Title: Systematic measurement of combination drug landscapes to predict *in vivo* treatment outcomes for tuberculosis

Authors: Jonah Larkins-Ford[1-4], Talia Greenstein[1-3], Nhi Van[1], Yonatan N. Degefu[1,4], Michaela C. Olson[1], Artem Sokolov[4], Bree B. Aldridge[1-5]*

Affiliations:

[1]Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA.

[2]Stuart B. Levy Center for Integrated Management of Antimicrobial Resistance, Boston, MA.

[3]Graduate School of Biomedical Sciences, Tufts University School of Medicine, Boston, MA.

[4]Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA.

[5]Department of Biomedical Engineering, Tufts University School of Engineering, Medford, MA.

*To whom correspondence should be addressed: Bree B. Aldridge, bree.aldridge@tufts.edu, 617-636-6703

One Sentence Summary: Signatures of *in vitro* potency and drug interaction measurements predict combination therapy outcomes in mouse models of tuberculosis.

Abstract: A lengthy multidrug chemotherapy is required to achieve a durable cure in tuberculosis. Variation in *Mycobacterium tuberculosis* drug response is created by the differing microenvironments in lesions, which create different bacterial drug susceptibilities. To better realize the potential of combination therapy to shorten treatment duration, multidrug therapy design should deliberately explore the vast combination space. We face a significant scaling challenge in making systematic drug combination measurements because it is not practical to use animal models for comprehensive drug combination studies, nor are there well validated high-throughput *in vitro* models that predict animal outcomes. We hypothesized that we could both prioritize combination therapies and quantify the predictive power of various *in vitro* models for drug development using a dataset of drug combination dose responses measured in multiple *in vitro* models. We systematically measured *M. tuberculosis* response to all 2- and 3-drug combinations among ten antibiotics in eight conditions that reproduce lesion microenvironments. Applying machine learning to this comprehensive dataset, we developed classifiers predictive of multidrug treatment outcome in a mouse model of disease relapse. We trained classifiers on multiple mouse models and identified ensembles of *in vitro* models that best describe *in vivo* treatment outcomes. Furthermore, we found that combination synergies are less important for predicting outcome than metrics of potency. Here, we map a path forward to rationally prioritize combinations for animal and clinical studies using systematic drug combination measurements with validated *in vitro* models. Our pipeline is generalizable to other difficult-to-treat diseases requiring combination therapies.

Introduction

Tuberculosis (TB), caused by infection with *Mycobacterium tuberculosis* (Mtb), remains a major global health issue. In 2019, an estimated ten million people fell ill with TB and about 1.4 million people died (*1*). Development of shorter treatment regimens is a key part of the third pillar of the WHO End TB Strategy (*2*). Multidrug treatment regimens were developed to treat active TB infections by shortening treatment duration, reducing disease relapse, and decreasing the development of antibiotic resistance (*3*). The standard TB treatment is six to nine months of multidrug treatment with an estimated 85% cure rate (*1, 4, 5*). The first two months of treatment (intensive, bactericidal phase) consist of four drugs (isoniazid, rifampicin, pyrazinamide, and ethambutol) that reduce sputum Mtb levels but are less effective against non-replicative bacilli (*3, 4, 6*). The following four to seven months of treatment (continuation phase) consist of two drugs (isoniazid and rifampicin) aimed at reducing disease relapse by treating persisting bacteria that survived the intensive phase (*3, 4, 6*). New regimens that can more efficiently treat Mtb are needed to shorten the intensive phase of treatment and reduce or eliminate the bacteria that persist and require continuation phase treatment (*4*).

Due, in large part, to the heterogeneity of TB lesions and treatment response among the Mtb population, combination therapy is required to treat active TB. Therapies should therefore be designed as combinations of antibiotics rather than single antibiotics alone. There are many drug options for new treatment regimens using existing drugs and drugs in development @(*7*), which creates an enormous number of possible drug combinations (*5*). Preliminary results from a Phase 3 clinical trial ("Study 31") demonstrated that treatment could be shortened using a novel combination of existing TB antibiotics (*8, 9*). Relatively new drugs that can target non-replicative bacteria (bedaquiline, pretomanid, delamanid, SQ019) (*10-12*) in combination with established drugs are in new, treatment-shorting regimens for multidrug resistant TB (MDR-TB) (*5, 13, 14*).

Treatment shortening potential in Phase 2b trials (*15, 16*)) led to the Phase 3 STAND clinical trial testing the use of pretomanid with pyrazinamide and moxifloxacin, PaMZ (Table 1) (*17*). Adding bedaquiline to PaMZ (BPaMZ, Table 1) in a Phase 2b trial (*18*) shortened culture conversion time of MDR-TB so dramatically that the STAND trial was put on permanent hold to start the Phase 3 SimpliciTB trial evaluating BPaMZ for treating both drug-sensitive TB and MDR-TB (*9, 19*). Together, these studies and the history of TB drug regimen design has demonstrated that there is treatment-shortening potential in the drug combination space. A critical step for developing new treatment regimens is to prioritize the thousands of drug combinations prior to clinical testing. However, it is not practical to evaluate thousands of combinations using the current preclinical regimen design pipeline, which combines *in vitro* and small animal studies. An efficient methodology is needed to systematically evaluate drug combinations and prioritize the thousands of multidrug combinations for their treatment-shorting potential.

Animal models are critical to regimen development, and mouse models are a primary tool in multidrug therapy design (*20-24*). Mouse strains where Mtb is primarily intracellular (e.g. BALB/c and C57BL/6) are the most widely used (*24*). Mouse strains that form mixed lesion types (e.g., C3HeB/FeJ) are used to study drug response because the disease pathology is more human-like, include granulomas with caseous necrotic cores (*21, 25, 26*). Mtb drug response differs between these two types of mouse models and both are important preclinical tools because the model-specific drug response is thought to result from the different lesion microenvironments present in each animal model (*24, 27-29*). Despite their utility for regimen development, comprehensive drug measurements in mice are not feasible. It is only practical to perform systematic drug combination studies *in vitro*, but *in vitro* studies do not clearly map to *in vivo* outcomes (*24, 30*). There are many *in vitro* models that mimic aspects of the host microenvironment encountered in the different

TB lesion types(*24, 30*). Many of these *in vitro* models are well suited for systematic drug combination studies, but none have been validated to prioritize drug combinations against preclinical animal models (*24, 30*).

We propose to realize the potential of drug combinations to improve treatment by developing a pipeline to map *in vitro* measurement of drug response to outcomes in mouse models. Here, we utilized the efficiency of an experimental design and analysis method called DiaMOND (diagonal measurement of n-way drug interactions) (*31*) to create a compendium of drug combination responses in Mtb using multiple *in vitro* models that were designed to reproduce aspects of the environments encountered in different lesion types. Applying machine learning to this comprehensive *in vitro* dataset, we identified signatures of drug potency and interaction that could predict whether combinations would outperform the standard of care. Classifiers based on these signatures also enabled us to establish a mapping between *in vitro* models and the different mouse models, which differ in lesion type (microenvironment) and outcome. Overall, our study establishes a logistical path to optimize combination therapies for TB using systematic measurement in validated *in vitro* growth models and computational modeling.

Results

Drug combination compendium construction

We developed a pipeline to efficiently prioritize drug combinations early in regimen development based on drug combination measurements from *in vitro* models. Using the DiaMOND methodology (*31*), we designed a compendium of drug combination measurements to survey informative drug-dose combinations (DiaMOND compendium). To compare *in vitro* data to

treatment responses in animal models, our DiaMOND compendium focused on (A) first- and second-line agents, for which there are abundant animal data and (B) measurements in *in vitro* growth conditions that model environments encountered during infection.

Mtb encounters a diversity of environmental niches during infection that influence response to drug treatment. We aimed to model drug response by aggregating measurements from a suite of *in vitro* models. We focused on modeling factors previously shown to influence Mtb growth and/or drug response, such as different carbon sources and abundance, low pH, low oxygen tension, and the intracellular environment (*21, 30, 32-40*). For this study, we developed or adapted eight *in vitro* models that were reproducible and scalable for systematic, high-throughput drug combination studies. We varied carbon source, with an emphasis on cholesterol and fatty acids, to model the lipid-rich environment in TB granulomas, using butyrate, valerate, cholesterol, and higher levels of cholesterol (cholesterol-high) as sole carbon sources. We used 7H9-based medium in order to compare against the most commonly utilized *in vitro* growth model with glycerol as a carbon source (standard). We also included *in vitro* models that mimic important factors encountered during infection: low pH (acidic), infection of J774 macrophages (intracellular), and developed a low-oxygen multi-stress model that induces dormancy using butyrate as a carbon source, sodium nitrate to respire (*41-43*), and plate seals to limit oxygen (dormancy). The doubling times varied considerably among the models, ranging from 16h to one week (Fig. 1A). We scaled the timing of the experiments relative to the doubling time of each model so that drug response measurements would not be biased by changes in growth rate (Table S1).

Drug combination dose response measurements

For the DiaMOND compendium, we selected ten antibiotics that are in first- and second-line TB treatment regimens and for which there are abundant *in vivo* (mouse) data (Table 1, Table S2).

These drugs include: cell wall synthesis inhibitors (ethambutol, isoniazid, and pretomanid), rifamycin transcriptional inhibitors (rifampicin, and rifapentine), protein synthesis inhibitor (linezolid), inhibitors of energy metabolism (bedaquiline, and clofazimine), DNA synthesis inhibitor (moxifloxacin), and the antimycobacterial agent pyrazinamide (Fig. 1A, Table S2). We treated the Mtb Erdman strain carrying an autoluminescent reporter and measured both optical density (OD_{600}) and luminescence at multiple time points after drug treatment. We observed a strong dependency in drug potency on *in vitro* model (Fig. 1A, inhibitory concentration to achieve 90% inhibition, IC_{90}. Table S3) consistent with the idea that drug efficacy is influenced by bacterial stress (*44*). We did not observe remarkable correlations in potency profiles by *in vitro* model. However, hierarchical clustering of drug potencies showed some groupings of drugs consistent with their target cell process (e.g., rifamycin transcriptional inhibitors group together, isoniazid and pretomanid - inhibitors of cell wall synthesis - group together). We also observed clustering of similar *in vitro* models. For example, growth media with short-chain fatty acids butyrate and valerate as the carbon source group together (Fig. 1A).

We observed condition-specific drug potencies consistent with previous reports, suggesting the models we adapted for high-throughput drug response measurements may be predictive of outcomes in animals. For example, activity of pyrazinamide in acidic and intracellular models and inactivity in the standard model was consistent with *in vitro* (*45, 46*) and animal studies (*47-49*). We also observed pyrazinamide activity with lipid carbon sources, which has not been previously reported. The rifamycins shared similar potency profiles with higher potency of rifapentine (Table S3), as previously described (*50*). Bedaquiline was more potent in medium with lipids as the carbon source compared to standard medium with sugars as previously described (*51*). Isoniazid potency was lower in the dormancy model, consistent with its inactivity towards non-replicating

bacilli (*52-54*) and previous studies showing decreased efficacy in the presence of nitrite (*41*). The wide range of single-drug responses and consistency with previous studies suggest that the *in vitro* models in this study produce non-redundant drug response data and form a validated set of conditions to model the lesion-specific variation in drug response.

Using these eight *in vitro* models, we constructed a compendium of systematic drug combination measurements by utilizing the efficiency of the DiaMOND method (Box). DiaMOND is a geometric optimization of the traditional checkerboard assay of drug-dose combinations. DiaMOND estimates the effect of combining drugs using a fraction of possible drug-dose combinations and focuses on the single drug and equipotent drug combination dose responses (*31*). We measured all 1-, 2-, and 3-drug combination dose responses (totaling 175 combinations) in at least biological duplicate (Fig. 1B), resulting in a compendium of over 51,000 individual dose response curves. To navigate this complex dataset, we focused our analysis on up to two time points per *in vitro* model. We chose the last time point (terminal, T) that is relative to the doubling rate (4 to 5 doublings for most models) and at a consistent treatment timepoint (constant, C) across *in vitro* models, ~7 days post treatment; Fig. 1A, Table S1. We also selected the measurement type that best benchmarks against colony forming units (OD_{600} for all models except intracellular and dormancy models, for which we used luminescence, Fig. S1). This selected dataset represents approximately one quarter of the total number of compendium dose responses.

We analyzed the single- and combination-drug treatments to derive potency and drug interaction information (see Box). With DiaMOND, we can quantify the degree and directionality of interactions at different levels of growth inhibition using common null models (e.g., Loewe additivity and Bliss independence). Drug combinations that are more, or less effective than expected based on single-drug behaviors are considered synergistic and antagonistic, respectively.

Drug interactions are quantified with fractional inhibitory concentrations (FICs) at different levels of growth inhibition (e.g., FIC_{50} and FIC_{90} are measured at the IC_{50} and IC_{90}, respectively). FIC measurements were log-transformed to represent synergistic and antagonistic combinations with negative and positive $\log_2(FIC)$ values, respectively. Drug interaction metrics based on Loewe additivity and Bliss independence were correlated (FIC_{50} and FIC_{90} for the constant and terminal time points, $r=0.81$, $p < 2.2x10^{-308}$, Pearson's correlation, Fig. S2); we therefore selected Loewe additivity as the null model for systematic analysis in the compendium (FIC_{50} and FIC_{90}). Dose response curves provide metrics of treatment potency at low dose (AUC_{25}; a normalized area under the curve to IC_{25}, see Box) or high dose (E_{inf}; the maximum achievable effect). To compare potency across models where Mtb have different growth properties, we calculated the maximum achievable inhibition of normalized growth rate (GR_{inf}; see Box) which allows direct comparison of treatment effects on cells with very different growth rates (*55*). Though many other metrics of drug response may be calculated from DiaMOND data, our analysis focused on these five metrics -- FIC_{50}, FIC_{90}, AUC_{25}, E_{inf}, GR_{inf} -- because they represent well-characterized and biologically interpretable aspects of drug interactions and potencies across low- and high-dose ranges.

