

# Compensatory evolution in NusG improves fitness of drug-resistant *M. tuberculosis*

<https://doi.org/10.1038/s41586-024-07206-5>

Received: 30 September 2023

Accepted: 19 February 2024

Published online: 20 March 2024

Open access

 Check for updates

Kathryn A. Eckart<sup>1,5</sup>, Madeleine Delbeau<sup>2,5</sup>, Vanisha Munsamy-Govender<sup>1</sup>, Michael A. DeJesus<sup>1</sup>, Zachary A. Azadian<sup>1</sup>, Abhijna K. Reddy<sup>1</sup>, Joshua Chandanani<sup>2</sup>, Nicholas C. Poulton<sup>1</sup>, Stefany Quiñones-García<sup>1</sup>, Barbara Bosch<sup>1</sup>, Robert Landick<sup>3,4</sup>, Elizabeth A. Campbell<sup>2</sup>✉ & Jeremy M. Rock<sup>1</sup>✉

Drug-resistant bacteria are emerging as a global threat, despite frequently being less fit than their drug-susceptible ancestors<sup>1–8</sup>. Here we sought to define the mechanisms that drive or buffer the fitness cost of rifampicin resistance (Rif<sup>R</sup>) in the bacterial pathogen *Mycobacterium tuberculosis* (Mtb). Rifampicin inhibits RNA polymerase (RNAP) and is a cornerstone of modern short-course tuberculosis therapy<sup>9,10</sup>. However, Rif<sup>R</sup> Mtb accounts for one-quarter of all deaths due to drug-resistant bacteria<sup>11,12</sup>. We took a comparative functional genomics approach to define processes that are differentially vulnerable to CRISPR interference (CRISPRi) inhibition in Rif<sup>R</sup> Mtb. Among other hits, we found that the universally conserved transcription factor NusG is crucial for the fitness of Rif<sup>R</sup> Mtb. In contrast to its role in *Escherichia coli*, Mtb NusG has an essential RNAP pro-pausing function mediated by distinct contacts with RNAP and the DNA<sup>13</sup>. We find this pro-pausing NusG–RNAP interface to be under positive selection in clinical Rif<sup>R</sup> Mtb isolates. Mutations in the NusG–RNAP interface reduce pro-pausing activity and increase fitness of Rif<sup>R</sup> Mtb. Collectively, these results define excessive RNAP pausing as a molecular mechanism that drives the fitness cost of Rif<sup>R</sup> in Mtb, identify a new mechanism of compensation to overcome this cost, suggest rational approaches to exacerbate the fitness cost, and, more broadly, could inform new therapeutic approaches to develop drug combinations to slow the evolution of Rif<sup>R</sup> in Mtb.

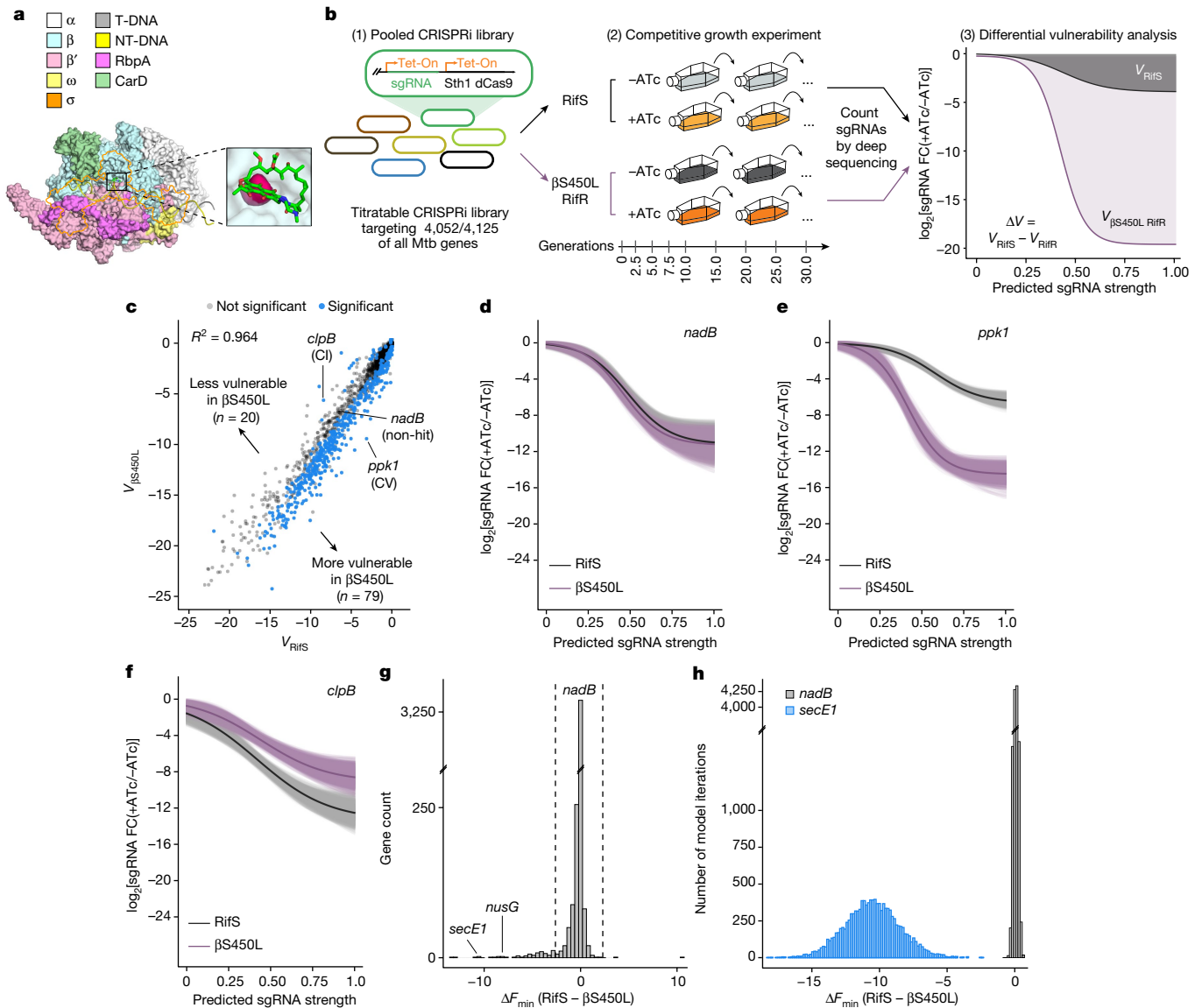
Antimicrobial resistance is a global threat, incurring both economic costs and loss of human lives. Although not generally discussed as part of the antimicrobial resistance crisis, drug-resistant Mtb accounts for one-quarter of all deaths due to antimicrobial resistance, and Mtb is the single leading cause of death due to infectious disease<sup>11</sup>. As is true for other bacterial pathogens<sup>1–4</sup>, drug resistance in Mtb is often associated with reduced fitness in the absence of drug pressure<sup>5–8</sup>. It was once thought that this reduced fitness might minimize patient-to-patient transmission and keep drug-resistant tuberculosis a localized problem, but this was not the case<sup>14</sup>. Mtb can acquire secondary ‘compensatory’ mutations that restore fitness and may promote more efficient pathogen transmission<sup>15–18</sup>.

One of the most potent first-line antibiotics to treat tuberculosis is rifampicin (Rif). Rif is the backbone of modern short-course tuberculosis treatment and its introduction effectively halved the duration of therapy<sup>9</sup>. Rif exerts its antibacterial effects by binding to the  $\beta$ -subunit of RNAP and blocking the extension of short RNA transcripts<sup>10</sup> (Fig. 1a). Decades of Rif use have selected for Rif<sup>R</sup> Mtb, which in 2021 caused up to 450,000 cases and 264,000 deaths worldwide<sup>11</sup>. Mtb develops resistance to Rif through mutations in the  $\beta$ -subunit that prevent Rif from binding, with the most frequent mutation being a single amino acid

substitution, serine 450 to leucine ( $\beta$ S450L; Fig. 1a).  $\beta$ S450L accounts for up to 70% of all Rif<sup>R</sup> Mtb in the clinic<sup>12</sup>. The  $\beta$ S450L mutation reduces Mtb fitness in the absence of Rif<sup>R</sup><sup>5–7,12</sup>, but Mtb can acquire compensatory mutations in the  $\alpha$ ,  $\beta$  or  $\beta'$  subunits that restore fitness to levels equivalent to the drug-sensitive ancestor<sup>15–17,19</sup>.

Rif<sup>R</sup> has pleiotropic effects on RNAP biochemistry and microbial physiology, but exactly how these differences reduce Mtb fitness are not fully understood. Rif<sup>R</sup> mutations, including  $\beta$ S450L, alter the shape and chemical nature of the Rif-binding pocket. The Rif-binding pocket helps form the elongating RNA exit pathway<sup>10</sup>, and thus Rif<sup>R</sup> mutations have the potential to affect the RNA elongation step of the transcription cycle<sup>20</sup>. Biochemical studies have shown that Rif<sup>R</sup> mutations in RNAP can positively or negatively alter the stability of promoter open complexes, elongation speed or termination efficiency<sup>20–22</sup>. In Mtb, Rif<sup>R</sup> decreases fitness and broadly impacts bacterial physiology. Several studies report alterations in the cell envelope in Rif<sup>R</sup> strains<sup>23</sup>, including increased levels of phthiocerol dimycocerosates<sup>24–26</sup>, decreased levels of acylated sulfoglycolipids<sup>25</sup>, and changes in the levels of up to 2–6% of all lipid species<sup>25</sup>. Rif<sup>R</sup> may also alter iron homeostasis, although this conclusion is muddled by conflicting reports about whether levels of mycobactin—the iron-scavenging siderophore—are higher<sup>27</sup> or lower<sup>25</sup>

<sup>1</sup>Laboratory of Host–Pathogen Biology, The Rockefeller University, New York, NY, USA. <sup>2</sup>Laboratory of Molecular Biophysics, The Rockefeller University, New York, NY, USA. <sup>3</sup>Department of Biochemistry, University of Wisconsin–Madison, Madison, WI, USA. <sup>4</sup>Department of Bacteriology, University of Wisconsin–Madison, Madison, WI, USA. <sup>5</sup>These authors contributed equally: Kathryn A. Eckart, Madeleine Delbeau. ✉e-mail: [campbee@rockefeller.edu](mailto:campbee@rockefeller.edu); [rock@rockefeller.edu](mailto:rock@rockefeller.edu)



**Fig. 1 | A genetic screen to identify differential vulnerabilities in βS450L**

**RifR Mtb.** **a**, Structural model of an Mtb RNAP transcription initiation complex bound to Rif. The sigma factor is shown as an orange outline. Inset, βSer450 is shown in pink. T-DNA, template DNA; NT-DNA, non-template DNA. **b**, Quantification of differential vulnerabilities ( $V$ ). (1) An anhydrotetracycline (ATc)-inducible CRISPRi library was transformed into RifS and βS450L Mtb. Genes essential for in vitro growth<sup>61</sup> were targeted with sgRNAs of varying predicted knockdown efficiencies. (2) Cultures were passaged and sgRNA abundance was assessed by deep sequencing at multiple timepoints. (3) A Bayesian model was then used to model the expression–fitness relationship (black and purple curved lines) for each gene. In brief, the x axis depicts predicted sgRNA strength (a proxy for the magnitude of target knockdown) and the y axis depicts bacterial fitness (predicted sgRNA  $\log_2$  transformed fold change ( $\log_2FC$ ) at 25 generations in the competitive growth experiment). The area above the expression–fitness curve was calculated to quantify gene

vulnerability. Genes were called differentially vulnerable when the 95% credible region of the difference in gene vulnerabilities did not overlap 0.  $\Delta V$  is differential vulnerability. Sth1 dCas9, *Streptococcus thermophilus* CRISPR1 dCas9. **c**, Scatter plot showing gene vulnerability (circles) in RifS and βS450L Mtb. Genes with significantly different vulnerabilities are shown in blue. CV, collateral vulnerability—that is, more vulnerable in βS450L; CI, collateral invulnerability—that is, less vulnerable in βS450L. **d–f**, Expression–fitness relationships for an example non-hit gene (*nadB*) (**d**), collateral vulnerability (*ppk1*) (**e**) and collateral invulnerability (*clpB*) (**f**). Light coloured lines represent the fits determined by 1,000 samples from the posterior distributions; dark lines represent the mean fits. **g**, Histogram showing the gene-level mean differences in the fitness cost imposed by the weakest possible sgRNAs ( $F_{\min}$ ) between RifS and βS450L Mtb. Dashed lines mark two s.d. from the mean. **h**, Histograms showing the  $F_{\min}$  distributions for *nadB* and *secE1* between RifS and βS450L Mtb.

in RifR Mtb, and the lack of such an iron-response signature across diverse RifR clinical isolates<sup>27</sup>.

Here we sought to define the mechanisms driving or buffering the fitness cost of βS450L Mtb. Given the often subtle and genetic background-dependent effects of RifR on the Mtb transcriptome and proteome<sup>27</sup>, we chose to use a functional genomics approach. Comparative genome-scale CRISPRi screening in RifS and βS450L Mtb identified

approximately 150 genes involved in diverse cellular processes that are more or less sensitive to genetic silencing in βS450L, which we refer to as differential vulnerabilities. Among these hit genes, we found that the universally conserved, essential transcription factor *nusG* is crucial for the fitness of βS450L Mtb. Combining this approach with a bacterial genome-wide association study, we further found that *nusG* is under positive selection in RifR clinical Mtb isolates. We show that clinically

observed NusG mutants reduce the pro-pausing and termination activity of NusG and compensate for the increased pausing and termination properties of  $\beta$ S450L RNAP. NusG mutants thereby increase the fitness of  $\beta$ S450L Mtb. Collectively, these results define excessive RNAP pausing and termination as molecular mechanisms that drive the fitness cost of  $\beta$ S450L, identify a new mechanism of compensation to overcome this cost, suggest rational approaches to exacerbate the fitness cost, and, more broadly, could inform new therapeutic approaches to develop drug combinations to slow the evolution of Rif<sup>r</sup> in Mtb.

### Identifying differential vulnerabilities

To identify differential vulnerabilities in Rif<sup>r</sup>  $\beta$ S450L (hereafter  $\beta$ S450L) Mtb, we used our tunable Mtb CRISPRi library<sup>28,29</sup> (Fig. 1b). This library consists of 96,700 unique single guide RNAs (sgRNAs) that target around 98% of all annotated Mtb genes. Knockdown tuning is achieved by varying the targeted protospacer adjacent motif (PAM), varying the length of the sgRNA targeting sequence, or both. Biosafety concerns dictated that we perform these screens in an isogenic pair of Rif<sup>s</sup> or  $\beta$ S450L Mtb biotin auxotrophs. The biotin auxotroph ( $\Delta$ *bioA*) grows similarly to prototrophic Mtb in the presence of 2  $\mu$ M biotin but cannot establish infection in mice. We confirmed that  $\beta$ S450L  $\Delta$ *bioA* had a growth defect relative to Rif<sup>s</sup>  $\Delta$ *bioA* Mtb (Supplementary Fig. 1a). All validation and follow-up experiments were conducted in the prototrophic parental H37Rv Rif<sup>s</sup> or  $\beta$ S450L strains.

