Quantitative measurement of antibiotic resistance in *Mycobacterium tuberculosis* reveals genetic determinants of resistance and susceptibility in a target gene approach

The CRyPTIC Consortium[1]

Abstract

The World Health Organization goal of universal drug susceptibility testing for patients with tuberculosis is most likely to be achieved through molecular diagnostics; however, to date these have focused largely on first-line drugs, and always on predicting binary susceptibilities. Here, we used whole genome sequencing and a quantitative microtiter plate assay to relate genomic mutations to minimum inhibitory concentration in 15,211 *Mycobacterium tuberculosis* patient isolates from 27 countries across five continents. This work identifies 449 unique MIC-elevating genetic determinants across thirteen drugs, as well as 91 mutations resulting in hypersensitivity for eleven drugs. Our results provide a guide for further implementation of personalized medicine for the treatment of tuberculosis using genetics-based diagnostics and can serve as a training set for novel approaches to predict drug resistance.

[126 words]

[1] See the CRyPTIC Author List at the end of manuscript

Introduction:

Mycobacterium tuberculosis (Mtb) caused an estimated 10 million new cases of tuberculosis (TB) and 1.4 million deaths in 2019[1]. Of particular concern are the estimated 465,000 rifampicin resistant (RR) cases, 78% of which were multi-drug resistant (MDR, resistant to both rifampicin and isoniazid)[1]. Drug resistance poses two major challenges to the successful treatment of TB, as it is both underdiagnosed (only 38% of RR/MDR cases in 2019)—leading to under-treatment—and has poor treatment success rates even when identified (57% globally in 2019)[1]. Despite attempts to move to shorter and all-oral MDR TB regimens using new drugs, most patients are still receiving toxic regimens that decrease patient adherence[1,2]. Collectively, the failure to identify and successfully treat these cases leads to onward transmission and amplification of drug resistant strains.

The WHO has identified better diagnosis and treatment of drug resistant tuberculosis as a key part of the global tuberculosis eradication strategy[1]. Rapid genetics-based diagnostic tools, such as GeneXpert, have been widely adopted as they are faster and cheaper than traditional culture-based diagnostic susceptibility testing (DST). However, outbreaks caused by drug-resistant strains with mutations not detected by such assays reveal the importance of developing assays that include a wider range of resistance determinants[3]. Some approaches incorporate whole-genome sequencing (WGS) or targeted next generation sequencing to identify all possible resistant variants and recently these methods have proven to be capable of replacing culture-based DST for

the first line drugs; however, implementation of this technology is not yet feasible globally due to cost and technical expertise constraints[4–6].

Most current culture and genetics-based DST approaches generate binary results—'resistant' or 'susceptible'—and thus fail to consistently report elevations in minimum inhibitory concentration (MIC) below or around the critical concentration[7]. These sub-threshold elevations in MIC may nevertheless be clinically meaningful, as the combination of significant interpatient pharmacokinetic variability and elevated MICs predisposes Mtb strains to development of higher-level resistance, risking treatment failure and worse patient outcomes[8,9]. A binary system also hampers the wider implementation of informed high-dose regimens which have been trialed to extend the clinical utility of relatively less toxic and more widely available drugs such as rifampicin and isoniazid[10–12]. While some previous efforts have attempted to use quantitative MICs to identify these lower-level resistance variants, they were limited by smaller sample sizes and combined heterogenous methods of resistance determination[13]. Additionally, relatively few studies have had adequate sample sizes to investigate drugs such as bedaquiline, linezolid, clofazimine and delamanid that are poised to become the new "front-line" drugs for the MDR-TB treatment.

To resolve these issues, we performed WGS and determined the MICs of thirteen drugs for 15,211 Mtb isolates selected from patient samples gathered from 27 countries over five continents using a previously validated microtiter plate[14]. This data covers all first-line drugs (except pyrazinamide), as well as eight drugs from the new MDR-TB

treatment guidelines (all Group A, one Group B, and four Group C)[15]. The results serve as guides for pharmacokinetic and dosing studies to extend the clinical utility of less toxic and more widely available drugs for the treatment of drug-resistant tuberculosis, as well as help to improve the design of genetics-based rapid diagnostics for MDR-TB and the recently published WHO genetic catalogue for tuberculosis[16]. They also provide a large, quality-controlled dataset for development of drug resistance prediction algorithms using machine-learning and other approaches.

Results:

Dataset description

Bacterial isolates were collected from patient samples from 27 different countries and were over-sampled for drug resistance. Of the 15,211 isolates included in the initial CRyPTIC dataset, 5,541 were phenotypically susceptible to isoniazid, rifampicin, and ethambutol, 5,602 were isoniazid mono-resistant, 5,261 were rifampicin resistant, and 4,125 were multidrug resistant (MDR, resistant to both rifampicin and isoniazid) based on previously published epidemiological cutoffs (ECOFF, MIC that encompasses 99% of wild type) for the microtiter plates used in this study[17]. Binary phenotypic resistance to the newer drugs was observed at a lower prevalence, with 71 bedaquiline resistant isolates, 106 clofazimine resistant isolates, 76 linezolid resistant isolates, and 85 delamanid resistant isolates (Table S1). Isolate lineages were determined using a published SNP-based protocol from WGS data and the lineage distribution across countries reflects previously described phylogeographic distributions[18–20]. Five out of six major lineages of Mtb were represented in the dataset, with most isolates mapping to L4 (6,066 isolates) and L2 (4,323 isolates), while L3 (1,059), L1 (677), and L6 (6) comprised the remainder. A full description of the CRyPTIC dataset and determination of the ECOFFs has been previously published (also see Methods)[17].

Genetic resistance determinants in Mycobacterium tuberculosis

Previous studies have shown that the majority of genetic determinants of resistance to most anti-tuberculosis drugs are related to a relatively small number of genes[6,21]. We thus employed a candidate gene approach and restricted our investigation of genomic

variation to previously identified genes and the 100bp upstream for each drug (Table 1). All unique variation in the target genes and upstream regions (SNPs, both synonymous and nonsynonymous, as well as insertions and deletions <50 base pairs in length) that occurred in isolates with matched high-quality phenotypic data was included in a separate multivariable linear mixed model controlling for population structure and technical variation between sites for each drug, after removing isolates with evidence for mixed allele calls at sites previously identified as resistant (e.g. rpoB S450X, Methods). Final sample sizes per drug ranged from 6,681 for moxifloxacin to 10,042 for rifabutin (mean sample size 8,353, Figure 1, Methods). Most isolates had less than five nonsynonymous mutations across all target genes for each drug (Table S2).

Across thirteen drugs, 540 mutations in 39 genes (out of 4,667 mutations and 49 genes tested) were found to have independent effects on MIC after correction for multiple testing (Benjamini-Hochberg correction, false discovery rate<0.05, Figure 2, Table 1, S3). Ethionamide had the most unique variants associated with reduced susceptibility (128), while linezolid had the least (8). Effect sizes were measured in log2MIC (where an increase in 1 log2MIC was equivalent to a doubling of the MIC) and positive effects for estimates derived from at least three observations ranged from a 0.22 increase in kanamycin log2(MIC) for *rrs* c492t to a 10.1 increase in isoniazid log2(MIC) for *katG* W477Stop. Multiple promoter mutations were implicated in resistance to isoniazid, ethionamide, amikacin, kanamycin, and ethambutol (Figure 2B). The effects of promoter mutations varied widely, with mutations upstream of *eis* and *embA* being almost exclusively associated with sub-ECOFF elevations in MIC for amikacin and ethambutol

respectively, while most promoter mutations for the isoniazid and ethionamide-related *fabG1* resulted in MICs above the ECOFF[17]. While a prior study found that common promoter mutations tended to be associated with lower-level resistance than their corresponding common gene-body counterparts (e.g. *fabG1* c-15 vs *inhA* I21), we found that intergenic vs gene body mutation was only associated with significantly different effects on MIC for isoniazid, ethambutol, and kanamycin (Table S4)[13]. In fact, we found that the widespread *fabG1* c-15t promoter mutation was associated with higher-level and equivalent-level resistance to its gene body counterparts *inhA* I21V and I21T respectively (Figure 2B, Wald test for equality of coefficients p=0.0006, p=0.24 respectively). Resistance-associated promoter mutations were enriched in the region around each gene's respective -10 element, which is consistent with the essentiality of the -10 hexamer and increased variability around the -35 position in mycobacterial promoters (Figure S1, +/- 5 nucleotides, Mantel-Haenszel common OR=4.5, p=0.0007)[22,23]. Multiple insertion/deletion mutations were associated with resistance to isoniazid, rifampicin, rifabutin, ethionamide, ethambutol, bedaquiline, clofazimine, and delamanid (Table S3, Figure 2). Homoplastic mutations (multiple evolutionarily independent occurrences) were more likely to be associated with resistance for all drugs except amikacin, clofazimine, linezolid, and delamanid (Woolf test for homogeneity of ORs p=0.0001, Table S5).

One notable advantage of quantitative MIC measurements is that they also enable investigation of variants associated with MIC decreases. We identified 63 increased susceptibility-associated mutations (with at least three occurrences) whose effect sizes ranged from -4.3 rifampicin log2(MIC) for Rv2752c H371Y to -0.23

kanamycin log2(MIC) for *eis* V163I (Figure 2A). Eight of these mutations were homoplastic with at least three independent occurrences, which raises the question of what selective pressure may drive increases in drug susceptibility (Methods).

First-line drugs

Rifampicin is a critical first line drug and resistance to it is almost entirely mediated by mutations within an 81-base pair region of the *rpoB* gene (rifampicin resistance determining region, RRDR). Most molecular assays target mutations in this region for rapid prediction of rifampicin resistance, however, mutations outside this region have been associated with outbreaks[24,25]. We identified 35 mutations in *rpoB* occurring at least 3 times whose effects collectively ranged from 1.0 to 9.0 increases in log2MIC (Figure 3A). Notably, seven unique resistance-associated mutations occurred outside the RRDR, at positions V170, Q172, I491, and L731; however, only V170F was associated with high level resistance (8.37 increased log2MIC). Although disparate in primary sequence from the RDRR, positions V170, Q172 and I491 are all near the drug-binding pocket structurally (Figure 3B). Interestingly, a homoplastic in-frame deletion 12 bp in size in the RRDR was also associated with rifampicin resistance (Figure 3C, Table S3). Several types of insertion/deletion mutations in the RDRR have previously been reported, although they are rare, consistent with their greater structural consequences for the essential RNA polymerase[26].

Prior studies have identified seven “borderline” mutations in *rpoB* (L430P, D435Y, H445L, H445N, H445S, L452P, and I491F) for rifampicin; resistant isolates with these mutations are often missed by phenotypic methods such as the Mycobacterial

Growth Indicator Tube (MGIT), possibly due to slower growth rates, which has led to a reduction in the critical concentration for MGIT in the latest WHO guidelines[27–29]. These mutations' MICs range on the plate from 5.1 log2MIC for H445L to 2.3 log2MIC for L430P (rifampicin ECOFF minus baseline MIC=3.3, Table S3). Here, we identify thirteen additional *rpoB* mutations independently associated with elevated MICs that are less than 5.1 log2MIC (8/13 located in the RDRR, Table S3). Sixteen *rpoB* mutations in total were independently associated with elevated MICs at or below the rifampicin ECOFF, including rpoB L430P, a variant that has been successfully treated with a high dose rifampicin-containing regimen clinically[30].Several rpoB positions (Q432, D435, H445) harbored both high and low-level resistance-associated alleles, while others (L430, L452, I491) were associated exclusively with lower-level resistance regardless of the amino acid substitution (Figure 3B,C orange and yellow shading respectively). Mapping these mutations onto the rpoB structure revealed that high--level resistance often involves disruption of the interactions with the rigid napthol ring while mutations at positions that contact the ansa bridge had more variable effects, potentially due to increased structural flexibility in this region of the drug (Figure 3B). Low-level resistance mutations often co-occurred with other low-level resistance mutations, producing high-level resistance additively.

