

Systems biology

A Gaussian process-based definition reveals new and bona fide genetic interactions compared to a multiplicative model in the Gram-negative *Escherichia coli*

Ashwani Kumar¹, Ali Hosseinnia², Alla Gagarinova³, Sadhna Phanse²,
Sunyoung Kim², Khaled A. Aly ², Sandra Zilles^{1,*} and Mohan Babu ^{2,*}

¹Department of Computer Science and ²Department of Biochemistry, University of Regina, Regina, SK S4S 0A2, Canada and
³Department of Biochemistry, University of Saskatchewan, Saskatoon, SK S7N 5E5, Canada

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on March 22, 2019; revised on July 24, 2019; editorial decision on August 20, 2019; accepted on August 23, 2019

Abstract

Motivation: A digenic genetic interaction (GI) is observed when mutations in two genes within the same organism yield a phenotype that is different from the expected, given each mutation's individual effects. While multiplicative scoring is widely applied to define GIs, revealing underlying gene functions, it remains unclear if it is the most suitable choice for scoring GIs in *Escherichia coli*. Here, we assess many different definitions, including the multiplicative model, for mapping functional links between genes and pathways in *E.coli*.

Results: Using our published *E.coli* GI datasets, we show computationally that a machine learning Gaussian process (GP)-based definition better identifies functional associations among genes than a multiplicative model, which we have experimentally confirmed on a set of gene pairs. Overall, the GP definition improves the detection of GIs, biological reasoning of epistatic connectivity, as well as the quality of GI maps in *E.coli*, and, potentially, other microbes.

Availability and implementation: The source code and parameters used to generate the machine learning models in WEKA software were provided in the Supplementary information.

Contact: sandra.zilles@uregina.ca or mohan.babu@uregina.ca

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Large-scale genetic (gene–gene, or epistatic) interaction (GI) maps in human cells (Bassik *et al.*, 2013; Boettcher *et al.*, 2018; Du *et al.*, 2017; Horlbeck *et al.*, 2018) and other organisms (Frost *et al.*, 2012; Horn *et al.*, 2011; Lehner *et al.*, 2006; Ryan *et al.*, 2012), especially the most comprehensive digenic (Costanzo *et al.*, 2010, 2016) and trigenic (Kuzmin *et al.*, 2018) GI maps of the budding yeast *Saccharomyces cerevisiae*, have revealed extensive insights into genetic redundancy, and identified cellular functionality of varied genes. GI maps in yeast have also revealed the functional wiring in response to DNA damage or autophagy stress (Bandyopadhyay *et al.*, 2010; Guenole *et al.*, 2013; Kramer *et al.*, 2017; Srivas *et al.*, 2013), metabolic adaptation (Szappanos *et al.*, 2011), cellular protein homeostasis (Rizzolo *et al.*, 2017) and the evolution of new traits (Taylor and Ehrenreich, 2014), as well as contributed to the discoveries of new gene functions and uncharacterized protein complexes or pathways (Collins *et al.*, 2007; Hoppins *et al.*, 2011; Jonikas *et al.*, 2009; Schuldiner *et al.*, 2005). Motivated by these

efforts in yeast, we (Butland *et al.*, 2008) and others (Typas *et al.*, 2008; van Opijnen *et al.*, 2009) have developed analogous high-throughput GI screening methods in the prokaryotic Gram-negative model organism, *Escherichia coli*.

Our eSGA (*E.coli* synthetic genetic array) approach allows for systematic creation and phenotypic scoring of double mutants by exploiting natural homologous recombination and bacterial conjugation to transfer a genetically marked ‘query’ deletion (or partial loss-of-function hypomorphic mutant allele) into a panel of *E.coli* essential hypomorphic or nonessential single gene deletion mutant ‘recipient’ strains. By employing this method, GIs pertaining to various biological processes (i.e. cell envelope, protein synthesis, genome integrity, nutrient stress), including antibiotic targets, were systematically mapped to reveal the epistatic dependencies and cross-talk (Babu *et al.*, 2014; Cote *et al.*, 2016), as well as the dynamic alteration of GI networks in response to environmental challenges or genotoxic insults (Babu *et al.*, 2011b; Gagarinova *et al.*, 2016; Kumar *et al.*, 2016).

The presence of an epistatic relationship is determined by comparing the growth fitness of double-mutant strains to their single-mutant counterparts, with colony size measurements reflecting strain fitness. In a multiplicative definition (a.k.a. model) used for *E. coli* and yeast GI mapping, for functionally unrelated genes, which represent the majority of possible gene–gene combinations, the product of the two single-mutant growth fitness measurements is expected to equal the fitness of the corresponding double mutant. Conversely, combining two functionally linked genes would be expected to produce more severe (negative or aggravating/synthetic sick or lethal effects, indicating parallel/redundant pathways) or less severe (positive or alleviating/buffering or suppression, suggesting a linear pathway) fitness defects than the predicted product of single-mutant fitness measurements. The strength and confidence of the GI is then determined using an interaction score for each *E. coli*-mutant gene pair based on the multiplicative rule originally developed for yeast (Baryshnikova *et al.*, 2010; Collins *et al.*, 2006).

In addition to the multiplicative model, other predefined functions for epistasis, such as minimum (double mutant has the same fitness as the less-fit single mutant), logarithmic (GIs measured on a logarithmic fitness scale) and linear additive (double mutant has the fitness equal the sum of two single-mutant fitness values), have been tested in yeast using the growth-rate measurements of single and double mutants (Mani *et al.*, 2008). This revealed that multiplicative and logarithmic functions are suitable for identifying functional relationships, specifically for growth-related phenotypes (i.e. colony fitness or growth rates), due to the nonlinear nature of cell growth (Mani *et al.*, 2008; Wang *et al.*, 2014). Conversely, additive model was shown to be appropriate for assessing phenotypes that change linearly (Wang *et al.*, 2014). Because these quantitative definitions diverge vividly and lead to different biological inferences when constructing GI networks (Mani *et al.*, 2008), the choice of a model should always depend on the type of phenotype and read-out (Walhout *et al.*, 2012). As with yeast studies (Costanzo *et al.*, 2010, 2016), large-scale *E. coli* genetic screens have typically used colony growth as a proxy for strain fitness measurement, and the multiplicative model for scoring GIs (Babu *et al.*, 2011, 2014; Butland *et al.*, 2008; Gagarinova *et al.*, 2016; Kumar *et al.*, 2016). While the choice of a model can have a determining effect on the interpretation of the results of a GI study, there has been no systematic evaluation of the GIs that would be differentially predicted in *E. coli* using other statistical models and machine learning algorithms.

Here, using our recently published large-scale GI map of *E. coli* genome integrity (Kumar *et al.*, 2016) as a case study, we compared multiple models based on their ability to identify functional relationships among gene pairs in *E. coli*. We show that a nonparametric machine learning-based Gaussian process (GP) can discriminate between genetically noninteracting and interacting *E. coli* gene pairs better than the published multiplicative and other machine learning and statistical models tested. We further experimentally verified this directly, testing predicted interactions for more than a dozen gene pairs. Using the GP model, we then re-evaluated existing GI data, identifying numerous interactions that have been missed by the multiplicative algorithm. We predict that the GP will improve the function prediction accuracy for the 16% (702 of 4432) of genes that still lack annotations, and thus can substantially increase the amount of gene functional information derived from the *E. coli* GI screens.

2 Materials and methods

2.1 Data collection and prediction of GIs

Colony size measurements (as proxies for strain fitness), normalized via the online SGAtool (Wagih *et al.*, 2013), were obtained for 102 255 viable digenic mutants and their corresponding 549 single gene mutants from our standard nutrient-rich condition *E. coli* GI network (Kumar *et al.*, 2016). In addition, we included 1767 gene pairs that were missed in the original genome integrity GI study (Kumar *et al.*, 2016), yielding quantitative values for a total of 104 022 gene pairs. For statistical reproducibility, each Hfr query

mutant in the genome integrity study was screened in quadruplicate against arrayed F[−]-recipient single-gene deletion mutants, generating four replicates per donor–recipient combination. The standard error (SE) measurements for the normalized colony sizes of the recipient single mutants (Wy; SE: 0 to 1.91×10^{-1}) and double mutants (Wxy; SE: 3.19×10^{-7} to 9.54×10^{-1}) were experimentally quantified (Kumar *et al.*, 2016), and fall within the SE range (Supplementary Table S1). Whereas the fitness measurement of each donor mutant (Wx) was represented as the median of normalized colony sizes for all double mutants arising from the corresponding donor's GI screen.