Following transformation of the CRISPRi library into Rif<sup>s</sup> and  $\beta$ S450L  $\Delta$ *bioA* Mtb, we carried out two separate competitive growth experiments. Triplicate cultures for each library were passaged for approximately 30 generations in the presence or absence of the CRISPRi inducer ATc (Fig. 1b). Every 2.5 or 5 generations, we collected genomic DNA, analysed sgRNA abundance by deep sequencing, and calculated the log<sub>2</sub>-transformed fold change of sgRNA read counts with or without ATc. Growth phenotypes were well correlated among triplicate screens (Supplementary Fig. 1b–e). Whole-genome sequencing of the final competitive growth timepoint confirmed that compensated  $\beta$ S450L mutants did not take over the culture at late timepoints. sgRNA depletion data were then used to calculate gene vulnerability using a multilevel Bayesian model, similar to that described previously<sup>28</sup>. As opposed to binary gene essentiality calls, the vulnerability model generates expression–fitness relationships for every targeted gene by relating the predicted magnitude of target knockdown (as inferred by the sgRNA strength) to the resulting fitness cost imposed on the bacteria. The model is iterated 12,000 times to generate distributions of vulnerability values for each gene. Differential vulnerabilities are defined as those genes for which the 95% credible region of differences in gene vulnerability do not overlap 0.

As expected, most genes were similarly vulnerable in the Rif<sup>s</sup> and  $\beta$ S450L strains ( $R^2 = 0.964$ ; Fig. 1c and Supplementary Table 1). For example, the nicotinamide adenine dinucleotide (NAD) biosynthetic enzyme *nadB* showed very similar expression–fitness relationships between the two Mtb strains (Fig. 1d). However, 99 genes displayed a statistically significant difference in vulnerability (Fig. 1c). Most differentially vulnerable genes were more sensitive to knockdown in  $\beta$ S450L ( $n = 79$  genes; Fig. 1c,e), but some were less sensitive ( $n = 20$  genes; Fig. 1c,f). Drawing parallels with collateral interactions between drugs and resistance mutations<sup>30</sup>, we refer to these differentially vulnerable genes as collateral vulnerabilities or collateral invulnerabilities, respectively.

We observed that a subset of genes did not exhibit a dose–response relationship between sgRNA strength and bacterial fitness, and the standard vulnerability model did not accurately capture the expression–fitness relationship for these genes (Supplementary Fig. 1f). These genes tended to be those that were already very vulnerable in the Rif<sup>s</sup> strain and became even more vulnerable in  $\beta$ S450L. To capture these hit genes, we identified the fitness cost imposed by

the weakest possible sgRNAs (defining the fitness minimum ( $F_{\min}$ )), or in effect where the expression–fitness relationship crossed the y axis (Supplementary Fig. 1g,h). This approach identified 73 hit genes (Fig. 1g,h and Supplementary Table 1), including 53 genes that were not identified in the original model.

### Translation is a collateral vulnerability in $\beta$ S450L

We next performed functional enrichment analysis to identify pathways that are more vulnerable to inhibition in  $\beta$ S450L Mtb (Fig. 2a). This approach identified genes involved in peptidoglycan, arabinogalactan and mycolic acid biosynthesis as more sensitive to CRISPRi inhibition in  $\beta$ S450L, consistent with the known cell envelope differences between Rif<sup>s</sup> and Rif<sup>r</sup> Mtb<sup>23</sup>. Many metabolic processes were also enriched as more vulnerable in  $\beta$ S450L, potentially consistent with prior observations of altered metabolism in Rif<sup>r</sup> Mtb<sup>27</sup>. Of note, many genes that are critical for translation were also identified as more vulnerable in  $\beta$ S450L, including aminoacyl-tRNA synthetases, tRNAs, ribosomal proteins, translation initiation and elongation factors, and several amino acid biosynthetic pathways (Fig. 2a).

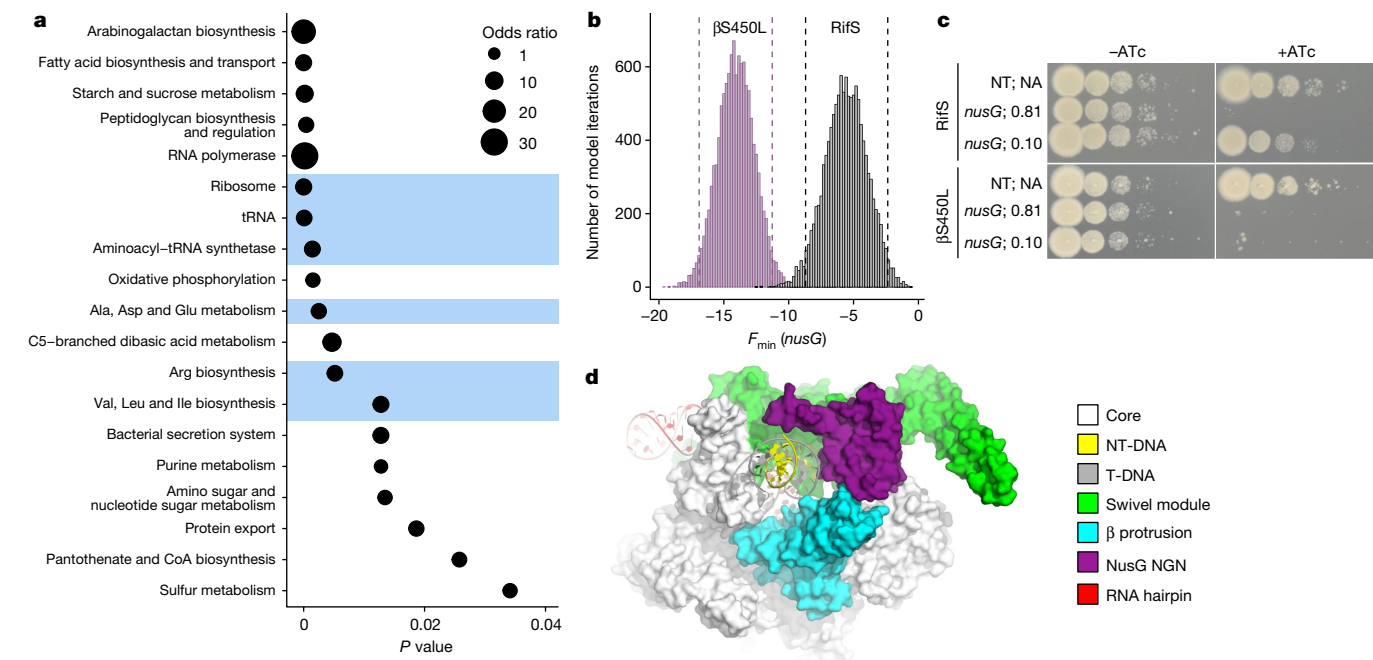
We next explored the reasons why translation might be more sensitive to CRISPRi inhibition in  $\beta$ S450L. Biochemical studies show that  $\beta$ S450L RNAP elongates more slowly and terminates more frequently than Rif<sup>s</sup> RNAP in vitro<sup>20</sup>. It is also known that during the process of transcription–translation coupling (TTC), the translating ribosome can stimulate transcription elongation by RNAP either directly<sup>31</sup> or mediated by the universally conserved transcription factor NusG<sup>32,33</sup>. Although TTC has yet to be directly demonstrated in Mtb, recent evidence suggests that, similar to *E. coli*, transcription and translation are coupled in Mtb<sup>34</sup>. Thus, we hypothesized that translation may promote fitness in  $\beta$ S450L by maintaining coupling of the ribosome to a slow, pause-prone RNAP to promote more efficient transcription elongation. *nusG* is a strong collateral vulnerability in  $\beta$ S450L (Figs. 1g and 2b,c), potentially consistent with the hypothesis that TTC may be important to promote  $\beta$ S450L fitness.

Mtb NusG is composed of two domains, a NusG N-terminal (NGN) domain and a C-terminal Kyprides, Ouzounis, Woese (KOW) domain. The NGN domain interacts with RNAP<sup>13</sup> and the KOW domain interacts potentially with several proteins, including the S10 ribosomal protein<sup>35</sup>. By interacting with the ribosome, NusG could serve as a bridge mediating TTC, at least as demonstrated in *E. coli*<sup>22,33,36</sup>. But the NGN domain also modulates transcription pausing via distinct contacts with RNAP and the DNA (Fig. 2d). In bacteria, pause-inducing sequences in the DNA and RNA reversibly halt RNAP elongation, thereby favouring RNAP kinetics towards termination<sup>37</sup>. RNAP pausing and termination are additionally aided by *trans*-acting factors such as NusG. Mtb NusG promotes pausing and termination by stabilizing a swivelled RNAP conformation, which slows nucleotide addition and thus RNA elongation<sup>13,38,39</sup>. This pro-pausing activity of Mtb NusG is mediated via unique interactions between the NusG NGN domain, the RNAP  $\beta$ -protrusion domain, and the NT-DNA, which stabilize the paused, swivelled RNAP state<sup>13</sup>.

*nusG* is a strong collateral vulnerability, and thus the presence of NusG promotes  $\beta$ S450L fitness. The positive effect of NusG on  $\beta$ S450L fitness may be explained by its anti-pausing role in mediating TTC, although other potential explanations exist. By contrast, the pro-pausing and termination activity of the NusG NGN domain might be expected to be detrimental to  $\beta$ S450L fitness. Thus, we explored whether the pro-pausing and termination activity of NusG might in fact be a source of the fitness cost in  $\beta$ S450L and, if so, whether it might be under selective pressure to mutate to restore fitness in  $\beta$ S450L Mtb.

### NusG is diversifying in Rif<sup>r</sup> Mtb

To answer this question, we sought to determine whether *nusG* is under positive selection in clinical Mtb isolates. Using a cohort of



**Fig. 2 | Pathway analysis identifies differentially vulnerable processes in  $\beta$ S450L Mtb.** **a**, Bubble plot of enriched ( $P \leq 0.05$ ) pathways for the genes identified as differential vulnerabilities in  $\beta$ S450L. Pathways highlighted in blue are related to translation.  $P$  values and odds ratios were determined by Fischer's exact test. **b**, The distribution of  $nusG$   $F_{min}$  values in RifS (grey) and  $\beta$ S450L (purple) Mtb. Dashed lines represent the 95% credible region.

Mtb clinical isolates from Peru<sup>40</sup>, we first performed a bacterial genome-wide association study to identify genetic variants associated with genotypically predicted RifR. This approach identified several well-known genes involved in Mtb drug resistance including *rpoB*, *pncA* and *embB*, as well as variants in the  $\beta'$  subunit (*rpoC*) that are known to compensate for the fitness cost of RifR<sup>15,16</sup> (Fig. 3a,b and Supplementary Tables 2 and 5). The association of RifR with resistance variants to other first and second-line drugs reflects the multiple resistance mutations present in many multidrug-resistant (MDR) Mtb isolates. Potentially consistent with our hypothesis, variants in *nusG* were also statistically associated with RifR in this cohort (Fig. 3a,b and Supplementary Fig. 2a). Although this association did not meet the significance cut-off when applying a genome-wide multiple test correction, *nusG* variants were nevertheless associated with RifR with a  $P$  value similar to that of well-known resistance-associated genes such as *gyrA*, *katG* and others, suggesting that *nusG* variants are indeed enriched in RifR Mtb.

To investigate this phenomenon further, we leveraged an in-house database of around 50,000 publicly available whole-genome sequences from Mtb clinical isolates from around the world<sup>41</sup>. Analysis of the ratio of nonsynonymous to synonymous mutations (dN/dS) showed *nusG* to be under purifying selection in RifS Mtb (Fig. 3c), consistent with the fact that *nusG* is an essential gene<sup>13</sup>. By contrast, *nusG* was under stronger diversifying selection in RifR Mtb, as evidenced by a higher dN/dS ratio in RifR as compared to RifS Mtb. Closer analysis revealed that *nusG* selection was specific to certain RifR *rpoB* alleles (Supplementary Fig. 2b,c and Supplementary Tables 2 and 3). For example, *nusG* is under comparatively diversifying selection in  $\beta$ S450X (where X indicates any amino acid except Ser) but under purifying selection in  $\beta$ H445X (Fig. 3d,e), a pattern very similar to that observed for variants in *rpoC*<sup>16,17</sup> (Fig. 3c–e). This phenomenon could be explained by the fact that some RifR RNAP mutants such as  $\beta$ S450L are slow and over-pause or over-terminate, whereas  $\beta$ H445Y RNAPs are fast and

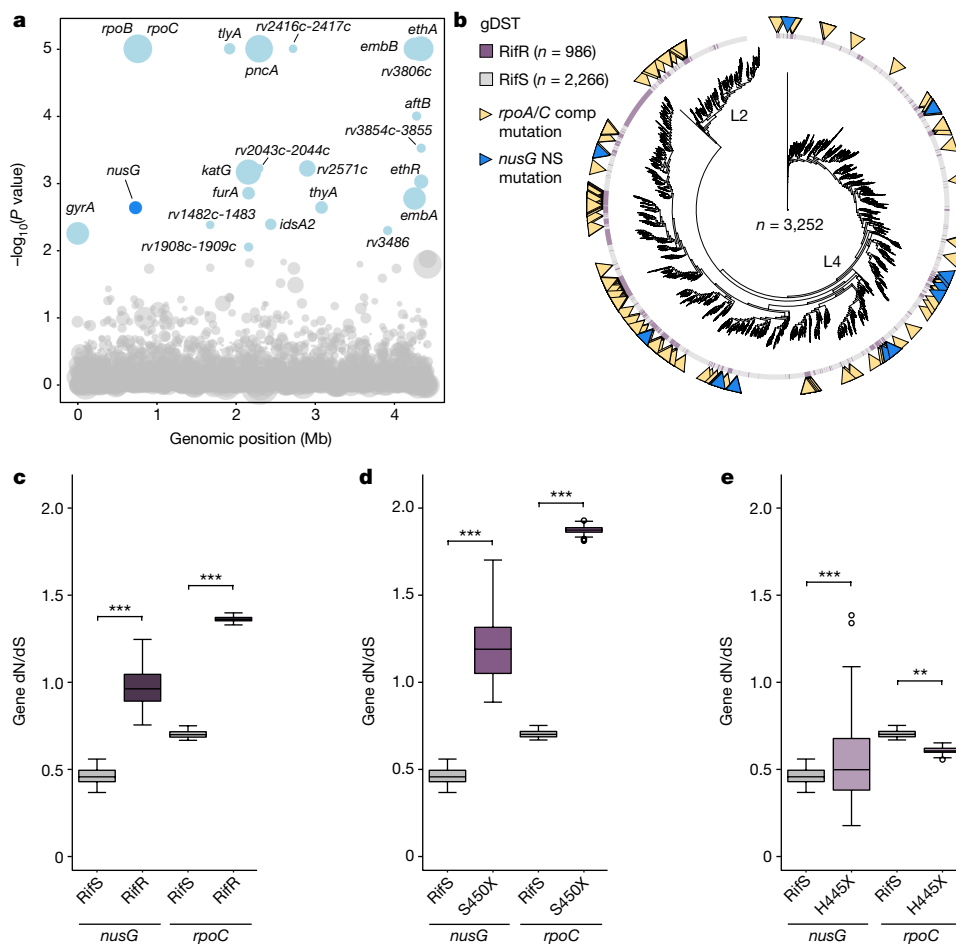
less likely to pause or terminate<sup>12,20,22</sup>. We would expect selective pressure to reduce the pro-pausing activity of NusG to be specific to slow, pause-prone RifR RNAP mutants. *nusG* variants are not commonly found in RifR strains with known compensatory mutations in *rpoA*, *rpoB* or *rpoC* (Supplementary Fig. 2d and Supplementary Tables 2 and 3). Collectively, these results show that *nusG* is evolving convergently and may represent a new class of compensatory mutations in RifR Mtb.