Rifabutin (a structural analogue to rifampicin) is associated with a lower ECOFF and mutations in *rpoB* were associated with lower elevations in rifabutin MIC compared to rifampicin MIC (paired Wilcoxon p=3.7e-9, Figure 3A, Table S3). Interestingly however, all structural features contacted by these mutations were shared between rifampicin and

rifabutin (Figure 3B). A single mutation, rpoB Q409R (n=24, p=5.0e-3 after Benjamini-Hochberg (BH) correction), was associated with decreased rifampicin and rifabutin MICs; interestingly, this mutation has been proposed as a compensatory mutation that may alter the rate of transcription initiation and resulting transcription efficiency for isolates that harbor other RDRR mutations[31].

Resistance to isoniazid is mediated primarily through loss-of-function mutations in the prodrug converting enzyme *katG*, with canonical high-level resistance caused by the S315T mutation, which was associated with a 6.2 log2 increase in MIC (Figure 4A). Not all *katG* mutations were associated with high level resistance, nearly half (15/31) being associated with increases in MIC at or below the ECOFF. No mutations likely to result in severe loss of function were associated with sub-ECOFF resistance, supporting the consensus of treating presumptive loss-of-function mutants in *katG* as resistant. The other canonical isoniazid-related genes, *inhA* and *fabG1*, tended to be associated with lower-level resistance, with 4/6 and 5/6 mutations associated with sub-ECOFF increases in MIC respectively (Figure 4A, Table S3). While *fabG1* was previously the only synonymous mutation known to be associated with resistance to isoniazid, here we identify a synonymous mutation in the first codon of *katG* that confers high-level resistance to isoniazid, likely by reducing the rate of translation initiation and subsequent production of katG enzyme required for activation of isoniazid (4.5 log2MIC, n=3, p=1.4e-8 after Benjamini-Hochberg (BH) correction, Table S3).

Most isoniazid resistance-associated mutations in katG occurred in the N-terminal lobe responsible for heme-binding and pro-drug conversion (Figure 4B). Most isolates harbored variation at position S315, located in the primary isoniazid-binding pocket on the δ edge of the heme; interestingly however, another cluster of resistance-associated mutations occurred in the helix made up of residues 138-155. Some structural evidence exists for promiscuous isoniazid binding at this site and mutations of this region in *Escherichia coli* cause reduced catalase/peroxidase activity and heme binding; however the precise mechanism of effect of these mutations in Mtb is unknown[32,33]. Intriguingly, one mutation in this region, KatG S140N, was associated with decreased isoniazid MIC (n=9, p=5.4e-4 after BH correction, Figure 4B).

Non-canonical isoniazid resistance-associated variants were identified in *ahpC*, *ndh*, and *Rv1258c* (*tap*) (Figure 4A). Mutations in *ahpC* were associated with increased MICs; however, these mutations almost always co-occurred with mutations in canonical isoniazid genes and investigation of the interaction between these co-occurring mutation pairs revealed that *ahpC* mutations did not result in additive resistance, consistent with their proposed compensatory role (Figure 4A). Several recent genome-wide association studies (GWAS) have implicated mutations in the ribonuclease/beta-lactamase *Rv2752c* in resistance and tolerance to both rifampicin and isoniazid; however, they also identified convergent mutations in drug susceptible strains[13,34]. While we identified nine nonsynonymous mutations with significant effects on log2MIC, only one, V218L, was shared between isoniazid and rifampicin, causing a 3.2 elevation in log2MIC for both drugs (Table S3). Only one other *Rv2752c* variant was associated

with elevated rifampicin MICs, while four variants in this gene were associated with elevated isoniazid MICs (Figure 4A).

Canonical ethambutol resistance is mediated by mutations in *embA* or *embB*. We identified 45 variants, 12 in the *embC-embA* intergenic region, five in *embA*, and 28 in *embB*, that were independently associated with elevated ethambutol MICs (Figure 4C). Mutations in the *embC-embA* intergenic region have been proposed to upregulate production of embA and embB by altered promoter structure. Most *embA* variants were in the upstream region from -16 to -8, however three were located upstream around the -35 element. All *embA* mutations were associated with MIC increases below the ECOFF (Figure 4C, Table S3). Interestingly, 22/28 mutations in *embB* were also associated with sub-ECOFF increases in MIC, including the canonical embB M306I. Low-level resistance mutations often co-occurred, resulting in high-level additive resistance, consistent with previous studies (Table S6)[35]. Mutations associated with resistance in embB were clustered around the drug binding pocket (Figure 4D)[36]. We also identified resistance-associated variants in *embC* and *ubiA*, although these occur less frequently and require further validation.

Group A and B MDR drugs

The principal mechanism of resistance to fluoroquinolones is mutations in either subunit of DNA gyrase (*gyrA* or *gyrB*). We identified 22 mutations (12 gyrA, 10 gyrB) and 19 mutations (10 gyrA, 9 gyrB) that were independently associated with increased levofloxacin and moxifloxacin MICs respectively (Figure 5A). Resistance-associated

mutations in *gyrB* occurred without an accompanying *gyrA* mutation ~65% of the time (29/44 isolates LEV, 35/54 isolates MXF) but were associated with lower overall—and in some cases sub-ECOFF—changes in MIC (Figure 5A, Table S3,S6). Most mutations associated with increased fluoroquinolone MICs were within 10 Å of the drug binding pocket (Figure 5B). Intriguingly, two positions—gyrB R446 and gyrB S447—each harbored two unique resistance-associated missense mutations despite being over 25 Å from the bound fluoroquinolone. Both residues make contacts with the gyrB protein backbone at positions 473-475, suggesting they may exert an allosteric effect by either influencing protein folding and/or the position of residues (notably D461 and R482) that make up part of the fluoroquinolone binding pocket (Figure 5B). Interestingly, while gyrB E501D was associated with resistance 1 log2MIC above the moxifloxacin ECOFF, it did not cause a similar elevation for levofloxacin (0.1 log2MIC above ECOFF), consistent with previous studies[7,37,38]. We speculate this could be due to alteration in the coordination of gyrB R482—which must shift to accommodate the bulkier side group of moxifloxacin—although this remains to be shown experimentally (Figure 5B).

While initial studies on bedaquiline and clofazimine resistance highlighted *atpE* (bedaquiline), *pepQ*, *Rv0678*, and *Rv1979c* as mediating resistance, surveillance of clinical samples has revealed the importance of the efflux mechanism mediated by the *mmpL5* membrane transporter, which is controlled by the transcriptional regulator *Rv0678*. Consistent with this, we identified sixteen and four mutations in *Rv0678* that were associated with elevated bedaquiline and clofazimine MICs respectively, of which four were shared (Figure S2, Table S3). We also identified two *mmpL5* mutations that

were associated with increased MICs for each drug which were not shared between the two drugs. Finally, we identified both the atpE E61D (n=3) drug binding site mutation associated with bedaquiline resistance and two mutations in *Rv1979c* associated with clofazimine resistance. No mutations in *pepQ* were associated with resistance to either drug. Importantly, 5 unique nonsense and frameshift mutations in *mmpL5* increased susceptibility to bedaquiline by -1.9 to -4.0 log2MIC, of which one, mmpL5 Y300Stop, was also shared with clofazimine (Figure 2A). Inactivating mutations in *mmpL5* abrogated resistance mediated by co-occurring *Rv0678* mutations, consistent with a hypothesis proposed by a prior study[39].

Resistance to linezolid is mediated by mutations in *rplC* and *rrl*, which tend to cause higher- and lower-level resistance respectively. We identified the classical rplC C154R (n=43) mutation and five variants in *rrl* associated with elevated linezolid MICs (Figure S2, Table S3).

Group C MDR drugs

Aminoglycoside resistance is canonically mediated by mutations in the 16s rRNA encoded by *rrs*. We identified five and six mutations in *rrs* that were independently associated with elevated MICs for amikacin and kanamycin respectively (Figure 5C). Multiple promoter mutations in *eis* were associated with elevated MICs to kanamycin (7) and amikacin (3). Interestingly, *eis* promoter mutations were associated with sub-ECOFF elevations in MIC for amikacin, while being associated with elevations in MIC comparable to *rrs* mutations for kanamycin. A deletion in *eis* leading to loss of function

was also associated with increased susceptibility to kanamycin, consistent with an epistatic interaction abrogating the resistance gained from *eis* overproduction[39]. Variants in *aftB*, *ccsA*, *whiB6* and *whiB7* were also associated with elevated MICs for at least one aminoglycoside, however they were infrequent and require further investigation (Figure 5C and Table S3).

Ethionamide is a prodrug that is activated by the monooxygenases *ethA* and *mymA* (*Rv3083*). More variants (128) were associated with increased ethionamide resistance than any other drug, with the majority (103) occurring in *ethA*. Notably however, most (97/103) MIC-elevating *ethA* variants did not raise the ethionamide MIC above the ECOFF. Variants in *fabG1* and *inhA* were common and strongly associated with elevated ethionamide MICs (Figure S2). Five resistance-associated variants were identified in the alternative activating enzyme for ethionamide, *Rv3083*, and three resistance-associated variants were found in the non-canonical ethionamide gene *mshA*. Two mutations in *ethR we*re associated with decreased ethionamide MICs, consistent with its role as a regulator of the prodrug activating enzyme *ethA*.

Resistance to delamanid is mediated by inactivating mutations in *ddn* or by mutations that affect the cofactor F_{420} biosynthesis pathway (namely *fgd1* and *fbiA-D*). We identified eleven mutations in *ddn*, seven in *fbiA*, and one in *fbiC* that were associated with increases in delamanid MIC (Figure S2, Table S3). Over half (6/11) of the mutations in *ddn* were nonsense or frameshift mutations.

Effect of genetic background on MIC

Several studies have noted that the strain genetic background can influence MICs in addition to primary resistance mutations[35,40,41]. In this study, we found that the effects of lineage on isolate MIC tended to be small compared to primary resistance allele effects for most drugs (mean lineage effect 0.41 log2MIC, mean lineage effect to median primary resistance allele effect ratio 0.15), yet still statistically significant (Figure S3). Notably however, lineage three was associated with a 1.5 lower moxifloxacin log2MIC compared to lineage four after controlling for primary resistance alleles in *gyrA* and *gyrB*.

Interactions beyond additivity

We also sought to identify whether there were any effects beyond additivity for co-occurring mutation pairs. Out of 33 pairs tested across 13 drugs, we identified three mutation pairs with greater than additive effects on ethambutol resistance and one pair (rpoB_L430P:rpoB_D435G) with greater than additive rifampicin resistance (Figure S4). The interaction of these mutations resulted in log2MICs increased beyond additivity by 1.4 to 2.4 log2MIC, which resulted in MICs well beyond that of the strongest individual mutations for ethambutol and rpoB S450L for rifampicin. The remaining significant mutation-pairs either consisted of a known resistance mutation with a putative compensatory mutation (such as rpoB with rpoC) or had additive MICs that were in the tails of the distribution, suggesting that interactions were reflecting assay thresholds, at least in part, as opposed to true effects.

Extension beyond the binary 2021 WHO catalogue

To assess how measurement of MICs improves our ability to detect meaningful genetic associations with resistance/susceptibility, we compared our MIC-based catalogue with the recently published 2021 WHO binary catalogue for tuberculosis (Table S7)[16]. 179 unique mutation-phenotype associations were matched across the two catalogues, with nearly a third (59/179) classified as “resistant – interim”. Our model finds that 61% (36/59) of these mutations are associated with significant elevations in MIC in our data, of which 14 were sub-ECOFF and therefore unlikely to be confidently identified by binary methods. The inability of binary methods to detect these smaller but significant elevations in MIC is also shown by the lack of associations in *Rv0678* for bedaquiline and clofazimine.

Discussion

In this study, we used WGS combined with high throughput MIC measurements to develop a quantitative catalogue of resistance to thirteen anti-tuberculosis drugs. Linking mutations to MICs allows for a rapid and reliable alternative to phenotypic DST for individual isolates that does not rely on critical concentrations that may be revised. These results can help to improve diagnostics and guide future study designs trialing high dose therapies of less toxic and more effective drugs (e.g. rifampicin, isoniazid and moxifloxacin)[10,11,24].

Notably, we identified 321 mutations whose effects on MIC are entirely or partially below their respective ECOFF. Further work is needed to understand whether these mutations lead to increased treatment failure and/or relapse rates as is the case for the "borderline" mutations in *rpoB* for rifampicin[27]. If so, rapid molecular assays should be employed to detect these variants.

