Promoters constrain 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 past selection created non-uniform mutational variability in promoters biased toward lower activities of random mutants now constrain future evolution of downstream essential genes toward higher expression levels.

Introduction

The rate and direction of phenotypic evolution largely depend on phenotypic variability, the tendency to vary phenotypically (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.

56 What molecular mechanisms account for gene-to-gene differences in transcriptional
57 variability? In principle, both *cis*- and *trans*-acting elements can influence transcriptional
58 variability(Wittkopp, et al. 2004; Tirosh, et al. 2006; Landry, et al. 2007; Payne and Wagner
59 2015; Tsuru and Furusawa 2024). Previous studies(Denver, et al. 2005; Rifkin, et al. 2005;
60 Landry, et al. 2007; Tsuru and Furusawa 2024) explored the transcriptional variability against
61 different genetic perturbations using mutation accumulation (MA) experiments(Halligan and
62 Keightley 2009), where independent mutants were generated by accumulating genomic
63 mutations through genetic drift, and their transcriptome profiles were obtained under identical
64 environmental conditions. Although MA experiments provide a powerful method to quantify
65 transcriptional variability in the absence of strong natural selection, the indiscriminate
66 accumulation of mutations on *cis*- and *trans*-acting elements(Tsuru, et al. 2015) complicates the
67 identification of the molecular mechanism underlying transcriptional variability. Accordingly,
68 to overcome this difficulty, it is necessary to consider the contributions of *cis*- and *trans*-acting
69 elements separately. Although such approaches have been extensively employed in
70 yeast(Hornung, et al. 2012; Metzger, et al. 2015; Duveau, et al. 2021), it remains largely
71 unknown which elements practically contribute to gene-to-gene differences in transcriptional
72 variability associated with functions of gene products.

73 Why have genes related to cellular growth and maintenance acquired low mutational
74 variability? The evolutionary origin of mutational variability is debatable (de Visser, et al. 2003).

75 Mutational variability could result from direct natural selection for mutational
76 robustness(Waddington 1942), or could be a byproduct of selection for robustness against
77 environmental perturbations(Meiklejohn and Hartl 2002; de Visser, et al. 2003). Several
78 biological traits, such as metabolic flux(Ho and Zhang 2016) and cellular morphology(Ho and
79 Zhang 2014), show signs of adaptive evolution through direct selection for mutational
80 robustness, where traits highly related to cellular growth fitness are likely to acquire lower
81 mutational variability through direct selection for mutational robustness. On the other hand,
82 other studies have identified congruence in trait variability between environmental and genetic
83 perturbations, such as in RNA structure(Szollosi and Derenyi 2009) and gene expression
84 levels(Landry, et al. 2007; Tsuru and Furusawa 2024), supporting the relevance of the byproduct
85 of selection for environmental robustness. As an alternative scenario, mutational variability may
86 spontaneously emerge with trait evolution, regardless of selection for robustness(Siegal and
87 Bergman 2002; de Visser, et al. 2003). To date, the most appropriate scenario for low
88 transcriptional variability of genes related to cellular growth and maintenance remains largely
89 unknown.

90 To address these questions, we focused on the mutational variability governed by cis-
91 acting elements, that is, promoter regions, of genes in *Escherichia coli*. We created a mutant
92 library for each promoter through random mutagenesis and analyzed the changes in promoter
93 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 past selection created a bias in mutational variability in promoters now constrains future 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 genomic mutations (**Fig. 1A**). Previously(Tsuru and Furusawa 2024), we obtained transcriptome profiles of six 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 nonmonotonic 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

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

135 In *E. coli*, genes with many chromosomal copies or genes located at near the origin of
136 replication (*oriC*) tend to have higher expression levels(Ying, et al. 2014). However, essential
137 genes shown in **Fig. 1** are present as only single copies on the chromosome. In addition, these
138 essential genes are widely spread on the chromosome and exhibited higher expression levels
139 than neighboring nonessential genes over the chromosome (**Fig. 1I**). Therefore, gene dosage
140 did not seem to explain the higher expression levels of essential genes. To explore whether the
141 higher expression levels of essential genes are caused by chromosomal position or by higher-
142 order chromosomal structures, we analyzed promoter-mediated gene expression from a plasmid
143 based on the experimental dataset obtained by Silander *et al*(Silander, et al. 2012). Using flow
144 cytometry, Silander *et al* quantified promoter activity of plasmid copies of different promoter
145 regions in *E. coli*, where green fluorescent protein (GFP) was transcriptionally fused to plasmid

146 copies of promoters. Based on fluorescence intensity from GFP, we compared promoter activity
147 between promoters controlling essential genes (termed essential promoters) and those
148 controlling nonessential genes only (termed nonessential promoters) (**supplementary fig. S1A**).
149 We found that essential promoters tended to have higher activities than nonessential promoters
150 in spite of plasmid copies. This result highlighted the importance of promoter regions in
151 accounting for higher expression levels of essential genes independent of chromosomal
152 positions or structures.

153 Like eukaryotes, bacteria often employ postreplicative DNA methylation for the
154 epigenetic control of gene expression. In *E. coli*, DNA adenine methyltransferase plays a major
155 role in the epigenetic control and the target GATC motifs are enriched in promoter
156 regions(Oshima, et al. 2002). Despite that adenine methylation at GATC sites in promoters can
157 inhibit transcription for some genes(Casadesus and Low 2006), we found that differences in
158 promoter activity between essential and nonessential promoters were almost independent of the
159 presence or absence of GATC motifs in the promoter regions (**supplementary fig. S1B**). These
160 results implied that the higher expression levels of essential genes reflected higher promoter
161 activities of essential promoters without known epigenetic controls.

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

163 To explore the molecular mechanisms and signs of natural selection underlying the lower
164 transcriptional variability of essential genes, we focused on the promoter regions of genes with

165 different essentialities. In particular, we explored how promoter activity varies in response to
166 new mutations in the promoter regions and how this variability relates to essentiality of the
167 downstream genes. To this end, we used 56 natural promoters from the *E. coli* Promoter
168 Collection(Zaslaver, et al. 2006). Importantly, to ensure an unbiased comparison of changes in
169 promoter activities between different essentialities and to avoid measurement of promoter
170 activity below the lower detection limit in flow cytometry, we avoided the use of natural
171 promoters with very low activity that were slightly enriched in the promoters of nonessential
172 genes. To achieve this, we selected natural promoters with moderate to high activity before
173 mutagenesis. This selection was based on the known promoter activities of the collection as
174 measured by Silander et al(Silander, et al. 2012). Using the RegulonDB database(Tierrafria, et
175 al. 2022), we retrieved information regarding known transcriptional units regulated by these
176 promoters. Based on the presence of essential genes in the transcription units, we categorized
177 natural promoters into nonessential and essential promoters, termed Nes and Ess, respectively
178 (**Fig. 2A**). Nonessential promoters control transcriptional units comprising only nonessential
179 genes, whereas essential promoters regulate those that include at least a single essential gene.
180 In addition to these natural promoters, we created synthetic promoters, termed Syn, as controls
181 that experienced no or minimal selection for mutational robustness. To this end, we placed
182 random sequences (~200 bp) upstream of the *gfp* gene on a low-copy-number plasmid, pUA66-
183 mCherry (**Fig. 2B**, **supplementary fig. S2**). The library of mixed transformants containing this

184 plasmid was analyzed using flow cytometry. The top 7% of cells showing higher expression
185 levels of GFP (red shaded area in **Fig. 2B**) were sorted, and 16 unique clones were subsequently
186 isolated. Importantly, these synthetic promoters experienced selection for their expression
187 levels only once, and no longer experienced further mutagenesis or selection for expression
188 level. Accordingly, they can be regarded as a control group that is almost free from selection
189 for mutational robustness. Consequently, in the upstream of the GFP gene on the plasmid, we
190 constructed three promoter groups, Nes, Ess, and Syn, comprising different sequences with low
191 homology (**supplementary fig. S3, supplementary note**). The green fluorescence from GFP
192 was used as a measure of promoter activity(Zaslaver, et al. 2006; Silander, et al. 2012). Despite
193 the diversity in sequence, we confirmed that the mean activities of the three promoter groups
194 were comparable to each other, as designed (**Fig. 2C**). We also confirmed that the obtained
195 promoter activities were highly consistent with those measured by Silander et al(Silander, et al.
196 2012) (Spearman's R=0.94, $p<0.05$, **supplementary fig. S4A**). Compared to the Env dataset,
197 the activities of the natural promoters were positively correlated with known mRNA expression
198 levels of the genes downstream of the chromosomal promoters in the wild-type strain
199 (Spearman's R=0.64, **supplementary fig. S4B**), supporting that the wild-type promoter
200 activities monitored by our method considerably reflect the activities of chromosomal
201 promoters and their impact on the expression levels of downstream genes.

202 For each of these promoters, we introduced mutations by error-prone PCR at a low

203 frequency (6.3 mutations/kb on average, **Fig. 2D**) and integrated them into pUA66-mCherry to
204 create a mutant library. Each mutant library had a genetic diversity of 10^5 or more genotypes.
205 The promoter activity of each mutant library was measured by flow cytometry based on the
206 green fluorescence derived from downstream *gfp*, and was compared with the green
207 fluorescence of the corresponding wild type (**Fig. 2E**). The introduction of mutations was
208 limited to the promoter regions of the plasmid, while the chromosomal copies of the natural
209 promoters were intact to avoid any confounding effects. The green fluorescence intensity was
210 $\log_{10}$-transformed after subtracting the background (**supplementary fig. S5**), and the mean and
211 standard deviation of the GFP distribution were calculated (**Fig. 2E, supplementary fig. S6**).

