Promoter constrains evolution of expression levels of essential genes in *Escherichia coli*

Saburo Tsuru[1,†,*], Naoki Hatanaka[1,†] and Chikara Furusawa[1,2,3,*]

[1]Universal Biology Institute, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[2]Department of Physics, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[3]Center for Biosystems Dynamics Research (BDR), RIKEN, 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan

[†]These authors are equally contributed.

*Corresponding authors:

Saburo Tsuru

E-mail: saburotsuru@gmail.com

Chikara Furusawa

chikara.furusawa@riken.jp

Abstract

Variability in expression levels in response to random genomic mutations varies among genes, influencing both the facilitation and constraint of phenotypic evolution in organisms. Despite its importance, both the underlying mechanisms and evolutionary origins of this variability remain largely unknown due to the mixed contributions of *cis*- and *trans*-acting elements. To address this issue, we focused on the mutational variability of *cis*-acting elements, that is, promoter regions, in *Escherichia coli*. Random mutations were introduced into the natural and synthetic promoters to generate mutant promoter libraries. By comparing the variance in promoter activity of these mutant libraries, we found no significant difference in mutational variability in promoter activity between promoter groups, suggesting the absence of a signature of natural selection for mutational robustness. In contrast, the promoters controlling essential genes exhibited a remarkable bias in mutational variability, with mutants displaying higher activities than the wild types being relatively rare compared to those with lower activities. Our evolutionary simulation on a rugged fitness landscape provided a rationale for this vulnerability. These findings suggest that bias in mutational variability in promoters may constrain the evolution of downstream essential genes toward higher expression levels.

Introduction

The rate and direction of phenotypic evolution largely depend on phenotypic variability(Lynch and Walsh 1998; Walker 2007). It is widely recognized that organisms often exhibit nonuniform phenotypic variability, with certain variants arising more frequently than others in response to different genetic and environmental perturbations, even in the absence of natural selection(Darwin 1859; Waddington 1957; Smith, et al. 1985; Noble, et al. 2019). Therefore, it is crucial to elucidate the mechanisms underlying the bias in phenotypic variability to understand why phenotypic evolution proceeds as it does(Uller, et al. 2018). Transcriptional variability of genes across different perturbations plays a significant role in phenotypic evolution because the evolution of many traits in organisms largely depends on the evolution of expression levels and their regulation(Wittkopp, et al. 2004; Zheng, et al. 2011). Interestingly, previous studies have identified differences in transcriptional variability from gene to gene in response to various perturbations(Denver, et al. 2005; Rifkin, et al. 2005; Landry, et al. 2007; Tsuru and Furusawa 2024), where genes related to cellular growth and maintenance tend to exhibit lower transcriptional variability against random genetic perturbations to the genome in yeast(Landry, et al. 2007) and bacteria(Tsuru and Furusawa 2024). These studies, supported by theoretical evidence(Kaneko 2009; Draghi and Whitlock 2012; Furusawa and Kaneko 2018), suggest that the transcriptional variability of genes is not a random occurrence devoid of biological significance but rather a product of natural selection.

What molecular mechanisms account for gene-to-gene differences in transcriptional variability? In principle, both *cis*- and *trans*-acting elements can influence transcriptional variability(Wittkopp, et al. 2004; Tirosh, et al. 2006; Landry, et al. 2007; Payne and Wagner 2015; Tsuru and Furusawa 2024). However, it remains largely unknown which elements practically contribute to gene-to-gene differences and how. Previous studies(Denver, et al. 2005; Rifkin, et al. 2005; Landry, et al. 2007; Tsuru and Furusawa 2024) explored the transcriptional variability against different genetic perturbations using mutation accumulation (MA) experiments(Halligan and Keightley 2009), where independent mutants were generated by accumulating genomic mutations through genetic drift, and their transcriptome profiles were obtained under identical environmental conditions. Although MA experiments provide a powerful method to quantify transcriptional variability in the absence of strong natural selection, the indiscriminate accumulation of mutations on *cis*- and *trans*-acting elements(Tsuru, et al. 2015) complicates the identification of the molecular mechanism underlying transcriptional variability. Accordingly, to overcome this difficulty, it is necessary to consider the contributions of *cis*- and *trans*-acting elements separately.

Why have genes related to cellular growth and maintenance acquired low mutational variability? The evolutionary origin of mutational variability is highly controversial(de Visser, et al. 2003). Mutational variability could result from direct natural selection for mutational robustness(Waddington 1942), or could be a byproduct of selection for robustness against

environmental perturbations(Meiklejohn and Hartl 2002; de Visser, et al. 2003). Several biological traits, such as metabolic flux(Ho and Zhang 2016) and cellular morphology(Ho and Zhang 2014), show signs of adaptive evolution through direct selection for mutational robustness, where traits highly related to cellular growth fitness are likely to acquire lower mutational variability through direct selection for mutational robustness. On the other hand, other studies have identified congruence in trait variability between environmental and genetic perturbations, such as in RNA structure(Szollosi and Derenyi 2009) and gene expression levels(Landry, et al. 2007; Tsuru and Furusawa 2024), supporting the relevance of the byproduct of selection for environmental robustness. As an alternative scenario, mutational variability may spontaneously emerge with trait evolution, regardless of selection for robustness(Siegal and Bergman 2002; de Visser, et al. 2003). To date, the most appropriate scenario for low transcriptional variability of genes related to cellular growth and maintenance remains largely unknown.

To address these questions, we focused on the mutational variability governed by cis-acting elements, that is, promoter regions, of genes in *Escherichia coli*. We created a mutant library for each promoter through random mutagenesis and analyzed the changes in promoter activity from the wild type to mutant using flow cytometry. To explore the relevant evolutionary scenarios for the mutational variability of promoters, we compared promoters controlling essential genes for cellular growth with those controlling nonessential genes. These groups were

expected to experience different natural selection pressures on their transcriptional variability in the past. In addition, random sequences were used to construct synthetic promoters that experienced no selection for mutational robustness. Interestingly, when comparing the variance in promoter activity of the mutant libraries, these three promoter groups exhibited similar mutational variability in promoter activity, indicating the absence of signatures of natural selection for mutational robustness. In contrast, the promoters for essential genes showed a remarkable vulnerability to mutations compared to the other promoter groups, where variants with higher activity than the wild type arose less frequently than variants with lower activity. Our evolutionary simulation on a rugged fitness landscape explains the higher vulnerability of the promoters for essential genes. These results suggest that biased mutational variability in promoters may constrain evolution of downstream essential genes toward higher expression levels.

Results

Higher expression levels and lower transcriptional variability of essential genes in *E. coli*

First, we analyzed the transcriptional variability of essential genes in *E. coli* in response to various genetic perturbations (**Fig. 1A**). Previously(Tsuru and Furusawa 2024), we obtained transcriptome profiles of independent mutants that had accumulated genomic mutations through genetic drift via the mutation accumulation (MA) experiment(Tsuru, et al. 2015), termed the Mut dataset. Using this dataset, we quantified the mean and standard deviation of the expression levels for each of thousands of genes (**Fig. 1B**). To compensate for the mean-dependency of the standard deviations, we calculated the vertical distance from the smoothed spline of the running median of the standard deviations, termed DM_{mut}. The DM_{mut} values reflect the mean transcriptional variability in response to different genetic perturbations(Tsuru and Furusawa 2024). We categorized thousands of genes into essential and nonessential genes for cell growth, following the definition of gene essentiality by Goodall *et al*(Goodall, et al. 2018). We confirmed that essential genes showed higher expression levels than nonessential genes across the different mutants (**Fig. 1C**). We also found that essential genes had lower DM_{mut} than nonessential genes (**Fig. 1D**).

Next, we explored the transcriptional variability of essential genes in response to different environmental perturbations (**Fig. 1E,F**) using a public dataset(Sastry, et al. 2019; Rychel, et al. 2021) of the transcriptome profiles of a single strain of *E. coli*, MG1655, cultured

124 under 76 different environmental conditions, termed the Env dataset. Previously(Tsuru and
125 Furusawa 2024), we quantified the mean-controlled standard deviations, termed DM_{env}, using
126 a method similar to DM_{evo}. The DM_{env} reflects transcriptional variability in response to different
127 environmental perturbations. We confirmed that the essential genes exhibited higher expression
128 levels (**Fig. 1G**) and lower variability (**Fig. 1H**) than the nonessential genes across different
129 environmental conditions. Taken together, essential genes exhibited higher expression levels
130 and lower transcriptional variability in response to different perturbations, which is similar to
131 the characteristics of essential genes in yeast(Tirosh, et al. 2006).

