Drug Resistance Amplification in Tuberculosis: Integrating Drug Sensitivity Testing with Whole Genome Sequencing

Drug Resistance Amplification in Tuberculosis: Integrating Drug Sensitivity Testing with Whole Genome Sequencing

Catherine Derner

A Thesis in the Field of Biology

for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

March 2024

Abstract

Tuberculosis (TB), a disease caused by *Mycobacterium tuberculosis* (*Mtb*), remains a global health threat, presenting significant challenges to effective treatment due to the emergence of drug-resistant strains. The prevalence and mechanisms of drug resistance amplification in TB remain areas of ongoing research and debate.

This study investigated the consequences of suboptimal treatment on the acquisition of drug resistance mutations in *Mtb*. We sought to assess the frequency of amplification while identifying potential risk factors; in particular, our goal was to understand the true prevalence of resistance amplification. Utilizing patient data collected during a comprehensive TB transmissibility and treatment study in Peru, we examined changes in drug susceptibility profiles through conventional drug sensitivity testing (phenotypic) and identifying genotypic changes via whole-genome sequencing. This dual examination provides a comprehensive understanding of drug resistance dynamics, going beyond conventional phenotypic assessments to explore the genomic landscape.

We investigated various factors that we hypothesized might contribute to suboptimal treatment outcomes, as suboptimal treatment has been associated with resistance amplification. Specifically, we examined variables such as HIV status, diabetes status, the presence of cavitary lesions, *Mtb* lineage, time to effective treatment, and adherence to treatment regimens.

While logistic regression analysis did not reveal statistically significant relationships between suboptimal treatment and resistance amplification, our study

provides valuable insight into the acquisition of drug-resistance amplification. It underscores the complexity of TB treatment outcomes and emphasizes the need for further investigation into the factors contributing to drug-resistance amplification. Our main finding was that drug-resistance amplification is rare. Thus, relying solely on phenotypic drug susceptibility testing (DST) data may lead to an overestimation of resistance amplification. This finding highlights the importance of considering the limitations of phenotypic DST data and the potential benefits of integrating advanced molecular techniques such as whole genome sequencing to gain a more accurate understanding of drug resistance dynamics in TB patients.

Acknowledgments

I am immensely grateful for the invaluable support and guidance I received during the journey of completing this thesis. I would like to express my heartfelt appreciation to the following individuals:

I extend my deepest gratitude to my thesis director, Dr. Megan Murray, and her dedicated team. Dr. Murray, your mentorship and expertise have been instrumental in shaping this research, and your support has been the cornerstone of this endeavor. I'd also like to specifically acknowledge Chuan Chin Huang and Zibiao Zhang for their willingness to go above and beyond in acquiring data and assisting me with challenges, which made this work possible.

To my advising team, Dr. James Morris and Dr. Joan Short, I am thankful for your wisdom, encouragement, and scholarly guidance. Your insights have been crucial in steering my research in the right direction.

I must acknowledge my family, without whom this journey would not have been possible. My husband, Phil, has been my unwavering support throughout this journey, constantly believing in me and providing the strength to persevere through the most challenging times. To my three boys, Zach, Hunter, and Luke, you were my inspiration and motivation to undertake this endeavor. Your love and encouragement were my driving force.

I am also grateful for the opportunities and resources provided by Harvard Extension School, which played a vital role in enabling this academic journey.

Table of Contents

List of Tables

List of Figures

Chapter I.

Introduction

Tuberculosis (TB) is an infectious disease caused by the bacillus *Mycobacterium tuberculosis* (*Mtb*). While the disease primarily affects the pulmonary system, it can also affect other sites in the body, such as the gastrointestinal (GI) system, the skin, the central nervous system, and the liver.[1] Pulmonary disease is the most common presentation, characterized by symptoms such as persistent coughing, chest pain, and hemoptysis (coughing up blood).[1] TB is highly contagious and spreads when actively infected individuals expel bacteria into the air, often through coughing.[1]

TB is a leading cause of infectious deaths, causing an estimated 1.4 million deaths worldwide, annually. [2] Approximately one quarter of the global population is estimated to have been infected with *Mtb*, and each year, roughly 10.6 million infected individuals develop active TB.[2] Tuberculosis surpasses HIV/AIDS as the leading cause of death from a single infectious agent.[2]

One of the most pressing challenges in global TB control is the ongoing spread of drug-resistant tuberculosis (DR-TB). Particularly, multidrug-resistant tuberculosis (MDR-TB) has emerged as a growing public health challenge due to the intrinsic

---

[1] Adigun R, Singh R. Tuberculosis. [Updated 2023 Jul 11]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK441916/

[2] World Health Organization (WHO). (2022) Global tuberculosis report 2022. World Health Organization, Geneva, Switzerland. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022

difficulty in treating these infections. [3] The presence of drug resistance increases the risk of adverse outcomes among TB patients, such as treatment failure and death. [4] Each year, nearly half a million new cases of MDR-TB are reported worldwide.[2]

MDR-TB strains exhibit resistance to the two most effective first-line drugs, isoniazid and rifampicin.[4] Extensively drug-resistant (XDR) strains are resistant not only to these first-line drugs but also to a fluoroquinolone and either a second-line injectable (amikacin, capreomycin, and kanamycin) or the drugs bedaquiline or linezolid.[4] The more drugs a strain of TB is resistant to, the more challenging it becomes to achieve effective and complete treatment.[4] For this reason, timely and effective treatment is crucial for stopping the disease's spread and preventing the emergence of drug-resistant strains.

Treatment for drug-susceptible TB requires a standard 4-drug antibiotic regimen lasting approximately 6 months.[4] When correctly prescribed and completed, this regimen can lead to cure rates as high as 95% when correctly administered to people with drug-susceptible TB. [5]

Treatment for DR-TB involves the utilization of first-line drugs that the patient's TB strain remains susceptible to, in addition to second-line drugs.[4] This treatment can vary in duration, lasting between six months to two years.[4] Drug resistance in *Mtb* is linked to lower treatment success rates. XDR-TB patients have been shown to face a

---

[3] Centers for Disease Control and Prevention. (2021). Tuberculosis. Division of Tuberculosis Elimination. Centers for Disease Control and Prevention. https://www.cdc.gov/tb/publications/default.htm

[4] Dheda, K., Gumbo, T., Gandhi, N. R., Murray, M., Theron, G., Udwadia, Z., Migliori, G. B., & Warren, R. (2014). Global control of tuberculosis: from extensively drug-resistant to untreatable tuberculosis. *The* Lancet. Respiratory medicine, 2(4), 321–338. https://doi.org/10.1016/S2213-2600(14)70031-1

[5] Shin, H. J., & Kwon, Y. S. (2015). Treatment of Drug Susceptible Pulmonary Tuberculosis. *Tuberculosis and Respiratory Diseases*, 78(3), 161–167. https://doi.org/10.4046/trd.2015.78.3.161

12.2-fold higher likelihood of experiencing an unsuccessful treatment outcome compared to MDR-TB patients.[6] Additionally, with each reduction in the number of potentially effective drugs in the treatment regimen, the odds of achieving treatment success has been shown to be decreased by a factor of 0.62. Conversely, for each additional effective drug included in the treatment regimen, the odds of treatment success are increased by a factor of 2.1.[7] Thus, increased drug resistance negatively impacts treatment success, whereas the inclusion of more effective drugs in the treatment plan improves the chances of a successful outcome.

