TITLE PAGE

Title

A flux-based machine learning model to simulate the impact of pathogen metabolic heterogeneity on drug interactions

Authors

Carolina H. Chung,[1] Sriram Chandrasekaran[1–4]*

Affiliations

[1]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, 48109, USA.
[2]Program in Chemical Biology, University of Michigan, Ann Arbor, MI, 48109, USA.
[3]Center for Bioinformatics and Computational Medicine, Ann Arbor, MI, 48109, USA.
[4]Rogel Cancer Center, University of Michigan Medical School, Ann Arbor, MI, 48109, USA.
*Corresponding author: csriram@umich.edu

Classification

Original research report

Keywords

Antibiotic resistance, drug combinations, genome-scale metabolic modeling, machine learning, heterogeneity

Author contributions

Conceptualization: SC
Methodology: CHC, SC
Investigation: CHC, SC
Visualization: CHC
Supervision: SC
Writing—original draft: CHC
Writing—review & editing: CHC, SC

ABSTRACT

Drug combinations are a promising strategy to counter antibiotic resistance. However, current experimental and computational approaches do not account for the entire complexity involved in combination therapy design, such as the effect of pathogen metabolic heterogeneity, changes in the growth environment, drug treatment order and time interval. To address these limitations, we present a comprehensive approach that uses genome-scale metabolic modeling and machine learning to guide combination therapy design. Our mechanistic approach (a) accommodates diverse data types, (b) accounts for time- and order-specific interactions, and (c) accurately predicts drug interactions in various growth conditions and their robustness to pathogen metabolic heterogeneity. Our approach achieved high accuracy (AUC = 0.83 for synergy, AUC = 0.98 for antagonism) in predicting drug interactions for *E. coli* cultured in 57 metabolic conditions based on experimental validation. The entropy in bacterial metabolic response was predictive of combination therapy outcomes across time scales and growth conditions. Simulation of metabolic heterogeneity using population-FBA identified two sub-populations of *E. coli* cells defined by regulation of cis- to trans- isomerization of unsaturated fatty acids with distinct effects on drug interactions. Analysis of the vast landscape of condition-specific drug interactions revealed a set of 24 robustly synergistic drug combinations with potential for clinical use.

SIGNIFICANCE

Worldwide, 700,000 people die each year from drug-resistant infections. Drug combinations have great potential to reduce the spread of drug-resistant bacteria. However, their potency is impacted by both the pathogen growth environment and the heterogeneity in pathogen metabolism. The metabolic heterogeneity in a pathogen population allows them to survive antibiotic treatment. Here we present a flexible machine-learning framework that utilizes diverse data types to effectively search through the large design space of both sequential and simultaneous combination therapies across hundreds of simulated growth conditions and pathogen metabolic states. Our approach can serve as a useful guide for the selection of robustly synergistic drug combinations.

MAIN TEXT

Introduction

Antimicrobial resistance (AMR) occurs due to extended exposure to antibiotics, which allows bacteria to evolve resistance mechanisms that render antibiotic treatments ineffective[1]. In the context of AMR, bacterial metabolism plays a key role. Cell-to-cell variation in metabolism within a population can be beneficial in responding to antibiotic stress[2,3], and several pathogens take on a distinct metabolic state *in vivo* to tolerate antibiotics[4,5]. It is important to note that tolerant cells are predicted to be the source of drug-resistant pathogens[5–7]. In addition to stochasticity in metabolic activity within a population, extrinsic factors such as the metabolic environment also influence antibiotic efficacy[8–10]. For example, the availability of oxygen and extracellular metabolites modulate potency of antibiotics[9]. Metabolism can thus promote pathogen survival through adaptable use of nutrients in the local environment [11,12]. Bacterial metabolism also impacts susceptibility to antibiotics through the production of reactive oxygen species[13,14] or changes in membrane permeability[9]. Of note, these metabolic responses are also tied to entropy (i.e., disorder) in the bacterial stress response, which has been shown to be a generalizable predictor for antibiotic sensitivity[15]. Altogether, these individual findings suggest that modeling bacterial metabolism in response to antibiotics may be insightful for the design of novel treatments that mitigate resistance.

Combination therapy, which involves the use of two or more therapeutics, holds great potential for combating resistant pathogens as it not only leverages already regulated therapeutics[16], but also offers room for improved efficacy[17]. Further, combination therapy could be optimized to selectively target resistant pathogens via collateral sensitivity, which has been shown to overcome multidrug resistance in cancer[18,19]. Collateral sensitivity entails the increased sensitivity to a therapeutic that results from initial treatment with another stress agent[20]. This phenomenon has been observed across various diseases and organisms[18,21–23], and in context of AMR, could be leveraged to prevent and mitigate resistance[24]. Theoretical studies have also predicted that antibiotic combinations can be effective in heterogeneous populations and reduce the rise of resistance more effectively than monotherapy[24,25]. However, these studies do not provide a guideline to identify promising combinations among thousands of possible candidates. Combination therapies are traditionally identified using experimental methods; however, this approach quickly becomes infeasible when considering the vast combinatorial search space, the effects of the growth media, pathogen metabolic heterogeneity, and time-/order-dependence for treatment efficacy.

With the advent of high-throughput omics data and application of machine learning (ML), it is now possible to expedite the search for effective combination therapies. ML has also been applied to reveal mechanistic insights into antibiotic mechanisms of action[26,27] and identify novel antibacterial compounds[28,29]. In the past decade, several groups have used these methods to computationally design combination therapies in context of cancer[30–35] and AMR[36–38]. For the latter case, prior models have been shown to generate predictions that accurately correspond to experimental and clinical efficacy against *Escherichia coli* and *Mycobacterium tuberculosis*, thus offering effective reduction of the search space for combination therapies against AMR[36,38]. However, these approaches are limited by the availability of omics data measuring the bacterial response to antibiotic treatment. The combined drug effect on bacterial growth has also only been assessed in a limited number

of growth environments[37]. Moreover, current models have primarily focused on simultaneous combinations; consequently, the potential of designing time- and order-dependent combination therapies that promote collateral sensitivity remains unexplored. Since combination therapy is increasingly used to treat many medical conditions such as tuberculosis (TB), gram negative-, and biofilm associated-infections[39–42], it is essential to consider how various metabolic factors (e.g., cell-to-cell heterogeneity, growth environment) influence the efficacy of different drug combinations. Computational tools are hence necessary to identify antibiotic treatments that are robustly efficacious across heterogeneous environments[43,44].

To address these limitations, we present an approach that integrates genome-scale metabolic models (GEMs) into ML model development to determine effective combination therapies. Using GEMs allows us to integrate diverse data types and account for different pathogen metabolic states and growth conditions. GEMs are computational models built from gene-protein-reaction associations of metabolic genes present in the genome of an organism[45]. Additionally, they include annotation of traditional antibiotic targets such as cell wall synthesis, DNA replication, and RNA transcription. Model constraints, such as from omics data or nutrient availability, can be imposed to simulate bacterial metabolism in response to different perturbations[46,47]. Our approach using GEMs and ML provides a systems-level perspective of the bacterial response to antibiotic treatment in condition-specific cases. This is critical for designing efficacious combination therapies, since experimentally measured susceptibility to antibiotics may not always translate into efficacy *in vivo*. We further extend our approach to predict outcomes for sequential combination therapies, which can be designed into cyclic antibiotic regimens that mitigate resistance[24]. Finally, we showcase how our models reveal mechanistic insights that explain treatment potency and can be leveraged to finetune data-driven combination therapy design.

Results

The CARAMeL approach for combination therapy design

Our approach, called *Condition-specific Antibiotic Regimen Assessment using Mechanistic Learning* (*CARAMeL*), involves a two-step modeling process: (a) simulating metabolic flux data using GEMs and (b) developing a ML model to predict combination therapy outcomes using flux from GEMs. For the first part, omics data and metabolite composition of the extracellular environment serve as GEM inputs to determine flux profiles in response to drug treatment and growth in defined media, respectively (***Fig. 1A***). For the second part, GEM-derived flux profiles and drug interaction data serve as inputs to train a ML model that predicts interaction outcomes for novel drug combinations (***Fig. 1B***). We developed ML models predictive of combination therapy outcomes for *E. coli* and *M. tb* using the Random Forests algorithm. We specifically chose this ML method as it can handle small datasets and determine feature importance, i.e., how much each feature contributes to the accuracy in model predictions. The feature importance can reveal mechanistic insights into the factors driving combination therapy outcomes.

We determined metabolic flux profiles in response to drug treatment and condition-specific growth by constraining the *E. coli* GEM iJO1366[48] and the *M. tb* GEM iEK1011[49]. For drug flux profiles, we imposed chemogenomic data for *E. coli*[50] and transcriptomic data for *M. tb*[38] as GEM constraints. Briefly, chemogenomic data measures single-gene knockout (KO) fitness while transcriptomics data measures genome-wide expression of genes. By selecting genes for which there was differential fitness or expression in response to a specific

157 treatment, we could infer a set of differentially regulated genes for individual drugs. For
158 transcriptomic data, positive and negative differential expression directly corresponded with
159 up- and down-regulation, respectively. For chemogenomic data, we surmised based on the
160 cost-benefit gene expression model[51] that gene KOs resulting in low fitness were up-
161 regulated while gene KOs that enhanced fitness were down-regulated. Direct comparison
162 of flux profiles simulated from a chemogenomic-based approach against flux profiles
163 simulated with transcriptomics and proteomics data confirmed that these assumptions were
164 valid (***Fig. S1***)[52–54]. To determine growth media flux profiles, the availability of metabolites
165 within a media condition was used to constrain the GEMs. Specifically, we modified the
166 uptake rate for exchange reactions providing key metabolites (e.g., glycerol exchange for
167 M9 glycerol media) to allow cellular intake (see ***Methods*** for further details).

168 Prior to ML model development, we processed drug and media flux profiles to determine
169 joint profiles for all combinations of interest. Joint profiles are comprised of four pieces of
170 information: (a) the combined effect of all treatments (i.e., sigma scores), (b) the unique
171 effect of individual treatments (i.e., delta scores), (c) the overall metabolic entropy (i.e.,
172 entropy scores), and (d) time interval (relevant for time- and order-dependent
173 combinations). To determine sigma and delta scores, we adapted a strategy previously used
174 for creating joint chemogenomic profiles[36,37]. Specifically, we binarized drug and media
175 flux profiles based on differential flux activity in comparison to baseline (i.e., GEM
176 simulation without additional constraints). Sigma scores were defined as the union of
177 binarized flux profiles for all treatments involved in a combination. Delta scores were
178 defined as the symmetric difference between flux profiles (see ***Methods*** for details). To
179 account for metabolic entropy, we first calculated entropy as defined by Zhu *et al.*[15] for each
180 drug and media flux profile. We then defined entropy scores as the mean and sum of entropy
181 among all treatments involved in a combination. Finally, the time feature was defined as the
182 time interval between the first and last treatments for a combination (see ***Methods*** and ***Fig.***
183 ***S2*** for further details).

184 Using feature (i.e., joint profiles) and outcome (i.e., interaction scores, IS) information for
185 a set of drug combinations, we trained ML models to associate feature patterns to drug
186 combination outcomes. Next, we used the trained ML models to predict outcomes for new
187 drug combinations based on their feature information alone. We then compared our
188 predictions against experimental data by calculating the Spearman correlation. We also
189 assessed model performance by calculating the area under the receiver operating curve
190 (AUROC) for both synergy and antagonism. High and positive values for both metrics
191 indicate that model predictions correspond well with actual drug interaction outcomes.

192 ***CARAMeL predicts drug interactions with high accuracy***
193 We benchmarked CARAMeL against previous approaches by directly comparing our
194 prediction accuracy against those reported in literature and those re-calculated using omics
195 data directly instead of using flux data. For these comparisons, we trained ML models and
196 evaluated their performance for five different cases:

197 1. Predicting novel pairwise drug interaction outcomes for *E. coli*[36]
198 2. Predicting novel three-way drug interaction outcomes for *E. coli*[37]
199 3. Predicting pairwise drug interaction outcomes for *E. coli* cultured in a novel nutrient
200 condition (M9 glycerol media)[37]
201 4. Predicting novel pairwise and three-way interaction outcomes for *M. tb*[38]

5. Predicting interaction outcomes for multi-drug TB regimens used in clinical trials[55]

Of note, the first, second, and fourth cases tested the model's ability to predict unseen combinations involving test drugs with new mechanisms of action. The third case assessed whether the model could predict drug interaction outcomes in a new growth environment, while the fifth case ascertained if predicted outcomes corresponded with clinical efficacy. ***Fig. 2*** summarizes our findings for all analyses listed above. For all these studies, the same train-test datasets were used for evaluating CARAMeL against the original methods to ensure direct comparison. The same thresholds for synergy and antagonism defined in the original studies were also used in all these comparisons. Further discussion on ML model development and results, including the specific train-test allocation of interaction data reported in literature for each case, is provided below.

For case 1, we used drug interaction data previously measured for 171 pairwise combinations involving 19 drugs that cover a diverse set of targets[36] (***Table S1***). Out of this total, 105 interactions involving 15 drugs were used for model training and the remaining 66 interactions, which involved four new drugs that introduced new mechanisms of action (e.g., RNA synthesis), were used for model validation. The CARAMeL model yielded significant correlations between experimental and predicted scores ($R = 0.71$, $p \sim 10^{-11}$, ***Fig. S3A***). Model predictions also yielded high AUROC values for classifying synergy ($IS < -0.5$, AUROC = 0.84) and antagonism ($IS > 2$, AUROC = 0.89) (***Fig. S3B***) based on thresholds defined in the original study. Of note, these results were considerably better than those reported in literature ($R = 0.52$)[36] and those re-calculated using the omic-based approach ($R = 0.63$) (***Fig. 2A***).

For case 2, we re-trained the CARAMeL model using 171 pairwise interactions to predict 56 three-way combinations involving eight antibiotics[37] (***Table S1***). Our model generated accurate predictions ($R = 0.62$, $p \sim 10^{-7}$, ***Fig. S3C***) and notably identified synergistic interactions ($IS < -0.2$, AUROC = 0.95, ***Fig. S3D***) with higher accuracy than the omics-based approach (AUROC = 0.76, ***Fig. 2B***).

For case 3, the CARAMeL model was once again re-trained with the 171 pairwise interactions and additional pairwise data measured for *E. coli* cultured in M9 glucose and lysogeny broth (LB) media. We then applied our model to predict 55 pairwise interaction outcomes for *E. coli* cultured in M9 glycerol media. Our model yielded results comparable to those from literature[37] and re-determined using omics data across all three performance measures (***Figs. 2, S3E, and S3F***).