We also found mutations associated with increased susceptibility to bedaquiline, clofazimine, and the aminoglycosides, which raises the intriguing possibility of optimizing regimens based on hypersensitivity as opposed to resistance. Given the relatively common rate of inactivating mutations in *mmpL5*, rapid molecular tests should be developed to ensure that these isolates are not falsely identified as resistant. Deletion of other transcriptional regulators has also been shown to increase bedaquiline susceptibility, suggesting other sensitizing mutations may also occur[42]. Further work to

understand the distribution and frequency of these mutations may help elucidate their clinical relevance globally.

Our new catalog was unable to explain most binary resistance to ethionamide, bedaquiline, clofazimine, linezolid and delaminid, implying that many new variants and loci remain to be discovered (Figure S5)[43]. More widespread use of these drugs clinically will facilitate collection of resistant strains for use in GWAS to identify other genetic loci involved in resistance; however, high levels of inactivating variation were observed in *ethA* (ethionamide), *ddn* (delamanid) and *Rv0678* (bedaquiline/clofazamine), suggesting that many isolates will need to be sampled to achieve saturation for these drugs, similar to pyrazinamide. Alternative approaches relying on random mutagenesis, directed evolution, and machine learning have been employed to generate predictions for mutations that have never been observed in a patient, however these may not always identify mutations that are competive *in vivo*[44–51]. The database generated by CRyPTIC can be used as a resource for these approaches by highlighting which mutations actually occur in patients and acting as a training set for machine learning algorithms.

Limitations to this study include the lower number of isolates resistant to newer drugs, potential misattribution of mutational effects outside our target genes or due to exclusion of insertions/deletions >50bp in size, and the use of ECOFFs that have not yet been extensively validated against other methods, although we have shown good concordance with MGIT results[17]. We have attempted to limit erroneous associations

through controlling for lineage and population structure in our modelling approach as well as by validating mutations through structural mapping and degree of homoplasy where possible. Finally, changes in transcription or translation may also mediate antibiotic tolerance and persistence states to impact the efficacy of antibiotics *in vivo*[52].

Figures and Tables

Table 1. Candidate genes used in this study.

Drug	Abbrev.	Candidate genes
Isoniazid	INH	*katG, fabG1, inhA, ahpC, ndh, kasA, Rv1258c, Rv2752c*
Ethionamide	ETH	*ethA, ethR, fabG1, inhA, mshA, Rv3083*
Rifampicin	RIF	*rpoA, rpoB, rpoC, rpoZ, Rv2752c*
Rifabutin	RFB	*rpoA, rpoB, rpoC, rpoZ, Rv2752c*
Ethambutol	EMB	*embA, embB, embC, embR, rmlD, iniA, iniC, manB, ubiA*
Amikacin	AMI	*rrs, eis, ccsA, whiB6, whiB7, aftB, fprA*
Kanamycin	KAN	*rrs, eis, ccsA, whiB6, whiB7, aftB, fprA*
Levofloxacin	LEV	*gyrA, gyrB*
Moxifloxacin	MXF	*gyrA, gyrB*
Bedaquiline	BDQ	*atpE, Rv0678, mmpL5, mmpS5, pepQ, Rv3249c*
Clofazimine	CFZ	*Rv1979c, pepQ, Rv0678, mmpL5, mmpS5, Rv3249c*
Linezolid	LZD	*rplC, rrl, Rv3249c*
Delamanid	DLM	*ddn, fgd1, fbiA, fbiB, fbiC, fbiD, Rv3249c*

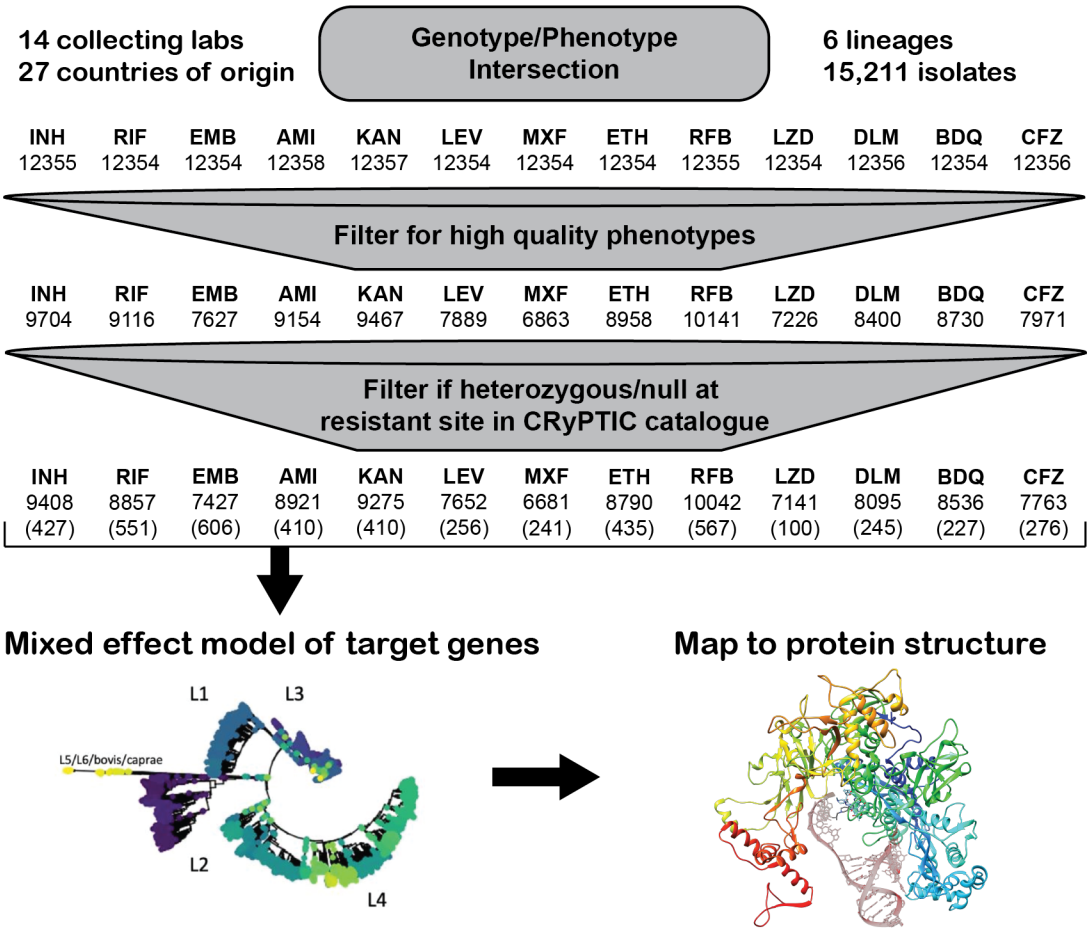


Figure 1: Data flow and sample sizes for CRyPTIC MIC models. Numbers in brackets represent the number of variables (mutations plus lineage and site effects) included in the final model.

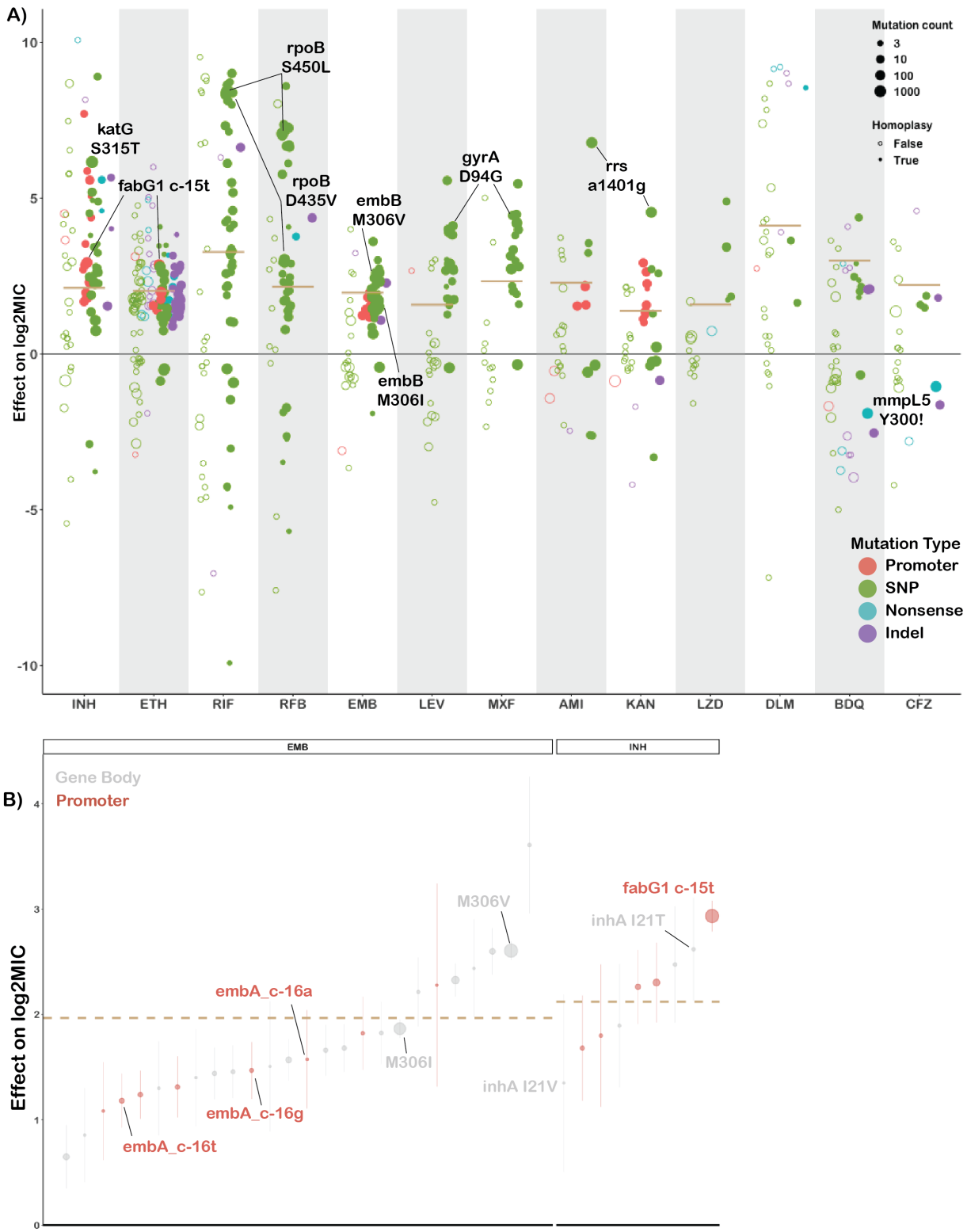


Figure 2: Variation in effect size by mutation type and drug. A) Effects on log2MIC for the 540 variants significant after false discovery rate correction using the Benjamini-

Hochberg method. Mutation types are delineated by color. Homoplastic mutations are shown as solid circles. ECOFF is shown as tan line. Common resistance mutations are highlighted. **B)** Comparison of effects on log2MIC for promoter and corresponding gene body variants for ethambutol (EMB) and isoniazid (INH). ECOFF (minus baseline MIC) is shown as a tan line.