212 **Mutational variability in promoter activity was similar between different promoter**
213 **groups**

214 We confirmed that all mutant libraries showed larger variations than the corresponding wild-
215 type clones, as designed (**Fig. 3A**), supporting that the difference in variations of GFP
216 distributions between wild types and the mutant libraries reflected new genetic variations
217 introduced by error-prone PCR rather than environmental changes or cell-to-cell heterogeneity.
218 In addition, we also confirmed the well-known negative correlation between mean expression
219 levels and standard deviations in GFP distributions regardless of the introduction of mutations
220 or whether the promoters were natural or synthetic (**Fig. 3B,C**), which is consistent with
221 previous observations(Zaslaver, et al. 2006; Wolf, et al. 2015). To compare different promoters,

222 we compensated for the mean-dependency of the standard deviations of both wild types and
223 mutant libraries by subtracting the regression line obtained from wild types (**Fig. 3B,C**). We
224 subsequently calculated MV_{prm}, a measure of mutational variability in promoter activity, by
225 subtracting the compensated standard deviation of wild types from that of mutant libraries.
226 Unexpectedly, MV_{prm}, showed similar values across the three promoter groups (Wilcoxon test,
227 adjusted p-values>0.05), even though each promoter might show unique values (**Fig. 3D**).
228 These results indicated that mutational variability of the focal essential promoters showed little
229 detectable signature of natural selection for mutational robustness despite their importance in
230 cellular growth, even though such selection might be active. Moreover, MV_{prm} showed no
231 significant positive correlations with DM_{env} and DM_{mut} of the downstream genes (**Fig. 3E,F**).
232 The DM values were defined in **Fig. 1**. Using the RegulonDB and EcoCyc(Keseler, et al. 2021)
233 databases to retrieve the known direct regulatory relationships between genes, we counted the
234 number of direct and unique regulations that each promoter receives from transcriptional
235 regulators and explored the impact of the number of regulations on mutational variability. We
236 found that a larger number of activatory transcriptional regulators facilitated mutational
237 variability in promoter activity (Spearman's R=0.31, **Fig. 3G**), probably due to a larger number
238 of mutational target sites. The relationship between MV_{prm} and the number of inhibitory
239 transcriptional regulators was unclear because promoters with null regulation dominated the
240 results (**Fig. 3H**). Despite that the number of activations was likely to make a difference in

241 MV_{prm}, the lack of significant correlations between MV_{prm} and DM values suggests that the
242 mutational variability of promoters provides no considerable explanation for the observed gene-
243 to-gene differences in transcriptional variability of the downstream genes in response to
244 different perturbations at the genome scale (**Fig. 1**). These results suggest that the gene-to-gene
245 differences in transcriptional variability in response to genome-wide perturbations may reflect
246 a certain *trans*-acting molecular mechanism, mainly other than the *cis*-acting one, consistent
247 with previous observations(Tsuru and Furusawa 2024). This implication was also rational
248 because promoter regions are shorter than trans-acting mutational sites for a single gene.

249 **Vulnerability in essential promotors against mutations**

250 Next, we explored the directionality of the mutational changes in promoter activity. Consistent
251 with the low mutation rate in error-prone PCR, the mean expression levels of mutant libraries
252 were similar to those of wild types (Spearman's R=0.99) (**Fig. 4A**). Nevertheless, we found that
253 log_{10}-fold changes in mean expression levels from wild types to mutant libraries followed a
254 global tendency over expression levels, where the promoters with higher activities in wild types
255 tended to decrease their activity in response to mutations (Spearman's R=–056) (**Fig. 4B**). This
256 tendency was retained even when the mean expression levels were calculated for the cells
257 within the narrow gates with the fixed size over expression levels (**supplementary figs. S6,7**),
258 supporting that the negative correlation was not a technical biproduct of our limitation of
259 detecting a decrease in expression levels. This tendency may reflect bias in the mode, activation

260 or inhibition, of transcriptional regulation(Kinney, et al. 2010; Belliveau, et al. 2018; Ireland,
261 et al. 2020). We hypothesized that promoters that receive more activation than inhibition might
262 tend to reduce their activity in response to mutations. To test this possibility, we defined
263 activation bias as the number of known activations subtracted by the number of inhibitions
264 (**supplementary table S1**). We found that the activation bias explained the observed mutational
265 changes (**Fig. 4C**). Furthermore, the activation bias also explained the $\log_{10}$-fold changes after
266 controlling the global trend using the regression line (**Fig. 4D**), indicating that a promoter
267 becomes more vulnerable as it receives more activation among promoters with similar activities.

268 Interestingly, synthetic and nonessential promoters showed similar mutational changes
269 around zero, whereas essential promoters showed biased directionality in mutational changes
270 toward negative values (**Fig. 4E**), even though the mean activities were almost the same among
271 the three promoter groups (**Fig. 2C**). There were no significant differences in promoter length
272 and GC content between the essential and nonessential promoters (Wilcoxon test, $p>0.05$). In
273 addition, there was no significant difference in the mutation rate in error-prone PCR between
274 the essential and nonessential promoters (**supplementary fig. S8A, supplementary table S2**).
275 The mutational spectrum in error-prone PCR was quite similar between the two promoter
276 groups (**supplementary fig. S8B**). These results supported the notion that the observed bias
277 was not derived from a technical bias. This bias in essential promoters was retained for the
278 normalized $\log_{10}$-fold changes (**Fig. 4F**). Accordingly, essential promoters were more

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

284 **Vulnerability in essential promotors against mutations may reflect past selection for**
285 **higher activity**

286 The difference in mutational vulnerability between essential and nonessential promoters (**Fig.**
287 **4E**, **F**) could reflect the difference in past selection pressures acting on higher expression levels
288 between essential and nonessential genes. Based on the observation that essential genes tended
289 to have higher expression levels than nonessential genes (**Fig. 1**), we hypothesized that the
290 remarkable vulnerability of essential promoters reflects local optima in terms of promoter
291 activity. That is, essential promoters might have evolved toward higher activity through past
292 evolution and might have been stuck in local optima because of the high essentiality of
293 downstream genes. In contrast, nonessential promoters might have experienced relatively
294 relaxed selection for higher expression levels because of the low essentiality of downstream
295 genes. To explore the relevance of our assumption, we compared the impact of knockdown on
296 growth fitness of *E. coli* between essential and nonessential genes using a public experimental
297 dataset reported by Hawkins *et al*(Hawkins, et al. 2020). We found that knockdown of essential

genes tended to be more deleterious for growth fitness than nonessential genes (**supplementary fig. S9A**), supporting our assumption that essential genes are under stronger selection pressure toward higher expression levels than nonessential genes. To investigate the emergence of differences in vulnerability depending on selection pressure, we performed a numerical simulation of promoter evolution on a simple rugged fitness landscape (**Fig. 5**). For simplicity, we modeled promoter genotype as 10-bit binary numbers such as 1101011011. That is, our model promoters are comprising of biallelic sites (0 or 1), which is simplified from real promoters comprising of quadallelic sites (A, T, G, and C). We assumed that the expression landscape is rugged and genotypes with higher expression levels are relatively rare, consistent with the previous experimental observations(Wolf, et al. 2015; Yona, et al. 2018). To satisfy this assumption, the mapping form genotype to expression level, equivalent to promoter activity, was defined by the Rough Mt. Fuji (RMF) model, as illustrated in **Fig. 5A, D**. In short, the genotype with the highest expression level, g^*, was first chosen at random. The expression levels of the other genotypes were decided as a decreasing linear function of the Hamming distance from g^*. Subsequently, stochastic noise was added to generate ruggedness. Finally, expression levels were set to range from 0 (lower bound) to 1 (upper bound) as detailed in the **Materials and Methods** section. Using this expression landscape, mutational changes were calculated for each genotype as the mean difference in the expression levels of all nearest neighbors with a single bit difference, or one Hamming distance, from the expression level of

317 the focal genotype (**Fig. 5B, E**). This simple expression landscape captured the negative
318 correlation between expression levels and mutational changes (**Fig. 5E**), which is consistent
319 with the global trend observed in our experiment (**Fig. 4B**). This trend was likely due to the
320 prevalence of local sinks (green dots) at lower expression levels and frequent local peaks (red
321 dots) at higher expression levels (**Fig. 5B, E**). To clarify this point, we calculated the fraction
322 of the nearest neighboring genotypes with higher expression levels for each genotype, termed
323 uphill ratio. Local sinks tend to have higher uphill ratios, while local peaks tend to have lower
324 uphill ratios. We confirmed that genotypes with higher uphill ratios dominated at lower
325 expression levels, while genotypes with lower uphill ratios dominated at higher expression
326 levels (**Fig. 5F**). Importantly, even at the same expression levels, local peaks tended to have
327 negative mutational changes, whereas local sinks tended to show positive changes. The fitness
328 landscape was then modeled by adding the basal fitness, F_b, to the expression levels multiplied
329 by $1 - F_b$, where F_b ranged from 0 to 1 (**Fig. 5C, G**). That is, expression level is under positive
330 directional selection in our model, where strength of selection pressure depends on F_b. A
331 smaller F_b corresponds to a steeper fitness landscape, while a larger F_b results in a flatter fitness
332 landscape. This simple model assumes a positive linear relationship between expression level
333 and fitness (**Fig. 5H**), which was relevant to the experimental evidence in most cases
334 (**supplementary fig. S9B**). Based on the fact that knockdown of essential genes tends to be
335 more deleterious for cell growth than nonessential genes (**supplementary fig. S9A**), we