Drug synergy is uncommon and does not distinguish effective combinations

To identify patterns in drug interactions, we clustered the compendium drug interactions at the terminal time point in all eight growth environments, using 90% growth inhibition ($\log_2(FIC_{90})$, Fig. 2A) and 50% growth inhibition ($\log_2(FIC_{50})$, Fig. S3). Clustering did not reveal obvious model-wide synergy for any combination. Instead, we observed that most drug interactions were antagonistic (70% of FIC_{90}s>0), consistent with a general trend towards antagonism in drug interactions observed in other organisms (*31, 56-60*) and cancer (*61*). The tendency towards antagonism was dependent on the growth model, with some conditions showing balance between

synergy and antagonism (intracellular and acidic), and others almost entirely antagonistic (cholesterol). Together these data suggest that synergy is a property of both drug and growth environment rather than an intrinsic property of drug.

To understand whether combinations that tend toward *in vitro* synergy are more effective *in vivo*, we compared selected combinations with differences in disease relapse from the most commonly used mouse strains (e.g., BALB/c, C56BL/6, Swiss). The relapsing mouse models (RMM) evaluate drug efficacy months after cessation of drug treatment, somewhat analogous to the clinical measurement of relapse (*62, 63*). We did not observe a rank-ordering by synergy in any growth condition that matched efficacy in the RMM; e.g., BPaL>MRZ>HRZ>RZ (Fig. 2B) (*22, 64-71*). Instead, we observed the 3-drug standard of care (HRZ) was the most synergistic drug combination, and BPaL was the most antagonistic among this subset (Fig. 2B). These examples suggest that drug interaction scores alone in the measured *in vitro* models were poor indicators of *in vivo* combination efficacy.

Synergistic drug combinations are not necessarily more effective than antagonistic combinations as maximum effect (or potency) of a combination can change independently of the drug interaction (*72*) (see Box). A tradeoff between synergy and potency appears to be important to consider when selecting effective drug combinations for treating other diseases (e.g., hepatitis C, HIV, and cancer), with potency often being more important than synergy (*73, 74*). To determine if maximum effect could be used to prioritize combinations from the DiaMOND compendium, we clustered the E_{inf} (a measure of maximum dose response effect, see Box) for all compendium drug combinations in all eight *in vitro* models at the terminal time point (Fig. 2C). We observed a high maximum effect (E_{inf}>0.9, Fig. 2C) in most combinations, consistent with the known anti-Mtb effects of the drugs. Dormancy and cholesterol-high models exhibited little variation in E_{inf}, suggesting that

neither condition had the dynamic range of maximum effect needed to discriminate among combinations or that all drug combinations are effective in these growth conditions for extended drug exposures. We compared E_{inf} profiles for the selected combinations we examined before, and we found that BPaL was more potent than HRZ or MRZ (Fig. 2D), consistent with animal outcomes of these regimens (*22, 64-71*). These examples suggest that maximum achievable effect *in vitro* may be a stronger predictor of outcomes in mouse models than synergy. As with E_{inf}, we observed correct rank ordering in some *in vitro* models by other potency metrics (AUC_{25} and GR_{inf}) (Fig. S4), though we identified no drug combinations in the DiaMOND compendium that were maximally potent across all eight models (Fig. S5). The correct ordering of selected drug combinations by mouse outcome suggests that the DiaMOND compendium contains information useful for identifying efficacious drug combinations.

DiaMOND metric signatures are predictive of treatment outcomes in the relapsing mouse model

We hypothesized that combinations of *in vitro* measurements could be compiled to model the *in vivo* microenvironments experienced by Mtb during drug treatment. We asked whether signatures of DiaMOND compendium measurements could distinguish drug combinations that were better than the standard of care in animal studies, HRZE or HRZ (Table 1). We classified 27 drug combinations that we measured in the compendium based on whether the treatment outcome in published RMM studies was better than standard of care (C1) or not (C0) (Table S4). Principal component analysis (PCA) demonstrated that linear combinations of *in vitro* features can separate C0 and C1 drug combinations (Fig. 3A, Wilcoxon rank-sum, $p<0.005$, Table S5). Inspection of feature contributions to the principal component (PC) that best separates C0 and C1 drug combinations revealed many features related to cholesterol, standard, and valerate growth models (Fig. 3B). We also observed that potency metrics (AUC_{25}, E_{inf}, GR_{inf}) are almost exclusively

represented in the top 20 contributing features (Fig. 3B). Together these results suggest that effective separation of C1 and C0 drug combinations requires measurement of drug combination potency in multiple growth environments.

To develop signatures of DiaMOND metrics that characterize C0 from C1 combinations, we trained binary classifiers with eight different machine learning (ML) methods to distinguish C0 and C1 drug combinations and compared their performance in 5-fold cross-validation (Table S6). We observed that nonlinear ensemble methods (Bayesian additive regression trees, random forest (RF), and gradient boosted trees) outperformed other ML algorithms, as measured by the area under the receiver operator characteristic (ROC) curve (AUC) and the F1 statistic, which is the harmonic mean of precision and recall (Table S6). We performed additional validation of the RF model by applying it to higher order (4- and 5-way) drug combinations commonly used in preclinical and clinical tests that were not considered during model training (Table S7). The RF model accurately predicted outcomes (Fig. 3C, AUC=1, F1=0.86) and exhibited performance similar to what was estimated in cross-validation. Overall, the performance of ML models demonstrates that there is a strong signal in the DiaMOND compendium that is predictive of RMM drug combination efficacy.

We observed that some of the *in vitro* models in the DiaMOND compendium are well-represented among the top ranked features in the classifying PCs (Fig. 3B), while other *in vitro* models are not present, suggesting that a subset of *in vitro* models may be sufficient to predict treatment outcome in the RMM. We asked whether classifiers using the DiaMOND compendium data from one *in vitro* model at a time were predictive of RMM outcome class. We observed that the data signal separating C0 and C1 drug combinations appeared in at least one PC for all eight *in vitro* models (Fig. 4A, Wilcoxon rank-sum test, $p<0.05$, Fig. S6). Furthermore, the five technically simpler

models to work with exhibited clear C0 and C1 separation (Fig. 4A, *in vitro* models cholesterol, butyrate, standard, valerate, Fig. S6 *in vitro* model acidic).

Though the single *in vitro* model classifiers were moderately predictive, they did not perform as well as the classifier trained using data from all eight *in vitro* models (Table S8). We asked whether another high-performing classifier could be derived using a subset of *in vitro* models. We systematically trained RF classifiers by considering all possible model combinations and observed that among the 255 possible combinations of *in vitro* models, 67 (26.3%) performed better than the classifier trained on all eight models. Furthermore, predictors including only the simpler *in vitro* models performed as well or better than those including the "complex" (intracellular, dormancy, cholesterol high) models (Fig. 4B, student's t-test, $p>0.05$). We further validated the highest performing classifiers trained on the simple *in vitro* models by applying them to the higher-order (4- and 5-way) drug combinations as well as drug combinations involving antibiotics (delamanid, sutezolid, and SQ109, Table S1) that were not included in the compendium's 10-drug set (Table S7). The high performance of classifiers on this validation set suggests that combining simple *in vitro* models computationally can produce classifiers that inform possible RMM outcomes (Fig. 4C). Additionally, the large number of classifiers that exhibit high accuracy and the shared DiaMOND compendium metrics among several *in vitro* models suggests that there may be multiple combinations of *in vitro* models that are predictive of outcomes in the RMM.

With many high performing RMM classifiers trained using subsets of the five simple *in vitro* models (Fig. 4C), we assessed whether the predicted RMM outcome for specific drug combinations would be consistent between these classifiers. The classifiers produce a probability that a drug combination belongs to each class (e.g., drug combination X belongs to C1 with 60% probability and C0 with 40% probability). The threshold probability is usually at 50% to assign

the classification but the probability can also be used to rank combination classification likelihood. We tabulated the predicted probabilities of outcome for all combinations in the compendium, as well as the higher-order and new drug combination validation set, using the top performing simple *in vitro* model classifiers shown in Fig. 4C. As we had previously observed, rank ordering the percent probabilities within each classifier shows high predictive performance when evaluating the validation set. Among all predictions made for the compendium and validation combinations, we noted that 36% of drug combinations had discordant predictions among the three classifiers. We did not observe a consistent pattern in which a classifier was discordant. We next tested whether a consensus prediction could be generated by simply averaging the probabilities of the top three classifiers. We observed that the discordant combinations were clustered in the second quartile (probability of C1 around 25-50%), suggesting that classifiers are most prone to error for combinations that are C0. This may be due to the mild class imbalance in the training set (11 C0 and 16 C1 combinations). The consensus prediction was highly accurate (84% of validation set and 93% overall). Incorrect consensus predictions were at the border between C0 and C1 at 42-47% C1, indicating that the misclassification was due to ambiguity near the 50% decision boundary, as opposed to strong classifier discordance. We conclude that a simple averaging of the probabilities generated by top classifiers is a practical means to construct an accurate consensus rank ordering for predicting drug combination response outcomes.

DiaMOND metrics describe efficacy of drug combination treatments in the C3HeB/FeJ mouse model

Given the success of ML classifiers to predict RMM outcomes, we next asked whether the DiaMOND compendium can be used to predict outcomes in other mouse models. Bactericidal activity in the most commonly used mouse strains (e.g., BALB/c, C56BL/6, Swiss) has been used

extensively to evaluate drug combination effectiveness. Bactericidal activity in these models (bactericidal mouse model, BMM, Table 1) measures the reduction in bacterial burden by drug treatment immediately following drug treatment and can be assessed more quickly than relapse. Using the same analysis pipeline, we trained ML classifiers to recognize C0 or C1 drug combinations for the BMM outcome (Table S9) but observed that the classifier performance was only mildly predictive (AUC = 0.67, F1 = 0.40) (Fig. S7). Additional analysis of *in vitro* model subsets identified many predictors with improved performance, but this improvement did not generalize to test data. Moderate model training performance and poor generalizability to new data suggest that the drug combination information needed for BMM outcome predictions may be difficult to capture with the *in vitro* models developed and used in this study.

The C3HeB/FeJ (HeB) mouse strain has become important for TB regimen development because the disease pathology is more similar to humans than other mouse strains (*26, 27, 75*). This includes the formation of caseous, necrotic granulomas that are characterized by low oxygen content (hypoxia) (*27, 75, 76*) and differential drug penetrance (*77, 78*). These lesions also contain large numbers of extracellular, non-replicating bacteria (*24, 76*). Like other mouse studies, those with HeB mice use bactericidal (BHeB) as well as relapse outcomes to determine drug effectiveness. Fewer drug combinations have been tested and published using HeB mice than other mouse strains. The DiaMOND compendium contained too few measured combinations to train ML classifiers. When we integrated the compendium combinations with higher-order drug combinations, we obtained a total of 16 combinations (Table S10) for the BHeB outcome, which was sufficient to train ML classifiers. However, we were not able to do the same for the relapse outcome, where we had four total combinations, even after augmenting with higher-order information.

To understand if DiaMOND metrics distinguish C0 and C1 BHeB combinations using this expanded training dataset, we evaluated class separation with PCA. We observed significant separation of BHeB outcome classes along the third PC (PC3) (Fig. 5A, $p<0.005$, Table S11). We then examined the top 10 features in PC3 by contribution (Fig. 5B) and found that the *in vitro* models and metrics were distinct from those we observed in the RMM analysis (Fig. 3B). Notably, the metrics for the BHeB were entirely drug interactions (Fig. 5B), and the presence of the dormancy model in the top ten features was of particular interest because we expected hypoxia-induced dormancy to be a microenvironment specific to the C3HeB/FeJ mice (*27, 75, 76*). Using the same approach described for RMM, we developed accurate RF models to classify BHeB C1 and C0 combinations (Fig 5C, all in vitro models, AUC=0.9, F1=0.80). Systematic evaluation of RF classifiers using all possible combinations of *in vitro* model subsets revealed that complex models did not improve performance (Fig. S8). Specifically, we found that models without dormancy perform as well as those with it (Fig. S9). As with the RMM classifiers, we identified *in vitro* model subsets that performed better than all models together trained for the BHeB outcome (37 (12.9%)). Lipid and acidic i*n vitro* models featured prominently among the most accurate classifiers. Together, these analyses demonstrate that the DiaMOND compendium data predicts outcomes in two pathologically distinct mouse models, suggesting that enough key information can be captured by simple *in vitro* models to help prioritize combination therapies for animal model tests.

Potency and antagonism are correlated with improved outcomes in mouse models

The signatures of DiaMOND data describing outcomes in RMM and BHeB highlighted that potency metrics were key predictors for RMM, while drug interactions were key for BHeB outcome classification. To understand whether C0 and C1 drug combinations showed significant

differences in these metrics, we examined the top five features from the most discriminatory PCs for both mouse models. Univariate analysis revealed significant differences between three of the top five features for the RMM outcome (Fig. 6A. Wilcoxon rank-sum, $p<0.05$). In each of these five features, potency was higher in C1 combinations compared to C0 combinations, which is consistent with expectations of increased potency for the most effective drug combinations. The top features describing BHeB outcomes are drug interactions (Fig. 6B), and three of the top five exhibited significant differences between C0 and C1 combinations (Wilcoxon rank-sum, $p<0.05$). For all five drug interaction features, C1 was the more antagonistic class. That antagonistic drug combinations may be more favorable is consistent with the results of our comparison of BPaL to the standard of care (HRZ, Fig. 2B). We found that different metric types (potency or interactions) may provide information that maps to different outcome types (bactericidal or relapse) in animal studies. Furthermore, our analysis suggests that high potency and antagonism in *in vitro* assays may be characteristics of favorable drug combinations.

Discussion

Our goal in this study was to develop a pipeline to efficiently prioritize drug combinations early in the TB regimen design process. Enthusiasm for use of *in vitro* models to design treatment regimens has been dampened because it has not been clear which *in vitro* conditions model the complex *in vivo* environment (*24, 30*). We hypothesized that treatment efficacy *in vivo* could be modeled as a "sum-of-parts" of the complex microenvironments. We, therefore, generated a dataset that profiles drug combination effects against Mtb in eight different *in vitro* growth environments. With this comprehensive drug combination data compendium, we identified signatures of potencies and drug interactions in specific *in vitro* models that distinguish whether

drug combinations are better than the standard of care in two important preclinical mouse models. We found that ML classifiers were accurate predictors of mouse disease relapse using data from only a few simple *in vitro* models. These classifiers were validated with higher order (4- and 5-drugs) combinations and had predictive power for combinations with drugs not included in the model training. Together, our study establishes a practical approach to prioritize combination therapies using economical, scalable, and expandable *in vitro* measurements.

