Clinically prevalent mutations in *Mycobacterium tuberculosis* alter propionate metabolism and mediate multidrug tolerance

Nathan D. Hicks¹,², Jian Yang²,⁶, Xiaobing Zhang²,⁶, Bing Zhao³,⁶, Yonatan H. Grad¹,⁴, Liguo Liu², Xichao Ou³, Zhili Chang², Hui Xia³, Yang Zhou³, Shengfen Wang³, Jie Dong², Lilian Sun³, Yafang Zhu², Yanlin Zhao³*, Qi Jin²* and Sarah M. Fortune¹,⁵*

The global epidemic of drug-resistant tuberculosis is a catastrophic example of how antimicrobial resistance is undermining the public health gains made possible by combination drug therapy. Recent evidence points to unappreciated bacterial factors that accelerate the emergence of drug resistance. In a genome-wide association study of *Mycobacterium tuberculosis* isolates from China, we find mutations in the gene encoding the transcription factor prpR enriched in drug-resistant strains. prpR mutations confer conditional drug tolerance to three of the most effective classes of antibiotics by altering propionyl-CoA metabolism. prpR-mediated drug tolerance is carbon-source dependent, and while readily detectable during infection of human macrophages, is not captured by standard susceptibility testing. These data define a previously unrecognized and clinically prevalent class of *M. tuberculosis* variants that undermine antibiotic efficacy and drive drug resistance.

Clinical isolates of the pathogens *Pseudomonas aeruginosa* and *Candida albicans* with increased production of persister cells have been identified, suggesting that this phenotype can emerge during treatment¹¹,¹². In *M. tuberculosis*, persister cells have been postulated to contribute to the long duration of antibiotic therapy required to treat infection¹¹,¹². In vitro studies in *M. tuberculosis* and other prokaryotes have also identified forms of drug tolerance including blunted drug efficacy due to population-wide slow-growth states in specific conditions¹⁶,¹⁷, and conditional drug efflux¹⁸. Furthermore, bacterial mutants that are specifically drug tolerant in vivo but drug susceptible in vitro have been isolated in laboratory studies; however, the mechanism(s) of in vivo drug tolerance has not been defined¹⁹. Importantly, studies in *E. coli* have shown that persistence and tolerance mutations can accelerate the subsequent emergence of drug resistance¹⁹,²¹.

In an effort to identify variants altering antibiotic susceptibility in natural populations of *M. tuberculosis*, we undertook a genome-wide association study of clinical isolates, correlating mutations with drug resistance. We identify several loci associated with resistance and demonstrate that variants in the gene encoding the bacterial transcription factor prpR confer conditional multidrug tolerance.

### Results

**Genetic variants associated with drug-resistant TB in China.**

To identify clinically relevant bacterial variants altering antibiotic efficacy in *M. tuberculosis*, we performed a bacterial genome-wide association analysis in strains isolated from TB patients. Antibiotic tolerance is not measured clinically but may serve as a stepping stone to the evolution of resistance²¹. We therefore...
reasoned that associations with drug resistance should also capture tolerance mutations that increase the probability that a strain will subsequently develop resistance. To perform this analysis, we leveraged a strain collection assembled during the 2007 national survey of drug-resistant TB in China. In this study, ~4,000 strains were collected from across China and resistance was measured for 6 first- and second-line antibiotics. A subset of 549 strains was selected for whole-genome sequencing including all isoniazid-resistant and MDR strains and a geographically matched set of drug-susceptible strains (Fig. 1a and Supplementary Table 1).

To define the phylogenetic structure of *M. tuberculosis* in China, we performed whole-genome sequencing and identified single nucleotide polymorphisms (SNPs) by comparing each strain with the laboratory strain H37Rv (Supplementary Table 1). The phylogenetic structure of *M. tuberculosis* in our cohort is broadly consistent with published phylogenies21-24 (Fig. 1b and Supplementary Fig. 1). As expected, drug-resistant *M. tuberculosis* in China is dominated by lineage 2 strains, accounting for 81% of drug-resistant isolates, the majority of which are part of sublineage L2.3 (also known as ‘modern’ Beijing). Chinese lineage 4 strains comprise 18% of drug-resistant isolates and form three distinct groups, which are generally restricted to China, as assessed by the distribution of publicly available *M. tuberculosis* genomes from other geographic locations (Fig. 1b and Supplementary Table 2). A genetic distance-based cluster analysis demonstrated that ~90% of isoniazid-resistant isolates in this sample reflect independent acquisitions of drug resistance (Supplementary Fig. 2). As the existing collections of sequenced drug-resistant *M. tuberculosis* are frequently dominated by epidemic clones25,26, this is a uniquely robust cohort for the identification of bacterial factors driving the acquisition of drug resistance.

To detect genetic associations in the context of the completely clonal *M. tuberculosis* population structure, we identified mutations occurring in isoniazid-resistant isolates more often than would be expected by chance based on phylogenetic convergence at a gene-wide level17, using an algorithm that we termed phyOverlap. Isoniazid resistance typically arises first in the progression to MDR28 and second-line drugs are used only in the context of pre-existing MDR, leading to programmatic linkage among drug-resistance phenotypes (Supplementary Table 3). Thus, by performing an association analysis with isoniazid resistance, we expected to capture mutations altering the efficacy of multiple clinically utilized antibiotics.

From this analysis, we identified 13 genes and inter-genic regions that are significantly (q < 0.05) associated with isoniazid resistance (Fig. 1c and Supplementary Table 4). The associations with the highest significance included the well-established isoniazid-resistance-determining genes and regulatory regions (*katG, inhA* promoter and *ahpC* promoter mutations). As expected, known resistance-determining and compensatory mutations for other first- and second-line antibiotics are also significantly associated with isoniazid resistance, reflecting the multiple resistances present in many MDR isolates. Among these high-confidence associations, an association was identified with *prpR* or Rv1129c (q = 0.018) for which there is no known mechanistic basis for the coincidence of mutations and drug resistance.