336 assigned lower and higher values of F_b for essential and nonessential promoters, respectively.
337 Thus, in our model, essential and nonessential promoters shared the same expression landscape
338 but experienced different strengths of selection pressures. Using the constructed fitness
339 landscape with different F_b values (0.3, 0.5, and 0.99), we performed an evolutionary
340 simulation based on the standard Wright-Fisher model(Otto and Day 2007) with asexual
341 haploid population without recombination and with reversible mutation. Initial genotypes were
342 randomly generated for each simulation. To mimic ongoing evolution of real organisms, the
343 evolutionary simulation was performed for relatively short generations. In practice, the most
344 abundant mutants within the evolved populations after 30 generations were isolated and
345 subjected to measurement of mutational changes (**Fig. 5 G,H**). Despite the short generations,
346 we confirmed that the evolved isolates in steeper fitness landscapes exhibited higher expression
347 levels (**Fig. 5I**), which is consistent with the designed selection pressure. Comparing the
348 mutational changes between different F_b values, we found that the mutants that evolved with
349 smaller F_b showed negatively larger mutational changes, even at the same expression levels
350 (**Fig. 5H**). These results suggest that higher activity and vulnerability in essential promoters are
351 linked, and both can emerge simultaneously through stronger positive selection acting on higher
352 expression levels of downstream essential genes.

353 Despite that our simple linear model basically agreed with the known relationship
354 between expression level and growth fitness in most cases (**supplementary fig. S9**), an

355 alternative model might be suitable for some genes such as *yidC* (**supplementary fig. S9B**).
356 These genes exhibited almost no changes in growth fitness in response to small knockdown
357 level, while growth fitness gradually decreases with increases in knockdown level beyond
358 certain knockdown levels. To deal with such biphasic cases, we constructed the linear-plateau
359 model (**supplementary fig. S9**). In short, the linear-plateau model has two phases: a positive
360 linear phase and a flat plateau phase, where the joint point of the two phases is characterized by
361 E_p ranging from 0 to 1 (**supplementary fig. S10D**). A larger E_p corresponds to a higher
362 optimal expression level. Using the linear-plateau model with different E_p values, we
363 performed an evolutionary simulation by following the same procedure as the linear model
364 (**supplementary fig. S10E, F**). We found that the evolved isolates under the larger E_p exhibited
365 higher expression levels (**supplementary fig. 10G**) and negatively larger mutational changes,
366 even at the same expression levels (**supplementary fig. 10F**). These results seemed to be
367 relevant by considering neutral evolution prevalent in the fitness landscapes with the smaller
368 E_p. These results highlighted an impact of optimal expression level on activity and vulnerability
369 in promoters, suggesting that higher optimal expression levels could also contribute to higher
370 activity and mutational vulnerability in essential promoters. Overall, our findings imply that
371 both higher optimal expression and stronger directional selection could contribute to mutational
372 vulnerability in expression levels for essential genes. Although the expression of essential
373 promoters tended to decrease due to mutations (**Fig. 4E**), a few essential promoters with lower

374 mean expression levels exhibited neutral to positive $\log_{10}$-fold changes in expression response
375 to mutations. A prominent example was the essential promoter controlling *tilS*, which exhibited
376 the lowest expression level among wild-type essential groups (**supplementary Table S2**). This
377 promoter showed a positive value in $\log_{10}$-fold change in response to mutations, indicating the
378 frequent occurrence of mutants with higher expression levels than wild-types. These essential
379 promoters might suggest weaker directional selection or lower optimal expression levels for the
380 downstream essential genes.

381 Consistent results were also obtained using the NK model(Kauffman and Weinberger
382 1989) for rugged expression landscapes (**supplementary figs. S11 and S12**, for the linear and
383 linear-plateau models, respectively). The NK model, similar to the RMF model, frequently
384 provided sinks (genotypes with higher uphill ratios) at lower expression levels and peaks
385 (genotypes with lower uphill ratios) at higher expression levels (**supplementary fig. S11C**).
386 However, compared to the RMF model (**Fig. 5F**), sinks were relatively rare at middle to high
387 expression levels (beyond 0.3 in expression level). Interestingly, peaks were also rare at middle
388 expression levels (0.3–0.6 in expression level) in the NK model. That is, genotypes with
389 moderate uphill ratios were relatively frequent at middle to high expression levels. The low
390 frequency of sinks and peaks indicates fewer genotypes with higher or lower mutational
391 changes. Consequently, the difference in mutational changes between different genotypes at the
392 same expression levels was relatively small in the NK model. Therefore, the difference in

393 mutational changes between different F_b (or E_p) values in the NK model was small relative to
394 the RMF model (**supplementary figs. S11G**,**12F**). Despite these minor differences, both
395 expression landscapes showed consistent outcomes for the linear (**Fig. 5J**,**K** and
396 **supplementary fig. S11G**,**H** for the RMF and NK model, respectively) and linear-plateau
397 models (**supplementary fig. S10F**,**G** and **supplementary fig. S12F**,**G** for the RMF and NK
398 model, respectively). These results support the broad applicability of our key assumptions
399 underlying rugged expression landscapes, suggesting the robustness of the outcomes derived
400 from our simple models.

Discussion

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

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

420 1942), congruent(Meiklejohn and Hartl 2002), and intrinsic(Siegal and Bergman 2002)
421 scenarios. The adaptive scenarios predict that essential traits have evolved to reduce mutational
422 variability and minimize the emergence of harmful mutants through direct selection for
423 mutational robustness. The congruent scenarios predict that mutational robustness is a
424 byproduct of evolution toward low variability against environmental perturbations. Contrary to
425 these two scenarios assuming adaptive evolution toward low variability, the intrinsic scenarios
426 predict that the mutational robustness of the traits emerges spontaneously and inherently along
427 with the emergence of traits without selection against variability. In the case of promoter activity,
428 an intrinsic scenario corresponds to the mutational robustness of promoters emerging
429 spontaneously with the emergence of promoter activity without selection for mutational
430 robustness, which corresponds to the mutational variability of synthetic promoters. Our results
431 demonstrated that both adaptive and congruent scenarios failed to explain the observed
432 mutational variability in natural promoters because essential promoters showed similar
433 mutational variability to the other promoters, and the natural promoters exhibited no significant
434 positive correlation with the transcriptional variability of downstream genes against different
435 environmental perturbations. We note that the adaptive scenario might explain the higher
436 expression levels, not transcriptional variability, of essential genes (**Fig. 1C, G**), which might
437 minimize the risk of mutational knock down for essential genes to a lethal level. In contrast, the
438 intrinsic scenario explained the observed mutational variability because the natural promoters

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

442 Why did essential promoters show no less mutational variability than other promoters,
443 despite the fact that the reduction in expression levels of essential genes is more likely to be
444 harmful to cell growth than nonessential genes(Cui, et al. 2018)? One possible reason is that
445 promoter sequences with significant mutational robustness are rare, although several possible
446 mechanisms underlying mutational robustness in promoter regions are hypothesized(Payne and
447 Wagner 2015). In other words, the expression landscape of the promoters might be rugged
448 everywhere as we modeled in **Fig. 5A**. In fact, synthetic promoters derived from different
449 random sequences achieved higher promoter activities equivalent to natural promoters, which
450 was consistent with previous observations(Wolf, et al. 2015; Yona, et al. 2018). A second
451 possible reason is that selection for higher expression levels on a rugged fitness landscape (**Fig.**
452 **5C**) prevents the essential promoters at local optima from traversing the expression landscape
453 to reach genotypes with higher mutational robustness. Consistently, essential promoters showed
454 biased mutational effects, where most mutations in essential promoters tended to reduce
455 promoter activity (**Fig. 4C**). Interestingly, this scenario agreed with the predictions from a
456 machine-learning model based on an empirical expression landscape of yeast
457 promoters(Vaishnav, et al. 2022). Thus, our results suggest several plausible hypotheses that

458 may be shared by different organisms. Ideally, future studies should explore evolutionary
459 scenarios for diverse promoters that were not tested in this study, and validate the hypothetical
460 landscape of promoters more thoroughly. The increase of the sample set of promoters would
461 also compensate for our small sample sets and would enhance power to detect a potential subtle
462 difference in mutational variability between different promoter groups.