Synergy is often assumed to be a property of optimized combination therapies because synergistic drugs are more effective together than expected based on single-drug efficacies alone. Our mapping of the DiaMOND compendium onto outcomes in two different mouse models challenges this notion. In the relapsing mouse model, drug interactions were not key features for classification; instead, the potency measures from the drug dose response curves were the most important predictors of outcome (Fig. 3B). Our findings are consistent with reports of treatment in hepatitis C, cancer, and HIV (*73, 74, 79*) that show a tradeoff between maximizing synergy and potency of a drug combination. Maximizing potency was often more important than synergy in treating these diseases with multidrug therapies (*73, 74, 79*). Antagonism was prevalent in our compendium (Fig. 2A), and we found that antagonism was characteristic of more efficacious drug combinations for the C3HeB/FeJ bactericidal model (Fig. 6B, C1 more antagonistic than C0). Partnering the most potent drugs together during regimen design may be generating highly potent combinations but biasing these combinations towards antagonistic drug interactions. Bedaquiline, pretomanid, and linezolid were recently found to be more potent in treating mice infected with the Mtb HN878 strain than the H37Rv strain (*80*). When combined, the drugs antagonized each other for treating Mtb strain HN878-infected mice. Despite this antagonism, the BPaL combination was highly effective at curing mice infected with either Mtb strain. These *in vivo* results are consistent

with our findings that BPaL is a highly potent but antagonistic drug combination for *in vitro* treatment of Mtb Erdman. One view of how drugs in combination exert their effect on cell populations is that each drug targets a different subpopulation rather than multiple drugs targeting the same cells (*73, 79*). Drug interactions would then explain how well a drug acts on the cellular population that was not susceptible to the other drugs in the multidrug treatment. This leads to the hypothesis that very potent drugs that alone can kill most of the cells in a population would achieve high maximum effect when combined but show no synergy or even antagonism. Study of the multidrug anticancer therapy R-CHOP (Table 1) supports this hypothesis (*73*) and an expanded study using more antibiotics could be used to test this hypothesis in tuberculosis. Our study suggests that for TB, potent drug combinations should be prioritized for further study and should not necessarily be deprioritized if they are antagonistic in *in vitro* assays.

Our approach enabled us to determine the relative importance of specific *in vitro* models to predict outcomes in mice, thereby serving to validate which growth conditions map to *in vivo* responses. We note that signatures including several *in vitro* models perform better than signatures using data from only one *in vitro* model; perhaps because the lesion microenvironment is complex and constitutes multiple stressors that affect Mtb drug response. It may also indicate that none of the physiological states imposed by the *in vitro* models we used in this study was so dominant that it drives Mtb *in vivo* drug treatment response. We chose several lipids to serve as carbon sources in our *in vitro* models because of the important role of lipids and cholesterol specifically for Mtb growth, survival, and infectivity (*81-90*). The cholesterol *in vitro* model was the top performing single *in vitro* model classifier for the RMM outcome and performed almost as well as the classifier with all *in vitro* models. This is consistent with the importance of cholesterol metabolism for Mtb survival and infectivity (*24, 85-87*). We also observed that other lipid-rich environments (modeled

by using short-chain fatty acids butyrate or valerate as carbon sources) induced distinct drug response patterns and that the best classifiers for both RMM and BHeB outcomes utilized metrics from multiple lipid-rich growth conditions. These findings suggest that it may be sufficient to measure drug combination responses with a suite of different simple growth environments to model the complex lipid environment encountered in TB lesions.

Mtb in the RMM mouse strains are thought to be primarily intracellular (*24*) and that intracellular residency has been shown to expose Mtb to the acidification of the phagolysosome (*36, 37*). We, therefore, expected the acidic growth environment to be a driver in classifiers for the RMM. We found that measurements from the acidic growth environment alone were not strongly predictive of outcomes in the RMM, but that these metrics were prominent in the best mixed-condition classifiers. These results indicate that response to acidic stress is important for Mtb survival to drug treatment *in vivo*, but that adaptation to other environments (such as lipid carbon sources) are important drivers of treatment response. We also observed that the acidic model was prominent among the best classifiers for the bactericidal outcome in the C3HeB/FeJ mouse strain (BHeB outcome). The C3HeB/FeJ mice are noted for the formation of the caseous necrotic granulomas (type II lesions, (*76*)) that have been shown to have a neutral pH (pH>7) (*91*) and with primarily extracellular Mtb (*76*). However, these animals have abundant intracellular bacteria in other lesion types and within macrophages that acidify the intracellular Mtb compartments (*77*) which may explain why acidic growth environments are important predictors of drug response in this mouse model.

The microenvironments in TB granulomas are complex, yet we were able to combine measurement in a "sum-of-parts" approach from relatively simple growth environments to model treatment outcomes (Fig. 4B, Figure S8). These results indicate that there is predictive drug combination

response information obtained from simple *in vitro* models that only needs to be combined correctly to predict drug treatment outcome in mice. The practical implication is that researchers can choose a subset of the most amenable *in vitro* models for performing drug combination experiments and still retain predictive capacity. We were able to successfully model RMM and BHeB outcomes using this compendium but not the BMM outcome, suggesting that there may be other important factors or emergent properties of complex environments encountered during infection that were not included in our measurement set. These factors may include other carbon sources, nutrient availability, iron limitation, oxygen tension, human serum, or other entry mechanisms into dormancy (e.g., via different lipids or other combinations of stressors) (*91-93*). Macrophage activation status has also been linked to Mtb drug susceptibility (*94*) and therefore models of cytokine-induced activation of macrophages may be more relevant to specific *in vivo* outcomes. Future studies measuring drug combination response in other and more complex *in vitro* models may permit accurate modeling of the BMM outcome and improve the accuracy of predictions of outcomes in RMM, BHeB.

We anticipate that our *in vitro*-to-*in vivo* pipeline for drug combination predictions may be applied to study treatment outcomes in other animals and clinical studies. Predictions using our RMM classifiers suggest the potential for using DiaMOND data to model responses in the clinic. The moxifloxacin containing regimens, HRZM and MRZE, were expected to shorten treatment time in humans by two months based on preclinical mouse studies but did not show non-inferiority to HRZE in the ReMOX clinical trial (*95*). A meta-analysis of mouse relapse outcome studies after the trial completion revealed that the treatment shortening of these combinations was expected to be smaller than initially thought and that perhaps could explain the ReMOX trial results (*62*). The consensus RMM classifier predictions from our study predicted that the MRZE and HRZM

combinations were not likely to show treatment improvement (47% and 53% chance of being C1, respectively), consistent with the meta-analysis and ReMOX trial outcome. Based on our prediction probabilities, MRZE and HRZM would not have been prioritized for further study compared to other combinations, like BPaMZ (83% chance of being C1), that are being evaluated in ongoing clinical trials (*96*). Another example of the potential utility of the DiaMOND compendium for clinical predictions comes from the "Study 31" clinical trial. The preliminary results from "Study 31" show treatment shortening of the continuation phase using PHZM compared with the standard treatment of HRZE (intensive) followed by HR (continuation) (*9*). A third treatment arm (PHZE) did not show improvement compared with the standard treatment. Similar to our ReMOX predictions, our consensus prediction indicated PHZE to have a low probability for treatment improvement over standard of care (42% chance of being C1) despite the mouse outcome indicating C1 classification. The ReMOX and "Study 31" examples suggest that the DiaMOND compendium contains information that, while contradictory to the mouse outcomes, is relevant to understanding clinical outcomes. Together these results suggest the DiaMOND compendium could be used in future modeling and predictions of clinical outcomes (*97, 98*).

Several changes to the experimental design may improve this pipeline. The importance of potency metrics in signatures of combination efficacy is perhaps surprising given that we design combination dose responses to have equipotent combinations of each drug. There is growing evidence that there is differential drug penetration into the lesions where Mtb is found (*78*) which would lead to non-equipotent levels of drug reaching Mtb cells. Utilizing pharmacokinetic data to design drug combinations may increase the utility and power of this approach and lead to a more predictive DiaMOND compendium dataset. The current standard of care and other new regimens

(e.g., "Study 31" and SimpliciTB) involve intensive and continuation phases of treatment. Including sequential treatments in an experimental approach could be used to understand how prior treatment sensitizes the bacterial population to future treatment regimens. One reason to use combination therapy for TB is to slow the acquisition of drug resistance. Systematic study of the drug combination space in different growth environments can also be used to investigate the evolution of drug resistance. For example, antagonistic drug interactions have been shown to suppress the evolution of drug resistance (*56, 73, 99, 100*) and the evolution of drug resistance can be tied to the growth rate and duration of drug exposure (*56, 80, 101, 102*). Finally, the depth of the DiaMOND compendium may be well-complemented with transcriptomic data of drug response to prioritize drug combinations based on predicted mechanisms of drug interaction (*103*).

TB is not the only disease that benefits from combination therapy. We expect that our pipeline may be adapted and applied to optimize multidrug regimens for other diseases including cancers, HIV/AIDs, and multi-drug resistant bacterial pathogens. Beyond our use of the TB DiaMOND compendium to describe combination efficacies in mouse models, we anticipate that this dataset may be used for other systematic studies of drug combination response.

Materials and Methods

Strains and media

M. tuberculosis Erdman strain was transformed with pMV306hsp+LuxG13 to generate an autoluminescent strain that was used for all experiments in this study (Addgene plasmid # 26161; http://n2t.net/addgene:26161; RRID:Addgene_26161) (*104*), see Supplemental Materials). We used the mouse cell line, J774 as a model of intracellular residency because J774 cells have been

used as a macrophage-like cell line to study early infection processes and Mtb drug response to complex host-like intracellular environment (*48, 105*).

Standard 7H9 Middlebrook medium supplemented with 0.2% glycerol, 10% OADC (0.5g/L oleic acid, 50g/L albumin, 20g/L dextrose and 0.04g/L catalase) and 0.05% Tween-80 with 25 µg/mL kanamycin was used for Mtb strain maintenance. Growth and culturing were performed at 37°C with aeration unless noted. All *in vitro* model media were buffered with 100 mM 3-(N-morpholino)propanesulfonic acid (MOPS, pH 7), unless noted, and filter-sterilized prior to use. The acidic model was based on the standard 7H9 Middlebrook media above and buffered with 100 mM 2-(N-morpholino)ethanesulfonic acid (MES) to pH 5.7. For acclimation to lipid carbon sources, a base medium consisting of 7H9 powder (4.7g/L), fatty acid-free BSA (0.5g/L), NaCl (100mM) and tyloxapol (0.05%) with 25 µg/mL kanamycin was used and the lipids sodium butyrate (5mM, final concentration), valeric acid (0.1% final concentration) or cholesterol (0.05mM or 0.2mM final concentration) were added to the base medium. For the cholesterol media, a cholesterol stock solution (100mM) was first prepared by dissolving cholesterol in a 1:1 (v/v) mixture of ethanol and tyloxapol and heated to 80°C for 30 minutes and added to pre-warmed (37°C) base medium (*32*). The dormancy media was based on the butyrate media with the addition of sodium nitrate (5mM) as a terminal electron acceptor (*38, 41-43*). J774 cells were cultured as previously described (*105*). Briefly, J774 cells were cultured in high glucose DMEM supplemented with 2mM L-glutamine, 1mM sodium pyruvate, and 10% heat-inactivated fetal bovine serum (FBS) at 37°C in 5% CO2. Media was changed every one-three days and cells passaged at ~80% confluence. Standard 7H10 Middlebrook agar plates supplemented with 0.5% glycerol, 10% OADC, 0.05% Tween-80 and 25 µg/mL kanamycin were used for enumerating colonies.

Mtb in vitro model acclimation

Mtb were inoculated into standard 7H9 Middlebrook medium, grown to mid-log phase (optical density, OD_{600} ~0.5-0.7) and were subcultured for less than two weeks prior to acclimation to assay medium. For acclimation to standard and acidic media, Mtb cells were diluted into assay media at a starting density of OD_{600} = 0.05, acclimated for 3-5 doubling times or until they reached mid-log phase (OD_{600} ~0.5-0.7), diluted to OD_{600} = 0.05 and grown back to mid-log phase before use in DiaMOND assays.

Similar to standard, and acidic conditions, Mtb were acclimated to butyrate, and valerate media and acclimated cells were frozen for use in assays. Frozen acclimated Mtb in butyrate and valerate media were inoculated into assay media, grown to mid-log phase (OD_{600} ~0.5-0.7), diluted into fresh lipid media at a starting concentration of OD_{600} = 0.05 and grown back to mid-log phase (OD_{600} ~0.5-0.7) and used for DiaMOND assays. The dormancy model used Mtb acclimated to butyrate medium grown to mid-log phase (OD_{600} ~0.5-0.7) and then diluted to a starting OD_{600} 0.05 in dormancy media. For the dormancy model (d), cells were incubated at 37°C without aeration for 28 days, which reduced autoluminescence close to media-only background levels, which we interpret as being dormant with very low metabolic activity.

Mtb growth on cholesterol media slowed without exchange of fresh medium. Cholesterol and cholesterol-high acclimation was similar to standard, and acidic conditions with fresh media exchanges every seven days to ensure continued growth. Mtb acclimated between 14 and 28 days were used for assays. Mtb growth rate on cholesterol-high was faster (four day doubling time) than cholesterol (seven day doubling time).

For the intracellular model, J774 cells were plated at 375,000 cells/mL in 384-well plates and cultured overnight, expecting ~one doubling prior to infection. Mtb grown to mid-log phase in

standard media was syringe-passed 8 times with a 25-gauge needle to reach a single-cell suspension, and J774s were infected with Mtb at MOI 2 for 24 hours followed by drug treatment for 5 days.