Convergence-based phyOverlap analysis of two additional publicly available data sets from a Chinese cohort29 and a mixed UK/African cohort30 (Supplementary Tables 5 and 6), as well as a meta-analysis of all three data sets, independently supported the association between *prpR* and isoniazid resistance (Fig. 1d). The meta-analysis also identified several additional known drug-resistance-associated loci, including *ethA*, *gid* and the *es* promoter. The most significant association in this analysis for which a mechanistic basis is unknown is Rv2752c, encoding a potentially bifunctional protein that has been demonstrated to have both RNase J-like activity and beta-lactamase activity31,32, and which frequently contains nonsense or INDEL mutations presumably abolishing function. Four additional genes were identified that failed to meet the corrected genome-wide significance threshold; however, they were nested within four known resistance-associated loci (Fig. 1d). These genes include *hcyY* (Rv2092c) and *dnaA*, both implicated in DNA replication and repair; Rv3402c, which encodes a conserved hypothetical protein; and Rv1830, which encodes another putative transcription factor (Fig. 1d).

The distribution of *prpR* mutations in global *Mycobacterium tuberculosis* populations. Although *prpR* mutants are strongly associated with drug resistance in our analysis, only a single previous genomic study, also performed using Chinese isolates, has implicated *prpR* in resistance29. To gain a broader context for the incidence of *prpR* mutation among clinical isolates, we assessed the frequency of *prpR* mutants within the three data sets included in Fig. 1d as well as in three additional large genomic data sets from other regions of the world23-29. In these three additional global data sets, we also performed targeted genetic analysis to measure the association between *prpR* mutations and mutations in *katG* and the *inhA* promoter regions as a proxy for isoniazid resistance (Table 1). Overall, the occurrence of *prpR* mutants in samples derived from outside China is markedly lower, representing just 2–3% of isolates overall as compared with 8–10% in Chinese isolates. In all data sets except for Casali et al.34 from Russia, the uncorrected frequency of *prpR* mutations is higher in the drug-resistant isolates than in drug-susceptible isolates. After correcting for phylogenetic structure as described above, *prpR* mutations are significantly associated with drug resistance in all data sets except the Malawian cohort, which has relatively little drug resistance overall. The discrepancies in the Russian and Malawian cohorts probably reflect the effects of clonal drug-resistant outbreaks, which strongly skew the frequency estimates.

Functional characterization of clinical *prpR* mutations. *prpR* has no previously known role in resistance to any antibiotic, although in an observational study examining the treatment of a single MDR patient over time, a *prpR* mutant was driven to near fixation while the patient was treated with a second-line drug regimen36. Based on parsimony analysis, the phylogenetic distribution of *prpR* mutants displays convergent evolution of particular SNPs indicative of positive selection (Fig. 2a and Supplementary Table 7). Although *prpR* mutations are significantly enriched in drug-resistant isolates, mutations including those identified as convergent also occur in 5% (10/200) of pan-susceptible isolates (Fig. 2b), suggesting that they do not directly cause drug resistance in standard culture conditions.

We therefore sought to define the basis of the association between *prpR* mutations and drug resistance. *prpR* is a transcription factor that is divergently transcribed from the *prpDC* operon in *M. tuberculosis*. The *prpD* and *prpC* genes encode proteins that mediate the first two steps in the methylcitrate cycle (MCC), which converts propionyl-CoA into pyruvate (Fig. 2c). Following exposure to propionate, or propionyl-CoA generating carbon sources such as cholesterol, which is essential for *M. tuberculosis* growth in vivo37, *prpDC* is strongly upregulated in a *prpR*-dependent fashion38,39. Deletion mutants of either *prpDC* or *prpR*, which are deficient in MCC activity, are unable to grow in medium containing cholesterol or propionate either as a sole carbon source or as components of mixed carbon sources, perhaps due to accumulation of unmetabolized propionyl-CoA40.

To directly assess the functional consequence of the identified *prpR* mutations, we constructed an isogenic panel of *prpR* SNP mutants by complementing a *prpR* deletion in the laboratory strain H37Rv with a chromosomally integrated copy of *prpR* containing the wild-type sequence, three common point mutations identified by our analysis or an empty vector. In liquid media containing
acetate and glycerol as carbon sources, all strains grew at similar rates (Fig. 2d). However, in media containing propionate and glycerol, ΔprpR::vector was strongly attenuated for growth while ΔprpR::D160G and ΔprpR::G361C demonstrated an extended lag phase before resuming growth approximately six days after exposure. In contrast, the deletion mutant complemented with wild-type prpD or E135G grew at the same rate as H37Rv (Fig. 2e). Consistent with these observations, prpD upregulation in response to propionate exposure was completely abrogated in the ΔprpR::vector strain and highly attenuated in strains expressing the D160G and G361C variants (Fig. 2f). We considered the possibility that the late growth of the D160G and G361C variants reflected the outgrowth
of suppressor mutants; however, when bacteria from propionate-containing media were recovered in standard 7H9 media and then re-exposed to propionate, they remained sensitive to propionate, exhibiting the same extended lag in growth (Supplementary Fig. 3). Taken together, these data demonstrate that at least a subset of common prpR mutants are hypomorphic for MCC activity and have altered growth characteristics in propionate-containing media.

**prpR mutants exhibit conditional drug tolerance in vitro.** We then assessed the effect of prpR mutations on both resistance and tolerance to isoniazid, rifampin and the second-line fluoroquinolone ofloxacin. Although propionate supplementation itself unexpectedly increased rifampin resistance in all strains, the antibiotic-dependent inhibition of growth in prpR mutants did not differ from wild-type *M. tuberculosis* for any drug regardless of the

---

**Table 1 | The frequency of prpR mutants among publicly available data sets and their association with isoniazid resistance**

| Location       | Sample size | Isoniazid resistant | Isoniazid sensitive | Total prpR mutants | Percentage of all isolates with a prpR mutation | Percentage of isoniazid-resistant isolates with a prpR mutation | Percentage of isoniazid-sensitive isolates with a prpR mutation | phyOverlap | P value |
|----------------|-------------|---------------------|---------------------|--------------------|-----------------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|-------------|--------|
| Current study  | China       | 549                 | 349                 | 200                | 55                                           | 10                                                          | 12.9                                                       | 5           | 0.00004 |
| Zhang et al.   | China       | 161                 | 117                 | 44                 | 13                                           | 8.1                                                         | 11.1                                                       | 0           | 0.0012 |
| Walker et al.  | Africa/UK   | 784                 | 415                 | 396                | 16                                           | 2                                                           | 3.1                                                        | 0.8         | 0.015  |
| Casali et al.  | Russia      | 1,253               | 756*                | 497                | 28                                           | 2.2                                                         | 2.2                                                        | 2.2         | 0.007  |
| Guerra-        | Malawi      | 1,847               | 147*                | 1,700*             | 54                                           | 2.9                                                         | 8.8                                                        | 2.4         | 0.089  |
| Assunção et al.| Vietnam     | 1,635               | 436*                | 1,199*             | 43                                           | 2.6                                                         | 5                                                          | 1.7         | 0.00152|