463 In this study, we focused only on mutations in promoter regions to explore the
464 mechanism underlying the lower transcriptional variability in essential genes. As a result, we
465 found that mutants with higher promoter activity were relatively rare for essential promoters.
466 However, this result does not exclude the contribution of any other mechanism underlying the
467 observed transcriptional variability in response to genome-wide perturbations (**Fig. 1)**. Notably,
468 *trans*-acting mechanisms, such as transcriptional regulators, may also play a role. As it is, the
469 previous studies reported that genes exhibiting greater transcriptional variability across diverse
470 environmental and genetic perturbations tend to be regulated by a larger number of unique
471 transcriptional regulators in *E. coli*(Tsuru and Furusawa 2024) and yeast(Landry, et al. 2007).
472 A higher number of regulators provides more *trans*-mutational target sites for the regulated
473 genes, thereby enhancing transcriptional variability in response to genomic perturbations.
474 Based on this evidence, it is plausible that essential genes have less transcriptional regulation.
475 To support this hypothesis, we used data from RegulonDB and EcoCyc to compare the number
476 of regulations between essential and nonessential genes (**supplementary fig. S13**). First, we

477 found a smaller fraction of essential genes (90 of 248 genes, 36%) lacking known
478 transcriptional regulators than nonessential genes (1976 of 3995, 49%) (**supplementary fig.**
479 **S13A**). This finding aligns with the positive activation bias values observed (**Fig. 4F**), probably
480 attributable to the rich activation by sigma factors aimed at enhancing expression levels. On the
481 other hand, among the genes regulated by at least one transcriptional regulator, we observed
482 that essential genes received fewer transcriptional regulations than nonessential genes
483 (**supplementary fig. S13B**), supporting this hypothesis. This observation is consistent with
484 previous findings in yeast(Bilu and Barkai 2005). Thus, our study underscores the dual
485 contributions of both *cis*- and *trans*-acting mechanisms including their epistatic interactions in
486 constraining the transcriptional variability of essential genes.

487 ## Materials and Methods

488 **Bacterial strains and medium**

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

504 **Selection of natural promoters**

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

506 was used to obtain the promoter regions of the natural promoters in *E. coli*. Initially, we filtered
507 out the promoters with low activities that were near our detection limit based on the promoter
508 activities of this library measured by Silander et al(Silander, et al. 2012). This filtration
509 excluded 12% and 17% of essential and nonessential promoters, respectively. Subsequently,
510 115 promoters were randomly selected from this subset library carrying the pUA66 plasmid.
511 The DNA sequences of these promoter regions were verified using Sanger sequencing (Azenta
512 Life Sciences, USA) and compared with the reference genome sequence of *E. coli* K-12
513 MG1655(Blattner, et al. 1997) (GenBank ID: U00096). Subsequently, 63 of the 115 strains
514 without mutations in their promoter sequences were identified, and plasmids were extracted
515 from these colonies. Among these, 56 strains were randomly selected for the subsequent
516 experiments (**supplementary table S2**). Each promoter was denoted by the "NP" (natural
517 promoter), followed by the name of the ORF present in the immediate downstream sequence in
518 the genome. For example, NPmetG represents the promoter regions upstream of the *metG* gene
519 on the chromosome.

520 **Plasmid construction**

521 Plasmid pUA66 was modified by replacing the promoter regions of *gfp*, flanked by the BamHI
522 and XhoI restriction sites, with a red fluorescence protein (mCherry) expression cassette. This
523 cassette comprised dual terminators RNAI and TSAL(Reynolds, et al. 1992), promoter PLtetO-
524 1(Lutz 1997), Shine-Dalgarno sequence SD8(Ringquist, et al. 1992) and codon-optimized

525 mCherry(Dunlop, et al. 2008) coding sequences. The terminators were obtained from pNS2-
526 σVL(Dunlop, et al. 2008)(Addgene, Plasmid #26756) and the assembled sequences of the other
527 elements were commercially synthesized (Fasmac Co., Ltd., Japan). The assembled cassette
528 was inserted between the BamHI and XhoI recognition sequences of the pUA66 plasmid in the
529 direction opposite to that of *gfp* (**supplementary fig. S2**), yielding plasmid pUA66-mCherry.
530 PCR amplification was performed using KOD One PCR Master Mix Blue (Toyobo, Japan,
531 KMM-201) and pUA66-mCherry as a template with primers A
532 (GCTTTCCAGGGATCCTCTAGATTTAAGAAGG) and B
533 (GAGCAGTACTCTCGAGGTGAAGACGAAAGG) to include all sequences except for the
534 mCherry expression cassette. The PCR product was column purified using the FastGene
535 Gel/PCR Extraction Kit (NIPPON Genetics, Japan, FG-91302). The purified PCR product was
536 double-digested with BamHI (Takara Bio, Japan, 1010A) and XhoI (Takara Bio, Japan, 1094A)
537 overnight at 37 °C, followed by column purification. The digested product is referred to as the
538 vector.

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

540 The construction of the synthetic promoters was based on a previous study(Wolf, et al. 2015).
541 Single-stranded oligonucleotides (CCTTTCGTCTTCACCTCGAG-(N200)-
542 GGATCCTCTAGATTTAAGAAGG) were synthesized by Eurofins Genomics (Ebersberg,
543 Germany). The 5' and 3' ends of the oligos contained XhoI and BamHI restriction sites,

544 respectively, and were homologous to primers C (CCTTTCGTCTTCACCTCGA) and D
545 (CCTTCTTAAATCTAGAGGATCC), respectively (**supplementary fig. S2**). A 200 bases
546 region of random nucleotides was present between them. PCR amplification using primers C
547 and D was performed using a mixture of single-stranded oligos as a template to create double-
548 stranded DNA. The PCR products were column purified and double digested using BamHI and
549 XhoI for 3 h at 37 °C, followed by column purification. Thus, an insert for the following ligation
550 reaction was obtained. The ligation reaction was performed using T4 DNA Ligase (Takara Bio,
551 Japan, 2011A) overnight at 16 °C involved mixing the vector and insert at a molar ratio of 1: 3.
552 The resulting ligation product solution was column purified and DNA was extracted using 20
553 μL of sterile water. All purified ligation product solutions were added to 50 μL of DH5α
554 Electro-Cells and electroporated at 1800 V (Eporator, Eppendroph, Germany, 4309000027).
555 Transformed cells were incubated at 37 °C for 1 h in 1 mL of SOC medium, followed by
556 overnight incubation at 37 °C in 1 mL of LB medium supplemented with 2 μL of 25 mg/mL
557 kanamycin solution.

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

559 Green and red fluorescence, as well as forward scatter of the cells in the DH5α-transformed
560 random promoter library were measured using a flow cytometer (FACSAria III, BD, USA).
561 Gates were set for the green and red channels to exclude cells with low GFP and high mCherry
562 fluorescence, respectively. Subsequently, 10,000 cells without red fluorescence but with green

563 fluorescence within the selected range (**Fig. 2B**) were collected by cell sorting, according to the
564 manufacturer's instructions. The sorted cells were spread on LB/Km agar plates and incubated
565 overnight at 37 °C. Bacterial colonies were randomly picked and their inserted synthetic
566 promoter regions were sequenced by Sanger sequencing. Consequently, 16 synthetic promoters
567 without improper mutations at the restriction sites were identified and named SP1–16,
568 (**supplementary table S1**).

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

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

577 **Construction of mutant promoter library**

578 The purified ligation product solution was added to 50 μL of DH5α Electro-Cells and
579 electroporated at 1800 V. Transformed cells were cultured at 37 °C for 1 h in 1 mL of SOC
580 medium and a portion of the culture was plated on LB/Km agar medium for overnight growth
581 at 37 °C. Bacterial colonies were counted to determine the transformation efficiency. If the

582 transformation efficiency exceeded 10^5 cells/reaction, the next step was performed; otherwise,
583 the process was repeated. The transformed culture was diluted with 5 mL of LB/Km medium
584 and grown overnight. Plasmid extraction from an overnight culture was performed using the
585 FastGene Plasmid Mini Kit (NIPPON Genetics, Japan, FG-90502). The obtained plasmid
586 solution was added to 50 μL of MG1655 electrocompetent cells and electroporated at 1800 V.
587 Transformed cells were cultured at 37 °C for 1 h in 1 mL of SOC medium, and a portion of the
588 solution was plated on LB/Km agar medium for overnight growth at 37 °C to count the number
589 of colonies and determine transformation efficiency. If the transformation efficiency exceeded
590 10^5 cells/reaction, the next step was performed; otherwise, the process was repeated. A sterile
591 2 mL 96 well deep well plate (Greiner Bio-One, Austria, 780271) was used to add 1 mL of the
592 synthetic medium to each well, followed by the addition of 10 μL of the overnight culture
593 containing the MG1655 transformants. The bacterial cells were grown at 37 °C with shaking at
594 800 rpm until confluent growth was achieved. Subsequently, another plate was prepared with 1
595 mL of fresh synthetic medium added to each well, and 10 μL of the confluently grown cultures
596 was diluted into each well. The cultures were grown for 2.5 h at 37 °C with shaking at 800 rpm
597 and were used as mutant libraries for flow cytometry.

598 **Verification of PCR random mutagenesis**

599 The number of mutations introduced into the promoters by error-prone PCR was determined
600 using Sanger sequencing. Three to four colonies were randomly selected from six randomly

601 selected mutant libraries, including essential and nonessential promoters (**supplementary table**
602 **S2**). Colony PCR was performed using KOD One PCR Master Mix Blue with primers E
603 (TTACTTTGCAGGGCTTCCCAA) and F (CCAGTCTTTCGACTGAGCCT) to amplify the
604 regions including the promoter regions (**supplementary fig. S2**). PCR products were column
605 purified and verified by Sanger sequencing. The average mutation rate was calculated to be 6.2
606 mutations/kb. Mutation rates and mutational spectrum for each promoter group were shown in
607 **supplementary fig. S8**.