### Mutations reduce NusG pro-pausing activity

Homoplastic *nusG* mutations occurred exclusively in the NGN domain (Supplementary Table 2). To predict the functional consequences of these mutations, we mapped them onto the Mtb NusG structure<sup>13</sup>. Notably, these mutations clustered at the unique, pro-pausing interfaces between the NusG NGN domain and either the  $\beta$ -protrusion or the NT-DNA<sup>13</sup> (Fig. 4a). These mutations would be expected to weaken the interaction between the NusG NGN domain and the  $\beta$ -protrusion or the NT-DNA and thereby weaken the pro-pausing and termination activity of NusG. Consistent with this interpretation, we also observed homoplastic mutations in the  $\beta$ -protrusion that are found exclusively in RifR Mtb (Fig. 4a and Supplementary Table 2). Similar to NusG NGN mutations, these  $\beta$ -protrusion mutations are anticipated to reduce the strength of the NusG– $\beta$ -protrusion interaction. Notably, none of these mutations would be predicted to prevent NusG from interacting with RNAP because of the extensive NusG–RNAP contacts via the swivel module<sup>13</sup>. Thus, these *nusG* and  $\beta$ -protrusion mutants would be expected to specifically weaken pro-pausing and termination activity but maintain other functions of the NusG–RNAP interaction.

To test the prediction that the homoplastic NusG mutants weaken its pro-pausing activity, we purified recombinant Mtb wild-type RNAP,  $\beta$ S450L RNAP, wild-type NusG, and three of the most common NusG variants observed in clinical RifR Mtb isolates. We then assessed





**Fig. 3 | NusG mutations are associated with RifR in clinical Mtb isolates.**

**a**, Manhattan plot showing the genetic association with RifR among 3,252 clinical Mtb isolates from Peru. Each circle represents a gene or intergenic region in the Mtb genome. The y axis represents uncorrected phyOverlap  $P$  values (blue,  $P \leq 0.05$ ). Circle sizes are scaled by the number of independent mutations observed for that gene or intergenic region.  $P$  values were derived from 50,000 permutations of mutations events, as described in Methods. **b**, Phylogenetic tree of Peruvian Mtb isolates from **a**. *Mycobacterium canettii* was included as an outgroup; isolates from lineages 2 and 4 (L2 and L4) are indicated. Purple lines represent isolates identified as RifR by genotypic drug susceptibility testing (gDST). Yellow triangles mark isolates with a known compensatory (comp) mutation in *rpoA/C* (Supplementary Table 2) and blue

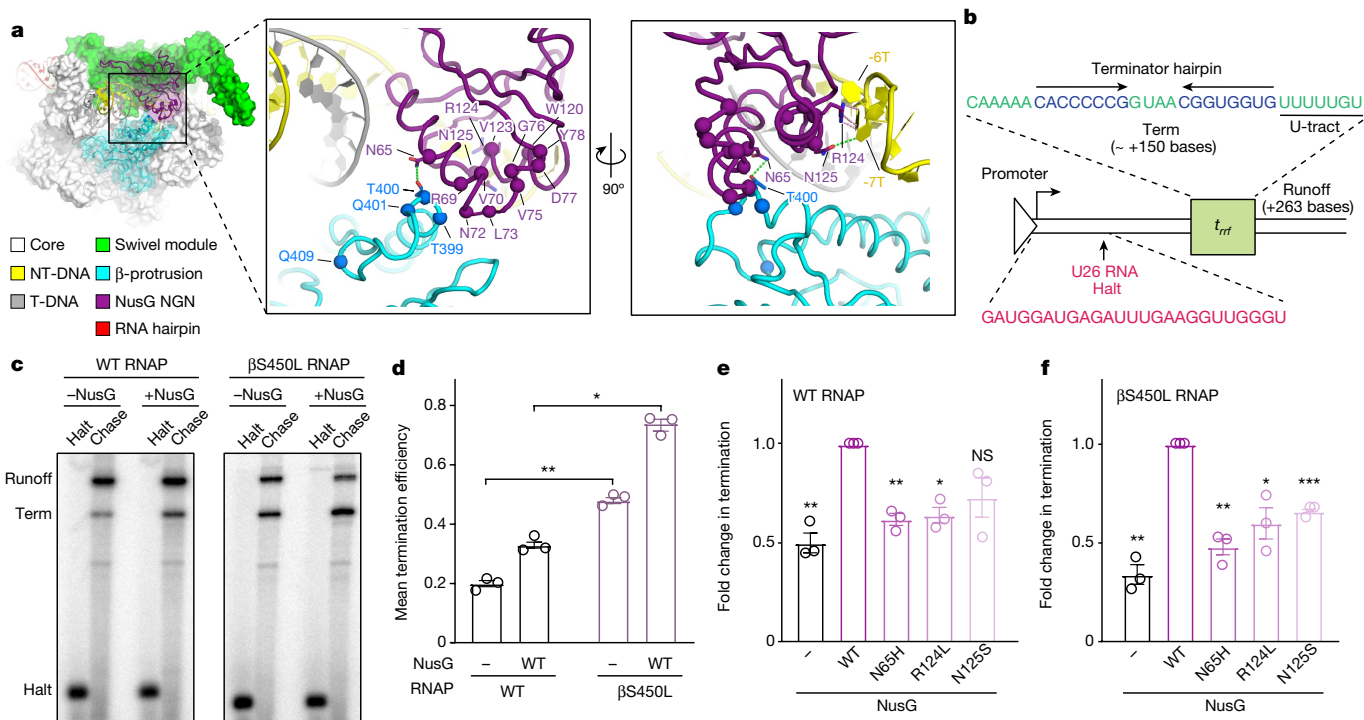
triangles mark isolates harbouring a nonsynonymous (NS) mutation in *nusG*. **c–e**, Box plots showing the ratio of nonsynonymous to synonymous mutations (dN/dS) for *nusG* and *rpoC* in genotypically predicted RifS (*nusG*,  $n = 1,365$ ; *rpoC*,  $n = 15,834$ ) and RifR (*nusG*,  $n = 350$ ; *rpoC*,  $n = 10,418$ ) Mtb (**c**), restricting the analysis to only RifR strains with mutations in Ser450 (*nusG*,  $n = 270$ ; *rpoC*,  $n = 7,898$ ) (**d**), or restricting the analysis to only RifR strains with mutations in His445 (*nusG*,  $n = 26$ ; *rpoC*,  $n = 996$ ) (**e**). X indicates any amino acid except Ser (S450) or His (H445). The centre line represents the mean, box edges delineate top and bottom quartiles, and whiskers indicate the minimum and maximum dN/dS values.  $P$  values by independent two-sample  $t$ -test. Two-sided  $P$  values: \*\* $P = 1.14 \times 10^{-3}$ , \*\*\* $P \leq 0.0001$ .

the effect of each variant on termination (and, by inference, RNAP pausing) using in vitro termination assays (Fig. 4b). In the absence of NusG,  $\beta$ S450L RNAP displayed increased termination compared to RifS RNAP<sup>20</sup> (Fig. 4c,d). In the presence of NusG, both RifS and  $\beta$ S450L RNAPs displayed increased levels of termination<sup>13</sup> (Fig. 4c,d). Consistent with the hypothesis that the clinical NusG variants weaken the interaction between NusG NGN and the  $\beta$ -protrusion and thereby decrease RNAP pausing and termination<sup>13</sup>, all three NusG mutants showed reduced levels of pro-pausing and termination activity when paired with either wild-type or  $\beta$ S450L RNAP (Fig. 4e,f). We note that NusG variants do not reduce  $\beta$ S450L RNAP termination to wild-type RNAP plus wild-type NusG levels in this assay, suggesting that RNAP pausing and termination levels may remain partially elevated in vivo. All NusG variants bind to RNAP similarly to wild-type NusG (Supplementary Fig. 2e,f). These data suggest that NusG and  $\beta$ -protrusion variants reduce RNAP swivelling dwell time and consequently decrease pausing and hyper-termination, rescuing the fitness defect in  $\beta$ S450L Mtb.

### NusG mutations boost $\beta$ S450L fitness

To test the prediction that *nusG* and  $\beta$ -protrusion variants may represent a new form of compensatory evolution to restore fitness in  $\beta$ S450L Mtb, we reconstructed these mutations in isogenic RifS and  $\beta$ S450L Mtb by single-stranded DNA (ssDNA) recombineering (Fig. 5a). These strains also contained unique molecular barcodes, thereby enabling competitive growth experiments and assessment of mutant abundance by deep sequencing. All strains were confirmed by whole-genome sequencing to harbour the targeted mutation and, in the case of  $\beta$ S450L strains, not to harbour untargeted compensatory mutations in *rpoA*, *rpoB* and *rpoC*.

To investigate the effect of these variants on the fitness of RifS and  $\beta$ S450L Mtb, we co-cultured all mutants in a competitive growth experiment, sequenced barcodes to assess mutant abundance, and calculated competitive indices to quantify relative fitness (Fig. 5b,c). As expected,  $\beta$ S450L was less fit than RifS Mtb (Supplementary Table 6). A control *nusG* mutation (T200A) that was found in both RifS and RifR clinical Mtb isolates and thus not expected to be compensatory had



**Fig. 4 | Clinical strain mutations decrease the pro-pausing activity of NusG.**

**a**, Cryo-EM structure of a NusG-bound paused elongation complex from *Mtb*. The boxed area shows the location of mutations observed in Rif<sup>R</sup> clinical *Mtb* isolates. Interactions between NusG and the  $\beta$ -protrusion or non-template DNA are shown in green dashed lines. **b**, Promoter template with the *Mtb rrf* termination sequence for in vitro transcription assays. The experimental schematic was followed for data shown in **c–f**. **c**, In vitro termination by *Mtb* RNAP on a promoter-initiated template with or without wild-type (WT) NusG. The U26, C-less halt site is labelled (Halt), as well as approximately +150 base termination site (Term) and +263 base runoff (Runoff). **d**, Termination efficiencies of wild-type and  $\beta$ S450L RNAP in the presence (two-sided adjusted  $P$  value = 0.039) or absence (two-sided adjusted  $P$  value = 0.002) of wild-type

NusG. Data are mean  $\pm$  s.d. from three experimental replicates. **e**, Fold change in termination ( $\Delta T$ ) relative to wild-type RNAP plus wild-type NusG. Data are mean  $\pm$  s.d. from three experimental replicates.  $\Delta T$  for each NusG protein was calculated from kinetics of the elongation versus termination process as described by von Hippel and Yager<sup>62,63</sup> based on the relationships between termination efficiency and the rates and activation energies of termination (Methods). Left to right, two-sided adjusted  $P$  value: 0.009, 0.006, 0.024 and 0.057. **f**, Fold change in termination as in **e**, relative to  $\beta$ S450L RNAP plus wild-type NusG. Left to right, two-sided adjusted  $P$  value: 0.009, 0.006, 0.038 and  $8.7 \times 10^{-4}$ . **(d–f)** Independent two-sample  $t$ -test, with multiple hypothesis correction by Benjamini, Krieger and Yekutieli method. \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ ; NS, not significant.

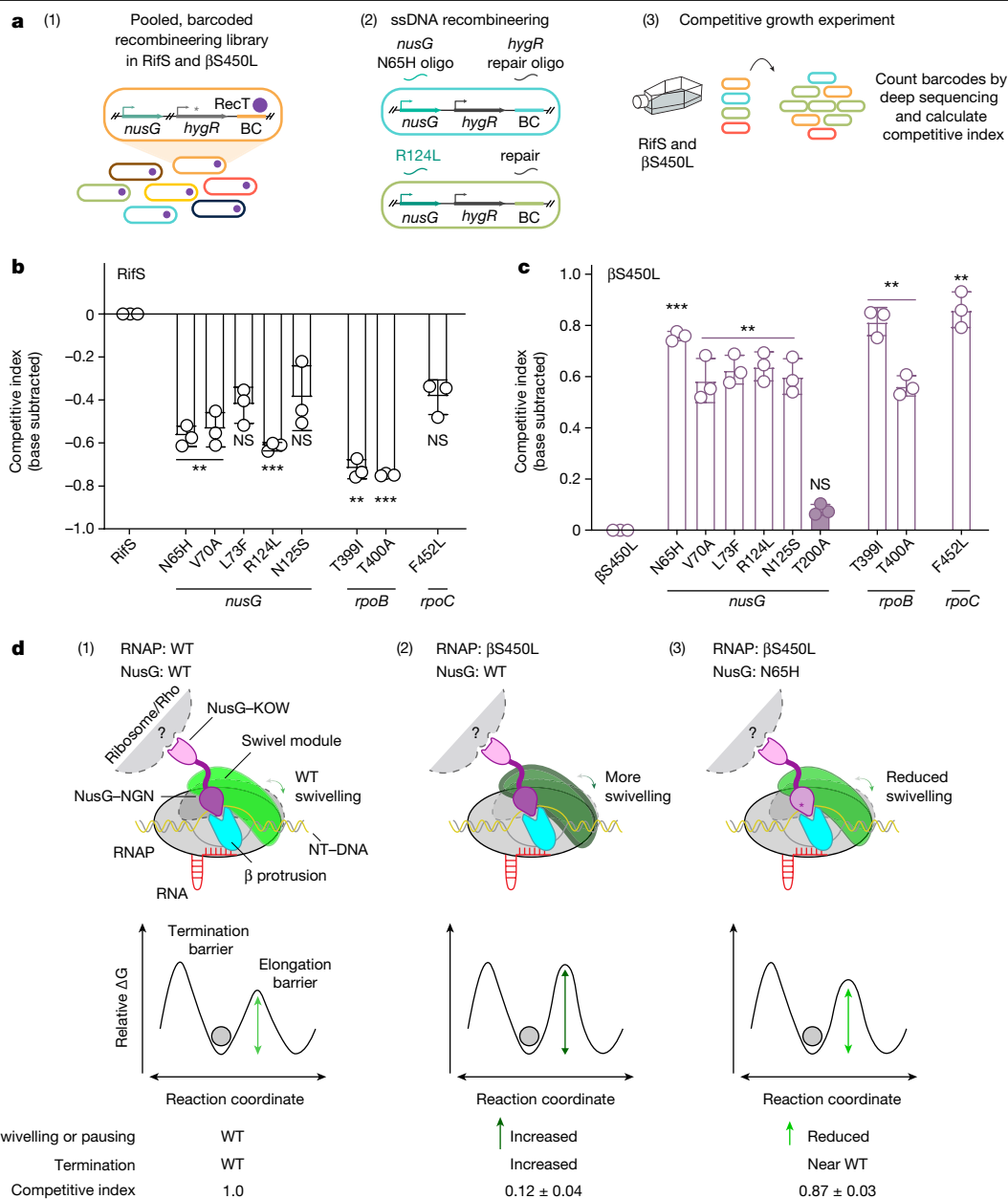
no effect on the fitness of  $\beta$ S450L *Mtb* (Fig. 5c). Consistent with its well-established compensatory behaviour, *rpoC*<sup>F452L</sup> increased the fitness of  $\beta$ S450L but decreased fitness of Rif<sup>S</sup> *Mtb*<sup>16,20</sup>. Similarly, all tested *nusG* and  $\beta$ -protrusion variants (except the T200A control) increased the fitness of  $\beta$ S450L but, if anything, decreased fitness of Rif<sup>S</sup> *Mtb*. Together, these data indicate that *nusG* and  $\beta$ -protrusion variants decrease the pro-pausing and termination activity of NusG and represent a new form of compensatory evolution to restore fitness in  $\beta$ S450L *Mtb* (Fig. 5d).