For case 4, we trained a CARAMeL model using combination data for *M. tb* treated with 196 pairwise to five-way interactions involving 40 drugs[38] (***Table S2***). We then used data for 36 unseen interactions for model validation. The CARAMeL model yielded predictions that significantly correlated with experimental data ($R = 0.55$, $p \sim 10^{-4}$, ***Fig. S4A***) and performed well in classifying synergistic ($IS < 0.9$, AUROC = 0.81) and antagonistic ($IS > 1.1$, AUROC = 0.83) interactions (***Fig. S4B***). Though the CARAMeL-based correlation is slightly lower than that reported in literature[38] ($R = 0.63$), our model classified both synergistic and antagonistic interactions with high accuracies that are comparable to a model trained on omics data (***Fig. 2***).

For case 5, we used the same CARAMeL model from case 4 to predict interaction outcomes for 57 multi-drug TB regimens involving nine drugs prescribed in separate clinical trials[55] (***Table S2***). Of note, interaction outcomes for this dataset measured regimen efficacy based on sputum clearance after two months of treatment. We found that model predictions were significantly correlated ($R = 0.56$, $p \sim 10^{-6}$, ***Fig. S4C***) with sputum clearance, and that model predictions classified as synergistic ($IS < 0.9$) captured most of the efficacious treatments (sputum clearance > 80%) amongst all 57 TB regimens (***Fig. S4D***). These results were comparable to both literature-[38] and omic-based results across all three performance measurements (***Fig. 2***). Overall, we found that our approach retained high accuracies in predicting combination therapy outcomes for a diverse set of test cases based on *E. coli* and *M. tb* data. This is striking considering that CARAMeL solely relies on simulated metabolic information, which was determined using only ~25–35% of available omics data.

Using CARAMeL to predict sequential interactions

Current approaches for predicting combination therapy outcomes focus on drug treatments that are given simultaneously. Here, we extended our approach to predict treatment efficacy for time- and order-dependent (i.e., sequential) interactions. In contrast to simultaneous combinations, the order and length of each drug treatment dictates how a pathogen adapts itself, and in turn, influences its sensitivity to successive drug treatments. As such, interaction outcomes are interpreted as leading to collateral sensitivity (analogous to synergy) or cross-resistance (analogous to antagonism). For this task, we used data for *E. coli* evolved in single drug treatments over three timespans (10, 21, and 90 days) then subsequently treated with a second drug[53,56,57]. To account for both time- and order-dependent drug effect, we re-defined the delta scores for sequential joint profiles. Briefly, delta scores were defined as the difference in binarized drug profiles scaled by the time interval between treatments (mathematically defined for pairwise sequences below):

$$\delta = \frac{t_2 v_2 - t_1 v_1}{t_2 - t_1} \qquad \textit{Eq. 1}$$

where δ = delta scores, t = length of treatment time, and v = binarized flux profile.

To initially assess how well the CARAMeL approach could predict sequential treatment outcomes, we first conducted a 10-fold cross-validation of the sequential data ($N = 628$), which involved 27 unique drugs (***Table S3***). We found that CARAMeL predictions moderately, but significantly, correlated with experimental outcomes ($R = 0.49$, $p < 10^{-16}$, ***Fig. 3A***). Further, the model performed well in determining whether a sequential interaction resulted in collateral sensitivity ($IS < -0.1$, AUROC = 0.73) or cross-resistance ($IS > 0.1$, AUROC = 0.75) (***Fig. 3D***).

We next evaluated the extent of our model's predictive power by conducting two types of leave-out analyses: (a) leave-first-drug-out and (b) leave-second-drug-out. The first case tested whether the model could generalize sequential treatment outcomes for an unknown evolved strain, while the second case assessed whether the model could generalize the immediate effect of a drug on strains evolved in other drugs. For a leave-out analysis, all interactions involving the drug of interest in the appropriate sequence position (first or second) were left out of model training and instead predicted for by the trained model. Similar to the cross-validation analysis, model performance was measured by the overall Spearman correlation and AUROC values for collateral sensitivity and cross-resistance. We

286 found that both leave-out analyses yielded accuracies similar to those attained from cross-
287 validation (***Fig. 3B, 3C, 3E, and 3F***). Overall, these results indicate that CARAMeL
288 generates robust and accurate predictions for sequential interactions.

289 To gain mechanistic insight into which factors influence combination therapy outcomes, we
290 trained a CARAMeL model using all interaction data available for *E. coli*. We then ranked
291 features by their predictive importance based on how the model accuracy decreases when a
292 feature is removed (see ***Methods*** for details). In total, we found that 580 features explained
293 95% of the variance in model predictions (***Data S1***). Of note, entropy features were amongst
294 the top 20, implying that metabolic disarray due to antibiotic stress is indicative of treatment
295 efficacy (***Fig. S5***). For the GEM-derived features (i.e., sigma and delta scores), we
296 determined that the differential flux through 221 metabolic reactions associated with the top
297 features significantly distinguished between synergistic and antagonistic interactions (two-
298 sample t-test, p-value < 0.05, ***Data S2***). We then deduced that 10 metabolic pathways were
299 enriched by this set of 221 reactions (***Table 1***). Differential activity through these pathways
300 aligned with the expected metabolic response to antibiotic treatments. For example,
301 increased flux through DNA repair systems (e.g., nucleotide salvage) is expected after
302 exposure to quinolones, which target DNA gyrase[58]. Differential flux through transport
303 reactions is also a common tactic that decreases drug concentrations within the bacterial
304 cell, therefore minimizing their adverse effects[59].

305 ***CARAMeL simulates the impact of intrinsic and extrinsic metabolic heterogeneity on***
306 ***drug interactions***
307 *In vivo* metabolic conditions span growth in diverse substrates such as sugars, nucleotides,
308 glycerol, lipids, and hypoxic conditions[44]. In contrast to existing approaches for drug
309 combination design, CARAMeL enables drug interaction predictions in diverse metabolic
310 conditions. This can help prioritize drug combinations for successful clinical translation
311 considering that the predominant nutrient source can change depending on where bacteria
312 reside inside the host[11]. By screening different conditions that are representative of *in vivo*
313 environments, we can identify drug combinations that target *E. coli* in diverse metabolic
314 conditions. Moreover, evaluating drug combinations based on efficacy across a large
315 compendium of metabolic network states will ensure robustness against heterogeneity.

316 To demonstrate the power of using CARAMeL in predicting condition-specific
317 combination therapy outcomes, we applied it to predict pairwise drug interactions in
318 multiple media conditions. For this task, we gathered experimental data for *E. coli* treated
319 with four single drug treatments (Aztreonam (AZT), Cefoxitin (CEF), Tetracycline (TET),
320 Tobramycin (TOB)) and two pairwise drug treatments (CEF + TET, CEF + TOB) (***Table***
321 ***S4***). Of note, this treatment panel evaluated the metabolic response in *E. coli* to bactericidal
322 (i.e., death-inducing) and bacteriostatic (i.e., growth-inhibiting) drugs, both individually and
323 in combination. Each drug treatment outcome was assessed in *E. coli* cultured in Biolog
324 phenotype microarray (PM)[60] plate-1, which measured metabolic respiration in 96 carbon
325 sources (***Fig. 4A***). Out of these 96 media conditions, 57 could be simulated based on the
326 metabolites annotated in the *E. coli* GEM (***Data S3***). As a result, ML model development
327 and all downstream analyses were conducted using the data subset pertaining to the 57
328 media conditions that were simulated.

329 We constructed a ML model using the following inputs: flux profiles for the four drug
330 treatments as well as the 57 media conditions, and interaction outcomes for 228 (4 * 57)

drug-media combinations. We then evaluated our model by predicting outcomes for 114 (2 * 57) drug-drug-media combinations (***Fig. 4B***). Overall, we found that model predictions significantly correlated with experimental outcomes ($R = 0.68$, $p < 10^{-16}$, ***Fig. 4C***). We also assessed correlations specific to each drug pair and found that model predictions still corresponded well with experimental data (CEF+TET: $R = 0.59$, $p \sim 10^{-6}$, CEF+TOB: $R = 0.81$, $p < 10^{-16}$). This large-scale inspection of combination therapy outcome in different growth environments was only possible with the CARAMeL approach, where flux profiles could be determined for 57 media conditions. A direct comparison of the same scale was not possible with the omic-based approach, as neither chemogenomic nor transcriptomics data was available for all these media conditions.

We next evaluated how cell-to-cell heterogeneity influenced combination therapy outcomes using population FBA[61], a modeling approach that simulates cell-specific metabolic heterogeneity based on single-cell proteomics data[62]. Specifically, information on the protein copy number levels measured for *E. coli* cultured in M9 glucose media is used to constrain the metabolic model[62]. To simulate heterogeneity between cells, the single-cell proteomics data is randomly sampled for each simulation. We used population FBA to simulate 1,000 *E. coli* cells cultured in M9 glucose media (***Methods***). We then generated pairwise predictions between all drugs for which the *E. coli* CARAMeL model was trained on ($N = 33$) against the 1,000 simulated cells (***Data S4***). For each drug pair, we evaluated three interaction cases: (a) simultaneous treatment ($D_1 + D_2$), (b) sequential treatment from D_1 to D_2, and (c) sequential treatment from D_2 to D_1. For sequential interactions, we set the duration for the first treatment to 14 days, based on the most commonly prescribed antibiotic treatment duration against bloodstream infection by *Enterobacteriaceae*[63], and one day for the second treatment.

Using the prediction landscape for the 1,000 cells, we determined the extent of cell-to-cell variability for each unique drug pair (***Data S5***). Overall, sequential predictions varied more largely between cells (up to 14% change in standard deviation relative to the mean) while there was less than 5% change in standard deviation compared to the mean for the simultaneous case (***Figure S6***). Interestingly, the variation among simultaneous predictions tended to follow a bimodal distribution (***Figs. 5A and S7***). Upon applying principal component analysis (PCA) onto the simultaneous prediction landscape for all 1,000 cells, we were able to group most cells into two distinct clusters that captured the previously observed bimodality (***Figs. 5B and 5C***). Using this PCA-based grouping, we subsequently identified 35 metabolic reactions through which flux activity was significantly different between the two clusters (***Fig. 5D***, unpaired Student's t-test, $p < 0.05$). Surprisingly, one reaction perfectly aligned with the PCA-based clustering: trans-2-decenoyl-ACP isomerase. In *E. coli*, this reaction carries out the conversion between the cis- and trans- form of 10-carbon enoyl-ACP, a key component for unsaturated fatty acid synthesis. The flux data shown in ***Fig. 5D*** indicates that cluster-1 cells (more sensitive to combination therapy) exhibited flux in the forward direction (trans- to cis- formation), while cluster-2 cells exhibited flux in the reverse direction (cis- to trans-). Of note, prior studies have proposed that Gram-negative bacteria can adapt to membrane perturbations by regulating cis-trans isomerization of unsaturated fatty acids[64]. Specifically, cis- to trans- conversion may occur in response to membrane stress to reduce membrane fluidity, and consequently, diminish its effect on the membrane. Although this mechanism has not been experimentally shown in *E. coli*, this may explain how cluster-2 cells (associated with cis- to trans- isomerization) are predicted to be more antagonistic to combination therapies.

Screening for robust combination therapies

Synergy observed in the lab may not result in synergy *in vivo* due to differences in growth conditions or drug pharmacokinetics, wherein drugs may reach the infection site at different times rather than simultaneously[65]. Considering these factors, combination therapies that show synergy across growth conditions and time scales hold the best potential for successful clinical translation. To discover such therapies, we predicted pairwise and three-way regimen outcomes for all drugs for which the *E. coli* CARAMeL model was trained on (N = 33) across 57 carbon sources (from Biolog PM01). For sequential interactions predictions, treatment duration for pairwise treatments was set to 14 days followed by 1 day, while three-way treatments were set to a 14-14-1-day prescription. In total, we generated predictions for 90,288 pairwise combinations ($_{33}C_2$ pairs x 3 interaction cases x 57 PM01 conditions, ***Data S6***) and 932,976 three-way combinations ($_{33}C_3$ combinations x 7 interaction cases x 57 PM01 conditions, ***Data S7***).

Out of 528 drug pairs and 5,456 three-way interactions, none was predicted to be synergistic across all media conditions and interaction cases. In fact, sustained synergy across media conditions seems to occur for only a small subset (< 10%) of drug interactions (***Fig. S8***). Specifically, 73 drug pairs and 165 three-way interactions were predicted to yield synergy both simultaneously and sequentially in at least one media condition (***Fig. 6 and S9***). Of note, all 73 drug pairs showed less than 5% cell-to-cell variation based on population FBA for all interactions cases (i.e., simultaneous and sequential interactions). Upon closer inspection of both pairwise and three-way sets, synergy was not found to be retained across a majority of media conditions for three-way drug interactions. On the other hand, several pairwise interactions were found to retain synergy well; specifically, 24 drug pairs out of 73 were found to be synergistic in more than 50% of conditions in both simultaneous and sequential cases.

Interestingly, several of these 24 drug pairs possess evidence for clinical use against bacterial infections. For example, amikacin-ampicillin treatment (AMK-AMP) has previously been shown to be clinically effective for a wide range of infections[66–68] including treatment of bacteremia in neutropenic patients[69] and neonatal bacterial infections[70]. Other drug interactions of note include: azithromycin-rifampicin (AZI-RIF), which has demonstrated clinical efficacy in treating arthritis induced by pathogenic Chlamydia (Gram-negative)[71]; fusidic acid-rifampicin (FUS-RIF), which has shown clinical efficacy against prosthetic joint infection caused by drug-resistant staphylococci (Gram-positive)[72]; and minocycline-rifampicin (MIN-RIF), which has been shown to prevent colonization by slime-producing staphylococci in catheters[73]. Additionally, rifampicin combined with other drugs has been advised as treatment for Gram-negative and non-mycobacterial infections[42,74]. Further investigation into the set of 24 drug pairs predicted to yield robust synergy may lead to the discovery of new combination therapies that could be put to clinical use.

