

# Phenotypic Landscape of a Bacterial Cell

Robert J. Nichols,<sup>1,2</sup> Saunak Sen,<sup>3</sup> Yoe Jin Choo,<sup>2</sup> Pedro Beltrao,<sup>4</sup> Matylda Zietek,<sup>2</sup> Rachna Chaba,<sup>2</sup> Sueyoung Lee,<sup>2</sup> Krystyna M. Kazmierczak,<sup>5</sup> Karis J. Lee,<sup>2,8</sup> Angela Wong,<sup>2,9</sup> Michael Shales,<sup>4</sup> Susan Lovett,<sup>6</sup> Malcolm E. Winkler,<sup>5</sup> Nevan J. Krogan,<sup>4</sup> Athanasios Typas,<sup>2,\*</sup> and Carol A. Gross<sup>2,7,\*</sup>

<sup>1</sup>Oral and Craniofacial Sciences Graduate Program, University of California, San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143, USA

<sup>2</sup>Department of Microbiology and Immunology, University of California, San Francisco, 600 16th Street, San Francisco, CA 94158, USA

<sup>3</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94107, USA

<sup>4</sup>Department of Cellular and Molecular Pharmacology and The California Institute for Quantitative Biomedical Research, University of California, San Francisco, San Francisco, CA 94158, USA

<sup>5</sup>Department of Biology, Indiana University, Bloomington, IN 47405, USA

<sup>6</sup>Department of Biology and Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454-9110, USA

<sup>7</sup>Department of Cell and Tissue Biology, University of California, San Francisco, San Francisco, CA 94158, USA

<sup>8</sup>Present address: UCLA School of Dentistry, 10833 Le Conte Avenue, Los Angeles, CA 90095-1762, USA

<sup>9</sup>Present address: Massachusetts General Hospital, 50 Staniford Street, Boston, MA 02114, USA

\*Correspondence: [athanasios.typas@ucsf.edu](mailto:athanasios.typas@ucsf.edu) (A.T.), [cgrossucsf@gmail.com](mailto:cgrossucsf@gmail.com) (C.A.G.)

DOI [10.1016/j.cell.2010.11.052](https://doi.org/10.1016/j.cell.2010.11.052)

## SUMMARY

The explosion of sequence information in bacteria makes developing high-throughput, cost-effective approaches to matching genes with phenotypes imperative. Using *E. coli* as proof of principle, we show that combining large-scale chemical genomics with quantitative fitness measurements provides a high-quality data set rich in discovery. Probing growth profiles of a mutant library in hundreds of conditions in parallel yielded > 10,000 phenotypes that allowed us to study gene essentiality, discover leads for gene function and drug action, and understand higher-order organization of the bacterial chromosome. We highlight new information derived from the study, including insights into a gene involved in multiple antibiotic resistance and the synergy between a broadly used combinatory antibiotic therapy, trimethoprim and sulfonamides. This data set, publicly available at <http://ecoliwiki.net/tools/chemgen/>, is a valuable resource for both the microbiological and bioinformatic communities, as it provides high-confidence associations between hundreds of annotated and uncharacterized genes as well as inferences about the mode of action of several poorly understood drugs.

## INTRODUCTION

Before the physical basis of genes was understood, associating phenotypes with a heritable unit laid the foundation of modern genetics. Following discovery of the genetic code, linking a

phenotype to the responsible gene remained the most expeditious way to unravel gene function. With the explosion of sequence information, the balance has shifted. We now have many genes of unknown function. To capitalize on the burgeoning sequence bank, it is imperative to develop high-throughput technologies that link genes to phenotypes and expedite discovery of gene function. This is particularly true for prokaryotes, which represent a major fraction of the sequenced genomes and are in the forefront of metagenomic efforts (Qin et al., 2010).

Chemical and environmental perturbations have traditionally linked phenotypes to genotypes through forward genetic screens, but reverse genetic approaches are being increasingly utilized (Barker et al., 2010). Phenotype microarrays utilize a high-resolution readout of cellular respiration to evaluate fitness of a strain in hundreds of conditions (Bochner, 2009). This approach is appropriate for studying a few strains but is difficult to expand to genome-scale screens. In pooled growth competitions, thousands of strains are assayed in a single culture environment. Fitness values are derived from measuring strain abundance in a test relative to control condition (Giaever et al., 2004; Girgis et al., 2009; Hillenmeyer et al., 2008; Hoon et al., 2008; Lee et al., 2005; Pan et al., 2004; Parsons et al., 2006; Warner et al., 2010; Xu et al., 2007). These approaches are very efficient, but competition between strains in each condition makes it difficult to determine relative strain growth across conditions (Girgis et al., 2009), especially for strains that grow slowly even in the absence of perturbation (Lee et al., 2005). Arraying mutant strains on solid media allows independent evaluation of strain fitness but has been used only for low-resolution measurements of entire libraries (Liu et al., 2010; Tamae et al., 2008) or for essential genes (Pathania et al., 2009). High-throughput genetic interaction studies, pioneered in yeast (Schuldiner et al., 2005; Tong et al., 2001), are complementary to chemical genomics approaches. Such analyses quantitatively

measure colony growth of double mutants in high-density format on agar surfaces and have led to numerous successes in identifying gene function and network organization (Beltrao et al., 2010). Similar methodology has been developed for *E. coli* (Butland et al., 2008; Typas et al., 2008).

We use *E. coli* to illustrate the power of applying the high-resolution quantitative fitness measurements of genetic interaction analysis to high-throughput phenotypic analysis of culturable microbes. “Phenomic profiling” provides a quantitative description of the response of all single-gene deletions to physiologically relevant stresses and drug challenges. By profiling ~4000 genes in > 300 perturbations, we identified thousands of phenotypes and a diverse suite of conditionally essential genes. This approach provides new insights into the chromosome organization, functional landscape, and evolutionary trajectory of *E. coli*. It facilitates high-confidence association of genes of unknown function to those of known function, as highlighted by discovery of the role of a gene involved in multiple antibiotic resistance in this manuscript and identification of two lipoproteins essential for peptidoglycan synthesis (Typas et al., 2010). Finally, the degree to which various gene deletions alter toxic drug effects has led to powerful insights regarding drug mode of action (Kohanski et al., 2008), and we demonstrate that our analysis generates numerous leads concerning drug function.

## RESULTS AND DISCUSSION

### Phenomic Profiling of *E. coli* K12 Yields a Robust, High-Quality Data Set

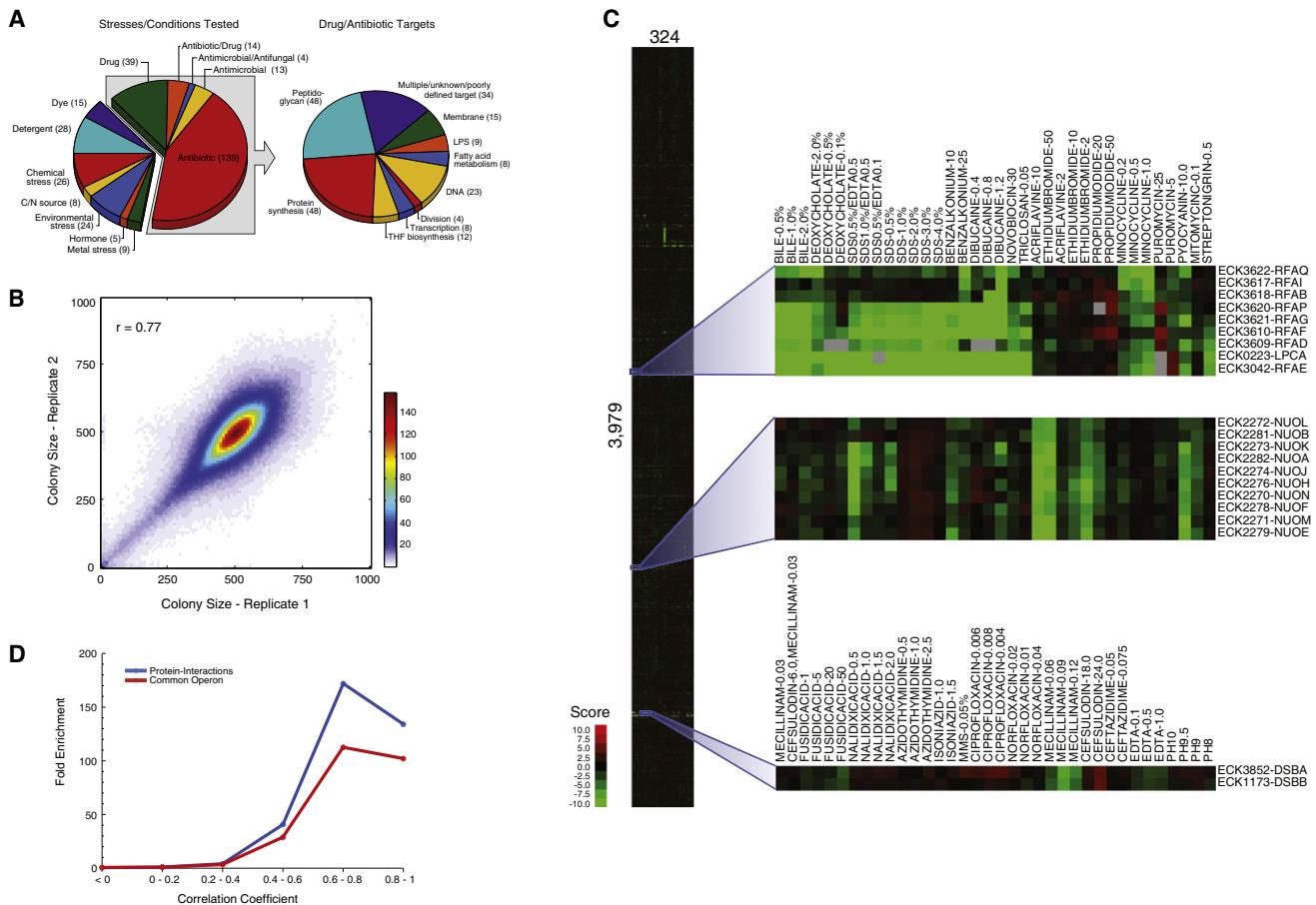
We determined quantitative growth scores for the Keio single-gene deletion library (Baba et al., 2006); essential gene hypomorphs (C-terminally tandem-affinity tagged [Butland et al., 2005] or specific alleles); and a small RNA/small protein knockout library (Hobbs et al., 2010) in conditions representing the range of stresses that *E. coli* encounters. Mutant strains arrayed in high density on agar plates (1536 colonies/plate) were grown in 324 conditions covering 114 unique stresses (Figure 1A and Figure S1A and Table S1 available online). Colony sizes were analyzed and converted to a drug-gene score using an approach developed for quantifying genetic interactions (see Extended Experimental Procedures). More than half of the conditions were antibiotic/antimicrobial treatments (Figure 1A). By using a subinhibitory concentration series that maximally inhibited growth of the wild-type (WT) strain  $\leq 50\%$ , we were able to search for specific drug-gene interactions (Figure S1A) and reduce the ability of spontaneous suppressor mutations to overtake the colony. Two independently derived clones of each mutant strain were analyzed (for sRNA mutants, a single isolate was arrayed twice), and screens were performed at least twice, enabling scores to be based on four to six independent measurements. Correlation between replicate colony size measurements was very high ( $r = 0.77$ ; Figure 1B). The final data set (Table S2) was comprised of scores for the 3979 mutant strains passing quality control (e.g., proper normalized colony size distribution and replicate reproducibility; see Extended Experimental Procedures). The entire data set is available in an interactive, searchable format and as a flat file on the *E. coli* wiki website (at <http://ecoliwiki.net/tools/chemgen/>).

The entire matrix (3979 mutants  $\times$  324 conditions) was subjected to two-dimensional (2D) hierarchical clustering (Figure 1C). Drugs with similar effects cluster on the x axis; mutants that behaved similarly cluster on the y axis. Notably, concentrations of the same drug, drugs of the same family, and/or similar conditions clustered together, as did mutants of genes known to be part of the same operon, biological pathway, and/or protein complex. Zoomed insets of our clustergram illustrate examples. Genes in the *rfa* operon (*rfaG*, *rfaP*, *rfaQ*, *rfaB* and *rfaI*), which encodes enzymes that synthesize the inner- and outer-lipopoly-saccharide (LPS) core, strongly cluster together with three/four genes that are responsible for the synthesis of one of the sugar building blocks, ADP-L glycero- $\beta$ -D-manno heptose (*rfaD*, *rfaE*, and *lpcA*). Importantly, clustering reflects their shared sensitivities to a concentration series of compounds known to perturb the envelope integrity of the cell, consistent with the role of LPS. *dsbA* and *dsbB*, encoded in different operons, also cluster. The DsbA/DsbB complex generates disulfide bridges in the periplasm.

The response of each mutant strain across all conditions is denoted as its “phenotypic signature.” High correlation between two phenotypic signatures is highly predictive of known indicators of functional connection between genes. Gene pairs with correlation coefficients ( $r$ ) between 0.6 and 0.8 ( $p < 10^{-34}$ ) are more than 100-fold enriched for genes sharing common operon membership and 150-fold enriched for genes with known protein interactions determined from low-throughput experiments (<http://www.ecocyc.org>; Figure 1D). This benchmarking analysis indicates that our phenomics data set is biologically meaningful. Correlated phenotypic signatures also reproduce connections between curated biological pathways (Figure S1B). For example, electron transfer components cluster tightly (e.g., *nuo* genes encoding NADH dehydrogenase I complex; Figure 1C). Their clustering reflects high sensitivity to membrane-perturbing stresses, including detergents, dyes, and metals, and increased resistance to aminoglycosides, in agreement with early studies that illustrate decreased aminoglycoside uptake in the absence of a fully functional electron transport chain (Girgis et al., 2009; Taber et al., 1987). All three examples described in Figure 1C are consistent with the expectation that highly correlated phenotypic signatures are biologically meaningful ( $r \geq 0.6–0.8$ ).

### Phenomic Profiling Defines Responsive and Conditionally Essential Genes

A central goal of this study was to systematically evaluate the impact of every gene deletion on *E. coli* fitness in diverse environments, as few gene deletions in *E. coli* have robust reported growth phenotypes, and only 8% of the genes are essential in rich media (Baba et al., 2006; Yamamoto et al., 2009). We used a statistical method to define a reliable phenotype. In brief, we standardized the interquartile range of the distribution of scores for each screen and then determined the probability that each condition-gene interaction represented a true phenotype using a normal cumulative distribution function (see Extended Experimental Procedures). Using a 5% probability that these phenotypes arose by chance as a cutoff (false discovery rate [FDR]  $\leq 5\%$ ), 49% of all strains tested (1957/3979 strains; Figure 2A) had one or more phenotypes. We refer



**Figure 1. Phenomic Profiling of the Enhanced Keio Collection Yields a Robust and Rich Data Set**

(A) Classification of the 324 stresses screened (left) and cellular targets of the antibiotic/antimicrobial/drug classes (right).

(B) Heat map representation of scatter plot comparing normalized colony sizes in pixels of plate replicates 1 and 2 across the entire data set. Bins indicate the square root of the number of replicate pairs within a  $10 \times 10$  pixel window as depicted by color scale. Note that the vast majority of the replicates have highly correlated colony sizes.