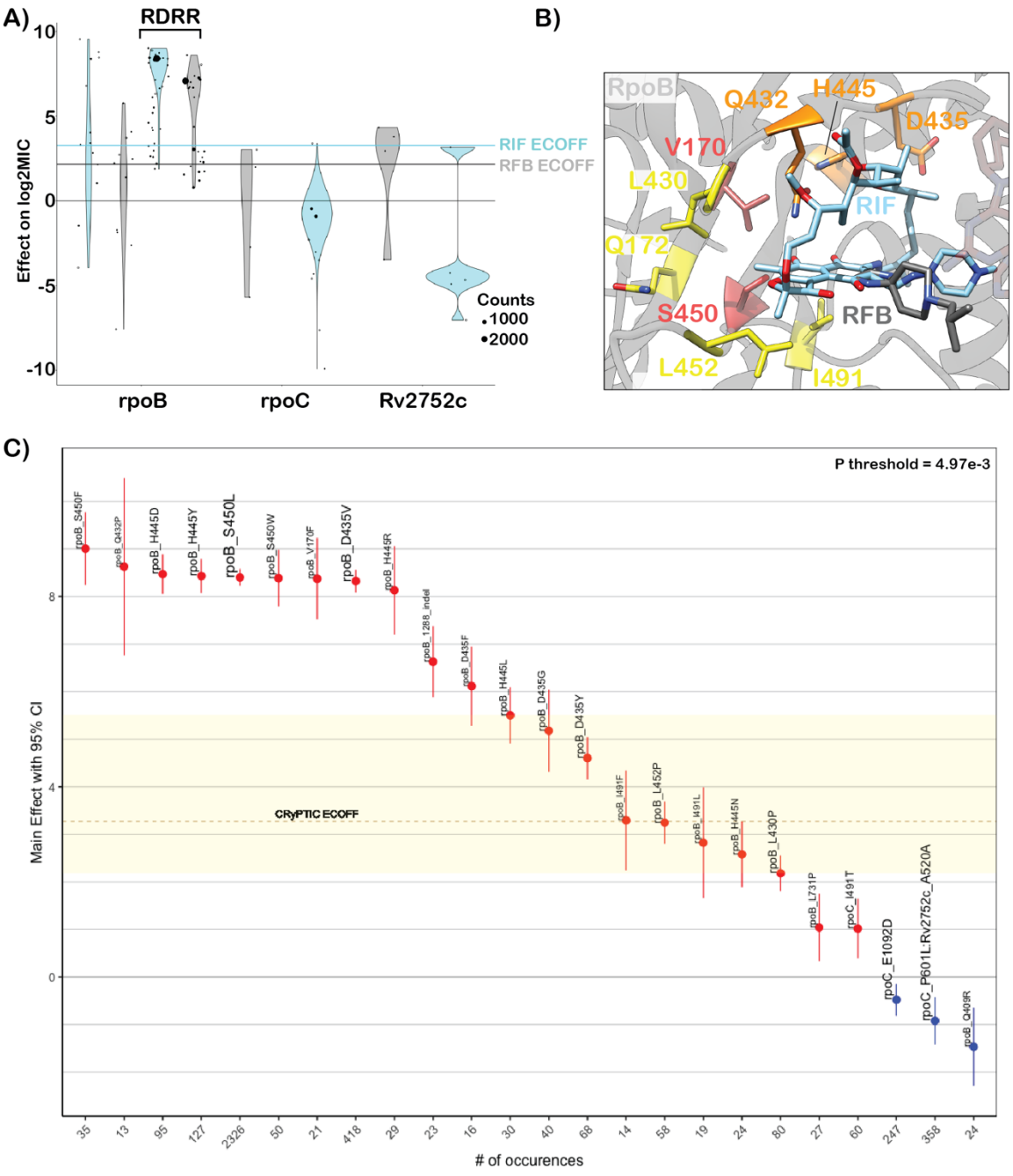


Figure 3: Heterogenous effects of *rpoB* mutations on rifampicin resistance. A) Effects of target gene variants on rifampicin (blue) and rifabutin (grey) log2MIC. ECOFFs are highlighted as lines. **B)** Rifampicin (blue) and rifabutin (grey) bound to rpoB with resistance-associated variants highlighted (red-high, orange-variable, yellow-

low). **C)** Effects on rifampicin MIC of mutations in rpoB. Text size is weighted by frequency of occurrence. Colored shading highlights “borderline” variants.

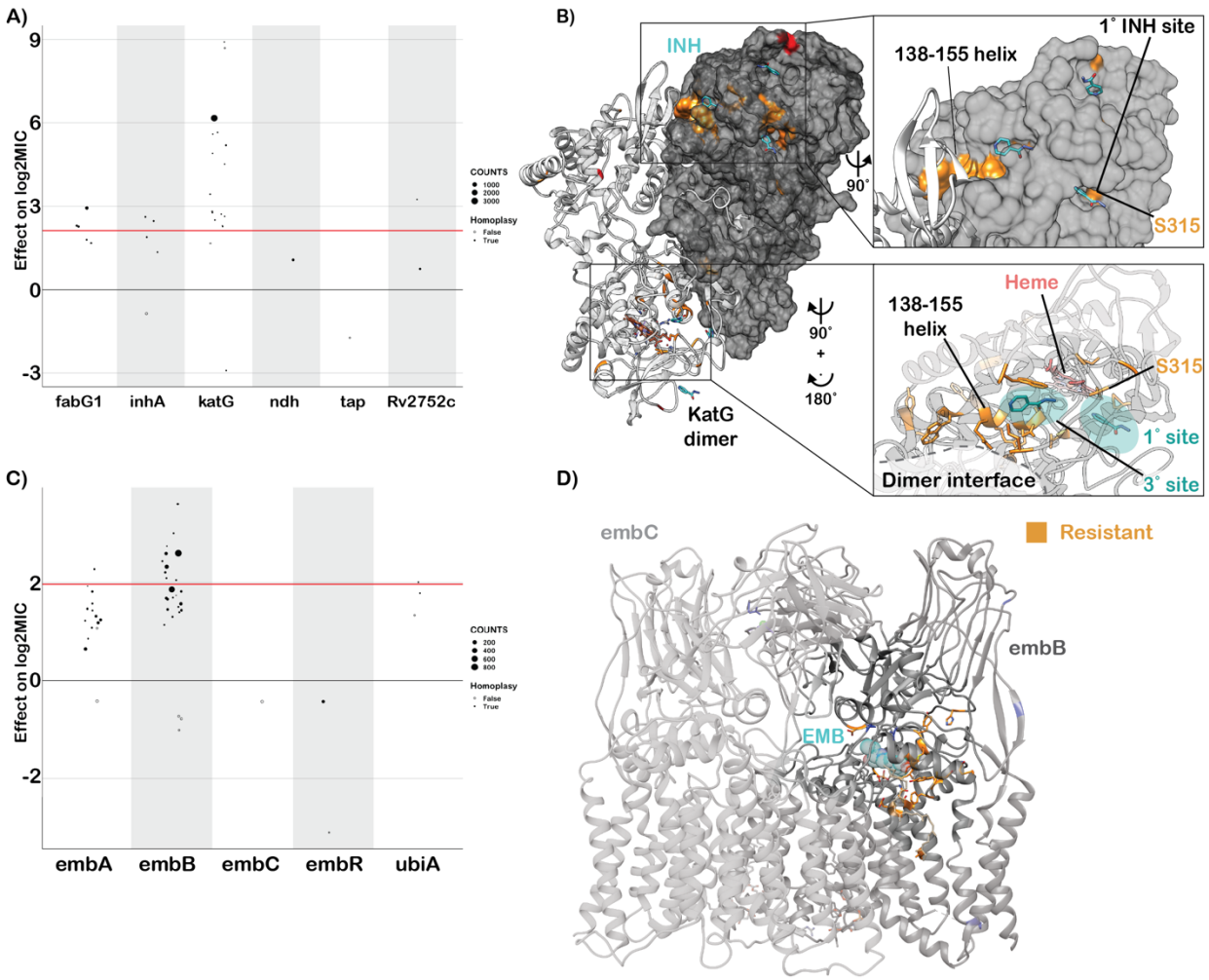


Figure 4: Resistance to isoniazid and ethambutol is a multi-gene phenomenon. A) Independent effects of variants in target genes on isoniazid log2MIC. ECOFF (minus baseline MIC) is denoted in red. **B)** KatG dimer with isoniazid (blue) modeled and resistance-associated positions highlighted in orange. **C)** Independent effects of variants in target genes on ethambutol log2MIC. ECOFF (minus baseline MIC) is denoted in red. **D)** EmbA-embB complex bound to ethambutol (blue) with resistant mutations highlighted in orange.

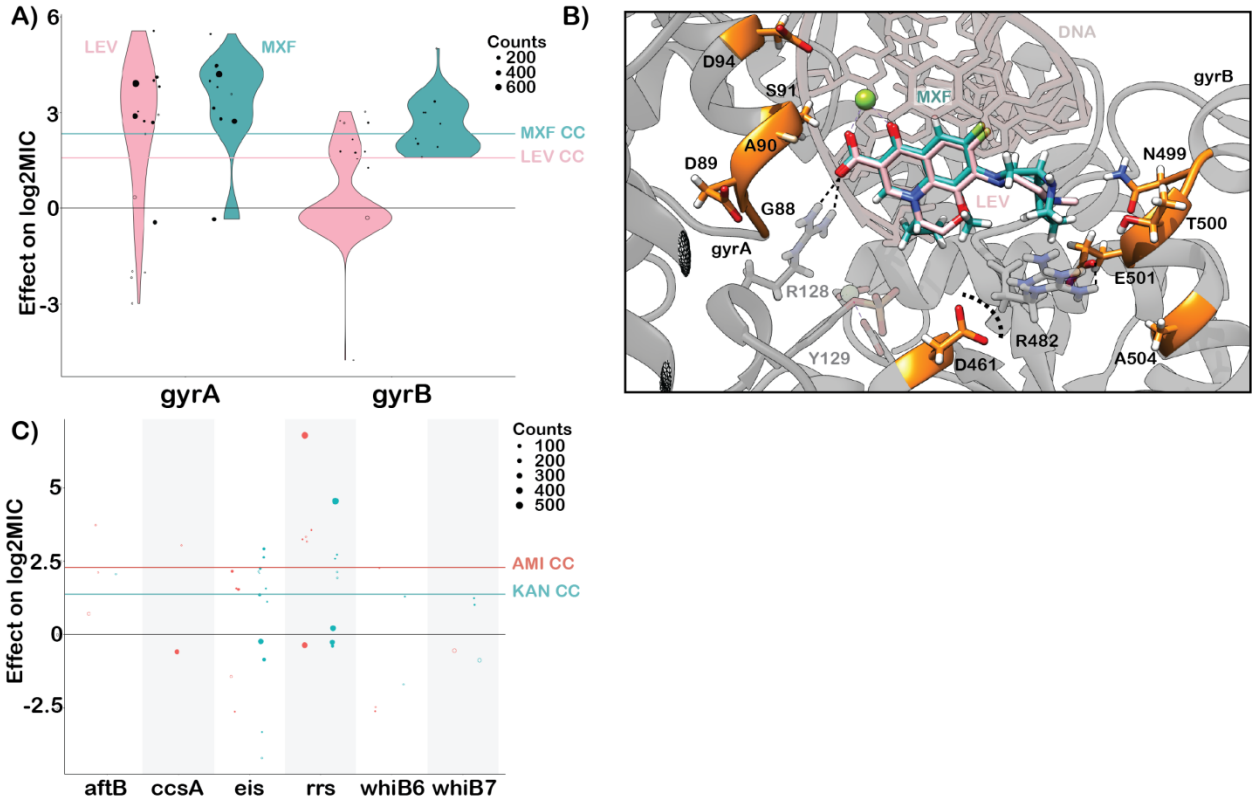


Figure 5: Resistance to second line drugs. A) Effects of mutations in gyrA and gyrB on levofloxacin (pink) and moxifloxacin (green) log2MIC. **B)** Structural mapping of fluoroquinolone resistance-associated variants reveal that majority lie within 10Å of the drug binding site. Positions gyrB R446 and S447 are not shown. **C)** Effects of mutations in aminoglycoside target genes on amikacin and kanamycin log2MIC.

