DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Elucidation of an essential genetic pathway under antibiotic selection in Mycobacterium tuberculosis

Citation
Liu, Yue Jane. 2024. Elucidation of an essential genetic pathway under antibiotic selection in Mycobacterium tuberculosis. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Permanent link
https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37378990

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Division of Medical Sciences

in the subject of Biological and Biomedical Sciences

have examined a dissertation entitled

Elucidation of an essential genetic pathway under antibiotic selection in Mycobacterium tuberculosis

presented by Yue Jane Liu

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature: Simon Dove
Simon Dove (Apr 22, 2024 13:59 EDT)

Typed Name: Dr. Simon Dove

Signature: David Rudner (Apr 24, 2024 13:30 EDT)

Typed Name: Dr. David Rudner

Signature:

Typed Name: Dr. Michael Starnbach

Signature: Shumin Tan
Shumin Tan (Apr 22, 2024 12:55 EDT)

Typed Name: Dr. Shumin Tan

Date: April 22, 2024

Elucidation of an essential genetic pathway under antibiotic selection in *Mycobacterium tuberculosis*

A dissertation presented

by

Yue Jane Liu

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

April 2024

© 2024 Yue Jane Liu

All rights reserved.

Dissertation Advisor: Sarah M. Fortune Yue Jane Liu

Elucidation of an essential genetic pathway under antibiotic selection in *Mycobacterium tuberculosis*

Abstract

Tuberculosis remains the world's deadliest infectious disease caused by a single agent. Although tuberculosis is curable, treatment success is limited by our narrow understanding of genetic factors allowing its causative agent, *Mycobacterium tuberculosis* (*Mtb*), to evade antibiotic clearance. Large-scale sequencing of clinical *Mtb* populations revealed ongoing selection on genetic variants that could confer fitness advantages in the presence of drug pressure. This unbiased approach allowed identification of genes with no previous link to drug resistance, including two essential genes *dnaA* and *resR*. Although initially investigated independently, DnaA and ResR share a common binding site at the *Rv0010c-Rv0011c* intergenic region (IGR) and this IGR itself is one of the highly mutated non-coding regions on *Mtb* genome. Clinical IGR variants overlap with DnaA and ResR binding sites and phenocopy *dnaA* and *resR* variants, revealing a genetic pathway under selection. Yet this genetic pathway and the function of *Rv0010c-Rv0011c* IGR remains uncharacterized.

Isogenic variants in the *Rv0010c-Rv0011c* IGR phenocopy *dnaA* and *resR* variants, showing similar increases in cell length, antibiotic resilience, and low-level isoniazid resistance. We found that DnaA and ResR bind at neighboring sites in the most conserved regions of this IGR, which paradoxically are where more recent clinical mutations accumulate. Knockout of the entire the *Rv0010c-Rv0011c* operon, including its 155bp IGR, resulted in shorter cells with

increased sensitivity to isoniazid. This defect can only be complemented with the entire operon, though this complementation does not require translation of the two coding genes. Complementation with the intact operon carrying clinically relevant IGR variants recapitulates isogenic variant phenotypes. Meanwhile, complementation with the intact operon carrying DnaA or ResR binding site deletions failed, highlighting the requirement of protein binding in its downstream function.

To understand the functional consequence of protein binding, we used biochemical approaches and found that DnaA and ResR bind cooperatively at this IGR. Clinical IGR variants increase the binding affinity of two proteins and binding site deletions reduce their affinity. Using transcriptomics, we identified genes that are differentially expressed in strains with clinical IGR variants versus strains with binding site deletions to pinpoint transcriptional changes correlating with divergent phenotypes in these strains. These genes included *whiB2* and its regulon of division related genes. Notably, the promoter of *whiB2* is a known direct target of ResR and is also highly mutated in clinical *Mtb* populations. We propose a model where clinical variations sequester ResR through its interaction with DnaA at the *Rv0010c-Rv0011c* IGR. This sequestration reduces ResR's ability to activate division related genes and alter division dynamics, resulting in morphology and drug phenotypes.

Together, the data in this dissertation provide functional insight into an essential and previously uncharacterized genetic pathway under selection in clinical *Mtb* populations. We propose that clinical mutations in this pathway alter dynamics of cell cycle events and contribute to changes in *Mtb* morphology and drug response. Understanding non-canonical drug determinants is critical to elucidate other mechanisms *Mtb* use to evade antibiotics killing and we

hope to inspire future studies on intergenic regions and unknown genetic pathways to better understand *Mtb* biology and improve treatment design.

Table of Contents

Acknowledgements

Graduate school was a journey and a marathon, and I would not have done it without the support of everyone that helped me along the way. I want to first thank my advisor, Dr. Sarah Fortune for her continuous support and enthusiasm. Sarah approaches every project and every data point with so much excitement and she's always down to brainstorm. We have come up with many "conspiracy stories" throughout our discussion and it was always so fun to speculate about science in general with Sarah. In addition to scientific wisdom, Sarah has also taught me how to be an effective communicator. Sarah is an amazing presenter, and I had the honor to learn from the best about how to engage the audience during a presentation. I want to also thank Sarah specifically for her support throughout this graduation process. She helped me keep on track in the last year working towards graduation and even took time on a Sunday to walk me through a last-minute breakdown prior to submitting this thesis. Sarah always has a smile on her face and she greets everyone she sees. I am really going to miss her energy and all the times I get to pop into Sarah's office and talk to her about all the random hypotheses I come up with.

Next, I want to thank the Fortune lab for their support throughout the last five years. I started in the lab during the 2020 pandemic, and it was a strange time looking back for a graduate student to start. When I first started, we were not allowed to talk within six feet and not allowed to be in the same room or the same bay in the lab. Despite these physical barriers, postdocs in the lab, Qingyun Liu, Xin Wang, Peter Culviner, and Junhao Zhu, still reached out to me and took time to discuss my project over zoom. They continued to support me throughout the entire time in the lab and always set aside time to discuss with me about my project. They helped me get my project up and going and I really appreciated them helping me thinking through experiment details. I would like to specially thank Junhao and Xin. Junhao was a postdoc in Eric's lab and yet he still took so much time mentoring me. I have learnt and grew so much as a microbiologist thank to Junhao. Both Junhao and Xin also helped me through my graduate school slump. I was feeling very lost and unmotivated at one point of my career and Junhao and Xin took time to have lunch with me and walk me through my mental blocks. It was a difficult time, and I would be so much more miserable if they weren't there. In the lab, I would also like to thank Shoko Wakabayashi for all her help in the BSL3 and for answering all my random questions about working with *Mtb*. I would also like to thank people in the Rubin lab for their feedback during our group meetings and also company on the 8th floor. I want to thank all previous and current graduate students on the floor in both Fortune and Rubin lab: Greg Babunovic, Samantha Giffen, Sydney Stanley, Francesca Tomasi, Harry Won, Kerry McGowen, Abigail Frey, Maia Mesyngier, and Sam Zinga. It was great to have a little student community and grow with all of you. I would like to specifically thank Greg for being my BBS and my rotation mentor, and for his energy that always reminds me to be optimistic about life.

I want to thank all the committee members that have supported me in various stages of graduate school. I want to thank my PQE committee, Dr. Simon Dove, Dr. Michael Starnbach, and Dr. Jonathan Abraham, who helped me formulate my project early on. I want to thank my DAC members, Dr. Simon Dove, Dr. Tom Bernhardt, and Dr. Bree Aldridge, who have given valuable feedback and suggestions on how to interpret and move forward with my data and hypotheses. Lastly, I want to thank my thesis defense committee, Dr. Simon Dove, Dr. David Rudner, Dr. Michael Starnbach, and Dr. Shumin Tan, for the insightful discussion during the private exam section. I want to specifically thank Dr. Simon Dove for making the time and being on every single committee of mine and for supporting me my entire graduate career.

I want to also thank Dr. Tom Bernhardt for having me as a rotation student in the fall of 2019, which allowed me to meet and build a friendship with the coryne crew. Despite me not joining the Bernhardt lab, postdocs on the coryne side of the lab, Dr. Amelia McKitterick, Dr. Betsy Hart, Dr. Anastacia Parks and Dr. Wanassa Beroual, continued to help and mentor me throughout my time in grad school. We have a standing meeting every Thursday and they have given me so many suggestions on how to do genetics in mycobacteria and have supported me whole heartedly while I navigated through this project. They are also just a fun group to hang out with. The coryne club is always excited about ideas with or without data, and I have learned so much on both how to do science and how to present a chalk talk. I really appreciated having coryne crew as my support group in grad school.

I also want to thank all the previous mentors that helped me along the way prior to graduate school. At Bryn Mawr, I was fortunate to join Dr. Joshua Shapiro's lab at the beginning of my sophomore year and was able to learn so many basic lab techniques while working with environmental yeast samples. In the summers of college, I got to work in Dr. Fred Winston's lab at HMS. In the spring of my sophomore year, I cold emailed Fred and he was kind enough to consider having me over in the summer and then having me back the summer of my junior year too. I learned so much about how to be a scientist in both the Shapiro lab and Winston lab. They really cultivated my initial interest in scientific research, which inspired me to go to grad school. Before grad school, I also got the opportunity to work and train in Dr. Michael Laub's lab at MIT. There, I was paired with a postdoc at the time in the lab, Dr. Michele LeRoux. During my time in the Laub lab, Michele has taught me how to read, think, and present like a scientist. I also learned numerous microbiology techniques from her. I want to think Mike and Michele for providing me with this learning experience. Mike had a great lab, and people in the lab supported me throughout my time as a technician and during grad school application. I would not have made it to where I am without the support of all these mentors I met along the way.

I want to also thank all my friends for their support during grad school. First, I want to thank the "Friendsgiving 2019" group: Rita Chan, Kristin Tsuo, Hugh Chen, Rojesh Shrestha, David Lowe, Della Syau, and Kat Warner. We met at the beginning of grad school and have become close friends very fast. We go out to eat or have little potlucks at home and we have our annual Friendsgiving and dumpling parties. Their company helped me get through the last few years and it was really nice to have friends who were going through the same process. I would also like to thank my rowing friends. I spent a lot of mornings on the Charles river since 2021. I got to train and compete with three other rowers Kevin Dutt, Libby McCann, and Jeff Nelson. Being on the river was the only time where I stopped worrying about everything and it was a very good de-stressing activity to have in grad school. Through my time rowing on the river with various groups, I also met another set of rowers, Muriel Drexler, Meredith Crenca, Erin O'Connell, Meg Lydon. Despite only rowed together for a few months, we developed a bond quickly and continued to hang out. We made time for each regardless of how busy we were. All of lives have changed a lot since we met and yet we remained close friends. I really appreciate the bond I built with them. I also want to give a shoutout to Francisco Lopez, Namir Jawdat, and Ana Clara Tolentino, who are friends I met during my time rowing at MIT. Namir held frequent weekend backyard barbeques at his place, and it was great hanging out with all of them outside and having them cheer me through this tough journey.

Next, I want to thank family members for their support. I want to first thank the Leopold family, especially Bill Leopold and Jeanne Goldberg-Leopold for making a trip to Boston to

attend my defense in person. The Leopold family has supported me throughout grad school, and I really appreciate them as my family here in the US. I want to thank my own family back in China. They have been supporting me and cheering me on from China since I left home for the US for college in 2013. I would like to specifically thank my parents, Ding Wen and Liu Chaoyang. I have learned from them since day 1 and I would be nowhere without them. When I was growing up, they have given me the freedom to learn and gain independence, which was extremely valuable for me at a young age. They also signed me up for English lessons when I was very young, and it really helped me learn the language and continue to use it daily for many years prior to moving to the US. When I was in college, my mom was the one who encouraged me to reach out to Fred at HMS by email, and I would not be here if it weren't her. My mom worked in a chemistry lab when she was pregnant with me, and I attribute my curiosity about science to early introduction in my life. Unfortunately, my mom passed away in 2017 and never got to see me go to grad school. Although she is not here physically, she has been with me this entire journey and I know she will always have my back and support me in every moment.

Lastly, I want to thank my husband Aaron Leopold and my baby Pepperoni Liu-Leopold. Aaron has to endure all the ups and downs of grad school with me and he had been there for me every time. Aaron moved to Boston for me away from his favorite city Philly and made sure I had everything I needed whenever I needed. Thesis writing was a very stressful period and Aaron made sure I had plenty of food and coffee. We had gone through a global pandemic and a PhD career together and I can't wait to continue conquering life with Aaron. I also want to thank my baby pup Pepperoni. Aaron and I picked pepper up two days before my first grad school interview and I can't imagine doing grad school without pepper. Walking her every day kept me in touch with the world during the lockdown and cuddling her made all the downs in grad school a lot more bearable. She had grown with me through all these years and she will always be my baby.

This PhD was a long journey, and I was lucky to have the help and support from a lot of people. Thank you!

This thesis is dedicated to my mom, Ding Wen, who taught me everything.

Chapter 1: Introduction

1.1 The ongoing endemic of Tuberculosis

Tuberculosis (TB) is one of the oldest infectious diseases affecting mankind (1). The disease was first named tuberculosis in 1839 and evidence of tuberculosis can be traced all the way back to Ancient Egypt (2, 3). In 1882, Robert Koch first identified the causative agent of TB, the bacterium *Mycobacterium tuberculosis* (*Mtb*). Despite broad advances in technology and medicine since this discovery, there are still approximately 10 million new TB cases and 1.5 million deaths globally each year (4). Estimations have found that roughly 25% of the global population has been infected by *Mtb,* but only 5-10% of the infected population continues to develop TB disease in their lifetime (5, 6).

As a bacterial disease, TB is treatable and curable by antibiotics. However, TB treatment is lengthy and requires patients to take combinations of four antibiotics, including isoniazid (INH), rifampicin (RIF), ethambutol (EMB) and pyrazinamide (PZA), daily for up to six months for drug susceptible TB (7). Drug-resistant TB requires patients to take antibiotics for up to 18 months (8). Since the introduction of isoniazid in the 1950s, combination therapy has been shown to have great treatment success and reduce the emergence of antibiotic resistance (9). Despite the effectiveness of the current chemotherapy since its introduction in the 1980s, treatment success rate remains at 85% on average for drug susceptible TB, and just 63% for drug resistant TB (4, 10-12). Treatment failure can be driven by multiple factors including: 1) the genetic heterogeneity of the infectious agent, *Mycobacterium tuberculosis* (*Mtb*), and its resistance to antibiotics used to treat TB; 2) the complexity of the immune environment and inconsistent penetration of antibiotics in granulomas; 3) patient non-compliance with lengthy treatment regimens and/or inadequate access to healthcare, which then fuels the emergence of resistant bacteria in the population and increases the difficulty of subsequent treatments (13-15).

Antibiotic resistance has been on a steady rise over the last decade and in 2022, 4% of total cases were drug resistant (4). Recent work has aimed to develop better TB treatments tailored to address the issue of bacterial resistance, but these efforts are hindered by our limited understanding of *Mtb* biology and the impact of bacterial genetic factors on drug susceptibility and treatment outcomes (14).

1.2 The complexity of *Mtb* drug resistance

The discovery of penicillin in the late 1920s revolutionized the treatment of bacterial diseases and led to the discovery of many other antibiotics including streptomycin in 1943, which was used to treat TB (16-18). However, the emergence of differentially susceptible *Mtb* population to streptomycin was reported within the first five years after the advent of streptomycin treatment. This provided the first evidence of the immense selective pressure that drives acquired resistance in *Mtb* (19). Over the last few decades, advances in molecular tools and sequencing methods allowed the discovery of the genetic basis of drug resistance to first-line TB drugs such as isoniazid (INH) and rifampicin (RIF) (20). Most drug resistance is driven by single nucleotide changes within coding regions or insertions or deletions of regulatory regions of target genes or genes required for drug function (21). For example, INH resistance largely associates with mutation in the *katG* gene, encoding the catalase enzyme KatG required to convert isoniazid from its pro-drug form to active form *in vivo* (22, 23). RIF resistance mostly arises from mutations in a specific region within the *rpoB* gene. *rpoB* encodes the target of RIF, the β-subunit of RNA polymerase, and mutations in this region are thought to induce conformational change and decrease binding affinity of RIF to the polymerase (20, 24). Both *katG* and *rpoB* mutations result in high level resistance, which increases the minimum inhibitory

concentration (MIC) of the strain above a critical threshold (25). MIC is currently defined as the lowest concentration that inhibits visible *Mtb* growth *in vitro* and is used as primary method for determining drug resistance and designing drug regimens in the clinic (26). Unfortunately, TB is not an easy-to-treat disease and drug resistance is less dichotomous than what is measured by MIC assays. In this new genomics era, the field has started to learn more about genetic changes associated with phenotypic susceptibility through whole genome sequencing, and it has become clear that high level resistance mutations are not the only clinically relevant way for *Mtb* to evade antibiotics (27, 28).

Advances in sequencing technologies in the last decade allowed researchers to perform whole genome sequencing on clinical *Mtb* isolates with known drug susceptibility profile and to perform genome-wide association studies (GWAS) to identify additional genetic determinants of drug resistance (27-31). Canonical high-level resistant mutations in genes such as *katG* and *rpoB* were identified along with a surprising set of genetic loci, coding and non-coding, previously unrelated to drug mechanism of action or metabolism (27). Genetic studies of these variants show that most do not result in MIC increase above the clinical breakpoint (28, 29, 32). However, recent clinical data has shown that even low-level MIC shift is sufficient to increase likelihood of treatment failure and TB relapse (33). This provides evidence for alternative mechanisms *Mtb* uses to evade antibiotic clearance. It is critical to dissect the functional consequences of these genetic variations to inform the design of comprehensive chemotherapy.

Over the last decade, genetic studies in *Mtb* have led to discoveries of many non-canonical drug resistance determinants. Bacteria are known to be able to survive extensive antibiotic treatment without acquiring high-level resistance through mechanisms such as tolerance and persistence (34). In *Mtb*, a subset of these non-canonical determinants results in

low level increase in resistance to first-line and second-line TB drugs (29, 35, 36). Though these small increases in MIC do not reach the clinical breakpoint to be formally categorized as antibiotic resistance, they do significantly increase the rate of relapse in patients (26, 33). The non-canonical determinants include *dnaA*, and mutation in *dnaA* confers low-level MIC shift to INH as a result of slight downregulation of the *katG* gene (29). *dnaA* encodes the replication initiator protein and is one of the most conserved and essential genes across bacteria, and yet mutations in *dnaA* are found in 3.2% of global *Mtb* isolates (29, 37). Another subset of novel drug determinants induce multi-drug tolerance that can be dependent or independent on host relevant carbon source (28, 38-42). Interestingly, and somewhat concerningly, existing mutations in *Mtb* already confer low-level resistance in new antitubercular drugs currently in clinical trials. This again shows how genetic heterogeneity in *Mtb* makes it notoriously difficult to treat (43).

In addition to low-level resistance and antibiotic tolerance, a new antibiotic phenotype, antibiotic resilience, was discovered in *Mtb* in 2022 (32). Genomic analysis of 50,000 global clinical *Mtb* isolates from patients identified enriching mutations in the gene encoding the regulator of resilience, *resR*. Genetic studies show that a single nucleotide change in *resR* resulted in faster recovery to both first and second-line TB antibiotics post antibiotic treatment compared to wild-type strain (32). Antibiotic recovery is highly clinically relevant but remains a very understudied component of *Mtb* drug response (15). Due to limits of pharmacokinetics, serum measurement of drug concentration following daily dose in patients with drug susceptible TB shows plasma concentrations of INH, RIF, and EMB fall below MIC for 21%, 42%, and 75% of treatment time respectively (44). Therefore, ability to regrow in a more permissive environment provides clinical advantage in the context of infection.

Recent technological advances have allowed scientists to better comprehend the complexity of *Mtb* drug resistance and to start untangling its genetic basis to aid in development of new chemotherapy. The rest of this chapter focuses on investigation of a subset of these variants, featuring *dnaA*, *resR*, and others in pathways controlling mycobacterial growth and division.

1.3 Clinical selection on mycobacterial replication and division

Genomic analyses of mutations enriched in *Mtb* clinical isolates identified coding and non-coding regions under positive selection, which refers to their ability to acquire mutations repeatedly in independent cases. *dnaA* and *resR* are among the top genes selected along with genes conferring high-level resistance such as *katG* and *rpoB* (32). Both *dnaA* and *resR* were first identified as novel genes strongly associated with INH resistant TB cases in a large strain collection from China (28, 29). *dnaA* (*Rv0001*) encodes the replication initiator protein DnaA and binding of DnaA to the origin of replication (*oriC*) directly regulates the progression of cell cycle (37, 45). DnaA is essential and its function in replication is well conserved across almost all bacteria (37, 46, 47). In *Bacillus subtills* (*B. subtills*), DnaA can also function as a transcriptional regulator and both directly and indirectly regulate expressions of cell cycle related genes (48, 49). Despite the essentiality of *dnaA* and its function in replication, *Mtb* accumulates many non-synonymous mutations in *dnaA*. These clinically relevant mutations in *dnaA* do not alter growth and replication dynamics but do increase steady state cell length. They also increase INH MIC by ~2 fold and this small MIC shift is thought to be a result of downregulation of *katG* (29). This low-level MIC difference is similar in magnitude to the MIC shift that is sufficient to increase risk of treatment failure in patients (33). Given the fact that DnaA is known to bind

regulatory regions and directly affect downstream gene expression in *B. subtills*, earlier studies used *in vitro* DNA affinity purification and sequencing (IDAP-seq) to identify direct binding sites of DnaA on the *Mtb* genome. They found that *DnaA* binds a cryptic intergenic region between *Rv0010c* and *Rv0011c* (*Rv0010c-Rv0011c*), in addition to its promoter and the promoter of gene encoding DNA helicase *dnaB*. *Rv0010c-Rv0011c* mutations phenocopy *dnaA* variants in both drug and morphology phenotype (29). The function of this intergenic region is unclear, but it has been associated with INH resistant TB cases in previous genomic analyses and is under positive selection in the large genomics study of 50,000 clinical isolates (27, 28, 32). The *Rv0010c-Rv0011c* intergenic region sits in between two co-transcribed genes, *Rv0010c* and *Rv0011c*. *Rv0011c* encodes a division associated protein CrgA, named after its homolog in streptomyces, and locates to the septum during early divisome formation. CrgA can directly interact with divisome components and deletion of *Rv0011c* leads to division defects (50, 51). *Rv0010c* encodes a conserved membrane protein of unknown function (52).

The other essential gene identified in both GWAS and the large genomics study is *resR* (*Rv1830*). *resR* function was unknown until a recent study demonstrated its link to antibiotic resilience and named it the regulator of resilience (ResR) (32). *resR* is a conserved and essential protein and has structural homology to *merR* like transcription factors. Clinical mutations in *resR* overlap with its predicted DNA binding domains in the protein structure and are thought to alter ResR affinity to its downstream targets and affect expression. Using *in vitro* DNA affinity purification and sequencing (IDAP-seq), ResR was shown to bind several genomic regions including upstream of *whiB2* (*whiB2-fbiA*), and mutations in *resR* directly alter *whiB2* expression (32). Interestingly, previous studies have shown mutations in *whiB2-fbiA* to be strongly associated with drug resistant TB cases, and these mutations directly overlap with the ResR

binding site (27, 32). *whiB2-fbiA* mutants largely phenocopy *resR* mutants, indicating convergent clinical selection on this pathway (32). WhiB2 (Rv3260c) is an essential transcriptional regulator governing growth and division in mycobacteria and altered *whiB2* expression leads to growth and division defect (53, 54). In a recent study, WhiB2 was shown to regulate expression of division related genes both directly through direct binding to promoters of its regulon and indirectly through regulation of *mtrA*, encoding an essential transcriptional regulator in the two-component system MtrA-B (55, 56). WhiB protein was also shown to function with another essential transcription factor WhiA in a closely related streptomyces species to control cell division (57). *whiA* (*Rv1423*) in *Mtb* is also under positive selection by genomic analysis and enriched in mutations in clinical populations, but remains less understood in mycobacteria (32). Clinical data has shown mutations in *resR*, *whiB2-fbiA*, and *whiA* are directly associated with patient relapse post antibiotic treatments (32). The study of *resR* led to the discovery of a downstream pathway surrounding regulation of cell division, which is thought to contribute to antibiotic resilience.

Our studies of clinical mutations in *dnaA* and *resR* showed two independent instances where essential cell cycle related processes in *Mtb*, replication and division, are under selection and highly mutated. In bacteria, replication and division are fundamental processes of the bacterial life cycle and coordination of these two processes is crucial to maintain the optimal physiology and metabolic state of bacterial populations. Therefore, mutations in these highly essential pathways are unexpected and modifications to regulation of growth and division can be detrimental. However, one beneficial instance of alteration to growth and division is the ability to adjust and allow bacterial adaptation to different environmental and stress conditions (58). In *Escherichia coli* (*E. coli*), altered cell size and physiology can directly impact bacterial fitness

and response to antibiotics (59, 60). The importance of growth and division in mycobacteria remains poorly understood. Given antibiotic resilience is a phenomenon driven primarily by bacterial growth and division, we speculated that selection on *dnaA* and *resR* could be an example of *Mtb*'s adaptation to antibiotic stress. Therefore, we hypothesized that investigating the growth and division pathway under selection in *Mtb* downstream of *dnaA* and *resR* could shine light on how *Mtb* might utilize altered physiology to gain advantage and survive antibiotic clearance.

1.4 Known roles of DnaA and ResR in mycobacterial cell cycle regulation

DnaA and ResR regulate replication and division, which are two essential processes in the bacterial cell cycle. Viability of bacterial cells heavily relies on the proper coordination of cell cycle events including chromosome replication, growth, and division. Unlike eukaryotes, bacteria do not have checkpoints to facilitate progression of the cell cycle and instead rely on regulatory elements to ensure proper timing of each step. Traditionally, the bacterial cycle consists of three main periods: 1) B period, referring to the period between division and initiation of replication; 2) C period, chromosome replication; and 3) D period, the period between the end of replication and division. The timing of these periods can be different especially between fast and slow growing bacteria, depending on the number of rounds of simultaneous replication needed in order to accommodate the faster growth rate (58, 61). Mycobacterial species follow similar organization of the cell cycle, with an extended B period in slow growing bacteria such as *Mtb* (62, 63). Studies on *dnaA* and *resR* are among the first set of evidence that essential processes such as cell cycle regulation could directly contribute to *Mtb* clinical outcomes.

However, our incomplete understanding makes it challenging to pinpoint the specific clinical advantage of altering such processes.

DnaA, encoded by *dnaA*, is the replication initiator protein and initiates the first step of the cell cycle: replication. DnaA is capable of initiating replication only when bound to ATP, and its binding and hydrolysis of ATP is tightly regulated to ensure proper timing of replication initiation. In *E. coli*, replication starts when ATP-bound DnaA (ATP-DnaA) binds to the origin of replication (*oriC*). Oligomerization of ATP-DnaA at *oriC* assists in loading DNA helicase to unwind double stranded DNA around *oriC* and facilitates binding of DNA polymerase to start replication. After initiation, ATP-DnaA is deactivated by either protein (Hda) or DNA (*datA*) induced ATP hydrolysis into its inactive form ADP-DnaA. Once the next round of replication is ready to initiate again, ADP-DnaA oligomerize at its re-charging sequences (*DARS*) along the genome. This oligomerization promotes dissociation of ADP and allows the apo enzyme to be released and available to bind ATP (37, 64).

In mycobacteria, replication is regulated by conserved DnaA and its binding to the origin (*oriC*), but mycobacterial genomes lack homologs of corresponding proteins and sequences for deactivating and activating of ATP-DnaA. It has been shown that *Mtb* ATP-DnaA go through active hydrolysis when oligomerizing on *oriC* and only oligomerization of ATP-DnaA can induce productive conformation change to DNA to initiate replication. Therefore, DnaA is actively deactivated at the *oriC* following initiation and misfiring of initiation is thought to be blocked by binding of DnaA elsewhere on the genome and binding of other proteins such as MtrA at *oriC* (37, 65, 66). In addition to initiating replication, DnaA has been shown in *B. subtills* to function as a transcriptional regulator both directly by binding upstream of its regulon genes or indirectly by acting through a nucleoid-associated protein Rok (48, 49). The analog of Rok in mycobacteria

is Lsr2, a histone-like nucleoid structuring (H-NS) protein, and deletion of Lsr2 directly affects replication dynamics (67, 68). A direct link between DnaA and Lsr2 has yet to be established in *Mtb*, though a previous study using chromatin immunoprecipitation (ChIP)-seq showed binding of Lsr2 in the *Rv0010c-Rv0011c* intergenic region where DnaA binds and clinical mutation accumulates (69). In *Caulobacter crescentus*, cell cycle coordination from replication to division is mediated by DnaA binding directly to the promoter of genes controlling growth and division, forming an expression gradient of genes initiating each period throughout cell cycle (70). It remains unclear how mycobacteria coordinate their cell cycle stages and if there is crosstalk between replication and division regulators.

The role of ResR in cell cycle regulation is less characterized and is mainly associated with WhiB2 encoded by one of the genes in ResR regulon. WhiB2 belongs to a family of essential transcription factors in mycobacteria, the WhiB family. Proteins in this family contain signature cysteine residues that bind oxygen and iron-sulfur clusters and are highly conserved in Actinobacteria. In *Mtb*, several WhiB proteins are associated with virulence, antibiotic resistance, and adaptation to environmental stress (71-73). ResR binds directly to the promoter of *whiB2*, and acts as a transcriptional activator of *whiB2* (32). WhiB2 and its homologs in other actinobacterial species have previously been shown to regulate expression of cell division related genes (53, 71, 74). This includes *ftsZ*, whose protein polymerization is required for Z-ring assembly and for initiation of division. The intrinsic GTPase activity of FtsZ drives polymerization and treadmilling, and forms the Z ring. Following FtsZ polymerization, SepF, whose expression is also directly regulated by WhiB2, guides placement of FtsZ polymers and ensures the proper formation of septum at correct division site in mycobacteria (55, 74-76). As an early divisome protein, SepF also helps recruit other proteins of the divisome including other

WhiB2 regulon members such as FtsQ and FtsW (77). CrgA, encoded by *Rv0011c*, is also part of the divisome and recruits another WhiB2 regulated protein, FtsI, which is a monofunctional peptidoglycan synthase required for septum assembly and known as PBP-B (50, 51, 78). At the last stage of division, septal hydrolysis is primarily carried out by peptidoglycan hydrolase RipA, which is also regulated by WhiB2 and finalizes the process of separation into two daughter cells (55, 79). After division, mother cell splits into daughter cells who will start growing independently in their next cell cycle. In contrast to *E. coli* and *B. subtills*, mycobacteria grow and elongate by adding new cell wall material only at the polar regions, and these two poles are intrinsically different in age and grow at different rates after birth. Daughters generate one nascent pole at division, which requires recruitment of elongasome proteins, and inherit one pole from the mother (80-83). Anchoring of elongation complexes at the poles requires Wag31, the mycobacterial homolog to DivIVA that is known as the membrane curvature sensing protein in *B. subtills* (84). Localization of Wag31 at the poles requires cell wall synthesis protein A (CwsA) and helps recruit other proteins in the elongation complex through direct protein interaction (51, 83). To ensure proper polar growth right after birth, Wag31, also under WhiB2 regulation, is known to interact directly with FtsI in the divisome and is recruited to the septum prior to daughter cell separation at the late stages of division (81, 85). As an essential regulator, WhiB2 regulates expression of genes across all stages of cell division. Since ResR directly regulates *whiB2* expression, ResR likely plays a direct role in cell division regulation.

Based on existing knowledge in the field, it is clear that DnaA and ResR regulate the two essential cell cycle processes in mycobacteria. However, there remains a missing link between these two processes. As a slow grower, *Mtb* is thought to replicate its chromosome once per cell cycle (62). Thus, communication between replication and division would be crucial to ensure

proper timing of machinery assembly and to avoid premature initiation of replication or division (86). Yet it remains unclear how *Mtb* facilitates transition between cell cycle stages. In addition, given WhiB proteins' ability to sense redox and environmental stress, there is a unique possibility that cell cycle processes can directly respond to stress conditions in *Mtb*. In this dissertation, we present functional characterization of a pathway connecting DnaA and ResR in *Mtb*. This pathway is under antibiotic selection in *Mtb* and provides an example of direct impact of cell cycle regulation in *Mtb* drug response.

1.5 Dissertation structure

Large-scale genomic sequencing of clinical *Mtb* populations identified *dnaA* and *resR* as two top-selected coding genes. *dnaA* and *resR* code for two cell cycle regulators DnaA and ResR, regulating replication and division. Though no previous link between these two proteins had been established, work in this dissertation identified a common binding site at an intergenic region *Rv0010c-Rv0011c,* which is also highly mutated in clinical *Mtb* isolates. Here, we aim to functionally characterize *Rv0010c-Rv0011c* operon and the consequences of variations in this genetic pathway under antibiotic selection. In Chapter 2, we present our data on discovery of this genetic pathway and genetic characterization of the *Rv0010c-Rv0011c* operon. Isogenic variants in *Rv0010c-Rv0011c* IGR phenocopy *dnaA* and *resR* variants and show similar increase in cell length and low-level increase in isoniazid (INH) resistance and antibiotic resilience. We found that DnaA and ResR bind at neighboring sites in the most conserved regions of this IGR and these sites are also where clinical mutations accumulate. To answer why these conserved regions are highly mutated, we first characterized the function of the *Rv0010c-Rv0011c* operon consisting of two coding genes *Rv0010c*, *Rv0011c* and the 155bp intergenic region. Knockout of

this operon in *Mtb* resulted in much shorter cells with increased sensitivity to isoniazid. This morphology and drug sensitivity defect can only be complemented with intact genomic sequence of all three elements and this complementation does not require translation of the two coding genes. Complementation with the intact operon carrying clinically relevant single nucleotide variations (SNVs) in IGR recapitulates phenotypes in isogenic variants while complementation with the intact operon carrying deletion of DnaA or ResR binding site failed to complement. Our data suggested a non-coding regulatory role of *Rv0010c-Rv0011c* IGR in morphology and drug phenotypes and that binding dynamics of DnaA and ResR to this region directly affect its regulatory function.

In the next chapter, we use biochemical and transcriptomic approaches to understand the consequences of variations in *Rv0010c-Rv0011c* IGR in the binding dynamics of DnaA and ResR and in transcriptional regulation. We found that DnaA and ResR bind cooperatively at this IGR *in vitro* and likely form a complex, and that clinical SNVs increase binding affinity of the complex while binding site deletions reduced affinity of the complex. After discovering the difference in binding dynamics and phenotypes in complementation strains, we sought to identify transcriptional signals downstream that could contribute to the phenotypes. We selected for genes that are differentially expressed in opposite directions in strains carrying clinical SNVs and strains carrying protein binding site deletions and identified a list of genes including transcriptional regulator *whiB2* and its regulon of division related genes. The promoter of *whiB2* is a known direct target of ResR and is also highly mutated in clinical *Mtb* populations. We propose a model where clinical variations sequester ResR through its interaction with DnaA at the *Rv0010c-Rv0011c* IGR. This sequestration reduces the ability of ResR to activate division

related genes and alters cell cycle dynamics, which produces the cell length and antibiotic phenotypes.