The normalized fitness measurements of the single (Wx: 0.91–1.20; Wy: 0.52–2.64) and double (Wxy: 0.00–1.96) mutants were then used to compute the expected double-mutant fitness according to predefined (i.e. additive, logarithmic, minimum, multiplicative) and machine learning (i.e. linear regression, LR; support vector regression, SVR; neural network, NN; GP) models, with normalized colony size values of 1.0 equaling wild-type (WT) fitness. To enable comparisons between the chosen predefined and machine learning models, we computed the GI score (ε) to define epistatic relationships among pairs of genes as the difference between expected and observed double-mutant fitness measurements, such that the score of zero would reflect a noninteraction, and most scores would be expected to center on zero. Conversely, aggravating and alleviating GIs would be defined as those whose double-mutant fitness significantly different from the fitness expected based on the two corresponding single-mutant fitness values.

2.2 Predefined GI models

The four predefined formulas employed to delineate GIs are detailed elsewhere (Mani *et al.*, 2008). Briefly, first, in the multiplicative definition GI score $Wxy - (Wx * Wy)$, the product of the two single-mutant (Wx, Wy) growth fitness values is deducted from the observed double-mutant (Wxy) fitness. Second, in the additive definition score $Wxy - (Wx + Wy - 1)$, the sum of single-mutant growth fitness values (minus one) is subtracted from the corresponding observed double-mutant fitness value. Third, in minimum definition GI score $Wxy - \text{Min}(Wx, Wy)$, the growth fitness value from the slowest-growing single mutant (Min [Wx, Wy]) is subtracted from the respective double-mutant fitness value. Fourth, the logarithmic definition GI scores $Wxy - \log_2((2Wx - 1)(2Wy - 1) + 1)$ reflect the fitness scale in logarithmic function.

2.3 Defining GIs by machine learning models

GIs measured using four different machine learning algorithms, along with chosen parameters under each definition, built within the WEKA (Smith and Frank, 2016) machine learning software (ver 3.9.3), are summarized, briefly, as follows:

LR: The training data are modeled in a way such that the predicted (dependent) variable is a linear function of the input (independent) variables. Here, the independent variables were Wx and Wy (signifying two single mutants' growth fitness), and the dependent variable was Wxy (double-mutant growth fitness). We used standard WEKA implementation of LR.

SVR: Kernel function is used to map the input data into a higher-dimensional space, in which a linear hyperplane is trained and then used as a model for prediction. Again, the input variables were Wx and Wy, and the output variable was Wxy. We used RegSMOImproved algorithm with the polynomial kernel function under standard parameter settings as defined in WEKA.

NN: Input values are linearly combined and fed through layers of so-called neurons, which pass on their output values (computed via a linear activation function) to the next layer, eventually leading to the output layer which produces the final prediction for the given inputs. In numerous iterations, internal weights are adjusted so that the output values approximate the ones given in the training data. We used the WEKA implementation for training a feed-forward NN (multilayer perceptron) with one hidden

layer with one unit, using the sigmoid activation function. For all other parameters, we used the standard setting in WEKA. As always, the input variables were W_x and W_y , and the output variable was W_{xy} .

GP: This definition that relies on Bayesian principles initializes a posterior probability distribution on the training dataset, and can be best explained by comparing simple linear regression to Bayesian linear regression. In the former, a linear relationship is assumed between the dependent (W_{xy}) and independent (W_x , W_y) variables, where the intercept and slope were calculated using the parameters θ_0 and θ_1 , and an error term (ϵ) is added.

$$Y = f(x) + \epsilon, \text{ where } f(x) = \theta_0 + \theta_1 x$$

In the latter, a probabilistic method is implemented where a distribution over the parameters θ_0 and θ_1 is initialized, and then restructured based on the observed data points (x , $f(x)$; here x is a pair of single-mutant values, and $f(x)$ the corresponding double-mutant value). While GP generalizes the Bayesian approach to a nonparametric method, it does not assume a certain shape of target function (such as linear function or polynomial of degree k) in advance, and thus does not maintain a probability distribution over parameter vectors as Bayesian linear regression does. Instead it maintains a distribution over potential functions $f(x)$ that are consistent with the observed data points (x , $f(x)$). Typically, one defines a probability distribution over functions $f(x)$, and a domain for the input variable x (i.e. the domain would be the set of possible pairs of single-mutant fitness values). Here, the functions $f(x)$ that are smooth over the chosen domain for x , with a mean, for example (1, 1), used for sampling. The smoothness is obtained from a covariance matrix, corresponding to a kernel function. The covariance matrix defines the shape of a multivariate Gaussian distribution. In case of a bivariate Gaussian distribution, if there are two normally distributed but independent variables x_1 and x_2 , the covariance between x_1 and x_2 is 0. However, if we assume 1 as the variance for each independent variable, the covariance matrix, in this case, will be:

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The joint probability of variables x_1 and x_2 [Equation (1)] that is used to infer the conditional probability of one variable to the other will allow GP to obtain a posterior distribution, which can be succinctly expressed as the joint probability of observed output values (i.e. output values of f) and unobserved output values (i.e. output values of the underlying target function f_*).

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right) \quad (1)$$

$$\begin{pmatrix} f \\ f_* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \mu_* \end{pmatrix}, \begin{pmatrix} K & K_* \\ K_*^T & K_{**} \end{pmatrix}\right) \quad (2)$$

Here K = covariance matrix expressing similarity between observed (training) x values calculated by a kernel function; K_* = covariance matrix expressing similarity between the training and test values for x ; K_{**} = covariance matrix expressing similarity between the test values for x ; and K_*^T = transpose of K_* . Finally, for training dataset x we have observed the corresponding outcome $f(x)$ (i.e. denoted as f), whereas for test dataset x_* we predict an outcome $f(x_*)$ (i.e. denoted as f_*). This is best represented by computing a probability $p(f_*|x_*, x, f)$, assuming that f and f_* jointly have a Gaussian distribution. Equation (1) means that (x_1, x_2) is normally distributed (\mathcal{N}) with mean $(\mu_1, \mu_2; \mu, \mu_*)$, and covariance matrix $(\sigma_{11}, \sigma_{12}, \sigma_{21}, \sigma_{22})$. To summarize, a GP is an infinite family of random variables with the property that every finite subset forms a vector of random variables that follows a Gaussian distribution. In machine learning with GPs, a kernel function is used for estimating the similarity between input data points so that similar input leads to comparable predictions. We used the WEKA implementation

with the radial basis function kernel. The only parameter of that kernel, gamma, was set to the default value 1.0 for the data from Kumar et al. (2016) or Costanzo et al. (2010), and to 3.0 for the data from Babu et al. (2014).

2.4 Training and test datasets

Unlike predefined models, supervised machine learning algorithms require training and test datasets. Since the normalized colony sizes of single and double mutants were subjected to predefined and machine learning models to predict epistatic relationships for evaluating the difference between these models, we used a fold change (FC) metric, as an alternate measure, to define our training and test datasets. As in any large-scale GI study, interacting gene pairs are defined using a cut-off assignment (e.g. $P \leq 0.05$; 5% significance), and gene pairs that pass the chosen statistical threshold suggest aggravating or alleviating interactions, while those that fall below this threshold are deemed as noninteracting pairs. Conversely, GI scores calculated for digenic mutant combinations that showed a normal distribution centered on zero (i.e. neutral or noninteracting) suggests that fitness of most double mutants (i.e. with combinations of functionally unrelated mutations) should be close to the fitness of individual single-mutant growth. However, in the case of our FC selection criteria (see below), the noninteracting gene pairs essentially range from FC of 0.90 to 1.11. Whereas with respect to interacting test pairs, aggravating interactions had FC of 0.00–0.49, and alleviating interactions had FC of 1.21–2.66. The FC was computed for gene pairs with double-mutant fitness (W_{xy}) that was similar to one or both of the two single mutants ($W_a = W_x$ or $W_a = W_y$).

$$FC(W_{xy}, W_a) = \frac{W_{xy}}{W_a} \text{ if } \frac{W_{xy}}{W_a} \leq 1, \text{ and}$$

$$FC(W_{xy}, W_a) = \frac{W_a}{W_{xy}} \text{ otherwise}$$

Our selection of training set (Supplementary Table S2) includes drawing random gene pairs using the cut-off indicated below (as long as sufficient number of pairs existed for training), where the normalized fitness of W_{xy} double mutants compared to W_x or W_y single mutants should have a $FC \geq 0.9$.

$$FC(W_{xy}, W_a) \geq 0.9$$

Otherwise, W_{xy} double mutants that are similar to W_x and W_y single mutants should have a $FC \geq 0.8$.

$$FC(W_{xy}, W_x) \geq 0.8, \text{ and } FC(W_{xy}, W_y) \geq 0.8$$

This led us to construct 12 bins with different normalized fitness values, with 20 distinct gene pairs chosen within each bin. The resulting 336 training gene pairs were subjected to 5-fold cross-validation by distributing them randomly into five sets. We then assessed the performance of each model using the remaining gene pairs in the other four sets. This process was repeated such that the model validation is iteratively performed, and the expected double-mutant fitness values from these iterations were then combined to generate the final model.