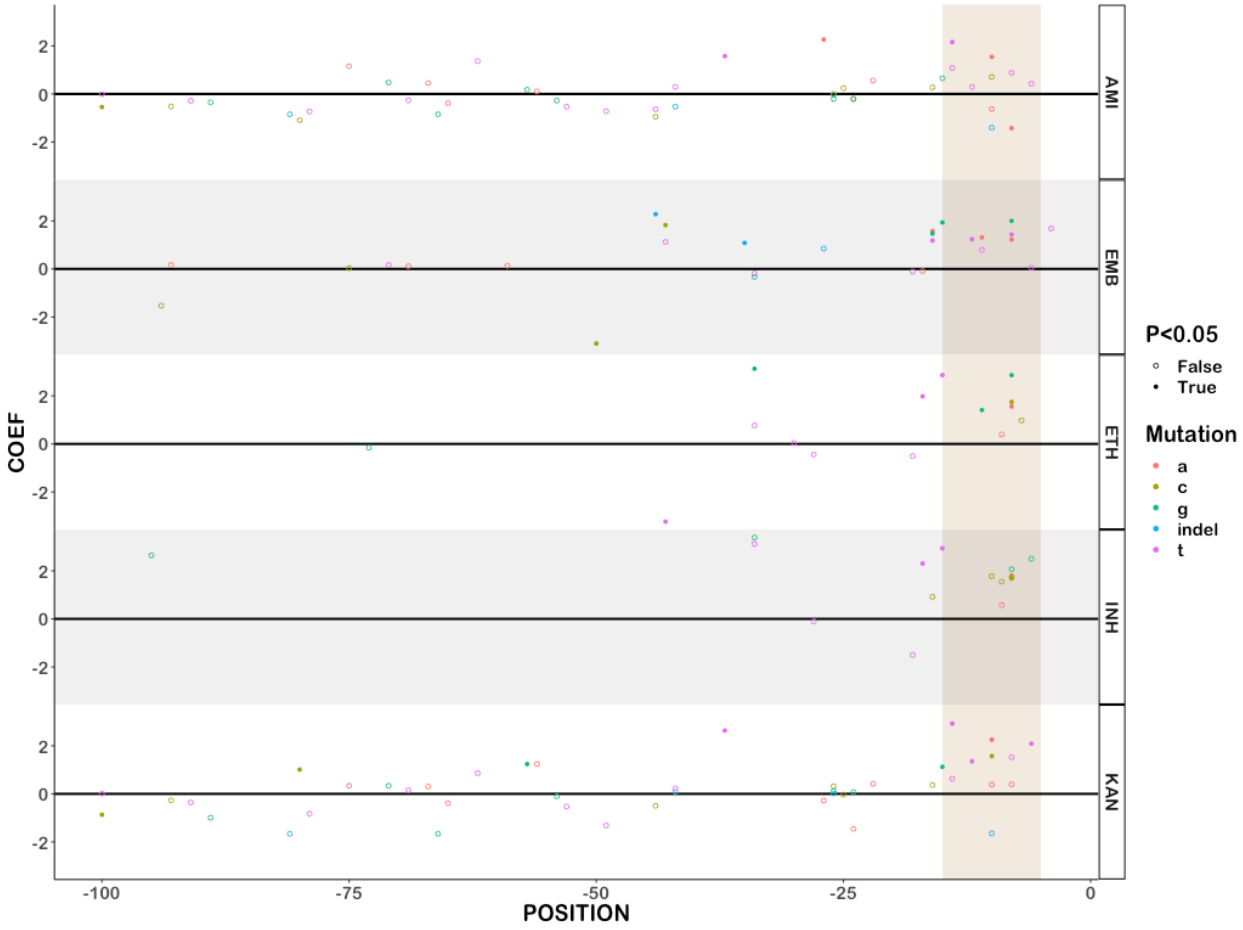
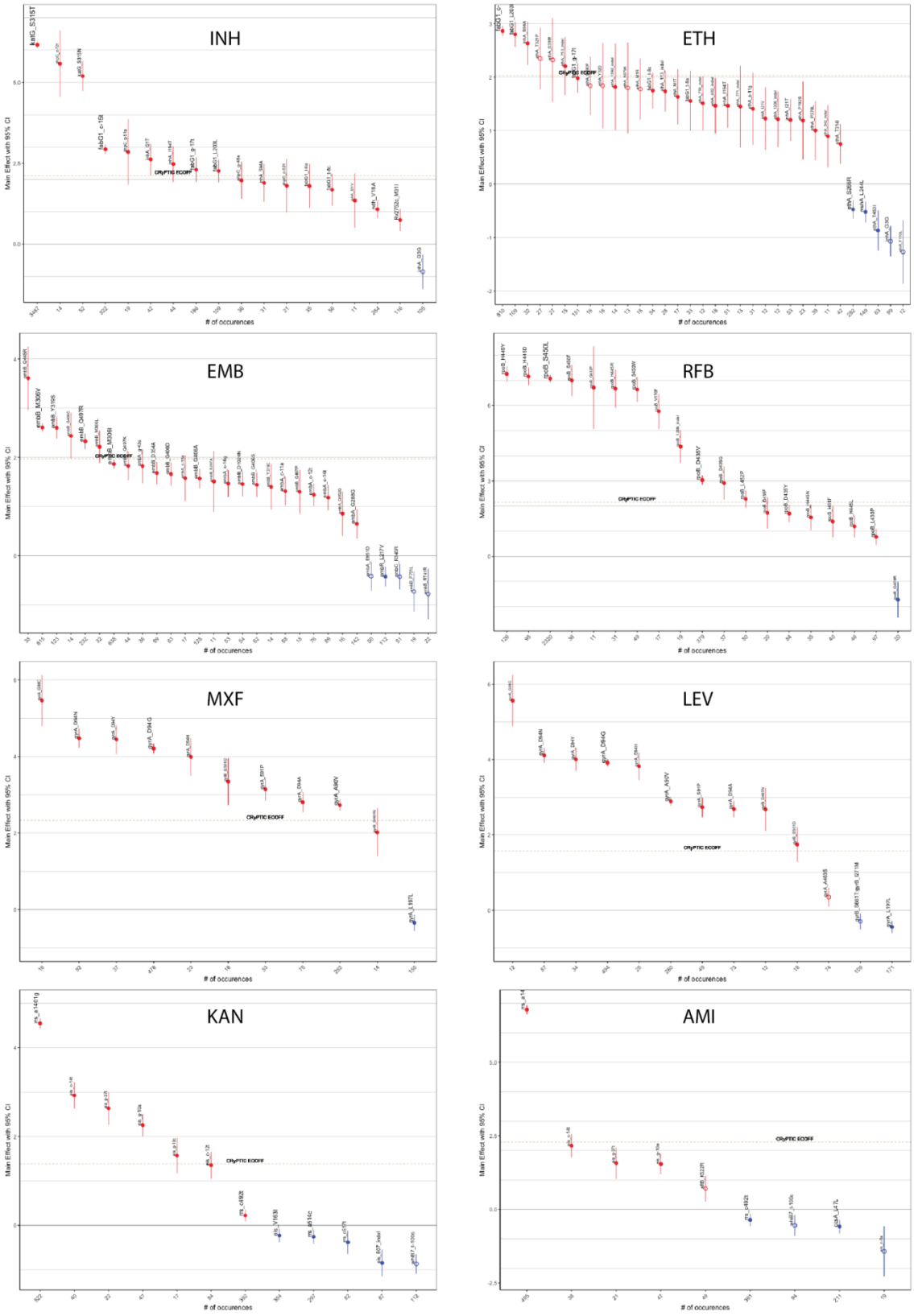


Figure S1: Resistant promoter mutations are enriched near the -10 element.

Promoter mutations are shown by position and colored by the mutant nucleotide. Mutations with significant effects ($p<0.05$) after Benjamini Hochberg correction are shown in solid circles. The -15 to -5 region is highlighted in tan.


INH
ETH
EMB
RFB
MXF
LEV
KAN
AMI
Main Effect with 95% CI
# of occurences
CRyPTIC ECOFF

1

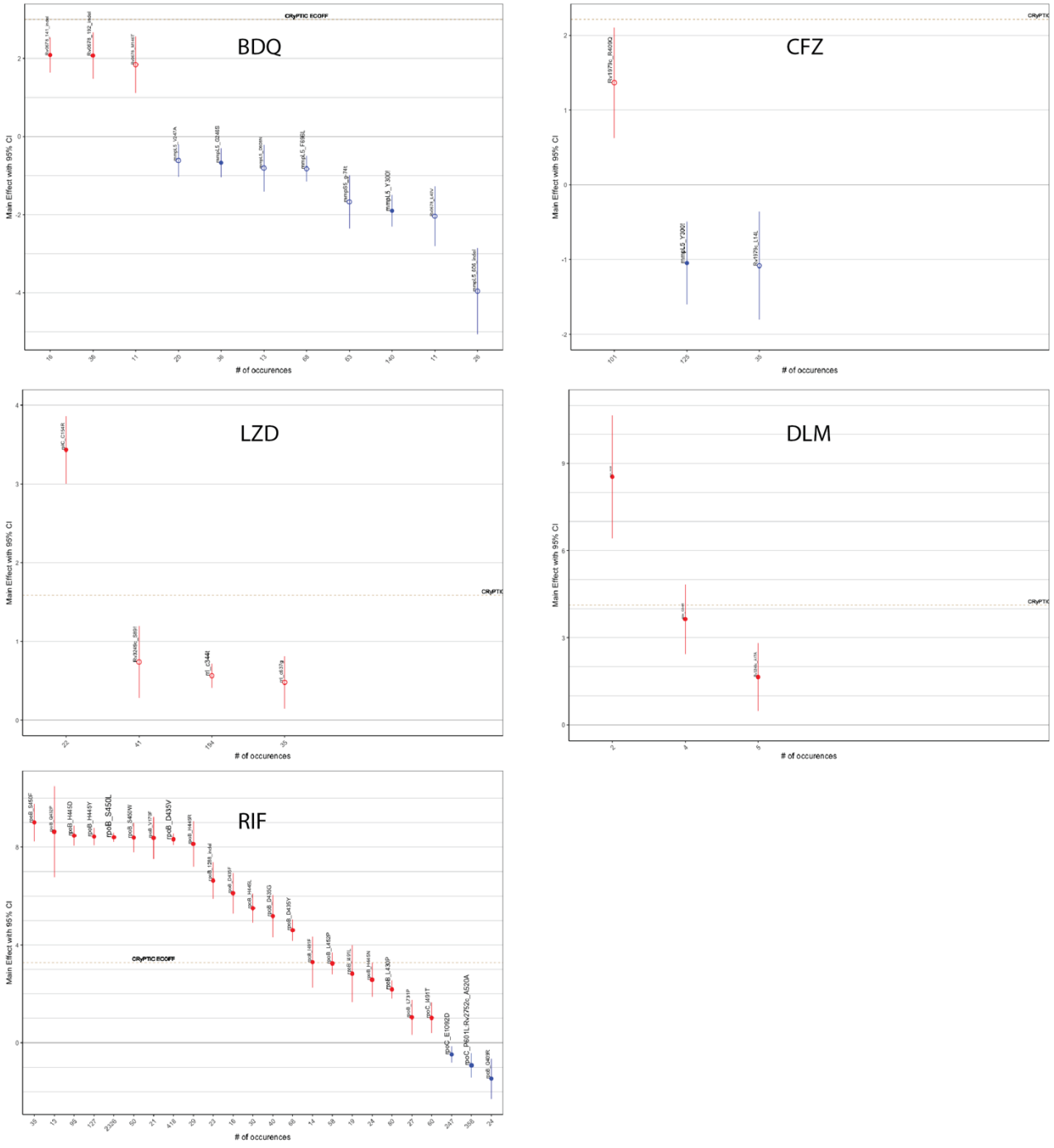


Figure S2: Effects of common (n>10) and/or homoplastic mutations on 13 antituberculosis drugs. Homoplastic mutations are solid circles and ECOFFs are shown as tan lines.

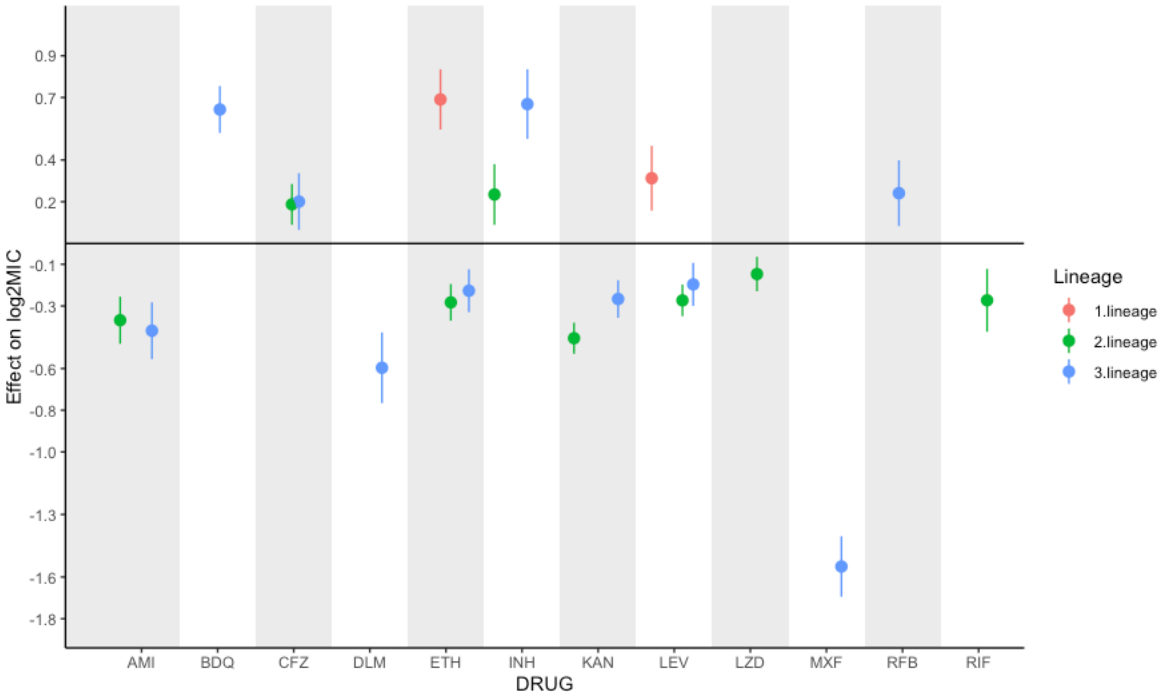


Figure S3: Genetic background has variable influences on strain MIC by drug.