Cases of DR-TB more frequently occur by transmission of drug-resistant *Mtb* strains from person to person, compared with the development of new resistance.[4] However, the acquisition of new drug resistance in response to inadequate treatment is more likely to occur through the selection of existing mutations rather than through the emergence of new spontaneous mutations in drug-resistance genes. This process of selection can be driven by the pressure imposed by minimally-effective or ineffective drugs, reinforcing the prevalence of mutations in genes associated with drug resistance.[8]

In the context of drug resistant TB, the differentiation between acquired resistance mutation vs primary transmission is a concern. XDR-TB can develop from MDR-TB

---

[6] Bhering, M., & Kritski, A. (2020). Primary and acquired multidrug-resistant tuberculosis: Predictive factors for unfavorable treatment outcomes in Rio de Janeiro, 2000-2016. *Revista panamericana de salud publica = Pan American Journal of Public Health*, 44, e178. https://doi.org/10.26633/RPSP.2020.178

[7] Cegielski, J. P., Dalton, T., Yagui, M., Wattanaamornkiet, W., Volchenkov, G. V., Via, L. E., Van Der Walt, M., Tupasi, T., Smith, S. E., Odendaal, R., Leimane, V., Kvasnovsky, C., Kuznetsova, T., Kurbatova, E., Kummik, T., Kuksa, L., Kliiman, K., Kiryanova, E. V., Kim, H., Kim, C. K., … Global Preserving Effective TB Treatment Study (PETTS) Investigators (2014). Extensive drug resistance acquired during treatment of multidrug-resistant tuberculosis. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 59(8), 1049–1063. https://doi.org/10.1093/cid/ciu572

[8] Shaikh, A., Sriraman, K., Vaswani, S., Oswal, V., Rao, S., & Mistry, N. (2021). Early phase of effective treatment induces distinct transcriptional changes in *Mycobacterium tuberculosis* expelled by pulmonary tuberculosis patients. *Scientific reports*, 11(1), 17812. https://doi.org/10.1038/s41598-021-96902-7

through a small number of additional resistance mutations that occur in patients undergoing drug treatment. In a study conducted in 2021 involving 286 MDR-TB patients, 63 patients developed additional resistance to both second-line injectable drugs and fluoroquinolones while undergoing standardized MDR-TB treatment.[9] Exogenous re-infection with a secondary drug-resistant strain was more common (65.1%) in these patients than the occurrence of acquired drug resistance through microevolution (28.6%). Furthermore, the study identified independent risk factors for the development of additional second-line drug resistance, with extensive disease on chest X-ray and the presence of type 2 diabetes mellitus being notable contributors.[9]

In pathogenic bacteria, whole genome sequencing (WGS) is able to capture genetic variation that can lead to drug resistance through the identification of single nucleotide polymorphisms (SNPs) in the bacterial genome. Through WGS, we are able to uncover SNPs that occur during various treatment scenarios and evaluate how these mutations may confer drug resistance to the pathogen.

In this study, our objective was to investigate whether drug resistance in *Mtb* develops through the acquisition of drug-resistance mutations during exposure to antimicrobial drugs and, if so, to quantify the extent to which it occurs. We aimed to shed light on the relationship between suboptimal treatment, often due to delays in DR-TB diagnosis, and the risk of acquiring resistance to second-line drugs. Our study's focus was on discerning how suboptimal treatment in DR-TB patients affects SNP acquisition,

[9] Hu, Y., Zheng, X., Davies Forsman, L., Ning, Z., Chen, C., Gao, Y., Zhang, Z., Lu, W., Werngren, J., Bruchfeld, J., Hoffner, S., & Xu, B. (2021). Emergence of additional drug resistance during treatment of multidrug-resistant tuberculosis in China: a prospective cohort study. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 27(12), 1805–1813. https://doi.org/10.1016/j.cmi.2021.04.001

and how SNPs in various regions of the *Mtb* genome contribute to the amplification of drug resistance.

We hypothesized that suboptimal treatment is a risk factor for drug resistance amplification and we anticipated that patients who received suboptimal treatment would exhibit increased SNP acquisition compared to those who received the most optimal treatment. Further, we expected that SNPs acquired in regions of the genome known to be associated with drug-resistance phenotypes would correlate with amplification of drug resistance.

Chapter II.

Materials and Methods

In this study, our objective was to gain insights into the true prevalence of resistance amplification within the context of TB treatment. Combining data from drug susceptibility testing and whole genome sequencing obtained from a prospective cohort study of TB patients, we systematically identified phenotypic and genotypic changes indicative of amplified drug resistance. Following this, we conducted statistical analyses, including univariate and multivariate logistic regression, to explore associations between various clinical variables and the likelihood of drug resistance amplification in TB patients. This analysis aimed to establish whether these identified changes genuinely indicated drug resistance amplification.

Participants

This study used patient data obtained during a prospective TB transmissibility and treatment study conducted in Lima, Peru, between 2009 and 2012. [10] The study cohort initially included 4,500 index patients, of whom 4,044 had microbiologically confirmed tuberculosis disease.

[10] Becerra, M. C., Huang, C. C., Lecca, L., Bayona, J., Contreras, C., Calderon, R., Yataco, R., Galea, J., Zhang, Z., Atwood, S., Cohen, T., Mitnick, C. D., Farmer, P., & Murray, M. (2019). Transmissibility and potential for disease progression of drug resistant *Mycobacterium tuberculosis*: prospective cohort study. *BMJ (Clinical research ed.)*, 367, l5894. https://doi.org/10.1136/bmj.l5894

Ethics statement

This research utilized previously collected data from human subjects who provided written informed consent for their data to be analyzed by the team of Dr. Megan Murray for research purposes. No new samples or personally-identifiable data were collected for this project. The dataset provided did not contain any identifiers, ensuring that individuals could not be re-identified through the data. IRB review was conducted for the main study; however, as this current study did not involve new human-subject research, additional IRB review was not required.

