggest the existence of a truncated transcript, albeit the signal was weak (Figure 4.11).



Scarlet Shell

Figure 4.8. 5' end mapping reveals processing sites in the *hupB* transcript. Processed 5' ends were distinguished from transcriptional start sites (TSS) by taking advantage of the difference in 5' phosphorylation status between these two types of transcripts. TSS's have 5' triphosphates, while processed transcripts have 5' monophosphates. For one library, the 5' triphosphates were enzymatically converted to monophosphates so that rna adaptors for 5'RACE could be ligated to all the 5' ends; this library would then contain TSS and processed 5' ends. In the second library, 5'RACE was done without first converting the 5' triphosphates. Since only monophosphorylated 5' ends can be ligated to the adaptor, this library would yield only the processed 5' ends. 5' ends were called processing sites if the mean ratio of their abundance in the converted/non-converted library was near 1.

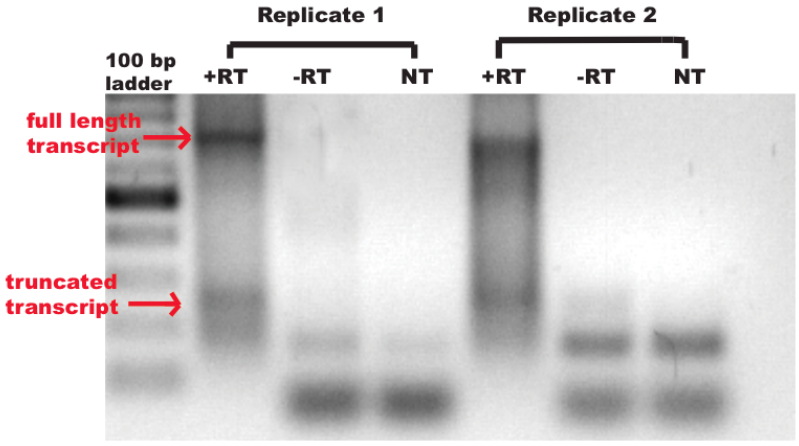


Figure 4.9. 3'RACE reveals the presence of a truncated hupB transcript encoding only the N-terminal HU domain. RNA was extracted from log phase cultures, and an adaptor was ligated to the 3' ends of the transcripts. Reverse transcription was then performed with a primer complementary to the adaptor sequence. PCR was done on the resulting cDNA, with the forward primer annealing to the beginning of the hupB transcript, and the reverse primer complementary to the adaptor. RT- reverse transcriptase, NT- no template control

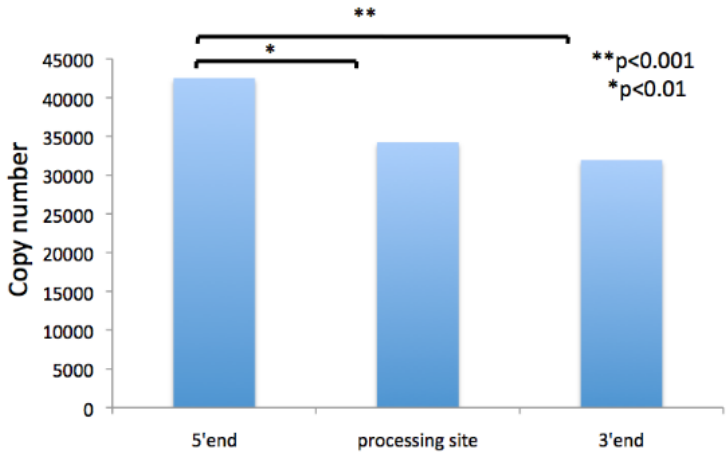


Figure 4.10. Expression of the 3' end of the *hupB* transcript is reduced by ~30% relative to the 5' end. qPCR was done with primers to the 5'end, primers spanning the processing site, and primers to the 3'end. Experiments were performed with biological duplicates and technical triplicates.

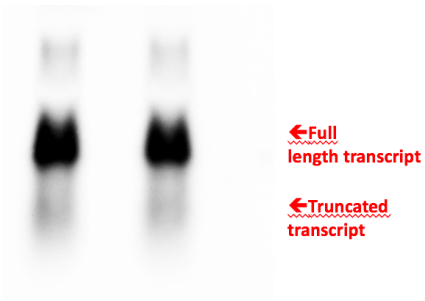


Figure 4.11. Northern blot of hupB. *M. tuberculsosis* was grown to log phase, total RNA was extracted and hybridized to a membrane and antisense probes to the N-terminus of the hupB transcript were used to detect native hupB forms.

Test the hypothesis that RNA processing of HupB creates 2 distinct proteins.

The preliminary 3' RACE data suggests that a truncated *hupB* transcript encoding only the N-terminal HU domain exists in the cell, raising the exciting possibility that RNA processing could create two distinct proteins from the same coding sequence. To investigate this possibility, we tagged hupB at the N-terminus with a flag tag and looked for evidence of two protein isoforms. Although I occasionally saw what looked like two protein isoforms, the results were not consistent across blots (Figure 4-12, data not shown). One explanation for the inconsistency of the results we observed is that processing of the hupB transcript varies with slight changes in culture conditions. More experiments will have to be done to directly test this hypothesis. However, we also tagged several additional proteins that contained RNA processing sites to ask if the creation of protein isoforms was a widespread event in mycobacteria, and we did not see any evidence of protein isoforms for the candidates we studied (data not shown).

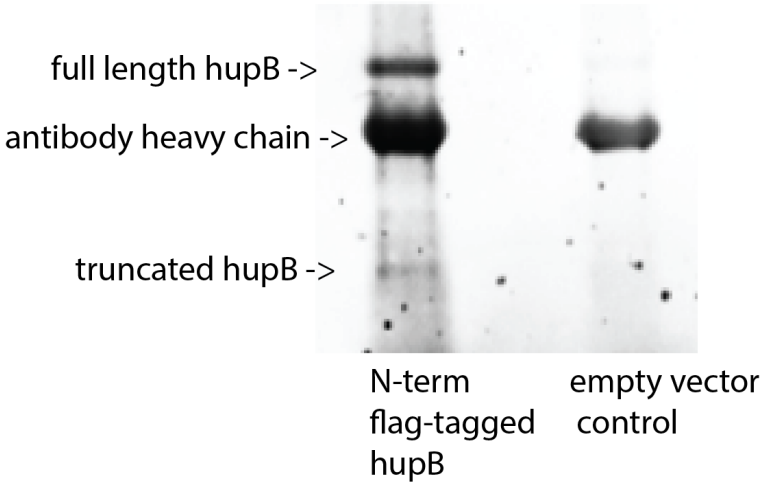


Figure 4.12. Western blot of flag-tagged hupB. hupB is tagged at the N-terminus with a flag tag. Flag-tag protein was purified from log phase *M. tuberculosis* cultures and run on an SDS-PAGE gel.