132 **Construction of mutant libraries of *E. coli* promoters to measure mutational effects**

133 To explore the molecular mechanisms and signs of natural selection underlying the lower
134 transcriptional variability of essential genes, we focused on the promoter regions of genes with
135 different essentialities. In particular, we explored how promoter activity varies in response to
136 new mutations in the promoter regions and how this variability relates to essentiality of the
137 downstream genes. To this end, we used 56 natural promoters from the *E. coli* Promoter
138 Collection(Zaslaver, et al. 2006). Importantly, to ensure an unbiased comparison of changes in
139 promoter activities between different essentialities, we avoided the use of natural promoters
140 with very low activity that were highly enriched in the promoters of nonessential genes. To
141 achieve this, we selected natural promoters with moderate to high activity before mutagenesis.
142 This selection was based on the known promoter activities of the collection as measured by

143 Silander et al(Silander, et al. 2012). Using the RegulonDB database(Tierrafria, et al. 2022), we
144 retrieved information regarding known transcriptional units regulated by these promoters.
145 Based on the presence of essential genes in the transcription units, we categorized natural
146 promoters into nonessential and essential promoters, termed Nes and Ess, respectively (**Fig.**
147 **2A**). Nonessential promoters control transcriptional units comprising only nonessential genes,
148 whereas essential promoters regulate those that include at least a single essential gene. In
149 addition to these natural promoters, we created synthetic promoters, termed Syn, as controls
150 that experienced no or minimal selection for mutational robustness. To this end, we placed
151 random sequences (~200 bp) upstream of the *gfp* gene on a low-copy-number plasmid, pUA66-
152 mCherry (**Fig. 2B**, **supplementary fig. S1**). The library of mixed transformants containing this
153 plasmid was analyzed using flow cytometry. The top 7% of cells showing higher expression
154 levels of GFP (red shaded area in **Fig. 2B**) were sorted, and 16 unique clones were subsequently
155 isolated. Importantly, these synthetic promoters experienced selection for their expression
156 levels only once, and no longer experienced further mutagenesis or selection for expression
157 level. Accordingly, they can be regarded as a control group that is almost free from selection
158 for mutational robustness. Consequently, in the upstream of the GFP gene on the plasmid, we
159 constructed three promoter groups, Nes, Ess, and Syn, comprising different sequences with low
160 homology (**supplementary fig. S2, supplementary note**). The green fluorescence from GFP
161 was used as a measure of promoter activity(Zaslaver, et al. 2006; Silander, et al. 2012). Despite

162 the diversity in sequence, we confirmed that the mean activities of the three promoter groups
163 were comparable to each other, as designed (**Fig. 2C**). We also confirmed that the obtained
164 promoter activities were highly consistent with those measured by Silander et al(Silander, et al.
165 2012) (Spearman's R=0.94, $p<0.05$, **supplementary fig. S3A**). Compared to the Env dataset,
166 the activities of the natural promoters were positively correlated with known mRNA expression
167 levels of the genes downstream of the chromosomal promoters in the wild-type strain
168 (Spearman's R=0.64, **supplementary fig. S3B**), supporting that the wild-type promoter
169 activities monitored by our method considerably reflect the activities of chromosomal
170 promoters and their impact on the expression levels of downstream genes.

171 For each of these promoters, we introduced mutations by error-prone PCR at a low
172 frequency (6.3 mutations/kb on average, **Fig. 2D**) and integrated them into pUA66-mCherry to
173 create a mutant library (MT). Each mutant library had a genetic diversity of 10^5 or more
174 genotypes. The promoter activity of each mutant library was measured by flow cytometry based
175 on the green fluorescence derived from downstream *gfp*, and was compared with the green
176 fluorescence of the corresponding wild-type (WT) (**Fig. 2E**). The introduction of mutations was
177 limited to the promoter regions of the plasmid, while the chromosomal copies of the natural
178 promoters were intact to avoid any confounding effects. The green fluorescence intensity was
179 $\log_{10}$-transformed after subtracting the background (**supplementary fig. S4**), and the mean and
180 standard deviation, SD, of the GFP distribution were calculated (**Fig. 2E, supplementary fig.**

181 **S5**).

182 **Mutational variability in promoter activity was similar between different promoter**
183 **groups**

184 We confirmed that all mutant libraries showed larger variations than the corresponding wild-
185 type clones, as designed (**Fig. 3A**). In addition, we also confirmed the well-known negative
186 correlation between mean expression levels and standard deviations in GFP distributions
187 regardless of the introduction of mutations or whether the promoters were natural or synthetic
188 (**Fig. 3B,C**), which is consistent with previous observations(Zaslaver, et al. 2006; Wolf, et al.
189 2015). To compare different promoters, we compensated for the mean-dependency of the
190 standard deviations of both WT and MT by subtracting the regression line obtained from WT
191 (**Fig. 3B,C**). We subsequently calculated MV_{prm}, a measure of mutational variability in
192 promoter activity, by subtracting the compensated SD of WT from that of MT. Unexpectedly,
193 MV_{prm}, showed similar values across the three promoter groups (Wilcoxon test, adjusted p-
194 values>0.05), even though each promoter might show unique values (**Fig. 3D**). These results
195 indicated that mutational variability of essential promoters showed no signature of natural
196 selection for mutational robustness despite their importance in the regulation of genes essential
197 for cellular growth. Moreover, MV_{prm} showed no significant positive correlations with DM_{env}
198 and DM_{mut} of the downstream genes (**Fig. 3E,F**). Using the RegulonDB and EcoCyc(Keseler,
199 et al. 2021) databases to retrieve the known regulatory relationships between genes, we counted

200 the number of regulations that each promoter receives and explored the impact of the number
201 of regulations on mutational variability. We found that a larger number of transcriptional
202 regulators, in particular activations (**Fig. 3G**), facilitated mutational variability in promoter
203 activity (**Fig. 3G,H**), probably due to a larger number of mutational target sites. Nevertheless,
204 the lack of significant correlations between MV_{prm} and DM values suggests that the mutational
205 variability of promoters provides no considerable explanation for the observed gene-to-gene
206 differences in transcriptional variability of the downstream genes in response to different
207 perturbations at the genome scale (**Fig. 1**). These results suggest that the gene-to-gene
208 differences in transcriptional variability in response to genome-wide perturbations may reflect
209 a certain *trans*-acting molecular mechanism, mainly other than the *cis*-acting one, consistent
210 with previous observations(Tsuru and Furusawa 2024). This implication was also rational
211 because promoter regions are relatively short in bacterial genomes, where the mutational target
212 sites within the promoter regions are small relative to the trans-acting mutational target sites for
213 genomic mutations.

214 **Vulnerability in essential promotors against mutations**

215 Next, we explored the directionality of the mutational changes in promoter activity. Consistent
216 with the low mutation rate in error-prone PCR, the mean expression levels of MT were similar
217 to those of WT (Spearman's R=0.99) (**Fig. 4A**). Nevertheless, we found that log_{10}-fold changes
218 in mean expression levels from WT to MT followed a global tendency over expression levels,

219 where the promoters with higher activities in WT tended to decrease their activity in response
220 to mutations (Spearman's R=–056) (**Fig. 4B**). This tendency may reflect bias in the mode,
221 activation or inhibition, of transcriptional regulation(Kinney, et al. 2010; Belliveau, et al. 2018;
222 Ireland, et al. 2020). We hypothesized that promoters that receive more activation than
223 inhibition might tend to reduce their activity in response to mutations. To test this possibility,
224 we defined activation bias as the number of known activations subtracted by the number of
225 inhibitions (**supplementary table S1**). We found that the activation bias explained the observed
226 mutational changes (**Fig. 4C**). Furthermore, the activation bias also explained the $\log_{10}$-fold
227 changes after controlling the global trend using the regression line (**Fig. 4D**), indicating that a
228 promoter becomes more vulnerable as it receives more activation among promoters with similar
229 activities.

230 Interestingly, synthetic and nonessential promoters showed similar mutational changes
231 around zero, whereas essential promoters showed biased directionality in mutational changes
232 toward negative values (**Fig. 4E**), even though the mean activities were almost the same among
233 the three promoter groups (**Fig. 2C**). In addition, there was no significant difference in the
234 mutation rate in error-prone PCR between the essential and nonessential promoters
235 (**supplementary fig. S6, supplementary table S2**), supporting the notion that the observed
236 bias was not derived from a technical bias. This bias in essential promoters was retained after
237 compensating for the global trend (**Fig. 4F**). Accordingly, essential promoters were more

238 vulnerable than other groups in terms of promoter activity. In other words, this biased
239 directionality indicated the low frequency of the variants with higher activities than the wild-
240 types in essential promoters. This biased variability might limit the transcriptional variability
241 of essential genes in response to mutations in cis-regulatory regions, independently from
242 transcriptional variability in response to genomic mutations (**Fig. 1**, **3F**).

243 **Vulnerability in essential promotors against mutations may reflect past selection for**
244 **higher activity**

245 The difference in vulnerability between essential and nonessential promoters could reflect the
246 variation in past selection pressures acting on higher expression levels between essential and
247 nonessential genes. Based on the observation that essential genes tended to have higher
248 expression levels than nonessential genes (**Fig. 1**), we hypothesized that the remarkable
249 vulnerability of essential promoters reflects local optima in terms of promoter activity. That is,
250 essential promoters might have evolved toward higher activity through past evolution and might
251 have been stuck in local optima because of the high essentiality of downstream genes. In
252 contrast, nonessential promoters might have experienced relatively relaxed selection for higher
253 expression levels because of the low essentiality of downstream genes. To investigate the
254 emergence of differences in vulnerability depending on selection pressure, we performed a
255 numerical simulation of promoter evolution on a simple rugged fitness landscape (**Fig. 5**). For
256 simplicity, we modeled promoter genotype as 10-bit binary numbers such as 1101011011, and

257 expression level, equivalent to promoter activity, was defined by the Rough Mt. Fuji (RMF)
258 model, as illustrated in **Fig. 5A, D**. Expression levels were set to range from 0 to 1. Using this
259 expression landscape, mutational changes were calculated for each genotype as the mean
260 difference in the expression levels of all nearest neighbors with a single bit difference, or one
261 Hamming distance, from the expression level of the focal genotype (**Fig. 5B, E**). This simple
262 expression landscape captured the negative correlation between expression levels and
263 mutational changes (**Fig. 5E**), which is consistent with the global trend observed in our
264 experiment (**Fig. 4B**). This trend was likely due to the prevalence of local sinks (green dots) at
265 lower expression levels and frequent local peaks (red dots) at higher expression levels.
266 Importantly, even at the same expression levels, local peaks tended to have negative mutational
267 changes, whereas local sinks tended to show positive changes. The fitness landscape was then
268 modeled by adding the basal fitness, F_b, to the expression levels multiplied by $1 - F_b$, where
269 F_b ranged from 0 to 1 (**Fig. 5C, F**). Consequently, a smaller F_b corresponds to a steeper fitness
270 landscape, while a larger F_b results in a flatter fitness landscape. Based on the fact that
271 knockdown of essential genes is more deleterious for cell growth than nonessential genes(Li,
272 et al. 2016), we assigned lower and higher values of F_b for essential and nonessential promoters,
273 respectively. Thus, in our model, essential and nonessential promoters shared the same
274 expression landscape but experienced different strengths of selection pressures. Using the
275 constructed fitness landscape with different F_b values (0.3, 0.5, and 0.99), we performed an

evolutionary simulation based on the standard Wright-Fisher model(Otto and Day 2007). Initial genotypes were randomly generated for each simulation. The most abundant mutants within the evolved populations after 30 generations were isolated and subjected to measurement of mutational changes (**Fig. 5 G,H**). We confirmed that the evolved isolates in steeper fitness landscapes exhibited higher expression levels (**Fig. 5I**), which is consistent with the designed selection pressure. Comparing the mutational changes between different F_b values, we found that the mutants that evolved with smaller F_b showed negatively larger mutational changes, even at the same expression levels (**Fig. 5H**). These results suggest that higher activity and vulnerability in essential promoters are linked, and both can emerge simultaneously through stronger positive selection acting on higher expression levels of downstream essential genes. Similar results were also obtained using the NK model(Kauffman and Weinberger 1989), suggesting the robustness of the trend (**supplementary fig. S7**).