Discussion

Here we introduced CARAMeL, a modeling approach to design condition-specific antibiotic regimens. CARAMeL offers multiple advantages over prior methods of similar nature. First, our approach enables use of diverse data sources (e.g., chemogenomics, transcriptomics) and/or their combined use, therefore maximizing the number of drugs that are screened. For instance, our *E. coli* CARAMeL model leveraged use of both chemogenomics data (for defining drug flux profiles) and proteomics data (for simulating

single cells), a feat that cannot be accomplished with prior methods of similar nature. Second, we extended our approach to simulate different interaction cases (simultaneous vs. sequential) when designing combination therapies. To our knowledge, no framework currently exists to incorporate these factors into drug discovery efforts against AMR. Third, the use of GEMs enables simulation of highly tunable metabolic conditions (as showcased with our analysis of the Biolog PM01 data), which may be leveraged to investigate combination therapy outcomes in the host environment. GEMs also enable the simulation of pathogen metabolic heterogeneity, due to both intrinsic stochasticity and the metabolic environment. Pathogen heterogeneity is a critical barrier in designing effective antibiotic therapies, and this must be factored into combination therapy design to mitigate the rise of resistant strains.

Additional advantages to the CARAMeL approach include its mechanistic model interpretability, its ability to simulate pathogen metabolic heterogeneity, and its use in generating predictions across numerous conditions in large-scale. With regard to model interpretation, the *E. coli* CARAMeL model revealed that entropy, or metabolic disarray, plays an important role in combination therapy outcome. The direct link between model features (i.e., sigma and delta scores) and GEM reactions also pinpointed pathways that are activate in response to drug treatment (***Table 1***), many of which align with the expected resistance mechanisms against antimicrobial stress[58,59]. Using population FBA[61], we investigated how drug interaction outcomes may differ from cell-to-cell. Interestingly, we found that level of cis- to trans- isomerization of unsaturated fatty acids, a response used by Gram-negative pathogens to overcome membrane perturbations[64], may influence the effect of drug combinations against *E. coli*. Pathogen metabolic heterogeneity may arise due to both intrinsic stochasticity and the local growth environment. Pathogen heterogeneity is a critical barrier in designing effective antibiotic therapies, and this must be factored into combination therapy design to mitigate the rise of resistant strains. Finally, analysis of the drug interaction landscape suggests that only a small set of 24 drug pairs show robust synergy across growth conditions and interaction cases, with some possessing clinical evidence for efficacy[42,67–74]. Further investigation into this list of drug interactions may lead to the discovery of new combination therapies for clinical application.

Although the use of GEMs in CARAMeL offers major advantages with data compatibility, condition tunability, and mechanistic insight, it also introduces some limitations. The level of accuracy and thoroughness in GEM annotation may influence CARAMeL model performance. Moreover, our current approach only provides a "snapshot" perspective of the metabolic response to a condition. This may be a potential reason for the diminished CARAMeL model performance in predicting sequential outcomes. Nevertheless, these are areas that can be addressed with continued curation of GEMs[75] and advances in dynamic metabolic modeling[76]. Overall, the ability to simulate specific growth environments and pathogen metabolic heterogeneity offers the potential to evaluate treatment efficacy *in vivo* and advance clinical translation of novel antibiotic regimens. Moreover, these combination therapies could restore use of defunct antibiotics against resistant pathogens while mitigating further resistance[24,77].

Methods

Experimental Design (Biolog Phenotype Microarray)

E. coli MG1655 was cultured in Biolog phenotype microarray (PM) 1, which screened bacterial growth in 95 carbon sources and a negative control (i.e., water)[60]. *E. coli* was subsequently treated with six distinct drug treatments in duplicate: aztreonam (0.03 ug/mL), cefoxitin (1.87 ug/mL), tetracycline (1.42 ug/mL), tobramycin (0.15 ug/mL), cefoxitin (1.87 ug/mL) + tetracycline (1.42 ug/mL), and cefoxitin (1.87 ug/mL) + tobramycin (1.42 ug/mL). Including a reference plate (*E. coli* growth in PM01 only), phenotype in each treatment was colorimetrically measured in duplicate using tetrazolium violet dye, which quantifies cellular respiration. All experimental procedures, data collection, and quality control were performed at Biolog, Inc. The area under the respiration curve was calculated using MATLAB and reported as the ratio of treatment to reference.

Simulating Metabolic Flux using GEMs

The *E. coli* GEM iJO1366[48] and the *M. tb* GEM iEK1008[49] were used to simulate metabolic fluxes at steady-state. To simulate drug flux profiles, chemogenomic data for *E. coli*[50] and transcriptomic data for *M. tb*[38] served as GEM constraints. Specifically, differential gene regulation in response to drug treatment was uniquely inferred from each dataset. For chemogenomic data, which measured single-gene knockout (KO) fitness, genes whose KOs promoted growth were assumed as dispensable while gene KOs that resulted in low fitness were assumed to be essential for growth in said condition. Based on these assumptions, genes corresponding with low ($z < -2$) and high ($z > 2$) fitness were inferred to be up- and down-regulated, respectively. For transcriptomic data, which measured single-gene expression, up- and down-regulation were directly inferred based on high ($z > 2$) and low ($z < -2$) expression values, respectively. These processes yielded individual sets of differentially regulated genes that were integrated into corresponding GEMs using a linear optimization version of the integrative metabolic analysis tool (iMAT)[78,79]. To determine media flux profiles, metabolite availability was computationally defined by constraining exchange reactions annotated in iJO1366. For each carbon source of interest (e.g., glycerol), the lower bound (i.e., uptake rate) for the corresponding exchange reaction (e.g., glycerol exchange) was set to -10 g/mmol to allow cellular intake.

Of note, use of the linear iMAT algorithm required constraint-based modeling (CBM) parameter fine-tuning for three variables: kappa, rho, and epsilon[80]. Kappa and rho serve as relative weights for "off" and "on" reactions associated with the differentially expressed genes, respectively, in their contribution to the objective function. Epsilon represents the minimum flux through "on" reactions. For the purposes of this research, we varied all three parameter values from 10^{-3} to 1 and determined the optimal parameter set based on three criteria: (1) maximizing the Spearman correlation between predicted and actual interaction scores after 10-fold cross-validation using a training dataset, (2) minimizing the number of conditions simulated to have no growth, and (3) ensuring non-zero variability in the simulated growth rates between conditions. ***Table S5*** provides results for all three assessments for all parameter sets of interest. The following optimal parameter values were obtained for each GEM using the training dataset: (1) iJO1366 – kappa $=10^{-2}$, rho $= 10^{-2}$, epsilon = 1, and (2) iEK1008 – kappa $= 10^{-2}$, rho $= 10^{-2}$, epsilon $= 10^{-2}$. These parameter values were used for all results when benchmarking CARAMeL against previous approaches based on *E. coli* and *M. tb* drug interaction datasets (***Table S6***).

To simulate cell-specific flux data, we applied population FBA[61], an approach that models metabolic heterogeneity within a cell population using proteomics data. For the purposes of this study, we defined a population of 1,000 *E. coli* cells to simulate using the default parameters for population FBA.

Data Processing to Determine Joint Profiles

Flux profiles were used to define joint profiles for each drug combination, which were comprised of four pieces of information: sigma scores, delta scores, cumulative entropy, and treatment time interval (***Fig. S2***). Sigma and delta scores were representative of the combined and unique effect of drugs involved in a combination, respectively. Both score types were determined after flux profiles were binarized based on differential flux activity (either positive or negative) in comparison to baseline, mathematically defined below:

$$v_{i,\ positive} = \begin{cases} 1 & \dfrac{v_{i,\ treatment}}{v_{i,\ baseline}} > 2 \\ 0 & otherwise \end{cases}, \ v \in \mathbb{R}^m \qquad \textit{Eq. 2a}$$

$$v_{i,\ negative} = \begin{cases} 1 & \dfrac{v_{i,\ treatment}}{v_{i,\ baseline}} < -2 \\ 0 & otherwise \end{cases}, \ v \in \mathbb{R}^m \qquad \textit{Eq. 3b}$$

where v = reaction flux and m = total number of GEM reactions. Sigma scores were mathematically defined for both simultaneous and sequential interactions using the following equation:

$$\sigma_i = \frac{2}{n} \sum_{j=1}^{n} v_{i,\ j}, \ v \in \mathbb{R}^{m \times n} \qquad \textit{Eq. 4}$$

where σ = sigma score, v = binarized flux profile, m = total number of GEM reactions, and n = total number of conditions in a combination. Delta scores were separately defined for simultaneous and sequential interactions based on *Eq. 6a* and *Eq. 6b*, respectively:

$$\delta_{i,\ simultaneous} = \begin{cases} 1 & \sum_{j=1}^{n} v_{i,j} = 1 \\ 0 & otherwise \end{cases}, v \in \mathbb{R}^{m \times n} \qquad \textit{Eq. 5a}$$

$$\delta_{i,\ sequential} = t_n v_{i,n} - \sum_{j=1}^{n-1} t_j v_{i,j}, t \in \mathbb{R}^n, v \in \mathbb{R}^{m \times n} \qquad \textit{Eq. 6b}$$

where δ = delta score, t = treatment time interval, v = binarized flux profile, m = total number of GEM reactions, and n = total number of conditions in a combination. Cumulative entropy features were determined by processing non-binarized flux profiles in two steps. First, metabolic entropy for each condition was mathematically defined by the following equation:

$$H_j = \ln\left(\sigma_j^2\right) \qquad \textit{Eq. 7}$$

where H_j = metabolic entropy due to condition j and σ_j^2 = variance in the non-binarized flux profile for condition j. Of note, this formulation was adapted from Zhu *et al.*, who quantified entropy of the bacterial stress response to antibiotics[15]. Next, the mean and sum in entropy for all conditions involved in an interaction were used to define two distinct entropy features. Finally, the time feature was defined as the time interval between the first and last treatment for a combination. For simultaneous interactions, the time feature was set to zero.

ML Model Development using Random Forests

All CARAMeL models were built in MATLAB (Mathworks, Inc.) using the regression-based Random Forests (RF) algorithm[81]. Briefly, RF is an ensemble method comprised of decision trees that learn to associate feature information to a target variable. For the regression-based approach, the RF model returns the mean prediction from all decision trees. To develop CARAMeL models, joint profiles served as feature information while drug interaction scores were used as the target variable. Interaction scores were quantified using the Loewe additivity model[82], which is based on drug concentrations (refer to the original sources of drug interaction datasets for further details in score calculation). Both joint profiles and interactions scores for drug combinations of interest were used as model inputs during training, while only joint profiles were provided as input during model testing. Default values for all other model parameters were used during both training and testing.

ML Model Performance Assessment

Model performance was evaluated based on two metrics: (1) the Spearman correlation between actual and predicted interaction scores and (2) the area under the receiver operating curve (AUROC) for classifying interactions as synergistic or antagonistic. Of note, model predictions for TB regimens used in clinical trials were negative transformed before being compared to clinical outcomes. Since these clinical trials reported percentage of patients that were cured, we would expect to see a negative correlation between interaction scores and clinical efficacy, with synergistic regimens (negative IS) performing better than antagonistic regimens. The sign for the scores were hence flipped to maintain a positive correlation indicating good model performance. Classification of simultaneous drug interactions was based on score threshold values reported in the original literature for a dataset. For both sequential interactions and the CARAMeL model trained on all interaction data for *E. coli*, interaction scores were first scaled by the maximum absolute value (***Eq. 2***). Interaction values below -0.1 and above 0.1 were then used to classify interactions as synergistic and antagonistic, respectively. For the 10-fold cross-validation analysis conducted for sequential interactions, the interaction data was randomly partitioned into ten subsets of similar size (N ~ 63). CARAMeL was then applied to predict each subset at a time, where the given subset was left out of the model training (i.e., the remaining 90% of the data was used to train the model). All model predictions were then compared to the sequential data as a whole to calculate the overall Spearman correlation and AUROC values.

CARAMeL Top Feature Extraction

Top features were determined based on their ranked importance in generating accurate predictions. To calculate feature importance, each feature was first left out of model training and testing. The mean squared error (MSE) between predicted and true interaction scores was then calculated for each model. Finally, feature importance was measured as the increase in MSE for a model lacking a feature compared to the model trained on all features. After ranking features according to decreasing importance, the first set of features amounting to a cumulative importance of 0.95 (corresponding to 95% variance explained) were selected for downstream model interpretation and analysis.

We trained a CARAMeL model using all interaction data available for *E. coli*. Broadly, this included three sets of simultaneous combinations[36,37] (pairwise, three-way, and media-specific treatments) and three sets of sequential interactions[53,56,57] (differing based on elapsed treatment time). To account for differing units of measurement between datasets, we scaled interaction scores according to the following formula:

$$x_{scaled} = \frac{x}{max|x|} \qquad Eq.\ 8$$

where x is a vector of interaction scores for a given dataset. This scaling constrained all interaction scores to range between ±1 while retaining the sign consensus for classifying interactions based on their score (negative IS → synergy, positive IS → antagonism). In total, we trained our model on 966 drug interactions and attained highly accurate predictions ($R = 0.45$, $p < 10^{-16}$) for both synergistic (IS < -0.1, AUROC = 0.67) and antagonistic (IS > 0.1, AUROC = 0.71) interactions based on a 10-fold cross-validation.

Statistical Analysis

A one-way analysis of variance (ANOVA) test was used to compare both the entropy mean and entropy sum of drug interactions grouped by their classification (synergy, neutral, antagonism). A multiple comparison test based on Tukey's honestly significant difference (HSD) was subsequently performed to identify statistically significant pairwise differences using a p-value threshold of 0.05. A two-sample Student's t-test with unequal variance was used to define which reactions distinguished between synergistic and antagonistic interactions based on differential flux activity. Lastly, a hypergeometric test was conducted to determine significantly enriched metabolic pathways based on GEM reactions associated with top CARAMeL predictors. For this test, the total number of reactions annotated in iJO1366 corresponded with the population size.