3.4 Discussion

In this chapter we investigated two mechanisms of post-transcriptional regulation of hupB. We found that hupB is post-translationally modified at lysine residues, and that mutation of one of these residues, K3, led to reduced protein and transcript levels. We showed that these effects are not due to an autoregulatory loop. We also tested the hypothesis that processing of the *hupB* transcript leads to the creation of two protein isoforms of hupB. Though we had evidence that a truncated *hupB* transcript, encoding only the N-terminal domain, was present, we did not find convincing evidence that the truncated transcript encoded protein. Importantly, the truncated transcript did not contain a stop codon. The lack of a stop codon can trigger the recruitment of the tmRNA system,, which monitors translational progress through the ribosome. In the event of a stalled ribosome, the tmRNA, which acts as both a tRNA and an mRNA, enters into the A site of the stalled complex and tags the incomplete protein product with an ssrA tag, which directs it to degradation by the ClpXP protease. tmRNA recruitment also stimulates release and degradation of the abberant transcript. The tmRNA system has been shown to be triggered by nonstop mRNAs, consecutive rare codons, weak termination signals, or depletion of a tRNA during amino acid starvation.

Interestingly, tmRNA deletion strains in model bacteria have been shown to be sensitive to stresses that cause protein misfolding, but a link between co-translational protein misfolding and the tmRNA pathway has not been established. Here we found that mutation of hupB leads to reduced protein stability and lowered transcript levels. Since we proved that these effects were not due to an autoregulatory loop, this leads to the alternative hypothesis that misfolding of the mutant protein is causing degradation of the corresponding transcript via co-translational surveillance of protein folding. In eukaryotes, it has been demonstrated that proteins can fold co-translationally and that

misfolding can trigger ubiquitination and degradation of the proteins. Additionally, it was just recently shown that the transcripts of misfolded secretory proteins can be co-translationally targeted for degradation(Karamyshev et al., 2014). However, an analogous mechanism in bacteria has not yet been demonstrated. A co-translational surveillance pathway for protein folding in bacteria would present a novel first line of defense against the multitude of stresses that can disrupt the protein folding environment. It would be interesting in future work to investigate this possibility.

Materials and Methods

Construction of HupB deletion and mutant strains. HupB deletion strain was constructed using previously published protocols to make unmarked deletion mutants. The complement constructs were made in pjEB402. The hupB ORF and 200 bp upstream of the ORF were amplified with primers and cut with XbaI and HindII, and ligated into a pjeB402 vector cut with the same enzymes. Mutations were made by doing site-directed mutagenesis PCR using pfu Ultra II HS polymerase.

5' end mapping 5'end mapping was performed as described in (Shell et al., 2015)

INH plating assays. Cultures were grown overnight from single colonies in triplicate in 5ml 7h9, shaking at 37 degrees. The next day cultures were all normalized and grown overnight again to an OD ~1.0. Cells were diluted as indicated and 100 ul aliquots were plated on 7h10 plates supplemented with OADC (Middlebrook) with or without INH at indicated concentrations.

Construction of HupB deletion and complemented mutant strains. HupB deletion strain was constructed using previously published protocols to make unmarked deletion mutants. The complement constructs were made in pjEB402. The hupB ORF and 200 bp upstream of the ORF were amplified with primers and cut with XbaI and HindII, and ligated into a pjeB402 vector cut with the same enzymes. The his-tag was added with the reverse primer. Mutations were made by doing site-directed mutagenesis PCR using pfu Ultra II HS polymerase.

3'RACE. An RNA adaptor was ligated to the 3'ends of total RNA using T4 RNA ligase. RT-PCR was then performed with a primer complementary to the adaptor sequence to create a single-stranded DNA template. The RT-PCR product was then amplified by PCR with a forward primer complementary to the start of the hupB sequence and the RT-PCR primer.

Western blot. N-terminally flag-tagged hupB was expressed from an integrating vector in M. tuberculosis. Recombinant cultures were grown to log phase, protein was extracted and flag-tagged proteins were purified by immunoprecipitation and run on an SDS-PAGE gel, then detected with an anti-flag antibody.

qPCR. RNA was extracted by phenol-chloroform extraction. cDNA was synthesized using superscript III reverse transcriptase (Invitrogen). qpCR was run using iTAQ Universal SYBR green Supermix (Bio-RAD). Primers were first tested for efficiency by running a standard curve using pre-defined cDNA concentrations.

CHAPTER FIVE: CONCLUSION

5.1 Epigenetically regulated phenotypic variants in mycobacteria

Phenotypic diversity enables the survival of uniquely adapted subpopulations in the presence of fluctuating environmental stressors. Though phenotypic diversity can be driven by genetic mutation, mutation is an extremely rare event. In mycobacteria, it is estimated that there are 2.4×10^{-10} mutations per base pair per generation. Still, higher rates of phenotypic variability in *M. tuberculosis* have been observed. Infection of an individual macaque with a clonal population of bacteria generates both sterile and active granuloma lesions, suggesting that phenotypic variability in the infecting population can drive the variable course of infection observed clinically. Similarly, persistent subpopulations are able to survive after treatment of a clonal population with antibiotics. These observations suggest that epigenetic factors are likely to be a primary driver of phenotypic variability in mycobacteria.

Here, we used dynamic and quantitative imaging to show that exposure of *M. smegmatis* to drug results in the selection of at least two distinct phenotypic variants that are able to grow in the presence of antibiotic. These variants form at a much higher rate than would be expected by a mutational mechanism, constituting ~3-10% of the total population. The variants showed an increased ability to grow on drug relative to the total population, and they could be distinguished by their distinct transcriptional signatures and growth properties. The observation of these distinct phenotypic variants at the colony level suggested the intriguing possibility that the variant phenotypes are inherited across generations. Using live-cell imaging we showed that the growth properties of the variant types were in fact sustained across generations. We also showed that the increased ability to grow on drug could be inherited across generations before reversion to the fully drug-susceptible form.

It is intriguing to note that the survival kinetics of *M. tuberculosis* in patients immediately following treatment with antibiotics are similar to the dynamics that we observed after exposing *M. smegmatis* to increasing concentrations of INH *in vitro.* Namely, in the first 2 days following treatment of patients with antibiotics, approximately 90-99% of the bacteria are eliminated, while the last 1-10% of bacteria persist at day 2 and then are more slowly eliminated over time (Jindani et al., 2003; Walter et al., 2015). The persistence of this subpopulation may contribute to the prolonged treatment time required to sterilize TB infections, which can last up to 6-9 months. We show similar effects when we expose *M. smegmatis* to a range of INH concentrations *in vitro*. Although the majority of the population is killed at lower concentrations of drug, approximately 1-10% of the population, which consists of the large and small variants, persists and is then more slowly eliminated with increasing concentration. This coincidence raises the possibility that slow elimination of Mtb in patients is due to the selection of epigenetically regulated phenotypic variants. In future work, it would be interesting to further explore this possibility.

