Chemogenomics and orthology-based design of antibiotic combination therapies

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Chemogenomics and orthology-based design of antibiotic combination therapies

Sriram Chandrasekaran$^{1,2,3,*}$, Melike Cokol-Cakmak$^4$, Nil Sahin$^4$, Kaan Yilancioglu$^5$, Hilal Kazan$^6$, James J Collins$^{2,3,7,8,9}$ & Murat Cokol$^{4,10,11,**}$

Abstract

Combination antibiotic therapies are being increasingly used in the clinic to enhance potency and counter drug resistance. However, the large search space of candidate drugs and dosage regimes makes the identification of effective combinations highly challenging. Here, we present a computational approach called INDIGO, which uses chemogenomics data to predict antibiotic combinations that interact synergistically or antagonistically in inhibiting bacterial growth. INDIGO quantifies the influence of individual chemical–genetic interactions on synergy and antagonism and significantly outperforms existing approaches based on experimental evaluation of novel predictions in Escherichia coli. Our analysis revealed a core set of genes and pathways (e.g., central metabolism) that are predictive of antibiotic interactions. By identifying the interactions that are associated with orthogonal genes, we successfully estimated drug-interaction outcomes in the bacterial pathogens Mycobacterium tuberculosis and Staphylococcus aureus, using the E. coli INDIGO model. INDIGO thus enables the discovery of effective combination therapies in less-studied pathogens by leveraging chemogenomics data in model organisms.

Keywords chemogenomics; combination therapy; drug resistance; Mycobacterium tuberculosis; Staphylococcus aureus

Subject Categories Genome-Scale & Integrative Biology; Methods & Resources; Pharmacology & Drug Discovery

DOI 10.15252/msb.20156777 | Received 31 December 2015 | Revised 20 April 2016 | Accepted 21 April 2016

Introduction

Antimicrobial combination therapy is utilized clinically to reduce treatment time, to prevent the emergence of resistance, and to expand the spectrum of pathogens targeted (Dancey & Chen, 2006; Hopkins, 2008; Fischbach & Walsh, 2009; Bush et al, 2011; Worthington & Melander, 2013). However, individual drugs in combination can enhance or interfere with other drugs’ actions leading to unexpected effects. Therefore, mapping antibiotic interactions is essential for designing effective combination therapies (Dancey & Chen, 2006; Nichols et al, 2011; Roemer & Boone, 2013; Worthington & Melander, 2013). Systematic screens for drug interactions require testing various ratios of the component drugs (Hopkins, 2008), and the experimental identification of combinations with robust effects across a broad dose range is resource- and time-intensive, even for a small number of compounds (Zimmermann et al, 2007). The large space of candidate drugs, their dosage and chemical properties, and the nature of the target pathogen all affect the choice of the drug combination, making it a challenging and complex problem (Dancey & Chen, 2006; Hopkins, 2008). Because the number of possible drug combinations is astronomically high, new in silico methods are crucial for determining the most promising combinations (Hopkins, 2008; Lehár et al, 2008). This mandates the integration of genomic and chemical data using a computational framework. An integrated computational platform for predicting antibiotic interactions could pave the way for the rapid assessment of novel antimicrobial combinations before entry into clinical usage.

Here, we present an approach entitled Inferring Drug Interactions using chemogenomics and Orthology (INDIGO), which predicts antibiotic combinations that interact synergistically or antagonistically in inhibiting bacterial growth based on the chemogenomic profiles of the individual antibiotics. Chemogenomic profiling measures fitness of gene-knockout strains treated with bioactive compounds...
and provides unbiased insights into the mechanism of action of drugs (Bredel & Jacoby, 2004; Lehar et al, 2008; Ho et al, 2009; Nichols et al, 2011; Lee et al, 2014). We developed and utilized INDIGO to analyze a large compendium of publicly available chemogenomic data in *Escherichia coli* (Nichols et al, 2011) to identify predictive genetic features of antibiotic synergy and antagonism, and subsequently infer novel drug interactions. By finding orthologs of genes identified by INDIGO to be predictive of drug interactions in *E. coli*, we successfully predicted drug synergy and antagonism in the bacterial pathogens *Mycobacterium tuberculosis* and *Staphylococcus aureus*. These two pathogens cause a significant mortality worldwide and are frequently treated clinically using combinations of antibiotics (Liu et al, 2011; Ramon-Garcia et al, 2011). INDIGO greatly expands the capability of current drug-interaction prediction approaches by estimating the interaction outcomes in pathogens using chemogenomic data from model organisms.