288 **Discussions**

289 Transcriptional variability differs between genes. Using public transcriptome datasets for *E.*
290 *coli*, we found that essential genes exhibited lower transcriptional variability across different
291 mutants and environmental conditions than nonessential genes. To explore the underlying
292 mechanism, we focused on the promoter regions. We quantified the changes in promoter
293 activity changes in response to mutations by constructing mutant libraries of natural and
294 synthetic promoters. As a result, essential promoters did not exhibit a significant difference in
295 mutational variability compared with nonessential and synthetic promoters. Unexpectedly, we
296 found that essential promoters showed remarkable vulnerability to mutations, where variants
297 with higher activities than wild-type were rare and those with lower activity were abundant.
298 These results imply that essential promoters have evolved to reach local optima in terms of
299 promoter activity through past selections for higher expression levels of downstream essential
300 genes. Our numerical evolutionary simulation confirmed the relevance of this scenario. Overall,
301 our results suggest the evolutionary impact of essentiality on the directionality of the mutational
302 variability of promoters and imply that the directional bias in mutational variability of promoter
303 regions might constrain further evolution of expression levels of essential genes toward higher
304 expression levels.

305 There are three well-known evolutionary scenarios for the emergence of mutational
306 robustness in many biological systems(de Visser, et al. 2003), namely adaptive(Waddington

307 1942), congruent(Meiklejohn and Hartl 2002), and intrinsic(Siegal and Bergman 2002)
308 scenarios. The adaptive scenarios predict that essential traits have evolved to reduce mutational
309 variability and minimize the emergence of harmful mutants through direct selection for
310 mutational robustness. The congruent scenarios predict that mutational robustness is a
311 byproduct of evolution toward low variability against environmental perturbations. Contrary to
312 these two scenarios assuming adaptive evolution toward low variability, the intrinsic scenarios
313 predict that the mutational robustness of the traits emerges spontaneously and inherently along
314 with the emergence of traits without selection against variability. In the case of promoter activity,
315 an intrinsic scenario corresponds to the mutational robustness of promoters emerging
316 spontaneously with the emergence of promoter activity without selection for mutational
317 robustness, which corresponds to the mutational variability of synthetic promoters. Our results
318 demonstrated that both adaptive and congruent scenarios failed to explain the observed
319 mutational variability in natural promoters because essential promoters showed similar
320 mutational variability to the other promoters, and the natural promoters exhibited no significant
321 positive correlation with the transcriptional variability of downstream genes against different
322 environmental perturbations. We note that the adaptive scenario might explain the higher
323 expression levels, not transcriptional variability, of essential genes (**Fig. 1C, G**), which might
324 minimize the risk of mutational knock down for essential genes to a lethal level. In contrast, the
325 intrinsic scenario explained the observed mutational variability because the natural promoters

326 showed similar mutational variability to that of synthetic promoters. Thus, the mutational
327 variability of essential promoters is unlikely to be a consequence of adaptive evolution toward
328 robustness against genetic and environmental perturbations.

329 Why did essential promoters show no less mutational variability than other promoters,
330 despite the fact that the reduction in expression levels of essential genes is more likely to be
331 harmful to cell growth than nonessential genes(Cui, et al. 2018)? One possible reason is that
332 promoter sequences with significant mutational robustness are rare, although several possible
333 mechanisms underlying mutational robustness in promoter regions are hypothesized(Payne and
334 Wagner 2015). In other words, the expression landscape of the promoters might be rugged
335 everywhere as we modeled in **Fig. 5A**. In fact, synthetic promoters derived from different
336 random sequences achieved higher promoter activities equivalent to natural promoters, which
337 was consistent with previous observations(Wolf, et al. 2015; Yona, et al. 2018). A second
338 possible reason is that selection for higher expression levels on a rugged fitness landscape (**Fig.**
339 **5C**) prevents the essential promoters at local optima from traversing the expression landscape
340 to reach genotypes with higher mutational robustness. Consistently, essential promoters showed
341 biased mutational effects, where most mutations in essential promoters tended to reduce
342 promoter activity (**Fig. 4C**). Interestingly, this scenario agreed with the predictions from a
343 machine-learning model based on an empirical expression landscape of yeast
344 promoters(Vaishnav, et al. 2022). Thus, our results suggest several plausible hypotheses that

345 may be shared by different organisms. Ideally, future studies should explore evolutionary
346 scenarios for diverse promoters that were not tested in this study, and validate the hypothetical
347 landscape of promoters more thoroughly.

348 In this study, we focused only on mutations in promoter regions to explore the
349 mechanism underlying the lower transcriptional variability in essential genes. As a result, we
350 found that mutants with higher promoter activity were relatively rare for essential promoters.
351 However, this result does not exclude the contribution of any other mechanism underlying the
352 observed transcriptional variability in response to genome-wide perturbations (**Fig. 1)**. Notably,
353 *trans*-acting mechanisms, such as transcriptional regulators, may also play a role. As it is, we
354 previously reported that genes exhibiting greater transcriptional variability across diverse
355 environmental and genetic perturbations tend to be regulated by a larger number of unique
356 transcriptional regulators in *E. coli*(Tsuru and Furusawa 2024). A higher number of regulators
357 provides more *trans*-mutational target sites for the regulated genes, thereby enhancing
358 transcriptional variability in response to genomic perturbations. Based on this evidence, it is
359 plausible that essential genes have less transcriptional regulation. To support this hypothesis,
360 we used data from RegulonDB and EcoCyc to compare the number of regulations between
361 essential and nonessential genes (**supplementary fig. S8**). First, we found a smaller fraction of
362 essential genes (90 of 248 genes, 36%) lacking known transcriptional regulators than
363 nonessential genes (1976 of 3995, 49%) (**supplementary fig. S8A**). This finding aligns with

364 the positive activation bias values observed (**Fig. 4F**), probably attributable to the rich
365 activation by sigma factors aimed at enhancing expression levels. On the other hand, among
366 the genes regulated by at least one transcriptional regulator, we observed that essential genes
367 received fewer transcriptional regulations than nonessential genes (**supplementary fig. S8B**),
368 supporting this hypothesis. This observation is consistent with previous findings in yeast(Bilu
369 and Barkai 2005). Thus, our study underscores the dual contributions of both *cis*- and *trans*-
370 acting mechanisms in constraining the transcriptional variability of essential genes.

371 ## Materials and Methods

372 **Bacterial strains and medium**

373 We used *E. coli* DH5α Electro-Cells (Takara Bio, Japan, 9027) to construct the pUA66-
374 mCherry plasmid and its derivatives. LB medium (LB Broth, Miller, BD Difco, USA, 244620)
375 supplemented with 25 μg/mL kanamycin (Kanamycin Sulfate, Wako, Japan, 115-00342),
376 referred to as LB/Km medium, served as the selective medium. The SOC medium, derived from
377 SOB medium (SOB Medium, BD Difco, USA, 244310) supplemented with 22 mM Glucose
378 (Wako, Japan, 049-31165), was used for recovery culture pos-transformation. The *E. coli* K-12
379 MG1655 strain served as the host for the constructed plasmids and was employed for measuring
380 promoter activity via flow cytometry. Bacterial cells for flow cytometry were cultured in a
381 synthetic medium, a derivative of M9 minimal medium (pH 7.2), supplemented with 47.7 mM
382 Na_2HPO_4 (Disodium Hydrogenphosphate 12-Water, Wako, Japan, 196-02835), 22.0 mM
383 KH_2PO_4 (Wako, Japan, 164-22635), 8.55 mM NaCl (Wako, Japan, 191-01665), 9.35 mM
384 NH_4Cl (Wako, Japan, 014-03005), 11 mM Glucose, 1 mM $MgSO_4$ (Magnesium Sulfate
385 Heptahydrate , Wako, Japan, 138-00415), 50 μM $CaCl_2$ (Wako, Japan, 038-24985), and 3.8 μM
386 thiamin (Thiamin Hydrochloride, Wako, Japan, 201-00852), with supplementation of 25 μg/mL
387 Kanamycin when required.

388 **Selection of natural promoters**

389 The *E. coli* Promoter Collection(Zaslaver, et al. 2006) (Horizon Discovery Ltd., UK, PEC3877)

390 was used to obtain the promoter regions of the natural promoters in *E. coli*. Initially, we filtered
391 out the promoters with low activities that were near our detection limit based on the promoter
392 activities of this library measured by Silander et al(Silander, et al. 2012). Subsequently, 115
393 promoters were randomly selected from this subset library carrying the pUA66 plasmid. The
394 DNA sequences of these promoter regions were verified using Sanger sequencing (Azenta Life
395 Sciences, USA) and compared with the reference genome sequence of *E. coli* K-12
396 MG1655(Blattner, et al. 1997) (GenBank ID: U00096). Subsequently, 63 of the 115 strains
397 without mutations in their promoter sequences were identified, and plasmids were extracted
398 from these colonies. Among these, 56 strains were randomly selected for the subsequent
399 experiments (**supplementary table S1**). Each promoter was denoted by the "NP" (natural
400 promoter), followed by the name of the ORF present in the immediate downstream sequence in
401 the genome. For example, NPmetG represents the promoter regions upstream of the *metG* gene
402 on the chromosome.