In addition to the training set, we generated a test dataset for both noninteracting and interacting gene pairs (Supplementary Table S2). To select noninteracting pairs, we set two criteria. First, we considered gene pairs whose double-mutant growth fitness were comparable to their respective single mutants.

$$FC(W_{xy}, W_x) \geq 0.9, \text{ and } FC(W_{xy}, W_y) \geq 0.9$$

Second, we chose gene pairs with double-mutant fitness close to the fitness of the most fit (i.e. fitness ≈ 1.0) parental single mutant.

$$FC(W_{xy}, \min(W_x, W_y)) \geq \{0.8 \text{ if } |W_x - W_y| \geq 0.4\}$$

Similarly, we set two analogous parameters for selecting interacting gene pairs to serve as a test set. In the first case, gene pairs with

double-mutant growth fitness distinct from their two single mutants were selected.

$$FC(W_{xy}, W_x) \leq 0.7, \text{ and } FC(W_{xy}, W_y) \leq 0.7$$

In the second case, we considered gene pairs with double-mutant growth fitness that were close to the more substantial constitutive single mutant's deletion fitness.

$$FC(W_{xy}, \max(W_x, W_y) \geq \{0.8 \text{ if } |W_x - W_y| \geq 0.4\}$$

These aforesaid factors yielded 400 noninteracting and 400 interacting test gene pairs, comparable in size to the training set, for proper evaluation of the performance of the selected models. Notably, we ensured that our noninteracting test gene pairs did not have any overlap with the training dataset.

2.5 Statistical measures to evaluate model performance

We adopted three widely used statistical metrics (mean absolute error, MAE; root mean square error, RMSE; coefficient of variation, CV) to measure the performance of the models. Using the MAE, the average magnitudes of deviation of double-mutant fitness from the expected neutral phenotype (i.e. error) in the training and test sets were measured. Specifically, MAE quantifies the average absolute differences between observed and expected double-mutant fitness values. Here, error (e) = deviation of the observed – expected double-mutant fitness for n samples ($e_k, k = 1, 2, 3, \dots, n$), e_k is the error of the k th gene pair, and n is the total number of gene pairs.

$$MAE = \frac{1}{n} \sum_{k=1}^n |e_k|$$

Because all errors in the MAE have equal weight, MAE is typically less sensitive to large prediction errors. Therefore, we applied RMSE, which is more sensitive in predicting large errors. Briefly, the average magnitude of prediction errors was measured on an entire training and test datasets using the square root of the average of squared differences between observed and expected double-mutant fitness values.

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n e_k^2}$$

For both MAE and RMSE, the average prediction error ranges from zero to infinity. As a measure of variance, for each model we applied $CV = \frac{\sigma}{\mu}$, where a standard deviation (SD) (σ) relative to the mean (μ) error (e) of all gene pairs was expressed as a percentage.

2.6 Decision tree-based classification

Using the C4.5 decision tree algorithm provided within the WEKA software, 800 test gene pairs were partitioned into interacting (aggravating, alleviating) or noninteracting and were incorporated independently into the decision tree classifier with training labels based on multiplicative or GP model. A 5-fold cross-validation was applied to these test pairs for each definition, whereby four-fifths of the data were used for training, and the other one-fifth was held out for testing. Briefly, each of the aggravating, alleviating and noninteracting pairs were assigned specific labels: -ve, +ve, and O, respectively. The resulting 3×3 confusion matrix was used to compute the percentage of interacting or noninteracting pairs that was classified correctly.

2.7 Examining GIs for shared functional links using gene set enrichment analysis

Functional enrichment analysis was performed as described previously (Subramanian *et al.*, 2005) using a gene set enrichment analysis (GSEA) (ver. 3.0) tool. Briefly, a pre-ranked gene list of normalized fitness values corresponding to the pairs defined as aggravating (1614), alleviating (2459) and noninteracting (4073) under the GP definition were investigated to determine if the GI pairs share similar Gene Ontology (GO) functions. All GO

annotation gene sets for *E. coli* were downloaded from EcoCyc, and GO terms with a minimum of 2 to a maximum of 500 annotated genes were considered for GSEA. Genes that were enriched for a GO term with a $P \leq 0.05$ and with a false discovery rate (FDR) of 25% or less were used to assess the functional relationship, as long as an enriched GO term function is held by both genes in a pair.

2.8 Bacterial strains and media

Strains used for experimental validation include the WT *E. coli* K-12 BW25113; kanamycin-marked F^- -recipient single-gene deletion mutant strains from the Keio knockout library (Baba *et al.*, 2006), including the functionally unrelated gene JW5028 (Gagarinova *et al.*, 2012); F^- single-mutant strains with hypomorphic alleles of essential genes generated by integrating a C-terminal sequential peptide affinity fusion tag and a kanamycin selection cassette into the 3'-UTR of the respective essential genes to perturb transcript abundance as described previously (Babu *et al.*, 2014; Butland *et al.*, 2008); and chloramphenicol-marked high-frequency recombination Cavalli (HfrC) donor single-mutant strains created using λ -Red recombination as reported earlier (Butland *et al.*, 2008). Strains were grown on Luria-Bertani (LB) medium at 32°C for 24 h. In addition, as indicated, kanamycin (F^- recipients and double mutants; 50 μ g/ml), chloramphenicol (HfrC donors and double mutants; 34 μ g/ml), bleomycin (BLM; 0.25 μ M) and methyl methanesulfonate (MMS; 0.05%) were added to the media.

2.9 Validation of predicted GI pairs

Double mutants were created by conjugating a chloramphenicol-marked HfrC donor gene deletion mutant with the kanamycin-marked F^- -recipient mutant strains, following the eSGA procedure (Butland *et al.*, 2008). The double mutants and the corresponding single mutants were serially diluted in LB medium and pinned onto LB plates with and without DNA-damaging agents (BLM, MMS) to assess the sensitivity of the strains to DNA damage. Phase-contrast cell morphology micrographs from the cultures of single or double mutants incubated for 2 h in the presence of DNA-damaging agents were captured using a Carl Zeiss LSM700 confocal microscope; and cell lengths were measured using ImageJ.

3 Results and discussion

3.1 GP identified more bona fide GIs than other models

To evaluate the performance of the selected models, we employed RMSE and MAE metrics on the training (336) and test (400) gene pairs based on predicted GI scores. Our hypothesis is that for any noninteracting gene pairs, the observed double-mutant fitness values that are not different from expectation should have a low RMSE and MAE values. Consistent with this notion, just as the noninteracting test gene pairs (Fig. 1A), GP showed smallest RMSE and MAE compared to other definitions (see Supplementary Table S2 for mean and SD from a 5-fold cross-validation of the training set shown for machine learning models in Fig. 1A). As well, in contrast to predefined models, low CV exhibited by GP was comparable to other machine learning models. We also attempted to modify the parameter settings (i.e. gamma and noise parameters) used in GP model to evaluate its performance by measuring the error metrics (RMSE, MAE) on the training dataset (Supplementary Table S2). Parameter sampling was performed manually by testing various gamma (or noise) values that are above or below the default setting. We found no substantial effect on the performance of the trained predictors even when parameter settings were modified. Similarly, since NN model on the training and test datasets exhibited error metrics (Fig. 1A) close to GP, we posited that fine-tuning the parameter settings for NN may achieve better performance than GP. However, even under different parameter settings, we were unable to achieve better error metrics for NN over GP (Supplementary Table S2).

Gene pairs with related biological functions connected by GIs are rare (Mani *et al.*, 2008). We therefore should expect most of the

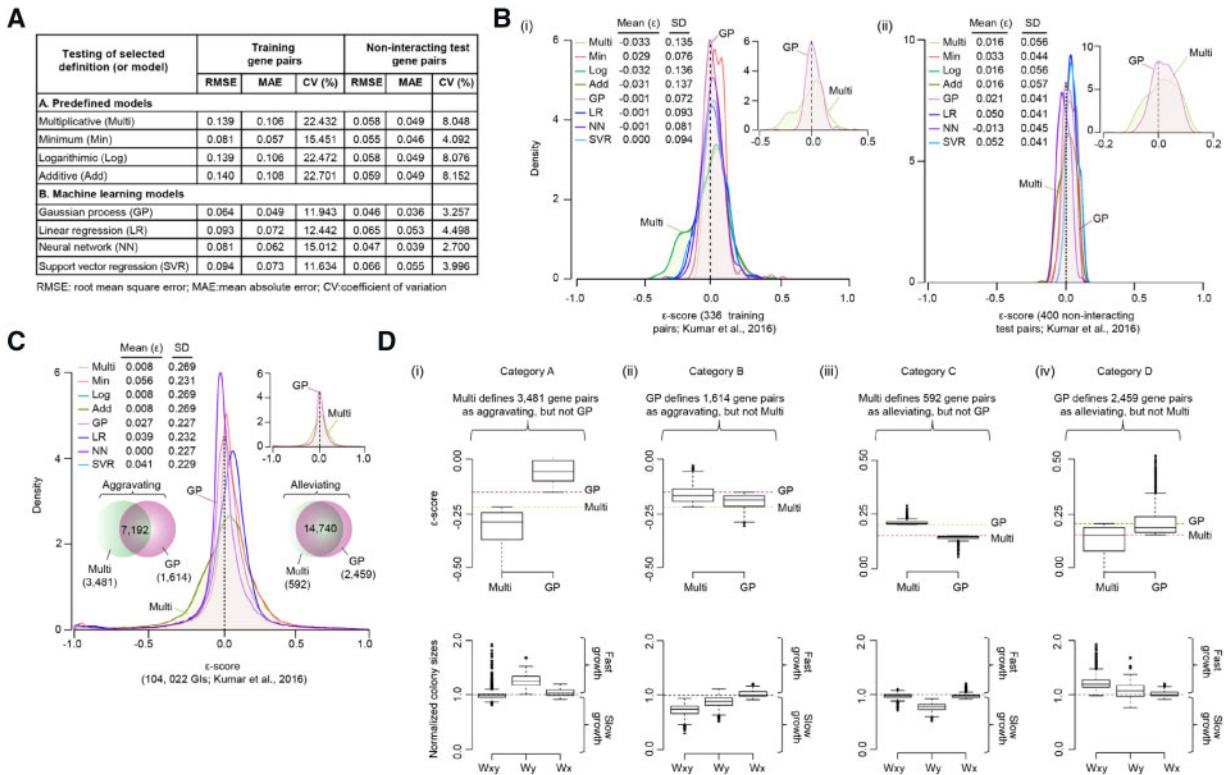


Fig. 1. Performance measurements of GP over other GI definitions. (A) Error values shown for the indicated definitions using training and test pairs compiled from the genome integrity study (Kumar et al., 2016). (B, C) GI (ϵ) score distributions of training (B-i) and test (B-ii) pairs and for all (C) gene pairs plotted against various definitions with normal distribution centered around zero (dotted line), and with tails signifying aggravating (-ve) and alleviating (+ve) interactions. The zoom in of the tails is shown as an inset in the top-right corner. The overlap of aggravating or alleviating GIs between GP and the multiplicative definition is shown as a Venn diagram (C) in the inset. (D) Box plots showing the distribution of GI (ϵ) scores for gene pairs classified differently by the GP and the multiplicative definitions (grouped under various categories, see text), and their respective normalized single (Wx: donor; Wy: recipient) and double (Wxy) mutant colony sizes that exhibit slow (< 1.0), normal (1.0) and fast (> 1.0) growth

gene pairs to be noninteracting. As a result, observed double-mutant fitness values will yield a distribution that approximates the expected distribution. Quantitative measurement of GIs for those noninteracting gene pairs would then have a tight normal distribution centered around zero (i.e. low dispersion and low bias), indicating that GIs are restricted only to certain pairs. Here, the left and right tails of the distribution represent gene pairs with negative (aggravating) and positive (alleviating) GIs, respectively. By applying this strategy to the training or test double-mutant strains, and to the 104 022 digenic mutants of the genome integrity network (Kumar et al., 2016), we examined the $|\epsilon$ -score distribution to discern which definition produced values close to zero. As with the training or test dataset (Fig. 1B), we found the distribution of GP for all digenic mutant pairs approximated a normal distribution that is tight and close to zero (i.e. mean = 0.02, SD = 0.22; with lowest bias and dispersion for ϵ scores; Fig. 1C), conforming to the expectation that most gene pairs are noninteracting. Alternately, other models showed a positive or negative shift in the distribution of $|\epsilon$ -score with a greater number of scores deviating from zero (i.e. with bigger SDs; Fig. 1B and C).

3.2 GP identifies new GIs missed by the multiplicative model

We next investigated if the performance of GP in terms of negligible deviation of observed double-mutant fitness from expectation and tight $|\epsilon$ -score distribution for genes involved in genome integrity was consistent when applied to a different *E. coli* GI dataset not focused on any particular function. To do so, we obtained the growth fitness measurements of 600 000 digenic mutant strains and their respective single mutants from our genome-wide GI study (Babu et al., 2014) in which we crossed 163 functionally unrelated 'donor' mutant alleles from diverse core bacterial processes against 3968 nonessential single

gene deletions and 149 hypomorphic mutant strains. Using this growth fitness evidence, 504 gene pairs were selected for training set based on the same selection criteria as described above (see Section 2). Here, GP once again exhibited insignificant deviation of double-mutant fitness from expectation as determined by 5-fold cross-validation (Supplementary Fig. S1A), as well as least bias and dispersion (i.e. normal distribution centered on neutrality) compared to other models (Supplementary Fig. S1B). Collectively, these results suggest that, as in yeast (Mani et al., 2008), the choice of model is vital since it can markedly impact our discovery of *E. coli* GIs, especially when 5% (235 076 of 4 889 920) of total digenic mutants examined to date are solely supported by large-scale genetic screens relying on the multiplicative function (Babu et al., 2011, 2014; Gagarinova et al., 2016; Kumar et al., 2016).

To further analyze the performance of the GP model in identifying GIs, we compared how well the gene pairs among the top 25% ranked (26 005 of 104 022 double mutants) by $|\epsilon$ -score thresholds for GP ($\epsilon \leq -0.152$; $\epsilon \geq 0.153$) and the multiplicative ($\epsilon \leq -0.221$; $\epsilon \geq 0.203$) model were in agreement with the identification of their putative GIs. Here, in order to compare ϵ_{GP} and ϵ_{Multi} , we defined gene pairs with $\epsilon_{GP} \leq -0.152$ or $\epsilon_{Multi} \leq -0.221$ as aggravating, $\epsilon_{GP} \geq 0.153$ or $\epsilon_{Multi} \geq 0.203$ as alleviating and those that do not pass the threshold as noninteracting. While the interaction type for most (84%, 21 932) of the digenic mutant gene pairs (14 740 alleviating, 7192 aggravating; Supplementary Table S1) displayed by GP and multiplicative definitions were similar, for one-sixth (16%, 4073) of the gene pairs they predicted inconsistently (Fig. 1C inset). The latter observation led us to investigate which model might have misclassified the interacting gene pairs.

A close examination of 3481 gene pairs (category A) that were defined as having aggravating interactions ($-2.064 \leq \epsilon_{Multi} \leq -0.221$) by the multiplicative, but not by the GP definition ($-0.152 \leq \epsilon_{GP} \leq 1.00$) showed no obvious growth fitness defects in the double mutants

relative to their respective single mutants (Fig. 1D-i), indicating that GP accurately followed the expectation that most gene pairs were not aggravating. For example, in contrast to the multiplicative model ($\epsilon_{\text{Multi}} = -0.224$), GP classified the association between *recA* and *ssb* as close to neutral ($\epsilon_{\text{GP}} = 0$). This observation is consistent with previous findings (Cox and Lehman, 1982; Eggler *et al.*, 2003; Kowalczykowski *et al.*, 1987), suggesting that both ssDNA binding proteins, Ssb and RecA, compete for ssDNA substrates *in vivo*; hence *ssb* suppresses the nucleation stage of *recA* filament formation on ssDNA and subsequent homologous recombination. Likewise, 1614 gene pairs (category B) were defined as aggravating by GP ($-0.308 \leq \epsilon_{\text{GP}} \leq -0.152$) but not by the multiplicative model ($-0.220 \leq \epsilon_{\text{Multi}} \leq -0.029$). While gene pairs in this category were assigned negative ϵ scores by both models, the strength and confidence of putative GIs are different. If a gene pair has a strong GI, then it is expected to be within the top-ranked list of gene pairs. In that sense, GIs predicted by the multiplicative model inconsistently classified the pairs as noninteracting (i.e. fell below the chosen cut-off; $\epsilon_{\text{Multi}} \leq -0.221$) despite many double mutants exhibited weak growth fitness compared to their single mutants (Fig. 1D-ii). For example, GP has successfully captured the aggravating interaction ($\epsilon_{\text{GP}} = -0.159$) previously reported for *recB priA* double mutants (Mccool and Sandler, 2001), whereas this interaction ($\epsilon_{\text{Multi}} = -0.108$) was precluded under the multiplicative definition as it was below the selected $|\epsilon\text{-score}|$ cut-off.

Additionally, among the 592 gene pairs (category C) predicted by the multiplicative model as alleviating ($0.203 \leq \epsilon_{\text{Multi}} \leq 0.291$), but were classified as noninteracting by the GP definition ($0.053 \leq \epsilon_{\text{GP}} \leq 0.151$) showed that when double mutants and their respective single mutants had moderate growth fitness defects, the GP model closely matches the expectation of these gene pairs being noninteracting (Fig. 1D-iii). Similarly, based on the growth fitness of double mutants that show better growth than their corresponding single mutants, GP has classified 2459 gene pairs (category D) as alleviating ($0.152 \leq \epsilon_{\text{GP}} \leq 1.00$) as expected; whereas the multiplicative model showed a greater tendency to place these gene pairs as aggravating or close to neutrality ($-1.941 \leq \epsilon_{\text{Multi}} \leq 0.203$; Fig. 1D-iv). An illustrative example is when compared to the multiplicative model ($\epsilon_{\text{Multi}} = 0$), GP ($\epsilon_{\text{GP}} = 0.215$) defines the interaction between the components of cellulose synthase complex, BcsF and BscQ, as alleviating, consistent with the notion that such interactions tend to occur between the subunits of a protein complex (Babu *et al.*, 2014; Bandyopadhyay *et al.*, 2008; Collins *et al.*, 2007; Costanzo *et al.*, 2019).

Although the above results suggested that the classification of gene pairs by GP to appropriate interaction class was consistent with colony growth measurements, we investigated further if the GI pairs (4073; sum of categories A and D) favored by the multiplicative model, that when not in agreement with GP, can be clearly demonstrated by their change in interaction direction, irrespective of their chosen threshold. We found that for less than a quarter (19%; 776 of 4073) of the pairs defined by the multiplicative model as aggravating, GP clearly classified these pairs either as alleviating (103 pairs) or noninteracting (673 pairs; Supplementary Table S1). The interaction trend (i.e. positive or negative) for the remaining three-fourths (81%; 3297 of 4073) of the pairs was reliable under two definitions, but these pairs had neutral (not significant) interactions according to GP, consistent with the observed neutral colony sizes (Fig. 1D) and with the expectation that most pairs would not interact.

Next, our rationale for using top 25% (26 005) of the ranked GI pairs by $|\epsilon\text{-score}|$ thresholds for GP and multiplicative model was to obtain large sets of gene pairs to make reasonable biological interpretations about model performance. While this method of utilizing top-ranked gene pairs is widely adopted for evaluating the performance of GI predictions (Ulitsky *et al.*, 2009), and in cross-study validation using multiple transcriptomic datasets to generate a robust prediction model (Kim *et al.*, 2016), we applied statistical thresholds corresponding to two SD ($|Z\text{-score}| \geq 1.96$; $P \leq 0.05$; Supplementary Fig. S1C) to the 104 022 digenic mutants of the genome integrity network to examine the predictive capability of GP and multiplicative model in identifying GIs. Of the 6980 statistically

significant ($P \leq 0.05$) GI pairs, 4502 had GI (ϵ) scores of -1.96 or lower indicating aggravating relationships, and 2478 with ϵ -scores of $+1.96$ or higher representing alleviating interactions (Supplementary Fig. S1C). Three-fourths (72%; 4998 of 6980) of the digenic mutant pairs were consistently captured by both GP and multiplicative models. Whereas for the rest of the gene pairs, GP approach has reliably classified 542 pairs as aggravating (category II) and 816 as alleviating (category IV), indicating that 601 pairs (category I) that have initially thought of as aggravating and 23 (category III) as alleviating by the multiplicative model were indeed inconsistent, based on the fitness of the double mutants relative to single mutants (Supplementary Fig. S1D; Supplementary Table S3). This trend persisted even when choosing different P -value-based cut-offs (data not shown), where most (95%; 6604) of the significant Z-transformed GI pairs in the 95th percentile ($P \leq 0.05$) were found in nearly all (99%; 25 991 of 26 005) of the top 25% of the ranked gene pairs (Supplementary Fig. S1E). Overall, these results suggest that the conclusions of our findings remain consistent, whether the selection of GI pairs was based on top-ranking or P -value based cut-off.

Since probabilistic decision tree-based classifiers can reveal the characteristics of a model in terms of its GI predictive accuracy for gene pairs (Wong *et al.*, 2004), we conducted an independent evaluation of the definitions based on training the decision tree using the 800-pair test dataset (classifying 200 as aggravating, 200 as alleviating and 400 as noninteracting). This trained classifier was then used to examine if GP and the multiplicative definition categorize gene pairs into appropriate interaction types, as expected. Based on the decision-tree classification score threshold of $\epsilon_{\text{GP}} \leq -0.440$ and $\epsilon_{\text{GP}} \geq 0.110$ (Supplementary Fig. S2A), GP achieved an accuracy (consistently classifying the GI type) of 99.63% (797 of 800 pairs) at an error (inconsistently classifying the GI type) rate of 0.375% (3 of 800 pairs) as determined by 5-fold cross-validation. Conversely, at the classification score threshold of $\epsilon_{\text{Multi}} \leq -0.385$ and $\epsilon_{\text{Multi}} > 0.126$ (Supplementary Fig. S2A), the accuracy of the multiplicative definition was 96.9% (775 of 800) at an error rate of 3.125% (25 of 800; Fig. 2A). Further investigation on these 25 gene pairs (Supplementary Table S4) revealed that the multiplicative model misclassified 21 alleviating interactions as noninteracting, 2 noninteracting pairs as alleviating and 2 alleviating pairs as aggravating or noninteracting. In fact, some of the alleviating interactions among the 21 gene pairs are well supported by literature evidence. For example, the multiplicative definition identified the association between helicase II (*uvrD*) and homologous recombination (*recA*, *recD*) mutants as noninteracting, whereas GP classified these pairs as being alleviating, in agreement with the observation that UvrD and RecA act in the same pathway and function in concert with RecBCD-dependent recombination intermediates (Bidnenko *et al.*, 2006). A similar high accuracy and negligible deviation of observed double-mutant fitness from expectation for the GP method was obtained in comparison to the multiplicative model, when the decision-tree classifier was used on 104 022 gene pairs (Supplementary Table S4). As with the decision tree approach (Fig. 2A), we also achieved a greater area under the receiver operating characteristic curve for the interacting (0.996) and noninteracting (0.998) test pairs predicted by GP model (versus 0.975 and 0.996 for the multiplicative model, respectively; Supplementary Table S4). While these differences appear modest, they correspond to hundreds of gene pairs that are differently classified and will have substantial impact on our biological interpretation of the identified pathways.

3.3 Validation of GI pairs identified by GP model

As error metrics, ϵ distribution, and cross-validation confirmed GP to be a better predictor than the multiplicative model for identifying GIs and functional relationships in *E. coli*, we next sought to systematically examine whether the aggravating (1614; category B) or alleviating (2459; category D) GIs identified under the GP model conform to the previous notion (Baryshnikova *et al.*, 2010) regarding a common biological function. To do so, we used GSEA (Subramanian *et al.*, 2005) to evaluate enriched functional

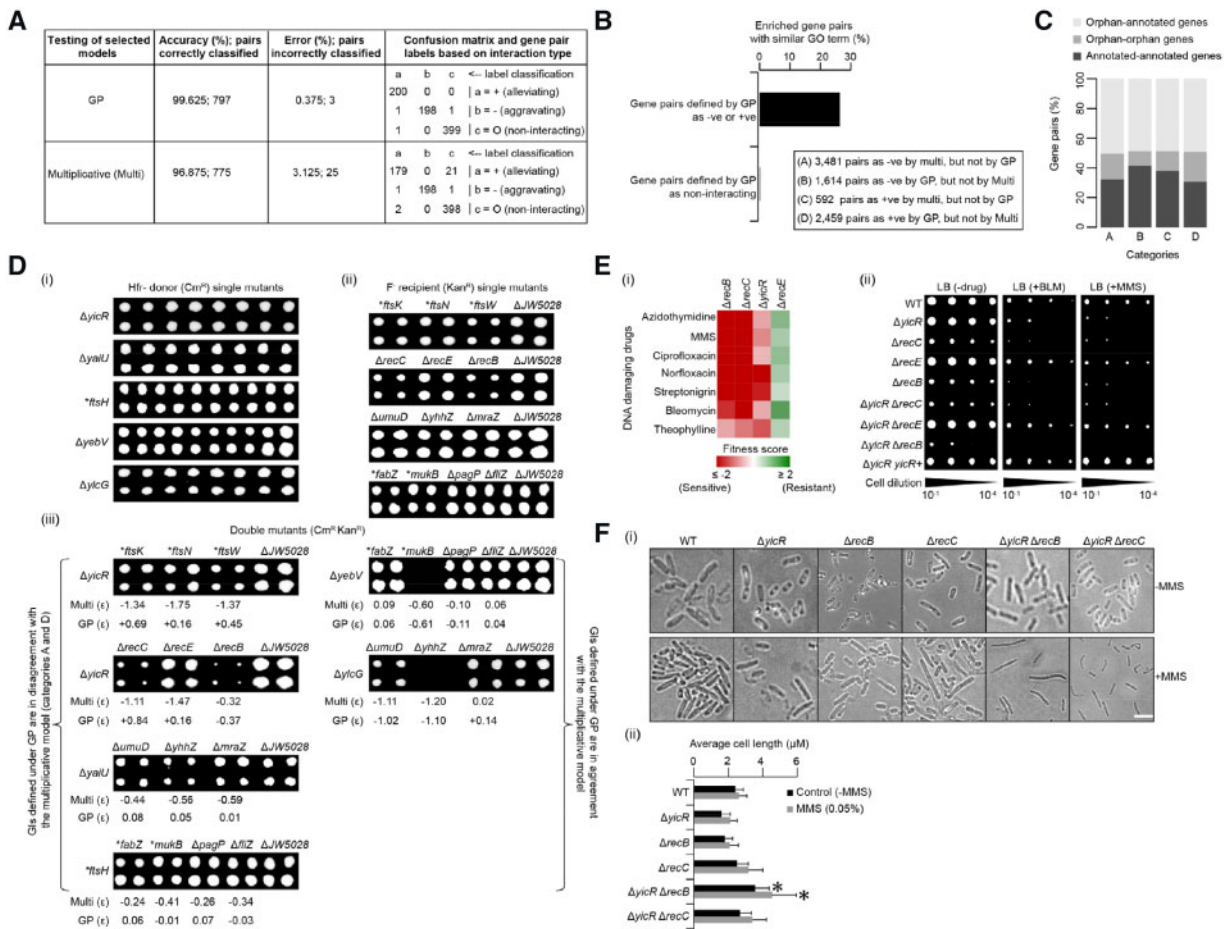


Fig. 2. Assessment and validation of GI pairs from GP definition. (A) Accuracy and error rate of the GI definitions estimated on 800 test pairs compiled from the genome integrity study (Kumar et al., 2016). (B) Percentage gene pairs defined as aggravating (-ve), alleviating (+ve) and noninteracting under GP definition enriched (FDR-corrected P -value ≤ 0.05 ; enriched GO term in the ranked gene set was assessed by Kolmogorov-Smirnov test) for a shared Gene Ontology (GO) term function. (C) Percentage gene pairs with epistatic relationships between orphan (gene of unknown function)-orphan, annotated-annotated or orphan-annotated genes. (D) Confirmation of gene pair classifications by the GP model that were either in agreement or in disagreement with the multiplicative model. For this comparison, the top 25% of the ranked GI pairs were selected. GI (b) score ≤ -0.152 for GP and ≤ -0.221 for the multiplicative model are considered as aggravating (-ve sign), and ≥ 0.153 for GP and ≥ 0.203 for multiplicative as alleviating (+ve). Scores that did not meet the chosen aggravating or alleviating threshold were considered as noninteracting (with no sign indicated). The fitness of the double-mutant colonies and their respective GI scores from the indicated model should dictate their predictive capability of identifying reliable GI type. Hfr Cavalli donor (nonessential or *essential hypomorphic) single mutant marked with chloramphenicol (Cm); (i) and F⁻ recipient single-mutant (nonessential or *essential hypomorphic) strains marked with kanamycin (Kan); (ii) were crossed and selected on plates with Cm and Kan to generate double mutants (iii), which are shown along with their respective scores from GP and multiplicative definitions below the colony images. Multiple colonies represent replicates. Functionally unrelated gene *JW5028* was used as a control. (E) Sensitivity or resistance (i) of the indicated deletion mutant strains to DNA-damaging agents that induce double-strand breaks, as reported in a chemical-genetic screen (Nichols et al., 2011). Growth sensitivity (ii) is shown for mutants and wild-type (WT) cells grown on bleomycin (BLM) and MMS. Phenotypic complementation shows the overexpression of *yicR* in trans. (F) Cell morphology micrographs (i) and cell lengths (ii) of WT and mutant strains before and after DNA damage with MMS ($t = 2$ h) treatment. Cell lengths of the WT or mutant strain is an average of ~ 100 different cells; * indicate significant (Student's t -test) difference between double mutants and their single mutants. Scale bar equals 10 μ m

annotations among interacting gene pairs. Using GSEA with the pre-ranked gene list of fitness data and all 3440 GO terms as input, we found enrichment for 9 significant ($P \leq 0.05$; false-discovery rate ≤ 0.25) GO terms. Among these terms, one-quarter (25%, 999 of 4073) of the enriched ($P = 1.04 \times 10^{-278}$ by hypergeometric test) gene pairs defined as aggravating or alleviating under the GP definition shared a GO term (Fig. 2B; Supplementary Table S5). Conversely, only very few (0.07%, 3 of 4073) disagreeing aggravating (3481; category A) or alleviating (592; category C) GI pairs that were defined by GP as noninteracting (Fig. 2B) shared the same GO functional attribute. Instead, 48-fold underrepresentation for pairs with a shared annotation ($P = 1.39 \times 10^{-232}$ by hypergeometric test) than interacting pairs observed under the GP definition. While the assumption is that most noninteracting gene pairs should not have a close functional relationship (Mani et al., 2008), shared functionality observed for certain enriched noninteracting pairs is expected because the case study (Kumar et al., 2016) was focused on genome integrity rather than random pairs, and because a low level

cross-talk is expected between processes due to indirect, common effects on fitness. In any case, since the enrichment for shared GO terms among the number of noninteracting pairs is several fold (220 times; 26.4% versus 0.12%; Fig. 2B) lower than among the number of interacting pairs under the GP definition, our results suggest that GP is a strong predictor of functional links, as indicated by aggravating or alleviating GIs.

Next, a detailed inspection of the pairs categorized as aggravating (1614; category B) and alleviating (2459; category D) by the GP model, and those pairs in disagreement with the multiplicative model (3481, category A; 592, category C), indicated that in comparison to interactions with annotated genes, over half (57%; 4673 of 8146) of the pairs had epistatic dependencies between unannotated *E. coli* genes (i.e. orphans or genes of unclear function) not previously linked to DNA repair, or between unannotated and known genomic integrity targets (Fig. 2C). To rigorously verify the reliability of the interactions defined using the GP model, we selected a subset of these gene pairs for independent experimental validation using

the mini-array genetic crosses. Specifically, our selection of ‘agreement’ gene pairs entails noninteracting or aggravating interactions under GP and multiplicative models (e.g. *ylcG* with *umuD/lybhZ/mraZ*; *yebV* with *fabZ/mukB/pagP/fliZ*; *yicR* with *recB*). Whereas the ‘disagreement’ gene pairs include aggravating interactions under multiplicative definition, but noninteracting according to the GP model (e.g. *yaiU* with *umuD/lybhZ/mraZ*; *ftsH* with *fabZ/mukB/pagP/fliZ*; category A in [Supplementary Table S1](#)), or pairs that were noninteracting in the multiplicative model, but alleviating under the GP definition (e.g., *yicR* with *ftsKNW* or *recCE*; category D in [Supplementary Table S1](#)). Notably, when GP and multiplicative definitions disagreed, GP predictions were experimentally verified ([Fig. 2D](#)). As an illustrative example, an uncharacterized gene, *yicR* displayed alleviating and aggravating GIs with homologous recombination genes (*recBCE*) under the GP definition, whereas the multiplicative model defined all these GIs as aggravating ([Fig. 2D](#)). To examine the biological implication of these GP or multiplicative assignments, we compared the sensitivity of the *E. coli* *yicR*-mutant strain to DNA damage induced by genotoxic agents (BLM, MMS) with the sensitivities of WT or *recBCE* single gene deletion strains.

BLM or MMS introduces clusters of lesions in DNA to form single- or double-strand breaks (SSBs, DSBs), or produces an assortment of N- and O-methylated bases in DNA by blocking the replication fork progression. In *E. coli*, when these genotoxins induce DNA damage, cell survival depends on *rec*-dependent homologous recombination and repair pathways ([Knezevic-Vukcevic and Simic, 1991](#); [Kosa et al., 2004](#); [Nowosielska et al., 2006](#)). It is known from previous evidence ([Knezevic-Vukcevic and Simic, 1991](#); [Nichols et al., 2011](#)) that *yicR* and the recombination-deficient *recBC*-mutant strains, but not *recE* deficient strains, are sensitive to the cytotoxic effects of BLM, MMS ([Fig. 2E](#)) or other genotoxic agents that induce DSBs ([Nichols et al., 2011](#)). Consistent with this observation, in our experiments, strains lacking *yicR* showed hypersensitivity ([Fig. 2E](#)), but inactivation of both *yicR* and *recC* did not enhance the effect compared to the respective single mutants. Thus, the lack of synergy in *yicR-recC* double mutants suggests, consistent with the GP prediction, that *yicR* functions with *recC*. Conversely, the increased sensitivity of *yicR-recB* double mutants to BLM or MMS is consistent with the GP and multiplicative definitions that *recB* is absolutely required for the *yicR* phenotype. Whereas, when compared to *yicR-recB/recC* phenotypes, *yicR-recE* double mutants were fully resistant like the *recE* single mutants, suggesting that *yicR* does not function redundantly with *recE*. This lack of additive effect is in agreement with the GP definition.

Recombination-deficient strains (e.g. *rec* mutants), sensitive to genotoxic agents, exhibit impaired cell division and chromosome segregation, forming aberrantly long, nonseparate, multi-nucleate filaments ([Babu et al., 2011](#); [Kumar et al., 2016](#)). Therefore, we ascertained if the observed *yicR-recB* and *yicR-recC* phenotypes suggested by hypersensitivity to MMS, as well as by aggravating and alleviating GIs by GP, respectively, indicate involvement in chromosome segregation following DNA damage. We found that the double mutants of *yicR* with *recB*, but not with *recC*, exhibited longer (~9.2 μ m avg. cell length) filamentous cells compared to WT or single mutants ([Fig. 2F](#)) after MMS treatment. This suggests that YicR appears to function in cooperation with RecC, while functioning in parallel with RecB in DSB repair pathway, which is consistent with the GP and multiplicative definition. Additionally, our findings decisively indicate that, like *yicR-recC* or *yicR-recE* growth fitness, other gene pairs that were tested by genetic crosses ([Fig. 2D](#)) were in fact misclassified as aggravating interactions under the multiplicative model.

4 Conclusions

Global epistatic *E. coli* connectivity maps based on colony growth measurements, under standard laboratory growth conditions or in the presence of genotoxic or environmental stressors ([Babu et al., 2011, 2014](#); [Cote et al., 2016](#); [Gagarinova et al., 2016](#);

[Kumar et al., 2016](#)) have been generated using the multiplicative function, following the framework established in yeast ([Baryshnikova et al., 2010](#); [Collins et al., 2006, 2007](#)). However, there was no comparable, rigorous evaluation performed on large-scale *E. coli* GI data to justify the use of the multiplicative model. As a result, it has not been clear if the multiplicative definition has had any ramifications on the interpretation of GIs generated in *E. coli* to date. Therefore, in the current study, using the growth fitness of viable digenic mutant combinations and their corresponding single-gene deletion mutants of functionally related protein-coding genes involved in genome integrity ([Kumar et al., 2016](#)), we provided predictions for pairs of genes with the interaction subtype for each of the selected machine learning and predefined models. This dataset was then employed to compare the underlying performance of the GIs between definitions, with a set of statistical and experimental measures, to reveal the practical differences among definitions. Notably, based on the error metrics and $|\epsilon$ -score| distributions of the training and genome integrity datasets, we found the GP definition to display lower RMSE and MAE values, with ϵ distributions centered around zero and more closely conforming to the ideal normal distribution (i.e. most gene pairs are noninteracting) than under other definitions that showed more bias and greater dispersion of ϵ -scores. Although the NN definition exhibited low CV on the noninteracting test gene pairs ([Fig. 1A](#)), as we have shown in [Figure 1B](#), NN on the test pairs exhibited a negative shift in the normal distribution of ϵ -score (i.e. deviating from zero) compared to the distribution of GP that approximated a normal distribution that is tight and close to zero (i.e. with lowest bias and dispersion for ϵ scores). In any case, our rationale for choosing GP over NN definition is not solely based on the error metrics of noninteracting test dataset, but rather on our observations with other sets of experimental results ([Fig. 1B](#); [Supplementary Fig. S1A](#); and [Table S2](#), sheets 4 and 5).

Other important observations came from the analyses was the gene pairs that were inconsistently classified by the GP and the multiplicative models. In the case of 3481 aggravating (category A) and 592 alleviating (category D) GIs defined under the multiplicative model, the double mutants and one of the two single-mutant fitness measurements appear to near one, or WT and not different from one another ([Fig. 1D](#)); yet, the multiplicative model has defined these pairs as aggravating or alleviating. In contrast, GP defined the majority of these pairs as noninteracting, which we confirmed to be consistent experimentally with the genetic crosses performed on more than a dozen gene pairs. Additional evidence from GSEA on these GP-classified noninteracting pairs revealed less functional enrichment, supporting the respective GP model assignments. Decision tree classifiers also further indicate that GP had the highest accuracy and lowest error in consistently classifying the test and genome integrity gene pairs to appropriate interaction type when compared to the multiplicative model. When we repeated the aforesaid analyses on an independent genome-wide GI study ([Babu et al., 2014](#)) that included genes broadly representing various bacterial processes, we found that GP similarly outperformed multiplicative and other definitions.

While through multiple lines of evidence we have demonstrated that GP is able to show comparable results with different *E. coli* datasets, for example, with functionally related ([Kumar et al., 2016](#)) or unrelated ([Babu et al., 2014](#)) groups of genes, we tested if this definition can likewise perform well in the budding yeast that examined 5.4 million digenic pairs for synthetic GIs ([Costanzo et al., 2010](#)). The examination of the $|\epsilon$ -score| distributions showed that, unlike in *E. coli*, the GP and multiplicative definitions in yeast displayed almost identical error metrics and $|\epsilon$ -score| distributions with lowest bias and dispersion, conforming to the expectation that most gene pairs are noninteracting ([Supplementary Fig. S2B](#)). This outcome raises the question of why the multiplicative model was not consistent in certain cases with the GP definition on *E. coli* GI datasets. We found that the discrepancy with the multiplicative definition occurred when the normalized colony sizes for one of the observed single mutants is much larger (W_x or $W_y > 1.0$), causing the multiplicative model to consider those gene pairs as aggravating instead of alleviating or noninteracting. Hence, it can be less sensitive to

defining alleviating or suppression phenotypes, which in fact is evident from one-tenth (13%; 311 of 2459) of the gene pairs (category D) that were classified as aggravating (Supplementary Table S1). The remaining (87%, 2148 of 2459) large proportion of gene pairs were misclassified as noninteracting because the multiplicative model cannot distinguish the small fitness differences observed between single and double mutants. While under certain circumstances the multiplicative model introduces disagreement with GP, for most (84%; 21 932 of 26 005) of the top-ranking gene pairs, we found both definitions to perform equally well for identifying epistatic relationships when the growth of the double mutants (better or worse) can be different from at least one single mutant with a moderate fitness defect or both single mutants with near WT fitness (Supplementary Fig. S2C). Taken together, we suggest those favoring the multiplicative model should exclude interactions, where the fitness of one of the single mutants is especially large, to place more confidence on biological conclusions.

To facilitate exploration, we have provided the source code used to generate the machine learning models in WEKA software environment, along with the supporting *E.coli* training dataset in WEKA's native format (Supplementary Material). In summary, we conclude that GP is a viable alternative for future epistatic studies or those underway in unicellular (Peters et al., 2016; Shapiro et al., 2018) or multicellular organisms (Billmann et al., 2018) using the CRISPR genetic screens that are in quest for identifying functional relationships. As with any other high-throughput screening framework, large-scale genetic screens can be noisy; hence it is beneficial to employ GP and the multiplicative model or any other definition, as long as there is a general consensus between definitions in classifying the consistent GI types.

Acknowledgements

We thank Gabe Musso (Harvard Medical School, Massachusetts, USA), members of the Babu lab (James Vlasblom, Matthew Jessulat) and anonymous reviewers for their helpful discussions and valuable comments. We also thank Mohammed Taha Moutaoufik for assisting with cell morphology micrographs. A.K. and S.Z. designed and conducted the machine learning-based analyses. S.P. compiled the raw and normalized fitness data from the large-scale genetic screens, and performed GSEA and other computational analyses. A.H., A.G. and S.K. created Hfr gene deletion mutants and F^- hypomorphic strains, and A.H. performed experimental validation, with feedback from K.A. and M.B. A.K., S.Z. and M.B. wrote computational methods, and M.B. wrote the article, with input from co-authors.

Funding

This work was supported by the National Sciences and Engineering Research Council of Canada [DG-05336 to S.Z., DG-20234 to M.B.]. A.G. was supported by a Canadian Institutes of Health Research (CIHR) Postdoctoral Fellowship. S.Z. holds Canada Research Chair in Computational Learning Theory, and M.B. is a CIHR New Investigator (MSH-130178).

Conflict of Interest: none declared.

References

Baba, T. et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006.0008.
 Babu, M. et al. (2011a) A dual function of the CRISPR-CAS system in bacterial antiviral immunity and DNA repair. *Mol. Microbiol.*, **79**, 484–502.
 Babu, M. et al. (2011b) Genetic interaction maps in *Escherichia coli* reveal functional crosstalk among cell envelope biogenesis pathways. *PLoS Genet.*, **7**, E1002377.
 Babu, M. et al. (2014) Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in *Escherichia coli*. *PLoS Genet.*, **10**, E1004120.
 Bandyopadhyay, S. et al. (2008) Functional maps of protein complexes from quantitative genetic interaction data. *PLoS Comput. Biol.*, **4**, E1000065.
 Bandyopadhyay, S. et al. (2010) Rewiring of genetic networks in response to DNA damage. *Science*, **330**, 1385–1389.
 Baryshnikova, A. et al. (2010) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods*, **7**, 1017–1024.

Bassik, M.C. et al. (2013) A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell*, **152**, 909–922.
 Bidnenko, V. et al. (2006) The *Escherichia coli* UvrD helicase is essential for Tus removal during recombination-dependent replication restart from Ter sites. *Mol. Microbiol.*, **62**, 382–396.
 Billmann, M. et al. (2018) Widespread rewiring of genetic networks upon cancer signaling pathway activation. *Cell Syst.*, **6**, 52–64.E54.
 Boettcher, M. et al. (2018) Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat. Biotechnol.*, **36**, 170–178.
 Butland, G. et al. (2008) eSGA: *E. coli* synthetic genetic array analysis. *Nat. Methods*, **5**, 789–795.
 Collins, S.R. et al. (2006) A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.*, **7**, R63.
 Collins, S.R. et al. (2007) Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature*, **446**, 806–810.
 Costanzo, M. et al. (2010) The genetic landscape of a cell. *Science*, **327**, 425–431.
 Costanzo, M. et al. (2016) A global genetic interaction network maps a wiring diagram of cellular function. *Science*, **353**, aaf1420.
 Costanzo, M. et al. (2019) Global genetic networks and the genotype-to-phenotype relationship. *Cell*, **177**, 85–100.
 Cote, J.P. et al. (2016) The genome-wide interaction network of nutrient stress genes in *Escherichia coli*. *MBio*, **7**, e01714-16.
 Cox, M.M. and Lehman, I.R. (1982) RecA protein-promoted DNA strand exchange. Stable complexes of RecA protein and single-stranded DNA formed in the presence of ATP and single-stranded DNA binding protein. *J. Biol. Chem.*, **257**, 8523–8532.
 Du, D. et al. (2017) Genetic interaction mapping in mammalian cells using CRISPR interference. *Nat. Methods*, **14**, 577–580.
 Egger, A.L. et al. (2003) The C terminus of the *Escherichia coli* RecA protein modulates the DNA binding competition with single-stranded DNA-binding protein. *J. Biol. Chem.*, **278**, 16389–16396.
 Frost, A. et al. (2012) Functional repurposing revealed by comparing *S. pombe* and *S. cerevisiae* genetic interactions. *Cell*, **149**, 1339–1352.
 Gagarinova, A. et al. (2012) Mapping bacterial functional networks and pathways in *Escherichia coli* using synthetic genetic arrays. *J. Vis. Exp.*, **69**, 4056.
 Gagarinova, A. et al. (2016) Systematic genetic screens reveal the dynamic global functional organization of the bacterial translation machinery. *Cell Rep.*, **17**, 904–916.
 Guenole, A. et al. (2013) Dissection of DNA damage responses using multiconditional genetic interaction maps. *Mol. Cell*, **49**, 346–358.
 Hoppins, S. et al. (2011) A mitochondrial-focused genetic interaction map reveals a scaffold-like complex required for inner membrane organization in mitochondria. *J. Cell Biol.*, **195**, 323–340.
 Horlbeck, M.A. et al. (2018) Mapping the genetic landscape of human cells. *Cell*, **174**, 953–967.E922.
 Horn, T. et al. (2011) Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nat. Methods*, **8**, 341–346.
 Jonikas, M.C. et al. (2009) Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science*, **323**, 1693–1697.
 Kim, S. et al. (2016) METAKTSP: a meta-analytic top scoring pair method for robust cross-study validation of omics prediction analysis. *Bioinformatics*, **32**, 1966–1973.
 Knezevic-Vukcevic, J. and Simic, D. (1991) RecBC promoted repair of bleomycin damage in *Escherichia coli*. *Biochimie*, **73**, 497–500.
 Kosa, J.L. et al. (2004) RecN and RecG are required for *Escherichia coli* survival of bleomycin-induced damage. *Mutat. Res.*, **554**, 149–157.
 Kowalczykowski, S.C. et al. (1987) Effects of the *Escherichia coli* Ssb protein on the binding of *Escherichia coli* RecA protein to single-stranded DNA. Demonstration of competitive binding and the lack of a specific protein-protein interaction. *J. Mol. Biol.*, **193**, 81–95.
 Kramer, M.H. et al. (2017) Active interaction mapping reveals the hierarchical organization of autophagy. *Mol. Cell*, **65**, 761–774.E765.
 Kumar, A. et al. (2016) Conditional epistatic interaction maps reveal global functional rewiring of genome integrity pathways in *Escherichia coli*. *Cell Rep.*, **14**, 648–661.
 Kuzmin, E. et al. (2018) Systematic analysis of complex genetic interactions. *Science*, **360**, eaao1729.
 Lehner, B. et al. (2006) Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nat. Genet.*, **38**, 896–903.
 Mani, R. et al. (2008) Defining genetic interaction. *Proc. Natl. Acad. Sci. USA*, **105**, 3461–3466.

- Mccooll,J.D. and Sandler,S.J. (2001) Effects of mutations involving cell division, recombination, and chromosome dimer resolution on a *priA2::kan* mutant. *Proc. Natl. Acad. Sci. USA*, **98**, 8203–8210.
- Nichols,R.J. *et al.* (2011) Phenotypic landscape of a bacterial cell. *Cell*, **144**, 143–156.
- Nowosielska,A. *et al.* (2006) Homologous recombination prevents methylation-induced toxicity in *Escherichia coli*. *Nucleic Acids Res.*, **34**, 2258–2268.
- Peters,J.M. *et al.* (2016) A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell*, **165**, 1493–1506.
- Rizzolo,K. *et al.* (2017) Features of the chaperone cellular network revealed through systematic interaction mapping. *Cell Rep.*, **20**, 2735–2748.
- Ryan,C.J. *et al.* (2012) Hierarchical modularity and the evolution of genetic interactomes across species. *Mol. Cell*, **46**, 691–704.
- Schuldiner,M. *et al.* (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell*, **123**, 507–519.
- Shapiro,R.S. *et al.* (2018) A CRISPR-Cas9-based gene drive platform for genetic interaction analysis in *Candida albicans*. *Nat. Microbiol.*, **3**, 73–82.
- Smith,T.C. and Frank,E. (2016) Introducing machine learning concepts with WEKA. *Methods Mol. Biol.*, **1418**, 353–378.
- Srivas,R. *et al.* (2013) A UV-induced genetic network links the RSC complex to nucleotide excision repair and shows dose-dependent rewiring. *Cell Rep.*, **5**, 1714–1724.
- Subramanian,A. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, **102**, 15545–15550.
- Szappanos,B. *et al.* (2011) An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat. Genet.*, **43**, 656–662.
- Taylor,M.B. and Ehrenreich,I.M. (2014) Genetic interactions involving five or more genes contribute to a complex trait in yeast. *PLoS Genet.*, **10**, E1004324.
- Typas,A. *et al.* (2008) High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nat. Methods*, **5**, 781–787.
- Ullitsky,L. *et al.* (2009) Towards accurate imputation of quantitative genetic interactions. *Genome Biol.*, **10**, R140.
- Van Opijnen,T. *et al.* (2009) Tn-Seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods*, **6**, 767–772.
- Wagih,O. *et al.* (2013) SGATools: one-stop analysis and visualization of array-based genetic interaction screens. *Nucleic Acids Res.*, **41**, W591–596.
- Walhout,M. *et al.* (2012) *Handbook of Systems Biology: Concepts and Insights*. Elsevier Science Publishing Co Inc, London, UK.
- Wang,X. *et al.* (2014) Widespread genetic epistasis among cancer genes. *Nat. Commun.*, **5**, 4828.
- Wong,S.L. *et al.* (2004) Combining biological networks to predict genetic interactions. *Proc. Natl. Acad. Sci. USA*, **101**, 15682–15687.