Measures

For index patients, pulmonary TB disease was diagnosed based on positive sputum smears or chest radiographs. Sputum samples collected at diagnosis and follow-up time points underwent bacteriological culture to assess treatment response. Baseline drug sensitivity testing targeted first-line drugs, expanding to second-line drugs upon detecting resistance. Throughout treatment, patients were regularly monitored, with routine cultures at the two-month mark. If persistent culture positivity was observed, additional drug sensitivity testing was administered in a predefined order, facilitating comprehensive resistance assessment.

At initiation, all samples from cultures that tested positive for *Mtb* were subjected to a baseline drug sensitivity test (DST), assessing resistance to first-line drugs: Pyrazinamide (pza), Isoniazid (inh), Ethambutol (eth), and Rifampin (rif). If resistance to any of these drugs was detected, further testing was conducted to assess resistance to second-line drugs: Streptomycin (sm), Para-aminosalicylic acid (pas), Capreomycin (cm), Ciprofloxacin (cpx), Cycloserine (cs), Ethionamide (eth), Kanamycin (km), Rifabutin

(rbu), Levofloxacin (lin), Moxifloxacin (mox), and Amikacin (amk). Whole genome sequencing was performed on a subset of Mtb isolates.

Patients diagnosed with drug-sensitive TB received a standard 6-month treatment course, consisting of a two-month phase involving isoniazid, Rifampicin, Pyrazinamide, and Ethambutol, followed by a four-month phase of isoniazid and Rifampicin alone. Patients with MDR-TB were initiated on a standard treatment regimen until drug resistance was confirmed. Delays in initiating appropriate treatment, specifically involving second-line drugs, for MDR-TB were due to the extended time required for drug resistance testing by culture, which could take several months. Follow-up data were collected at two, six, 12, and 24 months for all patients. Patients with DR-TB had additional follow-up at 36 or 48 months.

Procedures

To identify drug resistance amplification in patients, we used the process illustrated in Figure 1. This approach involved the systematic exclusion of cases that were not indicative of amplification. Our analysis encompassed both DST results to identify patients with phenotypic drug resistance changes and WGS results to pinpoint individuals exhibiting genotypic changes indicative of amplified drug resistance

We took a stepwise exclusion approach to classify participants as “amplified.” From the initial 353 index patients, 189 individuals whose DST results showed no change in resistance during treatment were eliminated. Fourteen participants with only one genomic file were also excluded, as two or more files were required for comparative analysis. Additional participants were eliminated if there was evidence of re-infection or

mixed infection, including 22 who were determined to exhibit *Mtb* lineage changes during treatment. After we analyzed SNPs in the remaining participants, we then eliminated 15 participants whose SNPs were not synonymous or were not unique between tests and 12 participants who had more than 50 SNP changes, a threshold we set to signify potential reinfection or mixed infection. Finally, 80 participants whose SNPs were in genes not associated with resistance were eliminated. The remaining 11 participants were categorized as amplified, while the 342 participants who were excluded were classified as non-amplified.

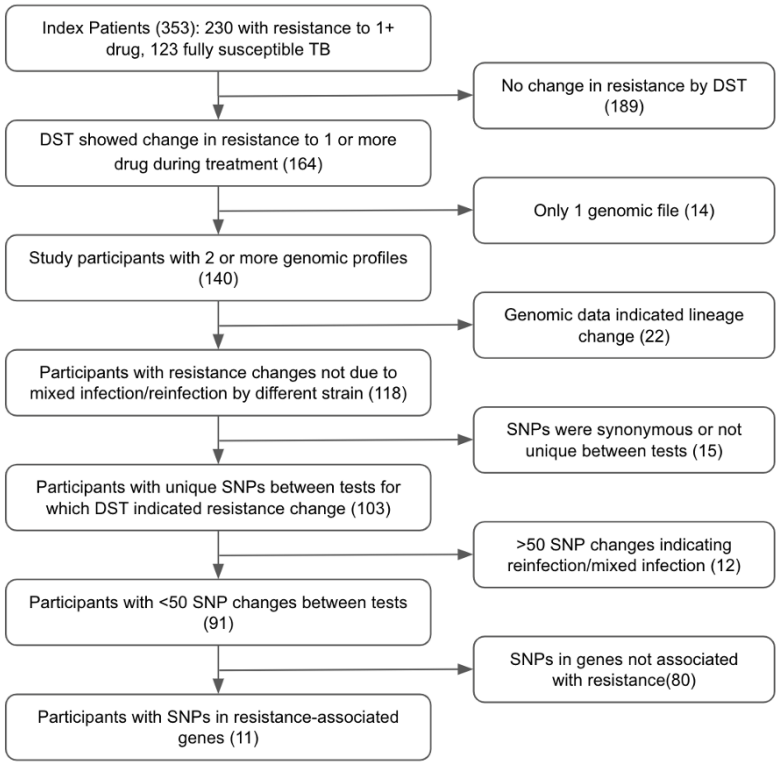

Figure 1. Process for Identifying Amplified Drug Resistance.

Criteria and analyses employed to pinpoint individuals exhibiting amplified drug resistance.

Analysis of drug susceptibility testing results

We included a subgroup of 353 index patients with DST results with at least two sequential positive sputum cultures. We conducted descriptive statistics for this group, particularly focusing on the drugs tested for resistance, and examined the number of individuals who exhibited either fully susceptible or drug-resistant TB at enrollment.

Subsequently, we identified patients whose DST results changed over time, focusing on samples that exhibited increased drug resistance or showed a transition from susceptibility to resistance during treatment. We identified when these changes occurred, noting whether transitions were between samples 1 and 2, 2 and 3, and so on. 15 patients had no result for test 1, but DST results that indicated resistance in test 2. This can likely be attributed to samples being compromised either before or during testing, leading to the inability to produce conclusive results.

Analysis of genomic data

To confirm the resistance changes observed through DST, we conducted an analysis of SNP changes identified through WGS data. The SNP calls for each sample were provided by the team of Dr. Megan Murray. Participants without two or more genomic files were omitted since comparisons could not be made in such cases.

We used lineage data to eliminate individuals whose genomic data indicated lineage changes. Such changes suggest potential reinfection or coinfection, and therefore, these individuals were excluded from our analysis.

Next, our objective was to distinguish whether individuals had actually acquired drug resistance or if their infections were instead the result of mixed infection or new strain infections. We concentrated on non-synonymous SNPs (coding SNPs) and retained

only those in coding regions. We excluded SNPs that were located within repeating elements because these elements can be duplicated throughout the genome, making it challenging to accurately identify their exact location. This exclusion ensured that our analysis remained focused on more reliably annotated and interpretable genomic regions.