5.2. Molecular mechanisms driving epigenetic inheritance in mycobacteria

Previously described forms of phenotypic variation in bacteria have ranged from extremely low frequency mutational events to high frequency switches generated during cell division (Aldridge et al., 2012). The form of phenotypic variation described here lies in between these two extremes, suggesting that mechanisms of epigenetic inheritance are driving the formation and maintenance of these phenotypic variants. DNA methylation is one mechanism by which bacteria can achieve epigenetic inheritance, and a recent study showed that DNA methylation occurs in mycobacteria and can alter gene expression(Shell et al., 2013). However, DNA methylation typically regulates one or a few genes. The phenotypic variants we described

exhibit widespread differences in gene expression. Furthermore, the distribution of the gene expression differences across the genome suggest that different nucleoid structure in the different phenotypic variants might underlie the transcriptional changes. In bacteria, nucleoid-associated proteins are the primary factors that maintain and alter the structure of the bacterial chromosome. HupB is a mycobacterial nucleoid-associated protein that had previously been shown to be important for survival in the presence of isoniazid (Niki et al., 2012; Whiteford et al., 2011). We found that deletion of hupB resulted in a significant loss of small and large colony variants. We also found that hupB is post-translationally modified by methylation and acetylation, and that mutation of a methylation site of hupB resulted in the specific loss of the small colony variants population. This finding suggests the existence of a novel form of epigenetic regulation in bacteria, where modification of nucleoid-associated proteins can drive heritable differences in nucleoid structure and gene expression. Variability in the levels of modification across cells could thereby drive the formation of phenotypically diverse subpopulations with different drug susceptibilities. Changes in modification state driven by fluctuating environments could also regulate

Other nucleoid-associated proteins (NAPs) in mycobacteria have been shown to affect survival in the presence of antibiotics. Lsr2 is one mycobacterial histone-like protein that has been shown to globally regulate gene expression. Deletion of Lsr2 reduces survival in the presence of multiple antibiotics (Colangeli et al., 2007). Similarly, napM, a newly described nucleoid-associated protein regulates the expression of ~150 genes and deletion of napM reduces survival in the presence of ethambutol and rifampicin. Interestingly, in our whole cell proteomics screen identified additional NAPs that were post-translationally modified, including Lsr2 (data not shown). This raises the intriguing possibility that post-translational modifications

of other NAPs could drive the formation of additional phenotypic variants that are specifically adapted to different antibiotics. Understanding the regulation of these NAPs along with elucidating the additional complexities of post-translational modification of hupB will give us a greater understanding of the mechanisms by which bacteria can modulate nucleoid architecture. This NAP-dependent regulation of nucleoid architecture may be a critical strategy by which bacterial populations achieve phenotypic diversity and survival across fluctuating environments.

5.3 Additional potential mechanisms of hupB post-transcriptional regulation.

HupB is unique among bacterial HU proteins because it has an additional C-terminal domain that is homologous to the eukaryotic histone H1. This C-terminal extension may increase the affinity of hupB for DNA (S. Kumar et al., 2010; Sharadamma et al., 2011). We found evidence of an RNA processing site in between the two hupB domains from global 5'end mapping data. We subsequently confirmed the existence of two *hupB* transcripts in the cell. We then tested the hypothesis that processing of the *hupB* transcript creates two protein isoforms in the cell. Though the results gave some indication of the existence of two protein isoforms, they were inconsistent. One explanation of the inconsistent results may be that processing of *hupB* varies with slight differences in culture conditions. Condition-dependent processing of the C-terminal extension of hupB could allow mycobacteria to modulate the affinity of hupB for DNA, and perhaps nucleoid architecture, in response to fluctuating environments.

5.4 Epigenetic drivers of the TB epidemic.

Tuberculosis recently surpassed HIV to become the world's leading infectious disease killer, causing approximately 1.5 million deaths per year. In 2014, 9.6 million people developed active TB disease. The extensive variability of infection and treatment outcomes have contributed to the intractability of the *M. tuberculosis* pathogen and hinders the effectiveness of

our current suite of antibiotics. The current drug regimen can last up to 9 months, resulting in many patients receiving incomplete treatment(World Health Organization, 2015). Even among those who have completed treatment, recurrent infections are common and may occur in up to 30% of treated cases. Although genetically encoded drug resistance contributes to a portion of treatment failures, the high failure rates suggest that epigenetic factors are likely to play a major role in driving treatment failures. Accordingly, it has been shown that strains responsible for relapsing often lack genetic mutations(Bryant et al., 2013). This dissertation describes epigenetic mechanisms that may contribute to the observed variability in infection outcomes and high percentage of treatment failures. Understanding these epigenetic drivers of *M. tuberculosis* diversification and survival could help facilitate the improved design of anti-tuberculosis therapies in the future.

Appendix I. Identification of novel riboswitches in *M. tuberculosis*

A riboswitch is a regulatory component of a mRNA, typically located in the 5'utr, which binds small molecules and implements changes in transcription or translation of the downstream gene in response to binding. Riboswitches allow for rapid responses to to fluctuating environments. Futhermore, riboswitches may be plausible drug targets, since many antibiotics target structural RNAs. There are two broad classes of riboswitches that have been described. The first class consists of translation attenuating riboswitches, where binding of the riboswitch to its cognate ligand results in the activation or termination of translation of the downstream ORF. Alternatively, transcription attenuating riboswitches regulate transcription of the downstream ORF upon binding their ligands.

Only one riboswitch has been fully described in *M. tuberculosis* thus far. This is the vitamin B_{12} riboswitch, which transcriptionally regulates the *metE* gene, which encodes a methionine synthase(Warner, Savvi, Mizrahi, & Dawes, 2007). Using the combination of genome-wide 5'end mapping and RNA-seq data, we sought to identify novel transcription attenuating riboswitches in *M. tuberculosis.* We reasoned that a transcription attenuating riboswitch, if set in the 'off' position, would have higher expression than the downstream ORF that it regulates. We therefore defined the 5'end of each ORF with the 5'end mapping data, calculated the expression of the first 100 bp following the 5'end, and then compared that to the expression level of the rest of the gene. We reasoned that the greater the difference between the first 100 bp with the rest of the gene, the more likely it was that those first 100 bp, or the entire 5'UTR was acting as a transcription attenuating riboswitch. Importantly, this technique only identifies riboswitches that are in the 'off' position in the condition in which the RNA-seq data was collected – in this case, in log phase Mtb. Performing these analyses with RNA-seq data

collected from a wide variety of conditions has the potential to more completely define the riboswitches in Mtb.

