Main Manuscript for

A continuous epistasis model for predicting growth rate given combinatorial variation in gene expression and environment

Ryan M. Otto[1], Agata Turska-Nowak[2‡], Philip M. Brown[1], Kimberly A. Reynolds[1,2*]

[1]Green Center for Systems Biology - Lyda Hill Department of Bioinformatics, The University of Texas Southwestern Medical Center, Dallas, TX 75230; [2]Department of Biophysics, The University of Texas Southwestern Medical Center, Dallas, TX 75230

[‡]Current affiliation: Polish Stem Cells Bank (PBKM), Warsaw, Poland

*Corresponding author
Kimberly A. Reynolds
Lyda Hill Department of Bioinformatics
UT Southwestern Medical Center, ND11.120D
6001 Forest Park Blvd
Dallas, TX 75390

Phone: 214-645-6157
Email: kimberly.reynolds@utsouthwestern.edu

Author Contributions: KAR and RMO designed the project. RMO, ATN, and PMB performed research, including both experiments and data analysis. RMO, PMB, and KAR wrote the paper. RMO, PMB, ATN, and KAR edited and revised the paper.

Competing Interest Statement: The authors have no competing interests to declare.

Classification: Biological Sciences – Systems Biology

Keywords: Genotype-to-phenotype, gene-by-environment, epistasis, gene expression, CRISPRi

This PDF file includes:
Main Text
Figures 1 to 6

Abstract

The ability to predict changes in cellular growth rate given variation in gene expression is critical to interpret disease-causing mutations, engineer biosynthetic pathways, and understand evolutionary constraints on mRNA abundance. However, the relationship between gene expression and growth rate is nonlinear and shaped by environmental and genetic context. As a consequence, genetic studies focused on discrete perturbations with strong effects like total knockouts or complete nutrient depletion miss critical information relevant to organismal variation, evolution, and human disease. To address these challenges, we developed a strategy in which we 1) experimentally measure the growth rate effects of titrated changes in enzyme expression under varied environmental and genetic contexts for *E. coli* and 2) use a subset of these data to train an interpretable, predictive model. Models trained using a sparse subset of our experimental data — sampling only pairwise perturbations over genes and environments — were sufficient to predict the effects of higher-order combinations of expression and environmental perturbations. This high-throughput, generalizable framework provides a strategy for characterizing the growth rate effects of altering gene expression across entire metabolic pathways, or even genomes, in varied environments using sparsely sampled, low-order measurements.

Significance Statement

A predictive understanding of the relationship between gene expression and growth rate is critical to engineer metabolic pathways with novel functions, relate mutations to disease progression, and discover new therapeutic opportunities. Here, we describe an interpretable, low-complexity modeling approach to predict bacterial growth rates given multiple changes in gene expression and environment. The model is trained using only single and pairwise perturbation data, yet it can predict the effects of higher-order combinations of perturbations. This approach drastically reduces the amount of data required for training and suggests a generalizable strategy for using well-sampled epistasis measurements to build quantitative genotype-environment-phenotype maps.

Main Text

Introduction

Variation in gene expression tunes organism-wide traits like biomass yield, cell division, and differentiation, but predicting these broad phenotypes based on molecular data remains challenging. In the context of infection, expression changes can lead to immune evasion (1), trigger a switch between commensal and pathogenic behavior (2), or confer antibiotic tolerance or resistance (3, 4). Similarly, dysregulated transcription in human cells can initiate tumorigenesis and metastasis (5, 6) and transcriptional diversity within tumors confers differential susceptibility to chemotherapeutics and radiation (7, 8). If transcriptional molecular markers can be leveraged to predict these phenotypes, we could drastically improve patient prognosis. More generally, the relationship between gene expression and cellular growth rate represents a well-defined instantiation of the broader problem of mapping genotype to phenotype.

However, two complications stand in the way of making such predictions. First, expression changes occur over a continuous range and are often non-linearly related to growth rate (9-11). Thus, a simple on/off mechanism is rarely sufficient to describe the compendium of expression levels observed in a population. While “endpoint” studies such as gene knockouts or near-complete knockdowns have classically proved useful for annotating gene function (12), discovering new regulatory networks (13), and revealing the organization of the cell (14-17), they do not offer sufficient information to predict the effects of intermediate changes in gene expression. Sampling a range of perturbation strengths is required to quantitate this continuous, nonlinear response. Second, expression-growth rate relationships are context-dependent (9, 10, 18). An organism’s genetic background and environment may alter the effect of gene expression changes, a phenomenon called epistasis (19, 20). The amount of data required to densely sample these context-dependent relationships scales exponentially with the complexity of the system. As a consequence, these experiments quickly become intractable. A new approach is required to approximate these landscapes using a reasonable amount of input data.

Interestingly, prior work on a related problem — predicting the effects of antibiotic and chemotherapeutic combination therapies on cell growth — has considered continuous and/or

29 context-dependent models (21-27). Of particular relevance is the dose-response model introduced
30 in Zimmer et al., 2016 (28). The authors examined cases where the presence of one drug altered
31 the effect of a second drug. This interaction resulted in a change in the "effective dose" of the
32 second drug and was summarized with two coupling constants – the effect of drug "A" on "B", and
33 the effect of drug "B" on "A". The authors found that a model constructed using only these pairwise
34 coupling terms could predict the effects of higher-order combinations of drugs. This model was unit-
35 agnostic – meaning the framework can be trained on any quantitative input, not just drug treatment.
36 However, to our knowledge, it has never been extended to make predictions of growth rate given
37 variation in expression or nutrient abundance. Following the rationale that a reduction in enzyme
38 expression is functionally analogous to inhibiting an enzyme with a drug, we hypothesized that a
39 similar coupling-based approach could capture context-dependence between genes and predict
40 the impact of multiple gene expression changes on bacterial growth rate.

41 In this work, we considered continuous expression-growth rate relationships for nine *E. coli*
42 metabolic enzymes drawn from diverse metabolic pathway architectures. We measured the growth
43 rate effects of 4,404 pairwise combinations of titrated gene knockdowns. Using 20% of these data,
44 we fit a continuous, coupling-sensitive model that was able to predict growth rate for the remaining
45 80% of pairwise perturbations. The model also predicted the effect of third-order gene knockdowns
46 without being trained on high-order data. Finally, we extended this model to predict the combined
47 effect of gene expression and growth media changes. Growth rate predictions for third- and fourth-
48 order genetic and environmental perturbations remained robust. Taken together, our experimental
49 and computational methods provide a strategy to incorporate sparsely sampled epistasis
50 measurements into a predictive model. This approach can be used to predict bacterial growth rate
51 following numerous intermediate gene expression changes in a variety of environments and is
52 scalable to quantify expression-growth rate coupling across entire metabolic pathways.

53

54 **Results**

55 **Growth rate titrates with gene expression following CRISPR interference**

56 Training and testing an expression-growth rate model requires high-throughput growth rate
57 measurements across cells harboring distinct, intermediate variations in gene expression. To
58 generate these data, we used mismatch CRISPR interference (CRISPRi). CRISPRi uses a
59 catalytically dead Cas9 endonuclease (dCas9) and a single guide RNA (sgRNA) to target a specific
60 DNA locus, sterically repressing transcription at the target site (29). In standard CRISPRi, the target
61 of the sgRNA is specified by a 20-nucleotide homology region designed to be fully complementary
62 to the target gene of interest. In contrast, mismatch CRISPRi uses sgRNAs containing nucleotide
63 mismatches in the target homology region (11, 30-32). These mismatches disrupt sgRNA-DNA
64 interactions, decreasing gene repression strength. By designing a group of sgRNAs targeting the
65 same DNA locus with different numbers of mismatches, we can tune a single gene's transcription
66 to numerous intermediate levels (Fig 1A). The growth rates resulting from this variation can be
67 quantified with excellent throughput using CRISPRi-seq (18, 30, 33-35). In our implementation of
68 this approach, these sgRNAs are transformed into *E. coli* containing a chromosomally encoded
69 dCas9 under control of a tet-inducible promoter. To ensure control of population size and
70 environmental conditions, we grow the transformed cells in a turbidostat (36). After inducing a cell's
71 CRISPRi machinery, we use next-generation sequencing to monitor the frequency of individual
72 sgRNAs in a mixed population over time. The frequency of individual sgRNAs exponentially
73 decreases (or increases) as the cell population grows. Thus, linear regression of the relationship
74 between the logarithm of the sgRNA relative frequency and time (in generations) allows us to
75 estimate the growth rate effects of thousands of sgRNAs in a single experiment (Fig 1B). By plotting
76 growth rate as a function of sgRNA repression strength, we can define continuous expression-
77 growth rate functions for genes of interest (Fig 1C).

78 In this study, we selected nine genes to target with mismatch CRISPRi, with an eye to
79 sampling distinct metabolic pathways and architectures (Fig 2A-D). Included were genes encoding
80 enzymes catalyzing consecutive metabolic steps (*dapA*/*dapB*, *purN*/*purL*), parallel reactions
81 (*gdhA*/*gltB*), and a three-enzyme loop (*folA*/*glyA*/*thyA*). All genes except *purN*, *gdhA* and *gltB* were
82 previously shown to be essential in our experimental conditions (OD_{600} < 0.01 after 24 hours in
83 minimal media, 37). Double knockout of *gdhA*/*gltB* was shown to be synthetically lethal (38). We

designed 9-12 distinct sgRNAs targeting each gene as well as a non-targeting sgRNA control to generate a 96 sgRNA library (Table S1). We quantified the effect of each CRISPRi perturbation on the mRNA concentration of its target gene using reverse transcription-quantitative PCR (RT-qPCR). As expected, our library of sgRNAs created a titrated range of knockdowns for every gene in our data set (Table S2).

Pairwise CRISPRi yields an expression-dependent epistatic landscape

To investigate expression-growth rate coupling between genes, we used Golden Gate cloning to assemble a library of constructs covering all pairwise sgRNA combinations (39). We constructed this library in three distinct barcoded plasmid vectors to generate internal replicates. Sequencing of our library prior to selection showed that it was complete and well-distributed across the designed sgRNAs (Fig S1). We quantified the growth rate effect of each CRISPRi knockdown (both singles and pairwise) on exponentially growing *E. coli* cells in M9 minimal media using continuous culture and next-generation sequencing. Growth rate was well-correlated between internal replicates (Fig S2A-C). In agreement with earlier work, changing the sgRNA order within pairwise constructs (sgRNA1-sgRNA2 vs. sgRNA2-sgRNA1) did not significantly alter growth rate (Fig S2D, 29). Thus, after pooling our three internal replicates for each of two sgRNA orders, we had six independent measurements for each pairwise CRISPRi treatment. As in prior work, we used these replicates to identify and filter out “escapers”: replicates whose growth rate was significantly faster than all other replicates of the same CRISPRi treatment, likely due to evasion of the CRISPRi machinery (Fig S3A, 32). Escaper correction removed 182 measurements (0.7% of the data, Fig S3B-C). Finally, we averaged the remaining replicates and rescaled our growth rate data, setting the minimum growth rate observed to 0 and the non-targeting CRISPRi control to 1. In total, we quantified the growth rate effect of 4,404 unique CRISPRi perturbations. We found that growth rates from this study correlated well with measurements from our previous work (Fig S4, 32).