In Chapter 4, we discuss the impact of this work in understanding fundamental processes of *Mtb* and the unique contribution of such essential processes to *Mtb* drug response. This work also highlights the importance of research on non-coding regions like *Rv0010c-Rv0011c*. *Rv0010c-Rv0011c* is one of many highly mutated intergenic regions in *Mtb*, many of which remain uncharacterized. We also discuss future work that could be inspired based on our model and conclusions. Together, we hope to provide new insights into *Mtb* biology and the essential regulatory processes that allow *Mtb* to survive antibiotics clearance. In the Appendix, we provide supplemental materials that are not included in the main text of this dissertation.

1.6 References

1. Hershkovitz I, Donoghue HD, Minnikin DE, May H, Lee OYC, Feldman M, et al. Tuberculosis origin: The Neolithic scenario. Tuberculosis. 2015;95:S122-S6.

2. Schönlein J. Zur Pathogenie der Impetigines. Auszug aus einer brieflichen Mitteilung an den Herausgeber. Archiv für Anatomie, Physiologie und wissenschaftliche Medicin. 1839:82.

3. Zink AR, Sola C, Reischl U, Grabner W, Rastogi N, Wolf H, et al. Characterization of Mycobacterium tuberculosis Complex DNAs from Egyptian Mummies by Spoligotyping. Journal of Clinical Microbiology. 2003;41(1):359-67.

4. World Health Organization. Global tuberculosis report 2023. Geneva: World Health Organization; 2023 2023.

5. Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLOS Medicine. 2016;13(10).

6. Vynnycky E, Fine PEM. Lifetime Risks, Incubation Period, and Serial Interval of Tuberculosis. American Journal of Epidemiology. 2000;152(3).

7. World Health Organization. WHO consolidated guidelines on tuberculosis: module 4: treatment: drug-susceptible tuberculosis treatment. Geneva: World Health Organization; 2022 2022.

8. World Health Organization. WHO consolidated guidelines on tuberculosis: module 4: treatment: drug-resistant tuberculosis treatment. 2022 update ed. Geneva: World Health Organization; 2022 2022.

9. Fox W. Various Combinations of Isoniazid with Streptomycin or with PAS in the Treatment of Pulmonary Tuberculosis. Group.1(1).

10. Service HKC, Council BMR. Controlled trial of 6-month and 8-month regimens in the treatment of pulmonary tuberculosis: the results up to 24 months. Tubercle. 1979;60(4):201-10.

11. Service HKC, Council BMR. Controlled trial of 4 three-times-weekly regimens and a daily regimen all given for 6 months for pulmonary tuberculosis second report: the results up to 24 months. Tubercle. 1982;63(2):89-98.

12. Council STSBMR. Clinical trial of six-month and four-month regimens of chemotherapy in the treatment of pulmonary tuberculosis. American Review of Respiratory Disease. 1979;119(4):579-85.

13. Saleem A, Azher M. The next Pandemic-Tuberculosis: The oldest disease of mankind rising one more time. British journal of medical practitioners. 2013;6(2):21-46.

14. Stanley S, Liu Q, Fortune SM. Mycobacterium tuberculosis functional genetic diversity, altered drug sensitivity, and precision medicine. Frontiers in Cellular and Infection Microbiology. 2022;12:1007958.

15. Zhu J, Liu YJ, Fortune SM. Spatiotemporal perspectives on tuberculosis chemotherapy. Curr Opin Microbiol. 2023;72:102266.

16. Fleming A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. British journal of experimental pathology. 1929;10(3):226.

17. Schatz A, Bugle E, Waksman SA. Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria.∗. Proceedings of the Society for Experimental Biology and Medicine. 1944;55(1):66-9.

18. Iseman M. Tuberculosis therapy: past, present and future. European Respiratory Journal. 2002;20(36 suppl):87S-94s.

19. Crofton J, Mitchison D. Streptomycin resistance in pulmonary tuberculosis. British medical journal. 1948;2(4588):1009.

20. Ramaswamy S, Musser JM. Molecular genetic basis of antimicrobial agent resistance in Mycobacterium tuberculosis: 1998 update. Tubercle and Lung disease. 1998;79(1):3-29.

21. Gillespie SH. Evolution of drug resistance in Mycobacterium tuberculosis: clinical and molecular perspective. Antimicrobial agents and chemotherapy. 2002;46(2):267-74.

22. Zhang Y, Heym B, Allen B, Young D, Cole S. The catalase—peroxidase gene and isoniazid resistance of Mycobacterium tuberculosis. Nature. 1992;358(6387):591-3.

23. Ramaswamy SV, Reich R, Dou S-J, Jasperse L, Pan X, Wanger A, et al. Single nucleotide polymorphisms in genes associated with isoniazid resistance in Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy. 2003;47(4):1241-50.

24. Telenti A, Imboden P, Marchesi F, Matter L, Schopfer K, Bodmer T, et al. Detection of rifampicin-resistance mutations in Mycobacterium tuberculosis. The Lancet. 1993;341(8846):647-51.

25. Palomino JC, Martin A. Drug resistance mechanisms in Mycobacterium tuberculosis. Antibiotics. 2014;3(3):317-40.

26. Schön T, Werngren J, Machado D, Borroni E, Wijkander M, Lina G, et al. Antimicrobial susceptibility testing of Mycobacterium tuberculosis complex isolates–the EUCAST broth microdilution reference method for MIC determination. Clinical Microbiology and Infection. 2020;26(11):1488-92.

27. Zhang H, Li D, Zhao L, Fleming J, Lin N, Wang T, et al. Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nat Genet. 2013;45(10):1255-60.

28. Hicks ND, Yang J, Zhang X, Zhao B, Grad YH, Liu L, et al. Clinically prevalent mutations in Mycobacterium tuberculosis alter propionate metabolism and mediate multidrug tolerance. Nat Microbiol. 2018;3(9):1032-42.

29. Hicks ND, Giffen SR, Culviner PH, Chao MC, Dulberger CL, Liu Q, et al. Mutations in dnaA and a cryptic interaction site increase drug resistance in Mycobacterium tuberculosis. PLOS Pathogens. 2020;16(11):e1009063.

30. Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, et al. Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. Nat Genet. 2018;50(2):307-16.

31. Consortium C, Allix-Béguec C, Arandjelovic I, Bi L, Beckert P, Bonnet M, et al. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. The New England journal of medicine. 2018;379(15):1403-15.

32. Liu Q, Zhu J, Dulberger CL, Stanley S, Wilson S, Chung ES, et al. Tuberculosis treatment failure associated with evolution of antibiotic resilience. Science. 2022;378(6624):1111-8.

33. Colangeli R, Jedrey H, Kim S, Connell R, Ma S, Chippada Venkata UD, et al. Bacterial Factors That Predict Relapse after Tuberculosis Therapy. N Engl J Med. 2018;379(9):823-33.

34. Brauner A, Fridman O, Gefen O, Balaban NQ. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. Nature Reviews Microbiology. 2016;14(5):320-30.

35. Lee AS, Teo AS, Wong S-Y. Novel mutations in ndh in isoniazid-resistant Mycobacterium tuberculosis isolates. Antimicrobial agents and chemotherapy. 2001;45(7):2157-9.

36. Hicks ND, Carey AF, Yang J, Zhao Y, Fortune SM. Bacterial genome-wide association identifies novel factors that contribute to ethionamide and prothionamide susceptibility in Mycobacterium tuberculosis. MBio. 2019;10(2):10.1128/mbio. 00616-19.

37. Hansen FG, Atlung T. The DnaA Tale. Frontiers in Microbiology. 2018;9.

38. Martini MC, Hicks ND, Xiao J, Alonso MN, Barbier T, Sixsmith J, et al. Loss of RNase J leads to multi-drug tolerance and accumulation of highly structured mRNA fragments in Mycobacterium tuberculosis. PLoS Pathogens. 2022;18(7):e1010705.

39. Andries K, Villellas C, Coeck N, Thys K, Gevers T, Vranckx L, et al. Acquired resistance of Mycobacterium tuberculosis to bedaquiline. PLOS one. 2014;9(7):e102135.

40. Almeida D, Ioerger T, Tyagi S, Li S-Y, Mdluli K, Andries K, et al. Mutations in pepQ confer low-level resistance to bedaquiline and clofazimine in Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy. 2016;60(8):4590-9.

41. Bellerose MM, Baek S-H, Huang C-C, Moss CE, Koh E-I, Proulx MK, et al. Common variants in the glycerol kinase gene reduce tuberculosis drug efficacy. MBio. 2019;10(4):10.1128/mbio. 00663-19.

42. Safi H, Gopal P, Lingaraju S, Ma S, Levine C, Dartois V, et al. Phase variation in Mycobacterium tuberculosis glpK produces transiently heritable drug tolerance. Proc Natl Acad Sci U S A. 2019;116(39):19665-74.

43. Poulton NC, Azadian ZA, DeJesus MA, Rock JM. Mutations in rv0678 confer low-level resistance to benzothiazinone DprE1 inhibitors in Mycobacterium tuberculosis. Antimicrobial Agents and Chemotherapy. 2022;66(9):e00904-22.

44. World Health Organization. Guidelines for treatment of drug-susceptible tuberculosis and patient care, 2017 update. 2017.

45. Rajagopalan M, Qin M-H, Nash DR, Madiraju M. Mycobacterium smegmatis dnaA region and autonomous replication activity. Journal of bacteriology. 1995;177(22):6527-35.

46. Mukai T, Miyamoto Y, Yamazaki T, Makino M. Identification of Mycobacterium species by comparative analysis of the dnaA gene. FEMS microbiology letters. 2006;254(2):232-9.

47. Sassetti CM, Boyd DH, Rubin EJ. Comprehensive identification of conditionally essential genes in mycobacteria. Proc Natl Acad Sci U S A. 2001;98(22):12712-7.

48. Washington TA, Smith JL, Grossman AD. Genetic networks controlled by the bacterial replication initiator and transcription factor DnaA in Bacillus subtilis. Molecular Microbiology. 2017;106(1):109-28.

49. Seid CA, Smith JL, Grossman AD. Genetic and biochemical interactions between the bacterial replication initiator DnaA and the nucleoid-associated protein Rok in Bacillus subtilis. Mol Microbiol. 2017;103(5):798-817.

50. Plocinski P, Ziolkiewicz M, Kiran M, Vadrevu SI, Nguyen HB, Hugonnet J, et al. Characterization of CrgA, a New Partner of the Mycobacterium tuberculosis Peptidoglycan Polymerization Complexes. Journal of Bacteriology. 2011;193(13):3246-56.

51. Plocinski P, Arora N, Sarva K, Blaszczyk E, Qin H, Das N, et al. Mycobacterium tuberculosis CwsA interacts with CrgA and Wag31, and the CrgA-CwsA complex is involved in peptidoglycan synthesis and cell shape determination. Journal of bacteriology. 2012;194(23):6398-409.

52. Kapopoulou A, Lew JM, Cole ST. The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. Tuberculosis. 2011;91(1):8-13.

53. Gomez JE, Bishai WR. whmD is an essential mycobacterial gene required for proper septation and cell division. Proceedings of the National Academy of Sciences. 2000;97(15):8554-9.

54. Raghunand TR, Bishai WR. Mycobacterium smegmatis whmD and its homologue Mycobacterium tuberculosis whiB2 are functionally equivalent. Microbiology. 2006;152(9):2735-47.

55. Giffen SR. The Essential whiB2 Transcriptional Network and its Role in Cell Division: Harvard University; 2021.

56. Fol M, Chauhan A, Nair NK, Maloney E, Moomey M, Jagannath C, et al. Modulation of Mycobacterium tuberculosis proliferation by MtrA, an essential two-component response regulator. Molecular microbiology. 2006;60(3):643-57.

57. Lilic M, Holmes NA, Bush MJ, Marti AK, Widdick DA, Findlay KC, et al. Structural basis of dual activation of cell division by the actinobacterial transcription factors WhiA and WhiB. Proceedings of the National Academy of Sciences. 2023;120(11):e2220785120.

58. Reyes-Lamothe R, Sherratt DJ. The bacterial cell cycle, chromosome inheritance and cell growth. Nature Reviews Microbiology. 2019;17(8):467-78.

59. Monds RD, Lee TK, Colavin A, Ursell T, Quan S, Cooper TF, et al. Systematic perturbation of cytoskeletal function reveals a linear scaling relationship between cell geometry and fitness. Cell reports. 2014;9(4):1528-37.

60. Harris LK, Theriot JA. Relative Rates of Surface and Volume Synthesis Set Bacterial Cell Size. Cell. 2016;165(6):1479-92.

61. Wang JD, Levin PA. Metabolism, cell growth and the bacterial cell cycle. Nature Reviews Microbiology. 2009;7(11):822-7.

62. Logsdon MM, Ho PY, Papavinasasundaram K, Richardson K, Cokol M, Sassetti CM, et al. A Parallel Adder Coordinates Mycobacterial Cell-Cycle Progression and Cell-Size Homeostasis in the Context of Asymmetric Growth and Organization. Curr Biol. 2017;27(21):3367-74 e7.

63. Nair N, Dziedzic R, Greendyke R, Muniruzzaman S, Rajagopalan M, Madiraju MV. Synchronous replication initiation in novel Mycobacterium tuberculosis dnaA cold-sensitive mutants. Molecular microbiology. 2009;71(2):291-304.

64. Katayama T, Kasho K, Kawakami H. The DnaA Cycle in Escherichia coli: Activation, Function and Inactivation of the Initiator Protein. Frontiers in Microbiology. 2017;8.

65. Madiraju MVVS, Moomey M, Neuenschwander PF, Muniruzzaman S, Yamamoto K, Grimwade JE, et al. The intrinsic ATPase activity of Mycobacterium tuberculosis DnaA promotes rapid oligomerization of DnaA on oriC. Molecular Microbiology. 2006;59(6):1876-90.

66. Purushotham G, Sarva KB, Blaszczyk E, Rajagopalan M, Madiraju MV. Mycobacterium tuberculosis oriC sequestration by MtrA response regulator. Molecular Microbiology. 2015;98(3):586-604.

67. Smits WK, Grossman AD. The transcriptional regulator Rok binds A+ T-rich DNA and is involved in repression of a mobile genetic element in Bacillus subtilis. PLoS genetics. 2010;6(11):e1001207.

68. Kołodziej M, Trojanowski D, Bury K, Hołówka J, Matysik W, Kąkolewska H, et al. Lsr2, a nucleoid-associated protein influencing mycobacterial cell cycle. Scientific Reports. 2021;11(1):2910.

69. Minch KJ, Rustad TR, Peterson EJR, Winkler J, Reiss DJ, Ma S, et al. The DNA-binding network of Mycobacterium tuberculosis. 2015;6:5829.

70. Collier J, Murray SR, Shapiro L. DnaA couples DNA replication and the expression of two cell cycle master regulators. EMBO J. 2006;25(2):346-56.

71. Bush MJ. The actinobacterial WhiB-like (Wbl) family of transcription factors. Molecular Microbiology. 2018;110(5):663-76.

72. Poulton NC, DeJesus MA, Munsamy-Govender V, Roberts CG, Azadian ZA, Bosch B, et al. Beyond antibiotic resistance: the whiB7 transcription factor coordinates an adaptive response to alanine starvation in mycobacteria. bioRxiv. 2023:2023.06. 02.543512.

73. Geiman DE, Raghunand TR, Agarwal N, Bishai WR. Differential gene expression in response to exposure to antimycobacterial agents and other stress conditions among seven Mycobacterium tuberculosis whiB-like genes. Antimicrobial agents and chemotherapy. 2006;50(8):2836-41.

74. Lee D-S, Kim Y, Lee H-S. The whcD gene of Corynebacterium glutamicum plays roles in cell division and envelope formation. Microbiology. 2017;163(2):131-43.

75. Du S, Lutkenhaus J. Assembly and activation of the Escherichia coli divisome. Molecular microbiology. 2017;105(2):177-87.

76. Gola S, Munder T, Casonato S, Manganelli R, Vicente M. The essential role of SepF in mycobacterial division. Molecular microbiology. 2015;97(3):560-76.

77. Baranowski C, Rego EH, Rubin EJ. The dream of a mycobacterium. Microbiology spectrum. 2019;7(2):10.1128/microbiolspec. gpp3-0008-2018.

78. Hett EC, Chao MC, Rubin EJ. Interaction and modulation of two antagonistic cell wall enzymes of mycobacteria. PLoS pathogens. 2010;6(7):e1001020.

79. Chao MC, Kieser KJ, Minami S, Mavrici D, Aldridge BB, Fortune SM, et al. Protein complexes and proteolytic activation of the cell wall hydrolase RipA regulate septal resolution in mycobacteria. PLoS pathogens. 2013;9(2):e1003197.

80. Aldridge BB, Fernandez-Suarez M, Heller D, Ambravaneswaran V, Irimia D, Toner M, et al. Asymmetry and Aging of Mycobacterial Cells Lead to Variable Growth and Antibiotic Susceptibility. Science. 2012;335(6064):100-4.

81. Santi I, Dhar N, Bousbaine D, Wakamoto Y, McKinney JD. Single-cell dynamics of the chromosome replication and cell division cycles in mycobacteria. Nat Commun. 2013;4:2470.

82. Hannebelle MTM, Ven JXY, Toniolo C, Eskandarian HA, Vuaridel-Thurre G, McKinney JD, et al. A biphasic growth model for cell pole elongation in mycobacteria. Nature Communications. 2020;11(1).

83. Kieser KJ, Rubin EJ. How sisters grow apart: mycobacterial growth and division. Nat Rev Microbiol. 2014;12(8):550-62.

84. Meniche X, Otten R, Siegrist MS, Baer CE, Murphy KC, Bertozzi CR, et al. Subpolar addition of new cell wall is directed by DivIVA in mycobacteria. Proc Natl Acad Sci U S A. 2014;111(31):E3243-51.

85. Plocinska R, Martinez L, Gorla P, Pandeeti E, Sarva K, Blaszczyk E, et al. Mycobacterium tuberculosis MtrB sensor kinase interactions with FtsI and Wag31 proteins reveal a role for MtrB distinct from that regulating MtrA activities. Journal of bacteriology. 2014;196(23):4120-9.

86. Willis L, Huang KC. Sizing up the bacterial cell cycle. Nat Rev Microbiol. 2017;15(10):606-20.

Chapter 2: Genetic characterization of an essential pathway under antibiotic selection in *Mycobacterium tuberculosis*

2.1 Abstract

Tuberculosis remains a global public health threat, claiming 1.5 million lives annually. Despite being treatable and curable, TB treatment outcomes are challenged by increasing antibiotic resistance in its causative agent, *Mycobacterium tuberculosis* (*Mtb*). *Mtb* clinical resistance are driven not only by canonical drug resistance but also by non-canonical mechanisms such as antibiotic tolerance and persistence. Our large-scale genomic analysis of 50,000 clinical *Mtb* isolates revealed ongoing bacterial evolution under selective pressure from host and drug treatments. We identified a genetic pathway under antibiotic selection, including genes including two cell cycle regulatory genes *dnaA* and *resR* and one intergenic region (IGR) *Rv0010c-Rv0011c* IGR. Clinical variants in *dnaA, resR,* and *Rv0010c-Rv0011c* IGR shared similar increase in low-level isoniazid resistance and antibiotic resilience characterized by faster regrowth post antibiotic exposure. These drug phenotypes were accompanied by an increase in cell length, suggestive of cell cycle regulation involvement. Knockout of *Rv0010c-Rv0011c* operon demonstrated the non-coding regulatory role of *Rv0010c-Rv0011c* IGR in morphology and drug susceptibility, which did not require translation of operon genes but required binding of DnaA and ResR. Complementation with clinically IGR variants recapitulated phenotypic changes in cell length and INH resistance, suggesting clinical mutations functioned by altering interaction dynamics with DnaA and ResR. These findings underscored the complex interplay between bacterial genetic elements, cell cycle regulation, and drug response, and offered insights for the role of *Mtb* physiology in drug resistance.

2.2 Introduction

Tuberculosis (TB) remains the world's deadliest infectious disease by a single agent and continues to claim lives of 1.5 million people every year (1). Despite being treatable and curable, TB treatment is lengthy and complicated, and treatment fails at an average rate of 10% in drug susceptible TB even in the context of clinical trials. TB treatment outcomes are partly complicated by the rise of antibiotic resistance in the population of the causative agent, *Mycobacterium tuberculosis* (*Mtb*) (2-4). To identify bacterial determinants of antibiotic resistance, genome-wide association studies (GWAS) were performed on drug resistant clinical strains and identified genes in canonical drug targeting pathways highly associated with first- and second-line antibiotics used to treat TB. These genes encoded proteins that are either drug targets or required for its mechanism of action (5-8). For example, mutations in *katG*, the converter enzyme for pro-drug form of INH to active form *in vivo*, directly confer antibiotic resistance in the definition of increasing minimal inhibitory concentration (MIC) (9, 10). Mutations in *rpoB* helps *Mtb* escape ribosomal inhibition caused by rifampicin by inducing confirmational change that reduces rifampicin affinity to RNA polymerase (5, 11). Both these types of mutation are evident in clinical populations of *Mtb* and directly correlate with treatment outcomes (12).

However, antibiotic resistance is not the only way bacteria can survive antibiotic treatments. Bacteria are known to evade antibiotic clearance through alternative methods such as tolerance and persistence (13). Treatment related pressure should select for mutations conferring additional antibiotic phenotypes in addition to resistance. These non-canonical drug resistance determinants have been identified by GWAS (7, 14, 15). With more isolates being whole-genome sequenced in the last decade, we reasoned we can use population genetics to identify signatures of genetic evolution driven by host and treatment induced selective pressure through genomic

comparison of clinical *Mtb* patient isolates. With 50,000 *Mtb* clinical isolates, we performed a large-scale genomic mutational analysis and identified coding and non-coding genomic regions accumulating mutations repeatedly as a signature of ongoing evolution. In addition to canonical drug resistance genes such as *katG*, and *rpoB*, we identified novel genetic determinants under selection such as *prpR (Rv1129c)*, *dnaA* and *Rv1830 (resR)* (16). Previous work found mutations in *prpR* promoted antibiotic tolerance by reducing metabolic activity in the presence of cholesterol in host relevant environments, while mutations in *dnaA* exhibited low level resistance to INH (7, 17). Although their MIC shifts did not meet clinical breakpoint, low level MIC shifts had been shown to increase treatment relapse significantly and can contribute directly to treatment failure (18). Phenotypic characterization of *resR* variants led us to discover a new antibiotic phenotype, named antibiotic resilience. Antibiotic resilience was characterized by faster regrowth post strong antibiotic exposure both on the single-cell level and bulk population level. Resilient populations, such as *resR* isogenic variants, showed faster incorporation of new cell wall material by fluorescent labeling and faster colony growth on solid media post strong exposure of a panel of first-line and second-line TB drugs (16). This is extremely relevant in the context of the spatiotemporal nature of TB treatment. Serum measurement of drug concentration following daily dose in patients with drug susceptible TB showed plasma concentrations of INH, RIF, and EMB fall below MIC for 21%, 42%, and 75% of treatment time respectively (19, 20). These measurements did not take into consideration of the complexity of the spatial niche and immune environment *Mtb* resides in, and these could complicate further drug penetration and accumulation (20, 21). Therefore, the ability to regrow at a more permissible antibiotic concentration could be beneficial to overall *Mtb* survival under antibiotic treatment. Having

discovered that, we tested *dnaA* variants for resilience phenotype and found signatures of resilience although at a much more modest level (Figure 2-1A).

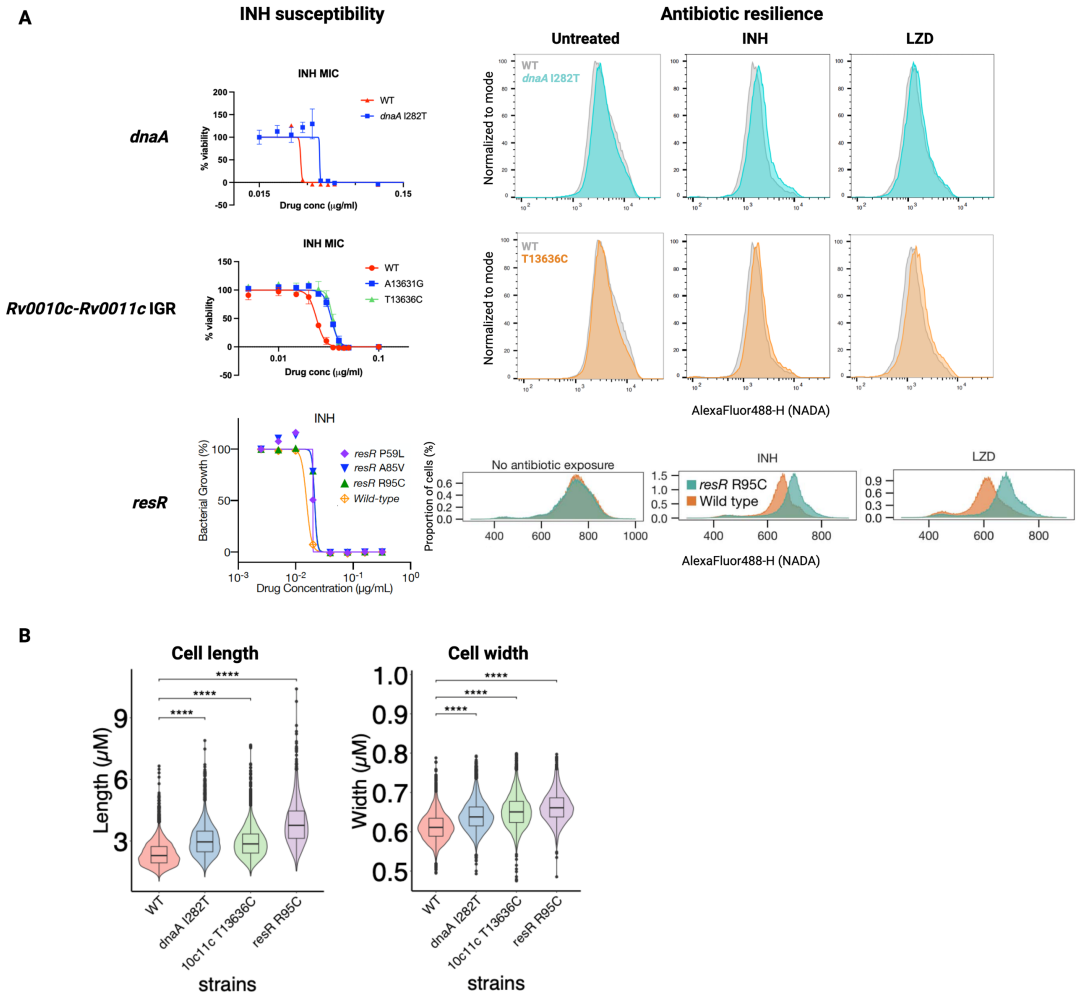


Figure 2-1 | Clinically relevant variations in *dnaA*, *resR*, and *Rv0010c-Rv0011c* IGR share comparable increase in INH resistance, antibiotic resilience, and cell size.

A) Quantification of INH susceptibility by MIC measurement with alarmar blue assay and antibiotic resilience by flow cytometry. Antibiotic resilience is characterized by increased membrane fluorescent labeling in variant strains after strong antibiotic exposure, a marker for faster resumption of cell wall synthesis. After treatment with no drug or 100X MIC of indicated drug, NADA-labeled WT and variant cells were harvest after 24 hours of recovery in media with no drug and fluorescent intensity of the population was measured by flow cytometry. *resR* INH MIC and resilience data was adapted from Liu *et al.* 2022. B) Morphology measurements were done using quantitative microscopy on cells grown to mid log in plain media and were analyzed through the MOMIA pipeline. Adjusted p-value by Mann Whitney U Test, *** < 0.001, **** < 0.0001.

In *resR* variants, the resilience phenotype was accompanied by a morphology phenotype similar to what we observed in *dnaA* variants (Figure 2-1B). Both *resR* and *dnaA* variants shared an increase in cell length and this increase was not due to impaired chromosomal replication or cellular growth rate (Supplementary Figure 2.1) (16, 17). This morphology phenotype was particularly of interest as *dnaA* and *resR* both encode cell cycle regulatory proteins and cell cycle regulation is tightly link to bacterial morphological control (22). *dnaA* encodes a highly conserved protein across all species of bacteria, DnaA, the replication initiator protein and master regulator of bacterial cell cycle. DnaA also acts as transcriptional factor directly and indirectly regulate expression of cell cycle related genes and cell cycle progression (17, 23, 24). *resR* encodes a conserved transcriptional factor, ResR, known to regulate cell division related genes (16, 25). Despite DnaA and ResR regulate two essential processes that are closely coordinated, there has been no previously proposed link between these two proteins. Here, we identified a common binding site of DnaA and ResR at *Rv0010c-Rv0011c* IGR, the intergenic region (IGR) between co-operonic genes *Rv0010c* and *Rv0011c*. Interestingly, *Rv0010c-Rv0011c* IGR was among the most highly mutated non-coding regions in clinical *Mtb* populations and mutations in this region perfectly aligned with both DnaA and ResR binding sites (16). Isogenic variants of *Rv0010c-Rv0011c* were shown to phenocopy *dnaA* and *resR* isogenic variants in both morphology and drug profiles including low-level INH resistance and antibiotic resilience (Figure 2-1 A) (17). This provided evidence for a genetic pathway under selection in *Mtb* clinical populations regulating morphological control and drug response. However, *Rv0010c-Rv0011c* operon was less studied, and its function remained unknown. We decided to take a genetic approach to investigate the relationships between these genetic elements and constructed a knockout of this operon. *Rv0010c-Rv0011c* KO produced shorter cells with increased sensitivity

to INH. Full complementation required the complete sequence integrity of this operon and did not require translation of coding genes *Rv0010c* or *Rv0011c*, suggesting a non-coding role of *Rv0010c-Rv0011c* operon in regulating downstream phenotypes of interest. This non-coding function was dependent on DnaA and ResR binding since removal of protein binding sites failed to restore wild type morphology and INH susceptibility. Complementation with the operon carrying clinically relevant variant in *Rv0010c-Rv0011c* IGR recapitulated increase in cell length and INH resistance as observed in isogenic strains. Together, we present here the identification and genetic characterization of an essential genetic pathway through *Rv0010c-Rv0011c* IGR and this pathway is under antibiotic selection in *Mtb*. *Rv0010c-Rv0011c* IGR functions by modulating DnaA and ResR interaction at this site and increased interaction evident in IGR variant strains can provide a survival advantage under antibiotic pressure.

2.3 Results

DnaA and ResR binds at *Rv0010c-Rv0011c* intergenic region and *Rv0010c-Rv0011c* IGR variants phenocopy *dnaA* and *resR* variants

Given their function as a DNA binding transcription regulator, previous works from the lab performed *in vitro* DNA affinity purification and deep sequencing (IDAPseq) of DnaA and ResR to identify direct regulon of these two proteins (16, 17). DnaA bound several genomic locations including its own promoter and promoter of *dnaB*, the gene encoding helicase required for origin unwinding during replication initiation (17, 23). DnaA also bound at the intergenic region between *Rv0010c* and *Rv0011c* with unknown function (Figure 2-2A). Interestingly, in our later study with ResR, we found abutting binding sites with DnaA and ResR at this intergenic

region (IGR) and they aligned with the most conserved nucleotide region in this IGR, emphasizing the importance of DnaA and ResR binding here in mycobacteria (Figure 2-2B, C) (16, 17). In the large-scale genomic analysis of 50,000 *Mtb* clinical isolates, *Rv0010c-*

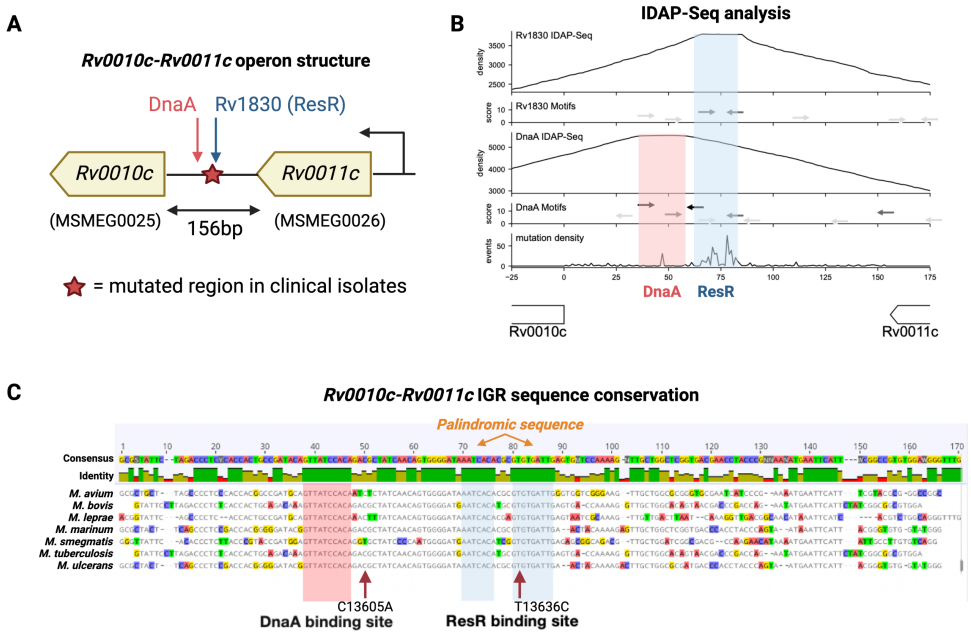


Figure 2-2 | DnaA and ResR both binds at *Rv0010c-Rv0011c* IGR and the binding sites overlaps with the most conserved and the most highly mutated nucleotides of this intergenic region.

A) Operon structure of *Rv0010c-Rv0011c.* B) *In vitro* DNA-affinity purification and deep sequencing (IDAP-seq) reveals binding sites of DnaA and ResR at *Rv0010c-Rv0011c* IGR and these binding sites overlap with sequences where mutations are enriched within this intergenic region in clinical *Mtb* populations. IDAP-seq tracing and predicted motif for each protein are shown in the top four panels. The bottom panel indicates the number of unfixed mutation events in clinical strains of *Mtb* from the analysis in Liu *et al.* 2022. C) *Rv0010c-Rv0011c* IGR is conserved across mycobacterial species and the most conserved nucleotides overlap with DnaA and ResR binding sites. These nucleotides are also the ones enriched in mutations in clinical strains. Red arrow indicates mutations that are characterized in downstream experiments.

Rv0011c IGR was one of the most highly mutated non-coding regions and mutations in this region were associated with drug resistance and treatment failure in multiple earlier genome-wide association studies (7, 16, 26). Mutations in *Rv0010c-Rv0011c* IGR clustered mainly in two loci and these two loci aligned with DnaA and ResR binding sites respectively (Figure 2-2B). Knowing *dnaA* and *resR* variants manifested in similar drug and morphology phenotypes, we

hypothesized that DnaA and ResR functioned through *Rv0010c-Rv0011c* and mutations in *Rv0010c-Rv0011c* IGR should phenocopy of *dnaA* and *resR* variants.