References

1. Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. V. Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology* vol. 13 42–51 (2015).
2. Balaban, N. Q., Gerdes, K., Lewis, K. & McKinney, J. D. A problem of persistence: still more questions than answers? *Nat. Rev. Microbiol. 2013 118* **11**, 587–591 (2013).
3. Kussell, E., Kishony, R., Balaban, N. Q. & Leibler, S. Bacterial PersistenceA Model of Survival in Changing Environments. *Genetics* **169**, 1807–1814 (2005).
4. Rittershaus, E. S. C., Baek, S. H. & Sassetti, C. M. The Normalcy of Dormancy: Common Themes in Microbial Quiescence. *Cell Host Microbe* **13**, 643–651 (2013).
5. Cohen, N. R., Lobritz, M. A. & Collins, J. J. Microbial Persistence and the Road to Drug Resistance. *Cell Host Microbe* **13**, 632–642 (2013).
6. Gill, E. E., Franco, O. L. & Hancock, R. E. W. Antibiotic Adjuvants: Diverse Strategies for Controlling Drug-Resistant Pathogens. *Chem. Biol. Drug Des.* **85**, 56–78 (2015).
7. Holmes, A. H. *et al.* Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet* **387**, 176–187 (2016).
8. Yang, J. H., Bening, S. C. & Collins, J. J. Antibiotic efficacy — context matters. *Curr. Opin. Microbiol.* **39**, 73–80 (2017).
9. Martínez, J. L. & Rojo, F. Metabolic regulation of antibiotic resistance. *FEMS Microbiol. Rev.* **35**, 768–789 (2011).
10. Yang, J. H. *et al.* Antibiotic-Induced Changes to the Host Metabolic Environment Inhibit Drug Efficacy and Alter Immune Function. *Cell Host Microbe* **22**, 757-765.e3 (2017).
11. Brown, S. A., Palmer, K. L. & Whiteley, M. Revisiting the host as a growth medium. *Nature Reviews Microbiology* vol. 6 657–666 (2008).
12. Fuchs, T. M., Eisenreich, W., Heesemann, J. & Goebel, W. Metabolic adaptation of human pathogenic and related nonpathogenic bacteria to extra- and intracellular habitats. *FEMS Microbiology Reviews* vol. 36 435–462 (2012).
13. Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A. & Collins, J. J. A Common Mechanism of Cellular Death Induced by Bactericidal Antibiotics. *Cell* **130**, 797–810

640 (2007).
641 14. Dharmaraja, A. T. Role of Reactive Oxygen Species (ROS) in Therapeutics and Drug
642 Resistance in Cancer and Bacteria. *J. Med. Chem.* **60**, 3221–3240 (2017).
643 15. Zhu, Z. *et al.* Entropy of a bacterial stress response is a generalizable predictor for fitness
644 and antibiotic sensitivity. *Nat. Commun.* **11**, 1–15 (2020).
645 16. Farha, M. A. & Brown, E. D. Drug repurposing for antimicrobial discovery. *Nature*
646 *Microbiology* vol. 4 565–577 (2019).
647 17. Tyers, M. & Wright, G. D. Drug combinations: a strategy to extend the life of antibiotics in
648 the 21st century. *Nature Reviews Microbiology* vol. 17 141–155 (2019).
649 18. Pluchino, K. M., Hall, M. D., Goldsborough, A. S., Callaghan, R. & Gottesman, M. M.
650 Collateral sensitivity as a strategy against cancer multidrug resistance. *Drug Resist. Updat.*
651 **15**, 98–105 (2012).
652 19. Vijayaraghavalu, S., Dermawan, J. K., Cheriyath, V. & Labhasetwar, V. Highly synergistic
653 effect of sequential treatment with epigenetic and anticancer drugs to overcome drug
654 resistance in breast cancer cells is mediated via activation of p21 gene expression leading
655 to G2/M cycle arrest. *Mol. Pharm.* **10**, 337–352 (2013).
656 20. Pál, C., Papp, B. & Lázár, V. Collateral sensitivity of antibiotic-resistant microbes. *Trends*
657 *in Microbiology* vol. 23 401–407 (2015).
658 21. Gadamski, G., Ciarka, D., Gressel, J. & Gawronski, S. W. Negative cross-resistance in
659 triazine-resistant biotypes of Echinochloa crus-galli and Conyza canadensis. *Weed Sci.* **48**,
660 176–180 (2000).
661 22. Deeks, S. G. Treatment of antiretroviral-drug-resistant HIV-1 infection. *Lancet* vol. 362
662 2002–2011 (2003).
663 23. Lukens, A. K. *et al.* Harnessing evolutionary fitness in Plasmodium falciparum for drug
664 discovery and suppressing resistance. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 799–804 (2014).
665 24. Baym, M., Stone, L. K. & Kishony, R. Multidrug evolutionary strategies to reverse
666 antibiotic resistance. *Science (80-.).* **351**, aad3292–aad3292 (2016).
667 25. Ankomah, P., Johnson, P. J. T. & Levin, B. R. The Pharmaco –, Population and
668 Evolutionary Dynamics of Multi-drug Therapy: Experiments with S. aureus and E. coli and
669 Computer Simulations. *PLOS Pathog.* **9**, e1003300 (2013).
670 26. Yang, J. H. *et al.* A White-Box Machine Learning Approach for Revealing Antibiotic
671 Mechanisms of Action. *Cell* **177**, 1649–1661 (2019).
672 27. Ribeiro da Cunha, B., Fonseca, L. P. & Calado, C. R. C. Simultaneous elucidation of
673 antibiotic mechanism of action and potency with high-throughput Fourier-transform
674 infrared (FTIR) spectroscopy and machine learning. *Appl. Microbiol. Biotechnol.* **105**,
675 1269–1286 (2021).
676 28. Stokes, J. M. *et al.* A Deep Learning Approach to Antibiotic Discovery. *Cell* **180**, 688–702
677 (2020).
678 29. El Zahed, S. S. & Brown, E. D. Chemical-Chemical Combinations Map Uncharted
679 Interactions in Escherichia coli under Nutrient Stress. doi:10.1016/j.isci.2018.03.018.
680 30. Zhao, X.-M. *et al.* Prediction of Drug Combinations by Integrating Molecular and
681 Pharmacological Data. *PLoS Comput. Biol.* **7**, e1002323 (2011).
682 31. Lee, J.-H. *et al.* CDA: Combinatorial Drug Discovery Using Transcriptional Response
683 Modules. *PLoS One* **7**, e42573 (2012).
684 32. Chua, H. E., Bhowmick, S. S. & Tucker-Kellogg, L. Synergistic target combination
685 prediction from curated signaling networks: Machine learning meets systems biology and
686 pharmacology. *Methods* **129**, 60–80 (2017).
687 33. Regan-Fendt, K. E. *et al.* Synergy from gene expression and network mining (SynGeNet)
688 method predicts synergistic drug combinations for diverse melanoma genomic subtypes.

689 *npj Syst. Biol. Appl.* **5**, 1–15 (2019).
690 34. Yuan, B. *et al.* CellBox: Interpretable Machine Learning for Perturbation Biology with
691 Application to the Design of Cancer Combination Therapy. *Cell Syst.* **12**, 128-140.e4
692 (2021).
693 35. Zhang, T., Zhang, L., Payne, P. R. O. & Li, F. Synergistic drug combination prediction by
694 integrating multiomics data in deep learning models. in *Methods in Molecular Biology* vol.
695 2194 223–238 (Humana Press Inc., 2021).
696 36. Chandrasekaran, S. *et al.* Chemogenomics and orthology-based design of antibiotic
697 combination therapies. *Mol. Syst. Biol.* **12**, 872 (2016).
698 37. Cokol, M., Li, C. & Chandrasekaran, S. Chemogenomic model identifies synergistic drug
699 combinations robust to the pathogen microenvironment. *PLOS Comput. Biol.* **14**, e1006677
700 (2018).
701 38. Ma, S. *et al.* Transcriptomic signatures predict regulators of drug synergy and clinical
702 regimen efficacy against tuberculosis. *MBio* **10**, (2019).
703 39. Solomkin, J. S. *et al.* Diagnosis and Management of Complicated Intra-Abdominal
704 Infection in Adults and Children: Guidelines by the Surgical Infection Society and the
705 Infectious Diseases Society of America. *Surg. Infect. (Larchmt).* **11**, 79–109 (2010).
706 40. Dheda, K., Barry, C. E. & Maartens, G. Tuberculosis. *Lancet* **387**, 1211–1226 (2016).
707 41. Boyd, N. & Nailor, M. D. Combination Antibiotic Therapy for Empiric and Definitive
708 Treatment of Gram-Negative Infections: Insights from the Society of Infectious Diseases
709 Pharmacists. *Pharmacother. J. Hum. Pharmacol. Drug Ther.* **31**, 1073–1084 (2011).
710 42. Forrest, G. N. & Tamura, K. Rifampin combination therapy for nonmycobacterial
711 infections. *Clinical Microbiology Reviews* vol. 23 14–34 (2010).
712 43. Bumann, D. Heterogeneous Host-Pathogen Encounters: Act Locally, Think Globally. *Cell*
713 *Host Microbe* **17**, 13–19 (2015).
714 44. Eisenreich, W., Dandekar, T., Heesemann, J. & Goebel, W. Carbon metabolism of
715 intracellular bacterial pathogens and possible links to virulence. *Nat. Rev. Microbiol. 2010*
716 *86* **8**, 401–412 (2010).
717 45. Price, N. D., Reed, J. L. & Palsson, B. Genome-scale models of microbial cells: Evaluating
718 the consequences of constraints. *Nature Reviews Microbiology* vol. 2 886–897 (2004).
719 46. Lewis, N. E., Nagarajan, H. & Palsson, B. O. Constraining the metabolic genotype-
720 phenotype relationship using a phylogeny of in silico methods. *Nature Reviews*
721 *Microbiology* vol. 10 291–305 (2012).
722 47. Dahal, S., Yurkovich, J. T., Xu, H., Palsson, B. O. & Yang, L. Synthesizing Systems
723 Biology Knowledge from Omics Using Genome-Scale Models. *Proteomics* vol. 20
724 1900282 (2020).
725 48. Orth, J. D. *et al.* A comprehensive genome-scale reconstruction of *Escherichia coli*
726 metabolism—2011. *Mol. Syst. Biol.* **7**, 535 (2011).
727 49. Kavvas, E. S. *et al.* Updated and standardized genome-scale reconstruction of
728 Mycobacterium tuberculosis H37Rv, iEK1011, simulates flux states indicative of
729 physiological conditions. *BMC Syst. Biol.* **12**, 25 (2018).
730 50. Nichols, R. J. *et al.* Phenotypic landscape of a bacterial cell. *Cell* **144**, 143–56 (2011).
731 51. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a
732 protein. *Nature* **436**, 588–592 (2005).
733 52. Maeda, T. *et al.* High-throughput laboratory evolution reveals evolutionary constraints in
734 Escherichia coli. *Nat. Commun.* **11**, 5970 (2020).
735 53. Suzuki, S., Horinouchi, T. & Furusawa, C. Prediction of antibiotic resistance by gene
736 expression profiles. *Nat. Commun.* **5**, 5792 (2014).
737 54. Mori, M. *et al.* From coarse to fine: the absolute Escherichia coli proteome under diverse

738 growth conditions. *Mol. Syst. Biol.* **17**, e9536 (2021).
739 55. Bonnett, L. J., Ken-Dror, G., Koh, G. C. K. W. & Davies, G. R. Comparing the Efficacy of
740 Drug Regimens for Pulmonary Tuberculosis: Meta-analysis of Endpoints in Early-Phase
741 Clinical Trials. *Clin. Infect. Dis.* **46**, 46–54 (2017).
742 56. Imamovic, L. & Sommer, M. O. A. Use of collateral sensitivity networks to design drug
743 cycling protocols that avoid resistance development. *Sci. Transl. Med.* **5**, 204ra132-
744 204ra132 (2013).
745 57. Oz, T. *et al.* Strength of selection pressure is an important parameter contributing to the
746 complexity of antibiotic resistance evolution. *Mol. Biol. Evol.* **31**, 2387–2401 (2014).
747 58. Fàbrega, A., Madurga, S., Giralt, E. & Vila, J. Mechanism of action of and resistance to
748 quinolones. *Microb. Biotechnol.* **2**, 40–61 (2009).
749 59. Levy, S. B. Active efflux, a common mechanism for biocide and antibiotic resistance. *J.*
750 *Appl. Microbiol.* **92**, 65S-71S (2002).
751 60. Bochner, B. R., Gadzinski, P. & Panomitros, E. Phenotype Microarrays for high-
752 throughput phenotypic testing and assay of gene function. *Genome Res.* **11**, 1246–1255
753 (2001).
754 61. Labhsetwar, P., Cole, J. A., Roberts, E., Price, N. D. & Luthey-Schulten, Z. A.
755 Heterogeneity in protein expression induces metabolic variability in a modeled Escherichia
756 coli population. *Proc. Natl. Acad. Sci.* **110**, 14006–14011 (2013).
757 62. Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule
758 sensitivity in single cells. *Science (80-.).* **329**, 533–538 (2010).
759 63. Tansarli, G. S., Andreatos, N., Pliakos, E. E. & Mylonakis, E. A Systematic Review and
760 Meta-analysis of Antibiotic Treatment Duration for Bacteremia Due to Enterobacteriaceae.
761 *Antimicrob. Agents Chemother.* **63**, (2019).
762 64. Heipieper, H. J., Fischer, J. & Meinhardt, F. Cis–Trans Isomerase of Unsaturated Fatty
763 Acids: An Immediate Bacterial Adaptive Mechanism to Cope with Emerging Membrane
764 Perturbation Caused by Toxic Hydrocarbons. *Handb. Hydrocarb. Lipid Microbiol.* 1605–
765 1614 (2010) doi:10.1007/978-3-540-77587-4_112.
766 65. Cicchese, J. M., Sambarey, A., Kirschner, D., Linderman, J. J. & Chandrasekaran, S. A
767 multi-scale pipeline linking drug transcriptomics with pharmacokinetics predicts in vivo
768 interactions of tuberculosis drugs. *Sci. Rep.* **11**, 5643 (2021).
769 66. Krause, K. M., Serio, A. W., Kane, T. R. & Connolly, L. E. Aminoglycosides: An
770 overview. *Cold Spring Harb. Perspect. Med.* **6**, (2016).
771 67. Guidelines for the management of adults with hospital-acquired, ventilator-associated, and
772 healthcare-associated pneumonia. *American Journal of Respiratory and Critical Care*
773 *Medicine* vol. 171 388–416 (2005).
774 68. Dellinger, R. P. *et al.* Surviving sepsis campaign: International guidelines for management
775 of severe sepsis and septic shock: 2012. *Crit. Care Med.* **41**, 580–637 (2013).
776 69. Palmblad, J. & Lönnqvist, B. Combination of Amikacin and either Ampicillin or
777 Cephalotin as Initial Treatment of Febrile Neutropenic Patients. *Acta Med. Scand.* **212**,
778 379–384 (1982).
779 70. Umaña, M. A., Odio, C. M., Castro, E., Salas, J. L. & McCracken, G. H. Evaluation of
780 aztreonam and ampicillin vs. amikacin and ampicillin for treatment of neonatal bacterial
781 infections. *Pediatr. Infect. Dis. J.* **9**, 175–180 (1990).
782 71. Carter, J. D. *et al.* Combination antibiotics as a treatment for chronic Chlamydia-induced
783 reactive arthritis: A double-blind, placebo-controlled, prospective trial. *Arthritis Rheum.*
784 **62**, 1298–1307 (2010).
785 72. Zimmerli, W., Trampuz, A. & Ochsner, P. E. Current concepts: Prosthetic-joint infections.
786 *N. Engl. J. Med.* **351**, 1645–1654 (2004).