*aResistant and sensitive are designated using mutations within katG or the inhA promoter as a proxy for resistance.
Fig. 3 | prpR mutants display conditional multidrug tolerance. a–c, Drug-resistance measurement of growth in varying concentrations of antibiotic normalized to a no-drug control for isoniazid (INH; a), ofloxacin (OFLX; b) and rifampin (RIF; c). The experiment was performed three times with similar results. d, Survival of prpR mutants in a library of strains over time in the indicated antibiotic and medium conditions. The fraction of survival represents the bulk library c.f.u. divided into the strain c.f.u. by relative abundance measurement. Each point represents the mean and standard deviation of three independent measurements. e–g, Bacterial survival after six days of treatment in single strain assays with isoniazid (e), ofloxacin (f) and rifampin (g) in media with acetate (clear background marked with A) or propionate (shaded background marked with P). Each dot represents the fraction of bacteria surviving six days of antibiotic treatment in a biological replicate with the mean and standard deviation indicated with bars. N = 3 for each strain in each condition. *P < 0.05 and **P < 0.01 indicate differences in log-transformed survival determined by Tukey’s multiple-comparison test after two-way ANOVA.
ΔprpR::G361C is >6 days in each case. The prolonged MDK<sub>90</sub> indicates population-wide drug tolerance and is consistent with visual inspection of the curves, where prpR mutants display slower overall killing kinetics rather than a higher plateau of persistor cells. Similarly, prpR mutants were also more tolerant of rifampin; however, the survival advantage was more modest, potentially reflecting the dampened killing of even the wild-type strains at this dose of rifampin in propionate. Under our assay conditions, the prpR mutants were eventually killed by high doses of isoniazid and ofloxacin at late time points that largely coincided with resumed growth of rifampin in propionate. Under our assay conditions, the prpR mutants displayed slower overgrowth and killing kinetics rather than a higher plateau of persistor cells. In human macrophages treated with isoniazid and ofloxacin at concentrations far above the minimum inhibitory concentration, the survival of the prpR mutants was significantly greater than that of ΔprpR::WT (Fig. 4b–d). Note that ΔprpR::WT was slightly more drug tolerant than H37Rv, probably reflecting incomplete complementation of the prpR deletion. Even the E135G prpR prpR mutation, which had little discernible effect during in vitro functional testing, was modestly but consistently more drug tolerant than ΔprpR::WT in human macrophages treated with isoniazid and rifampin (Fig. 4b,d). All of the prpR mutants were similarly more tolerant of all three antibiotics in the human-derived macrophage-like THP-1 cell line (Fig. 4e).}

**prpR tolerance depends on propionyl-CoA intoxication.** Although the best-defined role of prpR is regulation of prpC and prpD expression, prpR has been implicated in activation of icl-1, which mediates the conversion of methylisocitrate to succinate and pyruvate in the MCC, and ramB, which is itself a repressor of icl-1 expression in some conditions. prpR has also been proposed to have additional regulatory roles including as a transcriptional repressor of the lipid metabolism transcriptional regulator gene kstR, and the cell cycle regulator gene dnaA, which was independently associated with antibiotic resistance in our meta-analysis (Fig. 1d). Like prpC and prpD, icl-1 expression in propionate-containing media was strongly attenuated in the prpR mutants, as was ramB to a lesser extent. By contrast, expression of both kstR and dnaA was unaffected by prpR mutation under these conditions (Supplementary Fig. 6). Thus, the clearest consequence of the prpR mutations is altered expression of MCC components.

We therefore examined whether mutations in genes encoding other steps of the MCC are associated with drug resistance in our cohort. Mutations in prpC, which catalyses the first step of the MCC, are enriched in drug-resistant strains (8/9 prpC mutants are isoniazid resistant, phyOverlap P = 0.03) but because they are less common than prpR mutations, prpC failed to reach significance in the genome-wide screen. Mutations in prpD or icl-1 were not
Metabolic rescue of propionate sensitivity suppresses drug tolerance in prpR mutants. a, b, Growth of the indicated prpR mutants in propionate-containing media with or without vitamin B12. The data represent the mean and standard deviation of three biological replicates. c–f, Fitness of prpR mutants after antibiotic treatment in propionate-containing media with and without the supplementation of B12 normalized to input abundance and ΔprpR::WT. g, h, The normalized fitness of strains in THP-1 cells with or without vitamin B12 supplementation in the absence (g) or presence (h) of isoniazid. The dotted green line indicates ΔprpR::WT, which is set to 1 at each time point. The data represent the mean and standard deviation of three biological replicates. Significant differences in normalized fitness ± B12 supplementation were tested by Sidak’s multiple-comparison test after two-way ANOVA. *P < 0.05, **P < 0.01.

Discussion

Bacterial genomic association studies promise to provide unique insights into the key processes driving the progression and treatment response of human TB. In contrast to in vitro bacterial genetic screens, these approaches do not require a priori knowledge of the relevant environmental conditions or altered bacterial phenotypes that contribute to the final phenotypes measured in clinical settings. This represents both a strength and a weakness, as associations capture a wide array of mechanistic drivers that must be carefully dissected.

Here we performed a genome-wide association study with isoniazid resistance, leveraging clinical isolates from the Chinese national survey of drug-resistant TB and cohorts of M. tuberculosis strains collected in China, Europe and Africa. In addition to identifying many known genetic determinants of drug resistance, we define two statistically significant associations with drug resistance,
prpR and Rv2752c, putatively encoding RNase J, both of which had no known relationship with drug resistance. We validated our approach by defining the mechanistic basis for the association with prpR, a transcriptional regulator of propionate metabolism. Rather than causing drug resistance per se, we demonstrate that common prpR mutations confer multidrug tolerance against several mechanistically unrelated antibiotics.