608 **Construction of wild-type promoter control group**

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

614 **Flow cytometry**

615 Bacterial cells were dispensed in M9 buffer prior to analysis using a flow cytometer (BD, USA,
616 FACSAria III). Fluorescence intensity from GFP (GFP FI) and mCherry (mCherry FI) was
617 measured using a 488 nm laser with 515–545-nm emission filter, and a 561 nm laser with 563–
618 588-nm emission filter, respectively. The following PMT voltage settings were used: forward
619 scatter (FSC), 300; side scatter (SSC), 340; green fluorescence, 700; red fluorescence, 600. The

events exceeding the defined thresholds of FSC (200) and SSC (200) were recorded. Tota1,000,000 events were measured per library. The MG1655 strain was used as a control for autofluorescence of GFP FI.

Data analysis of flow cytometry

FCS files were converted to CSV format using the flowCore(Ellis, et al. 2024) package in R(R Core Team 2023). Subsequent analyses were conducted using custom R scripts. To compare green fluorescence among cells in similar physiological states, a narrow gate for the forward scatter channel was set to capture 40% of the events, including the most frequent values for MG1655 (**supplementary fig. S5A**). The top 0.5% value of the green fluorescence intensity distribution of MG1655 within the gate was determined as the autofluorescence intensity (**supplementary fig. S5B**). For the events of wild types and mutant libraries, the autofluorescence intensity was subtracted from the green fluorescence intensity of the events within the gate. Events with a negative green fluorescence intensity after subtracting the autofluorescence intensity were excluded from the analysis. In addition, events with a red fluorescence intensity greater than 10^3 [a.u.] were regarded as the cells carrying pUA66-mCherry and were excluded from the analysis. Clonal cell populations exhibit huge cell-to-cell heterogeneity in gene expression even under identical environmental conditions, resulting in long-tailed or log-normal shaped distributions in protein numbers(Tsuru, et al. 2009; Taniguchi, et al. 2010). This stochastic nature often results in rare cells exhibiting very highly expression

639 levels within cell populations, which greatly influences variance or standard deviation of
640 protein number distributions. To ensure robust measurement of variation in fluorescent
641 distributions, the green fluorescence intensity after subtracting the autofluorescence intensity
642 was $\log_{10}$-transformed and used to calculate the means and standard deviations. We note that
643 this log-transformation might limit detection of possible small differences in mutational
644 variability in promoter activities between different promoter groups.

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

646 Gene essentiality was based on a previous study(Goodall, et al. 2018), which identified 248
647 genes designated as essential across different datasets. In bacteria, clusters of genes are
648 controlled by the same promoter sequences and are transcribed as single mRNAs, termed
649 transcription units. Promoters controlling at least one essential gene in the transcription units
650 were defined as essential promoters (**Fig. 2A**). The known transcription units were retrieved
651 from the RegulonDB(Tierrafria, et al. 2022) database. Of the 56 promoters, 18 were classified
652 as essential and 38 as nonessential as listed in **supplementary table S1**.

653 **Estimation of transcriptional variability**

654 Transcriptional variability of each gene in *E. coli* in response to different genetic perturbations
655 was estimated using known transcriptome profiles obtained from six independent mutants
656 derive from MG1655 cultured under a single environmental condition, termed the Mut
657 dataset(Tsuru and Furusawa 2024). These mutants were subjected to mutation accumulation

658 (MA) experiments to accumulate genomic mutations via genetic drift(Halligan and Keightley
659 2009; Tsuru, et al. 2015). In addition, to quantify the transcriptional variability in response to
660 different environmental conditions, known transcriptome profiles of MG1655 cultured under
661 76 different environmental conditions(Sastry, et al. 2019; Rychel, et al. 2021; Lamoureux, et al.
662 2022), termed the Env dataset(Tsuru and Furusawa 2024), were utilized. The expression levels
663 were determined by $\log_2$-transformation and quantile normalization. The means and standard
664 deviations in expression levels for each gene across different genetic perturbations and
665 environmental conditions were quantified. To compensate for the mean-dependency of the
666 standard deviations (**Fig. 1B**,**G**), the DM values for each gene were calculated as the vertical
667 distance of each standard deviation from a smoothed running median of the standard deviations
668 for each dataset, as detailed previously(Tsuru and Furusawa 2024). The resulting DM values
669 were termed DM_{mut} and DM_{env} for the Mut and Env datasets, respectively. The DM values were
670 used as measures of transcriptional variability in response to genetic and environmental
671 perturbations.

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

673 Synthetic promoters underwent a blastn(Altschul, et al. 1990) search to confirm that there was
674 no considerable similarity to any natural sequences (**supplementary note**). Posterior sequences
675 of 73 bases were aligned using the msa(Bodenhofer, et al. 2015) package in R (Clustal Omega),
676 where 73 bases represented the shortest length of a synthetic promoter. A phylogenetic tree was

677 constructed using the Neighbor-Joining method.

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

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

684 **Identification of the number of transcriptional regulators regulating the natural**
685 **promoters**

686 Reported regulatory interactions between genes were compiled using all interactions with
687 experimental evidence from RegulonDB and EcoCyc(Keseler, et al. 2021) for both transcription
688 factors and sigma factors. The mode of regulation was classified as either activation or
689 inhibition, and the mode of regulation by sigma factors was classified as activation.

690 **Numerical simulation of promoter evolution**

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

694 $$E(g) = -cD(g, g^*) + \eta(g),$$

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. To construct an expression landscape, g^* was first chosen at random from the genotype space. $E(g)$ was subsequently calculated using the above equation. 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 positive selection, and nearly neutral evolution, respectively. It was assumed that essential genes underwent strong positive selection on their expression levels, while nonessential genes underwent relatively moderate selection for their expression levels, based on the empirical fact that knock-down of essential genes was more harmful for cell growth than nonessential genes(Li, et al. 2016). Ten independent fitness landscapes were constructed for each F_b.

714 Evolutionary simulation on the fitness landscapes were performed using the standard Wright-
715 Fisher model(Otto and Day 2007) assuming a haploid asexual population with a population size
716 of 10^5 and a mutation rate of 10^{-3} mutations per genotype per generation. The initial genotypes
717 were randomly selected for each simulation, and 600 independent simulations were examined
718 for each fitness landscape. For each generation, mutants with genotypes differing by one
719 Hamming distance from the current population were generated at the mutation rate. The
720 frequency of each genotype within the population was updated based on the fitness landscape.
721 Subsequently, a new population for the next generation was formed by sampling mutants, with
722 replacement, from the current population using a binomial distribution with the given
723 population size. The most frequent mutants in a population after 30 generations were isolated
724 in each simulation, resulting in 6,000 mutants for each F_b. Mutational changes in expression
725 levels were calculated for each isolate by averaging the expression levels of all nearest
726 neighboring genotypes that differed by one Hamming distance from the focal isolate. Custom
727 R scripts were used for the numerical simulations, modified from scripts constructed in a
728 previous study(Obolski, et al. 2018; Song and Zhang 2021).

729 **Data visualization**

730 All figures were generated in R and the following packages were used. Phylogenetic trees of
731 the promoters were visualized using the ape(Paradis and Schliep 2019) and ggtree(Yu, et al.
732 2017) packages. Other illustrative plots were generated using the ggplot2(Wickham 2016) and

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 (wild types for https://doi.org/10.5281/zenodo.11108555; mutant libraries 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.).

752 **References**

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

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

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

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

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

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

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

771 Casadesus J, Low D. 2006. Epigenetic gene regulation in the bacterial world. Microbiol
772 Mol Biol Rev 70:830-856.
773 Cui L, Vigouroux A, Rousset F, Varet H, Khanna V, Bikard D. 2018. A CRISPRi screen in
774 *E. coli* reveals sequence-specific toxicity of dCas9. Nature Communications 9:1-10.
775 Darwin C. 1859. On the origin of species by means of natural selection, or preservation of
776 favoured races in the struggle for life: London : John Murray.
777 de Visser JA, Hermisson J, Wagner GP, Ancel Meyers L, Bagheri-Chaichian H, Blanchard
778 JL, Chao L, Cheverud JM, Elena SF, Fontana W, et al. 2003. Perspective: Evolution and
779 detection of genetic robustness. Evolution 57:1959-1972.
780 Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK. 2005. The
781 transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. Nat
782 Genet 37:544-548.
783 Draghi JA, Whitlock MC. 2012. Phenotypic plasticity facilitates mutational variance,
784 genetic variance, and evolvability along the major axis of environmental variation. Evolution
785 66:2891-2902.
786 Dunlop MJ, Cox RS, Levine JH, Murray RM, Elowitz MB. 2008. Regulatory activity
787 revealed by dynamic correlations in gene expression noise. Nature Genetics 40:1493-1498.