(C) Clustergram of fitness scores for 3979 mutant strains in response to all 324 conditions. Zoomed insets demonstrate coclustering of conditions (x axis) and genes (y axis) for a common pathway (*rfa* cluster) and protein complexes encoded in the same operon (*nuo*) or in different operons (*dsbA* and *dsbB*). Gray boxes indicate missing data.

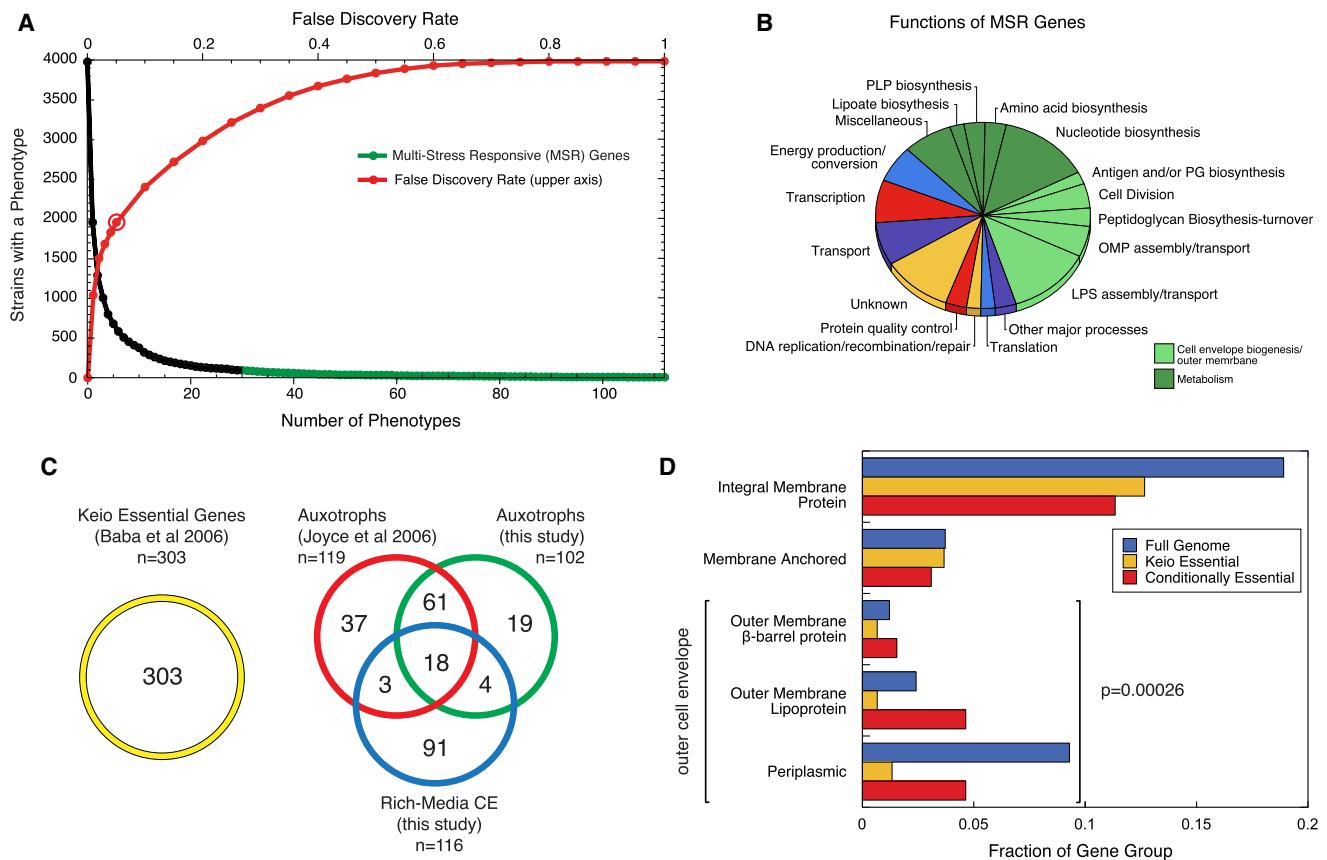
(D) High correlation between a pair of phenotypic signatures is predictive of shared protein interaction and/or operon membership.

See also Figure S1, Table S1, and Table S2.

to these genes as the “responsive genome.” This responsive genome is a work in progress, as it is limited to genes whose removal causes growth phenotypes in response to the stresses tested. Expanding the stresses tested and/or the readout (e.g., motility) will certainly increase this number (Girgis et al., 2007). A cumulative plot of the number of individual phenotypes per strain shows that very few genes have many phenotypes. Multi-stress responsive (MSR) strains ( $\geq 30$  phenotypes; Table S3) participate in many cellular processes, suggesting that our stresses encompassed diverse cellular challenges (Figure 2B). With a stringent cutoff of 5% FDR, the maximum number of phenotypes from a single screen was 173 (~4% strains; Figure S2A), and the total number of phenotypes (13497) represents ~1% of all condition-gene pairs tested. Overall, 80% of the phenotypes were negative (gene deletion more sensitive) and 20% positive (gene deletion more resistant), consistent with

recent genetic interaction analyses in *S. cerevisiae* (Fiedler et al., 2009) and *S. pombe* (Roguev et al., 2008). This suggests that removal of a gene product is more likely to decrease than enhance resistance to stress (Figure S2B). In summary, our analysis captured numerous highly specific condition-gene responses. Clearly, this data set can be used to assign more phenotypes at a lower confidence level. Indeed, a recent chemical genomics data set in *S. cerevisiae* reported phenotypes for more than 95% of gene deletions tested, many stemming from a handful of severe stresses (Hillenmeyer et al., 2008).

Conditionally essential (CE) genes are essential for growth in a particular condition. Deletions of such genes show very small colony sizes and high-confidence negative scores in particular conditions (see Extended Experimental Procedures). We identified 197 CE genes, comprised of auxotrophs, which exhibit no growth in minimal media, and rich-media CE genes, which



**Figure 2. Identification of Responsive and Conditionally Essential Genes**

(A) Using a 5% false discovery rate (FDR), 49% of strains tested had at least one phenotype (open circle on the red line). As the FDR is relaxed, more phenotypes are identified (red line). At 5% FDR, some strains have several phenotypes (black), and very few (2.3%; 94 strains) have 30 or more phenotypes (green, multi-stress responsive [MSR] genes).

(B) MSR genes participate in a wide variety of cellular processes, particularly those related to metabolism and the cell envelope. Genes were manually curated to COG-based functions; each gene was allowed to belong only to a single function.

(C) 196 genes are conditionally essential (CE) in this study. Of these, roughly half have been previously described as CE due to auxotrophy. Note that some auxotrophic genes also display a no growth phenotype in at least one rich medium condition and are classified jointly as auxotroph and rich media CE.

(D) CE gene products are enriched in the outer cell envelope (periplasm and outer membrane) relative to Keio essential genes ( $p = 0.00026$ ), highlighting the importance of this compartment in tolerating stress. The cytoplasmic gene category is not displayed here but is not enriched for CE gene products.

See also Figure S2, Table S3, Table S4, and Table S5.

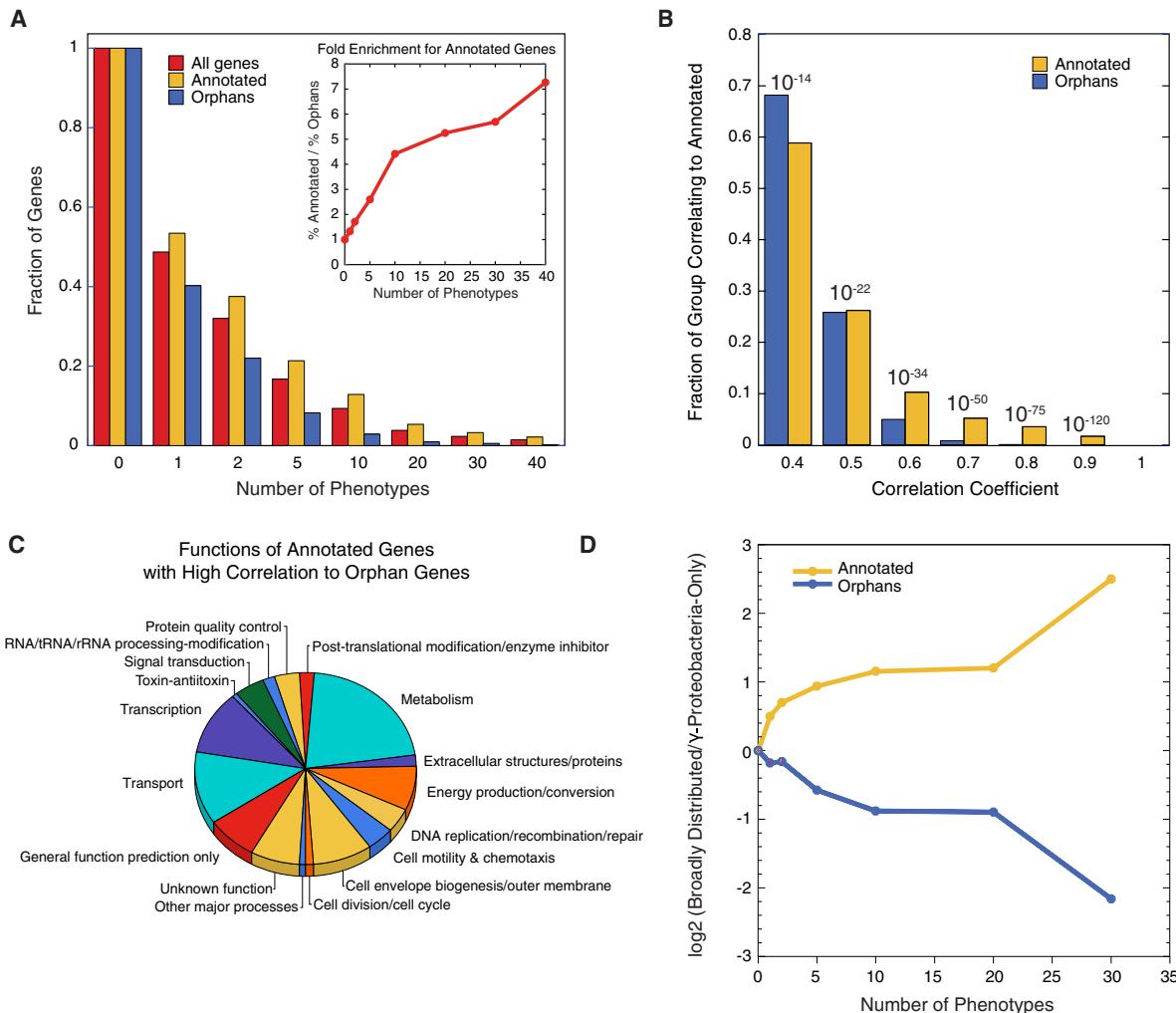
exhibit no growth in at least one rich media-based stress (Table S4). Importantly, our data set had 70% overlap with a previous study of Keio Collection auxotrophs (Figure 2C; Joyce et al., 2006) despite significant experimental differences (e.g., growth in liquid versus solid media). Many of the remaining 30% were extremely sick but above the stringent threshold that we used to define no growth. We also identified 23 additional auxotrophs specific to alternative carbon/nitrogen sources that were not tested in the Joyce et al. study.

Genes that are essential for survival in natural environments are likely to extend beyond those required for laboratory growth and could be targets for new antimicrobials (D'Elia et al., 2009). The 116 rich-media CE genes that we identified (Figure 2C) result from physiologically relevant stresses and increase the current number of essential genes by roughly 30%. Interestingly, many of these gene products are located in the outer cell envelope

(Figure 2D), a selective permeability barrier for Gram-negative bacteria that is severely underrepresented for known essential genes (Figure 2D). Many of the stresses generating CE phenotypes are part of the natural environment of *E. coli*, e.g., bile salts (Table S5), indicating that these genes are likely indispensable for *E. coli* to survive in vivo. Similarly, using the largest metagenomic data set to date, Qin et al. reported that envelope-specific functions, such as adhesion, were commonly required for life in the gut (Qin et al., 2010).

#### Phenomic Profiling Helps Assign Function to Uncharacterized Genes

A key motivation for our study was to provide phenotypes for mutants of genes without functional annotation. Using a recently assembled compendium of such “orphan genes” in *E. coli* (Hu et al., 2009), we find that the fraction of mutant orphan genes

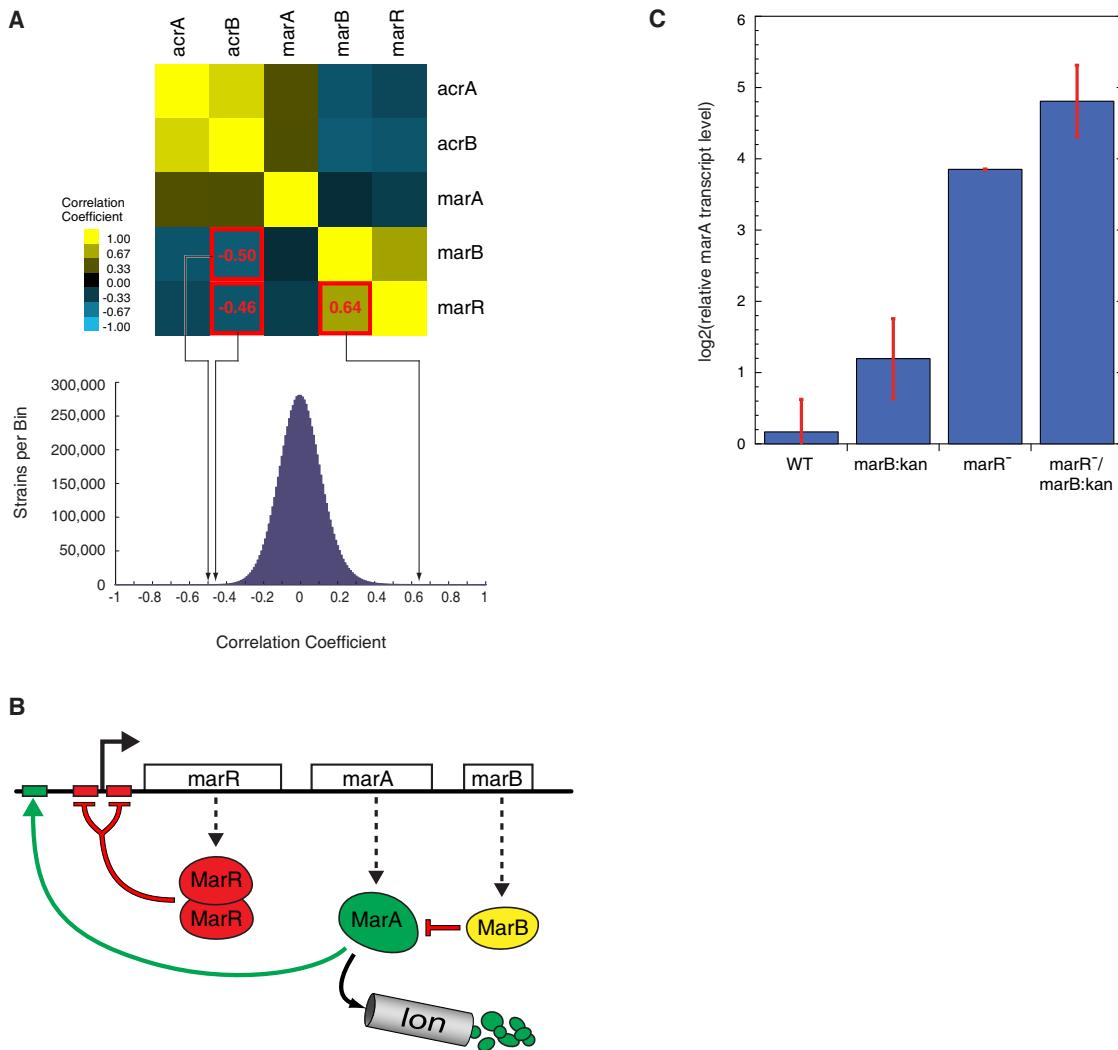


**Figure 3. Phenomic Profiling Identifies Phenotypes for Orphan Gene Mutants**

- (A) Cumulative distribution of phenotypes indicating the fraction of gene mutants in each class having at least the number of phenotypes shown on the x axis. The plot reveals that orphan gene mutants have phenotypes but tend to have fewer phenotypes than annotated gene mutants. The inset quantifies phenotype deficit of orphan mutants.
- (B) Cumulative distribution of highly correlated pairs identifies many orphan genes that correlate highly to an annotated gene, providing high-confidence clues to the function of the orphan gene. Values shown above each pair of bars are the p values associated with pairwise correlation of any two strains at the indicated correlation coefficients.
- (C) High-confidence correlations between orphans and annotated genes ( $r \geq 0.5$ ) provide leads that are related to many different cellular functions. Procedure for functional assignment is described in Figure 2. Note that several “annotated genes” were classified as genes of “unknown function” or “general function prediction only” after manual curation.
- (D) Annotated genes that are responsible for many phenotypes tend to be broadly conserved, whereas the most responsive orphan genes tend to be restricted to  $\gamma$ -proteobacteria. For a list of the orphans-annotated pairs ( $r > 0.5$ ), see Table S6.