403 **Plasmid construction**

404 Plasmid pUA66 was modified by replacing the promoter regions of *gfp*, flanked by the BamHI
405 and XhoI restriction sites, with a red fluorescence protein (mCherry) expression cassette. This
406 cassette comprised dual terminators RNAI and TSAL(Reynolds, et al. 1992), promoter PLtetO-
407 1(Lutz 1997), Shine-Dalgarno sequence SD8(Ringquist, et al. 1992) and codon-optimized
408 mCherry(Dunlop, et al. 2008) coding sequences. The terminators were obtained from pNS2-

409 σVL(Dunlop, et al. 2008)(Addgene, Plasmid #26756) and the assembled sequences of the other
410 elements were commercially synthesized (Fasmac Co., Ltd., Japan). The assembled cassette
411 was inserted between the BamHI and XhoI recognition sequences of the pUA66 plasmid in the
412 direction opposite to that of *gfp* (**supplementary fig. S1**), yielding plasmid pUA66-mCherry.
413 PCR amplification was performed using KOD One PCR Master Mix Blue (Toyobo, Japan,
414 KMM-201) and pUA66-mCherry as a template with primers A
415 (GCTTTCCAGGGATCCTCTAGATTTAAGAAGG) and B
416 (GAGCAGTACTCTCGAGGTGAAGACGAAAGG) to include all sequences except for the
417 mCherry expression cassette. The PCR product was column purified using the FastGene
418 Gel/PCR Extraction Kit (NIPPON Genetics, Japan, FG-91302). The purified PCR product was
419 double-digested with BamHI (Takara Bio, Japan, 1010A) and XhoI (Takara Bio, Japan, 1094A)
420 overnight at 37 °C, followed by column purification. The digested product is referred to as the
421 vector.

422 **Construction of promoter library comprising of random sequences**

423 The construction of the synthetic promoters was based on a previous study(Wolf, et al. 2015).
424 Single-stranded oligonucleotides (CCTTTCGTCTTCACCTCGAG-(N200)-
425 GGATCCTCTAGATTTAAGAAGG) were synthesized by Eurofins Genomics (Ebersberg,
426 Germany). The 5' and 3' ends of the oligos contained XhoI and BamHI restriction sites,
427 respectively, and were homologous to primers C (CCTTTCGTCTTCACCTCGA) and D

428 (CCTTCTTAAATCTAGAGGATCC), respectively (**supplementary fig. S1**). A 200 bases
429 region of random nucleotides was present between them. PCR amplification using primers C
430 and D was performed using a mixture of single-stranded oligos as a template to create double-
431 stranded DNA. The PCR products were column purified and double digested using BamHI and
432 XhoI for 3 h at 37 °C, followed by column purification. Thus an insert for the following ligation
433 reaction was obtained. The ligation reaction was performed using T4 DNA Ligase (Takara Bio,
434 Japan, 2011A) overnight at 16 °C involved mixing the vector and insert at a molar ratio of 1: 3.
435 The resulting ligation product solution was column purified and DNA was extracted using 20
436 μL of sterile water. All purified ligation product solutions were added to 50 μL of DH5α
437 Electro-Cells and electroporated at 1800 V (Eporator, Eppendroph, Germany, 4309000027).
438 Transformed cells were incubated at 37 °C for 1 h in 1 mL of SOC medium, followed by
439 overnight incubation at 37 °C in 1 mL of LB medium supplemented with 2 μL of 25 mg/mL
440 kanamycin solution.

441 **Selection of synthetic promoters using flow cytometry**

442 Green and red fluorescence, as well as forward scatter of the cells in the DH5α-transformed
443 random promoter library were measured using a flow cytometer (FACSAria III, BD, USA).
444 Gates were set for the green and red channels to exclude cells with low GFP and high mCherry
445 fluorescence, respectively. Subsequently, 10,000 cells without red fluorescence but with green
446 fluorescence within the selected range (**Fig. 2B**) were collected by cell sorting, according to the

447 manufacturer's instructions. The sorted cells were spread on LB/Km agar plates and incubated
448 overnight at 37 °C. Bacterial colonies were randomly picked and their inserted synthetic
449 promoter regions were sequenced by Sanger sequencing. Consequently, 16 synthetic promoters
450 without improper mutations at the restriction sites were identified and named SP1–16,
451 (**supplementary table S1**).

452 **Random mutagenesis using error-prone PCR**

453 Error-prone PCR was conducted using 0.1 ng of each promoter region as a template, with 35
454 cycles of amplification performed using the GeneMorph II Random Mutagenesis Kit (Agilent
455 Technologies, USA, 200550) and primers C and D. The resulting PCR products underwent gel
456 purification and were double digested with BamHI and XhoI for 3 hours at 37 °C, followed by
457 column purification to create mutant inserts. Ligation was performed overnight at 16 °C using
458 T4 DNA Ligase, with the vector and mutant inserts mixed at a molar ratio of 1:3. The ligation
459 reaction solution was column purified, and DNA was extracted using 20 μL of sterile water.

460 **Construction of mutant promoter library**

461 The purified ligation product solution was added to 50 μL of DH5α Electro-Cells and
462 electroporated at 1800 V. Transformed cells were cultured at 37 °C for 1 h in 1 mL of SOC
463 medium and a portion of the culture was plated on LB/Km agar medium for overnight growth
464 at 37 °C. Bacterial colonies were counted to determine the transformation efficiency. If the
465 transformation efficiency exceeded 10^5 cells/reaction, the next step was performed; otherwise,

466 the process was repeated. The transformed culture was diluted with 5 mL of LB/Km medium
467 and grown overnight. Plasmid extraction from an overnight culture was performed using the
468 FastGene Plasmid Mini Kit (NIPPON Genetics, Japan, FG-90502). The obtained plasmid
469 solution was added to 50 μL of MG1655 electrocompetent cells and electroporated at 1800 V.
470 Transformed cells were cultured at 37 °C for 1 h in 1 mL of SOC medium, and a portion of the
471 solution was plated on LB/Km agar medium for overnight growth at 37 °C to count the number
472 of colonies and determine transformation efficiency. If the transformation efficiency exceeded
473 10^5 cells/reaction, the next step was performed; otherwise, the process was repeated. A sterile
474 2 mL 96 well deep well plate (Greiner Bio-One, Austria, 780271) was used to add 1 mL of the
475 synthetic medium to each well, followed by the addition of 10 μL of the overnight culture
476 containing the MG1655 transformants. The bacterial cells were grown at 37 °C with shaking at
477 800 rpm until confluent growth was achieved. Subsequently, another plate was prepared with 1
478 mL of fresh synthetic medium added to each well, and 10 μL of the confluently grown cultures
479 was diluted into each well. The cultures were grown for 2.5 h at 37 °C with shaking at 800 rpm
480 and were used as MT for flow cytometry.

481 **Verification of PCR random mutagenesis**

482 The number of mutations introduced into the promoters by error-prone PCR was determined
483 using Sanger sequencing. Three to four colonies were randomly selected from six randomly
484 selected mutant libraries, including essential and nonessential promoters (**supplementary table**

485 **S2**). Colony PCR was performed using KOD One PCR Master Mix Blue with primers E
486 (TTACTTTGCAGGGCTTCCCAA) and F (CCAGTCTTTCGACTGAGCCT) to amplify the
487 regions including the promoter regions (**supplementary fig. S1**). PCR products were column
488 purified and verified by Sanger sequencing. The average mutation rate was calculated to be 6.2
489 mutations/kb.

490 **Construction of WT promoter control group**

491 To ensure accurate measurement of GFP fluorescence from cells with wild-type promoters,
492 wild-type plasmids were transformed into MG1655 cells by electroporation. Transformed
493 cultures were plated on LB/Km agar medium and incubated overnight, and the resulting
494 colonies were cultured in LB/Km medium. Cell cultures of these control groups in the synthetic
495 medium were prepared and used as WT for flow cytometry.

496 **Flow cytometry**

497 Bacterial cells were dispensed in M9 buffer prior to analysis using a flow cytometer (BD, USA,
498 FACSAria III). Fluorescence intensity from GFP (GFP FI) and mCherry (mCherry FI) was
499 measured using a 488 nm laser with 515–545-nm emission filter, and a 561 nm laser with 563–
500 588-nm emission filter, respectively. The following PMT voltage settings were used: forward
501 scatter (FSC), 300; side scatter (SSC), 340; green fluorescence, 700; red fluorescence, 600. The
502 events exceeding the defined thresholds of FSC (200) and SSC (200) were recorded.
503 Tota1,000,000 events were measured per library. The MG1655 strain was used as a control for

504 autofluorescence of GFP FI.

505 **Data analysis of flow cytometry**

506 FCS files were converted to CSV format using the flowCore(Ellis, et al. 2024) package in R(R
507 Core Team 2023). Subsequent analyses were conducted using custom R scripts. To compare
508 green fluorescence among cells in similar physiological states, a narrow gate for the forward
509 scatter channel was set to capture 40% of the events, including the most frequent values for
510 MG1655 (**supplementary fig. S4A**). The top 0.5% value of the green fluorescence intensity
511 distribution of MG1655 within the gate was determined as the autofluorescence intensity
512 (**supplementary fig. S4B**). For the events of WT and MT, the autofluorescence intensity was
513 subtracted from the green fluorescence intensity of the events within the gate. Events with a
514 negative green fluorescence intensity after subtracting the autofluorescence intensity were
515 excluded from the analysis. In addition, events with a red fluorescence intensity greater than
516 10^3 [a.u.] were regarded as the cells carrying pUA66-mCherry and were excluded from the
517 analysis. The green fluorescence intensity after subtracting the autofluorescence intensity was
518 $\log_{10}$-transformed and used to calculate the means and standard deviations.

519 **Classification of promoter based on essentiality of transcribed unit**

520 Gene essentiality was based on a previous study(Goodall, et al. 2018), which identified 248
521 genes designated as essential across different datasets. Promoters containing at least one
522 essential gene in the transcription units were defined as essential promoters. The known

523 transcription units were retrieved from the RegulonDB(Tierrafria, et al. 2022) database. Of the
524 56 promoters, 18 were classified as essential and 38 as nonessential as listed in **supplementary**
525 **table S1**.