73. Raad, I., Darouiche, R., Hachem, R., Sacilowski, M. & Bodey, G. P. Antibiotics and prevention of microbial colonization of catheters. *Antimicrob. Agents Chemother.* **39**, 2397 (1995).
74. Drapeau, C. M. J., Grilli, E. & Petrosillo, N. Rifampicin combined regimens for Gram-negative infections: data from the literature. *Int. J. Antimicrob. Agents* **35**, 39–44 (2010).
75. Bernstein, D. B., Sulheim, S., Almaas, E. & Segrè, D. Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. *Genome Biology* vol. 22 1–22 (2021).
76. Saa, P. A. & Nielsen, L. K. Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnology Advances* vol. 35 981–1003 (2017).
77. Kim, S., Lieberman, T. D. & Kishony, R. Alternating antibiotic treatments constrain evolutionary paths to multidrug resistance. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 14494–14499 (2014).
78. Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. & Ruppin, E. Network-based prediction of human tissue-specific metabolism. *Nature Biotechnology* vol. 26 1003–1010 (2008).
79. Shen, F., Boccuto, L., Pauly, R., Srikanth, S. & Chandrasekaran, S. Genome-scale network model of metabolism and histone acetylation reveals metabolic dependencies of histone deacetylase inhibitors. *Genome Biol.* **20**, 49 (2019).
80. Shen, F., Cheek, C. & Chandrasekaran, S. Dynamic Network Modeling of Stem Cell Metabolism. *Methods Mol. Biol.* **1975**, 305–320 (2019).
81. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
82. Loewe, S. & Muischnek, H. Über Kombinationswirkungen - Mitteilung: Hilfsmittel der Fragestellung. *Arch. für Exp. Pathol. und Pharmakologie* **114**, 313–326 (1926).

Acknowledgments

We thank Brendan Lewis at Biolog for carrying out the Phenotype Microarray assays.

Funding

National Institutes of Health grant R35 GM13779501 (SC)
National Institutes of Health NIAID R56AI150826 (SC)
University of Michigan faculty start-up fund (SC)

Competing interests

The authors declare that they have no competing interests.

Data and materials availability

All datasets and code used within this work are provided through the CARAMeL GitHub repository (https://github.com/sriram-lab/CARAMeL). Additional information on datasets and results are available in the supplementary materials.

FIGURES AND TABLES

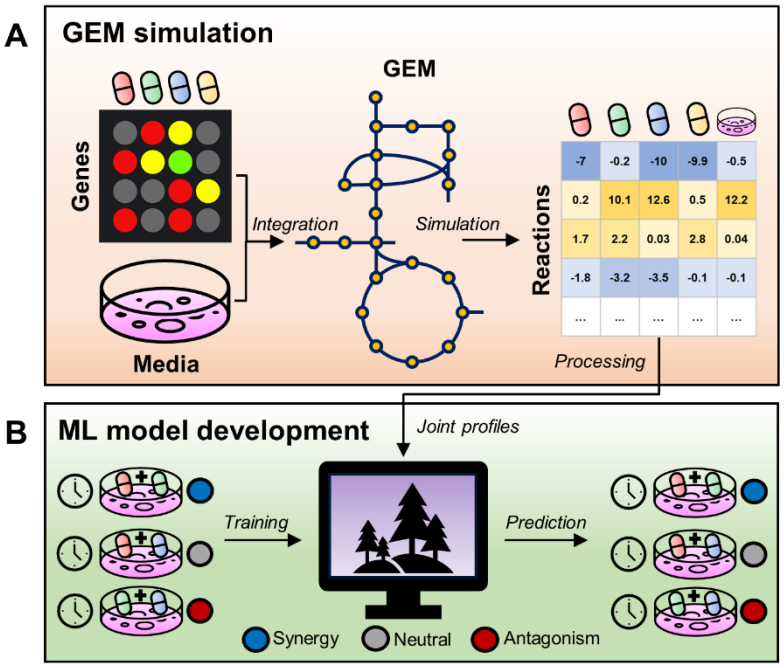

Fig. 1. CARAMeL approach schematic. The *Condition-specific Antibiotic Regimen Assessment using Mechanistic Learning* (CARAMeL) approach involves a two-step process: (**A**) omics data (e.g., transcriptomics) measured for single drug treatments and information on growth media composition are integrated into a genome-scale metabolic model (GEM) to simulate metabolic flux changes. (**B**) This information, along with drug interaction data, serve as inputs to train a machine learning (ML) model to predict outcomes for novel drug interactions.

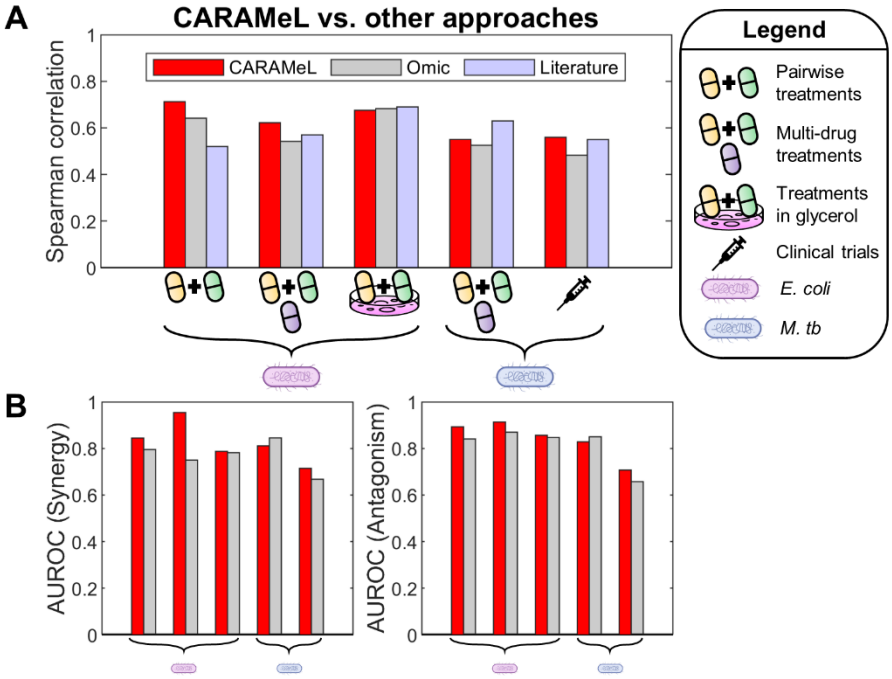

Fig. 2. CARAMeL was benchmarked against other predictive approaches. Model results for three approaches are reported: CARAMeL (this study), omics (based on chemogenomic or transcriptomic data as input), and literature[36–38] (reported in literature). Model results were quantified based on (**A**) the Spearman correlation between actual outcomes and model predictions, and (**B**) the area under the receiver operating curve (AUROC) for classifying interactions as synergistic or antagonistic.

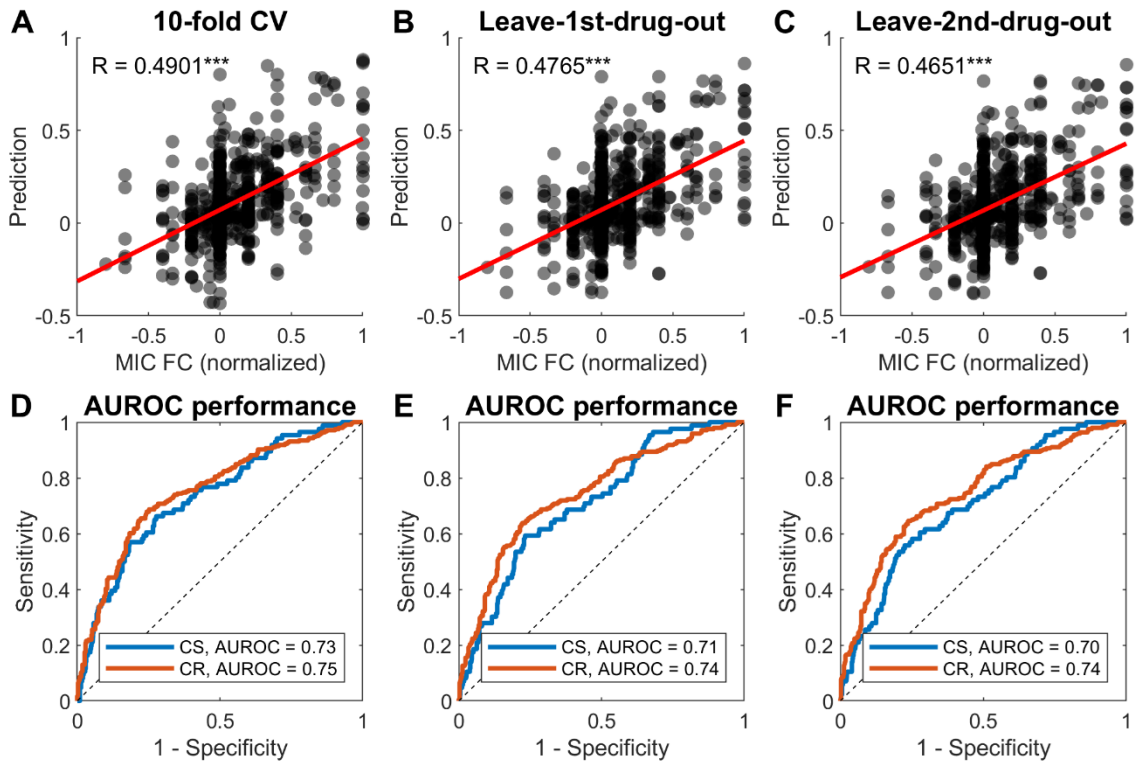

Fig. 3. 10-fold cross-validation and leave-out analysis results for sequential drug interactions. CARAMeL model performance for (**A, D**) 10-fold cross-validation, (**B, E**) leave-first-drug-out, and (**C, F**) leave-second-drug-out analyses. MIC: minimal inhibitory concentration; FC: fold-change, CS: collateral sensitivity, CR: cross-resistance. ***p-value < 10^{-3}.

Table 1. Metabolic pathways enriched amongst top predictors for the *E. coli* CARAMeL model. Pathway enrichment was determined based on 580 features explaining 95% of the variance in model predictions. These features mapped to 333 reactions in the *E. coli* GEM iJO1366, out of which 221 had differential flux that significantly distinguished between synergy and antagonism (two-sample t-test, p-value < 0.05). Based on this 221-reaction list, 10 pathways were found to be significantly enriched (hypergeometric test, p-value < 0.05). N = number of reactions in pathway, Ratio = N / total reactions in pathway, P-value = hypergeometric test p-value.

Pathway	N	Ratio	P-value
Glycine and Serine Metabolism	8	0.57	3E-07
Pyruvate Metabolism	6	0.60	3E-06
Purine and Pyrimidine Biosynthesis	9	0.36	2E-05
Transport, Inner Membrane	47	0.14	8E-05
Inorganic Ion Transport and Metabolism	19	0.17	1E-03
Nucleotide Salvage Pathway	22	0.16	1E-03
Glycolysis/Gluconeogenesis	6	0.27	2E-03
Cell Envelope Biosynthesis	20	0.15	4E-03
Folate Metabolism	3	0.33	5E-03
Methylglyoxal Metabolism	2	0.22	4E-02

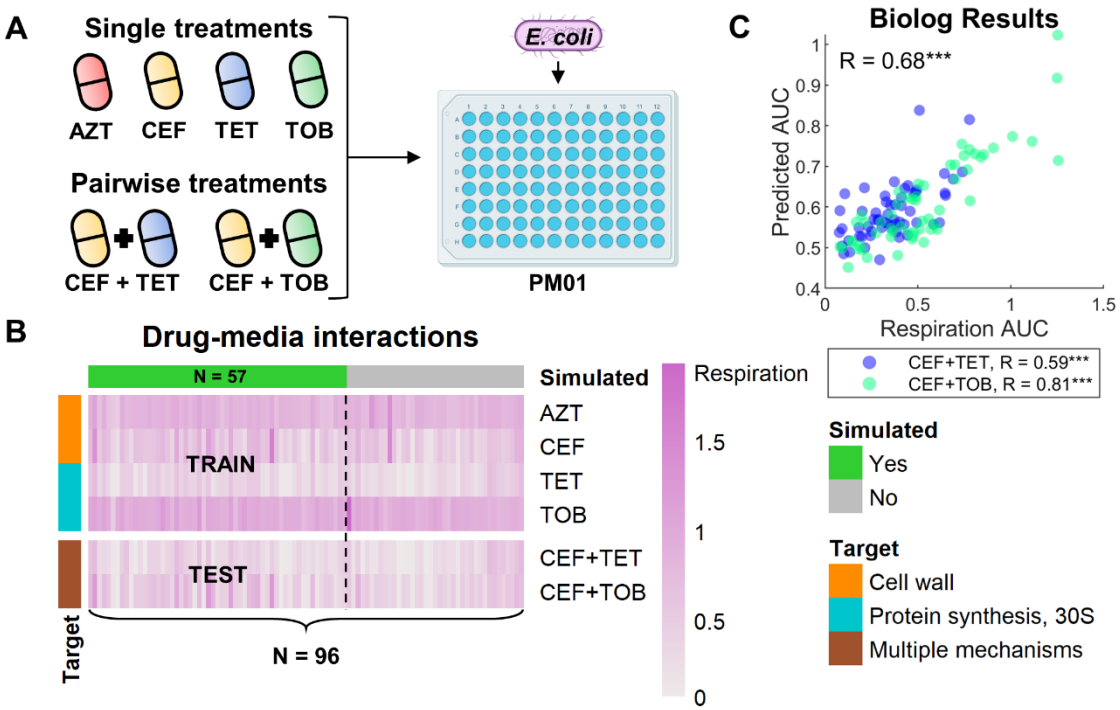

Fig. 4. CARAMeL accurately predicted drug interaction outcomes in 57 carbon sources. (**A**) *E. coli* was cultured in 96 carbon sources (Biolog PM01 plate), then treated with four single drug treatments (AZTreonam, CEFoxitin, TETtracycline, TOBramycin) and two pairwise treatments (CEF + TET and CEF + TOB). (**B**) Heatmap of metabolic activity (measured based on the respiration ratio between treatment vs. control) in response to all experimental perturbations. (**C**) Spearman correlation between experimental outcome and model predictions for all combinations in the test set are shown. *** p-value < 10^{-3}.