with phenotypes is close to that of annotated genes (Figure 3A), but the former tend to have fewer phenotypes, indicating the power of phenomic analysis for identifying their phenotypes. Importantly, the phenotypic profiles of  $> 25\%$  of all orphan genes correlate strongly with those of annotated genes ( $r \geq 0.5$ ; Figure 3B and Table S6), providing high-confidence leads ( $p < 10^{-22}$ ) for discovery of their function. As these orphans are tied to a wide variety of cellular processes (Figure 3C), the data set will be of broad utility.

A small fraction of orphan gene knockouts have many phenotypes. Whereas annotated genes that are responsible for many phenotypes are broadly distributed among bacteria, the most responsive orphans tend to be narrowly distributed (Figure 3D). This result suggests that evolutionary conservation is not a reliable indicator of the importance of an orphan gene to the organism and that annotating them solely by homology has limitations. Such orphans may have evolved to fulfill an important but specialized function required by the niche of the organism.



**Figure 4. A Function for *marB***

(A) *marR*<sup>-</sup> and *marB*<sup>-</sup> phenotypic signatures are highly correlated with each other and are highly anticorrelated with that of *acrB*<sup>-</sup> (top). The bottom graph positions these correlations in a histogram showing all pair-wise correlation coefficients between the 3979 mutants.

(B) Schematic of the *E. coli* multiple antibiotic resistance (*mar*) operon. *marB* is a gene of unknown function, but our results suggest that it encodes a protein that inhibits MarA.

(C) RT-PCR analysis shows that *marA* transcription is derepressed in *marB*<sup>-</sup> cells. Derepression is independent of and additive with that of *marR*<sup>-</sup>.

In support of this idea, a multiresponsive orphan identified in this study (*lpoB*) is restricted to enterobacteria and regulates peptidoglycan synthesis, a conserved process that is ubiquitous among bacteria (Typas et al., 2010).

#### Using Phenotypic Signatures to Identify Gene Function

Both correlated phenotypic signatures (Figures 1C and 1D; Typas et al., 2010) and anticorrelated phenotypic signatures have functional significance. For example, the phenotypic signatures of deletions of a transcriptional repressor and important target genes are likely to be anticorrelated. We find that *marR*<sup>-</sup> and *marB*<sup>-</sup> were highly anticorrelated with *acrB*<sup>-</sup>, whereas *marR*<sup>-</sup> and *marB*<sup>-</sup> were highly correlated (Figure 4A). *marB* is a gene of unknown function in the multiple antibiotic resistance

operon (*marRAB*), which also includes the operon repressor *marR* and its activator, *marA*. MarA also activates genes involved in antibiotic resistance, most importantly *acrAB*, encoding the major antibiotic efflux pump in *E. coli* (Figure 4B) (Martin and Rosner, 2002). We explored whether MarB, like MarR, repressed MarA. Because of the inherent problems of high-throughput collections (suppressors, gene duplications, and cross-contamination), we always apply stringent quality control procedures to any follow-up investigations, including PCR validation of Keio isolates and verification that retransduced strains maintain their phenotype. As *mar* is a hot spot for adaptive mutations, we also sequenced the entire operon and promoter region of single deletions and the double mutants that we constructed.

Deletion of either *marB* or *marR* resulted in higher MarA levels, and the double-*marRB* mutant showed additive effects on MarA transcript level (Figure 4C) and protein level (data not shown). These effects were observed in both Kan-marked and clean deletions (Kan cassette excised, leaving an 82 nt scar). The  $\Delta marR$  strain exhibits  $\sim 2\times$  more increase in MarA transcript levels than *marR::kan* (data not shown), arguing for a small polar effect of the cassette. Both *marB::kan* and  $\Delta marB$  exhibit the same 2-fold increase in MarA levels (data not shown). These data suggest that MarB represses MarA independently of MarR. MarB does not have the signature of a DNA-binding protein, suggesting that it acts posttranscriptionally. MarA level is controlled by the Lon protease (Griffith et al., 2004), but *lon*<sup>-</sup> and *marB*<sup>-</sup> effects are additive, indicating that MarB does not function through Lon (data not shown). MarA has been proposed to scan for activation sites while bound to RNA polymerase; by direct binding to either partner, MarB could disrupt complex formation. Alternatively, MarB may function in the periplasm. As MarB has a predicted periplasmic signal sequence, it could titrate an activating ligand for *mar* (e.g., salicylate).

Although *mar* is highly studied (~200 primary publications; PubMed), our screen provided the first lead for MarB function. MarA targets ~40 genes, many of which are also coregulated by the SoxS and Rob activators, with similar DNA-binding motifs as MarA (Martin et al., 2008; Martin and Rosner, 2002). The rules of engagement are poorly understood, but each activator responds to different environmental cues, and overexpression of each leads to distinct phenotypes (Warner and Levy, 2010). It is likely that tight control of each activator has an impact on the final gene expression output, which is crucial for cellular proliferation. MarB may be an important player in fine-tuning the expression of MarA, especially because it is a conserved member of the *mar* operon, which has only recently emerged in selected enterobacteria. Strong evidence for the importance of *mar* operon regulation in these organisms is that *mar* is a hot spot for mutations conferring higher drug resistance in *E. coli* (Nicoloff et al., 2006, 2007).

### Phenomic Profiling Reveals Metabolic Network Behaviors under Antifolate Drug Stress

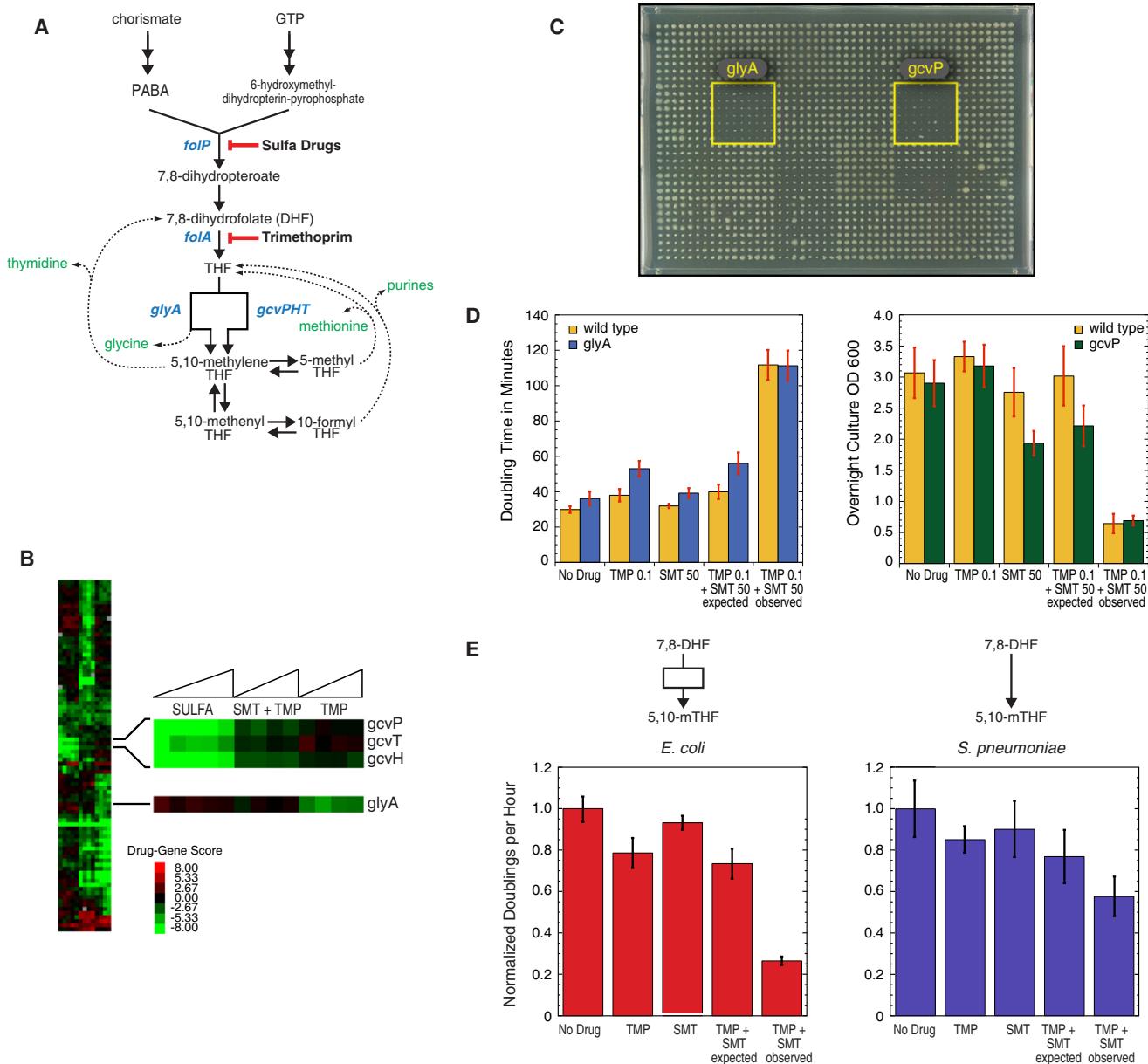
Tetrahydrofolate (THF) and its methyl/formylated derivatives are key molecules in all kingdoms of life for one-carbon metabolism. THF is used to synthesize glycine, methionine, purines, and dTTP in a process that leads to recycling of the THF species back to THF or dihydrofolate (DHF) (Figure 5A). The bacterial THF biosynthesis pathway is targeted by two drugs: sulfonamides (Sulfa) target FolP, and trimethoprim (TMP) targets FolA (Figure 5A). Dual inhibition by Sulfa and TMP is strongly synergistic, and therefore these drugs are almost exclusively administered in combination for treatment of ear, urinary tract, and bronchial infections. Despite extensive clinical use and years of laboratory investigation, we lack a complete mechanistic understanding of why these drugs are strongly synergistic. A network feature identified by phenomic profiling could contribute to synergy.

We find that the two drug classes have major phenotypic differences. Sulfa and TMP treatments are highly correlated within their class ( $r = 0.57$  for Sulfa,  $0.67$  for TMP) but poorly

correlated with each other ( $r = 0.15 \pm 0.04$ ), just slightly more than the correlation observed between all screens ( $r = 0.025 \pm 0.12$ ). Thus, subinhibitory TMP and Sulfa treatments have fundamentally different effects on the cell even though both partially block THF biosynthesis. Importantly, removing enzymes acting directly downstream of THF production resulted in opposite drug sensitivities: the serine hydroxymethyltransferase mutant (*glyA::kan*) was sensitive only to TMP; conversely, glycine cleavage (GCV) mutants (*gcvP::kan*, *gcvH::kan*, and *gcvT::kan*) were sensitive only to Sulfa (Figure 5B). The mutant results were reproduced in liquid culture (Figure 5D), in which *glyA*<sup>-</sup> TMP sensitivity is manifested as a growth rate phenotype (left), and *gcvP*<sup>-</sup> sulfamethizole (SMT) sensitivity is registered as a low stationary phase density (right).

GlyA and GCV lie on opposite sides of a branched pathway that converts THF to 5,10-methylene THF (5,10-mTHF; Figure 5A). As *glyA* and *gcv* mutants exhibit synthetic lethality, they are the only routes to production of this essential metabolite (Figure 5C). A simple explanation for the differential responses of *glyA*<sup>-</sup> and *gcvP*<sup>-</sup> is that 5,10-mTHF is predominantly produced via different branches under each drug treatment. A corollary is that combination drug treatment inhibits both branches, resulting in synergistic limitation for 5,10-mTHF, before the pools of THF are depleted. In support of this idea, despite the increased sensitivity of *glyA*<sup>-</sup> and *gcvP*<sup>-</sup> to single drugs, these strains grew no more poorly than WT under the drug combination (Figure 5D). Thus, genetically eliminating either branch of the pathway reduced but did not eliminate synergy. The downstream biosynthetic reactions are also differentially affected by TMP and Sulfa (Figure S3A), and we are currently testing whether they partially account for the residual synergy. *Streptococcus pneumoniae* lacks the GCV system and exhibits significantly less drug synergy than *E. coli* across different growth conditions (Figure 5E and data not shown). We performed our comparison using concentrations of TMP and SMT that caused the same relative growth defect in each species (Figure 5E). These data together support the hypothesis that simultaneous inhibition of the branched pathway for production of 5,10-mTHF contributes to the observed antifolate synergy in *E. coli*.