We quantified single gene expression-growth rate relationships by relating target gene repression strength to growth rate for sgRNA constructs consisting of one targeting and one non-targeting control sgRNA (Fig 2E-M, Table S2). All genes showed steep, monotonic expression-

112 growth rate relationships except *gdhA* and *gltB*, which were known to be nonessential in our growth
113 conditions. Notably, our approach to sgRNA library design reliably sampled the steep, intermediate
114 growth regimes for each gene, allowing us to precisely identify the knockdown level where each
115 gene's transcription became growth-limiting.

116 In addition to these single gene relationships, we densely sampled pairwise expression-
117 growth rate landscapes (Fig 3, lower triangle, Table S3). To quantify interactions between CRISPRi
118 perturbations cooccurring within a cell, we quantified epistasis between every pair of CRISPRi
119 knockdowns using a multiplicative (Bliss) model, where epistasis between sgRNAs 1 and 2 is equal
120 to $gr_{1,2} – gr_1 * gr_2$ (Fig 3, upper triangle). For the majority of sgRNA pairs, we observed minimal
121 epistasis. This is consistent with expectation given that: 1) many of the gene pairs are drawn from
122 unrelated metabolic pathways and 2) a number of the sgRNAs yielded only modest changes in
123 gene expression. We observed both positive and negative epistasis, particularly for sgRNAs with
124 the strongest effects on gene repression. For example, our data recapitulated the synthetic lethality
125 previously reported for *gdhA* and *gltB* as negative epistasis: we observed minimal growth rate
126 defects when these genes were repressed individually, but severe growth rate defects when both
127 genes were repressed together (Fig 3, upper triangle, purple pixels, 38).

128 However, even for gene pairs where epistasis was prevalent, the intensity of epistasis
129 varied drastically within different expression regimes. For the *gdhA*/*gltB* example, moderate
130 knockdown of each gene was required before epistasis could be quantified. Following pairwise
131 repression of *dapA* and *glyA*, however, epistasis strength appears to correlate with *glyA* knockdown
132 intensity, irrespective of *dapA* knockdown. And some gene pairs, such as *gdhA*/*folA*, showed both
133 positive and negative epistasis throughout their pairwise expression-growth rate landscape. Given
134 how widely gene-gene epistasis varied based on the specific perturbations explored, we sought to
135 discover a new way to describe and model gene-gene epistasis across entire pairwise expression-
136 growth rate landscapes.

137

138 **A continuous, coupling-sensitive model of expression-growth rate relationships**

139 In developing the model, we sought to move beyond a matrix of many discrete epistasis
140 measurements to a continuous, predictive description of expression-growth rate relationships. To
141 begin, we modeled the effect of single gene knockdowns on growth rate. We found that a sigmoidal
142 model relating relative repression (R) to growth rate (g) using two parameters, steepness (n) and
143 repression at half-maximal growth rate (R_0) effectively modeled all single gene knockdown data
144 (Equation 1, Fig 2E-M).

$$g(R) = \frac{1}{1 + e^{n(R-R_0)}} \tag{1}$$

145

146 These sigmoidal fits enabled continuous prediction of the effects of gene repression on
147 growth rate, improving upon our discrete experimental measurements. We wanted to retain this
148 continuous predictive power as we expanded our growth rate predictions to pairwise CRISPRi
149 knockdowns. Assuming independence between CRISPRi perturbations, we constructed a Null
150 model for pairwise knockdowns by multiplicatively combining the growth rate effects of two single
151 gene knockdown functions (Equation 2).

$$g(R_1, R_2) = g_1(R_1) * g_2(R_2) \tag{2}$$

152

153 This initial model enabled continuous prediction of growth rate following pairwise gene
154 knockdown, but it did not explicitly account for the possibility of gene-gene epistasis. To introduce
155 epistasis into this modeling framework, we fit coupling constants a_{ij} and a_{ji} to modulate the effective
156 repression of a given gene (R_{ieff}) based on the intensity of a secondary perturbation (R_{jeff}) and vice
157 versa, as in Zimmer et al. (Equation 3, 28).

$$R_{ieff} = R_i \prod_{j \neq i} \left(1 + a_{ij} \frac{R_{jeff}/R_{0j}}{1 + R_{jeff}/R_{0j}}\right)^{-1} \tag{3}$$

158

159 Based on the assumption of sparse gene-gene coupling, we introduced a regularization
160 term to penalize the absolute values of a_{ij} and a_{ji} (Methods). Using mild regularization, we reduced
161 the spread of coupling constant values while retaining our model's predictive power (Fig S5, Table
162 S4). The resulting epistatic model effectively predicted growth rates following knockdown of all gene

163 pairs, as quantified by the root-mean-square deviation (RSMD) between the model's predictions
164 and experimentally determined pairwise growth rates (Fig 4A). These decreases in RMSD
165 (visualized as a point falling below y = x) illustrated that our model outperformed the nonepistatic
166 Null model for most gene pairs, but we wanted to ensure that these improvements were sufficient
167 to warrant the additional complexity introduced by fitting coupling constants. Thus, we computed
168 the Akaike Information Criterion (AIC) for each model, which supported the inclusion of these
169 additional parameters (Null AIC: -13,580, Model AIC: -15,114). We refer to this regularized,
170 coupling-sensitive model as the "continuous epistasis" model.

171 To investigate these pairwise expression-growth rate landscapes further, here we consider
172 several example gene pairs. In the background of reduced expression of *dapA*, the growth rate
173 effect of *purN* repression is reduced, and vice versa. This positive epistatic interaction is well-
174 described by a pair of positive gene-gene coupling terms (a_{ij} = 0.061, a_{ji} = 0.197, Fig 4B). Notably,
175 the couplings between *dapA*/*purN* and *purN*/*dapA* are asymmetric: in this instance, it reflects the
176 fact that *purN* repression more strongly modulates the effect of *dapA* repression than the converse.
177 In contrast, the synthetic lethal *gdhA* and *gltB* show negative epistasis, which is captured with one
178 negative gene-gene coupling term and one zero coupling term (a_{ij} = 0, a_{ji} = -0.425, Fig 4C). In this
179 case, the apparent coupling directionality for *gdhA*/*gltB* is the result of our regularization strategy.
180 We observed that other sufficiently negative coupling constants (and swapping which coupling
181 constant is zero) fit the data with comparable accuracy. The gene pair *dapA*/*dapB* shows positive
182 coupling, which we infer is because their gene products catalyze consecutive metabolic reactions
183 (a_{ij} = 0.361, a_{ji} = 0.327, Fig 4D). Metabolic flux is limited by only a single knockdown, and the effect
184 of an additional knockdown is dampened by this first flux restriction. However, we see almost no
185 interaction between *purN* and *purL*, even though the products of these genes are also consecutive
186 (a_{ij} = 0.053, a_{ji} = 0.009, Fig 4E). Thus, the continuous epistasis model summarizes gene-gene
187 coupling across entire pairwise expression-growth rate landscapes without incorporating bias from
188 prior expectations.

189 Next, we sought to understand the relationship between our approach and more standard
190 "binary" genetic interaction measurements. We compared the gene-gene coupling values (a_{ij}/a_{ji}) fit

191 using our model to epistatic couplings calculated from a Bliss model considering only maximal-
192 strength knockdown, mimicking a classical double-knockout study (Fig 4F-G). We find that
193 coupling, when considered over an entire pairwise expression-growth rate landscape, is vastly less
194 common than coupling between extreme perturbations alone. This insight highlights how seemingly
195 disparate epistasis measurements can be concisely summarized when treated as a continuous
196 landscape.
197
198 **The expression-growth rate model can be reliably inferred using a small number of**
199 **measurements**
200 A key strength of this modeling strategy is its ability to represent an entire pairwise expression-
201 growth rate landscape using a short list of parameters (n and R_0 for each of two genes, a_{ij}, and a_{ji}).
202 However, if the pairwise landscape must be densely sampled to reliably fit these parameters, a
203 predictive model will be of little additional use. As such, we evaluated the effect of subsampling our
204 pairwise expression data on our coupling constant fits. To accomplish this, we randomly removed
205 a subset of our pairwise expression-growth rate data from our training set, fit our model using the
206 remaining data, and evaluated the model's growth rate predictions on our holdout data over 100
207 subsampling iterations (Fig S6A). We determined that the model is reliably fit with 20% of our
208 original input data with little loss of prediction quality, indicating that future experiments can be
209 performed more efficiently with a compact training set (Fig 4H, Fig S6B, Fig S7). These compact
210 libraries could be designed intentionally by reducing the number of sgRNAs targeting each gene,
211 or randomly by bottlenecking a dense CRISPRi library.
212
213 **High-order predictions can be made using low-order training data**
214 We next wanted to evaluate whether a model trained on only single and pairwise expression-growth
215 rate data could predict the effects of higher-order CRISPRi perturbations. This would vastly dampen
216 the effect of the combinatorial explosion problem, enabling more complex expression-growth rate
217 predictions. As such, we constructed a compact third-order sgRNA library and quantified growth
218 rate following these third-order CRISPRi perturbations (Fig 5A). The library was designed to include

219 a number of epistatic interactions from our pairwise dataset, with the goal of enriching for nontrivial
220 gene-gene interactions. The library contained 225 distinct constructs with variable numbers of
221 targeting sgRNAs (Fig 5B, Table S5). We established that single and pairwise CRISPRi growth
222 rates measured in this experiment were well-correlated to those from our pairwise data set
223 (RMSD = 0.132, Fig S8). While growth rates across constructs from the third-order experiment
224 were globally slightly faster than those from our pairwise experiment, we have reasonable
225 experiment-to-experiment concordance.

226 We used the gene-gene coupling values fit on our pairwise data to predict growth rates
227 following all third-order CRISPRi perturbations. These predictions outperformed the coupling-
228 insensitive Null model (RMSD = 0.154 vs RMSD = 0.247 for the epistatic and Null models,
229 respectively, Fig 5C-D). This showed that our model's predictions can extend to higher-order
230 expression perturbations without an exponential increase in training. By truncating epistasis
231 measurements at pairwise interactions, future studies of gene-gene epistasis will be able to span
232 entire metabolic or transcriptional networks using CRISPRi libraries comparable to those used in
233 this work.

234

235 **The continuous epistasis model captures gene-by-environment interactions**

236 Importantly, the continuous epistasis model is unit-agnostic — in principle, couplings need not be
237 limited to gene-gene epistasis, but might more generally describe interactions between genes,
238 environments, and other "continuous" perturbations. To assess the ability of the model to capture
239 gene-by-environment interactions, we used mismatch CRISPRi to repress transcription of the
240 folate metabolic genes *folA* and *thyA* while supplementing culture growth media with two products
241 of folate metabolism: thymidine and/or methionine (Figure 6A, red text). Previous studies
242 demonstrated that these perturbations have higher-order growth rate interactions. In our own
243 prior work, we observed that *folA* and *thyA* form an environmentally-dependent synthetic rescue.
244 That is, loss of function mutations to DHFR (the gene encoded by *folA*) are rescued by loss of
245 function mutations to TYMS (the gene encoding *thyA*) in the presence (but not the absence) of
246 thymidine (40). Moreover, the balance between thymidine starvation and amino acid limitation

247 influences whether DHFR inhibition by the antibiotic trimethoprim results in growth inhibition in the
248 absence of thymidine (41, 42). Thus, these combined genetic and environmental perturbations
249 serve as an excellent test of our model's ability to leverage pairwise couplings to predict high-
250 order phenotypes.