## Results

### Experimental measurement of 105 interactions among 15 drugs as training data

To train INDIGO, 15 compounds with available chemogenomic profiles were selected (Nichols et al, 2011) (Table 1). These compounds, consisting of 14 antibiotics and a stress agent (henceforth referred to as drugs), covered a variety of classes including cell wall, transcription, translation, and DNA metabolism inhibitors. The interaction outcome of all pairwise combinations of these 15 drugs (105 pairs) was measured in duplicate. For each drug pair, individual drugs were combined in a 4 × 4 2-dimensional checkerboard assay, with each drug’s dose linearly increasing from 0 to near minimum inhibitory concentration (MIC) in each axis (Fig 1; Table 1; Appendix Fig S1; Materials and Methods).

The Loewe additivity model was used to quantify the drug interactions (Loewe, 1953). The Loewe additivity model defines a drug as additive or noninteracting with itself. Antagonistic and synergistic interactions are inferred based on the deviations from additivity. Synergy implies that the same amount of growth inhibition is achieved with a lower dose when both drugs are combined. For the Loewe model, we used as input the growth of the bacteria at different dose combinations of the individual drugs to generate a quantitative interaction score (α score) between the drug pair, with low scores (α < −0.5) corresponding to synergy and high scores (α > 1) corresponding to antagonism (Fig 1; Dataset EV1; Materials and Methods). This sensitive framework approximately corresponds to a bliss interaction score of −0.3 for α < −0.5 and +0.3 for α > 1 and provides a robust measure of drug interaction across doses (Appendix Fig S2). The measured interaction scores had strong correlation between the two biological replicates (rank correlation \( R = 0.81 \), \( P = 10^{-26} \)) and were consistent with existing interaction data (Appendix Fig S3; Appendix Supplementary Methods). Overall, among the 105 drug combinations, antagonistic drug interactions were more common than synergistic interactions. Further,

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Abbreviation</th>
<th>Target process</th>
<th>Symbol</th>
<th>Drug class</th>
<th>MIC (µg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amikacin</td>
<td>AMK</td>
<td>Protein synthesis, 30S</td>
<td><img src="image" alt="Aminoglycoside" /></td>
<td>Aminoglycoside</td>
<td>4</td>
</tr>
<tr>
<td>Gentamicin</td>
<td>GEN</td>
<td>Protein synthesis, 30S</td>
<td><img src="image" alt="Aminoglycoside" /></td>
<td>Aminoglycoside</td>
<td>3</td>
</tr>
<tr>
<td>Tobramycin</td>
<td>TOB</td>
<td>Protein synthesis, 30S</td>
<td><img src="image" alt="Aminoglycoside" /></td>
<td>Aminoglycoside</td>
<td>2.5</td>
</tr>
<tr>
<td>Tetracycline</td>
<td>TET</td>
<td>Protein synthesis, 30S</td>
<td><img src="image" alt="Tetracycline" /></td>
<td>Tetracycline</td>
<td>18</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>CHL</td>
<td>Protein synthesis, 50S</td>
<td><img src="image" alt="Phenylpropanoid" /></td>
<td>Phenylpropanoid</td>
<td>5</td>
</tr>
<tr>
<td>Clarithromycin</td>
<td>CLA</td>
<td>Protein synthesis, 50S</td>
<td><img src="image" alt="Macrolide" /></td>
<td>Macrolide</td>
<td>22</td>
</tr>
<tr>
<td>Erythromycin</td>
<td>ERY</td>
<td>Protein synthesis, 50S</td>
<td><img src="image" alt="Macrolide" /></td>
<td>Macrolide</td>
<td>13</td>
</tr>
<tr>
<td>Ciprofloxacin</td>
<td>CIP</td>
<td>DNA gyrase</td>
<td><img src="image" alt="Quinolone" /></td>
<td>Quinolone</td>
<td>0.01</td>
</tr>
<tr>
<td>Levofloxacin</td>
<td>LEV</td>
<td>DNA gyrase</td>
<td><img src="image" alt="Quinolone" /></td>
<td>Quinolone</td>
<td>0.013</td>
</tr>
<tr>
<td>Nalidixic acid</td>
<td>NAL</td>
<td>DNA gyrase</td>
<td><img src="image" alt="Quinolone" /></td>
<td>Quinolone</td>
<td>5</td>
</tr>
<tr>
<td>Trimethoprim</td>
<td>TRI</td>
<td>Folic acid biosynthesis</td>
<td><img src="image" alt="Pyrimidine" /></td>
<td>Pyrimidine</td>
<td>0.35</td>
</tr>
<tr>
<td>Oxacillin</td>
<td>OXA</td>
<td>Cell wall</td>
<td><img src="image" alt="ROS" /></td>
<td>Stress</td>
<td>250</td>
</tr>
<tr>
<td>Cefoxitin</td>
<td>CEF</td>
<td>Cell wall</td>
<td><img src="image" alt="ROS" /></td>
<td>Stress</td>
<td>250</td>
</tr>
<tr>
<td>H₂O₂</td>
<td>H22</td>
<td>Oxidative stress</td>
<td><img src="image" alt="ROS" /></td>
<td>Stress</td>
<td>250</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>NIT</td>
<td>Multiple mechanisms</td>
<td><img src="image" alt="Furan" /></td>
<td>Furan</td>
<td>12</td>
</tr>
<tr>
<td>Fusidic acid</td>
<td>FUS</td>
<td>Elongation factor—protein synthesis</td>
<td><img src="image" alt="Fusidane" /></td>
<td>Fusidane</td>
<td>800</td>
</tr>
<tr>
<td>Rifampicin</td>
<td>RIF</td>
<td>RNA synthesis</td>
<td><img src="image" alt="Rifampin" /></td>
<td>Rifampin</td>
<td>2.5</td>
</tr>
<tr>
<td>Vancomycin</td>
<td>VAN</td>
<td>Cell wall</td>
<td><img src="image" alt="Glycopeptide" /></td>
<td>Glycopeptide</td>
<td>180</td>
</tr>
<tr>
<td>Spectinomycin</td>
<td>SPE</td>
<td>30S protein synthesis</td>
<td><img src="image" alt="Aminocyclitol" /></td>
<td>Aminocyclitol</td>
<td>5.6</td>
</tr>
</tbody>
</table>