Drugs, dose responses, and dispensing

The drugs used in this study are listed in Table 1. All drugs were reconstituted and diluted in DMSO with the exception of PZA for the intracellular model; to avoid exceeding the DMSO limit (0.5%) in the intracellular condition, PZA was diluted in 1x phosphate-buffered saline with 0.01% Triton-X. Drugs were dispensed with an HP D300e digital dispenser and locations were randomized to reduce plate effects. For each *in vitro* model, the concentration to achieve 90% inhibition (IC_{90}) was determined. IC_{90} were used to design combination dose responses with equipotent mixtures of drugs (*31*). A ten-dose resolution with 1.5- or 2-fold dose spacing was used for all experiments.

Treatment and DiaMOND assays

Mtb were acclimated to *in vitro* model media prior to drug treatment as described above. For acidic, butyrate, cholesterol, cholesterol-high, standard and valerate models: 50µL of acclimated Mtb at the indicated density was added to each well in 384-well plates containing freshly dispensed drugs and incubated at 37°C in humidified bags to prevent evaporation. Edge wells contained media but were not used for assays. For the dormancy model: Mtb were acclimated as described above, gently resuspended and 20µL of dormant Mtb culture was transferred to each well on the assay plates. Plates were sealed with PCR seals to reduce oxygen exposure during drug treatment and incubated for seven days. We measured regrowth after drug treatment as a readout of drug effect during dormancy. Therefore, after drug treatment, plate seals were removed, 80µL of standard media was added to each well, and plates were incubated at 37°C in humidified bags to prevent evaporation.

For the intracellular model: drugs were printed into media-only plates and transferred onto infected J774 cells 24 hours after Mtb infection. To accommodate quality control assessment, we included multiple untreated and positive drug treatment controls in each plate as well as uninfected J774 cells for the intracellular model. (See Supplementary Materials).

Plate measurements

Luminescence and OD_{600} measurements were made at 3-5 time points per sample on a Synergy Neo2 Hybrid Multi-Mode Reader. Time points were based on the approximate doubling time of each model. To simplify the analysis, we generally compare time points at either a relatively similar time point (constant) or time ~4-5x doubling times after drug exposure (terminal time point). Constant and terminal time points correspond to the same set of measurements for the standard and intracellular *in vitro* models (constant/terminal, CT). For the dormancy model, plate readings were made during recovery in standard media and time points were selected based on doubling time in standard media. For the dormancy and intracellular models, OD_{600} measurements could not reflect Mtb biomass alone, so only luminescence measurements are used. Autoluminescence has been demonstrated as a proxy for Mtb cell growth (*104*) and viability in response to drug treatment (*106, 107*). In order to benchmark changes in luminescence to changes in growth in our conditions, we performed a series of drug treatment experiments in the dormancy and intracellular models (see Supplementary Materials and Fig. S1). Briefly, cells were treated as described above followed by plating treated cells on 7H10 plates with and incubating to enumerate colony forming units (CFU). Portions of the luminescence dose response curve that correlated with CFU changes were considered indicative of growth inhibition and metrics derived from these portions of the curve were used for analysis.

Data processing and metric calculation

Data processing and dose response metric calculation were performed using custom MATLAB scripts. In brief, raw data were background-subtracted using the median of media-only wells and normalized to the mean of untreated wells within each plate. For the intracellular model, uninfected macrophages provided the background (rather than media only) for subtraction from raw data and subsequently data was normalized to (infected) untreated within each plate. A 3-parameter Hill function was fit to each dose response (single drug or combination). Inhibitory concentrations (ICs) were calculated based on the Hill curve parameters. The area under the curve at 25% inhibition (AUC_{25}) was calculated using the integral of the fit curves from 0 to the 25% inhibitory concentration (IC_{25}) and normalized to the IC_{25}, allowing comparisons between drug combinations. Drug interactions scores were quantified by the fractional inhibitory concentration (FIC) using Loewe additivity and Bliss independence (See Box). FICs calculated by Loewe additivity and Bliss independence were well correlated, and neither model was observed to suffer from significant bias relative to the order of the drug combination (*108*) (Supplementary Material); therefore, we proceeded to analyze drug interactions based on Loewe additivity. The growth rate inhibition (GR) metrics were calculated as previously described (*55*). See Supplementary Materials for details on data processing and analysis.

Data quality

Experiments were performed in a minimum of biological triplicate. Comparisons of data between plates and between experimental days required data quality control assessment. Each dose response was assigned a quality score that takes into account the overall quality of the data from a plate, the quality of fit of the Hill function, the single drug dose responsiveness from an experiment and in the case of drug combinations, the equipotency in the drug combination dose responses (Supplementary Materials). In brief, plate data quality was assessed with a Z'-score using multiple

untreated (negative) and complete inhibition treatment (positive) wells in each plate. The fitting of the Hill function was assessed by coefficient of determination (R^2) of the fit as well as the closeness of the E_{inf} for each fit to the maximum observed effect for each dose response curve. Drug combination equipotency was assessed by comparing the proportional combinations normalized to their respective MICs and the idealized combination of drugs if they were perfectly equipotent. Dose responses with poor quality scores were excluded from analysis.

Computational Analyses

Biological replicate dose response and drug interaction data passing quality control were averaged. Means of replicate data were used for all downstream analyses unless noted. Hierarchical clustering was performed using cosine distance, and heatmaps with complete linkage dendrograms were generated using MATLAB. Other data preparation and visualizations were performed in R studio (version 3.5.3) using the tidyverse environment packages and ggplot2 and ggpubr packages for visualization. Data table import and export were performed in R using the openxl and readxls packages

PCA was performed in R using the prcomp function from the stats package with each feature scaled to have unit variance before PCA. Some features were missing data; e.g., FIC_{90} metrics were missing because single drugs didn't achieve IC_{90}. Features with more than 35% missing data points were excluded from PCA. The remaining missing values were imputed using the mean of the corresponding input features (mean imputation) (*109*).

Machine learning

The machine learning in R (mlr v2.17.0) package was used for all machine learning tasks involving projections of the original features onto the principal component (PC) space as the input features and classified drug combination outcome (C0 or C1) as labels.

Feature selection, feature number optimization, and model validation

The Kruskal-Wallis test was used to rank order the PC input features for ML based on the ability to discriminate outcome classes C0 and C1. As there were a limited number of drug combinations, we aimed to reduce the number of features used in the model. A Monte-Carlo resampling strategy was used to split the training data into 70/30% training/test partitions, to which we applied grid search to find the number of features that produced the largest test AUC. This feature number optimization was repeated five times for each training set and the smallest feature set from the five iterations was chosen as the final training feature set. Models were trained on the full set of training data and performance on new data was estimated using standard 5-fold cross validation. Validation was performed by projecting new data onto the PC space used for the model training and testing model classification performance.

Machine learner packages

Upon feature selection, machine learning algorithms were compared using standard 5-fold cross validation. The performance was evaluated using the AUC and the F-score (F1). The mlr package made possible on-demand loading of learners from other R packages, including Bayesian additive regression tree (bartMachine, v1.2.5.1), random forest (randomForestSRC, v2.9.3), extreme gradient boosting (xgboost, v1.1.1.1), logistic regression (stats), naive bayes (e1071, v1.7-3), support vector machine (e1071, v1.7-3), and weighted k-nearest neighbors (kknn, v1.3.1).

Statistical Analysis

Differences between outcome class groups for DiaMOND features or PCs were assessed by means (IC_{90} averages), medians (class comparisons), and standard deviation of drug combinations from each outcome group in each *in vitro* model. Because data normality could not easily be assessed with small numbers of drug combinations in each group, the Wilcoxon rank-sum test was used to compare outcome group means for statistical significance. Student's t-test were used for testing hypotheses of differences between model performance distributions. The hypothesis that Loewe and Bliss interaction (FIC) scores were correlated was tested using Pearson correlations. Statistical analyses were performed using the stats, ggpubr, or the rstatrix packages in R version 3.5.3.

Supplementary Materials

Materials and Methods

Fig. S1. Benchmarking cell viability with luminescence measurements.

Fig. S2. Comparing null reference models for drug interaction scoring.

Fig. S3. Intermediate potency drug interaction profiles.

Fig. S4. Alternative potency metric profiles for selected drug combinations.

Fig. S5. Alternative potency metric profiles for DiaMOND compendium.

Fig. S6. Outcome class separation in single *in vitro* model Principal components analyses (PCAs).

Fig. S7. BMM classifier performance on training and test data.

Fig. S8. BHeB *in vitro* model subset model performance distributions.

Fig. S9. BHeB *in vitro* model with and without dormancy model performance distributions.

Table S1. Experiment time points and estimated growth amounts for *in vitro* models.

Table S2. Drug information table.

Table S3. Drug IC_{90} for *in vitro* models.

Table S4. Drug combinations with RMM outcomes for 2- and 3-way drug combinations.

Table S5. RMM PC class separation.

Table S6. Machine learning algorithm benchmarking performance metrics.

Table S7. RMM model validation set.

Table S8. RMM single *in vitro* model classifier performance.

Table S9. Drug combinations with BMM outcomes for 1-, 2- and 3-way drug combinations.

Table S10. Drug combinations with BHeB outcomes for 1-, 2- and 3-way drug combinations.

Table S11. BHeB PC class separation.

References and Notes:

1. "Global tuberculosis report 2020," (World Health Organization, Geneva, 2020).https://apps.who.int/iris/handle/10665/336069
2. "The End TB Strategy," (World Health Organization, 2014).https://www.who.int/tb/strategy/End_TB_Strategy.pdf?ua=1
3. W. Fox, G. A. Ellard, D. A. Mitchison, Studies on the treatment of tuberculosis undertaken by the British Medical Research Council Tuberculosis Units, 1946-1986, with

relevant subsequent publications. *The International Journal of Tuberculosis and Lung Disease* **3**, S231-S279 (1999).

4. C. A. Kerantzas, W. R. Jacobs, Origins of Combination Therapy for Tuberculosis: Lessons for Future Antimicrobial Development and Application. *mBio* **8**, e01586-01516 (2017).

5. S. Tiberi, N. du Plessis, G. Walzl, M. J. Vjecha, M. Rao, F. Ntoumi, S. Mfinanga, N. Kapata, P. Mwaba, T. D. McHugh, G. Ippolito, G. B. Migliori, M. J. Maeurer, A. Zumla, Tuberculosis: progress and advances in development of new drugs, treatment regimens, and host-directed therapies. *The Lancet Infectious Diseases* **18**, e183-e198 (2018).

6. D. A. Mitchison, Modern methods for assessing the drugs used in the chemotherapy of mycobacterial disease. *Journal of Applied Bacteriology* **81**, 72S-80S (1996).

7. J. C. Evans, V. Mizrahi, Priming the tuberculosis drug pipeline: new antimycobacterial targets and agents. *Curr Opin Microbiol* **45**, 39-46 (2018).

8. S. E. Dorman, P. Nahid, E. V. Kurbatova, S. V. Goldberg, L. Bozeman, W. J. Burman, K. C. Chang, M. Chen, M. Cotton, K. E. Dooley, M. Engle, P. J. Feng, C. V. Fletcher, P. Ha, C. M. Heilig, J. L. Johnson, E. Lessem, B. Metchock, J. M. Miro, N. V. Nhung, A. C. Pettit, P. P. J. Phillips, A. T. Podany, A. E. Purfield, K. Robergeau, W. Samaneka, N. A. Scott, E. Sizemore, A. Vernon, M. Weiner, S. Swindells, R. E. Chaisson, High-dose rifapentine with or without moxifloxacin for shortening treatment of pulmonary tuberculosis: Study protocol for TBTC study 31/ACTG A5349 phase 3 clinical trial. *Contemp Clin Trials* **90**, 105938 (2020).

9. CDC, Media Statement: Landmark TB Trial Identifies Shorter-Course Treatment Regimen | CDC, 2020. https://www.cdc.gov/nchhstp/newsroom/2020/landmark-tb-trial-media-statement.html

10. A. Iacobino, G. Piccaro, F. Giannoni, A. Mustazzolu, L. Fattorini, Fighting tuberculosis by drugs targeting nonreplicating Mycobacterium tuberculosis bacilli. *International Journal of Mycobacteriology* **6**, 213-221 (2017).

11. Y. Liu, M. Matsumoto, H. Ishida, K. Ohguro, M. Yoshitake, R. Gupta, L. Geiter, J. Hafkin, Delamanid: From discovery to its use for pulmonary multidrug-resistant tuberculosis (MDR-TB). *Tuberculosis* **111**, 20-30 (2018).

12. J. van den Boogaard, G. S. Kibiki, E. R. Kisanga, M. J. Boeree, R. E. Aarnoutse, New Drugs against Tuberculosis: Problems, Progress, and Evaluation of Agents in Clinical Development. *Antimicrobial Agents and Chemotherapy* **53**, 849-862 (2009).

13. F. Conradie, A. H. Diacon, N. Ngubane, P. Howell, D. Everitt, A. M. Crook, C. M. Mendel, E. Egizi, J. Moreira, J. Timm, T. D. McHugh, G. H. Wills, A. Bateson, R. Hunt, C. Van Niekerk, M. Li, M. Olugbosi, M. Spigelman, Treatment of Highly Drug-Resistant Pulmonary Tuberculosis. *New England Journal of Medicine* **382**, 893-902 (2020).

14. E. Pontali, R. Centis, L. D'Ambrosio, F. Toscanini, G. B. Migliori, Recent evidence on delamanid use for rifampicin-resistant tuberculosis. *J Thorac Dis* **11**, S457-S460 (2019).