To test this hypothesis, we created isogenic variants in the *Rv0010c-Rv0011c* intergenic regions and assessed their drug and morphology phenotype. Previous work from the lab had started to characterize isogenic variants of *Rv0010c-Rv0011c* IGR. *Rv0010c-Rv0011c* variants in the ResR binding region resulted in similar low-level INH resistance and cell length increase as *dnaA* and *resR* variants without any change in growth and replication rate (17). *Rv0010c-Rv0011c* IGR variants also showed subtle resilience phenotype at a degree similar to *dnaA* variants (Figure 2-1, Supplementary Figure 2.1). Given the existing variants were within the ResR binding site, we constructed two additional isogenic SNP variants at the two other clusters of mutations in this operon, one within the DnaA binding region at genomic position 13605 bp and one with *Rv0010c* coding region at genomic position 13490 bp. New SNP variants in *Rv0010c-Rv0011c* also shared similar morphology and INH susceptibility phenotype as other IGR variants (Supplementary figure 2.2). Phenocopying of *Rv0010c-Rv0011c* variants to *dnaA* and *resR* variants provided evidence for clinical evolution selecting on this convergent genetic pathway in *Mtb*.

Knowing the increase in cell length, we tested polar growth dynamics in these isogenic variants. Mycobacteria grows only from the poles and in polar growing bacteria, increase in cell length along the horizontal axis directly correlates with polar growth increments. Mycobacteria cell poles also grow at different rates and result in the unique asymmetrical growth and division pattern (27-29). Asymmetrical growth has been thought to be a way for mycobacteria to generate phenotypic heterogeneity that facilitate intrinsic drug resistance and enhanced survival in the face of host and drug pressure (27, 30, 31). We observed that morphology differences in *dnaA*,

resR, and *Rv0010c-Rv0011c* variants were driven primarily with cell length with relatively smaller increase in cell width (Figure 2-3). We then used an established pulse-chase labeling method combined with fluorescent microscopy to assess polar growth dynamics (27). By using a high-throughput quantitative microscopy platform, we were able to quantify growth dynamics from both poles of the cell: the nascent pole and the mature pole (32).

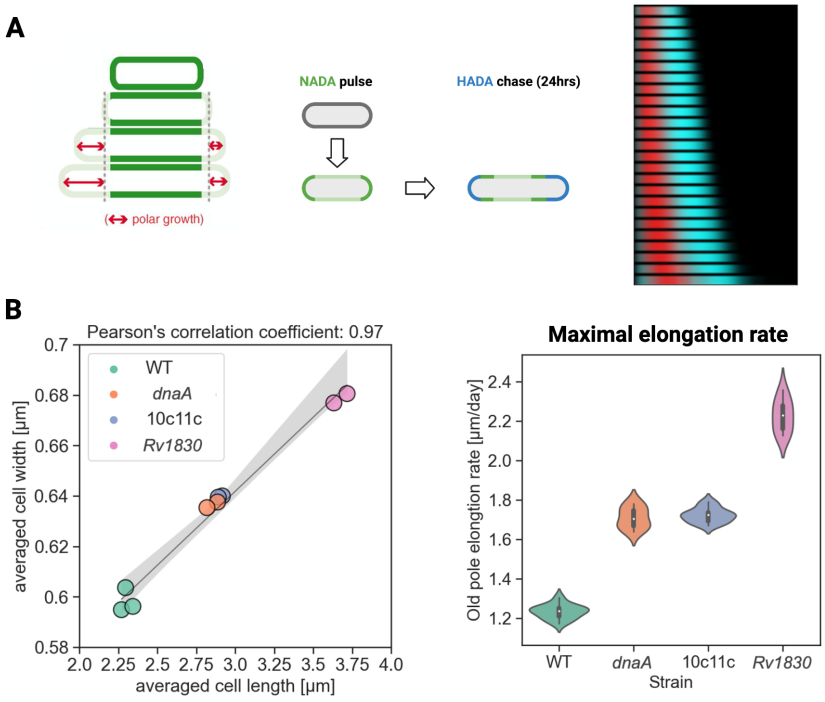


Figure 2-3 | Cell size increase shared by isogenic variants is driven by increase in polar elongation rate.

A) Schematic of *Mtb* polar growth (left) and pulse chase microscopy (middle). The polar growth diagram on the left shows how mycobacteria grow from the poles at different rates. Mature pole grows faster (old pole, left red arrows) and nascent pole grows slower (new pole, right red arrows). Demograph on the right shows an example of a labeled population ranked from shorter to longest top to bottom with their labeled cell wall during pulse in red and labeled cell wall during chase in cyan. B) Measurements of cell length, width, and mature pole elongation outgrowth were done using quantitative microscopy on cells grown to mid log in plain media and were analyzed through the MOMIA pipeline. Maximal pole elongation rate was calculated by the absolute amount of elongation at mature pole (old pole) divided by time. Nascent pole elongation measurements were not significant and were not included.

The nascent pole growth rate is largely dependent on the maturation period and changes in nascent pole growth rate reflects difference in maturation timing. On the other hand, mature pole

dynamics are reflections of maximal polar growth rates (29). We observed that *Rv0010c-Rv0011c*, *dnaA*, and *resR* variants shared an increase in mature pole elongation rate while maintained a similar nascent pole elongation rate (Figure 2-3). Changes in mature pole growth rate correlated directly with and likely drove the differences in cell length among these variants. As we know from *resR* variants, antibiotic resilience was driven by faster regrowth post strong antibiotic pressure on both single-cell and population levels. One of the main phenotypes was the amount of polar growth immediately following recovery (16). Polar elongation rate could be an important link to faster growth during recovery, but more work is required to further characterize the change in polar growth dynamics in resilience. Together, with shared drug and morphology phenotype, we further emphasized the importance of this intergenic region in clinical *Mtb* populations and continued to solidify the possibility of all three elements being in the same genetic pathway under convergent evolution.

***Rv0010c-Rv0011c* operon is well conserved across mycobacteria**

So, what is *Rv0010c-Rv0011c* and what makes this operon important? *Rv0010c-Rv0011c* operon consists of two genes encoding two small and conserved membrane proteins and 155 bp intergenic region. *Rv0011c*, also known as *crgA*, encodes a protein with homology to *Streptomyces* CrgA proteins involved in cell division, has been shown to localize to membrane and mid-cell and directly interact with division proteins such as FtsZ and cell wall synthesis protein (CwsA) in mycobacteria (33). *Rv0010c* encodes a conserved membrane protein with unknown function. AlphaFold predicted Rv0010c to be a small protein with N-terminal transmembrane helices and homology search revealed structural similarities to eukaryotic phosphatidylinositol transferases (Supplementary Figure 2.3) (34, 35). Phosphatidylinositol is an

important lipid anchor in mycobacteria and mainly are involved in maintaining membrane lipids such as phosphatidylinositol mannosides (PIMs) and derivatives of PIMs: lipomannans (LMs) and lipoarabinomannans (LAMs). These membrane lipids are important for maintaining membrane integrity and immune manipulation in mycobacteria (36-40). PIM synthesis is relatively well characterized in mycobacteria and yet there has been no implication of Rv0010c in any step and Rv0010c function remains a mystery. *Rv0010c* and *Rv0011c* are both non-essential and yet are well conserved across mycobacteria and Rv0011c homology extends further into neighboring actinobacterial species (Supplementary Figure 2.3) (41).

Rv0010c-Rv0011c intergenic region also is well conserved across mycobacterial species and its function remains unclear (42). *Rv0010c-Rv0011c* IGR is relatively long at 155 bp and yet is thought not to contain additional promoter as the predicted transcriptional start site of this operon locates in front of *Rv0011c* and both genes are thought to be co-transcribed (Figure 2-2A) (43). Alignment of mycobacterial *Rv0010c-Rv0011c* IGRs identified two regions with highest conservation and interestingly they overlapped with DnaA and ResR binding sites (Figure 2-2C). This indicated that binding of these two proteins at this region serves conserved and functionally important processes for mycobacteria. As DnaA regulates onset of DNA replication and ResR regulates expression of division regulators, we speculated that IGR mutations in protein binding sites could alter dynamics of these two proteins at this site and have impacts on coordination of cell cycle processes. Conservation of this intergenic region across mycobacteria likely implicates its function in fundamental processes and yet this region is accumulating more mutations than most other non-coding sites on the *Mtb* genome. This is likely a result of immense selective pressure *in vivo*, which could be both host and drug related, and it is important to functionally characterize these mutations to better understand how they facilitate *Mtb* clearance by

antibiotics. In summary, we showed conservation of *Rv0010c-Rv0011c* operon in mycobacteria, both in the coding and non-coding regions. Function of this operon remains understudied and would be the focus of our next step to understand the clinical importance of this operon in the face of host and antibiotic pressure.

Rv0010c-Rv0011c **IGR mutations does not affect expression of** ***Rv0010c*** **or** ***Rv0011c***

Based on a previous whole genome mapping of transcriptional start sites in *Mtb*, *Rv0010c* is co-transcribed with *Rv0011c* with the TSS upstream of *Rv0011c* (43). As mentioned above, the two genes in the operon are separated by a 155 bp intergenic region. A recent expression profile study showed co-expression level between two co-directional genes drops significantly as the intergenic region goes beyond 100 bp (44). This raised the possibility of another layer of transcriptional regulation within the intergenic region and led us to hypothesize that binding site mutations of DnaA and ResR, two transcriptional regulators, in front of *Rv0010c* inside the intergenic region directly alter *Rv0010c* expression. To test this hypothesis, we did transcriptional profiling (RNA-seq) of *Rv0010c-Rv0011c* IGR variants and asked if *Rv0010c* was differentially expressed in the presence of the variants compared to WT. Analysis of differentially expressed genes showed a slight upregulation of *Rv0010c,* but this upregulation was not statistically significant (adjusted p-value > 0.05) after multiple hypothesis correction. *Rv0011c* expression remained largely unchanged (Figure 2-4). To further test if altering *Rv0010c* expression directly resulted in drug and morphology phenotype observed in IGR variants, we created aTC-inducible overexpression of *Rv0010c* and knockdown of *Rv0010c* by CRISPR interference in *Mtb*. Neither overexpression nor knockdown of *Rv0010c* resulted in difference in morphology or INH susceptibility. Overexpression of *Rv0011c* also showed no change in morphology (Figure 2-4 B).

Additionally, RNAseq in *dnaA* and *resR* variants done previously did not show significant changes in expression of *Rv0010c-Rv0011c* either (Supplementary figure 2.4). This led us to conclude that mutations in *Rv0010c-Rv0011c* IGR did not function downstream through changing expression of *Rv0010c* or *Rv0011c*.

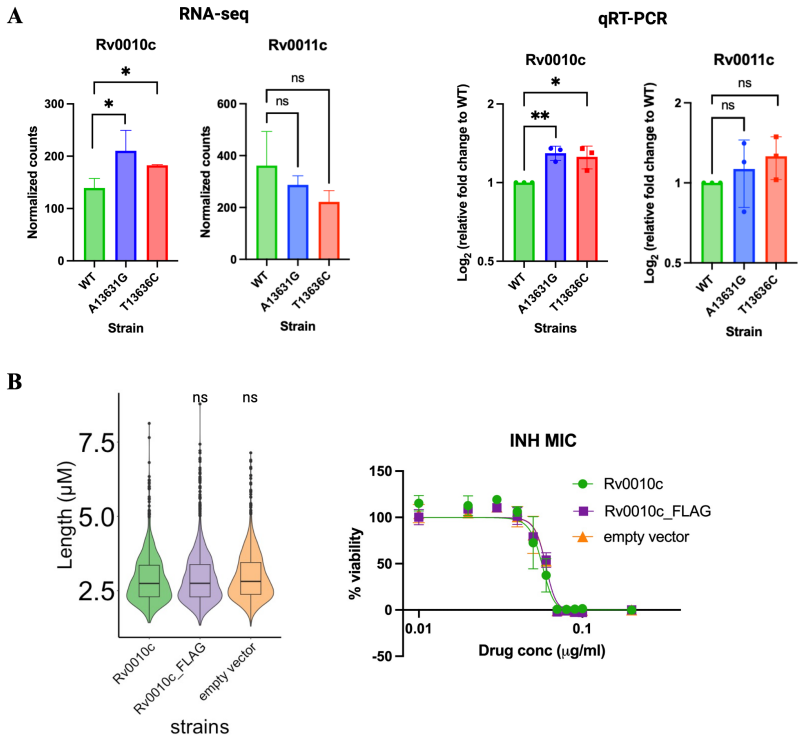


Figure 2-4 | *Rv0010c-Rv0011c* IGR variants slightly upregulate *Rv0010c* but overexpression of *Rv0010c* did not affect morphology or INH susceptibility.

A) Quantification of *Rv0010c* and *Rv0011c* expression by RNA-seq and qRT-PCR. Fold change in qRT-PCR was calculated using the ΔΔCt method normalized to housekeeping gene *sigA*. Adjusted p-value by t-test, * < 0.05, ** < 0.01. B) *Rv0010c* overexpression is constructed by insertion of a plasmid carrying *Rv0010c* under a constitutively and highly expressed promoter into WT H37Rv. Morphology measurements were done using quantitative microscopy on cells grown to mid log in plain media and were analyzed through the MOMIA pipeline. INH susceptibility was determined using the alamar blue assay and percentage bacteria survival at a given concentration is shown here. Adjusted p-value by Mann Whitney U Test, ns > 0.05.

Rv0010c-Rv0011c IGR functions as a non-coding regulatory element

Knowing the *Rv0010c-Rv0011c* IGR mutations did not directly alter expression of *Rv0010c* and *Rv0011c*, we decided to take a genetic approach to understand how this operon functioned. To do this, we wanted to test the function of individual component of this operon (two coding genes and one intergenic region). Instead of making deletions independently, we decided to first create a knockout (KO) of this operon from its native locus and complemented with various version of this operon at a phage integration site (Figure 2-5 B).

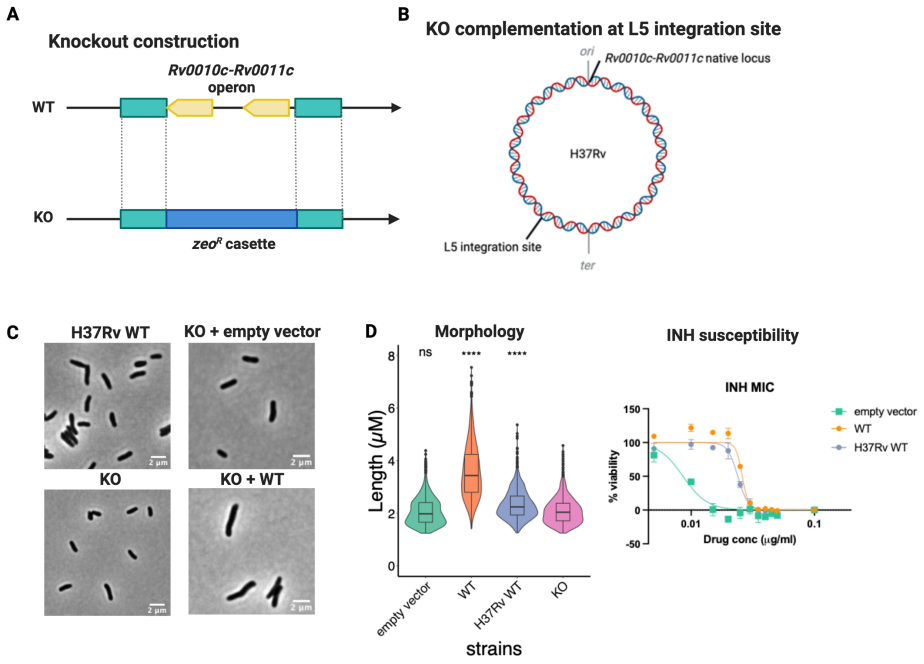


Figure 2-5 | Knockout of *Rv0010c-Rv0011c* operon decreases cell length and can be complemented by WT operon at a non-native site.

A) Schematic of KO construction. B) Schematic of complementation at L5 phage integration site, located at a distant location from *Rv0010c-Rv0011c* native locus. C) and D) Phenotype profiling and quantification of *Rv0010c-Rv0011c* KO strains compared to H37Rv WT and *Rv0010c-Rv0011c* KO strain complemented with empty vector or with WT operon. Morphological quantification was done by quantitative microscopy on strains were grown to mid-log before harvesting into PFA for microscopy. INH susceptibility test was performed with alamar blue assay and percentage bacteria survival at a given concentration is shown here. Imaging analysis was done using the MOMIA pipeline. Adjusted p-value by Mann Whitney U Test , *** < 0.001, **** < 0.0001.

First, we noticed that KO of this operon in *Mtb* resulted in much smaller cells, distinct from wild type *Mtb*. This KO morphology phenotype was accompanied by increased sensitivity to INH (Figure 2-5D). Morphology of KO strain provided further evidence that this operon regulated cell morphology and absence of this operon directly resulted in morphological defect. The small cell phenotype was consistent with blocking of IGR binding sites by CRISPR interference, which indicated this phenotype could be driven by loss of DnaA or ResR interaction at this site (Supplementary Figure 2.5). Complementation of wild type operon with its native TSS restored cell length and INH susceptibility phenotype, although this complementation was imperfect (Figure 2-5 D). We observed a further increase in cell length compared to H37Rv WT and this could be due to differences in genomic location that affected copy number or any dosage effect or other biological reasons we did not understand yet. Although morphology didn't fully complement, we observed complement WT was able to restore INH sensitivity to similar level as H37Rv WT, which suggested our operon was functioning properly (Figure 2-5D). Despite the two genes being non-essential, KO did result in slower growth and such defect can be complemented with WT operon (Supplementary Figure 2.6).

Given the morphological change seen in KO cells, we decided to use morphology as the main phenotype to profile complementation of various versions of this operon and assess the genetic requirement for restoring normal morphology (Figure 2-6). We first complemented back *Rv0010c, Rv0011c,* and intergenic region individually (Figure 2-6 A). Single gene or IGR alone failed to complement KO morphology phenotype and remained hypersensitive to INH. Next, we created combinations of single gene and IGR: *Rv0010c*+IGR, *Rv0011c*+IGR, and *Rv0010c+Rv0011c*. All combinations failed to complement KO morphology and INH sensitivity (Figure 2-6 A). Therefore, maintaining normal morphology required the structural integrity of all

three components of this operon. However, this did not distinguish the requirement of the DNA sequence of the operon or the requirement of translation of the two coding genes. To test if

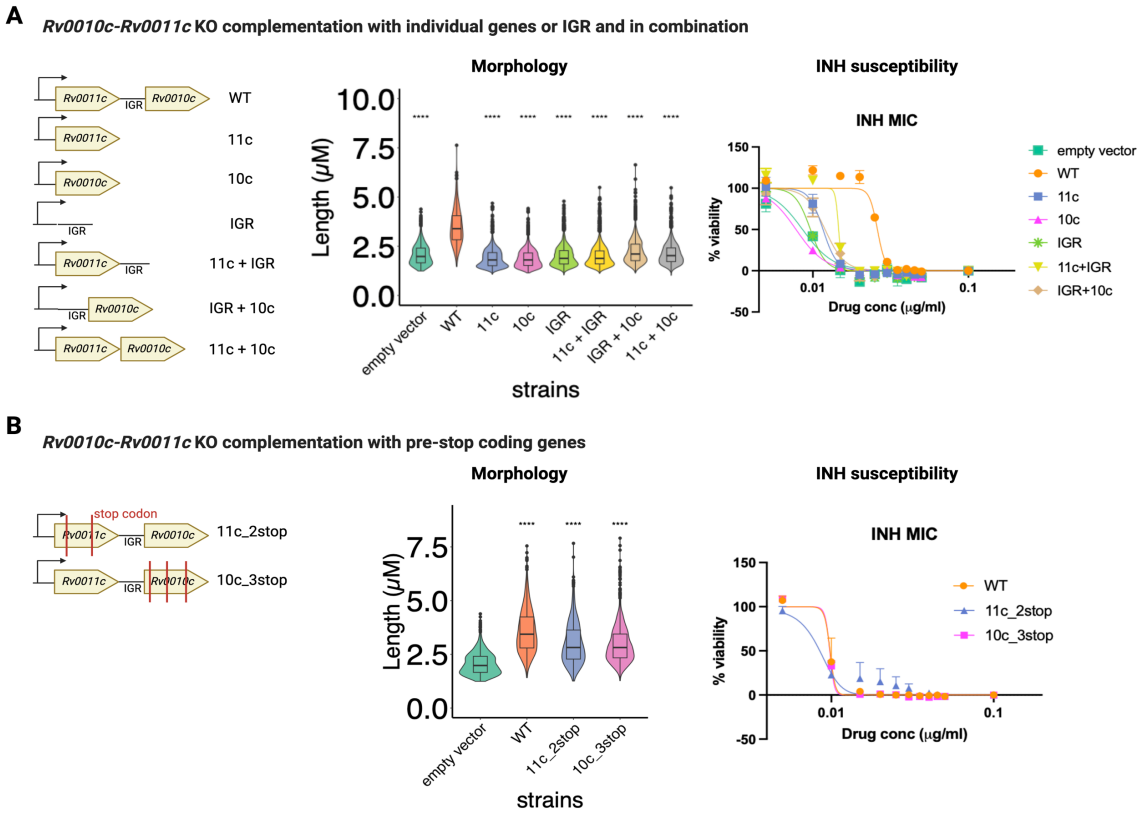


Figure 2-6 | Complete sequencing integrity of *Rv0010c-Rv0011c* operon is required for complementation while translation of *Rv0010c* or *Rv0011c* is not.

A) Schematic of *Rv0010c-Rv0011c* KO complementation with single and combinations of operon elements. Cell length is quantified by microscopy and INH susceptibility is measured by alamar blue assay. B) Schematic of *Rv0010c-Rv0011c* KO complementation with pre stop codons inserted in *Rv0010c* or *Rv0011c* coding regions to test necessity of translation in complementation. All strains were grown to mid-log before harvesting into PFA for microscopy. Cell length is quantified by microscopy using MOMIA pipeline. INH susceptibility is measured by alamar blue assay and percentage bacteria survival at a given concentration is shown here. Adjusted p-value by Mann Whitney U Test , *** < 0.001, **** < 0.0001.

translation of the two coding genes of this operon was required, we complemented with the complete operon that carried either *Rv0011c* or *Rv0010c* with mutations to insert pre-stop codons in the amino acid sequence (Figure 2-6 B). Complemented with presumably truncated Rv0011c or Rv0010c restored wild type morphology and INH susceptibility. Together, we concluded that

complete genomic sequence of this operon was required for its function at maintaining morphology and drug sensitivities and it was independent of translation of the two coding genes.

Next, to assess how clinically relevant variants affected morphology and drug phenotypes in our complementation system, we complemented with the operon carrying single nucleotide mutations in the intergenic region and showed they recapitulated the increased cell length and INH resistance in isogenic IGR variants (Figure 2-7).

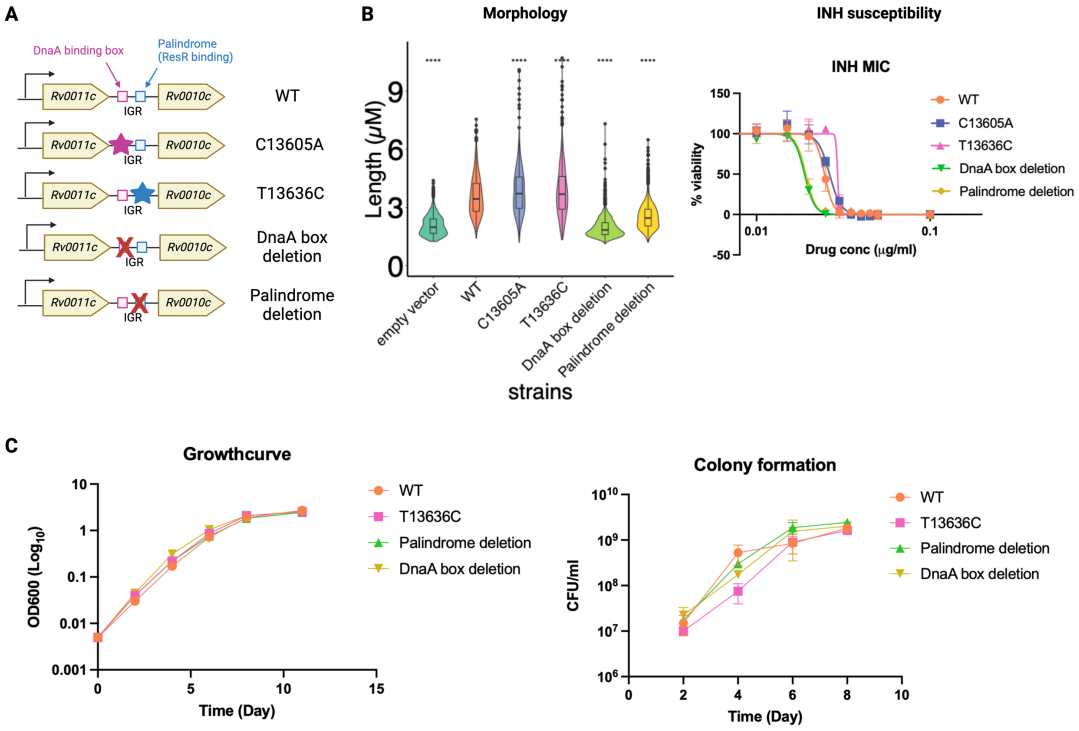


Figure 2-7 | Complete operon with clinical single nucleotide variants re-capitulates INH and morphology phenotypes in isogenic variants, while complete operon with protein binding site deletion fails complementation.

A) Schematic of *Rv0010c-Rv0011c* KO complementation with complete operon with clinical SNP or deletion of DnaA or ResR binding sites. B) Phenotyping of complementation strain constructed as shown in A). All strains were grown to mid-log before harvesting into PFA for microscopy. Imaging analysis was done using the MOMIA pipeline. INH susceptibility is measured by alamar blue assay and percentage bacteria survival at a given concentration is shown here. B) Measurement of complementation strain growth in liquid by OD600 and on solid media by colony formation. Adjusted p-value by Mann Whitney U Test , *** < 0.001, **** < 0.0001.

Knowing the non-coding nature of this operon's regulatory function and binding of two DNA binding proteins DnaA and ResR at the mutational hotspot, we reasoned that clinical IGR variants altered protein binding and including a deletion of protein binding at this site could help us determine how clinical IGR variants affected DnaA and ResR binding. Therefore, we constructed and complemented the KO strain with the operon carrying deletions of either the DnaA binding box or the palindromic sequences that ResR binds to. Loss of DnaA and ResR binding sites failed to complement the morphology phenotype and remained more sensitive to INH (Figure 2-7). Observed morphology differences were not a result of growth defect in these complementation strains as they grew and replicate at the same rate (Figure 2-7B). Together, this suggested that the single nucleotide mutations in *Rv0010c-Rv0011c* IGR were gain of function and resulted in increased cell length and INH resistance, and loss of protein binding was loss of function and resulted in decreased cell length and INH resistance.

In summary, using a genetic system constructed based on a KO of the Rv0010c-Rv0011c operon, we were able to identify that *Rv0010c-Rv0011c* operon regulates morphology and INH susceptibility independent of the production of full length Rv0010c and Rv0011c proteins. We showed that binding of DnaA and ResR at *Rv0010c-Rv0011c* IGR is specifically required in regulating morphology and drug phenotypes and clinically relevant single nucleotide changes in IGR is likely gain of function, possibility increasing affinity of DnaA and ResR proteins.

2.4 Discussion

TB is a treatable and curable disease, but treatment fails in part due to the inability of antibiotics to clear *Mtb* (1, 5). In this new genomic age, investigation of *Mtb* clinical evolution by whole-genome sequencing can provide valuable insights into genetic determinants of

bacterial adaptation to selective pressure in the face of host and antibiotics. Under the hypothesis that parallel evolution of genetic mutations was likely to correlate with fitness advantage *in vivo*, a large genomic analysis of *Mtb* clinical isolates from patient samples identified genomic loci repeatedly and independently enriched with mutations. Among these genomic loci were genes encoding two essential DNA binding proteins, *dnaA* and *resR*, and one intergenic region, *Rv0010c-Rv0011c* IGR. Although initially investigated as independent genes, we found that clinically relevant variants of *dnaA* and *resR* shared similar drug and morphology phenotypes. Single nucleotide changes in these two genes both promoted low-level INH resistance as well as a new antibiotic-related phenomenon: antibiotic resilience. Antibiotic resilience was characterized by faster regrowth post antibiotics pressure (16, 17). Although variant cells showed signs of resilience, the phenotype we observed was much more minor in *dnaA* variants. We also needed a more sensitive assay to test resilience, as cell cycle regulated phenotypes can be subtle and hard to detect in an asynchronous population. Interestingly, the differences in the degree of resilience observed in *dnaA* and *resR* correlated with their shared morphology and associated polar growth phenotype. Variations in *dnaA* and *resR* were found to primarily increase cell length, and this morphological difference was not due to impaired chromosomal replication rate or overall cell growth rate. Future experiment could further explore the role of polar growth dynamics in antibiotic resilience.

dnaA encodes the essential cell cycle regulator DnaA and is required for replication initiation. DnaA also doubles as a transcription factor, directly controlling expression of regulatory genes in later stages of cell cycle (23). *resR* encodes an essential and conserved DNA binding protein with unknown function. ResR was predicted to be a transcription factor based on structural similarity to MerR-like transcription factor family protein and function to control

division related genes through cell division regulator WhiB2 (16, 25). Given DnaA and ResR's role in replication and division, respectively, we speculate that coordination of progression of major cell cycle events likely plays a role in antibiotic resilience, as the regrowth phenotype is tightly linked to regulation of growth and division. It would be interesting to investigate other processes of ResR to shine more lights on processes behind resilience. Previous binding data showed ResR interacted strongly with multiple other sites in addition to *whiB2* promoter. Most of these areas were next to genes encoding proteins of unknown function and it was difficult to pinpoint which biological processes they play in. Future study is needed to prove the connection between fundamental cellular processes and antibiotics survival in mycobacteria.

Before our study, there was no link between DnaA and ResR despite their roles in regulating two processes, replication and division, known to be coupled to ensure optimal progression of cell cycle events (22). Given the DNA binding nature of DnaA and ResR, we analyzed their *in vitro* genomic binding profiles and identified a common binding site at the intergenic region between two genes *Rv0010c* and *Rv0011c*. *Rv0010c-Rv0011c* intergenic region (IGR) was one of the most highly mutated non-coding regions on the genome and we found single nucleotide variations in the IGR led to similar drug and morphology phenotype as *dnaA* and *resR* variants. Mutational hotspots within the IGR aligned with DnaA and ResR binding sites. Phenocopying of *Rv0010c-Rv0011c* IGR variants provided evidence for clinical selection on a genetic pathway with DnaA and ResR that acted its regulatory role through *Rv0010c-Rv0011c*. However, the function of the *Rv0010c-Rv0011c* operon remained unclear. Through literature search, we found that Rv0011c encoded a cell division protein also known as CrgA and was anchored on the membrane and septum and thought play a role in early stages of divisome formation (33, 45). Rv0010c was less studied and annotated as a conserved membrane protein.

Homology search identified similar proteins in eukaryotes and implicated a role of Rv0010c in synthesis of membrane lipids such as PIMs. However, no existing literature has identified the role of Rv0010c in any lipid pathway. Both *Rv0010c* and *Rv0011c* are non-essential but highly conserved across mycobacteria. *Rv0010c-Rv0011c* intergenic region is also conserved in mycobacteria and the most conserved regions overlap with DnaA and ResR binding sites. This indicates binding of DnaA and ResR to this region is a conserved process in mycobacteria and yet this region is being mutated more than most genomic regions. We tried to identify functional consequences of isogenic IGR variants by transcriptional profiling. Although we found a slight regulation of *Rv0010c*, we were unable to prove the role of *Rv0010c* in regulating morphology and drug susceptibility by overexpression. Although we proved later in the results section that *Rv0010c-Rv0011c* IGR functioned independently from Rv0010c and Rv0011c, it would be interesting to characterize these two genes as they are well conserved across mycobacterial species.

To genetically understand the function of *Rv0010c-Rv0011c* operon, we constructed a knockout (KO) strain of this operon and showed KO results in morphology defect and increased sensitivity to first-line TB drug. KO phenotypes can be complemented by insertion of WT operon at phage integration site, although this complementation was imperfect. As the phage integration site was close to the terminal (*ter*) on the genome, this could be an effect on difference in copy number and potentially accessibility of DnaA (23, 46). Although morphology did not perfectly complement, complement WT did restore H37Rv WT level INH susceptibility, suggesting the role of IGR in drug response was unaffected. By proceeding to test requirement of elements to restore length and INH susceptibility, we found integrity of the genomic sequence, instead of the translatable coding regions, was required for maintaining normal morphology and

provides evidence for the non-coding regulatory role of *Rv0010c-Rv0011c* operon. Despite both coding genes being non-essential, Rv0010c-Rv0011c KO had a growth defect compared to WT. This is likely due to loss of *Rv0011c* (*crgA*) since the only other complementation strains that had similar growth defects had *Rv0011c* either deleted or disrupted with pre-mature stop codons. The encoded protein CrgA plays an important role in early divisome formation and cells might have to resort to alternative and less preferred methods to stabilize division (33, 47).

Complementation with operon carrying clinically single nucleotide changes in the IGR recapitulated the morphology and drug phenotype seen in isogenic variants, proving the relevance of our system. Interestingly, it shows opposite phenotypes in both morphology and drug susceptibility to operon carrying DnaA or ResR binding site deletions in the IGR. Loss of either one of the two protein binding sites failed to complement the KO phenotype, implying the requirement of binding of both proteins at the IGR for its regulatory role. Together, the genetic data identified a non-coding regulatory function of *Rv0010c-Rv0011c* operon tightly linked to DnaA and ResR binding at the intergenic region. This is an exciting finding because such non-coding regulatory elements are understudied in *Mtb* and studying it in the context of clinical selection could provide interesting insights into additional methods *Mtb* utilizes to regulate its processes. Interestingly, DnaA and ResR are not the only transcription factors that could bind this region. Previous chromatin immunoprecipitation sequencing (ChIP-seq) dataset on *Mtb* transcription factors showed that multiple transcriptional factors, including nucleoid associated proteins, bind in this region near the mutational hotspots (48). Interaction of DnaA or ResR with unknown protein partner is evident by the fact that complementation with IGR alone was not sufficient to restore WT morphology. However, DnaA and ResR are the dominant proteins in this region as complementation fails when one of these proteins cannot associate with this IGR. As a

highly trafficked site by essential regulatory proteins, *Rv0010c-Rv0011c* IGR potentially exerts a conserved regulatory function which remains an unknown part of *Mtb* biology.

Another interesting aspect we observed along the way was the difference in the function of *Rv0010c-Rv0011c* operon in *Mtb* compared to in *Mycobacterium smegmatis* (*Msm*), the faster growing model mycobacterium used to study *Mtb* outside Biosafety Level-3 labs. We created a knockout of this operon in *Msm* and it also resulted in a morphology defect but in an opposite way to *Mtb* KO. *Mtb* KO is smaller than wild tyoe while *Msm* KO is much longer than wild type. One of the possible explanations is that coordination of cell cycle events such as replication and division is different in slow-growing *Mtb* and fast-growing *Msm*. A recent study showed that in one doubling *Mtb* overall spends more time in the period between previous division (birth) and replication initiation and less time in the period with active chromosomal replication than *Msm* (49). Coordination of cell cycle events requires tight regulation of regulatory proteins for each stage and contribute to differences in *Rv0010c-Rv0011c* operon function in these two bacteria.