We found some interesting candidates for riboswitches after doing this analysis (Supplementary figure 6.1). We followed up on two particular candidates, narG (Rv1161) and narK(Rv1737). These genes are involved in the metabolism of nitrate(Malm et al., 2009; Sohaskey & Wayne, 2003). narK is a nitrate transporter, while narG is a nitrate reductase. Nitrate metabolism is important for survival during anaerobic conditions, and studies have shown that nark and narG are in fact upregulated during hypoxia and during macrophage infection (Jung et al., 2013) (Sohaskey & Wayne, 2003). Intriguingly, mutations in the upstream regions of both narG and nark2 have been shown to affect transcription the downstream ORF and nitrate reductase activity(Chauhan, Singh, & Tyagi, 2010; Stermann, Bohrssen, Diephaus, Maass, & Bange, 2003), consistent with the idea that the 5'UTR is a potential riboswitch. A polymorphism in the narG upstream region distinguishes *M. tuberculosis* from *M. bovis* (BCG) and correspondingly, differences in expression of narG and nitrate reduction activity between *M. tuberculosis* and *M. bovis* have been observed(Sohaskey & Modesti, 2009).

We found that both narG and nark2 were top candidates for riboswitch activity , with their UTR's upregulated relative to their downstream ORFs by ~5 fold and ~14 fold, respectively (Notably, narG is in an operon with narGHJI, meaning the riboswitch likely regulates the entire operon, but we limited our analysis to the first gene following the transcription start site). We validated the RNA-seq data by doing qPCR (Figure 6.1, narK2 data not shown). We then wondered whether hypoxia might upregulation narG expression via a 5'utr riboswitch. If this was true, we would expect that during hypoxia, the ratio of expression between the 5'utr and the downstream ORF would be reduced due to either loss of inhibition or activation of the

riboswitch, leading to read through of the downstream transcript. We in fact did see the expected reduction in this ratio when we collected RNA from hypoxic conditions (Figure 6.2). These results suggest the possibility of an oxygen-sensing riboswitch, which would be a novel class of riboswitches in bacteria. Further studies must be done to test this hypothesis.

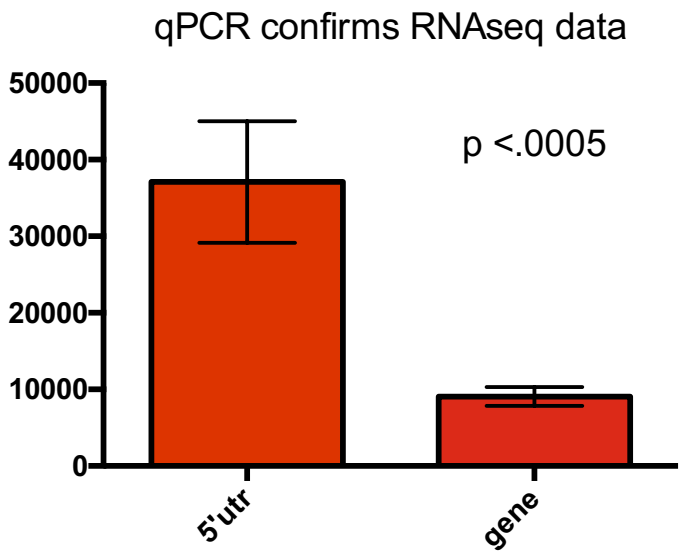


Figure 6.1. Expression of the 5'UTR and the ORF of narG, Rv1161

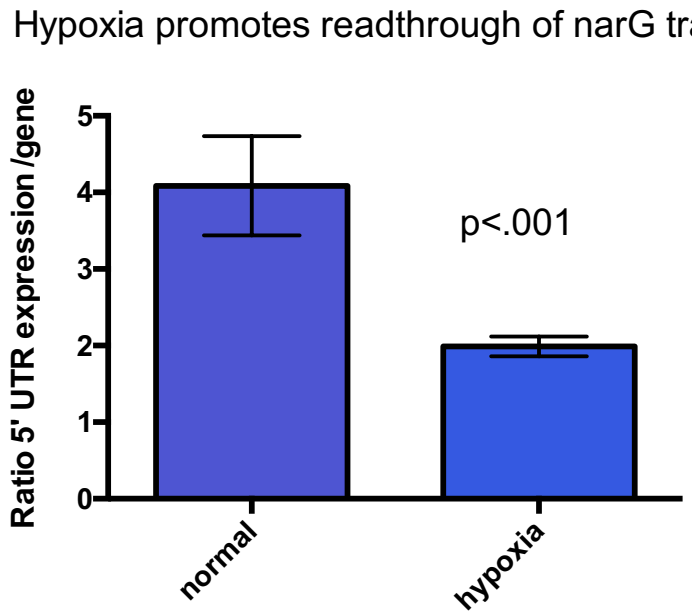


Figure 6.2. Ratio of 5'UTR expression relative to the downstream ORF for narG in normal vs. hypoxic conditions.

Supplementary table 6.1. List of top riboswitch candidates

coord_of_tss	*associated_gene*	*5'_utr_length*	*ratio 5'utr expression/gene*
4411270	Rv0001	263	183
104959	Rv0094c	296	58
4099474	Rv3660c	326	39
4411177	Rv0001	356	37
2047600	Rv1806	472	37
2047594	Rv1806	478	36
2544383	Rv2270	315	34
79313	Rv0070c	417	25
4083020	Rv3644c	299	24
1165681	Rv1042c	182	22
678896	Rv0584	333	18
309839	Rv0256c	292	18
4083048	Rv3644c	327	17
706075	Rv0612	249	16
577169	Rv0488	495	15
3926504	Rv3507	65	14
1965571	Rv1737c	201	14

706243	Rv0610c	334	13
1374266	Rv1230c	69	13
1347105	Rv1203c	200	12
3446005	Rv3080c	17	12
1985024	Rv1753c	249	12
909581	Rv0815c	263	12
3725971	Rv3340	156	11
1298962	Rv1167c	270	11
2982041	Rv2661c	465	11
3446041	Rv3080c	53	11
3545368	Rv3176c	68	10
1469899	Rv1313c	394	10
434762	Rv0355c	83	10
676031	Rv0578c	115	10
1696858	Rv1507A	498	10
1735508	Rv1535	468	10
2226211	Rv1983	33	9
4400438	Rv3912	432	9
4322175	Rv3848	151	9
3540860	Rv3171c	115	9
74921	Rv0068	380	9
501057	Rv0415	91	9
181022	Rv0152c	126	9
501051	Rv0415	97	8
1302855	Rv1173	76	8
1090328	Rv0977	45	8
361275	Rv0297	59	8
193358	Rv0162c	223	8
361265	Rv0297	69	8
1523934	Rv1355c	-98	8
4052867	Rv3611	83	8
1545892	Rv1373	120	8
4052688	Rv3611	262	8
916423	Rv0822c	112	8
133845	Rv0111	105	7
2402719	Rv2142c	209	7
1935139	Rv1706A	490	7
3287971	Rv2943	493	7
3267425	Rv2935	312	7
3025432	Rv2713	9	7