Our data do not indicate whether differential effects of TMP and Sulfa on GlyA and GCV result from differential inhibition of expression or activity or the intrinsic properties of each enzyme. We favor the idea that differential metabolite accumulation and subsequent feed-forward enzymatic regulation make a contribution to the distinct cellular responses to these two drugs. Recent metabolomic flux analyses indicate that high doses of TMP lead to accumulation and depletion of select metabolites, as well as to protein-level regulation of portions of the network (Kwon et al., 2008, 2010). Although a comparable analysis has not been performed for Sulfa drugs, deletion of the predicted 5-formyl-THF cycloligase, *ygfA*, which likely degrades 5-formyl-THF (Jeanguenin et al., 2010), clusters tightly with the *gcv* mutants and exhibits sensitivity only to Sulfa drugs (Figure S3A). That 5-formyl-THF degradation is critical only under Sulfa stress suggests differential accumulation (or requirement) of THF species under Sulfa and TMP treatments. 5-formyl-THF is a known inhibitor of several enzymes in the THF network of other organisms (Field et al., 2006; Stover and Schirch, 1993)

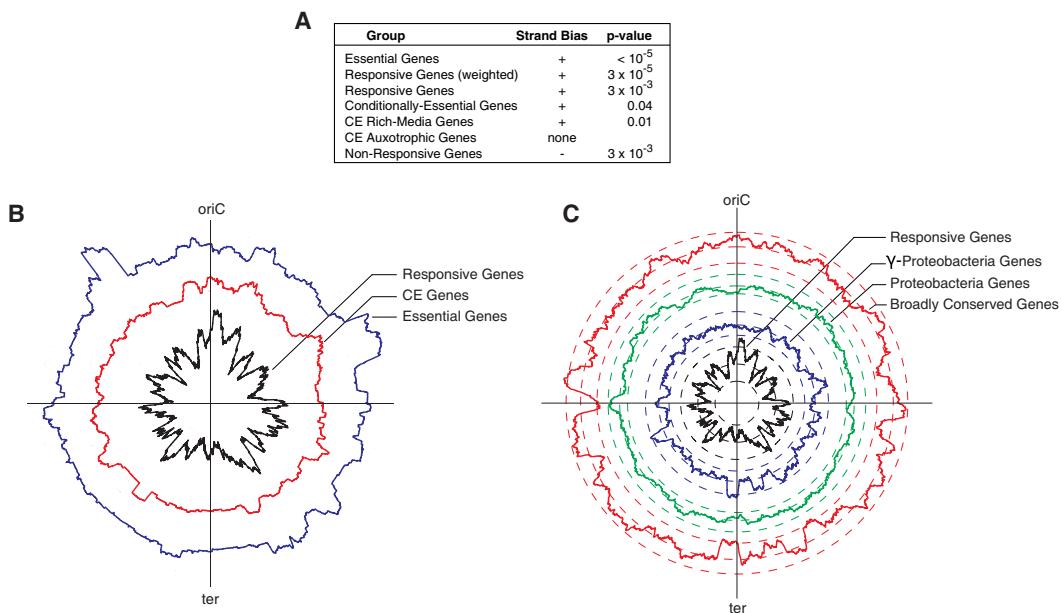


**Figure 5. A New Network Feature Contributing to Antifolate Drug Synergy**

- (A) Schematic of the *E. coli* tetrahydrofolate (THF) biosynthesis pathway and the enzymatic steps inhibited by Sulfa and TMP.
- (B) Clustergram of genes that respond to Sulfa, TMP, or the combination. Zoomed image indicates that *gcv* mutants are sensitive to Sulfa, that *glyA*<sup>-</sup> is sensitive to TMP, and that these four mutants exhibit essentially wild-type growth in response to the drug combination.
- (C) *glyA*<sup>-</sup> and *gcvP*<sup>-</sup> are a synthetic lethal pair. Image of a plate mating between the donor Hfr *gcvP*:cat and 24 *kan*<sup>R</sup> recipients (arrayed in boxes of 8×8 colonies) grown on kanamycin/chloramphenicol medium to select for double-mutant strains; position of the *glyA*:*kan* and *gcvP*:*kan* recipients is highlighted.
- (D) Liquid culture experiments verify growth phenotypes on agar plates shown in Figure 5B. The deviation of the observed from the expected value for the TMP/SMT combination denotes the degree of synergy of the two drugs, which is lower for *glyA*<sup>-</sup> and *gcvP*<sup>-</sup> cells compared to wild-type cells. Concentrations shown for TMP and SMT are in µg/ml.
- (E) Quantification of synergy in *E. coli* and *S. pneumoniae*, which lacks the branched pathway for generating 5,10-mTHF present in *E. coli*. Comparisons were performed using single drug concentrations giving equivalent inhibition of both organisms. *S. pneumoniae* has reduced synergy compared to *E. coli*. See also Figure S3.

and could act as an effective protein-level regulator. Similarly, a strain lacking a predicted alanine racemase, *yggS*, is sensitive only to Sulfa; D-alanine is known to inactivate GlyA (Schirch

et al., 1985), and *yggS*<sup>-</sup> and *glyA*<sup>-</sup> form a synthetic lethal pair (Figure S3B). Thus, the different cellular responses to these two drugs may be due, in part, to metabolite-based enzymatic



**Figure 6. Phenomic Profiling Generates Insights into Genome Organization**

(A) Essential and responsive genes are biased to the plus strand of DNA (transcription direction coincident with replication), and the nonresponsive genes are biased to the minus strand of DNA.

(B and C) For each panel, circular plots depict gene position, adjusting coordinates so that the chromosome starts at the origin of replication (*oriC* = 0 bp); the terminus region (*ter*) is opposite the *oriC*. Each trace represents spatial enrichment for the variable plotted based on a 100 kb sliding window. Three dashed lines of the same color accompany each trace, indicating the minimum permutation threshold, the baseline representing zero enrichment, and the maximum permutation threshold (inside to outside). Permutation thresholds are the result of 1000 randomizations of gene class assignments (see [Extended Experimental Procedures](#)) and indicate significant negative and positive spatial enrichment at a p value of 0.05.

(B) Responsive and CE genes are concentrated around the *oriC* and scarce around the terminus.

(C) The terminus is positively enriched for genes restricted to the  $\gamma$ -proteobacteria and negatively enriched for broadly conserved genes.

See also [Figure S4](#).

regulation. An extension is that the synergy of combination therapy could rest primarily on complementary inhibition of different one-carbon biosynthesis reactions and therefore recycling of THFs. This model would allow for synergy even with the expected additive limitation of THF production.

In summary, our results illustrate the power of phenomic profiling to yield insights into drug action and the ability of a networks view to provide new paradigms for analysis of drug interaction mechanisms, which can facilitate hypothesis-driven research on drug interactions (Bollenbach et al., 2009). This type of analysis may be generally useful in predicting drug synergies and in explaining variable drug-drug interactions across species.

### Phenomic Profiling Gives Insights into Genomic Organization

The *E. coli* genome is encoded on a single, circular chromosome, with a single origin of replication, *oriC*. Essential genes are biased to the plus (+) strand, where transcription proceeds in the same direction as DNA replication. This may avert head-on collisions between RNA and DNA polymerases that would result in aborted transcripts or truncated or frame-shifted proteins (Rocha and Danchin, 2003). Here, we show that responsive and CE genes, which are important for optimal growth of the organism, also show + strand bias (Figure 6A). Indeed, the weighted

responsive genome (responsive genes weighted by number of phenotypes identified) is heavily biased to the + strand, indicating great selective pressure to place genes that are important for rapid growth on the + strand. Conversely, the nonresponsive genome is biased to the minus (-) strand. As our approaches expand to incorporate additional phenotypic readouts that are more important for cells with reduced division and DNA replication (e.g., biofilm formation), the + strand bias of responsive genes will presumably be reduced.

The chromosome is massively compacted in the cell to create the nucleoid, which is thought to contribute significantly to the organization of gene expression (Travers and Muskhelishvili, 2005; Vora et al., 2009). Chromosomal loci have spatial addresses in the cell, corresponding closely to their chromosomal position (Toro and Shapiro, 2010). In addition, highly expressed genes that are associated with transcription and translation are located near the origin of replication (*oriC*), presumably to benefit from the “gene dosage” effect created when rapidly growing cells initiate multiple rounds of DNA replication per division (Couturier and Rocha, 2006). Projecting the spatial distribution of the responsive genes onto the circular chromosome (Figure 6B, black trace) provides us with a snapshot of the *E. coli* genome from a functional perspective. This projection is based on a 100 kb sliding window and therefore captures organization above the operon level (Figure S4A and

**Extended Experimental Procedures.** A pattern of alternating peaks and valleys is clearly evident, indicating that responsive genes cluster spatially into large chromosomal regions that are separated by regions generally devoid of responsive genes. “Valleys” are comprised of spatially separated operons that are often transcribed from different strands, indicating that low responsiveness is a regional characteristic rather than an artifact due to large nonresponsive operons. Our finding of clustering above the operon level is in accord with other studies showing that gene expression is broadly correlated across certain regions of the chromosome (Carpentier et al., 2005; Jeong et al., 2004).

The responsive genome is most enriched around *oriC*, which has the highest concentration of responsive genes (Figure 6B, black trace). This area is also enriched for the most responsive genes (Figure S4B) and for CE genes (Figure 6B, red trace), providing strong support for the idea that the *E. coli* chromosome tends to store genes of high functional importance near the *oriC*. In contrast, the terminus region is relatively devoid of responsive genes (Figure 6B, black trace) and has a paucity of broadly conserved genes (Figure 6C, red trace) and a corresponding enrichment for genes restricted to  $\gamma$ -proteobacteria (Figure 6C, blue trace). We postulate that the terminus region contains newly acquired genes that have yet to fully integrate into the cellular network and tend to lack phenotypes. This could enable cells to avoid unnecessarily high expression of such genes as a consequence of the gene dosage effect. Should this result prove true across bacterial species, it could point to a general organizing principle of circular chromosomes.

### Phenomic Profiling Describes Drug Action

“Drug-centric” analyses are more complex than “gene-centric” analyses. Whereas genes mostly participate in a single biological process, many parameters are required to describe drug action: uptake, primary/secondary targets, and efflux. Therefore, pairwise relationships between drugs are more complex than those between genes. For example, two drugs may cluster based on drug uptake even though their primary targets differ. In addition, drug signatures are an order of magnitude larger than gene signatures (3979 versus 324). To reduce the complexity of drug signatures, we calculated drug-gene ontology (GO) scores, which represent the probability that a given GO group specifically interacts with a given drug (i.e., number of phenotypes associated with genes in the GO group versus across the entire data set). We used these drug-GO scores to explore drug mode of action through a network-based clustering strategy (see **Extended Experimental Procedures**). The position of drugs in the network (Figure 7A) is based both on the magnitude of their drug-gene ontology (GO) scores (gray) and on drug-drug correlations (yellow). Of the 719 significant drug-GO interactions ( $p \leq 10^{-3}$ ), which include 64 drugs and 218 GO groups, 657 were negative, and only 62 were positive. Thus, disrupting a linked biological process was very likely to increase drug sensitivity (Table S7). Drug-drug correlations increased the resolution of the network and captured drug similarities that escaped the drug-GO analysis.

We found that drugs with the same cellular target tend to cluster. For example, drugs targeting DNA (orange) fall in the lower right, those targeting THF-biosynthesis (light green) fall

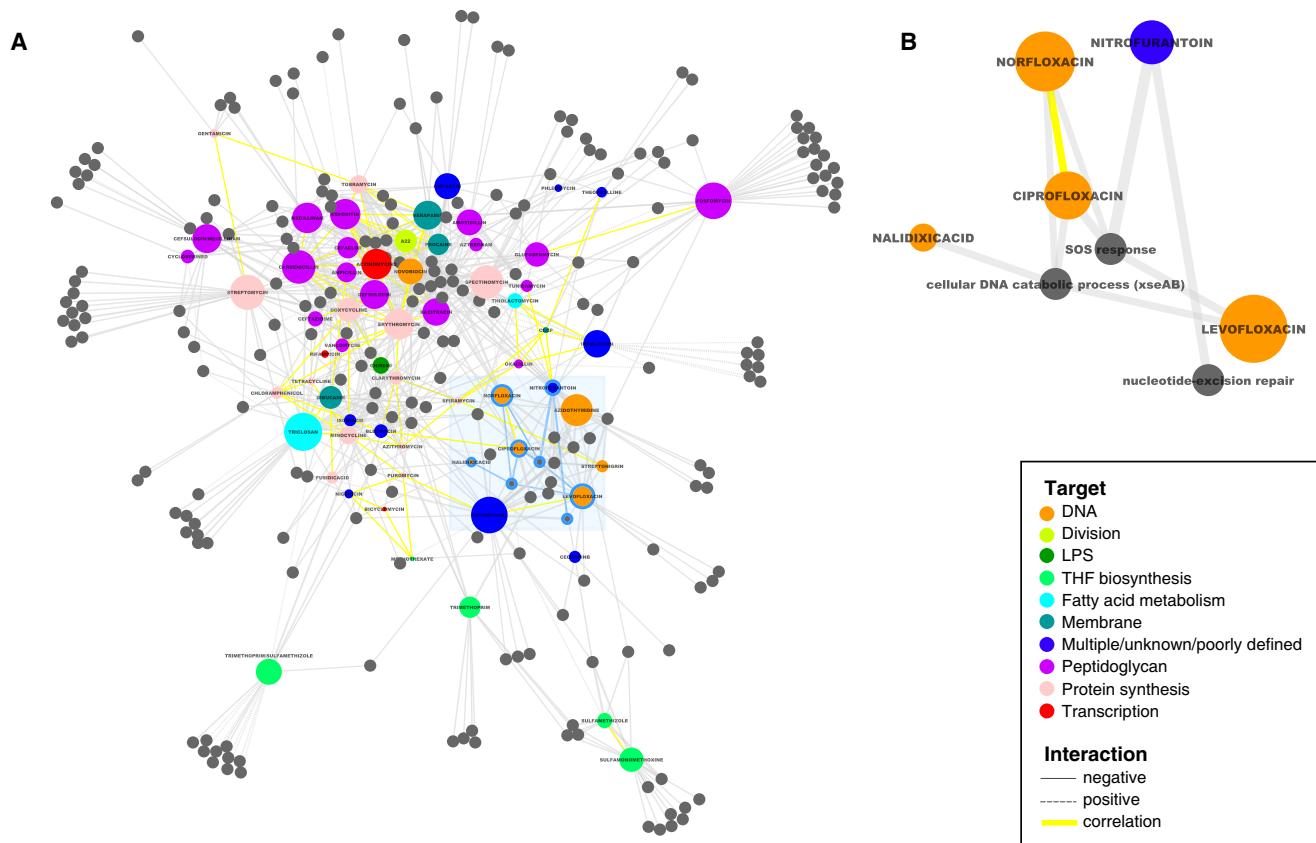
on the bottom edge, and those targeting peptidoglycan (PG; purple) predominantly cluster in the upper left. Interestingly,  $\beta$ -lactams cover the center of the PG cluster, whereas drugs targeting other steps of PG synthesis are located at the periphery. The correlation coefficients between  $\beta$ -lactams reveal that the similarity of their phenotypic signatures is related to their respective primary target penicillin-binding proteins (Figure S5A). Interestingly, known synergistic double-drug combinations (TMP/sulfonamides and mecillinam/cefsulodin) occupy spaces that are distinct from either individual drug, arguing that the combination elicits a different cellular response from the individual drugs. It will be interesting to determine whether this holds true for antagonistic or neutral interactions or whether these combinations elicit responses closer to one or both drugs.