The antibiotics used in the test set for validation are highlighted in italics.
unsupervised clustering of interactions grouped the drugs based on their mechanism of action, as well as their bacteriostatic and bactericidal properties, consistent with previous studies (Yeh et al., 2006; Ocampo et al., 2014). In sum, these observations support the accuracy and robustness of the drug-interaction data used for training INDIGO.

**Framework for predicting drug–drug interactions using chemogenomic profiles**

The training data (input) for INDIGO consists of (i) chemogenomic profiles of individual drugs of interest and (ii) experimental interaction scores for the combinations of drugs. A chemogenomic profile of a drug is an array of fitness scores for gene-deletion strains treated with the drug of interest compared to the wild-type strain. We transformed the chemogenomic profile of a drug into a binary sensitivity profile by identifying the deletion strains that are significantly sensitive to a drug (Materials and Methods). The sensitivity profiles of individual drugs in a drug combination are combined by INDIGO using Boolean operations to create a joint profile. In this framework, the union (sigma score) and intersection (delta score) operations capture similarity and uniqueness in the mechanism of action of the individual drugs. A machine learning algorithm called random forests is then used to build a predictive model that links the interaction outcome of drug combinations to the joint chemogenomic profile of the drug pair (Materials and Methods; Fig 2). The random forest algorithm builds an ensemble of decision trees using the training dataset and outputs the mean prediction of the individual trees; it also identifies genes in the chemogenomics data that are most predictive of drug interactions. INDIGO learns the mechanism of drug interactions from the chemogenomics data in an unbiased fashion by using the random forest algorithm.

INDIGO thus takes a systems approach to predict drug interactions; the approach quantifies the contribution of individual genes with a chemical–genetic interaction on the overall drug-interaction
To benchmark INDIGO, the predictions were compared to a chemogenomics-based approach used to infer antifungal interactions, which predicted synergy based on chemogenomic profile similarity of the individual drugs (Jansen et al., 2009; Materials and Methods). In this approach, the similarity of two drugs was evaluated by measuring the Pearson’s correlation between the individual drugs’ chemogenomic profile, or through the total overlap between the genes present in the chemogenomic profile of individual drugs. These approaches performed poorly in predicting experimental interaction scores compared to INDIGO ($R = 0.14$, $P$-value $= 0.25$, Fig 3D–F; Appendix Fig S4). INDIGO also significantly outperformed the O2M algorithm (Brown et al., 2014), which is an extension of the Jansen et al.’s overlap-based approach (Appendix Fig S5).

Similarity and overlap-based approaches are less effective in predicting interaction outcomes for new classes of drugs and lack a model for antagonism. Thus, taking into account the identity of the individual genes in the chemogenomic profile and accounting for both similarity and dissimilarity increases the ability to predict drug interactions by INDIGO. Our data show that drugs with similar targets and chemogenomic profiles can have both synergistic and antagonistic outcomes. For example, combinations of tobramycin–spectinomycin, tobramycin–gentamicin, and fusidic acid–clarithromycin share similar chemogenomic profiles and target processes, yet have antagonistic, neutral, and synergistic outcomes, respectively (Fig EV1).