3872851	Rv3450c	355	7
282641	Rv0235c	27	7
2397210	Rv2136c	372	7
532306	Rv0442c	92	7
468245	Rv0388c	244	7
3564305	Rv3195	59	7
3849277	Rv3430c	472	7
1342979	Rv1199c	374	7
2498779	Rv2226	53	7
3945735	Rv3514	59	7
3005761	Rv2688c	111	6
3569031	Rv3197A	352	6
3849290	Rv3430c	485	6
1078545	Rv0969	198	6
2632124	Rv2351c	49	6
1933132	Rv1705c	478	6
1455270	Rv1299	225	6
3126479	Rv2818c	335	6
2800788	Rv2488c	-92	6
4405547	Rv3916c	380	6
1933138	Rv1705c	484	6
3972481	Rv3533c	28	5
3806046	Rv3391	-425	5
105268	Rv0096	56	5
612908	Rv0522	130	5
3795119	Rv3379c	252	5
3068307	Rv2754c	362	5
375746	Rv0305c	35	5
3958096	Rv3522	352	5
2618971	Rv2340c	63	5
131340	Rv0109	42	5
3930950	Rv3508	55	5
3799988	Rv3386	104	5
2934953	Rv2608	93	5
1830902	Rv1628c	341	5
3423215	Rv3060c	2	5
271804	Rv0226c	240	5
1123602	Rv1006	112	5
714190	Rv0621	12	5
2966120	Rv2642	413	5

412695	Rv0343	62	5
333417	Rv0277c	281	5
1287124	Rv1161	204	5
3877132	Rv3455c	310	5
3939558	Rv3511	59	5
92206	Rv0084	122	5
186429	Rv0158	356	5
2039833	Rv1800	-380	4
3886280	Rv3468c	211	4
538945	Rv0449c	397	4
3535413	Rv3166c	62	4
463394	Rv0386	17	4
1123591	Rv1006	123	4
1855710	Rv1646	54	4
1701052	Rv1510	243	4
1521874	Rv1354c	-2	4
1169280	Rv1046c	52	4
1701064	Rv1510	231	4
3494567	Rv3129	93	4
3985358	Rv3546	199	4
2562953	Rv2291	221	4
1371225	Rv1227c	400	4
1998811	Rv1765c	296	4
1325686	Rv1184c	75	4
3784774	Rv3370c	34	4
2966085	Rv2642	448	4
3778521	Rv3367	47	4
3752231	Rv3347c	-953	4
110206	Rv0101	-205	4
330375	Rv0274	47	4
3166652	Rv2856	32	4
2529385	Rv2255c	211	4
4400809	Rv3912	61	4
2170139	Rv1918c	-473	4
2062749	Rv1818c	75	4
1243672	Rv1121	35	4

Appendix 2. RNA-seq in *M. tuberculosis* across different conditions.

One defining feature of the macrophage environment in which *M. tuberculosis* lives is its lack of carbohydrate nutrients, which forces the bacterium to rely on fatty acid/lipid metabolism for growth. Additionally, *M. tuberculosis* in a macrophage is starved for iron and experiences acidic conditions. Global transcriptional responses likely underlie mycobacterial persistence under these conditions. Although microarray studies have been done to assess changes in gene expression under different conditions, this experiments were not able to define potentially important features of the mycobacterial transcriptome, including noncoding RNAs and transcriptional start sites. Here, we used RNA-seq to generate high resolution maps of the mycobacterial transcriptome under different metabolic states.

We grew *M. tuberculosis* to log phase in media containing either .4% glucose, .1% butyrate, .2% glucose + .1% butyrate, low iron and low pH. RNA was harvested and RNAseq libraries were made and sequenced with the help of Jonathan Livny (Broad).

The data show that global remodeling of the TB transcriptome occurs under these conditions (Figure 3.1). We reassuringly saw genes we expected to change in certain conditions, such as the upregulation of the mycobactin iron regulatory proteins in low iron. Additionally, we were able to clearly see start sites in the data that lined up with previous 5'end mapping (Figure 3.2). Intriguingly, we observed instances of multiple start sites for a given ORF in our RNA-seq data. A systematic analysis of the plasticity of start site usage could provide compelling evidence for start site regulation as a key component of transcriptional plasticity. Also, analyzing the relationship between start site usage and expression of the corresponding ORF might provide insight as to whether alternating start sites may function to regulate the expression of the downstream gene.

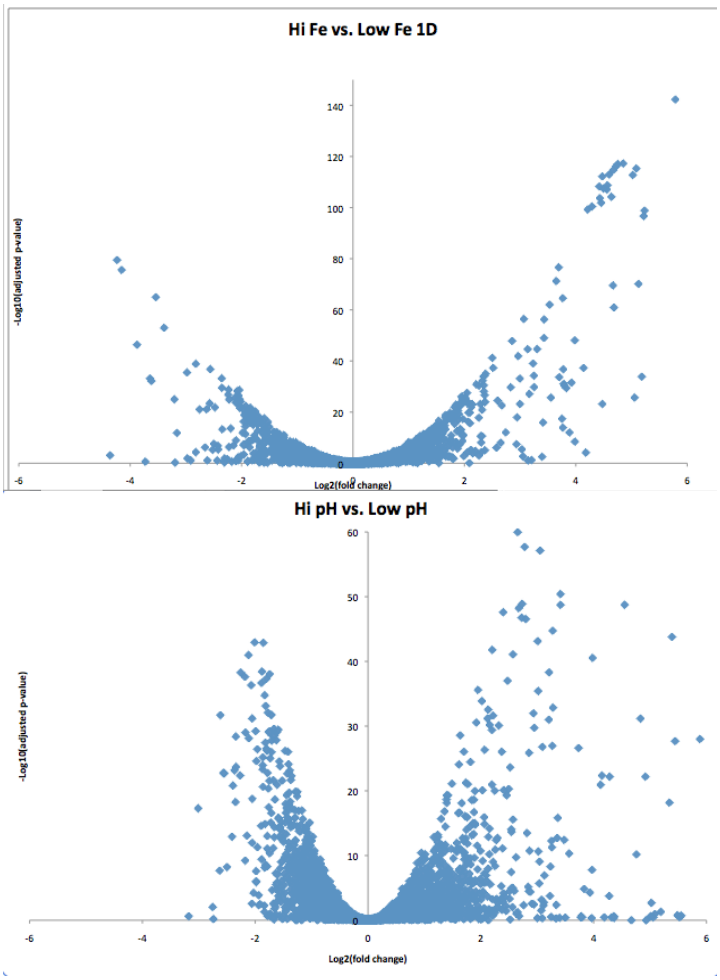


Figure 7.1 Representative volcano plots for High vs Low iron and High vs Low pH.

A.

B.

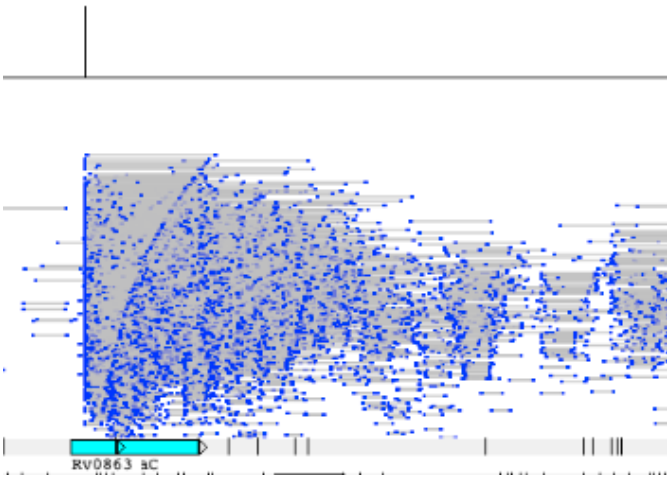


Figure 7.2. Visualization of 5'ends by RNAseq aligns with 5' end mapping data. A. Location of 5' end from 5' end mapping data (Scarlet Shell). B. Representation of reads in RNA-seq data.