In summary, in this chapter, we showed our identification and characterization of an essential genetic pathway under selection in clinical *Mtb* population. This pathway is tightly linked to cell cycle event and functions downstream through *Rv0010c-Rv0011c* intergenic region. Variations in this pathway directly contribute to changes in morphology and drug susceptibility in the form of low-level resistance and antibiotic resilience to facilitate *Mtb* escape from drug. Through genetic characterization, we showed that *Rv0010c-Rv0011c* IGR functions as a non-coding regulatory element and directly binding of DnaA and ResR is required for its function. This result supports the hypothesis that clinical single nucleotide polymorphisms (SNPs) in IGR are a gain of function of protein binding. In the next chapter we will directly investigate the functional consequences of clinical SNPs *in vitro* and *in vivo*.

2.5 Material and Methods

Bacterial strains and growth conditions

Mtb strains used in this study are constructed in the background of wild type lab strain H37Rv and *Msm* strains used are constructed in the background of wild type mc^2155. *Mtb* strains were grown in liquid in Middlebrook 7H9 media (BD) supplemented with 10% oleic acid albumin-dextrose-catalase (OADC) (BD), 0.2% glycerol, and 0.05% Tween-80 (referred to as 7H9 in following sections) at 37 ºC. Solid media for *Mtb* were made with Middlebrook 7H10 Agar (BD) supplemented with 10% oleic acid albumin-dextrose-catalase (OADC) (BD) and 0.2% glycerol (referred to as 7H10 in following sections). *Msm* strains were grown in Middlebrook 7H9 media (BD) supplemented with 10% albumin-dextrose-catalase (ADC containing a final concentration of 5g/L bovine serum albumin, 2g/L dextrose, and 3mg/L catalase), 0.2% glycerol, and 0.05% Tween-80 at 37 ºC. Solid media for *Msm* were made with Middlebrook 7H10 Agar (BD) supplemented with 10% albumin-dextrose-catalase (ADC) (BD) and 0.2% glycerol. Antibiotics were supplemented to media at the following final concentration when needed: Kan (20μg/mL), Hyg (50μg/mL), Nat (25μg/mL), Zeo (20μg/mL). Induction by anhydrotetracycline (aTC) were done at final concentration of 100ng/ml. Bacterial growth were measured by optical density at 600nm wavelength (OD_{600}).

Minimum inhibitory concentration measurement

Minimum inhibitory concentration (MIC) for anti-tuberculosis drugs were measured by alamar blue reduction assay. *Mtb* strains were started from frozen stocks and grown to mid-late log phase at OD_{600} = 0.5-0.8 before back diluting to OD = 0.05 in 10ml 7H9 media supplemented with appropriate antibiotics. Cells were allowed to grow for another two days to OD = 0.3. Serial or specific dilutions of antibiotics were prepared in 96 well plates and *Mtb* cells

were added to a final OD = 0.00015 in 200µl total volume per well in technical duplicates. Plates were incubated in an airtight container for 7-8 days total at 37 ºC. On day 6, 0.0002% final concentration of alamarBlue (BioRad) reagent were added to each well and allowed to incubate for 24-48 hours before reading on a plate reader. Absorbance profile was measured at OD = 600nm and OD = 570nm and subtraction of OD_{600} from OD_{570} was used to calculate percentage of viability in each well compared to no-drug control. Final growth inhibition profiles were plotted, and minimum inhibitory concentration (IC_{50}) was determined after performing linear regression on viability data at each antibiotic concentration in Prism Graphpad.

Antibiotic resilience assay

Antibiotic resilience was measured by two metrics: 1) immediate recovery by fluorescent D-amino acid (FDAA) labeling and 2) by colony formation on solid media. *Mtb* strains were started from frozen stocks and grown to mid-late log phase at OD_{600} = 0.5-0.8 before back diluting to OD = 0.05 in 10ml 7H9 media supplemented with appropriate antibiotics. Two parallel cultures were prepared for each strain. After the strains grew to 0.2, 1ml of each culture was seeded into 24-well plates and added 100X MIC of each strain based on IC_{50} calculations. After 1 day, cells were washed with equal volume of 7H9 media two times before being resuspended in 1ml 7H9 media + 10nM NADA (FDAA, Tocris Bioscience). For colony formation, washed cells were diluted in 5-fold serial dilutions and 10µL of each dilution was spotted on 7H10 agar in square petri dish. Plates were incubated for 18-26 days and images of colonies were taken to quantify colony growth.

For immediate recovery, labeling of NADA was quantified by flow cytometry. Washed cells were allowed to recover at 37ºC for 24 hours and were mixed equal volume with 4% paraformaldehyde (PFA) and incubated for at least 1 hour at room temperature. Cells were spun

down and quenched with 200mM Tris-HCl (pH 7.5), and resuspended in PBS + 0.1% Triton X-100. Flow cytometry of NADA labeling was performed on a MACSQuant Analyzer 10 using the FITC fluorescent setting (channel B1, excitation at 488nm and emission at 525/50nm). Specific trigger setting and manual gating were used to select singlets and FITC fluorescence profile were plotted using FlowJo.

Pulse-chase labeling

Pulse-chase labeling was performed using NADA as pulse and HADA (Tocris Bioscience) as chase in *Mtb*. Strains were started from frozen stocks and grown to mid-late log phase at OD_{600} = 0.5-0.8 before back diluting to OD = 0.05 in replicates in 10ml 7H9 media + 10nM NADA supplemented with appropriate antibiotics. Cells were grown for two days to OD_{600} ~ 0.3 and washed two times with equal volume of plain 7H9. Cells were resuspended in equal volume of 7H9 + 25nM HADA with appropriate antibiotics. After 24 hours, cells were harvested by mixing equal volume with 4% PFA and incubated for at least 1 hour at room temperature. PFA-fixed cells are stored at 4^{o}C prior to microscopy.

Quantitative microscopy

As mycobacterial cells are prone to aggregates, PFA-fixed cells were first processed to remove aggregates to improve imaging quality. 100µl of PFA-fixed cells were centrifuged at 3200g for 15min and then resuspend in 200µl disaggregation buffer (200mM Tric-HCl pH 7.5, 1% (w/v) Triton X-100, 0.67% (w/v) Xylenes and 0.33% (w/v) Heptane). Cells were immediately centrifuged at the same setting for 15min. Then cells were washed one time in 200µl PBS + 0.1% Triton X-100 before resuspending in 20-25µl PBS + 0.1% Triton X-100 to ideal turbidity. 1.5µl of resuspended cells were spotted onto a custom-made 96 well pedestal array made from 2% agarose in 1X + PBS supplemented with optional Nile Red and DAPI stain.

If using DAPI, the assembled agarose cartridge was incubated at room temperature for at least 30min for DAPI to fully penetrate cells. Imaging was performed on a Nikon Ti-E inverted, widefield microscopy equipped with a Plan Apo 100x 1.45 NA objective lens and an Andor Zyla sCMOS camera. A custom Nikon JOBS script on the NIS Elements software was designed to perform automated imaging in 96 well format and captured 2 images each at the pre-designed 12 fields of interest for each sample. Fluorescent signal was captured using a 6-channel Spectra X LED light source and the Sedat Quad filter set. Excitation (Ex) and emission (Em) filters were set at Ex 395/25nm and Em 435/25nm for HADA/DAPI, Ex 470/24nm and Em 515/25nm for NADA (FITC), and Ex 550/15nm and Em 595/25nm for Nile Red. Exposure time was set at 100ms for all channels. Post imaging morphological profiling was performed using a previous described microscopy analysis pipeline MOMIA (32). Single cells were filtered and traced along the cellular midline and their fluorescent profile was used to calculate polar outgrowth.

Transcriptional profiling

For RNA extraction, *Mtb* strains were started from frozen stocks and grown to mid-late log phase at OD_{600} = 0.5-0.8 before back diluting to OD = 0.05 in triplicates in 10ml 7H9 media supplemented with appropriate antibiotics. After 3 days, cells are grown to mid log OD = 0.5-0.6 and centrifuged at room temperature to collect pellet. Cell pellets were resuspended in 1ml TRIzol Reagent (Thermo Fisher) and lysed by bead beating to release nucleic acid that was further extraction by adding 300µL chloroform. Total RNA was then isolated using the Direct-zol RNA Miniprep Kit (Zymo Research). DNA digestion was done using TURBO DNase kit (Thermo Fisher) and post depletion total RNA was re-purified using RNA Clean and Concentrator kit (Zymo Research).

For real time qRT-PCR, cDNA was generated from total RNA using SuperScript IV First Strand cDNA synthesis kit with random hexamers (Thermo Fisher). qPCR reactions were prepared with iTaq Universal SYBR-Green master mix (BioRad) and were run on a ViiA7 real time PCR system (Thermo Fisher). Raw CT values were analyzed with the ΔΔCT methods and all experimental gene expressions were normalized to *sigA* expression. For RNA-seq, RNA library was prepped using the KAPA RNA HyperPrep kit (KAPA Biosystems) according to manufacturer's instructions. rRNA was depleted using the KAPA RiboErase kit (KAPA Biosystems) with custom *Mtb* specific rRNA targeting oligos. RNA libraries were sequenced on an Illumina MiSeq sequencer with a 100bp single-ended reads. FASTQ files generated form sequencing had adapters removed using cutadapt and were mapped to the *Mtb* genome (NC_000962) using bowtie2 using the –very-sensitive argument set (50, 51). Using a custom script, counts were applied to genes if any part of the fragment defined by the paired end read overlapped with the gene. Analysis of transcriptomic expression changes based on gene counts was done using an R package, DEseq2 (52). For each variant strain, $\log_2$ fold change and adjusted p-value following multiple hypothesis correction was generated for each gene compared to its expression in WT strain.

Mutant construction

Mtb isogenic point variants were constructed using oligo recombineering (53). *Mtb* WT H37Rv was first transformed with two plasmids. One plasmid was an integrating plasmid that has an inactive hygromycin resistance gene with pre-stop codon inserted, and the other was an episomal plasmid carrying the phage recombinase gene *recT* whose expression was controlled by an anhydrotetracycline (aTC) -inducible promoter. This recombineering strain was grown to mid-log phase before inducing with aTC and was transformed with two oligos, one oligo with

sequencing to correct the prestop codon within the hygromycin resistance gene and one oligo with sequencing to create the desired single nucleotide change. The transformants recovered in plain 7H9 media for 72 hours at 37ºC before plating on 7H10 agar plate supplemented with hygromycin. Once single colonies came up on the plate, colonies were check by PCR and sanger sequencing to confirm mutant identify. Correct colonies were then patched to one plate with hygromycin and one plate with kanamycin, and only colonies that have lost the episomal plasmid and regained kanamycin sensitivity were stocked. Colonies in transformation with WT oligo were used in future experiments as WT for each set of corresponding isogenic variants.

Knockout mutants were generated in H37Rv using dsDNA recombineering with phase RecET recombinase as previously described (54). The knockout cassette included 500 bp upstream and downstream of the *Rv0011c-Rv0010c* coding region. *Rv0010c-Rv0011c* was replaced with a zeocin-resistance cassette at its native site. Recombinants were selected on 7H10 agar with zeocin and correct colonies confirmed by PCR were also patched onto zeocin and kanamycin plates to make sure they have lost the episomal RecET plasmid. Complementation of *Rv0010c-Rv0011c* knockout was constructed by adding operon elements in combinations after its native promoter region (the entire 5' UTR region of *Rv0011c*) to a L5 integrating plasmid marked by nourseothricin (nat) resistance. Inducible *Rv0010c* and *Rv0011c* were constructed by inserting *Rv0010c* or *Rv0011c* downstream of an aTC-inducible cassette on a L5 integrating plasmid carrying the TetOR system (55). Overexpression of *Rv0010c* was done by placing *Rv0010c* downstream of a constitutively active *tetO* promoter on a plasmid without *tetR*.

CRISPR interference (CRISPRi)

CRISPRi plasmid was designed and constructed as described previously (56). Guide RNA for specific gene knockdown was designed using the online sgRNA design tool from the Rock lab at Rockefeller University (https://pebble.rockefeller.edu/tools/sgrna-design/) (57). sgRNA sequences was ordered from IDT and cloned into CRISPRi plasmid backbone by restriction digest and Gibson assembly. CRISPRi system was induced by aTC at 100ng/ml and strains were induced for five day prior to phenotyping to ensure proper knockdown.

Growthcurve and colony formation measurement

Growth was measured by both liquid growth and solid media colony formation. *Mtb* strains were started from frozen stocks and grown to mid-late log phase at OD_{600} = 0.5-0.8 before back diluting to OD = 0.005 in biological duplicates in 10ml 7H9 media supplemented with appropriate antibiotics. Cells were then sampled every 2 days for liquid growth measurement with OD_{600} and colony formation by serial diluting and plating.

gDNA sequencing and *ori/ter* ratio

gDNA was extracted in biological triplicate cultures of strains grown to mid-log. Cells were lysed by bead beating and nucleic acid was extracted with phenol-chloroform. DNA library prep and Illumina whole genome sequencing were performed by SeqCenter in Pittsburgh, PA. After sequencing, Illumina sequencing reads were quality filtered and adapters trimmed with fastp (58). Sequencing reads were aligned to the H37Rv (NC_000962.3) reference genome using bwa and duplicates marked with picard MarkDuplicates (59, 60). Depth of coverage was calculated for each base using the GATK DepthOfCoverage tool with a coverage threshold of 10 (60). Coverage was averaged into 1000 bp non overlapping windows across the entire genome with a custom perl script. Coverage across the genome was then plotted using Prism Graphpad and owess smoothing was performed. LOWESS smoothed coverage values were used to

calculated ori/ter ratio as peak values were close to *ori*-proximal region and the lowest values were near *ori*-distal region halfway through the genome. Unpaired t-test was performed in Prism on biological triplicates of each variant compared to WT.

2.6 References

1. World Health Organization. Global tuberculosis report 2023. Geneva: World Health Organization; 2023 2023.

2. World Health Organization. WHO consolidated guidelines on tuberculosis: module 4: treatment: drug-susceptible tuberculosis treatment. Geneva: World Health Organization; 2022 2022.

3. World Health Organization. WHO consolidated guidelines on tuberculosis: module 4: treatment: drug-resistant tuberculosis treatment. 2022 update ed. Geneva: World Health Organization; 2022 2022.

4. Farhat M, Cox H, Ghanem M, Denkinger CM, Rodrigues C, Abd El Aziz MS, et al. Drug-resistant tuberculosis: a persistent global health concern. Nature Reviews Microbiology. 2024:1-19.

5. Ramaswamy S, Musser JM. Molecular genetic basis of antimicrobial agent resistance in Mycobacterium tuberculosis: 1998 update. Tubercle and Lung disease. 1998;79(1):3-29.

6. Gillespie SH. Evolution of drug resistance in Mycobacterium tuberculosis: clinical and molecular perspective. Antimicrobial agents and chemotherapy. 2002;46(2):267-74.

7. Hicks ND, Yang J, Zhang X, Zhao B, Grad YH, Liu L, et al. Clinically prevalent mutations in Mycobacterium tuberculosis alter propionate metabolism and mediate multidrug tolerance. Nat Microbiol. 2018;3(9):1032-42.

8. Zhang H, Li D, Zhao L, Fleming J, Lin N, Wang T, et al. Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nature genetics. 2013;45(10):1255-60.

9. Zhang Y, Heym B, Allen B, Young D, Cole S. The catalase—peroxidase gene and isoniazid resistance of Mycobacterium tuberculosis. Nature. 1992;358(6387):591-3.

10. Ramaswamy SV, Reich R, Dou S-J, Jasperse L, Pan X, Wanger A, et al. Single nucleotide polymorphisms in genes associated with isoniazid resistance in Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy. 2003;47(4):1241-50.

11. Telenti A, Imboden P, Marchesi F, Matter L, Schopfer K, Bodmer T, et al. Detection of rifampicin-resistance mutations in Mycobacterium tuberculosis. The Lancet. 1993;341(8846):647-51.

12. Palomino JC, Martin A. Drug resistance mechanisms in Mycobacterium tuberculosis. Antibiotics. 2014;3(3):317-40.

13. Brauner A, Fridman O, Gefen O, Balaban NQ. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. Nature Reviews Microbiology. 2016;14(5):320-30.

14. Martini MC, Hicks ND, Xiao J, Alonso MN, Barbier T, Sixsmith J, et al. Loss of RNase J leads to multi-drug tolerance and accumulation of highly structured mRNA fragments in Mycobacterium tuberculosis. PLoS Pathogens. 2022;18(7):e1010705.

15. Safi H, Gopal P, Lingaraju S, Ma S, Levine C, Dartois V, et al. Phase variation in Mycobacterium tuberculosis glpK produces transiently heritable drug tolerance. Proc Natl Acad Sci U S A. 2019;116(39):19665-74.

16. Liu Q, Zhu J, Dulberger CL, Stanley S, Wilson S, Chung ES, et al. Tuberculosis treatment failure associated with evolution of antibiotic resilience. Science. 2022;378(6624):1111-8.

17. Hicks ND, Giffen SR, Culviner PH, Chao MC, Dulberger CL, Liu Q, et al. Mutations in dnaA and a cryptic interaction site increase drug resistance in Mycobacterium tuberculosis. PLOS Pathogens. 2020;16(11):e1009063.

18. Colangeli R, Jedrey H, Kim S, Connell R, Ma S, Chippada Venkata UD, et al. Bacterial Factors That Predict Relapse after Tuberculosis Therapy. N Engl J Med. 2018;379(9):823-33.

19. World Health Organization. Guidelines for treatment of drug-susceptible tuberculosis and patient care, 2017 update. 2017.

20. Zhu J, Liu YJ, Fortune SM. Spatiotemporal perspectives on tuberculosis chemotherapy. Curr Opin Microbiol. 2023;72:102266.

21. Prideaux B, Via LE, Zimmerman MD, Eum S, Sarathy J, O'Brien P, et al. The association between sterilizing activity and drug distribution into tuberculosis lesions. Nature Medicine. 2015;21(10):1223-7.

22. Willis L, Huang KC. Sizing up the bacterial cell cycle. Nat Rev Microbiol. 2017;15(10):606-20.

23. Katayama T, Kasho K, Kawakami H. The DnaA Cycle in Escherichia coli: Activation, Function and Inactivation of the Initiator Protein. Frontiers in Microbiology. 2017;8.

24. Seid CA, Smith JL, Grossman AD. Genetic and biochemical interactions between the bacterial replication initiator DnaA and the nucleoid-associated protein Rok in Bacillus subtilis. Mol Microbiol. 2017;103(5):798-817.

25. Zhou W, Huang S, Cumming BM, Zhang Y, Tang W, Steyn AJ, et al. A feedback regulatory loop containing McdR and WhiB2 controls cell division and DNA repair in mycobacteria. mbio. 2022;13(2):e03343-21.

26. Zhang H, Li D, Zhao L, Fleming J, Lin N, Wang T, et al. Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nat Genet. 2013;45(10):1255-60.

27. Aldridge BB, Fernandez-Suarez M, Heller D, Ambravaneswaran V, Irimia D, Toner M, et al. Asymmetry and Aging of Mycobacterial Cells Lead to Variable Growth and Antibiotic Susceptibility. Science. 2012;335(6064):100-4.

28. Kieser KJ, Rubin EJ. How sisters grow apart: mycobacterial growth and division. Nat Rev Microbiol. 2014;12(8):550-62.

29. Hannebelle MTM, Ven JXY, Toniolo C, Eskandarian HA, Vuaridel-Thurre G, McKinney JD, et al. A biphasic growth model for cell pole elongation in mycobacteria. Nature Communications. 2020;11(1).

30. Logsdon MM, Aldridge BB. Stable regulation of cell cycle events in mycobacteria: Insights from inherently heterogeneous bacterial populations. Frontiers in Microbiology. 2018;9:348921.

31. Poulton NC, Rock JM. Unraveling the mechanisms of intrinsic drug resistance in Mycobacterium tuberculosis. Frontiers in Cellular and Infection Microbiology. 2022;12:1573.

32. Zhu J, Wolf ID, Dulberger CL, Won HI, Kester JC, Judd JA, et al. Spatiotemporal localization of proteins in mycobacteria. Cell reports. 2021;37(13).

33. Plocinski P, Ziolkiewicz M, Kiran M, Vadrevu SI, Nguyen HB, Hugonnet J, et al. Characterization of CrgA, a New Partner of the Mycobacterium tuberculosis Peptidoglycan Polymerization Complexes. Journal of Bacteriology. 2011;193(13):3246-56.

34. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9.

35. Van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. Nature Biotechnology. 2024;42(2):243-6.

36. Dulberger CL, Rubin EJ, Boutte CC. The mycobacterial cell envelope - a moving target. Nat Rev Microbiol. 2020;18(1):47-59.

37. Bansal-Mutalik R, Nikaido H. Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. Proceedings of the National Academy of Sciences. 2014;111(13):4958-63.

38. Fukuda T, Matsumura T, Ato M, Hamasaki M, Nishiuchi Y, Murakami Y, et al. Critical roles for lipomannan and lipoarabinomannan in cell wall integrity of mycobacteria and pathogenesis of tuberculosis. MBio. 2013;4(1):10.1128/mbio. 00472-12.

39. Koul A, Herget T, Klebl B, Ullrich A. Interplay between mycobacteria and host signalling pathways. Nature Reviews Microbiology. 2004;2(3):189-202.

40. Abrahams KA, Besra GS. Mycobacterial cell wall biosynthesis: a multifaceted antibiotic target. Parasitology. 2018;145(2):116-33.

41. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Research. 2014;43(D1):D447-D52.

42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.

43. Shell SS, Wang J, Lapierre P, Mir M, Chase MR, Pyle MM, et al. Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. PLOS Genetics. 2015;11(11):e1005641.

44. Junier I, Unal EB, Yus E, Lloréns-Rico V, Serrano L. Insights into the mechanisms of basal coordination of transcription using a genome-reduced bacterium. Cell systems. 2016;2(6):391-401.

45. Plocinski P, Arora N, Sarva K, Blaszczyk E, Qin H, Das N, et al. Mycobacterium tuberculosis CwsA interacts with CrgA and Wag31, and the CrgA-CwsA complex is involved in peptidoglycan synthesis and cell shape determination. Journal of bacteriology. 2012;194(23):6398-409.

46. Charbon G, Schei Haugan M, Frimodt-Møller N, Løbner-Olesen A. Counting Replication Origins to Measure Growth of Pathogens. Antibiotics. 2020;9(5):239.

47. Baranowski C, Rego EH, Rubin EJ. The dream of a mycobacterium. Microbiology spectrum. 2019;7(2):10.1128/microbiolspec. gpp3-0008-2018.

48. Minch KJ, Rustad TR, Peterson EJR, Winkler J, Reiss DJ, Ma S, et al. The DNA-binding network of Mycobacterium tuberculosis. 2015;6:5829.

49. Chung ES, Kar P, Kamkaew M, Amir A, Aldridge BB. Mycobacterium tuberculosis grows linearly at the single-cell level with larger variability than model organisms. BioRxiv. 2023:2023.05. 17.541183.

50. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011;17(1):10-2.

51. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9(4):357-9.

52. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15:1-21.

53. Murphy KC. Oligo-mediated recombineering and its use for making SNPs, knockouts, insertions, and fusions in Mycobacterium tuberculosis. Mycobacteria Protocols. 2021:301-21.

54. Murphy KC, Papavinasasundaram K, Sassetti CM. Mycobacterial recombineering. Mycobacteria Protocols. 2015:177-99.

55. Ehrt S, Guo XV, Hickey CM, Ryou M, Monteleone M, Riley LW, et al. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. Nucleic acids research. 2005;33(2):e21-e.

56. Rock JM, Hopkins FF, Chavez A, Diallo M, Chase MR, Gerrick ER, et al. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. Nat Microbiol. 2017;2:16274.

57. Bosch B, DeJesus MA, Poulton NC, Zhang W, Engelhart CA, Zaveri A, et al. Genome-wide gene expression tuning reveals diverse vulnerabilities of M. tuberculosis. Cell. 2021;184(17):4579-92. e24.

58. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i90.

59. Li H. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;38:1767.

60. Van der Auwera GA, O'Connor BD. Genomics in the cloud: using Docker, GATK, and WDL in Terra: O'Reilly Media; 2020.

Chapter 3: Biochemical and functional characterization of variations in *Rv0010c-Rv0011c* intergenic region

3.1 Abstract

Tuberculosis (TB) remains a global health crisis, claiming 1.5 million lives annually despite being treatable. Treatment outcomes vary in patients, in part due to drug resistance rising in its causative agent *Mycobacterium tuberculosis* (*Mtb*). Large genomic studies in clinical *Mtb* isolates identified genetic elements under selection, including novel determinants like *dnaA*, *resR*, and the intergenic region (IGR) of *Rv0010c-Rv0011c* operon. *dnaA* variants exhibited low-level resistance to isoniazid (INH) and *resR* variants displayed both low-level INH resistance and faster regrowth post-antibiotic exposure, namely antibiotic resilience. Notably, both proteins regulated cell cycle events and had binding sites at *Rv0010c-Rv0011c* IGR, prompting investigation into their coordination. Genetic characterization of *Rv0010c-Rv0011c* demonstrated a non-coding regulatory role of *Rv0010c-Rv0011c* IGR dependent on DnaA and ResR protein binding and suggested altered interaction dynamics at this site when mutated. *in vitro* binding assays revealed cooperative binding of DnaA and ResR to *Rv0010c-Rv0011c* IGR, with clinical SNPs enhancing this interaction. Transcriptomic analysis unveiled downstream effects, notably on cell cycle related genes and regulatory genes such as *whiB2*, a known regulatory target downstream of ResR. Combinations of *in vitro* binding and *in vivo* transcriptional response led us to propose a model where clinical variations at *Rv0010c-Rv0011c* sequestered DnaA and ResR and led to differences in cell cycle dynamics, influencing both cell morphology and drug susceptibility. Together these data provided a novel connection of cell cycle regulation to *Mtb* drug response.

3.2 Introduction

Tuberculosis remains the world's deadliest infectious disease, claiming on average 1.5 million lives every year. Although TB is treatable and curable, treatment outcome varies. On average 15% of patient undergoing treatment for drug susceptible TB and 37% of patient with drug resistant TB fail treatment (1). TB treatment is lengthy and complicated, with treatment outcome depending on a series of host and bacterial factors (2-4). The causative agent of TB, *Mycobacterium tuberculosis* (*Mtb*), has been shown to survive antibiotics treatment by directly mutating drug target genes to confer high-level drug resistance (5, 6). With the advances in sequencing technologies, genome-wide studies have revealed non-canonical pathways that enhance *Mtb* antibiotic survival through low-level resistance, tolerance, and persistence (7-9). Given the repeated occurrence of mutations that allowed *Mtb* to adapt to drug pressure, we hypothesized by comparing of a large number of *Mtb* clinical populations isolated from patients, we can identify mutations that accumulate independently and reveal ongoing evolution in genetic elements that conferred fitness advantage under antibiotic pressure *in vivo* (10). Analysis of 50,000 clinical isolates from patients' sputum identified coding and non-coding regions under selection. In addition to previous characterized genes conferring high-level resistance, we identified interesting genetic elements that had no known links to drug resistance and yet were frequently mutated. This included an essential genetic pathway involving two essential DNA binding proteins DnaA and ResR, and one intergenic region *Rv0010c-Rv0011c* (10).

All three elements of this pathway were first identified as novel drug resistance determinants in genome-wide association studies with clinical resistance (7, 8). *dnaA* encodes the conserved replication initiator protein in mycobacteria, and isogenic strains carrying *dnaA* clinical variants showed low level resistance to first-line TB drug isoniazid (INH) (7). This low-

level resistance to first line TB drugs has been shown to sufficiently increase the likelihood of treatment failure in patients (11). *dnaA* variants also have increased cell length, and this morphology phenotype was not due to changes in replication or growth rate (7). *resR* was a conserved hypothetical protein before being characterized in 2022. It encodes a conserved and essential transcriptional regulator regulating cell division and DNA repair in mycobacteria (12). Our study on *resR* variants revealed a new antibiotic phenotype characterized by faster regrowth post antibiotic exposure (10). This phenotype is clinically relevant as plasma concentrations of INH, RIF, and EMB fall below MIC for 21%, 42%, and 75% of treatment time respectively (13). ResR directly binds to the promoter of *whiB2*, an essential transcription regulator of cell division genes, and regulates *whiB2* expression (10, 12). Having discovered that, we tested *dnaA* variants for resilience phenotype and found signatures of resilience although at a much more modest level. In *resR* variants, the resilience phenotype was accompanied by similar morphology phenotype observed in *dnaA* variants. Given the regulatory role of DnaA in replication and role of ResR in division, we proposed that the resilience phenotype can be manifested through coordination of replication and division. However, there has been no existing link between these two proteins and it is unclear how such coordination between two essential regulatory proteins.

In the previous chapter, we analyzed their *in vitro* binding profiles knowing both proteins bind DNA and identified abutting binding sites in the intergenic region between gene *Rv0010c* and *Rv0011c*. Interestingly, the *Rv0010c-Rv0011c* IGR was one of the most highly mutated non-coding regions in our genomic analysis and DnaA and ResR binding site directly overlapped with clinical mutational hotspots. Single nucleotide changes in the IGR shared similar resilience and morphology phenotype as *dnaA* and *resR* variants. This provided evidence for a genetic pathway under selection through *Rv0010c-Rv0011c*, that likely coordinated DnaA and ResR in

their function downstream. However, little was known about this operon and its function. *Rv0010c-Rv0011c* operon consists of two co-transcribed genes and a 155 bp intergenic region. *Rv0011c* encodes a protein named CrgA, and CrgA has been shown to play a role in division and thought to anchor the divisome (14, 15). *Rv0010c* encodes a conserved membrane protein with unknown function. AlphaFold prediction showed Rv0010c to be a small protein with N-terminal transmembrane helices and homology search revealed structural similarities to eukaryotic phosphatidylinositol transferases (16, 17). Both genes are non-essential and yet well conserved across mycobacteria (18, 19). The intergenic region is also well conserved and alignment of this region across mycobacteria shows the most conserved nucleotides are the DnaA and ResR binding sites (19). However, this intergenic region was mutating more rapidly than most other sites on the genomes of clinical *Mtb*.

To answer why this intergenic region was accumulating mutations, we needed to understand the function of this operon. We took a genetic approach and created a knockout (KO). KO cells were much shorter in cell length and more sensitive to INH. Complementation experiments with components of this operon revealed requirement of complete genomic sequence of this operon independent of translation of the two coding genes for regulating morphology and drug sensitivity. Direct overexpression and knockdown of coding genes also did not affect morphology or drug susceptibility. This led us to conclude that *Rv0010c-Rv0011c* operon serves a non-coding regulatory role in regulating morphology and drug susceptibility. Complementation with clinical single nucleotide changes (SNPs) in IGR not only complemented the KO shorter cell phenotype but also recapitulated the increase in cell length in isogenic variants compare to wild type. This was opposite to phenotypes of KO strains complemented with DnaA or ResR binding sites deletion, which indicated the requirement of DnaA and ResR

binding to this region in morphological control. However, it remained unclear how these IGR variations, both SNPs and protein binding site deletions, affected downstream phenotypes.

In this chapter, we aim to functionally characterize the consequences of mutations in the intergenic region both *in vitro* and *in vivo*. By using *in vitro* DNA binding assays, we were able to test how protein affinity changes in the presence of SNP and binding site deletions. We reported that both DnaA and ResR can bind the IGR *in vitro* alone and yet they likely bound in a cooperative manner. Clinical SNPs in this region presumably increased affinity of this complex while binding deletions reduced affinity of this complex. As DnaA and ResR both can act as transcriptional regulators, we reasoned that sequestration of these two proteins at *Rv0010c-Rv0011c* IGR could lead to transcriptomic changes. To test this, we performed bulk RNA-seq in complementation strains with operon carrying WT, SNP, and binding site deleted IGR. Analysis of differentially expressed genes revealed a set of genes with opposite expression patterns in SNP and binding site deletion strains, and majority of these genes were cell wall related. Interestingly, one of these genes was *whiB2*, a known target of ResR. Most division genes were either under direct control of WhiB2 or under control of MtrA, a response regulator of two component system MtrAB whose expression was controlled by WhiB2. *whiB2* expression was downregulated in SNP strain while upregulated in binding deletion strain. Knowing ResR can act as a transcriptional activator, we proposed that clinical variations sequestered DnaA and ResR at *Rv0010c-Rv0011c*, leading to downregulation of its regulon. This sequestration was subtle and likely cell cycle dependent to ensure normal progression of cell cycle events but was sufficient to potentially alter morphology and regrowth dynamics.

3.3 Results

DnaA and ResR can bind at *Rv0010c-Rv0011c* intergenic region *in vitro*

Through genetic manipulation, we learned that *Rv0010c-Rv0011c* IGR clinical single nucleotide variants had opposite phenotypes to *Rv0010c-Rv0011c* IGR with DnaA or ResR binding site deletions. We hypothesized these clinical SNPs were gain of function and likely increased protein affinity. To test this, we used an *in vitro* DNA binding assay, electrophoretic mobility shift assay (EMSA), to directly measure affinity of DnaA and ResR individually to WT and variant *Rv0010c-Rv0011c* IGR (20). Conserved across bacteria, DnaA is known to bind adenosine nucleotides and ATP-bound DnaA (DnaA-ATP) is required for replication initiation. DnaA contains a AAA+ ATPase domain and requires exchange of bound ATP to ADP to inactivate the protein and prevent untimely initiation during other stages of cell cycle (21-23). In *Escherichia coli*, DnaA-ATP has higher affinity to the DnaA recognition sequences (DnaA boxes) at the origin of replication (*oriC*). In contrast, in mycobacteria, DnaA-ATP and DnaA-ADP have similar binding kinetics to *oriC* (23, 24). It was unclear if DnaA required a specific adenosine nucleotide for its binding at *Rv0010c-Rv0011c* intergenic region (IGR). To test this, we incubated DnaA with ATP or ADP prior to and supplied only ATP or ADP during incubation with wild type (WT) sequence of *Rv0010c-Rv0011c* IGR. DnaA can bind *Rv0010c-Rv0011c* IGR *in vitro* as expected. Quantification of EMSA traces shows there is no significant difference in binding kinetics in the presence of ATP compared to ADP (Supplementary Figure 3.1). This led us to conclude that DnaA was capable of specifically binding *Rv0010c-Rv0011c* IGR *in vitro*, and its binding to the region was not dependent on the type of nucleotide bound. Using the same assay, we were also able to show that ResR can bind *Rv0010c-Rv0011c* IGR specifically (Figure 3-1). Notably, the binding kinetics of these two proteins were different at *Rv0010c-Rv0011c* IGR

in vitro. DnaA was able to bind at lower concentration than ResR, implicating DnaA had higher affinity to this region than ResR alone. This correlated well with the fact that DnaA had

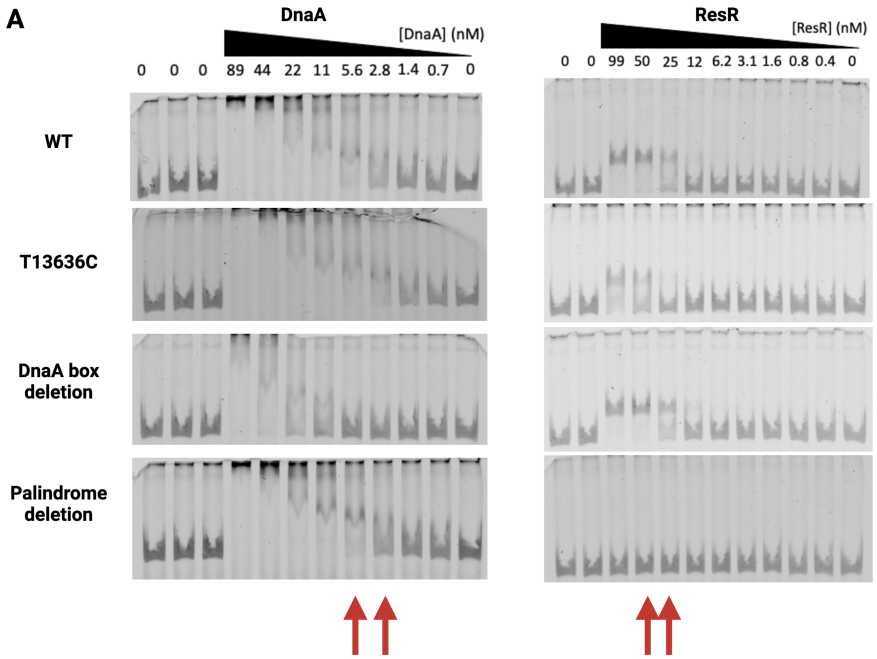


Figure 3-1 | In vitro DNA binding kinetics of DnaA and ResR with Rv0010c-Rv0011c IGR confirms altered binding affinity in the presence of SNP and binding site deletions.