## Discussion

It has long been known that Rif<sup>R</sup> produces a fitness cost in bacteria. Here we show that the fitness cost of the most common Rif<sup>R</sup> mutation in *Mtb*— $\beta$ S450L—is driven at least in part by excessive RNAP pausing and termination. We further show that mutations in NusG or the  $\beta$ -protrusion reduce RNAP pausing and termination and increase  $\beta$ S450L fitness, and therefore represent a new form of compensatory evolution in Rif<sup>R</sup> *Mtb*.

We hypothesize that NusG has dual roles in regulating  $\beta$ S450L fitness. Whereas the pro-pausing function of NusG is detrimental for  $\beta$ S450L fitness, another aspect of NusG function promotes  $\beta$ S450L fitness. One way in which NusG could promote  $\beta$ S450L fitness is by mediating TTC. Consistent with a role in translation promoting  $\beta$ S450L fitness, we identify numerous tRNAs, tRNA synthetases, ribosomal proteins,

translation factors and amino acid biosynthetic genes as collateral vulnerabilities (Supplementary Table 1). These translation-related collateral vulnerabilities are not differentially expressed at the RNA level and the majority (22 out of 26 detected proteins) are not differentially expressed at the protein level between a panel of paired Rif<sup>S</sup> and  $\beta$ S450L Rif<sup>R</sup> *Mtb* clinical isolates<sup>27</sup> (Supplementary Fig. 3a,b). These results argue against a role for pervasive under-expression of translation-related genes as the sole explanation for increased vulnerability in  $\beta$ S450L. That *Mtb* NusG interacts with the ribosomal protein S10<sup>35</sup> and the existence of translationally coupled attenuation mechanisms<sup>34,42,43</sup> suggests that transcription and translation are coupled in *Mtb*, but direct demonstration of TTC in *Mtb* remains an important area of future research. Although TTC in *Mtb* is speculative, it is a reasonable hypothesis that is consistent with the available data. Alternatively, NusG could promote  $\beta$ S450L fitness for other reasons. For example, should *Mtb* NusG be capable of interacting with Rho, NusG could influence Rho-mediated termination<sup>36</sup>. However, the fact that Rho is not a differential vulnerability (Supplementary Fig. 3c,d) argues against a model in which NusG-stimulated Rho termination is a critical mechanism by which NusG promotes  $\beta$ S450L fitness. The recent *Mtb* NusG–RNAP structure is potentially consistent with a role for *Mtb* NusG in preventing RNAP backtracking<sup>13</sup>, which in turn could promote  $\beta$ S450L fitness by promoting RNAP elongation. Although the precise mechanisms by which NusG promotes  $\beta$ S450L fitness require further investigation, our work definitively shows that both the NusG



**Fig. 5 | Compensatory mutations in *nusG* increase the fitness of  $\beta$ S450L**

**Mtb.** **a**, Experimental design to quantify the competitive fitness of *nusG*, *rpoB* and *rpoC* mutants in RifS and  $\beta$ S450L Mtb. (1) Generation of barcode (BC) library; asterisk indicates premature stop codon. (2) ssDNA recombineering to generate mutants of interest. (3) Pooled, competitive growth experiment. Oligo, oligonucleotide. **b,c**, Competitive index (mean  $\pm$  s.d. from 3 biological replicates) of each mutant relative to the RifS base strain (no mutations in *nusG*, *rpoB* or *rpoC*). The competitive index of each mutant was normalized to its respective base strain, RifS or  $\beta$ S450L, by subtracting the competitive index of the base strain. *nusG*-T200A is a control mutation not associated with RifR in clinical isolates and is thus not expected to restore fitness to  $\beta$ S450L Mtb. For reference, the fitness of  $\beta$ S450L relative to RifS in this experiment is 0.96.

Paired two-sided *t*-test. \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ . **d**, Model explaining the mechanism of reduced fitness of  $\beta$ S450L Mtb and its compensation. The  $\beta$ S450L RNAP elongates more slowly and is thus more likely to swivel and enter the paused and termination states. This effect is exacerbated by the pro-pausing wild-type NusG, as illustrated by an increase in the activation barrier of elongation. These interactions decrease  $\beta$ S450L Mtb fitness. Note that the competitive index values shown here are not base-subtracted as they are in **b,c**. Compensatory mutations at the NusG- $\beta$ -protrusion or NusG-NT-DNA interface reduce the ability of NusG to stabilize the swivelled state, as illustrated by a relative decrease in the activation barrier of elongation. These mutations restore  $\beta$ S450L RNAP pausing and termination levels closer to wild-type RNAP and increase  $\beta$ S450L Mtb fitness.

NGN domain and  $\beta$ -protrusion are under selective pressure to mutate to dampen pro-pausing and termination activity to restore  $\beta$ S450L fitness.

Our identification of a genetic association between *nusG* variants and RifR is consistent with other recent work that observed signs of selection in Mtb *nusG*<sup>44,45</sup>. We found that the NusG NGN domain and the RNAP  $\beta$ -protrusion are under positive selection in specific RifR *rpoB* mutants. The vast majority (>95%) of *nusG* or  $\beta$ -protrusion-compensated RifR

Mtb strains are  $\beta$ S450L, as is also true for strains harbouring compensatory mutations in *rpoC* or *rpoA*<sup>12</sup> (Supplementary Fig. 2b and Supplementary Table 2). In a seminal paper, Gagneux and colleagues<sup>15</sup> performed a laboratory evolution experiment of RifR Mtb to identify mechanisms of fitness compensation. We reanalysed their data and—in addition to their discovery of compensatory mutations in *rpoC* and *rpoA*—we identified a previously overlooked compensatory mutation

(T399A) in the  $\beta$ -protrusion. The frequency of compensation by *nusG* or  $\beta$ -protrusion mutations is similar to that seen for *rpoA*, but compensation by *rpoC* mutations is more common (Fig. 3b, Supplementary Fig. 2b and Supplementary Table 2). The higher frequency of compensation by *rpoC* mutations probably reflects at least in part a larger mutational target size in *rpoC* (Supplementary Table 2), but it could also result from more efficient biochemical compensation of the  $\beta$ S450L mutation by secondary mutations in *rpoC*.

Our findings can be understood when considering the recent structures of Mtb RNAP bound to NusG<sup>13</sup>. The bridge helix and the trigger loop are two conserved structural modules in the active site of all cellular RNAPs. Concerted conformational changes of the bridge helix and folding and unfolding transitions of the trigger loop are essential for progression through the nucleotide addition cycle. RNAP swivelling leads to bridge helix distortions that interfere with: (1) the positioning of the DNA base templating the incoming NTP substrate in the active site; and (2) trigger loop folding, which collectively inhibits RNAP elongation. Thus, swivelling is likely to be an off-line RNAP conformation that kinetically competes with normal elongation<sup>39,46</sup>. Whether by cause or effect, the slower elongation rate of the  $\beta$ S450L RNAP is associated with a higher propensity for swivelling, explaining the elevated pause half-lives and termination rates for the mutant RNAP<sup>20</sup> (Figs. 4c,d and 5d). Mtb NusG promotes RNAP pausing and termination by stabilizing the swivelled conformation<sup>13</sup>. We show here that the  $\beta$ S450L mutation results in selection for additional mutations in either of the two NusG interfaces used to stabilize pausing: the NusG- $\beta$ -protrusion interface or the NusG-NT-DNA interface (Fig. 5d). These compensatory mutations reduce the stabilizing effects of NusG on swivelling and temper the hyper-termination effects of the  $\beta$ S450L substitution. Thus, increased pausing and termination is at least one source of the fitness cost of  $\beta$ S450L, and one form of compensation is to reduce these activities. Whether compensatory mutations within the RNAP also work by reducing the propensity of  $\beta$ S450L RNAP to swivel and pause and could thus represent a 'unifying principle' of  $\beta$ S450L compensation represents an important area of future research.

The hyper-termination phenotype of  $\beta$ S450L RNAP presumably leads to widespread transcriptome changes, but whether expression dysregulation of a few or many genes underlies the fitness cost of  $\beta$ S450L remains an open question. RNA-sequencing and quantitative proteomic studies have shown subtle but pleiotropic differences between the transcriptomes and proteomes of  $\beta$ S450L and Rif<sup>R</sup> Mtb<sup>24,27,47</sup>. One might expect that genes whose under-expression contributes to the fitness cost of  $\beta$ S450L might be regulated by RNAP pausing and termination and be enriched as collateral vulnerabilities in the  $\beta$ S450L screen, a hypothesis that will be tested in future studies. Should such genes be identified, they could in principle serve as additional sources of fitness compensating mutations outside of NusG and RNAP.

Rif<sup>R</sup> RNAP compensatory mutations have been known for more than two decades<sup>12,15,16,19,48</sup>. To our knowledge, our work represents the first biochemically defined mechanism of compensation for RNAP. Intriguingly, compensation works late during the transcription cycle, at the level of transcription elongation. Our work builds on foundational studies that first demonstrated the fitness cost of chromosomal drug-resistance mutations<sup>1,4,5</sup> and its compensation<sup>1,2,4,7,48-51</sup>. Beyond Rif and RNAP, one of the best studied drug targets is the bacterial ribosome and translation. Several studies in diverse bacteria have shown that resistance to translation-perturbing antibiotics frequently comes with a fitness cost and that there exist diverse mechanisms to compensate for that cost. For example, resistance to aminoglycoside and tuberculinomycin antibiotics can occur through mutations in the 16S ribosomal RNA (rRNA), and compensation can occur by intragenic secondary mutations that restore 16S rRNA secondary structure or non-mutational mechanisms that increase 16S rRNA methylation<sup>8,52,53</sup>. Aminoglycoside resistance can also be caused by mutations in *rpsL*, which encodes the ribosomal protein S12. Although low-cost mutations are by far the

most common in clinical Mtb isolates<sup>8</sup>, high-cost mutations can be compensated by a constellation of secondary mutations in *rpsL* or in genes encoding other ribosomal proteins. At least some of these mutations may serve to restore ribosomal translation fidelity altered by the original resistance mutation<sup>4,54</sup>. Peptide deformylase inhibitors target peptide deformylase, a bacterial enzyme that removes the formyl group from the N-terminal methionine in nascent polypeptides. Resistance to peptide deformylase inhibitors commonly occurs by reducing formylation of the methionyl initiator tRNA (Met-tRNAi)—for example, as a result of loss-of-function mutations in formyl-methionine-transferase<sup>55</sup> (*fnt*). As a consequence of *fnt* mutations, both translation and growth rates are reduced<sup>56</sup>. Compensation for the fitness cost of *fnt* mutants can occur by a myriad of mechanisms, many of which ultimately enable the ribosome to initiate translation without a formylated methionyl initiator tRNA<sup>57-59</sup>. Given these precedents and the evolutionary flexibility of bacteria, it will be important to determine potential fitness costs and mechanisms of compensation for newly developed antibiotics.

The fitness cost of drug resistance serves as the basis for antibiotic combination approaches that aim to limit the evolution and spread of resistant mutants<sup>30</sup>. For example, collateral sensitivity represents an evolutionary trade-off whereby resistance to one antibiotic confers increased sensitivity to another antibiotic. One can imagine several ways in which the findings presented here could be used to this end. For example, the fact that the pro-pausing activity of NusG reduces  $\beta$ S450L fitness suggests that compounds that promote RNAP pausing may be particularly effective in preventing  $\beta$ S450L evolution. Alternatively, small molecule inhibitors of alternative collateral vulnerabilities identified in the screen might serve as good partner drugs with Rif to slow the evolution of  $\beta$ S450L.

We show that excessive pausing and termination contributes to the fitness cost of Rif<sup>R</sup> in Mtb, and that Rif<sup>R</sup> Mtb selects for compensatory mutations in NusG and the  $\beta$ -protrusion that partially alleviate this defect. Compensatory mutations may improve patient-to-patient Mtb transmission<sup>17</sup>, and indeed Mtb clinical strains harbouring mutations in *nusG* were previously associated with increased Mtb transmission in Pakistan<sup>60</sup>. Thus, molecular diagnostic approaches to detect fitness compensated drug-resistant tuberculosis could be used to identify patients with strains that are more likely to be transmitted. Our findings suggest rational approaches to exacerbate the fitness cost of Rif<sup>R</sup>, which could inform new therapeutic approaches to develop drug combinations to slow the evolution of Rif<sup>R</sup> Mtb. Our approach is generalizable to other organisms, setting the stage for expanded differential vulnerability studies to better understand and ultimately target unique vulnerabilities of drug-resistant pathogens.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-024-07206-5>.

1. Andersson, D. I. & Hughes, D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat. Rev. Microbiol.* **8**, 260–271 (2010).
2. Björkholm, B. et al. Mutation frequency and biological cost of antibiotic resistance in *Helicobacter pylori*. *Proc. Natl Acad. Sci. USA* **98**, 14607–14612 (2001).
3. Andersson, D. I. & Levin, B. R. The biological cost of antibiotic resistance. *Curr. Opin. Microbiol.* **2**, 489–493 (1999).
4. Björkman, J., Hughes, D. & Andersson, D. I. Virulence of antibiotic-resistant *Salmonella typhimurium*. *Proc. Natl Acad. Sci. USA* **95**, 3949–3953 (1998).
5. Billington, O. J., McHugh, T. D. & Gillespie, S. H. Physiological cost of rifampin resistance induced in vitro in *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **43**, 1866–1869 (1999).
6. Mariam, D. H., Mengistu, Y., Hoffner, S. E. & Andersson, D. I. Effect of *rpoB* mutations conferring rifampin resistance on fitness of *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **48**, 1289–1294 (2004).