15. clinicaltrials.gov. National Library of Medicine. (2011). NCT01215851. Evaluation of Early Bactericidal Activity in Pulmonary Tuberculosis With(J-M-Pa-Z) (NC-001). https://ClinicalTrials.gov/show/NCT01215851

16. clinicaltrials.gov. National Library of Medicine. (2013). NCT01498419. Evaluation of 8 Weeks of Treatment With the Combination of Moxifloxacin, PA-824 and Pyrazinamide

in Patients With Drug Sensitive and Multi Drug-Resistant Pulmonary Tuberculosis (TB) (NC-002). https://ClinicalTrials.gov/show/NCT01498419

17. clinicaltrials.gov. National Library of Medicine. (2018). NCT02342886. Shortening Treatment by Advancing Novel Drugs. https://ClinicalTrials.gov/show/NCT02342886

18. clinicaltrials.gov. National Library of Medicine. (2016). NCT02193776. A Phase 2 to Evaluate the Efficacy, Safety and Tolerability of Combinations of Bedaquiline, Moxifloxacin, PA-824 and Pyrazinamide in Adult Subjects With Drug-Sensitive or Multi Drug-Resistant Pulmonary Tuberculosis. https://ClinicalTrials.gov/show/NCT02193776

19. clinicaltrials.gov. National Library of Medicine. (2020). NCT02410772. TBTC Study 31: Rifapentine-containing Tuberculosis Treatment Shortening Regimens. https://ClinicalTrials.gov/show/NCT02410772

20. K. E. Dooley, P. P. J. Phillips, P. Nahid, M. Hoelscher, Challenges in the clinical assessment of novel tuberculosis drugs. *Advanced Drug Delivery Reviews* **102**, 116-122 (2016).

21. T. Gumbo, A. J. Lenaerts, D. Hanna, K. Romero, E. Nuermberger, Nonclinical Models for Antituberculosis Drug Development: A Landscape Analysis. *The Journal of Infectious Diseases* **211**, S83-S95 (2015).

22. E. Nuermberger, S. Tyagi, R. Tasneen, K. N. Williams, D. Almeida, I. Rosenthal, J. H. Grosset, Powerful Bactericidal and Sterilizing Activity of a Regimen Containing PA-824, Moxifloxacin, and Pyrazinamide in a Murine Model of Tuberculosis. *Antimicrobial Agents and Chemotherapy* **52**, 1522-1524 (2008).

23. R. Tasneen, S. Y. Li, C. A. Peloquin, D. Taylor, K. N. Williams, K. Andries, K. E. Mdluli, E. L. Nuermberger, Sterilizing activity of novel TMC207- and PA-824-

containing regimens in a murine model of tuberculosis. *Antimicrob Agents Chemother* **55**, 5485-5492 (2011).

24. E. L. Nuermberger, Preclinical Efficacy Testing of New Drug Candidates. *Microbiology Spectrum* **5**, (2017).

25. A. Apt, I. Kramnik, Man and mouse TB: contradictions and solutions. *Tuberculosis (Edinb)* **89**, 195-198 (2009).

26. I. Kramnik, G. Beamer, Mouse models of human TB pathology: roles in the analysis of necrosis and the development of host-directed therapies. *Semin Immunopathol* **38**, 221-237 (2016).

27. E. R. Driver, G. J. Ryan, D. R. Hoff, S. M. Irwin, R. J. Basaraba, I. Kramnik, A. J. Lenaerts, Evaluation of a mouse model of necrotic granuloma formation using C3HeB/FeJ mice for testing of drugs against Mycobacterium tuberculosis. *Antimicrob Agents Chemother* **56**, 3181-3195 (2012).

28. J. P. Lanoix, A. J. Lenaerts, E. L. Nuermberger, Heterogeneous disease progression and treatment response in a C3HeB/FeJ mouse model of tuberculosis. *Dis Model Mech* **8**, 603-610 (2015).

29. A. Lenaerts, C. E. Barry, 3rd, V. Dartois, Heterogeneity in tuberculosis pathology, microenvironments and therapeutic responses. *Immunol Rev* **264**, 288-307 (2015).

30. T. Parish, In vitro drug discovery models for Mycobacterium tuberculosis relevant for host infection. *Expert Opin Drug Discov* **15**, 349-358 (2020).

31. M. Cokol, N. Kuru, E. Bicak, J. Larkins-Ford, B. B. Aldridge, Efficient measurement and factorization of high-order drug interactions in Mycobacterium tuberculosis. *Science Advances* **3**, e1701881 (2017).

32. W. Lee, B. C. VanderVen, R. J. Fahey, D. G. Russell, Intracellular Mycobacterium tuberculosis exploits host-derived fatty acids to limit metabolic stress. *J Biol Chem* **288**, 6788-6800 (2013).

33. J. V. Early, A. Casey, M. A. Martinez-Grau, I. C. Gonzalez Valcarcel, M. Vieth, J. Ollinger, M. A. Bailey, T. Alling, M. Files, Y. Ovechkina, T. Parish, Oxadiazoles Have Butyrate-Specific Conditional Activity against Mycobacterium tuberculosis. *Antimicrob Agents Chemother* **60**, 3608-3616 (2016).

34. K. Pethe, P. C. Sequeira, S. Agarwalla, K. Rhee, K. Kuhen, W. Y. Phong, V. Patel, D. Beer, J. R. Walker, J. Duraiswamy, J. Jiricek, T. H. Keller, A. Chatterjee, M. P. Tan, M. Ujjini, S. P. Rao, L. Camacho, P. Bifani, P. A. Mak, I. Ma, S. W. Barnes, Z. Chen, D. Plouffe, P. Thayalan, S. H. Ng, M. Au, B. H. Lee, B. H. Tan, S. Ravindran, M. Nanjundappa, X. Lin, A. Goh, S. B. Lakshminarayana, C. Shoen, M. Cynamon, B. Kreiswirth, V. Dartois, E. C. Peters, R. Glynne, S. Brenner, T. Dick, A chemical genetic screen in Mycobacterium tuberculosis identifies carbon-source-dependent growth inhibitors devoid of in vivo efficacy. *Nat Commun* **1**, 57 (2010).

35. V. Guerrini, B. Prideaux, L. Blanc, N. Bruiners, R. Arrigucci, S. Singh, H. P. Ho-Liang, H. Salamon, P. Y. Chen, K. Lakehal, S. Subbian, P. O'Brien, L. E. Via, C. E. Barry, 3rd, V. Dartois, M. L. Gennaro, Storage lipid studies in tuberculosis reveal that foam cell biogenesis is disease-specific. *PLoS Pathog* **14**, e1007223 (2018).

36. J. J. Baker, S. J. Dechow, R. B. Abramovitch, Acid Fasting: Modulation of Mycobacterium tuberculosis Metabolism at Acidic pH. *Trends Microbiol* **27**, 942-953 (2019).

37. O. H. Vandal, C. F. Nathan, S. Ehrt, Acid resistance in Mycobacterium tuberculosis. *J Bacteriol* **191**, 4714-4721 (2009).

38. B. Gold, C. Nathan, Targeting Phenotypically Tolerant Mycobacterium tuberculosis. *Microbiol Spectr* **5**, (2017).

39. C. de Miranda Silva, A. Hajihosseini, J. Myrick, J. Nole, A. Louie, S. Schmidt, G. L. Drusano, Effect of Moxifloxacin plus Pretomanid against Mycobacterium tuberculosis in Log Phase, Acid Phase, and Nonreplicating-Persister Phase in an In Vitro Assay. *Antimicrob Agents Chemother* **63**, (2019).

40. G. L. Drusano, S. Kim, M. Almoslem, S. Schmidt, D. Z. D'Argenio, J. Myrick, B. Duncanson, J. Nole, D. Brown, C. A. Peloquin, M. Neely, W. Yamada, A. Louie, The Funnel: a Screening Technique for Identifying Optimal Two-Drug Combination Chemotherapy Regimens. *Antimicrob Agents Chemother* **65**, (2021).

41. A. Cunningham-Bussel, F. C. Bange, C. F. Nathan, Nitrite impacts the survival of Mycobacterium tuberculosis in response to isoniazid and hydrogen peroxide. *Microbiologyopen* **2**, 901-911 (2013).

42. A. Cunningham-Bussel, T. Zhang, C. F. Nathan, Nitrite produced by Mycobacterium tuberculosis in human macrophages in physiologic oxygen impacts bacterial ATP consumption and gene expression. *Proc Natl Acad Sci U S A* **110**, E4256-E4265 (2013).

43. C. D. Sohaskey, Nitrate enhances the survival of Mycobacterium tuberculosis during inhibition of respiration. *Journal of bacteriology* **190**, 2981-2986 (2008).

44. D. F. Warner, V. Mizrahi, Tuberculosis chemotherapy: the influence of bacillary stress and damage response pathways on drug efficacy. *Clin Microbiol Rev* **19**, 558-570 (2006).

45. Y. Zhang, W. W. Yew, M. R. Barer, Targeting Persisters for Tuberculosis Control. *Antimicrobial Agents and Chemotherapy* **56**, 2223-2230 (2012).

46. E. A. Lamont, N. A. Dillon, A. D. Baughn, The Bewildering Antitubercular Action of Pyrazinamide. *Microbiol Mol Biol Rev* **84**, (2020).

47. E. A. Lamont, A. D. Baughn, Impact of the host environment on the antitubercular action of pyrazinamide. *EBioMedicine* **49**, 374-380 (2019).

48. D. Pires, E. Valente, M. F. Simoes, N. Carmo, B. Testa, L. Constantino, E. Anes, Esters of Pyrazinoic Acid Are Active against Pyrazinamide-Resistant Strains of Mycobacterium tuberculosis and Other Naturally Resistant Mycobacteria In Vitro and Ex Vivo within Macrophages. *Antimicrob Agents Chemother* **59**, 7693-7699 (2015).

49. K. Rohde, R. M. Yates, G. E. Purdy, D. G. Russell, Mycobacterium tuberculosis and the environment within the phagosome. *Immunol Rev* **219**, 37-54 (2007).

50. O. Alfarisi, W. A. Alghamdi, M. H. Al-Shaer, K. E. Dooley, C. A. Peloquin, Rifampin vs. rifapentine: what is the preferred rifamycin for tuberculosis? *Expert Rev Clin Pharmacol* **10**, 1027-1036 (2017).

51. A. Koul, L. Vranckx, N. Dhar, H. W. H. Göhlmann, E. Özdemir, J.-M. Neefs, M. Schulz, P. Lu, E. Mørtz, J. D. McKinney, K. Andries, D. Bald, Delayed bactericidal response of Mycobacterium tuberculosis to bedaquiline involves remodelling of bacterial metabolism. *Nature Communications* **5**, 3369 (2014).

52. Z. Xie, N. Siddiqi, E. J. Rubin, Differential antibiotic susceptibilities of starved Mycobacterium tuberculosis isolates. *Antimicrob Agents Chemother* **49**, 4778-4780 (2005).

53. J. C. Betts, P. T. Lukey, L. C. Robb, R. A. McAdam, K. Duncan, Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. *Mol Microbiol* **43**, 717-731 (2002).

54. L. G. Wayne, H. A. Sramek, Metronidazole is bactericidal to dormant cells of Mycobacterium tuberculosis. *Antimicrob Agents Chemother* **38**, 2054-2058 (1994).

55. M. Hafner, M. Niepel, M. Chung, P. K. Sorger, Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nature Methods* **13**, 521-527 (2016).

56. P. Yeh, A. I. Tschumi, R. Kishony, Functional classification of drugs by properties of their pairwise interactions. *Nature Genetics* **38**, 489-494 (2006).

57. M. Cokol, H. N. Chua, M. Tasan, B. Mutlu, Z. B. Weinstein, Y. Suzuki, M. E. Nergiz, M. Costanzo, A. Baryshnikova, G. Giaever, C. Nislow, C. L. Myers, B. J. Andrews, C. Boone, F. P. Roth, Systematic exploration of synergistic drug pairs. *Mol Syst Biol* **7**, 544 (2011).

58. S. Chandrasekaran, M. Cokol-Cakmak, N. Sahin, K. Yilancioglu, H. Kazan, J. J. Collins, M. Cokol, Chemogenomics and orthology-based design of antibiotic combination therapies. *Molecular systems biology* **12**, 872-872 (2016).

59. D. J. Mason, I. Stott, S. Ashenden, Z. B. Weinstein, I. Karakoc, S. Meral, N. Kuru, A. Bender, M. Cokol, Prediction of Antibiotic Interactions Using Descriptors Derived from Molecular Structure. *J Med Chem* **60**, 3902-3912 (2017).

60. M. Cokol, C. Li, S. Chandrasekaran, Chemogenomic model identifies synergistic drug combinations robust to the pathogen microenvironment. *PLOS Computational Biology* **14**, e1006677 (2019).

61. R. Richards, H. R. Schwartz, M. E. Honeywell, M. S. Stewart, P. Cruz-Gordillo, A. J. Joyce, B. D. Landry, M. J. Lee, Drug antagonism and single-agent dominance result from differences in death kinetics. *Nat Chem Biol* **16**, 791-800 (2020).

62. J.-P. Lanoix, R. E. Chaisson, E. L. Nuermberger, Shortening Tuberculosis Treatment With Fluoroquinolones: Lost in Translation? *Clinical Infectious Diseases* **62**, 484-490 (2015).

63. D. A. Mitchison, G. R. Davies, Assessment of the Efficacy of New Anti-Tuberculosis Drugs. *Open Infect Dis J* **2**, 59-76 (2008).

64. R. Tasneen, F. Betoudji, S. Tyagi, S. Y. Li, K. Williams, P. J. Converse, V. Dartois, T. Yang, C. M. Mendel, K. E. Mdluli, E. L. Nuermberger, Contribution of Oxazolidinones to the Efficacy of Novel Regimens Containing Bedaquiline and Pretomanid in a Mouse Model of Tuberculosis. *Antimicrob Agents Chemother* **60**, 270-277 (2016).

65. J. Xu, S. Y. Li, D. V. Almeida, R. Tasneen, K. Barnes-Boyle, P. J. Converse, A. M. Upton, K. Mdluli, N. Fotouhi, E. L. Nuermberger, Contribution of Pretomanid to Novel Regimens Containing Bedaquiline with either Linezolid or Moxifloxacin and Pyrazinamide in Murine Models of Tuberculosis. *Antimicrob Agents Chemother* **63**, (2019).

66. M. A. De Groote, J. C. Gilliland, C. L. Wells, E. J. Brooks, L. K. Woolhiser, V. Gruppo, C. A. Peloquin, I. M. Orme, A. J. Lenaerts, Comparative studies evaluating mouse models used for efficacy testing of experimental drugs against Mycobacterium tuberculosis. *Antimicrob Agents Chemother* **55**, 1237-1247 (2011).

67. E. L. Nuermberger, T. Yoshimatsu, S. Tyagi, R. J. O'Brien, A. N. Vernon, R. E. Chaisson, W. R. Bishai, J. H. Grosset, Moxifloxacin-containing regimen greatly reduces

time to culture conversion in murine tuberculosis. *Am J Respir Crit Care Med* **169**, 421-426 (2004).