Importantly, INDIGO quantitatively predicts drug interactions and its predictive ability was not influenced by any specific threshold for inferring synergy or antagonism (Appendix Fig S6), it was robust to the choice of methods used to evaluate predictions (Materials and Methods; Appendix Fig S4), and its predictions were more likely to be synergistic than a random screen ($P < 10^{-19}$; t-test; Appendix Fig S7). Further, the results were robust to the removal of the outlier compound hydrogen peroxide that had promiscuous antagonistic interactions (Appendix Fig S8), and led to similar accuracy based on cross-validation analysis (Appendix Figs S9, S10, S11 and S12). Cross-validation analysis also revealed that predictions were consistently accurate across most drugs and identified
hydrogen peroxide as leading to the largest errors in predictions (Appendix Figs S8 and S9). This suggests that the chemogenomic profile of hydrogen peroxide may not accurately reflect its mechanism of interaction. In sum, INDIGO successfully predicts novel drug interactions as assessed by several metrics.

Global interaction landscape of compound combinations

To further strengthen our model, INDIGO was re-trained on the entire experimental drug–drug interaction dataset, including the test-set interactions (171 drug pairs). This model was then applied to the entire compendium of 73 drugs and 53 stress-agent chemogenomic profiles from Nichols et al. By mining this dataset (2,628 drug–drug interactions and 3,869 drug–stress agent interactions), we observed the well-known antagonistic drug interactions between bactericidal antibiotics targeting DNA synthesis (such as ciprofloxacin and other quinolones) and bacteriostatic antibiotics targeting translation (such as tetracycline) (Ocampo et al., 2014; Lobritz et al., 2015). Our interaction dataset also included nonantibiotic drugs that could be repurposed as antibiotic adjuvants (Gill et al., 2015). Interestingly, membrane stress agents, such as the detergents sodium dodecyl sulfate (SDS) and triclosan, were found to be synergistic with a large number of antibiotics (Fig 4). These predictions are consistent with a previous high-throughput drug combination screen in E. coli (Farha & Brown, 2010).

INDIGO correctly predicted synergy between commonly used, clinically validated antibiotic combinations. Vancomycin was synergetic with beta-lactam antibiotics (ampicillin and aztreonam), as was the combination of ampicillin and gentamicin, which has been a mainstay combination treatment for intra-abdominal infections (Solomkin et al., 2010; Liu et al., 2011). Rifampicin, which is frequently used in combination with other antibiotics, particularly in the case of biofilm-associated infections, is predicted to be synergetic with oxacillin, macrolide, and tetracycline antibiotics, affirming the clinical efficacy of this drug with other companion antibiotics (Forrest & Tamura, 2010). Our approach thus consistently and rapidly identifies synergy among antibiotic combinations that required clinical validation over decades of study.

Genetic predictors of synergy and antagonism

Surprisingly, a small set of genes account for most of the predictive ability of our model. Genes were assigned an importance score by INDIGO proportional to their relative contribution in predicting drug interactions. The top 81 genes accounted for 50%, the top 222 accounted for 75%, and the top 581 accounted for 95% of variance.
in the predicted data (Fig EV2; Dataset EV3). We defined the set of top 250 genes that contributed to over 75% of the variance predicted by the model as drug interaction-related (DIR) genes.

The presence of these DIR genes in the sensitivity profile of the drugs also correlated with either synergy or antagonism (Table 2A). Overall, the interaction outcomes for each drug combination depend on complex interactions between many genes. Nevertheless, the presence of these top genes can be a strong predictor of synergy or antagonism. For example, we find genes associated with synergy to be enriched in the chemogenomic profile of triclosan (P-value = 10^{-10}, hypergeometric test), explaining its promiscuous synergy. By ranking genes used by INDIGO, we can determine the dominant pathways that are commonly used across drug classes. KEGG pathway enrichment analysis among the DIR genes identified processes that are targeted by the drugs, and pathways known to be involved in mediating drug interactions and resistance such as efflux pumps [Webber & Piddock, 2003] (Table 2B). Interestingly, in addition to known targets of antibiotic action—translation, cell wall synthesis, DNA replication, folic acid biosynthesis, and RNA metabolism—genes related to central energy metabolism and drug transport were also enriched in the DIR genes. Notably, lipopolysaccharide biosynthesis and oxidative phosphorylation were found as two of the top ten pathways associated with drug interactions. This is in agreement with a recent study that measured antibiotic interactions in all nonessential gene-deletion strains of E. coli and found that strains harboring deletions in these pathways resulted in altered interaction outcomes (Chevereau & Bollenbach, 2015). This further affirms the validity and significance of the DIR genes identified by INDIGO.