Following this, SNP calls against a reference genome were made by the team, and we analyzed them here to identify instances of new, unique SNPs emerging between sequential tests. These unique SNPs were defined based on their distinct genomic positions and/or sequence alterations. To clarify, a SNP was considered unique if it appeared in one test but had not been present in the previous test. If the SNP occurred at the same genomic position but exhibited a different sequence alteration between tests, it was also classified as unique. In contrast, SNPs that repeated between tests without alterations or with identical alterations were not considered new and were excluded from our analysis. This approach allowed us to identify and track the development of novel SNPs that arose over the course of treatment.

We applied a threshold for exclusion, removing individuals who exhibited more than 50 SNP changes between consecutive tests. This threshold was indicative of substantial genetic variation that raised the possibility of mixed infection or reinfection with a new TB strain, which would make it challenging to definitively attribute drug resistance amplification solely to suboptimal treatment.

Using the remaining SNPs, we compiled a list of genes in which we identified SNPs associated with amplification. In order to identify the genes known to be associated with resistance and filter out irrelevant ones, we consulted the 'Catalogue of mutations in *Mycobacterium tuberculosis* complex and their association with drug resistance'

published by the World Health Organization. [11] From the variant names in the document, we compiled a list of genes and drugs associated with resistance. We retained genes from the WHO dataset labeled as 'Associated with Resistance' ('Assoc w R'), 'Associated with Resistance - Interim' ('Assoc w R - Interim'), or 'Combination' ('Combo') and excluded those labeled as 'Not Associated with Resistance' ('Not assoc w R'), 'Not Associated with Resistance - Interim' ('Not assoc w R - Interim'), or 'Uncertain Significance.' The WHO dataset contained numerous entries with "NA" in the “genome.position” column; these entries were omitted.

Through a comparison of the gene list derived from the WHO dataset and our own list of genes containing SNPs, we established a final list of genes potentially associated with drug resistance. Subsequently, we identified individuals in our cohort with SNP changes in these genes. This process enabled us to categorize participants into one of two outcomes: amplified or not amplified. It also offered insights into the frequency of amplification.

Statistical analysis

To explore the relationship between various independent variables and the likelihood of amplification, we conducted a series of univariate logistic regression analyses, followed by multivariate logistic regression analysis. The primary dependent variable in our analyses was the amplification status: 'amplified' or 'not amplified.’ The

[11] WHO Catalogue of mutations in *Mycobacterium tuberculosis* complex and their association with drug resistance. 2021. https://apps.who.int/iris/handle/10665/341981

independent variables of interest were:

- HIV Status: A binary variable indicating the presence or absence of HIV infection. Information on HIV status was obtained from a questionnaire completed by index patients at enrollment. Patients who did not know their HIV status had blood drawn for HIV and cluster of differentiation 4 (CD4) count.
- Diabetes Status: A binary variable indicating the presence or absence of diabetes. Diabetes status was self-reported by patients in the enrollment questionnaire.
- Cavitary Lesions: A binary variable representing the presence or absence of cavitary lesions. Cavitary lesions were identified by chest radiography.
- *Mtb* Lineage: A categorical variable representing different *M. tuberculosis* lineage groups. These include lineage2 and lineage4.
- Time to Effective Treatment: A continuous variable measured in days, signifying the duration from the initiation of treatment to the point when it became tailored to the patient's DST results.
- Treatment Adherence: A continuous variable representing the degree of treatment adherence, measured on a scale from 0 to 100. This scale reflects the proportion of prescribed doses actually taken by patients during the relevant time period. Higher values indicate a higher percentage of adherence to the prescribed treatment regimen.

Prior to conducting the logistic regression analysis, data cleaning and preprocessing were performed to address missing values. Participants with missing values were temporarily excluded from specific univariate analyses relevant to the

missing data. For instance, if there were five individuals with missing HIV data, they were excluded from the HIV analysis but were included in other univariate analyses like cavitary lesions. However, all individuals with missing data were ultimately excluded from the multivariate analysis to ensure a complete dataset for this comprehensive analysis.

To minimize the risk of reverse causality, in which the presumed cause-and-effect relationships could be reversed, we implemented a cutoff point for the 'time to effective treatment' variable, capping its values at 60 days. Cases where values exceeded this threshold were standardized to 60 days for analytical purposes. Beyond 60 days, the likelihood of resistance changes may become less pertinent to our research objectives. It also aligns with the critical timeframe for obtaining drug susceptibility testing results, during which decisions about treatment adjustments are typically made.

We pooled all of the sublineages within lineage4 before comparing them with lineage2. We used lineage2 as the reference category for this categorical variable. The coefficients for lineage4 are interpreted in relation to this chosen baseline, with the reference category assigned a value of 1.

Each logistic regression analysis was conducted using the 'glm' function in the R programming environment, specifying 'binomial' as the family parameter. This approach allowed us to model the probability of amplification status as a function of the independent variables mentioned above, including HIV status, diabetes status, presence of cavitary lesions, *Mtb* lineage, time to effective treatment, and adherence to treatment regimens.

In our initial analysis, each variable was individually examined in a univariate logistic regression analysis, assessing its potential association with drug resistance amplification. We then proceeded to conduct a multivariate logistic regression analysis, where all variables were included simultaneously to explore potential interactions and dependencies among them.

For these analyses, we applied standard statistical significance thresholds ($p < 0.05$) and reported the corresponding coefficient estimates, odds ratios, standard errors, and p-values for each predictor variable.

Chapter III.

Results

The main objectives of this study were to investigate the consequences of suboptimal treatment on the acquisition of drug resistance mutations in *Mtb*, assess the frequency of amplification, and identify potential risk factors. Here, we present the comprehensive results of our investigation, encompassing cohort descriptive statistics, phenotypic and genotypic amplification findings, and a detailed examination of independent variables related to study participants, including HIV status, diabetes status, cavitary lesions, *Mtb* lineage, adherence, and time to effective treatment.

Descriptive Statistics

Of the initial 4,500 index patients from the Peru study, we identified 353 participants who remained culture positive 2 months or more after the initiation of TB treatment. Among this subset, 119 individuals (33.7%) had drug-susceptible TB whereas 234 individuals (66.3%) had DR-TB, displaying resistance to at least one drug. One participant demonstrated resistance to 12 drugs at baseline, marking the highest degree of drug resistance observed within the cohort (Figure 2).