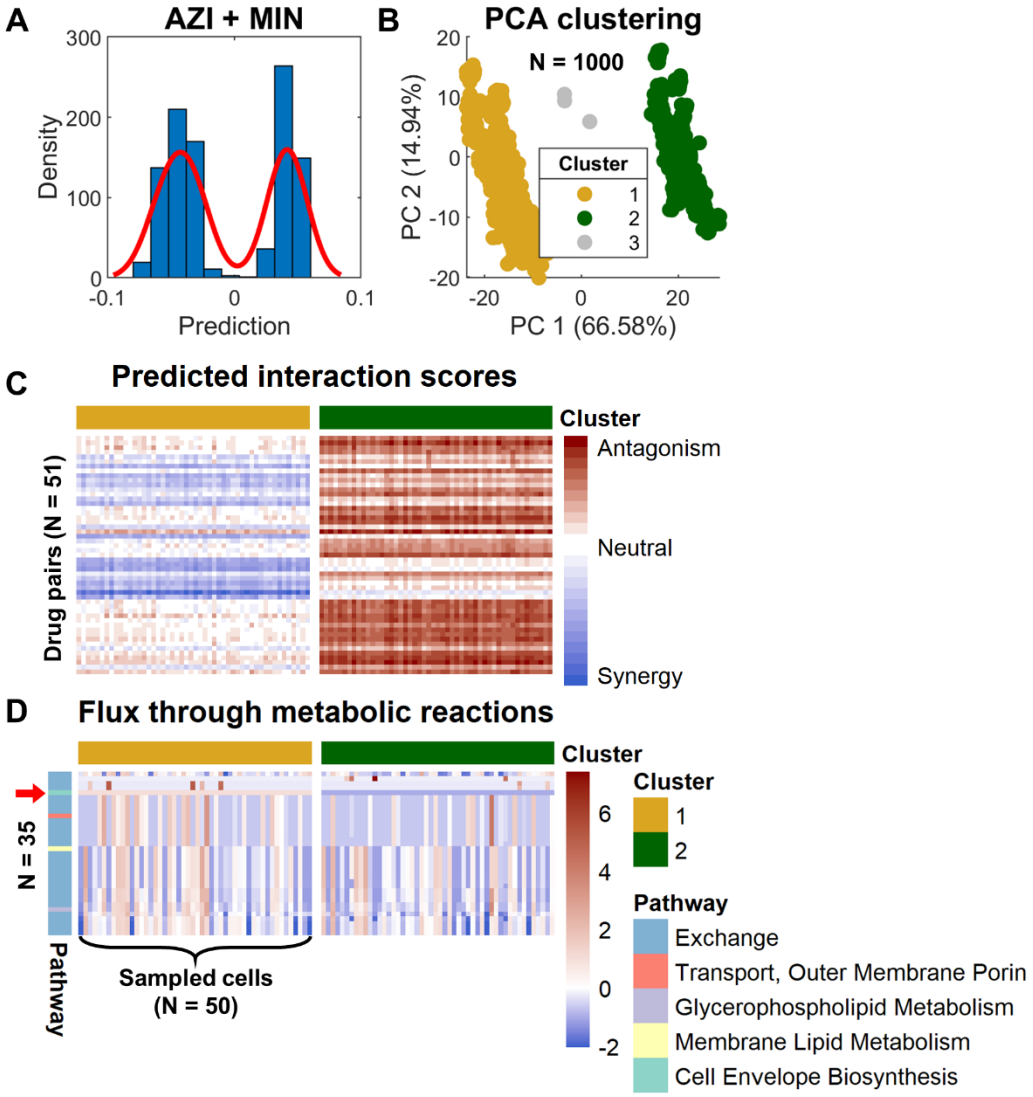

Fig. 5. Cell-specific pairwise combination therapy predictions. (**A**) Predictions for the vast majority of simultaneous drug interactions were found to be bimodally distributed across the cells (distribution for AZIthromycin + MINocycline shown as an example). (**B**) The cells were found to cluster into two distinct groups after applying principal component analysis (PCA) to the simultaneous prediction data. (**C**) A heatmap of 51 drug interactions predicted to yield synergy in at least one cell for both simultaneous and sequential cases reveals that cluster 1 cells were predicted to be more sensitive to combined drug treatments. (**D**) Out of 35 metabolic reactions with significantly different flux activity between the two clusters, the flux through one reaction (trans-2-decenoyl-ACP isomerase, cell envelope biosynthesis) perfectly aligned with the PCA-based cell grouping.

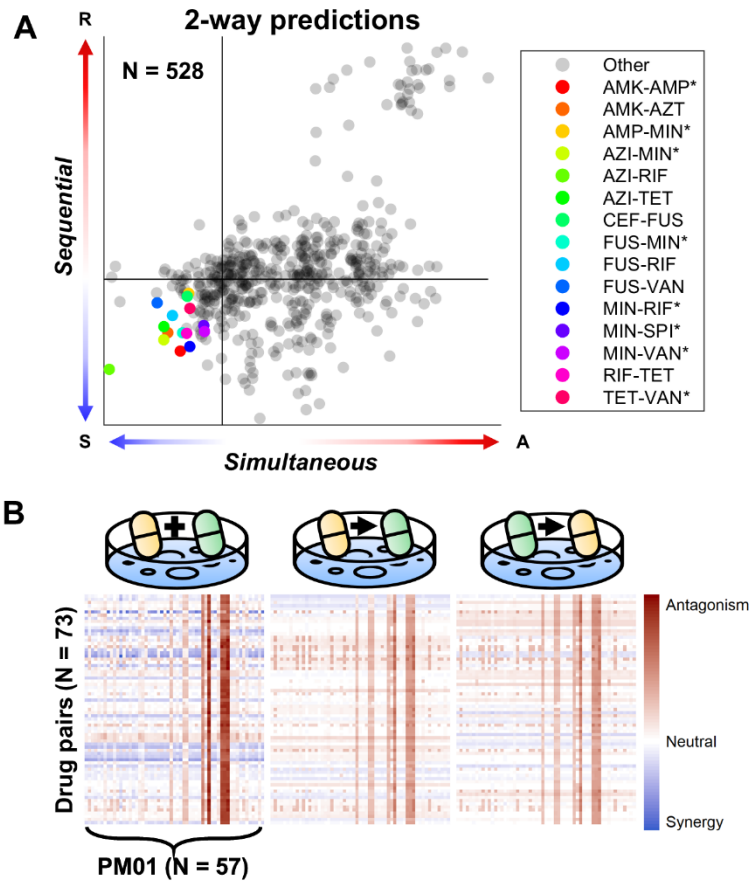

Fig. 6. Pairwise combination therapy prediction across 57 media conditions. (**A**) Out of 528 drug pairs, 73 were predicted to yield synergy (IS < 0) in at least one media condition for both the simultaneous and sequential cases (the top 15 robustly synergistic drug pairs are listed in the legend). (**B**) Heatmap of the predicted interaction scores for the 73 drug pairs across 57 media conditions and three interaction types ($D_1 + D_2$, $D_1 \rightarrow D_2$, $D_2 \rightarrow D_1$). Refer to ***Table S1*** for full descriptions on antibiotics used for *E. coli*.

Supplementary Materials for

A flux-based machine learning model to simulate the impact of pathogen metabolic heterogeneity on drug interactions

Carolina H. Chung & Sriram Chandrasekaran*

*Corresponding author. Email: csriram@umich.edu

This PDF file includes:

Figs. S1 to S9
Tables S1 to S6
Data S1 to S7

Other Supplementary Materials for this manuscript include the following:

Data S1 to S7 (Available through Dropbox at https://www.dropbox.com/s/859ebsx1drri1xc/data.xlsx?dl=0)

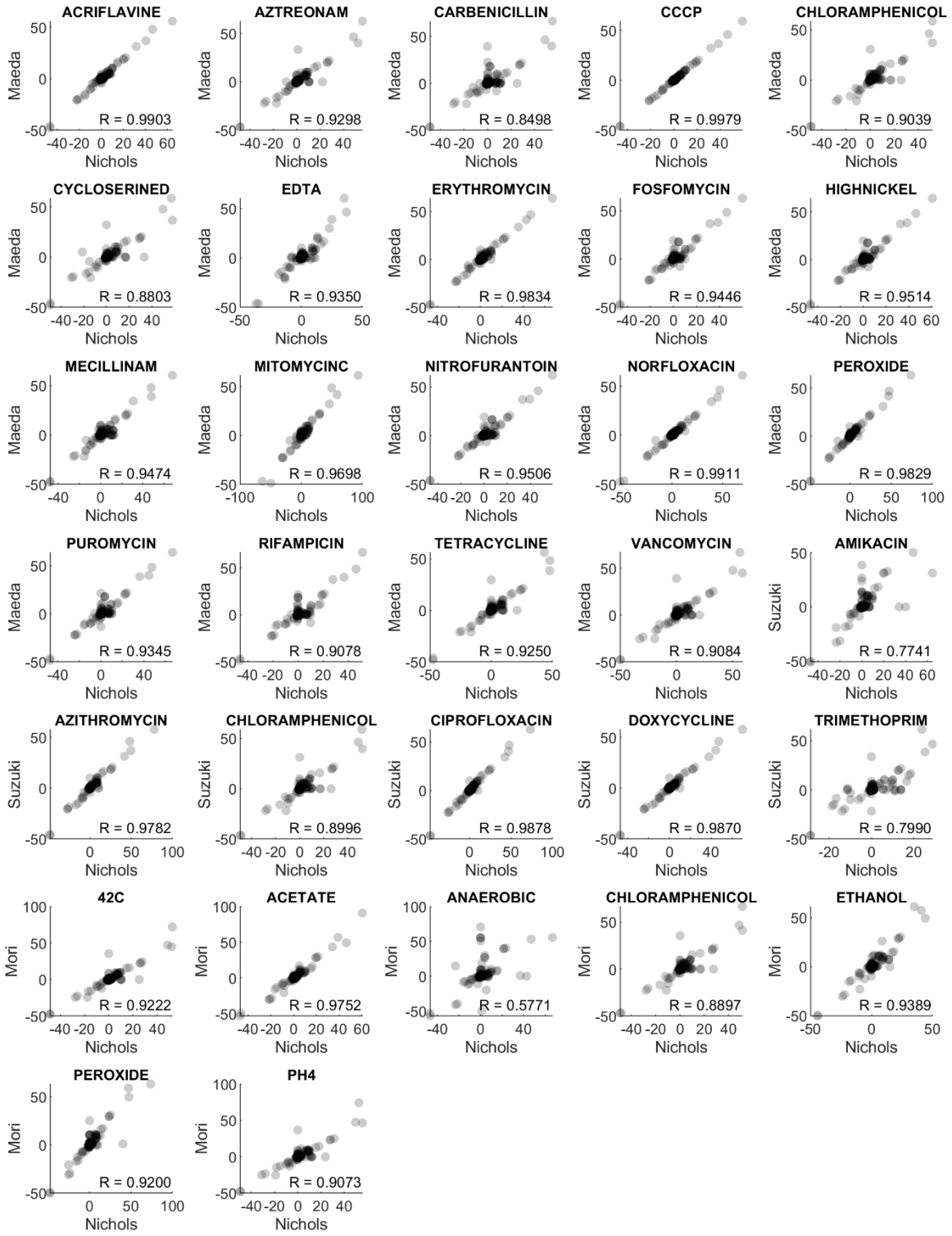

Fig. S1.

Flux profile comparison between different omic-based simulations. Correlations based on Pearson's method (all yielded $p << 10^{-3}$). All plots possess the same number of points (i.e., reactions, N = 2583). Nichols: chemogenomic-based (*50*), Maeda: transcriptomic-based (*52*), Suzuki: transcriptomic-based (*53*), Mori: proteomic-based (*54*).

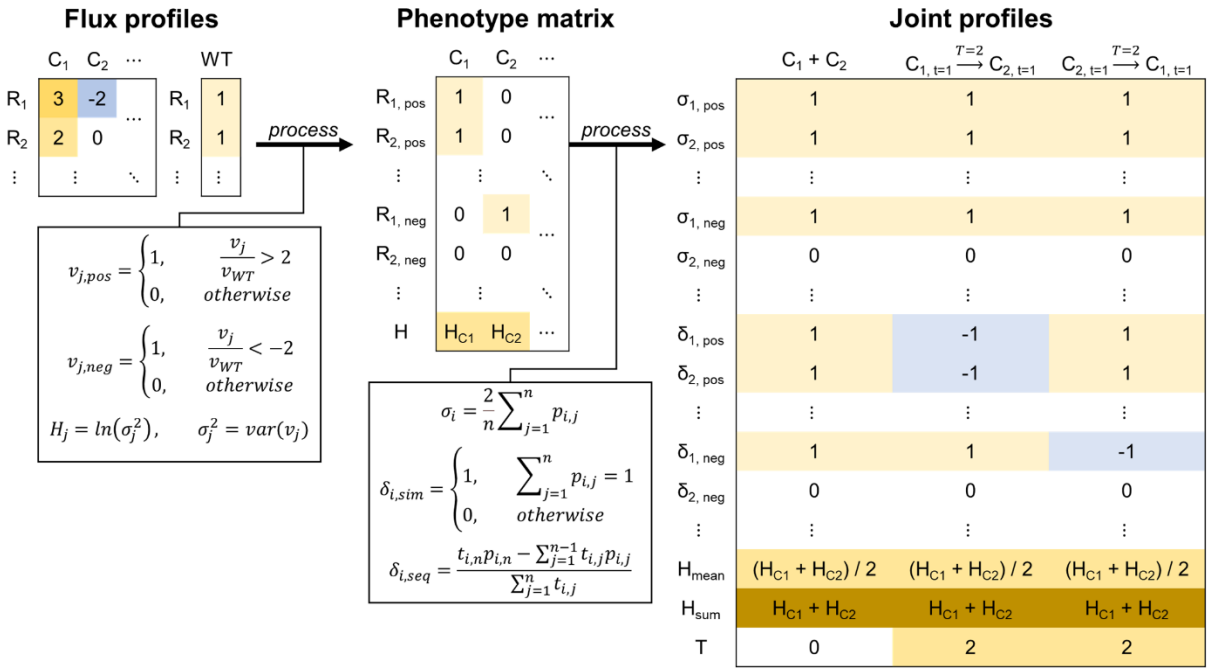

Fig. S2.