Predicting drug interactions across species using orthology mapping

While chemogenomic profiling data are widely available for E. coli, such datasets are often not available for other microbes, including many pathogens. We hypothesized that INDIGO could be applied to clinically relevant pathogens by identifying orthologs of E. coli DIR genes. We tested this hypothesis using the Gram-positive pathogen S. aureus. Genes that were orthologous between E. coli and S. aureus were mapped onto the INDIGO model (Materials and Methods). Interestingly, the top 250 DIR genes in the model were significantly enriched for genes orthologous between E. coli and S. aureus (P-value = 0.001, hypergeometric test; total orthologs = 844 genes). This observation suggests that interaction outcomes in S. aureus can be predicted using the E. coli INDIGO model. We hypothesized that interactions that depend on orthologous genes
would be conserved between the two species, while those that are predicted by INDIGO to depend on non-orthologous genes would not be conserved. Therefore, we used the *E. coli* chemogenomic profiles of the orthologous genes to predict drug-interaction scores for *S. aureus* (Materials and Methods). To quantify the relative contribution of non-orthologous genes on each drug-interaction outcome, the chemogenomic scores of non-orthologous genes in the *E. coli* INDIGO model were set to take minimal value (Materials and Methods). This process corresponds to deleting these genes from the *E. coli* INDIGO model. The predictions from this modified model were then compared with the original *E. coli* model predictions. The difference in scores for each interaction, called the deviation score, allowed us to identify drug interactions most sensitive to the state of the non-orthologous genes and therefore less likely to be conserved between *E. coli* and *S. aureus*.

We predicted interaction and deviation scores for all 171 drug combinations that were experimentally tested in *E. coli* using this orthology framework. To experimentally validate the predictions in *S. aureus*, 10 drugs from the *E. coli* dataset were chosen; these drugs belong to different classes that were predicted by INDIGO to have the greatest range of variation in their interactions between *E. coli* and *S. aureus* (Materials and Methods; Fig 5, Appendix Fig 12).
S. aureus using the same setup used for E. coli (Fig SB). Overall, we found that there is a small yet significant correlation between the experimentally measured drug-interaction scores of the two species \( R = 0.39; P\text{-value} = 0.008 \), Appendix Fig S14). This is in agreement with the observation that the top predictive genes are conserved between the two species. Consistent with our hypothesis, interactions that are associated with non-orthologous genes as quantified by the deviation score have stronger differences in their interaction outcome between the two species \( R = 0.52; P\text{-value} = 10^{-4} \); Appendix Table S4; Fig 5C). Our analysis indicates that although some drug interactions are conserved between the two species, there are many exceptions. These exceptions can be identified by INDIGO based on the deviation score. By taking into account these deviations, interaction outcomes in S. aureus can be directly estimated using INDIGO \( R = 0.47 \) between experimental and predicted S. aureus interactions; \( P\text{-value} = 10^{-4} \); Figs 5D and EV3). For example, INDIGO correctly discovered that the oxacillin–tetracycline combination is synergistic in E. coli, but antagonistic in S. aureus. These results are noteworthy given that S. aureus interaction data were not used for training the model, and chemogenomics data from a different species were used to predict interactions.

To further validate the applicability of our approach to other organisms, INDIGO was applied to predict antibiotic interaction outcomes in the pathogen Mycobacterium tuberculosis using E. coli chemogenomics data; 909 genes in M. tuberculosis were orthologous with genes in our E. coli dataset. Again, these conserved genes are enriched among the top 250 DIR genes in our experimental dataset. Similar to S. aureus predictions, we found strong correlation between the predicted and experimentally observed interaction outcomes in M. tuberculosis \( R = 0.54; P\text{-value} = 0.006 \); Figs 5E and EV3; Appendix Table S5). Many combinations that are synergistic in E. coli are also predicted to be synergistic in M. tuberculosis. For example, combinations of spectinomycin with clarithromycin or azithromycin are synergistic in both species. Remarkably, 20 out of the 24 antibiotic combinations for which predictions were made for M. tuberculosis were not in our experimental dataset for E. coli and

**Figure 5.** Predicting drug-interaction outcomes in Staphylococcus aureus and Mycobacterium tuberculosis using INDIGO.

- **A** Overview of our framework for predicting drug-interaction data in S. aureus using Escherichia coli drug-interaction and chemogenomics data. The model predicts interactions that would differ between the two species.
- **B** Experimental S. aureus drug-interaction data. Average interaction scores of two replicates are shown. The target process of the antibiotics is depicted as cartoons and is described in Table 1.
- **C** Scatter plot of predicted difference in interaction score between the two species and experimentally measured difference in the interaction score. Interactions that have the largest differences in interaction scores between the two species are highlighted.
- **D** Box plots show predicted interaction score by INDIGO for experimentally measured synergistic and antagonistic interactions in Staphylococcus aureus \( R = 0.47; P\text{-value} = 0.001 \); one-way ANOVA \( P\text{-value} = 0.03 \).
- **E** Box plots show predicted interaction score by INDIGO for synergistic and antagonistic interactions in Mycobacterium tuberculosis \( R = 0.54; P\text{-value} = 0.006 \); one-way ANOVA \( P\text{-value} = 0.02 \). Mycobacterium tuberculosis interaction data were curated from literature and are qualitative in nature. See Fig EV3 for a more detailed plot with drug labels.