In the case of prpR, these tolerance mutations have remained unrecognized because the effects of prpR mutations on antibiotic efficacy are not captured in standard drug susceptibility studies, which measure drug resistance, not drug tolerance, and are performed under only one growth condition. Nevertheless, the population-based association of prpR mutations with drug resistance provides a real-world demonstration of the stepping-stone nature of drug tolerance that has been described in E. coli in vitro.

However, from these data, it is difficult to determine which antibiotics might be most affected by prpR mutations in clinical practice. It is possible that, in TB patients, prpR mutations broadly slow bacterial clearance by the first-line regimen. Alternatively, these mutations may have greater effects on the efficacy of certain antibiotics, perhaps because of the in vivo pharmacology of the individual drugs. Moreover, there may be host or disease state factors that influence the impact of prpR mutations perhaps by altering the availability of propionate or vitamin B12. Our in vitro data further suggest that prpR point mutants may be most protective during acute exposure to metabolic stress, and during chronic exposure may eventually resume growth and become re-sensitized to antibiotics, although in mouse models, cholesterol metabolism appears most important during chronic infection. The extent to which M. tuberculosis moves in and out of these metabolic stresses in M. tuberculosis is unknown and could determine the degree of protection.

Mutations in prpR may also contribute directly to treatment failure even in the absence of acquired drug resistance. In our study, we hypothesize that prpR mutations may be most protective during acute exposure to metabolic stress, and during chronic exposure may eventually resume growth and become re-sensitized to antibiotics, although in mouse models, cholesterol metabolism appears most important during chronic infection. The extent to which M. tuberculosis moves in and out of these metabolic stresses in humans is unknown and could determine the degree of protection. This effect may account for the notable lack of any INDEL or nonsense mutations in prpR across any data set examined as part of our analyses, where mutations providing a balance between tolerance and growth are optimal for fitness during treatment. The extent to which these factors impact the frequency of prpR mutants in the population may also explain the uneven distribution and strength of association with prpR mutants across global data sets.

Regional differences in second-line drug regimens may also contribute to the discords between the published M. tuberculosis genome-wide association studies. This type of effect can clearly be seen for eas promoter mutations, which are known to cause kanamycin resistance. eas promoter mutations evolved only twice in our Chinese cohort and thus are not significantly associated with any drug resistances. By contrast, these mutations evolved 31 times in the European/African cohort where they were significantly associated with isoniazid resistance without being a causal determinant (Fig. 1d). Thus, regional differences in second-line drug usage may contribute to different associations, as could geographic differences in public health infrastructure that could impact drug access and quality. In either case, measuring the association with the true phenotype, rather than a correlate, would presumably decrease variability among study groups. Moreover, it is possible that certain mutations may be more common in some mycobacterial lineages or strains due to genetic epistasis, although we observe prpR mutations in both lineage 2 and lineage 4 backgrounds in our data set.

Finally, on the basis of the blunted efficacy of several antibiotics in our study, we hypothesize that prpR mutations may also contribute directly to treatment failure even in the absence of acquired drug resistance. As noted above, prpR mutations are found in 5% of drug-susceptible strains in China and 1–2% in other regions of the world. Other genes identified in this study also contain convergent mutations in drug-susceptible strains; for example, Rv2752c mutations are present in 4% of drug susceptible strains. Testing the hypothesis that these mutations may also contribute directly to treatment failure will require improved clinical end points for genetic association, ideally including metrics such as relapse and failure to convert to diagnostic culture negativity during treatment. While assaying drug tolerance under a range of growth conditions is not feasible in clinical practice, further characterization of these genetic associations will permit the identification of tolerant strains through molecular diagnostics and allow for targeted therapies and monitoring to help prevent the emergence of de novo drug resistances and improve treatment outcomes.

Methods

Strain selection and culture. A national survey of drug-resistant TB was conducted in China in 2007, which collected a total of 5,929 clinical M. tuberculosis strains from 70 sites in mainland China. The survey was performed under the authority of the National TB Control Program and was approved by the Tuberculosis Research Ethics Review Committee of the China CDC.

Written informed consent was obtained from each participant. Three subsets of the collection were included in the current study based on the following criteria: all MDR strains (defined as isolates that were resistant to at least isoniazid and rifampin); all isoniazid mono-resistant strains (defined as isolates that were resistant to isoniazid but sensitive to all of the other 5 drugs including rifampicin, streptomycin, ethambutol, ofloxacin and kanamycin); 3–5 randomly selected drug-sensitive strains (defined as isolates that were sensitive to all of the six drugs aforementioned) from each of the 70 collection sites. All isolates were cultured on Lowenstein–Jensen slants at 37 °C for 4–6 weeks. Strains that failed in bacterial revival or produced aborted/mixed colonies were excluded from further analysis.

In all, 288 MDR strains, 61 isoniazid mono-resistant strains and 200 drug-sensitive strains were available for follow-up whole-genome sequencing.

Sequencing, alignment and SNP calling. Genomic DNA was extracted from each isolate using standard methods, and was then sequenced on the Illumina GAIIx or HiSeq 2500 platforms, generating single-end reads of 72 base pairs (bp) or 101 bp in length. For each sample, duplicate reads were retained by custom Perl scripts. Further quality control was conducted using the NGSQC Toolkit with a cutoff of Q20. Valid reads were then aligned to the reference genome sequence of H37Rv (GenBank accession NC_000962) using the Burrows–Wheeler algorithm as implemented in BWA. SNPs were identified with a minimum depth of 10x and a consensus quality score of 50 using SAMtools. The publicly available complete/draft genomes involved in our phylogenetic analysis (Supplementary Table 2), the wgsim program from the SAMtools package was used to generate three million 72-bp single reads for each to generate an ~50x depth of simulated sequencing data for each genome and SNPs were identified as above. SNPs located within repetitive regions, including transposases, PE/PPE genes, prophages and exact sequence repeats of the H37Rv genome previously proposed were identified by RepeatMasker (http://www.repeatmasker.org/) or PhageFinder, were excluded. Mixed base calls were considered valid only if the numbers of the most abundant and second most abundant (n2) nucleotides at each SNP in each strain satisfied the criterion n1/n2 ≥ 5. The detailed workflow of the SNP-calling procedure of data sets for follow-up analysis is given in Supplementary Fig. 7.