Schematic of flux data processing into joint profiles. Flux data (*v*) simulated from genome-scale metabolic models is binarized according to differential flux (either positive or negative) in comparison to wild type (WT, i.e., reference). These binarized flux profiles, along with the entropy (H) calculated for each condition (C), define the phenotype matrix which is subsequently processed into joint profiles. The sigma (σ) definition is the same between simultaneous (sim) and sequential (seq) interactions, while the delta (δ) definition differs depending on the interaction type. R = reaction, T = time interval, n = number of conditions in a combination.

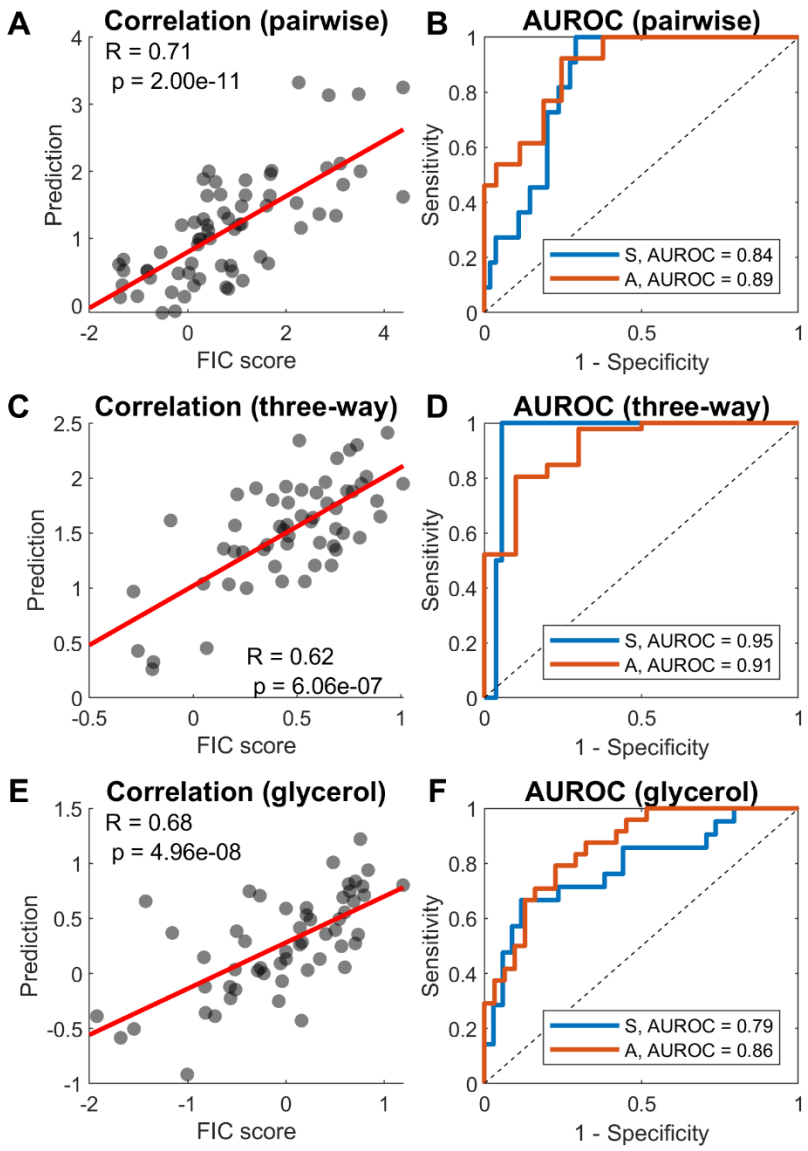

Fig. S3.

CARAMeL results for *E. coli* drug interaction data. Model performance results visualized as scatter and receiver operating curve (ROC) plots are shown for predicting (**A-B**) pairwise interactions, (**C-D**) three-way interactions, and (**E-F**) pairwise interactions in M9 glycerol. AUROC: area under the receiver operating curve. FIC: fractional inhibitory concentration, S: synergy, A: antagonism.

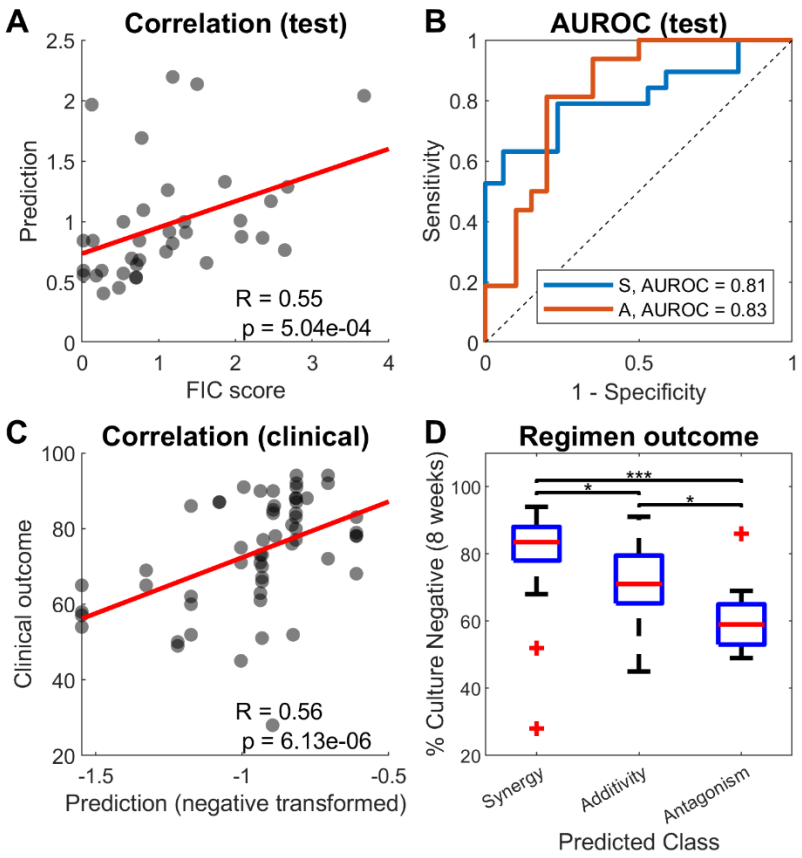

Fig. S4.

CARAMeL results for M. tb drug interaction data. (**A-B**) Model performance results visualized as scatter and receiver operating curve (ROC) plots are shown for predicting multi-drug interactions measured experimentally. (**C**) The model predictions for 57 TB regimens prescribed in clinical trials correlate with clinical efficacy. (**D**) Predictions classified as synergistic capture most of the efficacious treatments (sputum clearance > 80%). AUROC: area under the receiver operating curve, ** p-value < 0.01, *** p-value < 0.001 (unpaired t-test). FIC: fractional inhibitory concentration, AUROC: area under the receiver operating curve, S: synergy, A: antagonism.

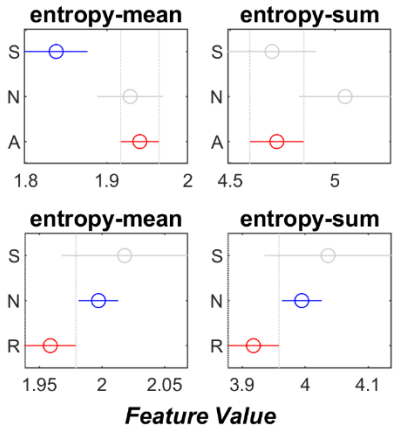

Fig. S5.

Distribution of cell-specific drug interaction predictions. The distribution for the top ten drug interactions with the largest variation across cells are shown for all time cases ($D_1 + D_2$, $D_1 \rightarrow D_2$, $D_2 \rightarrow D_1$). Refer to Table S1 for full descriptions on antibiotics used for *E. coli*.

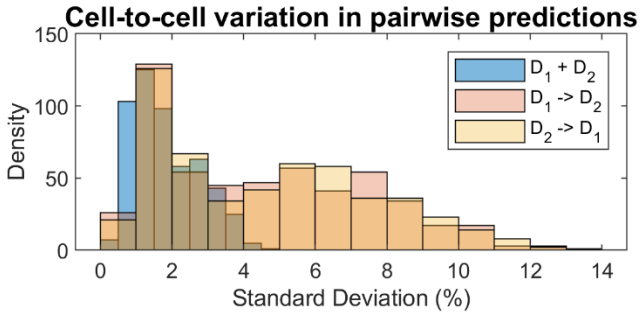

Fig. S6.

Distribution of cell-specific drug interaction predictions. The distribution for the top ten drug interactions with the largest variation across cells are shown for all time cases ($D_1 + D_2$, $D_1 \rightarrow D_2$, $D_2 \rightarrow D_1$). Refer to Table S1 for full descriptions on antibiotics used for *E. coli*.

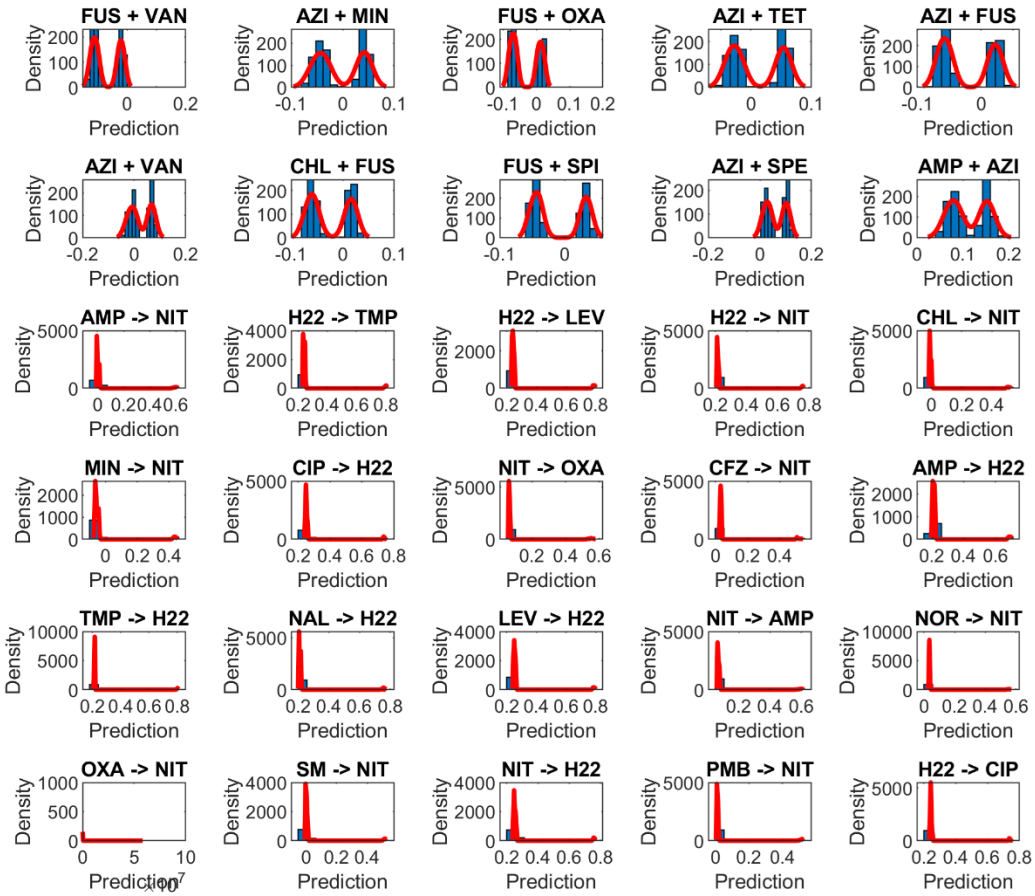

Fig. S7.

Distribution of cell-specific drug interaction predictions. The distribution for the top ten drug interactions with the largest variation across cells are shown for all time cases ($D_1 + D_2$, $D_1 \rightarrow D_2$, $D_2 \rightarrow D_1$). Refer to Table S1 for full descriptions on antibiotics used for *E. coli*.

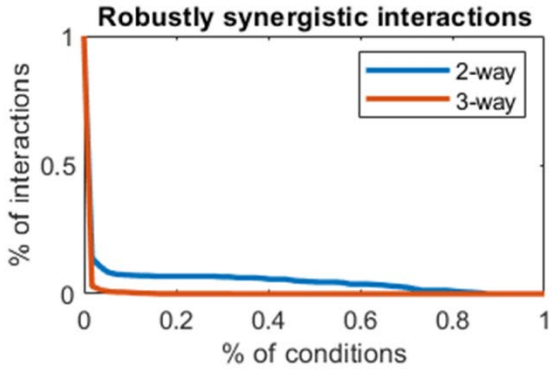

Fig. S8.

Distribution of cell-specific drug interaction predictions. The distribution for the top ten drug interactions with the largest variation across cells are shown for all time cases ($D_1 + D_2$, $D_1 \rightarrow D_2$, $D_2 \rightarrow D_1$). Refer to Table S1 for full descriptions on antibiotics used for *E. coli*.