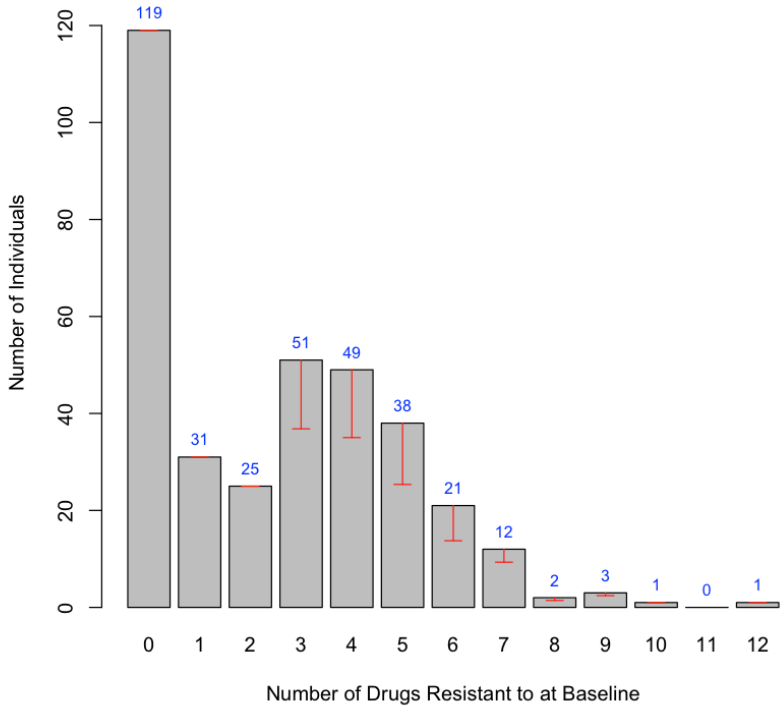

Figure 2. Distribution of Drug Resistance in Study Participants at Baseline.

Distribution of drug resistance among study participants at baseline. Red error bars represent standard deviations.

Phenotypic Amplification

Phenotypic amplification was determined by analysis of DST results. Within the cohort of 353 individuals, 164 participants (46.5%) exhibited changes in resistance for at least one drug during the study. This included 99 individuals with changes from susceptible to resistant (S to R), 97 individuals with changes from resistant to susceptible (R to S). Among those with multiple resistance changes, 32 individuals had both S to R and R to S changes for different drugs. The remaining 189 individuals (53.5%) exhibited no changes in resistance to any drug (Table 1).

Table 1. Changes in Drug Resistance Profiles by DST.

Resistance Changes Observed	Number of Individuals	Percentage
S to R	99	28.05%
R to S	97	24.48%
Both S to R and R to S	32	9.07%
No changes in resistance	189	53.54%

Changes in drug resistance profiles observed among the cohort of 353 individuals with two or more DST results. Indicates the number of people with changes from susceptible to resistant (S to R), resistant to susceptible (R to S), and individuals who had both types of changes (to different drugs) or no changes during the course of the study,

Among those who amplified (exhibited S to R changes) by DST, 59 individuals had changes for just one drug, 21 individuals for 2 different drugs, 9 individuals for 3 different drugs, 6 individuals for 4 different drugs, and 3 individuals for 5 different drugs. Conversely, among the individuals with R to S changes, 64 had changes for just one drug, 22 for 2 different drugs, 7 for 3 different drugs, and 2 for 4 different drugs (Figure 3).

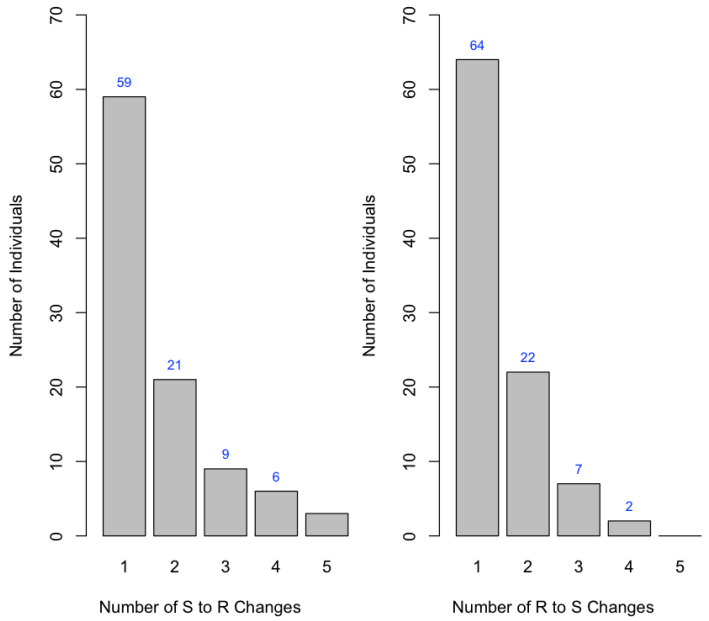

Figure 3. Comparative Analysis of S to R and R to S Transitions.

The number of individuals exhibiting differing degrees of phenotypic changes from susceptible to resistant (S to R) or resistant to susceptible (R to S) during treatment.

Among individuals who exhibited resistance amplification, the largest proportion acquired new phenotypic resistance to first-line drugs, including rifampin (21 individuals), pyrazinamide (20 individuals), isoniazid (19 individuals), and ethambutol (19 individuals) (Table 2). Additionally, there was a noteworthy incidence of resistance amplification in streptomycin, with 31 individuals exhibiting changes. Among individuals who lost resistance, several drugs within both first and second-line categories showed alterations in resistance profiles, including streptomycin (25 individuals), ethambutol (21 individuals), pyrazinamide (9 individuals), and rifampin (9 individuals).

Table 2. Phenotypic Changes in Drug Resistance Profiles.

Drug Name	Individuals With Isolates That Gained Resistance	Individuals With IsolatesThat Lost Resistance
Pyrazinamide	20	9
Isoniazid	19	8
Ethambutol	19	21
Rifampin	21	9
Streptomycin	31	25
Para-aminosalicylic acid	0	1
Capreomycin	5	7
Ciprofloxacin	3	0
Cycloserine	1	1
Ethionamide	5	9
Kanamycin	6	5

Summary of the phenotypic changes in drug resistance profiles among individuals in the study, indicating the number of individuals whose bacterial isolates acquired new phenotypic resistance (S to R) and the number of individuals whose isolates lost resistance (R to S) to each anti-tuberculosis drug.

Genotypic Amplification

After evaluating phenotypic resistance changes in drug susceptibility testing and identifying individuals who exhibited resistance amplification, we further validated these findings using WGS data.