Phylogenetic analysis. A concatenated superset of SNPs relative to H37Rv was generated across all 549 sequenced strains and 130 published M. tuberculosis genomes (Supplementary Table 1). SNP sites with missing data over 5% of the strains within the data set were removed. To avoid the potential effects of homoplasoy of drug-resistance-associated mutations in phylogeny, SNPs located in known drug-resistance-related genes available from TBDReaMD were further excluded from the data set for phylogenetic tree construction (Supplementary Fig. 7). A strain of Mycobacterium canetti was included as an outgroup (GenBank accession NC_015848). The refined SNP set was used to construct the maximum-likelihood phylogeny using RAxML under the GTRgamma substitution model. The reliability of each node was tested via a bootstrap analysis on 100 resampled data sets. The iTOL server and MEGA5 software were employed for the final visualization and presentation of the phylogenetic trees. To define clades, we first employed a threshold of 100 SNPs among isolates based on molecular clock estimates of ~0.5 SNPs accumulating per strain per year where larger distances would place the most recent common ancestor of a set of isolates before the widespread use of isoniazid. We also assessed the isoniazid-resistance-determining SNP present within each cluster to discriminate true isoniazid-resistant clusters from closely related isolates with distinct resistance-causing mutations.

Drug-resistance associations. Genomic associations were performed using a method similar to that of Farhat et al. to that we refer to as phyOverlap. We maximized pairwise parsimony ancestral sequence reconstruction to determine the state of each SNP residue at the common ancestor of our strain collection using the Mesquite software package (http://mesquiteproject.org/). In cases where the ancestral sequence could not be unambiguously determined, we excluded the site from further analysis. Each isolate was then scored as ancestral or derived for a given SNP locus. Isolates with ambiguous base calls were considered ancestral for this calculation to minimize the potential of error driving associations with drug
resistance. The overlap with drug resistance was scored by dividing the number of isolates containing a derived allele that also were resistant to isoniazid by the total number of isolates with a derived allele at a given SNP locus. To generate a gene-wide score, we excluded synonymous SNPs and averaged the individual SNP scores, weighting the scores by the number of times derived alleles evolved across the phylogenetic tree. The number of times a site mutated across the tree was calculated as the parsimony score at each SNP position using the Fitch algorithm as implemented in the R package phangorn. This was performed with mixed bases treated as missing data so that low-quality base calls would not appear highly convergent.

Significance testing was performed by redistributing the mutation events in each SNP in each gene randomly across the phylogenetic tree, with the probability of a mutation occurring on any branch proportional to the branch length and all offspring of a branch being assigned the derived allele for calculation of the parsimony score. The gene-wide score was then calculated and compared with the actual value. This process was repeated 50,000 times to derive an empirical P value. For all statistical calculations, associations that achieved a P value of 0.05 were considered significant.

Antibiotic-susceptibility testing. Minimum inhibitory concentrations were measured using the Alamar blue reduction assay. Strains were grown to mid-log phase, spun down and resuspended in 7H12 media with the indicated short-chain fatty acid. After two days of growth, each strain was diluted to OD\textsubscript{600} 0.0015 in 200 μl of the appropriate media with antibiotics at the indicated concentrations. Twenty microtiter plates of Alamar blue reagent was added at four days post drug exposure, and reduction was measured by OD\textsubscript{570} at eight days post drug exposure. For each strain, the MIC was defined as the lowest concentration that inhibited ≥95% growth. The minimum inhibitory concentration for each strain was determined by subtracting the OD\textsubscript{570} of a no-drug well and scaling by the reduction measurement of a no-drug control. The minimum inhibitory concentration data are representative of two independent replicates.

A library containing H37Rv and all prpR mutants described above was constructed by mixing all strains in equal proportion, freezing to −80°C, and then lyophilizing to ensure the library would remain consistent between experiments. Antibiotic tolerance was measured using time-dependent killing assays. The library was grown to mid-log phase, spun down and resuspended in 7H12 media with the indicated short-chain fatty acid to OD\textsubscript{600} 0.05 in triplicate for each condition. Antibiotic tolerance was measured by plating serial dilutions onto solid media after 21 days of growth. After the initial c.f.u. measurement, antibiotics were applied in the indicated concentrations. c.f.u. determination was performed for each replicate every 2 days for the next six days. To measure the relative abundance of each strain during treatment, a bulk sample representing >1,000 c.f.u. was plated onto drug-free plates at each time point. After 21 days of growth on solid media, gDNA was extracted from the entire resulting biomass. Amplicon sequencing libraries spanning the variable region of each unique molecular barcode were generated using PCR primers binding to regions common among all vectors as previously described. During this PCR, random numbers were incorporated into the sequence to allow for the accurate determination of input templates. The abundance of each barcode was then measured by deep sequencing on the Illumina MiSeq platform. Each library represented at least 40,000 unique combinations of barcodes and template counters. The fraction of the c.f.u. in each replicate/condition attributable to each strain was calculated by dividing the total c.f.u. by the representation of each strain as determined by sequencing. To calculate fraction survival, the c.f.u. at each point was divided by the measured starting c.f.u. that was normalized to 1. To test for differences in the fraction of surviving bacteria, survival fractions were log transformed and the differences among mean survival values were compared using Tukey’s honestly significant difference post-hoc test within time points after two-way ANOVA.