7. Gagneux, S. et al. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science* **312**, 1944–1946 (2006).
8. Sander, P. et al. Fitness cost of chromosomal drug resistance-conferring mutations. *Antimicrob. Agents Chemother.* **46**, 1204–1211 (2002).
9. Mitchison, D. A. Role of individual drugs in the chemotherapy of tuberculosis. *Int. J. Tuberc. Lung Dis.* **4**, 796–806 (2000).
10. Campbell, E. A. et al. Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* **104**, 901–912 (2001).
11. WHO. *Global Tuberculosis Report 2023* (2023); <https://www.who.int/publications/i/item/9789240083851>.
12. Borrell, S. & Trauner, A. Strain diversity and the evolution of antibiotic resistance. *Adv. Exp. Med. Biol.* **1019**, 263–279 (2017).
13. Delbeau, M. et al. Structural and functional basis of the universal transcription factor NusG pro-pausing activity in *Mycobacterium tuberculosis*. *Mol. Cell* **83**, 1474–1488.e8 (2023).
14. Cohen, T. & Murray, M. Modeling epidemics of multidrug-resistant *M. tuberculosis* of heterogeneous fitness. *Nat. Med.* **10**, 1117–1121 (2004).
15. Comas, I. et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat. Genet.* **44**, 106–110 (2011).
16. Song, T. et al. Fitness costs of rifampicin resistance in *Mycobacterium tuberculosis* are amplified under conditions of nutrient starvation and compensated by mutation in the  $\beta'$  subunit of RNA polymerase. *Mol. Microbiol.* **91**, 1106–1119 (2014).
17. Goig, G. A. et al. Effect of compensatory evolution in the emergence and transmission of rifampicin-resistant *Mycobacterium tuberculosis* in Cape Town, South Africa: a genomic epidemiology study. *Lancet Microbe* **4**, e506–e515 (2023).
18. Sherman, D. R. et al. Compensatory ahpC gene expression in isoniazid-resistant *Mycobacterium tuberculosis*. *Science* **272**, 1641–1643 (1996).
19. Brandis, G. & Hughes, D. Genetic characterization of compensatory evolution in strains carrying rpoB Ser531Leu, the rifampicin resistance mutation most frequently found in clinical isolates. *J. Antimicrob. Chemother.* **68**, 2493–2497 (2013).
20. Stefan, M. A., Ugur, F. S. & Garcia, G. A. Source of the fitness defect in rifampicin-resistant *Mycobacterium tuberculosis* RNA polymerase and the mechanism of compensation by mutations in the  $\beta'$  subunit. *Antimicrob. Agents Chemother.* **62**, e00164–18 (2018).
21. Yanofsky, C. & Horn, V. Rifampin resistance mutations that alter the efficiency of transcription termination at the tryptophan operon attenuator. *J. Bacteriol.* **145**, 1334–1341 (1981).
22. Jin, D. J., Walter, W. A. & Gross, C. A. Characterization of the termination phenotypes of rifampicin-resistant mutants. *J. Mol. Biol.* **202**, 245–253 (1988).
23. Campodónico, V. L., Rifat, D., Chuang, Y.-M., Joergler, T. R. & Karakousis, P. C. Altered *Mycobacterium tuberculosis* cell wall metabolism and physiology associated with RpoB mutation H526D. *Front. Microbiol.* **9**, 494 (2018).
24. Bisson, G. P. et al. Upregulation of the phthiocerol dimycocerosate biosynthetic pathway by rifampin-resistant, rpoB mutant *Mycobacterium tuberculosis*. *J. Bacteriol.* **194**, 6441–6452 (2012).
25. Lahiri, N. et al. Rifampin resistance mutations are associated with broad chemical remodeling of *Mycobacterium tuberculosis*. *J. Biol. Chem.* **291**, 14248–14256 (2016).
26. Howard, N. C. et al. *Mycobacterium tuberculosis* carrying a rifampicin drug resistance mutation reprograms macrophage metabolism through cell wall lipid changes. *Nat. Microbiol.* **3**, 1099–1108 (2018).
27. Trauner, A. et al. Expression dysregulation as a mediator of fitness costs in antibiotic resistance. *Antimicrob. Agents Chemother.* **65**, e0050421 (2021).
28. Bosch, B. et al. Genome-wide gene expression tuning reveals diverse vulnerabilities of *M. tuberculosis*. *Cell* **184**, 4579–4592.e24 (2021).
29. Rock, J. M. et al. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat. Microbiol.* **2**, 16274 (2017).
30. Baym, M., Stone, L. K. & Kishony, R. Multidrug evolutionary strategies to reverse antibiotic resistance. *Science* **351**, aad3292 (2016).
31. Wee, L. M. et al. A trailing ribosome speeds up RNA polymerase at the expense of transcript fidelity via force and allostery. *Cell* **186**, 1244–1262.e34 (2023).
32. Saxena, S. et al. *Escherichia coli* transcription factor NusG binds to 70S ribosomes. *Mol. Microbiol.* **108**, 495–504 (2018).
33. Washburn, R. S. et al. *Escherichia coli* NusG Links the lead ribosome with the transcription elongation complex. *iScience* **23**, 101352 (2020).
34. Ju, X. et al. Incomplete transcripts dominate the *Mycobacterium tuberculosis* transcriptome. *Nature* <https://doi.org/10.1038/s41586-024-07105-9> (2024).
35. Kalyani, B. S., Kunamneni, R., Wal, M., Ranjan, A. & Sen, R. A NusG paralogue from *Mycobacterium tuberculosis*, Rv0639, has evolved to interact with ribosomal protein S10 (Rv0700) but not to function as a transcription elongation-termination factor. *Microbiology* **161**, 67–83 (2015).
36. Wang, B. & Artsimovitch, I. NusG, an ancient yet rapidly evolving transcription factor. *Front. Microbiol.* **11**, 619618 (2020).
37. Ray-Soni, A., Bellecourt, M. J. & Landick, R. Mechanisms of bacterial transcription termination: all good things must end. *Annu. Rev. Biochem.* **85**, 319–347 (2016).
38. Czyz, A., Mooney, R. A., Iaconi, A. & Landick, R. Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *mBio* **5**, e00931 (2014).
39. Kang, J. Y. et al. RNA polymerase accommodates a pause RNA hairpin by global conformational rearrangements that prolong pausing. *Mol. Cell* **69**, 802–815.e5 (2018).
40. Huang, C.-C. et al. Spatial scale of tuberculosis transmission in Lima, Peru. *Proc. Natl Acad. Sci. USA* **119**, e2207022119 (2022).
41. Li, S. et al. CRISPRi chemical genetics and comparative genomics identify genes mediating drug potency in *Mycobacterium tuberculosis*. *Nat. Microbiol.* **7**, 766–779 (2022).
42. Lee, J.-H., Lee, E.-J. & Roe, J.-H. uORF-mediated riboregulation controls transcription of *whiB7/wblC* antibiotic resistance gene. *Mol. Microbiol.* **117**, 179–192 (2022).
43. D'Halluin, A. et al. Premature termination of transcription is shaped by Rho and translated uORFs in *Mycobacterium tuberculosis*. *iScience* **26**, 106465 (2023).
44. Liu, Q. et al. Tuberculosis treatment failure associated with evolution of antibiotic resilience. *Science* **378**, 1111–1118 (2022).
45. Green, A. G. et al. Analysis of genome-wide mutational dependence in naturally evolving *Mycobacterium tuberculosis* populations. *Mol. Biol. Evol.* **40**, msad131 (2023).
46. Kang, J. Y., Mishanina, T. V., Landick, R. & Darst, S. A. Mechanisms of transcriptional pausing in bacteria. *J. Mol. Biol.* **431**, 4007–4029 (2019).
47. Xu, Z. et al. Transcriptional approach for decoding the mechanism of rpoC compensatory mutations for the fitness cost in rifampicin-resistant *Mycobacterium tuberculosis*. *Front. Microbiol.* **9**, 2895 (2018).
48. Reynolds, M. G. Compensatory evolution in rifampin-resistant *Escherichia coli*. *Genetics* **156**, 1471–1481 (2000).
49. Schrag, S. J. & Perrot, V. Reducing antibiotic resistance. *Nature* **381**, 120–121 (1996).
50. Björkman, J., Nagaev, I., Berg, O. G., Hughes, D. & Andersson, D. I. Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. *Science* **287**, 1479–1482 (2000).
51. Nagaev, I., Björkman, J., Andersson, D. I. & Hughes, D. Biological cost and compensatory evolution in fusidic acid-resistant *Staphylococcus aureus*. *Mol. Microbiol.* **40**, 433–439 (2001).
52. Shcherbakov, D. et al. Directed mutagenesis of *Mycobacterium smegmatis* 16S rRNA to reconstruct the in vivo evolution of aminoglycoside resistance in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **77**, 830–840 (2010).
53. Freihofer, P. et al. Nonmutational compensation of the fitness cost of antibiotic resistance in mycobacteria by overexpression of tlyA rRNA methylase. *RNA* **22**, 1836–1843 (2016).
54. Maisnier-Patin, S., Berg, O. G., Liljas, L. & Andersson, D. I. Compensatory adaptation to the deleterious effect of antibiotic resistance in *Salmonella typhimurium*. *Mol. Microbiol.* **46**, 355–366 (2002).
55. Margolis, P. S. et al. Peptide deformylase in *Staphylococcus aureus*: resistance to inhibition is mediated by mutations in the formyltransferase gene. *Antimicrob. Agents Chemother.* **44**, 1825–1831 (2000).
56. Guillon, J. M., Mechulam, Y., Schmitter, J. M., Blanquet, S. & Fayat, G. Disruption of the gene for Met-tRNA(fMet) formyltransferase severely impairs growth of *Escherichia coli*. *J. Bacteriol.* **174**, 4294–4301 (1992).
57. Guillon, J. M. et al. Interplay of methionine tRNAs with translation elongation factor Tu and translation initiation factor 2 in *Escherichia coli*. *J. Biol. Chem.* **271**, 22321–22325 (1996).
58. Nilsson, A. I. et al. Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. *Proc. Natl Acad. Sci. USA* **103**, 6976–6981 (2006).
59. Zorzet, A., Pavlov, M. Y., Nilsson, A. I., Ehrenberg, M. & Andersson, D. I. Error-prone initiation factor 2 mutations reduce the fitness cost of antibiotic resistance. *Mol. Microbiol.* **75**, 1299–1313 (2010).
60. Napier, G. et al. Characterisation of drug-resistant *Mycobacterium tuberculosis* mutations and transmission in Pakistan. *Sci. Rep.* **12**, 7703 (2022).
61. DeJesus, M. A. et al. Comprehensive essentiality analysis of the *Mycobacterium tuberculosis* genome via saturating transposon mutagenesis. *mBio* **8**, e02133-16 (2017).
62. von Hippel, P. H. & Yager, T. D. Transcript elongation and termination are competitive kinetic processes. *Proc. Natl Acad. Sci. USA* **88**, 2307–2311 (1991).
63. von Hippel, P. H. & Yager, T. D. The elongation–termination decision in transcription. *Science* **255**, 809–812 (1992).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

## Methods

### Bacterial strains

Mtb strains are derivatives of H37Rv unless otherwise noted. *ΔbioA* Mtb was obtained from the Schnappinger laboratory<sup>64</sup>. *E. coli* strains are derivatives of DH5α (NEB), Rosetta2, or BL21(DE3) (Novagen).

### Mycobacterial cultures

Mtb was grown at 37 °C in Difco Middlebrook 7H9 broth or on 7H10 agar supplemented with 0.2% glycerol (7H9) or 0.5% glycerol (7H10), 0.05% Tween-80, 1× oleic acid-albumin-dextrose-catalase (OADC) and the appropriate antibiotics (kanamycin 10–20 μg ml<sup>-1</sup> and/or hygromycin 25–50 μg ml<sup>-1</sup>). ATc was used at 100 ng ml<sup>-1</sup>. Mtb cultures were grown standing in tissue culture flasks (unless otherwise indicated) with 5% CO<sub>2</sub>. Note that both 7H9 and 7H10 medium are normally supplemented with biotin (0.5 mg l<sup>-1</sup>; ~2 μM), thereby allowing growth of the *ΔbioA* Mtb auxotroph.

### Selection of Rif-resistant Mtb isolates

For the selection of Rif<sup>R</sup> H37Rv and *ΔbioA* Mtb, 5 independent 5-ml cultures were started at a density of ~2,000 cells per ml (to minimize the number of preexisting Rif<sup>R</sup> bacteria) and grown to stationary phase (OD<sub>600</sub> > 1.5). Cultures were pelleted at 4,000 rpm for 10 min, resuspended in 30 μl remaining medium per pellet and plated on 7H10 agar supplemented with Rif at 0.5 μg ml<sup>-1</sup>. After outgrowth, colonies were picked into 7H9 medium. After 1 week of outgrowth, an aliquot was heat-inactivated and the Rif resistance determining region of *rpoB*, *rpoA* and *rpoC* were amplified by PCR and Sanger sequenced. See Supplementary Table 4 for primer sequences.

### Generation of structural models

The structural model of Mtb RNAP transcription initiation complex bound to Rif in Fig. 1a was generated by modelling *Mycobacterium smegmatis* RNAP bound to Rif (PDB: 6CCV)<sup>65</sup> on to the transcription initiation complex structure (PDB: 6EDT)<sup>66</sup>.

The cryo-EM structures of a NusG-bound paused elongation complex from Mtb (PDB: 8E74) in Fig. 2d, and the location of clinical isolate mutations in Fig. 4a are derived from Delbeau et al.<sup>13</sup>.

### Generation of individual CRISPRi strains

Individual CRISPRi plasmids were cloned as described<sup>67</sup> using Addgene plasmid 166886. In brief, the CRISPRi plasmid backbone was digested with BsmBI-v2 (NEB R0739L) and gel-purified. sgRNAs were designed to target the non-template strand of the target gene open reading frame (ORF). For each individual sgRNA, two complementary oligonucleotides with appropriate sticky end overhangs were annealed and ligated (T4 ligase NEB M0202 M) into the BsmBI-digested plasmid backbone. Successful cloning was confirmed by Sanger sequencing.

Individual CRISPRi plasmids were then electroporated into Mtb. Electrocompetent cells were obtained as described<sup>68</sup>. In brief, an Mtb culture was expanded to an OD<sub>600</sub> = 0.4–0.6 and treated with glycine (final concentration 0.2M) for 24 h before pelleting (4,000g for 10 min). The cell pellet was washed three times in sterile 10% glycerol. The washed bacilli were then resuspended in 10% glycerol in a final volume of 5% of the original culture volume. For each transformation, 100 ng plasmid DNA and 100 μl electrocompetent mycobacteria were mixed and transferred to a 2 mm electroporation cuvette (Bio-Rad 1652082). Where necessary, 100 ng plasmid pIRL19 (Addgene plasmid 163634) was also added. Electroporation was performed using the Gene Pulser X cell electroporation system (Bio-Rad 1652660) set at 2,500 V, 700 Ω and 25 μF. Bacteria were recovered in 7H9 for 24 h. After the recovery incubation, cells were plated on 7H10 agar supplemented with the appropriate antibiotic to select for transformants.

### CRISPRi library transformation

CRISPRi libraries were generated as described previously<sup>28</sup>. In brief, fifty transformations were performed to generate Rif<sup>S</sup> and βS450L *ΔbioA* libraries. For each transformation, 1 μg of RLC12 plasmid DNA was added to 100 μl electrocompetent cells. The cells:DNA mix was transferred to a 2 mm electroporation cuvette (Bio-Rad 1652082) and electroporated at 2,500 kV, 700 Ω, and 25 μF. Each transformation was recovered in 2 ml 7H9 medium supplemented with OADC, glycerol and Tween-80 (100 ml total) for 16–24 h. The recovered cells were collected at 4,000 rpm for 10 min, resuspended in 400 μl remaining medium per transformation and plated on 7H10 agar supplemented with kanamycin (see 'Mycobacterial cultures') in Corning Bioassay dishes (Sigma CLS431111-16EA).

After 21 days of outgrowth on plates, transformants were scraped and pooled. Scraped cells were homogenized by two dissociation cycles on a gentleMACS Octo Dissociator (Miltenyi Biotec 130095937) using the RNA\_01 program and 30 gentleMACS M tubes (Miltenyi Biotec 130093236). The library was further declumped by passaging 1 ml of homogenized library into 100 ml of 7H9 supplemented with kanamycin (see Mycobacterial cultures) for between 5 and 10 generations. Final Rif<sup>S</sup> and βS450L *ΔbioA* Mtb library stocks were obtained after passing the cultures through a 10-μm cell strainer (Pluriselect SKU 43-50010-03). Genomic DNA was extracted from the final stocks and library quality was validated by deep sequencing (see 'Genomic DNA extraction and library preparation for Illumina sequencing').

### Pooled CRISPRi screen

Pooled CRISPRi screens were performed as described<sup>28</sup>. In brief, 20-ml cultures were grown in vented tissue culture flasks (T-75; Falcon 353136) and 7H9 medium supplemented with kanamycin (see 'Mycobacterial cultures') and maintained at 37 °C, 5% CO<sub>2</sub> in a humidified incubator.