788 Duveau F, Vande Zande P, Metzger BP, Diaz CJ, Walker EA, Tryban S, Siddiq MA, Yang
789 B, Wittkopp PJ. 2021. Mutational sources of trans-regulatory variation affecting gene
790 expression in Saccharomyces cerevisiae. Elife 10.
791 Ellis B, Haaland P, Hahne F, Meur L, Nolwenn, Gopalakrishnan N, Spidlen J, Jiang M,
792 Finak G, Granjeaud S. 2024. flowCore: flowCore: Basic structures for flow cytometry data. R
793 package version 2.14.2:https://bioconductor.org/packages/flowCore/.
794 Furusawa C, Kaneko K. 2018. Formation of dominant mode by evolution in biological
795 systems. Phys Rev E 97:042410.
796 Goodall ECA, Robinson A, Johnston IG, Jabbari S, Turner KA, Cunningham AF, Lund PA,
797 Cole JA, Henderson IR. 2018. The essential genome of *Escherichia coli* K-12. mBio 9:1-18.
798 Halligan DL, Keightley PD. 2009. Spontaneous mutation accumulation studies in
799 evolutionary genetics. Annual Review of Ecology, Evolution, and Systematics 40:151-172.
800 Hawkins JS, Silvis MR, Koo BM, Peters JM, Osadnik H, Jost M, Hearne CC, Weissman
801 JS, Todor H, Gross CA. 2020. Mismatch-CRISPRi Reveals the Co-varying Expression-
802 Fitness Relationships of Essential Genes in *Escherichia coli* and *Bacillus subtilis*. Cell Syst
803 11:523-535 e529.
804 Ho WC, Zhang J. 2016. Adaptive Genetic Robustness of *Escherichia coli* Metabolic
805 Fluxes. Molecular Biology and Evolution 33:1164-1176.

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

Hornung G, Oren M, Barkai N. 2012. Nucleosome organization affects the sensitivity of gene expression to promoter mutations. Mol Cell 46:362-368.

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

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

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

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

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

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

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

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

Levenshtein VI editor. Soviet physics doklady. 1966.

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

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

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

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

Metzger BP, Yuan DC, Gruber JD, Duveau F, Wittkopp PJ. 2015. Selection on noise constrains variation in a eukaryotic promoter. Nature 521:344-347.

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.

Oshima T, Wada C, Kawagoe Y, Ara T, Maeda M, Masuda Y, Hiraga S, Mori H. 2002. Genome-wide analysis of deoxyadenosine methyltransferase-mediated control of gene expression in Escherichia coli. Mol Microbiol 45:673-695.

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.

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

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

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

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

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

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

878 Smith JM, Burian R, Kauffman S, Alberch P, Campbell J, Goodwin B, Lande R, Raup D,
879 Wolpert L. 1985. Developmental constraints and evolution: a perspective from the Mountain
880 Lake conference on development and evolution. The Quarterly Review of Biology 60:265-
881 287.
882 Song S, Zhang J. 2021. Unbiased inference of the fitness landscape ruggedness from
883 imprecise fitness estimates. Evolution 75:2658-2671.
884 Szollosi GJ, Derenyi I. 2009. Congruent evolution of genetic and environmental robustness
885 in micro-RNA. Mol Biol Evol 26:867-874.
886 Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS. 2010.
887 Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single
888 cells. Science 329:533-538.
889 Tierrafria VH, Rioualen C, Salgado H, Lara P, Gama-Castro S, Lally P, Gomez-Romero L,
890 Pena-Loredo P, Lopez-Almazo AG, Alarcon-Carranza G, et al. 2022. RegulonDB 11.0:
891 Comprehensive high-throughput datasets on transcriptional regulation in *Escherichia coli* K-
892 12. Microb Genom 8.
893 Tirosh I, Weinberger A, Carmi M, Barkai N. 2006. A genetic signature of interspecies
894 variations in gene expression. Nat Genet 38:830-834.
895 Tsuru S, Furusawa C. 2024. Genetic properties underlying transcriptional variability in
896 *Escherichia coli*. bioRxiv:2024.2004.2015.589659.

Tsuru S, Ichinose J, Kashiwagi A, Ying BW, Kaneko K, Yomo T. 2009. Noisy cell growth rate leads to fluctuating protein concentration in bacteria. Phys Biol 6:036015.

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

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

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

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

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

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

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

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

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

Ying BW, Tsuru S, Seno S, Matsuda H, Yomo T. 2014. Gene expression scaled by distance to the genome replication site. Mol Biosyst 10:375-379.

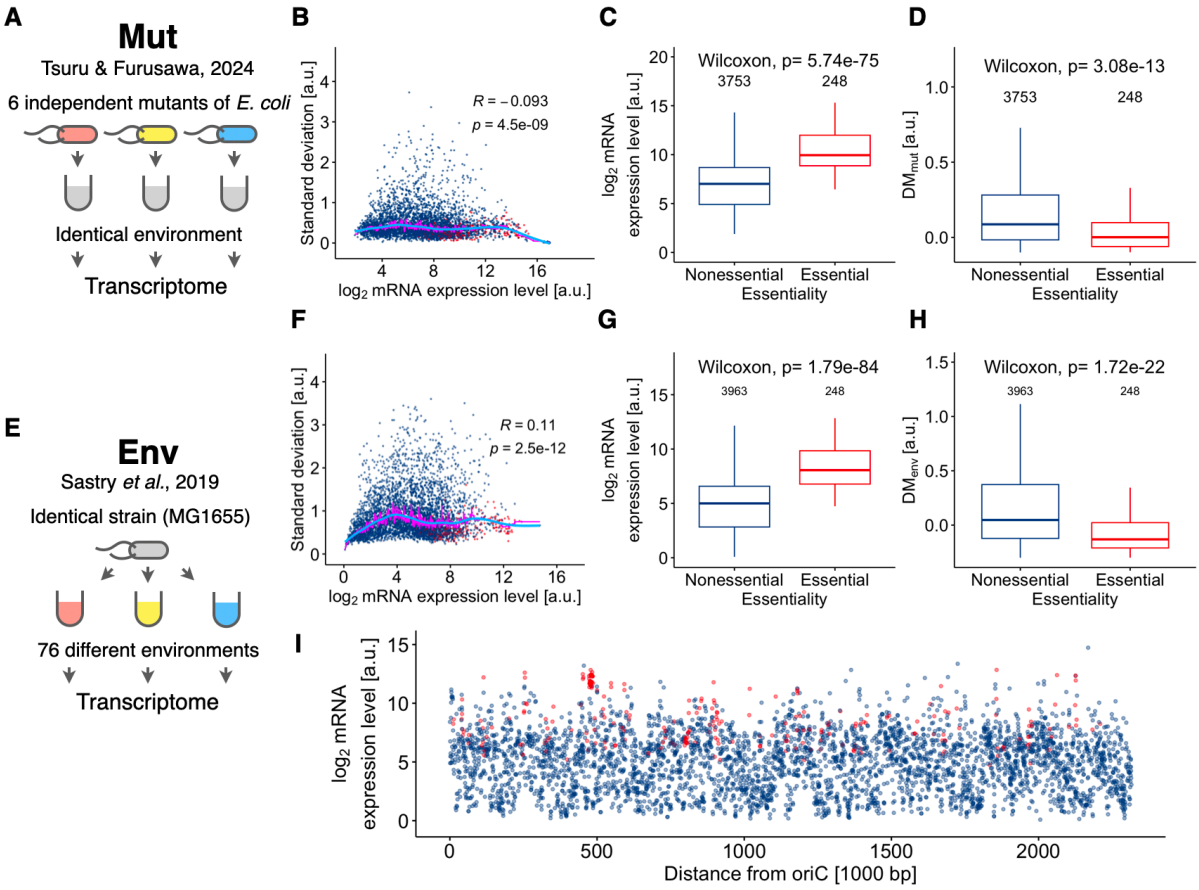
Yona AH, Alm EJ, Gore J. 2018. Random sequences rapidly evolve into de novo promoters. Nat Commun 9:1530.

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

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

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

931 **Figures and Figure legends**



932

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

934 (**A**–**D**) Transcriptional profiles of six independent mutants of *E. coli* cultured under identical
935 environmental conditions, termed the Mut dataset. The mutants were created through the MA
936 experiment to accumulate random genomic mutations by genetic drift. (**E**–**I**) Transcriptional
937 profiles of a single strain, MG1655, cultured under 76 different environmental conditions,
938 termed the Env dataset. (**A**, **E**) Schematics of the datasets. (**B**, **F**) Relationship between the
939 means and standard deviations of mRNA expression levels for thousands of genes. The magenta
940 and cyan lines represent running medians and their smoothed splines, respectively. Spearman's
941 rank correlation coefficients (R) and p-values are shown. (**C**, **G**) Mean expression levels of

942 essential and nonessential genes. (**D**, **H**) Transcriptional variability, represented as either DM_{mut}

943 or DM_{env}, of essential and nonessential genes. The DM values were calculated as the vertical

944 distance of the standard deviations from the smoothed splines of the running median of standard

945 deviations in panels **B** and **F**. The lower and upper edges of the boxes in panels **C**, **D**, **G**, **H**

946 represent the first (q1) and third (q3) quartiles, respectively. The horizontal lines in the boxes

947 represent the medians (m). The whiskers from the boxes extend to the most extreme observed

948 values inside inner fences, m±1.5(q3–q1). (**I**) Relationship between mean expression levels and

949 the shorter distance from *oriC* on the *E. coli* chromosome. Red and blue dots in panels **B**, **F**,

950 and **I** represent essential and nonessential genes, respectively.