References

Adams, K. N., Takaki, K., Connolly, L. E., Wiedenhoft, H., Winglee, K., Humbert, O., et al. (2011). Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism. *Cell*, *145*(1), 39–53. http://doi.org/10.1016/j.cell.2011.02.022

Aldridge, B. B., Fernandez-Suarez, M., Heller, D., Ambravaneswaran, V., Irimia, D., Toner, M., & Fortune, S. M. (2012). Asymmetry and aging of mycobacterial cells lead to variable growth and antibiotic susceptibility. *Science (New York, N.Y.)*, *335*(6064), 100–104. http://doi.org/10.1126/science.1216166

Arnoldini, M., Vizcarra, I. A., Peña-Miller, R., Stocker, N., Diard, M., Vogel, V., et al. (2014). Bistable expression of virulence genes in salmonella leads to the formation of an antibiotic-tolerant subpopulation. *PLoS Biology*, *12*(8), e1001928. http://doi.org/10.1371/journal.pbio.1001928

Balaban, N. Q. (2011). Persistence: mechanisms for triggering and enhancing phenotypic variability. *Current Opinion in Genetics & Development*, *21*(6), 768–775. http://doi.org/10.1016/j.gde.2011.10.001

Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., & Leibler, S. (2004). Bacterial persistence as a phenotypic switch. *Science (New York, N.Y.)*, *305*(5690), 1622–1625. http://doi.org/10.1126/science.1099390

Berger, M., Farcas, A., Geertz, M., Zhelyazkova, P., Brix, K., Travers, A., & Muskhelishvili, G. (2010). Coordination of genomic structure and transcription by the main bacterial nucleoid-associated protein HU. *EMBO Reports*, *11*(1), 59–64. http://doi.org/10.1038/embor.2009.232

Browning, D. F., Grainger, D. C., & Busby, S. J. (2010). Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Current Opinion in Microbiology*, *13*(6), 773–780. http://doi.org/10.1016/j.mib.2010.09.013

Bryant, J. M., Harris, S. R., Parkhill, J., Dawson, R., Diacon, A. H., van Helden, P., et al. (2013). Whole-genome sequencing to establish relapse or re-infection with Mycobacterium tuberculosis: a retrospective observational study. *The Lancet. Respiratory Medicine*, *1*(10), 786–792. http://doi.org/10.1016/S2213-2600(13)70231-5

Chao, M. C., & Rubin, E. J. (2010). Letting sleeping dos lie: does dormancy play a role in tuberculosis? *Annual Review of Microbiology*, *64*, 293–311. http://doi.org/10.1146/annurev.micro.112408.134043

Chauhan, S., Singh, A., & Tyagi, J. S. (2010). A single-nucleotide mutation in the -10 promoter region inactivates the narK2X promoter in Mycobacterium bovis and Mycobacterium bovis BCG and has an application in diagnosis. *FEMS Microbiology Letters*, *303*(2), 190–196. http://doi.org/10.1111/j.1574-6968.2009.01876.x

Chodavarapu, S., Felczak, M. M., Yaniv, J. R., & Kaguni, J. M. (2008). Escherichia coli DnaA interacts with HU in initiation at the E. coli replication origin. *Molecular Microbiology*, *67*(4), 781–792. http://doi.org/10.1111/j.1365-2958.2007.06094.x

Claudi, B., Spröte, P., Chirkova, A., Personnic, N., Zankl, J., Schürmann, N., et al. (2014). Phenotypic variation of Salmonella in host tissues delays eradication by antimicrobial chemotherapy. *Cell*, *158*(4), 722–733. http://doi.org/10.1016/j.cell.2014.06.045

Colangeli, R., Helb, D., Vilchèze, C., Hazbón, M. H., Lee, C.-G., Safi, H., et al. (2007). Transcriptional regulation of multi-drug tolerance and antibiotic-induced responses by the histone-like protein Lsr2 in M. tuberculosis. *PLoS Pathogens*, *3*(6), e87. http://doi.org/10.1371/journal.ppat.0030087

Cox, H., Kebede, Y., Allamuratova, S., Ismailov, G., Davletmuratova, Z., Byrnes, G., et al. (2006). Tuberculosis recurrence and mortality after successful treatment: impact of drug resistance. *PLoS Medicine*, *3*(10), e384. http://doi.org/10.1371/journal.pmed.0030384

Dame, R. T. (2005). The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Molecular Microbiology*, *56*(4), 858–870. http://doi.org/10.1111/j.1365-2958.2005.04598.x

Dhar, N., & McKinney, J. D. (2007). Microbial phenotypic heterogeneity and antibiotic tolerance. *Current Opinion in Microbiology*, *10*(1), 30–38. http://doi.org/10.1016/j.mib.2006.12.007

Dillon, S. C., & Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews. Microbiology*, *8*(3), 185–195. http://doi.org/10.1038/nrmicro2261

Fridman, O., Goldberg, A., Ronin, I., Shoresh, N., & Balaban, N. Q. (2014). Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature*, *513*(7518), 418–421. http://doi.org/10.1038/nature13469

Ghatak, P., Karmakar, K., Kasetty, S., & Chatterji, D. (2011). Unveiling the role of Dps in the organization of mycobacterial nucleoid. *PloS One*, *6*(1), e16019. http://doi.org/10.1371/journal.pone.0016019

Ghosh, S., Padmanabhan, B., Anand, C., & Nagaraja, V. (2016). Lysine acetylation of the Mycobacterium tuberculosis HU protein modulates its DNA binding and genome organization. *Molecular Microbiology*, n/a–n/a. http://doi.org/10.1111/mmi.13339

Huisman, O., Faelen, M., Girard, D., Jaffé, A., Toussaint, A., & Rouvière-Yaniv, J. (1989). Multiple defects in Escherichia coli mutants lacking HU protein. *Journal of Bacteriology*, *171*(7), 3704–3712.

Jaffé, A., Vinella, D., & D'Ari, R. (1997). The Escherichia coli histone-like protein HU affects DNA initiation, chromosome partitioning via MukB, and cell division via MinCDE. *Journal of Bacteriology*, *179*(11), 3494–3499.