68. E. L. Nuermberger, T. Yoshimatsu, S. Tyagi, K. Williams, I. Rosenthal, R. J. O'Brien, A. A. Vernon, R. E. Chaisson, W. R. Bishai, J. H. Grosset, Moxifloxacin-containing regimens of reduced duration produce a stable cure in murine tuberculosis. *Am J Respir Crit Care Med* **170**, 1131-1134 (2004).

69. I. M. Rosenthal, M. Zhang, K. N. Williams, C. A. Peloquin, S. Tyagi, A. A. Vernon, W. R. Bishai, R. E. Chaisson, J. H. Grosset, E. L. Nuermberger, Daily dosing of rifapentine cures tuberculosis in three months or less in the murine model. *PLoS Med* **4**, e344 (2007).

70. M. A. De Groote, V. Gruppo, L. K. Woolhiser, I. M. Orme, J. C. Gilliland, A. J. Lenaerts, Importance of confirming data on the in vivo efficacy of novel antibacterial drug regimens against various strains of Mycobacterium tuberculosis. *Antimicrob Agents Chemother* **56**, 731-738 (2012).

71. B. C. Mourik, G. J. de Knegt, A. Verbon, J. W. Mouton, H. I. Bax, J. E. M. de Steenwinkel, Assessment of Bactericidal Drug Activity and Treatment Outcome in a Mouse Tuberculosis Model Using a Clinical Beijing Strain. *Antimicrob Agents Chemother* **61**, (2017).

72. C. T. Meyer, D. J. Wooten, B. B. Paudel, J. Bauer, K. N. Hardeman, D. Westover, C. M. Lovly, L. A. Harris, D. R. Tyson, V. Quaranta, Quantifying Drug Combination Synergy along Potency and Efficacy Axes. *Cell Syst* **8**, 97-108 e116 (2019).

73. A. C. Palmer, C. Chidley, P. K. Sorger, A curative combination cancer therapy achieves high fractional cell killing through low cross-resistance and drug additivity. *eLife* **8**, e50036 (2019).

74. P. Sen, A. Saha, N. M. Dixit, You Cannot Have Your Synergy and Efficacy Too. *Trends Pharmacol Sci* **40**, 811-817 (2019).

75. J. Harper, C. Skerry, S. L. Davis, R. Tasneen, M. Weir, I. Kramnik, W. R. Bishai, M. G. Pomper, E. L. Nuermberger, S. K. Jain, Mouse model of necrotic tuberculosis granulomas develops hypoxic lesions. *J Infect Dis* **205**, 595-602 (2012).

76. S. M. Irwin, E. Driver, E. Lyon, C. Schrupp, G. Ryan, M. Gonzalez-Juarrero, R. J. Basaraba, E. L. Nuermberger, A. J. Lenaerts, Presence of multiple lesion types with vastly different microenvironments in C3HeB/FeJ mice following aerosol infection with Mycobacterium tuberculosis. *Disease Models & Mechanisms* **8**, 591-602 (2015).

77. S. M. Irwin, B. Prideaux, E. R. Lyon, M. D. Zimmerman, E. J. Brooks, C. A. Schrupp, C. Chen, M. J. Reichlen, B. C. Asay, M. I. Voskuil, E. L. Nuermberger, K. Andries, M. A. Lyons, V. Dartois, A. J. Lenaerts, Bedaquiline and Pyrazinamide Treatment Responses Are Affected by Pulmonary Lesion Heterogeneity in Mycobacterium tuberculosis Infected C3HeB/FeJ Mice. *ACS Infect Dis* **2**, 251-267 (2016).

78. V. Dartois, The path of anti-tuberculosis drugs: from blood to lesions to mycobacterial cells. *Nat Rev Microbiol* **12**, 159-167 (2014).

79. A. C. Palmer, P. K. Sorger, Combination Cancer Therapy Can Confer Benefit via Patient-to-Patient Variability without Drug Additivity or Synergy. *Cell* **171**, 1678-1691.e1613 (2017).

80. K. M. Bigelow, R. Tasneen, Y. S. Chang, K. E. Dooley, E. L. Nuermberger, Preserved Efficacy and Reduced Toxicity with Intermittent Linezolid Dosing in Combination with

Bedaquiline and Pretomanid in a Murine Tuberculosis Model. *Antimicrobial Agents and Chemotherapy* **64**, e01178-01120 (2020).

81. K. M. Wilburn, R. A. Fieweger, B. C. VanderVen, Cholesterol and fatty acids grease the wheels of Mycobacterium tuberculosis pathogenesis. *Pathog Dis* **76**, (2018).

82. G. Gago, L. Diacovich, H. Gramajo, Lipid metabolism and its implication in mycobacteria-host interaction. *Curr Opin Microbiol* **41**, 36-42 (2018).

83. J. E. Griffin, A. K. Pandey, S. A. Gilmore, V. Mizrahi, J. D. McKinney, C. R. Bertozzi, C. M. Sassetti, Cholesterol catabolism by Mycobacterium tuberculosis requires transcriptional and metabolic adaptations. *Chem Biol* **19**, 218-227 (2012).

84. D. G. Russell, B. C. VanderVen, W. Lee, R. B. Abramovitch, M. J. Kim, S. Homolka, S. Niemann, K. H. Rohde, Mycobacterium tuberculosis wears what it eats. *Cell Host Microbe* **8**, 68-76 (2010).

85. A. K. Pandey, C. M. Sassetti, Mycobacterial persistence requires the utilization of host cholesterol. *Proc Natl Acad Sci U S A* **105**, 4376-4380 (2008).

86. J. C. Chang, M. D. Miner, A. K. Pandey, W. P. Gill, N. S. Harik, C. M. Sassetti, D. R. Sherman, igr Genes and Mycobacterium tuberculosis cholesterol metabolism. *J Bacteriol* **191**, 5232-5239 (2009).

87. R. R. Lovewell, C. M. Sassetti, B. C. VanderVen, Chewing the fat: lipid metabolism and homeostasis during M. tuberculosis infection. *Curr Opin Microbiol* **29**, 30-36 (2016).

88. S. Puckett, C. Trujillo, H. Eoh, J. Marrero, J. Spencer, M. Jackson, D. Schnappinger, K. Rhee, S. Ehrt, Inactivation of fructose-1,6-bisphosphate aldolase prevents optimal co-catabolism of glycolytic and gluconeogenic carbon substrates in Mycobacterium tuberculosis. *PLoS Pathog* **10**, e1004144 (2014).

89. T. Beites, K. O'Brien, D. Tiwari, C. A. Engelhart, S. Walters, J. Andrews, H. J. Yang, M. L. Sutphen, D. M. Weiner, E. K. Dayao, M. Zimmerman, B. Prideaux, P. V. Desai, T. Masquelin, L. E. Via, V. Dartois, H. I. Boshoff, C. E. Barry, 3rd, S. Ehrt, D. Schnappinger, Plasticity of the Mycobacterium tuberculosis respiratory chain and its impact on tuberculosis drug development. *Nat Commun* **10**, 4970 (2019).

90. S. Ehrt, D. Schnappinger, Mycobacterium tuberculosis virulence: lipids inside and out. *Nat Med* **13**, 284-285 (2007).

91. J. P. Sarathy, V. Dartois, Caseum: a Niche for Mycobacterium tuberculosis Drug-Tolerant Persisters. *Clin Microbiol Rev* **33**, (2020).

92. M. Pai, M. A. Behr, D. Dowdy, K. Dheda, M. Divangahi, C. C. Boehme, A. Ginsberg, S. Swaminathan, M. Spigelman, H. Getahun, D. Menzies, M. Raviglione, Tuberculosis. *Nature Reviews Disease Primers* **2**, 16076 (2016).

93. J. P. Sarathy, L. E. Via, D. Weiner, L. Blanc, H. Boshoff, E. A. Eugenin, C. E. Barry, 3rd, V. A. Dartois, Extreme Drug Tolerance of Mycobacterium tuberculosis in Caseum. *Antimicrob Agents Chemother* **62**, (2018).

94. Y. Liu, S. Tan, L. Huang, R. B. Abramovitch, K. H. Rohde, M. D. Zimmerman, C. Chen, V. Dartois, B. C. VanderVen, D. G. Russell, Immune activation of the host cell induces drug tolerance in Mycobacterium tuberculosis both in vitro and in vivo. *J Exp Med* **213**, 809-825 (2016).

95. S. H. Gillespie, A. M. Crook, T. D. McHugh, C. M. Mendel, S. K. Meredith, S. R. Murray, F. Pappas, P. P. Phillips, A. J. Nunn, R. E. Consortium, Four-month moxifloxacin-based regimens for drug-sensitive tuberculosis. *N Engl J Med* **371**, 1577-1587 (2014).

96. clinicaltrials.gov. National Library of Medicine. (2020). NCT03338621. Trial to Evaluate the Efficacy, Safety and Tolerability of BPaMZ in Drug-Sensitive (DS-TB) Adult Patients and Drug-Resistant (DR-TB) Adult Patients. https://ClinicalTrials.gov/show/NCT03338621

97. K. E. Dooley, D. Hanna, V. Mave, K. Eisenach, R. M. Savic, Advancing the development of new tuberculosis treatment regimens: The essential role of translational and clinical pharmacology and microbiology. *PLoS Med* **16**, e1002842 (2019).

98. I. H. Bartelink, N. Zhang, R. J. Keizer, N. Strydom, P. J. Converse, K. E. Dooley, E. L. Nuermberger, R. M. Savic, New Paradigm for Translational Modeling to Predict Long-term Tuberculosis Treatment Response. *Clin Transl Sci* **10**, 366-379 (2017).

99. J. B. Michel, P. J. Yeh, R. Chait, R. C. Moellering, Jr., R. Kishony, Drug interactions modulate the potential for evolution of resistance. *Proc Natl Acad Sci U S A* **105**, 14918-14923 (2008).

100. A. R. M. Coates, Y. Hu, J. Holt, P. Yeh, Antibiotic combination therapy against resistant bacterial infections: synergy, rejuvenation and resistance reduction. *Expert Rev Anti Infect Ther* **18**, 5-15 (2020).

101. J. Liu, O. Gefen, I. Ronin, M. Bar-Meir, N. Q. Balaban, Effect of tolerance on the evolution of antibiotic resistance under drug combinations. *Science* **367**, 200-204 (2020).

102. B. K. Greenfield, S. Shaked, C. F. Marrs, P. Nelson, I. Raxter, C. Xi, T. E. McKone, O. Jolliet, Modeling the Emergence of Antibiotic Resistance in the Environment: an Analytical Solution for the Minimum Selection Concentration. *Antimicrob Agents Chemother* **62**, (2018).

103. S. Ma, S. Jaipalli, J. Larkins-Ford, J. Lohmiller, B. B. Aldridge, D. R. Sherman, S. Chandrasekaran, Transcriptomic Signatures Predict Regulators of Drug Synergy and Clinical Regimen Efficacy against Tuberculosis. *mBio* **10**, (2019).

104. N. Andreu, A. Zelmer, T. Fletcher, P. T. Elkington, T. H. Ward, J. Ripoll, T. Parish, G. J. Bancroft, U. Schaible, B. D. Robertson, S. Wiles, Optimisation of bioluminescent reporters for use with mycobacteria. *PLoS One* **5**, e10777 (2010).

105. S. A. Stanley, S. S. Grant, T. Kawate, N. Iwase, M. Shimizu, C. Wivagg, M. Silvis, E. Kazyanskaya, J. Aquadro, A. Golas, M. Fitzgerald, H. Dai, L. Zhang, D. T. Hung, Identification of novel inhibitors of M. tuberculosis growth using whole cell based high-throughput screening. *ACS Chem Biol* **7**, 1377-1384 (2012).

106. S. Sharma, E. Gelman, C. Narayan, D. Bhattacharjee, V. Achar, V. Humnabadkar, V. Balasubramanian, V. Ramachandran, N. Dhar, N. Dinesh, Simple and rapid method to determine antimycobacterial potency of compounds by using autoluminescent Mycobacterium tuberculosis. *Antimicrob Agents Chemother* **58**, 5801-5808 (2014).

107. T. Zhang, S.-Y. Li, E. L. Nuermberger, Autoluminescent Mycobacterium tuberculosis for Rapid, Real-Time, Non-Invasive Assessment of Drug and Vaccine Efficacy. *PLOS ONE* **7**, e29774 (2012).

108. D. Russ, R. Kishony, Additivity of inhibitory effects in multidrug combinations. *Nat Microbiol* **3**, 1339-1345 (2018).

109. S. Dray, J. Josse, Principal component analysis with missing values: a comparative survey of methods. *Plant Ecology* **216**, 657-667 (2014).

Acknowledgments: We thank members of the Aldridge laboratory, V. Dartois, R. Isberg, K. Mdluli, J. Mecsas, A. Palmer, J. Silverman, and S. Tan for insightful discussion. Bedaquiline and pretomanid were provided by the NIH AIDS Reagent Program and the TB Alliance, respectively. pMV306hsp+LuxG13 was a gift from Brian Robertson and Siouxsie Wiles. **Funding**: This work is funded by BMGF OPP1189457 to BA and NIH grant 1U54CA225088: Systems Pharmacology of Therapeutic and Adverse Responses to Immune Checkpoint and Small Molecule Drugs for AS. **Author contributions**: JLF, TG, NV, YD, and BBA conceived and designed the experiments. JLF, TG, NV, and YD performed the experiments. JLF, TG, NV, YD, MO, AS, and BBA conceived and designed the computational analysis. JLF, TG, NV, YD, and MO performed the computational analysis. The manuscript was written by JLF and BBA. All authors contributed to interpretation of the results and editing of the manuscript. **Competing interests**: All authors declare that they have no competing interests.

Figures:

Fig. 1.