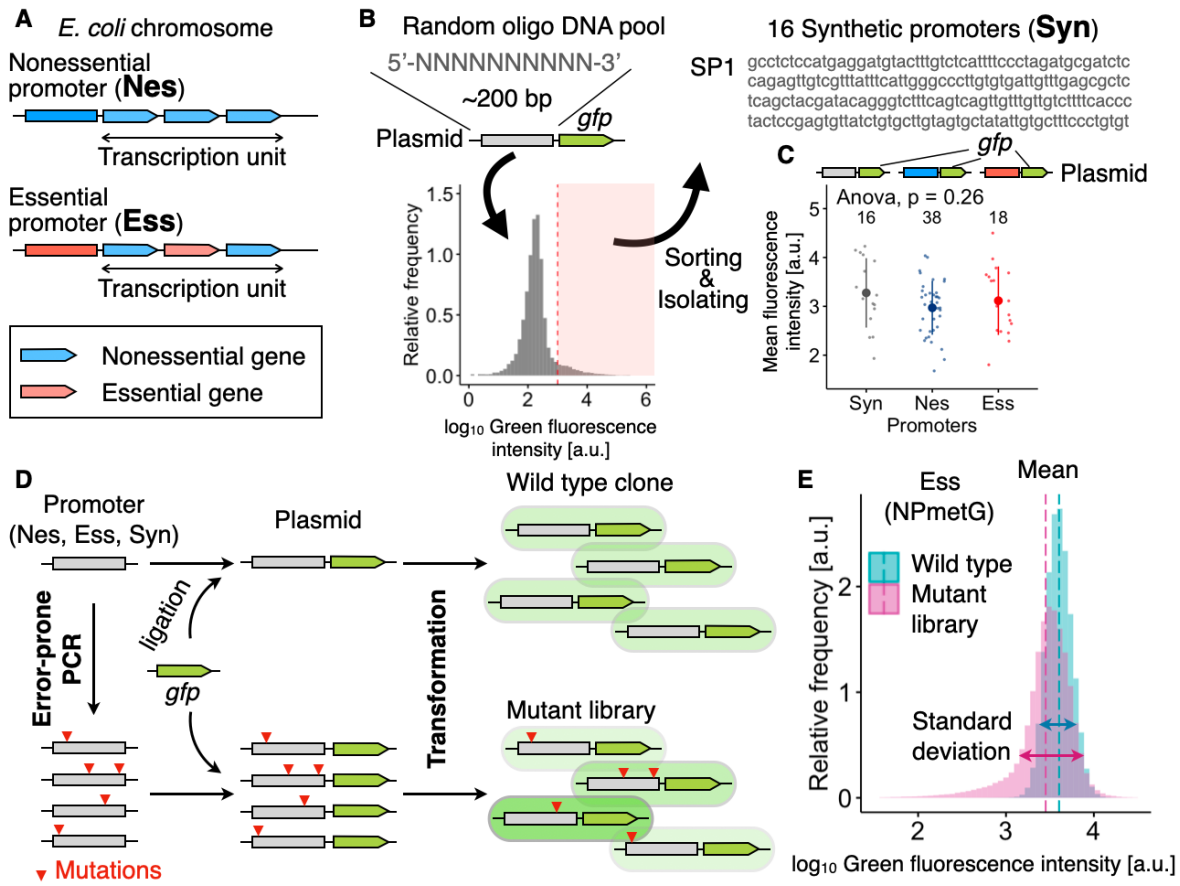


Fig. 2: Experimental design to measure expression levels of mutant promoter library.

(**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 cytometry. The 16 unique transformants, satisfying a defined green fluorescence (red shaded

962 area), were isolated. The obtained synthetic promoters were named SP1–16. (**C**) Natural and
963 synthetic promoters were placed upstream of *gfp* on a plasmid. Transformants harboring each
964 of these plasmids were analyzed using flow cytometry. Mean green fluorescence of each
965 population is shown as small dots. The large dots and error bars represent the means and
966 standard deviations within the groups, respectively. The number of promoters is indicated inside
967 the panel. The p-value corresponds to ANOVA. (**D**) Schematic showing construction of mutant
968 library from a given promoter region. The promoter region was subjected to random
969 mutagenesis by error-prone PCR. A mutant library was created through transformation to yield
970 more than 100,000 genetically-different mutants in a population. Unmutagenized plasmid was
971 also transformed to create wild type clones. (**E**) Representative examples of $\log_{10}$-transformed
972 green fluorescence intensity distributions of wild type and mutant library for the promoter
973 region of the *metG* gene in *E. coli* (NPmetG). The means (vertical lines) and standard deviations
974 (horizontal arrows) were calculated in each population.

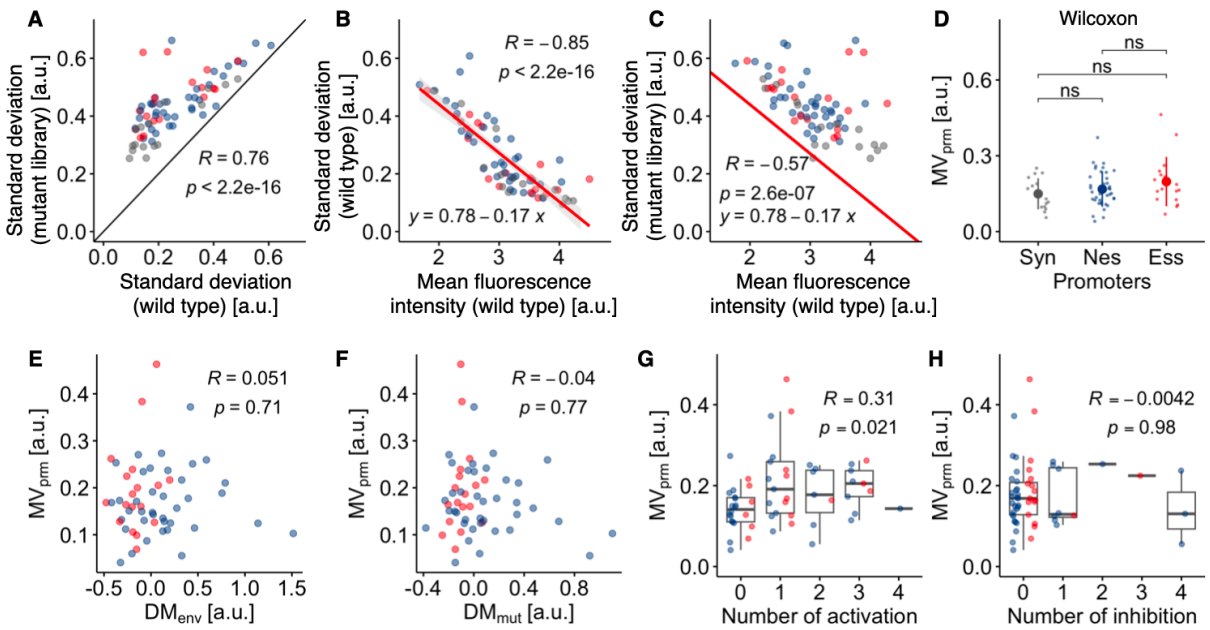
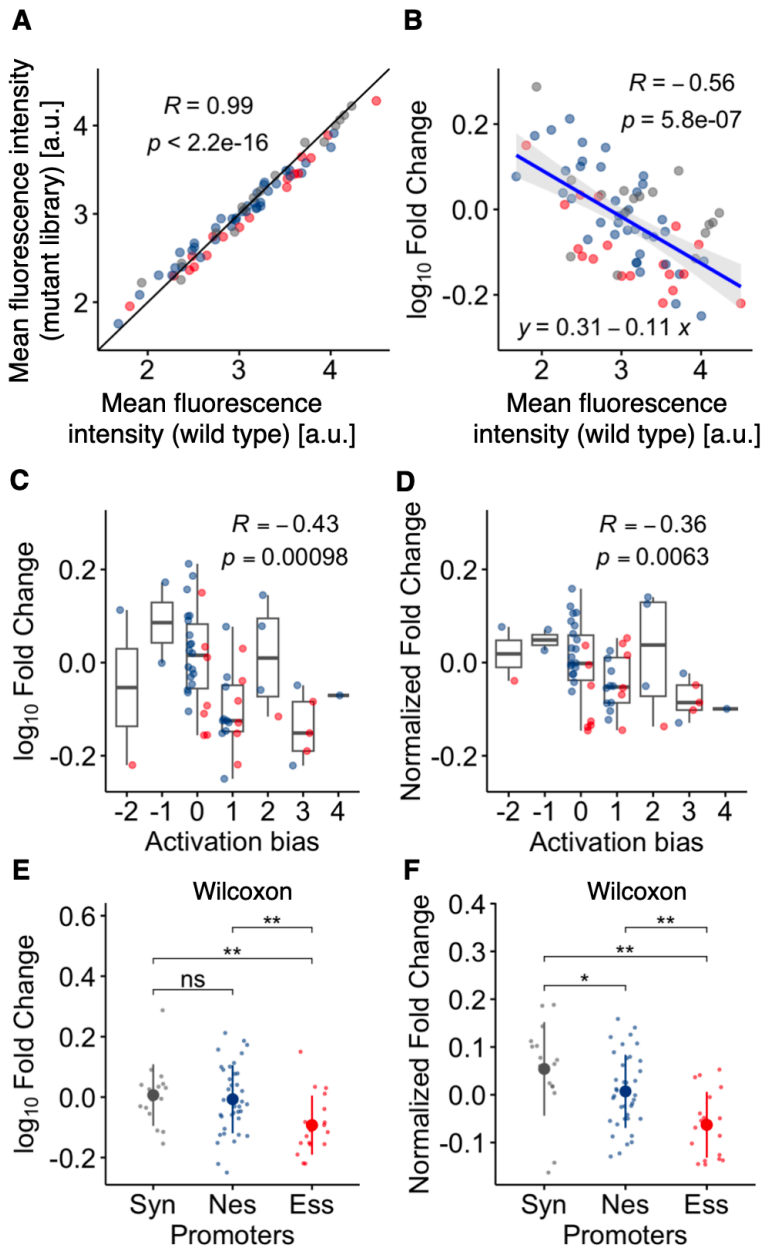


Fig. 3: Mutational variability in promoter activity.