DnaA and ResR binding kinetics to WT and variant *Rv0010c-Rv0011c* IGR measured with electrophoretic mobility shift assay (EMSA). Concentrations of proteins used are indicated on top of the gel in reaction loaded into each lane. DnaA were pre-incubated with 5mM ATP and reaction for both proteins were carried out at 2.5mM ATP. Probes for specific interaction (WT and variant *Rv0010c-Rv0011c* IGR) labeled with IRD800 dye and non-specific interaction (*sigA*, not shown) labeled with IRD700 dye were added at 0.1nM. Cold competitor Poly (dI-dC) was added at final concentration of 100 ng/µl. Proteins and probes were incubated at room temperature for 20min prior to loading into 6% DNA retardation gel. Gel images shown here was captured wcith LI-COR imagining system and analyzed in Empiria Studio. Red arrows indicate potential concentrations of interest. Computational tracing of these gels can be found in supplemental figure 3.2.

more predicted binding motifs at *Rv0010c-Rv0011c* IGR based on conserved binding sequence than ResR. ResR binding was concentrated around the palindromic sequence within the IGR while DnaA was capable of binding near the DnaA binding boxes and near the palindromes. Binding of ResR to a palindromic sequence was consistent with its structural homology to MerR-type transcriptional regulators, which normally bind DNA in dimers at palindromic sequences

(10, 25). Here, we demonstrated through *in vitro* DNA binding assay that DnaA and ResR can directly bind *Rv0010c-Rv0011c* IGR. The next step is to investigate whether variations in this IGR alter protein binding affinity.

Variations in the *Rv0010c-Rv0011c* IGR alter DnaA and ResR affinity *in vitro*

In the previous chapter, we have phenotypically characterized clinically relevant single nucleotide variants and protein binding site deletions in the forms of complementation strains. We discovered that single nucleotide variants resulted in opposite drug and morphology phenotypes to protein binding site deletions, implicating single nucleotide variations could be gain of function in protein binding. To test if variations in *Rv0010c-Rv0011c* IGR can affect binding affinity of DnaA and ResR, we incubated each protein with IGR variants in addition to wild type IGR sequence and use EMSA to quantify binding kinetics of both proteins. We found that single nucleotide variations increased affinity of DnaA while decreased affinity of ResR (Figure 3-1). For binding site deletions, we found that deletion of DnaA binding box reduced DnaA affinity and did not affect ResR affinity (Figure 3-1). Deletion of the palindrome abolished ResR's ability to bind at this IGR while it did not change affinity of DnaA. Our *in vitro* results confirmed that variations in the *Rv0010c-Rv0011c* IGR directly altered affinity of DnaA and ResR. The changes in binding affinity were opposite, consistent with earlier phenotypic data *in vivo*. However, one puzzling piece was the fact that single nucleotide difference in IGR did not affect DnaA and ResR in the same direction and deletion of binding sites only affected the affinity of the corresponding protein. These results suggested that binding relationships are likely more complicated than anticipated, and testing this in single protein format might not provide a complement understanding of the binding relationships.

In our previous complementation experiment, deletion of DnaA binding box and deletion of the palindrome produced similar drug and morphology phenotype, while our EMSA showed that each deletion only affected affinity of the corresponding protein. We hypothesized that these proteins could bind this IGR cooperatively as a complex and individual protein affinity could affect the affinity of the complex. To test the cooperative binding of two proteins at this IGR, we co-incubated both DnaA and ResR with WT IGR sequence. Here we kept a fixed concentration of one protein and added a range of concentrations of the second protein and asked if the second protein could bind at lower concentration in the presence of the first protein.

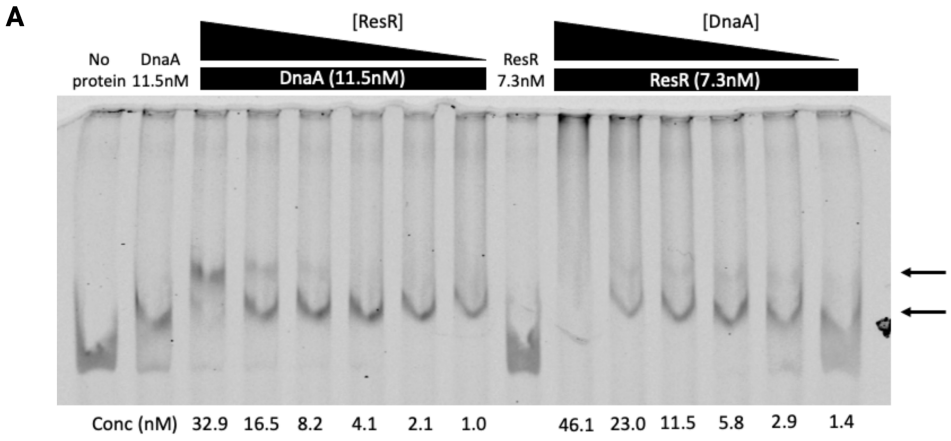


Figure 3-2 | Co-incubation shows cooperativity of DnaA and ResR binding at *Rv0010c-Rv0011c* IGR.

DnaA and ResR were co-incubated with WT Rv0010c-Rv0011c IGR across a gradient of concentrations and the formation of a supershift band is evident by EMSA and is concentration-dependent. Protein concentrations are listed both on top (fixed concentration) and at the bottom (gradient concentration) of the gel. DnaA were pre-incubated with 5mM ATP and reaction was carried out at 2.5mM ATP. Probes for specific interaction (WT Rv0010c-Rv0011c IGR) labeled with IRD800 dye and non-specific interaction (sigA, not shown) labeled with IRD700 dye were added at 0.1nM. Cold competitor Poly (dI-dC) was added at final concentration of 100 ng/µl. Proteins and probes were incubated at room temperature for 20min prior to loading into 6% DNA retardation gel. Gel images shown here was captured with LI-COR imagining system and analyzed in Empiria Studio. The protein bound bands pre and post supershift are indicated by the lower and upper black arrow respectively.

What we observed was the presence of a supershift band, likely corresponding to the upshift of the band due to both proteins bound. We also observed based on the concentration where the

supershift band appeared, individual protein affinity was different in the presence of the other protein. In the presence of DnaA, ResR can bind *Rv0010c-Rv0011c* IGR at approximately one-third of the concentration required for ResR to bind alone. In the presence of low concentration of ResR, DnaA was also able to shift the band at slightly lower concentration (Figure 3-2). The increase in DnaA affinity aided by ResR was much more modest than the increase in ResR affinity in the presence of DnaA. This suggested DnaA could be the dominant protein in this complex. In single protein EMSA, we observed that clinical single nucleotide polymorphism in IGR increased DnaA affinity while decreased ResR affinity. Despite the change in ResR affinity, the complex likely has a higher affinity in the presence of IGR SNP due to its effect on DnaA binding. This was consistent with the fact that IGR SNP strain led to opposite phenotypes as DnaA or ResR binding site deletion strains. Therefore, we concluded that clinical SNPs in the *Rv0010c-Rv0011c* IGR region increase affinity of protein complex formed by DnaA and ResR, while loss of binding sites of either protein results in loss of complex formation. These findings provided the first evidence in *Mtb* that DnaA functions with a partner protein ResR and cooperative binding of these two proteins at this specific genomic site can have functional consequences.

Functional characterization of *Rv0010c-Rv0011c* IGR variants *in vivo*

Through biochemical approaches, we revealed that clinically relevant SNP in the *Rv0010c-Rv0011c* IGR increased binding of DnaA and ResR complex while loss of binding site of either protein decreased affinity of this complex. In previous chapter, genetic characterization of the same variants showed that clinical SNPs in IGR increased cell length and decreased drug sensitivity, while deletion of binding site resulted in opposite cell length and drug sensitivity

phenotypes. Knowing both DnaA and ResR can act as transcriptional regulators, we hypothesized that the difference in binding affinity at *Rv0010c-Rv0011c* IGR led to downstream gene expression changes and contributed to the distinct phenotypic profiles in these strains.

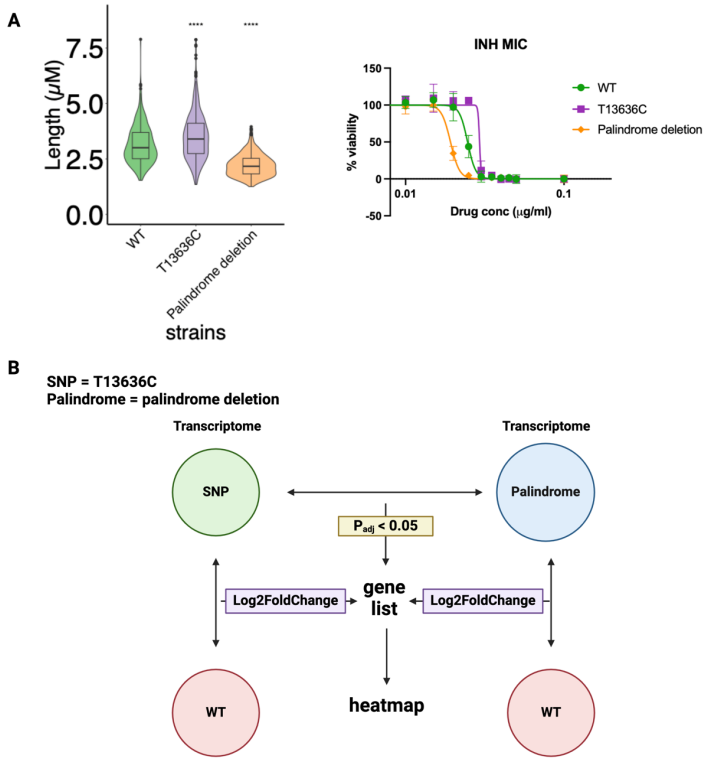


Figure 3-3 | Transcriptomic profiling to reveal expression changes that could drive phenotype bifurcation in complementation strains.

A) Phenotypes of strains used in RNA-sequencing. T13636C (SNP) and palindrome deletion (Palindrome) show opposite morphology and INH phenotype. Morphology measurements were done using quantitative microscopy on cells grown to mid log in plain media and were analyzed through the MOMIA pipeline. INH susceptibility was determined using the alamar blue assay. B) Analysis scheme for identifying genes with opposite expression changes in SNP and ΔPalindrome strain when compared to WT. We took significantly differentially expressed genes (adjusted p-value < 0.05) in SNP vs Palindrome strain and asked how expression of each gene changes in SNP vs WT and in Palindrome vs WT. Our final gene pathway analysis focused on only genes that changed in opposite direction in SNP vs WT compared to Palindrome vs WT. This includes genes upregulated in SNP vs WT and downregulated in Palindrome vs WT, or *vice versa*. Log_2 fold change of each gene in SNP or Palindrome compared to WT was used to make the final heatmaps.

To test this, we performed population transcriptomic profiling using RNA sequencing (RNA-seq) on three complementation strains: wild type (WT), clinical SNP (SNP), and palindrome deletion (ΔPalindrome) strains (Figure 3-3).

To identify gene expression changes of interest, we first selected genes that were significantly differentially expressed in SNP strain compared to ΔPalindrome strain. Then we selected a list of genes who were differentially expressed in opposite direction in SNP and ΔPalindrome strain when compared to WT strain (Figure 3-3). By these criteria, we were able to identify two sets of genes: 1) genes that were downregulated in SNP strain and upregulated in ΔPalindrome strain, and 2) genes that were upregulated in SNP strain and downregulated in ΔPalindrome strain. Analysis of the first set of genes identified an interesting cluster of genes. Several cell division related genes, including *sepF*, *ripA*, *ripB*, *ftsQ*, *ftsW*, *murG*, and *wag31*, were found to be downregulated in SNP strain and upregulated in ΔPalindrome strain. (Figure 3-4) Most of these genes encode proteins participating in main stages of cell division in mycobacteria. Conserved across bacteria, mycobacterial cell division initiation is marked by polymerization of FtsZ in mid cell (26-28). In mycobacteria, SepF and FtsQ are among the first proteins recruited to the divisome following Z-ring formation and are required for proper divisome assembly and recruitment of other proteins participating in division (26, 29, 30). FtsW is the mycobacterial divisome-specific homolog of conserved SEDS protein RodA participating in septal PG synthesis and recruitment of other PG synthase to the septum (31-33). RipA-B are two PG hydrolases and are shown to be required for septal PG hydrolysis and daughter cell separation (34). Together, these proteins encoded by the list of downregulated division genes covered almost every step of cell division, indicating an overall reduction in cell division.

Interestingly, one gene following the same trend as the set of division genes was *whiB2*, encoding the cell division transcriptional regulator WhiB2 (Figure 3-4).

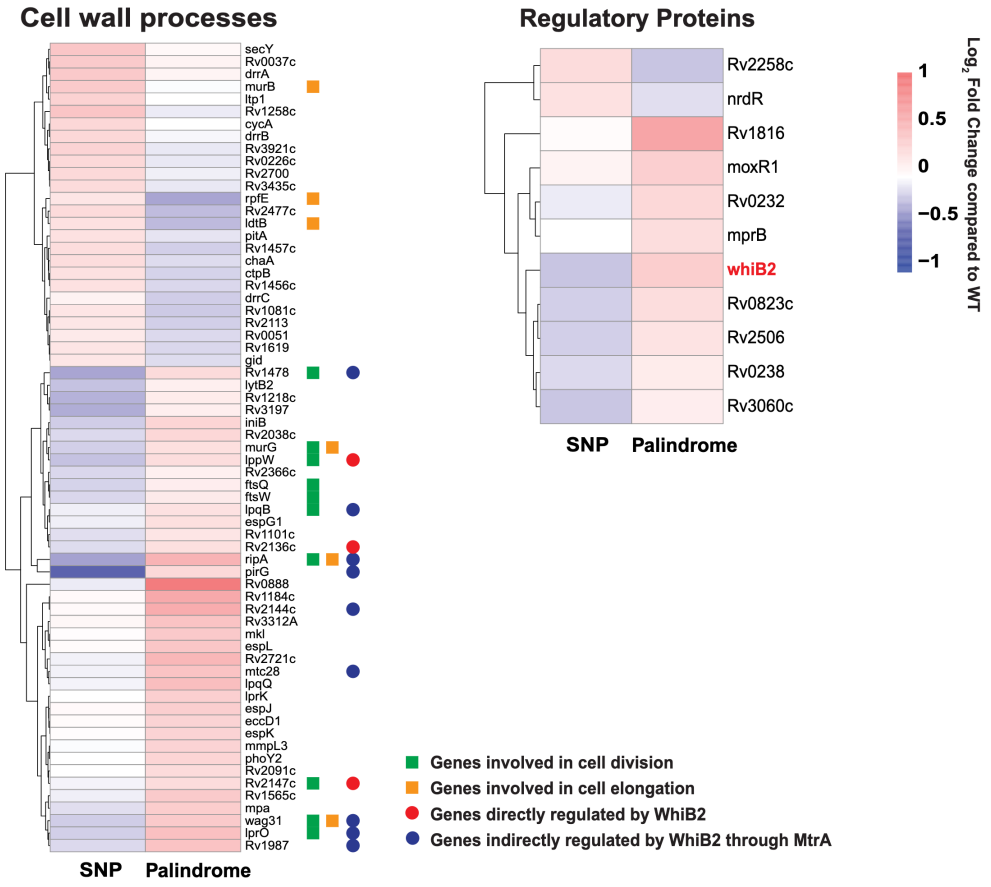


Figure 3-4 | Cell wall related genes are differentially expressed in SNP vs ΔPalindrome strain and most are under transcriptional control by WhiB2.

A) Differentially expressed genes in the RNAseq was selected as described in figure 3.3. We assigned broad pathway categorization to all selected genes and identified many pathways of interest, including cell wall related processes and regulatory pathways shown here in heatmaps with $\log_2$ fold change of each gene in SNP or Palindrome compared to WT. *whiB2* is highlighted in red, and green and orange boxes indicate genes linked to cell division and cell elongation respectively. Red and blue circle marks genes directly regulated by WhiB2 and MtrA, a downstream mediator regulated by WhiB2. Binding data was adapted from ChIP-seq analysis from Minch *et al.* 2015. Table of differentially expressed genes in opposite directions can be found in supplementary table 3.1 and heatmaps of other pathways can be found in supplementary figure 3.3.

Most of the genes mentioned above were direct regulatory targets of WhiB2 based on direct protein binding of WhiB2 itself and its downstream effector MtrA, another transcriptional

regulator (35). This network of genes controlled by WhiB2 and *whiB2* itself was upregulated in ΔPalindrome strain, indicating upregulation of cell division events. This data suggested altered division dynamics as a potential driver of phenotypic differences in these strains. Cell division timing has been tightly linked to morphological control in bacteria, while its effect on drug response remained unknown (36).

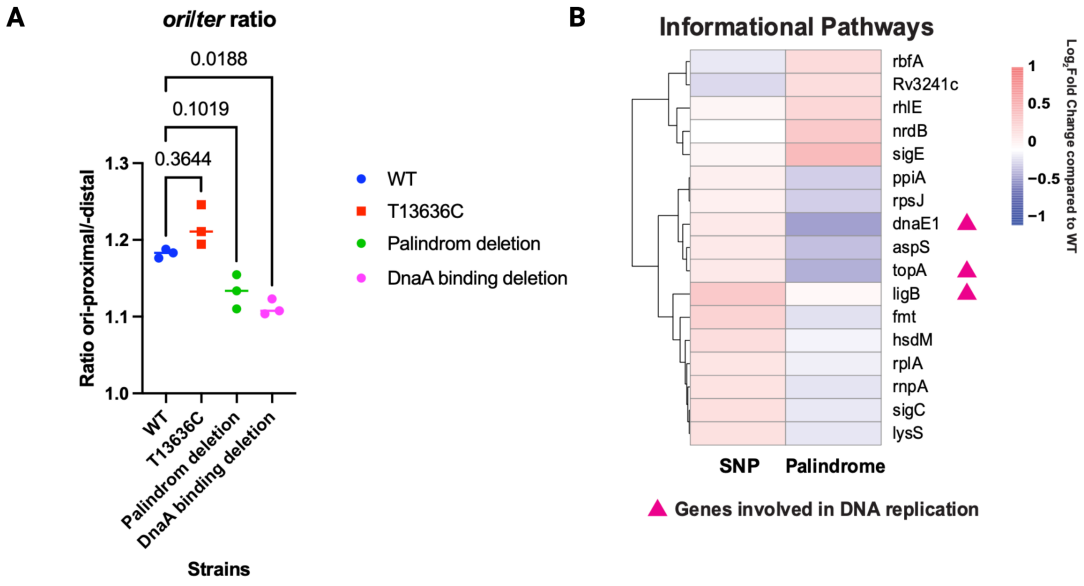


Figure 3-5 | *ori/ter* ratio reveals potential differences in replication dynamics, which is evident in transcriptional profiling of SNP and ΔPalindrome strain.

A) *ori/ter* ratio of complementation strains measured and quantified by whole genome sequencing. Cells were grown to mid-log and genomic DNA was harvested. After sequencing, reads were mapped and average coverage was calculated across the genome. Ratio of *ori*-proximal and *ori*-distal region is used as proxy for *ori/ter* ratio. P-value by unpaired t-test shown. B) Heatmap of genes in information pathways with opposite expression changes in SNP and palindrome deletion (Palindrome) strain. Genes of interested from the RNAseq were selected as described in figure 3.3 and 3.4. Replication related genes are highlighted by magenta triangles.

Interestingly, in the other set of genes (upregulated in SNP and downregulated in ΔPalindrome strain), we identified expression changes in other steps of cell cycle: replication and cell elongation (Figure 3-4& Figure 3-5). We observed increase in expression of cell elongation related genes such as *murB*, *ldtB*, and *rpfE* involved in PG synthesis, crosslinking, and recycling

(Figure 3-4) (37). This was consistent with the increased in maximal polar growth rate phenotype we observed in isogenic IGR variants and pulse-chase labeling of these complementation strains showed similar increase in polar elongation rate at the old pole (mature pole) and almost no change in polar elongation rate at the new pole (nascent pole) (Supplementary Figure 3.4). We also found increased expression of replication related genes such as *dnaE1*, *topA*, and *ligB*, encoding DNA polymerase III DnaE1, topoisomerase I TopA, and DNA ligase LigB respectively (38-40) (Figure 3-5). This suggested higher replication activity in SNP strains, and we tested this directly by quantifying *ori*/*ter* ratio using whole genome sequencing. Consistently, we observed a slight increase in *ori*-proximal contents than *ori*-distal, which potentially indicated altered replication dynamics (Figure 3-5). This change in replication dynamics was also accompanied by an increase in sensitivity to ofloxacin, a fluoroquinone targeting DNA gyrase used in TB treatment, observed in both complementation and isogenic strains with *Rv0010c-Rv0011c* IGR SNP (Supplementary figure 3.5) (41). Gyrases and topoisomerases are required to unwind and rewind DNA as replication forks travel, and sensitivity to inhibitors of these protein could potentially be another proxy for more active replication (42).

Our transcriptional analysis discovered differential regulation of *whiB2* and its regulon as potential contributor to the morphology and antibiotics phenotype seen in SNP and ΔPalindrome strain. The promoter of *whiB2* was a direct target of ResR, and this promoter region (*whiB2-fbiA*) was also one of the most highly mutated non-coding regions on the *Mtb* genome and has been strongly associated with *Mtb* drug resistance (8, 10, 43). These mutations were thought to alter ResR binding, although the direction of change is unknown (10). Isogenic variations in *whiB2-fbiA* showed similar morphology and INH susceptibility as *dnaA*, *resR*, and *Rv0010c-Rv0011c* IGR variants, adding another downstream factor in this essential genetic pathway under selection

in clinical *Mtb* populations. Knowing the binding affinity differences in SNP and ΔPalindrome strains, we propose a model where in the presence of single nucleotide variations in *Rv0010c-Rv0011c* IGR, ResR is sequestered at this IGR by DnaA and unable activate *whiB2* and downstream regulon expression as ResR activates *whiB2* expression (Figure 3-6) (10).

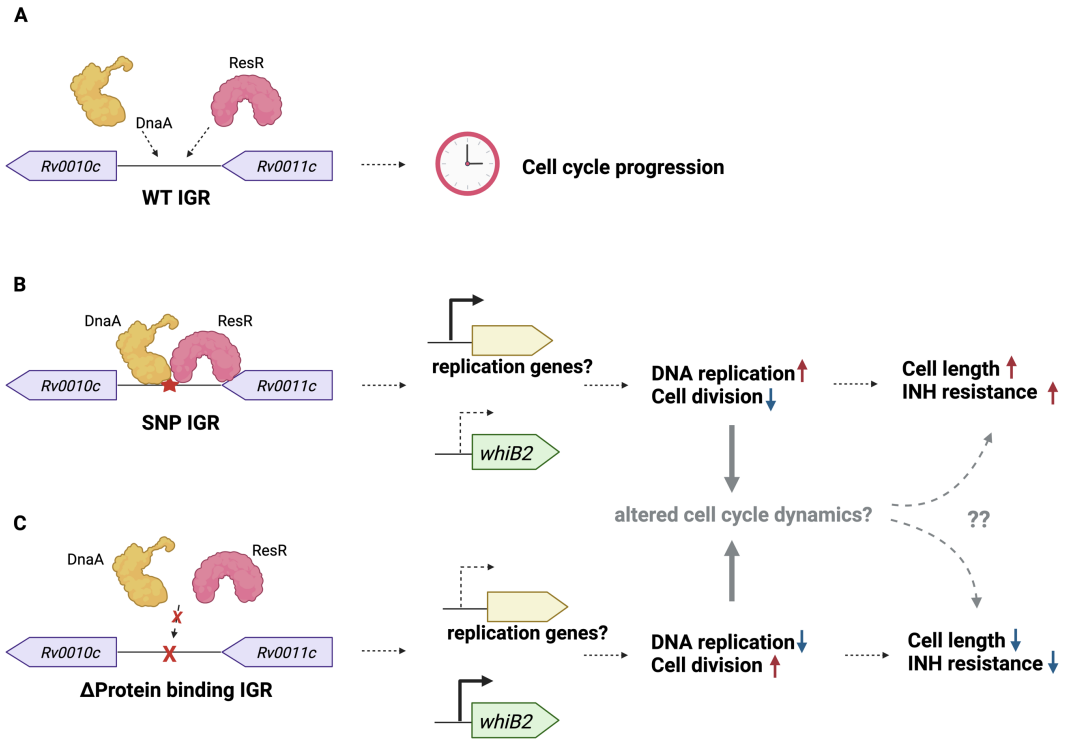


Figure 3-6| Proposed model of genetic pathway through *Rv0010c-Rv0011c* IGR.

We propose a model where DnaA and ResR bind at *Rv0010c-Rv0011c* IGR to regulate temporal events leading to cell cycle progression. There is likely a balance of these two proteins bound at this site at a given time and changes to this balance at different cell cycle stages regulates progression of cell cycle events. Clinical SNP with increased affinity results in sequestration of DnaA and ResR at this site, which downregulate *whiB2* and upregulated replication related genes. The binding site deletion strains allow freer DnaA and ResR to function downstream and results in upregulation of *whiB2* and downregulate replication related genes. These expression changes potentially alter cell cycle dynamics along with observed drug and morphology phenotypes. These effects are thought to be cell cycle dependent likely to avoid dysregulation of growth.

Increased of DnaA affinity at this site the presence of SNP resulted in changes in expression of other cell cycle related genes through an unknown mechanism. Together these transcriptional

changes contributed to altered cell cycle dynamics and the terminal drug and morphology phenotypes. In the presence of binding deletions, ResR is free to bind at *whiB2* promoter and activate division regulon, while replication genes are more repressed. This results in opposite cell cycle dynamics changes to SNP strains and opposite phenotypes. In our model, the interaction of DnaA and ResR at this site is cell cycle dependent to relay information between and coordinate DNA replication and division. There is likely a balance of these two proteins bound at this site that regulates progression of cell cycle events and could change dependent on cell cycle stage variation. Variations in *Rv0010c-Rv0011c* IGR alter replication and division dynamics subtly as both processes are extremely essential and requires extreme fine tuning to ensure progression of cell cycle events and population growth.

In this chapter, we presented data showing biochemically *Rv0010c-Rv0011c* IGR mutations directly altered binding dynamics of DnaA and ResR and functionally these variations led to differential expression of *whiB2*, a known target of ResR regulation, and its regulon of division related genes. Clinically relevant mutations in *whiB2* promoter region phenocopied *dnaA*, *resR*, and *Rv0010c-Rv0011c* clinical variants, implicating a downstream role of this genetic pathway in *Mtb*. Our data provided functional characterization of a previously unstudied genomic region under clinical selection, *Rv0010c-Rv0011c* IGR, and provided the first evidence of this essential genetic pathway coordinating cell cycle events as important contributor to evasion from antibiotic clearance in clinical *Mtb* populations.

3.4 Discussion

Using biochemical and transcriptomic approaches, we learned clinical variants in *Rv0010c-Rv0011c* IGR increased binding affinity of DnaA and ResR, two essential cell cycle

regulators. Altered binding relationships of DnaA and ResR to this intergenic region affected expression of downstream cell cycle related genes. We proposed a model where *Rv0010c-Rv0011c* IGR served a regulatory role in coordination of replication and division and altered binding dynamics of DnaA and ResR resulted in changes in cell cycle dynamics. Altered growth and division ultimately contributed to differences in cell morphology and drug response in clinical variants of *Mtb*. Our study provided the first genetic evidence for regulation of division by other parts of cell cycle and its direct impact on cell size control in mycobacteria. In our model, we attributed the downstream effect partially to the function of WhiB2. WhiB2 is a well conserved cell cycle regulator and is bound by ResR at its promoter where clinical mutations are accumulating (10, 44). We observed *whiB2* promoter mutants to share similar phenotypes in INH resistance, antibiotic resilience, and morphology. Our data was consistent with our previous study that showed ResR acts as a positive regulator of *whiB2* transcription (10). However, we do not know the directionality of ResR affinity changes in the presence of this promoter mutation and it remains to be tested. In streptomycetes, WhiB interact with WhiA and these two proteins bind its regulatory targets in the presence of each other (45, 46). Mycobacterial WhiA is an essential regulator, but its function remains largely unknown. *whiA* was one of the highly selected genes in our genomics screen along with *dnaA* and *resR*, and isogenic *whiA* mutants shared similar increase in drug and morphology phenotypes. We suspect WhiA plays a role in our cell cycle regulatory pathway and are actively investigating the connection between our pathway and WhiA.

In mycobacteria, cell cycle coordination remains a mystery. In slow growing mycobacteria like *Mtb*, chromosome replicate once per cell cycle, and division follows soon after replication is complete (47). The duration of chromosomal replication affects other parts of cell

cycle and such dynamics can be altered by deletion of *lsr2*, encoding a nucleoid associated protein Lsr2 that interacts with the replisome (48). Interestingly, Lsr2 is the functional analog of a DnaA interactor protein Rok in *B. subtills* and it has binding site at *Rv0010c-Rv0011c* IGR adjacent to DnaA and ResR (35, 49). Another nucleoid associated protein that also binds *Rv0010c-Rv0011c* IGR and has previously characterized link with DnaA is NapM. NapM is thought to sequester DnaA and induce growth arrest after sending environmental stress signals during *in vivo* infection (50). Nucleoid associated protein play an important role in regulating chromosomal dynamics during active replication and transcription and in altering genetic expressions following chromosomal organization (49, 51). It is evident that in our EMSA and in our complementation strains from the previous chapter, DnaA and ResR play a dominant role at this IGR but are not the only protein interactors in this region. Multiple other transcription regulator and nucleoid associated proteins bind in neighboring regions and the role of the protein binding landscape at *Rv0010c-Rv0011c* IGR in our model remains to be explored (35).

Replication and division are connected to cell size control in a recent modeling of mycobacterial cell cycle. Model bacteria such as *E. coli* and *B. subtilis* follows the adder model where cells add constant increments between every round of division (52). Mycobacteria adopt a unique parallel adder model and cell size is determined by two sets of constant growth increments, one between two consecutive rounds of replication initiation and the other between replication initiation and division (47). In the parallel adder model, replication initiation plays a central role in determining the increments added within the two parallel sets of cell cycle events. Our data on altered replication and division dynamics could play directly into how these two increments are set in *Mtb* and likely a contributor to the upshift in cell size we observed in our genetic variants. In all bacteria, cell size control is tightly regulated and is crucial in maintaining

optimal physiology in any given growth condition. Such regulation on morphology is even more important for mycobacteria in adaptation to different environment it encounters during infection (53, 54). Within our lab collection of clinical strains, we observed an increase in cell size and overall shift in distribution of size towards longer cells when grown in ambient lab condition (Supplementary Figure 3.6). Such increase in cell size was also evident in clinically drug resistant *Mtb* cultured in host relevant conditions (55). It remains largely a mystery how cell morphology can contribute to *Mtb* pathogenesis and would be an interesting future area to explore and learn from in *Mtb* physiology.

Under the parallel adder model, our data suggests increase in overall increments between initiation and division with the changes in replication and division dynamics. This shift in cell length is not a result of changes in growth rate in both liquid and in colony formation, which implies variant strains need to increase their elongation rate to lay down more cell wall material within the same inter cell cycle period as WT. This is consistent with the increase in polar growth rate we have observed both in isogenic strains and complementation strains. Mycobacteria has a unique multi-layer membrane structure that contains a peptidoglycan layer (PG), an arabinogalactan layer (AG), and a mycolic acid layer (MA) (56). Increase in polar growth requires the coordination of synthesis in all three layers. We have observed in our RNA-seq that increase in expression of elongation genes involved synthesis of components in these layers such as *ldtB*, *rpfE*, *ripA*, and *murB* for PG synthesis, *drrA/B*, *mas* for surface lipids related to the AG and MA layer, and *fbpA*, *mmaA1*, and *inhA* for MA synthesis (56). InhA is a direct target of active form of INH and increased *inhA* expression is correlated with low level INH resistance (5, 57, 58). This could explain the increased INH resistance in complementation strains carrying SNP in *Rv0010c-Rv0011c* IGR as *inhA* expression increased in SNP strain and decreased in

palindrome deletion strains. Here, we propose that change in *inhA* could be a result of feedback regulation following changes in polar growth rate, but we don't rule out the possibility of an unknown molecular mechanisms that links *Rv0010c-Rv0011c* IGR with *inhA* regulation.

In conclusion, together with chapter 2, we identified a novel and essential genetic pathway under antibiotic in *Mtb*. This pathway directly regulated cell cycle dynamics, and variations within it could contribute to regrowth post antibiotic exposure. These findings provided a genetic foundation for connecting *Mtb* physiology with drug response and an important bacterial factor to be considered in future designs of TB chemotherapy.