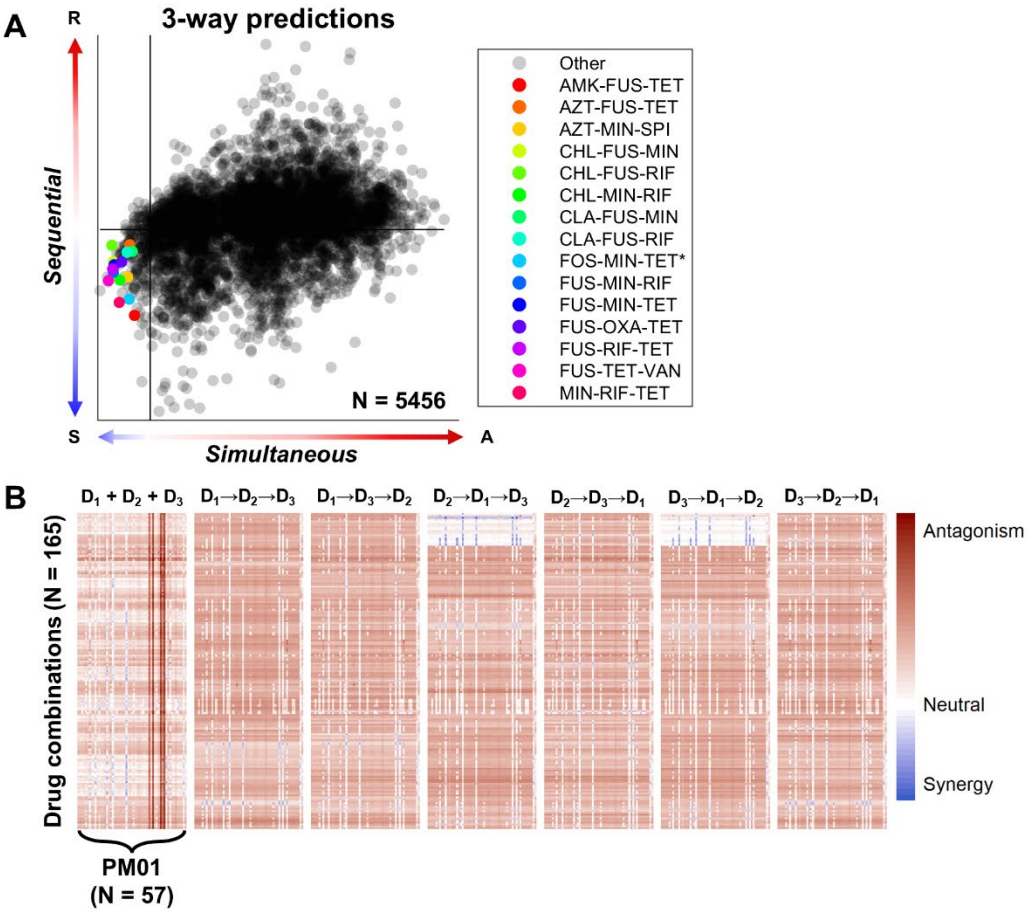

Fig. S9.

CARAMeL predictions for three-way combination therapy landscape. (**A**) Out of 5,456 unique three-way drug interactions, 165 were predicted to yield synergy (IS < 0) in at least one media condition for both the simultaneous and sequential cases (the top 15 robustly synergistic drug combinations are listed in the legend). (**B**) Heatmap of the predicted interaction scores for the 165 drug combinations across 57 media conditions and seven interaction types. Refer to Table S1 for full descriptions on antibiotics used for *E. coli*.

Table S1.

List of antibiotics used in *E. coli* drug interaction datasets. Abb.: abbreviation.

				Dataset					
Compound	**Abb.**	**Target**	**Class**	**Pair (train)**	**Pair (test)**	**Three-way**	**LB**	**Glucose**	**Glycerol**
Amikacin	AMK	Protein synthesis, 30S	Aminoglycoside	✓	✓				✓
Gentamicin	GEN			✓	✓				
Spectinomycin	SPE				✓		✓	✓	✓
Tobramycin	TOB			✓	✓				
Minocycline	MIN		Tetracycline			✓			
Tetracycline	TET			✓	✓		✓	✓	✓
Azithromycin	AZI	Protein synthesis, 50S	Macrolide			✓	✓	✓	✓
Chlarythromycin	CLA			✓	✓				
Erythromycin	ERY			✓	✓				
Chloramphenicol	CHL		Phenylpropanoid	✓	✓	✓	✓	✓	✓
Ciprofloxacin	CIP	DNA gyrase	Quinolone	✓	✓	✓			
Levofloxacin	LEV			✓	✓				
Nalidixic acid	NAL			✓	✓	✓			✓
Ampicillin	AMP	Cell wall	Beta-lactam			✓	✓	✓	✓
Aztreonam	AZT						✓	✓	✓
Cefoxitin	CEF			✓	✓				✓
Oxacillin	OXA			✓	✓				
Vancomycin	VAN		Glycopeptide		✓				
Fusidic acid	FUS	Elongation factor	Fusidane		✓				
Trimethoprim	TMP	Folic acid biosynthesis	Pyrimidine	✓	✓				
Rifampicin	RIF	RNA synthesis	Rifampin		✓	✓	✓	✓	✓
Nitrofurantoin	NIT	Multiple mechanisms	Furan	✓	✓	✓			
Triclosan	TRI		Phenol				✓	✓	✓
Hydrogen peroxide	H22	Oxidative stress	Stress	✓	✓				

Table S2.

List of antibiotics used in *M. tb* drug interaction datasets. Abb.: abbreviation, PTM: post-translational modification. * Putative mechanism.

				Dataset		
Compound	**Abb.**	**Target**	**Class**	**Train**	**Test**	**Clinical**
Sutezolid	SUTx	Protein synthesis, 23S	Oxazolidinone	✓		
Amikacin	AMK	Protein synthesis, 30S	Aminoglycoside	✓		
Kanamycin	KAN			✓		
Spectinomycin	SPE			✓	✓	
Streptomycin	SM			✓	✓	✓
Minocycline	MIN		Tetracycline	✓		
Tetracycline	TET			✓		
Azithromycin	AZI	Protein synthesis, 50S	Macrolide	✓		
Chlarythromycin	CLA			✓	✓	
Erythromycin	ERY			✓		
Roxithromycin	ROX			✓		
Linezolid	LZDx		Oxazolidinone	✓		
Chloramphenicol	CHL		Phenylpropanoid	✓		
Ciprofloxacin	CIP	DNA gyrase	Quinolone	✓	✓	
Levofloxacin	LEV			✓	✓	
Moxifloxacin	MOX			✓	✓	✓
Norfloxacin	NFX				✓	
Ofloxacin	OFX1			✓		✓
Novobiocin	NOV		Glycoside	✓		
Ampicillin	AMP	Cell wall	Beta-lactam	✓		
Oxacillin	OXA			✓		
Vancomycin	VAN		Glycopeptide	✓		
Cefaclor	CFL		Cephalosporin	✓		
SQ109	SQ109		Ethylenediamine	✓		
Isoniazid*	INH		Hydrazine	✓	✓	✓
Econazole	ECO		Imidazole	✓		
Pretomanid	PA824			✓		✓
Ethionamide*	ETH		Isonicotinic acid	✓		
Cycloserine D	CSD		Serine	✓		
PBTZ169	PBTZ169x		Thiazine	✓		
Capreomycin	CAP	Multiple mechanisms	Peptide	✓	✓	
Clofazimine*	CFZ		Phenazine	✓	✓	
Fusidic acid	FUS	Elongation factor	Fusidane	✓		
Ethambutol	EMBx	RNA synthesis	Ethylenediamine	✓		✓
Rifampicin	RIF		Rifampin	✓	✓	✓
Rifapentine	RIFP		Rifamycin			✓
Bedaquiline	BDQ	ATP synthase	Diarylquinoline	✓	✓	

Ethium bromide	EB	DNA structure	Phenanthridine	✓		
Pyrazinamide*	PZA	Fatty acid synthase	Pyrazine			✓
Menadione	MEN	PTM	Vitamin	✓		
Verapamil	VERx	Calcium channels	Phenethylamine	✓		
Thioridazine	THZ	Synaptic activity	Phenothiazine	✓		
Chlorpromazine	CPZ			✓	✓	

Table S3.

List of antibiotics used in sequential drug interaction datasets for *E. coli*. Abb.: abbreviation.

				Time scale		
Compound	**Abb.**	**Target**	**Class**	*T = 10*	*T = 21*	*T = 90*
Amikacin	AMK	Protein synthesis, 30S	Aminoglycoside	✓	✓	✓
Gentamicin	GEN			✓		✓
Spectinomycin	SPE				✓	
Streptomycin	SM			✓	✓	✓
Tobramycin	TOB				✓	
Doxycycline	DOX		Tetracycline		✓	✓
Minocycline	MIN			✓		✓
Tetracycline	TET			✓	✓	✓
Azithromycin	AZI	Protein synthesis, 50S	Macrolide	✓		✓
Erythromycin	ERY				✓	
Spiramycin	SPI				✓	
Chloramphenicol	CHL		Phenylpropanoid	✓	✓	✓
Ciprofloxacin	CIP	DNA gyrase	Quinolone	✓	✓	✓
Levofloxacin	LEV			✓		✓
Nalidixic acid	NAL			✓	✓	✓
Norfloxacin	NOR					✓
Ampicillin	AMP	Cell wall	Beta-lactam	✓	✓	
Cefoxitin	CEF				✓	
Ceftazidime	CFZ					✓
Amoxicillin	AMX			✓		
Sulfamonomethoxine	SMM	Folic acid biosynthesis	Sulfonamide		✓	
Trimethoprim	TMP		Pyrimidine	✓	✓	✓
Nitrofurantoin	NIT	Multiple mechanisms	Furan	✓	✓	
Fosfomycin	FOS	Cell wall biogenesis	Phosphonic acid	✓		
Fusidic acid	FUS	Elongation factor	Fusidane		✓	
Polymyxin B	PMB	Lipopolysaccharide	Peptide	✓		
Rifampicin	RIF	RNA synthesis	Rifampin	✓		✓

Table S4.

Drug information for Biolog experiment. Abb.: abbreviation, Conc.: drug concentration.

					Conc. (µg/mL)	
Compound	**Abb.**	**Target**	**Class**	**Type**	*Single*	*Pairwise*
Aztreonam	AZT	Cell wall	Beta-lactam	Bactericidal	0.03	-
Cefoxitin	CEF			Bactericidal	1.87	1.87
Tetracycline	TET	Protein synthesis, 30S	Tetracycline	Bacteriostatic	1.42	1.42
Tobramycin	TOB		Aminoglycoside	Bactericidal	0.15	0.15

Table S5.

Constraint-based modeling (CBM) parameter optimization results. * Chosen parameters for *M. tb*, + chosen parameters for *E. coli*. CV-R: 10-fold cross-validation correlation in the training dataset, GR-V: variance in growth rate, NG-P: percentage of no growth (GR = 0) conditions.

	CBM parameters			***E. coli* results**			***M. tb* results**		
	Kappa	*Rho*	*Epsilon*	*CV-R*	*GR-V*	*NG-P*	*CV-R*	*GR-V*	*NG-P*
	0.001	0.001	0.001	0.4055	0	0	0.3640	0	0
	0.01	0.01	0.001	0.3260	0.0074	0	0.5024	0.0003	0.0233
	0.1	0.1	0.001	0.4077	0.0634	0	0.4858	0.0005	0.0698
	1	1	0.001	0.4313	0.2022	0.3636	0.4409	0.0005	0.0698
	0.001	0.001	0.01	0.4260	0	0	0.4124	0	0
*	**0.01**	**0.01**	**0.01**	0.3739	0.0091	0	0.5207	0.0003	0.1395
	0.1	0.1	0.01	0.389	0.0670	0	0.5149	0.0004	0.6977
	1	1	0.01	0.4189	0.1866	0.4848	0.5242	0.0003	0.7674
	0.001	0.001	0.1	0.6406	0	0	0.5231	0	0
	0.01	0.01	0.1	0.4090	0.0095	0	0.4959	0.0001	0.5581
	0.1	0.1	0.1	0.3978	0.1056	0.0606	0.4792	0.0001	0.7907
	1	1	0.1	0.3869	0.1294	0.5152	0.4471	0.0001	0.8837
	0.001	0.001	1	0.3860	0	0	0.4620	0	0
+	**0.01**	**0.01**	**1**	0.6512	0.0113	0	0.5131	0.0003	0.6047
	0.1	0.1	1	0.6150	0.0994	0.0909	0.5270	0	0.8140
	1	1	1	0.6294	0.0765	0.5758	0.5099	0	0.9070

Table S6.

Benchmarking correlation results based on different constraint-based modeling (CBM) parameter choices. * Chosen parameters for *M. tb*, + chosen parameters for *E. coli*. R1: pairwise interactions (*27*), R2: three-way interactions (*28*), R3: glycerol interactions (*28*), R4: pairwise and three-way interactions (*29*), R5: pairwise to five-way TB clinical regimens (*40*).

	CBM parameters			***E. coli* results**			***M. tb results***	
	Kappa	*Rho*	*Epsilon*	*R1*	*R2*	*R3*	*R4*	*R5*
	0.001	0.001	0.001	0.2884	0.4441	0.5781	0.6370	0.5535
	0.01	0.01	0.001	0.5032	0.3516	0.5092	0.5873	0.4369
	0.1	0.1	0.001	0.4676	0.3666	0.5715	0.4946	0.2304
	1	1	0.001	0.5115	0.4001	0.5421	0.4717	0.4135
	0.001	0.001	0.01	0.3726	0.3636	0.5544	0.4638	0.5361
*	**0.01**	**0.01**	**0.01**	0.4525	0.2447	0.5900	0.5256	0.5445
	0.1	0.1	0.01	0.5390	0.2599	0.5514	0.4858	0.2642
	1	1	0.01	0.3899	0.3460	0.5915	0.4730	0.4124
	0.001	0.001	0.1	0.5421	0.5809	0.5669	0.5382	0.4483
	0.01	0.01	0.1	0.5829	0.5023	0.6281	0.6335	0.4263
	0.1	0.1	0.1	0.3313	0.3762	0.6382	0.5942	0.5140
	1	1	0.1	0.1545	0.4375	0.6745	0.5560	0.4982
	0.001	0.001	1	0.1772	0.1372	0.2668	0.3253	0.5380
+	**0.01**	**0.01**	**1**	0.6445	0.6216	0.6641	0.4884	0.4870
	0.1	0.1	1	0.6057	0.6536	0.6650	0.4947	0.4788
	1	1	1	0.6352	0.6169	0.6101	0.4939	0.3726

Data S1 to S7 (Available at https://www.dropbox.com/s/859ebsx1drri1xc/data.xlsx?dl=0)

Data S1. (separate file)
Top CARAMeL features explaining 95% of the variance in model predictions.

Data S2. (separate file)
iJO1366 reactions explaining 95% of the variance between actual and predicted interaction outcomes.

Data S3. (separate file)
Biolog phenotype microarray 1 (PM01) conditions.

Data S4. (separate file)
CARAMeL predictions for 2-way drug interactions for 1,000 individual cells and three treatment strategies (1 x simultaneous, 2 x sequential) (N = 1,584,000).

Data S5. (separate file)
Cell-to-cell variation in predictions for pairwise drug interaction outcomes (N = 528).

Data S6. (separate file)
CARAMeL predictions for 2-way drug interactions in 57 media conditions and three treatment strategies (1 x simultaneous, 2 x sequential) (N = 90,288).

Data S7. (separate file)
CARAMeL predictions for 2-way drug interactions in 57 media conditions and seven treatment strategies (1 x simultaneous, 6 x sequential) (N = 932,976).