Source data are available online for this figure.
hence could not have been predicted from experimental data directly.

We have thus established that we can reliably estimate drug-interaction outcomes in E. coli, S. aureus, and M. tuberculosis. To assess the organism-specific variability of interactions, the entire compendium of interactions between 73 drugs and 53 stress agent from Nichols et al was compared across the three species (Dataset EV2). Combinations involving tetracyclines and quinolones were found to be more likely to vary across species (Fig EV4). Consistent with this observation, no significant enrichment was found for orthologous genes among the DIR genes in the chemogenomic profile of these antibiotics (P-value > 0.05, hypergeometric test). Among all predictions, 3 and 5% were predicted to exhibit strong broad-spectrum synergy and antagonism in all three species, respectively. In contrast, only 1% of the drug pairs showed a mix of synergy and antagonism in the three species.

Discussion

Despite an alarming rise in resistance, few new antibiotics have been developed in the past decade, necessitating the increased use of antibiotic combination therapies. In this study, we developed and validated a computational framework that can successfully predict synergistic and antagonistic drug interactions in three bacterial species. In addition to the 171 drug pairs experimentally assessed in E. coli, drug-interaction predictions were validated for 45 drug pairs in S. aureus. In agreement with previous studies in bacteria and yeast, antagonistic interactions were more prevalent than synergistic interactions. This underscores the difficulty of selecting drug combinations that do not adversely affect clinical outcomes.

An important finding from this study is that drug interactions can be predicted based on chemogenomic profiles of individual drugs. This suggests that cellular sensitivity to drug combinations can be abstracted by a linear model of individual drug-sensitivity profiles (Pritchard et al., 2013). This is consistent with the observation that protein and transcriptional dynamics of cells treated with drug combinations are a linear combination of the individual drug effects (Geva-Zatorsky et al., 2010; Chevereau & Bollenbach, 2015). A large resource of chemogenomic profiling is now available for E. coli, S. cerevisiae, and human cancer cell lines; INDIGO can thus be applied to a wide range of drugs across different systems (Flores et al., 2005; Tamae et al., 2008; Girgis et al., 2009; Ho et al., 2009; Mueller et al., 2011; Nichols et al., 2011; Lee et al., 2014).

INDIGO assesses the influence of hundreds of individual chemi-genetic interactions on drug-drug interaction outcomes. In contrast, existing chemogenomics-based approaches determine synergy based on a single aggregate metric of drug similarity. Our gene-centric model has also enabled us to apply our E. coli drug-interaction model to predict outcomes in other bacterial species. Genes identified to be predictive of drug interactions by INDIGO (DIR genes) were involved in the processes targeted by the drugs or processes such as drug transport and bacterial metabolism that are known to influence drug interactions (Chevereau & Bollenbach, 2015). DIR genes connected to energy metabolism might be necessary for facilitated transport of the drugs (Allison et al., 2011) or may be connected to cellular damage induced by antibiotic-related redox changes (Kohanski et al., 2007, 2010; Dwyer et al., 2014; Lobritz et al., 2015). Overall, we found that 15% of the E. coli genome account for 95% of the predictive ability of our model, and this core set is enriched for genes that are conserved across species.

Our study goes beyond existing drug combination discovery approaches by leveraging chemogenomic data in E. coli to infer antibiotic interactions in S. aureus and M. tuberculosis. Our approach may be especially relevant for M. tuberculosis, where combination therapy is critical and experimental screens are difficult, expensive, and time-consuming. Identifying synergistic combinations can help reduce the duration of tuberculosis combination therapy, which currently takes 6 months to 2 years (Ramon-Garcia et al., 2011).

INDIGO identified combinations that are synergistic in E. coli, but are surprisingly antagonistic in S. aureus. The discovery of such combinations that are selectively synergistic (narrow-spectrum synergy) could be used to improve specificity of antibiotic combination therapies to target pathogen strains without significantly affecting beneficial microbes. Strong synergistic combinations in model organisms that are predicted by INDIGO to be conserved in pathogens could be prioritized for testing as “broad-spectrum synergistic combinations”. Species-specific interaction outcomes have also been observed for drug combinations against fungal pathogens, similar to our observation in bacteria (Wildenhain et al., 2015).