The screen was initiated by thawing 4×1-ml aliquots of the Mtb *ΔbioA* (Rif<sup>S</sup> or βS450L) CRISPRi library (RLC12) and inoculating each aliquot into 24 ml 7H9 medium supplemented with kanamycin in a T-75 flask (starting OD<sub>600</sub> ~ 0.06). The cultures were expanded to approximately OD<sub>600</sub> = 1.0, pooled and passed through a 10-μm cell strainer (pluriSelect 43-50010-03) to obtain a single cell suspension. The single cell suspension (flow-through) was used to set up six 'generation 0' cultures: three replicate cultures with ATc (+ATc) and three replicate control cultures without ATc (-ATc). From each generation 0 culture, we collected 10 OD<sub>600</sub> units of bacteria (~3 × 10<sup>9</sup> bacteria; ~30,000X coverage of the CRISPRi library) for genomic DNA extraction. The remaining culture volume was used to initiate the pooled CRISPRi fitness screen. Cultures were periodically passaged in pre-warmed medium in order to maintain log phase growth. At generation 2.5, 5, and 7.5, cultures were back-diluted 1:6 (to a starting OD<sub>600</sub> = 0.2) and cultivated for approximately 2.5 doublings. At generation 10, 15, 20, and 25, cultures were back-diluted 1:24 (to a starting OD<sub>600</sub> = 0.05) and expanded for 5 generations before reaching late-log phase. ATc was replenished at every passage. By keeping the OD<sub>600</sub> of the 20 ml cultures ≥ 0.05, we guaranteed sufficient coverage of the library (3,000X) at all times. At set time points (approximately 2.5; 5; 7.5; 10; 15; 20; 25 and 30 generations), we collected bacterial pellets (10 OD<sub>600</sub> units) to extract genomic DNA.

### Genomic DNA extraction and library preparation for Illumina sequencing of CRISPRi libraries

Genomic DNA was isolated from bacterial pellets using the CTAB-lysozyme method described previously<sup>69</sup>. Genomic DNA concentration was quantified using the DeNovix dsDNA high sensitivity assay (KIT-DSDNA-HIGH-2; DS-11 Series Spectrophotometer/Fluorometer).

Illumina libraries were constructed as described<sup>28</sup>. In brief, the sgRNA-encoding region was amplified from 500 ng genomic DNA using NEBNext Ultra II Q5 master Mix (NEB M0544L). PCR cycling conditions were: 98 °C for 45 s; 17 cycles of 98 °C for 10 s, 64 °C for

# Article

30 s, 65 °C for 20 s; 65 °C for 5 min. Each PCR reaction a unique indexed forward primer (0.5 μM final concentration) and a unique indexed reverse primer (0.5 μM) (Supplementary Table 4). Forward primers contain a P5 flow cell attachment sequence, a standard Read1 Illumina sequencing primer binding site, custom stagger sequences to ensure base diversity during Illumina sequencing, and unique barcodes to allow for sample pooling during deep sequencing. Reverse primers contain a P7 flow cell attachment sequence, a standard Read2 Illumina sequencing primer binding site, and unique barcodes.

Following PCR amplification, each ~230 bp amplicon was purified using AMPure XP beads (Beckman–Coulter A63882) using two-sided selection (0.75X and 0.12X). Eluted amplicons were quantified with a Qubit 2.0 Fluorometer (Invitrogen), and amplicon size and purity were quality controlled by visualization on an Agilent 4200 TapeStation (Instrument- Agilent Technologies G2991AA; reagents- Agilent Technologies 5067-5583; tape- Agilent Technologies 5067-5582). Next, individual PCR amplicons were multiplexed into 20 nM pools and sequenced on an Illumina sequencer according to the manufacturer's instructions. To increase sequencing diversity, a PhiX spike-in of 2.5–5% was added to the pools (PhiX sequencing control v3; Illumina FC-110-3001). Samples were run on the Illumina NextSeq 500 or NovaSeq 6000 platform (single-read 1 × 85 cycles, 8 × i5 index cycles, and 8 × i7 index cycles).

## Differential vulnerability analysis of Rif-resistant versus Rif-sensitive strains

Gene vulnerability in the RifS and βS450L Mtb strains was determined using an updated vulnerability model based on the one previously described<sup>28</sup>. In the updated model, read counts for a given sgRNA in the minus ATc conditions were modelled using a negative binomial distribution with a mean proportional to the counts in the plus ATc condition, plus a factor representing the log<sub>2</sub> fold change:

$$y_i^{-\text{ATc}} \sim \text{NegBinom}(\eta_i, \phi)$$
$$\eta_i = \log(y_i^{+\text{ATc}} + \lambda_i) + \text{TwoLine}(x_i, \alpha_i, \beta_i, \gamma, \beta_e)$$

where  $\lambda_i$  is an sgRNA-level correction factor estimated by the model,  $x_i$  represents the generations analysed for the  $i$ th guide, and the TwoLine function represents the piecewise linear function previously described, which models sgRNA behaviour over the logistic function describing gene-level vulnerabilities was simplified by setting the top asymptote of the curve (previously  $K$ ) equal to 0, representing the fact that weakest possible sgRNAs are expected to impose no effect on bacterial fitness, that is:

$$\text{Logistic}(s) = \frac{\beta_{\max}}{(1 + e^{(-H \cdot (s - M))})}$$

The Bayesian vulnerability model was run for each condition independently, and samples for all the parameters were obtained using Stan running 4 independent chains with 1,000 warmup iterations and 3,000 samples each (for a total of 12,000 posterior samples for each parameter in the model after discarding warmup iterations).

Differential vulnerabilities were estimated by two approaches. First, for each gene, the difference in pairwise (guide-level) vulnerability estimates was obtained, resulting in posterior samples of the differential vulnerability (delta-vulnerability). This effectively estimated the difference in the integrals of the vulnerability functions. If the 95% credible region did not overlap 0.0 those were taken as significant differential vulnerabilities between the strains.

Next, to identify differences between genes which may not exhibit the expected dose–response curve, we estimated the fitness cost (log<sub>2</sub>FC) predicted by our model for a (theoretical) sgRNA of strength 0.0 (that

is, Logistic( $s = 0$ )). This represented the weakest phenotype theoretically possible with our CRISPRi system, which we call  $F_{\min}$ . The difference between this value was estimated for each gene ( $\Delta F_{\min}$ ) and those where the 95% credible region did not overlap 0.0 were identified as significant differential vulnerabilities by this approach.

## Pathway analysis

First, all annotated Mtb genes were associated with a pathway as defined by the Kyoto Encyclopedia of Genes and Genomes (KEGG) database<sup>70–72</sup>. If necessary, annotations were manually curated to update or correct pathway assignments. To quantify pathway enrichment, the query set was defined as the union of the upper quartile of differential vulnerabilities defined by both the original gene vulnerability calling method ( $\Delta V$ ) and the  $F_{\min}$  approach. The background set was defined as all annotated Mtb genes. Enrichment of the pathways identified as differentially vulnerable was calculated by an odds ratio and significance was determined with a Fisher's exact test.

## phyOverlap

To detect associations between gene variants and Rif resistance, we employed a phylogenetic convergence test using the phyOverlap algorithm<sup>73</sup> (<https://github.com/Nathan-d-hicks/phyOverlap>). In brief, FASTQ files were aligned to H37Rv genome (NC\_018143.2) using bwa (version 0.7.17-r1188). FASTQ accession numbers are provided in Supplementary Table 3. Single-nucleotide polymorphisms (SNPs) were called and annotated using the HaplotypeCaller tool Genome Analysis Toolkit (version 3.5) using inputs from samtools (version 1.7). SNP sites with less than 10x coverage or missing data in >10% strains were removed from the analysis. Repetitive regions of the genome (PE/PPE genes, transposases, and prophage genes) are excluded from the analysis. Known drug-resistance regions were further excluded so as not to bias phylogenetic tree construction. *M. canetti* was provided as an outgroup (NC\_015848). We performed Maximum Likelihood Inference using RAxML (v8.2.11) to construct the ancestral sequence and determine the derived state of each allele. Overlap with Rif resistance was scored by dividing the number of genotypically predicted (Mykrobe v0.9.012) RifR isolates containing a derived allele by the total number of isolates with a derived allele at a given genomic position. To generate a gene-wide score, we excluded synonymous SNPs and averaged the individual nonsynonymous SNP scores, weighting the scores by the number of times derived alleles evolved across the phylogenetic tree. The significance of the overlap is then tested by redistributing mutation events for each SNP randomly across the tree and recalculating the score. This permutation is done 50,000 times to derive the  $P$  value. This analysis additionally used FastTree (version 2.1.11) and figTree (v1.4.4).

## dN/dS calculations

The ratio of nonsynonymous (dN) to synonymous (dS) nucleotide substitutions was used to quantify selective pressure acting on *nusG* and *rpoC*. A dN/dS value less than one suggests negative or purifying selection whereas a dN/dS value greater than one suggests positive or diversifying selection. For this analysis, we used a collection of ~50,000 Mtb clinical isolate whole-genome sequences, as described<sup>41</sup>. Isolates were grouped based on the presence of genotypically predicted Rif resistance (Mykrobe v0.9.012), as well as the identity of the *rpoB* mutation (S450X or H445X; where X indicates any amino acid other than Ser or His, respectively) conferring RifR. The number of samples used in the *nusG* dN/dS analysis shown in Fig. 3 are as follows: 1,365 RifS, 350 RifR, 270 S450X, and 26 H445X. The number of samples used in the *rpoC* dN/dS analysis shown in Fig. 3 are as follows: 23,024 RifS, 13,993 RifR, 11,067 S450X, and 1,215 H445X. Insertions and deletions were necessarily excluded from this analysis. A bootstrap-analysis was performed to calculate the dN/dS ratios to reduce any potential effects of recent clonal expansion events or convergent evolution of a specific site, like acquired drug-resistance mutations, as performed previously<sup>44</sup>.

The analysis was performed by sub-sampling 80% of total variants in each group. The sub-sampling was repeated 100 times. dN/dS values were calculated for each subset of samples using a python script obtained from the github repository: [https://github.com/MtbEvolution/resR\\_Project/tree/main/dNdS](https://github.com/MtbEvolution/resR_Project/tree/main/dNdS).

### SNP calling and upset plot

SNP information for all Mtb clinical isolate whole-genome sequences were called as follows. FASTQ reads were aligned to the H37Rv genome (NC\_018143.2) and SNPs were called and annotated using Snippy9 (version 3.2-dev) using default parameters (minimum mapping quality of 60 in BWA, samtools base quality threshold of 20, minimum coverage of 10, minimum proportion of reads that differ from reference of 0.9). Mapping quality and coverage was further assessed using QualiMap with the default parameters (version 2.2.2-dev). Samples with a mean coverage < 30, mean mapping quality  $\leq 45$ , or GC content  $\leq 50\%$  or  $\geq 70\%$  were excluded. Drug resistance-conferring SNPs were annotated using Mykrobe (v0.9.012). The resulting SNP and drug-resistance calls were used to generate the values depicted in the upset plot.

### Phylogenetic trees

Phylogenetic trees based on SNP calls described above were built using FastTree (version 2.1.11 SSE3). A list of SNPs in essential genes was concatenated to build phylogenetic trees. Indels, drug resistance-conferring SNPs, and SNPs in repetitive regions of the genome (PE/PPE genes, transposases and prophage genes) were excluded. Tree visualization was performed in iTol (<https://itol.embl.de/>).

### Barcode library production

The barcode library was designed to include over 100,000 random 18-mer sequences cloned into an Giles-integrating backbone (*attP* only, no Integrase) containing a hygromycin resistance cassette with a premature stop codon (pINP472). Oligonucleotides were synthesized as a gBlocks Library by IDT, containing 104,976 fragments.

pINP472 (1.6  $\mu\text{g}$ ) was digested with PciI (NEB R0655) and gel-purified (QIAGEN 28706). The library was PCR amplified using NEBNext High-Fidelity 2X PCR Master Mix (NEB M0541L). One 50- $\mu\text{l}$  reaction was prepared, containing 25  $\mu\text{l}$  of PCR master mix, 0.0125 pmol of the gBlock library, and a final concentration of 0.5  $\mu\text{M}$  of the appropriate forward and reverse primers (Fwd: 5'-TTACGCGTTTCACTGGCCGATTG-3' + Rev: 5'-TTTTGCTGGCCTTTTGCTCAAC-3'). PCR cycling conditions were: 98 °C for 30 s; 15 cycles of 98 °C for 10 s, 68 °C for 10 s, 72 °C for 15 s; 72 °C for 120 s. The PCR amplicon were purified using the QIAGEN MinElute PCR purification kit (QIAGEN 28004). One Gibson assembly reaction (NEB E2621) was prepared with 0.01 pmol  $\mu\text{l}^{-1}$  digested pINP472 backbone, 0.009 pmol  $\mu\text{l}^{-1}$  cleaned PCR amplicon, and master mix, representing a 1:2 molar ratio of vector:insert.

Following incubation at 50 °C for 1 h, 7  $\mu\text{l}$  the Gibson product was dialysed to remove salts and transformed into 100  $\mu\text{l}$  MegaX DH10B T1R Electrocomp Cells (Invitrogen C640003) diluted with 107  $\mu\text{l}$  10% glycerol. For each of three total transformations, 75  $\mu\text{l}$  of the cells:DNA mix was transferred to a 0.1 cm electroporation cuvette (Bio-Rad 1652089) and electroporated at 2,000 V, 200 ohms, 25  $\mu\text{F}$ . Transformations were washed twice with 300  $\mu\text{l}$  provided recovery medium and recovered in a total of 3 ml medium. Cells were allowed to recover at 37 °C with gentle rotation. Recovered cells were plated across three plates of LB agar supplemented with zeocin. After 1 d incubation at 37 °C, transformants were scraped and pooled. One fourth of the pellet (3.2 g dry mass) was used to perform 24 minipreps using a QIA prep Spin Miniprep Kit (Qiagen 27104).

### Transformation of barcode library into Mtb

The barcode library was transformed into Rif<sup>S</sup> and  $\beta\text{S450L}$  Mtb expressing RecT (mycobacteriophage recombinase) similarly to the CRISPRi library (see CRISPRi library transformation), with minor modifications.

In brief, cultures for competent cells were grown in 7H9 supplemented with kanamycin to retain the episomal *recT* encoding plasmid (pRL4). Twenty-millilitre cultures were concentrated ten times and transformed with 250 ng of library and 100 ng of non-replicating, Giles integrase containing plasmid (pRL40). Additionally, after recovery cells were plated on 7H10 agar supplemented with kanamycin and zeocin. Transformants were scrapped after 29 days of outgrowth.

### ssDNA recombineering and validation of strains

Clinical *nusG*, *rpoB* and *rpoC* mutants were introduced into Rif<sup>S</sup> and  $\beta\text{S450L}$  Mtb using oligonucleotide-mediated (ssDNA) recombineering, as described previously<sup>68</sup>. In brief, 70-mer oligonucleotides were designed to correspond to the lagging strand of the replication fork, with the desired mutation in the middle of the sequence. Alterations were chosen to avoid recognition by the mismatch-repair machinery of RecT expression was induced -16 h before transformation by addition of ATc to a final concentration of 0.5  $\mu\text{g ml}^{-1}$ . 400  $\mu\text{l}$  of competent cells were transformed with 5  $\mu\text{g}$  of mutation containing oligonucleotide and 0.1  $\mu\text{g}$  of hygromycin resistance cassette repair oligonucleotide (1:50 ratio of mutant oligonucleotide to repair oligonucleotide) and recovered in 5 ml 7H9 medium.