526 **Estimation of transcriptional variability**

527 Transcriptional variability of each gene in *E. coli* in response to different genetic perturbations
528 was estimated using known transcriptome profiles obtained from six independent mutants
529 derive from MG1655 cultured under a single environmental condition, termed the Mut
530 dataset(Tsuru and Furusawa 2024). These mutants were subjected to mutation accumulation
531 (MA) experiments to accumulate genomic mutations via genetic drift(Halligan and Keightley
532 2009; Tsuru, et al. 2015). In addition, to quantify the transcriptional variability in response to
533 different environmental conditions, known transcriptome profiles of MG1655 cultured under
534 76 different environmental conditions(Sastry, et al. 2019; Rychel, et al. 2021; Lamoureux, et al.
535 2022), termed the Env dataset(Tsuru and Furusawa 2024), were utilized. The expression levels
536 were determined by $\log_2$-transformation and quantile normalization. The means and standard
537 deviations in expression levels for each gene across different genetic perturbations and
538 environmental conditions were quantified. To compensate for the mean-dependency of the
539 standard deviations (**Fig. 1B**,**G**), the DM values for each gene were calculated as the vertical
540 distance of each standard deviation from a smoothed running median of the standard deviations
541 for each dataset, as detailed previously(Tsuru and Furusawa 2024). The resulting DM values

542 were termed DM_{mut} and DM_{env} for the Mut and Env datasets, respectively. The DM values were
543 used as measures of transcriptional variability in response to genetic and environmental
544 perturbations.

545 **Phylogenetic analysis of promoter sequence of length 73 bases**

546 Synthetic promoters underwent a blastn(Altschul, et al. 1990) search to confirm that there was
547 no considerable similarity to any natural sequences (**supplementary note**). Posterior sequences
548 of 73 bases were aligned using the msa(Bodenhofer, et al. 2015) package in R (Clustal Omega),
549 where 73 bases represented the shortest length of a synthetic promoter. A phylogenetic tree was
550 constructed using the Neighbor-Joining method.

551 **Phylogenetic analysis of core elements of sigma70 promoters**

552 We identified sigma70 promoters from our natural promoters using RegulonDB. The core
553 elements comprising the –10/–35 elements of these promoters were concatenated and subjected
554 to phylogenetic analysis. The Levenshtein distance(Levenshtein 1966) between the focal
555 elements and the consensus elements, which consist of TTGACA and TATAAT, was calculated.
556 A phylogenetic tree was constructed as previously described.

557 **Identification of the number of transcriptional regulators regulating the natural**
558 **promoters**

559 Reported regulatory interactions between genes were compiled using all interactions with
560 experimental evidence from RegulonDB and EcoCyc(Keseler, et al. 2021) for both transcription

factors and sigma factors. The mode of regulation was classified as either activation or inhibition, and the mode of regulation by sigma factors was classified as activation.

Numerical simulation of promoter evolution

The promoter genotype, g, was modeled by a bit string comprising either zeros or ones of length N=10 (e.g., 1110001100). The expression level, E, was defined by the Rough Mt. Fuji (RMF) model(Aita, et al. 2000; Neidhart, et al. 2014) as:

$$E(g) = -cD(g, g^*) + \eta(g),$$

where c is a scaling coefficient equal to one in this study, g^* is the genotype with the highest expression level before adding noise, $D(g, g^*)$ is the Hamming distance between g and g^*, and η is an independent and identically distributed random variable. A standard normal distribution with a mean equal to zero and a standard deviation equal to one for η was used to generate a rugged expression landscape. The expression landscape was then normalized using the minimal and maximal values of $E(g)$ to range from zero to one, as follows:

$$\hat{E}(g) = \frac{E(g) - min\big(E(g)\big)}{max\big(E(g)\big) - min\ \big(E(g)\big)}\ .$$

The fitness of a given genotype g was defined as:

$$F(g) = (1 - F_b)\hat{E}(g) + F_b\ ,$$

where $1 \geq F_b \geq 0$. F_b represents the basal fitness when the expression level is zero and determines the strength of the selection pressure acting on the expression level. Different values of F_b (0.3, 0.5, and 0.99) were used to demonstrate the evolution under strong positive selection, moderate

580 positive selection, and nearly neutral evolution, respectively. It was assumed that essential
581 genes underwent strong positive selection on their expression levels, while nonessential genes
582 underwent relatively moderate selection for their expression levels, based on the empirical fact
583 that knock-down of essential genes was more harmful for cell growth than nonessential
584 genes(Li, et al. 2016). Ten independent fitness landscapes were constructed for each F_b.
585 Evolutionary simulation on the fitness landscapes were performed using the standard Wright-
586 Fisher model(Otto and Day 2007) assuming a haploid asexual population with a population size
587 of 10^5 and a mutation rate of 10^{-3}. The initial genotypes were randomly selected for each
588 simulation, and 600 independent simulations were examined for each fitness landscape. For
589 each generation, mutants with genotypes differing by one Hamming distance from the current
590 population were generated at the mutation rate. The frequency of each genotype within the
591 population was updated based on the fitness landscape. Subsequently, a new population for the
592 next generation was formed by sampling mutants, with replacement, from the current
593 population using a binomial distribution with the given population size. The most frequent
594 mutants in a population after 30 generations were isolated in each simulation, resulting in 6,000
595 mutants for each F_b. Mutational changes in expression levels were calculated for each isolate
596 by averaging the expression levels of all nearest neighboring genotypes that differed by one
597 Hamming distance from the focal isolate. Custom R scripts were used for the numerical
598 simulations, modified from scripts constructed in a previous study(Obolski, et al. 2018; Song

599 and Zhang 2021).

600 **Data visualization**

601 All figures were generated in R and the following packages were used. Phylogenetic trees of

602 the promoters were visualized using the ape(Paradis and Schliep 2019) and ggtree(Yu, et al.

603 2017) packages. Other illustrative plots were generated using the ggplot2(Wickham 2016) and

604 ggpubr(Kassambara 2023) packages.

Data Availability

Custom R scripts used in the manuscript are available at GitHub (https://github.com/tsuruubi/Variability_PromoterAct_E_coli).

The raw fcs data are available at Zenodo (WT for https://doi.org/10.5281/zenodo.11108555; MT for https://doi.org/10.5281/zenodo.11108786).

Acknowledgments

The authors thank Hiroyo Koike for technical assistance of Sanger sequencing. The authors also appreciate Dr. Olin Silander and Dr. Luise Wolf for fruitful technical advice on the construction of a synthetic promoter library. The pNS2-σVL plasmid was a gift from Dr. Michael Elowitz (Addgene, plasmid # 26756).

Funding

This study was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (18H02427, 22H05403 to S.T.;17H06389 to C.F. and S.T.; 22K21344 to C.F.), and the Japan Science and Technology Agency (JST) ERATO (JPMJER1902 to S.T. and C.F.).

References

Aita T, Uchiyama H, Inaoka T, Nakajima M, Kokubo T, Husimi Y. 2000. Analysis of a local fitness landscape with a model of the rough Mt. Fuji-type landscape: application to prolyl endopeptidase and thermolysin. Biopolymers 54:64-79.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403-410.

Belliveau NM, Barnes SL, Ireland WT, Jones DL, Sweredoski MJ, Moradian A, Hess S, Kinney JB, Phillips R. 2018. Systematic approach for dissecting the molecular mechanisms of transcriptional regulation in bacteria. Proc Natl Acad Sci U S A 115:E4796-E4805.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57:289-300.

Bilu Y, Barkai N. 2005. The design of transcription-factor binding sites is affected by combinatorial regulation. Genome Biol 6:R103.

Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, et al. 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277:1453-1462.

639 Bodenhofer U, Bonatesta E, Horejs-Kainrath C, Hochreiter S. 2015. msa: an R package for
640 multiple sequence alignment. Bioinformatics 31:3997-3999.

641 Cui L, Vigouroux A, Rousset F, Varet H, Khanna V, Bikard D. 2018. A CRISPRi screen in *E.*
642 *coli* reveals sequence-specific toxicity of dCas9. Nature Communications 9:1-10.

643 Darwin C. 1859. On the origin of species by means of natural selection, or preservation of
644 favoured races in the struggle for life: London : John Murray.

645 de Visser JA, Hermisson J, Wagner GP, Ancel Meyers L, Bagheri-Chaichian H, Blanchard JL,
646 Chao L, Cheverud JM, Elena SF, Fontana W, et al. 2003. Perspective: Evolution and
647 detection of genetic robustness. Evolution 57:1959-1972.

648 Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK. 2005. The
649 transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*.
650 Nat Genet 37:544-548.

651 Draghi JA, Whitlock MC. 2012. Phenotypic plasticity facilitates mutational variance, genetic
652 variance, and evolvability along the major axis of environmental variation. Evolution
653 66:2891-2902.

654 Dunlop MJ, Cox RS, Levine JH, Murray RM, Elowitz MB. 2008. Regulatory activity
655 revealed by dynamic correlations in gene expression noise. Nature Genetics 40:1493-1498.

656 Ellis B, Haaland P, Hahne F, Meur L, Nolwenn, Gopalakrishnan N, Spidlen J, Jiang M, Finak
657 G, Granjeaud S. 2024. flowCore: flowCore: Basic structures for flow cytometry data. R
658 package version 2.14.2:https://bioconductor.org/packages/flowCore/.

659 Furusawa C, Kaneko K. 2018. Formation of dominant mode by evolution in biological
660 systems. Phys Rev E 97:042410.

661 Goodall ECA, Robinson A, Johnston IG, Jabbari S, Turner KA, Cunningham AF, Lund PA,
662 Cole JA, Henderson IR. 2018. The essential genome of *Escherichia coli* K-12. mBio 9:1-
663 18.

664 Halligan DL, Keightley PD. 2009. Spontaneous mutation accumulation studies in
665 evolutionary genetics. Annual Review of Ecology, Evolution, and Systematics 40:151-172.

666 Ho WC, Zhang J. 2016. Adaptive Genetic Robustness of *Escherichia coli* Metabolic Fluxes.
667 Molecular Biology and Evolution 33:1164-1176.