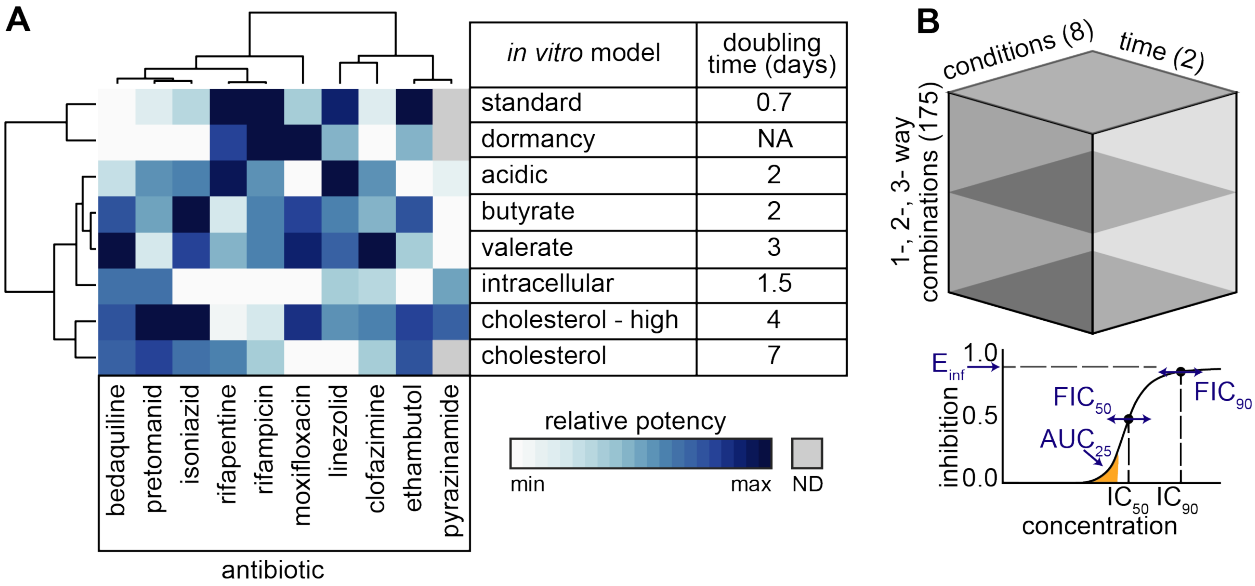


Fig. 1. 10-drug DiaMOND compendium of Mtb response to drug combination treatment. (A) Relative potencies of the ten compendium drugs in eight *in vitro* conditions (IC_{90}, terminal time point; left) with doubling times for each condition in untreated Mtb (right). Hierarchical clustering of potencies as calculated with cosine distances and average linkage. IC_{90} is color scaled (log10 transformation) within each drug (Table S3). ND = Not determined. NA = not applicable. (B) Metrics from DiaMOND dose response curves. IC_{50} and IC_{90} are used to calculate drug interactions at the 50% and 90% levels of growth inhibition (FIC_{50} and FIC_{90}, respectively). Three potency metrics are derived: AUC_{25} = normalized area under the curve until 25% inhibition, E_{inf} = theoretical maximum inhibition, and GR_{inf} = theoretical maximum normalized growth rate inhibition (Box and Materials and Methods). (C) Schematic data cube of the DiaMOND compendium. Mtb were treated with all 1-, 2-, and 3-way drug combinations (175 combinations) among 10-

drugs in dose responses measured in 10-dose resolution in at least biological duplicate. Dose response measurements were made in eight *in vitro* models and at 3-4 time points, but we focus on 1-2 time points for analysis; therefore, this data cube represents ~25% of the total measurements made.

Fig. 2.

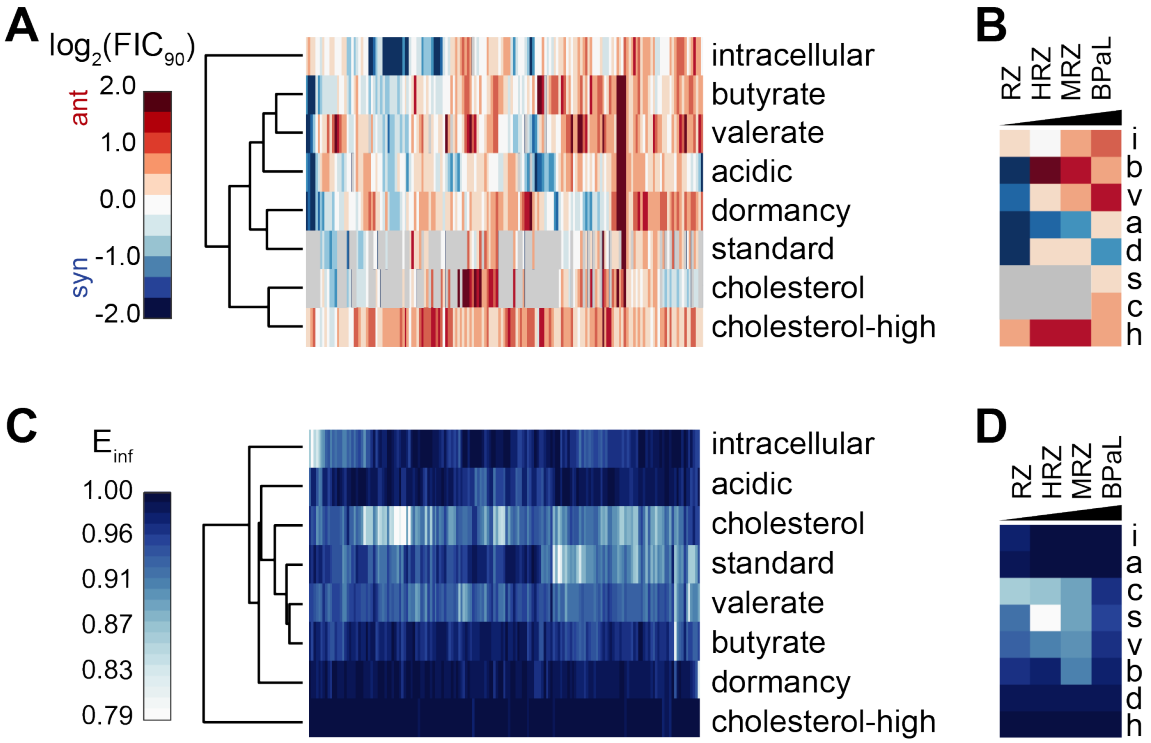
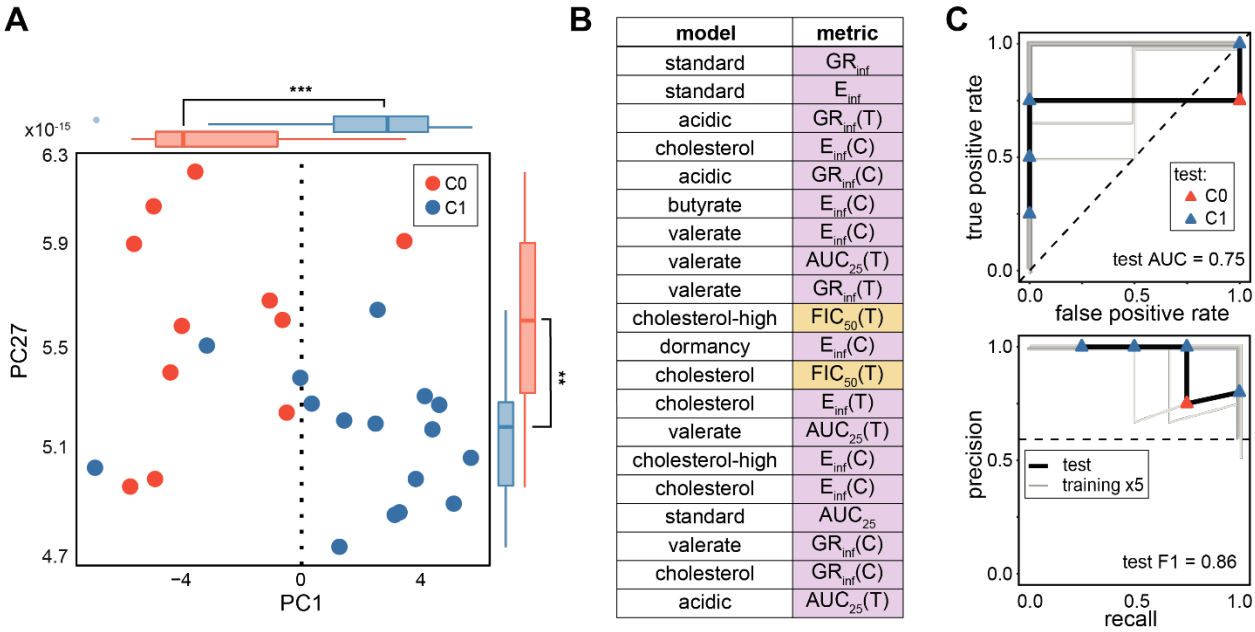


Fig. 2. Drug interaction and potency patterns in the DiaMOND compendium. **(A)** Drug interaction profiles of all 2- and 3-drug combinations among the ten compendium drugs across the *in vitro* models ($\log_2(FIC_{90})$ at the terminal time point, clustered based on cosine distance). **(B)** Drug interaction profiles of selected drug combinations ordered by mouse relapse outcome efficacy (*22, 64-71*). See Table 1 for drug combination abbreviations. **(C)** Drug combination potency profiles of all 2- and 3-drug combinations among the ten compendium drugs across the *in vitro* models (E_{inf} at the terminal time point, clustered based on cosine distance). **(D)** Drug interaction profiles of selected drug combinations ordered by mouse relapse outcome efficacy (*22, 64-71*). See Table 1 for drug combination abbreviations. gray = ND.

Fig. 3.



model	metric
standard	GR_{inf}
standard	E_{inf}
acidic	$GR_{inf}(T)$
cholesterol	$E_{inf}(C)$
acidic	$GR_{inf}(C)$
butyrate	$E_{inf}(C)$
valerate	$E_{inf}(C)$
valerate	$AUC_{25}(T)$
valerate	$GR_{inf}(T)$
cholesterol-high	$FIC_{50}(T)$
dormancy	$E_{inf}(C)$
cholesterol	$FIC_{50}(T)$
cholesterol	$E_{inf}(T)$
valerate	$AUC_{25}(T)$
cholesterol-high	$E_{inf}(C)$
cholesterol	$E_{inf}(C)$
standard	AUC_{25}
valerate	$GR_{inf}(C)$
cholesterol	$GR_{inf}(C)$
acidic	$AUC_{25}(T)$

Fig. 3. Prediction of combination treatment outcomes in the RMM with DiaMOND data. **(A)** PCA of DiaMOND data labeled by outcome in the RMM (C1 is better than the standard of care, blue; C0 is standard of care or worse, red) with the most discriminating two PCs shown. Outside the scatter plot are box and whisker plots of the distributions of C1 and C0 combinations along PC1 and PC27 (Wilcoxon rank-sum test: *** $p<0.005$. ** $p<0.01$). **(B)** Highest weighted features in PC1 with *in vitro* model (abbreviations in Fig. 1A) and metric type indicated. Metrics are classified and shaded according to whether they are related to drug combination potency (purple: AUC_{25}, E_{inf} and GR_{inf}) or drug interaction (orange: FIC_{50} and FIC_{90}). **(C)** ROC curves (top panel, Table 1) and PR curves (bottom panel, Table 1) of a random forest-based classifier trained on all eight conditions in the DiaMOND compendium. The model is tested with high-order combinations (4- and 5-drug combinations) that were excluded from training. Training (gray lines each show one of five cross validations; lines are slightly offset to aid

visualization) and test (black) performances are shown with lines. Test combinations are colored by outcome class as in (A). Performance metrics are shown on plots for test data (Area Under the ROC curve (AUC) and F1, harmonic mean of precision and recall, Table 1). Dashed lines indicate theoretical “no-skill” model performance.

Fig. 4.

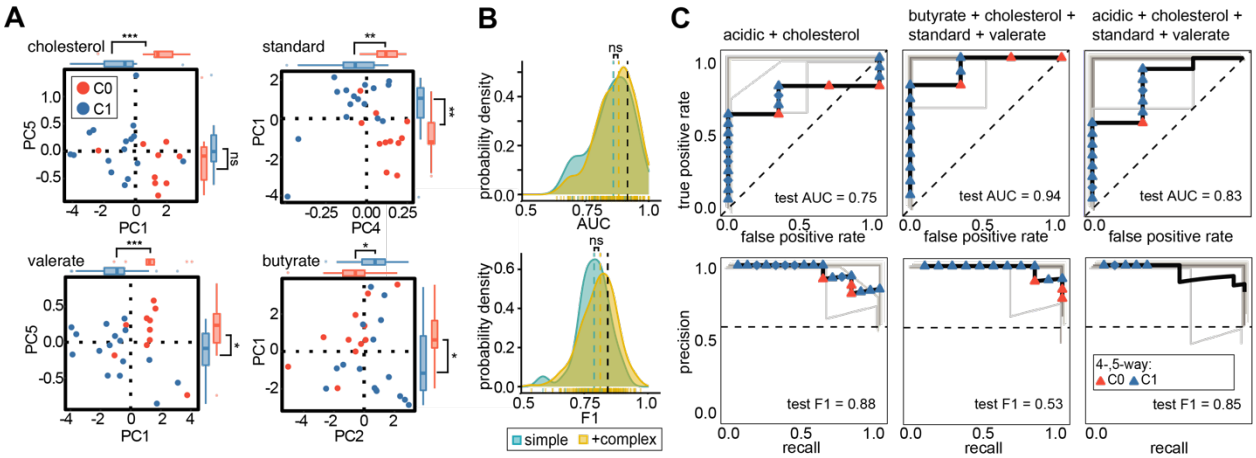


Fig. 4. Prediction of combination therapy outcomes in the RMM using fewer *in vitro* models. **(A)** PCA plot of DiaMOND data labeled by outcome in the RMM (plots are labeled as in Fig. 3A). Each subplot is DiaMOND data from one *in vitro* condition plotted in the PC space with the most discriminating two PCs shown for each model. Outside the scatter plot are box and whisker plots of the distributions of C1 and C0 combinations along PC1 and PC2 (Wilcoxon rank test: *** $p<0.005$. ** $p<0.01$. * $p<0.05$. ns $p>0.05$). **(B)** Density distribution plots of estimated classifier performances from systematic survey of all possible *in vitro* model subsets. Distributions of ROC AUC (top) and F1 (bottom) are separated based on whether technically complex models (intracellular, cholesterol-high, dormancy) are included (yellow) or whether only simple conditions (acidic, butyrate, cholesterol, standard, valerate) are considered. Colored dashed lines indicate mean value for distribution. The estimated performances when using all *in vitro* models (as in Fig. 3) is shown with black dashed lines. Distributions are compared with a Wilcoxon rank sum test (ns = not significant). **(C)** Comparison of classification performances of three high-performance random forest classifiers using subsets of simple *in vitro* models. Training

(gray lines each show one of five cross validations; lines are slightly offset when they are on top of each other) and test (black) performance is demonstrated with ROC (top) and PR (bottom) curves. Test combinations are colored by outcome class as in panel (A). Plot shapes indicate whether a test combination contained higher-order 4- and 5- drug combinations (triangle) or a combination containing a new drug (diamond) not included in the compendium described in Fig. 1. Dashed lines indicate theoretical “no-skill” model performance.