After 24 h of recovery, 200  $\mu\text{l}$  of cells were plated on 7H10 plates supplemented with hygromycin. After 21 days of outgrowth, 12 colonies per construct were picked into 100  $\mu\text{l}$  7H9 medium supplemented with hygromycin in a 96 well plate (Fischer Scientific 877217). 50  $\mu\text{l}$  of culture were heat-inactivated at 80 °C for 2 h in a sealed microamp 96 well plate (Fischer Scientific 07200684; Applied Biosystems N8010560). Fifty microlitres of heat-inactivated culture was mixed with 50  $\mu\text{l}$  of 25% DMSO and lysed at 98 °C 10 min.

Mutations of interest and unique barcodes were confirmed with PCR amplification and Sanger sequencing. The region of interest was PCR amplified with NEBNext High-Fidelity 2X PCR Master Mix (NEB M0541L) using 0.5  $\mu\text{l}$  of heat-lysed product with the appropriate primers, annealing temperatures and extension times (see Supplementary Table 4). Residual PCR primers were removed with NEB Shrimp Alkaline Phosphatase (rSAP) and exonuclease I (exo) (rSAP- NEB M0371; exo- NEB M0293) per manufacturer's instructions. Amplicons were then submitted for Sanger sequencing. One to three unique independent isolates were generated for all tested mutations.

### Pooled barcode competitive growth assay

Validated mutants were first grown in 1 ml 7H9 with hygromycin and after 3 days, expanded to 5 ml 7H9 with hygromycin. Strains were pooled to contain approximately  $1.2 \times 10^7$  cells for each mutant. The pool was then diluted to a starting  $\text{OD}_{600}$  of 0.01 in 7H9 supplemented with hygromycin. At this point, three 20 ml cultures in vented tissue culture flasks (T-75; Falcon 353136) were expanded to late-log phase and used as input for the competitive growth experiment. Sixteen  $\text{OD}_{600}$  units of cells were collected from flask as the input culture (generation 0). Triplicate cultures were then diluted back to  $\text{OD}_{600} = 0.05$  and grown for -4.5 generations, back-diluted again to  $\text{OD}_{600} = 0.05$  and grown for an additional 4 generations. After this, cultures were collected for a cumulative 8.5 generations of competitive growth.

Genomic DNA extraction and library preparation for next-generation sequencing followed the same protocol as that of the CRISPRi libraries (see above), with minor modifications. In brief, the barcode region was amplified from 100 ng genomic DNA using NEBNext Ultra II Q5 master Mix (NEB M0544L). PCR cycling conditions were: 98 °C for 45 s; 16 cycles of 98 °C for 10 s, 64 °C for 30 s, 65 °C for 20 s; 65 °C for 5 min. Each PCR reaction contained a unique indexed forward primer (0.5  $\mu\text{M}$  final concentration) and a unique indexed reverse primer (0.5  $\mu\text{M}$ ) (see Supplementary Table 4). Additionally, individual PCR amplicons were multiplexed into a 1 nM pool and sequenced on an Illumina sequencer according to the manufacturer's instructions. To increase sequencing diversity, a PhiX spike-in of 20% was added to the pool (PhiX sequencing



# Article

control v3; Illumina FC-110-3001). Samples were run on the Illumina MiSeq Nano platform (paired-read 2 × 150 cycles, 8 × i5 index cycles, and 8 × i7 index cycles).

## WGS and SNP calling for passaging timepoints and ssDNA recombinants

Genomic DNA (gDNA) was extracted as described above. gDNA was diluted and subjected to Illumina whole-genome sequencing by SeqCenter. In brief, Illumina libraries were generated through tagmentation-based and PCR-based Illumina DNA Prep kit and custom IDT 10 bp unique dial indices, generating 320 bp amplicons. Resulting libraries were sequenced on the Illumina NovaSeq 6000 platform (2 × 150 cycles). Demultiplexing quality control, and adapter trimming was performed with bcl-convert (v4.1.5).

Reads were aligned to the Mtb (H37Rv; CP003248.2) reference genome using bwa (v1.3.1) with default parameters. Variant detection was performed by Snippy (v4.6.0)/freebayes (v1.3.1). Resulting vcf files were inspected for compensatory mutations (Supplementary Table 2) in *rpoABC* and/or the presence of the desired mutation.

## Definition of putative compensatory *nusG*, *rpoA*, *rpoB*, *rpoC* variants

Compensatory mutations in *rpoA*, *rpoB* and *rpoC* were taken from published sources and are described in Supplementary Table 2. Inclusion as a putative compensatory mutation in our list required that each reported variant in *rpoA*, *rpoB*, or *rpoC* was found specifically in Rif-resistant strains, defined here as meaning that ≥90% of all strains harbouring the putative compensatory mutation were genotypically predicted (gDST) Rif<sup>R</sup>. The use of the ≥90% gDST Rif<sup>R</sup> cut-off allows for presumptive instances of incorrect gDST calls for strains harbouring rare compensatory variants. The strains used for this analysis are the approximately 50,000 Mtb WGS strain collection described previously<sup>41</sup>.

The rules to define putative compensatory *nusG* mutations are as follows. Each *nusG* variant observed was assessed according to the following three rules and, if it met one of them, was deemed a putative compensatory variant.

- (1) The *nusG* variant was found in ≥80% genotypically predicted (gDST) Rif<sup>R</sup> strains and was present in at least two distinct Mtb (sub)lineages. The use of the ≥80% gDST Rif<sup>R</sup> cut-off allows for presumptive instances of incorrect gDST calls for strains harbouring rare *nusG* variants.
- (2) The *nusG* variant was found in 100% gDST Rif<sup>R</sup> strains but only present in a single Mtb sublineage, but the same or nearby NusG site (±5 amino acids) was also mutated to an alternative amino acid that met the criteria stated in rule 1.
- (3) Residues based on the Mtb NusG–RNAP structure<sup>13</sup> that were predicted to be important for the NusG pro-pausing activity (for example, NusG Trp120).

The rules to define a putative compensatory mutation in the *rpoB* β-protrusion were similar to those described for *nusG*, except that only *rpoB* β-protrusion residues at or near the NusG interface (RpoB Arg392–Thr410) were included in the analysis. Note that two such β-protrusion mutations (Thr400Ala and Gln409Arg) were previously identified as putative compensatory mutation<sup>17,74,75</sup> (Supplementary Table 2).

## Rif<sup>R</sup> *rpoB* allele frequency distribution calculations

To check whether the observed distribution of Rif<sup>R</sup> *rpoB* mutations was different for each of the three groups (all Rif<sup>R</sup> strains in our clinical strain genome database, those harbouring known compensatory mutations in *rpoA* or *rpoC*, or those harbouring compensatory mutations in *nusG* or the β-protrusion), we performed a chi-squared test on the observed Rif<sup>R</sup> *rpoB* mutant frequencies. Specifically, we take the Rif<sup>R</sup> *rpoB* mutant frequencies observed in all Rif<sup>R</sup> samples as representing

an estimate of the base probabilities under the null hypothesis. We then use these base probabilities to calculate the frequency of mutations that would be expected in the other groups, based on the null hypothesis. That is:

For each mutation (*m*):

$$p(m) = \frac{\text{Number of times } m \text{ occurs in Rif}^R \text{ samples}}{\text{Total number of Rif}^R \text{ samples}}$$

For each group (*G*) and mutation (*m*),

$$E[m|G] = p(m) \times \text{total number of samples in } G$$

## Protein expression and purification

**Mtb RNAP.** Mtb RNAP was purified as previously described<sup>66,76</sup>. In brief, plasmid pMP61 (wild-type RNAP) or pMP62 (S450L RNAP) was used to overexpress Mtb core RNAP subunits *rpoA*, *rpoZ*, a linked *rpoBC* and a His<sub>8</sub> tag. pMP61/pMP62 was grown in *E. coli* Rosetta2 cells in LB with 50 μg ml<sup>-1</sup> kanamycin and 34 μg ml<sup>-1</sup> chloramphenicol at 37 °C to an OD<sub>600</sub> of 0.3, transferred to room temperature and left shaking to an approximate OD<sub>600</sub> of 0.6. RNAP expression was induced by adding IPTG to a final concentration of 0.1 mM, grown for 16 h, and collected by centrifugation (8,000g, 15 min at 4 °C). Collected cells were resuspended in 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM PMSF, 1 mM protease inhibitor cocktail, 5% glycerol and lysed by sonication. The lysate was centrifuged (27,000g, 15 min, 4 °C) and polyethyleneimine (PEI, Sigma-Aldrich) added to the supernatant to a final concentration of 0.6% (w/v) and stirred for 10 min to precipitate DNA binding proteins including target RNAP. After centrifugation (11,000g, 15 min, 4 °C), the pellet was resuspended in PEI wash buffer (10 mM Tris-HCl, pH 7.9, 5% v/v glycerol, 0.1 mM EDTA, 5 mM DTT, 300 mM NaCl) to remove non-target proteins. The mixture was centrifuged (11,000g, 15 min, 4 °C), supernatant discarded, then RNAP eluted from the pellet into PEI Elution Buffer (10 mM Tris-HCl, pH 7.9, 5% v/v glycerol, 0.1 mM EDTA, 5 mM DTT, 1 M NaCl). After centrifugation, RNAP was precipitated from the supernatant by adding (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> to a final concentration of 0.35 g l<sup>-1</sup>. The pellet was dissolved in Nickel buffer A (20 mM Tris pH 8.0, 5% glycerol, 1 M NaCl, 10 mM imidazole) and loaded onto a HisTrap FF 5 ml column (GE Healthcare Life Sciences). The column was washed with Nickel buffer A and then RNAP was eluted with Nickel elution buffer (20 mM Tris, pH 8.0, 5% glycerol, 1 M NaCl, 250 mM imidazole). Eluted RNAP was subsequently purified by gel filtration chromatography on a HiLoad Superdex 26/600 200 pg in 10 mM Tris pH 8.0, 5% glycerol, 0.1 mM EDTA, 500 mM NaCl, 5 mM DTT. Eluted samples were aliquoted, flash frozen in liquid nitrogen and stored in -80 °C until usage.

**Mtb σ<sup>A</sup>–RbpA.** Mtb σ<sup>A</sup>–RbpA was purified as previously described<sup>76,77</sup>. The Mtb σ<sup>A</sup> expression vector pAC2 contains the T7 promoter, ten histidine residues, and a precision protease cleavage site upstream of Mtb σ<sup>A</sup>. The Mtb RbpA vector is derived from the pET-20B backbone (Novagen) and contains the T7 promoter upstream of untagged Mtb RbpA. Both plasmids were co-transformed into *E. coli* Rosetta2 cells and selected on medium containing kanamycin (50 μg ml<sup>-1</sup>), chloramphenicol (34 μg ml<sup>-1</sup>) and ampicillin (100 μg ml<sup>-1</sup>). Protein expression was induced at OD<sub>600</sub> of 0.6 by adding IPTG to a final concentration of 0.5 mM and leaving cells to grow at 30 °C for 4 h. Cells were then collected by centrifugation (4,000g, 20 min at 4 °C). Collected cells were resuspended in 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.1 mM PMSF, 1 mM protease inhibitor cocktail, and 1 mM β-mercaptoethanol, then lysed using a continuous-flow French press. The lysate was centrifuged twice (15,000g, 30 min, 4 °C) and the proteins were purified by Ni<sup>2+</sup>-affinity chromatography (HisTrap IMAC HP, GE Healthcare Life Sciences) via elution at 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 500 mM imidazole, and 1 mM β-mercaptoethanol. Following elution,

the complex was dialysed overnight into 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5 mM imidazole, and 1 mM  $\beta$ -mercaptoethanol and the His<sub>10</sub> tag was cleaved with precision protease overnight at a ratio of 1:30 (protease mass:cleavage target mass). The cleaved complex was loaded onto a second Ni<sup>2+</sup>-affinity column and was retrieved from the flow-through. The complex was loaded directly onto a size-exclusion column (SuperDex-200 16/16, GE Healthcare Life Sciences) equilibrated with 50 mM Tris-HCl, pH 8, 500 mM NaCl, and 1 mM DTT. The sample was concentrated to 4 mg ml<sup>-1</sup> by centrifugal filtration and stored at -80 °C until usage.

**Mtb CarD.** Mtb CarD was purified as previously described<sup>66,76</sup>. In brief, Mtb CarD was overexpressed from pET SUMO (Invitrogen) in *E. coli* BL21(DE3) cells (Novagen) and selected on medium containing 50  $\mu$ g ml<sup>-1</sup> kanamycin. Protein expression was induced by adding IPTG to a final concentration of 1 mM when cells reached an apparent OD<sub>600</sub> of 0.6, followed by 4 h of growth at 28 °C, then collected by centrifugation (4,000g, 15 min at 4 °C). Collected cells were resuspended in 20 mM Tris-HCl, pH 8.0, 150 mM potassium glutamate, 5 mM MgCl<sub>2</sub>, 0.1 mM PMSF, 1 mM protease inhibitor cocktail, and 1 mM  $\beta$ -mercaptoethanol, then lysed using a continuous-flow French press. The lysate was centrifuged twice (16,000g, 30 min, 4 °C) and the proteins were purified by Ni<sup>2+</sup>-affinity chromatography (HisTrap IMAC HP, GE Healthcare Life Sciences) via elution at 20 mM Tris-HCl, pH 8.0, 150 mM potassium glutamate, 250 mM imidazole, and 1 mM  $\beta$ -mercaptoethanol. Following elution, the complex was dialysed overnight into 20 mM Tris-HCl, pH 8.0, 150 mM potassium glutamate, 5 mM MgCl<sub>2</sub>, and 1 mM  $\beta$ -mercaptoethanol and the His<sub>10</sub> tag was cleaved with ULP-1 protease (Invitrogen) overnight at a ratio of 1/30 (protease mass/cleavage target mass). The cleaved complex was loaded onto a second Ni<sup>2+</sup>-affinity column and was retrieved from the flow-through. The complex was loaded directly onto a size-exclusion column (SuperDex-200 16/16, GE Healthcare Life Sciences) equilibrated with 20 mM Tris-HCl, pH 8, 150 mM potassium glutamate, 5 mM MgCl<sub>2</sub> and 2.5 mM DTT. The sample was concentrated to 5 mg ml<sup>-1</sup> by centrifugal filtration and stored at -80 °C.