While this capability to predict drug interactions by mapping orthologous genes is a significant advance for this field and can be extended to any organism with genome sequence information, a limitation of this approach is that INDIGO cannot explicitly model the contribution of pathogen DIR genes that are not present in E. coli. Our interaction estimates for these systems could be further improved with the availability of chemogenomics data for these pathogens and by performing drug-interaction measurements directly in pathogens as training data for INDIGO. Predictive accuracy of INDIGO can be enhanced in the future through new technologies for the accurate measurement of drug-drug and chemical–genetic interactions, performing chemogenomic screens with essential genes (Cameron & Collins, 2014), and by harnessing drug physicochemical properties (Yilancioglu et al., 2014) and chemical structure (Wildenhain et al., 2015). INDIGO can complement theoretical models for predicting multidrug interactions from pairwise interactions (Wood et al., 2012), and kinetic modeling-based approaches, which are currently restricted to small pathways due to a lack of known kinetic parameters and drug targets (Singh et al., 2015). By transforming drug responses and interactions into the genomic space, our study provides a framework for genomics-driven drug combinations discovery.

Materials and Methods

Drug-interaction assays

All drug-interaction experiments were conducted using a 4 × 4 checkerboard assay, where the concentration of each drug increases linearly in each axis starting with 0 (no drug). Escherichia coli MG1655 or S. aureus ATCC 29213 were grown in tryptic soy buffer media in the presence of antibiotics. In E. coli experiments, optical density (OD) measurements were done every 15 min for 12 h in a TECAN Infinite F200 microplate reader. Growth rate was then
estimated based on the area under the growth curve. Since the E. coli area under the growth curve was highly correlated with the end-point readings, only end-point OD measurement after 12 h was used for S. aureus to expedite experiments. For each drug-interaction experiment, we ensured that for each individual drug, there were > 50% inhibition at the highest dose used and < 50% inhibition at the lowest dose used as a quality control check. All drugs were purchased from Sigma. MICs for each drug are provided in Table 1 and Appendix Table S2.

### Quantifying drug interactions

Interactions were quantified based on the isophenotypic growth contour method described in Cokol et al (2011) based on the Loewe’s additivity model (Loewe 1953). The growth contour (line of constant growth) is linear for noninteracting drugs. Depending on the concavity or convexity of the growth contour, interactions are classified as synergistic or antagonistic. Our null hypothesis based on the Loewe’s interaction model is that a drug is noninteracting with itself. Deviations from this null model lead to either synergy or antagonism. The advantage of this approach is that there is no underlying assumption that the data should be normally distributed or that neutral interactions should be more common than synergy or antagonism.

### INDIGO cross-validation and controls

#### Alternate metrics to evaluate the predictive ability

We assessed the overall accuracy of the predictions by measuring the area under the receiver-operating characteristics (ROC) curve. INDIGO quantitatively predicted drug interactions with high precision, significantly better than random, and with equal accuracy for both synergy and antagonism (AUC for synergy = 0.79, P-value = 10^{-15}; AUC for antagonism = 0.8, P-value = 10^{-15}). Both the correlation-based and profile overlap-based approaches from Jansen et al were significantly less accurate than INDIGO in predicting drug interactions (AUC for synergy = 0.6; P-value = 0.003 for profile correlation, and AUC synergy = 0.64; P-value = 10^{-5} for profile overlap; Fig 3E and F). In addition to these metrics for evaluating accuracy, we also measured the probabilistic concordance index, PCC (Bansal et al, 2014), which quantifies the concordance between the ranking of drug pairs in the predicted and measured data, accounting for variance in the experimental measurement of the drug interactions. INDIGO’s predictions were significantly better than predictions based on profile similarity and random permutation of the data (INDIGO PCC = 0.68, P-value < 10^{-6}; profile similarity PCC = 0.53; P-value = 0.15). We also tested our model in a single-dose study involving 21 antibiotic pairs (Yeh et al, 2006) and found that INDIGO outperformed similarity-based approaches in predicting drug interactions [rank correlation R = 0.59, P-value < 10^{-15} for INDIGO; no significant correlation for correlation and profile overlap (Appendix Fig S15)].