Jaini, S., Lyubetskaya, A., Gomes, A., Peterson, M., Park, S. T., Raman, S., et al. (2014). Transcription Factor Binding Site Mapping Using ChIP-Seq. *Microbiology Spectrum*, *2*(2), 161–181. http://doi.org/10.1128/microbiolspec.MGM2-0035-2013

Jindani, A., Doré, C. J., & Mitchison, D. A. (2003). Bactericidal and sterilizing activities of antituberculosis drugs during the first 14 days. *American Journal of Respiratory and Critical Care Medicine*, *167*(10), 1348–1354. http://doi.org/10.1164/rccm.200210-1125OC

Jung, J.-Y., Madan-Lala, R., Georgieva, M., Rengarajan, J., Sohaskey, C. D., Bange, F.-C., & Robinson, C. M. (2013). The intracellular environment of human macrophages that produce nitric oxide promotes growth of mycobacteria. *Infection and Immunity*, *81*(9), 3198–3209. http://doi.org/10.1128/IAI.00611-13

Kahl, B. C. (2014). Small colony variants (SCVs) of Staphylococcus aureus--a bacterial survival strategy. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *21*, 515–522. http://doi.org/10.1016/j.meegid.2013.05.016

Kamashev, D., & Rouvière-Yaniv, J. (2000). The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *The EMBO Journal*, *19*(23), 6527–6535. http://doi.org/10.1093/emboj/19.23.6527

Kar, S., Edgar, R., & Adhya, S. (2005). Nucleoid remodeling by an altered HU protein: reorganization of the transcription program. *Proceedings of the National Academy of Sciences*, *102*(45), 16397–16402. http://doi.org/10.1073/pnas.0508032102

Karamyshev, A. L., Patrick, A. E., Karamysheva, Z. N., Griesemer, D. S., Hudson, H., Tjon-Kon-Sang, S., et al. (2014). Inefficient SRP interaction with a nascent chain triggers a mRNA quality control pathway. *Cell*, *156*(1-2), 146–157. http://doi.org/10.1016/j.cell.2013.12.017

Kester, J. C., & Fortune, S. M. (2014). Persisters and beyond: mechanisms of phenotypic drug resistance and drug tolerance in bacteria. *Critical Reviews in Biochemistry and Molecular Biology*, *49*(2), 91–101. http://doi.org/10.3109/10409238.2013.869543

Kiviet, D. J., Nghe, P., Walker, N., Boulineau, S., Sunderlikova, V., & Tans, S. J. (2014). Stochasticity of metabolism and growth at the single-cell level. *Nature*, *514*(7522), 376–379. http://doi.org/10.1038/nature13582

Kumar, M., Khan, F. G., Sharma, S., Kumar, R., Faujdar, J., Sharma, R., et al. (2011). Identification of Mycobacterium tuberculosis genes preferentially expressed during human infection. *Microbial Pathogenesis*, *50*(1), 31–38. http://doi.org/10.1016/j.micpath.2010.10.003

Kumar, S., Sardesai, A. A., Basu, D., Muniyappa, K., & Hasnain, S. E. (2010). DNA clasping by mycobacterial HU: the C-terminal region of HupB mediates increased specificity of DNA binding. *PloS One*, *5*(9), e12551. http://doi.org/10.1371/journal.pone.0012551

Leimer, N., Rachmühl, C., Palheiros Marques, M., Bahlmann, A. S., Furrer, A., Eichenseher, F., et al. (2016). Nonstable Staphylococcus aureus Small-Colony Variants Are Induced by Low pH and Sensitized to Antimicrobial Therapy by Phagolysosomal Alkalinization. *The Journal of Infectious Diseases*, *213*(2), 305–313. http://doi.org/10.1093/infdis/jiv388

Levin-Reisman, I., Gefen, O., Fridman, O., Ronin, I., Shwa, D., Sheftel, H., & Balaban, N. Q. (2010). Automated imaging with ScanLag reveals previously undetectable bacterial growth phenotypes. *Nature Methods*, *7*(9), 737–739. http://doi.org/10.1038/nmeth.1485

Lewin, A., Baus, D., Kamal, E., Bon, F., Kunisch, R., Maurischat, S., et al. (2008). The mycobacterial DNA-binding protein 1 (MDP1) from Mycobacterium bovis BCG influences various growth characteristics. *BMC Microbiology*, *8*(1), 91. http://doi.org/10.1186/1471-2180-8-91

Lin, P. L., Ford, C. B., Coleman, M. T., Myers, A. J., Gawande, R., Ioerger, T., et al. (2014). Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. *Nature Medicine*, *20*(1), 75–79. http://doi.org/10.1038/nm.3412

Luijsterburg, M. S., Noom, M. C., Wuite, G. J. L., & Dame, R. T. (2006). The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *Journal of Structural Biology*, *156*(2), 262–272. http://doi.org/10.1016/j.jsb.2006.05.006

Malm, S., Tiffert, Y., Micklinghoff, J., Schultze, S., Joost, I., Weber, I., et al. (2009). The roles of the nitrate reductase NarGHJI, the nitrite reductase NirBD and the response regulator GlnR in nitrate assimilation of Mycobacterium tuberculosis. *Microbiology (Reading, England)*, *155*(Pt 4), 1332–1339. http://doi.org/10.1099/mic.0.023275-0

Manina, G., Dhar, N., & McKinney, J. D. (2015). Stress and host immunity amplify Mycobacterium tuberculosis phenotypic heterogeneity and induce nongrowing metabolically active forms. *Cell Host & Microbe*, *17*(1), 32–46. http://doi.org/10.1016/j.chom.2014.11.016

Mirani, Z. A., Aziz, M., & Khan, S. I. (2015). Small colony variants have a major role in stability and persistence of Staphylococcus aureus biofilms. *The Journal of Antibiotics*, *68*(2), 98–105. http://doi.org/10.1038/ja.2014.115

Mukherjee, A., Bhattacharyya, G., & Grove, A. (2008). The C-terminal domain of HU-related histone-like protein Hlp from Mycobacterium smegmatis mediates DNA end-joining. *Biochemistry*, *47*(33), 8744–8753. http://doi.org/10.1021/bi800010s

Niki, M., Niki, M., Tateishi, Y., Ozeki, Y., Kirikae, T., Lewin, A., et al. (2012). A novel mechanism of growth phase-dependent tolerance to isoniazid in mycobacteria. *The Journal of Biological Chemistry*, *287*(33), 27743–27752. http://doi.org/10.1074/jbc.M111.333385

Oberto, J., Nabti, S., Jooste, V., Mignot, H., & Rouviere-Yaniv, J. (2009). The HU regulon is composed of genes responding to anaerobiosis, acid stress, high osmolarity and SOS induction. *PloS One*, *4*(2), e4367. http://doi.org/10.1371/journal.pone.0004367

Ogawa, T., Wada, M., Kano, Y., Imamoto, F., & Okazaki, T. (1989). DNA replication in Escherichia coli mutants that lack protein HU. *Journal of Bacteriology*, *171*(10), 5672–5679.