After assessing resistance changes detected by DST and identifying 164 individuals who displayed phenotypic amplification, we aimed to validate these findings using WGS data. To distinguish individuals with genuine genotypic amplification, we implemented a multi-step approach. As detailed in Chapter 2, Figure 1, the following individuals were systematically eliminated from the "amplified" group and subsequently categorized as "non-amplified":

1. 14 individuals who did not have more than one genomic file for analysis
2. 22 people whose genomic data indicated *Mtb* lineage changes
3. 15 participants whose SNPs were synonymous
4. 12 participants with >50 SNP changes in their genomic data
5. 80 individuals with SNPs not known to be associated with resistance

After completing step 4, we identified 91 individuals in our cohort with SNP changes across 95 distinct genes. In step 5, we aimed to refine our dataset by filtering out SNPs not associated with drug resistance. To accomplish this, we cross-referenced our gene list with the World Health Organization (WHO) catalog, which featured 17 genes known to be linked to resistance. Through this comparative analysis, we identified the following final set of six genes associated with resistance amplification:

- ethA
- gid

- gyrA
- katG
- pncA
- rpoB

We used this list of six genes to classify the remaining participants as either 'amplified' or 'not amplified,' assigning the 'amplified' designation to individuals who had SNPs within these genes.

Our analysis identified a total of 11 participants (3.1% of the cohort) who exhibited phenotypic and genotypic amplification, with a total of 15 SNPs found within resistance-associated genes. The remaining 342 individuals (96.9%) in the cohort were categorized as 'not amplified.

Importantly, the identification of only 11 cases of amplification among the total 4,500 participants equates to 99.76% who were not amplified and only 0.24% exhibited amplification, indicating the true rarity of this occurrence.

Statistical Analysis

Characteristics of Independent Variables Related to Study Participants.

HIV status was recorded as negative for 335 individuals, positive for 13 individuals, and five records contained no information (NA). Notably, one individual from the cohort classified as 'amplified' had NA for HIV status. Consequently, this individual had to be excluded from the analysis of this variable.

Diabetes status was recorded as negative for 328 individuals, positive for 21 individuals, and four records contained no information (NA). None of the people with positive diabetes status were among the amplified subset. Each individual with NA was eliminated prior to regression analysis.

Cavitary lesions were identified in 109 individuals and absent in 236 individuals. Eight records contained no information (NA). None of the individuals with cavitary lesions exhibited amplification. The 8 individuals with NA were eliminated from the analysis before logistic regression.

The cohort consisted of individuals whose *Mtb* isolates belonged to two lineages – lineage 2 and lineage 4. Of the 353 participants, 36 (10.2%) had isolates that were identified as belonging to lineage 2, while the remaining 317 (89.8%) were classified under lineage 4. Within the lineage 4 category, participants were further classified into distinct sublineages, with the following distribution: :

- Lineage4.1: 111 participants
- Lineage4.3: 148 participants
- Lineage4.4: 3 participants
- Lineage4.7: 3 participants
- Lineage4.8: 11 participants
- Lineage4.x: 18 participants

Treatment-adherence levels ranged from 0 to 100 in the cohort. Eight individuals had missing adherence data (NA), and all of these individuals were removed from the dataset. Consequently, these 8 individuals were classified as not amplified. Among the

remaining 345 participants, one individual exhibited 0% adherence, while 302 individuals displayed 100% adherence. Among the final amplified group of 11 individuals, 10 had 100% adherence, while 1 person had an adherence level of 96.7%.

Time to effective treatment values ranged from 0 to 1194 days in the dataset, where zero days indicated immediate administration of effective treatment. Among individuals classified as amplified, the days ranged from 0 to 849, with 5 individuals receiving effective treatment on the same day. Detailed descriptive statistics for time to effective treatment are provided in Table 3, including the mean, median, and standard deviation, for both the full cohort of 353 individuals and the subset of amplified individuals.

Table 3. Descriptive Statistics for Time to Effective Treatment.

Subset	Mean	Median	SD
Full Cohort	196.92	144.50	205.93
Amplified Subset	224.00	128.00	287.45

Descriptive statistics for time to effective treatment, including the mean, median, and standard deviation, for both the full cohort of 353 individuals and the subset of amplified individuals.

Univariate Logistic Regression Analyses

We conducted a series of univariate regression analyses to examine the relationships between independent variables and resistance amplification. Applying the

conventional threshold for statistical significance, none of the variables demonstrated a significant association with resistance amplification (Table 4).

Table 4: Summary of Univariate Logistic Regression Analyses.

Predictor Variable	Log Odds	Odds Ratio	Standard Error	P-Value
HIV Status	-15.090	0.000	1809.050	0.993
Diabetes Status	0.457	1.580	1.074	0.670
Cavitary Lesions	0.211	1.234	0.638	0.740
Lineage2	0.000	1.000	NA	NA
Lineage4	0.138	1.147	1.064	0.897
Time to Effective Treatment	0.004	1.004	0.011	0.730
Adherence at 2 mos	0.079	1.082	0.130	0.543

Summary of the univariate analyses, examining the relationship between individual predictor variables, including lineage, HIV status, diabetes status, cavitary lesions, diabetes status, time to effective treatment, and adherence at 2 months, in relation to drug resistance amplification. The table includes coefficient estimates, odds ratios, standard errors, and p-values.

Multivariate Logistic Regression Analysis

Subsequently, we performed a multivariate logistic regression to examine the interrelationships among these variables. similar to the univariate analyses, the results

from the multivariate logistic regression also demonstrated no statistical significance for any of the independent variables (Table 5).

Table 5: Summary of Multivariate Logistic Regression Analysis.

Predictor Variable	Log Odds	Odds Ratio	Standard Error	P-Value
HIV Status	-15.150	2.633	1963.000	0.994
Diabetes Status	0.657	1.929	1.094	0.548
Cavitary Lesions	0.402	1.495	0.663	0.545
Lineage2	0.000	1.000	NA	NA
Lineage4	0.125	1.133	1.082	0.908
Time to Effective Treatment	0.000	1.000	0.011	0.994
Adherence at 2 mos	0.076	1.079	0.130	0.556

Summary of the multivariate analysis, examining the relationship between individual predictor variables, including lineage, HIV status, diabetes status, cavitary lesions, time to effective treatment, and adherence at 2 months, in relation to drug resistance amplification. The table includes coefficient estimates, odds ratios, standard errors, and p-values

Chapter IV.