668 Ho WC, Zhang J. 2014. The genotype-phenotype map of yeast complex traits: basic
669 parameters and the role of natural selection. Mol Biol Evol 31:1568-1580.

670 Ireland WT, Beeler SM, Flores-Bautista E, McCarty NS, Roschinger T, Belliveau NM,
671 Sweredoski MJ, Moradian A, Kinney JB, Phillips R. 2020. Deciphering the regulatory
672 genome of *Escherichia coli*, one hundred promoters at a time. Elife 9.

673 Kaneko K. 2009. Relationship among phenotypic plasticity, phenotypic fluctuations,
674 robustness, and evolvability; Waddington's legacy revisited under the spirit of Einstein. J
675 Biosci 34:529-542.

676 Kassambara A. 2023. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version
677 0.6.0:https://rpkgs.datanovia.com/ggpubr/.

678 Kauffman SA, Weinberger ED. 1989. The NK model of rugged fitness landscapes and its
679 application to maturation of the immune response. J Theor Biol 141:211-245.

680 Keseler IM, Gama-Castro S, Mackie A, Billington R, Bonavides-Martinez C, Caspi R,
681 Kothari A, Krummenacker M, Midford PE, Muniz-Rascado L, et al. 2021. The EcoCyc
682 Database in 2021. Front Microbiol 12:711077.

683 Kinney JB, Murugan A, Callan CG, Jr., Cox EC. 2010. Using deep sequencing to characterize
684 the biophysical mechanism of a transcriptional regulatory sequence. Proc Natl Acad Sci U
685 S A 107:9158-9163.

686 Lamoureux CR, Decker KT, Sastry AV, Rychel K, Gao Y, McConn JL, Zielinski DC, Palsson
687 BO. 2022. A multi-scale transcriptional regulatory network knowledge base for *Escherichia*
688 *coli*. bioRxiv:2021.2004.2008.439047.

689 Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. 2007. Genetic properties
690 influencing the evolvability of gene expression. Science 317:118-121.

691 Levenshtein VI editor. Soviet physics doklady. 1966.

692 Li XT, Jun Y, Erickstad MJ, Brown SD, Parks A, Court DL, Jun S. 2016. tCRISPRi: tunable
693 and reversible, one-step control of gene expression. Sci Rep 6:39076.

694 Lutz R. 1997. Independent and tight regulation of transcriptional units in *Escherichia coli* via
695 the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Research
696 25:1203-1210.

697 Lynch M, Walsh B. 1998. Genetics and analysis of quantitative traits: Sinauer Sunderland,
698 MA.

699 Meiklejohn CD, Hartl DL. 2002. A single mode of canalization. Trends in Ecology &
700 Evolution 17:468-473.

Neidhart J, Szendro IG, Krug J. 2014. Adaptation in tunably rugged fitness landscapes: the rough Mount Fuji model. Genetics 198:699-721.

Noble DWA, Radersma R, Uller T. 2019. Plastic responses to novel environments are biased towards phenotype dimensions with high additive genetic variation. Proc Natl Acad Sci U S A 116:13452-13461.

Obolski U, Ram Y, Hadany L. 2018. Key issues review: evolution on rugged adaptive landscapes. Rep Prog Phys 81:012602.

Otto SP, Day T. 2007. A Biologist's Guide to Mathematical Modeling in Ecology and Evolution: Princeton University Press.

Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526-528.

Payne JL, Wagner A. 2015. Mechanisms of mutational robustness in transcriptional regulation. Front Genet 6:322.

R Core Team. 2023. R: A Language and Environment for Statistical Computing.

Reynolds R, Bermúdez-Cruz RM, Chamberlin MJ. 1992. Parameters affecting transcription termination by *Escherichia coli* RNA polymerase. Journal of Molecular Biology 224:31-51.

717 Rifkin SA, Houle D, Kim J, White KP. 2005. A mutation accumulation assay reveals a broad
718 capacity for rapid evolution of gene expression. Nature 438:220-223.

719 Ringquist S, Shinedling S, Barrick D, Green L, Binkley J, Stormo GD, Gold L. 1992.
720 Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site.
721 Molecular Microbiology 6:1219-1229.

722 Rychel K, Decker K, Sastry AV, Phaneuf PV, Poudel S, Palsson BO. 2021. iModulonDB: a
723 knowledgebase of microbial transcriptional regulation derived from machine learning.
724 Nucleic Acids Res 49:D112-D120.

725 Sastry AV, Gao Y, Szubin R, Hefner Y, Xu S, Kim D, Choudhary KS, Yang L, King ZA,
726 Palsson BO. 2019. The *Escherichia coli* transcriptome mostly consists of independently
727 regulated modules. Nat Commun 10:5536.

728 Siegal ML, Bergman A. 2002. Waddington's canalization revisited: developmental stability
729 and evolution. Proc Natl Acad Sci U S A 99:10528-10532.

730 Silander OK, Nikolic N, Zaslaver A, Bren A, Kikoin I, Alon U, Ackermann M. 2012. A
731 genome-wide analysis of promoter-mediated phenotypic noise in *Escherichia coli*. PLoS
732 Genet 8:e1002443.

Smith JM, Burian R, Kauffman S, Alberch P, Campbell J, Goodwin B, Lande R, Raup D, Wolpert L. 1985. Developmental constraints and evolution: a perspective from the Mountain Lake conference on development and evolution. The Quarterly Review of Biology 60:265-287.

Song S, Zhang J. 2021. Unbiased inference of the fitness landscape ruggedness from imprecise fitness estimates. Evolution 75:2658-2671.

Szollosi GJ, Derenyi I. 2009. Congruent evolution of genetic and environmental robustness in micro-RNA. Mol Biol Evol 26:867-874.

Tierrafria VH, Rioualen C, Salgado H, Lara P, Gama-Castro S, Lally P, Gomez-Romero L, Pena-Loredo P, Lopez-Almazo AG, Alarcon-Carranza G, et al. 2022. RegulonDB 11.0: Comprehensive high-throughput datasets on transcriptional regulation in *Escherichia coli* K-12. Microb Genom 8.

Tirosh I, Weinberger A, Carmi M, Barkai N. 2006. A genetic signature of interspecies variations in gene expression. Nat Genet 38:830-834.

Tsuru S, Furusawa C. 2024. Genetic properties underlying transcriptional variability in *Escherichia coli*. bioRxiv:2024.2004.2015.589659.

749 Tsuru S, Ishizawa Y, Shibai A, Takahashi Y, Motooka D, Nakamura S, Yomo T. 2015.
750 Genomic confirmation of nutrient-dependent mutability of mutators in *Escherichia coli*.
751 Genes Cells 20:972-981.

752 Uller T, Moczek AP, Watson RA, Brakefield PM, Laland KN. 2018. Developmental Bias and
753 Evolution: A Regulatory Network Perspective. Genetics 209:949-966.

754 Vaishnav ED, de Boer CG, Molinet J, Yassour M, Fan L, Adiconis X, Thompson DA, Levin
755 JZ, Cubillos FA, Regev A. 2022. The evolution, evolvability and engineering of gene
756 regulatory DNA. Nature 603:455-463.

757 Waddington CH. 1942. Canalization of development and the inheritance of acquired
758 characters. Nature 150:563-565.

759 Waddington CH. 1957. The strategy of the genes: Routledge.

760 Walker JA. 2007. A general model of functional constraints on phenotypic evolution. Am Nat
761 170:681-689.

762 Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis.https://ggplot2.tidyverse.org.

763 Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in cis and trans gene
764 regulation. Nature 430:85-88.

765 Wolf L, Silander OK, van Nimwegen E. 2015. Expression noise facilitates the evolution of
766 gene regulation. Elife 4:1-48.

767 Yona AH, Alm EJ, Gore J. 2018. Random sequences rapidly evolve into de novo promoters.
768 Nat Commun 9:1530.

769 Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an r package for visualization and
770 annotation of phylogenetic trees with their covariates and other associated data. Methods in
771 Ecology and Evolution 8:28-36.

772 Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, Liebermeister W, Surette MG,
773 Alon U. 2006. A comprehensive library of fluorescent transcriptional reporters for
774 *Escherichia coli*. Nature Methods 3:623-628.

775 Zheng W, Gianoulis TA, Karczewski KJ, Zhao H, Snyder M. 2011. Regulatory variation
776 within and between species. Annu Rev Genomics Hum Genet 12:327-346.

777

Figures and Figure legends

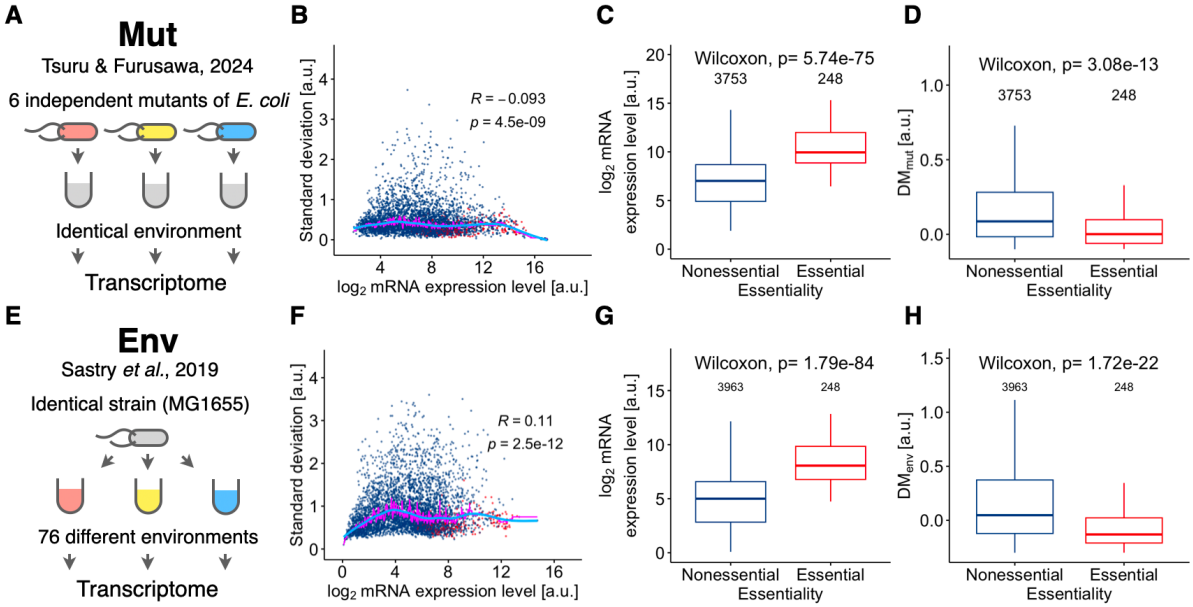

Fig. 1: Transcriptional variability of essential genes in *E. coli*.