Mouse and macrophage infection and competition measurement. All mouse experiments were approved by the Harvard University Institutional Animal Care and Use Committee as complying with ethical regulations. The programme and facility is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International and the Office of Laboratory Animal Welfare assurance number is A3431-01. Six- to eight-week-old female C57BL/6J mice were obtained from Jackson Laboratories. Infection was initiated by tail-vein injection of the prpR library at 1×10^9 c.f.u. in a 100 μl injection. On day 1 post infection, 3 mice were euthanized, spleens were collected and homogenized in 7H12 media.图书馆的感染后第3天，2×10^7 c.f.u. that were stored at −80°C were used for competition measurement on 7H10 agar. Concurrently with c.f.u. determination, additional homogenate was frozen in 10% glycerol at −80°C. Colonies were counted after 21 days of incubation at 37°C. At 14 days post infection, an additional 5 mice were euthanized and the remaining mice were administered antibiotics or served as no-drug controls (n = 5 per time point per treatment). Mice were treated by ad libitum administration of drugs through their drinking water containing either isoniazid at 0.1 g l\textsuperscript{−1}, rifampin at 0.1 g l\textsuperscript{−1} or moxifloxacin at 0.4 g l\textsuperscript{−1}. The sample size was selected on the basis of the initial validation of the strain tagging and pooling system described by Blumenthal et al. where groups of 4–5 mice were sufficient to identify roughly 10-fold changes in vivo abundance of mutants. After c.f.u. determination, bacteria were scraped from the agar surface, suspended in gDNA was extracted. In cases where <500 colonies were sampled, additional frozen material was plated to increase the sample size of recovered bacteria. Sequencing library preparation was performed as described above. Relative abundance was determined by dividing the number of reads from each molecular barcode by the number of reads from the molecular barcode corresponding to ΔprpR::WT.
Articles

NATuRe MICRobIology

3. Weis, S. E. et al. The effect of directly observed therapy on the rates of drug resistance and relapse in tuberculosis. N. Engl. J. Med. 371, 1577–1587 (2014).

4. Paspasanody, J. G. & Gumbo, T. A meta-analysis of self-administered vs directly observed therapy effect on microbiologic failure, relapse, and acquired drug resistance in tuberculosis patients. Clin. Infect. Dis. 57, 21–31 (2013).

5. Gillespie, S. H. et al. Four-month moxifloxacin-based regimens for drug-sensitive tuberculosis. N. Engl. J. Med. 371, 1599–1608 (2014).

6. Jindani, A. et al. High-dose rifampin with moxifloxacin for pulmonary tuberculosis. N. Engl. J. Med. 371, 1599–1608 (2014).

7. Merle, C. S. et al. A four-month gatifloxacin-containing regimen for treating tuberculosis. N. Engl. J. Med. 371, 1588–1598 (2014).

8. Fridman, O., Goldberg, A., Ronin, I., Shores, N. & Balaban, N. Q. Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. Nature 513, 418–421 (2014).

9. Van den Bergh, B. et al. Frequency of antibiotic application drives rapid evolutionary adaptation of Escherichia coli persistence. Nat. Microbiol. 1, 1620 (2016).

10. Brauner, A., Fridman, O., Gefen, O. & Balaban, N. Q. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. Nat. Rev. Microbiol. 14, 320–330 (2016).

11. Dour, K. Persister cells. Annu. Rev. Microbiol. 64, 357–372 (2010).

12. Mulcahy, L., Burns, J., Lory, S. & Lewis, K. Emergence of Pseudomonas aeruginosa strains producing high levels of persister cells in patients with cystic fibrosis. J. Bacteriol. 192, 6191–6199 (2010).

13. Lafleur, M. D., Qi, Q. & Lewis, K. Patients with long-term oral carriage and high-persistence mutants of Candida albicans. Antimicrob. Agents Chemother. 54, 39–44 (2010).

14. Keren, I., Minami, S., Rubin, E. & Lewis, K. Characterization and transcriptome analysis of Mycobacterium tuberculosis persister. mBio 2, e00100–11 (2011).

15. Gomez, J. E. & McKinney, J. D. M. tuberculosis persistence, latency, and drug tolerance. Tuberculosis 94, 29–44 (2014).

16. Wayne, L. G. & Hayes, L. G. An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence. Infect. Immun. 64, 2062–2069 (1996).

17. Xie, Z., Siddiqui, N. & Rubin, E. J. Differential antibiotic susceptibilities of starved Mycobacterium tuberculosis isolates. Antimicrob. Agents Chemother. 49, 4778–4780 (2005).

18. Adams, K. N. et al. Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism. Cell 145, 39–53 (2011).

19. Dhar, N. & McKinney, J. D. Mycobacterium tuberculosis persistence mutants identified by screening in isoniazid-treated mice. Proc. Natl. Acad. Sci. USA 107, 12725–12728 (2010).

20. Levin, B. R. & Rozen, D. E. Non-inherited antibiotic resistance. Nat. Rev. Microbiol. 4, 556–562 (2006).

21. Levin-Reisman, I. et al. Antibiotic tolerance facilitates the evolution of resistance. Science 355, 826–830 (2017).

22. Zhao, Y. et al. National survey of drug-resistant tuberculosis in China. N. Engl. J. Med. 366, 2161–2170 (2012).

23. Comas, I. et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat. Genet. 45, 1176–1182 (2013).

24. Luo, T. et al. Southern East Asian origin and coexpansion of Mycobacterium tuberculosis Beijing family with Han Chinese. Proc. Natl. Acad. Sci. USA 112, 8316–41 (2015).

25. Casali, N. et al. Microevolution of extensively drug-resistant tuberculosis in Russia. Genome Res. 22, 735–745 (2012).

26. Cohen, K. A. et al. Evolution of extensively drug-resistant tuberculosis over four decades: whole genome sequencing and dating analysis of Mycobacterium tuberculosis isolates from KwaZulu-Natal. PLoS Med. 12, e1001880 (2015).

27. Farhat, M. R. et al. Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. Nat. Genet. 45, 1183–1189 (2013).

28. Manson, A. L. et al. Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and spread of multidrug resistance. Nat. Genet. 49, 395–402 (2017).

29. Zhang, H. et al. Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nat. Genet. 45, 1525–1526 (2013).

30. 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, 1193–202 (2015).

31. Sun, L., Zhang, L., Zhang, H. & He, Z.-G. Characterization of a bifunctional β-lactamase/bonclonuclease and its interaction with a chaperone-like protein in the pathogen Mycobacterium tuberculosis H37Rv. Biochem. 76, 350–358 (2011).