3.5 Material and Methods

Bacterial strains and growth conditions

Mtb strains used in this study are constructed in the background of wild type lab strain H37Rv and *Msm* strains used are constructed in the background of wild type mc^2155. *Mtb* strains were grown in liquid in Middlebrook 7H9 media (BD) supplemented with 10% oleic acid albumin-dextrose-catalase (OADC) (BD), 0.2% glycerol, and 0.05% Tween-80 (referred to as 7H9 in following sections) at 37 ^{o}C. Solid media for *Mtb* were made with Middlebrook 7H10 Agar (BD) supplemented with 10% oleic acid albumin-dextrose-catalase (OADC) (BD) and 0.2% glycerol (referred to as 7H10 in following sections). *Msm* strains were grown in Middlebrook 7H9 media (BD) supplemented with 10% albumin-dextrose-catalase (ADC containing a final concentration of 5g/L bovine serum albumin, 2g/L dextrose, and 3mg/L catalase), 0.2% glycerol, and 0.05% Tween-80 at 37 ^{o}C. Solid media for *Msm* were made with Middlebrook 7H10 Agar (BD) supplemented with 10% albumin-dextrose-catalase (ADC) (BD) and 0.2% glycerol. Antibiotics were supplemented to media at the following final concentration

when needed: Kan (20µg/mL), Hyg (50µg/mL), Nat (25µg/mL), Zeo (20µg/mL). Induction by anhydrotetracycline (aTC) were done at final concentration of 100ng/ml. Bacterial growth were measured by optical density at 600nm wavelength (OD_{600}).

Minimum inhibitory concentration measurement

Minimum inhibitory concentration (MIC) for anti-tuberculosis drugs were measured by alamar blue reduction assay. *Mtb* strains were started from frozen stocks and grown to mid-late log phase at OD_{600} = 0.5-0.8 before back diluting to OD = 0.05 in 10ml 7H9 media supplemented with appropriate antibiotics. Cells were allowed to grow for another two days to OD = 0.3. Serial or specific dilutions of antibiotics were prepared in 96 well plates and *Mtb* cells were added to a final OD = 0.00015 in 200µl total volume per well in technical duplicates. Plates were incubated in an airtight container for 7-8 days total at 37 ^{o}C. On day 6, 0.0002% final concentration of alamarBlue (BioRad) reagent were added to each well and allowed to incubate for 24-48 hours before reading on a plate reader. Absorbance profile was measured at OD = 600nm and OD = 570nm and subtraction of OD_{600} from OD_{570} was used to calculate percentage of viability in each well compared to no-drug control. Final growth inhibition profiles were plotted, and minimum inhibitory concentration (IC_{50}) was determined after performing linear regression on viability data at each antibiotic concentration in Prism Graphpad.

Pulse-chase labeling

Pulse-chase labeling was performed using NADA as pulse and HADA (Tocris Bioscience) as chase in *Mtb*. Strains were started from frozen stocks and grown to mid-late log phase at OD_{600} = 0.5-0.8 before back diluting to OD = 0.05 in replicates in 10ml 7H9 media + 10nM NADA supplemented with appropriate antibiotics. Cells were grown for two days to OD_{600} ~ 0.3 and washed two times with equal volume of plain 7H9. Cells were resuspended in

equal volume of 7H9 + 25nM HADA with appropriate antibiotics. After 24 hours, cells were harvested by mixing equal volume with 4% PFA and incubated for at least 1 hour at room temperature. PFA-fixed cells are stored at 4ºC prior to microscopy.

Quantitative microscopy

As mycobacteria cells are prone to aggregates, PFA-fixed cells were first processed to remove aggregates to improve imaging quality. 100µl of PFA-fixed cells were centrifuged at 3200g for 15min and then resuspend in 200µl disaggregation buffer (200mM Tric-HCl pH 7.5, 1% (w/v) Triton X-100, 0.67% (w/v) Xylenes and 0.33% (w/v) Heptane). Cells were immediately centrifuged at the same setting for 15min. Then cells were washed one time in 200µl PBS + 0.1% Triton X-100 before resuspending in 20-25µl PBS + 0.1% Triton X-100 to ideal turbidity. 1.5µl of resuspended cells were spotted onto a custom-made 96 well pedastral array made from 2% agarose in 1X + PBS supplemented with optional Nile Red and DAPI stain. If using DAPI, the assembled agarose cartridge was incubated at room temperature for at least 30min for DAPI to fully penetrate cells. Imaging was performed on a Nikon Ti-E inverted, widefield microscopy equipped with a Plan Apo 100x 1.45 NA objective lens and an Andor Zyla sCMOS camera. A custom Nikon JOBS script on the NIS Elements software was designed to perform automated imaging in 96 well format and captured 2 images each at the pre-designed 12 fields of interest for each sample. Fluorescent signal was captured using a 6-channel Spectra X LED light source and the Sedat Quad filter set. Excitation (Ex) and emission (Em) filters were set at Ex 395/25nm and Em 435/25nm for HADA/DAPI, Ex 470/24nm and Em 515/25nm for NADA (FITC), and Ex 550/15nm and Em 595/25nm for Nile Red. Exposure time was set at 100ms for all channels. Post imaging morphological profiling was performed using a previous

described microscopy analysis pipeline MOMIA (59). Single cells were filtered and traced along the cellular midline and their fluorescent profile was used to calculate polar outgrowth.

Transcriptional profiling

For RNA extraction, *Mtb* strains were started from frozen stocks and grown to mid-late log phase at OD_{600} = 0.5-0.8 before back diluting to OD = 0.05 in triplicates in 10ml 7H9 media supplemented with appropriate antibiotics. After 3 days, cells are grown to mid log OD = 0.5-0.6 and centrifuged at room temperature to collect pellet. Cell pellets were resuspended in 1ml TRIzol Reagent (Thermo Fisher) and lysed by bead beating to release nucleic acid that was further extraction by adding 300µL chloroform. Total RNA was then isolated using the Direct-zol RNA Miniprep Kit (Zymo Research). DNA digestion was done using TURBO DNase kit (Thermo Fisher) and post depletion total RNA was re-purified using RNA Clean and Concentrator kit (Zymo Research).

For real time qRT-PCR, cDNA was generated from total RNA using SuperScript IV First Strand cDNA synthesis kit with random hexamers (Thermo Fisher). qPCR reactions were prepared with iTaq Universal SYBR-Green master mix (BioRad) and were run on a ViiA7 real time PCR system (Thermo Fisher). Raw CT values were analyzed with the ΔΔCT methods and all experimental gene expressions were normalized to *sigA* expression. For RNA-seq, RNA library was prepped using the KAPA RNA HyperPrep kit (KAPA Biosystems) according to manufacturer's instructions. rRNA was depleted using the KAPA RiboErase kit (KAPA Biosystems) with custom *Mtb* specific rRNA targeting oligos. RNA libraries were sequenced on an Illumina MiSeq sequencer with a 100bp single-ended reads. FASTQ files generated form sequencing had adapters removed using cutadapt and were mapped to the *Mtb* genome (NC_000962) using bowtie2 using the –very-sensitive argument set (60, 61). Using a custom

script, counts were applied to genes if any part of the fragment defined by the paired end read overlapped with the gene. Analysis of transcriptomic expression changes based on gene counts was done using a R package, DEseq2 (62). For each variant strain, $\log_2$ fold change and adjusted p-value following multiple hypothesis correction was generated for each gene compared to its expression in WT strain.

gDNA sequencing and *ori/ter* ratio

gDNA was extracted in biological triplicate cultures of strains grown to mid-log. Cells were lysed by bead beating and nucleic acid was extracted with phenol-chloroform. DNA library prep and Illumina whole genome sequencing were performed by SeqCenter in Pittsburgh, PA. After sequencing, Illumina sequencing reads were quality filtered and adapters trimmed with fastp (63). Sequencing reads were aligned to the H37Rv (NC_000962.3) reference genome with bwa and duplicates marked with picard MarkDuplicates (64, 65). Depth of coverage was calculated for each base using the GATK DepthOfCoverage tool with a coverage threshold of 10 (65). Coverage was averaged into 1000bp non overlapping windows across the entire genome with a custom perl script. Coverage across the genome was then plotted using Prism Graphpad and owess smoothing was performed. LOWESS smoothed coverage values were used to calculated ori/ter ratio as peak values were close to *ori*-proximal region and the lowest values were near *ori*-distal region halfway through the genome. Unpaired t-test was performed in Prism on biological triplicates of each variant compared to WT.

Electrophoretic mobility shift assay (EMSA)

EMSA probes were generated by PCR and were labeled with either IRD800 or IRD700 dye for imaging on the LICOR Odyssey system. *Rv0010c-Rv0011c* IGR WT and variants were ~500 bp long labeled with IRD800 and non-specific control probe from sigA region was ~700bp

labeled with IRD700. Prior to setting up EMSA reactions, DnaA was incubated with 5mM ATP for 2 hours at 4ºC. While still at 4ºC, Proteins and probes were mixed with 15µl EMSA binding buffer (40mM HEPES KOH pH 7.6, 10mM potassium chloride, 140mM potassium glutamate, 10mM magnesium acetate, 2.5mM ATP, 0.5mM EDTA, 1mM DTT, 50ug/ml bovine serum albumin fraction V, 20% glycerol) and 100ng/µl Poly dI-dC (Thermo Fisher) to desired concentration and ~0.1 nM final DNA probe concentration to a final volume of 21µl. Reactions were then incubated for 20 minutes at 23ºC and 15ul was loaded onto a 6% DNA retardation gel (Thermo Fisher). Gels were run with 0.5X TBE buffer at 80 volts for 1.5 hours in the dark at room temperature. The resulting complexes were visualized by imaging on a LICOR Odyssey imager to simultaneously detect both *Rv0010c-Rv0011c IGR* and *sigA* on the 800 and 700 channels respectively.

3.6 References

1. World Health Organization. Global tuberculosis report 2023. Geneva: World Health Organization; 2023 2023.

2. Saleem A, Azher M. The next Pandemic-Tuberculosis: The oldest disease of mankind rising one more time. British journal of medical practitioners. 2013;6(2):21-46.

3. Stanley S, Liu Q, Fortune SM. Mycobacterium tuberculosis functional genetic diversity, altered drug sensitivity, and precision medicine. Frontiers in Cellular and Infection Microbiology. 2022;12:1007958.

4. Zhu J, Liu YJ, Fortune SM. Spatiotemporal perspectives on tuberculosis chemotherapy. Curr Opin Microbiol. 2023;72:102266.

5. Ramaswamy S, Musser JM. Molecular genetic basis of antimicrobial agent resistance in Mycobacterium tuberculosis: 1998 update. Tubercle and Lung disease. 1998;79(1):3-29.

6. Gillespie SH. Evolution of drug resistance in Mycobacterium tuberculosis: clinical and molecular perspective. Antimicrobial agents and chemotherapy. 2002;46(2):267-74.

7. Hicks ND, Giffen SR, Culviner PH, Chao MC, Dulberger CL, Liu Q, et al. Mutations in dnaA and a cryptic interaction site increase drug resistance in Mycobacterium tuberculosis. PLOS Pathogens. 2020;16(11):e1009063.

8. Hicks ND, Yang J, Zhang X, Zhao B, Grad YH, Liu L, et al. Clinically prevalent mutations in Mycobacterium tuberculosis alter propionate metabolism and mediate multidrug tolerance. Nat Microbiol. 2018;3(9):1032-42.

9. Torrey HL, Keren I, Via LE, Lee JS, Lewis K. High persister mutants in Mycobacterium tuberculosis. PLoS One. 2016;11(5):e0155127.

10. Liu Q, Zhu J, Dulberger CL, Stanley S, Wilson S, Chung ES, et al. Tuberculosis treatment failure associated with evolution of antibiotic resilience. Science. 2022;378(6624):1111-8.

11. Colangeli R, Jedrey H, Kim S, Connell R, Ma S, Chippada Venkata UD, et al. Bacterial Factors That Predict Relapse after Tuberculosis Therapy. N Engl J Med. 2018;379(9):823-33.

12. Zhou W, Huang S, Cumming BM, Zhang Y, Tang W, Steyn AJ, et al. A feedback regulatory loop containing McdR and WhiB2 controls cell division and DNA repair in mycobacteria. mbio. 2022;13(2):e03343-21.

13. World Health Organization. Guidelines for treatment of drug-susceptible tuberculosis and patient care, 2017 update. 2017.

14. Plocinski P, Ziolkiewicz M, Kiran M, Vadrevu SI, Nguyen HB, Hugonnet J, et al. Characterization of CrgA, a New Partner of the Mycobacterium tuberculosis Peptidoglycan Polymerization Complexes. Journal of Bacteriology. 2011;193(13):3246-56.

15. Plocinski P, Arora N, Sarva K, Blaszczyk E, Qin H, Das N, et al. Mycobacterium tuberculosis CwsA interacts with CrgA and Wag31, and the CrgA-CwsA complex is involved in peptidoglycan synthesis and cell shape determination. Journal of bacteriology. 2012;194(23):6398-409.

16. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9.

17. Van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. Nature Biotechnology. 2024;42(2):243-6.

18. Sassetti CM, Boyd DH, Rubin EJ. Comprehensive identification of conditionally essential genes in mycobacteria. Proc Natl Acad Sci U S A. 2001;98(22):12712-7.

19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.

20. Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. Nature protocols. 2007;2(8):1849-61.

21. Katayama T, Kasho K, Kawakami H. The DnaA Cycle in Escherichia coli: Activation, Function and Inactivation of the Initiator Protein. Frontiers in Microbiology. 2017;8.

22. Yamamoto K, Muniruzzaman S, Rajagopalan M, Madiraju MV. Modulation of Mycobacterium tuberculosis DnaA protein–adenine-nucleotide interactions by acidic phospholipids. Biochemical Journal. 2002;363(2):305-11.

23. Madiraju MVVS, Moomey M, Neuenschwander PF, Muniruzzaman S, Yamamoto K, Grimwade JE, et al. The intrinsic ATPase activity of Mycobacterium tuberculosis DnaA promotes rapid oligomerization of DnaA on oriC. Molecular Microbiology. 2006;59(6):1876-90.

24. Grimwade JE, Rozgaja TA, Gupta R, Dyson K, Rao P, Leonard AC. Origin recognition is the predominant role for DnaA-ATP in initiation of chromosome replication. Nucleic acids research. 2018;46(12):6140-51.

25. Brown NL, Stoyanov JV, Kidd SP, Hobman JL. The MerR family of transcriptional regulators. FEMS microbiology reviews. 2003;27(2-3):145-63.

26. Baranowski C, Rego EH, Rubin EJ. The dream of a mycobacterium. Microbiology spectrum. 2019;7(2):10.1128/microbiolspec. gpp3-0008-2018.

27. Adams DW, Errington J. Bacterial cell division: assembly, maintenance and disassembly of the Z ring. Nature Reviews Microbiology. 2009;7(9):642-53.

28. Bi E, Lutkenhaus J. FtsZ ring structure associated with division in Escherichia coli. Nature. 1991;354(6349):161-4.

29. Gola S, Munder T, Casonato S, Manganelli R, Vicente M. The essential role of SepF in mycobacterial division. Molecular microbiology. 2015;97(3):560-76.

30. Wu KJ, Zhang J, Baranowski C, Leung V, Rego EH, Morita YS, et al. Characterization of conserved and novel septal factors in Mycobacterium smegmatis. Journal of Bacteriology. 2018;200(6):10.1128/jb. 00649-17.

31. Meeske AJ, Riley EP, Robins WP, Uehara T, Mekalanos JJ, Kahne D, et al. SEDS proteins are a widespread family of bacterial cell wall polymerases. Nature. 2016;537(7622):634-8.

32. Datta P, Dasgupta A, Bhakta S, Basu J. Interaction between FtsZ and FtsW of Mycobacterium tuberculosis. Journal of Biological Chemistry. 2002;277(28):24983-7.

33. Datta P, Dasgupta A, Singh AK, Mukherjee P, Kundu M, Basu J. Interaction between FtsW and penicillin‐binding protein 3 (PBP3) directs PBP3 to mid‐cell, controls cell septation and mediates the formation of a trimeric complex involving FtsZ, FtsW and PBP3 in mycobacteria. Molecular microbiology. 2006;62(6):1655-73.

34. Böth D, Schneider G, Schnell R. Peptidoglycan remodeling in Mycobacterium tuberculosis: comparison of structures and catalytic activities of RipA and RipB. Journal of molecular biology. 2011;413(1):247-60.

35. Minch KJ, Rustad TR, Peterson EJR, Winkler J, Reiss DJ, Ma S, et al. The DNA-binding network of Mycobacterium tuberculosis. 2015;6:5829.

36. Westfall CS, Levin PA. Bacterial cell size: multifactorial and multifaceted. Annual review of microbiology. 2017;71:499-517.

37. Maitra A, Munshi T, Healy J, Martin LT, Vollmer W, Keep NH, et al. Cell wall peptidoglycan in Mycobacterium tuberculosis: An Achilles' heel for the TB-causing pathogen. FEMS microbiology reviews. 2019;43(5):548-75.

38. Gui W-J, Lin S-Q, Chen Y-Y, Zhang X-E, Bi L-J, Jiang T. Crystal structure of DNA polymerase III β sliding clamp from Mycobacterium tuberculosis. Biochemical and Biophysical Research Communications. 2011;405(2):272-7.

39. Yang F, Lu G, Rubin H. Cloning, expression, purification and characterization of DNA topoisomerase I of Mycobacterium tuberculosis. Gene. 1996;178(1-2):63-9.

40. Gong C, Martins A, Bongiorno P, Glickman M, Shuman S. Biochemical and genetic analysis of the four DNA ligases of mycobacteria. Journal of Biological Chemistry. 2004;279(20):20594-606.

41. Todd PA, Faulds D. Ofloxacin: a reappraisal of its antimicrobial activity, pharmacology and therapeutic use. Drugs. 1991;42:825-76.

42. Levine C, Hiasa H, Marians KJ. DNA gyrase and topoisomerase IV: biochemical activities, physiological roles during chromosome replication, and drug sensitivities. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression. 1998;1400(1-3):29-43.

43. Zhang H, Li D, Zhao L, Fleming J, Lin N, Wang T, et al. Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nature genetics. 2013;45(10):1255-60.

44. Bush MJ. The actinobacterial WhiB-like (Wbl) family of transcription factors. Molecular Microbiology. 2018;110(5):663-76.

45. Bush MJ, Chandra G, Bibb MJ, Findlay KC, Buttner MJ. Genome-wide chromatin immunoprecipitation sequencing analysis shows that WhiB is a transcription factor that

cocontrols its regulon with WhiA to initiate developmental cell division in Streptomyces. MBio. 2016;7(2):10.1128/mbio. 00523-16.

46. Lilic M, Holmes NA, Bush MJ, Marti AK, Widdick DA, Findlay KC, et al. Structural basis of dual activation of cell division by the actinobacterial transcription factors WhiA and WhiB. Proceedings of the National Academy of Sciences. 2023;120(11):e2220785120.

47. Logsdon MM, Ho PY, Papavinasasundaram K, Richardson K, Cokol M, Sassetti CM, et al. A Parallel Adder Coordinates Mycobacterial Cell-Cycle Progression and Cell-Size Homeostasis in the Context of Asymmetric Growth and Organization. Curr Biol. 2017;27(21):3367-74 e7.

48. Kołodziej M, Trojanowski D, Bury K, Hołówka J, Matysik W, Kąkolewska H, et al. Lsr2, a nucleoid-associated protein influencing mycobacterial cell cycle. Scientific Reports. 2021;11(1):2910.

49. Seid CA, Smith JL, Grossman AD. Genetic and biochemical interactions between the bacterial replication initiator DnaA and the nucleoid-associated protein Rok in Bacillus subtilis. Mol Microbiol. 2017;103(5):798-817.

50. Liu Y, Xie Z, Zhou X, Li W, Zhang H, He Z-G. NapM enhances the survival of Mycobacterium tuberculosis under stress and in macrophages. Communications Biology. 2019;2(1).

51. Dillon SC, Dorman CJ. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. Nature Reviews Microbiology. 2010;8(3):185-95.

52. Taheri-Araghi S, Bradde S, Sauls JT, Hill NS, Levin PA, Paulsson J, et al. Cell-size control and homeostasis in bacteria. Current biology. 2015;25(3):385-91.

53. Logsdon MM, Aldridge BB. Stable regulation of cell cycle events in mycobacteria: Insights from inherently heterogeneous bacterial populations. Frontiers in Microbiology. 2018;9:348921.

54. Willis L, Huang KC. Sizing up the bacterial cell cycle. Nat Rev Microbiol. 2017;15(10):606-20.

55. Vijay S, Vinh DN, Hai HT, Ha VTN, Dung VTM, Dinh TD, et al. Influence of Stress and Antibiotic Resistance on Cell-Length Distribution in Mycobacterium tuberculosis Clinical Isolates. Front Microbiol. 2017;8:2296.

56. Dulberger CL, Rubin EJ, Boutte CC. The mycobacterial cell envelope - a moving target. Nat Rev Microbiol. 2020;18(1):47-59.

57. Rozwarski DA, Grant GA, Barton DH, Jacobs Jr WR, Sacchettini JC. Modification of the NADH of the isoniazid target (InhA) from Mycobacterium tuberculosis. Science. 1998;279(5347):98-102.

58. Zhang M, Yue J, Yang Y-p, Zhang H-m, Lei J-q, Jin R-l, et al. Detection of mutations associated with isoniazid resistance in Mycobacterium tuberculosis isolates from China. Journal of clinical microbiology. 2005;43(11):5477-82.

59. Zhu J, Wolf ID, Dulberger CL, Won HI, Kester JC, Judd JA, et al. Spatiotemporal localization of proteins in mycobacteria. Cell reports. 2021;37(13).

60. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011;17(1):10-2.

61. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9(4):357-9.

62. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15:1-21.

63. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i90.

64. Li H. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;38:1767.

65. Van der Auwera GA, O'Connor BD. Genomics in the cloud: using Docker, GATK, and WDL in Terra: O'Reilly Media; 2020.

Chapter 4: Discussion

4.1 Genetic regulation of cell cycle dynamics

Tuberculosis is a notoriously difficult disease to treat in part because treatment is impacted by bacterial genetic factors of its causative agent, *Mycobacterium tuberculosis* (*Mtb*) [1, 2]. Thanks to advances in sequencing technologies, we now have the ability to identify evolving genetic loci that enable *Mtb* to evade host immune and antibiotics clearance *in vivo* [3-6]. In this dissertation, we presented a case study elucidating one novel *Mtb* genetic pathway under clinical selection that involves multiple essential proteins and intergenic regions, and that contributes to antibiotic resilience. We identified two coding genes, *dnaA* and *resR,* and their shared protein interaction site *Rv0010c-Rv0011c* intergenic region (IGR) that are enriched in mutations in clinical *Mtb* populations by using a large-scale genomics screen to conduct an unbiased search for genetic factors under selection. Although they were studied independently, *dnaA*, *resR*, and *Rv0010c-Rv0011c* IGR variants share similar drug phenotypes in the form of low-level resistance to first-line tuberculosis drug isoniazid (INH) and share morphology phenotypes with shared increase in cell length.

After identifying the convergence in phenotype, we proposed that *dnaA*, *resR*, and *Rv0010c-Rv0011c* IGR participate in a previously uncharacterized genetic pathway and that this pathway is under clinical selection. Genetic and phenotypic characterization of the *Rv0010c-Rv0011c* operon led us to discover that this IGR serves as a non-coding regulatory element in control of *Mtb* morphology and drug response, and that binding of DnaA and ResR is required for its function. Clinical mutations at *Rv0010c-Rv0011c* IGR directly alter the binding dynamics of these two proteins and result in expression changes of cell cycle related genes. Based on this discovery, we propose a model where clinical IGR mutations increase sequestration of ResR by

DnaA and alter dynamics between replication and division, leading to clinical phenotypes in the forms of antibiotic resistance and resilience.

Antibiotic resilience is characterized as faster bacterial regrowth after exposure to strong antibiotics, which means resilience is tightly linked to fundamental cell cycle processes such as growth and division [6]. DnaA, a highly conserved protein across bacteria, is required for replication initiation and marks the onset of a new round of cell cycle events. Given DnaA's essential role in one of the most fundamental processes, altering its activity and function can be fatal to bacteria [7]. *dnaA* is essential in most bacteria, including mycobacteria, for *in vitro* growth by transposon sequencing and it is highly vulnerable to CRISPR interference [8, 9]. Surprising, *dnaA* was found to accumulate many mutations in clinical strains of *Mtb* and these mutations are highly associated with drug resistance. *dnaA* mutations were enriched in its AAA+ ATPase domain and DNA binding domain, the two highly conserved functional domains, and yet these mutations did not significantly change growth or chromosomal replication rate. This GWAS was the first study linking DnaA to drug response in *Mtb* [10]. In a different and attenuated *Mtb* strain, it was shown that DnaA can be sequestered by a nucleoid associate protein to slow down growth under host stress *in vivo* [11]. These studies, taken together with our data, show that DnaA plays an unexpected role in *Mtb* drug survival, which could be unique to the environment in which *Mtb* thrives. However, the role of essential proteins remains understudied in most pathogenic bacteria due in part to the difficulty of genetic manipulation .

In addition to DnaA, the other key cell-cycle regulatory protein in antibiotic resilience is ResR, which is an essential protein regulating cell division. ResR was only characterized recently and was first identified as the essential transcriptional regulator of WhiB2, a redox-sensitive protein in the conserved WhiB family protein in actinobacteria [12, 13]. Clinical

mutations in *resR* are enriched in the predicted DNA binding domain of the encoded protein ResR and are thought to alter binding at the *whiB2* promoter and affect *whiB2* expression [6]. WhiB2, a transcriptional regulator itself, regulates expression of multiple genes participating in cell division [14]. Proteins whose expression is regulated by WhiB2 include SepF, an early divisome protein, and later stage divisome proteins such as FtsI, FtsQ, and Wag31. WhiB2 also regulates expression of cell wall hydrolase genes *ripA* and *ripB*, which are required for daughter cell separation at the last step of division. WhiB2 is thought to positively regulate its regulon genes as knockdown of *whiB2* results in downregulation of the regulon expression and severe morphological defect due to cells' inability to divide [12, 14, 15].

Our data led us to propose a model in which clinical mutations that alter binding dynamics of DnaA and ResR at *Rv0010c-Rv0011c* IGR lead to small changes in cell cycle dynamics between replication and division. In contrast to fast growers, slow growing mycobacteria like *Mtb* initiate replication only once per cell cycle and require coordination between replication initiation and division in order to avoid untimely initiation or division [16]. However, the mechanisms by which mycobacteria coordinate progression of cell cycle events remain unknown. Through biochemical and transcriptomic approaches, we found that clinical variants in the *Rv0010c-Rv0011c* IGR increase affinity of ResR through sequestration by DnaA and downregulate division related genes that are under regulatory control by WhiB2. We propose that a subtle increase in sequestration of ResR during cell cycle progression is sufficient to delay the onset of division and increase the gap between initiation and division. This delay has a direct impact on cell size control in mycobacteria as they adopt a unique growth model that relies on timing of initiation and division. Unlike *E. coli* and *B. subtilis*, mycobacteria grow at a constant increment per cycle between two sets of cell cycle events: 1) between two rounds of replication

initiation, and 2) between replication initiation and division [17]. Our data points to a delay in division and indicates this delay could extend the initiation to division interval, resulting in the longer cells that we observed in our clinical variants. Despite reaching a longer length every cell cycle, clinical variants did not affect growth rate or colony duplication rate and therefore the inter cell cycle time remained the same as WT. This implies variant cells needed to elongate faster to reach a longer final cell length, which is consistent with our pulse-chase labeling data. In our RNA-seq, we also see evidence of upregulation in genes involved in elongation and synthesis of the mycomembrane. *inhA* is one of the upregulated genes in mycolic acid synthesis and is a direct target of INH. *inhA* upregulation is associated with increased INH resistance, which is consistent with the INH susceptibility changes we observed [18-20]. However, we recognize that we have not rigorously tested cell cycle dynamics to confirm that they change as predicted by our model. Further experimental validation with a synchronizable *Mtb* strain is needed to confirm the proposed changes in cell cycle dynamics [21].

In the mycobacterial cell cycle, replication precedes division and division only initiates after chromosomal duplication is nearly complete. Towards the end of replication, the terminal region (*ter*) moves to near mid-cell and the location of the *ter* marks the division site [16, 17]. Following our model, we speculate that DnaA sequesters ResR during active replication initiation to prevent onset of division prior to chromosomal duplication. Such sequestration of ResR during active replication should not only delay division, but also should subtly increase replication. In our complementation strain, we observed that complementation with clinical IGR variants that were predicted to sequester ResR resulted in a slight increase in *ori/ter* ratio while strains with protein binding site deletions in IGR had lower *ori/ter* ratio. *Ori/ter* ratio is a reflection of chromosomal duplication rates and is correlated with an increase in replication and

growth rate [22]. However, in our strains with different *ori/ter* ratios, growth rate in both liquid and solid media did not change. The increase in replication dynamics supports our model where the active form of DnaA sequesters ResR and the invariant growth rate suggests that this sequestration is likely temporal in order to not affect population growth dynamics. This could be experimentally tested in the future with finer resolution techniques to measure single cell dynamics in *Mtb*.

4.2 MtrAB: a hidden player in this pathway?

Although we do not have direct measurements of cell cycle events in our genetically manipulated strains, we can spot transcriptional differences that point to altered replication and division dynamics, in addition to the cell division genes mentioned in previous chapters. These gene expression changes are linked to MtrAB, one of the two essential two-component systems (TCS) in *Mtb* [23, 24]. Two-component systems are important and conserved regulatory systems that allow bacteria to sense and respond to environmental changes. Two-component systems consist of a membrane bound histidine kinase (HK) and a cytosolic response regulator (RR). Upon stimulation by a certain environmental signal, HK autophopshorylates and transfers phosphate to its cognate RR. The phosphorylated, and therefore active, RR alters expression of its regulon genes to mount a cellular response [25, 26]. In this paired TCS, MtrB is a membrane bound histidine kinase and MtrA is the cytosolic response regulator whose activation and regulon expression requires phosphorylation by MtrB [23, 27].

MtrAB is conserved across Actinobacteria and is the functional analog of the YycFG TCS in Gram-positive Firmicutes that regulates cell wall homeostasis and controls cell growth and division [28-30]. In mycobacteria, altered expression of MtrAB showed defects in cell wall

synthesis and division, and resulted in differences in morphology, antibiotic susceptibility, and virulence [31-34]. Both MtrA and MtrB have been associated with clinical drug resistance and tolerance in *Mtb*, but the mechanism by which they affect drug susceptibility remains unknown [35, 36]. *Mtb* MtrB localizes to the septum and that septal localization is required for its activation. MtrB associates with the septum through interaction with divisome protein FtsI (PBPb) and Wag31, which implies a role for MtrB in cell division [33, 37]. Due to its role in MtrB activation, septal association is also required for MtrA activation and expression of MtrA regulon genes. Consistently, MtrA directly regulates expression of genes encoding proteins that are involved in cell division such as *ftsI*, *wag31*, and *ripA* [37]. Notably, we found in our data that many differentially expressed genes (including most of the genes related to cell division) in the complementation strains are under direct control of MtrA based on evidence from a previous ChIP-seq dataset [38].

In previously unpublished work from our lab, we found that WhiB2 is a direct regulator of the *mtrAB* operon and binds at the promoter region. Downregulation of *whiB2* by CRISPR interference results in downregulation of *mtrA* and of its regulon genes such as *ripA* and *wag31* [15]. However, in our transcriptomic data, we did not observe significant changes to *mtrA* expression level. We speculate that a decrease in phosphorylation of MtrA by MtrB due to delay in division and septum formation are driving these expression changes. In our data, we did observe downregulation of *lpqB*, a gene immediately downstream of the *mtrAB* operon encoding a putative lipoprotein [28]. LpqB interacts with MtrB and disruption of *lpqB* decreases MtrA phosphorylation significantly [39]. Therefore, downregulation of *lpqB* is consistent with the suspected decrease in MtrA phosphorylation.

Although MtrB activation is dependent on division and the MtrA regulon mainly consists of division related genes, phosphorylated MtrA plays an additional role in replication and *dnaA* expression [31]. Phosphorylated MtrA binds replication origin *oriC* and likely occupies the origin post-replication initiation to avoid unwanted interaction of DnaA to *oriC* and prevent untimely initiation [40, 41]. Our transcriptional data supports a less active MtrA in the presence of clinical IGR variant given the downregulation of its regulon genes and therefore less phosphorylated MtrA. This is consistent with our model of delayed division and slightly prolonged replication where less phosphorylated MtrA allows the *oriC* to be available for DnaA binding [31, 41]. Together with our data and existing literature, we suspect that the MtrAB system plays an important role in the downstream regulon, but future experiments are required to prove the connection of MtrAB to the pathway through *Rv0010c-Rv0011c* IGR.

4.3 *Rv0010c-Rv0011c* IGR: a busy traffic intersection on the *Mtb* chromosome

Although little was known when we first started to work on the *Rv0010c-Rv0011c* operon, it quickly became evident that this operon is highly trafficked by many regulatory proteins and elements [38]. These includes several nucleoid associated proteins, such as Lsr2 and NapM, which are known to interact with DnaA either in mycobacteria or in homologous forms in other bacteria [11, 42]. Other proteins that have binding sites at this region based on an existing ChIP-seq dataset include CRP and EspR [38]. CRP is characterized as a cAMP-dependent transcription factor and regulates gene expression in response to cAMP signaling [43]. cAMP signaling transduction is an important sensing mechanism for prokaryotes and is critical for virulence and pathogenesis in pathogenic bacteria [44, 45]. In *Mtb*, CRP binds at the *ori* region and regulates expression of *sucC* and *sucD*, two differentially regulated genes in our

complementation strain RNAseq [43]. EspR transcriptionally regulates a key virulence system in *Mtb*, the ESX-1 secretion system, which is required for *Mtb* pathogenesis [46, 47]. Potential interaction with either the nucleoid associated proteins or these transcriptional regulators could be a reason why complementation requires the complete sequence integrity of this operon as some of them bind within coding regions of *Rv0010c* or *Rv0011c*.

Another genetic element that was connected to *Rv0010c-Rv0011c* IGR is the mobile genetic element IS*6110*. IS*6110* is a unique transposable element in the *Mycobacterium tuberculosis* complex (MTBC) that includes *Mtb*, along with several closely related pathogenic mycobacteria [48]. IS*6110*'s copy number, insertion frequency, and location differ between lineages of *Mtb* and the effect of its insertion on genome organization and expression is thought to drive evolution in lineages of *Mtb*. In our previous GWAS, IS*6110* were found to insert at the *Rv0010c-Rv0011c* IGR near the palindrome at high frequency [10]. A recent pangenome study of *Mtb* also identified the same region within the *Rv0010c-Rv0011c* IGR as a hotspot for IS*6110* insertion [49]. With a promoter at the 3' end of the element, IS*6110* insertion in regulatory regions can change gene expression and its insertion upstream or intragenic of drug resistance related genes is known to increase drug resistance in *Mtb* [50, 51]. In preliminary analysis of IS*6110* insertion sites on the genomes of clinical strains in our lab collection, IS*6110* insertions appear at the same spot as previously characterized in the *Rv0010c-Rv0011c* IGR and they do not seem to have a directionality bias. This likely links back to the fact that, in our data, this IGR functions independently of coding regions in this operon. IS*6110* was also known to insert at the replication origin between *dnaA* and *dnaN* and these insertions were found to be highly prevalent in *Mtb* strains of Beijing lineage [52]. Notably, our original GWAS that found links between *dnaA* and drug resistance was done in a collection of mainly strains of Beijing lineage [10].