Pethe, K., Bifani, P., Drobecq, H., Sergheraert, C., Debrie, A.-S., Locht, C., & Menozzi, F. D. (2002). Mycobacterial heparin-binding hemagglutinin and laminin-binding protein share antigenic methyllysines that confer resistance to proteolysis. *Proceedings of the National Academy of Sciences*, *99*(16), 10759–10764. http://doi.org/10.1073/pnas.162246899

Proctor, R. A., Eiff, von, C., Kahl, B. C., Becker, K., McNamara, P., Herrmann, M., & Peters, G. (2006). Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nature Reviews. Microbiology*, *4*(4), 295–305. http://doi.org/10.1038/nrmicro1384

Sala, C., Forti, F., Magnoni, F., & Ghisotti, D. (2008). The katG mRNA of Mycobacterium tuberculosis and Mycobacterium smegmatis is processed at its 5' end and is stabilized by both a polypurine sequence and translation initiation. *BMC Molecular Biology*, *9*(1), 33. http://doi.org/10.1186/1471-2199-9-33

Salerno, P., Larsson, J., Bucca, G., Laing, E., Smith, C. P., & Flärdh, K. (2009). One of the two genes encoding nucleoid-associated HU proteins in Streptomyces coelicolor is developmentally regulated and specifically involved in spore maturation. *Journal of Bacteriology*, *191*(21), 6489–6500. http://doi.org/10.1128/JB.00709-09

Sánchez-Romero, M. A., & Casadesús, J. (2014). Contribution of phenotypic heterogeneity to adaptive antibiotic resistance. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(1), 355–360. http://doi.org/10.1073/pnas.1316084111

Sharadamma, N., Khan, K., Kumar, S., Patil, K. N., Hasnain, S. E., & Muniyappa, K. (2011). Synergy between the N-terminal and C-terminal domains of Mycobacterium tuberculosis HupB is essential for high-affinity binding, DNA supercoiling and inhibition of RecA-promoted strand exchange. *The FEBS Journal*, *278*(18), 3447–3462. http://doi.org/10.1111/j.1742-4658.2011.08267.x

Sharma, S. V., Lee, D. Y., Li, B., Quinlan, M. P., Takahashi, F., Maheswaran, S., et al. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*, *141*(1), 69–80. http://doi.org/10.1016/j.cell.2010.02.027

Shell, S. S., Prestwich, E. G., Baek, S.-H., Shah, R. R., Sassetti, C. M., Dedon, P. C., & Fortune, S. M. (2013). DNA methylation impacts gene expression and ensures hypoxic survival of Mycobacterium tuberculosis. *PLoS Pathogens*, *9*(7), e1003419. http://doi.org/10.1371/journal.ppat.1003419

Shell, S. S., Wang, J., Lapierre, P., Mir, M., Chase, M. R., Pyle, M. M., et al. (2015). Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. *PLoS Genetics*, *11*(11), e1005641. http://doi.org/10.1371/journal.pgen.1005641

Shires, K., & Steyn, L. (2001). The cold-shock stress response in Mycobacterium smegmatis induces the expression of a histone-like protein. *Molecular Microbiology*, *39*(4), 994–1009.

Sohaskey, C. D., & Modesti, L. (2009). Differences in nitrate reduction between Mycobacterium tuberculosis and Mycobacterium bovis are due to differential expression of both narGHJI and narK2. *FEMS Microbiology Letters*, *290*(2), 129–134. http://doi.org/10.1111/j.1574-6968.2008.01424.x

Sohaskey, C. D., & Wayne, L. G. (2003). Role of narK2X and narGHJI in hypoxic upregulation of nitrate reduction by Mycobacterium tuberculosis. *Journal of Bacteriology*, *185*(24), 7247–7256. http://doi.org/10.1128/JB.185.24.7247-7256.2003

Stermann, M., Bohrssen, A., Diephaus, C., Maass, S., & Bange, F.-C. (2003). Polymorphic nucleotide within the promoter of nitrate reductase (NarGHJI) is specific for Mycobacterium tuberculosis. *Journal of Clinical Microbiology*, *41*(7), 3252–3259. http://doi.org/10.1128/JCM.41.7.3252-3259.2003

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550. http://doi.org/10.1073/pnas.0506580102

Sureka, K., Ghosh, B., Dasgupta, A., Basu, J., Kundu, M., & Bose, I. (2008). Positive feedback and noise activate the stringent response regulator rel in mycobacteria. *PloS One*, *3*(3), e1771. http://doi.org/10.1371/journal.pone.0001771

Swinger, K. K., & Rice, P. A. (2004). IHF and HU: flexible architects of bent DNA. *Current Opinion in Structural Biology*, *14*(1), 28–35. http://doi.org/10.1016/j.sbi.2003.12.003

Tuchscherr, L., Medina, E., Hussain, M., Völker, W., Heitmann, V., Niemann, S., et al. (2011). Staphylococcus aureus phenotype switching: an effective bacterial strategy to escape host immune response and establish a chronic infection. *EMBO Molecular Medicine*, *3*(3), 129–141. http://doi.org/10.1002/emmm.201000115

Wakamoto, Y., Dhar, N., Chait, R., Schneider, K., Signorino-Gelo, F., Leibler, S., & McKinney, J. D. (2013). Dynamic persistence of antibiotic-stressed mycobacteria. *Science (New York, N.Y.)*, *339*(6115), 91–95. http://doi.org/10.1126/science.1229858

Walter, N. D., Dolganov, G. M., Garcia, B. J., Worodria, W., Andama, A., Musisi, E., et al. (2015). Transcriptional Adaptation of Drug-tolerant Mycobacterium tuberculosis During Treatment of Human Tuberculosis. *The Journal of Infectious Diseases*, *212*(6), 990–998. http://doi.org/10.1093/infdis/jiv149

Wang, W., Li, G.-W., Chen, C., Xie, X. S., & Zhuang, X. (2011). Chromosome organization by a nucleoid-associated protein in live bacteria. *Science (New York, N.Y.)*, *333*(6048), 1445–1449. http://doi.org/10.1126/science.1204697

Warner, D. F., Savvi, S., Mizrahi, V., & Dawes, S. S. (2007). A riboswitch regulates expression of the coenzyme B12-independent methionine synthase in Mycobacterium tuberculosis: implications for differential methionine synthase function in strains H37Rv and CDC1551. *Journal of Bacteriology*, *189*(9), 3655–3659. http://doi.org/10.1128/JB.00040-07

Weng, X., & Xiao, J. (2014). Spatial organization of transcription in bacterial cells. *Trends in Genetics : TIG*, *30*(7), 287–297. http://doi.org/10.1016/j.tig.2014.04.008

Whiteford, D. C., Klingelhoets, J. J., Bambenek, M. H., & Dahl, J. L. (2011). Deletion of the histone-like protein (Hlp) from Mycobacterium smegmatis results in increased sensitivity to UV exposure, freezing and isoniazid. *Microbiology (Reading, England)*, *157*(Pt 2), 327–335. http://doi.org/10.1099/mic.0.045518-0

World Health Organization. (2015). Global tuberculosis report 2015.