32. Taeverniti, V., Forti, F., Ghisotti, D. & Putzer, H. Mycobacterium smegmatis RNase J is a 5′-3′ exonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. Mol. Microbiol. 82, 1260–1276 (2011).
33. Guerra-Assunção, J. A. et al. Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. Elife 4, e05166 (2015).
34. Casali, N. et al. Evolution and transmission of drug-resistant tuberculosis in a Russian population. Nat. Genet. 46, 279–286 (2014).
35. Holt, K. É. et al. Frequent transmission of the Mycobacterium tuberculosis Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. Nat. Genet. 50, 849–856 (2018).
36. Liu, Q. et al. Within patient microevolution of Mycobacterium tuberculosis correlates with heterogeneous responses to treatment. Sci. Rep. 5, 17507 (2015).
37. Pandey, A. K. & Sassetti, C. M. Mycobacterial persistence requires the utilization of host cholesterol. Proc. Natl Acad. Sci. USA 105, 4376–4380 (2008).
38. Griffin, J. L. et al. Cholesterol catabolism by Mycobacterium tuberculosis requires transcriptional and metabolic adaptations. Chem. Biol. 19, 218–227 (2012).
39. Masiewicz, P., Brzostek, A., Wołarski, M., Dziadek, J. & Zakrzewska-Czerwińska, J. A novel role of the PrpR as a transcription factor involved in the regulation of methylcitrate pathway in Mycobacterium tuberculosis. PLoS ONE 7, e36561 (2012).
40. Muñoz-Elias, E. I., Upton, A. M., Cherian, J. & McKinney, J. D. Role of the methylcitrate cycle in Mycobacterium tuberculosis metabolism, intracellular growth, and virulence. Mol. Microbiol. 60, 1109–1122 (2006).
41. KATO, N. The free and bound forms of the serum vitamin B12 in various animal species. J. Vitaminol. (Kyoto) 6, 132–138 (1960).
42. Micklinghoff, J. C. et al. Role of the transcriptional regulator RamB (Rv0465c) in the control of the glyoxylate cycle in Mycobacterium tuberculosis. J. Bacteriol. 191, 7260–7269 (2009).
43. Eoh, H. & Rhee, K. Y. Methylcitrate cycle defines the bactericidal essentiality of isocitrate lyase for survival of Mycobacterium tuberculosis on fatty acids. Proc. Natl Acad. Sci. USA 111, 4976–4981 (2014).
44. Savvi, S. et al. Functional characterization of a vitamin B12-dependent methylmalonyl pathway in Mycobacterium tuberculosis: implications for propanoate metabolism during growth on fatty acids. J. Bacteriol. 190, 3886–3895 (2008).
45. Somerville, W., Thübert, L., Schwartzman, K. & Behr, M. A. Extraction of Mycobacterium tuberculosis DNA: a question of containment. J. Clin. Microbiol. 43, 2996–2997 (2005).
46. Patel, R. & Jain, M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE 7, e30619 (2012).
47. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26, 589–595 (2010).
48. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
49. Comas, I. et al. Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. Nat. Genet. 42, 498–503 (2010).
50. Pouts, D. PhageFinder: automated identification and classification of prophage regions in complete bacterial genome sequences. Nucleic Acids Res. 34, 5839–5851 (2006).
51. Sandgren, A. et al. Tuberculosis drug resistance mutation database. PLoS Med. 6, e1000002 (2009).
52. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313 (2014).
53. Letunic, I. & Bork, P. Interactive tree of life (iTOl) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242–W245 (2016).
54. Tatusova, K. et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739 (2011).
55. Schliep, K. P. phangorn: phylogenetic analysis in R. Bioinformatics 27, 592–593 (2011).
56. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57, 289–300 (1995).
57. McMackin, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010).
58. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. 26, 1641–1650 (2009).
59. Martin, C. J. et al. Digitally barcoding Mycobacterium tuberculosis reveals in vivo infection dynamics in the macaque model of tuberculosis. mBio 8, e00312–17 (2017).
60. Blumthenthal, A., Trujillo, C., Ehrt, S. & Schnappinger, D. Simultaneous analysis of multiple Mycobacterium tuberculosis knockdown mutants in vitro and in vivo. PLoS ONE 5, e15667 (2010).

Acknowledgements
Support was generously provided by the National Major Science and Technology Project of China (2012ZX100030001-001; 2014ZX10003002) to Y. Z.; the National Program on Key Basic Research Project of China (2014CB7444037) and the CAMS Innovation Fund for Medical Sciences (2016-I2M-1-013) to Q. J.; the National Institute of Health (5U19AI109755-04) to S. M. F.; and NIH pre-doctoral training grant support (5T32AI007638-15 to E. J. Rubin and 5T32-AI049928-14 to D. Wirth) to N. D. H.

Author contributions
X.Z., R.Z., X.C.O., S.F.W., H.X. and Y. Zhou were responsible for cohort study and strain set assembly. Q.J. and Y. Zhao were responsible for genome sequencing study design and supervision. J.Y., Z.S., L.L., Z.C., I.D., L.S. and Y. Zhu were responsible for genome sequencing and variant analysis. N.D.H., J.Y., Y.G. and S.M.F. were responsible for genome-wide association study. N.D.H. and S.M.F. were responsible for experimental validation. N.D.H., J.Y. and S.M.F. were responsible for manuscript preparation.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41564-018-0218-3.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to Y.Z. or Q. J. or S.M.F.
Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- **n/a**
- **Confirmed**
- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  - *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
- Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
- Clearly defined error bars
- State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection
- Genomes were sequenced on the Illumina GAIIx or HiSeq 2500 platforms, generating single end reads of 72 bp or 101 bp in length.

Data analysis
- The code used to analyze the data in this study include commercially available and custom softwares as described in detail in the methods. We have described it briefly below.

For each sample duplicate reads were removed by custom Perl scripts. Further quality control was conducted using the NGSQCToolkit with a cutoff of Q20. Valid reads were then aligned to the reference genome sequence of H37Rv (GenBank accession NC_000962) using the Burrows-Wheeler algorithm as implemented in BWA. SNPs were identified with a minimum depth of 10X and a consensus quality score of 50 using SAMtools. SNPs located within repetitive regions, including transposases, PE/PPE genes, prophages and exact sequence repeats of the H37Rv genome previously proposed or identified by RepeatMasker (http://www.repeatmasker.org/) or PhageFinder 40, were excluded. Mixed base calls were considered valid only if the numbers of the most abundant (n1) and the second most abundant (n2) nucleotides at each SNP in each strain satisfied the criteria n1/n2 > 5. The detailed workflow of the SNP-calling procedure of datasets for follow-up analysis is given in Supplementary Figure 7.