Importantly, specific drug-GO interactions suggest hypotheses for mechanism of action even for well-studied drugs. Quinolones inhibit DNA gyrase by trapping it as a quinolone adduct, whose mechanism of resolution is poorly understood (Drlica et al., 2008). One GO category, “cellular DNA catabolic process” (*xseAB*), selectively and specifically ( $p = 10^{-6}$ ) interacted negatively with all four quinolones screened (Figure 7B and Figure S5B), expanding on a previous report of *xseAB* mutant sensitivity to fluoroquinolones (Tamae et al., 2008). We suggest that XseAB (exonuclease VII) is the enzyme that cleaves quinolone-bound DNA gyrase from the DNA to allow repair to proceed, a possibility that we are currently exploring.

Our drug network also provides clues for the mode of action of poorly described drugs and, conversely, suggests that additional factors are required to explain the action of other drugs. Nitrofurantoin (NTF; Figure 7B) is reported to have a multifaceted impact on cells (McOske and Fitzpatrick, 1994; Tu and McCalla, 1975), but our data suggest that its cytotoxic effects reflect DNA damage, as it causes lesions requiring nucleotide excision repair (NER) and activates the SOS response. NTF is the only DNA-targeting drug requiring NER, but not double-strand break repair, suggesting that its primary toxic lesion is associated with the replication fork (Figure S5B). In addition, our network analysis validates the idea that indolicidin, a neutrophil antimicrobial peptide, mediates its effects by compromising the inner-membrane permeability of *E. coli* in a manner that is similar to the proton motive force uncoupler, CCCP (Falla et al., 1996). Finally, phleomycin and bleomycin do not cluster with DNA response drugs, suggesting that they have broader cellular impact (Hecht, 2000; Yeh et al., 2006) from inducing DNA scissions (Giloni et al., 1981). These insights suggest that this phenomic data set is a rich source for discovery of drug function and interrelationships.

### Perspectives

To keep pace with exploding sequence information, cost-effective, high-throughput phenotyping technologies must be developed. Here, we show that phenomic profiling in *E. coli* fulfills this goal. Our data set is of great utility in identifying the function of orphan genes. Three cases (*marB*, *lpoA*, and *lpoB*) were investigated here or in a study based on this data set (Typas et al., 2010), and we are actively pursuing functional discovery of numerous (>20) orphan genes, as well as annotated genes with previously unsuspected roles in collaboration with others. Because > 25% of the orphan genes are highly correlated to



**Figure 7. Network View Reveals New Insights into Drug Action**

(A) Colored nodes represent all drugs profiled in this study that were found to have significant interactions with gene ontology (GO) biological process groups (gray nodes). Connections between nodes represent significant drug-GO interactions ( $p \leq 10^{-3}$ , gray) or high drug-drug correlation ( $r \geq 0.32$ ;  $p \leq 10^{-97}$ , yellow). Drug node size is based on the number of connections associated with that node, i.e., larger nodes have more drug-GO interactions. Spatial clustering is driven by the  $p$  values of drug-GO interactions and drug-drug correlations, resulting in drugs with similar cellular action lying near each other in the network. Drugs with multiple, unknown, or poorly defined targets are shown in dark blue.

(B) Zoomed view of subnetwork shadowed by light blue box in (A). All four quinolones screened (orange) interact negatively with *xseAB* (exonuclease VII) and are the only drugs that require the exonuclease;  $p = 10^{-6}$ . NTF is found to activate the SOS response and create lesions requiring nucleotide-excision repair (Figure S5B). Connections between nodes represent significant drug-GO interactions ( $p < 10^{-3}$ , gray). For all drug-GO interactions, see Table S6. See also Figure S5.

an annotated gene ( $r \geq 0.5$ ), this data set provides a rapid method for function discovery.

An important finding is that the most responsive orphan genes tend to be narrowly distributed among bacteria. Interestingly, our results mirror initial observations from human microbiome studies. These studies found that: (1) roughly half of the functions encoded in the minimal gut metagenome (ubiquitously present in all 124 individuals screened) are both unknown and of limited evolutionary conservation (Qin et al., 2010), and (2) across four pan-genome species analyzed, the vast majority of noncommon genes were of either unknown function (~70%) or were unique family members of functions that were part of the core gene set (Nelson et al., 2010). The latter are probably species-specific additions to conserved biological processes of the pan-genome. Together, these studies argue that, when computational methods based on gene conservation fail, large-scale phenomic analyses can be a second tier for assigning function. To make this approach a reality, low-cost methods for developing deletion

libraries must be developed (Goodman et al., 2009). Single-gene deletion ordered libraries are currently available for only a handful of organisms (Cameron et al., 2008; de Berardinis et al., 2008; Gallagher et al., 2007; Goodman et al., 2009; Kim et al., 2010; Liu et al., 2008; Noble et al., 2010; references in Barker et al., 2010), but advances in transposon mutagenesis make it feasible to create ordered mutant libraries in most organisms. In *E. coli*, expansion of this work will rest on the ability to assess additional phenotypes through deeper exploration of phenotypic space. The greatest potential resides at the intersection of screening more diverse stresses and incorporating additional cellular readouts. Colorimetric readouts would enable measurement of transcriptional activity or biofilm formation on solid agar surfaces and represent an immediate potential advance for phenomic profiling. High-throughput microscopy would provide a new avenue for such approaches (Werner et al., 2009).

Our data set provides information on a substantial collection of antibiotics/antimicrobial compounds that cover a broad

spectrum of drug targets, structural classes, and drug generations, providing a platform for future studies focused on natural products or antimicrobials with unknown targets. Our data set can also provide a platform for studying the mechanism behind drug interactions (Yeh et al., 2009), as shown here for the case of sulfonamides and TMP. Understanding the mechanism underlying known drug interactions may help to predict novel interactions and manipulate existing drug combinations to increase their effectiveness in the clinic.

In summary, we have generated a valuable resource for microbiologists studying a wide range of biology and demonstrated the numerous and diverse applications of this data set to infer information on both gene and drug function. As, to our knowledge, the most comprehensive prokaryotic chemical genomic study to date (3979 strains  $\times$  324 conditions), our data set will serve as a base for future studies that aim to increase information and/or resolution on both the gene and drug fronts. We hope that the usefulness of this resource will trigger analogous studies in other organisms, bringing us a step nearer to closing the gene sequence-function gap.

## EXPERIMENTAL PROCEDURES

Experimental procedures are partially elaborated upon in the text and figure legends and are fully explicated in the *Supplemental Information*.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, and seven tables and can be found with this article at doi:10.1016/j.cell.2010.11.052.

## ACKNOWLEDGMENTS

We thank R. Kishony, J. Weissman, A. Hochschild, and R.G. Martin for critically reading this manuscript; J. Hu and P. Thomas for hosting these data on *E. coli* Wiki; C. Raetz for CHIR-090 and M. Gottesman for Bicyclomycin; H. Mori for the Keio Collection; G. Storz for sharing the sRNA deletion library prior to publication; J. Greenblatt and A. Emili for SPA-tagged alleles; W. Margolin, R. Misra, T. Silhavy, and B. Palsson for mutants; and T. Baker and R. Sauer for controllable degradation plasmids. This work was supported by NIH R01 GM085697 and ARRA GM085697-01S1 to C.A.G. and N.J.K.; NIH R01 GM036278 to C.A.G.; NIH K99GM092984 to A.T.; NIH AI060744 to M.E.W.; NIH GM078338 to S.S.; NIH F31 DE020206-01 and NIH T32 DE007306 (R.J.N. support); European Molecular Biology Organization long-term fellowship (to A.T.); and Human Frontier Science Program Long-Term Postdoctoral Fellowship (to P.B.). In loving memory of S.R. Bartlett and J.T. Blair.

A.T., R.J.N., N.J.K., and C.A.G. conceived this study; A.T. and R.J.N. designed research; A.T., R.J.N., Y.J.C., S.L., J.L., and A.W. performed the screen; R.J.N. did the data processing; R.J.N., S.S., P.B., and A.T. analyzed the data; A.T., R.J.N., C.A.G., and S.L. interpreted the data; R.J.N., A.T., and M.Z. performed the follow-up work in the antifolate drug story, except for the work in *Streptococcus pneumoniae*, which was conducted by K.M.K.; R.J.N., R.C., A.T., and M.Z. performed the follow-up work in the *mar* story; R.J.N., A.T., and C.A.G. wrote the manuscript; R.J.N. and M.S. prepared the figures; P.B., R.C., M.E.W., K.M.K., S.L., S.S., and N.J.K. edited the manuscript; A.T. and C.A.G. supervised all aspects of this project.

Received: August 31, 2010

Revised: November 7, 2010

Accepted: November 24, 2010

Published online: December 23, 2010

## REFERENCES

- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 2006.0008.
- Barker, C.A., Farha, M.A., and Brown, E.D. (2010). Chemical genomic approaches to study model microbes. *Chem. Biol.* 17, 624–632.
- Beltrao, P., Cagney, G., and Krogan, N.J. (2010). Quantitative genetic interactions reveal biological modularity. *Cell* 141, 739–745.
- Bochner, B.R. (2009). Global phenotypic characterization of bacteria. *FEMS Microbiol. Rev.* 33, 191–205.
- Bollenbach, T., Quan, S., Chait, R., and Kishony, R. (2009). Nonoptimal microbial response to antibiotics underlies suppressive drug interactions. *Cell* 139, 707–718.
- Butland, G., Peregrín-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadian, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., et al. (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* 433, 531–537.
- Butland, G., Babu, M., Díaz-Mejía, J.J., Bohdane, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A.G., Pogoutse, O., et al. (2008). eSGA: *E. coli* synthetic genetic array analysis. *Nat. Methods* 5, 789–795.
- Cameron, D.E., Urbach, J.M., and Mekalanos, J.J. (2008). A defined transposon mutant library and its use in identifying motility genes in *Vibrio cholerae*. *Proc. Natl. Acad. Sci. USA* 105, 8736–8741.
- Carpentier, A.S., Torrésani, B., Grossmann, A., and Hénaut, A. (2005). Decoding the nucleoid organisation of *Bacillus subtilis* and *Escherichia coli* through gene expression data. *BMC Genomics* 6, 84.
- Couturier, E., and Rocha, E.P. (2006). Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol. Microbiol.* 59, 1506–1518.
- D'Elia, M.A., Pereira, M.P., and Brown, E.D. (2009). Are essential genes really essential? *Trends Microbiol.* 17, 433–438.
- de Berardinis, V., Vallenet, D., Castelli, V., Besnard, M., Pinet, A., Cruaud, C., Samair, S., Lechaplain, C., Gyapay, G., Richez, C., et al. (2008). A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* 4, 174.
- Drlica, K., Malik, M., Kerns, R.J., and Zhao, X. (2008). Quinolone-mediated bacterial death. *Antimicrob. Agents Chemother.* 52, 385–392.
- Falla, T.J., Karunaratne, D.N., and Hancock, R.E. (1996). Mode of action of the antimicrobial peptide indolicidin. *J. Biol. Chem.* 271, 19298–19303.
- Fiedler, D., Braberg, H., Mehta, M., Chechik, G., Cagney, G., Mukherjee, P., Silva, A.C., Shales, M., Collins, S.R., van Wageningen, S., et al. (2009). Functional organization of the *S. cerevisiae* phosphorylation network. *Cell* 136, 952–963.
- Field, M.S., Szebenyi, D.M., and Stover, P.J. (2006). Regulation of de novo purine biosynthesis by methenyltetrahydrofolate synthetase in neuroblastoma. *J. Biol. Chem.* 281, 4215–4221.
- Gallagher, L.A., Ramage, E., Jacobs, M.A., Kaul, R., Brittnacher, M., and Manoil, C. (2007). A comprehensive transposon mutant library of *Francisella novicida*, a bioweapon surrogate. *Proc. Natl. Acad. Sci. USA* 104, 1009–1014.
- Giaever, G., Flaherty, P., Kumm, J., Proctor, M., Nislow, C., Jaramillo, D.F., Chu, A.M., Jordan, M.I., Arkin, A.P., and Davis, R.W. (2004). Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc. Natl. Acad. Sci. USA* 101, 793–798.
- Giloni, L., Takeshita, M., Johnson, F., Iden, C., and Grollman, A.P. (1981). Bleomycin-induced strand-scission of DNA. Mechanism of deoxyribose cleavage. *J. Biol. Chem.* 256, 8608–8615.
- Girgis, H.S., Hottes, A.K., and Tavazoie, S. (2009). Genetic architecture of intrinsic antibiotic susceptibility. *PLoS ONE* 4, e5629.
- Girgis, H.S., Liu, Y., Ryu, W.S., and Tavazoie, S. (2007). A comprehensive genetic characterization of bacterial motility. *PLoS Genet.* 3, 1644–1660.

- Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A., Knight, R., and Gordon, J.I. (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* 6, 279–289.
- Griffith, K.L., Shah, I.M., and Wolf, R.E., Jr. (2004). Proteolytic degradation of *Escherichia coli* transcription activators SoxS and MarA as the mechanism for reversing the induction of the superoxide (SoxRS) and multiple antibiotic resistance (Mar) regulons. *Mol. Microbiol.* 51, 1801–1816.
- Hecht, S.M. (2000). Bleomycin: new perspectives on the mechanism of action. *J. Nat. Prod.* 63, 158–168.
- Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D., et al. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* 320, 362–365.
- Hobbs, E.C., Astarita, J.L., and Storz, G. (2010). Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in *Escherichia coli*: analysis of a bar-coded mutant collection. *J. Bacteriol.* 192, 59–67.
- Hoon, S., Smith, A.M., Wallace, I.M., Suresh, S., Miranda, M., Fung, E., Proctor, M., Shokat, K.M., Zhang, C., Davis, R.W., et al. (2008). An integrated platform of genomic assays reveals small-molecule bioactivities. *Nat. Chem. Biol.* 4, 498–506.
- Hu, P., Janga, S.C., Babu, M., Díaz-Mejía, J.J., Butland, G., Yang, W., Pogoutse, O., Guo, X., Phanse, S., Wong, P., et al. (2009). Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biol.* 7, e96.
- Jeanguenin, L., Lara-Nunez, A., Pribat, A., Hamner Mageroy, M., Gregory, J.F., III, Rice, K.C., de Crecy-Lagard, V., and Hanson, A.D. (2010). Moonlighting glutamate formiminotransferases can functionally replace 5-formyltetrahydrofolate cyclohydrolase. *J. Biol. Chem.*, in press. Published online October 15, 2010. 10.1074/jbc.M110.190504.
- Jeong, K.S., Ahn, J., and Khodursky, A.B. (2004). Spatial patterns of transcriptional activity in the chromosome of *Escherichia coli*. *Genome Biol.* 5, R86.
- Joyce, A.R., Reed, J.L., White, A., Edwards, R., Osterman, A., Baba, T., Mori, H., Lesely, S.A., Palsson, B.O., and Agarwalla, S. (2006). Experimental and computational assessment of conditionally essential genes in *Escherichia coli*. *J. Bacteriol.* 188, 8259–8271.
- Kim, D.U., Hayles, J., Kim, D., Wood, V., Park, H.O., Won, M., Yoo, H.S., Duhig, T., Nam, M., Palmer, G., et al. (2010). Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. *Nat. Biotechnol.* 28, 617–623.
- Kohanski, M.A., Dwyer, D.J., Wierzbowski, J., Cottarel, G., and Collins, J.J. (2008). Mistranslation of membrane proteins and two-component system activation trigger antibiotic-mediated cell death. *Cell* 135, 679–690.
- Kwon, Y.K., Higgins, M.B., and Rabinowitz, J.D. (2010). Antifolate-induced depletion of intracellular glycine and purines inhibits thymineless death in *E. coli*. *ACS Chem. Biol.* 5, 787–795.
- Kwon, Y.K., Lu, W., Melamud, E., Khanam, N., Bognar, A., and Rabinowitz, J.D. (2008). A domino effect in antifolate drug action in *Escherichia coli*. *Nat. Chem. Biol.* 4, 602–608.
- Lee, W., St Onge, R.P., Proctor, M., Flaherty, P., Jordan, M.I., Arkin, A.P., Davis, R.W., Nislow, C., and Giaever, G. (2005). Genome-wide requirements for resistance to functionally distinct DNA-damaging agents. *PLoS Genet.* 1, e24.
- Liu, O.W., Chun, C.D., Chow, E.D., Chen, C., Madhani, H.D., and Noble, S.M. (2008). Systematic genetic analysis of virulence in the human fungal pathogen *Cryptococcus neoformans*. *Cell* 135, 174–188.
- Liu, A., Tran, L., Becket, E., Lee, K., Chinn, L., Park, E., Tran, K., and Miller, J.H. (2010). Antibiotic sensitivity profiles determined with an *Escherichia coli* gene knockout collection: generating an antibiotic bar code. *Antimicrob. Agents Chemother.* 54, 1393–1403.
- Martin, R.G., and Rosner, J.L. (2002). Genomics of the marA/soxS/rob regulon of *Escherichia coli*: identification of directly activated promoters by application of molecular genetics and informatics to microarray data. *Mol. Microbiol.* 44, 1611–1624.
- Martin, R.G., Bartlett, E.S., Rosner, J.L., and Wall, M.E. (2008). Activation of the *Escherichia coli* marA/soxS/rob regulon in response to transcriptional activator concentration. *J. Mol. Biol.* 380, 278–284.
- McOske, C.C., and Fitzpatrick, P.M. (1994). Nitrofurantoin: mechanism of action and implications for resistance development in common uropathogens. *J. Antimicrob. Chemother.* 33 Suppl A, 23–30.
- Nelson, K.E., Weinstock, G.M., Highlander, S.K., Worley, K.C., Creasy, H.H., Wortman, J.R., Rusch, D.B., Mitreva, M., Sodergren, E., Chinwalla, A.T., et al; Human Microbiome Jumpstart Reference Strains Consortium. (2010). A catalog of reference genomes from the human microbiome. *Science* 328, 994–999.
- Nicoloff, H., Perreten, V., McMurry, L.M., and Levy, S.B. (2006). Role for tandem duplication and Ion protease in AcrAB-TolC-dependent multiple antibiotic resistance (Mar) in an *Escherichia coli* mutant without mutations in marRAB or acrRAB. *J. Bacteriol.* 188, 4413–4423.
- Nicoloff, H., Perreten, V., and Levy, S.B. (2007). Increased genome instability in *Escherichia coli* Ion mutants: relation to emergence of multiple-antibiotic-resistant (Mar) mutants caused by insertion sequence elements and large tandem genomic amplifications. *Antimicrob. Agents Chemother.* 51, 1293–1303.
- Noble, S.M., French, S., Kohn, L.A., Chen, V., and Johnson, A.D. (2010). Systematic screens of a *Candida albicans* homozygous deletion library decouple morphogenetic switching and pathogenicity. *Nat. Genet.* 42, 590–598.
- Pan, X., Yuan, D.S., Xiang, D., Wang, X., Sookhai-Mahadeo, S., Bader, J.S., Hieter, P., Spencer, F., and Boeke, J.D. (2004). A robust toolkit for functional profiling of the yeast genome. *Mol. Cell* 16, 487–496.
- Parsons, A.B., Lopez, A., Givoni, I.E., Williams, D.E., Gray, C.A., Porter, J., Chua, G., Sopko, R., Brost, R.L., Ho, C.H., et al. (2006). Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell* 126, 611–625.
- Pathania, R., Zlitni, S., Barker, C., Das, R., Gerritsma, D.A., Lebert, J., Awuah, E., Melacini, G., Capretta, F.A., and Brown, E.D. (2009). Chemical genomics in *Escherichia coli* identifies an inhibitor of bacterial lipoprotein targeting. *Nat. Chem. Biol.* 5, 849–856.
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al; MetaHIT Consortium. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65.
- Rocha, E.P., and Danchin, A. (2003). Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat. Genet.* 34, 377–378.
- Rouquerol, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S.R., Qu, H., Shales, M., Park, H.O., Hayles, J., et al. (2008). Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science* 322, 405–410.
- Schirch, V., Hopkins, S., Villar, E., and Angelaccio, S. (1985). Serine hydroxymethyltransferase from *Escherichia coli*: purification and properties. *J. Bacteriol.* 163, 1–7.
- Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., et al. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniaarray profile. *Cell* 123, 507–519.
- Stover, P., and Schirch, V. (1993). The metabolic role of leucovorin. *Trends Biochem. Sci.* 18, 102–106.
- Taber, H.W., Mueller, J.P., Miller, P.F., and Arrow, A.S. (1987). Bacterial uptake of aminoglycoside antibiotics. *Microbiol. Rev.* 51, 439–457.
- Tamae, C., Liu, A., Kim, K., Sitz, D., Hong, J., Becket, E., Bui, A., Solaimani, P., Tran, K.P., Yang, H., and Miller, J.H. (2008). Determination of antibiotic hypersensitivity among 4,000 single-gene-knockout mutants of *Escherichia coli*. *J. Bacteriol.* 190, 5981–5988.

- Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Pagé, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., et al. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* 294, 2364–2368.
- Toro, E., and Shapiro, L. (2010). Bacterial chromosome organization and segregation. *Cold Spring Harb. Perspect. Biol.* 2, a000349.
- Travers, A., and Muskhelishvili, G. (2005). DNA supercoiling - a global transcriptional regulator for enterobacterial growth? *Nat. Rev. Microbiol.* 3, 157–169.
- Tu, Y., and McCalla, D.R. (1975). Effect of activated nitrofurans on DNA. *Biochim. Biophys. Acta* 402, 142–149.
- Typas, A., Nichols, R.J., Siegele, D.A., Shales, M., Collins, S.R., Lim, B., Braberg, H., Yamamoto, N., Takeuchi, R., Wanner, B.L., et al. (2008). High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nat. Methods* 5, 781–787.
- Typas, A., Banzhaf, M., van den Berg van Saparoea, B., Verheul, J., Biboy, J., Nichols, R.J., Zietek, M., Beilharz, K., Kannenberg, K., von Rechenberg, M., et al. (2010). Regulation of peptidoglycan synthesis by outer-membrane proteins. *Cell* 143, 1097–1109.
- Vora, T., Hottes, A.K., and Tavazoie, S. (2009). Protein occupancy landscape of a bacterial genome. *Mol. Cell* 35, 247–253.
- Warner, D.M., and Levy, S.B. (2010). Different effects of transcriptional regulators MarA, SoxS and Rob on susceptibility of *Escherichia coli* to cationic antimicrobial peptides (CAMPs): Rob-dependent CAMP induction of the marRAB operon. *Microbiology* 156, 570–578.
- Werner, J.N., Chen, E.Y., Guberman, J.M., Zippilli, A.R., Irgon, J.J., and Gitai, Z. (2009). Quantitative genome-scale analysis of protein localization in an asymmetric bacterium. *Proc. Natl. Acad. Sci. USA* 106, 7858–7863.
- Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B., and Gill, R.T. (2010). Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat. Biotechnol.* 28, 856–862.
- Xu, D., Jiang, B., Ketela, T., Lemieux, S., Veillette, K., Martel, N., Davison, J., Sillaots, S., Trosok, S., Bachewich, C., et al. (2007). Genome-wide fitness test and mechanism-of-action studies of inhibitory compounds in *Candida albicans*. *PLoS Pathog.* 3, e92.
- Yamamoto, N., Nakahigashi, K., Nakamichi, T., Yoshino, M., Takai, Y., Touda, Y., Furabayashi, A., Kinjo, S., Dose, H., Hasegawa, M., et al. (2009). Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. *Mol. Syst. Biol.* 5, 335.
- Yeh, P., Tschumi, A.I., and Kishony, R. (2006). Functional classification of drugs by properties of their pairwise interactions. *Nat. Genet.* 38, 489–494.
- Yeh, P.J., Hegreness, M.J., Aiden, A.P., and Kishony, R. (2009). Drug interactions and the evolution of antibiotic resistance. *Nat. Rev. Microbiol.* 7, 460–466.

## EXTENDED EXPERIMENTAL PROCEDURES

### Bacterial Strains

The screened collection includes 4334 deletions of non-essential genes (Keio Collection; Baba et al., 2006), 148 SPA-tagged derivatives of essential genes (Butland et al., 2005), 9 point-mutant alleles of essential genes (plus corresponding “linked” strains for each of the 9 alleles, in which the antibiotic-resistance cassette was linked to the wild-type allele as a control), 5 DAS-tagged essential genes (McGinnness et al., 2006), 2 truncations of essential genes, and 100 deletions of small RNAs (sRNAs) and small proteins (Hobbs et al., 2010). All mutant strains are marked with or linked to kanamycin. All linked alleles and controls, DAS-tags, and truncations were constructed in BW25113 (Keio Collection strain background) using standard recombineering methods. Primers are available upon request. SPA-tagged essential genes, originally constructed in W3110 (Butland et al., 2005), were transduced into BW25113 using P1 phage. sRNA and small protein deletions were constructed in MG1655 (Hobbs et al., 2010) and used in that background. For all mutants, except for the sRNA/small protein library, two independently derived clones were screened and evaluated for reproducibility of growth. For the sRNA/small protein mutants a single isolated clone was arrayed twice. Of the 4607 strains screened, 628 were eliminated from the study either because of consistently discrepant growth between clones or because updated curation of the Keio collection reclassified them as incorrect strains, leaving a total of 3979 strains in the final dataset (3737 Keio, 117 SPA, 5 DAS, 9 alleles, 9 linked controls, 2 truncations, and 100 sRNAs). A list of these strains is available online at <http://ecoliwiki.net/tools/chemgen/>.

All individual mutants used for follow up biology stories were re-transduced to BW25113 before further experiments. During the entire study, BW25113 is referred to as wild-type. Double mutants used in the experiments dissecting the function of MarB were constructed using P1 transduction or recombineering.

### Growth Conditions for Drug Synergy Experiments

Wild-type and mutant strains were grown overnight to stationary phase in LB Lennox broth, and diluted into fresh drug-containing LB broth (0.1 µg/ml TMP and/or 50 µg/ml Sulfamethizole) at 1:1000 v/v. Cultures were then grown aerobically in 150 ml flasks at 37°C and optical densities were recorded at 600nm every 20–30 minutes. Final culture density was recorded at a 24-hour endpoint. The *Streptococcus pneumoniae* experiments used serotype 2 strain D39 (IU1690; Lanie et al., 2007), which was cultured statically in Brain Heart Infusion broth (Bacto BHI, Becton Dickinson) at 37°C in an atmosphere of 5% CO<sub>2</sub>. Growth was monitored by optical density (OD<sub>620</sub>). Bacterial cultures were started from frozen stocks and propagated overnight in BHI as described (Ramos-Montanez et al., 2008). Overnight cultures in exponential phase (OD<sub>620</sub> = 0.1–0.3) were diluted back to a density of OD<sub>620</sub>~0.001 to start final cultures. Final cultures were grown to a density of OD<sub>620</sub>~0.1 and divided among 16 mm glass tubes containing no antibiotic, 7.8 µg/ml TMP, 1000 µg/ml SMT or both antibiotics.

### Screen Conditions and Data Processing

The collection was arrayed in 1536-format onto six LB-agar plates using a Singer Rotor robot. The independently derived or duplicate clones of each strain were arrayed side-by-side. Glycerol stocks of each plate were kept at –80°C, and re-arrayed at least once every four weeks in effort to reduce the passage and accumulation of spontaneous suppressor mutants throughout the study. For the same reasons and to minimize the exposure time of a given plate at 4°C, the collection was passaged to fresh plates at least once a week.

For each screen, the chemical/stress was mixed into LB agar (except where indicated) at the concentrations listed in Table S1. Plates were allowed to dry for two days, and then pinned with the collection. For the majority of conditions, plates were removed from the incubator following 14–16 hrs growth at 37°C. At this timepoint, fitness differences were apparent but growth had not saturated. Several stresses, e.g. low temperature, required longer incubation times. Plates were then imaged using a Canon G10 digital camera, and the images processed into colony size data, which was used to generate Condition-Gene scores. All image processing, colony size measurements, and score generation was carried out as described previously for genetic interactions (Collins et al., 2006; Typas et al., 2008).

Two key quality control principles utilized in the S-score fitness measurement procedure for genetic interaction analysis are not applicable to our chemical profiling analysis. First, in genetic interaction analysis, each gene-gene interaction is constructed twice, once with each mutation as the donor, and those scores are averaged. Here, condition-gene interactions are constructed only one way, (i.e. stress applied to single mutants), so the scores generated here are analogous to “unaveraged” S-scores (Collins et al., 2006). Second, a key quality control principle underlying the S-score procedure is based on genetic linkage, which is not applicable to Condition-Gene interactions. Therefore, two alternative quality control measures were implemented: (1) the interquartile range (IQR) of the distribution of normalized colony sizes for each screen; and (2) the correlation between replicates of the same screen. Screens with high IQR, or with poor correlation within a series of technical replicates were eliminated from the dataset to minimize noise. High IQR or low correlation of technical replicates was found to be due to a variety of factors including high rates of spontaneous suppressor mutants (due to selection under high stress), poor or saturated growth and human error in the generation of the data.

The sRNA/protein deletion library was constructed in MG1655, whereas the rest of the collection was constructed in BW25113. As these strains were reasonably close, we did not transduce the sRNA/small protein deletions into BW25113. Therefore, raw colony

sizes of all sRNA/small protein mutants (which were arrayed on the same plate of the collection) were normalized separately to eliminate the relative comparisons with the BW25113-based mutants arrayed on the plate. This approach reduced complications from strain background differences, but it should be noted that sRNA/small protein clustering patterns are less reliable, unless the particular mutant was found to have strong phenotypes.

### Phenotype Analysis

For each screen, we estimated the mean and IQR of Condition-Gene scores and re-scaled the data to a standard normal distribution ( $IQR=1.35$ ). We reasoned that if there were no phenotypes (outliers) in a given screen, all scores would conform to a standard normal distribution. For example, we expect 5% of the scores to be less than the 5<sup>th</sup> normal percentile. More generally, we expect a fraction of scores,  $p$ , to be less than the  $p^{\text{th}}$  quantile of the normal distribution. If a fraction of scores,  $x$  ( $x>p$ ), is found to be less than the  $p^{\text{th}}$  quantile, then the false discovery rate (FDR) associated with the  $p^{\text{th}}$  quantile would be  $p/x$  (Efron et al., 2001).