These insertions at the *ori* were at a specific location and also can be found in both directions [52]. Knowing IS*6110* inserts at two locations where DnaA bind to on the *Mtb* genome, the *ori* and the *Rv0010c-Rv0011c* IGR, we wonder what role IS*6110* insertions play on the function and evolution of the *Rv0010c-Rv0011c* IGR and its phenotypic consequences. In summary, there are many other potential proteins and genetic elements that could play a role in how *Rv0010c-Rv0011c* IGR functions and it would be an interesting part of biology at this IGR to follow up on in future studies.

4.4 Linking morphology to fitness in Mycobacteria

This study on clinical variants highlighted one key phenotype, morphology, that remains unexplored and overlooked in mycobacteria. In *B. subtills* and *E. coli*, morphology has been tightly linked to bacterial fitness especially in response to environmental stresses. Unlike mycobacteria, *B. subtills* and *E. coli* add new cell wall material laterally throughout their cell length, dependent on the bacterial actin homolog MreB. Targeted evolution of MreB variants directly increase cell size and relative fitness compared to their ancestral counterparts [53, 54]. A later study characterizing these *mreB* variants shows that they alter cell size without affecting growth rate, providing another example of essential protein perturbation that drives cell morphology phenotypes without large impacts on cell cycle [55]. A long-term experimental evolution in *E. coli* also showed a positive correlation between cell volume and relative fitness [56].

Though it has been studied in other bacteria, the correlation between cell size and fitness remains unexplored in mycobacteria. With our collection of clinical strains in the lab, we sought to understand how cell length varies in ambient growth condition and observed an overall shift to

longer cells across the clinical strains. Mycobacteria does not have a MreB homolog and the determinants of cell size control remain uncharacterized, complicating our ability to investigate the link between size and fitness [54]. One possible benefit of increased cell size is decreased surface-to-volume ratio (SA/V) in the presence of drug. Decreased SA/V reduces antibiotic influx and dilutes concentration of intracellular antibiotics [57, 58]. Phenotyping of drug resistant *Mtb* collected directly from patient sputum or grown in macrophages has previously demonstrated significant cell size differences compared to the same strain grown *in vitro* [59]. Future experiments exploring the link between morphology and *Mtb* pathogenesis would be beneficial to provide more insight into how *Mtb* regulates its physiology to maximize survival inside human hosts.

4.5 Concluding remarks

Our study provides the first genetic characterization of a novel cell cycle regulatory pathway under clinical selection. Variation in this pathway alters cell cycle dynamics to potentially allow faster recovery post-antibiotic pressure. Our data provided the first evidence that cell cycle regulation can play a role in *Mtb* evasion from antibiotic clearance and showed the importance of bacterial physiology in response to drug pressure. The discovery of antibiotic resilience also emphasized the spatiotemporal nature of TB disease, which we need to take into consideration in future improvements of TB chemotherapy.

Several components of this genetic pathway were uncharacterized prior to our studies in the last few years. A large proportion of the *Mtb* genome remains uncharacterized, even though many coding and non-coding regions of unknown function are under clinical selection in our genomics screen. A more comprehensive understanding of these unknown regions could allow us

to discover additional mechanisms *Mtb* use to evade clearance and could help us pinpoint the aspects of treatment that *Mtb* may use to their survival advantage. We hope to inspire more genetic exploration into the unknowns of *Mtb* biology in the future and, together with our data, we hope to design more comprehensive antitubercular therapy and contribute to eradicate TB one day.

4.6 References

1. Farhat, M., et al., *Drug-resistant tuberculosis: a persistent global health concern.* Nature Reviews Microbiology, 2024: p. 1-19.
2. World Health Organization, *Global tuberculosis report 2023*. 2023, Geneva: World Health Organization.
3. Zhang, H., et al., *Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance.* Nature genetics, 2013. **45**(10): p. 1255-1260.
4. Hicks, N.D., et al., *Clinically prevalent mutations in Mycobacterium tuberculosis alter propionate metabolism and mediate multidrug tolerance.* Nat Microbiol, 2018. **3**(9): p. 1032-1042.
5. Safi, H., et al., *Phase variation in Mycobacterium tuberculosis glpK produces transiently heritable drug tolerance.* Proc Natl Acad Sci U S A, 2019. **116**(39): p. 19665-19674.
6. Liu, Q., et al., *Tuberculosis treatment failure associated with evolution of antibiotic resilience.* Science, 2022. **378**(6624): p. 1111-1118.
7. Hansen, F.G. and T. Atlung, *The DnaA Tale.* Frontiers in Microbiology, 2018. **9**.
8. Sassetti, C.M., D.H. Boyd, and E.J. Rubin, *Genes required for mycobacterial growth defined by high density mutagenesis.* Molecular Microbiology, 2003. **48**(1): p. 77-84.
9. Bosch, B., et al., *Genome-wide gene expression tuning reveals diverse vulnerabilities of M. tuberculosis.* Cell, 2021. **184**(17): p. 4579-4592. e24.
10. Hicks, N.D., et al., *Mutations in dnaA and a cryptic interaction site increase drug resistance in Mycobacterium tuberculosis.* PLOS Pathogens, 2020. **16**(11): p. e1009063.

11. Liu, Y., et al., *NapM enhances the survival of Mycobacterium tuberculosis under stress and in macrophages.* Communications Biology, 2019. **2**(1).

12. Zhou, W., et al., *A feedback regulatory loop containing McdR and WhiB2 controls cell division and DNA repair in mycobacteria.* mbio, 2022. **13**(2): p. e03343-21.

13. Bush, M.J., *The actinobacterial WhiB-like (Wbl) family of transcription factors.* Molecular Microbiology, 2018. **110**(5): p. 663-676.

14. Gomez, J.E. and W.R. Bishai, *whmD is an essential mycobacterial gene required for proper septation and cell division.* Proceedings of the National Academy of Sciences, 2000. **97**(15): p. 8554-8559.

15. Giffen, S.R., *The Essential whiB2 Transcriptional Network and its Role in Cell Division*. 2021, Harvard University.

16. Logsdon, M.M. and B.B. Aldridge, *Stable regulation of cell cycle events in mycobacteria: Insights from inherently heterogeneous bacterial populations.* Frontiers in Microbiology, 2018. **9**: p. 348921.

17. Logsdon, M.M., et al., *A Parallel Adder Coordinates Mycobacterial Cell-Cycle Progression and Cell-Size Homeostasis in the Context of Asymmetric Growth and Organization.* Curr Biol, 2017. **27**(21): p. 3367-3374 e7.

18. Rozwarski, D.A., et al., *Modification of the NADH of the isoniazid target (InhA) from Mycobacterium tuberculosis.* Science, 1998. **279**(5347): p. 98-102.

19. Ramaswamy, S. and J.M. Musser, *Molecular genetic basis of antimicrobial agent resistance in Mycobacterium tuberculosis: 1998 update.* Tubercle and Lung disease, 1998. **79**(1): p. 3-29.

20. Zhang, M., et al., *Detection of mutations associated with isoniazid resistance in Mycobacterium tuberculosis isolates from China.* Journal of clinical microbiology, 2005. **43**(11): p. 5477-5482.

21. Nair, N., et al., *Synchronous replication initiation in novel Mycobacterium tuberculosis dnaA cold-sensitive mutants.* Molecular microbiology, 2009. **71**(2): p. 291-304.

22. Charbon, G., et al., *Counting Replication Origins to Measure Growth of Pathogens.* Antibiotics, 2020. **9**(5): p. 239.

23. Zahrt, T.C. and V. Deretic, *An essential two-component signal transduction system in Mycobacterium tuberculosis.* Journal of bacteriology, 2000. **182**(13): p. 3832-3838.

24. Haydel, S.E., et al., *The prrAB two-component system is essential for Mycobacterium tuberculosis viability and is induced under nitrogen-limiting conditions.* Journal of bacteriology, 2012. **194**(2): p. 354-361.

25. Hoch, J.A., *Two-component and phosphorelay signal transduction.* Current opinion in microbiology, 2000. **3**(2): p. 165-170.

26. Stock, A.M., V.L. Robinson, and P.N. Goudreau, *Two-component signal transduction.* Annual review of biochemistry, 2000. **69**(1): p. 183-215.

27. Al Zayer, M., et al., *Mycobacterium tuberculosis mtrA merodiploid strains with point mutations in the signal-receiving domain of MtrA exhibit growth defects in nutrient broth.* Plasmid, 2011. **65**(3): p. 210-218.

28. Hoskisson, P.A. and M.I. Hutchings, *MtrAB–LpqB: a conserved three-component system in actinobacteria?* Trends in microbiology, 2006. **14**(10): p. 444-449.

29. Winkler, M.E. and J.A. Hoch, *Essentiality, bypass, and targeting of the YycFG (VicRK) two-component regulatory system in gram-positive bacteria.* Journal of bacteriology, 2008. **190**(8): p. 2645-2648.

30. Dubrac, S., et al., *A matter of life and death: cell wall homeostasis and the WalKR (YycGF) essential signal transduction pathway.* Molecular microbiology, 2008. **70**(6): p. 1307-1322.

31. Fol, M., et al., *Modulation of Mycobacterium tuberculosis proliferation by MtrA, an essential two -component response regulator.* Molecular microbiology, 2006. **60**(3): p. 643-657.

32. Cangelosi, G.A., et al., *The two-component regulatory system mtrAB is required for morphotypic multidrug resistance in Mycobacterium avium.* Antimicrobial agents and chemotherapy, 2006. **50**(2): p. 461-468.

33. Plocinska, R., et al., *Mycobacterium tuberculosis MtrB sensor kinase interactions with FtsI and Wag31 proteins reveal a role for MtrB distinct from that regulating MtrA activities.* Journal of bacteriology, 2014. **196**(23): p. 4120-4129.

34. Gorla, P., et al., *MtrA response regulator controls cell division and cell wall metabolism and affects susceptibility of mycobacteria to the first line antituberculosis drugs.* Frontiers in microbiology, 2018. **9**: p. 424715.

35. Farhat, M.R., et al., *Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis.* Nature genetics, 2013. **45**(10): p. 1183-1189.

36. Peterson, E.J., et al., *MtrA modulates Mycobacterium tuberculosis cell division in host microenvironments to mediate intrinsic resistance and drug tolerance.* Cell Reports, 2023. **42**(8).

37. Plocinska, R., et al., *Septal localization of the Mycobacterium tuberculosis MtrB sensor kinase promotes MtrA regulon expression.* Journal of Biological Chemistry, 2012. **287**(28): p. 23887-23899.

38. Minch, K.J., et al., *The DNA-binding network of Mycobacterium tuberculosis.* 2015. **6**: p. 5829.

39. Nguyen, H.T., et al., *A lipoprotein modulates activity of the MtrAB two-component system to provide intrinsic multidrug resistance, cytokinetic control and cell wall homeostasis in Mycobacterium.* Molecular microbiology, 2010. **76**(2): p. 348-364.

40. Purushotham, G., et al., *Mycobacterium tuberculosis oriC sequestration by MtrA response regulator.* Molecular Microbiology, 2015. **98**(3): p. 586-604.

41. Rajagopalan, M., et al., *Mycobacterium tuberculosis origin of replication and the promoter for immunodominant secreted antigen 85B are the targets of MtrA, the essential response regulator.* Journal of biological chemistry, 2010. **285**(21): p. 15816-15827.

42. Seid, C.A., J.L. Smith, and A.D. Grossman, *Genetic and biochemical interactions between the bacterial replication initiator DnaA and the nucleoid-associated protein Rok in Bacillus subtilis.* Mol Microbiol, 2017. **103**(5): p. 798-817.

43. Bai, G., L.A. McCue, and K.A. McDonough, *Characterization of Mycobacterium tuberculosis Rv3676 (CRPMt), a cyclic AMP receptor protein-like DNA binding protein.* Journal of bacteriology, 2005. **187**(22): p. 7795-7804.

44. Botsford, J.L. and J.G. Harman, *Cyclic AMP in prokaryotes.* Microbiological reviews, 1992. **56**(1): p. 100-122.

45. Wolfgang, M.C., et al., *Coordinate regulation of bacterial virulence genes by a novel adenylate cyclase-dependent signaling pathway.* Developmental cell, 2003. **4**(2): p. 253-263.

46. Rosenberg, O.S., et al., *EspR, a key regulator of Mycobacterium tuberculosis virulence, adopts a unique dimeric structure among helix-turn-helix proteins.* Proceedings of the National Academy of Sciences, 2011. **108**(33): p. 13450-13455.

47. Wong, K.-W., *The role of ESX-1 in Mycobacterium tuberculosis pathogenesis.* Microbiology spectrum, 2017. **5**(3): p. 10.1128/microbiolspec. tbtb2-0001-2015.

48. McEvoy, C.R., et al., *The role of IS6110 in the evolution of Mycobacterium tuberculosis.* Tuberculosis, 2007. **87**(5): p. 393-404.

49. Marin, M.G., et al., *Analysis of the limited M. tuberculosis accessory genome reveals potential pitfalls of pan-genome analysis approaches.* bioRxiv, 2024: p. 2024.03. 21.586149.

50. Gonzalo-Asensio, J., et al., *New insights into the transposition mechanisms of IS 6110 and its dynamic distribution between Mycobacterium tuberculosis complex lineages.* PLoS genetics, 2018. **14**(4): p. e1007282.

51. Antoine, R., C. Gaudin, and R.C. Hartkoorn, *Intragenic distribution of IS 6110 in clinical Mycobacterium tuberculosis strains: bioinformatic evidence for gene disruption leading to underdiagnosed antibiotic resistance.* Microbiology spectrum, 2021. **9**(1): p. 10.1128/spectrum. 00019-21.

52. Kurepina, N., et al., *Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS6110elements inMycobacterium tuberculosis: non-random integration in thednaA–dnaNregion.* Tubercle and Lung Disease, 1998. **79**(1): p. 31-42.

53. Monds, R.D., et al., *Systematic perturbation of cytoskeletal function reveals a linear scaling relationship between cell geometry and fitness.* Cell reports, 2014. **9**(4): p. 1528-1537.

54. Cabeen, M.T. and C. Jacobs-Wagner, *Skin and bones: the bacterial cytoskeleton, cell wall, and cell morphogenesis.* The Journal of Cell Biology, 2007. **179**(3): p. 381.

55. Shi, H., et al., *Deep Phenotypic Mapping of Bacterial Cytoskeletal Mutants Reveals Physiological Robustness to Cell Size.* Curr Biol, 2017. **27**(22): p. 3419-3429 e4.

56. Lenski, R.E. and M. Travisano, *Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations.* Proceedings of the National Academy of Sciences, 1994. **91**(15): p. 6808-6814.

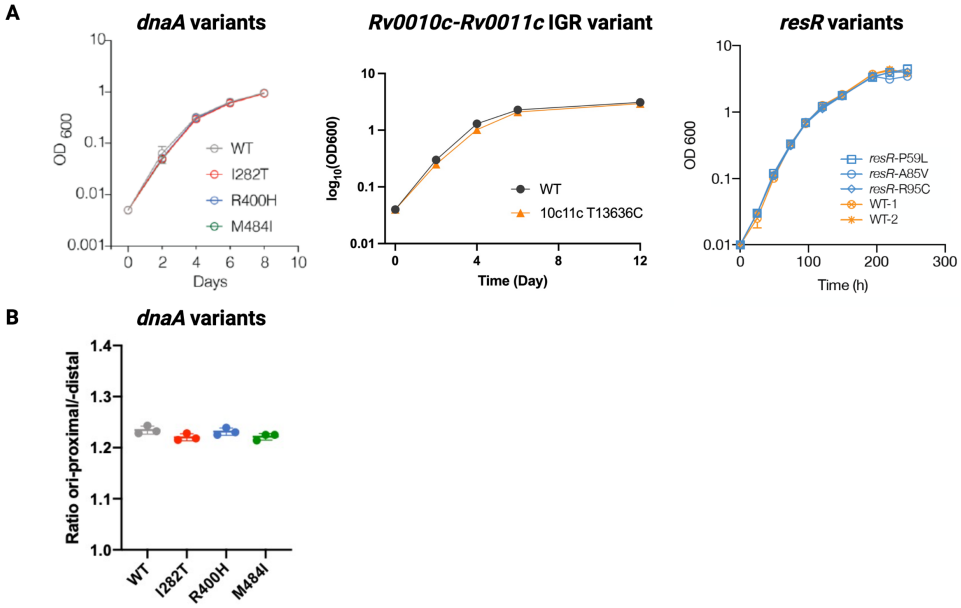
57. Harris, L.K. and J.A. Theriot, *Relative Rates of Surface and Volume Synthesis Set Bacterial Cell Size.* Cell, 2016. **165**(6): p. 1479-1492.

58. Ojkic, N., D. Serbanescu, and S. Banerjee, *Antibiotic resistance via bacterial cell shape-shifting.* Mbio, 2022. **13**(3): p. e00659-22.

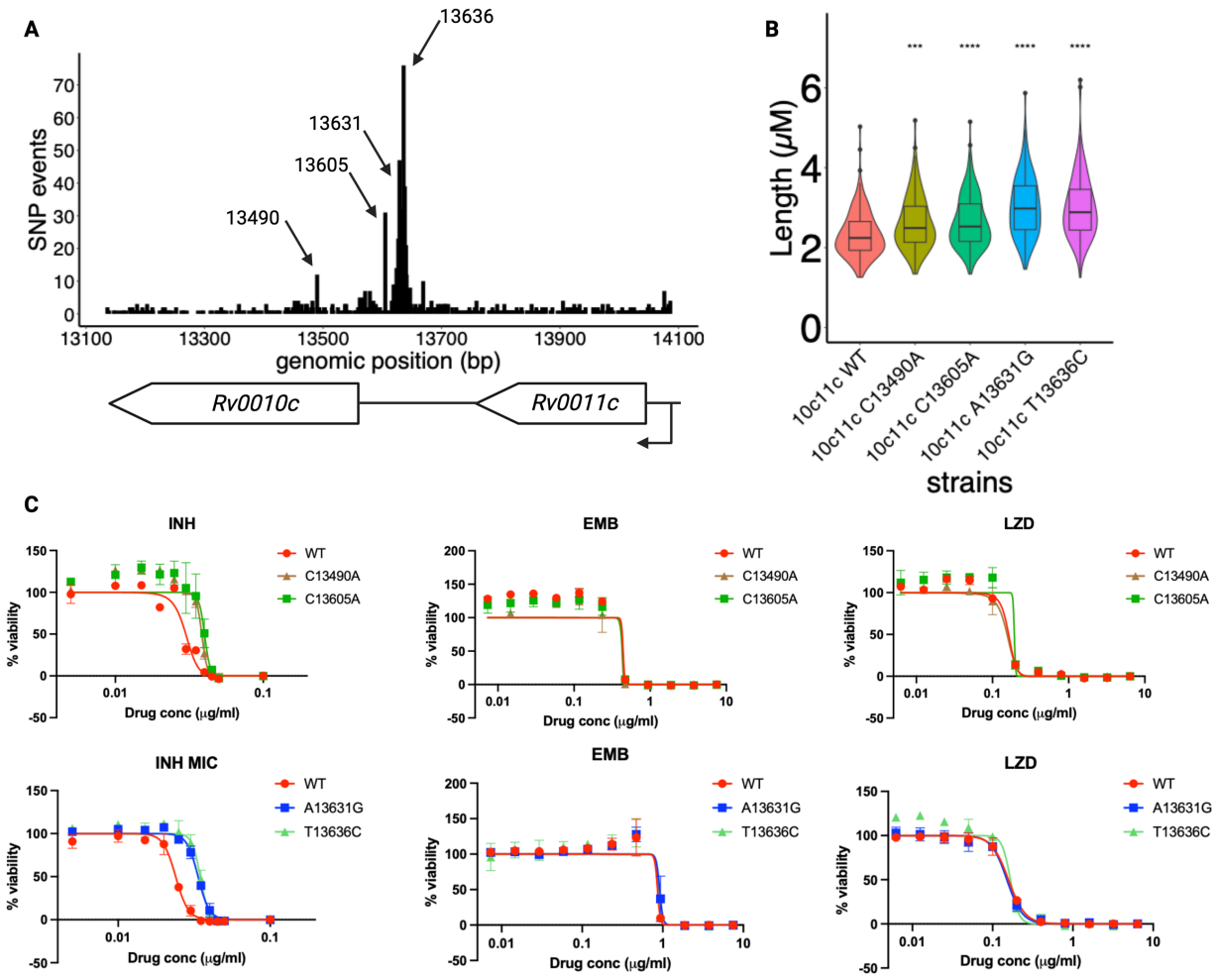
59. Vijay, S., et al., *Influence of Stress and Antibiotic Resistance on Cell-Length Distribution in Mycobacterium tuberculosis Clinical Isolates.* Front Microbiol, 2017. **8**: p. 2296.

Chapter 5: Appendix

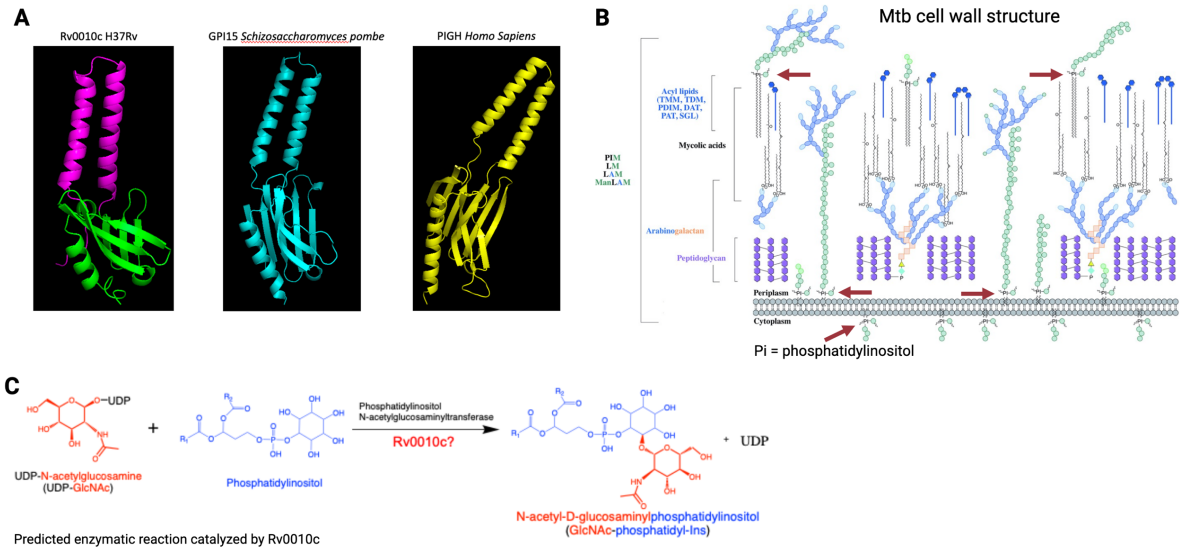
5.1 Supplementary figures for chapter 2



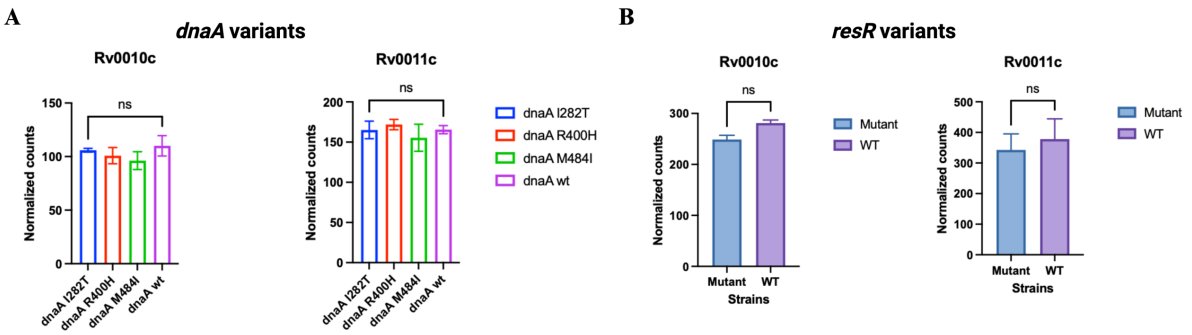
Supplemental figure 2.1 | Isogenic variants in *dnaA*, *Rv0010c-Rv0011c* IGR, *resR* do not alter *in vitro* growth rate or chromosomal replicate rate. A) *in vitro* growth measurement in liquid culture in clinically relevant isogenic mutants of *dnaA*, *Rv0010c-Rv0011c* IGR, and *resR*. Growthcurves of *dnaA* and *resR* variants are adapted from Hicks *et al.* 2020 and Liu *et al.* 2022 respectively. B) Coverage comparison of *ori*-proximal region and *ori*-distal region by whole-genome sequencing of *dnaA* variants. Adapted from Hicks *et al.* 2020.



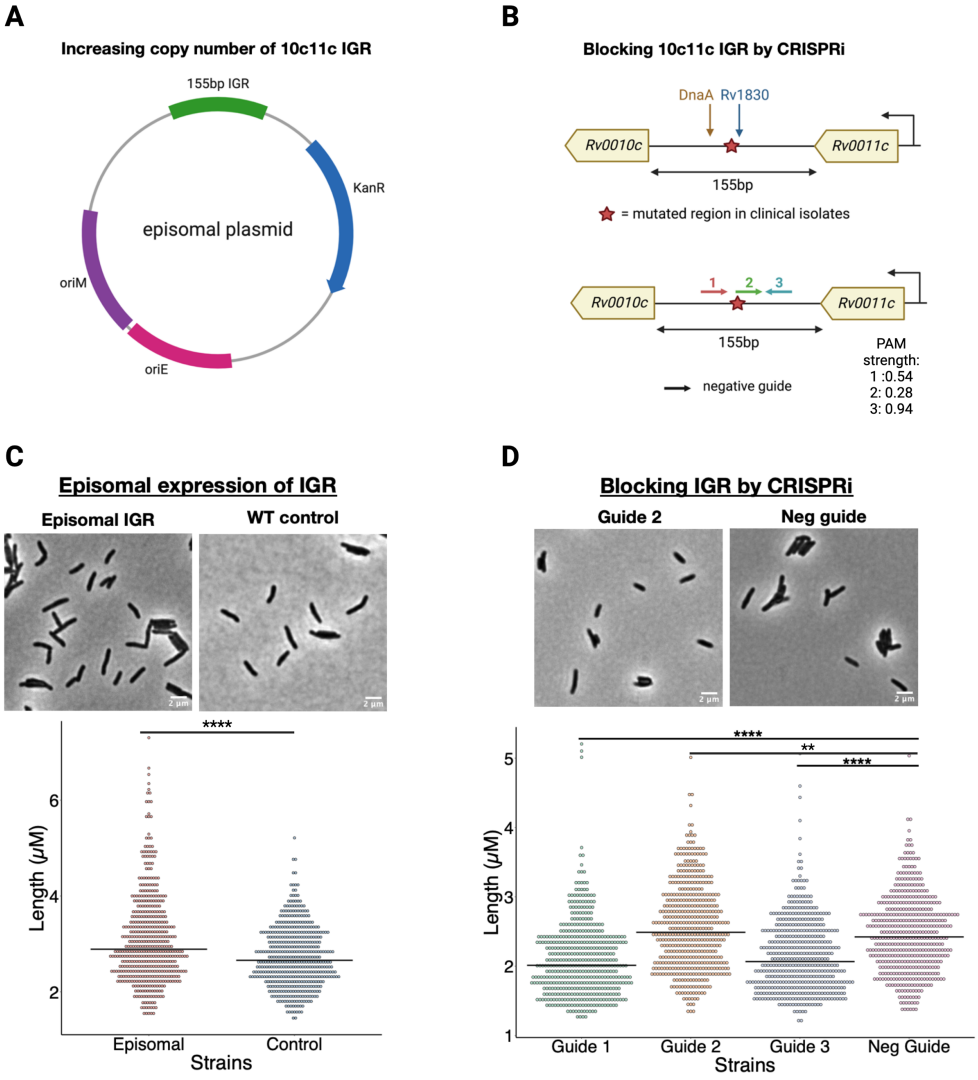
Supplemental figure 2.2 | Single nucleotide variants across different mutational spots in *Rv0010c-Rv0011c* operon share similar morphology phenotypes and drug susceptibility differences in addition to INH. A) Mutational events across Rv0010c-Rv0011c operon. Adapted from Liu *et al.* 2022. B) and C) Morphology measurements were done using quantitative microscopy on cells grown to mid log in plain media and were analyzed through the MOMIA pipeline. INH susceptibility was determined using the alamar blue assay. Adjusted p-value by Mann Whitney U Test , *** < 0.001, **** < 0.0001.



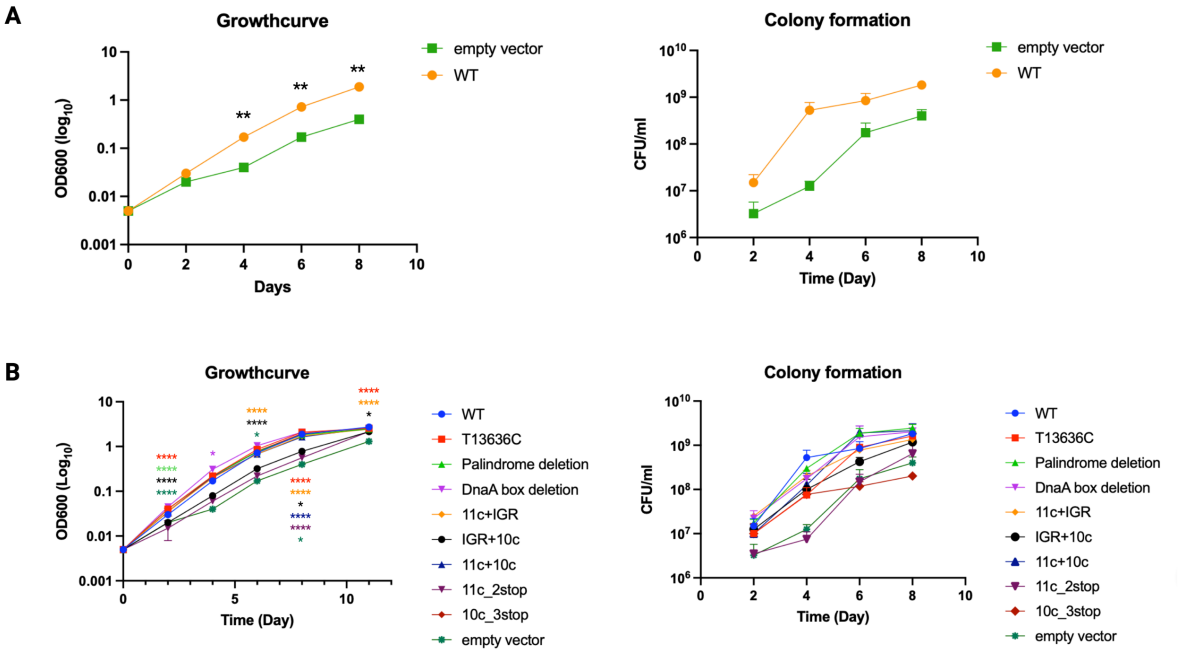
Supplementary figure 2.3 | Rv0010c shares homology to eukaryotic Phosphatidylinositol N-acetylglucosaminyltransferases and phosphatidylinositol are important lipid anchors in mycobacterial cell wall. A) Alphafold predicted structure of Rv0010c and its structural homologs in eukaryotes. B) Rv0010c potentially adds N-acetylglucosamine group onto phosphatidylinositol, base of many surface glycolipids (red arrows). Adapted from Abrahams and Besra, 2018. C) Predicted enzymatic reaction catalyzed by Rv0010c based on the function of its eukaryotic homologs shown in A).



Supplementary figure 2.4 | Transcription of *Rv0010c* and *Rv0011c* are unaffected in *dnaA* and *resR* mutant strains. A) Normalized counts from previous RNA-seq in *dnaA* variants and B**)** in *resR* variants show no change in transcription of *Rv0010c* and *Rv0011c.* Adapted from RNA seq data from Hicks *et al.* 2020 and Liu *et al.* 2022.

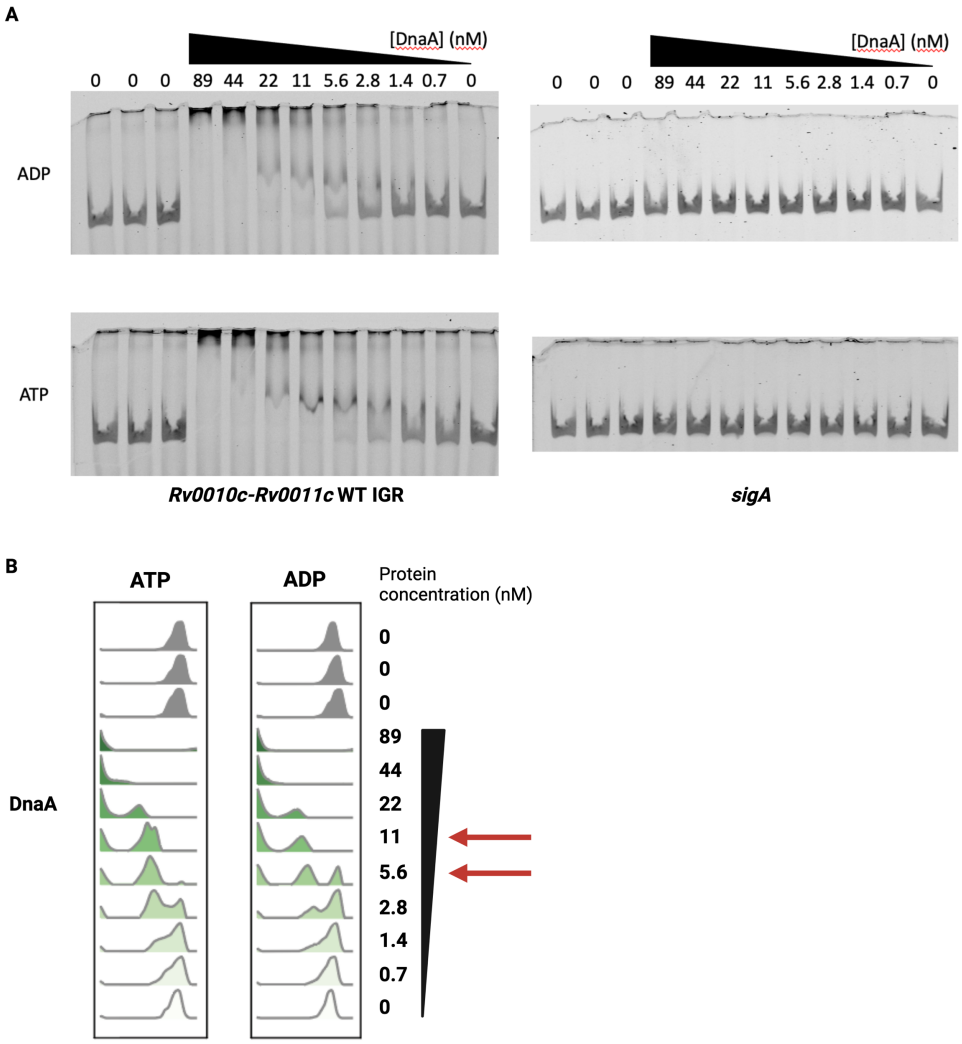


Supplementary figure 2.5 | Varying availability of *Rv0010c-Rv0011c* in *Mtb* directly affect cell morphology. A) and B) Schematic of vectors used to increase copy number of *Rv0010c-Rv0011c* by episomal expression and to block *Rv0010c-Rv0011c* by CRISPR interference. C) Increasing copy number of *Rv0010c-Rv0011c* IGR in a low-copy episomal plasmid (~3-5 copies/cell) increases cell length when compared to scrambled control. Any promoter element in the region on the plasmid was removed prior to insertion of IGR. D) Blocking *Rv0010c-Rv0011c* by CRISPR interference decreases cell length when compared to strain with negative control guide. Cells were harvested after induction with 100ng/ml aTC for five days at mid-log OD for microscopy. Image analysis was done using the MOMIA pipeline. Adjusted p-value by Mann Whitney U Test , *** < 0.001, **** < 0.0001.

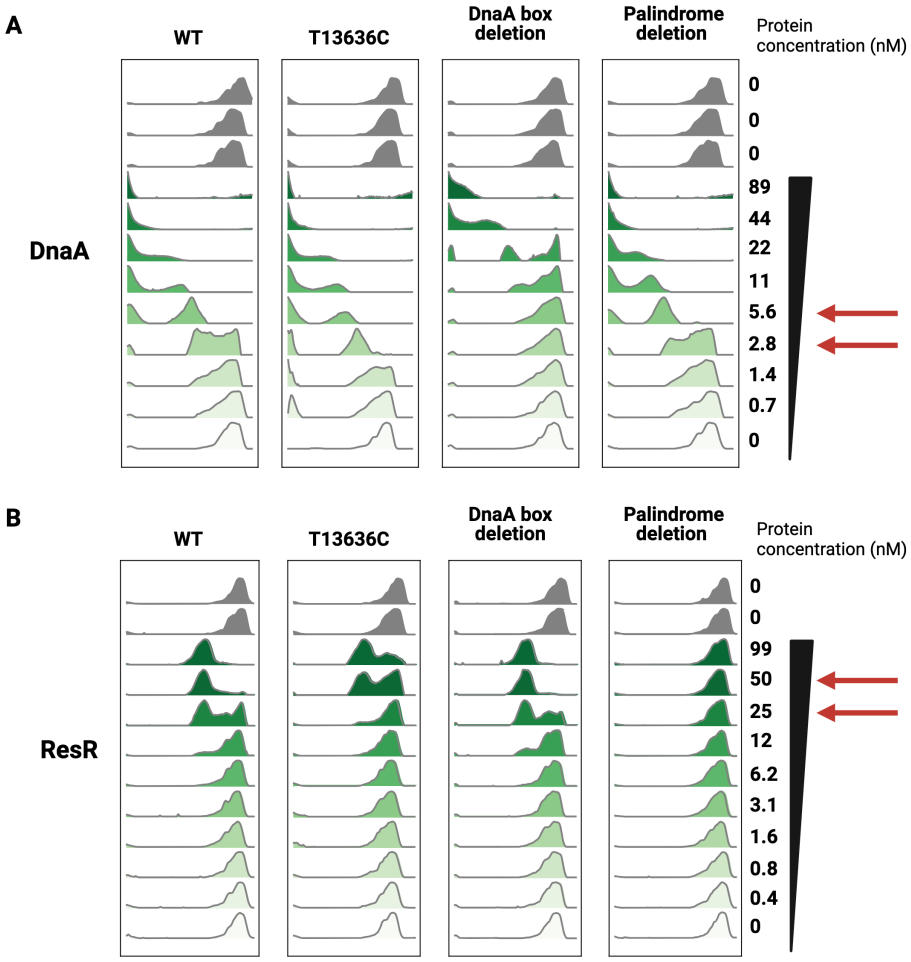


Supplementary figure 2.6 | *Rv0010c-Rv0011c* KO grows slower than WT and this growth defect can be complemented and is largely driven by loss of *Rv0011c*. A) and B) Measurement of complementation strain growth in liquid by OD600 and on solid media by colony formation. Notably, complementation strains without *Rv0011c* or with nonfunctional Rv0011c (empty vector, IGR + 10c, and 11c_2stop respectively) grow slower and loss of *Rv0011c* likely explains the growth difference. Adjusted p-value by two-way ANOVA shown, * < 0.05, ** < 0.01, **** < 0.0001.

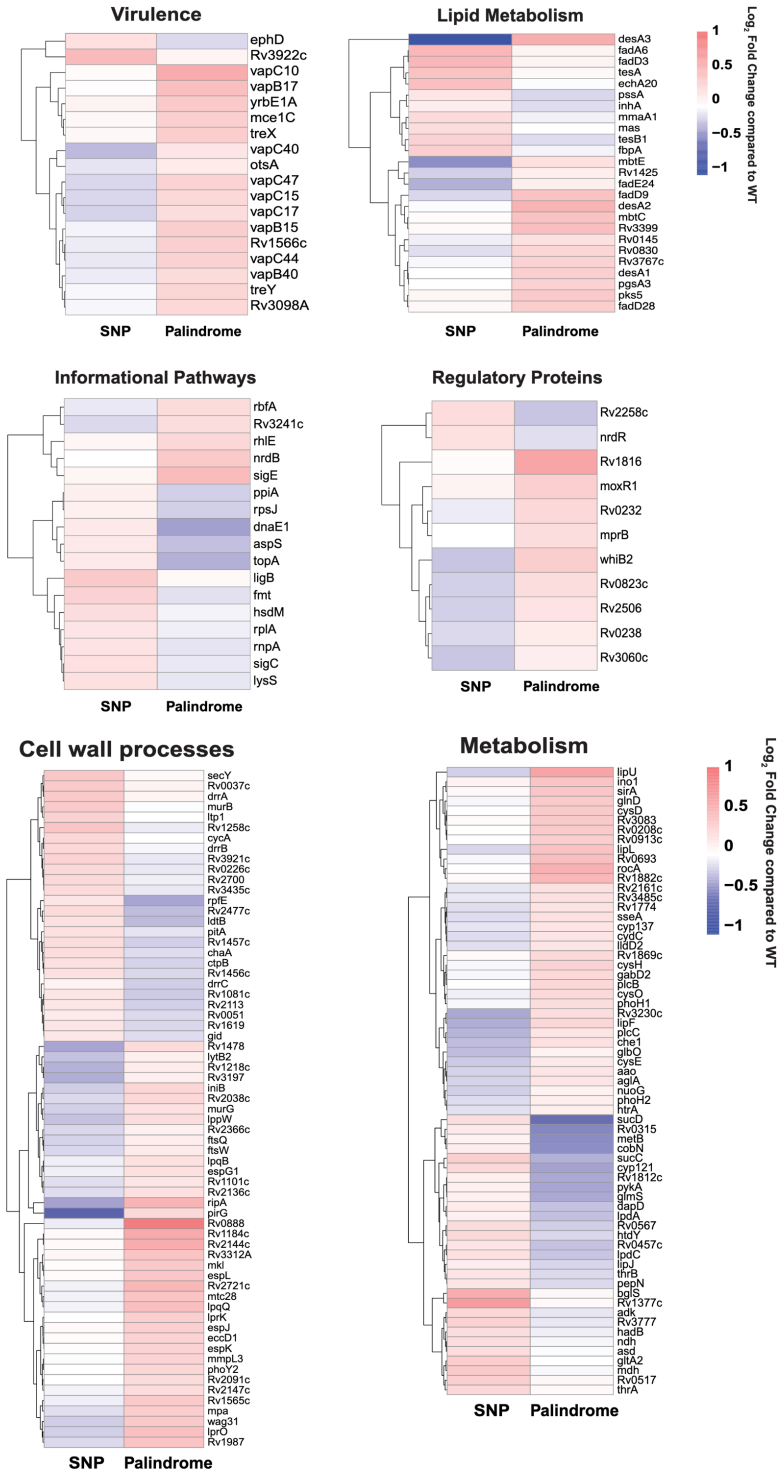
5.2 Supplementary figures for chapter 3



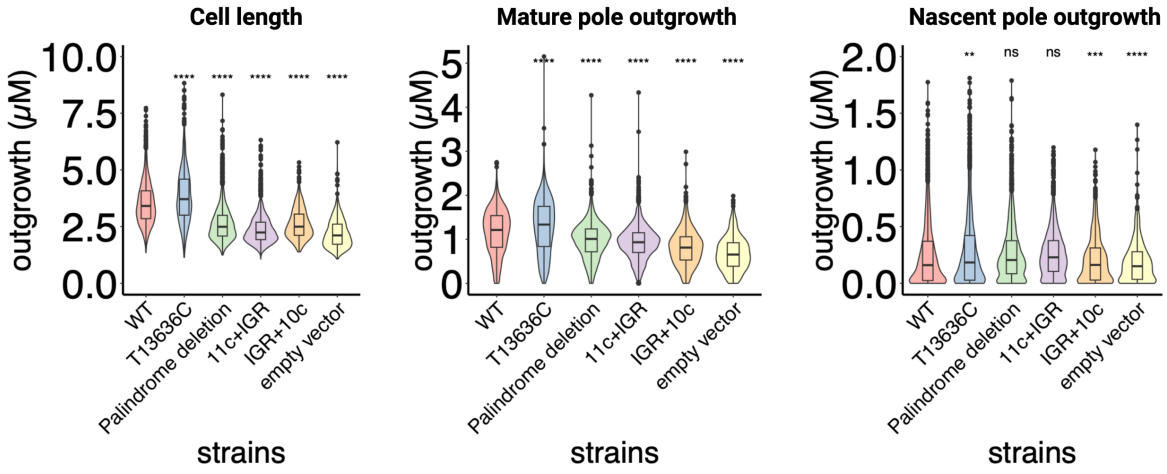
Supplementary figure 3.1 | The presence of ADP vs ATP does not dramatically affect DnaA binding to *Rv0010c-Rv0011c* IGR. A) DnaA binding kinetics to WT *Rv0010c-Rv0011c* IGR in the presence of ATP and ADP measured with electrophoretic mobility shift assay (EMSA). DnaA and ResR binding kinetics to WT and variant *Rv0010c-Rv0011c* IGR measured with electrophoretic mobility shift assay (EMSA). Concentrations of DnaA used are indicated on top of the gel in reaction loaded into each lane. DnaA were pre-incubated with 5mM ATP or ADP and reactions were carried out at 2.5mM ATP or ADP. Probes for specific interaction (WT *Rv0010c-Rv0011c* IGR) labeled with IRD800 dye and non-specific interaction (*sigA*, not shown) labeled with IRD700 dye were added at 0.1nM. Cold competitor poly dI-dC was added at final concentration of 100 ng/µl. Proteins and probes were incubated at room temperature for 20min prior to loading into 6% DNA retardation gel. Gel images shown here was captured with LI-COR imagining system and analyzed in Empiria Studio. Red arrows indicate potential concentrations of interest. B) Computational tracing of EMSA gel results based on band intensity generated in ImageJ using a custom python script. Concentrations where we see the most difference are indicated by red arrows.



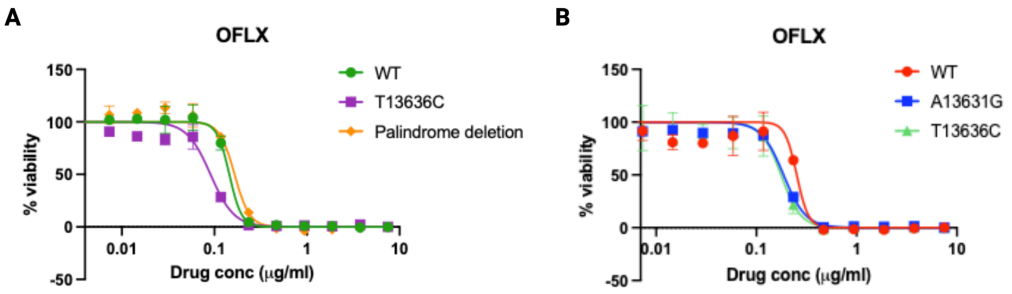
Supplementary figure 3.2 | Computational tracing of EMSA results for Figure 3.1. A custom python script was used to generate tracing of band intensity following quantification in ImageJ. Red arrows indicate potential concentrations of interest.



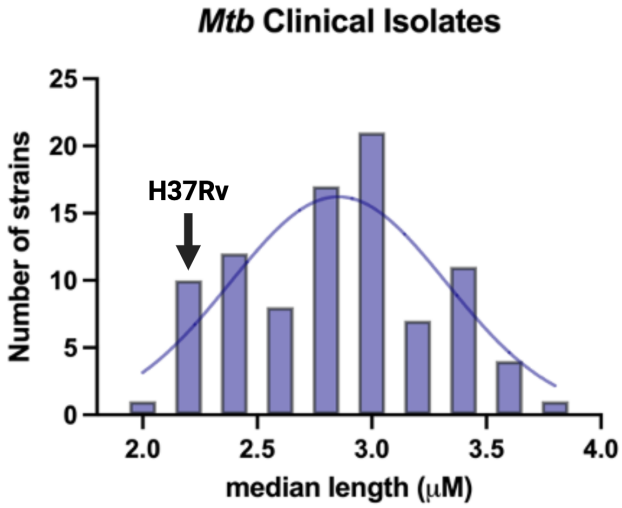
Supplementary figure 3.3 | Heatmaps of differentially expression genes in major pathways. A few pathways were excluded from this graph: insertion sequences, PE/PPE genes, and conserved hypotheticals because of the small number of genes and the uncertainty of their functions. Genes in these pathways are included in the final gene list in supplementary table 3.1.



Supplementary figure 3.4 | Pulse-chase labeling reveals positive relationship between cell length and mature pole elongation in complementation strains as observed in isogenic variants. Pulse chase labeling was performed as described in figure 2.3. Measurements of cell length, width, and mature pole elongation outgrowth were done using quantitative microscopy on cells grown to mid log in plain media and were analyzed through the MOMIA pipeline. Adjusted p-value by Mann Whitney U Test , ** < 0.01, *** < 0.001, **** < 0.0001.



Supplementary figure 3.5 | *Rv0010c-Rv0011c* IGR variants are more sensitive to ofloxacin (OFLX), a fluoroquinone used in TB treatment. Measurements of OFLX susceptibility in A) complementation variants and B) isogenic variants are performed using the alamar blue assay.



Supplementary figure 3.6 | Morphological phenotyping to clinical strains reveal an increase in cell length compared to lab strain H37Rv. Cell length was measured from clinical strains at mid-log OD by quantitative microscopy and MOMIA imaging analysis pipeline. Median cell length was used to bin strains and a Gaussian curve was plotted to fit the distribution.

Supplementary table 3.1 | Differentially expressed genes in SNP and ΔPalindrome strain.

Selection criteria are described in figure 3.3.

gene	Log2FoldChange (SNP/WT)	Log2FoldChange (Palindrome/WT)	Padj (SNP/Palindrome)	gene_name
Rv0009	0.02944438	-0.2895904	0.00751054	ppiA
Rv0037c	0.30274979	-0.00714	0.04730612	Rv0037c
Rv0040c	-0.1595029	0.36810043	7.0653E-05	mtc28
Rv0046c	-0.0071305	0.35760471	0.02544087	ino1
Rv0051	0.06454768	-0.2474074	0.03492669	Rv0051
Rv0103c	0.13336067	-0.2712346	0.00242722	ctpB
Rv0145	-0.1656363	0.12639251	0.00367513	Rv0145
Rv0164	0.13481854	-0.2222063	0.03534823	TB18.5
Rv0167	-0.0022407	0.32098695	0.02118005	yrbE1A
Rv0171	-0.0385396	0.26242883	0.03196703	mce1C
Rv0173	-0.1004385	0.26757586	0.00750189	lprK
Rv0179c	-0.2842148	0.38238385	5.4764E-09	lprO
Rv0184	0.38424365	-0.029151	0.00468647	Rv0184
Rv0186	0.53978829	-0.0353307	1.5553E-05	bglS
Rv0206c	-0.1158704	0.2346296	0.00052557	mmpL3
Rv0207c	-0.1136205	0.31873645	0.00043078	Rv0207c
Rv0208c	-0.1027431	0.30454263	0.01415961	Rv0208c
Rv0226c	0.1997728	-0.1881839	0.02098973	Rv0226c
Rv0232	-0.1741388	0.20145537	0.01660899	Rv0232
Rv0233	-0.0820468	0.32064563	0.00466983	nrdB
Rv0238	-0.2442967	0.05150402	0.03461612	Rv0238
Rv0312	-0.4386999	0.16597724	1.1595E-05	Rv0312
Rv0315	0.07330358	-0.5503934	6.3686E-06	Rv0315
Rv0341	-0.2924007	0.22564242	0.00097509	iniB
Rv0436c	0.10780983	-0.2584655	0.0251576	pssA
Rv0457c	0.09138898	-0.3301873	0.02309908	Rv0457c
Rv0462	0.12000013	-0.3221457	0.00038576	lpdC
Rv0482	0.30902288	-0.1106502	0.01340894	murB
Rv0517	0.27912939	-0.0378747	0.04958269	Rv0517
Rv0545c	0.12305188	-0.2074023	0.01184761	pitA
Rv0567	0.1858965	-0.2967299	0.00147895	Rv0567
Rv0597c	-0.6244162	0.11989448	5.1001E-06	Rv0597c
Rv0636	0.15547866	-0.1665044	0.01818254	hadB

Rv0641	0.09673364	-0.1687058	0.02166158	rplA
Rv0645c	0.18908221	-0.1526329	0.03315383	mmaA1
Rv0655	-0.0758158	0.32496178	0.00011649	mkl
Rv0693	-0.128501	0.39997277	1.0944E-06	Rv0693
Rv0698	-0.2332167	0.155443	0.01720703	Rv0698
Rv0700	0.01300435	-0.2867577	0.04817839	rpsJ
Rv0732	0.33346379	-0.0400863	0.01037944	secY
Rv0733	0.21760407	-0.1984601	0.00400759	adk
Rv0821c	-0.0922798	0.22209106	0.03866969	phoY2
Rv0822c	-0.2373412	0.20268313	1.565E-05	Rv0822c
Rv0823c	-0.2819398	0.15528291	0.00093184	Rv0823c
Rv0824c	-0.1053032	0.26900898	0.00147895	desA1
Rv0830	-0.2208409	0.22951027	0.00750189	Rv0830
Rv0835	-0.1417459	0.40674592	0.0018403	lpqQ
Rv0883c	-0.2905987	0.03946688	0.02544131	Rv0883c
Rv0888	-0.1700432	0.90746535	1.2464E-08	Rv0888
Rv0896	0.29451706	-0.0880861	0.00024467	gltA2
Rv0913c	-0.0864709	0.29146884	0.02931893	Rv0913c
Rv0951	0.28750308	-0.4075643	1.5578E-08	sucC
Rv0952	0.16773505	-0.6568892	1.522E-09	sucD
Rv0982	-0.097066	0.16659062	0.02167655	mprB
Rv0997a	-0.266976	0.06237076	0.03085786	Rv0997a
Rv1076	-0.2839151	0.6011988	1.5578E-08	lipU
Rv1078	0.46419977	-0.0359265	4.2062E-05	Rv1078
Rv1079	0.00271311	-0.5336268	0.00096803	metB
Rv1081c	0.09659476	-0.304255	0.02968323	Rv1081c
Rv1094	-0.1143833	0.47230971	5.1001E-06	desA2
Rv1095	-0.2313951	0.03109172	0.04932322	phoH2
Rv1101c	-0.2298334	0.09833839	0.0323255	Rv1101c
Rv1110	-0.3369858	0.02631161	0.00378678	lytB2
Rv1157c	-0.313296	0.37024937	7.0451E-07	Rv1157c
Rv1158c	-0.0279235	0.41892037	0.01797825	Rv1158c
Rv1184c	-0.0587427	0.55868599	1.9583E-06	Rv1184c
Rv1187	-0.0954609	0.5064122	0.00049414	rocA
Rv1195	-0.1976521	0.34498358	3.5324E-05	PE13
Rv1196	-0.1132403	0.50726387	5.7216E-08	PPE18
Rv1201c	0.05109496	-0.3320082	0.01385384	dapD
Rv1218c	-0.3931014	0.05336158	0.02264135	Rv1218c

Rv1221	-0.0277336	0.42058006	0.00448058	sigE
Rv1223	-0.2170304	0.02432585	0.02169518	htrA
Rv1240	0.3026345	-0.127774	0.00116406	mdh
Rv1258c	0.34152868	-0.1747918	0.00180995	Rv1258c
Rv1278	0.34106645	-0.0597825	0.00466983	Rv1278
Rv1285	-0.0825731	0.31290276	0.03315383	cysD
Rv1294	0.213564	-0.0571078	0.02982671	thrA
Rv1296	0.09274542	-0.2430027	0.03565216	thrB
Rv1335	-0.1680453	0.19694061	0.02182745	cysO
Rv1377c	0.65861618	-0.0562921	0.00054952	Rv1377c
Rv1396c	-0.1940299	0.26242246	0.00107637	PE_PGRS25
Rv1397c	-0.0615575	0.53776231	1.758E-05	vapC10
Rv1406	0.24608661	-0.2113027	0.00712967	fmt
Rv1425	-0.2832982	0.0367643	0.04648755	Rv1425
Rv1430	0.3617762	-0.0207161	0.03315383	PE16
Rv1456c	0.136659	-0.2536429	0.02896445	Rv1456c
Rv1457c	0.16934329	-0.2888033	0.00604513	Rv1457c
Rv1477	-0.4541193	0.4709383	4.1968E-12	ripA
Rv1478	-0.4400808	0.18000972	2.4288E-05	Rv1478
Rv1479	-0.0092253	0.26935881	0.02104979	moxR1
Rv1484	0.04891849	-0.2386905	0.03700451	inhA
Rv1497	-0.2594026	0.37000932	2.5488E-07	lipL
Rv1527c	-0.0268244	0.31343394	0.01340894	pks5
Rv1547	0.07183416	-0.4621478	1.5043E-05	dnaE1
Rv1563c	-0.1280313	0.2203544	0.02547367	treY
Rv1564c	-0.0363399	0.29834529	0.02349788	treX
Rv1565c	-0.1671526	0.30547002	4.2256E-05	Rv1565c
Rv1566c	-0.1743438	0.26532152	0.00143784	Rv1566c
Rv1607	0.15911609	-0.2348223	0.01252725	chaA
Rv1617	0.01480766	-0.443696	0.009991	pykA
Rv1618	0.24867161	-0.2387448	0.00044768	tesB1
Rv1619	0.09127585	-0.2500586	0.03991062	Rv1619
Rv1620c	-0.2310313	0.09046464	0.04828703	cydC
Rv1697	-0.2229207	0.17286091	0.00107637	Rv1697
Rv1704c	0.2143951	-0.1095352	0.04458244	cycA
Rv1728c	-0.0195224	0.35853359	0.03534823	Rv1728c
Rv1731	-0.1340095	0.20741342	0.02274402	gabD2
Rv1754c	-0.0913379	0.33489631	0.01921692	Rv1754c

Rv1772	-0.4578039	0.3826324	7.575E-08	Rv1772
Rv1774	-0.221858	0.11677796	0.04309371	Rv1774
Rv1809	-0.2679201	0.16430472	0.00770578	PPE33
Rv1810	-0.0938381	0.47209083	0.00046834	Rv1810
Rv1812c	0.03214438	-0.4315582	0.01921692	Rv1812c
Rv1816	-0.0600993	0.59759034	2.6775E-06	Rv1816
Rv1830	-0.0264725	0.28586331	0.02169518	Rv1830
Rv1836c	-0.1031857	0.19562631	0.00810533	Rv1836c
Rv1854c	0.15361282	-0.1326368	0.01769602	ndh
Rv1869c	-0.050313	0.19770243	0.02714414	Rv1869c
Rv1870c	-0.0570941	0.43997259	3.3297E-06	Rv1870c
Rv1871c	-0.1537893	0.2060115	0.00012931	Rv1871c
Rv1872c	-0.2289749	0.09928945	0.00978832	lldD2
Rv1882c	-0.0598818	0.45031201	0.00011978	Rv1882c
Rv1900c	0.04116267	-0.2673001	0.04647182	lipJ
Rv1905c	-0.2794274	0.092021	0.00184196	aao
Rv1906c	-0.1406654	0.23657448	0.00431988	Rv1906c
Rv1927	-0.2821501	0.09374508	0.00957398	Rv1927
Rv1987	-0.2450595	0.36408012	3.5498E-05	Rv1987
Rv2009	-0.142646	0.2833268	0.00957398	vapB15
Rv2010	-0.2543305	0.19551532	0.00070686	vapC15
Rv2038c	-0.2491469	0.21862389	0.00727134	Rv2038c
Rv2052c	-0.341009	0.13760387	0.00012394	Rv2052c
Rv2062c	0.0410207	-0.5277737	0.00011923	cobN
Rv2069	0.14666271	-0.1872961	0.04485534	sigC
Rv2091c	-0.0938512	0.18166506	0.01822285	Rv2091c
Rv2108	0.71103685	-0.1349121	7.0653E-05	PPE36
Rv2113	0.08211312	-0.2837382	0.04971655	Rv2113
Rv2115c	-0.2226929	0.29502273	1.8488E-05	mpa
Rv2135c	-0.2278782	0.15503431	0.01036748	Rv2135c
Rv2136c	-0.2312439	0.1413437	0.0024937	Rv2136c
Rv2144c	-0.051295	0.53160483	3.0315E-05	Rv2144c
Rv2145c	-0.2901066	0.31389386	2.8145E-08	wag31
Rv2147c	-0.1415701	0.17987472	0.01252725	Rv2147c
Rv2151c	-0.2762441	0.04253043	0.02198737	ftsQ
Rv2153c	-0.2839296	0.14920665	0.00234129	murG
Rv2154c	-0.2578333	0.06079198	0.02220912	ftsW
Rv2161c	-0.1960195	0.12514627	0.02714414	Rv2161c

Rv2214c	0.14715251	-0.2430765	0.00210991	ephD
Rv2257c	0.06483805	-0.3345078	0.00892111	Rv2257c
Rv2258c	0.17742024	-0.3225137	0.0001307	Rv2258c
Rv2275	0.28905373	-0.3287295	3.8521E-05	Rv2275
Rv2276	0.19132088	-0.4563916	7.8046E-05	cyp121
Rv2293c	-0.1277934	0.36888863	0.00438728	Rv2293c
Rv2295	-0.0973257	0.33345549	0.00208219	Rv2295
Rv2327	-0.0802282	0.43988149	0.00056849	Rv2327
Rv2331A	-0.0192435	0.77611368	5.8153E-07	Rv2331A
Rv2335	-0.3019288	0.07247644	0.01095865	cysE
Rv2349c	-0.3894319	0.10008727	0.00052557	plcC
Rv2350c	-0.1029382	0.21757656	0.01900757	plcB
Rv2366c	-0.252956	0.01291165	0.0323255	Rv2366c
Rv2368c	-0.1249164	0.15945871	0.01622703	phoH1
Rv2380c	-0.5291727	0.14787967	7.885E-06	mbtE
Rv2382c	-0.0641308	0.37823679	0.02166158	mbtC
Rv2391	-0.0446819	0.34399776	0.00170321	sirA
Rv2392	-0.1283384	0.24252421	0.01149402	cysH
Rv2393	-0.3583439	0.13175503	0.00065876	che1
Rv2395A	-0.2506178	0.24232124	0.00011486	aprA
Rv2395B	-0.492553	0.09597858	6.3648E-06	aprB
Rv2396	-0.2756607	0.0582754	0.02132072	PE_PGRS41
Rv2406c	-0.0220789	0.34175922	0.02098973	Rv2406c
Rv2450c	0.09474791	-0.4437639	0.00321958	rpfE
Rv2467	0.10351141	-0.2437041	0.02538416	pepN
Rv2470	-0.3540626	0.00169865	0.028222	glbO
Rv2471	-0.2711316	0.10971139	0.00278389	aglA
Rv2477c	0.18504299	-0.3508727	2.1357E-06	Rv2477c
Rv2506	-0.2841446	0.11276767	0.02169518	Rv2506
Rv2510c	0.01290442	-0.3822583	0.02676394	Rv2510c
Rv2518c	0.12491369	-0.3605142	0.00205525	ldtB
Rv2519	0.2025703	-0.1234542	0.01443191	PE26
Rv2522c	0.03524754	-0.3464619	0.04433655	Rv2522c
Rv2526	-0.0702951	0.39390892	0.01270185	vapB17
Rv2527	-0.2708162	0.16729739	0.01034093	vapC17
Rv2572c	0.07682968	-0.3479768	0.01584703	aspS
Rv2590	-0.2571554	0.37106203	6.5796E-08	fadD9
Rv2591	-0.3271469	0.30307926	4.4675E-06	PE_PGRS44

Rv2595	-0.1755115	0.18205954	0.03358174	vapB40
Rv2596	-0.365436	0.09315593	0.00712967	vapC40
Rv2629	-0.1736229	0.20950788	0.00294446	Rv2629
Rv2700	0.19058924	-0.1697147	0.03602642	Rv2700
Rv2718c	0.12279237	-0.2250511	0.03326966	nrdR
Rv2721c	-0.1524491	0.45542219	2.5047E-09	Rv2721c
Rv2746c	-0.1048052	0.25917673	0.02897613	pgsA3
Rv2756c	0.16417152	-0.1488484	0.0334252	hsdM
Rv2790c	0.25772354	-0.0957609	0.00680199	ltp1
Rv2791c	0.29133323	-0.0280446	0.01720207	Rv2791c
Rv2818c	-0.2008953	0.18690952	0.02538416	Rv2818c
Rv2838c	-0.1828608	0.18258876	0.01393776	rbfA
Rv2905	-0.3352879	0.16518814	6.3648E-06	lppW
Rv2918c	-0.1372631	0.30420219	0.00056849	glnD
Rv2926c	-0.2010606	0.10288465	0.02160028	Rv2926c
Rv2928	0.37011751	-0.0446424	0.00237032	tesA
Rv2929	0.09541321	-0.3670891	0.00628414	Rv2929
Rv2936	0.32931217	-0.0127065	0.0098889	drrA
Rv2937	0.20891848	-0.1365363	0.01443191	drrB
Rv2938	0.00263219	-0.2941492	0.03534823	drrC
Rv2940c	0.15339366	-0.091781	0.02581984	mas
Rv2941	-0.0487948	0.28529943	0.01006749	fadD28
Rv2960c	0.52589409	-0.0681949	0.00374079	Rv2960c
Rv3060c	-0.3149148	0.04199527	0.01247447	Rv3060c
Rv3062	0.32090732	-0.0411015	0.04402178	ligB
Rv3083	-0.0659629	0.30628443	0.04633773	Rv3083
Rv3098A	-0.1370071	0.20364225	0.03427604	Rv3098A
Rv3136	0.04241341	-0.5439807	2.2688E-05	PPE51
Rv3139	-0.4023736	0.02891362	0.00202089	fadE24
Rv3151	-0.2825422	0.00498689	0.03693861	nuoG
Rv3197	-0.4172279	0.03761763	5.1001E-06	Rv3197
Rv3207c	-0.2781589	0.04592402	0.01040663	Rv3207c
Rv3209	-0.4102535	0.06219425	0.00020005	Rv3209
Rv3211	-0.0283726	0.21546765	0.04023597	rhlE
Rv3229c	-1.0055185	0.52426441	3.0776E-16	desA3
Rv3230c	-0.4359928	0.17351204	3.1519E-05	Rv3230c
Rv3241c	-0.243995	0.16996458	0.00445113	Rv3241c
Rv3244c	-0.1602027	0.12224198	0.04864999	lpqB

Rv3260c	-0.3244903	0.28377086	4.6936E-08	whiB2
Rv3283	-0.231878	0.13385037	0.0031859	sseA
Rv3284	-0.2434182	0.10406241	0.00885839	Rv3284
Rv3303c	0.01696799	-0.3407788	0.04745989	lpdA
Rv3312A	-0.0349148	0.36677409	0.03218869	Rv3312A
Rv3312b	-0.2945124	0.08912602	0.0337756	Rv3312b
Rv3320c	-0.1816849	0.25151804	0.00340173	vapC44
Rv3389c	0.1918571	-0.2593107	0.00064927	htdY
Rv3399	-0.0594933	0.40591059	0.01623781	Rv3399
Rv3408	-0.2501183	0.23067124	0.00025948	vapC47
Rv3413c	-0.2658974	0.45391533	8.0553E-09	Rv3413c
Rv3435c	0.18187246	-0.1826426	0.01095865	Rv3435c
Rv3436c	0.01678307	-0.4304612	0.00466983	glmS
Rv3485c	-0.1844005	0.14213338	0.03534823	Rv3485c
Rv3487c	-0.4192588	0.21017586	4.7554E-07	lipF
Rv3490	-0.2099386	0.05892483	0.02250622	otsA
Rv3550	0.38929188	-0.1170324	0.00242722	echA20
Rv3556c	0.45962786	-0.0075267	0.00039517	fadA6
Rv3561	0.44218745	-0.0084253	0.00294949	fadD3
Rv3598c	0.13573437	-0.1976515	0.03602642	lysS
Rv3646c	0.07458034	-0.4032038	2.2602E-05	topA
Rv3659c	0.20552722	-0.193737	0.02182745	Rv3659c
Rv3685c	-0.20313	0.10019227	0.0189516	cyp137
Rv3708c	0.17533111	-0.1165589	0.02654012	asd
Rv3750c	-0.1397393	0.24411209	0.00680933	Rv3750c
Rv3767c	-0.126003	0.24282902	0.01004288	Rv3767c
Rv3777	0.25404296	-0.1802119	0.00409611	Rv3777
Rv3785	-0.1287589	0.20363239	0.03244471	Rv3785
Rv3804c	0.26536427	-0.147899	0.00012109	fbpA
Rv3810	-0.7366587	0.19236403	7.0395E-11	pirG
Rv3811	-0.2718058	0.12182528	0.00225373	Rv3811
Rv3818	0.1298649	-0.263488	0.00361681	Rv3818
Rv3822	-0.0195601	0.58144296	1.1539E-07	Rv3822
Rv3863	-0.2486748	0.3658098	1.01E-06	Rv3863
Rv3866	-0.1680249	0.14920127	0.02250622	espG1
Rv3877	-0.0659586	0.2371954	0.01647338	eccD1
Rv3878	-0.0563625	0.273852	0.03189715	espJ
Rv3879c	-0.0747431	0.23993149	0.00497486	espK

Rv3880c	-0.0610827	0.33335702	0.00201107	espL
Rv3919c	0.08801953	-0.23922	0.0251576	gid
Rv3920c	0.28418384	-0.0773459	0.01643251	Rv3920c
Rv3921c	0.23514132	-0.1869885	7.7899E-05	Rv3921c
Rv3922c	0.43167919	-0.0161618	0.00105343	Rv3922c
Rv3923c	0.11549234	-0.2078556	0.03656983	rnpA
RVnc0024	-0.2765425	0.23331233	4.5496E-05	mcr7
RVnc0036a	0.01516153	-0.3529692	0.00630148	MTS2823
Rvnr01	0.23001657	-0.1275696	0.0334252	rrs
Rvnr03	0.05843421	-0.7434419	7.4387E-06	rrf