(**A**) Comparison of the standard deviations in the GFP distributions between wild types and mutant libraries. The solid line represents y=x. (**B**, **C**) Relationship between the mean and standard deviation of the GFP distributions for wild types (**B**) and mutant libraries (**C**). Red lines represent a regression line for wild types and were used to compensate for the mean-dependency of the standard deviations of both wild types and mutant libraries 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 wild types from those of mutant libraries. 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**). The DM values were defined in **Fig. 1**. (**G**, **H**) Relationship between MV_{prm} and

988 the number of known activation (**G**) and inhibition (**H**) of natural promoters by transcriptional
989 regulators. The relationship between promoters and known transcriptional regulators were
990 based on RegulonDB(Tierrafria, et al. 2022) and EcoCyc(Keseler, et al. 2021). The grey, blue,
991 and red dots in each panel represent synthetic (Syn), nonessential (Nes), and essential (Ess)
992 promoters (defined in **Fig. 2**), respectively. Spearman's R and p-values are shown in panes **A**–
993 **C** and **E**–**H**.



994

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

996 (**A**) Comparison of the means in the GFP distributions between wild types and mutant libraries.

997 The solid line represents y=x. (**B**) Relationship between $\log_{10}$-fold change of mean expression

998 levels from wild types to mutant libraries and the expression levels of wild types. The blue line

999 represents a regression line used to compensate for the mean-dependency of the $\log_{10}$-fold

1000 changes by subtraction. (**C**, **D**) Relationship between activation bias in natural promoters and

1001 the unnormalized (**C**) and normalized (**D**) $\log_{10}$-fold changes. The activation bias was calculated

1002 by subtracting the number of known inhibitions from those of activations by transcriptional

1003 regulators. Spearman's R and p-values are shown in panes **A–D**. Grey, blue, and red dots in

1004 each panel represent synthetic (Syn), nonessential (Nes), and essential (Ess) promoters (defined

1005 in **Fig. 2**), respectively. (**E**, **F**) Unnormalized (**E**) and normalized (**F**) $\log_{10}$-fold changes for

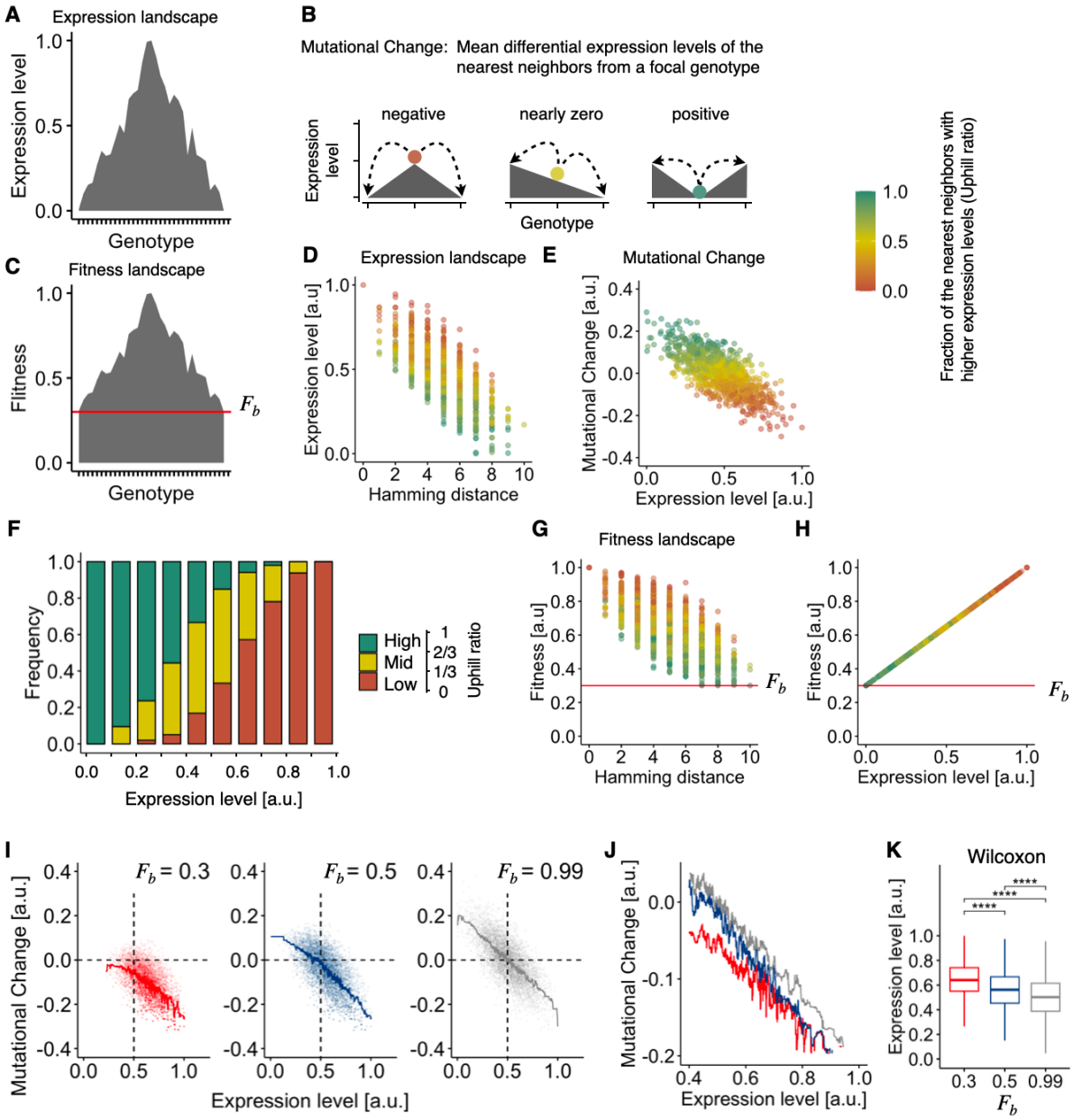
1006 each promoter group. The large dots and error bars are the means and standard deviations,

1007 respectively. Asterisks represent significance levels in pairwise Wilcoxon test (ns: $p > 0.05$; *:

1008 $p \leq 0.05$; **: $p \leq 0.01$). The p-values were adjusted using the BH method(Benjamini and

1009 Hochberg 1995).

1010



1011

1012 **Fig. 5: Numerical simulation of promoter evolution on a ragged fitness landscape.**

1013 (**A**–**C**) Schematics of landscapes of expression levels (**A**, **B**) and fitness (**C**). (**A**) Expression

1014 landscape was constructed by the rough Mt. Fuji (RNF) model. (**B**) Mutational change of each

1015 genotype was calculated as the mean of differential expression levels of the nearest neighboring

1016 genotypes (arrow heads, mutant types) from the focal genotype (colored circles, wild type). The

1017 nearest neighbors differ from the focal genotype by one Hamming distance. The color bar
1018 represents the fraction of the genotypes with higher expression levels than the focal genotype
1019 among the nearest neighbors, termed uphill ratio. (**C**) Fitness landscape was modeled by adding
1020 basal fitness (F_b , red line) to the expression landscape and normalizing it. (**D**, **E**)
1021 Representatives of expression landscape (**D**) and mutational change (**E**). Genotype was
1022 modeled as 10-bit binary numbers. Hamming distance represents the distance from the
1023 genotype with the global optima in expression level and fitness. (**F**) Relationship between
1024 expression level and uphill ratio. The uphill ratio was equally divided into three classes (Low,
1025 Mid, and High). Expression level was also equally divided into ten classes. Frequency of each
1026 class of uphill ratio was calculated as the number of corresponding genotypes divided by the
1027 total number of genotypes within the class. (**G**, **H**) Representatives of fitness landscape (**G**) and
1028 relationship between fitness and expression level (**H**) (F_b= 0.3). (**I**) Mutational changes of the
1029 evolved isolates obtained through evolutionary simulation using the standard Wright-Fisher
1030 model. The basal fitness was varied by mimicking strong positive (F_b=0.3, left), moderate
1031 positive (F_b=0.5, middle), and nearly neutral selection (F_b=0.99, right) acting upon higher
1032 expression levels. Ten independent expression/fitness landscapes were generated for each
1033 condition of basal fitness. In each fitness landscape, 600 independent evolutionary simulations
1034 were examined where initial genotypes were randomly selected. Dots represents the most
1035 abundant mutants within the population after 30 generations in each run. Solid lines represent

1036 the running medians. Dashed lines are guides for the eye. (**J**) Enlarged figure of the running

1037 medians in (**I**). (**K**) Expression level of the evolved isolates. Asterisks represent significance

1038 levels of the adjusted p-value (the BH method) in the Wilcoxon test (****: $p \leq 0.0001$).