Significant effects of lineage on MIC compared with lineage 4 as reference. No significant effects were identified for lineage 6.

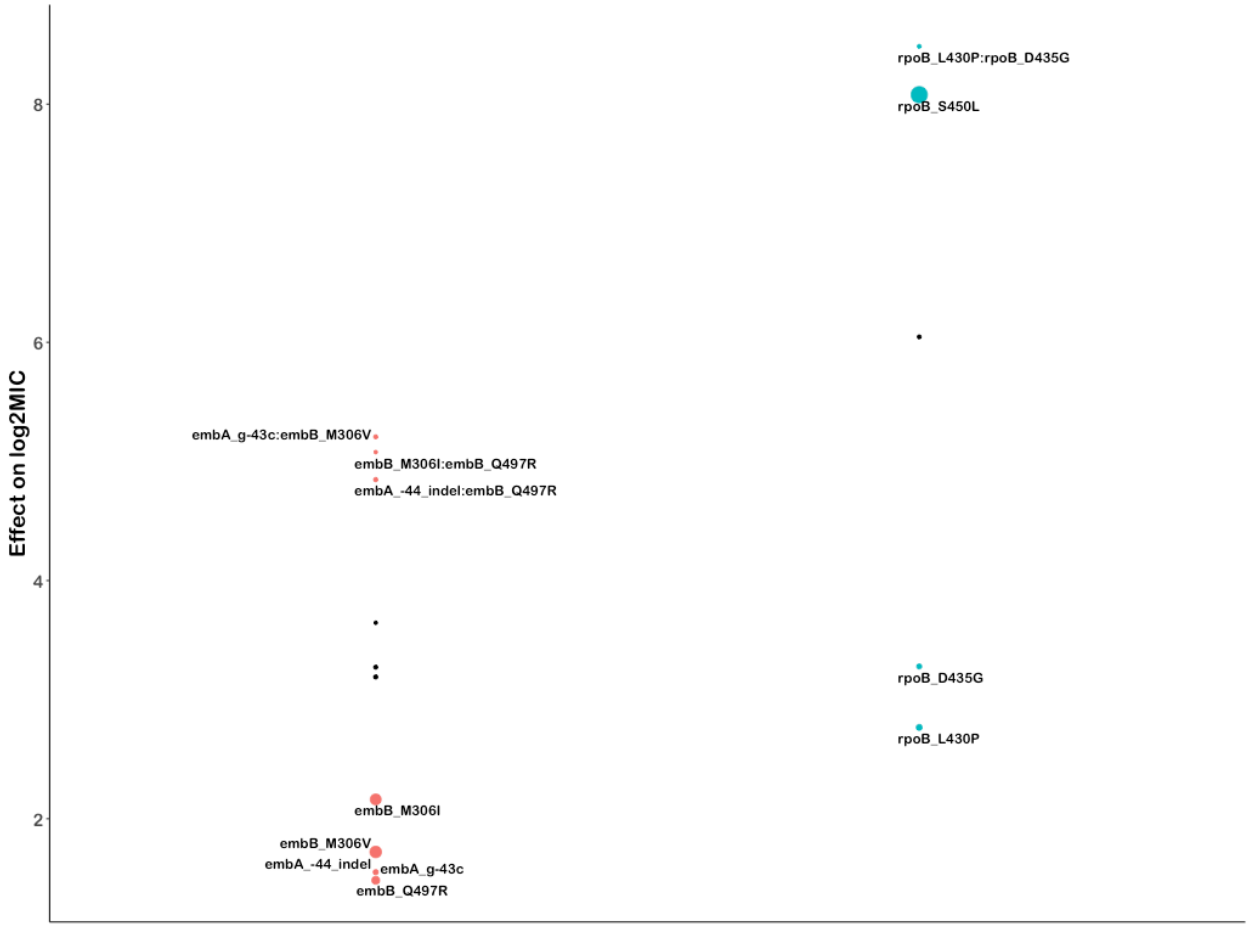


Figure S4: Mutation pairs cause interaction beyond additivity in ethambutol and rifampicin. Individual, predicted additive (black), and actual combined effects of mutation pairs on log2MIC. Common single resistance mutations are provided for reference.

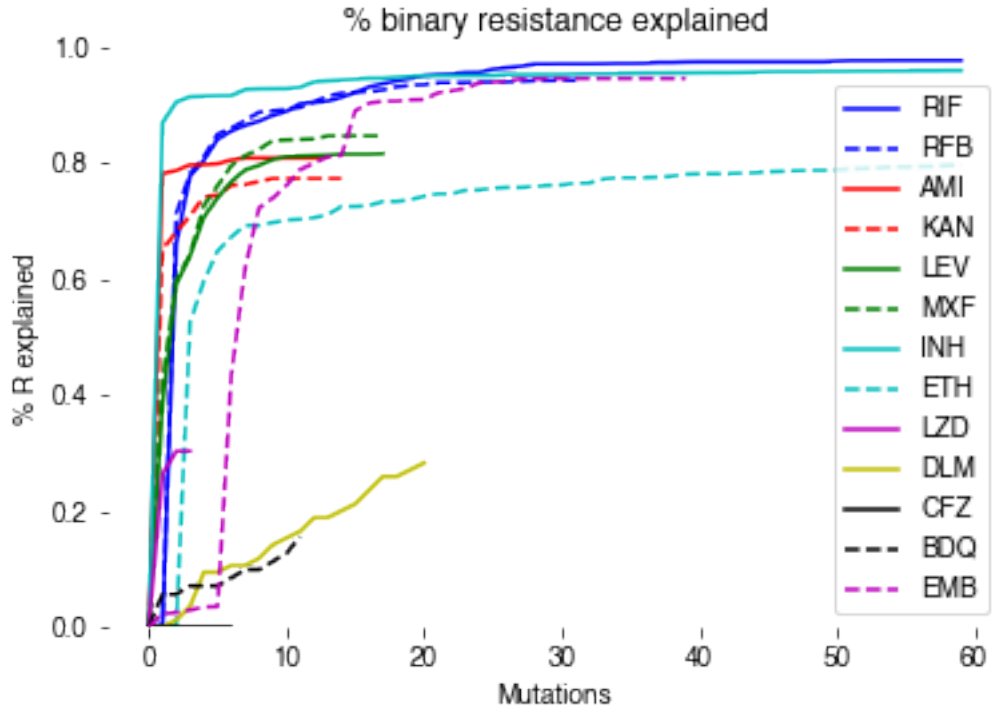


Figure S5: Heritability of binary phenotype for 13 antituberculosis drugs. The proportion of binary resistance explained by each additional mutation is shown by drug. Curves asymptotically approaching horizontal indicate that few major mutations in catalog genes remain to be described.

Methods:

Dataset collection

The CRyPTIC dataset collection and processing has been previously described in detail[17]. Briefly, clinical isolates were sub-cultured before inoculation into CRyPTIC-designed 96-well microtiter plates manufactured by ThermoFisher. Plates contained doubling-dilution ranges for 14 different antibiotics (para-aminosalicylic acid was excluded from the study due to poor-quality results on the plate). Isolate MICs were read after 14 days by a laboratory scientist using a Thermo Fisher Sensititre Vizion digital MIC viewing system and an image of the plate was also uploaded to a bespoke web server, allowing for additional MIC measurements by an automated computer vision system (AMYGDA) and by citizen science volunteers (Bash the Bug Zooniverse project) as previously described[53,54]. MIC measurements were classified as high (all three methods agree), medium (only two methods agree), or low (no methods agree). While sequencing processes differed slightly between CRyPTIC laboratories, all sequencing was performed using Illumina. A bespoke processing pipeline took in paired FASTQ files before filtering, mapping, and providing variant calls for each isolate. Isolates that had both phenotypic and whole-genome sequencing data were used as a starting dataset for this study (Clockwork availble from: https://github.com/iqbal-lab-org/clockwork, more detailed description of pipeline available in [16]). ECOFFs were defined elsewhere and are provided in Table S1[17].

Target gene selection

Target genes were selected based on the results of a prior study and through a literature search for each drug[55].

Statistical modeling

All genetic variation smaller than 50bp occurring in the target genes for each drug (Table S1) were included as candidates for effects in this study. Large insertions, deletions, and other structural changes larger than 50bp were not included in this study. Insertion/deletion mutations that occurred at the same position were pooled as one candidate effect. Mutations that always co-occurred in the dataset were combined into one candidate effect with all mutations named. Isolates were excluded from analysis if they contained evidence for mixed alleles at positions previously associated with resistance to that drug (i.e. a mixed allele call for position S450 in *rpoB* for rifampicin)[21]. Interval regression was performed in Stata version 16.1 with a genomic cluster variable (cluster ID where clustering was based on the SNP distance calculated from the whole genome sequence) as a random effect to control for population structure. A sensitivity analysis was performed to compare the effects of clustering at 12, 25, 50, and 100 single nucleotide polymorphism (SNP) distances (100 used for all results shown). Lineage and laboratory performing the MICs (SITEID variable) were included as factor variables to control for genetic and technical variation in each individual drug model. MICs were encoded as the interval with upper bound log2(MIC) and lower bound log2(MIC minus 1 doubling dilution). The bottom and top wells were extended by 3 doubling dilutions to account for censoring.

The Benjamini-Hochberg correction was used to adjust raw p-values and the false discovery rate was set at 5% for each drug based on the number of variants considered, including all variants in one mutually adjusted multivariable model. Pairs of mutations that occurred at least three times with each individual mutation occurring at least 5 times were subsequently tested for interactions in a mixed effect interval regression model containing all other variants for that drug reaching the significance threshold (Benjamini-Hochberg adjusted p-value < 0.05).

Data preparation, analysis and figure-making

Data was prepared for analysis using Python, statistical outputs were analyzed using R, and figures were made using ggPlot2 in R[56]. Homoplasy was calculated by the number of unique sub-lineages (predicted by SNP-IT and MYKROBE), with a mutation considered homoplastic if it had evolved in at least 2 independent sublineages[18–20]. A file that recapitulates all the post-model analysis and figures is available in the Supplemental material (Supplemental Code). Structural modeling was done using UCSF Chimera[57].

Author Contributions:

Members of the CRyPTIC consortium collected, phenotyped, and sequenced all isolates in the CRyPTIC dataset. JC, ASW, and TMW designed this study, JC performed statistical analyses and structural mapping, JC wrote the manuscript, JC, PWF, TEAP, TMW, and ASW revised the manuscript with all partners providing feedback, and PWF, TMW, ASW, and DWC supervised the work.