251 To construct our model of gene-by-environment interactions, we first quantified the effect
252 of media supplementation on bacterial growth rate in the absence of gene expression perturbations.
253 To do this, we measured growth rates for *E. coli* harboring a non-targeting CRISPRi control in a
254 titrated range of thymidine (0, 0.05, 1, 2, 5, 10, and 50 ng/μL) and methionine (0, 0.01, 0.02, 0.05,
255 0.1, 0.3, and 1 mM) concentrations, as well as select combinations of both additives. For wildtype
256 *E. coli* in the absence of gene repression, the addition of methionine to growth media accelerated
257 growth, while the addition of thymidine had a minimal effect (Fig S9). These data were used to
258 model the growth rate effect of each individual environment. Next, to investigate gene-by-
259 environment interactions within folate metabolism, we created a compact mismatch CRISPRi
260 library containing 110 pairwise sgRNA combinations targeting *folA* and *thyA*. We transformed this
261 library into our CRISPRi strain and grew the transformed population in a continuous culture device
262 that allowed for parallel control of multiple supplemented media conditions (Fig 6B, 36, 43). We
263 again quantified relative growth rates following CRISPRi by next-generation sequencing and
264 rescaled these relative growth rates such that the growth rate of *E. coli* harboring the non-targeting
265 CRISPRi control in unsupplemented media was equal to one. We then identified and removed
266 escapers (125 measurements, 0.9% of data, Fig S10) and averaged across the remaining growth
267 rate replicates, yielding 2,394 unique growth rate measurements across 22 growth conditions (Fig
268 6C, Table S6). This generated a rich dataset to train and evaluate the continuous epistasis model.

269 We separated our pairwise (second-order) perturbation data into training and test sets, and
270 fit single perturbation growth curves and pairwise couplings between perturbations as with our other
271 data (Table S7). We found that pairwise growth rate predictions again worked well on this
272 independent test set, with an epistatic model RMSD of 0.185 and a Null RMSD of 0.291. As with
273 our gene-gene model, subsampling as little as 20% of our gene-by-environment data was sufficient
274 to reliably fit the continuous epistasis model (RMSD = 0.200, Fig S11A). Thus, we see that our unit-

275 agnostic model can capture not only expression-growth rate interactions, but also gene-by-
276 environment and environment-by-environment interactions.

277 Next, we evaluated the ability of our model — trained exclusively on second-order
278 perturbations — to predict the growth rate effect of third- and fourth-order perturbations. The
279 continuous epistasis model strongly outperformed the Null model on these higher-order predictions:
280 including pairwise couplings yielded a prediction RMSD of 0.303 between the model and the data,
281 while Null model predictions had a RMSD of 0.590. To investigate where our model's accuracy
282 gains emerged, we separated our data by genetic and environmental perturbations and plotted
283 each model's error (Fig 6D). The continuous epistasis model outperformed the Null model the most
284 when *thyA* was repressed and thymidine was added to the media. As thymidine replaces the key
285 metabolite made by the *thyA* gene product, growth is rescued, resulting in a positive pairwise
286 coupling (a_{ij} = 0, a_{ji} = 0.721). This asymmetric coupling is indicative of a hierarchy between
287 thymidine and *thyA* expression, where the addition of thymidine to growth media fully decouples
288 growth rate from *thyA* expression. This epistasis, as well as other gene-gene and gene-by-
289 environment interactions, are captured even after subsampling as little as 20% of our pairwise input
290 data (RMSD = 0.322, Fig S11B). We performed subsampling using the same regularization used
291 in our pairwise study, supporting the generalizability of this mild regularization (Fig S12). These
292 results indicate that the continuous epistasis model can be applied beyond low-order perturbation
293 studies to predict the combined effects of complex gene-by-environment and environment-by-
294 environment interactions. This now opens the door to creating continuous, coupling sensitive
295 pharmacogenomic models that describe the interaction between drug, genotype, and environment.
296

297 **Discussion**

298 In this work, we generated and trained a continuous, coupling-sensitive model relating gene
299 expression to growth rate using sparsely sampled single and pairwise perturbation measurements.
300 The model's parameters are interpretable and relevant across gene expression levels, and this
301 unitless approach enables the investigation of context-dependence between disparate
302 perturbations, such as CRISPRi and metabolite supplementation. We observed that gene-gene

303 epistasis, when viewed across a continuous expression landscape, is less common than epistasis
304 observed at the level of knockout or extreme knockdown. In addition, while it is possible to infer
305 why some gene pairs show positive or negative coupling based on prior knowledge, we see that
306 biochemical pathway structure is not sufficient to predict all gene-gene interactions, and it does not
307 easily offer a clear perspective on coupling strength. Certainly, the relationship between coupling
308 strength, asymmetry, and functional interactions between enzymes will be an interesting avenue
309 for further study.

310 While we limited our experiments to transcriptional and environmental perturbations in this
311 work, one can imagine integrating proteomic, metabolomic, and other measurements into this
312 unitless modeling framework. We modeled the growth rate effect of each perturbation with a
313 monotonic function, however there are known examples of nonmonotonic growth rate relationships,
314 limiting some applications of the model as it is currently constructed (10). Finally, while we showed
315 that accounting for pairwise coupling alone captured most variation in our third- and fourth-order
316 perturbation landscapes, sufficiently strong high-order epistasis in a system could require a
317 different approach.

318 To expand this workflow, the limiting experimental step is quantifying the repression
319 strength of individual sgRNAs. RT-qPCR is relatively low-throughput, precluding genome-level
320 studies. An alternative, single-cell approach to quantify CRISPRi repression, such as microSPLiT
321 (bacterial cells, 44), PETRI-seq (bacterial cells, 45) or perturb-seq (mammalian cells, 46) alongside
322 next-generation sequencing-based growth rate measurements would remove these throughput
323 restrictions and permit significantly larger CRISPRi studies. Alternatively, one could utilize a
324 predictive model relating sgRNA homology sequence to CRISPRi repression strength (11, 31).

325 Applications of this continuous, epistatic model are numerous. The reduced sampling
326 requirements make pathway-level epistatic analyses feasible in single experiments. Prior work has
327 indicated that metabolic enzymes are often produced in evolutionarily conserved stoichiometric
328 ratios (47). The experiments presented here could be used to better understand the constraints on
329 enzyme abundance in metabolic systems that shape these ratios. The continuous epistasis model
330 could also be applied to optimize biosynthetic pathway performance and/or cell growth conditions.

331 As the approach is flexible and unitless, it could be used to quantify interactions between
332 expression variation and metabolite concentrations to improve model systems describing metabolic
333 disorders. We hope that interpretable, continuous, and coupling-sensitive models like the one
334 described in this work can continue to improve computational predictions to extract more
335 information from the data-rich -omics experiments we now regularly perform.
336

337 **Materials and Methods**

338 **Pairwise CRISPRi Library Construction**

339 We designed a pool of CRISPRi sgRNAs targeting nine *E. coli* genes. For each gene, we selected
340 9-12 distinct sgRNAs to provide a titrated range of mRNA knockdown. We previously characterized
341 CRISPRi knockdown of seven of these genes (*dapA*, *dapB*, *folA*, *thyA*, *glyA*, *purN*, and *purL*) (32),
342 giving us an initial list of sgRNAs for each gene with mutually distinguishable growth rate effects.
343 We supplemented these sgRNAs with up to two fully on-target sgRNAs per gene and added
344 sgRNAs by hand to populate underrepresented growth regimes. The final two genes (*gdhA* and
345 *gltB*) were previously uncharacterized by our lab. We designed two on-target sgRNAs for each
346 gene using previously published design principles. In addition, we introduced 4, 6, 8, 10, and 14
347 mismatches into the homology region of each on-target sgRNA. For a full list of sgRNA sequences,
348 see Table S1.

349 All sgRNAs were synthesized as a mixed oligo pool by Integrated DNA Technologies. We
350 resuspended the library to a final concentration of 50 nM in MB H_2O and added BsaI flanking sites
351 using PCR. We performed two separate PCRs to add two distinct sets of flanking sites onto the
352 library for Golden Gate cloning. We used the GG_UnivF/GG_R1 primers to generate the insert 1
353 library and the GG_F2/GG_UnivR primers to generate the insert 2 library (for primer sequences,
354 see Table S8). The PCR mix contained 10.75 µL MB H_2O, 5 µL Q5 Reaction Buffer (NEB), 5 µL
355 High GC Enhancer (NEB), 0.5 µL 10 mM dNTPs, 1.25 µL 10 µM forward primer, 1.25 µL 10 µM
356 reverse primer, 0.25 µL Q5 High-Fidelity DNA Polymerase (NEB, #M0491), and 1 µL of the 50 nM
357 oligo pool. Initial denaturation was carried out at 98°C for 30 seconds, followed by 25 cycles of
358 98°C for 30 seconds, 58°C for 10 seconds, and 72°C for 20 seconds. A final elongation was

359 performed at 72°C for 60 seconds, followed by an indefinite hold at 4°C. We ran the PCR product
360 on a 1% agarose 1X TAE gel with EtBr staining at 80 V for 45 minutes and excised the 176 bp
361 band. We extracted the band using a DNA Clean & Concentrator-5 Kit (Zymo, #D4014), eluted in
362 30 µL MB H_2O, and quantified dsDNA concentration using a PicoGreen assay (Thermo, #P7581).
363 To ensure representation of the non-targeting sgRNA sequence (Nont) in the final library, we
364 isolated this control sgRNA using the same protocol illustrated above, replacing the 50 nM oligo
365 pool with 1 µL of purified 20-50 ng/µL plasmid containing the Nont sgRNA (pCRISPR2, 32). We
366 spiked Nont inserts into each of the mixed libraries at a 1:100 Nont:library ratio.

367 We performed three Golden Gate cloning reactions to ligate the CRISPRi inserts into three
368 distinct barcoded CRISPRi vectors (pCRISPR3, AddGene ID: XXX). The reaction mix contained
369 75 ng of barcoded vector, 6 ng of insert 1 mix, 6 ng of insert 2 mix, 1.5 µL of 10X T4 DNA Ligase
370 Buffer (NEB), 1 µL T4 DNA Ligase (2,000,000 U/mL, NEB, #M0202M), 1 µL BsaI-HF v2 (NEB,
371 #R3733), and MB H_2O to 15 µL. Reaction conditions were 40 cycles of 37°C for 2 minutes and
372 16°C for 3 minutes. These cycles were followed by 50°C for 10 minutes, 80°C for 20 minutes, and
373 an indefinite 4°C hold. A no-insert control, where both inserts were left out of the reaction, was
374 performed identically.

375 We cleaned and concentrated each reaction using a DNA Clean & Concentrator-5 Kit
376 (Zymo, #D4014), eluting in 20 µL MB H_2O. We transformed 2 µL of each purified library into 100
377 µL of XL1 Blue electrocompetent cells followed by a 1 hour recovery at 37°C with shaking. We
378 plated 50 µL of a 1:10 and 1:100 dilution to quantify library coverage and added the remaining cells
379 to 9 mL SOB + 35 µg/mL kanamycin. We incubated this culture overnight at 37°C with shaking. We
380 also transformed and plated the no-insert control and a no DNA control and adjusted library
381 coverage estimates accordingly. Library coverage ranged from 10-14x for each barcoded vector.
382 In addition to these mixed CRISPRi libraries, we synthesized constructs containing two Nont
383 sgRNAs in six distinct barcoded vectors using the same protocol described above, substituting
384 each of the mixed insert libraries for 6 ng of purified Nont sgRNA insert.

385 We isolated plasmids using the GeneJET Plasmid Miniprep Kit (Thermo, #K0503), eluted
386 DNA in 50 µL MB H_2O, and quantified dsDNA concentration with a PicoGreen assay (Thermo,

387 #P7581). We combined the three barcoded CRISPRi libraries in an equimolar ratio. We also mixed
388 the Nont+Nont controls in an equimolar ratio, then spiked this control into the library at a 1:5,000
389 Nont+Nont:library ratio. We transformed 3 ng of this final mixture into 45 µL MG1655 +
390 chromosomal dCas9 electrocompetent cells (gift from the Bikard lab) using the same
391 transformation protocol as above, plating 50 µL of 1:100 and 1:1,000 dilutions. We performed a no
392 DNA control with zero breakthrough colonies. Coverage exceeded 200x, and we grew the
393 remaining cells in 9 mL SOB + 35 µg/mL kanamycin overnight at 37°C with shaking.

394 The pairwise *folA*/*thyA* CRISPRi library was constructed as described above, with only
395 Nont, *folA*-targeting, and *thyA*-targeting sgRNAs included in the Integrated DNA Technologies-
396 synthesized oligo pool. Library coverage calculated after the first transformation event in the three
397 barcoded vectors ranged from 14x to 89x. We mixed three barcoded Nont+Nont CRISPRi
398 constructs in equimolar ratios and spiked them into the *folA/thyA* library at a 1:400
399 Nont+Nont:library ratio. As two continuous culture experiments were required to investigate all
400 environments of interest, we freshly transformed the library two separate times. For each
401 transformation, we added 1 ng of mixed library to 45 µL MG1655 + dCas9 electrocompetent cells
402 and achieved 583x and 127x library coverage.

403

404 **Continuous culture**

405 Following transformation of the MG1655 dCas9 electrocompetent *E. coli*, cells were grown
406 overnight in LB + 35 µg/mL kanamycin. We spun saturated outgrowth cultures at 1,000 rcf for 5
407 minutes, decanted media, and resuspended cells in 1 mL M9 minimal media pH 6.5, 0.4% glucose,
408 35 µg/mL kanamycin (M9+Kan). We repeated this wash, then resuspended 10 µL of concentrated
409 cells in 8 mL M9+Kan. We incubated this culture for 12 hours at 37°C with shaking.

410 Following outgrowth, we spun the culture at 1,000 rcf for 5 minutes, decanted, and
411 resuspended cells in 1 mL fresh M9+Kan. We repeated this wash, diluted the culture to OD_{600} =
412 0.05 in 15 mL M9+Kan, and grew these cells in a continuous culture device at 37°C, clamping the
413 optical density at 0.15, with a spin rate of 2. Our continuous culture device follows the design
414 specifications from 36 and 43. Following an overnight adaptation, we added anhydrotetracycline

415 (Cayman Chemical Company, #10009542) to a final concentration of 50 ng/mL to induce CRISPRi.
416 We took the first timepoint (T_0) after 3 hours of CRISPRi by extracting 1 mL of cells from the culture
417 and took additional timepoints every 2 hours until T_{14}. We centrifuged timepoints at 3,000 rcf for 5
418 minutes to pellet cells, decanted by pipetting, and stored them at -20°C.

419 We performed continuous culture for the third-order CRISPRi and environmental variation
420 experiments with slight modifications. We terminated these experiments after T_{10}, as the final two
421 timepoints were shown to be unnecessary for growth rate quantification (Fig S13). We inoculated
422 the third-order CRISPRi library into the continuous culture device at a starting OD_{600} = 0.2. Media
423 conditions were altered during the *folA*/*thyA* gene-by-environment experiment. Thymidine (Sigma,
424 #T1895) and/or methionine (Sigma, #M5308) were supplemented into M9+Kan to create the
425 following conditions: No additive, 0.05 ng/μL thymidine, 1 ng/μL thymidine, 2 ng/μL thymidine, 5
426 ng/μL thymidine, 10 ng/μL thymidine, 50 ng/μL thymidine, 0.01 mM methionine, 0.02 mM
427 methionine, 0.05 mM methionine, 0.1 mM methionine, 0.3 mM methionine, 1 mM methionine, 0.05
428 ng/μL thymidine + 0.01 mM methionine, 0.05 ng/μL thymidine + 0.05 mM methionine, 0.05 ng/μL
429 thymidine + 0.1 mM methionine, 1 ng/μL thymidine + 0.01 mM methionine, 1 ng/μL thymidine +
430 0.05 mM methionine, 1 ng/μL thymidine + 0.1 mM methionine, 2 ng/μL thymidine + 0.01 mM
431 methionine, 2 ng/μL thymidine + 0.1 mM, and 5 ng/μL thymidine + 0.01 mM methionine.
432

433 **Sample preparation and next-generation sequencing**

434 We resuspended each frozen cell pellet (each corresponding to a sample – one time point and
435 environmental condition) in 100 μL MB H_2O with vortexing and lysed them at 95°C for 3 minutes.
436 Following a 10 minute, 20,000 rcf centrifugation to pellet debris, we took a 50 μL working stock for
437 use in downstream reactions.

438 We then added adaptors for Illumina sequencing in two consecutive PCR steps. First, we
439 added flanking sequences to each timepoint. The reaction mix was composed of 10.5 μL MB H_2O,
440 5 μL Q5 Reaction Buffer (NEB), 5 μL 50% glycerol, 0.5 μL 10 mM dNTPs, 1.25 μL 10 μM TruSeqF
441 primer, 1.25 μL 10 μM TruSeqR_trunc primer, 0.5 μL Q5 High-Fidelity DNA Polymerase (NEB,
442 #M0491), and 1 μL timepoint working stock. Reaction conditions were 98°C for 30 seconds, 7

443 cycles of 98°C for 30 seconds, 61°C for 10 seconds, and 72°C for 15 seconds, followed by a final
444 elongation at 72°C for 45 seconds and an indefinite hold at 4°C.
445 We added Illumina sequencing adapters and i5/i7 barcoding sequences to each timepoint in a
446 second PCR. The reaction mix contained 10.5 µL MB H_2O, 5 µL Q5 Reaction Buffer (NEB), 5 µL
447 50% glycerol, 0.5 µL 10 mM dNTPs, 1.25 µL 10 µM i5 primer, 1.25 µL 10 µM i7 primer, 0.5 µL Q5
448 High-Fidelity DNA Polymerase (NEB, #M0491), and 1 µL template from the previous reaction.
449 Reaction conditions were 98°C for 30 seconds, 20 cycles of 98°C for 30 seconds, 61°C for 10
450 seconds, and 72°C for 20 seconds, followed by a final elongation at 72°C for 60 seconds and an
451 indefinite hold at 4°C.
452 Note: While the glycerol was used in the preparation of timepoints for the pairwise CRISPRi library
453 experiment and the *folA*/*thyA* sublibrary experiment, we have observed future amplification
454 reactions that failed in the presence of glycerol.
455 We quantified dsDNA concentration from each reaction using a PicoGreen assay (Thermo,
456 #P7581), and mixed timepoints in an equimolar ratio to create the mixed sequencing library. We
457 ran the library on a 1% agarose 1X TAE gel with EtBr staining at 90 V for 60 minutes and excised
458 the 486 bp construct. We performed gel extraction using the DNA Clean & Concentrator-5 Kit
459 (Zymo, #D4014) and eluted in 20 µL MB H_2O. Following a Qubit Assay (Thermo, #Q32851) to
460 quantify final dsDNA quantification, we submitted the library to Azenta for Illumina HiSeq
461 Sequencing on a 300-cycle paired-end run.
462
463 **Isolation of single CRISPRi knockdown strains**
464 To quantify the concentration of mRNA following each individual CRISPRi perturbation, we isolated
465 plasmids expressing a single targeting sgRNA paired with the Nont sgRNA. For this purpose, we
466 constructed a CRISPRi:Nont library using Golden Gate cloning as described previously, adding the
467 Nont sgRNA insert in place of the mixed library in the second sgRNA position. We screened
468 individual colonies from this mixed library to isolate the majority of CRISPRi:Nont constructs.
469 We individually synthesized plasmids not isolated during this screen using inverse PCR
470 (iPCR). First, we phosphorylated iPCR primers with a PNK reaction. The reaction mix consisted of

471 6.5 μL MB H_2O, 1 μL 10X T4 DNA Ligase Buffer (NEB), 1 μL 100 μM iPCR_sgRNA_F primer, 1 μL
472 100 μM iPCR_UnivR primer, and 0.5 μL T4 PNK (NEB, #M0201S). Thermocycler conditions were
473 37°C for 30 minutes, 65°C for 20 minutes, and an indefinite 4°C hold. Then, we performed iPCR
474 with a reaction mix of 8.25 μL MB H_2O, 5 μL Q5 Reaction Buffer (NEB), 0.5 μL 10 mM dNTPs, 0.5
475 μL Q5 High-Fidelity DNA Polymerase (NEB, #M0491), 10 μL primer phosphorylation reaction, and
476 1 μL 30-50 ng/μL template plasmid (pCRISPR2). Reaction conditions consisted of 98°C for 30
477 seconds, followed by 25 cycles of 98°C for 30 seconds, 58°C for 30 seconds, and 72°C for 90
478 seconds. A final 3-minute elongation at 72°C preceded an indefinite hold at 4°C. We cleaned and
479 concentrated the reaction product using a DNA Clean & Concentrator-5 Kit (Zymo, #D4014) and
480 eluted in 25 μL MB H_2O. We removed the template plasmid using a DpnI digestion. Reaction
481 conditions were 13 μL cleaned and concentrated product, 2.5 μL 10x Cutsmart Buffer (NEB,
482 #B6004), 8.25 μL MB H_2O, and 1.25 μL DpnI (NEB, #R0176). The reaction was held at 37°C for 1
483 hour, then held at 4°C. Finally, we circularized the PCR product in a ligation reaction consisting of
484 8 μL DpnI product, 1 μL 10X T4 DNA Ligase Buffer (NEB), and 1 μL T4 DNA Ligase (400,000
485 U/mL, NEB, #M0202S). Ligation occurred at 25°C for 2 hours, followed by 10 minutes at 65°C and
486 a hold at 4°C. We transformed this product into XL1 Blue chemically competent cells (1 μL DNA,
487 45 μL competent cells) with 1 hour recovery at 37°C with shaking. We purified plasmids using a
488 GeneJET Plasmid Miniprep Kit (Thermo, #K0503) and transformed into CRISPRi MG1655 + dCas9
489 cells as described previously.
490
491 **RNA Extraction and RT-qPCR**
492 To collect RT-qPCR data following CRISPRi, we grew CRISPRi-Nont strains from glycerol stocks
493 in 4 mL LB + 35 μg/mL kanamycin for 6-8 hours at 37°C with shaking. We washed these cultures
494 into M9+Kan as described previously and incubated them for 8-14 hours at 37°C with shaking. We
495 then washed cultures into M9+Kan+50 ng/mL anhydrotetracycline (Cayman Chemical Company,
496 #10009542) and incubated them for 5 hours at 37°C with shaking at a starting OD_{600} = 0.05. After
497 treatment, we isolated RNA using the Qiagen RNeasy Protect Bacteria Mini Kit (QIAGEN, #74524),
498 with on-column DNase digestion (QIAGEN, #79254), as described in their protocol. We extracted

499 RNA from control Nont+Nont strains using the same procedure. We performed RT-qPCR on a CFX
500 Opus 384 Real-Time PCR System (Bio-Rad, #12011452) using the Luna Universal One-Step RT-
501 qPCR Kit (NEB, #E3005). No template and gDNA controls were included in all experiments. We
502 analyzed raw results using the CFX Maestro Software (Bio-Rad), and calculated changes in mRNA
503 concentration using the $\Delta\Delta C_t$ method with the *hcaT* gene as reference in a custom Python script.
504

505 **Third-order sgRNA library construction and next-generation sequencing**

506 We constructed the third-order sgRNA library in six unique barcoded vectors using Golden Gate
507 cloning (pCRISPR3, AddGene ID: XXX). We individually amplified sgRNA inserts using the same
508 strategy used to generate Nont inserts. We amplified sgRNAs targeting *dapA* and *gdhA* and the
509 Nont sgRNA using GG_UnivF and GG_R1. We amplified sgRNAs targeting *dapB* and *gltB* and the
510 Nont sgRNA using GG_F2 and GG_R2. Finally, we amplified sgRNAs targeting *folA*, *thyA*, *purN*,
511 and *purL* and the Nont sgRNA using GG_F3 and GG_UnivR. We mixed all inserts with shared
512 amplification primers in equimolar ratios to generate three final insert mixes. Golden Gate cloning
513 conditions were 75 ng barcoded vector, 6 ng insert 1 mix, 6 ng insert 2 mix, 6 ng insert 3 mix, 2 µL
514 10X T4 DNA Ligase Buffer (NEB), 1 µL T4 DNA Ligase (2,000,000 U/mL, NEB, #M0202M), 1 µL
515 BsaI-HF v2 (NEB, #R3733), and MB H_2O to 20 µL. Reaction conditions were adjusted to 60 cycles
516 of 2 minutes at 37°C and 3 minutes at 16°C. These cycles were followed by 10 minutes at 50°C,
517 20 minutes at 80°C, and an indefinite 4°C hold. We transformed libraries and estimated library
518 coverage as described previously. Library coverages ranged from 15x – 62x. To ensure the
519 presence of Nont control constructs, we individually generated Nont+Nont+Nont constructs in six
520 barcoded vectors. We spiked the control plasmid mix into the library at a 1:225
521 Nont+Nont+Nont:library ratio. Transformed library coverage was 964x.

522 We altered the sampled preparation protocol to accommodate the increased amplicon size
523 of the third-order sgRNA CRISPRi library. For the first PCR, we used a reverse primer with 4
524 additional base pairs of homology (TruSeqR). The reaction conditions were changed to 30 seconds
525 at 98°C, 17 cycles of 30 seconds at 98°C and 60 seconds at 76°C, followed by 180 seconds at
526 72°C and an indefinite 4°C hold. In the second PCR, we altered thermal cycler conditions to 30

527 seconds at 98°C, 10 cycles of 30 seconds at 98°C and 75 seconds at 72°C, followed by 180
528 seconds at 72°C and an indefinite 4°C hold. We quantified dsDNA from these reactions using a
529 Qubit Assay (Thermo, #Q32851), mixed them in an equimolar ratio, and concentrated DNA using
530 a DNA Clean & Concentrator-5 Kit (Zymo, #D4014) with a 30 µL MB H_2O elution. We ran this
531 concentrated library on a 2% agarose 1X TAE gel with an EtBr stain at 90 V for 85 minutes and
532 extracted the resulting 636 bp band using a DNA Clean & Concentrator-5 Kit (Zymo, #D4014) with
533 a 20 µL MB H_2O elution. We quantified dsDNA with a Qubit Assay (Thermo, #Q32851) and
534 sequenced using an Illumina MiSeq Nano v2 500-cycle paired-end run.
535
536 **Plate reader growth rate assay**
537 We cultured *E. coli* strains harboring a non-targeting CRISPRi control overnight in 4 mL LB + 35
538 µg/mL kanamycin at 37°C with shaking. The next day, we washed these cultures twice into M9+Kan
539 as described previously. We diluted cultures 1:200 into each of the 22 media conditions described
540 in Continuous Culture and adapted cells for 4 hours at 37°C with shaking. Following adaptation,
541 we washed cells into 4 mL of their respective supplemented media + 50 ng/mL anhydrotetracycline
542 at OD_{600} = 0.005. Three 200 µL replicates of each culture were grown in a Synergy Neo2 Hybrid
543 Multi Mode Reader (Agilent, #BTNEO2) overnight at 37°C while taking regular OD_{600}
544 measurements. Growth rates were fit across replicates using a log-linear fit of exponential phase
545 growth and averaged across triplicate measurements.
546
547 **Calculating CRISPRi growth rates**
548 We wrote all analysis code in Python 3.9.12 using Jupyter Notebook (available on github at
549 https://github.com/reynoldsk/ContinuousEpistasis). We extracted sgRNA sequences from Illumina
550 FASTQ files using regular expressions, searching for a 20 bp region between flanking sequences
551 CTAGCTCTAAAAC and A. Up to 3 mismatches and no gaps were permitted in the flanking
552 sequences. The identified homology region was required to match a desired sequence from the
553 CRISPRi library. In addition, to increase confidence, we imposed a filter requiring two specific base
554 calls in each sequence to have a Q-score above 30. The filtered bases were the final "on-target"

555 and first “mismatch” base in the homology region, as this two-nucleotide combination is unique to
556 a single sgRNA within each sgRNA family. We identified the plasmid barcode similarly, with flanking
557 sequences GTACAGCGAGGCAAC and ACGGATCCCCAC, imposing a maximum of 3
558 mismatches and allowing no gaps. The barcode sequence needed to match a target barcode
559 exactly, and no Q-score filter was imposed.

560 For the three-sgRNA library, the following modifications were made. As sequencing quality
561 decreases for a 500-cycle kit, the sgRNA flanking sequences were extended to CTAGCTCTAAAAC
562 and ACTAGTATTATAC. These flanking sequences were permitted to have up to 10 mismatches
563 for sgRNAs 1 and 2, and 6 mismatches for sgRNA 3. Flanking sequences for the barcode were the
564 same as above and were permitted to have up to 6 mismatches. The barcode itself was required
565 to match a target exactly, however the cutoff to call a specific sgRNA homology region was lowered.
566 Sequences in position 1 were allowed to have as many as 2 mismatches, and sequences in position
567 2 were allowed to have as many as 8 mismatches. This permissive filter was required due to the
568 sharp decrease in sequencing confidence at the end of the sequencing run. If a sequenced sgRNA
569 ambiguously mapped to multiple library sequences with this less stringent filter, no identity was
570 called. We performed next-generation sequencing analysis from HiSeq data on a high-performance
571 computing cluster, while all subsequent analysis was performed locally.

572 We calculated relative growth rates following each CRISPRi treatment using the change in
573 sgRNA counts over time. At each timepoint (t), we normalized counts for a specific barcoded
574 CRISPRi construct (sgRNA) to the control construct (Nont) and our initial timepoint (t0) using
575 Equation 4.

$$Relative\ frequency_{sgRNA,t} = \frac{counts_{sgRNA,t}/counts_{Nont,t}}{counts_{sgRNA,t0}/counts_{Nont,t0}} \qquad (4)$$

576

577 We then $\log_2$-transformed this relative frequency prior to growth rate fitting. For the pairwise sgRNA
578 libraries, if fewer than 10 counts were identified for a construct at a given timepoint, no future
579 timepoints were considered for that construct. Growth rates were calculated for CRISPRi constructs
580 with sufficient counts over at least the first three timepoints. For pairwise and third-order CRISPRi
581 experiments, the time axis was rescaled from hours to generations based on the overall culture

582 generation time in each continuous culture vial. We then fit a line to our $\log_2$(Relative frequency)
583 vs. generation data using scipy.stats.linregress and defined the slope of the best fit line as the raw
584 growth rate. At this step, two sgRNAs (gdhA_1_42_B_MM14 and gdhA_3_216_B_MM8) were
585 determined to have off-target effects and were removed from future analysis (see Jupyter Notebook
586 3_Pairwise_Growth_Rates.ipynb for details). For gene-by-environment analysis, continuous
587 culture growth rates for each media and sgRNA combination were rescaled using plate reader
588 absolute growth data using Equation 5.

$$Relative\ GR_{media,sgRNA} = \frac{Raw\ GR_{sgRNA,media} + Absolute\ GR_{media}}{Absolute\ GR_{M9}} \quad (5)$$

589

590 We then removed escapers, CRISPRi constructs that abnormally lacked a growth rate
591 deficit, using a one-way Dixon Q-test for outliers at 95% confidence. If at least four growth rates
592 were successfully calculated across barcodes and sgRNA orders after removing escapers, the
593 replicates were averaged to determine a mean relative growth rate. Finally, we normalized growth
594 rates so the fully Nont control had a growth rate of 1 and the most severe growth rate perturbation
595 observed (glyA_1_26_C + thyA_1_60_B_MM2, Min GR) had a growth rate of 0 using Equation 6.

$$Normalized\ GR = \frac{Relative\ GR - Min\ GR}{abs(Min\ GR)} \quad (6)$$

596

597

598 **Fitting the Coupling Model**

599 We fit single-gene expression-growth rate curves using scipy.optimize.least_squares, bounding R_0
600 to non-negative values with an initial guess of 0.5 for R_0 and 0 for n. We fit gene-gene coupling
601 values for each gene pair using the strategy described in Zimmer et al., 2016, with some
602 modifications (28). We optimized coupling constants a_{ij} and a_{ji} using scipy.optimize.least_squares,
603 bounding between -1 and 10 with an initial guess of 0 for both. To limit the spread of coupling
604 constant values, we calculated the RMSD between growth rate predictions and experimental
605 measurements with an added regularization term $\lambda * \Sigma abs(a_{ij})$. We used a regularization value of λ
606 $= 10^{-1.25}$ after evaluating a range of values on subsampled data for both the pairwise and *folA*/*thyA*

607 libraries (Fig S5 and Fig S12). We calculated pairwise and third-order predicted growth rates using
608 each gene's individual repression-growth rate parameters and relevant coupling constants.
609

610 **Acknowledgments**

611 The authors would like to thank Olivier Rivoire for insightful discussions and data analysis ideas.
612 We are grateful to Prashant Mishra for thoughtful comments on our manuscript. We also thank the
613 Reynolds lab for feedback on the project and manuscript. Research reported in this publication was
614 supported by the National Institute of General Medical Sciences of the National Institutes of Health
615 under award number R01GM136842. R.O. was also partly supported by a National Institutes of
616 Health training grant 5T32GM007062-46. This content is solely the responsibility of the authors and
617 does not necessarily represent the official views of the National Institutes of Health.

References

1. Q. Bernard *et al.,* Plasticity in early immune evasion strategies of a bacterial pathogen. *Proc. Natl. Acad. Sci.* **115**, E3788-E3797 (2018).
2. M. C. Abt, P. T. McKenney, E. G. Pamer, Clostridium difficile colitis: pathogenesis and host defence. *Nat. Rev. Microbiol.* **14**, 609–620 (2016).
3. F. Depardieu, I. Podglajen, R. Leclercq, E. Collatz, P. Courvalin, Modes and Modulations of Antibiotic Resistance Gene Expression. *Clin. Microbiol. Rev.* **20**, 79–114 (2007).
4. N. Woodford, M. J. Ellington, The emergence of antibiotic resistance by mutation. *Clin. Microbiol. Infect.* **13**, 5–18 (2007).
5. R. Sager, Expression genetics in cancer: Shifting the focus from DNA to RNA. *Proc. Natl. Acad. Sci.* **94**, 952–955 (1997).
6. A. C. Chiang, J. Massagué, Molecular Basis of Metastasis. *N. Engl. J. Med.* **359**, 2814–2823 (2008).
7. M. Uhlen *et al.,* A pathology atlas of the human cancer transcriptome. *Science* **357**, eaan2507 (2017).
8. A. Magen *et al.,* Beyond Synthetic Lethality: Charting the Landscape of Pairwise Gene Expression States Associated with Survival in Cancer. *Cell Rep.* **28**, 938-948.e6 (2019).
9. D. E. Dykhuizen, A. M. Dean, D. L. Hartl, Metabolic Flux and Fitness. *Genetics* **115**, 25–31 (1987).
10. L. Keren *et al.,* Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell* **166**, 1282-1294.e18 (2016).
11. J. S. Hawkins *et al.,* Mismatch-CRISPRi Reveals the Co-varying Expression-Fitness Relationships of Essential Genes in Escherichia coli and Bacillus subtilis. *Cell Syst.* **11**, 523-535.e9 (2020).
12. M. N. Price *et al.,* Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* **557**, 503–509 (2018).
13. M. M. Metzstein, G. M. Stanfield, H. R. Horvitz, Genetics of programmed cell death in C. elegans: past, present and future. *Trends Genet.* **14**, 410–416 (1998).

14. A. Typas *et al.,* High-throughput, quantitative analyses of genetic interactions in E. coli. *Nat. Methods* **5**, 781–787 (2008).

15. G. Butland *et al.,* eSGA: E. coli synthetic genetic array analysis. *Nat. Methods* **5**, 789–795 (2008).

16. M. Costanzo *et al.,* The Genetic Landscape of a Cell. *Science* **327**, 425–431 (2010).

17. M. Costanzo *et al.,* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420–aaf1420 (2016).

18. M. Jaffe *et al.,* Improved discovery of genetic interactions using CRISPRiSeq across multiple environments. *Genome Res.* **29**, 668–681 (2019).

19. F. J. Poelwijk, V. Krishna, R. Ranganathan, The Context-Dependence of Mutations: A Linkage of Formalisms. *PLOS Comput. Biol.* **12**, e1004771 (2016).

20. J. Domingo, P. Baeza-Centurion, B. Lehner, The Causes and Consequences of Genetic Interactions (Epistasis). *Annu. Rev. Genomics Hum. Genet.* **20**, 433–460 (2019).

21. K. Wood, S. Nishida, E. D. Sontag, P. Cluzel, Mechanism-independent method for predicting response to multidrug combinations in bacteria. *Proc. Natl. Acad. Sci.* **109**, 12254–12259 (2012).

22. K. B. Wood, K. C. Wood, S. Nishida, P. Cluzel, Uncovering Scaling Laws to Infer Multidrug Response of Resistant Microbes and Cancer Cells. *Cell Rep.* **6**, 1073–1084 (2014).

23. A. Zimmer, A. Tendler, I. Katzir, A. Mayo, U. Alon, Prediction of drug cocktail effects when the number of measurements is limited. *PLOS Biol.* **15**, e2002518 (2017).

24. I. Katzir, M. Cokol, B. B. Aldridge, U. Alon, Prediction of ultra-high-order antibiotic combinations based on pairwise interactions. *PLOS Comput. Biol.* **15**, e1006774 (2019).

25. A. Tendler, A. Zimmer, A. Mayo, U. Alon, Noise-precision tradeoff in predicting combinations of mutations and drugs. *PLOS Comput. Biol.* **15**, e1006956 (2019).

26. S. Pandey *et al.,* Pairwise Stimulations of Pathogen-Sensing Pathways Predict Immune Responses to Multi-adjuvant Combinations. *Cell Syst.* **11**, 495-508.e10 (2020).

27. J. Larkins-Ford *et al.,* Systematic measurement of combination-drug landscapes to predict in vivo treatment outcomes for tuberculosis. *Cell Syst.* **12**, 1046-1063.e7 (2021).

28. A. Zimmer, I. Katzir, E. Dekel, A. E. Mayo, U. Alon, Prediction of multidimensional drug dose responses based on measurements of drug pairs. *Proc. Natl. Acad. Sci.* **113**, 10442–10447 (2016).

29. L. S. Qi *et al.,* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).

30. J. M. Peters *et al.,* A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* **165**, 1493–1506 (2016).

31. M. Jost *et al.,* Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. *Nat. Biotechnol.* **38**, 355–364 (2020).

32. A. D. Mathis, R. M. Otto, K. A. Reynolds, A simplified strategy for titrating gene expression reveals new relationships between genotype, environment, and bacterial growth. *Nucleic Acids Res.* **49**, e6 (2021).

33. M. H. Larson *et al.,* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**, 2180-2196 (2013).

34. F. Rousset *et al.*, Genome-wide CRISPR-dCas9 screens in *E. coli* identify essential genes and phage host factors. *PLOS Genet.* **14**, e1007749 (2018).

35. T. Wang *et al*., Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. *Nat. Commun.* **9**, 2475 (2018).

36. E. Toprak *et al.,* Building a morbidostat: an automated continuous-culture device for studying bacterial drug resistance under dynamically sustained drug inhibition. *Nat. Protoc.* **8**, 555–567 (2013).

37. T. Baba *et al.,* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).

38. J.-P. Côté *et al.,* The Genome-Wide Interaction Network of Nutrient Stress Genes in Escherichia coli. *mBio* **7**, e01714-16 (2016).

39. W. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).

40. A. F. Schober *et al.,* A Two-Enzyme Adaptive Unit within Bacterial Folate Metabolism. *Cell Rep.* **27**, 3359-3370.e7 (2019).

41. Y. K. Kwon, M. B. Higgins, J. D. Rabinowitz, Antifolate-Induced Depletion of Intracellular Glycine and Purines Inhibits Thymineless Death in *E. coli*. *ACS Chem. Biol.* **5**, 787–795 (2010).

42. D. P. Sangurdekar, Z. Zhang, A. B. Khodursky, The association of DNA damage response and nucleotide level modulation with the antibacterial mechanism of the anti-folate drug Trimethoprim. *BMC Genomics* **12**, 583 (2011).

43. E. Toprak et al., Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* **18**, 101-105 (2011).

44. A. Kuchina *et al.,* Microbial single-cell RNA sequencing by split-pool barcoding. *Science* **371**, eaba5257 (2021).

45. S. B. Blattman, W. Jiang, P. Oikonomou, S. Tavazoie, Prokaryotic Single-Cell RNA Sequencing by *In Situ* Combinatorial Indexing. *Nat. Microbiol* **5** 1192-1201 (2020).

46. A. Dixit *et al.,* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866.e17 (2016).

47. J. C. Taggart, J.-B. Lalanne, G.-W. Li, Quantitative Control for Stoichiometric Protein Synthesis. *Annu. Rev. Microbiol.* **75**, 243–267 (2021).

Figures and Tables

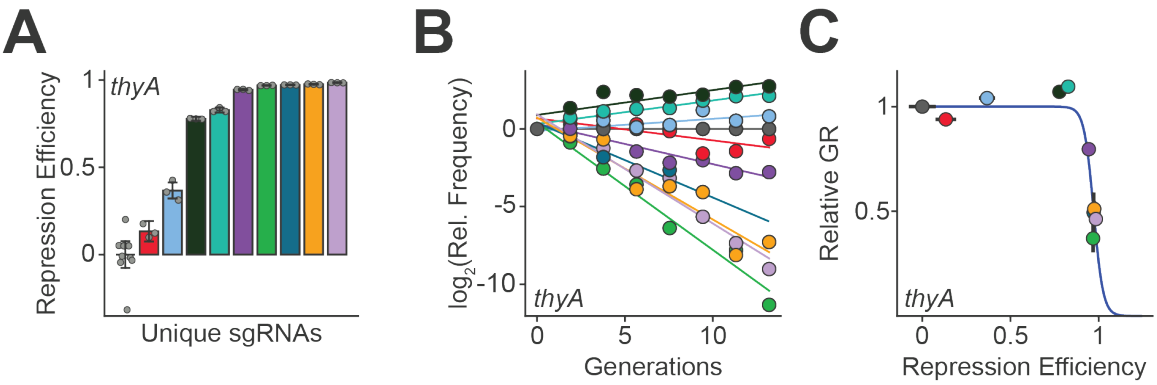

Figure 1. Mismatch CRISPRi enables characterization of gene expression/growth rate relationships. (A) The concentration of mRNA in log-phase *E. coli* for each titrating sgRNA was quantified by RT-qPCR. A repression efficiency of 1 corresponds to near-total knockdown. Individual colored bars represent different titrating sgRNAs targeting *thyA,* with error bars describing the standard error of the mean (SEM) across technical replicates. Individual measurements are shown as gray dots. (B) Growth rate was quantified by next-generation sequencing. We measured the $\log_2$ relative frequency of each sgRNA relative to a non-targeting control (gray dots along y = 0) over eight time points spanning 14 hours. Color coding of sgRNAs is identical to panel A. The slope of the line of best fit represents the growth rate of each knockdown relative to the non-targeting control. (C) The measurements of relative growth rate and repression efficiency show a sigmoidal relationship (blue line), with significant growth rate deficits emerging at severe repression levels. Color coding of sgRNAs is identical to panel A, individual points reflect average relative growth rate across n ≥ 4 replicates, and error bars represent SEM of growth rate measurements (vertical) and RT-qPCR measurements (horizontal).

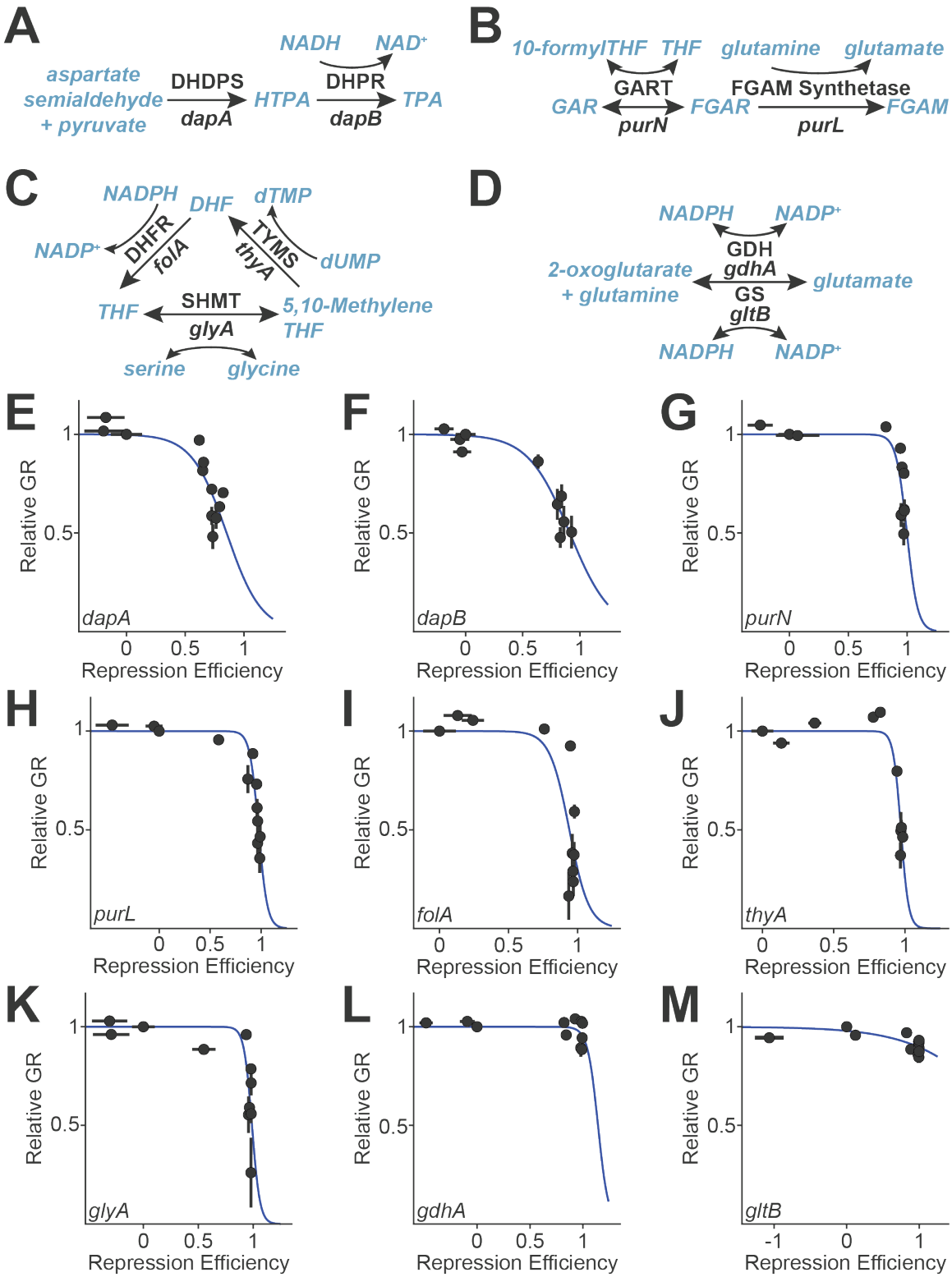

Figure 2. Mismatch CRISPRi for a diverse set of metabolic enzymes. sgRNAs were designed to target 9 *E. coli* genes from (A) lysine, (B) purine, (C) folate, and (D) glutamate metabolism. Gene names are listed in black italic text underneath the corresponding metabolic enzyme (bold black

text). Metabolite names are in blue. (E-M) Repression efficiency of each CRISPRi sgRNA, as measured by RT-qPCR, correlated with relative growth rate, as measured by next-generation sequencing. Error bars represent SEM of growth rate measurements (vertical) and RT-qPCR measurements (horizontal). Blue lines represent the two-parameter logistic fit to each data set. The genes *gdhA* and *gltB* are nonessential in our experimental conditions and show no significant growth rate defect following repression.

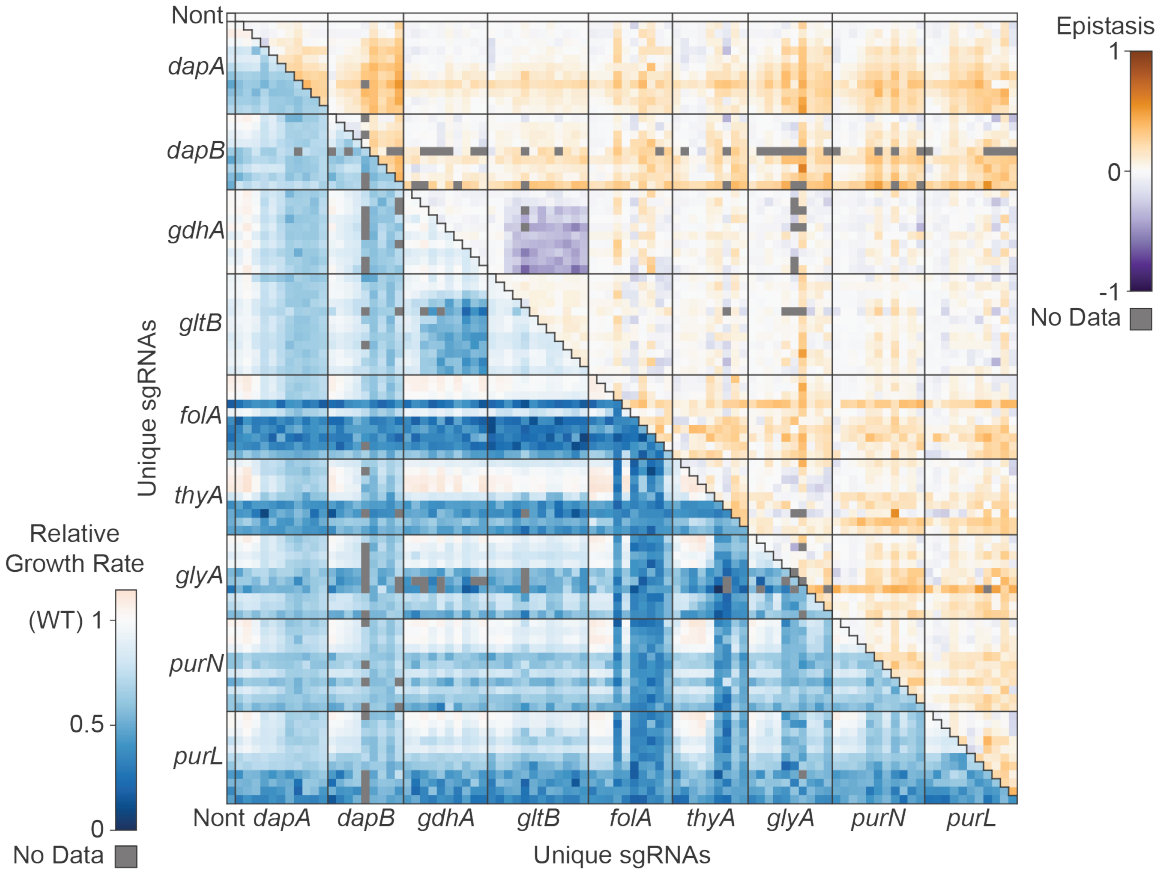

Figure 3. Pairwise CRISPRi growth rate and epistasis measurements. Each column and row represents a unique sgRNA perturbation. Gene names denote sgRNA groups targeting each gene, and sgRNAs are sorted within each group by increasing CRISPRi repression strength (left to right and top to bottom). Nont indicates the non-targeting control sgRNA. The lower triangle of the matrix describes pairwise relative growth rate measurements calculated relative to the Nont-Nont control across 14-hours, averaged across n ≥ 4 experimental replicates. The upper triangle of the matrix describes pairwise growth rate epistasis. Epistasis was calculated as the difference between the observed growth rate and the multiplicative (Bliss) expected growth rate of pairwise sgRNA perturbations.

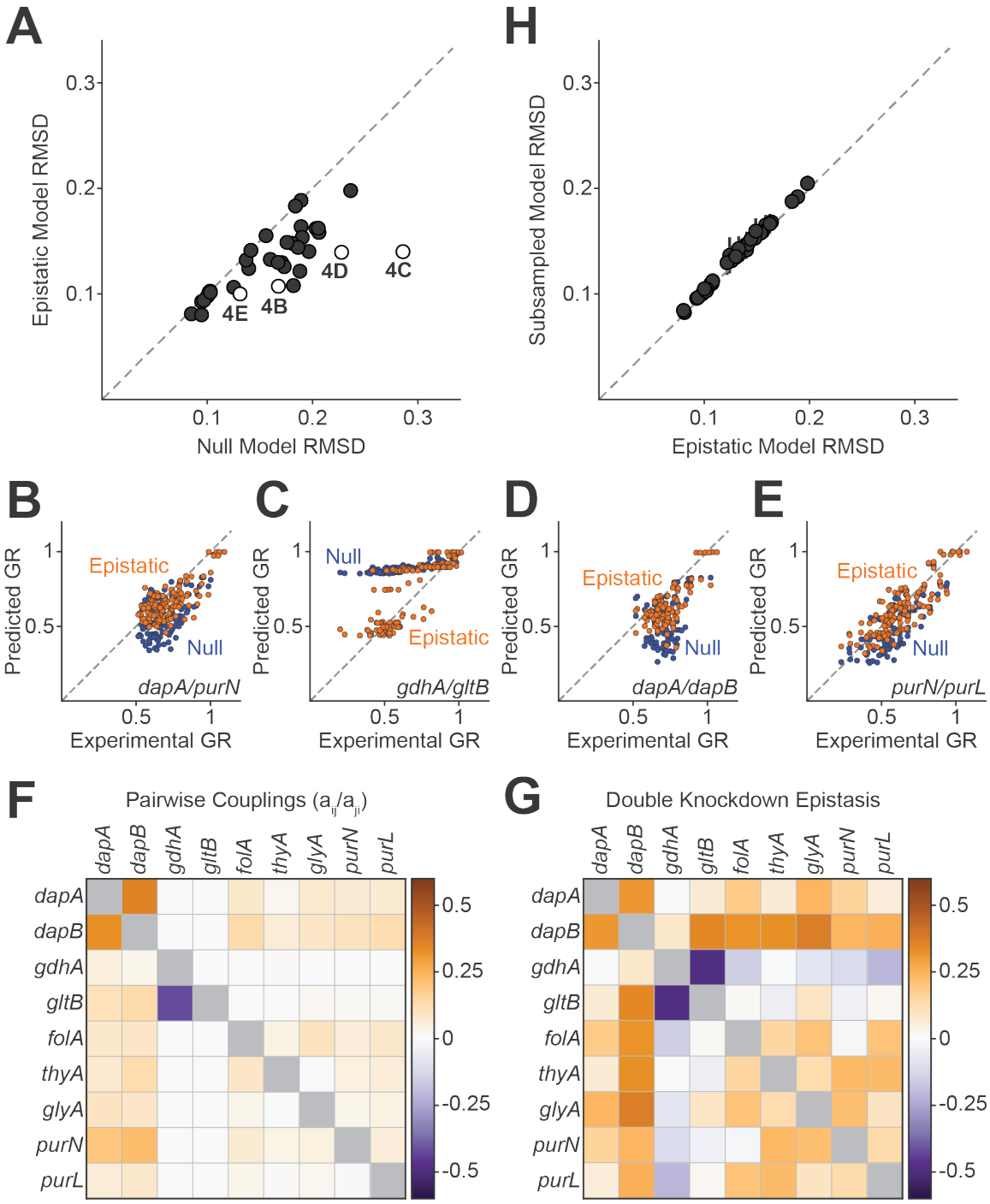

Figure 4. The continuous epistasis model captures gene-gene coupling at all expression levels. (A) The performance of the coupling-sensitive model compared to the Null model. Each data point represents model RMSD across all pairwise sgRNA combinations for a gene pair. Dotted gray line is y = x. (B-E) Measured and predicted growth rates for four pairwise gene knockdowns indicated in A. Dotted gray line is y = x. (F) Coupling constants fit across all pairwise growth rate data. (G) Expression-growth rate epistasis of the most severe double knockdown for each gene pair, calculated as the difference between the experimentally determined growth rate and the product of each single-gene knockdown growth rate. (H) Subsampled model performance

compared to the epistatic model without subsampling, as in A. Dotted gray line is y = x. Vertical error bars are SEM of subsampled model RMSD across 100 subsampling iterations.

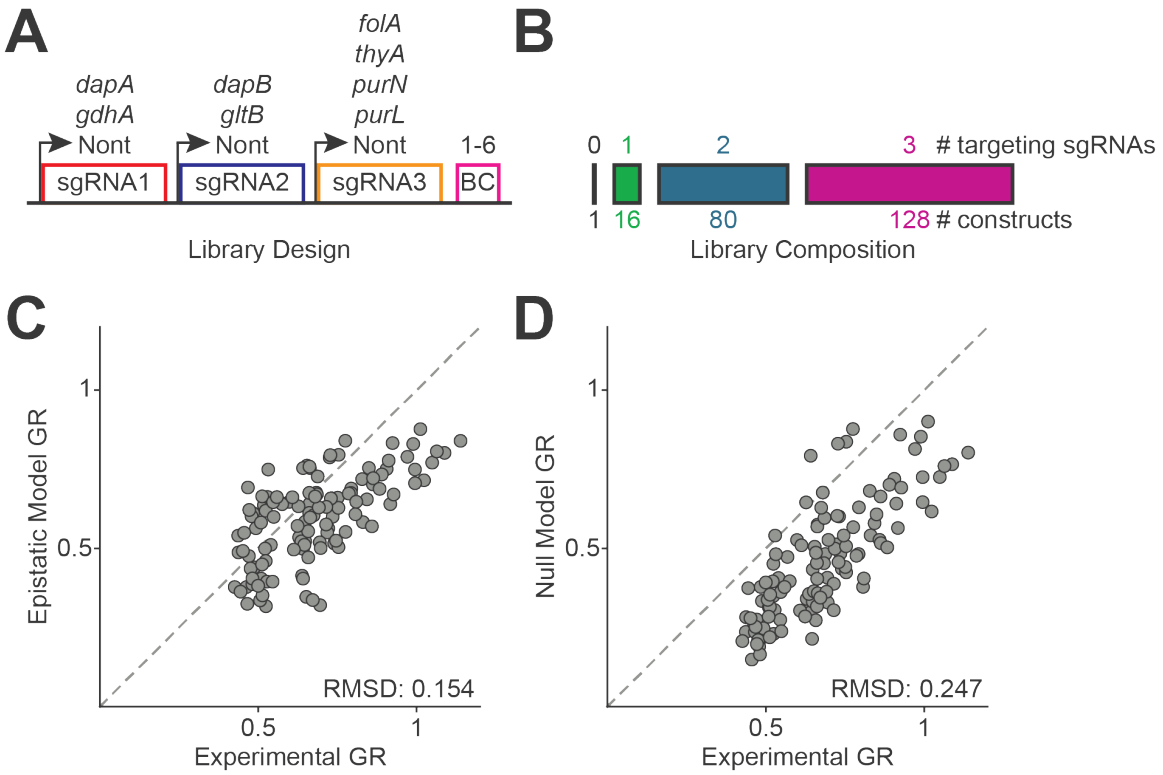

Figure 5. A model trained on pairwise measurements predicts growth rate following higher-order transcriptional changes. (A) Diagram of the third-order sgRNA library, with targeted genes indicated at each sgRNA position. Each position contained an sgRNA targeting one of the genes indicated or the non-targeting (Nont) control sgRNA. (B) Distribution of targeting sgRNAs in the library. The library contained zero- through third-order targeting sgRNA combinations, with bars scaled to illustrate each order's representation in the library. (C-D) Model performance on third-order sgRNA combinations, where experimentally determined growth rates are plotted against predicted growth rates for the (C) continuous epistasis and (D) Null model. Dotted gray line is y = x.

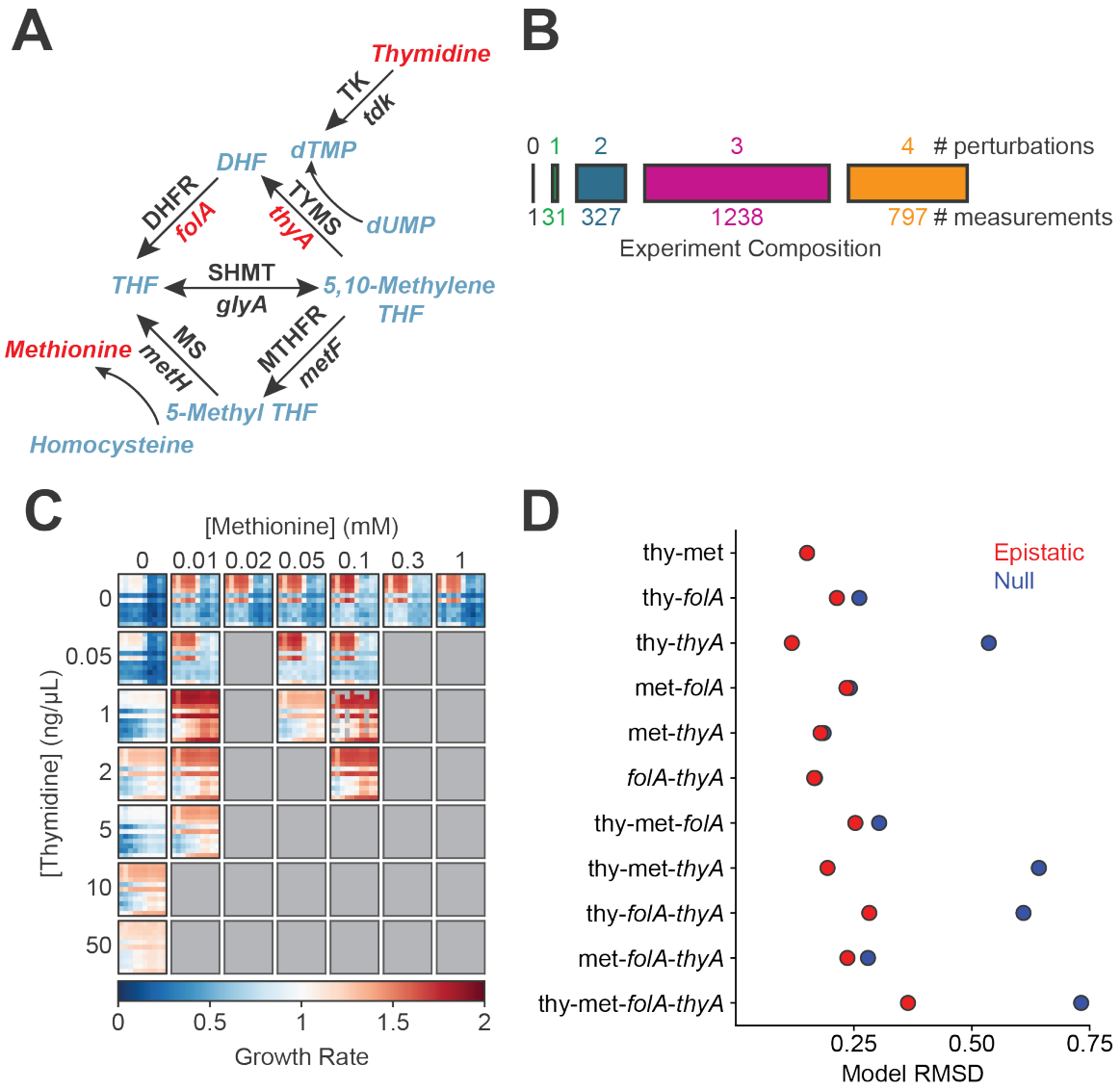

Figure 6. The continuous epistasis model captures gene-by-environment interactions. (A) Simplified schematic of a portion of *E. coli* folate metabolism. Gene names are listed in italics under each arrow, and enzyme abbreviations are listed in bold, black text above each arrow. Red text indicates a gene or metabolite perturbed over the course of the experiment. (B) Distribution of measurements in the gene-by-environment experiment. The experiment consisted of up to fourth-order (two genes and two environments) perturbations. Bars are scaled to illustrate the representation of each perturbation order within the data set. (C) Gene-by-gene-by-environment-by-environment relative growth rates, normalized to unsupplemented *E. coli* harboring a non-targeting control sgRNA construct. Within each heatmap, rows represent distinct *folA*-targeting sgRNAs, and columns represent distinct *thyA*-targeting sgRNAs. Thymidine and methionine supplementation is listed on axes, and gray boxes represent pairwise media combinations not explored in this study. (D) Continuous epistasis model performance, split by perturbation combination. For each second- and higher-order combination of expression and environmental perturbations (y-axis labels), the RMSD of growth rate measurements for the continuous epistasis model (red) and Null model (blue) are indicated.