Using this approach, we classified all scores with FDR 5% or less as phenotypes. Responsive genes were defined as strains for which at least one phenotype was identified. It is important to note that the 5% FDR attached to each phenotype was calculated individually for each phenotype. As each phenotype was deemed to be of independent interest, we did not apply a multiple testing correction across phenotypes.

Conditionally-essential genes were identified using a combination of the probability scores and raw colony sizes. Because cells are always transferred to the screen plate during the pinning process, lethal events cannot be scored as colony sizes of zero, unless the stress/drug is bactericidal. Therefore, we defined conditionally-essential events as those in which the probability score was 1 (definite outlier), and raw colony size was less than 60 pixels. These stringent parameters were empirically defined based on known lethal interactions, and likely exclude other true conditionally-lethal events.

### Gene Annotation

The COG annotation index (<http://www.ncbi.nlm.nih.gov/COG/>; Tatusov et al., 2003) was used as the basis for grouping genes according to function. We applied minimal changes to the main COG function families, manually curating genes of interest based on existing COG annotations, published literature (we sampled Pubmed and Ecocyc) and whether evidence was experimental/computational. Genes with purely computationally predicted functions were mostly assigned to the “general function prediction only” or the “unknown function” families. Additionally, each gene was assigned to only a single function.

Protein localization of *E. coli* K12 genes has been reported previously (Riley et al., 2006). Conditionally-essential gene enrichment for envelope proteins was calculated relative to the Keio Collection essential genes (Baba et al., 2006) using Fisher's exact test.

Orphan genes were identified previously, as were ortholog counts across bacterial clades for all *E. coli* K12 genes (Hu et al., 2009). These counts were used to assign all strains in our dataset to  $\gamma$ -proteobacteria-only, proteobacteria-only, or broadly-conserved categories. Genes assigned to the  $\gamma$ -proteobacteria-only category were those for which no orthologs were identified outside the  $\gamma$ -proteobacteria. The proteobacteria-only class contains genes with orthologs outside the  $\gamma$ -proteobacteria, but not outside the proteobacteria. Broadly conserved genes were defined as those for which orthologs had been identified outside the proteobacteria.

### RNA Isolation, cDNA Preparation, and Quantitative RT-PCR

Cultures were grown to OD<sub>600</sub> of 0.3. Samples (8 ml) were harvested and added to ice-cold 5% water-saturated phenol in ethanol solution, centrifuged for 2 min, and the cell pellets were flash-frozen in liquid nitrogen before storing at -80 °C. RNA was extracted using the hot-phenol technique, with modifications. Briefly, cell pellets were resuspended in 500  $\mu$ l of lysis solution (320 mM Na acetate, pH 4.6, 8% SDS, 16 mM EDTA), and mixed with 1 ml of water-buffered phenol. The samples were incubated at 65 °C for 5 min with intermittent mixing. The samples were then placed on ice for 5 min and centrifuged for 10 min at 4 °C. The supernatant was extracted twice with phenol-chloroform, precipitated with 2.5 volumes of 100% ethanol and washed with 70% ethanol. The RNA pellet was air dried and resuspended in 85  $\mu$ l of RNase free water. Genomic DNA was removed from the samples using Turbo DNA-free DNase Treatment according to the manufacturer's directions for rigorous DNase treatment (Applied Biosystems, Foster City, CA, USA). cDNA was prepared for qRT-PCR as previously described using 5  $\mu$ g of input RNA (Cummings et al., 2006).

Quantitative RT-PCR reactions were carried out using Stratagene Brilliant II Sybrgreen master mix according to the manufacturer's directions (Agilent Technologies, La Jolla, CA, USA), and 6 pmol of each forward and reverse primer (Integrated DNA Technologies). Primer sequences are available upon request. Real-time PCR was performed with a Stratagene Mx3000P sequence-detection system (Agilent Technologies). Data were analyzed as described (Vandesompele et al., 2002) with *recA* and *gyrA* as references.

### Genomic Context Analysis

For all genomic context plots, base pair coordinates for each gene were adjusted to center the chromosome around the origin of replication ( $oriC = 0$  base pairs). Each trace (solid line) represents relative enrichment for the designated gene class in 100 kb windows beginning at the origin of replication and sliding by a 1 kb interval. Enrichment values were calculated using the following approach:

Let  $x$  be the number of genes in a given class (i.e. responsive genes) in a 100 kb window out of  $m$  genes in that window; let  $f$  be the fraction of genes in that class in the whole genome. Our enrichment value is  $(x/m - f)$ . If we wish to test the hypothesis that  $x/m = f$ , the chi-squared test statistic would be  $m^*(x/m-f)^2$ . Thus, we used it as our enrichment measure. High (or low) enrichment values would

indicate association between class membership and genomic position. Note that this is based on a one-sample test for a proportion. For visual purposes, we plotted the square root of the enrichment values.

The significance of the enrichment is not straightforward to establish since there are correlations between neighboring enrichment values (because of overlapping windows). To find out how high (or low) enrichment values would be under the null hypothesis of no association between genomic position and class membership, we used a permutation test. Under the null hypothesis, class membership and genomic position would be exchangeable. Therefore, we created 1000 random permutations of the data where class membership of genes was permuted. The max (and min) enrichment values from each permutation were recorded. These provide us with the distribution of the max (and min) enrichment values under the null hypothesis. We used the 95th and 5<sup>th</sup> percentiles of the max and min enrichment values, respectively, to establish our thresholds of positive and negative spatial enrichment corresponding to a p-value of 0.05.

Coding strand assignments were reported previously (Baba et al., 2006), and were inverted for all genes on the left arc of the *oriC*-centered chromosome to standardize the meaning of the strand designation for all genes. Plus strand indicates the gene is transcribed in the same direction as DNA replication, whereas minus strand indicates the gene is transcribed in the opposite direction of DNA replication. Strand bias values were calculated using the following approach:

Let x be the fraction of genes in a given class (i.e., responsive genes) on the plus strand, and y be the fraction of genes in that category on the minus strand. Also let m be the number of genes on the plus strand in the whole genome, and n be the number of genes on the minus strand in the whole genome. To test for strand bias in a given class, we used the formula  $(x-y) / \sqrt{(1/m)+(1/n)}$ . This statistic is based on a two-sample test of a proportion.

To establish statistical significance of the strand bias value, we used a permutation test similar to the one described above for genomic context enrichment values. P-values were calculated directly based on 100,000 permutations of the class designations as described above for genomic context enrichment calculations.

### Drug Network Analysis

Gene Ontology Biological Process associations were obtained from the Gene Ontology and EcoliWiki collaborative annotation for *E. coli* K-12 version 1.21 (<http://www.ecocyc.org>). Drug-GO scores used to build the network in Figure 7 were generated using the hypergeometric probability density function in Matlab. Using the binary phenotype/no phenotype data produced as described above (Phenotype Analysis section), this procedure tested for relative enrichment of phenotypes within a GO biological process relative to the entire dataset for each of the 324 screens. Scores were produced by calculating the negative natural logarithm of the resulting p-values, and the sign of the interaction (positive or negative) was re-attached by calculating the mean of all condition-gene scores underlying the phenotypes in a given group. All conditions tested in a concentration series were then collapsed into a single value by taking the mean of Drug-GO scores for each GO group across the concentration series. All connections used to build the network represent Drug-GO interactions with a p-value of  $10^{-3}$  or lower.

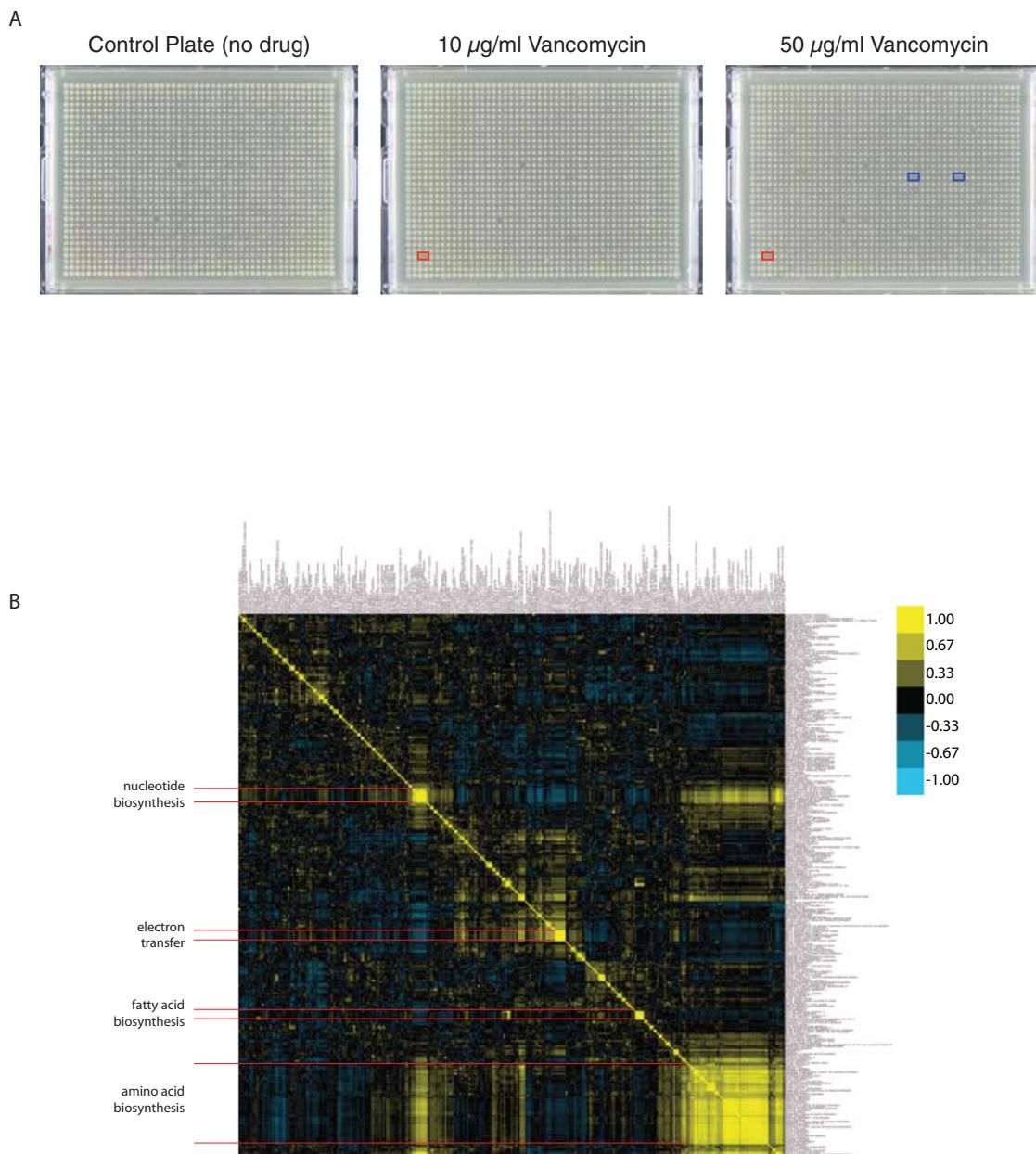
Drug-Drug correlations in the network represent Pearson correlation coefficients of 0.32 or higher, which corresponds to a threshold of +2 standard deviations in the distribution of all pair-wise Drug-Drug correlation coefficients. Here, condition-gene scores for drugs screened in a concentration series were averaged to generate a composite signature for each drug. These composite signatures were used for calculation of the correlation coefficients.

The network in Figure 7 was generated using Cytoscape version 2.6.3 (Shannon et al., 2003). Clustering was based on the Edge-weighted Spring-Embedded algorithm, using the P-values of Drug-GO interactions and high Drug-Drug correlations to drive the positions of nodes in the network.

### SUPPLEMENTAL REFERENCES

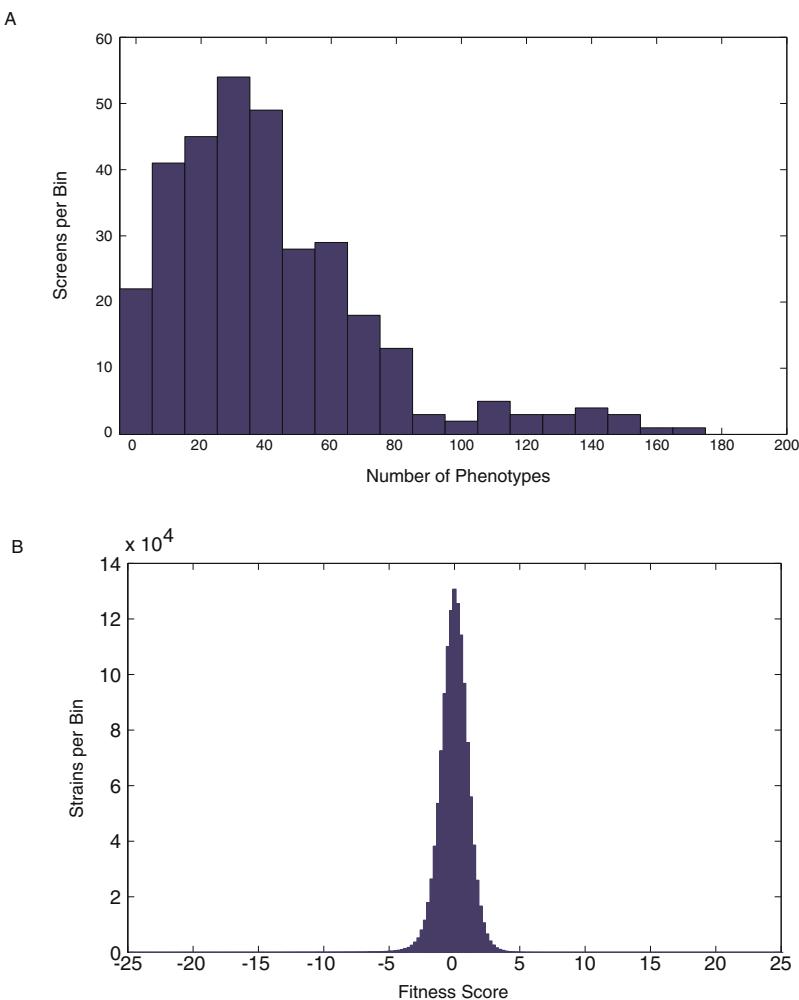
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 2006.0008.
- Butland, G., Peregrín-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., et al. (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* 433, 531–537.
- Collins, S.R., Schuldiner, M., Krogan, N.J., and Weissman, J.S. (2006). A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.* 7, R63.
- Cummings, C.A., Bootsma, H.J., Relman, D.A., and Miller, J.F. (2006). Species- and strain-specific control of a complex, flexible regulon by *Bordetella* BvgAS. *J. Bacteriol.* 188, 1775–1785.
- Efron, B., Tibshirani, R., Storey, J.D., and Tusher, V. (2001). Empirical Bayes Analysis of a Microarray Experiment. *J. Am. Stat. Assoc.* 96, 1151–1160.
- Girgis, H.S., Hottes, A.K., and Tavazoie, S. (2009). Genetic architecture of intrinsic antibiotic susceptibility. *PLoS ONE* 4, e5629.
- Hobbs, E.C., Astarita, J.L., and Storz, G. (2010). Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in *Escherichia coli*: analysis of a bar-coded mutant collection. *J. Bacteriol.* 192, 59–67.
- Hu, P., Janga, S.C., Babu, M., Díaz-Mejía, J.J., Butland, G., Yang, W., Pogoutse, O., Guo, X., Phanse, S., Wong, P., et al. (2009). Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biol.* 7, e96.
- Jeanguenin, L., Lara-Nunez, A., Pribat, A., Hamner Mageroy, M., Gregory, J.F., III, Rice, K.C., de Crecy-Lagard, V., and Hanson, A.D. (2010). Moonlighting glutamate formiminotransferases can functionally replace 5-formyltetrahydrofolate cycloligase. *J. Biol. Chem.*, in press. Published online October 15, 2010. 10.1074/jbc.M110.190504.