(**A**–**D**) Transcriptional profiles of six independent mutants of *E. coli* cultured under identical environmental conditions, termed the Mut dataset. The mutants were created through the MA experiment to accumulate random genomic mutations by genetic drift. (**E**–**H**) Transcriptional profiles of a single strain, MG1655, cultured under 76 different environmental conditions, termed the Env dataset. (**A**, **E**) Schematics of the datasets. (**B**, **F**) Relationship between the means and standard deviations of mRNA expression levels for thousands of genes. Red and blue dots represent essential and nonessential genes, respectively. The magenta and cyan lines represent running medians and their smoothed splines, respectively. Spearman's rank correlation coefficients (R) and p-values are shown. (**C**, **G**) Mean expression levels of essential and nonessential genes. (**D**, **H**) Transcriptional variability, represented as either DM_{mut} or DM_{env}, of essential and nonessential genes. The DM values were calculated as the vertical distance of

the standard deviations from the smoothed splines of the running median of standard deviations in panels **B** and **F**. The lower and upper edges of the boxes in panels **C**, **D**, **G**, **H** represent the first (q1) and third (q3) quartiles, respectively. The horizontal lines in the boxes represent the medians (m). The whiskers from the boxes extend to the most extreme observed values inside inner fences, m±1.5(q3–q1).

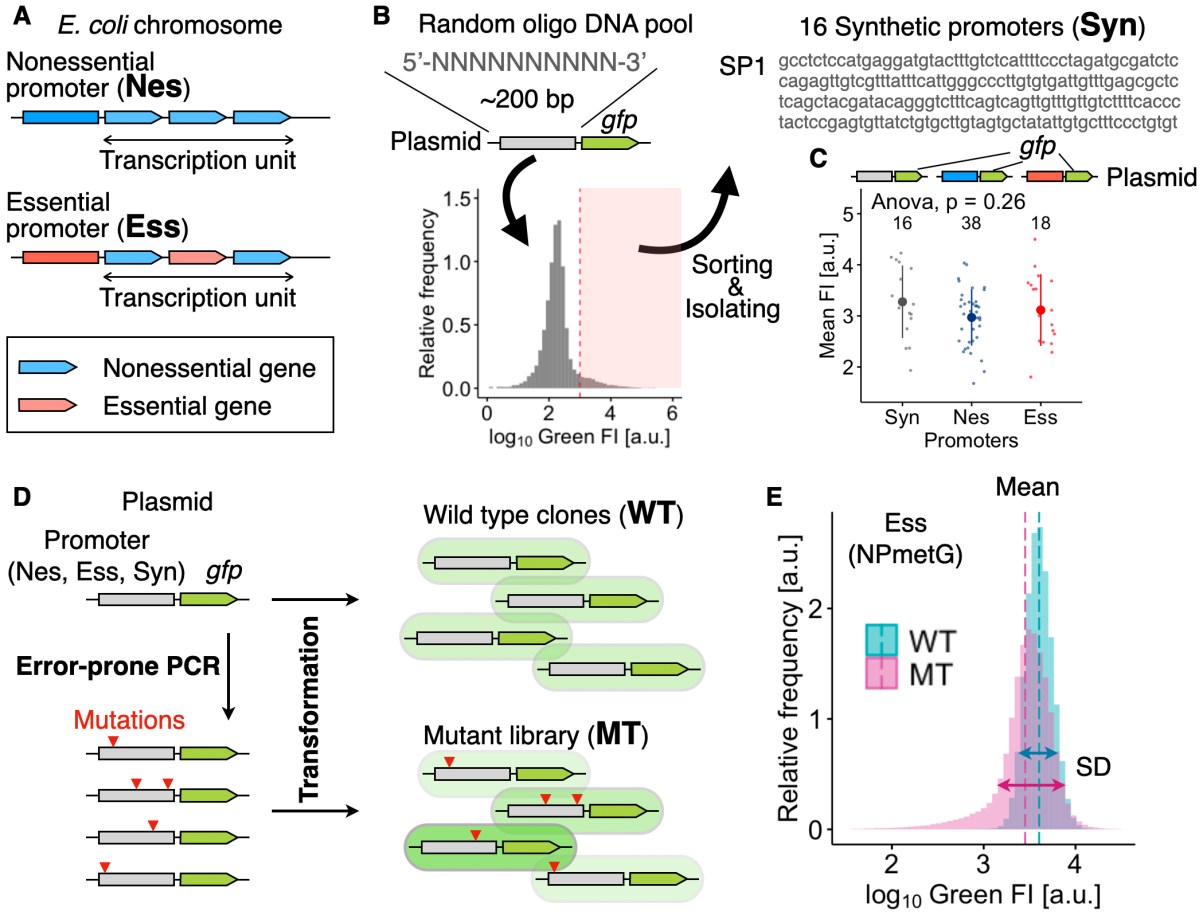

Fig. 2: Experimental design to measure expression levels of mutant promoter library. (**A**) Natural promoters of *E. coli* were categorized into nonessential and essential promoters, termed Nes and Ess, respectively. Nonessential promoters control transcriptional units comprising only nonessential genes, while essential promoters regulate transcriptional units containing at least one essential gene. Gene essentiality was based on the requirement for cell growth, as defined by Goodall et al(Goodall, et al. 2018). The promoter regions were obtained from the *E. coli* Promoter Collection(Zaslaver, et al. 2006). (**B**) Synthetic promoters, termed Syn, were derived from a random oligo DNA pool. Transformants harboring a plasmid on which the random sequences were placed upstream of the *gfp* gene were analyzed using flow

807 cytometry. The 16 unique transformants, satisfying a defined green fluorescence (red shaded
808 area), were isolated. The obtained synthetic promoters were named SP1–16. (**C**) Natural and
809 synthetic promoters were placed upstream of *gfp* on a plasmid. Transformants harboring each
810 of these plasmids were analyzed using flow cytometry. Mean green fluorescence of each
811 population is shown as small dots. The large dots and error bars represent the means and
812 standard deviations within the groups, respectively. The number of promoters is indicated inside
813 the panel. The p-value corresponds to ANOVA. (**D**) Schematic showing construction of mutant
814 library from a given promoter region. The promoter region was subjected to random
815 mutagenesis by error-prone PCR. A mutant library (MT) was created through transformation to
816 yield more than 100,000 genetically-different mutants in a population. Unmutagenized plasmid
817 was also transformed to create wild type clones (WT). (**E**) Representative examples of $\log_{10}$-
818 transformed green fluorescence intensity distributions of WT and MT for the promoter region
819 of the *metG* gene in *E. coli* (NPmetG). The means (vertical lines) and standard deviations
820 (horizontal arrows) were calculated in each population.

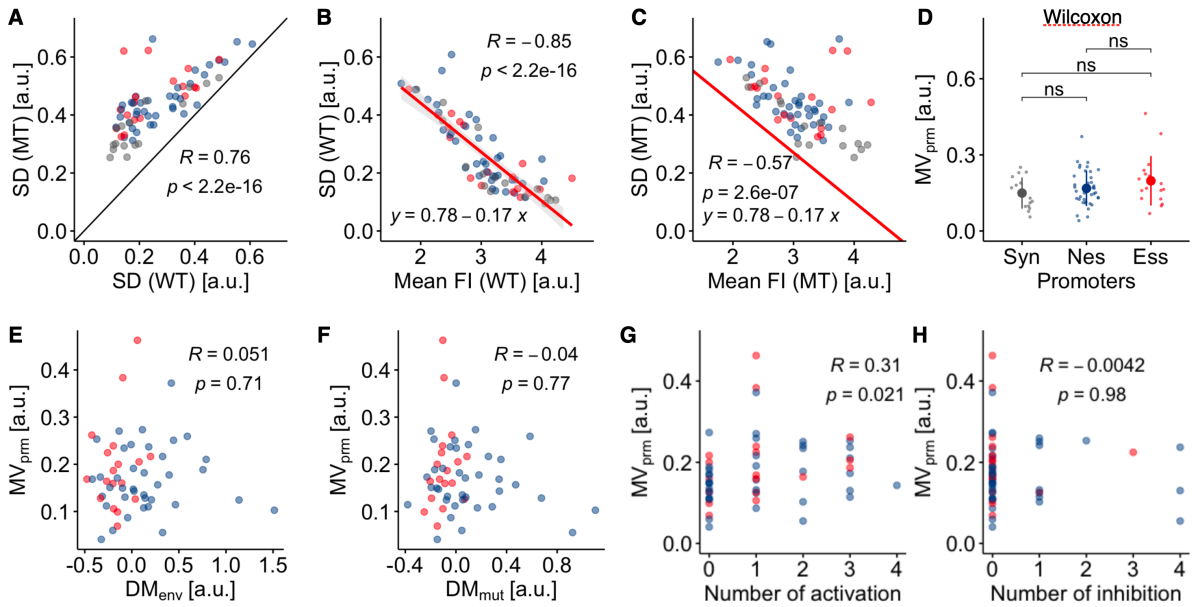

Fig. 3: Mutational variability in promoter activity.