#### Estimating robustness to training data

In addition to the test-set predictions, the model’s predictive ability was also assessed through cross-validation. Two types of cross-validation analyses were performed. In tenfold cross-validation, 10% of the interactions were randomly blinded and predicted by the model based on information from the remaining 90% of the interactions (Appendix Fig S10). This represents a scenario where new interactions are inferred for drugs for which some interactions with other drugs are already known. We also performed another kind of cross-validation where each drug was removed from the network and its interactions were inferred based on its chemogenomic profile and interactions of the remaining drugs. This is a more stringent test than tenfold cross-validation and represents the situation where interactions are inferred for a new drug based on its chemogenomic profile (Appendix Figs S9 and S10). Through these analyses, we found that INDIGO could accurately predict interactions with compounds that belong to novel chemical classes or with distinct mechanisms of action. Nevertheless, we found that the prediction accuracy could be further improved by choosing drugs representative of different classes in the training set. In addition, the presence of a few known interactions about a drug of interest can greatly increase predictive ability. Finally, INDIGO was able to predict synergistic interactions with high accuracy in the test set, which had significantly more synergistic interactions (13 synergies and 25 antagonisms) compared to the training data (14 synergies and 56 antagonisms), suggesting that it is robust to the biased distribution of interactions in the training set.

#### Negative and positive controls for INDIGO

As a negative control, we shuffled the interaction scores for the training data and found that it removed correlation and predictive ability of the model. Swapping drug names, which results in switching the chemogenomic profiles of individual drugs, also led to a significant decrease in correlation compared to a random model, confirming that chemogenomic profiles provide unique drug-specific information (Appendix Figs S11 and S12). To determine the role of interaction network topology in influencing the interaction predictions ("hub effect"), we also compared predictions with a simple model that outputs for a query drug, the average interaction score of the partner drug with other drugs in the network. This model had a Pearson’s correlation of only 0.3 based on cross-validation, ruling out the possibility that the predictions were primarily influenced by few synergistic and antagonistic hubs in the drug-interaction network. As a positive control, we found that INDIGO correctly identified interactions between the same drug at different doses to be linear (median score = 0 for self-interactions involving 73 drugs; Appendix Fig S16). This indicates that INDIGO can differentiate chemogenomic profiles of the same drug at different doses from other drugs. This is significant given that the model was trained only on interactions between different drugs.

#### Chemogenomic profile similarity-based approaches

For the correlation method, we measured the Pearson’s correlation between the chemogenomic profiles of two drugs of interest. Interactions were defined to be synergistic, antagonistic, or linear if the correlation passed a specific threshold. The best threshold was determined based on the training data using the random forest algorithm. Similarly, profile overlap was defined as the overlap in the number of sensitivity genes between the chemogenomic profiles of two drugs of interest. We used a hypergeometric test to determine whether the total overlap was significantly higher than that
expected by random chance. The best $P$-value threshold for differentiating different interaction types was determined based on the training data using the random forest algorithm.

**Orthology mapping to *S. aureus* and *M. tuberculosis***

Orthologous genes in *E. coli* were obtained from OrtholugeDB (Whiteside *et al.*, 2013). The database uses the reciprocal-best-BLAST hit (RBBH) procedure to generate the initial set of ortholog predictions. Genes are declared orthologs if they are each other's top BLAST hit when each genome is BLASTed against the other (Whiteside *et al.*, 2013). This resulted in 844 and 909 genes that were predicted to be orthologs of *S. aureus* and *M. tuberculosis* among the *E. coli* genes in our model. Enrichment for orthologs among the top predictors was done using the hypergeometric test in MATLAB.

**Predicting interactions that differ between *E. coli* and other organisms**

To predict interactions that differ between the two species, we identified genes in the *E. coli* model that were not orthologous between the two species. The sigma and delta scores for these genes were then set to be at the minimum value (0). The resulting interaction scores for all the drug-interaction pairs were recalculated using the modified scores. Theoretically, there are millions of possible solutions for changing the sigma score matrix to match *S. aureus* or *M. tuberculosis*. To simplify the problem, we simulate the extreme case and set the states of all non-orthologous genes to be zero. Choosing either the maximal (2) or minimal (0) sigma score value led to the same correlation with the predicted difference ($R = 0.52$). Changing the sigma scores would allow us to identify drug interactions most sensitive to the state of the non-orthologous genes. By sequentially removing interactions that have large predicted deviation scores, we found that we could predict *S. aureus* interactions using INDIGO with a much higher correlation ($R > 0.55$) than experimental *E. coli* data (Appendix Fig S14).

**Algorithm and data availability**

The following datasets are provided: (i) experimental drug-interaction scores for all 171 combinations in *E. coli* (Dataset EV1); (ii) interaction predictions for 2628 drug combinations of 73 drugs (highest dose in chemogenomics data was chosen) in all three species (*E. coli*, *S. aureus*, and *M. tuberculosis*) (Dataset EV2), and (iii) ranked list of all genes used by INDIGO for predicting drug-interaction outcomes (Dataset EV3). The MATLAB implementation of INDIGO along with associated experimental data is provided as Computer Code EV1 and is also freely available on Synapse, the online portal for sharing computational biology data and software: https://www.synapse.org/#!Synapse:syn3880435/wiki/.

**References**