Fig. 5.

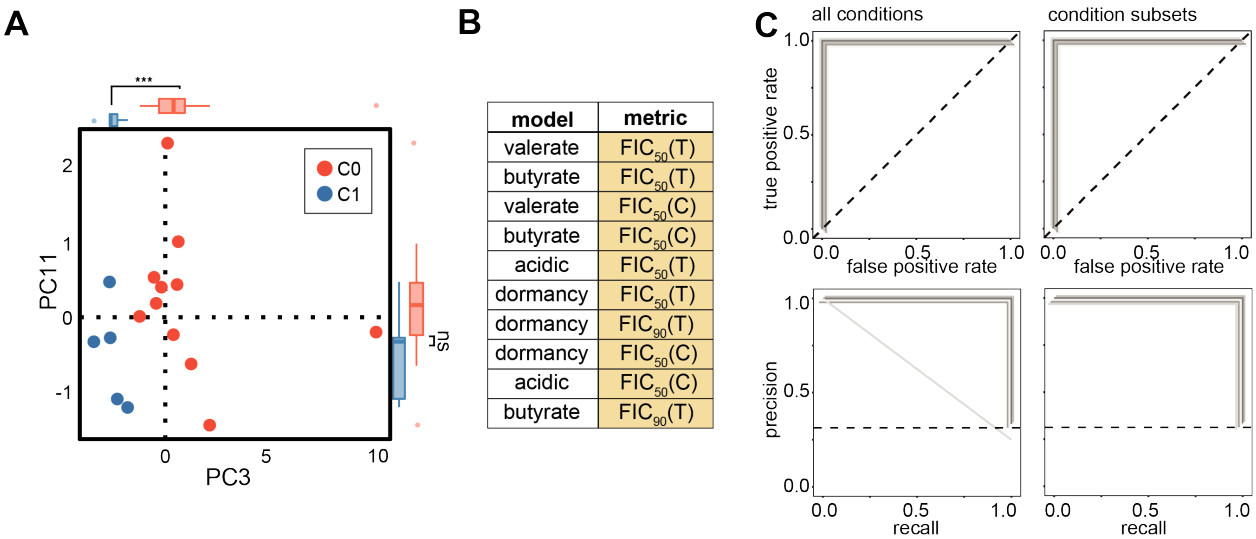


Fig. 5. Signatures of DiaMOND data to describe outcome in the C3HeB/FeJ (BHeB) mouse model. **(A)** PCA plot of DiaMOND data labeled by outcome in the BHeB (plot labels are as in Fig. 3A). **(B)** Highest weighted features in PC3 with *in vitro* model and metric type indicated (features are as described in Fig. 3B). **(C)** Machine learning performance plots for training with 5-fold cross validation (each in a gray line) with ROC (top) and PR (bottom) curves for models trained using all 8 conditions (left) and three high-performing subsets of conditions (right: acidic + butyrate + valerate, cholesterol + valerate, and standard + valerate). Subsets had perfect training performance (AUC = 1.0, Table S7). Dashed lines indicate theoretical "no-skill" model performance.

Fig. 6.

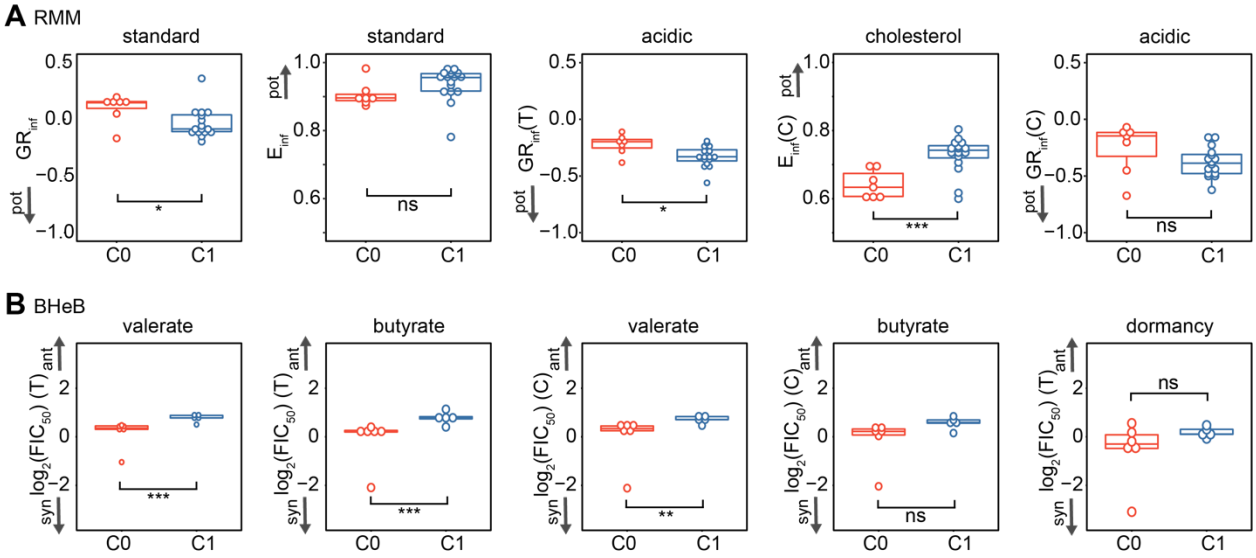
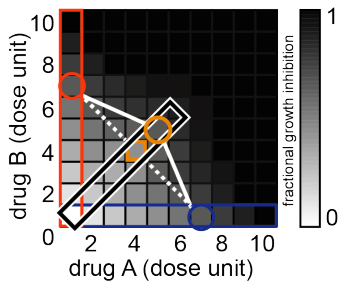
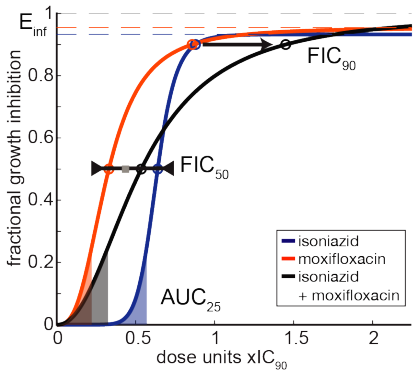


Fig. 6. Properties of signature potency and drug interaction characteristics that describe RMM and BHeB combination treatment outcomes. Values of the five highest weighted features in the most discriminatory PC are compared for C1 and C0 combinations in the RMM **(A)** and BHeB models **(B)** using dot and box plots. The top features in RMM are potency metrics whereas the top features are drug interaction metrics in BHeB. High vs. low potency (pot) and synergy (syn) vs antagonism (ant) is indicated with arrows on each subplot. (Wilcoxon rank test: *** $p<0.005$. ** $p<0.01$. * $p<0.05$. ns $p>0.05$).



DiaMOND (Diagonal measurement of n-way drug interactions) is a quantitative framework to efficiently measure drug interactions . The method is based on geometric sampling of traditional combination checkerboards and can be applied to any number of drugs in combination. Optical density (OD_{600}) or luminescence measures are normalized to untreated controls and subtracted from 1 to obtain fractional growth inhibition. The concentrations to achieve a particular effect (e.g., concentration to achieve 90% growth inhibition, IC_{90}, depicted in blue and orange circles) are experimentally determined for all single drugs so that dosing in subsequent measurement of drug combinations is equipotent (e.g., the IC_{90} should be dose #~7 for all drugs). Doses may be spaced linearly or logarithmically, but the spacing must be consistent between drugs. The single-drug dose responses (blue and orange boxes) and the equipotent drug combination dose response (black box) are highlighted. Drug interactions can be estimated using only the measurements from these boxes rather than the entire checkerboard by approximating the shape of isoboles (contours of equal effect). In the diagram, the isobole for IC_{75} is traced by the circles. If drug A and B are additive, the isobole would be a straight diagonal, and we calculate the expected IC_{75}on the combination dose response (orange square) where the dotted line intersects with the diagonal (combination) dose response curve. In this illustration, the combination reaches an IC_{75} at higher dose levels (orange circle) than the expected IC_{75}, indicating an antagonistic interaction. The ratio of observed and expected doses (observed/expected) is the fractional inhibitory concentration (FIC):



$$FIC = \frac{\text{observed combination dose}}{\text{expected combination dose}}$$

The DiaMOND methodology was used to obtain dose response data for every drug and drug combination measured over multiple time points. A Hill function was fit to these data and several potency and drug interactions metrics were derived from these dose response curves.

DiaMOND dose response metrics:

E_{inf} (the maximum achievable effect): derived from the fitted Hill function (lower pane, dashed lines, colored by single drug or drug combination), E_{inf} describes the maximal achievable effect (upper asymptote, dashed lines) of a given drug or drug combination at a particular time point, where the maximum possible effect is 1.

AUC_{25}: the area under the curve (AUC) simultaneously captures variation in potency and effect of a drug or drug combination, i.e., sensitivity to drug. AUC_{25}captures sensitivity to drug at concentrations with low growth inhibition. To compare low does potency to other drugs or drug combinations with different concentration ranges, we normalize the area by dividing by the IC_{25}. The resulting AUC_{25}values range from 0 (no effect) and 1 (potent).

FIC: drug interactions measure the effect of combining drugs on drug potency, i.e., the dose required to achieve a specific effect. The fractional inhibitory concentration (FIC) is the ratio of the observed combination dose (black circle) to achieve X effect over the expected combination dose (grey square), where FIC < 1 indicates synergy, FIC > 1 indicates antagonism, and FIC = 1 indicates additivity. In this example, the FIC_{50} is approximately additive whereas the FIC_{90} is antagonistic, which is indicated by the relative position of the combination dose response (black) near (IC_{50}) and to the right (IC_{90}) of the single dose response curves. We log transformation FICs to balance such that $\log_2 FIC < 0$ is synergistic and $\log_2 FIC > 0$ is antagonistic.

GR_{inf}: Derived from the growth rate curve (not shown here, see (55) for details), GRinf describes the maximal achievable effect of a drug or drug combination on the normalized growth rate, ranging between 1 and -1, where GR(c) is between 0 and 1 in the case of partial growth inhibition, GR(c) = 0 in the case of complete cytostasis, and GR < 0 indicates cell death. This unitless metric describes the effect of a drug on cells independent of doubling time, enabling comparison of drug effect on cells in different growth conditions.

Table 1.

	drug and drug combinations:
B	bedaquiline
C	clofazimine
E	ethambutol
H	isoniazid
L	linezolid
M	moxifloxacin
Pa	pretomanid
Z	pyrazinamide
R	rifampicin
P	rifapentine
D	delamanid
Su	sutezolid
Sq	SQ109
PaMZ	bedaquiline + pretomanid + moxifloxacin
BPaMZ	bedaquiline + pretomanid + moxifloxacin + pyrazinamide
HRZE	isoniazid + rifampicin + pyrazinamide + ethambutol - four drug standard of care
HRZ	isoniazid + rifampicin + pyrazinamide - three drug standard of care
BPaL	bedaquiline + pretomanid + linezolid

MRZ	moxifloxacin + rifampicin + pyrazinamide
RZ	rifampicin + pyrazinamide
HRZM	isoniazid + rifampicin + pyrazinamide + moxifloxacin – ReMOX trial combination
MRZE	moxifloxacin + rifampicin + pyrazinamide + ethambutol – ReMOX trial combination
PHZM	rifapentine + isoniazid + pyrazinamide + moxifloxacin – Study 31 trial combination
PHZE	rifapentine + isoniazid + pyrazinamide + ethambutol – Study 31 trial combination
R-CHOP	rituximab + cyclophosphamide + doxorubicin hydrochloride + vincristine sulfate + prednisone – anti-cancer drug combination
	treatment outcome classification:
C0	as good or worse than standard of care (HRZE or HRZ)
C1	better than standard of care
	mouse models:
RMM	relapsing mouse model[1]
BMM	bactericidal mouse model[2]
BHeB	bactericidal outcome in C3HeB/FeJ mouse strain[3]
	***in vitro* models:**

a	acidic
b	butyrate
c	cholesterol (0.05mM)
d	dormancy
h	cholesterol-high (0.2mM)
i	intracellular
s	standard
v	valerate
	data and model metrics:
C	constant time point
T	terminal time point
CT	constant and terminal time point are the same
IC_n	inhibitory concentration at n % growth inhibition
FIC_n	fractional inhibitory concentration at n % growth inhibition
AUC_{25}	normalized area under the curve to the 25% inhibition point
E_{inf}	effect at infinite drug concentration (maximum achievable effect)
GR_{inf}	normalized growth inhibition effect at infinite drug concentration (maximum achievable effect)
ROC	receiver operator characteristic
AUC	area under the ROC curve
PR	precision-recall
F1	harmonic mean of the precision and recall

Table 1: Abbreviations used in this study. Abbreviations along with brief descriptions are listed. [1]The RMM outcome assesses lasting cure months after cessation of drug treatment in the most commonly used mouse strains (e.g., BALB/c, C56BL/6, Swiss). [2]The BMM outcome assesses reduction of bacterial burden immediately following drug treatment in the most commonly used mouse strains. [3]The BHeB assesses reduction of bacterial burden immediately following drug treatment but in the pathologically distinct C3HeB/FeJ mouse strain.