(**A**) Comparison of the standard deviations in the GFP distributions between WT and MT. The solid line represents y=x. (**B**, **C**) Relationship between the mean and standard deviation of the GFP distributions for WT (**B**) and MT (**C**). Red lines represent a regression line for WT and were used to compensate for the mean-dependency of the standard deviations of both WT and MT by subtraction. (**D**) Mutational variability in promoter activity, term MV_{prm}, is shown for each promoter group. The MV_{prm} was calculated by subtracting the compensated standard deviations of WT from those of MT. The large dots and error bars are the means and standard deviations, respectively. Pairwise Wilcoxon test was examined, where ns represents no significance. (**E**, **F**) The relationships between MV_{prm} and the DM values (DM_{env} for **E** and DM_{mut} for **F**). (**G**, **H**) Relationship between MV_{prm} and the number of known activation (**G**) and inhibition (**H**) of natural promoters by transcriptional regulators. The relationship between promoters and known transcriptional regulators were based on RegulonDB(Tierrafria, et al.

835 2022) and EcoCyc(Keseler, et al. 2021). The grey, blue, and red dots in each panel represent

836 synthetic (Syn), nonessential (Nes), and essential (Ess) promoters, respectively. Spearman's R

837 and p-values are shown in panes **A**–**C** and **E**–**H**.

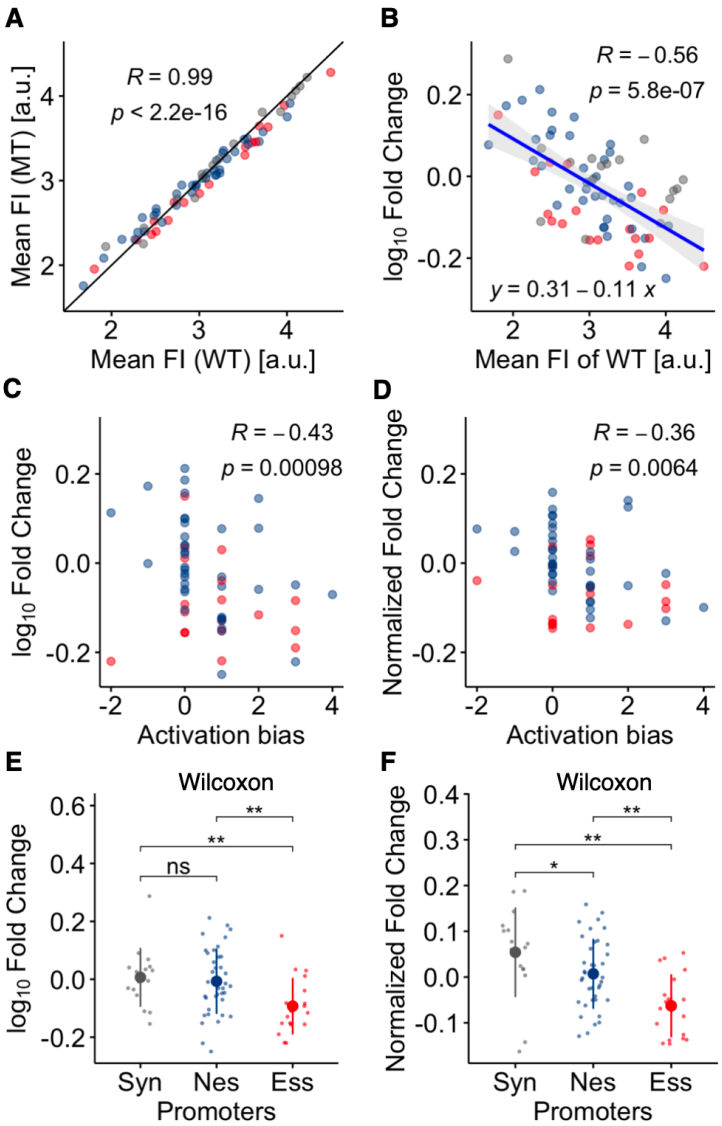

Fig. 4: Biased mutational changes in activity of essential promoters.

(**A**) Comparison of the means in the GFP distributions between WT and MT. The solid line represents y=x. (**B**) Relationship between $\log_{10}$-fold change of mean expression levels from WT to MT and the expression levels of WT. The blue line represents a regression line used to compensatie the mean-dependency of the $\log_{10}$-fold changes by subtraction. (**C**, **D**) Relationship between activation bias in natural promoters and the unnormalized (**C**) and normalized (**D**) $\log_{10}$-fold changes. The activation bias was calculated by subtracting the

number of known inhibitions from those of activations by transcriptional regulators. Spearman's R and p-values are shown in panes **A–D**. Grey, blue, and red dots in each panel represent synthetic (Syn), nonessential (Nes), and essential (Ess) promoters, respectively. (**E**, **F**) Unnormalized (**E**) and normalized (**F**) $\log_{10}$-fold changes for each promoter group. The large dots and error bars are the means and standard deviations, respectively. Asterisks represent significance levels in pairwise Wilcoxon test (ns: $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$). The p-values were adjusted using the BH method(Benjamini and Hochberg 1995).

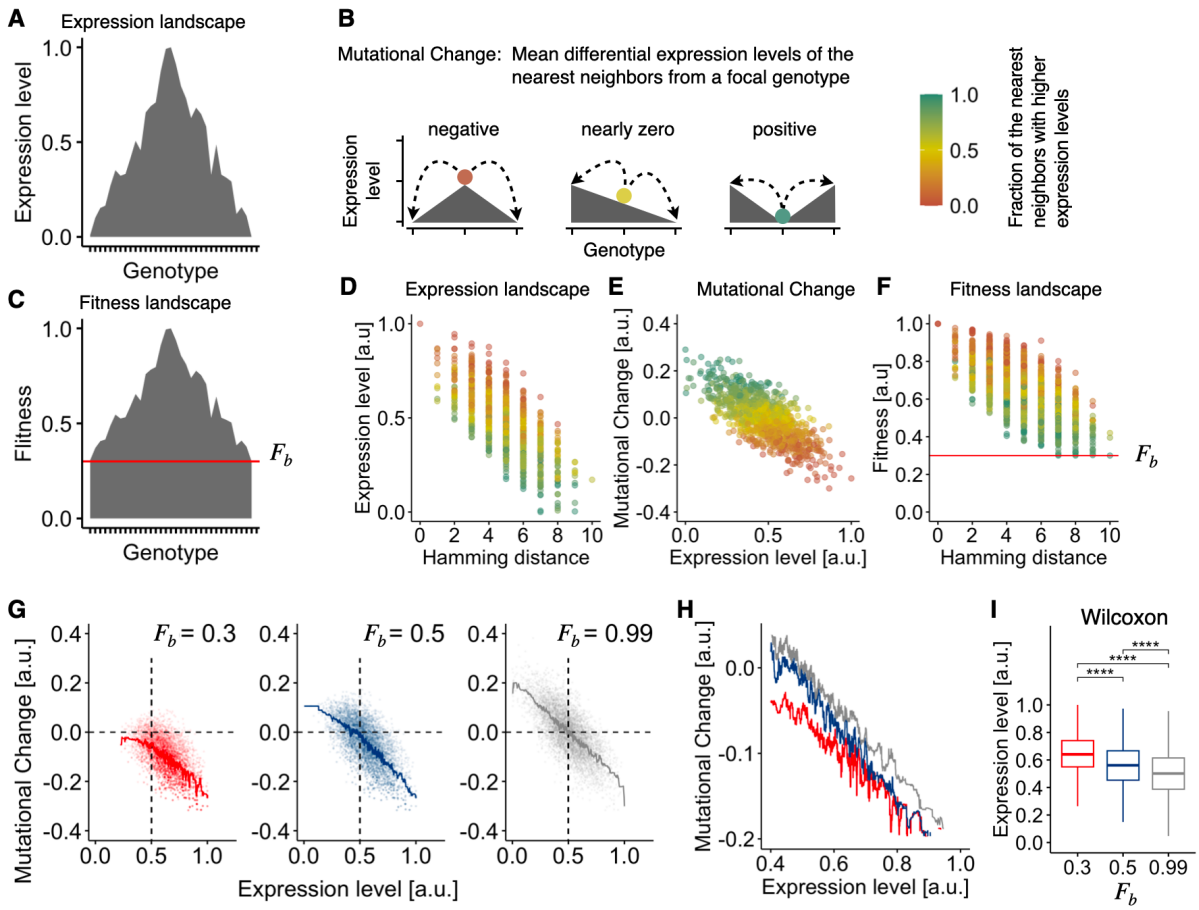

Fig. 5: Numerical simulation of promoter evolution on a ragged fitness landscape.

(**A**–**C**) Schematics of landscapes of expression levels (**A**, **B**) and fitness (**C**). (**A**) Expression landscape was constructed by the rough Mt. Fuji (RNF) model. (**B**) Mutational change of each genotype was calculated as the mean of differential expression levels of the nearest neighboring genotypes (arrow heads, MT) from the focal genotype (colored circles, WT). The nearest neighbors differ from the focal genotype by one Hamming distance. The color bar represents the fraction of the genotypes with higher expression levels than the focal genotype among the nearest neighbors. (**C**) Fitness landscape was modeled by adding basal fitness (F_b, red line) to the expression landscape and normalizing it. (**D**–**F**) Representatives of expression landscape (**D**), mutational change (**E**), and fitness landscape (**F**). Genotype was modeled as 10-bit binary

numbers. Hamming distance represents the distance from the genotype with the global optima in expression level and fitness. (**G**) Mutational changes of the evolved isolates obtained through evolutionary simulation using the standard Wright-Fisher model. The basal fitness was varied by mimicking strong positive (F_b=0.3, left), moderate positive (F_b=0.5, middle), and nearly neutral selection (F_b=0.99, right) acting upon higher expression levels. Ten independent expression/fitness landscapes were generated for each condition of basal fitness. In each fitness landscape, 600 independent evolutionary simulations were examined where initial genotypes were randomly selected. Dots represents the most abundant mutants within the population after 30 generations in each run. Solid lines represent the running medians. Dashed lines are guides for the eye. (**H**) Enlarged figure of the running medians in (**G**). (**I**) Expression level of the evolved isolates. Asterisks represent significance levels of the adjusted p-value (the BH method) in the Wilcoxon test (****: $p \leq 0.0001$).