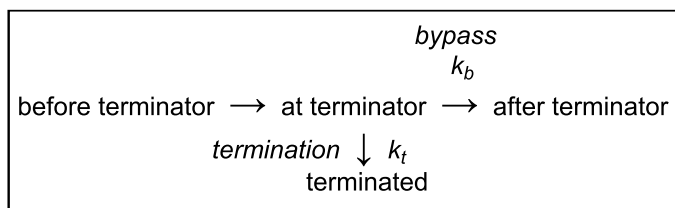
**Wild-type Mtb NusG (+ mutants N65H, R124L and N125S).** Plasmid pAC82 (or mutant variation) was used to overexpress wild-type Mtb NusG<sup>13</sup>. Plasmids encoding NusG mutants were generated using Q5 Site-directed mutagenesis (NEB) and sequenced to confirm the presence of target mutations. *E. coli* BL21 cells containing plasmids encoding different versions of Mtb NusG were grown in LB with 50  $\mu$ g ml<sup>-1</sup> kanamycin at 37 °C to an OD<sub>600</sub> of 0.4, then transferred to room temperature and left shaking to an OD<sub>600</sub> of 0.67. Protein expression was induced by adding IPTG to a final concentration of 0.1 mM, grown for an additional 4 h, then collected by centrifugation (4,000g, 20 min at 4 °C). Collected cells were resuspended in 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5 mM imidazole, 10% glycerol, 1 mM PMSF, 1 mM protease inhibitor cocktail (Roche), 2 mM  $\beta$ -mercaptoethanol, and lysed by French press. The lysate was centrifuged (4,000 rpm for 20 min, 4 °C) and the supernatant was removed and applied to a HisTrap column pre-washed with 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% glycerol, 15 mM imidazole, and 2 mM  $\beta$ -mercaptoethanol. After loading the sample, the column was washed with five volumes of the same buffer, before gradient elution with 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% glycerol, 250 mM imidazole, and 2 mM  $\beta$ -mercaptoethanol. The eluted protein was mixed with precision protease and dialysed overnight at 4 °C in 20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10 mM  $\beta$ -mercaptoethanol to cleave the N-terminal His<sub>10</sub> tag before applying to a HisTrap column to remove the uncleaved protein. The flow-through was collected and glycerol was added to a final concentration of 20% (v/v). Aliquots were flash frozen in liquid nitrogen and stored in -80 °C until use.

**Promoter-based in vitro termination assays.** The DNA sequence for the Mtb H37Rv 5 S rRNA (*rrf* gene) intrinsic terminator was taken

from Mycobrowser (MTB000021), with genomic coordinates of 1,476,999 to 1,477,077 basepairs. The intrinsic terminator was found by predicting its RNA structure using mfold (RNA folding form v2.3) via the UNAFold Web Server. The intrinsic terminator was cloned downstream of a cytidine-less halt cassette in plasmid pAC70<sup>38</sup>, a gift of the R. Landick laboratory, using Q5 site-directed mutagenesis (following manufacturer's protocol - NEB) at an annealing temperature of 59 °C with GC enhancer for the PCR step, with primers 5'-TGGTGTTTTGTATGTTTATATCGACTCAGCCGCTCGGCCATGGACGCTCTCTCGA-3' and 5'-CCGTTACCGGGGTGTTTTTGTATGTTTCGGCGGTCTCTGGATCCTGGCAGTCCCT-3' (synthesized by IDT), to create plasmid pJCl. The 323 base pairs linear DNA fragment used for in vitro transcription assays was PCR amplified using Accuprime Pfx DNA polymerase (Invitrogen) at an annealing temperature of 56.5 °C, with primers 5'-GAATCAAATATTTGTTGTTAACTCTTGACAAAAGTGTAAAAGC-3' and 5'-GTTGCTTCGCAACGTTCAAATCC-3' (synthesized by IDT), following manufacturer instructions, and PCR purified (using the QIAquick PCR Purification Kit (QIAGEN)) to remove protein contents and buffer exchange into 10 mM Tris-HCl pH 8.5.

pJCl contains the *rrf* termination site at approximately +150 bp. This template also contained a C-less cassette (+1 to +26). Core RNAP was incubated for 15 min at 37 °C with  $\sigma$ A/RbpA in transcription buffer (20 mM Tris, 25 mM KCl, 10 mM MgOAc, 1 mM DTT, 5  $\mu$ g ml<sup>-1</sup> BSA) to form holo-RNAP, followed by 10 min incubation with 500 nM CarD at 37 °C. Holo-RNAP (200 nM) was then incubated with template DNA (10 nM) for 15 min at 37 °C. To initiate transcription, the complex was incubated with ATP + GTP (both at 16  $\mu$ M), UTP (2  $\mu$ M), and 0.1  $\mu$ l per reaction [ $\alpha$ -<sup>32</sup>P]UTP for 15 min at 37 °C to form a halted complex at U26. Transcription was restarted by adding a master mix containing NTP mix (A + C + G + U), heparin, and NusG at a final concentration of 150  $\mu$ M (each NTP), 10  $\mu$ g ml<sup>-1</sup> (heparin), and 1  $\mu$ M NusG at 23 °C. The reaction was allowed to proceed for 30 min, followed by a 'chase' reaction in which all 4 nucleotides were added to a final concentration of 500  $\mu$ M each. After 10 min, aliquots were removed and added to a 2 $\times$  Stop buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). Samples were analysed on an 8% denaturing PAGE (19:1 acrylamide: bis acrylamide, 7 M urea, 1X TBE pH= 8.3) for 1.25 h at 400 V, and the gel was exposed on a Storage Phosphor Screen and imaged using a Typhoon PhosphorImager (GE Healthcare).

**Quantification of termination and changes in termination.** Synthesized RNA bands on the gel image were quantified using ImageJ software (NIH). Each lane from below the *rrf* termination site (-150 nt) to above the runoff RNA products (263 nt) was converted to a pseudodensitometer plot using the ImageJ line function and the relative areas of the termination and runoff bands were measured. Termination efficiency (TE) was calculated as the fraction of the termination (term) peak area relative to total of the termination and runoff (term + runoff) peak areas. Fold changes in termination attributable to each NusG ( $\Delta T$ ) were determined as the aggregate of changes in the termination rates  $k_b$  and  $k_t$ , as defined by von Hippel and Yager (equations (1) and (2))<sup>62,63</sup>. Multiple algebraic transforms can yield the aggregate fold changes in termination,  $\Delta T$ , based on the following equations.



$$TE = \frac{k_t}{k_t + k_b} \quad (1)$$

$$TE = [1 + e^{-\Delta\Delta G^\ddagger / -RT}]^{-1}, \quad (2)$$

where  $\Delta\Delta G^\ddagger$  is the difference in activation barriers between termination and bypass, which is most directly related to the energies of RNAP–NusG and internal RNAP interactions that govern termination.

$$\Delta\Delta G^\ddagger = -RT \times \ln((1/TE) - 1) \quad (3)$$

(equation (2) rearranged).

$$\Delta T = e^{(\Delta\Delta G_2^\ddagger - \Delta\Delta G_1^\ddagger)} \quad (4)$$

(fold change in aggregate termination rates for two conditions, 1 and 2).

$$\Delta T = \frac{\left(\frac{1}{TE_2}\right) - 1}{\left(\frac{1}{TE_1}\right) - 1} \quad (5)$$

(alternative calculation derived from equation (1) assuming NusG only affects  $k_b$ ).

Calculating  $\Delta T$  using either the combinations of equations (3) and (4) or using equation (5) gives the same results because the  $\Delta T$  is the same whether conditions differ by aggregate effects on both  $k_b$  and  $k_t$  or an effect on only one of them. We calculate  $\Delta T$  using these approaches rather than the simple difference in energies of activation ( $\Delta\Delta G_2^\ddagger - \Delta\Delta G_1^\ddagger$ ) because it allows a clearer graphical depiction of effects without changing the results. Errors in  $\Delta T$  were calculated using a two-sided, unpaired  $t$ -test with no assumptions on variance.

**Electrophoretic mobility shift assay.** RNAP–NusG complexes were assembled and run on an electrophoretic mobility shift assay to test proper binding of all mutant NusGs. Core RNAP (200 nM) was incubated with the template strand of elongation scaffold DNA<sup>13</sup> (50 nM) for 15 min at room temperature. Next, the complex was incubated with the complementary non-template strand (50 nM) for 15 min at room temperature. Finally, the complex was incubated with 1  $\mu$ M wild-type NusG, N65H NusG, R124L NusG, or N125S NusG for 10 min at room temperature. All complexes were assembled in the following transcription buffer: 20 mM Tris, 25 mM potassium glutamate, 10 mM magnesium acetate, 1 mM DTT, 5  $\mu$ g ml<sup>-1</sup> BSA. Samples were immediately loaded and run on a native PAGE (4.5% acrylamide:bis solution 37.5:1, 4% glycerol, 1 $\times$  TBE) for 1 h at 15 mA. The gel was run at 4 °C. The gel was first stained with GelRed (Biotium) followed by Coomassie blue for visualization of DNA and protein respectively.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

Raw sequencing data are deposited to the NCBI Short Read Archive under project number PRJNA1021243. The H37Rv reference genome

(CP003248.2) was applied for alignments and SNP calling. Manually curated pathway calls were derived from KEGG ([https://www.genome.jp/kegg-bin/show\\_organism?org=mtu](https://www.genome.jp/kegg-bin/show_organism?org=mtu)) and PATRIC databases ([https://www.bv-brc.org/search/?keyword\(tuberculosis\)](https://www.bv-brc.org/search/?keyword(tuberculosis))) databases.

## Code availability

All source code is available at [https://github.com/rock-lab/nusg\\_paper\\_2023](https://github.com/rock-lab/nusg_paper_2023).

64. Woong Park, S. et al. Evaluating the sensitivity of *Mycobacterium tuberculosis* to biotin deprivation using regulated gene expression. *PLoS Pathog.* **7**, e1002264 (2011).
65. Peek, J. et al. Rifamycin congeners kanglemycins are active against rifampicin-resistant bacteria via a distinct mechanism. *Nat. Commun.* **9**, 4147 (2018).
66. Boyaci, H., Chen, J., Jansen, R., Darst, S. A. & Campbell, E. A. Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature* **565**, 382–385 (2019).
67. Wong, A. I. & Rock, J. M. CRISPR Interference (CRISPRi) for targeted gene silencing in *Mycobacteria*. *Methods Mol. Biol.* **2314**, 343–364 (2021).
68. Murphy, K. C., Papavinasundaram, K. & Sasseti, C. M. Mycobacterial recombineering. *Methods Mol. Biol.* **1285**, 177–199 (2015).
69. Larsen, M. H., Biermann, K., Tandberg, S., Hsu, T. & Jacobs, W. R. Genetic manipulation of *Mycobacterium tuberculosis*. *Curr. Protoc. Microbiol.* **10**, A.2 (2007).
70. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
71. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
72. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
73. Hicks, N. D. et al. Clinically prevalent mutations in *Mycobacterium tuberculosis* alter propionate metabolism and mediate multidrug tolerance. *Nat. Microbiol.* **3**, 1032–1042 (2018).
74. Ma, P. et al. Compensatory effects of *M. tuberculosis* *rpoB* mutations outside the rifampicin resistance-determining region. *Emerg. Microbes Infect.* **10**, 743–752 (2021).
75. Loiseau, C. et al. The relative transmission fitness of multidrug-resistant *Mycobacterium tuberculosis* in a drug resistance hotspot. *Nat. Commun.* **14**, 1988 (2023).
76. Hubin, E. A. et al. Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. *eLife* **6**, e22520 (2017).
77. Boyaci, H. et al. Fidaxomicin jams *Mycobacterium tuberculosis* RNA polymerase motions needed for initiation via RbpA contacts. *eLife* **7**, e34823 (2018).

**Acknowledgements** The authors thank members of the Rock laboratory, N. Hicks, Q. Liu, S. Darst and D. Schnappinger for comments on the manuscript and/or helpful discussions; D. Schnappinger for sharing the  $\Delta$ bioA Mtb strain; and the Weill Cornell and Rockefeller University Genomics Core for sequencing. This work was supported by a joint NIH Tuberculosis Research Units Network (TBRU-N) grant (U19AI162584, J.M.R.), an NIH/NIAID New Innovator Award (1DP2AI144850-01, J.M.R.), R01 GM038660 (R.L.), R01 GM114450 (E.A.C.), and the Stavros Niarchos Foundation (SNF) as part of its grant to the SNF Institute for Global Infectious Disease Research at The Rockefeller University (J.M.R. and E.A.C.).

**Author contributions** Conceptualization: K.A.E., M.A.D., M.D., E.A.C. and J.M.R. Investigation: K.A.E., V.M.-G., M.D., N.C.P., S.Q.-G., J.C., E.A.C. and J.M.R. Data analysis: M.A.D., K.A.E., Z.A.A., A.K.R., M.D., R.L., E.A.C. and J.M.R. Writing—original draft: K.A.E., M.D., E.A.C. and J.M.R. Writing—review and editing: K.A.E., M.D., B.B., R.L., E.A.C. and J.M.R. Funding acquisition: J.M.R. and E.A.C. Supervision: J.M.R. and E.A.C. V.M.-G., M.A.D. and Z.A.A. contributed equally to this work.

**Competing interests** The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-024-07206-5>.

**Correspondence and requests for materials** should be addressed to Elizabeth A. Campbell or Jeremy M. Rock.

**Peer review information** Nature thanks Dan Andersson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

## Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the 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
- A statement on 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 statistical parameters 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

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection | Data collection was done with custom scripts dependent on third-party tools. All source code will be made publicly available online(GitHub: [https://github.com/rock-lab/nusg\\_paper\\_2023](https://github.com/rock-lab/nusg_paper_2023)).

Data analysis | GraphPad Prism (version 10.0.2)  
Microsoft Excel(365)  
bwa (version 0.7.17-r1188 or v1.3.1, depending on operating system requirements)  
phyOverlap algorithm (<https://github.com/Nathan-d-hicks/phyOverlap>, Commit 20ae57d)  
HaplotypeCaller tool Genome Analysis Toolkit (version 3.5)  
samtools (version 1.7)  
Mykrobe (version 0.9.012)  
FastTree (version 2.1.11)  
figTree (version 1.4.4)  
Snippy9 (version 3.2-dev or v4.6.0, depending on operating system requirements)  
QualiMap (version 2.2.2-dev)  
FastTree (version 2.1.11 SSE3)  
iTol (<https://itol.embl.de/>)  
freebayes (version 1.3.1)  
Image J software (NIH)  
viridis (version 0.6.3)  
reshape (version 0.8.9)  
Epi (version 2.47.1)



readxl (version 1.4.2)  
 RColorBrewer (version 1.1-3)  
 tidyverse (version 2.0.0)  
 ggplot2 (version 3.4.2)  
 subread-align (version 1.6.0)  
 Python (version 3.11.2)  
 SciPy (version 1.10.1)  
 numpy (version 1.23.5)  
 R (version 4.2.1)  
 Stan (version 2.21.8)  
 Rstan (version 2.21.8)

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Raw sequencing data are deposited to the NCBI Short Read Archive under project number PRJNA1021243. The H37Rv reference genome (CP003248.2) was applied for alignments and SNP calling. Manually curated pathway calls were derived from KEGG ([https://www.genome.jp/kegg-bin/show\\_organism?org=mtu](https://www.genome.jp/kegg-bin/show_organism?org=mtu)) and PATRIC databases ([https://www.bv-brc.org/search/?keyword\(tuberculosis\)](https://www.bv-brc.org/search/?keyword(tuberculosis))) databases.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

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

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size	The passaging and competitive growth data presented in the manuscript were conducted in biological triplicate. The termination efficiency and termination rates presented in Fig 4D,E,F consist of three experimental replicates. Previous work on libraries these size (Bosch et al., 2021 and Li et al., 2022) showed that 3 biological replicates are enough to detect statistically significant differences between conditions.
Data exclusions	None
Replication	We have indicated the number of times experiments were independently preformed in the figure legends and their corresponding methods.
Randomization	Strains were sequenced to determine that no other SNPs were different between the WT and S450L strains. As this was the only relevant variable, one was assigned as control (WT) and one as experimental (S450L) and no randomization was needed.

Blinding was not performed for any experiments as measurements of optical density, spot assays, and in vitro termination quantification do not require researcher-based judgments and therefore blinding was not deemed necessary.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

### Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging