

# Large-scale chemical–genetics yields new *M. tuberculosis* inhibitor classes

Eachan O. Johnson<sup>1,2,3</sup>, Emily LaVerriere<sup>1,8</sup>, Emma Office<sup>1</sup>, Mary Stanley<sup>1,9</sup>, Elisabeth Meyer<sup>1,10</sup>, Tomohiko Kawate<sup>1,2,3</sup>, James E. Gomez<sup>1</sup>, Rebecca E. Audette<sup>4,11</sup>, Nirmalya Bandyopadhyay<sup>1</sup>, Natalia Betancourt<sup>5,12</sup>, Kayla Delano<sup>1</sup>, Israel Da Silva<sup>5</sup>, Joshua Davis<sup>1,13</sup>, Christina Gallo<sup>1,14</sup>, Michelle Gardner<sup>4</sup>, Aaron J. Golas<sup>1</sup>, Kristine M. Guinn<sup>4</sup>, Sofia Kennedy<sup>1</sup>, Rebecca Korn<sup>1</sup>, Jennifer A. McConnell<sup>5</sup>, Caitlin E. Moss<sup>6,15</sup>, Kenan C. Murphy<sup>6</sup>, Raymond M. Nietupski<sup>1</sup>, Kadamba G. Papavinasundaram<sup>6</sup>, Jessica T. Pinkham<sup>4</sup>, Paula A. Pino<sup>5</sup>, Megan K. Proulx<sup>6</sup>, Nadine Ruecker<sup>5</sup>, Naomi Song<sup>5</sup>, Matthew Thompson<sup>1,16</sup>, Carolina Trujillo<sup>5</sup>, Shoko Wakabayashi<sup>4</sup>, Joshua B. Wallach<sup>5</sup>, Christopher Watson<sup>1,17</sup>, Thomas R. Ioerger<sup>7</sup>, Eric S. Lander<sup>1</sup>, Brian K. Hubbard<sup>1</sup>, Michael H. Serrano-Wu<sup>1</sup>, Sabine Ehrts<sup>5</sup>, Michael Fitzgerald<sup>1</sup>, Eric J. Rubin<sup>4</sup>, Christopher M. Sasseti<sup>6</sup>, Dirk Schnappinger<sup>5</sup> & Deborah T. Hung<sup>1,2,3\*</sup>

**New antibiotics are needed to combat rising levels of resistance, with new *Mycobacterium tuberculosis* (Mtb) drugs having the highest priority. However, conventional whole-cell and biochemical antibiotic screens have failed. Here we develop a strategy termed PROSPECT (primary screening of strains to prioritize expanded chemistry and targets), in which we screen compounds against pools of strains depleted of essential bacterial targets. We engineered strains that target 474 essential Mtb genes and screened pools of 100–150 strains against activity-enriched and unbiased compound libraries, probing more than 8.5 million chemical–genetic interactions. Primary screens identified over tenfold more hits than screening wild-type Mtb alone, with chemical–genetic interactions providing immediate, direct target insights. We identified over 40 compounds that target DNA gyrase, the cell wall, tryptophan, folate biosynthesis and RNA polymerase, as well as inhibitors that target EfpA. Chemical optimization yielded EfpA inhibitors with potent wild-type activity, thus demonstrating the ability of PROSPECT to yield inhibitors against targets that would have eluded conventional drug discovery.**

The WHO (World Health Organization)<sup>1</sup> has declared antibiotic resistance one of the greatest threats to health, with tuberculosis in particular causing more than 1.6 million deaths every year<sup>2</sup>. Despite the recent approval of two new drugs<sup>3,4</sup>, tuberculosis drug discovery has not kept pace with increasing levels of drug resistance<sup>5</sup>.

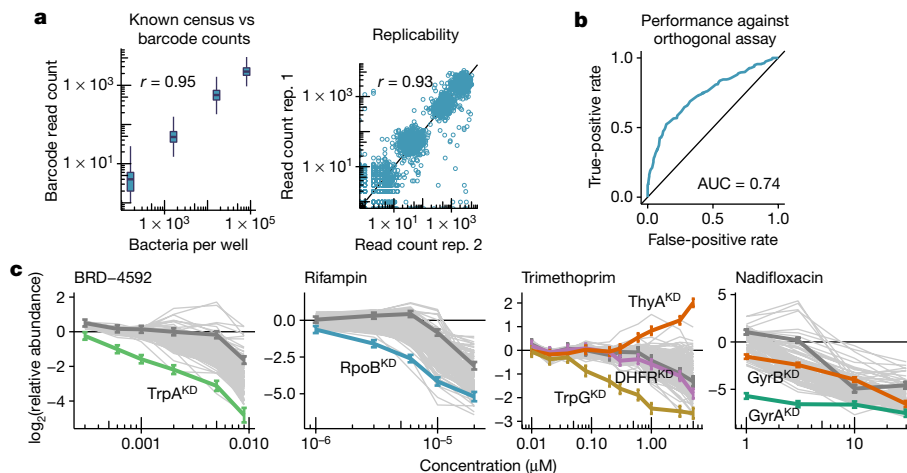
A fundamental challenge in antibiotic discovery is finding new compound classes that kill the causative pathogen, especially by inhibiting essential targets without host toxicity. Primary chemical screening approaches using biochemical, target-based assays have yielded compounds that lack whole-cell activity, and hit compounds derived from conventional whole-cell assays using wild-type bacteria make it difficult to determine the mechanism of action (MOA), thus delaying compound prioritization and progression. The all-too-few successful cases for Mtb, the causative agent of tuberculosis, underscore this challenge, with hit compounds repeatedly targeting two proteins (MmpL3 and DprE1), leaving most of the approximately 625 essential proteins of Mtb unexploited<sup>6–8</sup>.

We aimed to develop an antimicrobial discovery paradigm that simultaneously identifies whole-cell active compounds and predicts their MOA from the primary screening data, thereby incorporating putative target information into hit prioritization. This strategy enables the exploration of broader target space and the discovery of new chemical scaffolds that could not be identified by conventional

whole-cell screening. We performed primary chemical screening with hundreds of mutant strains that were depleted in essential targets (known as hypomorphs), and exploit their hypersensitivity to generate large-scale chemical–genetic interaction profiles (CGIPs). Although such drug hypersensitivity is well established in yeast<sup>9–11</sup>, depletion of essential targets of haploid bacteria is challenging. Consequently, chemical screening of bacterial hypomorphs has been limited to a single hypomorph of *Staphylococcus aureus*<sup>12</sup> and Mtb<sup>13,14</sup>, or to determining the MOA of a single compound against small collections of *S. aureus* hypomorphs<sup>15,16</sup> or a genome-wide non-essential gene-deletion library in *Escherichia coli*<sup>17</sup>. By contrast, generating CGIPs by screening large hypomorph pools (100–150 hypomorphs) against large chemical libraries (50,000 compounds)—an approach we term PROSPECT (primary screening of strains to prioritize expanded chemistry and targets)—markedly increases the detection of active compounds and allows immediate target annotation, informing hit prioritization.

PROSPECT yielded tenfold more hit compounds, each associated with CGIPs, than conventional whole-cell screening approaches. Using primary CGIPs, we rapidly identified and validated more than 40 new scaffolds against established targets including DNA gyrase, RNA polymerase (RNAP), and the biosynthesis of the cell wall, folate and tryptophan. We also discovered an inhibitor of a new target—the essential efflux pump EfpA. Because hits may have limited activity

<sup>1</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>2</sup>Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA, USA. <sup>3</sup>Department of Genetics, Harvard Medical School, Boston, MA, USA. <sup>4</sup>Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, MA, USA. <sup>5</sup>Department of Microbiology and Immunology, Weill Cornell Medical College, New York, NY, USA. <sup>6</sup>Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, MA, USA. <sup>7</sup>Department of Computer Science, Texas A&M University, College Station, TX, USA. <sup>8</sup>Present address: Biological and Biomedical Sciences Program, Harvard Medical School, Boston, MA, USA. <sup>9</sup>Present address: Rush University Medical Center, Chicago, IL, USA. <sup>10</sup>Present address: Department of Biochemistry, Stanford University, Stanford, CA, USA. <sup>11</sup>Present address: Saint Louis University School of Medicine, St Louis, MO, USA. <sup>12</sup>Present address: Regeneron Pharmaceuticals, Tarrytown, NY, USA. <sup>13</sup>Present address: Regulatory and Quality Solutions LLC, Braintree, MA, USA. <sup>14</sup>Present address: Boston University School of Medicine, Boston, MA, USA. <sup>15</sup>Present address: Yale School of Medicine, New Haven, CT, USA. <sup>16</sup>Present address: Caribou Biosciences, Berkeley, CA, USA. <sup>17</sup>Present address: College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, USA. \*e-mail: [dhung@broadinstitute.org](mailto:dhung@broadinstitute.org)



**Fig. 1 | Generating large-scale chemical–genetic interaction profiles from primary screening.** **a**, Defined mixtures of barcoded wild-type Mtb strains were subjected to census enumeration by sequencing-based barcode counting. The method is accurate across several orders of magnitude (left,  $n = 768$  biologically independent samples), and reproducible between replicates (right). **b**, ROC curve showing that primary data were predictive of activity in a confirmatory secondary growth assay ( $n = 4$  biologically independent experiments). We retested 112 compounds predicted to have activity in the primary screen in an orthogonal resazurin-based colorimetric growth assay, which demonstrated the primary assay as predictive of real activity that could

be detected by more conventional growth methods. **c**, CGIPs showed expected hypersensitivity for compounds of known MOA ( $n = 2$  biologically independent experiments). Profiles show the mean fold change (that is,  $\log_2(\text{abundance relative to DMSO negative control})$ ) of each strain at each concentration tested, with wild-type Mtb in dark grey, and mutants of interest highlighted. Error bars of highlighted strains show 95% confidence interval of the maximum likelihood estimate of the mean. Examples shown are the compound–hypomorph pairs of BRD-4592 with TrpA<sup>KD</sup>, rifampin with RpoB<sup>KD</sup>, trimethoprim with TrpG<sup>KD</sup>, DHFR and ThyA, and the fluoroquinolone nadifloxacin with GyrA and GyrB. KD, knockdown.

### Large-scale CGIPs from primary screening

against wild-type bacteria, we demonstrated the ability to optimize the EfpA inhibitor to drug-like potency against wild-type Mtb. PROSPECT thus uncovers molecules against new targets with potent wild-type activity that could not have been found by conventional screening.

We genetically engineered hypomorphs by controlling target levels using conditional proteolysis<sup>18,19</sup> or transcriptional control. For proteolytic control, essential genes were fused to a C-terminal tag, which targets the protein for degradation (Extended Data Fig. 1a). We recognized that some proteins may not be amenable to this strategy because the tag disrupts function or is destabilizing enough to prohibit viability. To control the level of target depletion, we attempted to construct for each gene up to five hypomorphs, with varying degrees of knockdown (Extended Data Fig. 1b, c). At the time of publication, we have constructed 2,014 strains representing the first 474 of approximately 625 essential genes defined in Mtb<sup>20</sup> (Supplementary Data 1), each carrying a 20-nucleotide barcode with conserved flanking regions.

We established a multiplexed assay to measure the abundance of each strain in a pool by amplifying and sequencing strain barcodes<sup>21–24</sup> (Extended Data Fig. 1d, e). By creating strain mixtures with abundances that span three orders of magnitude, we confirmed that sequencing read counts of the barcodes were an accurate and reproducible proxy for abundance (Fig. 1a). Using this assay to determine the fitness of each strain in a pool containing 100–150 strains, including barcoded wild-type strain in 384-well plates over two weeks, we selected a single mutant for each gene that corresponds to the greatest degradation of the essential target while maintaining similar growth to the wild-type strain (Extended Data Fig. 2a) for inclusion in the final screening pool; this balanced the need for reliable strain behaviour and increased the likelihood of finding hypersensitized strain inhibitors. We anticipated that we would not discover inhibitors for some hypomorphs in which the achieved degree of target depletion does not confer strain hypersensitivity, or because their slower growth resulted in lower statistical power. Nevertheless, using our defined screening pool and rifampin as a positive control, we observed excellent assay performance across the dynamic range for all strains ( $Z'$  factors  $> 0.5$ ; Extended Data Fig. 2b).

Finally, we developed a barcode counting (ConCensusMap) and inference (ConCensusGLM) computational pipeline that calculated

the fold change (that is,  $\log_2(\text{abundance relative to control})$ ) of strain abundance after compound exposure compared with dimethylsulfoxide (DMSO) control and the associated  $P$  value. The fold change vector across all strains for a compound is termed the CGIP.

Genetic interactions of Mtb–bioactive compounds

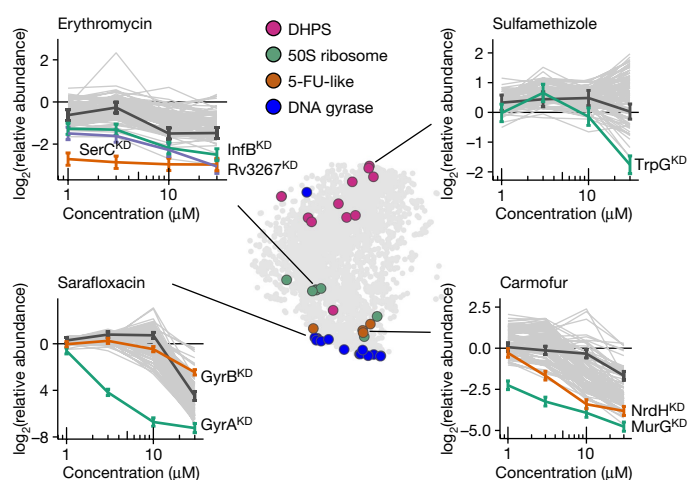
### Genetic interactions of Mtb–bioactive compounds

We assembled a library of 3,226 bioactive small molecules, enriched for compounds with activity against wild-type Mtb based on literature reports (Extended Data Fig. 2c; see Methods), confirmed their activity by testing for inhibition of green fluorescent protein (GFP)-expressing wild-type Mtb, and found that 1,312 (45%) had an MIC<sub>90</sub> value (the minimum inhibitory concentration required to inhibit growth by 90% of its maximal rate) of less than 64  $\mu\text{M}$  (Extended Data Fig. 2d). We then screened this chemical library at 1.1, 3.3, 10 and 30  $\mu\text{M}$  against a pool of the first 100 successfully created Mtb hypomorphs in duplicate (Pearson's  $r = 0.93$ ), generating 1,290,400 potential chemical–genetic interactions.

Most interactions (927,025, 71%) were inhibitory (fold change  $< 1$ ) (Extended Data Fig. 2e); of these, 55,508 interactions (6%) representing 940 compounds (29%) were strong ( $P < 10^{-10}$ ). In a minority, protein depletion conferred resistance to inhibitors of wild-type Mtb; for example, the mycothiol cysteine ligase (MshC) hypomorph was resistant to tuberculosis drugs isoniazid (INH) and ethionamide (ETH), known inhibitors of enoyl-[acyl-carrier-protein] reductase (InhA)<sup>25</sup>.

Using an orthogonal simplex growth assay, we retested 112 identified hits that displayed some specificity (activity against less than 10 strains) for a subset of mutants ( $P < 10^{-10}$ ) against their corresponding hypomorph interactor, wild-type Mtb, and several other hypomorphs as negative controls. Using a receiver operating characteristic (ROC) curve to assess the ability of the multiplexed assay to predict activity in the orthogonal assay, the area under the ROC curve (AUROC; 0.74) indicated a high true-positive rate in the primary assay with a well-controlled false-positive rate (Fig. 1b). Given the complexity of the primary screen, reassuringly 1,375 (52%) of the 2,664 strong interactions were confirmed in the secondary assay.

We recovered interactions between well-characterized inhibitors and hypomorphs of established targets (Fig. 1c), including between the fluoroquinolones and the DNA gyrase  $\alpha$ -subunit (GyrA), rifampin and the RNAP  $\beta$ -subunit (RpoB), and BRD-4592 and the tryptophan



**Fig. 2 | Chemical-genetic interactions of bioactive compounds with known MOA.** Using *t*-distributed stochastic neighbour embedding (*t*-SNE) to visualize the 400-dimensional dataset reveals the MOA-based clustering of compounds. Four exemplar MOAs are illustrated, including the complex mechanism of 5-fluorouracil (5-FU) derivatives<sup>45</sup>. DHPS, dihydropteroate synthase. Grey circles represent compounds, and coloured circles represent a subset of compounds with known MOAs. Representative CGIPs are shown for compounds that represent each of the four exemplar MOAs ( $n = 2$  biologically independent experiments). Profiles show the mean fold change (that is,  $\log_2(\text{abundance relative to DMSO control})$ ) of each strain at each concentration tested, with wild-type *Mtb* in dark grey, and mutants of interest highlighted.

synthase  $\alpha$ -subunit (TrpA)<sup>26</sup>. Notably, trimethoprim (TMP)—a folate biosynthesis inhibitor that targets dihydrofolate reductase (DHFR)—demonstrated a strong interaction with the folate pathway enzyme glutamine amidotransferase (TrpG) rather than DHFR. Although the DHFR hypomorph demonstrated hypersensitization to TMP when tested in isolation (Extended Data Fig. 1f), its slow growth provided low statistical power in primary screening. By contrast, the ThyA hypomorph was resistant to TMP, consistent with the previous observation that DHFR is non-essential on ThyA loss-of-function<sup>27</sup>. Projecting the 400-dimensional CGIPs (100 strains  $\times$  4 concentrations) for each compound onto two dimensions, we found that compounds with the same MOA clustered together, independent of their chemical structures (Fig. 2).

### New inhibitor classes of validated targets

We sought new compound scaffolds that inhibited well-validated, clinical targets based solely on primary screening data by training on a reference ground-truth set of 107 CGIPs of known antimicrobials (Supplementary Data 2). Using Lasso classification models<sup>28</sup>—a supervised machine learning method—we identified 39 new inhibitors of DNA gyrase (training set  $n = 14$ ) and of the biosynthesis of mycolic acid ( $n = 6$ ), folate ( $n = 12$ ) and tryptophan.

### DNA gyrase inhibitors

Training on CGIPs of the fluoroquinolones, our model identified hypersensitivity of the GyrA hypomorph as the most discriminatory feature of gyrase inhibition, which we term a sentinel strain (Extended Data Fig. 3a). The model predicted 55 non-quinolone DNA gyrase inhibitors (Fig. 3a), including the structurally distinct inhibitor novobiocin.

Using a *Mtb* DNA gyrase in vitro assay, we confirmed 27 (52%) out of 52 predicted new DNA gyrase inhibitors (Extended Data Fig. 4a), whereas 25 randomly selected compounds showed no activity (classifier enrichment  $P = 2 \times 10^{-7}$ , Fisher's exact test; AUROC = 0.89). Of the validated compounds, tryptanthrin (Extended Data Fig. 4b) is an anti-infective with a target that has eluded extensive antibacterial and antitrypanosomal research, with current results suggesting that it may

be a gyrase inhibitor<sup>29,30</sup>. Furthermore, the acridine scaffold of ethacridine (Fig. 3c) has been previously reported to inhibit DNA gyrase<sup>31</sup>. All remaining scaffolds were previously undiscovered.

### Inhibitors of mycolic acid biosynthesis

The cornerstone clinical antitubercular prodrugs INH and ETH both inhibit InhA<sup>32</sup>—a key enzyme in the biosynthesis of mycolic acid. We sought new inhibitors of mycolic acid biosynthesis by training on CGIPs of these drugs. Although the strain pool did not include an InhA hypomorph, our model successfully used increased relative fitness of the MshC sentinel strain (Fig. 3c, Extended Data Fig. 3b). MshC catalyses the incorporation of cysteine into mycothiol—an antioxidant unrelated to biosynthesis of mycolic acid—and has been shown to confer resistance to INH and ETH when depleted<sup>25</sup>.

The model predicted six hydrazone derivatives of INH that are inactive against a loss-of-function mutant of the catalase KatG, indicating that they require KatG activation in the same way as the INH prodrug<sup>33</sup>. Notably, the model also predicted one completely new scaffold, the indenedione BRD-9942 (Fig. 3d) that was active against the KatG mutant. Taking advantage of their wild-type activity, we found that in wild-type *Mtb*, like INH and ETH, the INH hydrazones and BRD-9942 inhibited the incorporation of <sup>14</sup>C-acetate into mycolic acids (Extended Data Fig. 5).

### Inhibitors of folate and tryptophan biosynthesis

Folate biosynthesis has been an effective antimicrobial target, although not for *Mtb*. We sought new folate inhibitor classes by training on the CGIPs of the sulfonamides, which target dihydropteroate synthase. The most discriminatory feature was inhibition of the TrpG sentinel strain (Extended Data Fig. 3c). TrpG supports the biosynthesis of both folate and tryptophan (Extended Data Fig. 6a), catalysing the formation of the folate precursor 4-amino-4-deoxychorismate and the tryptophan precursor 2-amino-2-deoxyisochorismate.

We tested whether supplementation with tryptophan, folate or the folate pathway intermediate *para*-amino benzoic acid (PABA) suppressed the activity of 7 of the 43 predicted compounds that had wild-type *Mtb* activity and spanned several chemotypes (Fig. 3e, f). The effects of the nitrothiophene compounds BRD-2550, BRD-3387, BRD-5592 and BRD-9737 (Fig. 3f, Extended Data Fig. 6d), and a *para*-aminosalicylic acid (PAS) derivative BRD-9819, were abolished by supplementation with both PABA and folate; this pattern is the same for the known DHFR inhibitors methotrexate and PAS (Extended Data Fig. 6b). By contrast, BRD-8884 had effects that could only be rescued by folate and not by PABA (Extended Data Fig. 6c), suggesting a new folate target. Thus, PROSPECT uncovered inhibitors that target several different steps in the folate pathway.

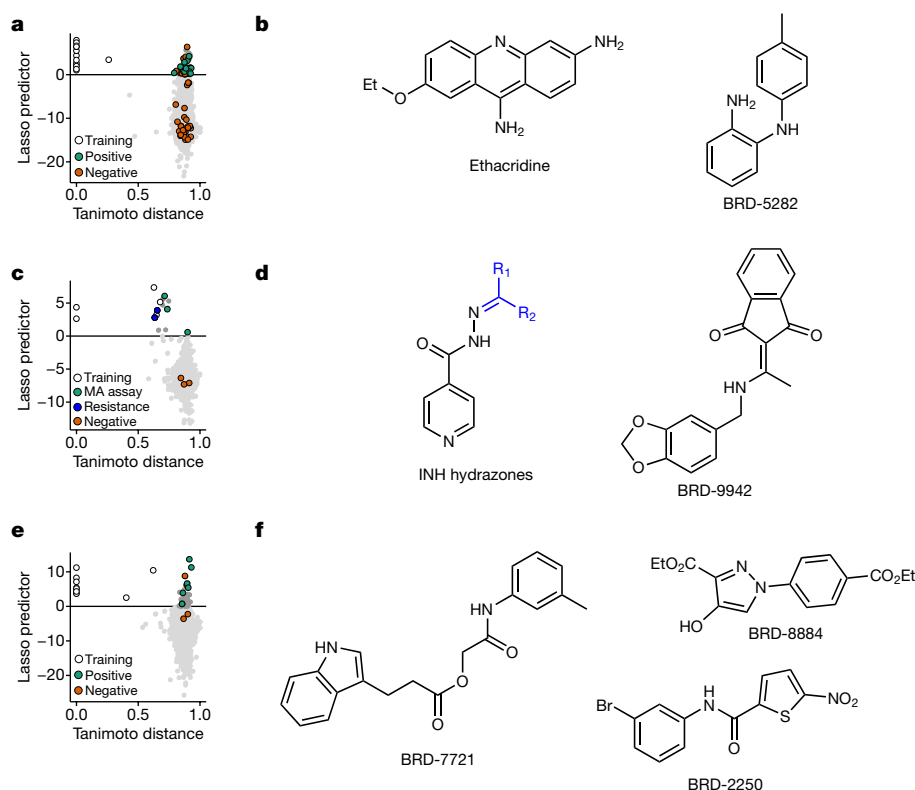
Finally, the effects of BRD-7721—a 3-indolepropionic acid (3-IPA) ester—were only abrogated by supplementation with tryptophan (Extended Data Fig. 6c), suggesting that tryptophan biosynthesis is its target. 3-Indolepropionic acid was recently identified as having antimycobacterial activity in a fragment-based screen<sup>34,35</sup> and its activity was also abolished by supplementation with tryptophan (Extended Data Fig. 6b).

Empirical validation of new folate and tryptophan inhibitors allowed retraining of updated models for each pathway. The refined classifier learned to weigh FolB hypomorph behaviour in conjunction with TrpG in its discrimination of folate inhibitors (Extended Data Fig. 3d, e).

### Screening a larger, unbiased compound library

We applied the PROSPECT approach to a large, unbiased library of approximately 50,000 compounds at 50  $\mu$ M and an expanded pool of 152 strains available at the initiation of this screen, to demonstrate the scalability of the model and its ability to identify compounds with potential activity against wild-type *Mtb* that could be revealed by hypersensitive hypomorphs and to discover inhibitors of new targets, which, after chemical optimization, could have potent wild-type activity.





**Fig. 3 | New inhibitor classes of well-validated targets in *M. tuberculosis*.**

**a**, Lasso predictor, derived from a DNA gyrase inhibitor classifier applied to the mean fold change ( $n = 2$ ) values from the bioactive screen, plotted against Tanimoto distance (chemical structure dissimilarity) to the fluoroquinolones, known DNA gyrase inhibitors. Each point is a compound in the screen; those above the horizontal line were predicted DNA gyrase inhibitors. White points were known fluoroquinolone DNA gyrase inhibitors in the training set; compounds confirmed in an in vitro assay for DNA gyrase inhibitory activity are shown in green; compounds with no in vitro activity are shown in orange. **b**, Examples of new DNA gyrase inhibitor chemotypes predicted by the Lasso classifier and confirmed in vitro. **c**, As in **a**, but for the mycolic acid biosynthesis inhibitor classifier applied to the mean fold change ( $n = 2$ ) values from the bioactive screen. White points were known inhibitors of mycolic acid

Of the 7,245,009 potential chemical–genetic interactions tested, 95,685 (1.3%) were strongly inhibitory. By selecting 1,331 compounds for retesting, we confirmed 78% of the inhibitory chemical–genetic interactions (fold change  $< 1$ ), resulting in an AUROC of 0.74, similar to that observed for the bioactive screen. In an orthogonal assay, the AUROC was 0.69 using 75% inhibition as the ground-truth assumption of true activity (Extended Data Fig. 7a).

The hit rate against wild-type Mtb alone was 0.9% (436 compounds), which is typical for an unbiased whole-cell compound screen; by contrast, tenfold more compounds (4,403; 9%) were active against at least one of the 152 hypomorph strains (Extended Data Fig. 7b). In addition, 92 of the 152 hypomorphs had the lowest abundance relative to DMSO control for at least three compounds, suggesting latent diversity in the MOA that is inaccessible by conventional screening. Of the 4,403 active compounds, 3,967 had no activity against wild-type Mtb. However, 73% were highly specific against the hypomorphs (1–10 strains hit) and 16% were relatively non-specific (more than 50 strains hit); these latter compounds will be retested at lower concentrations to delineate their MOAs because wild-type activity is advantageous as a chemical optimization starting point); compounds showed greater strain specificity than the compounds in the bioactive library (35% and 34%), probably because the smaller library is enriched for compounds with wild-type Mtb activity (Extended Data Fig. 7c). The larger library yielded a greater diversity of CGIPs, suggesting greater target diversity. Clustering of the CGIPs produced 1,864 distinct clusters in the unbiased library

compared to 235 in the bioactive library<sup>36</sup>; unbiased library clusters were meaningful as more than 10% of them were enriched for structurally similar compounds (Extended Data Fig. 7d, e). We applied the folate biosynthesis inhibitor classifier derived from training on the bioactive library to data from the unbiased library. Despite being trained on a dataset generated from several concentrations of each compound and a smaller hypomorph set, the folate model showed excellent transferability, predicting 60 compounds from this larger screen (Fig. 4a), including 12 sulfonamides, 1 PAS derivative, and 2 nitrothiophenes that are now shown to inhibit the folate pathway (Fig. 3f). Three novel scaffolds (Fig. 4b) had activity that could be suppressed by supplementation with PABA or folate (Extended Data Fig. 7f). These results demonstrate the scalability and generalizability of the PROSPECT model, and suggest that revealing latent target diversity could allow the discovery of inhibitors of new targets.

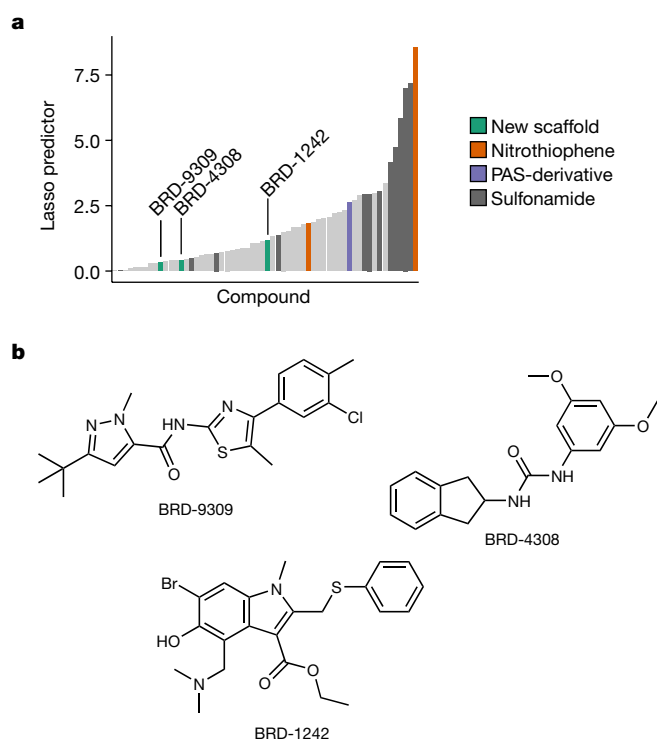
compared to 235 in the bioactive library<sup>36</sup>; unbiased library clusters were meaningful as more than 10% of them were enriched for structurally similar compounds (Extended Data Fig. 7d, e).

We applied the folate biosynthesis inhibitor classifier derived from training on the bioactive library to data from the unbiased library. Despite being trained on a dataset generated from several concentrations of each compound and a smaller hypomorph set, the folate model showed excellent transferability, predicting 60 compounds from this larger screen (Fig. 4a), including 12 sulfonamides, 1 PAS derivative, and 2 nitrothiophenes that are now shown to inhibit the folate pathway (Fig. 3f). Three novel scaffolds (Fig. 4b) had activity that could be suppressed by supplementation with PABA or folate (Extended Data Fig. 7f). These results demonstrate the scalability and generalizability of the PROSPECT model, and suggest that revealing latent target diversity could allow the discovery of inhibitors of new targets.

### Inhibitor discovery without reference data

Having rapidly identified new chemotypes against established targets using reference data, we turned to the discovery of inhibitors with completely new MOAs in the absence of reference data. We developed an approach to identify inhibitors of a target of interest from specific inhibition of the corresponding hypomorph.

Initially, we sought inhibitors of RNAP—a target of the chemotherapy-anchoring rifamycins—for which there is increasing resistance.



**Fig. 4 | Predicting inhibitors of folate biosynthesis from a 50,000-compound screen.** **a**, Lasso predictor scores, derived from a folate biosynthesis inhibitor classifier applied to the mean fold change ( $n = 2$ ) values for the 50,000-compound screen, of the 60 compounds predicted to be inhibitors of folate biosynthesis. The highest-scoring compounds were known folate inhibitors (sulfonamides and nitrothiophene compounds), thus validating the approach. New scaffolds were also identified. **b**, Three new folate biosynthesis inhibitor chemotypes predicted by the Lasso classifier and confirmed by metabolite supplementation.

Because concentrations of rifamycin in our screen inhibited all growth, their CGIPs from the primary screen were uninformative.

Taking an alternative approach, we prioritized 20 compounds that showed strong chemical–genetic interactions with the RpoB hypomorph ( $P < 10^{-10}$ ), which corresponds to RNAP, and requiring it to be among the two most inhibited strains for at least one compound concentration. Testing these compounds in an in vitro RNAP

assay<sup>37,38</sup>, three compounds—including the antineoplastic human RNAP inhibitor actinomycin D—showed direct inhibition of the *E. coli* RNAP (Fig. 5a, Extended Data Fig. 8). Although the positive predictive value was lower than that using machine learning, PROSPECT readily identified new scaffolds against this important target, which has proved recalcitrant to whole-cell inhibitor discovery<sup>39</sup>.

### Discovering inhibitors of a new target

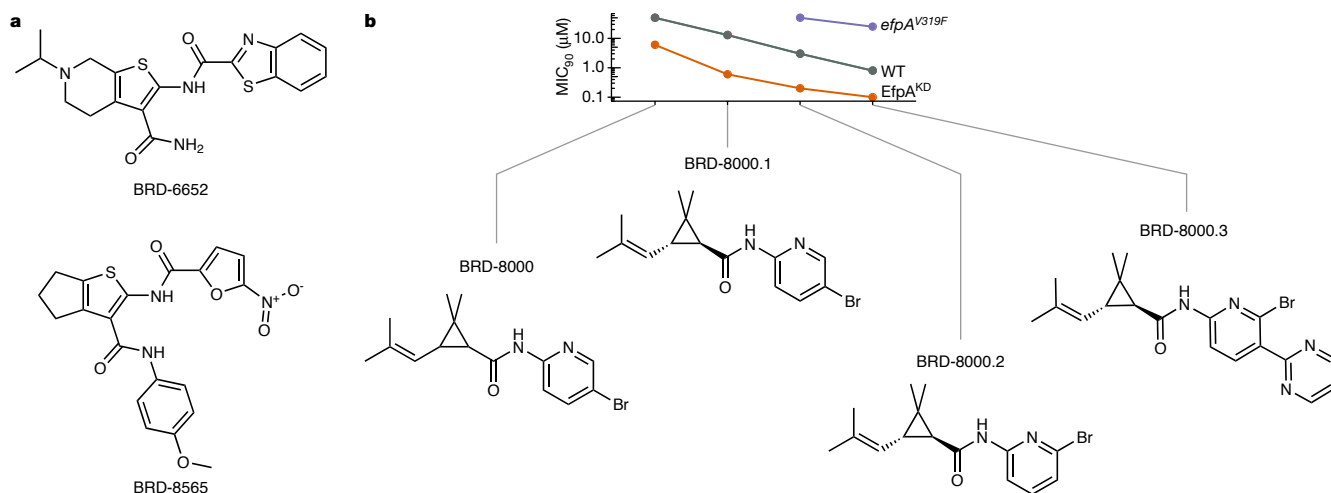
Finally, we demonstrated that PROSPECT can identify inhibitors of a new target that is inaccessible to conventional strategies. Specifically, we identified a compound that inhibits the target EfpA, and optimized it to achieve potent wild-type activity.

We prioritized compounds that did not strongly inhibit wild-type Mtb, but were strongly active against at least one hypomorph at all screening concentrations. Chemically attractive scaffolds were ranked by how few hypomorphs were significantly inhibited, which was used as a proxy for specificity. The highest-ranked interaction was between BRD-8000 and the hypomorph of EfpA—an uncharacterized essential efflux pump. A broth microdilution assay confirmed that BRD-8000 was active against the EfpA hypomorph with little wild-type activity (Extended Data Fig. 9a).

We optimized BRD-8000, first resolving the stereoisomer mixture (see Supplementary Note 7) to yield the pure active (*S,S*-*trans* stereoisomer (BRD-8000.1; Fig. 5b). Migration of the pyridyl bromine from the 6 to the 5 position furnished BRD-8000.2, maintaining hypomorph hypersensitivity while improving wild-type potency. Further chemical optimization yielded BRD-8000.3, a methyl-pyrazole derivative of BRD-8000.1 with an at least 60-fold overall improvement in activity from the original hit ( $MIC_{90} = 800$  nM).

BRD-8000.3 showed good mouse plasma exposure by oral dosing, and low risk for drug–drug interactions, prolongation of the QT interval, or hepatotoxicity based on minimal inhibition of cytochrome P450 enzyme activity, hERG channel binding, and cellular viability of HepG2 cells (Extended Data Fig. 9c). Thus, BRD-8000.3 represents a good lead optimization candidate, because liabilities that typically require extensive optimization are currently absent.

The BRD-8000 series is bactericidal (Extended Data Fig. 9b), kills non-replicating phenotypically drug-tolerant Mtb<sup>40</sup> (BRD-8000.3  $MBC_{50} < 390$  nM), and is narrow-spectrum, inactive against Gram-positive *S. aureus*, Gram-negative *Klebsiella pneumoniae*, and wild-type or efflux-deficient strains of *E. coli* and *Pseudomonas aeruginosa*; it has activity against *M. marinum* (Mmar) but not *M. smegmatis* (Msm) (Extended Data Table 1).



**Fig. 5 | Discovery of inhibitors in the absence of reference data, including inhibitors of a novel target in *M. tuberculosis*.** **a**, New RNAP inhibitor chemotypes validated in an in vitro assay. **b**, Optimization of wild-type activity of the BRD-8000 series, discovered by hypersensitivity

of the EfpA hypomorph (EfpA<sup>KD</sup>). EfpA was validated as the target, partly through evolution of *efpA* mutants (*efpA*<sup>V319F</sup>) displaying high-level resistance.

**Table 1 | Best-fit kinetic parameters for inhibition of efflux of fluorescent dyes in *M. smegmatis***

Msm strain	Complement	Inhibitor	Dye	$K_i$ ( $\mu$ M)	$K_i'$ ( $\mu$ M)	$K_i/K_i'$
WT	—	BRD-8000.2	BisH	10,000	31.3	300
		BRD-8000.2	EtBr	53	15.9	3.3
		BRD-8000.3	EtBr	300	6.33	47
		BisH	EtBr	0.09	11	0.008
$\Delta$ <i>efpA</i>	—	BRD-8000.3	EtBr	12.0	1,000	0.01
		<i>efpA</i> <sup>Mtb</sup>		19.9	8.8	2.3
		<i>efpA</i> <sup>Mtb</sup> <sub>V319F</sub>		7.09	22.6	0.31

BisH, bisbenzimidazole-H; EtBr, ethidium bromide; WT, wild type.  $K_i$  and  $K_i'$  are the aggregate in vivo parameters for inhibitor affinity to substrate-free and substrate-bound efflux pumps, respectively.

We explored two alternative explanations for the sensitivity of the EfpA hypomorph to BRD-8000: (i) the compound directly targets and inhibits EfpA, which is essential in Mtb, and (ii) the compound has a different target, but is pumped out of cells by EfpA, affording higher intracellular concentrations in the EfpA hypomorph. Several lines of evidence strongly support the first hypothesis.

First, we showed that BRD-8000 inhibits efflux by EfpA uncompetitively rather than competitively—the latter being the expectation if it were a substrate being effluxed by EfpA. Specifically, we studied efflux of the known EfpA substrate ethidium bromide<sup>41</sup> (EtBr) in Msm, in which the pump is non-essential and thus impervious to the bactericidal effects of BRD-8000<sup>41</sup> (Extended Data Fig. 9e, f). Although another known EfpA substrate, bisbenzimidazole-H, showed competitive inhibition of EtBr efflux in wild-type Msm, the analogues BRD-8000.2 and BRD-8000.3 showed uncompetitive inhibition (free enzyme inhibitor constant ( $K_i$ )/enzyme-substrate inhibitor constant ( $K_i'$ )  $\geq 3.3$ ), with the trend in potency matching their corresponding MIC<sub>90</sub> values in Mtb (BRD-8000.2  $K_i' = 15.9$ , BRD-8000.3  $K_i' = 6.33$ ; Table 1). In an *efpA* deletion strain of Msm (Msm $\Delta$ *efpA*), BRD-8000.2 shows competitive inhibition (Table 1), presumably of the low-level efflux of other, background efflux pumps.

Second, genetic evidence showed that Mtb mutants resistant to the compound all have the same amino acid substitution in EfpA. Specifically, more than 30 independent resistant Mtb mutants raised against BRD-8000.2 (resistance frequency of approximately  $10^{-8}$ ) all contained a C955A mutation in *efpA*, resulting in a Val319Phe substitution and conferring a more than 30-fold increase in MIC<sub>90</sub> (Fig. 5b). Consistent with this mutation disrupting the compound binding site, expression of the resistant Mtb *efpA*<sup>V319F</sup> allele in Msm $\Delta$ *efpA* abolished the uncompetitive inhibition by BRD-8000.2 of EtBr efflux (Table 1).

Third, we excluded the possibility that the resistance conferred by the EfpA(V319F) mutant is due to hyperactivation of the EfpA efflux pump. There was no difference in EtBr efflux in Mtb between strains expressing wild-type EfpA and mutant EfpA(V319F), but, as expected, there was increased efflux in a Mtb strain overexpressing wild-type EfpA (Extended Data Fig. 9g). We also measured the sensitivity of Mtb to INH, because INH resistance is known to be mediated by EfpA efflux<sup>42,43</sup>; we found no difference between Mtb strains expressing wild-type EfpA and mutant EfpA(V319F) (Extended Data Fig. 9h). Finally, intracellular accumulation of the fluorescent analogue BRD-8000.3 was no higher in the Msm $\Delta$ *efpA* strain expressing EfpA(V319F) versus wild-type EfpA (Extended Data Fig. 9i).

Having confirmed that the BRD-8000 series inhibits the target EfpA, we returned to the primary screening data to identify additional EfpA inhibitors based on CGIP similarity. We tested 11 prioritized compounds for their ability to inhibit EtBr efflux and identified 3 new scaffolds, encompassing 6 molecules that inhibited EtBr efflux (Extended Data Fig. 9j–l). Although not as specific for the EfpA efflux pump as BRD-8000, the whole-cell activity of these compounds is probably due to their activity on EfpA, because EfpA is the only essential efflux pump

in Mtb<sup>20</sup>. Thus, the PROSPECT method can be applied iteratively to expand the chemical diversity of small-molecule candidates against a novel target quickly, even ones with unknown function.

## Discussion

By enabling the discovery of many compounds against previously unexploited targets for Mtb drug development, PROSPECT is a powerful and rapid CGIP strategy. Immediate integration of potential target information empowers hit selection and prioritization by moving beyond simply compound potency. Interpreting multi-dimensional data guided either by the CGIPs of known compounds or by hypomorph specificity, we identified MOA for 45 new molecules. Importantly, these hits include new scaffolds against known targets—a valuable strategy to overcome antimicrobial resistance<sup>44</sup>—and the first inhibitors against a new target, exploiting the enormous underexplored target space in Mtb. Identifying approximately tenfold more hits than by screening wild-type Mtb alone, PROSPECT greatly expands the chemical and target space of identified compounds with the potential to reveal synergistic genetic interactions that inform the development of Mtb combination therapy.

With the discovery and optimization of an EfpA inhibitor with potent wild-type activity, we demonstrated that PROSPECT uncovers inhibitors of targets that would not have resulted from conventional approaches. Furthermore, iterative mining of the entire CGIP dataset enables the discovery of additional inhibitors of the same target. Although these large datasets have only been superficially analysed so far, we have provided examples of how integrating insight about the MOA into primary whole-cell screening can transform the targets and molecules that emerge and are prioritized, even without a clear understanding of a target's function. To accelerate the community's discovery of new inhibitor classes and their respective targets, we have made primary data publicly available (<https://broad.io/cgtb>). Finally, we propose that PROSPECT is widely applicable to important pathogens beyond Mtb.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-019-1315-z>.

Received: 14 December 2018; Accepted: 21 May 2019;

Published online: 19 June 2019

- World Health Organization. *Antimicrobial Resistance: Global Report on Surveillance 2014*. (WHO, 2014).
- World Health Organization. *Global Tuberculosis Report 2018*. (WHO, 2018).
- Andries, K. et al. A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* **307**, 223–227 (2005).
- Matsumoto, M. et al. OPC-67683, a nitro-dihydro-imidazooxazole derivative with promising action against tuberculosis in vitro and in mice. *PLoS Med.* **3**, e466 (2006).
- Dheda, K. et al. The epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable tuberculosis. *Lancet Respir. Med.* **5**, 291–360 (2017).
- Christophe, T. et al. High content screening identifies decaprenyl-phosphoribose 2' epimerase as a target for intracellular antimycobacterial inhibitors. *PLoS Pathog.* **5**, e1000645 (2009).
- Grzegorzewicz, A. E. et al. Inhibition of mycolic acid transport across the *Mycobacterium tuberculosis* plasma membrane. *Nat. Chem. Biol.* **8**, 334–341 (2012).
- Stanley, S. A. et al. Diarylcoumarins inhibit mycolic acid biosynthesis and kill *Mycobacterium tuberculosis* by targeting FadD32. *Proc. Natl Acad. Sci. USA* **110**, 11565–11570 (2013).
- Giaever, G. et al. Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat. Genet.* **21**, 278–283 (1999).
- Nelson, J. et al. MOSAIC: a chemical-genetic interaction data repository and web resource for exploring chemical modes of action. *Bioinformatics* **34**, 1251–1252 (2018).
- Xu, D. et al. Genome-wide fitness test and mechanism-of-action studies of inhibitory compounds in *Candida albicans*. *PLoS Pathog.* **3**, e92 (2007).
- Wang, J. et al. Platensimycin is a selective FabF inhibitor with potent antibiotic properties. *Nature* **441**, 358–361 (2006).
- Evans, J. C. et al. Validation of CoaBC as a bactericidal target in the coenzyme A pathway of *Mycobacterium tuberculosis*. *ACS Infect. Dis.* **2**, 958–968 (2016).

14. Abrahams, G. L. et al. Pathway-selective sensitization of *Mycobacterium tuberculosis* for target-based whole-cell screening. *Chem. Biol.* **19**, 844–854 (2012).
15. Donald, R. G. et al. A *Staphylococcus aureus* fitness test platform for mechanism-based profiling of antibacterial compounds. *Chem. Biol.* **16**, 826–836 (2009).
16. Huber, J. et al. Chemical genetic identification of peptidoglycan inhibitors potentiating carbapenem activity against methicillin-resistant *Staphylococcus aureus*. *Chem. Biol.* **16**, 837–848 (2009).
17. Typas, A. et al. High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nat. Methods* **5**, 781–787 (2008).
18. Kim, J. H. et al. Protein inactivation in mycobacteria by controlled proteolysis and its application to deplete the beta subunit of RNA polymerase. *Nucleic Acids Res.* **39**, 2210–2220 (2011).
19. Kim, J. H. et al. A genetic strategy to identify targets for the development of drugs that prevent bacterial persistence. *Proc. Natl Acad. Sci. USA* **110**, 19095–19100 (2013).
20. DeJesus, M. A. et al. Comprehensive essentiality analysis of the *Mycobacterium tuberculosis* genome via saturating transposon mutagenesis. *mBio* **8**, e02133-16 (2017).
21. Yu, C. et al. High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. *Nat. Biotechnol.* **34**, 419–423 (2016).
22. Han, T. X., Xu, X.-Y., Zhang, M.-J., Peng, X. & Du, L.-L. Global fitness profiling of fission yeast deletion strains by barcode sequencing. *Genome Biol.* **11**, R60 (2010).
23. Smith, A. M. et al. Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic Acids Res.* **38**, e142 (2010).
24. Smith, A. M. et al. Quantitative phenotyping via deep barcode sequencing. *Genome Res.* **19**, 1836–1842 (2009).
25. Vilchèze, C. et al. Coresistance to isoniazid and ethionamide maps to mycothiol biosynthetic genes in *Mycobacterium bovis*. *Antimicrob. Agents Chemother.* **55**, 4422–4423 (2011).
26. Wellington, S. et al. A small-molecule allosteric inhibitor of *Mycobacterium tuberculosis* tryptophan synthase. *Nat. Chem. Biol.* **13**, 943–950 (2017).
27. Moradigaravand, D. et al. *dhfA thyA* double deletion in *para*-aminosalicylic acid-resistant *Mycobacterium tuberculosis* Beijing strains. *Antimicrob. Agents Chemother.* **60**, 3864–3867 (2016).
28. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. B* **58**, 267 (1994).
29. Scovill, J., Blank, E., Konnick, M., Nenortas, E. & Shapiro, T. Antitrypanosomal activities of tryptanthrins. *Antimicrob. Agents Chemother.* **46**, 882–883 (2002).
30. Hwang, J. M. et al. Design, synthesis, and structure–activity relationship studies of tryptanthrins as antitubercular agents. *J. Nat. Prod.* **76**, 354–367 (2013).
31. Medapi, B., Meda, N., Kulkarni, P., Yogeewari, P. & Sriram, D. Development of acridine derivatives as selective *Mycobacterium tuberculosis* DNA gyrase inhibitors. *Bioorg. Med. Chem.* **24**, 877–885 (2016).
32. Banerjee, A. et al. *inhA*, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. *Science* **263**, 227–230 (1994).
33. Yu, S., Giroto, S., Lee, C. & Magliozzo, R. S. Reduced affinity for isoniazid in the S315T mutant of *Mycobacterium tuberculosis* KatG is a key factor in antibiotic resistance. *J. Biol. Chem.* **278**, 14769–14775 (2003).
34. Negatu, D. A. et al. Whole-cell screen of fragment library identifies gut microbiota metabolite indole propionic acid as antitubercular. *Antimicrob. Agents Chemother.* **62**, e01571-17 (2018).
35. Negatu, D. A. et al. Gut microbiota metabolite indole propionic acid targets tryptophan biosynthesis in *Mycobacterium tuberculosis*. *mBio* **10**, e02781-18 (2019).
36. Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *J. Roy. Stat. Soc. B* **63**, 411–423 (2001).
37. Kool, E. T. Circular DNA vectors for synthesis of RNA and DNA. US patent WO1998038300A1 (2000).
38. Daubendiek, S. L. & Kool, E. T. Generation of catalytic RNAs by rolling transcription of synthetic DNA nanocircles. *Nat. Biotechnol.* **15**, 273–277 (1997).
39. Silver, L. L. Challenges of antibacterial discovery. *Clin. Microbiol. Rev.* **24**, 71–109 (2011).
40. Grant, S. S. et al. Identification of novel inhibitors of nonreplicating *Mycobacterium tuberculosis* using a carbon starvation model. *ACS Chem. Biol.* **8**, 2224–2234 (2013).
41. Li, X. Z., Zhang, L. & Nikaido, H. Efflux pump-mediated intrinsic drug resistance in *Mycobacterium smegmatis*. *Antimicrob. Agents Chemother.* **48**, 2415–2423 (2004).
42. Li, G. et al. Efflux pump gene expression in multidrug-resistant *Mycobacterium tuberculosis* clinical isolates. *PLoS ONE* **10**, e0119013 (2015).
43. Rodrigues, L., Machado, D., Couto, I., Amaral, L. & Viveiros, M. Contribution of efflux activity to isoniazid resistance in the *Mycobacterium tuberculosis* complex. *Infect. Genet. Evol.* **12**, 695–700 (2012).
44. Silver, L. L. in *Antibacterials* Vol. 1 (eds Fisher, J. F. et al.) Ch. 24, 31–67 (Springer, 2017).
45. Singh, V. et al. The complex mechanism of antimycobacterial action of 5-fluorouracil. *Chem. Biol.* **22**, 63–75 (2015).

**Acknowledgements** Funding was provided by the Bill and Melinda Gates Foundation, Broad Institute TB Gift Donors and Pershing Square Foundation.

**Reviewer information** Nature thanks Terry Roemer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** The manuscript was written by E.O.J., E.S.L. and D.T.H. Statistical analysis was carried out by E.O.J. Computational pipelines were written by E.O.J. and N. Bandyopadhyay. Experiments were designed as follows. Strain construction: K.M.G., K.C.M., T.R.I., S.E., E.J.R., C.M.S., D.S. Assay development and screening: E.O.J., J.E.G., M.F., D.T.H. Medicinal chemistry: T.K., B.K.H., M.H.S.-W., D.T.H. Mechanism of action follow-up: E.O.J., J.E.G., D.T.H. Experiments were carried out as follows. Strain construction: R.E.A., N. Betancourt, I.D.S., M.G., J.A.M., C.E.M., K.G.P., J.T.P., P.A.P., M.K.P., N.R., N.S., C.T., S.W., J.B.W. Assay development: E.O.J., M.T. Compound screening: E.O.J., E.L., M.S., K.D., J.D., C.G., A.J.G., R.K., R.M.N., M.T., C.W. Mechanism of action follow-up: E.O.J., J.E.G., E.O., E.M., S.K.

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

#### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-019-1315-z>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-019-1315-z>.

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

**Correspondence and requests for materials** should be addressed to D.T.H. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019



## METHODS

**Data reporting.** No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

**Strain selection and construction.** The bacterial strains we used and designated as wild-type were Mtb H37Rv, Msm mc2155<sup>46</sup>, Mmar M, *E. coli* K-12, and *P. aeruginosa* PAO1.

To create Mtb hypomorphs, we used a protein degradation system that has been previously described<sup>18</sup>. In brief, a DAS+4 tag (abbreviated as DAS-tag) was recombined into the chromosome of Mtb H37Rv, at the 3' end of the target gene, before transformation with an ATC-inducible *sspB* plasmid. SspB delivers DAS-tagged protein to the native ClpXP protease. We developed plasmids that produced graded SspB induction (Extended Data Fig. 1a) by varying the strengths of both the promoter driving transcription of *sspB* and the translational initiation signal required to produce SspB protein. Regulation was achieved by repression of the *sspB* promoter by a reverse tetracycline repressor (revTetR); we thus refer to these mutants as TetON mutants (because the presence of ATC represses degradation of the DAS-tagged target protein).

To facilitate the large scale, a sequence-design program was developed (<http://orca2.tamu.edu/tom/U19/seqtool.html>). Every cassette comprises 500-base-pair (bp) flanking sequences around the stop codon of the target, the DAS-tag (inserted at the 3' end of the target gene), a *loxP* site, a unique nucleotide sequence ('molecular barcode'), and a hygromycin-selectable marker. If the target gene was located less than 21 bp upstream of the 5' end of an adjacent open reading frame (ORF), a new ribosomal-binding site was inserted. Designed DNA fragments were synthesized (Gen9 or GenScript) in plasmid pUC57 with flanking PmeI sites. PmeI excision yielded double-stranded recombinering substrates<sup>47</sup>.

Molecular barcodes enabled the identification and quantification of each strain among a pool of strains. Each barcode region was 74 nucleotides long, with common flanking regions on each end that include a PacI site (underlined) and primers for PCR amplification (capitalized), and a unique barcode sequence of 20 nucleotides (<20N>): ttaatTAATCTTGTGGAAAGGACGA<20N>ACGCTATGTG GATACGCTGCTTTAattaa. Each barcode is unique to each target, thus only one SspB version strain of any target can be included in a given pool.

Strains were validated at several steps. DAS-tagged recombinants were selected by hygromycin resistance and screened for target specificity by PCR for the presence of both junctions. To confirm the presence of inserted sequence, PCR confirmation amplified the 3' region of the targeted gene across the insert, showing increased length of the amplicon (approximately 2,500 bp), and absence of a wild-type amplicon (1,100 bp). After curing by plating recombinants on 3% sucrose, loss of the recombinase plasmid was validated by kanamycin sensitivity and by PCR screens for plasmid absence. The final step was introduction of the *sspB* plasmid by selection for streptomycin resistance; strains were validated by a PCR screen for the presence of the *sspB* plasmid.

The MsmΔ*efpA* strain was constructed as previously described<sup>48</sup>. Episomal expression constructs for Mtb *efpA* and *efpA*<sup>V319P</sup> were assembled in the pUV15-tetOrn plasmid<sup>49</sup>, with GFP and kanamycin resistance under control of a strong, leaky ATC-inducible P<sub>smyc</sub> promoter. PCR primers were designed to linearize the pUV15 backbone, excluding the GFP-coding region, and PCR-amplified *efpA* from Mtb genomic DNA was inserted using the InFusion molecular cloning kit (Clontech). The Val315Phe mutation was introduced using primer-directed mutagenesis<sup>50</sup>. Constructs confirmed by Sanger sequencing (Eton or Genewiz) were transformed into Msm by electroporation as previously described<sup>51</sup>.

**Multiplexed screening of compound libraries.** Bioactive compounds were identified from screening efforts in our laboratory<sup>26,40,52</sup> and previously reported in the literature<sup>53–56</sup>, whereas the 50,000-compound library comprised Broad Institute and external collections<sup>56</sup>; see also Supplementary Note 1.

We optimized screening assay parameters in 384-well format, including compound exposure time (14 days), genomic DNA extraction (10% DMSO and 95°C for 15 min), and PCR conditions (20 cycles, annealing at 65°C) that maximized the robust Z'-score of log-transformed counts for a rifampin dilution series and minimized random noise as determined by strain-wise coefficient of variance. Optimal sequencing depth is discussed in Supplementary Note 2.

Final screening pool strains were grown separately in Middlebrook 7H9 (Difco) supplemented with oleic albumin dextrose catalase (OADC, from Becton Dickinson) and 10 mM sodium acetate, appropriate antibiotics, and 1 μg ml<sup>-1</sup> ATC. At mid-exponential growth phase, bacteria were combined equally into a single pooled culture, which was then diluted in Middlebrook 7H9-OADC-acetate to an approximate OD<sub>600</sub> of 0.005. After washing in Middlebrook 7H9-OADC-acetate, 40 μl dilute culture was distributed into clear polystyrene 384-well plates (Corning), containing 120–200 nl of screening compound solution per well as prepared by Broad Institute Compound Management. On every plate, rows A, B, O and P and columns 1 and 24 were left empty. Columns 2 and 23 were occupied by alternating DMSO (negative) and rifampin (positive) on-board controls. Each

batch included eight control plates containing 12-point twofold serial dilutions of rifampin and trimethoprim, and for the larger screen BRD-4592 and methotrexate.

Plates were incubated for 14 days in humidified containers at 37°C. Then, 40 μl of 10% (v/v) aqueous DMSO was added to each well, heating at 80°C for 2 h. PCR (20 cycles) was performed in 384-well PCR plates (Eppendorf) using Q5 Master Mix (NEB) and 1 μl heat-inactivated culture in 10 μl. Primer 5' overhangs added plate and well identification barcodes and sequences for Illumina NGS (Extended Data Fig. 1c).

PCR products were combined and cleaned-up using AMPure XP beads (Beckmann) at twice the volume, eluting in 200 μl MilliQ water.

Sequencing was carried out at the Broad Institute Genomics Platform using Illumina HiSeq 2500 at an average sequencing depth of at least 500 reads per strain per well. Clarification of statistical analysis (<http://www.R-project.org/>)<sup>57,58</sup>, unsupervised machine learning (<https://github.com/jkrijthe/Rtsne>)<sup>56,59,60</sup>, compound annotation<sup>61</sup>, and supervised machine learning<sup>62</sup> are provided in Supplementary Notes 3–6.

**Bactericidal compound kill kinetics.** Mtb cultures at OD<sub>600</sub> of 0.6–1.0 were diluted to OD<sub>600</sub> of 0.4 and split into three 10-ml cultures. Then, 100 μl of 2.4 mM BRD-8000.2 in DMSO or 100 μl DMSO (negative control) was added to each culture, and incubated with shaking at 37°C. At 0, 1, 2, 3, 6 and 13 days, 500-μl samples were taken from each culture, diluted 10<sup>4</sup>-fold, 10<sup>5</sup>-fold and 10<sup>6</sup>-fold in Middlebrook 7H9 in series and plated on Middlebrook 7H10 agar.

Agar plates were incubated at 37°C for 21 days, before colony-forming units were counted.

**Broth microdilution assays.** The minimum inhibitory concentration of compounds was determined in a 96-well plate (Corning), filled with 49 μl of appropriate medium (Middlebrook 7H9-OADC-acetate for Mycobacteria or Lysogeny Broth (LB) for *E. coli* or *P. aeruginosa*), and 1 μl 100× compound DMSO stock. Then, 50 μl exponential-phase bacterial culture diluted to an OD<sub>600</sub> of 0.005 was added. Final concentration was 0.1–50.0 μM. Plates were incubated at 37°C in a humidified container for 24 h for non-mycobacteria, 3 days for Msm, 7 days for Mmar, and 14 days for Mtb. OD<sub>600</sub> was measured using a SpectraMax M5 plate reader (Molecular Dimensions). Normalized percentage outgrowth (NPO) was reported using  $NPO = (x_i - \mu_n) / (\mu_p - \mu_n)$ , in which  $\mu_p$  is the mean positive control value,  $\mu_n$  is the mean negative control value, and  $x_i$  is the value of compound *i*.

To determine the MBC<sub>50</sub> of BRD-8000.3, Mtb bacteria at OD<sub>600</sub> of 0.6–1.0 were washed twice with PBS containing 0.05% (w/v) tyloxapol and resuspended at a final OD<sub>600</sub> of 0.2 in starvation media (Middlebrook 7H9 with 0.05% (w/v) tyloxapol and no other supplementation). The carbon-starved, non-replicating drug-tolerant state was induced by incubation at 37°C for 5 weeks<sup>40</sup>.

Starved cultures were treated with test compounds as above, and viability was determined by plating for colony-forming units as above.

**DNA gyrase assays.** All compounds were tested at 160 μM in duplicate in a 96-well PCR plate (Axygen). A linear pBR322 control was made via digestion of pBR322 by EcoRI-HF (NEB) for 1 h at 37°C. Mtb DNA gyrase (Inspiralis) was diluted in 50 mM Tris-HCl at pH 7.9, 5 mM dithiothreitol (DTT), and 30% (w/v) glycerol and 2.5 U diluted DNA gyrase were added to each reaction mixture; pBR322 supercoiling and kDNA decatenation assays were carried out as recommended by the manufacturer, additionally digesting with 50 μg ml<sup>-1</sup> proteinase K (Qiagen) before 1% (w/v) agarose gel electrophoresis. Pixel density of EtBr-stained bands was measured using ImageJ, fold change was determined by a Gamma-family generalized linear model (GLM) with a log link, using DMSO untreated control as the intercept and the relaxed band density as an offset to normalize for sample loading.

**Mycolic acid assay.** Thin-layer chromatography detection of the incorporation of <sup>14</sup>C-acetic acid into mycolic acids by Mtb was carried out as previously described<sup>63</sup>, using 10 ml exponential-phase culture treated with 100 μl of a 100× compound stock in DMSO to give a final concentration of 10 × MIC<sub>90</sub> as measured by broth microdilution. Pixel density of bands was measured as above, using DMSO untreated control as the GLM intercept and the FAMES band density as an offset to normalize for sample loading.

**Metabolite supplementation assay.** Broth microdilution was performed as above, but each dose response series of a compound of interest was repeated three times, each time with a different metabolite in the growth medium: PABA at 200 μM from a 20 mM stock in ethanol; folic acid at 200 μM from a 20 mM stock in DMSO; and L-tryptophan at 1 mM from a 100 mM stock in 500 mM hydrochloric acid.

**RNA polymerase assay.** Inhibition of in vitro RNA synthesis by *E. coli* RNAP was carried out as previously described<sup>38</sup>, using 2 pmol of 5'-phosphorylated, 45-nucleotide single-stranded DNA (ssDNA) oligomer (5'-GGAGCCCTTATTTA GACTTAAATAAGTTCCTCAACATCCTTCGAT-3'; Integrated DNA Technologies), which had been circularized using CircLigase ssDNA Ligase (Lucigen), *E. coli* RNA polymerase (NEB), 10 μM rifampin (positive control) or 100 μM test compound. All compounds were tested in quadruplicate. RNA was quantified using RiboGreen fluorescent dye (Thermo Fisher). Fold change was determined by a GLM with a log link, using DMSO untreated control wells as the intercept and modelling plate-to-plate variation.



**Evolution of resistant mutants.** Mid-exponential growth phase bacterial cultures were pelleted and resuspended as a slurry in 1 ml Middlebrook 7H9-OADC. Then, 50  $\mu$ l was plated on 6 ml agar containing  $2 \times$ ,  $4 \times$  or  $8 \times$  MIC<sub>90</sub> of compound. Plates were incubated at 37°C in a humidified container for more than 21 days.

Colonies were picked and cultured in 1 ml Middlebrook 7H9-OADC-acetate in a 96-well 2 ml well volume culture block and incubated at 37°C for 7 days in a humidified container.

**Whole-genome sequencing of mycobacteria.** In brief, 10  $\mu$ l of culture was combined with 10  $\mu$ l 10% (v/v) DMSO in a 96-well clear round-bottom plate (Corning). Plates were heat-inactivated at 80°C for 2 h. Genomic DNA (gDNA) was separated from intact cells and cell debris using AMPure XP (Beckman), eluting in 40  $\mu$ l MilliQ water. gDNA (1.5  $\mu$ l) was amplified using 6  $\mu$ M random primers (Invitrogen) and  $\phi$ 29 DNA Polymerase (NEB) in a 10  $\mu$ l reaction volume at 30°C for 24 h.

Amplified gDNA was purified using AMPure XP and subjected to NextEra XT NGS library construction (Illumina) before 150-cycle paired-end sequencing on the Illumina MiSeq platform. Reads were aligned to the AL123456 reference sequence<sup>64</sup> using the BWA-mem algorithm and mutations were called using the deepSNV package<sup>65</sup> for R.

**Kinetic solubility.** In brief, 10  $\mu$ l of a 10 mM stock solution of test compound was added to 490  $\mu$ l 50 mM phosphate buffer, pH 7.4. The solution was shaken at room temperature for 24 h, before filtering and detecting filtrate concentration using ultraviolet spectroscopy and comparing to 1, 20 and 200  $\mu$ M standards.

**Mouse pharmacokinetics.** Experiments were performed in accordance with relevant guidelines and regulations. Oral dosing of 12-h fasted female 7–9-week-old C57BL/6 mice ( $n = 12$ ) was performed, using up to 150 mg kg<sup>-1</sup> test compound in 5% (v/v) corn oil, 10% (v/v) labrasol, 5% (v/v) solutol in water. The plasma concentration of test compounds was measured at nine time points up to 8 h. Internal standard (120  $\mu$ l of 100 ng ml<sup>-1</sup> labetalol, 100 ng ml<sup>-1</sup> dexamethasone, 100 ng ml<sup>-1</sup> tolbutamide, 100 ng ml<sup>-1</sup> verapamil, 100 ng ml<sup>-1</sup> glyburide, 100 ng ml<sup>-1</sup> celecoxib in acetonitrile) was added to 6  $\mu$ l plasma samples. Protein precipitate was centrifuged out; 4  $\mu$ l supernatant was subjected to liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis, on a UPLC protein BEH C4 1.7  $\mu$ m 2.1  $\times$  50 mm column (Acquity) with gradient of 22–95% acetonitrile in water, and detecting using a positive-mode Triple Quad 6500+ mass spectrometer (AB SCIEX Instruments).

**HepG2 toxicity.** HepG2 cells (50  $\mu$ l) were seeded at a density of 1,000 cells per well. Test compounds dissolved in DMSO were threefold serially diluted for ten doses ( $n = 1$ ). Compound stocks (125 nl) were transferred into a 384-well plate, which was incubated for 72 h at 37°C. The plate was allowed to reach room temperature for 30 min, after which 25  $\mu$ l CellTiterGlo reagent (Promega) was added. After 10 min, luminescence was detected using an EnVision plater reader (PerkinElmer). Normalized percentage inhibition was calculated using DMSO vehicle control values, and 1  $\mu$ M paclitaxel. The half-maximum inhibitory concentration (IC<sub>50</sub>) values were calculated using the four-parameter logistic model in Prism statistical analysis software (GraphPad).

**Microsomal stability.** Test compound (1  $\mu$ M) was incubated with pooled microsomes in 100 mM potassium phosphate buffer, pH 7.4, supplemented with 2 mM NADPH and 3 mM magnesium chloride at 37°C. At 0, 5, 15, 30 and 45 min, samples were taken ( $n = 1$ ), and stopped by adding an equal volume ice-cold methanol, and incubating on ice for 10 min. After removing precipitated protein by centrifugation, remaining compound was measured using LC–MS/MS as above. Half-life was determined by fitting a first-order decay model to the time course.

**Isoform-specific cytochrome P450 (CYP) inhibition.** Seven concentrations ( $n = 1$ ) of test compound, of which the DMSO stock solution was diluted tenfold in acetonitrile, were added to pooled human liver microsomes in 100 mM potassium phosphate buffer, pH 7.4, supplemented with 2 mM NADPH and 3 mM magnesium chloride at 37°C. Conversion of isoform-specific substrate probes to their products (tacrine to hydroxytacrine for CYP1A2; bupropion to hydroxybupropion for CYP2B6; amodiaquine to desethylamodiaquine for CYP2C8; tolbutamide to  $\alpha$ -hydroxytolbutamide for CYP2C9; mephenytoin to 4-hydroxymephenytoin for CYP2C19; and dextromethorphan to dextrorphan for CYP2D6) was stopped after 10 min incubation (15 min for CYP2C9; 60 min for CYP2C19) by adding an equal volume of methanol, and incubating on ice for 10 min. After removing precipitated protein by centrifugation, products were measured using LC–MS/MS. Reduction in substrate conversion was normalized to the maximal effects of isoform-specific positive controls ( $\alpha$ -naphthoflavone for CYP1A2; ticlopidine for CYP2B6; quercetin for CYP2C8; sulfaphenazole for CYP2C9; ticlopidine for CYP2C19; and quinidine for CYP2D6).

**Mouse plasma protein binding.** Protein-free and plasma protein-containing compartments were separated with a semi-permeable membrane, and allowed to equilibrate with 5  $\mu$ M test compound at 37°C ( $n = 3$ ). Compound concentration on both sides of the membrane was measured using LC–MS/MS, and the bound fraction was calculated as  $1 - F_c/T_c$ , in which  $F_c$  is the free compound concentration

in the dialysate, and  $T_c$  is the compound concentration in the retentate. Warfarin was used as a positive control.

**hERG inhibition.** hERG inhibition ( $n = 2$ ) was assayed using the Predictor hERG Fluorescence Polarization Kit (Thermo) according to the manufacturer's instructions.

**Efflux assay.** Efflux rates were measured as previously described<sup>66–68</sup>. Msm or Mtb strains were grown in Middlebrook 7H9 medium to an OD<sub>600</sub> of 0.4–0.6. Cultures were then centrifuged for 5 min at 3,500 r.p.m. The pellet was washed once with PBS at 37°C and resuspended in PBS to give a final OD<sub>600</sub> of 0.4. Cultures were split into eight and EtBr (or bisbenzimidazole-H) was added at a final concentration of 0.2–1.95  $\mu$ g ml<sup>-1</sup> (or for bisbenzimidazole-H, 0.6–6.3  $\mu$ g ml<sup>-1</sup>), and bacteria were incubated for 30 min (Msm) or 2 h (Mtb) at 37°C. After EtBr treatment, cells were centrifuged for 5 min at 3,500 r.p.m. and resuspended in 37°C PBS to give a final OD<sub>600</sub> of 0.8. A white 96-well plate (Corning) was prepared with serially diluted compound and 50  $\mu$ l PBS containing 0.8% (w/v) glucose. Dye-loaded bacteria (50  $\mu$ l) were added to each well of the plate. Fluorescence was read at 37°C in a SpectraMax M5 plate reader using 530 nm excitation and 585 nm emission wavelengths for EtBr (or 346 nm and 460 nm for bisbenzimidazole-H) and was recorded every 30 s for 2 h (Msm) or 4 h (Mtb).

To infer kinetic parameters, we modelled the rate of fluorescence decay as a modified Michaelis–Menten<sup>67</sup> equation, which included a term for Fick diffusion<sup>69</sup> between the cytoplasm and extracellular milieu. This approach is discussed in Supplementary Note 8.

**Compound accumulation assay.** Accumulation of BRD-8000.3 inside Msm bacteria was assayed in triplicate as previously described<sup>70</sup>, using the intrinsic fluorescence of the compound in PBS containing 0.4% (w/v) glucose, and 1.5  $\mu$ l of 5 mM BRD-8000.3 in DMSO. Bacteria were then incubated for 3 h at 37°C. Suspensions were centrifuged for 10 min at 3,500 r.p.m. Cell pellets were resuspended in 60  $\mu$ l PBS containing 5% (v/v) DMSO then frozen at –80°C for 20 min, heated to 98°C for 15 min, before centrifugation at 10,000 r.p.m. for 10 min.

Lysate (60  $\mu$ l) was added to wells of a black 384-well plate. Excitation and emission wavelengths of BRD-8000.3 were 355 nm and 419 nm. Change in compound accumulation in lysate was determined by subtracting mean untreated control fluorescence from the treated sample fluorescence.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

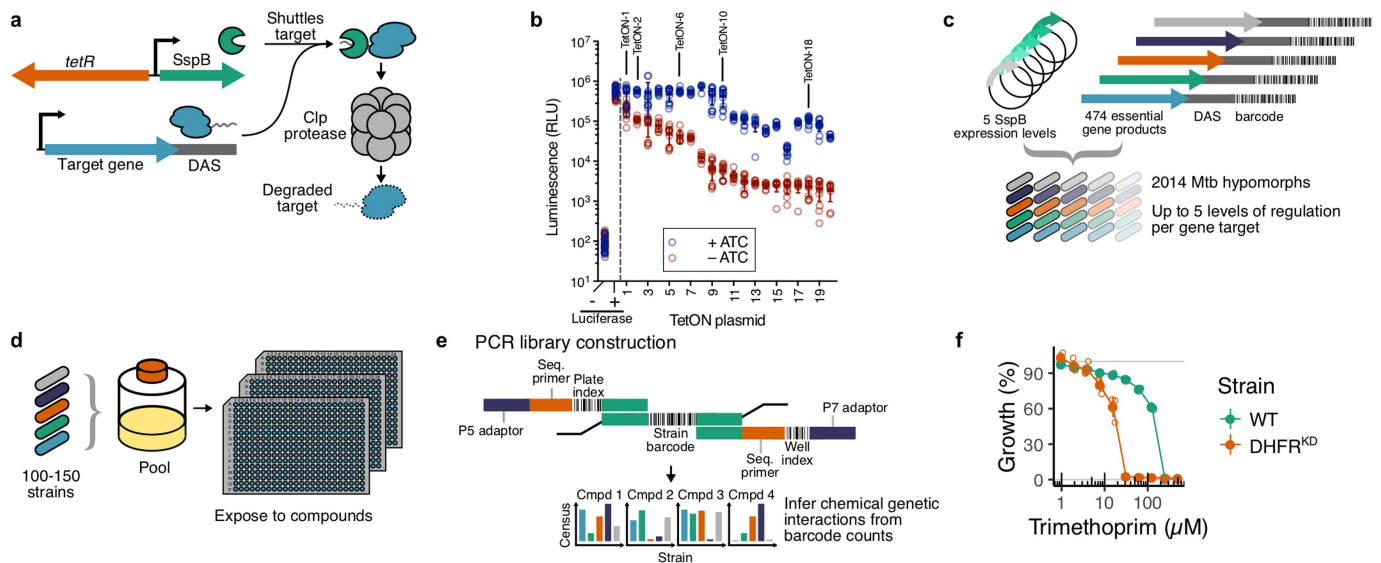
Source Data for Figs. 1a, b, 2, 3a, c, e, 4a, 5b and Extended Data Figs. 1b, 2a, c, d, 4a, 5, 7e, 8, 9a, b, c, f, i, l are provided with the online version of the paper. The raw primary screen data, calculated fold changes and  $P$  values, and compound annotations, are available online at <https://broad.io/cgtp>.

## Code availability

ConcensusGLM is available at <https://doi.org/10.5281/zenodo.3235787> and on GitHub at <http://github.com/broadinstitute/concensusGLM>. Other computer code is available from the corresponding author upon request.

- Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T. & Jacobs, W. R. Jr. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol. Microbiol.* **4**, 1911–1919 (1990).
- Murphy, K. C., Papavinasandaram, K. & Sasseti, C. M. Mycobacterial recombineering. *Methods Mol. Biol.* **1285**, 177–199 (2015).
- Murphy, K. C. et al. ORBIT: a new paradigm for genetic engineering of mycobacterial chromosomes. *mBio* **9**, e01467-18 (2018).
- Ehrt, S. et al. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res.* **33**, e21 (2005).
- Carey, M. F., Peterson, C. L. & Smale, S. T. PCR-mediated site-directed mutagenesis. *Cold Spring Harb. Protoc.* **2013**, 738–742 (2013).
- Wards, B. J. & Collins, D. M. Electroporation at elevated temperatures substantially improves transformation efficiency of slow-growing mycobacteria. *FEMS Microbiol. Lett.* **145**, 101–105 (1996).
- Stanley, S. A. et al. Identification of novel inhibitors of *M. tuberculosis* growth using whole cell based high-throughput screening. *ACS Chem. Biol.* **7**, 1377–1384 (2012).
- Reynolds, R. C. et al. High throughput screening of a library based on kinase inhibitor scaffolds against *Mycobacterium tuberculosis* H37Rv. *Tuberculosis (Edinb.)* **92**, 72–83 (2012).
- Ananthan, S. et al. High-throughput screening for inhibitors of *Mycobacterium tuberculosis* H37Rv. *Tuberculosis (Edinb.)* **89**, 334–353 (2009).
- Maddy, J. A. et al. Antituberculosis activity of the molecular libraries screening center network library. *Tuberculosis (Edinb.)* **89**, 354–363 (2009).
- Ballell, L. et al. Fueling open-source drug discovery: 177 small-molecule leads against tuberculosis. *ChemMedChem* **8**, 313–321 (2013).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* 4th edn (Springer, 2002).

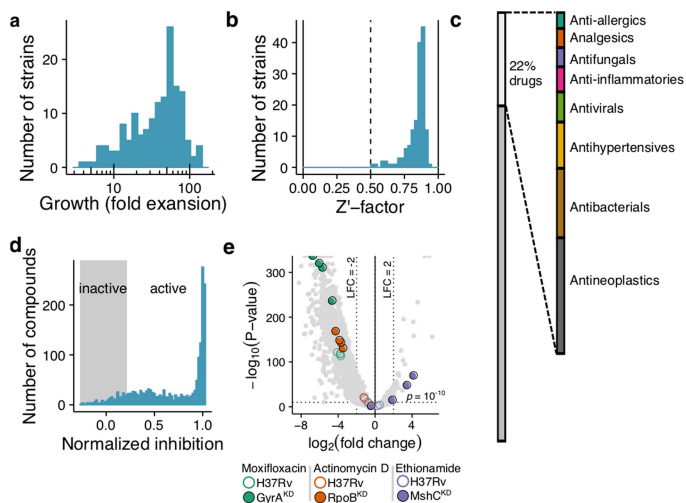
59. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
60. Murtagh, F. & Legendre, P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J. Classif.* **31**, 274–295 (2014).
61. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
62. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
63. Vilcheze, C. & Jacobs, W. R. Isolation and analysis of *Mycobacterium tuberculosis* mycolic acids. *Curr. Protoc. Microbiol.* **5**, 10A.3.1–10A.2.11 (2007).
64. Cole, S. T. et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).
65. Gerstung, M. et al. Reliable detection of subclonal single-nucleotide variants in tumour cell populations. *Nat. Commun.* **3**, 811 (2012).
66. Paixão, L. et al. Fluorometric determination of ethidium bromide efflux kinetics in *Escherichia coli*. *J. Biol. Eng.* **3**, 18 (2009).
67. Michaelis, L., Menten, M. L., Johnson, K. A. & Goody, R. S. The original Michaelis constant: translation of the 1913 Michaelis–Menten paper. *Biochemistry* **50**, 8264–8269 (2011).
68. Nelder, J. A. & Mead, R. A simplex method for function minimization. *Comput. J.* **7**, 308–313 (1965).
69. Fick, A. Ueber diffusion. *Ann. Phys.* **170**, 59–86 (1855).
70. Vergalli, J. et al. Spectrofluorimetric quantification of antibiotic drug concentration in bacterial cells for the characterization of translocation across bacterial membranes. *Nat. Protocols* **13**, 1348–1361 (2018).



**Extended Data Fig. 1 | Mtb hypomorph strain creation.** **a**, Hypomorph strains were constructed by introducing a DAS tag at the 3' end of the gene of interest, with concomitant introduction of a 20-nucleotide barcode and an episomally encoded, regulated *SspB* gene to control the level of protein depletion. **b**, Degradation of a DAS-tagged target gene product was mediated by *SspB*, the expression of which was driven by an ATC-inducible TetON promoter. To allow individualized degrees of knockdown for each gene product, a series of TetON promoters with varying strengths was generated. Regulated promoter strength was quantified by fusion to a luciferase gene and measuring luminescence in the presence and absence of the ATC inducer. In the screen, strains containing the TetON-1, -2, -6, -10 and -18 promoters were used for subsequent strain construction. Independent biological replicates ( $n = 8$ ) are shown as open circles; means are shown as horizontal bars; error bars denote 95% confidence intervals.

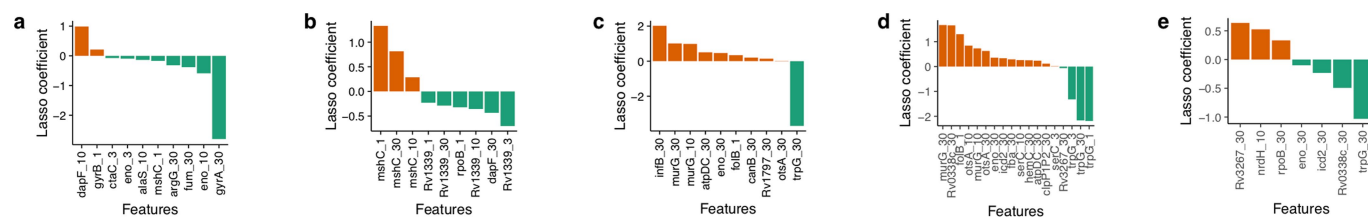
**c**, A range of up to five different knockdown levels was attempted using five different promoters for each target gene, allowing the generation of 2,014 hypomorphs. **d**, Barcoded hypomorph strains were pooled and distributed into 384-well plates containing the compound library and incubated for 14 days. **e**, Chromosomal strain barcodes were inserted into each engineered hypomorph, thus allowing PCR amplification by an array of primers containing 5' overhangs encoding screen location (well and plate) barcodes. For census enumeration of pooled strains, PCR products were combined and subjected to Illumina NGS. **f**, Dose response of trimethoprim, a DHFR inhibitor, against wild-type Mtb and the DHFR hypomorph, showing the hypersensitivity of the hypomorph. Growth was calculated as  $OD_{600}$  normalized to untreated controls. Independent biological replicates ( $n = 4$ ) are shown as open circles; means are shown as filled circles; error bars denote 95% confidence intervals.





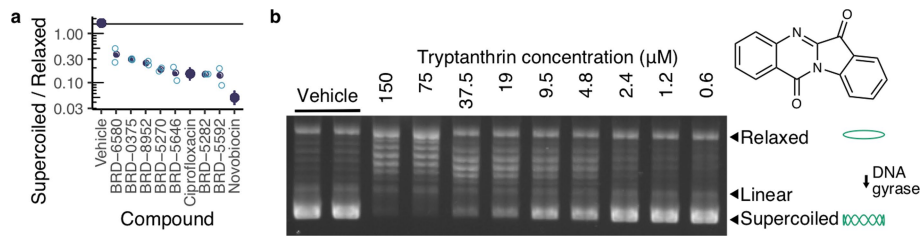
### Extended Data Fig. 2 | Strain and compound collection

**characterization.** **a**, Histogram of growth of the individual hypomorphs in the screening pool over the 14-day duration of the screen. Most hypomorphs were within a tenfold window of growth rates. **b**, Histogram of hypomorph  $Z'$  factors between the untreated wells and rifampin control wells of the screening assay for the hypomorphs. All  $Z'$  factors were greater than 0.5, indicating an excellent screening assay. **c**, The composition of the bioactive compound library. The right bar shows the broad classes within the segment of known drugs. **d**, The bioactive compound library was tested up to 50  $\mu\text{M}$  against a GFP-expressing Mtb strain. Most compounds had detectable activity for at least one concentration tested. **e**, Volcano plot (maximum likelihood mean fold change from  $n = 2$  biologically independent samples against the unadjusted two-sided Wald test  $P$  value) of chemical-genetic interactions from the bioactive library. Each point represents a single strain-compound interaction at a single concentration. Some interactions of interest are highlighted: compounds are designated by colour, with wild-type Mtb interactions shown as open circles and hypomorphs interactions of interest shown as solid circles. Most interactions were inhibitory, because the compound library was confirmed to be enriched for antitubercular activity.



**Extended Data Fig. 3 | Feature weights of Lasso classifiers.** **a**, Feature weights of the hypomorphs used by the Lasso binary classifier to predict DNA gyrase inhibitors. We trained on the fluoroquinolones, with the GyrA hypomorph being the most predictive strain. Features are denoted as the target gene of the hypomorph and compound concentration separated by an underscore. **b**, As in **a** for the Lasso binary classifier to predict inhibitors of mycolic acid biosynthesis. We trained on known InhA inhibitors, with the MshC hypomorph being a prominent discriminator.

**c**, As in **a** for the Lasso binary classifier to predict inhibitors of folate biosynthesis. We trained on the sulfonamides, with the TrpG hypomorph being a prominent discriminator. **d**, As in **a** but trained on chemical-genetic interaction profiles from the sulfonamides and confirmed new folate biosynthesis inhibitors. **e**, As in **a** but trained on chemical-genetic interaction profiles of the newly identified inhibitors of tryptophan biosynthesis.



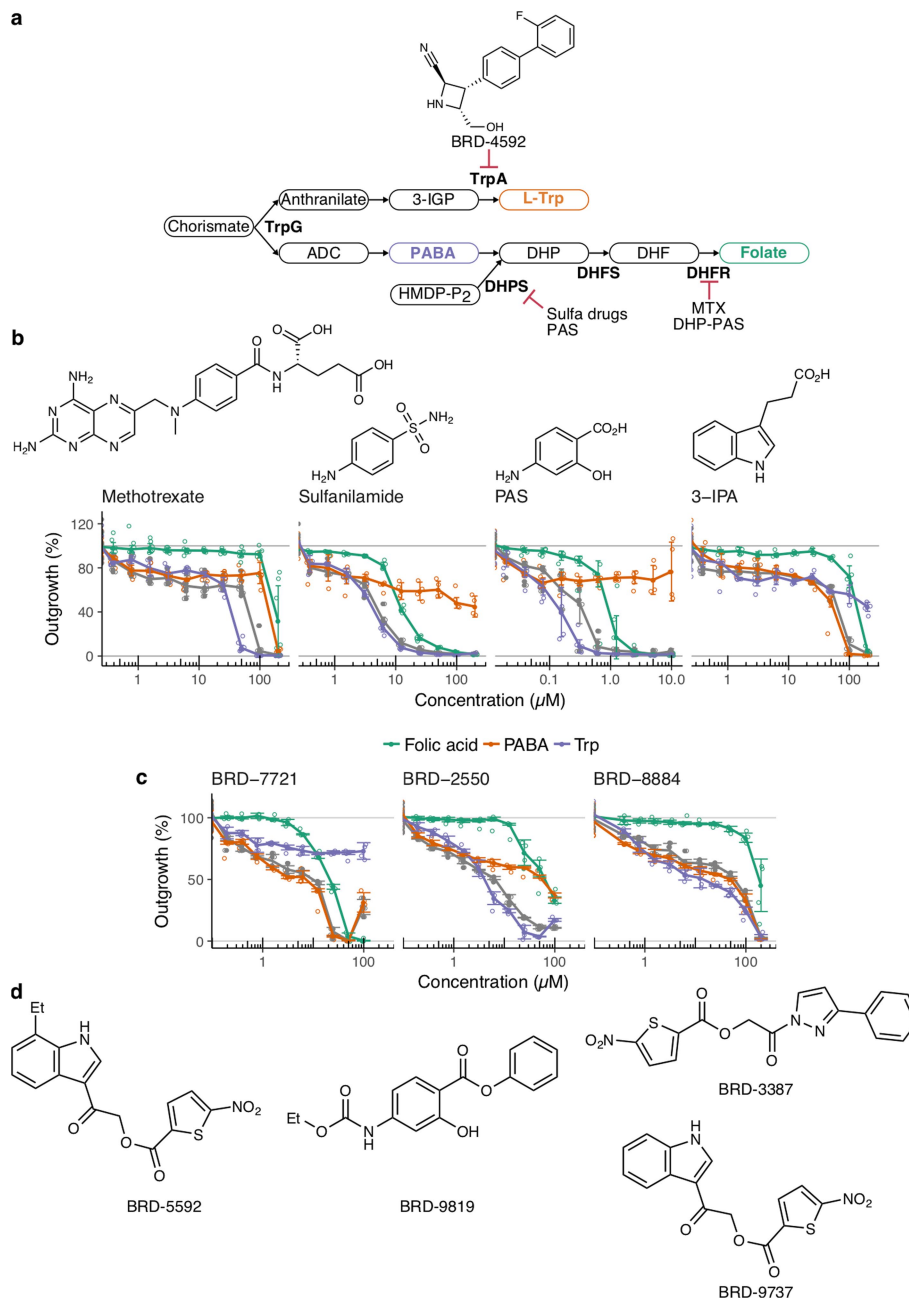
#### Extended Data Fig. 4 | Validation of predicted DNA gyrase inhibitors.

**a.** Actual compound performance of predicted DNA gyrase inhibitors in an agarose gel-based in vitro assay of DNA gyrase supercoiling inhibition. The ratio of imaged pixel intensities (Supplementary Fig. 1a–l) for supercoiled and relaxed bands was indicative of inhibition, shown by the ciprofloxacin control. Eight of the twenty-seven compounds with the greatest effect sizes that showed statistically significant ( $P < 0.05$ , two-sided Wald test) inhibition are shown. Open circles show the independent

samples ( $n = 23$  for ciprofloxacin and novobiocin,  $n = 2$  for all other conditions); filled circles indicate the mean pixel intensity ratio; error bars denote the 95% confidence interval of the mean. **b.** Agarose gel showing increasing inhibition of Mtb DNA gyrase supercoiling activity with increasing tryptanthrin concentration. DNA gyrase catalyses supercoiling of pBR322; inhibitors prevent the accumulation of supercoiled gel bands. This experiment was repeated independently once with similar results. The uncropped version is shown in Supplementary Fig. 1m.

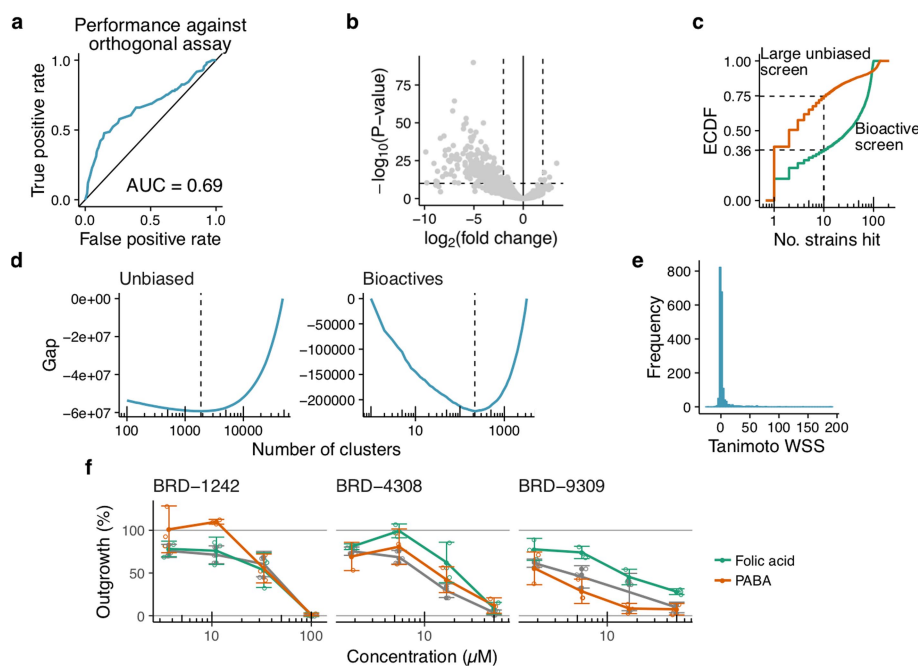






**Extended Data Fig. 6 | New classes of inhibitors of folate and tryptophan biosynthesis.** **a**, Schematic of the folate and tryptophan biosynthesis pathways. TrpG is an amphibolic enzyme, upstream of both PABA and tryptophan. Biosynthetic enzymes mentioned in the text are indicated in their metabolic context. 3-IGP, 3-indoleglycerol phosphate; ADC, 4-amino-4-deoxychorismate; DHF, dihydrofolate; DHFR, dihydrofolate reductase; DHFS, dihydrofolate synthase; DHP, dihydropteroate; DHP-PAS: adduct of DHP and PAS; DHPS, dihydropteroate synthase; HMDP-P<sub>2</sub>, 6-hydroxymethyl-7,8-dihydropterin diphosphate; MTX, methotrexate; PABA, *para*-amino-benzoic acid; PAS, *para*-aminosalicylic acid; Sulfa drugs, sulfonamide antibiotics. **b**, Dose–response curves of known inhibitors of folate biosynthesis and the validated tryptophan biosynthesis inhibitor scaffold, 3-indole propionic acid, supplemented with PABA, folic acid or tryptophan. Chemical structures of the known inhibitors are shown. Independent biological

replicates ( $n = 4$ ) are shown as open circles; means are shown as filled circles; error bars show 95% confidence intervals. **c**, Actual performance of predicted inhibitors of folate biosynthesis in a metabolite rescue assay. Mtb was treated with predicted inhibitors in the presence or absence of tryptophan, folic acid or PABA. The effect of BRD-7721, a 3-indole propionic acid ester, is abolished by supplementation with tryptophan, indicating it is an inhibitor of tryptophan biosynthesis. By contrast, the effect of the nitrothiophene BRD-2550 is abolished by folate and PABA, and that of BRD-8884 is abolished by folate alone, showing that they are inhibitors of folate biosynthesis with distinct mechanisms. Independent biological replicates ( $n = 4$ ) are shown as open circles; means are shown as filled circles; error bars denote the 95% confidence intervals. **d**, Chemical structures of predicted and subsequently validated inhibitors of folate biosynthesis, including the nitrothiophenes and *para*-aminosalicylic acid derivative BRD-9819.

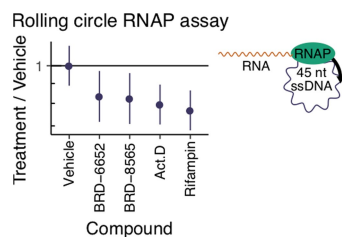


### Extended Data Fig. 7 | Performance of a large, unbiased compound library.

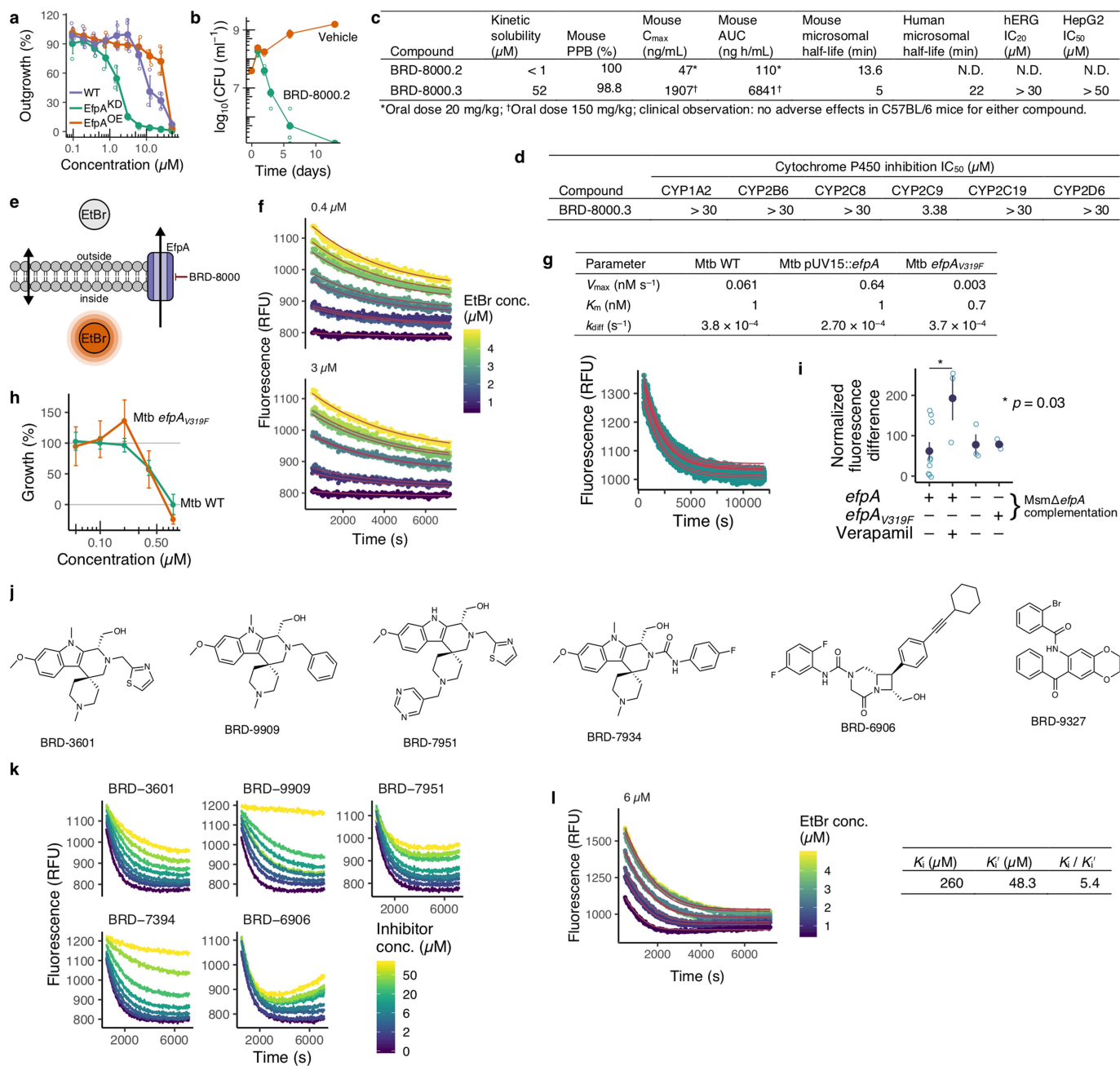
**a**, ROC curve showing that primary data were predictive of activity in a confirmatory secondary growth assay ( $n = 4$  biologically independent experiments). We retested 454 compound–strain interactions using a resazurin, growth-based colorimetric assay. Taking 75% inhibition in the secondary assay as the ground truth, we demonstrated the primary assay as predictive of real activity that could be detected by conventional methods. Using 50% and 90% inhibition as the ground-truth assumption, the ROC AUC values were 0.61 and 0.69. **b**, Volcano plot (maximum likelihood mean fold change from  $n = 2$  biologically independent samples against the unadjusted two-sided Wald test  $P$  value) of chemical–genetic interactions from the larger unbiased chemical library. Each point represents a single strain–compound interaction at a single concentration. **c**, Compounds in the library of bioactive compounds generally hit more strains than compounds in the unbiased library. Empirical cumulative distribution functions of number of hypomorphs hit by compounds in the two screens is plotted. Shown by the dotted lines, 36% of compounds in the bioactive library and 75% of compounds in the larger library hit 10 strains or fewer, suggesting that activity detected in the larger screen was generally more hypomorph-specific. **d**, Clustering of chemical–genetic interaction profiles. The number of chemical–genetic interaction profile clusters in

the two libraries was determined by finding the minimum Gap statistic, a measure of within-cluster similarity compared to clustering at random. The minimum, denoted by the dotted lines, shows the estimate of the true number of clusters in the unbiased and bioactive libraries, with the unbiased library containing many more unique chemical–genetic clusters. **e**, Chemical–genetic clusters ( $n = 1,864$ ) are enriched for chemically similar compounds in the unbiased library. The  $y$  axis shows the frequency of chemical–genetic clusters with a particular Tanimoto WSS Z score ( $x$  axis), which is an indicator of in-cluster chemical similarity. A total of 221 (12%) of the clusters have meaningful structure–activity relationships—that is, compounds within a cluster have significantly greater chemical similarities than by chance—as indicated by a one-sided permutation test (unadjusted  $P < 0.05$ , 10,000 permutations). **f**, Actual compound performance of predicted inhibitors of folate biosynthesis in a metabolite rescue assay. *Mtb* was treated with predicted inhibitors in the presence or absence of folate or PABA. The effect of BRD-1242 is abolished by PABA alone, and the effects of BRD-4308 and BRD-9309 are abolished by folate alone, suggesting that they are inhibitors of folate biosynthesis with distinct mechanisms. Independent biological replicates ( $n = 3$ ) are shown as open circles; means are shown as filled circles; error bars denote 95% confidence intervals.



**Extended Data Fig. 8 | Validation of predicted RNAP inhibitors.**

Actual compound performance of predicted RNAP inhibitors in an in vitro assay for inhibition of RNA synthesis by *E. coli* RNAP. Three compounds that showed statistically significant inhibition are shown with a rifampin control ( $P < 0.05$ , two-tailed Wald test). Open circles show the independent replicates ( $n = 36$  for vehicle control,  $n = 4$  for BRD-6652,  $n = 8$  for BRD-8565 and rifampin,  $n = 12$  for actinomycin D), filled circles indicate the mean ratio of treatment-to-vehicle fluorescence, and error bars show the 95% confidence interval of the mean. Act. D, actinomycin D.

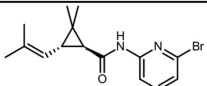
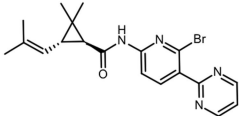


Extended Data Fig. 9 | See next page for caption.

**Extended Data Fig. 9 | Inhibitors of EfpA, in *M. tuberculosis*.** **a**, Dose–response analysis of BRD-8000 on the growth of wild-type Mtb, the EfpA hypomorph, and a mutant overexpressing EfpA (pUV15::*efpA*), demonstrating hypersensitivity of the hypomorph. Independent biological replicates ( $n = 4$ ) are shown as open circles; means are shown as filled circles; error bars denote 95% confidence intervals. **b**, BRD-8000.2 is bactericidal as demonstrated by reducing colony-forming units over time. Independent biological replicates ( $n = 8$ ) are shown as open circles; means are shown as filled circles; error bars denote 95% confidence intervals. **c**, Toxicity and bioavailability measurements of BRD-8000.2 and BRD-8000.3. hERG, human ether a-go-go related gene. **d**, Cytochrome P450 (CYP) inhibition measurements of BRD-8000.2 and BRD-8000.3. **e**, Schematic of the EtBr efflux assay. Bacteria were loaded with EtBr and its efflux was monitored by change in fluorescence. **f**, Example kinetic time courses of EtBr fluorescence decay for Msm with BRD-8000.3. The concentration of EtBr used for pre-incubation is indicated in colour, with two different inhibitor concentrations shown. Numerically integrated Michaelis–Menten best-fit time courses are shown in red. Experiments were repeated independently once with similar results. **g**, Kinetic time course of EtBr fluorescence decay for Mtb. Numerically integrated Michaelis–Menten best-fit time courses are shown in red. The table shows best-fit Michaelis–Menten parameters and Fick's diffusion constant for wild-type Mtb, an EfpA-overexpressor (pUV15::*efpA*), and the BRD-8000 resistant mutant (*efpA*<sup>V319F</sup>). Although the in vivo apparent maximal efflux rate ( $V_{max}$ ) of the EfpA overexpressor is higher than the wild-type Mtb, that of the BRD-8000 resistant mutant is not, indicating that the resistant mutant is not hyperactive for efflux. Experiments were

repeated independently once with similar results. **h**, Dose–response analysis of isoniazid against wild-type Mtb and the BRD-8000 resistant mutant (*efpA*<sup>V319F</sup>). Because isoniazid is a substrate of EfpA, no shift in the MIC<sub>90</sub> value for isoniazid with the BRD-8000 resistant mutant indicates that EfpA(V319F) is not hyperactive for efflux. Mean growth ( $n = 4$  biologically independent replicates) is shown, with error bars indicating 95% confidence intervals. **i**, Results of an assay for intracellular accumulation of BRD-8000.3 in MsmΔ*efpA* complemented with either Mtb wild-type *efpA* or *efpA*<sup>V319F</sup>. Fluorescence of BRD-8000.3 in bacterial lysates was measured, and lysate background fluorescence was subtracted. Although verapamil, a general efflux pump inhibitor, caused statistically significant intracellular accumulation of BRD-8000.3, there was no significant difference in accumulation between the different strains in the absence of verapamil, indicating that EfpA(V319F) is not hyperactive for efflux, and that BRD-8000.3 is not a substrate of EfpA. Independent biological replicates ( $n = 9$  for wild-type control without verapamil;  $n = 3$  for other conditions) are shown as open circles; means are shown as filled circles; error bars denote s.e.m. **j**, Additional compounds predicted and validated to be efflux inhibitors. **k**, Dose–response analysis of kinetic time courses of EtBr fluorescence decay for compounds in **h**. Increasing inhibitor concentration is denoted by colour. Experiments were repeated independently once with similar results. **l**, Example kinetic time courses of EtBr fluorescence decay for Msm at one concentration (6 μM) of BRD-9327. Numerically integrated Michaelis–Menten best-fit time courses are shown in red. The table shows global best-fit kinetic inhibition parameters across ten concentrations for this inhibitor. Experiments were repeated independently once with similar results.

**Extended Data Table 1 | Minimum inhibitory concentrations of BRD-8000 series compounds against non-tubercular strains**

Compound	Structure	Eco WT	Msm WT	Mmar WT	Eco $\Delta tolC$	Psa WT	Psa PAO397	Kpn WT	Acb WT	SA WT
BRD-8000.2		> 50	> 50	12.5	N.D.	> 50	> 50	> 50	> 50	> 50
BRD-8000.3		> 50	> 50	6.25	> 50	> 50	> 50	> 50	> 50	> 50

Concentrations shown are micromolar. Acb, *Acinetobacter baumannii*; Eco, *Escherichia coli*; Kpn, *Klebsiella pneumoniae*; N.D., not determined; Psa, *Pseudomonas aeruginosa*; SA, *Staphylococcus aureus*.

## Reporting Summary

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

### Statistical parameters

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

n/a | Confirmed

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

*Our web collection on [statistics for biologists](#) may be useful.*

### Software and code

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

Data collection

For processing raw sequence data from Illumina, PicardTools and Samtools. Fluorescence and absorbance plate reader data was collected using Molecular Dimensions SoftMax Pro.

Data analysis

GraphPad Prism versions 7.03, 8.1.1  
 Microsoft Excel 16.16.8  
 For gel and TLC image analysis, ImageJ version 1.50i  
 R versions 3.2, 3.5  
 Python versions 2.7, 3.6 with RDKit version 2018\_03\_4  
 For nucleotide sequence analysis, Samtools, GATK.  
 ConsensusGLM is available at <http://github.com/eachanjohnson/consensusGLM>. Other custom computer code is available from the corresponding author upon request.

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



## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The raw primary screen data and calculated log fold changes and p-values are available online at <https://broad.io/cgtb> and from the corresponding author upon request.

## Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://nature.com/authors/policies/ReportingSummary-flat.pdf)

## Life sciences study design

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

Sample size	Sample sizes (n=2) for screening were chosen as standard for high throughput compound screening as a balance of cost and accuracy. The assay was optimized to allow 2 replicates to provide statistical power. All results from the screening were confirmed by orthogonal methods, whose sample size was chosen (n=3-4) from experience to provide power and accuracy.
Data exclusions	No exclusions
Replication	Primary data were generated from at least 2 independent replicates, which gave similar results. Results were confirmed using orthogonal methods, which demonstrated the reliability of the primary data as described in the manuscript. Follow-up mechanism of action studies were performed at least twice with similar results.
Randomization	Plate-to-plate variation in screening and RNA polymerase assays, and gel-to-gel variation in DNA gyrase assays was accounted for by modelling covariate effects (i.e. batch effects) with a generalized linear model as described in the manuscript.
Blinding	Compounds in our study were assigned ID numbers, essentially blinding their identity until after collection and analysis was complete. Follow-up mechanism of action studies were not blinded.

## Reporting for specific materials, systems and methods

### Materials & experimental systems

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

### Methods

n/a	Involvement 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

## Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials      Biological materials, including hypomorphic mutants, are available from the corresponding author upon reasonable request.

## Eukaryotic cell lines

---

Policy information about [cell lines](#)

Cell line source(s)	HepG2 (human liver hepatocellular carcinoma) cells: WuXi AppTec
Authentication	None of the cell lines were authenticated in our lab.
Mycoplasma contamination	The cell lines tested negative for mycoplasma contamination.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	None.

## Animals and other organisms

---

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Female mice, C57BL/6, 7-9 weeks old
Wild animals	The study did not involve wild animals
Field-collected samples	The study did not involve samples collected from the field