The refined SNP set was used to construct the maximum-likelihood phylogeny using RAxML under the GTRgamma substitution model. The reliability of each node was tested via a bootstrap analysis on 100 resampled datasets. The iTOL server and MEGA5 software were employed for the manipulation and presentation of the phylogenetic trees.

Genomic associations were performed using a method similar to that of Farhat et al which we termed phyOverlap. We performed
maximum parsimony ancestral sequence reconstruction to determine the state of each SNP residue at the common ancestor of our strain collection using the Mesquite software package (http://mesquiteproject.org/). In cases where the ancestral sequence could not be unambiguously determined, we excluded the site from further analysis. Each isolate was then scored as ancestral or derived for a given SNP locus. Isolates with ambiguous base calls were considered ancestral for this calculation to minimize the potential of error driving associations with drug resistance. The overlap with drug resistance was scored by dividing the number of isolates containing a derived allele which also were resistant to isoniazid by the total number of isolates with a derived allele at a given SNP locus. To generate a gene-wide score, we excluded synonymous SNPs and averaged the individual SNP scores, weighting the scores by the number of times derived alleles evolved across the phylogenetic tree. The number of times a site mutated across the tree was calculated as the parsimony score at each SNP position using the Fitch algorithm as implemented in the R package phangorn. This was performed with mixed bases treated as missing data so that low quality base-calls would not appear highly convergent. Significance testing was performed by redistributing the mutation events in each SNP in each gene randomly across the phylogenetic tree, with the probability of a mutation occurring on any branch proportional to the branch length and all offspring of a branch being assigned the derived allele for calculation of the permuted overlap score. The gene-wide score was then recalculated and compared with the actual value. This process was repeated 50,000 times to derive an empirical p-value. For all statistical calculations, associations which achieved a p-value of 0 were considered <2x10⁻⁵. False discovery rate q-values were then calculated to account for multiple hypothesis testing using the Benjamini-Hochberg procedure.

Scripts for this method are available online at: https://github.com/nathan-d-hicks/phyOverlap

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing reads have been submitted to the NCBI Sequence Read Archive (SRA) under accession PRJNA268900. A complete list of the new TB strains analyzed in this study together with phenotypic and sequencing information is given in Supplementary Table 1. The accession numbers for previously sequenced strains used in Figures 1B, 1D, and Table 1 are available at in Supplementary Table 2 and Supplementary Table 8.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | Sample size calculations are relevant for experiments shown in Figures 2-5. The growth curves and expression data presented in the manuscript consist of three biological replicates (Figure 2). We used three biologically independent replicates for in vitro killing assays (Figures 3, 4, 5) which our preliminary data indicated would be sufficient to measure 2-fold or greater changes in bacterial numbers. Mouse experiments included 5 mice per group which was sufficient to identify roughly 10-fold changes in in vivo abundance of mutants in previously published experiments. |
|---|---|
| Data exclusions | We excluded three datapoints in the entire in the entire study as described below. Figure 3D: A single datapoint was excluded each from the ofloxacin acetate data on day 2 and propionate day 2 due to contamination of the CFU plate. A single datapoint was excluded from the high-dose INH Day 6 propionate in figure 3D where the prpR WT complement had substantially lower reads than in the two other replicates (while true WT H37Rv was unaffected). This exclusion weakened the statistical significance reported at this timepoint, rather than increasing the magnitude of the effect. Supplemental Figure 5: RIF Day 57 only has 4 mice because one spleen was contaminated during extraction. |
| Replication | We have indicated the number of times experiment was independently performed as described below and in the figure legends. The growth curves shown in figure 2 were performed three times, each with three replicates, and yielded similar results. The MIC assays presented in Figure 3 a-c were performed 3 times for the prpR complement and deletion strain and in all cases we did not see prpR dependent antibiotic protection. The library format antibiotic killing in figure 3D was performed once, and then the propionate mediated protection of strains was confirmed in the single strain CFU assays presented in figure 3e-g, which were repeated >3 times. In all cases the prpR mutants displayed... |
increased antibiotic tolerance. We assessed prpR mutant antibiotic susceptibility in primary macrophages in two independent experiments and confirmed these findings in THP1 cells in two independent experiments. The mouse experiment was performed once. B12 mediated reversal of antibiotic protection in vitro shown in Figure 5 was performed twice and in both cases demonstrated re-sensitization of prpR mutants. The THP-1 infection B12 reversal of protection was performed once.

Randomization
There was no randomization in this study because there were no features for which randomization was deemed appropriate.

Blinding
Similarly blinding was not performed for any of the in vitro or in vivo experimentation. The measurements of optical density, CFU, RNA expression, and representation of strains in competition assays do not require researcher-based judgments and therefore we deemed blinding not necessary.

Materials & experimental systems

Policy information about availability of materials

n/a Involved in the study

☐ ☒ Unique materials
☐ ☒ Antibodies
☐ ☒ Eukaryotic cell lines
☐ ☒ Research animals
☐ ☒ Human research participants

Unique materials

Obtaining unique materials
All bacterial strains generated in this study will be made available by the investigators upon request.

Research animals

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Animals/animal-derived materials
6-8 week old female C57BL/6J mice were obtained from Jackson Laboratories.

Human research participants

Policy information about studies involving human research participants

Population characteristics
The population characteristics of the Chinese CDC cohort are described in detail in: Zhao et al, NEJM 2012. The sub-selection of strains for sequencing is described in detail in the methods.

Human research participants were used to acquire blood components for the macrophage experiments. Human blood components (buffy coats) for macrophage experiments were procured through Blood Transfusion Services (BTS) at Massachusetts General Hospital (MGH) for in vitro research by Investigators at the Ragon Institute of MGH, MIT and Harvard. These specimens were collected from donations for clinical use and only excess blood or blood products that are not needed are then processed for research purposes. The specimens supplied for these purposes are not identifiable. Donors in the Blood Transfusion Services at MGH sign a Donor Consent within a Registration Form that includes the following sentence, “I give permission for my blood to be used for transfusion to patients or for research.”

Method-specific reporting

n/a Involved in the study

☒ ☒ ChIP-seq
☒ ☒ Flow cytometry
☒ ☒ Magnetic resonance imaging