Discussion

The primary objective of this study was to investigate how suboptimal treatment may contribute to the amplification of drug resistance in individuals diagnosed with drug-resistant tuberculosis, and to measure the frequency of such amplification. We examined a range of potential risk factors, including clinical characteristics and treatment-related variables, to determine their associations with the likelihood of resistance amplification. Our key finding was that drug-resistance amplification is a rare occurrence. Therefore, reliance on phenotypic drug susceptibility testing data alone may lead to an overestimation of resistance amplification.

Initially, we anticipated that time to effective therapy would serve as a key indicator of suboptimal care for DR-TB patients. Thus, suboptimal treatment was defined by the time to effective therapy variable. Other variables such as HIV status, diabetes status, the presence of cavitary lesions, *Mtb* lineage, and adherence to treatment regimens were considered potential risk factors for drug resistance amplification.

Individuals with DR-TB typically received first-line drugs until drug susceptibility test results became available at around the two-month mark. During this period, they might not be receiving the most effective treatment, constituting a form of suboptimal care. We aimed to quantify the incremental impact of time to effective therapy on resistance amplification, assessing how each day of delayed effective therapy contributed to resistance amplification.

Findings of Logistic Regression Analyses

Our investigation included both univariate and multivariate logistic regression analyses, aiming to assess the influence of specific variables on the likelihood of experiencing resistance amplification.

The first variable that we examined was time to effective therapy. Despite considering time to effective therapy as a key indicator of suboptimal care for DR-TB patients, our logistic regression analysis revealed that the duration from diagnosis to the initiation of effective therapy showed no statistically significant association with drug-resistance amplification.

Another risk factor we investigated in this study was the presence of cavitary lesions. These lesions, a distinctive characteristic of advanced and severe forms pulmonary TB disease, manifest as hollow spaces that develop within the lung tissue, often as a response to result of tissue destruction caused by infection with *Mtb*. These cavities can harbor populations of *Mtb*, potentially sheltering the bacteria from the effects of anti-TB drugs, thus influencing the overall treatment response. In this way, they may contribute to suboptimal treatment outcomes. Our univariate analysis suggested some evidence of a potential association, with an odds ratio of 1.23 indicating a 23.4% higher likelihood of resistance amplification in individuals with cavitary lesions.

As a comorbidity, diabetes may complicate tuberculosis by weakening the immune system, increasing bacterial levels, and influencing drug effectiveness. In our univariate analysis, diabetes status revealed a noticeable trend, suggesting a potential association with resistance amplification.

Additionally, we examined *Mtb* lineage as a potential risk factor in our study. Genetic diversity within *Mtb* can influence how the bacterium interacts with the host and responds to treatment. Distinct lineages have been associated with varying patterns of disease severity, transmissibility, and response to anti-TB drugs. However, our analyses revealed no significant association between Lineage4 and the occurrence of resistance amplification when compared to Lineage2, which served as the baseline.

HIV status was another key risk factor investigated in this study. Patients with HIV-positive status face an elevated likelihood of developing active TB disease. The coexistence of HIV can profoundly influence the host's response to TB treatment, potentially contributing to suboptimal treatment outcomes. In our univariate analysis, the absence of cases where both HIV and resistance amplification co-occurred made it challenging to estimate a meaningful odds ratio, consequently resulting in a higher p-value.

Similarly, adherence to treatment during the first two months did not exhibit a statistically significant association with resistance amplification.

Phenotypic and Genotypic Data Integration

The primary finding of our study was the low frequency of resistance amplification observed among study participants. This finding highlights an issue concerning the accurate estimation of resistance amplification. Relying solely on phenotypic DST data may lead to overestimation. When our analysis focused solely on these DST results and compared the shifts from drug-susceptible (S) to drug-resistant (R) phenotypes with the reverse transitions from R to S, we noted that both types of shifts occurred with similar frequencies among individuals. This observation underscores the

need for caution when interpreting DST results, as it carries the inherent risk of overestimating the prevalence of resistance amplification, potentially leading to misinformed decisions in tuberculosis treatment and intervention development. However, the incorporation of WGS analysis presents a useful approach that overcomes the limitations of DST and provides a more accurate understanding of drug resistance dynamics in TB patients.

It is valuable to contrast our findings with the research conducted by Cegielski et al., whose focus on DST data revealed high prevalence of resistance amplification.[7] Their research did not include WGS analysis, which we emphasize can provide a more accurate and comprehensive perspective on drug resistance dynamics in TB patients. This highlights the potential for overestimation in studies relying solely on DST data, making a strong case for integrating advanced molecular techniques like WGS to improve our understanding of resistance dynamics.

By acknowledging the limitations of DST and embracing advanced molecular techniques like WGS, we can make more informed decisions in the treatment and intervention strategies for drug-resistant tuberculosis."

Considerations

While our study explored several potential risk factors, it is important to recognize the possible involvement of other contributors not considered here, such as socioeconomic or environmental influences. Combating the rise of DR-TB requires a multifaceted approach that considers not only treatment optimization but also an exploration of additional contributing factors. As such, a more comprehensive

examination into a wider range of variables may be necessary to better understand the dynamics of resistance amplification.

One important consideration is the limited size of our dataset. While the Peru study included over 3,500 participants, making it a substantial study, our specific subset consisted of only 353 people, which, when viewed in the context of the entire study, may still be considered relatively small. It is reasonable to consider that a larger dataset could potentially yield more significant findings, particularly for the variables where we observed trends toward significance but didn't reach statistical significance. A larger dataset could allow for a better representation of the population and reduce the impact of random variability. This would enhance statistical power and increase the likelihood of detecting significant relationships or associations that might be more subtle or context-specific with a smaller sample size. With a larger dataset, we might have more statistical power to detect smaller effect sizes, which could lead to significant relationships that might not be apparent in a smaller dataset.

Future research efforts may include the incorporation of advanced data analytics, including machine learning and artificial intelligence, to provide enhanced modeling of resistance amplification. These advanced approaches offer the potential to uncover subtle interactions that might not be easily apparent through the traditional statistical methods used in our study. Through advanced computational approaches, researchers could gain deeper insights into the complex dynamics of resistance amplification and develop more accurate predictive models. This, in turn, can inform more targeted interventions and strategies for effectively managing DR-TB.

Appendix 1.

Supplemental Tables

In this appendix, we present supplemental tables that offer a detailed exploration of key aspects related to drug resistance in our study cohort. Table 6 displays the distribution of TB drug resistance among study participants, categorizing individuals based on the number of drugs to which they exhibited resistance. Tables 7 and 8 provide insights into the distribution of S to R and R to S changes in DST resistance profiles, offering a nuanced view of the dynamics of resistance amplification during treatment. Lastly, Table 9 offers a comprehensive breakdown of sensitivity changes in resistance profiles by drug, shedding light on the variations observed in our study cohort. These supplemental tables enrich our understanding of the multifaceted factors influencing drug resistance, providing context to the primary findings presented in the main results section.