Acknowledgements – funders

This work was supported by Wellcome Trust/Newton Fund-MRC Collaborative Award (200205/Z/15/Z); and Bill & Melinda Gates Foundation Trust (OPP1133541). Oxford CRyPTIC consortium members are funded/supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC), the views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health, and the National Institute for Health Research (NIHR) Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, a partnership between Public Health England and the University of Oxford, the views expressed are those of the authors and not necessarily those of the NIHR, Public Health England or the Department of Health and Social Care. J.M. is supported by the Wellcome Trust (203919/Z/16/Z). Z.Y. is supported by the National Science and Technology Major Project, China Grant No. 2018ZX10103001. K.M.M. is supported by EMBL's EIPOD3 programme funded by the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska Curie Actions. T.C.R. is funded in part by funding from Unitaid Grant No. 2019-32-FIND MDR. R.S.O. is supported by FAPESP Grant No. 17/16082-7. L.F.

received financial support from FAPESP Grant No. 2012/51756-5. B.Z. is supported by the National Natural Science Foundation of China (81991534) and the Beijing Municipal Science & Technology Commission (Z201100005520041). N.T.T.T. is supported by the Wellcome Trust International Intermediate Fellowship (206724/Z/17/Z). G.T. is funded by the Wellcome Trust. R.W. is supported by the South African Medical Research Council. J.C. is supported by the Rhodes Trust and Stanford Medical Scientist Training Program (T32 GM007365). A.L. is supported by the National Institute for Health Research (NIHR) Health Protection Research Unit in Respiratory Infections at Imperial College London. S.G.L. is supported by the Fonds de Recherche en Santé du Québec. C.N. is funded by Wellcome Trust Grant No. 203583/Z/16/Z. A.V.R. is supported by Research Foundation Flanders (FWO) under Grant No. G0F8316N (FWO Odysseus). G.M. was supported by the Wellcome Trust (098316, 214321/Z/18/Z, and 203135/Z/16/Z), and the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation (NRF) of South Africa (Grant No. 64787). The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of this report. The opinions, findings and conclusions expressed in this manuscript reflect those of the authors alone. L.G. was supported by the Wellcome Trust (201470/Z/16/Z), the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number 1R01AI146338, the GOSH Charity (VC0921) and the GOSH/ICH Biomedical Research Centre (www.nihr.ac.uk). A.B. is funded by the NDM Prize Studentship from the Oxford Medical Research Council Doctoral Training Partnership and the Nuffield Department of Clinical Medicine. D.J.W. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant No. 101237/Z/13/B) and by the Robertson

Foundation. A.S.W. is an NIHR Senior Investigator. T.M.W. is a Wellcome Trust Clinical Career Development Fellow (214560/Z/18/Z). A.S.L. is supported by the Rhodes Trust. R.J.W. receives funding from the Francis Crick Institute which is supported by Wellcome Trust, (FC0010218), UKRI (FC0010218), and CRUK (FC0010218). T.C. has received grant funding and salary support from US NIH, CDC, USAID and Bill and Melinda Gates Foundation. The computational aspects of this research were supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z and the NIHR Oxford BRC. Parts of the work were funded by the German Center of Infection Research (DZIF). The Scottish Mycobacteria Reference Laboratory is funded through National Services Scotland. The Wadsworth Center contributions were supported in part by Cooperative Agreement No. U60OE000103 funded by the Centers for Disease Control and Prevention through the Association of Public Health Laboratories and NIH/NIAID grant AI-117312. Additional support for sequencing and analysis was contributed by the Wadsworth Center Applied Genomic Technologies Core Facility and the Wadsworth Center Bioinformatics Core. SYNLAB Holding Germany GmbH for its direct and indirect support of research activities in the Institute of Microbiology and Laboratory Medicine Gauting. N.R. thanks the Programme National de Lutte contre la Tuberculose de Madagascar.

Competing Interest

E.R. is employed by Public Health England and holds an honorary contract with Imperial College London. I.F.L. is Director of the Scottish Mycobacteria Reference Laboratory. S.N. receives funding from German Center for Infection Research, Excellenz Cluster Precision Medicine in Chronic Inflammation, Leibniz Science Campus Evolutionary Medicine of the LUNG

(EvoLUNG)tion EXC 2167. P.S. is a consultant at Genoscreen. T.R. is funded by NIH and DoD and receives salary support from the non-profit organization FIND. T.R. is a co-founder, board member and shareholder of Verus Diagnostics Inc, a company that was founded with the intent of developing diagnostic assays. Verus Diagnostics was not involved in any way with data collection, analysis or publication of the results. T.R. has not received any financial support from Verus Diagnostics. UCSD Conflict of Interest office has reviewed and approved T.R.'s role in Verus Diagnostics Inc. T.R. is a co-inventor of a provisional patent for a TB diagnostic assay (provisional patent #: 63/048.989). T.R. is a co-inventor on a patent associated with the processing of TB sequencing data (European Patent Application No. 14840432.0 & USSN 14/912,918). T.R. has agreed to "donate all present and future interest in and rights to royalties from this patent" to UCSD to ensure that he does not receive any financial benefits from this patent. S.S. is working and holding ESOPs at HaystackAnalytics Pvt. Ltd. (Product: Using whole genome sequencing for drug susceptibility testing for Mycobacterium tuberculosis).

Wellcome Trust Open Access

*This research was funded in part, by the Wellcome Trust/N*ewton Fund-MRC Collaborative Award [200205/Z/15/Z]. F*or the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.*

This research was funded, in part, by the Wellcome Trust [214321/Z/18/Z, and 203135/Z/16/Z]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Acknowledgements – people

JC would like to thank Spencer Dunleavy (Columbia Medical School, New York City, USA). We thank Faisal Masood Khanzada and Alamdar Hussain Rizvi (NTRL, Islamabad, Pakistan), Angela Starks and James Posey (Centers for Disease Control and Prevention, Atlanta, USA), and Juan Carlos Toro and Solomon Ghebremichael (Public Health Agency of Sweden, Solna, Sweden).

Ethics Statement

Approval for CRyPTIC study was obtained by Taiwan Centers for Disease Control IRB No. 106209, University of KwaZulu Natal Biomedical Research Ethics Committee (UKZN BREC) (reference BE022/13) and University of Liverpool Central University Research Ethics Committees (reference 2286), Institutional Research Ethics Committee (IREC) of The Foundation for Medical Research, Mumbai (Ref nos. FMR/IEC/TB/01a/2015 and FMR/IEC/TB/01b/2015), Institutional Review Board of P.D. Hinduja Hospital and Medical Research Centre, Mumbai (Ref no. 915-15-CR [MRC]), scientific committee of the Adolfo Lutz Institute (CTC-IAL 47-J / 2017) and in the Ethics Committee (CAAE: 81452517.1.0000.0059) and Ethics Committee review by Universidad Peruana Cayetano Heredia (Lima, Peru) and LSHTM (London, UK).

Members of the CRyPTIC consortium (in alphabetical order)

Ivan Barilar[29], Simone Battaglia[1], Emanuele Borroni[1], Angela Pires Brandao[2,3], Alice Brankin[4], Andrea Maurizio Cabibbe[1], Joshua Carter[5], Daniela Maria Cirillo[1], Pauline Claxton[6], David A

Clifton[4], Ted Cohen[7], Jorge Coronel[8], Derrick W Crook[4], Viola Dreyer[29], Sarah G Earle[4], Vincent Escuyer[9], Lucilaine Ferrazoli[3], Philip W Fowler[4], George Fu Gao[10], Jennifer Gardy[11], Saheer Gharbia[12], Kelen Teixeira Ghisi[3], Arash Ghodousi[1,13], Ana Luíza Gibertoni Cruz[4], Louis Grandjean[33], Clara Grazian[14], Ramona Groenheit[44], Jennifer L Guthrie[15,16], Wencong He[10], Harald Hoffmann[17,18], Sarah J Hoosdally[4], Martin Hunt[19,4], Zamin Iqbal[19], Nazir Ahmed Ismail[20], Lisa Jarrett[21], Lavania Joseph[20], Ruwen Jou[22], Priti Kambli[23], Rukhsar Khot[23], Jeff Knaggs[19,4], Anastasia Koch[24], Donna Kohlerschmidt[9], Samaneh Kouchaki[4,25], Alexander S Lachapelle[4], Ajit Lalvani[26], Simon Grandjean Lapierre[27], Ian F Laurenson[6], Brice Letcher[19], Wan-Hsuan Lin[22], Chunfa Liu[10], Dongxin Liu[10], Kerri M Malone[19], Ayan Mandal[28], Mikael Mansjö[44], Daniela Matias[21], Graeme Meintjes[24], Flávia de Freitas Mendes[3], Matthias Merker[29], Marina Mihalic[18], James Millard[30], Paolo Miotto[1], Nerges Mistry[28], David Moore[31,8], Kimberlee A Musser[9], Dumisani Ngcamu[20], Hoang Ngoc Nhung[32], Stefan Niemann[29, 48], Kayzad Soli Nilgiriwala[28], Camus Nimmo[33], Nana Okozi[20], Rosangela Siqueira Oliveira[3], Shaheed Vally Omar[20], Nicholas Paton[34], Timothy EA Peto[4], Juliana Maira Watanabe Pinhata[3], Sara Plesnik[18], Zully M Puyen[35], Marie Sylvianne Rabodoarivelo[36], Niaina Rakotosamimanana[36], Paola MV Rancoita[13], Priti Rathod[21], Esther Robinson[21], Gillian Rodger[4], Camilla Rodrigues[23], Timothy C Rodwell[37,38], Aysha Roohi[4], David Santos-Lazaro[35], Sanchi Shah[28], Thomas Andreas Kohl[29], Grace Smith[21,12], Walter Solano[8], Andrea Spitaleri[1,13], Philip Supply[39], Utkarsha Surve[23], Sabira Tahseen[40], Nguyen Thuy Thuong Thuong[32], Guy Thwaites[32,4], Katharina Todt[18], Alberto Trovato[1], Christian Utpatel[29], Annelies Van Rie[41], Srinivasan Vijay[42], Timothy M Walker[4,32], A Sarah Walker[4], Robin Warren[43], Jim Werngren[44], Maria Wijkander[44], Robert J Wilkinson[45,46,26], Daniel J Wilson[4], Penelope Wintringer[19], Yu-Xin Xiao[22], Yang Yang[4], Zhao Yanlin[10], Shen-Yuan Yao[20], Baoli Zhu[47]

Institutions

1 IRCCS San Raffaele Scientific Institute, Milan, Italy

2 Oswaldo Cruz Foundation, Rio de Janeiro, Brazil

3 Institute Adolfo Lutz, São Paulo, Brazil

4 University of Oxford, Oxford, UK

5 Stanford University School of Medicine, Stanford, USA

6 Scottish Mycobacteria Reference Laboratory, Edinburgh, UK

7 Yale School of Public Health, Yale, USA

8 Universidad Peruana Cayetano Heredia, Lima, Perú

9 Wadsworth Center, New York State Department of Health, Albany, USA

10 Chinese Center for Disease Control and Prevention, Beijing, China

11 Bill & Melinda Gates Foundation, Seattle, USA

12 UK Health Security Agency, London, UK

13 Vita-Salute San Raffaele University, Milan, Italy

14 University of New South Wales, Sydney, Australia

15 The University of British Columbia, Vancouver, Canada

16 Public Health Ontario, Toronto, Canada

17 SYNLAB Gauting, Munich, Germany

18 Institute of Microbiology and Laboratory Medicine, IMLred, WHO-SRL Gauting, Germany

19 EMBL-EBI, Hinxton, UK

20 National Institute for Communicable Diseases, Johannesburg, South Africa

21 Public Health England, Birmingham, UK

22 Taiwan Centers for Disease Control, Taipei, Taiwan

23 Hinduja Hospital, Mumbai, India

24 University of Cape Town, Cape Town, South Africa

25 University of Surrey, Guildford, UK

26 Imperial College, London, UK

27 Université de Montréal, Canada

28 The Foundation for Medical Research, Mumbai, India

29 Research Center Borstel, Borstel, Germany

30 Africa Health Research Institute, Durban, South Africa

31 London School of Hygiene and Tropical Medicine, London, UK

32 Oxford University Clinical Research Unit, Ho Chi Minh City, Viet Nam

33 University College London, London, UK

34 National University of Singapore, Singapore

35 Instituto Nacional de Salud, Lima, Perú

36 Institut Pasteur de Madagascar, Antananarivo, Madagascar

37 FIND, Geneva, Switzerland

38 University of California, San Diego, USA

39 Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 9017 - CIIL - Center for Infection and Immunity of Lille, F-59000 Lille, France

40 National TB Reference Laboratory, National TB Control Program, Islamabad, Pakistan

41 University of Antwerp, Antwerp, Belgium

42 University of Edinburgh, Edinburgh, UK

43 Stellenbosch University, Cape Town, South Africa

44 Public Health Agency of Sweden, Solna, Sweden

45 Wellcome Centre for Infectious Diseases Research in Africa, Cape Town, South Africa

46 Francis Crick Institute, London, UK

47 Institute of Microbiology, Chinese Academy of Sciences, Beijing, China

48 German Center for Infection Research (DZIF), Hamburg-Lübeck-Borstel-Riems, Germany

References:

1. WHO. *Global Tuberculosis Report 2020*. (2020).
2. Schnippel, K., Firnhaber, C., Berhanu, R., Page-Shipp, L. & Sinanovic, E. Adverse drug reactions during drug-resistant TB treatment in high HIV prevalence settings: A systematic review and meta-analysis. *J. Antimicrob. Chemother.* **72**, 1871–1879 (2017).
3. André, E. *et al.* Novel rapid PCR for the detection of Ile491Phe rpoB mutation of Mycobacterium tuberculosis, a rifampicin-resistance-conferring mutation undetected by commercial assays. *Clin. Microbiol. Infect.* **23**, 267.e5-267.e7 (2017).
4. Cox, H., Hughes, J., Black, J. & Nicol, M. P. Precision medicine for drug-resistant tuberculosis in high-burden countries: is individualised treatment desirable and feasible? *Lancet Infect. Dis.* **18**, e282–e287 (2018).
5. Pankhurst, L. J. *et al.* Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: A prospective study. *Lancet Respir. Med.* **4**, 49–58 (2016).
6. Walker, T. M. *et al.* Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: A retrospective cohort study. *Lancet Infect. Dis.* **15**, 1193–1202 (2015).
7. World Health Organization, (WHO). Technical report on critical concentrations for TB drug susceptibility testing of medicines used in the treatment of drug-resistant TB. *Who* 106 (2018).
8. Colangeli, R. *et al.* Bacterial factors that predict relapse after tuberculosis therapy.

N. Engl. J. Med. **379**, 823–833 (2018).

9. McCallum, A. D. & Sloan, D. J. The importance of clinical pharmacokinetic–pharmacodynamic studies in unraveling the determinants of early and late tuberculosis outcomes. *Int. J. Pharmacokinet.* **2**, 195–212 (2017).

10. Grobbelaar, M. *et al.* Evolution of rifampicin treatment for tuberculosis. *Infect. Genet. Evol.* **74**, 103937 (2019).

11. Walsh, K. F. *et al.* Improved Outcomes with High-dose Isoniazid in Multidrug-resistant Tuberculosis Treatment in Haiti. *Clin. Infect. Dis.* **69**, 717–719 (2019).

12. Decroo, T. *et al.* High-dose first-line treatment regimen for recurrent rifampicin-susceptible tuberculosis. *Am. J. Respir. Crit. Care Med.* **201**, 1578–1579 (2020).

13. Farhat, M. R. *et al.* GWAS for quantitative resistance phenotypes in Mycobacterium tuberculosis reveals resistance genes and regulatory regions. *Nat. Commun.* **10**, (2019).

14. Rancoita, P. M. V. *et al.* Validating a 14-drug microtiter plate containing bedaquiline and delamanid for large-scale research susceptibility testing of mycobacterium tuberculosis. *Antimicrob. Agents Chemother.* (2018). doi:10.1128/AAC.00344-18

15. Falzon, D. *et al.* World Health Organization treatment guidelines for drug-resistant tuberculosis, 2016 update. *Eur. Respir. J.* **49**, (2017).

16. WHO. *Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance*. (2021).

17. The CRyPTIC Consortium. Epidemiological cutoffs for a 96-well broth microtitre plate for high-throughput research antibiotic susceptibility testing of M .

tuberculosis. *medRxiv. doi:10.1101/2021.02.24.21252386* (2021). doi:10.1101/2021.02.24.21252386

18. Lipworth, S. *et al.* SNP-IT tool for identifying subspecies and associated lineages of Mycobacterium tuberculosis complex. *Emerg. Infect. Dis.* **25**, 482–488 (2019).

19. Bradley, P. *et al.* Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. *Nat. Commun.* **6**, 1–14 (2015).

20. Hunt, M. *et al.* Antibiotic resistance prediction for Mycobacterium tuberculosis from genome sequence data with Mykrobe. *Wellcome Open Res.* **4**, 191 (2019).

21. The CRyPTIC Consortium & The 100000 Genomes Project. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. *N. Engl. J. Med.* **379**, 1403–1415 (2018).

22. Newton-Foot, M. & Gey Van Pittius, N. C. The complex architecture of mycobacterial promoters. *Tuberculosis* **93**, 60–74 (2013).

23. Zhu, D. X., Garner, A. L., Galburt, E. A. & Stallings, C. L. CarD contributes to diverse gene expression outcomes throughout the genome of Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 13573–13581 (2019).

24. Makhado, N. A. *et al.* Outbreak of multidrug-resistant tuberculosis in South Africa undetected by WHO-endorsed commercial tests: an observational study. *Lancet Infect. Dis.* (2018). doi:10.1016/S1473-3099(18)30496-1

25. Beckert P, Sanchez-Padilla P, Merker M, Dreyer V, Kohl TA, Utpatel C, Köser CU, Barilar I, Ismail N, Omar SV, Klopper M, Warren RM; Hoffmann H, Maphalala G, Ardizzoni E, de Jong BC, Kerschberger B, Schramm B, Andres S, Kranzer K,

Maurer FP, Bonnet M, N. S. MDR M. tuberculosis outbreak clone in Eswatini missed by Xpert has elevated bedaquiline resistance dated to the pre-treatment era. *Genome Med.* **12**, (2020).

26. Brandis, G. & Hughes, D. Mechanisms of fitness cost reduction for rifampicin-resistant strains with deletion or duplication mutations in rpoB. *Sci. Rep.* **8**, 1–6 (2018).

27. World Health Organization. *Technical Report on critical concentrations for drug susceptibility testing of isoniazid and the rifamycins (rifampicin, rifabutin and rifapentine)*. (2021).

28. Torrea, G. *et al.* Variable ability of rapid tests to detect Mycobacterium tuberculosis rpoB mutations conferring phenotypically occult rifampicin resistance. *Sci. Rep.* **9**, 1–9 (2019).

29. Miotto, P., Cabibbe, A. M., Borroni, E., Degano, M. & Cirilloa, D. M. Role of disputed mutations in the rpoB gene in interpretation of automated liquid MGIT culture results for rifampin susceptibility testing of mycobacterium tuberculosis. *J. Clin. Microbiol.* **56**, 1–9 (2018).

30. Jeong, D. H. *et al.* Successful treatment with a high-dose rifampin-containing regimen for pulmonary tuberculosis with a disputed rpob mutation. *Intern. Med.* **57**, 3281–3284 (2018).

31. Ma, P. *et al.* Compensatory effects of M. tuberculosis rpoB mutations outside the rifampicin resistance-determining region. *Emerg. Microbes Infect.* **10**, 743–752 (2021).

32. Loewen, P. C., Switala, J., Smolenski, M. & Triggs-Raine, B. L. Molecular

characterization of three mutations in katG affecting the activity of hydroperoxidase I of Escherichia coli. *Biochem. Cell Biol.* **68**, 1037–1044 (1990).

33. Munir, A. *et al.* Using cryo-EM to understand antimycobacterial resistance in the catalase-peroxidase (KatG) from Mycobacterium tuberculosis. *Structure* **29**, 1–14 (2020).

34. Hicks, N. D. *et al.* Clinically prevalent mutations in Mycobacterium tuberculosis alter propionate metabolism and mediate multidrug tolerance. *Physiol. Behav.* **176**, 139–148 (2019).

35. Safi, H. *et al.* Evolution of high-level ethambutol-resistant tuberculosis through interacting mutations in decaprenylphosphoryl-β-D-Arabinose biosynthetic and utilization pathway genes. *Nat. Genet.* **45**, 1190–1197 (2013).

36. Zhang, L. *et al.* Structures of cell wall arabinosyltransferases with the anti-tuberculosis drug ethambutol. *Science (80-.).* **9102**, eaba9102 (2020).

37. Disratthakit, A. *et al.* Role of gyrB mutations in pre-extensively and extensively drug-resistant tuberculosis in Thai clinical isolates. *Antimicrob. Agents Chemother.* **60**, 5189–5197 (2016).

38. Maruri, F. *et al.* A systematic review of gyrase mutations associated with fluoroquinolone-resistant Mycobacterium tuberculosis and a proposed gyrase numbering system. *J. Antimicrob. Chemother.* **67**, 819–831 (2012).

39. Vargas Jr, R. *et al.* The role of epistasis in amikacin, kanamycin, bedaquiline, and clofazimine resistance in Mycobacterium tuberculosis complex. *bioRxiv* (2020).

40. Farhat, M. R. *et al.* Rifampicin and rifabutin resistance in 1003 Mycobacterium tuberculosis clinical isolates. *J. Antimicrob. Chemother.* **74**, 1477–1483 (2019).

41. Nebenzahl-Guimaraes, H., Jacobson, K. R., Farhat, M. R. & Murray, M. B. Systematic review of allelic exchange experiments aimed at identifying mutations that confer drug resistance in Mycobacterium tuberculosis. *J. Antimicrob. Chemother.* **69**, 331–342 (2014).
42. Peterson, E. J. R., Ma, S., Sherman, D. R. & Baliga, N. S. Network analysis identifies Rv0324 and Rv0880 as regulators of bedaquiline tolerance in Mycobacterium tuberculosis. *Nat. Microbiol.* **1**, 6–12 (2016).
43. Kadura, S. *et al.* Systematic review of mutations associated with resistance to the new and repurposed Mycobacterium tuberculosis drugs bedaquiline , clofazimine , linezolid , delamanid and pretomanid. *J. Antimicrob. Chemother.* 2031–2043 (2020). doi:10.1093/jac/dkaa136
44. Lee, B. M. *et al.* Predicting nitroimidazole antibiotic resistance mutations in Mycobacterium tuberculosis with protein engineering. *PLoS Pathog.* **16**, 1–27 (2020).
45. Sonnenkalb, L. *et al.* Deciphering Bedaquiline and Clofazimine Resistance in Tuberculosis : An Evolutionary Medicine Approach. *bioRxiv doi:10.1101/2021.03.19.436148* 1–38 (2021). doi:10.1101/2021.03.19.436148
46. Yadon, A. N. *et al.* A comprehensive characterization of PncA polymorphisms that confer resistance to pyrazinamide. *Nat. Commun.* **8**, 1–10 (2017).
47. Portelli, S. *et al.* Prediction of rifampicin resistance beyond the RRDR using structure-based machine learning approaches. *Sci. Rep.* **10**, 1–13 (2020).
48. Carter, J. J. *et al.* Prediction of pyrazinamide resistance in Mycobacterium tuberculosis using structure-based machine learning approaches. *bioRxiv* (2019).

doi:10.1101/518142

49. Karmakar, M., Rodrigues, C. H. M., Horan, K., Denholm, J. T. & Ascher, D. B. Structure guided prediction of Pyrazinamide resistance mutations in pncA. *Sci. Rep.* **10**, 1–10 (2020).

50. Karmakar, M. *et al.* Empirical ways to identify novel Bedaquiline resistance mutations in AtpE. *PLoS One* **14**, 1–14 (2019).

51. Battaglia, S. *et al.* Characterization of genomic variants associated with resistance to bedaquiline and delamanid in naïve Mycobacterium tuberculosis clinical strains. *J. Clin. Microbiol.* **58**, 1–16 (2020).

52. Harms, A., Maisonneuve, E. & Gerdes, K. Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science (80-.).* **354**, (2016).

53. Rancoita, P. M. V. *et al.* Validating a 14-drug microtitre plate containing bedaquiline and delamanid for large-scale research susceptibility testing of *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* AAC.00344-18 (2018). doi:10.1128/AAC.00344-18

54. Fowler, P. W. *et al.* Automated detection of bacterial growth on 96-well plates for high-throughput drug susceptibility testing of Mycobacterium tuberculosis. *Microbiology* (2018). doi:10.1099/mic.0.000733

55. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. *N. Engl. J. Med.* **379**, 1403–1415 (2018).

56. Wickam, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

57. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory

research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).