- Lanie, J.A., Ng, W.L., Kazmierczak, K.M., Andrzejewski, T.M., Davidsen, T.M., Wayne, K.J., Tettelin, H., Glass, J.I., and Winkler, M.E. (2007). Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. *J. Bacteriol.* 189, 38–51.
- McGinnes, K.E., Baker, T.A., and Sauer, R.T. (2006). Engineering controllable protein degradation. *Mol. Cell* 22, 701–707.
- Ramos-Montañez, S., Tsui, H.C., Wayne, K.J., Morris, J.L., Peters, L.E., Zhang, F., Kazmierczak, K.M., Sham, L.T., and Winkler, M.E. (2008). Polymorphism and regulation of the spxB (pyruvate oxidase) virulence factor gene by a CBS-HotDog domain protein (SpxR) in serotype 2 *Streptococcus pneumoniae*. *Mol. Microbiol.* 67, 729–746.
- Riley, M., Abe, T., Arnaud, M.B., Berlyn, M.K., Blattner, F.R., Chaudhuri, R.R., Glasner, J.D., Horiuchi, T., Keseler, I.M., Kosuge, T., et al. (2006). *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. *Nucleic Acids Res.* 34, 1–9.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4, 41.
- Typas, A., Nichols, R.J., Siegele, D.A., Shales, M., Collins, S.R., Lim, B., Braberg, H., Yamamoto, N., Takeuchi, R., Wanner, B.L., et al. (2008). High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nat. Methods* 5, 781–787.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3, RESEARCH0034.



**Figure S1. Colony Plates Illustrate High-Density Arrayed Screening, and Phenotypic Signatures Reproduce Known Connections between Biological Pathways, Related to Figure 1**

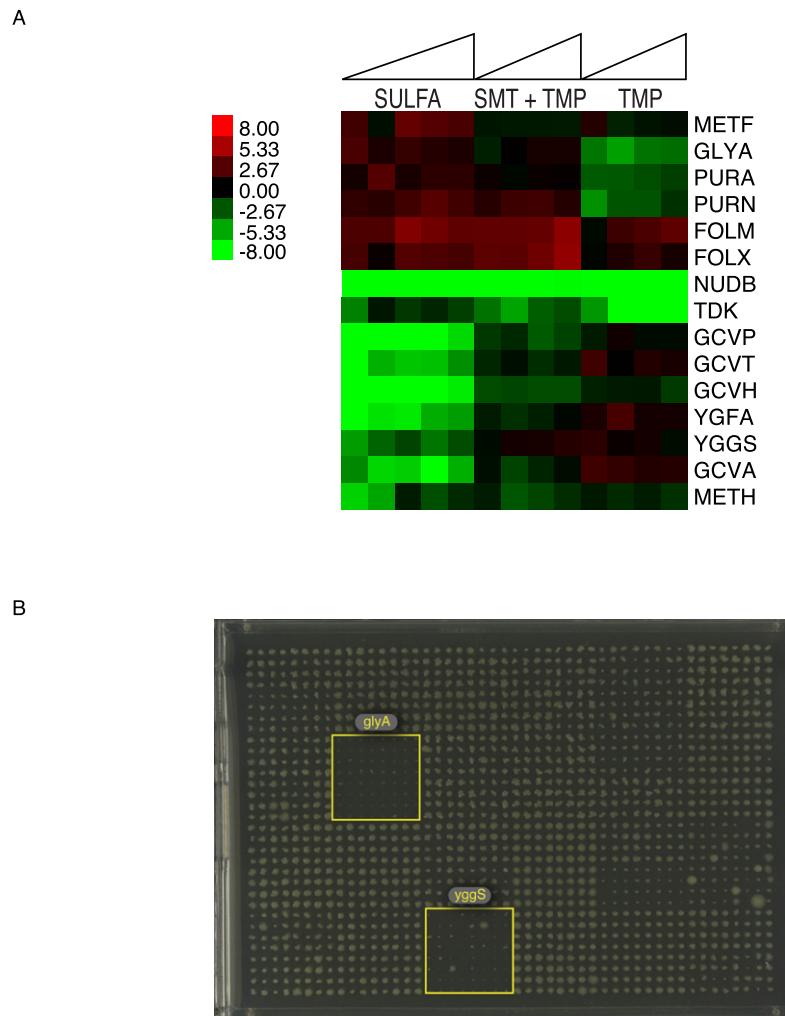
(A) Representative plates show the high-density array format used in our screens (1536 colonies/plate) and demonstrate the high-resolution condition-gene interactions (condition-gene interactions may be detected only under specific concentrations) identified by screening a range of drug concentrations (Vancomycin). (B) Phenotypic signatures reproduce known connections between biological pathways. Pathway signatures are generated by averaging the phenotypic signatures of all strains whose mutated gene is assigned to the given pathway in Ecocyc (<http://www.ecocyc.org>). Correlation coefficients are plotted for all pairs between 299 pathways. Clusters of known related pathways are highlighted. High off-axis correlation is observed between nucleotide and amino acid biosynthesis pathways, reflecting their shared essentiality under minimal media stresses screened in this study. Download figure for higher resolution.



**Figure S2. Phenotypes Reported Here Are Highly Specific, and the Majority Reflect a Conditional Loss of Fitness, Related to Figure 2**

(A) Histogram of the number of phenotypes identified at 5% FDR per screen. Bin size = 10 phenotypes. The number of phenotypes identified for any screen ranged from 0-4% of all strains tested.

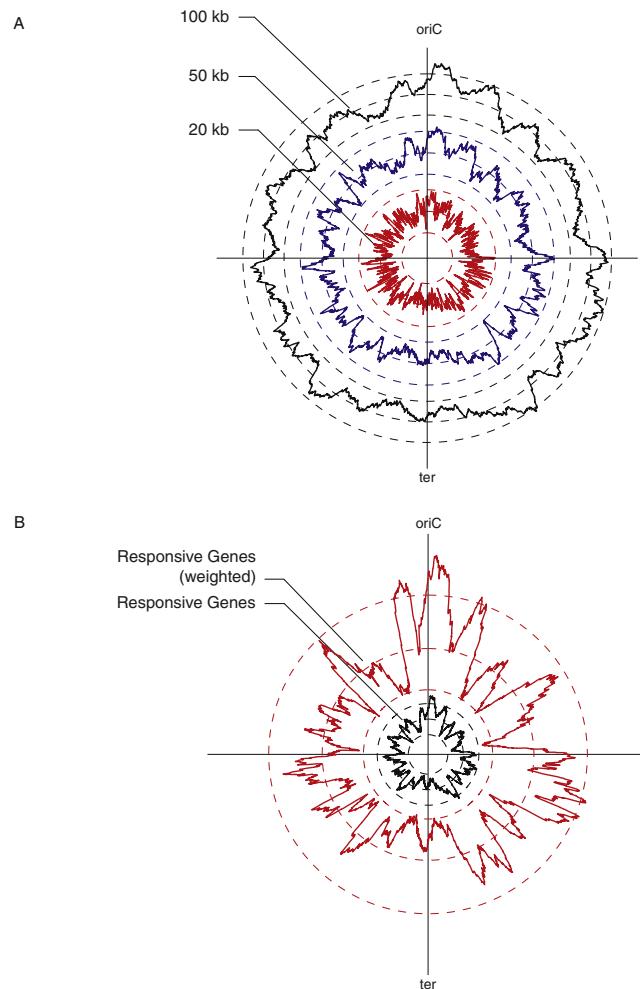
(B) Histogram of all fitness scores contained in the full clustergram in Fig. 1C (3979 strains  $\times$  324 conditions). The negative tail is larger than the positive tail, reflecting the  $\sim 4:1$  ratio of negative to positive phenotypes identified in this study.



**Figure S3. THF and 1C Mutant Responses Suggest Differential Network Effects of Sulfa and TMP, Related to Figure 5**

(A) Clustergram of select mutants associated with THF and one-carbon biosynthesis with multiple responses to sulfonamides (Sulfa), trimethoprim (TMP), or the combination. Profiles reveal that in addition to *gcv* and *glyA* mutants, knockouts of several other known and predicted network components exhibit differential responses to Sulfa or TMP. Mutants of *folM* and *folX* exhibit increased resistance to Sulfa, TMP (slightly), and the combination, in agreement with a previously published report (Girgis et al., 2009). Removal of *nudB*, which encodes the enzyme catalyzing the committed step in folate biosynthesis, renders the cell hypersensitive to all folate stresses tested, as expected. A deletion of *ygfA*, like the *gcv* mutants, is sensitive only to Sulfa stress. Interestingly, a recent study demonstrates that 5-formyl-THF accumulates in a *ygfA*<sup>-</sup> background only in conditions of excess glycine (Jeanguenin et al., 2010), conditions in which the GCV system is presumably active.

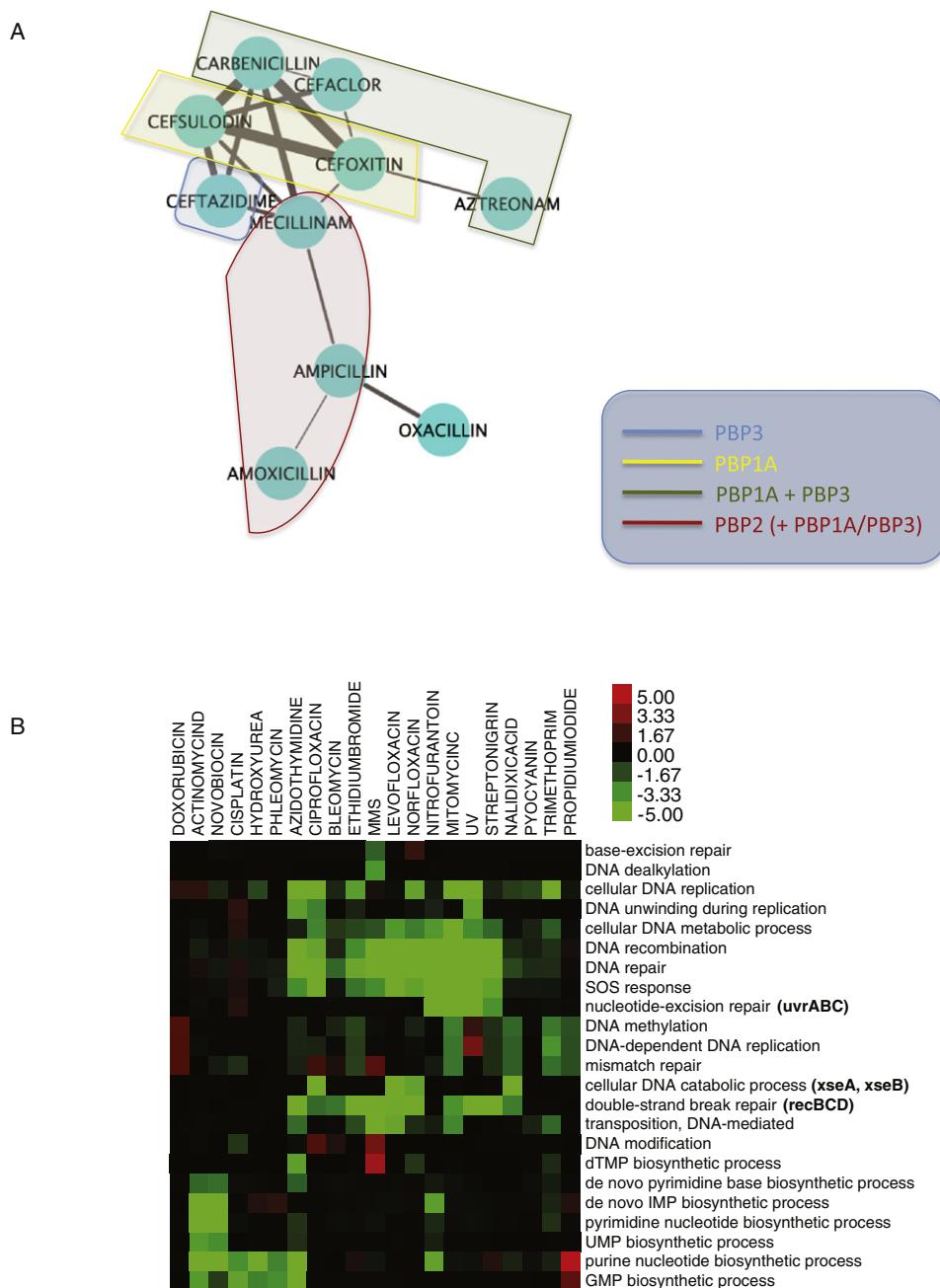
(B) *glyA*<sup>-</sup> and *yggS*<sup>-</sup> are a synthetic lethal pair. Image of a plate mating between the donor Hfr *yggS*::cat and 24 kan recipients (arrayed in boxes of 8x8 colonies), grown on kanamycin/chloramphenicol to select for double mutant strains; position of the *glyA*::kan and *yggS*::kan recipients is highlighted.



**Figure S4. 100 kb Window Size Is Optimal for Avoiding Operon Effects, and Weighted Responsive Genome Exhibits Highly Similar Spatial Enrichment to the Responsive Genome, Related to Figure 6**

Weighted Responsive Genome reflects Responsive Genome. Circular plots represent spatial enrichment for the variable plotted. Each trace (color) is accompanied by three dashed lines of the same color indicating the 5th percentile of all minimum negative enrichment values stored from 1000 permutations of the data, baseline representing zero enrichment, 95th percentile of all maximum positive enrichment values stored from 1000 permutations of the data (inside to outside).

(A) Three plots of the responsive genome illustrate the operon effects associated with smaller window sizes, especially the 20kb window (red trace).  
 (B) Black shows spatial enrichment for responsive genes, while red shows spatial enrichment for the weighted responsive genes. Here, responsive genes are weighted according to how many phenotypes were identified for each.



**Figure S5. Drug-Drug Correlations Overlap with Primary Target Specificity for B-Lactams, and Drug-GO Interactions Describe Cellular Responses to DNA-Damaging Agents, Related to Figure 7**

(A) Correlation of drug signatures roughly overlaps with PBP target affinity for the β-lactams. Connections between nodes represent positive correlation of drug signatures ( $r > 0.16$ , or 1 SD). Line width is based on the magnitude of positive correlation. Color shading corresponds to the highest affinity PBP target for each drug, based on  $IC_{50}$  data (W. Vollmer, personal communication).

(B) Drug-GO clustergram shows interactions between DNA-damaging stresses and DNA-related GO biological processes.