Table 6. Distribution of TB Drug Resistance Among Study Participants.

Number of Drugs Resistant to	Number of Individuals	Percentage	Standard Deviation
0	119	33.71%	0
1	31	8.78%	0
2	25	7.08%	0
3	51	14.45%	14.17

Number of Drugs Resistant to	Number of Individuals	Percentage	Standard Deviation
4	49	13.88%	13.98
5	38	10.76%	12.64
6	21	5.95%	7.25
7	12	3.40%	2.67
8	2	0.57%	0.57
9	3	0.85%	0.57
10	1	0.28%	0.0
11	0	NA	NA
12	1	0.28%	0.0

Distribution of drug susceptibility profiles among study participants, including the number of drugs to which individuals exhibited susceptibility with the corresponding number of participants in each category. The percentages in each cell reflect the proportion of participants within each susceptibility category relative to the total study cohort of 353 individuals. Participants with resistance to zero drugs had fully susceptible TB.

Table 7. Distribution of S to R Changes in DST Resistance Profiles.

Number of Drugs with S to R changes	Number of Participants	Percentage of Total Participants (of 353)
1	59	0.167
2	21	0.059
3	9	0.025
4	6	0.017
5	3	0.008

The distribution of S to R changes in resistance profiles who had two or more DST results. Participants are categorized based on the number of drugs for which they exhibited changes in resistance profiles. Percentage represents the proportions of participants within each category relative to the total study cohort of 353 individuals,

Table 8. Distribution of R to S Changes in DST Resistance Profiles.

Number of Drugs with R to S changes	Number of Participants	Percentage of Total Participants
1	64	0.181
2	22	0.062
3	7	0.020
4	2	0.006

The distribution of R to S changes in resistance profiles who had two or more DST results.

Table 9. Distribution of Sensitivity Changes in Resistance Profiles by Drug.

Drug	Number of Tests	Number of S to R Changes	Number of R to S Changes	Percent S to R	Percent R to S
pza	350	26	15	0.074	0.043
inh	350	19	19	0.054	0.026
emb	349	24	31	0.069	0.089
rif	350	24	12	0.069	0.034
sm	349	40	29	0.115	0.083
pas	149	5	1	0.034	0.007
cm	188	9	8	0.048	0.043
cpx	149	7	2	0.047	0.013
cs	149	4	2	0.027	0.013
eth	188	6	16	0.032	0.085
km	188	9	7	0.048	0.037
rbu	0	0	0	NA	NA
lin	0	0	0	NA	NA
mox	0	0	0	NA	NA

Drug	Number of Tests	Number of S to R Changes	Number of R to S Changes	Percent S to R	Percent R to S
amk	0	0	0	NA	NA

The distribution of sensitivity changes in resistance profiles for various drugs within the study cohort. Including the number of changes from susceptible to resistant (S to R), the number of changes from resistant to susceptible (R to S). Percentages are relative to the total number of tests conducted for each drug.

References

Adigun, R., & Singh, R. (2023). *Tuberculosis*. In StatPearls. StatPearls Publishing.

Becerra, M. C., Huang, C. C., Lecca, L., Bayona, J., Contreras, C., Calderon, R., Yataco, R., Galea, J., Zhang, Z., Atwood, S., Cohen, T., Mitnick, C. D., Farmer, P., & Murray, M. (2019). Transmissibility and potential for disease progression of drug resistant *Mycobacterium tuberculosis*: prospective cohort study. *BMJ (Clinical research ed.)*, 367, l5894. https://doi.org/10.1136/bmj.l5894

Bhering, M., & Kritski, A. (2020). Primary and acquired multidrug-resistant tuberculosis: Predictive factors for unfavorable treatment outcomes in Rio de Janeiro, 2000-2016. *Revista panamericana de salud publica = Pan American journal of public health*, 44, e178. https://doi.org/10.26633/RPSP.2020.178

Cegielski, J. P., Dalton, T., Yagui, M., Wattanaamornkiet, W., Volchenkov, G. V., Via, L. E., Van Der Walt, M., Tupasi, T., Smith, S. E., Odendaal, R., Leimane, V., Kvasnovsky, C., Kuznetsova, T., Kurbatova, E., Kummik, T., Kuksa, L., Kliiman, K., Kiryanova, E. V., Kim, H., Kim, C. K., … Global Preserving Effective TB Treatment Study (PETTS) Investigators (2014). Extensive drug resistance acquired during treatment of multidrug-resistant tuberculosis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 59(8), 1049–1063. https://doi.org/10.1093/cid/ciu572

Centers for Disease Control and Prevention. (2021). Tuberculosis. In Division of Tuberculosis Elimination. Centers for Disease Control and Prevention. https://www.cdc.gov/tb/publications/default.htm

Dheda, K., Gumbo, T., Gandhi, N. R., Murray, M., Theron, G., Udwadia, Z., Migliori, G. B., & Warren, R. (2014). Global control of tuberculosis: from extensively drug-resistant to untreatable tuberculosis. *The Lancet. Respiratory medicine*, 2(4), 321–338. https://doi.org/10.1016/S2213-2600(14)70031-1

Hu, Y., Zheng, X., Davies Forsman, L., Ning, Z., Chen, C., Gao, Y., Zhang, Z., Lu, W., Werngren, J., Bruchfeld, J., Hoffner, S., & Xu, B. (2021). Emergence of additional drug resistance during treatment of multidrug-resistant tuberculosis in China: a prospective cohort study. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 27(12), 1805–1813. https://doi.org/10.1016/j.cmi.2021.04.001

Shaikh, A., Sriraman, K., Vaswani, S., Oswal, V., Rao, S., & Mistry, N. (2021). Early phase of effective treatment induces distinct transcriptional changes in

Mycobacterium tuberculosis expelled by pulmonary tuberculosis patients. Scientific reports, 11(1), 17812. https://doi.org/10.1038/s41598-021-96902-7

Shin, H. J., & Kwon, Y. S. (2015). Treatment of Drug Susceptible Pulmonary Tuberculosis. Tuberculosis and respiratory diseases, 78(3), 161–167. https://doi.org/10.4046/trd.2015.78.3.161

World Health Organization (WHO). (2021) Catalogue of mutations in *Mycobacterium tuberculosis* complex and their association with drug resistance. https://apps.who.int/iris/handle/10665/341981

World Health Organization (WHO). (2022) Global tuberculosis report 2022. World Health Organization, Geneva, Switzerland. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022